# GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates 

KARL N. KIRSCHNER, AUSTIN B. YONGYE, SARAH M. TSCHAMPEL, JORGE GONZÁLEZOUTEIRIÑO, CHARLISA R. DANIELS, B. LACHELE FOLEY, and ROBERT J. WOODS Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, Georgia 30602


#### Abstract

A new derivation of the GLYCAM06 force field, which removes its previous specificity for carbohydrates, and its dependency on the AMBER force field and parameters, is presented. All pertinent force field terms have been explicitly specified and so no default or generic parameters are employed. The new GLYCAM is no longer limited to any particular class of biomolecules, but is extendible to all molecular classes in the spirit of a small-molecule force field. The torsion terms in the present work were all derived from quantum mechanical data from a collection of minimal molecular fragments and related small molecules. For carbohydrates, there is now a single parameter set applicable to both $\alpha$ - and $\beta$-anomers and to all monosaccharide ring sizes and conformations. We demonstrate that deriving dihedral parameters by fitting to QM data for internal rotational energy curves for representative small molecules generally leads to correct rotamer populations in molecular dynamics simulations, and that this approach removes the need for phase corrections in the dihedral terms. However, we note that there are cases where this approach is inadequate. Reported here are the basic components of the new force field as well as an illustration of its extension to carbohydrates. In addition to reproducing the gas-phase properties of an array of small test molecules, condensed-phase simulations employing GLYCAM06 are shown to reproduce rotamer populations for key small molecules and representative biopolymer building blocks in explicit water, as well as crystalline lattice properties, such as unit cell dimensions, and vibrational frequencies.


## Keywords

carbohydrate; force field; GLYCAM; AMBER; molecular dynamics; parameter development

[^0]
## Introduction

Carbohydrates are involved in numerous biological functions, such as recognition in axonal growth or path-finding, ${ }^{1}$ blood anticoagulation, ${ }^{2}$ cell-cell recognition, ${ }^{3}$ antibody-antigen inter-actions, ${ }^{4}$ structure factors in extra-cellular matrices, ${ }^{5}$ and post- or cotranslational modifications of polypeptides. ${ }^{6}$ Correct glycosylation patterns are essential for normal cell and organism function, and aberrant glycosylation is associated with numerous human diseases. ${ }^{7,8}$ To provide a structural basis for the multitude of biological roles played by carbohydrates, it is imperative that their spatial and dynamic properties be accurately determined. Experimental structure determination methods such as X-ray crystallography, ${ }^{9-11}$ NMR spectroscopy, ${ }^{12-15}$ and fluorescence energy-transfer spectroscopy ${ }^{16}$ have been applied in studies of carbohydrate conformation, either free or complexed with proteins. While NMR spectroscopy has been extensively used to characterize the dynamics of glycans in solution, ${ }^{17}$ interglycosidic linkage conformations are notoriously difficult to determine by NMR spectroscopy because of the paucity of nuclear Overhauser effects (nOes), ${ }^{18}$ the uncertainties in the Karplus-type equations employed to interpret scalar $J$-coupling constants ${ }^{19}$ and the potential for the linkage to populate multiple rotamer states. ${ }^{20}$ Moreover, NMR techniques employed to determine the structural properties of polysaccharides or protein-carbohydrate complexes are limited by molecular weight constraints. Alternatively, X-ray crystallography can be a powerful source of structural information. However, the presence of multiple glycoforms often prevents crystallization of glycoproteins, and the inherent flexibility of oligosaccharides is the presumed reason for the notable absence of X-ray structures for any but the smallest systems.

Theoretical methods, such as Monte Carlo and molecular dynamics (MD) simulations, are employed increasingly to augment the experimental approaches in determining the conformational properties of carbohydrates, and biomolecules in general. The level of interest in applying classical simulations to oligosaccharides arises from experimental limitations and is demonstrated by the numerous force fields and parameter sets that have been derived for carbohydrates (See ref. 17 for a review). The performance of several parameter sets, which included the first version of GLYCAM (GLYCAM_93 ${ }^{21}$ ), has been quantitatively evaluated against data from quantum mechanical calculations ${ }^{22}$ as well as on a relative basis, using a chemometric analysis from 20 second-generation carbohydrate force fields. ${ }^{23}$ The conclusion from those comparisons was that no single parameter set or force field consistently out-performed the others. In a study that focused on the abilities of three prominent biomolecular force fields to reproduce experimental solution data for two related disaccharides, the second version of GLYCAM (GLYCAM2000 ${ }^{24}$ ) with AMBER performed well in terms of conformational predictions. ${ }^{25}$ Although the GLYCAM_93 ${ }^{21}$ and $2000^{24}$ parameter sets, augmented by terms from the AMBER parameters, ${ }^{26}$ have been employed frequently in the study of oligosaccharides ${ }^{5,22-25,27-32}$ and oligosaccharideprotein complexes, ${ }^{10,14,33-39}$ they have some serious limitations. Earlier GLYCAM versions performed poorly at reproducing diffusion rates in explicit solvent MD simulations, and differed substantially from other carbohydrate parameter sets in the prediction of putative radial pair distribution functions (RPD) between hydroxyl groups and TIP3P ${ }^{40}$ water. ${ }^{25}$ For
example, the RPD for the O3-atom in methyl $a$-D-isomaltoside computed with GLYCAM2000a resulted in a less structured secondary hydration shell, and a first hydration shell at $3.0 \AA$, which was higher than the average distance between $\mathrm{O}_{\text {carbohydrate }} \ldots \mathrm{O}_{\text {water }}$ atoms found in carbohydrate crystals $(2.77 \AA) .{ }^{41}$ The first hydration shell of the $\mathrm{O}_{3} \ldots \mathrm{O}_{\text {water }}$ RPD had a peak density value of 1.0 , which compared unfavorably with the experimental value of about 3.0 for the RPD of $\mathrm{O}_{\text {water }} \ldots \mathrm{O}_{\text {water }}{ }^{25}$ The computed $\mathrm{O}_{3} \ldots \mathrm{O}_{\text {water }}$ RPD employing the present version of GLYCAM (GLYCAM06) resulted in a first hydration shell at $2.75 \AA$ with a peak density of 1.6 and a well-defined secondary hydration layer at about 5.0-5.5 $\AA$. This improvement in GLYCAM06 was due to the utilization of AMBERconsistent van der Waals parameters for the hydroxyl oxygen. In the case of self-diffusion rates of $a$-D-isomaltose GLYCAM2000a gave a value that was about twice the experimental value. ${ }^{25}$ The self-diffusion rate computed with GLYCAM06 was $3.6 \times 10^{-26} \mathrm{~cm}^{2} / \mathrm{s}$, which is in agreement with the experimental value of $4.2 \times 10^{-26} \mathrm{~cm}^{2} / \mathrm{s} .{ }^{25}$

The previous versions of GLYCAM were designed with the intention that they would introduce the minimal parameters necessary to add carbohydrate functionality to AMBER, and otherwise to maintain consistency with the AMBER force field development philosophy. ${ }^{21}$ Regarding transferability, GLYCAM development followed the general approach employed in biomolecular force fields of defining a single dihedral angle term for each molecular-class-specific linkage. Analogous to the treatment of peptide backbone dihedral angles, the rotational properties associated with glycosidic linkages were characterized by a single dihedral angle term specific to a unique atomic sequence, present only in carbohydrates. All other potentially contributing terms associated with that linkage were explicitly eliminated. ${ }^{21}$ That approach facilitated precise fitting to QM rotational data, but was specific to six-membered ring forms (pyranosides) and to the anomeric configuration. This limited the ability to readily introduce new chemical functionality into GLYCAM, to study ring conformational interconversions, and to apply it to other ring sizes (furanosides, in particular). ${ }^{42}$

The derivation of a highly consistent and transferable parameter set for modeling carbohydrates and glycoconjugates (GLYCAM06) is the focus of the present work. In light of accumulated applications of GLYCAM_93, ${ }^{10,22-25,27-31,43-48}$ several areas were targeted for improvement. In particular, the parameter set should (1) be transferable to all carbohydrate ring conformations and sizes, (2) be self-contained and therefore readily transferable to many quadratic force fields, (3) not require specific atom types for $\alpha$ - and $\beta$ anomers, (4) be readily extendible to carbohydrate derivatives and other biomolecules, (5) be applicable to monosaccharides and complex oligosaccharides, (6) be rigorously assessed in terms of the relative accuracy of its component terms, and (7) avoid the use of $1-4$ electrostatic or non-bonded scaling factors. ${ }^{24}$ In a study of the $\omega$-angle rotation $\left(\mathrm{O}_{5}-\mathrm{C}_{5}-\right.$ $\mathrm{C}_{6}-\mathrm{O}_{6}$ ) in monosaccharides, we observed that $\mathrm{O}_{6}$ may interact with either $\mathrm{O}_{4}$ (in a $1-5$ relationship) or $\mathrm{O}_{5}$ (in a $1-4$ relationship) and the use of $1-4$ scaling therefore unbalanced these interactions leading to an inability to correctly predict rotamer populations. ${ }^{24}$

Parameterization of GLYCAM06 employed training and test sets of $\sim 100$ molecules from the chemical families of hydrocarbons, alcohols, ethers, amides, esters, carboxylates, molecules of mixed functional groups as well as simple ring systems related to cyclic
carbohydrates, as outlined in Table 1. If a molecule had the potential of forming an internal hydrogen bond during a torsion rotation, two rotational energy curves were examined, one in which a hydrogen bond was allowed, and a second in which the hydrogen bond was disallowed.

To maintain consistency with the current AMBER protein parameters, the B3LYP/6-31+ $+\mathrm{G}(2 \mathrm{~d}, 2 \mathrm{p}) / / \mathrm{HF} / 6-31 \mathrm{G}^{*}$ level of theory was selected as the reference for all valence quantum calculations. ${ }^{49}$ Energies at this level may be expected to display minimal basis set superposition error (BSSE), ${ }^{50}$ which is not only important for determining accurate interaction energies, but also for locating minima and establishing barrier heights along the valence reaction coordinates. It has been shown that the B3LYP functional performs well in carbohydrates and related molecules, ${ }^{50-52}$ and was therefore selected for our force field development. Of particular importance for dynamics, this level of theory leads frequently to a reduction of barrier heights for internal rotations relative to those obtained with the HF/ 6-31G* method, resulting in increased conformational sampling for oligosaccharides.

## Methods

## Quantum Mechanical Calculations

Quantum mechanical calculations were performed using the Gaussian98 software package. ${ }^{53}$ Geometries were optimized using the Gaussian 98 default optimization criteria, unless otherwise noted. The HF/6-31G* level of theory was employed for neutral fragments, whereas for anionic molecules diffuse functions were added. Rotational energy curves were generated by varying the relevant torsion angle from $0-360^{\circ}$ in $30^{\circ}$ increments while optimizing all other variables. Single point energies were computed at the B3LYP/6-31+ $+G(2 d, 2 p)$ level of theory.

## Molecular Mechanics Calculations

All molecular mechanics (MM) calculations were performed using either the AMBER754 or AMBER $8^{55}$ software packages, with nonbonded and electrostatic scaling factors (SCEE, SCNB) set to unity. The valence parameters were determined by minimizing the error between the energies obtained from the $a b$ initio and molecular mechanics calculations in the traditional manner.

## Molecular Dynamics

Molecular dynamics (MD) simulations were carried out with the explicit inclusion of solvent (TIP3P) ${ }^{40}$ under isothermal-isobaric (nPT) conditions. Charged systems were neutralized by adding the appropriate number of counter ions. In general, the initial solvent configurations were optimized through 50 cycles of steepest descent and 950 cycles of conjugate gradient energy minimization with the SANDER module, followed by minimization of the entire system through the same protocol. The entire system was then annealed by heating from 5 to 300 K in 50 ps , followed by cooling to 5 K in another 50 ps . Initial velocities were assigned from a Boltzmann distribution at 5 K . Prior to the production dynamics stage the entire system was thermally equilibrated by heating again from 5 to 300 K in 150 ps . A 2-fs time step was used to integrate the equations of motion, using the Verlet

Algorithm. ${ }^{56}$ Long-range electrostatic interactions were treated using the particle mesh Ewald summation. A dielectric constant of unity was employed in all MD and MM calculations. Bonds containing hydrogen were constrained to their equilibrium lengths using the SHAKE algorithm. ${ }^{57}$

## Potential of Mean Force Calculations

The weighted histogram analysis method (WHAM) ${ }^{58}$ was used to calculate potentials of mean force (PMF). ${ }^{59}$ To ensure maximum sampling along the reaction coordinate, harmonic force constants of 10 and $5 \mathrm{kcal} / \mathrm{mol}$ were used for regions of high (maxima) and low (local and global minima) potential energies, respectively. Prior to the PMF calculations, the system was equilibrated following the protocol outlined in the MD section. Umbrella sampling was performed for 140 ps under nPT conditions with a time-step of 2 fs .

## Single Molecule and Ensemble-Averaged Charge Calculations

The atomic partial charges of all molecules employed in the gas phase for the development of valence bond, angle, and torsional parameters were derived from the HF/cc-pVTZ//HF/ 6-31G* molecular electrostatic potential (MEP), employing the restrained electrostatic potential (RESP) ${ }^{60}$ charge fitting methodology, with a hyperbolic charge restraint weight of 0.0005 , as indicated by Bayly et al. ${ }^{60}$ However, partial charge sets destined for condensedphase simulations were generated from the HF/6-31G*//HF/6-31G* MEP, appropriate for TIP3P ${ }^{40}$ water simulations. For the condensed-phase charge sets, a RESP charge restraint weight of 0.01 was employed, based on earlier MD simulations of experimental crystal lattices of $a$-D-glucopyranose, ${ }^{31}$ wherein this value led to optimal reproduction of unit-cell dimensions of the crystal.

In GLYCAM06, the issue of charge-conformation coupling was addressed by employing ensemble-averaged (EA) charge sets. ${ }^{61}$ For any given monosaccharide, an MD simulation, typically 50-100 ns was performed in TIP3P water, with 100-200 structures being selected from the trajectory for individual charge calculations. For each of these snapshots partial charges were calculated by fitting to the HF/6-31G* MEP. Prior to the charge calculations, each structure was optimized at the $\mathrm{HF} / 6-31 \mathrm{G}^{*}$ level, with the rotatable exocyclic bonds constrained to their MD conformations. ${ }^{61}$ The charges were then averaged for each monosaccharide to afford a final EA charge set, as exemplified in Table 2. These sets of charges were thus weighted by the actual occurrence of a particular solution conformation in an MD simulation. For these terminal monosaccharides, the average RESP-derived charge on the aglycon was -0.194 au , with the total charge on the glycoside equal to +0.194 au ( $Q_{\text {term }}$ ), for overall charge neutrality. To generate charges on nonterminal residues, the charge on the linking hydroxyl proton is added to that of the linking oxygen atom to give a new charge, $Q_{\mathrm{ol}}$. Next, $Q_{\text {term }}$ is subtracted from $Q_{\mathrm{ol}}$ to afford the final charge on the linking oxygen atom. This results in an overall charge of zero for nonterminal residues.

Nonbonded van der Waals parameters were taken directly from the PARM94 parameter set, which are also employed in the current distribution of AMBER parameter sets AMBER9. ${ }^{62}$

## Vibrational Analysis

Vibrational frequencies were computed for $a$-D-glucopyranose using the NMODE facility in AMBER9 with the GLYCAM parameter set. The calculation was performed on a monomer and on a 64-unit crystal. In both calculations, the initial structure was adapted from neutron scattering crystallographic data (PDB ID: GLUCSA10). ${ }^{63}$ The NMODE calculations were performed with default settings except for the values of the $1-4$ nonbonded and 1-4 electrostatic scale factors, which were each set equal to unit as suggested by earlier work. ${ }^{31}$ Both monomer and crystal were minimized to $\triangle$ RMS $=1 \times$ $10^{-5}$. Since periodic boundary conditions are not implemented in NMODE, the 64-mer was minimized as if it were a discrete nanoparticle. Frequency analyses were also performed on the monomer at the HF/6-31G*//6-31G* and B3LYP/6-31++G(2d,2p)//6-31++G(2d,2p) levels of theory.

Because of the complexity of the molecular structures, particularly in the 64-mer, a program (DMODES, for "describe modes") was written to automatically detect and assign vibrational modes. Although assignments for glucopyranose have often ${ }^{64-68}$ been made based on the potential energy distribution (PED) within each vibrational mode, NMODE facilitates use of the eigenvectors (collection of atomic motion vectors). Primarily for this reason, DMODES bases its assignments on the motions and relative kinetic energies of the atoms. However, there are other reasons to make assignments based on motion rather than PED's. While PED assignments are limited to bonded and near-neighbor nonbonded interactions, motion analysis can easily consider concerted motions of atoms far removed from each other. For example, motion analysis can assign ring expansions and deformations, assignments that are far more difficult to make using PED analysis. In addition, motion analysis yields assignments that are intuitive and that use language familiar to spectroscopists. The use of a computer program in general is, of course, also preferred to visual inspection of the eigenvector data because of increases in speed, objectivity, and reliability. At present, DMODES is in the prototype stage and is of limited general usefulness, but a copy of the code can be obtained by contacting the authors.

In brief, DMODES employs the relationships of individual atomic motions with respect to each other and to the molecular geometry to determine an appropriate description of the vibrational motion for a mode. For example, the dot product of a unit motion vector for atom
$i$, $\widehat{\mathbf{v}}_{i}$, with a corresponding unit bond vector, $\widehat{\mathbf{B}}_{i j}$, indicates the extent to which the motion is a stretch along that bond. Since the dot product of two unit vectors yields the cosine of the angle between them, the result of this calculation will be nearly $\pm 1$ if the motion is parallel to the bond (stretching) and $\sim 0$ if the motion is perpendicular (bending or torsioning). Similarly, bending and torsional motions of a system $i-j-k$ can be determined using eq. (1):

$$
\begin{equation*}
\widehat{\mathbf{v}}_{i} \bullet\left(\widehat{\mathbf{B}}_{i j} \times \widehat{\mathbf{B}}_{j k}\right) \tag{1}
\end{equation*}
$$

If this relationship is nearly zero, the motion lies within the plane formed by $i, j$, and $k$ and is therefore a bend (if also determined not to be a stretch). If the magnitude of the relationship is nearly one, the motion is nearly perpendicular to the plane and is torsional. Intermediate
values, of course, correspond to mixtures of torsion and bending. DMODES employs a set of cutoff values to distinguish between motions.

The significance of an individual atom's motion was determined using the relative kinetic energy of the atom. Each atom's motion vector represents its relative velocity at its vibrational potential energy minimum. The vector magnitudes can therefore be converted into relative kinetic energies, which correspond to the significances of the individual atomic motions within the vibrational mode. Although DMODES considered each atomic motion, the primary vibrational modes, which are presented here, were assigned on the basis of the most significant motion within each motion class. The implications of this choice are discussed in more detail in the results and conclusions section.

Assignments for the 64-mer required additional considerations. Since NMODE does not employ crystal-based symmetry arguments when calculating frequencies, 4608 normal modes ( $3 N$, where $N=64$ units $\times 24$ atoms per unit) were generated. Also, since an accurate frequency calculation requires a minimization, the positions of the atoms were necessarily displaced somewhat from their perfect-crystal locations. Because of this, the assignments for the 64 units, while generally similar, were each unique. To reduce the 294,912 different motion assignments, the data were treated in the following manner. First, the motions of all 64 units were assigned as usual for each vibrational frequency. During the assignment, the overall importance of each unit's motion to each mode was scored by summing the magnitudes of the motion vectors for each of the atoms in the unit. The motion vectors used for this score had previously been altered to remove any overall translational component (due to lattice-wide motions). The maximum magnitude sum for each frequency was then determined. If, within any frequency, a unit's magnitude was at least $50 \%$ of the largest magnitude, the results of its motion analysis were included in the results presented here. This procedure removes from our immediate attention those motions that occur with low intensity, being primarily due to resonance with more energetic motions of nearby neighbors. However, it does not discount resonances that result in motions that are more energetic, if somewhat unexpected. An example of the latter appears in the data and is discussed below.

Modern characterizations of the infrared and Raman spectra of carbohydrates typically employ some form of calculation of the PED when making assignments, ${ }^{64-71}$ though they occasionally use more traditional methods. ${ }^{69,71}$ Those that rely significantly ${ }^{71}$ or exclusively ${ }^{72}$ on traditional methods, as well as the occasional computation, ${ }^{70}$ will often employ more descriptive terminology in their assignments. While the assignment method employed here was designed to translate more readily into standard spectroscopic descriptions, it is important, also, to compare with the less-intuitive PED-based systems. To present a balanced comparison with experimental data, we have chosen to compare with four experimental studies. ${ }^{67,68,70,71}$ Since the various assignment naming methodologies do not always translate in a one-to-one correspondence (for an illustration, see Fig. 20 in Wells and Atalla ${ }^{71}$ ), we have chosen categories that are sufficiently broad that they facilitate comparisons, but are also narrow enough that the comparisons remain useful.

Of the four experimental studies considered here, two ${ }^{70,71}$ employed traditional assignment nomenclature and two ${ }^{67,68}$ assigned based on PEDs. All four used one or more computational method to make assignments, though one ${ }^{71}$ also used traditional methods, e.g., isotopic substitution, to guide their assignments. Three ${ }^{67,68,70}$ compared their calculations with spectra from earlier studies, ${ }^{64-66,69,72}$ but their results are being used here instead of the earlier work because of the detailed assignments they provided. Since our interest is comparison with experiment, only the experimentally observed frequencies from these studies are considered here. Additionally, we have only considered those experimental frequencies for which assignments were offered.

Since our assignments are based on the most energetic motion class within a given mode, we compared with only the most significant (first on the list) potential energy contribution within any PED assignment. ${ }^{67,68}$ With no hierarchy of significance apparent in the work of Gregurick and Kafafi, ${ }^{70}$ each assignment within a frequency was counted. The ranges from Wells and Atalla ${ }^{71}$ come mostly from their Figure 20, with occasional modifications based on the accompanying text, the latter being necessary for proper correlation to the set of motion classes being considered here.

Since IR and Raman intensities are not calculated natively by NMODE, and since the focus of the present study is the extent to which the force field reproduces vibrational frequencies we will not consider them here. In our comparisons with experimental data below, where both IR and Raman frequencies are available, we have compared with only the former, since the frequencies usually differ by only a few wavenumbers, which is not significant in this work. For modes that are not IR active, we have compared with the Raman frequency.

## Results and Discussion

## Atomic Partial Charges

The accurate reproduction of the electrostatic properties of a molecule is essential in a classical force field that seeks to quantitatively model intermolecular interactions of polar molecules, such as carbohydrates, proteins, and nucleic acids in water. The complexity of this requirement is reflected by the number of partial atomic charge derivation schemes reported in the literature. These include empirical methods, in which charges are fit to reproduce liquid densities and heats of vaporization, ${ }^{73}$ or a popular alternative, and that used in AMBER, in which the charges are fit to reproduce quantum mechanical MEPs. ${ }^{26}$ It should be noted that the concept of a partial charge is artificial but convenient, thus there is no definitive approach to their assignment.

A common drawback of several empirical methods is that they rely on reproducing bulk liquid properties and cannot be applied readily to solutes. In all MEP partial charge-based models, there are limitations that arise from the omission of a mechanism for incorporating the dependence of partial charge on molecular conformation. ${ }^{74,75}$ In addition, it is not necessarily ideal to locate the partial charges at the nuclear positions, ${ }^{40,76}$ nor do all atoms require partial charges. ${ }^{77}$ Close examination of ESP-charges computed for biomolecular fragments ${ }^{74}$ revealed that aliphatic hydrogen atoms tend to have small magnitude charges, as expected from the low bond polarity of most $\mathrm{C}-\mathrm{H}$ bonds. Further, charges on all
hydrogen atoms showed wide fluctuations coupled to the central carbon atom. However, the net charge of the methyl or methylene groups displayed minimal fluctuations. In GLYCAM_93, ${ }^{21}$ the effects of basis set and ESP-fitting algorithm on partial charges in carbohydrates were examined and again indicated the same behavior for aliphatic groups.

A second problematic area is that of charge partitioning. While MEP-fitting has the advantage of reproducing intermolecular interaction energies, caution must be exercised in the selection of atoms that will be included in the fitting procedure. Francl et al. ${ }^{77}$ performed singular value decomposition analyses on the least-squares matrices used to assign partial charges in a series of compounds, which frequently showed that an optimal fit to the ESP was achievable with fewer charges than the total number of atoms, suggesting that consistent reproduction of the MEP can be obtained by assigning zero net charges to some atoms. Consequently, in GLYCAM06, restraints were employed in the ESP fitting to ensure that the charges on all aliphatic hydrogen atoms were zero, leading to consistent charge sets with minimal degradation in the accuracy of the fit. For example, the omission of partial atomic charges on aliphatic hydrogens in methyl $a$-D-mannopyranoside had essentially no effect on the molecular dipole moment $(4.589 \mathrm{D})$ relative to the all atom charge model (4.581 D), and resulted in only an approximate $2.5 \%$ increase in the relative error in the fit to the potential.

In contrast to intermolecular interactions, which are well reproduced by MEP-derived partial charges, intramolecular interactions or solution properties are often poorly reproduced, unless conformationally averaged charges are employed. This issue becomes particularly significant when considering the relative energies of conformational isomers. For example, in the case of cyclohexanol, slightly different partial charge sets result from MEP fitting for both the equatorial and axial forms. When separate charge sets are utilized, incorrect relative energies of the ring conformations were obtained. However, accurate prediction of the relative energies between the chair forms, ${ }^{4} \mathrm{C}_{1} \leftrightarrow \mathrm{C}_{4}(<1 \mathrm{kcal} / \mathrm{mol})^{78}$ necessitated employing a common charge set for both the equatorial and axial species.

Extension of this logic to monosaccharides is challenged by notable variations among partial charge sets as a function of both anomeric and ring-carbon configuration. From the data in Table 2, both changes in anomericity $(\alpha, \beta)$ and hydroxyl group configuration (methyl $\alpha, \beta$ -D-glucopyranoside, $\alpha, \beta$-D-Glc $p \mathrm{OMe}$; methyl $\alpha, \beta$-D-mannopyranoside, $\alpha, \beta$-D-Man $p \mathrm{OMe}$; and methyl $\alpha, \beta$-D-galacotopyranoside, $\alpha, \beta$ - D -Gal $p \mathrm{OMe}$ ) resulted in partial charge sets that were statistically indistinguishable. Therefore, it is reasonable to generate a single charge set that can be used for either anomer of a given monosaccharide. However, a limitation in employing anomer-averaged charge sets is the loss of precision in the calculated intermolecular properties. Nevertheless, with a single charge set for a given monosaccharide, the relative intramolecular energies of ring conformations in which the anomeric configuration becomes inverted $\left({ }^{4} \mathrm{C}_{1} \leftrightarrow{ }^{1} \mathrm{C}_{4}\right)$ can now be effectively computed. For an example of the effect of charge protocol on internal energies, consider the relative energies of the $\alpha$ - and $\beta$-anomers of GlcpOMe. At the HF/6-31G*//HF/6-31G* level of theory the $\alpha$-anomer is predicted to be $1.46 \mathrm{kcal} / \mathrm{mol}$ more stable than the $\beta$-anomer, while at the B3LYP/6-31++G(2d,2p)//HF/6-31G* level, this value is $0.77 \mathrm{kcal} / \mathrm{mol}$, which is in good agreement with other recently reported theoretical results. ${ }^{79}$ Using GLYCAM06 and the $\alpha, \beta$

EA (RESP $<\alpha, \beta>$ ) charge set derived from Table 2, the $\alpha$-anomer was predicted to be 1.15 $\mathrm{kcal} / \mathrm{mol}$ more stable, a value between the $\mathrm{HF} / 6-31 \mathrm{G}^{*}$ and B3LYP/6-31++G(2d,2p) results. When the same calculations were performed using the individual EA charge sets, ( $\mathrm{RESP}<\alpha>$ ) for $\alpha$-D- and ( $\mathrm{RESP}<\beta>$ ) $\beta$-D-Glc $p \mathrm{OMe}$, the relative energy was reversed, with the $\beta$-anomer predicted to be the more stable by $4.19 \mathrm{kcal} / \mathrm{mol}$. Illustrating further the concept of a common charge set, the relative energies of 12 low-energy conformers of D GlcpOMe and D-GalpOMe, computed using GLYCAM06 were compared with their respective $\mathrm{HF} / 6-31 \mathrm{G}^{*} / / \mathrm{HF} / 6-31 \mathrm{G}^{*}$ values, Table 3. In the case of $\mathrm{D}-\mathrm{Glc} p \mathrm{OMe}$, when anomer-specific charges were used, the predicted relative energies were once more reversed relative to the QM -predicted values, with the most stable $\alpha$-conformer being $4.45 \mathrm{kcal} / \mathrm{mol}$ less stable than the most stable $\beta$-conformer. When $\langle\alpha, \beta\rangle$ EA charge sets were employed, the resulting relative energies were comparable once more with the QM values; the $\alpha$ anomer being preferred by $1.29 \mathrm{kcal} / \mathrm{mol}$. It can be inferred that this energy difference has an electrostatic origin, since a common parameter set was employed for both anomers. Among the $\beta$-D-Glc $p \mathrm{OMe}$ conformers, the trend in the GLYCAM06 predicted relative energies $(\beta 1<\beta 3<\beta 2)$ did not reproduce that determined by $\mathrm{QM}(\beta 2<\beta 1<\beta 3)$. This variance, however, is insignificant given that the higher energy conformers in both the QM and MM calculations are all within $0.6 \mathrm{kcal} / \mathrm{mol}$ of the global minima. In the case of DGalpOMe, the GLYCAM06-computed relative energies employing anomer-specific charge sets resulted in both large energy differences $(11.71 \mathrm{kcal} / \mathrm{mol})$ and incorrect anomer preferences $(\beta<\alpha)$. Once more, the correct trends were predicted when the $\operatorname{RESP}<\alpha, \beta>$ charge set was used, with the most stable $\alpha$-anomer being $1.58 \mathrm{kcal} / \mathrm{mol}$ more stable than the lowest energy $\beta$-anomer, which compared favorably with the QM value of $1.88 \mathrm{kcal} / \mathrm{mol}$. This strategy of averaging anomer- specific charges could be extrapolated to obtain a common set of charges for all the glycosides (RESP<overall>). Employing this charge set in $\mathrm{D}-\mathrm{Glc} p \mathrm{OMe}$ and $\mathrm{D}-\mathrm{Gal} p \mathrm{OMe}$, the $\alpha$-anomer was correctly predicted to be more stable by 1.12 and $1.79 \mathrm{kcal} / \mathrm{mol}$, respectively.

## Bond Lengths and Angles Parameters

The stereoelectronic effects that give rise to valence bond and angle variations at the anomeric carbon atom in pyranoses are well known. ${ }^{80,81}$ These variations are not only dependent on anomericity, but also influenced by the size of the aglycon. ${ }^{82}$ The stereoelectronic effects may be incorporated into Type II classical force-fields through stretch-bend and stretch-torsion cross terms. ${ }^{83}$ However, these cross terms do not exist in well-established macromolecular force fields, such as AMBER, ${ }^{26}$ CHARMM, ${ }^{84}$ and GROMOS. ${ }^{85}$ In GLYCAM_93, these geometrical variations were incorporated by defining anomer-specific atom types, each with unique values for the bond ( $K_{\mathrm{r}}$ ) and angle force constants ( $K_{\theta}$ ), and the equilibrium bond lengths ( $r_{\text {eq }}$ ) and angles $\left(\theta_{\text {eq }}\right)$. While the use of unique atom types for the different anomeric carbon atoms facilitated the inclusion of these stereoelectronic effects, it inhibited the ability of the force field to model processes involving ring inversions, which are equivalent to anomer interconversions. In addition, maintaining nongeneralizable anomer-dependent parameters prevented facile extension of the force field to nonstandard glycosyl residues. In the present parameterization, a common set of parameters has been derived for both $\alpha$ - and $\beta$-anomers employing a single $s p^{3}$ atom type (CG). For example, in the previous scheme, a 2-deoxymonosaccharide would require
new parameters to be derived, whereas in the current approach all such derivatives are implicitly allowed.

In principle harmonic stretching and bending force constants may be derived from spectroscopic data. However, such data are not available for all molecules of interest to this work. In addition, to be consistent with the use of QM data in the derivation of rotational barriers, we wished to employ only gas-phase force constants. Thus all force constants were derived by fitting a classical harmonic function to a distortion energy curve, computed at the B3LYP/6-31++G(2d,2p) level.

The accurate determination of the bending force constants for atomic sequences involved in glycosidic angles ensures that the flexing and rotational dynamics of the molecule will be correctly reproduced. Values of $K_{\theta}$ for the $\mathrm{C}-\mathrm{O}-\mathrm{C}$ and $\mathrm{O}-\mathrm{C}-\mathrm{O}$ angles were estimated from distortion analyses of the applicable angles in methoxymethane and 1,1dimethoxyethane, respectively. Fitting to the QM energy curves, computed over the range $\theta_{\text {eq }} \pm 10^{\circ}$, gave $K_{\theta}$ values of $50 \mathrm{kcal} / \mathrm{rad}^{2}$ and $100 \mathrm{kcal} / \mathrm{rad}^{2}$ for the $\mathrm{C}-\mathrm{O}-\mathrm{C}$ and $\mathrm{O}-\mathrm{C}-\mathrm{O}$ angles, respectively. The average relative errors in these curve fits were 0.73 and $0.16 \mathrm{kcal} /$ mol, respectively. These angle-bending constants are slightly softer than those employed in GLYCAM_93 C—O—C (60.0 kcal/ $\mathrm{rad}^{2}$ ) and $\mathrm{O}-\mathrm{C}-\mathrm{O}\left(110.0 \mathrm{kcal} / \mathrm{rad}^{2}\right) .{ }^{21}$

It should be noted that the experimentally observed bond lengths and valence angles in molecules such as carbohydrates and amino acids should not be employed as equilibrium values. These complex molecules contain internal strain forces that perturb the bond lengths and angles away from their theoretical equilibrium values. Equilibrium values, however, may be estimated from small carefully selected molecular fragments. These equilibrium values may differ significantly from the observed valence angles in larger molecules. In addition, effects associated with intermolecular interactions can lead to considerable differences between the QM values for bonds and angles, and those observed experimentally. For example, in the case of peptide bonds, the average experimental $\mathrm{C}-\mathrm{N}$ bond length ( $1.33 \AA$ ) from a survey of crystal structures from the Cambridge Crystallographic Data Bank differs markedly from both the gas-phase electron diffraction, and MP2/6-31G* values of 1.386 and $1.365 \AA$, respectively, computed for $N$ methylacetamide. ${ }^{86,87}$ Similarly, the $\theta_{\text {eq }}$ value for the C-O—C angle in GLYCAM06 $\left(111.6^{\circ}\right)$ is smaller than the average values observed in $\alpha-(2 \rightarrow 8)$ trisialoside $\left(117^{\circ}\right),{ }^{88}$ and generally in saccharides $\left(116.5^{\circ}\right) .{ }^{89,90}$ Nonetheless, using tetrahydro-2-methoxy-5-(tetrahydro-2H-pyran-2-yloxy)-2H-pyran, Figure 1, the GLYCAM06 minimized structure gave a glycosidic valence angle of $116.8^{\circ}$, comparable with that found in experimental structures of glycosides. ${ }^{90}$ When methyl $\alpha$-D-Glc $p(1 \rightarrow 4) \alpha$-D-Glc $p$ and methyl $\beta$-D-Glc $p$ $(1 \rightarrow 4) \beta$-D-Glc $p$, Figure 1, were employed the glycosidic angles were $117.3^{\circ}$ and $116.9^{\circ}$, respectively, which compared favorably with experimental observations. ${ }^{90}$

The bond and angle parameters developed in this work are presented in Table 4. Only six out of 35 bond, and 20 out of 91 angle parameters were eventually transferred from the existing GLYCAM_93 parameter set.

## Torsion Parameters

The GLYCAM_93 force field, as well as other current biomolecular force fields, such as the PARM $99{ }^{91}$ parameterization in AMBER, frequently employ a single atomic sequence to define the torsion properties for a given bond. For example, in GLYCAM_93, the rotation associated with the $\phi$-angle was defined by the terms for the $\mathrm{O}_{5}-\mathrm{C}_{1}-\mathrm{O}_{X}-\mathrm{C}_{X}$ sequence only; with a specific torsion term required for each $\alpha$ - and $\beta$-anomer. The related terms $\mathrm{C}_{2}$ -$\mathrm{C}_{1}-\mathrm{O}_{X}-\mathrm{C}_{X}$ and $\mathrm{H} 1-\mathrm{C}_{1}-\mathrm{O}_{X}-\mathrm{C}_{X}$ were explicitly set to zero. While this approach may lead to accurate reproduction of the QM rotational energy curves, it suffers from a lack of transferability. In view of making the parameters in GLYCAM06 fully transferable we adopted a more general approach wherein all atomic sequences had an explicitly defined set of torsion terms. Each term was derived by fitting to rotational energy curves for a molecular training set composed of relevant small molecules. Because of the symmetry of these small molecules, no phase angles were required. The omission of phase angles further facilitates parameter transferability. However, as each rotational energy is now the result of the sum of the contributions from the constituent atomic sequences, the overall accuracy is limited by the sum of errors in each term. The benefits in terms of applying to novel molecular classes is particularly important for carbohydrates as they frequently exist in highly derivatized forms in vivo.

Together with the nonbonded terms, torsion terms are crucially important for force fields to correctly model the conformational and particularly, the dynamic properties of macromolecules. As opposed to bond and angle parameters, in which only small excursions from equilibrium values are observed, torsional rotations may cover the entire range between 0 and $360^{\circ}$, and frequently exhibit more than one energy minimum. To achieve correct thermodynamic and kinetic behavior during MD simulations, it is important that both the relative energies of the minima and the barrier heights between them be reproduced. Thus, we not only sought to locate the minima and maxima, but also to reproduce all barrier heights. To assess the extent to which this had been achieved for each molecular class, the errors between the QM and MM rotational energies were computed over the entire range of the curves, $\left\langle E r r o r>_{\text {curve }}\right.$, and the minima, $\left\langle E r r o r>{ }_{\text {minima }}\right.$, are presented in Table 5.

## Hydrocarbons

Hydrocarbons provide the foundation for any biomolecular force field. As can be seen from the data in Table 5, the force field reproduced the gas-phase relative energies for these simple molecules very well, with an average error in the energies over the entire rotational energy curve of $0.15 \mathrm{kcal} / \mathrm{mol}$ and an error in the $\langle\text { Error }\rangle_{\text {minima }}$ of $0.11 \mathrm{kcal} / \mathrm{mol}$.

Rotating about the central $\mathrm{C}_{\mathrm{sp} 3}-\mathrm{C}_{\mathrm{sp} 3}(\mathrm{CG}-\mathrm{CG})$ bond in butane allowed us to examine how several torsion parameters performed collectively, namely those associated with the CG - CG-CG-CG, CG-CG-CG-HC, and HC-CG-CG-HC sequences. GLYCAM06 predicted the stabilization of the trans conformation, relative to the gauche conformation, of $1.16 \mathrm{kcal} / \mathrm{mol}$, in reasonable agreement with both the B3LYP/6-3111G(2d,2p) value computed here of $0.95 \mathrm{kcal} / \mathrm{mol}$ and the experimental range of $0.74-0.90 \mathrm{kcal} / \mathrm{mol} .{ }^{92-94}$ For this molecule, the GLYCAM06 rotational barriers were determined to be $5.51 \mathrm{kcal} / \mathrm{mol}$ at $0^{\circ}$ and $3.15 \mathrm{kcal} / \mathrm{mol}$ at $120^{\circ}$, which may be compared with the B3LYP/6-31++G(2d,2p) values
of 5.80 and $3.33 \mathrm{kcal} / \mathrm{mol}$, respectively, and with the experimental values ${ }^{92}$ of 4.56 and 3.62 $\mathrm{kcal} / \mathrm{mol}$, respectively.

The hydroxyl groups of alcohols make this class of compounds particularly relevant in the development of force fields for nucleic acids and carbohydrates. The overall errors in the fits for alcohols were <Error> ${ }_{\text {curve }} 50.45 \mathrm{kcal} / \mathrm{mol}$ and $\left\langle\right.$ Error $>_{\text {minima }}=0.28 \mathrm{kcal} / \mathrm{mol}$. The alcohols with a single hydroxyl group were easier to fit and resulted in $\left\langle\right.$ Error $>_{\text {curve }}=0.20$ $\mathrm{kcal} / \mathrm{mol}$, while the fit was slightly degraded for the diols with $\langle\text { Error }\rangle_{\text {curve }}=0.73 \mathrm{kcal} / \mathrm{mol}$.

## Mono-Alcohols

Mono-alcohols were used for the derivation of parameters that involved hydroxyl groups without the inclusion of other electronegative groups. As relates to the rotational behavior of the hydroxyl protons, the $\mathrm{H} 1-\mathrm{CG}-\mathrm{OH}-\mathrm{HO}$ rotational barrier in methanol was determined at the B3LYP/6-31++G(2d,2p) level and with GLYCAM06 to be 1.03 and 1.08 $\mathrm{kcal} / \mathrm{mol}$, respectively, both in excellent agreement with the experimental value ${ }^{95}$ of 1.07 $\mathrm{kcal} / \mathrm{mol}$. For ethanol, GLYCAM06 determined the trans conformation about the CG-CG $-\mathrm{OH}-\mathrm{HO}$ torsion to be $0.34 \mathrm{kcal} / \mathrm{mol}$ more stable than the gauche conformation, which may be compared with the experimental value ${ }^{95-97}$ of $0.40 \mathrm{kcal} / \mathrm{mol}$, while the B3LYP/ $6-31++G(2 d, 2 p)$ level overestimated the stability of the gauche conformer by $0.30 \mathrm{kcal} / \mathrm{mol}$. The barrier to rotation between the gauche conformations was determined to be $1.11 \mathrm{kcal} /$ mol, which compared well with the B3LYP/6-31++G(2d,2p) value and experimental range ${ }^{98}$ of $1.01 \mathrm{kcal} / \mathrm{mol}$ and $0.8-1.2 \mathrm{kcal} / \mathrm{mol}$, respectively. Similarly, for propan-2-ol, GLYCAM06 determined the gauche (defined by H1-CG-OH—HO) conformer to be more stable than the trans by $0.11 \mathrm{kcal} / \mathrm{mol}$, which may be compared with the experimental ${ }^{96,97}$ and QM values of $0.45 \pm 0.21$ and $0.24 \mathrm{kcal} / \mathrm{mol}$, respectively. The GLYCAM06 value for the $\mathrm{HC}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ rotational barrier in ethanol was found to be $3.09 \mathrm{kcal} / \mathrm{mol}$, compared with the B3LYP/6-31++G(2d,2p) value of $3.07 \mathrm{kcal} / \mathrm{mol}$, which both slightly underestimated the experimental range ${ }^{95,99}$ of $3.32-3.55 \mathrm{kcal} / \mathrm{mol}$.

Developing parameters for the $\mathrm{CG}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ sequence was challenging due primarily to the different conformational preferences of this torsion in open-chain and cyclic systems. In open-chain systems, using $n$-propanol as an example, the gauche conformation computed here at the MP2/6-31++G(2d,2p)//MP2/6-31G** level, was favored over the trans by $0.21 \mathrm{kcal} / \mathrm{mol}$. In contrast, the equatorial conformer of cyclohexanol ${ }^{100}$ in which the sequence CG-CG-CG-OH is trans, is more stable than the axial conformation (CG-$\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ is gauche) by $0.56 \mathrm{kcal} / \mathrm{mol}$. The differences between the open-chain and the cyclic preferences appear to reside in induction effects, which are difficult to account for in a classical mechanical treatment lacking polarizability. ${ }^{101}$ Therefore, GLYCAM06 was derived to slightly overestimate the trans conformation in $n$-propanol ( $0.77 \mathrm{kcal} / \mathrm{mol}$ ) over the gauche to optimize the performance of this parameter in glycan rings. As a result, optimal agreement was achieved for the rotational energy profile of the equatorial conformation of hydroxymethylcyclohexane Figure 2, as well as in highly substituted openchain alcohols, such as 2,2-dimethylpropanol and 1,1,2,2-tetramethylpropanol.

## Diols

Diols mimic the atomic sequence, $\mathrm{O}-\mathrm{C}-\mathrm{C}-\mathrm{O}$ prevalent in the vicinal hydroxyl fragments on the rings of common glycosyl residues. The torsional preferences for this linkage affect ring conformations as well as the rotamer preferences of exocyclic side chains, such as in sialic acid. This torsion term is most strongly influenced by the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ parameters as well as by internal hydrogen bonding. To establish the relative contributions from each of these effects, rotational energy curves for the $\mathrm{O}-\mathrm{C}-\mathrm{C}-\mathrm{O}$ linkage were computed both allowing and prohibiting the formation of internal H -bonds. ${ }^{24}$ The ability of GLYCAM06 to reproduce these relative energies and trends was illustrated by the average error values <Error> ${ }_{\text {curve }}$ and <Error> ${ }_{\text {minima }}$, in the rotational energy profiles. For example, in 2,3-butanediol $\left\langle\right.$ Error ${ }_{\text {curve }}=0.40 \mathrm{kcal} / \mathrm{mol}$ and $\langle\text { Error }\rangle_{\text {minima }}=0.35 \mathrm{kcal} / \mathrm{mol}$ when internal H-bonding was allowed, while $\langle\text { Error }\rangle_{\text {curve }}=0.34 \mathrm{kcal} / \mathrm{mol}$ and $\langle\text { Error }\rangle_{\text {minima }}=$ $0.21 \mathrm{kcal} / \mathrm{mol}$ when internal H -bonding was disallowed.

The presence of the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ dihedral angle in the exocyclic glyceryl side chains of sialic acids such as 5- N -acetyl-neuraminic acid (Neu5Ac) is significant in determining the conformational properties of the side chain, Figure 3. The side chain features two dihedral angles, which bear the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ sequence: $\omega_{8}$ $\left(\mathrm{O}_{7} \mathrm{C}_{7} \mathrm{C}_{8} \mathrm{O}_{8}\right)$ and $\omega_{9}\left(\mathrm{O}_{8} \mathrm{C}_{8} \mathrm{C}_{9} \mathrm{O}_{9}\right)$. The tendency for the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ angle to adopt the gauche conformation in simple diols has been referred to as the gauche effect. ${ }^{24}$ However, in Neu5Ac, while $\omega_{9}$ populates all three staggered rotamers, $\omega_{8}$ predominantly exists in the trans conformation, as ascertained from both solution $\mathrm{NMR}^{88,102-104}$ and X-ray crystallography ${ }^{105,106}$ studies. A pair of intramolecular hydrogen bonds, between H 8 O and the oxygen atoms of the carboxylate group, and H7O and the carbonyl group of the amido moiety, appears to stabilize this rotamer. However, the magnitude of this stabilization is unclear, as the same trans preference for $\omega_{8}$ is also observed in the $\beta$-anomer, ${ }^{106}$ in which the intramolecular H8O-carboxylate H -bond is not possible. Application of the initial parameters for the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ torsion term, generated by fitting to the QM rotational energy curves for compounds in the initial training set of diols, in solvated MD simulations, did not quantitatively reproduce the rotamer populations for the glyceryl side chain in Neu5Ac. In contrast to experimental data, the initial parameters led to a strong predilection for the gauche rotamer. Consequently a new approach was adopted in developing torsional parameters for the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ atomic sequence.

The initial values for the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ torsion term were modified by systematically adjusting the $V_{1}$ coefficient until explicitly solvated MD simulations of Neu5Ac were able to reproduce the solution populations for $\omega_{8}$. Because our goal was to increase the sampling of the trans rotamer of $\omega_{8}$, and only the energy function of the $V_{1}$ coefficient has a single minimum at $180^{\circ}$, only this term was modified. Next, the effect of the different $V_{1}$ values on the stability of the ring conformations in selected monosaccharides was evaluated, because the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ atomic sequence is also present in ring systems. Because the entire range of $V_{1}$ values were consistent with predicting the ring conformational preferences, and a series of MD simulations of Neu5Ac indicated that different $V_{1}$ values resulted in a predominant occurrence of the $\omega_{8}$ trans rotamer, the most consistent value was finally
determined by performing explicitly solvated MD simulations of 1,2-ethanediol, and selecting the value that best reproduced its experimentally determined rotamer populations.

The rotamer sampling of $\omega_{8}$ in the glyceryl side chain, employing different $V_{1}$ values are presented in Table 6 . Traversing these values from -1.0 to $1.0 \mathrm{kcal} / \mathrm{mol}$ resulted in an overall increase in the occurrence of the trans rotamer, indicating that an adjustment of this coefficient was necessary to reproduce the solution conformational properties of $\omega_{8}$.

As mentioned earlier, the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ term is also present in monosaccharide rings, most commonly existing in gauche relationships, Figure 4. Increasing the $V_{1}$ coefficient could induce a preference for the trans relationships of $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ dihedral angles present in ring systems, subsequently introducing artifactual ring flips during MD simulations. To determine whether the different $V_{1}$ coefficients had any influence on the stabilities of pyranose chair conformations when the parameters were employed in MD simulations, the MM generalized Born surface area (MMGBSA) implicit solvation model was used to compute the average MM energies of $\alpha$-D-Manp and $\alpha$-D-idopyranose ( $\alpha$-DIdop) in both ${ }^{4} \mathrm{C}_{1}$ and ${ }^{1} \mathrm{C}_{4}$ conformations. In the experimental solution conformations, $a$-DManp displays a single trans and two gauche $\mathrm{O}-\mathrm{C}-\mathrm{C}-\mathrm{O}$ conformations in both the ${ }^{4} \mathrm{C}_{1}$ and ${ }^{1} \mathrm{C}_{4}$ chair forms. In contrast, $a$-D-Idop displays three trans and three gauche $\mathrm{O}-\mathrm{C}-\mathrm{C}$ - O conformations in the ${ }^{4} \mathrm{C}_{1}$ and ${ }^{1} \mathrm{C}_{4}$ chair forms, respectively.

For each monosaccharide, 1-ns explicit solvent MD simulations were performed in which the ring was either in the ${ }^{4} \mathrm{C}_{1}$ or ${ }^{1} \mathrm{C}_{4}$ conformation. Snapshots were selected from the last 500 ps at 1 ps intervals from each simulation, and water molecules were removed. Using different $V_{1}$ coefficients the average MM energies of the individual chair conformers of each monosaccharide were determined by averaging the MMGBSA computed energies of the 500 selected snapshots. Over the entire range of $V_{1}$ values, the calculations predicted a consistent preference for the ${ }^{4} \mathrm{C}_{1}$ over the ${ }^{1} \mathrm{C}_{4}$ conformations in $a$-D-Man $p$, Figure 3. This insensitivity of the $a$-D-Manp ring conformation relative to the $V_{1}$ coefficient can be attributed to the similar distribution of the trans and gauche $\mathrm{O}-\mathrm{C}-\mathrm{C}-\mathrm{O}$ conformations in both chair forms. The preference for the ${ }^{4} \mathrm{C}_{1}$ over the ${ }^{1} \mathrm{C}_{4}$ conformer is also expected due to the unfavorable axial configuration of the hydroxymethy1 group at C 5 , in addition to the energetically disallowed 1-3 interaction between the axial hydroxyl group at C3 and the C5 hydroxymethyl group. On the other hand, traversing the range of the $V_{1}$ coefficients in the $a$-D-Ido $p$ system, there was a preference for the ${ }^{1} \mathrm{C}_{4}$ conformation at the negative extreme $(-2.0 \mathrm{kcal} / \mathrm{mol})$ with a progressive increase in the stabilization of the ${ }^{4} \mathrm{C}_{1}$ conformation toward the positive extreme $(2.0 \mathrm{kcal} / \mathrm{mol})$. This trend is expected because for negative $V_{1}$ coefficients, the $V_{1}$ energy profile as a function of dihedral angle has a maximum barrier at $180^{\circ}$, the reverse being true for positive coefficients. Thus the exclusive presence of the O-$\mathrm{C}-\mathrm{C}-\mathrm{O}$ trans conformations in the ${ }^{4} \mathrm{C}_{1}$ structure leads to a destabilization for negative $V_{1}$ values, but becomes a stabilizing factor for positive values of $V_{1}$.

From MD simulations of methyl $a$-Neu5Ac, observation of the $\omega_{8}$ trans and gauche rotamer populations as a function of $V_{1}$, Table 6 , indicated that the following values: $-0.2,0.0,0.4$, $0.6,0.8$, and $1.0 \mathrm{kcal} / \mathrm{mol}$, would lead to rotamer preferences that were consistent with experimental data. In addition, the MMGBSA calculations showed that these coefficients
would not cause spurious ring flips during MD simulations. Therefore, the most consistent coefficient was determined by performing explicit solvent MD simulations of 1,2ethanediol, with $V_{1}$ ranging from -0.5 to $0.5 \mathrm{kcal} / \mathrm{mol}$, in increments of $0.1 \mathrm{kcal} / \mathrm{mol}$. The conformational properties of 1,2-ethanediol have been extensively investigated by both experimental and theoretical studies, ${ }^{107,118}$ which predict a predominance of the gauche conformation, albeit to different extents, ranging from gauche:trans 99:1 ${ }^{116}$ to 67:33. ${ }^{115}$ Experimental NMR data ${ }^{117}$ indicate a mixture of $80: 20 \%$ gauche:trans populations in solvents with low dielectric constants. Chidichimo et al. ${ }^{110}$ reported the existence of only the gauche conformation. However, their studies were done using nematic-lyotropic crystalline solutions. A $V_{1}$ value of $-0.1 \mathrm{kcal} / \mathrm{mol}$ best reproduced the experimental populations obtained in conditions most similar to those of the simulation, with a gauche:trans ratio of $81: 19 \%$. Using this value of $-0.1 \mathrm{kcal} / \mathrm{mol}$ a PMF-WHAM analysis of methyl $a$-Neu5Ac with $\omega_{8}$ as the reaction coordinate was carried out. This analysis produced an energy profile with the trans rotamer being more stable by $\sim 1.0 \mathrm{kcal} / \mathrm{mol}$ relative to the + gauche conformer, and $1.2 \mathrm{kcal} / \mathrm{mol}$ with respect to the -gauche conformer, Figure 5. The rotational barriers to transition between +gauche and -gauche conformers were $\sim 2.4$ and $3.4 \mathrm{kcal} / \mathrm{mol}$, respectively, indicating some transition to the + gauche rotamer. The MD simulations correctly predicted the solution behavior of the glyceryl $\omega_{8}$ torsion, Figure 5. To assess further the extent to which these new parameters were able reproduce the B3LYP/ $6-31++G(2 d, 2 p)$ rotational energies, the rotational energy curves for the $\mathrm{O}-\mathrm{C}-\mathrm{C}-\mathrm{O}$ torsions for 2,3-butanediol were determined for comparison, Figure 6. As expected, when intramolecular hydrogen bonds are allowed, there is a stabilization of the gauche conformers over the trans conformer. When intramolecular hydrogen bonds are excluded, by constraining the hydroxyl proton to a trans orientation (HO-OH-CG-CG $=180^{\circ}$ ), or by substituting the hydroxyl protons with methyl groups, electrostatic repulsions between the oxygen atoms reverse the relative stabilities of the rotamers, resulting in the trans rotamer being the most stable. The $\langle\text { Error }\rangle_{\text {curve }}=0.87 \mathrm{kcal} / \mathrm{mol}$ and $\left\langle\right.$ Error $>_{\text {minima }}=0.35 \mathrm{kcal} / \mathrm{mol}$ when internal H-bonding was allowed, and $\left\langle\right.$ Error $>_{\text {curve }}=1.04$ and $\langle\text { Error }\rangle_{\text {minima }}=0.61$ $\mathrm{kcal} / \mathrm{mol}$ when internal H -bonding was disallowed. Comparing these error values to those determined using the initial $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ parameters, indicated that the present parameters resulted in fits that were slightly less accurate. However, this slight degradation was necessary for the correct prediction of the solution conformational properties of the glyceryl side chain of sialic acid.

## Ethers

The parameters derived from ethers are important to the accurate description of carbohydrate ring conformational properties as well as the rotational properties of glycosidic linkages. The overall accuracy of the derived torsion terms is indicated in the <Error> ${ }_{\text {curve }}$ and $\left\langle\right.$ Error> ${ }_{\text {minima }}$ of 0.56 and $0.35 \mathrm{kcal} / \mathrm{mol}$, respectively. In methoxymethane, GLYCAM06 yielded a rotational barrier for the H1-CG-OS—CG torsion of $2.31 \mathrm{kcal} /$ mol , in reasonable agreement with the B3LYP/6-31++G(2d,2p) and experimental values ${ }^{119}$ of 2.43 and $2.62 \mathrm{kcal} / \mathrm{mol}$, respectively.

The OS—CG—CG—OS torsion term is particularly important for glycans having a $1 \rightarrow 6$ linkage. The simplest molecule exhibiting this torsion term is 1,2-dimethoxyethane (DME).

A single $V_{2}$ coefficient of $0.82 \mathrm{kcal} / \mathrm{mol}$ for this term leds to good overall reproduction of the gas-phase data for the entire set of diethers, with a value of $\langle\text { Error }\rangle_{\text {curve }}=0.30 \mathrm{kcal} / \mathrm{mol}$ (Table 5) for the OS—CG—CG-OS and associated parameters. This term is also significant in branched glycans, which may exhibit glycosidic linkages between vicinal hydroxyl groups within the glycan ring. Incorrect parameters for this term may alter the conformational stability of the glycan ring. To quantify the behavior of the parameters for this linkage, we examined the rotational properties of the central $\mathrm{C}_{\mathrm{sp} 3}-\mathrm{C}_{\mathrm{sp} 3}(\mathrm{CG}-\mathrm{CG})$ bond in 2,3-dimethoxybutane. As can be seen from the data in Figure 6, the parameters well reproduced the QM data for 2,3-dimethoxybutane.

To assess further the applicability of the parameters for the OS-CG-CG-OS sequence to solution phase studies, the conformer populations of DME from a 20-ns explicit solvent MD simulation were compared with those determined from solution phase experimental data. The conformers were characterized by the rotational preferences of the dihedral angles composed of the $\mathrm{C}_{\mathrm{sp} 3}-\mathrm{O}_{\mathrm{sp} 3}$ (CG—OS) bonds, and the central CG-CG bond; with $T$ and $G$ denoting the trans and gauche conformations, respectively. Theoretical ${ }^{120-122}$ and experimental ${ }^{123-125}$ techniques have been employed to elucidate these conformational properties in aqueous solutions at different DME concentrations, and in the gas phase. There was a general preference for the G rotamer around the $\mathrm{CG}-\mathrm{CG}$ bond, and for the T rotamer around the CG-OS bonds, with the major conformers identified as the TTT, TGT, $\mathrm{TGG}^{\prime}$, TGG, and TTG. ${ }^{124}$ Notable, were the observations that the populations of the various conformers differed in aqueous solutions relative to those present in neat liquid or in the gas phase. The populations also displayed a dependence on temperature. In the gas phase, the TTT, TGT, and $\mathrm{TGG}^{\prime}$ were the major rotamers, while in neat liquid and aqueous solutions, the TGT and TGG were predominant, ${ }^{123,124}$ indicating that DME-DME and DME- $\mathrm{H}_{2} \mathrm{O}$ interactions influenced rotamer populations. Thus, our 20-ns MD-derived conformer populations were compared with those observed experimentally at low DME concentrations. The populations determined from the MD simulations could be ranked as: TGT (56\%), TGG (13\%), $\operatorname{TTT}(11 \%), \mathrm{TGG}^{\prime}(10 \%)$, $\mathrm{TTG}(6 \%)$, and others (4\%), which concurred qualitatively with experimental observations that the TGT and TGG predominate. ${ }^{123,124}$

## Anomeric Linkages

Anomeric linkages, represented by the $\mathrm{C}-\mathrm{O}-\mathrm{C}-\mathrm{O}$ torsion angle common to all carbohydrates, are often modeled in pyranosides by tetrahydro-2-methoxy-2H-pyran (axial) and tetrahy-dro-2-methoxy-2H-pyran (equatorial), corresponding to $\alpha$ - and $\beta$-pyranosides, respectively. ${ }^{21,126}$ In GLYCAM_93, reproduction of the $\mathrm{C}-\mathrm{O}-\mathrm{C}-\mathrm{O}$ rotational energies for the $\alpha$ - and $\beta$-link-ages was accomplished by using unique torsion parameters for each anomer, which were distinguished by different atom types for the anomeric carbon atom. Here, we use a single parameter for both anomeric configurations. Both GLYCAM06 and GLYCAM_93 correctly reproduced the B3LYP/6-31++G(2d,2p) rotational energy profiles, with <Error> ${ }_{\text {curve }}$ values of $0.45 \mathrm{kcal} / \mathrm{mol}$ (GLYCAM06) and $0.3 \mathrm{kcal} / \mathrm{mol}$ (GLYCAM_93), for the $a$-glycosides, while for the $\beta$-glycosides the <Error> ${ }_{\text {curve }}$ values were $0.36 \mathrm{kcal} / \mathrm{mol}$ (GLYCAM06) and $0.57 \mathrm{kcal} / \mathrm{mol}$ (GLYCAM_93). Therefore, employing a torsion term common to both $\alpha$ - and $\beta$-glycosides did not lead to degradation in the fit to the relevant rotational energy curves. As depicted in Figure 7, GLYCAM06 reproduced the rotational
topologies for both $a$ - and $\beta$-anomers, showing good quantitative agreement with the B3LYP/6-31++G(2d,2p) energies. The HF/6-31G* rotational energy curve for tetrahydro-2-methoxy-2H-pyran (axial) displayed a global minimum at $60^{\circ}$, and a less stable local minimum at $300^{\circ}$, while for tetrahydro-2-methoxy-2H-pyran (equatorial), the global minimum was at $300^{\circ}$ with a local minimum $5 \mathrm{kcal} / \mathrm{mol}$ higher at $\sim 60^{\circ}$. Three of these minima are consistent with the minima based on expectations due to the stabilizing exoanomeric effect, ${ }^{127}$ and the repulsive steric interactions. However, the minimum of tetrahydro-2-methoxy-2H-pyran (axial) located at $300^{\circ}$ has not been observed experimentally, even though it displays a stabilizing exo-anomeric effect, presumably due to large steric repulsions between the hydrogen atoms of the aglycon and that of C 2 . This structure was subjected to a subsequent optimization at the MP2/6-311++G(2d,2p) level of theory, during which it collapsed to the global minimum energy structure with the $\mathrm{C}-\mathrm{O}-\mathrm{C}$ -O torsion of $\sim 60^{\circ}$.

## Amides and Esters

Amide and ester parameters are relevant to a carbohydrate force field since many monosaccharides contain N - and O -acetyl moieties. In addition, N -linked glycans are attached via an amide bond to asparagine side chains in glycoproteins.

The exact extent of deviation from planarity in the ground-state amido groups, and the origin of the large barrier to rotations about the $\mathrm{N}_{\mathrm{sp} 2}-\mathrm{C}_{\mathrm{sp} 2}(\mathrm{~N}-\mathrm{C})$ bond in amides have been the subject of an extensive study. ${ }^{128}$ From their highest level calculation (CCSD(T)/PVTZ) Fogarasi and Szalay found that the optimized structure of methanamide was exactly planar. It is accepted that the large barrier to rotation is associated with the breaking of $\pi$ delocalization at $90^{\circ}$. Langley and Allinger. ${ }^{129}$ observed that the rotational barrier is sensitive to which proton is rotated and the out-of-plane distortion of the proton not constrained by the torsional rotation. When either $\mathrm{H}-\mathrm{N}-\mathrm{C}-\mathrm{O}$ torsion angle for one proton is driven above $90^{\circ}$, the unconstrained proton lags behind and follows an out-ofplane, rather than a purely torsional, energetic pathway. ${ }^{129}$ Consequently, in acetamide the two different $\mathrm{H}-\mathrm{N}-\mathrm{C}-\mathrm{O}$ dihedral angles were both varied counterclockwise from 0 to $90^{\circ}$, and from $180^{\circ}$ to $90^{\circ}$, resulting in two transition states at the $90^{\circ}$ angle having rotational barriers of 13.96 and $10.87 \mathrm{kcal} / \mathrm{mol}$, corresponding to the $0 \rightarrow 90^{\circ}$ and $180 \rightarrow 90^{\circ}$ paths, respectively. In both transition states the nitrogen atom is pyramidal. In the lower energy state the unconstrained proton is closer to the carbonyl oxygen $\left(\mathrm{H}_{\text {unconst }}-\mathrm{N}-\mathrm{C}-\mathrm{O}\right.$ $=22^{\circ}$ ), while in the higher energy state this proton is farther from the oxygen atom $\left(\mathrm{H}_{\text {unconst }}-\mathrm{N}-\mathrm{C}-\mathrm{O}=-154^{\circ}\right)$. The rotational energies in GLYCAM06 were in line with the B3LYP/6-31++G(2d,2p) values, although GLYCAM06 underestimated the barrier heights by $\sim 2.0 \mathrm{kcal} / \mathrm{mol}$ in each case, Figure 8 . In general, the amide parameters performed well, giving $\left\langle\right.$ Error $>_{\text {curve }}=0.64 \mathrm{kcal} / \mathrm{mol}$ and $\left\langle\right.$ Error $_{\text {minima }}=0.38 \mathrm{kcal} / \mathrm{mol}$.

To determine the torsion terms relevant to substituted amides, the rotational energy curves of N -methylacetamide and $\mathrm{N}, \mathrm{N}$-dimethylacetamide were utilized. The latter molecule enabled the derivation of a rotational profile consisting solely of the two coupled torsion terms ( $\mathrm{CG}-\mathrm{N}-\mathrm{C}-\mathrm{O} / \mathrm{CG}-\mathrm{N}-\mathrm{C}-\mathrm{CG}$ ), while $N$-methylacetamide allowed us to examine the transferability of the $\mathrm{CG}-\mathrm{N}-\mathrm{C}-\mathrm{O} / \mathrm{CG}-\mathrm{N}-\mathrm{C}-\mathrm{CG}$ and the $\mathrm{H}-\mathrm{N}-\mathrm{C}-\mathrm{O} / \mathrm{H}-\mathrm{N}-$

C-CG coupled torsion pairs. As it is impossible to separate these terms, coupled through a
$180^{\circ}$ phase-shift, the energy contributions were derived by simultaneously fitting to the QMderived rotational energy curve, employing torsional angles of $\theta^{\circ}(\mathrm{CG}-\mathrm{N}-\mathrm{C}-\mathrm{O})$ and $\theta+$ $180^{\circ}(\mathrm{CG}-\mathrm{N}-\mathrm{C}-\mathrm{CG})$, in a nonlinear least-squares approach. For $N, N$-dimethylacetamide GLYCAM06 gave the rotational barrier about the $\mathrm{N}_{\mathrm{sp} 2}-\mathrm{C}_{\mathrm{sp} 2}$ bond to be $14.6 \mathrm{kcal} /$ mol, which is lower than the B3LYP/6-31++G(2d,2p) value of $16.9 \mathrm{kcal} / \mathrm{mol}$. However, GLYCAM06 is in better agreement with the NMR determined ${ }^{130}$ effective barrier of 15.3 $\mathrm{kcal} / \mathrm{mol}$, Figure 8. The transferability of the GLYCAM06 parameters was illustrated by the reproduction of the B3LYP/6-31++G(2d,2p) rotational energy curve for the $\mathrm{N}_{\mathrm{sp} 2}-\mathrm{C}_{\mathrm{sp} 2}$ bond in N -methylacetamide, in which the trans rotamer was correctly predicted to be less stable than the cis rotamer, Figure 8. $\mathrm{N}, \mathrm{N}$-dimethylacetamide was also used to parameterize the $\mathrm{CG}-\mathrm{N}-\mathrm{CG}-\mathrm{H} 1 / \mathrm{C}-\mathrm{N}-\mathrm{CG}-\mathrm{H} 1$ coupled torsion pair. The $\mathrm{C}-\mathrm{N}-\mathrm{CG}-\mathrm{H} 1$ torsion is present in $N$-acetylated carbohydrates, and its parameters are relevant to the $\mathrm{C}-\mathrm{N}-\mathrm{CG}$ -H2 linkage present in N -linked sugars. GLYCAM06 gave a rotational barrier of 0.64 $\mathrm{kcal} / \mathrm{mol}$ for the $\mathrm{N} —$ CG bond, which may be compared with the B3LYP/6-31++G(2d,2p) value of $0.59 \mathrm{kcal} / \mathrm{mol}$.

Esters were employed for generating parameters necessary for extending the force field to model the properties of $O$-acetylated carbohydrates such as those common in certain bacterial capsular polysaccharides, and in the sialic acid residues of higher animals. ${ }^{131}$ The overall <Error> curve computed for the parameters of the following atomic sequences H1-$\mathrm{CG}-\mathrm{OS}-\mathrm{C}, \mathrm{HC} \mathrm{CG} \mathrm{C}-\mathrm{OS}, \mathrm{CG}-\mathrm{OS}-\mathrm{C}-\mathrm{O}, \mathrm{CG}-\mathrm{OS}-\mathrm{C}-\mathrm{CG}$ and $\mathrm{CG}-\mathrm{CG}-\mathrm{OS}-$ C were $0.03,0.02,0.18,0.62$, and $0.45 \mathrm{kcal} / \mathrm{mol}$, respectively.

The GLYCAM06 rotational energy barrier for the $O$-methyl group (H1-CG-OS-C) in methyl acetate was $0.91 \mathrm{kcal} / \mathrm{mol}$ which can be compared with the B3LYP/6-31++G(2d,2p) and microwave ${ }^{132}$ values of 0.87 and $1.22 \mathrm{kcal} / \mathrm{mol}$, respectively. The negligible rotational energy barrier for the acetyl-methyl group (HC-CG—C-OS) of $0.25 \mathrm{kcal} / \mathrm{mol}$ was also in good agreement with both B3LYP/6-31++G(2d,2p) ( $0.23 \mathrm{kcal} / \mathrm{mol}$ ) and experimental ( 0.29 $\mathrm{kcal} / \mathrm{mol})$ values. ${ }^{132}$ In esters, the CG—OS—C-O and CG—OS—C—CG dihedral angles are coupled through a $180^{\circ}$ phase-shift. The parameters for these atomic sequences were derived by following the same simultaneous fitting approach employed in developing the coupled parameters $(\mathrm{CG}-\mathrm{N}-\mathrm{C}-\mathrm{O} / \mathrm{CG}-\mathrm{N}-\mathrm{C}-\mathrm{CG})$ in amides, with torsional angles of $\theta^{\circ}(\mathrm{CG}-\mathrm{OS}-\mathrm{C}-\mathrm{O})$ and $\theta+180^{\circ}(\mathrm{CG}-\mathrm{OS}-\mathrm{C}-\mathrm{CG})$. This simultaneous fit resulted in $V_{2}$ values of -3.20 and $3.0 \mathrm{kcal} / \mathrm{mol}$ for $\mathrm{CG}-\mathrm{OS}-\mathrm{C}-\mathrm{O}$ and $\mathrm{CG}-\mathrm{OS}-\mathrm{C}-\mathrm{CG}$ atomic sequence, respectively. The GLYCAM06 rotational energy curve for the CG-OS-C-O sequence resulted in a local minimum at $180^{\circ}$, which was $7.52 \mathrm{kcal} / \mathrm{mol}$ less stable than the global minimum at $0^{\circ}$. This relative energy underestimated the IR spectroscopy ${ }^{133}$ value of $8.5 \pm 1.0 \mathrm{kcal} / \mathrm{mol}$, but was comparable to the B3LYP/6-31++G(2d, 2 p$)$ value of $7.48 \mathrm{kcal} /$ mol.

Lastly, $N$-tert-butylacetamide and $N$-(tetrahydro-2H-pyran-3-yl) acetamide were utilized to determine the coupled torsion parameters (CG-CG-N-H/CG-CG-N-C) and (H1-$\mathrm{CG}-\mathrm{N}-\mathrm{H} / \mathrm{CG}-\mathrm{CG}-\mathrm{N}-\mathrm{C}$ ), while ethyl acetate and tetrahydro-2H-pyran-3-yl acetate were employed in deriving the $\mathrm{C}-\mathrm{OS}-\mathrm{CG}-\mathrm{CG}$ parameters, Figure 9. The B3LYP/6-31+ $+\mathrm{G}(2 \mathrm{~d}, 2 \mathrm{p})$ rotational energy curve for the $\mathrm{CG}-\mathrm{CG}-\mathrm{N}-\mathrm{C}$ sequence in N -tert-
butylacetamide resulted in a threefold $120^{\circ}$ periodic profile, which GLYCAM06 reproduced with <Error> ${ }_{\text {curve }}=0.05 \mathrm{kcal} / \mathrm{mol}$. For this rotation in tetrahydro- 2 H -pyran-3-yl acetate, however, both GLYCAM06 and the B3LYP/6-31++G(2d,2p) energies resulted in asymmetric profiles with a broad low energy global minimum at $90^{\circ}$, which was $\sim 1.75$ $\mathrm{kcal} / \mathrm{mol}$ more stable than a local minimum at $300^{\circ}$. The minimum energy structure at $90^{\circ}$ is consistent with the conformational properties of the acetamido moiety in the solid-phase structures of Neu5Ac methyl ester and its monohydrate, ${ }^{106,134}$ as well as the solution-phase NMR structure of $\beta$-D-2-deoxy-2- $N$-acetylglucopyranoside, ${ }^{135}$ in which the carbonyl group almost eclipses the ring $\mathrm{C}-\mathrm{H}$ bond. This spatial orientation results in an approximate trans dihedral angle between the amide proton and the ring aliphatic hydrogen atom in the H 1 -$\mathrm{CG}-\mathrm{N}-\mathrm{H}$ sequence. It should be noted that at the global minimum of the $\mathrm{CG}-\mathrm{CG}-\mathrm{N}-$ C sequence being driven, the $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{H}$ dihedral angle is approximately at the trans conformation. The ability of these parameters to model their corresponding rotational energy curves was reflected in the <Error> ${ }_{\text {curve }}$ and <Error> minima values of 0.52 and $0.08 \mathrm{kcal} / \mathrm{mol}$, respectively, Figure 9. However, during a 10-ns MD simulation of $N$-acetyl 2-deoxy- $N$ -acetyl- $\beta$-D-glucopyranoside employing the QM -derived parameters, rotation about the $\mathrm{C}_{\mathrm{sp} 3}-\mathrm{N}_{\mathrm{sp} 2}(\mathrm{CG}-\mathrm{N})$ bond to the local minimum ( $\tau_{\mathrm{C}-\mathrm{N}-\mathrm{C} 2-\mathrm{H} 2} \approx 180^{\circ}$ ) occurred after $\sim 1 \mathrm{~ns}$. This rotamer remained stable throughout the remainder of the simulation. A very recent experimental and QM-NMR study of the conformational properties of this group in GlcNAc indicated that the $\mathrm{H}-\mathrm{C}_{\mathrm{sp} 3}-\mathrm{N}_{\mathrm{sp} 2}-\mathrm{H}$ atomic sequence exists only in the trans low energy con-formation $\left(\tau_{\mathrm{C}-\mathrm{N}-\mathrm{C} 2-\mathrm{H} 2} \approx 0^{\circ}\right)$, but displays broad librations around that minimum. ${ }^{136}$ Therefore, to increase the stability of the $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{H}$ trans rotamer, and ultimately prevent this conformational transitions from occurring, $V_{1}$ terms were introduced for the H1 $-\mathrm{CG}-\mathrm{N}-\mathrm{H}$ and $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{C}$ parameters to determine a reasonable value. The $V_{1}$ coefficients were varied from 0.0 to $2.0 \mathrm{kcal} / \mathrm{mol}$ in increments of $0.5 \mathrm{kcal} / \mathrm{mol}$, and PMF calculations were performed for each value, employing $N$-(tetrahydro-2H-pyran-3-yl) acetamide with the $\mathrm{C} 1-\mathrm{C} 2-\mathrm{N}-\mathrm{C}$ dihedral angle as the reaction coordinate. Based on the $\sim 2 \mathrm{kcal} / \mathrm{mol}$ increase in the relative energy of the local minimum, Figure 9, the pair of values 1.0 and $-1.0 \mathrm{kcal} / \mathrm{mol}$ was selected for the $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{H}$ and $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{C}$ torsion parameters, respectively. Employing these values, a $50-\mathrm{ns} \mathrm{MD}$ simulation performed with $N$-acetyl 2-deoxy- $N$-acetyl- $\beta$-D-glucopyranoside did not display any conformational transitions of the acetamido moiety. The origin of the observed solution-phase transitions, when employing the purely QM-derived torsion terms, remains undetermined.

In the case of ethyl acetate, rotating the about the $\mathrm{C}_{\mathrm{sp} 3}-\mathrm{O}_{\mathrm{sp} 3}$ bond in the atomic sequence (CG-CG-OS-C) resulted in a symmetric B3LYP/6-31++G(2d,2p) energy curve, which produced a large barrier $(7.45 \mathrm{kcal} / \mathrm{mol})$ at $0^{\circ}$, with minima at the trans configuration as well as at $90^{\circ}$ and $270^{\circ}$. A negligible QM-determined rotational energy barrier ( $0.83 \mathrm{kcal} / \mathrm{mol}$ ) separated the global trans minimum from the local minima at $90^{\circ}$ and $270^{\circ}$. These results were consistent with previous theoretical studies of ethyl acetate. ${ }^{137}$ GLYCAM06 reproduced the B3LYP/6-31++G(2d,2p) relative energies over the entire curve with <Error> ${ }_{\text {curve }}=0.45 \mathrm{kcal} / \mathrm{mol}$. As for the CG-CG-OS—C rotation in tetrahydro- 2 H -pyran-3-yl acetate, both B3LYP/6-31++G(2d,2p) and GLYCAM06 produced similar asymmetric energy profiles, Figure 9 . The <Error> ${ }_{\text {curve }}$ for this atomic sequence was 0.54 $\mathrm{kcal} / \mathrm{mol}$. To assess the extent to which GLYCAM06 could model other $O$-acetylated
pyranosyl moieties, rotations were performed around the $\mathrm{C}_{\mathrm{sp} 3}-\mathrm{O}_{\mathrm{sp} 3}$ bond in both tetrahydro-2-methoxy-2H-pyran-3-yl acetate and (tetra-hydro-2H-pyran-2-yl)methyl acetate, Figure 10. Unlike the rotational energy profile of tetrahydro- $2 H$-pyran-3-yl acetate, the energy profile of tetrahydro-2-methoxy-2H-pyran-3-yl acetate resulted in a barrier to rotation at $\sim 30^{\circ}$, which could arise from Coulombic repulsions between the carbonyl oxygen and the oxygen atom of the methoxy group. Compared to the QM-value, GLYCAM06 underestimated this repulsion by about $2.31 \mathrm{kcal} / \mathrm{mol}$. Overall, there was a good reproduction of the B3LYP/6-31++G(2d,2p) data (<Error> curve values of 0.89 , and 0.50 , $\mathrm{kcal} / \mathrm{mol}$, respectively).

## Carboxylates

Under physiological conditions, the extent to which carboxylate groups are ionized depends on their pKa . In the case of the ionic state it should be noted that it will never exist in the absence of a counter ion, such as sodium. It is generally presumed that uronic and ulosonic acids exist in the ionized form, and parameters were derived accordingly, with exceptions noted below. The B3LYP/6-3111G(2d,2p) rotational energy curve about the $\mathrm{C}_{\mathrm{sp} 2}-\mathrm{C}_{\mathrm{sp} 3}(\mathrm{C}$ -CG) bond in 2-methylpropanoate displayed minima at $120^{\circ}$ and $300^{\circ}$. For this rotation, GLYCAM06 was fit to the MP2/cc-pVDZ rotational energies, Figure 11, with <Error> ${ }_{\text {curve }}$ and 〈Error> ${ }_{\text {minima }}$ values of 0.72 and $0.60 \mathrm{kcal} / \mathrm{mol}$, respectively. The QM-determined rotational energy barrier of the carboxyl group was $0.8 \mathrm{kcal} / \mathrm{mol}$. This is a very low rotational barrier, which is less than the room temperature equipartition energy of a molecule $(k T)$. Thus during an MD simulations, free rotations of the carboxylic group about the $\mathrm{C}-\mathrm{CG}$ bond are expected. To this extent, it was more important to ensure reasonable barriers to rotation, than to identify the minima.

In addition to the $\mathrm{O} 2-\mathrm{C}-\mathrm{CG}-\mathrm{CG}$ torsion term, modeling uronic acids also required the development of parameters for the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{C}$ and $\mathrm{OS}-\mathrm{CG}-\mathrm{CG}-\mathrm{C}$ sequences. The $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{C}$ parameters were derived using 2-hydroxy-proponoate as the model compound, by simultaneously fitting to two sets of curves: with and without the presence of an internal hydrogen bond between the hydroxyl and carboxyl groups. For the $\mathrm{OH}-\mathrm{CG}-$ CG—C rotations, the <Error> ${ }_{\text {curve }}$ was $0.79 \mathrm{kcal} / \mathrm{mol}$ and <Error> ${ }_{\text {minima }}$ was $0.31 \mathrm{kcal} / \mathrm{mol}$ (with the inclu sion of intramolecular H-bond), while <Error> ${ }_{\text {curve }}$ and <Error> minima were $0.96 \mathrm{~mol}^{-1}$ and $0.16 \mathrm{kcal} / \mathrm{mol}$, respectively (when the H -bond was disallowed). The <Error> ${ }_{\text {curve }}$ and <Error> ${ }_{\text {minima }}$ for the OS—CG-CG-C rotations were 1.15 and 0.09 $\mathrm{kcal} / \mathrm{mol}$, respectively.

## Mixed Functional Groups

Mixed functional groups that are present in carbohydrates include the alcohol-ethers, alcohol-amides, and ether-amides. alcohol-ethers are required in the derivation of torsion terms that model the torsion angles involving the ring oxygen and the C 2 hydroxyl group as well as the glycosidic ether oxygen and other ring hydroxyl groups. These molecules with mixed functionalities can form intramolecular H-bonds, hence rotational energy curves were derived in the presence and absence of internal H-bonds, as appropriate. The B3LYP/6-31+ $+\mathrm{G}(2 \mathrm{~d}, 2 \mathrm{p})$ and GLYCAM06 $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OS}$ rotational energy curves for 2-
methoxybutan-3-ol are shown in Figure 12. Overall for ether-alcohols, the <Error> ${ }_{\text {curve }}$ was $0.61 \mathrm{kcal} / \mathrm{mol}$ with a value of $0.59 \mathrm{kcal} / \mathrm{mol}$ for the $\langle\text { Error }\rangle_{\text {minima }}$.

The highly polar mixed functionality of ether-amides was challenging to parameterize (Fig. S1, supplementary materials), as indicated in the overall values of $\langle\text { Error }\rangle_{\text {curve }}=1.12$ $\mathrm{kcal} / \mathrm{mol}$ and $\langle\text { Error }\rangle_{\text {minima }}=1.10 \mathrm{kcal} / \mathrm{mol}$. In terms of carbohydrates, the $\mathrm{OS}-\mathrm{CG}-\mathrm{N}-$ C parameter is relevant to the modeling carbohydrate-protein linkages. In this set of parameters, the OS-CG-N-C parameter was the least accurate with $<$ Error ${ }_{\text {curve }}=1.93$ $\mathrm{kcal} / \mathrm{mol}$. The other ether amide parameters performed better as can be seen in Table 5.

To determine the torsion parameters associated with the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{N}$ rotation, molecules having both alcohol and amide functionality were employed. The overall values for <Error> ${ }_{\text {curve }}$ and <Error> ${ }_{\text {minima }}$ were 0.49 and $0.41 \mathrm{kcal} / \mathrm{mol}$, respectively. The performance of these parameters is exemplified by $N$-ethanolacetamide as shown in Figure 13.

## Special Cases

Ulosonic Acids—For GLYCAM06 to correctly model the glycosidic conformational properties of ulosonic acids such as 5-N-acetylneuraminic acid (Neu5Ac), a class of sugars frequently present at the termini of eukaryotic cell surface glycans, accurate parameterization of the anomeric angle $\phi\left(\mathrm{C}_{x} \mathrm{C}_{x} \mathrm{O}_{x-1} \mathrm{C}_{x-1}\right)$ is required, Figure 3. Unlike most glycosyl residues, ulosonic acids contain a carboxyl functional group attached to the anomeric carbon atom, which alters the rotational preferences for the $\phi$-angle in this class of compounds. In solution, the $\phi$-angle predominantly exists in two conformations, -gauche and trans, in an $\sim 1: 1$ ratio. ${ }^{104,138}$ The torsion term (C-CY-OS-CG) associated with the carboxylate group for Neu5Ac contributes significantly to the $\phi$-angle conformation in these carbohydrates. Generating parameters for this torsion term proved challenging; exhibiting ether-carboxylate mixed functionality. Initial MD simulations of the glycoside of methyl $a$ Neu5Ac employing parameters generated by fitting to gas-phase QM rotational energy profiles of either ionized or protonated 2-methoxypropanoate resulted in the sampling of only the -gauche rotamer about the $\phi$-angle of methyl $\alpha$-Neu5Ac. The inability of the initial parameters to reproduce experimental $\phi$ populations may originate from the omission of charge polarizabilities in the force field. Thus, we took the approach of empirically varying the $V_{1}$ coefficient of this term until we obtained the proper rotamer distribution for $\phi$ from explicitly solvated simulations of methyl $\alpha$-Neu5Ac. By changing the $V_{1}$ value from 1.0 to $8.0 \mathrm{kcal} / \mathrm{mol}$ it was possible to shift the preference of the sampling from predominantly gauche to exclusively trans. An intermediate value of $3.5 \mathrm{kcal} / \mathrm{mol}$ reproduced the experimental rotamer distribution, Figure 14. Employing this $V_{1}$ term, a PMF analysis was carried out to determine the energy profile of the $\phi$-angle in methyl $\alpha$-Neu5Ac in solution, Figure 14. The PMF results indicated that the -gauche and trans conformers now displayed similar relative energies in solution, with a stabilization energy of $\sim 2.3 \mathrm{kcal} / \mathrm{mol}$ over the + gauche. The rotational energy barrier between the -gauche and trans conformers was $\sim 1.4$ $\mathrm{kcal} / \mathrm{mol}$. The small barrier to rotation results in the frequent transitions between the two rotamers, as observed in the MD simulation trajectory, Figure 14. To assess whether this new torsional coefficient could model a mixed system of ionized and neutral carboxylic acid
groups, the GLYCAM06 rotational energy curve for the rotation around the exocyclic $\mathrm{C}_{\mathrm{sp} 3}-\mathrm{O}_{\mathrm{sp} 3}\left(\mathrm{C}_{1}-\mathrm{O}_{1}\right)$ bond was compared with the B3LYP/6-31++G(2d,2p) curves for both ionized and neutral $(R)$-tetrahydro-2-methoxy-2H-pyran-2-carboxylic acid, Figure 15. Notably, the GLYCAM06 rotational behavior appeared to be a compromise between each of the QM curves in ionized and neutral ( $R$ )-tetrahydro-2-methoxy-2H-pyran-2-carboxylic acid.

Force Field Validation-In the course of developing torsion parameters for molecules in the training set, most of the parameters were validated concurrently by determining how well GLYCAM06 was able to reproduce the B3LYP/6-31++G(2d,2p) rotational energy curves for test molecules, or by performing MD simulations and comparing MD-computed rotamer populations to those observed experimentally. In this section, we extend the validation by examining how the parameters collectively perform in selected glycans.

Unlike the earlier parameterization of GLYCAM, ${ }^{21}$ which was based on carbohydratespecific analogues, only small representative molecules of a broad diversity were employed in the derivation of GLYCAM06. Therefore, validation of the force field by comparison with carbohydrate molecules provides an independent assessment of the ability of this small-molecule-based approach to model carbohydrate properties. Explicit solvent MD simulations ( 50 ns ) were performed for $a$-D-Glc $p \mathrm{OMe}, a$-D-Gal $p \mathrm{OMe}$ and $a$-D-Man $p \mathrm{OMe}$, and compared with NMR experimental data, such as rotamer populations and scalar ${ }^{3} J$ couplings. While scalar ${ }^{3} J$-couplings can rarely provide a complete picture of the structural properties of carbohydrates, when combined with nOe distances both datasets can provide valuable insights about oligosaccharide conformational properties. Average ring puckering parameters were also computed from the MD simulation data and compared with solid phase neutron diffraction data. The comparison of ring puckering parameters between gas-phase and crystallographic data is not ideal. However, Momany and Willet ${ }^{139}$ have shown that the geometrical properties of pyranosyl rings from gas-phase B3LYP/6-31G* calculations were remarkably close to those determined in the solid-phase, suggesting that the packing environment may not have a significant effect on the geometry of the ring.

In solution, the rotamer preferences of the $\omega$-angle $\left(\mathrm{O}_{5}-\mathrm{C}_{5}-\mathrm{C}_{6}-\mathrm{O}_{6}\right)$ involving the exocyclic hydroxymethyl group are greatly influenced by the stereochemistry at the C 4 position and the polarity of the solvent. ${ }^{28}$ The three accessible staggered rotamers about the $\mathrm{C}_{5}-\mathrm{C}_{6}$ bond are generally designated gauche-gauche (gg), gauche-trans (gt), and transgauche ( $t g$ ) with respect to the dihedral angles $\mathrm{O}_{5}-\mathrm{C}_{5}-\mathrm{C}_{6}-\mathrm{O}_{6}$ and $\mathrm{C}_{4}-\mathrm{C}_{5}-\mathrm{C}_{6}-\mathrm{O}_{6}$, respectively. A significant test of a carbohydrate force field is its ability to reproduce the rotamer populations about the $\mathrm{C}_{5}-\mathrm{C}_{6}$ exocyclic bonds because the populations are sensitive to the energetic balance between the formation of intramolecular and solute-solvent hydrogen bonds. ${ }^{140}$ The ability of GLYCAM06 to reproduce the NMR $J$-couplings for the $\mathrm{C}_{5}-\mathrm{C}_{6}$ bond is reflective of the performance of its dihedral, nonbonded, and electrostatic components.

Homonuclear scalar ${ }^{3} J$-couplings between the $\mathrm{H}_{5}-\mathrm{C}_{5}-\mathrm{C}_{6}-\mathrm{H}_{6 \mathrm{R}} / \mathrm{H}_{6 \mathrm{~S}}$ hydrogen atoms from a 50-ns explicit solvent MD simulation of $a$-D-GlcpOMe were computed, employing an experimentally parameterized Karplus equation. ${ }^{141}$ Values computed over the entire
trajectory resulted in coupling constants of $5.4 \pm 1.7 \mathrm{~Hz}$ and $2.9 \pm 2.0 \mathrm{~Hz}$ for ${ }^{3} J_{\mathrm{H} 5 \mathrm{H} 6 \mathrm{R}}$ and ${ }^{3} J_{\mathrm{H} 5 \mathrm{H} 6 \mathrm{~S}}$, respectively, which compared favorably with experimental values, ${ }^{142-144}$ Table 7. In the course of the MD simulation, rotations about the exocyclic $\mathrm{C}_{5}-\mathrm{C}_{6}$ bond resulted in each of the rotational states of the $\omega$-angle being populated. The MD gg:gt:tg populations, $62: 36: 2$, were in good agreement with the experimentally observed populations of 57:38:5 ${ }^{144}$ and 53:47:0. ${ }^{142}$

Similar $J$-calculations were performed for $\alpha$-D-GalpOMe, the C-4 epimer of Glcp. From the MD data, the ${ }^{3} J_{\mathrm{H} 5 \mathrm{H} 6 \mathrm{R}}$ and ${ }^{3} J_{\mathrm{H} 5 \mathrm{H} 6 \mathrm{~s}}$ coupling constants were $7.9 \pm 1.6$ and $3.7 \pm 1.8 \mathrm{~Hz}$, respectively. Several experimental values have been reported for the ${ }^{3} J_{\mathrm{H} 5 \mathrm{H} 6 \mathrm{R} / \mathrm{S}}$ coupling constants and the $g g: g t: t g$ populations of $a$-D-GalpOMe, Table 7. The MD-computed values for ${ }^{3} J_{\mathrm{H} 5 \mathrm{H} 6 \mathrm{R}}$ and ${ }^{3} J_{\mathrm{H} 5 \mathrm{H} 6 \mathrm{~S}}$ fell within the experimental ranges of $7.8-8.6 \mathrm{~Hz}$ and $3.7-6.0 \mathrm{~Hz}$, respectively. ${ }^{142-144}$ With respect to the gg:gt:tg populations, the MD values of 6:76:18 showed good agreement with the more recent data.

In addition to bond rotational properties, pyranose ring puckering parameters, ring interproton torsion angles and distances provide another means of assessing the accuracy of the ring geometry. Analyses of the Cremer and Pople ${ }^{145}$ ring puckering parameters over the 50-ns MD simulations of $\alpha$-D-Glc $p \mathrm{OMe}, \alpha$-D-Gal $p \mathrm{OMe}$ and $\alpha$-D-Man $p \mathrm{OMe}$ indicated stable ${ }^{4} \mathrm{C}_{1}$ ring chair forms, with $(Q, \theta)$ values of $(0.55 \pm 0.04,6 \pm 10),(0.57 \pm 0.04,24 \pm 8)$ and $(0.55 \pm 0.04,1 \pm 10)$, respectively, which were in good agreement with the values determined from solid state neutron diffraction experiments of $(0.57,2.3),{ }^{146}(0.57,4.9),{ }^{147}$ and $(0.56,0)$, respectively. ${ }^{146}$

Ring interproton distances and dihedral angles are a source of NMR-restraints employed in determining the ring geometries of glycans. In addition, the distances provide internal calibration references employed in the elucidation of glycan conformational properties from nOe data. Thus, the feasibility of employing GLYCAM06 to predict glycan ring conformational properties was determined by its ability to reproduce experimental interproton distances and dihedral angles. A comparison of interproton distances and dihedral angles computed using GLYCAM06, and those determined from neutron diffraction data ${ }^{146,147}$ for $a$-D-Glc $p \mathrm{OMe}, ~ a$-D-GalpOMe and $\alpha$-D-Man $p \mathrm{OMe}$ is presented in Table 8. To make a direct comparison between experimental and theoretical data, the coupling constants between vicinal ring-hydrogen atoms were computed from solvated MD simulations of $\alpha / \beta$-D-GlcpOMe, $\alpha / \beta$-D-GalpOMe, and $\alpha / \beta$-D-Man $p \mathrm{OMe}$, and compared with available solution-phase NMR experimental data, Table 9. The overall agreement between the GLYCAM06-derived and experimental data are quite good. However, theoretical and experimental values for the ${ }^{3} J_{\mathrm{H} 4 \mathrm{H} 5}$ and ${ }^{3} J_{\mathrm{H} 1 \mathrm{H} 2}$ constants in $\alpha / \beta$-D-Gal $p \mathrm{OMe}$ and $\beta$-D-Man $p \mathrm{OMe}$, respectively, were in weaker agreement. Superimpositions of the crystal structures with energy-minimized and average ring geometries of these pyranosides, computed from the MD simulations did not reveal any structural distortions (Fig. S2, supplementary materials). This error may be arising from electronic hyperconjugative effects between the lone pair of electrons on the ring oxygen atom and the antibonding orbital of the $\mathrm{C} 5-\mathrm{H} 5$ or the $\mathrm{C} 1-\mathrm{H} 1$ bonds when the anomeric center is in the $\beta$ configuration, which are not included in classical force fields. Using the B3LYP functional with the cc-pV5ZT and 5s2p1d basis sets for hydrogen and heavy atoms, respectively, QM

NMR predictions from the B3LYP/aug-cc-pVTZ optimized neutron diffraction structures of $\alpha$ - and $\beta$-D-Gal $p$ OMe resulted in ${ }^{3} J_{\mathrm{H} 4 \mathrm{H} 5}$ values of 1.9 and 1.2 Hz , respectively.

Another indication of the performance of a biomolecular force field is its ability to reproduce experimentally determined unit cell geometries. In general, unit cell dimensions are influenced by internal torsional rotations as well as by van der Waals and electrostatic intermolecular interactions. However, it is recognized that in polar molecules such as carbohydrates, the treatment of electrostatics is crucial in reproducing experimental crystal geometries. ${ }^{31}$ The suitably of the present treatment of electrostatics in GLYCAM06 was assessed by comparing the average difference between the unit cell geometries of a methyl $a$-D-Glc $p$ crystal computed from a 1-ns MD simulation to those determined experimentally. ${ }^{63}$ The average difference computed with GLYCAM06 was $1.2 \AA$, which showed better performance when compared with the average differences computed with GLYCAM_93, ${ }^{21}$ CHARMM HGFB, ${ }^{148}$ and GROMOS (93) ${ }^{149}$ of $1.7,1.8$, and $1.6 \AA$, respectively.

The results of the frequency calculations, assignments, and comparisons to the various experimental results are summarized in Figure 16. The character of the assignments generated by our program is illustrated in Table 10, where we have listed the primary motion assigned by DMODES to each vibrational frequency for the monomer calculated by NMODE. The notations are described in the caption. The assignments for the monomer were also checked by visually inspecting corresponding plots of the atom positions with energy-weighted motion vectors. For the comparisons with experiment, the DMODESgenerated assignments were grouped into the broader classes used in Figure 16, Figure 17 and Table 11.

In general, the agreement between the calculated frequencies and the experimental results is good, with the only notable exceptions being the $\mathrm{H}-\mathrm{O}$ stretches, which are about $5 \%$ blueshifted. This deviation is likely due to the use of the gas-phase $\mathrm{H}-\mathrm{O}$ stretching frequency for methanol in the initial derivation of the AMBER HO-OH stretching constant. ${ }^{150}$ Since the gas-phase $\mathrm{H}-\mathrm{O}$ stretching frequency of methanol is about $360 \mathrm{~cm}^{-1}$ higher than the liquid phase value, it is reasonable to expect that a smaller force constant would produce better results in the condensed phases.

We have chosen to use the standard AMBER values for the $\mathrm{H}-\mathrm{O}$ stretches, and also the H -C, for two reasons. First, we wish to minimize the introduction of new atom types, and associated complications, into the AMBER force field. Second, fitting force constants to quantum-derived separation energies for these bond stretches does not improve the function of the force field. Instead, the quantum derived force constants increase the normal mode frequency disagreement to $10-20 \%$. However, although the current values work well, the AMBER force field is used primarily for condensed-phase simulations, so it may be valuable to reevaluate the use of the current $\mathrm{HO}-\mathrm{OH}$ stretching constant.

Although the $\mathrm{O}-\mathrm{C}$ stretches seem in Figure 16 to differ by $\sim 50 \%$, the issue is more complex, and is partly one of assignment. As with the experimental work, we also see $\mathrm{O}-\mathrm{C}$ stretches in the region $\sim 1000-1200 \mathrm{~cm}^{-1}$, as is illustrated for the 64-mer in Figure 17.

However, according to our molecular mechanics analysis of the monomer and the 64-mer, as well as the quantum analyses of the monomer (Tables 10 and 11 for the monomer data), the O - C stretches do not account for the major portion of the vibrational kinetic energy at these frequencies. In Figure 17, we present all frequencies calculated for the crystal that corresponded to a $\mathrm{C}-\mathrm{C}$ or $\mathrm{O}-\mathrm{C}$ stretch. The figure shows that the calculation yields many $\mathrm{C}-\mathrm{C}$ and $\mathrm{O}-\mathrm{C}$ stretches in the range of "heavy atom stretches" given by Wells and Atalla. ${ }^{71}$ The more intense stretches at the lower frequencies are all parts of larger, more complex motions that would not typically be assigned as stretches, so frequencies in these ranges are not expected to correspond to experiment. On the other hand, the presence of the cluster of high-frequency C-C stretches, about 1400-1600, is puzzling. The cause of the discrepancy is not known. Kuttel et al., ${ }^{68}$ reported similar calculated frequencies, although theirs also included $\mathrm{O}-\mathrm{C}$ stretches. They attributed the high frequencies to the influence of ring-atom bond and angle force constants that are not mitigated by the presence of cross terms in the force field. Although our outlying frequencies might have a similar cause, further investigation is needed to determine the precise cause of the discrepancy.

The agreement with experimental results for bending, distortion, and torsion modes, on the other hand, is very good (Figure 16). Our calculated ranges significantly overlap experimental ranges for these motion classes, though the experimental ranges do not always agree with each other. For example, consider the data for the $\mathrm{H}-\mathrm{O}-\mathrm{C}$ torsions. These show the least consensus among the other studies. This may in part be due to difficulties translating between assignment conventions, but it may more likely represent a wide variability in those torsional frequencies, as is found for the 64-unit calculation. Similar arguments might be made for the "other C6 \& O6" class. It should be noted that the Gregurick and Kafafi ${ }^{70}$ study used a different assignment method than what was used in the other studies to which we compare our results, which might account for the difference between their results and those of the others within these classes.

Two features in Figure 16 that are specific to the results of this study require further discussion. The first is the presence of the highest frequency - C 6 - distortion modes. There were no assignments offered by the other studies that satisfactorily correlated with this motion, which is essentially the movement of the C6 atom along a line between the C5 and O6 atoms. There is accompanying motion in C5 and in the two hydrogens attached to C6, but the primary energy is in the motion of the C6. Note that this discussion only applies to the - C6- distortion modes above $1500 \mathrm{~cm}^{-1}$ (the lower frequency modes are other stretching motions of the C6 atom, for example, along the bond to one of the aliphatic hydrogens). Considering the strong stretching character of the motion, it is likely that the frequency reported here is only approximate. Normal mode analysis of the HF/6-31G* data for the monomer revealed a similar mode at $1635 \mathrm{~cm}^{-1}$ (Table 11). The density functional calculation generated a similar mode at $1458 \mathrm{~cm}^{-1}$. This mode is likely to be real, but may be difficult to observe since the motion is highly symmetric. In all of the quantum calculations in which the vibrational kinetic energy was primarily coupled to the motion of C6 between C5 and O6, the IR intensity calculated for that mode was small.

The other feature of note is the small cluster of very high frequency $\mathrm{H}-\mathrm{O}-\mathrm{C}$ torsions noted with $(*)$ in Figure 10. Note that the relative energy associated with the motion is
small, but it was not small enough to be filtered out by the procedures described earlier. The motions responsible for these anomalous assignments make physical sense, and are likely to occur in nature, but may be too weak to be observed. The cluster of torsions arises due to an accidental resonance. In the 64-unit crystal used for the calculation, the minimization resulted in a configuration such that a stretch of the $\mathrm{O} 2-\mathrm{HO} 2$ of a single molecule would necessarily excite torsion of the hydroxyl HO6 of another neighboring molecule. Since one of the molecules was situated at one a face of the crystal, and since NMODE does not employ periodic boundaries, the isolated behavior was not representative of the bulk properties.

## Conclusions

This work describes the derivation of a parameter set for classical quadratic force fields that accurately models carbohydrates, but which can also be generalized to more diverse molecules by virtue of the noncarbohydrate specific nature of the parameters. QM calculations were employed to compute properties that are difficult or impossible to access experimentally, such as, bond and valence angle deformation force constants, dihedral angle rotational barriers, and electrostatic properties. The structures of the biomolecular building blocks (monosaccharides in the present case) were taken from experimental neutron diffraction data. Consistent with the AMBER force field, the PARM94 parameter set for van der Waals terms was employed. Particular attention was given to those properties that most impact the overall 3D structure and dynamics of biopolymers; namely, internal electrostatic interactions, solute-solvent interactions, and internal rotational barriers. While we derived partial atomic charges by fitting to the QM molecular electrostatic potentials (so called ESPfitting), we departed from the general practice of assigning partial atomic charges to every atom in the molecule adopted in our earlier work (GLYCAM_93, GLYCAM2000) by not fitting partial charges to aliphatic hydrogen atoms. In addition, the current parameterization removes the need to treat $1-4$ non-bonded interactions as a special case; that is, $1-4$ scaling has been removed.

Although the parameter set is complete for carbohydrates, the methodology and the force field may be readily extended to lipids, proteins, and nucleic acids. The generality of the parameters is exemplified by the utilization of a common set of terms for $\alpha$-and $\beta$ carbohydrate anomers. When combined with appropriate charge sets, these common terms provide a means of predicting the relative energies of interconverting ring forms. Not surprisingly, for nonpolar molecules the parameters derived by fitting to gas-phase B3LYP/ $6-31++G(2 d, 2 p)$ data successfully reproduced the QM rotational energies. In addition, the force field was able to reproduce solid state ring puckering parameters, ring inter-proton torsion angles and distances, solution phase populations of the $\omega$-angle, and scalar ${ }^{3} \mathrm{~J}$ coupling constants for the $\mathrm{H}_{5}-\mathrm{C}_{5}-\mathrm{C}_{6}-\mathrm{H}_{6 \mathrm{R} / \mathrm{S}}$ atoms in representative carbohydrates. In highly polar molecules, however, the percentage errors in the torsional energies were occasionally high, and in the case of carboxylate ions the gas-phase QM-derived parameters failed to reproduce solution phase conformational properties. The expectation that condensed phase simulations, employing dihedral terms fit to gas-phase QM data, can reproduce solution rotamer populations is founded on the assumption that internal electrostatic polarization is relatively independent of internal rotation, either in the gas or
condensed phases. This is frequently a reasonable assumption for relatively nonpolar molecules, but we found that as molecular polarity increased the observed condensed phase rotamer populations began to deviate from experimental expectations. This is clearly the case for ionic systems, for which it was necessary, in the absence of an explicit treatment of polarizability, to adjust the dihedral terms empirically to achieve accurate behavior in the aqueous solution. However, the empirically-corrected rotational energy curves were shown to differ only modestly from the gas-phase QM data. For such systems an iterative procedure, which involved explicit solvent MD simulations, was adopted to generate the torsion parameters. The inadequacy of the gas-phase QM-derived torsion terms when applied to highly polar molecules may be attributed to the absence of charge polarizability in the classical force field. The development of a polarizable version of GLYCAM is underway.

## Acknowledgments

RJW would like to thank Dr. Dusan Uhrin for insightful discussions.

Contract/grant sponsor: National Institutes of Health; contract/grant numbers: RR05357, GM55230

Contract/grant sponsor: Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory (PNNL); contract/grant numbers: 9607

## References

1. Weinhold B, Seidenfaden R, Rockle I, Muhlenhoff M, Schert-zinger F, Conzelmann S, Marth JD, Gerardy-Schann R, Hildebrandt H. J Biol Chem. 2005; 280:42971. [PubMed: 16267048]
2. Jin L, Abrahams JP, Skinner R, Petitou M, Pike R, Carrel RW. Proc Natl Acad Sci USA. 1997; 94:14683. [PubMed: 9405673]
3. Haltiwanger RS, Lowe JB. Annu Rev Biochem. 2004; 73:491. [PubMed: 15189151]
4. Sanders RW, Venturi M, Schiffner L, Kalyanaraman R, Katinger H, Lloyd KO, Kwong PD, Moore JP. J Virol. 2002; 76:7293. [PubMed: 12072528]
5. Almond A, Sheehan JK. Glycobiology. 2000; 10:329. [PubMed: 10704532]
6. Karaveg K, Siriwardena A, Tempel W, Liu ZJ, Glushka J, Wang BC, Moremen KW. J Biol Chem. 2005; 280:16197. [PubMed: 15713668]
7. Varki A. Glycobiology. 1993; 3:97. [PubMed: 8490246]
8. Dwek RA. Chem Rev. 1996; 96:683. [PubMed: 11848770]
9. Imberty A, Perez S. Chem Rev. 2000; 100:4567. [PubMed: 11749358]
10. Tempel W, Tschampel S, Woods RJ. J Biol Chem. 2002; 277:6615. [PubMed: 11714721]
11. Vyas NK, Vyas MN, Chervenak MC, Johnson MA, Pinto BM, Bundle DR, Quiocho FA. Biochemistry. 2002; 41:13575. [PubMed: 12427018]
12. Kogelberg H, Solis D, Jimenez-Barbero J. Curr Opin Struct Biol. 2003; 13:646. [PubMed: 14568621]
13. Landersjo C, Widmalm G. Biopolymers. 2002; 64:283. [PubMed: 12124846]
14. Sayers EW, Prestegard JH. Biophys J. 2002; 82:2683. [PubMed: 11964255]
15. Wormald MR, Petrescu AJ, Pao Y-L, Glithero A, Elliott T, Dwek RA. Chem Rev. 2002; 102:371. [PubMed: 11841247]
16. Rice KG, Wu P, Brand L, Lee YC. Biochemistry. 1991; 30:6646. [PubMed: 2065052]
17. Vliegenthart, JFG.; Woods, RJ., editors. NMR Spectroscopy and Computer Modeling of Carbohydrates: Recent Advances. American Chemical Society; Washington, DC: 2006. p. 235-257.
18. Wooten EW, Edge CJ, Bazzo R, Dwek RA, Rademacher T. W Carbohydr Res. 1990; 203:13.
19. Cano FH, Foces-Foces C, Jimenez-Barbero J, Alemany A, Bernabe M, Martin-Lomas M. J Org Chem. 1987; 52:3367.
20. Gonzalez-Outeirino J, Kadirvelraj R, Woods RJ. Carbohydr Res. 2005; 340:1007. [PubMed: 15780265]
21. Woods RJ, Dwek RA, Edge CJ, Fraser-Reid B. J Phys Chem. 1995; 99:3832.
22. Hemmingsen L, Madsen DE, Esbensen AL, Olsen L, Engelsen SB. Carbohydr Res. 2004; 339:937. [PubMed: 15010301]
23. Perez S, Imberty A, Engelsen SB, Gruza J, Mazeau K, Jimenez-Barbero J, Poveda A, Espinosa J-F, van Eyek BP, Johnson G, French AD, Kouwijzer MLCE, Grootenusi PDJ, Bernardi A, Raimondi L, Senderowitz H, Durier V, Vergoten G, Rasmussen K. Carbohydr Res. 1998; 314:141.
24. Kirschner KN, Woods RJ. Proc Natl Acad Sci USA. 2001; 98:10541. [PubMed: 11526221]
25. Corzana F, Motawia MS, Du Penhoat CH, Perez S, Tschampel SM, Woods RJ, Engelsen SB. J Comput Chem. 2004; 25:573. [PubMed: 14735575]
26. Weiner JS, Kollman PA. J Comput Chem. 1986; 7:230.
27. Almond A, Sheehan JK. Glycobiology. 2003; 13:255. [PubMed: 12626381]
28. Gonzalez-Outeirino J, Kirschner KN, Thobhani SJ, Woods R. Can J Chem Rev Can Chim. 2006; 84:569.
29. Umemura M, Yuguchi Y, Hirotsu T. J Mol Struct (THEO-CHEM). 2005; 130:1.
30. Vishnyakov A, Widmalm G, Kowalewski J, Laaksonen A. J Am Chem Soc. 1999; 121:5403.
31. Woods RJ, Chappelle R. J Mol Struct (THEOCHEM). 2000; 527:149.
32. Woods RJ, Pathiaseril A, Wormald MR, Edge CJ, Dwek RA. Eur J Biochem. 1998; 158:372. [PubMed: 9874202]
33. Biarnes X, Nieto J, Planas A, Rovira C. J Biol Chem. 2006; 281:1432. [PubMed: 16260784]
34. Ford MG, Weimar T, Kohli T, Woods R. J Proteins: Struct Funct Genet. 2003; 53:229.
35. Kadirvelraj R, Gonzalez-Outeirino J, Foley BL, Beckham ML, Jennings HJ, Foote S, Ford MG, Woods R. J Proc Natl Acad Sci USA. 2006; 103:8149.
36. Kawatkar SP, Kuntz DA, Woods RJ, Rose DR, Boons GJ. J Am Chem Soc. 2006; 128:8310. [PubMed: 16787095]
37. Krieger E, Geretti E, Brandner B, Goger B, Wells TN, Kungl A. J Proteins: Struct Funct Bioinf. 2004; 54:768.
38. Michel G, Pojasek K, Li Y, Sulea T, Linhardt RJ, Raman R, Prabhakar V, Sasisekharan R, Cygler M. J Biol Chem. 2004; 279:32882. [PubMed: 15155751]
39. Pratap JV, Bradbrook GB, Reddy A, Surolia A, Raftery J, Helliwell JR, Vijayan M. Acta Crystallogr Sect D: Biol Crystallogr. 2001; 57:1584. [PubMed: 11679723]
40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J Chem Phys. 1983; 79:926.
41. Jeffrey GA. J Mol Struct. 1994; 322:21.
42. Gruza J, Koca J, Perez S, Imberty A. J Mol Struct (THEO-CHEM). 1998; 424:269.
43. Bosques CJ, Tschampel SM, Woods RJ, Imperiali B. J Am Chem Soc. 2004; 126:8421. [PubMed: 15237998]
44. Chen S-Y, Lin T-H. J Phys Chem B. 2005; 109:9764. [PubMed: 16852176]
45. Chevalier F, Lopez-Prados J, Perez S, Martin-Lomas M, Nieto PM. Eur J Org Chem. 2005; 16:3489.
46. Nguyen DH, Colvin ME, Yeh Y, Feeney RE, Fink W. H Biophys J. 2002; 82:2892.
47. Perera L, Darden TA, Pedersen LG. J Comput Chem. 2002; 23:35. [PubMed: 11913388]
48. Shroll RM, Straatsma TP. Biophys J. 2003; 84:1765. [PubMed: 12609878]
49. Kirschner KN, Woods RJ. J Phys Chem A. 2001; 105:4150. [PubMed: 16518456]
50. Lii JH, Ma BY, Allinger N. L J Comput Chem. 1999; 20:1593.
51. Momany FA, Willet J. L Carbohydr Res. 2000; 326:210.
52. Momany FA, Willet JL. Carbohydr Res. 2000; 326:194. [PubMed: 10903029]
53. Frisch, MJ.; T, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Zakrzewski, VG.; Montgomery, JA.; Stratmann, RE., Jr; Burant, JC.; Dapprich, S.; Millam, JM.; Daniels, AD.;

Kudin, KN.; Strain, MC.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, GA.; Ayala, PY.; Cui, Q.; Morokuma, K.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Cioslowski, J.; Ortiz, JV.; Baboul, AG.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Andres, JL.; Gonzalez, C.; Head-Gordon, M.; Replogle, ES.; Pople, JA. Gaussian '98. Gaussian, Inc; Pittsburgh, PA: 2002.
54. Case, DA.; Pearlman, DA.; Caldwell, JW.; Cheatham, TE., III; Wang, J.; Ross, WS.; Simmerling, CL.; Darden, TA.; Merz, KM.; Stanton, RV.; Cheng, AL.; Vincent, JJ.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, RJ.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, GL.; Singh, UC.; Weiner, PK.; Kollman, PA. AMBER 7. University of California; San Francisco, CA: 2002.
55. Case, DA.; Darden, TA.; Cheatham, TE., III; Simmerling, CL.; Wang, J.; Duke, RE.; Lou, R.; Merz, KM.; Wang, B.; Pearlman, DA.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, P.; Caldwell, JW.; Ross, WS.; Kollman, PA. AMBER 8. University of California; San Francisco, CA: 2004.
56. Verlet L. Phys Rev. 1967; 159:98.
57. van Gunsteren WF, Berendsen HJC. Mol Phys. 1977; 34:1311.
58. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. J Comput Chem. 1992; 13:1011.
59. Roux B. Comput Phys Commun. 1995; 91:275.
60. Bayly CI, Cieplak P, Cornell WD, Kollman PA. J Phys Chem. 1993; 97:10269.
61. Basma M, Sundara S, Calgan D, Vernali T, Woods RJ. J Comput Chem. 2001; 22:1125. [PubMed: 17882310]
62. Case, DA.; Darden, TA.; Cheatham, TE., III; Simmerling, CL.; Wang, J.; Duke, RE.; Lou, R.; Merz, KM.; Pearlman, DA.; Crowley, M.; Walker, RC.; Zhang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, KF.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, DH.; Schafmeister, C.; Ross, WS.; Kollman, PA. AMBER 9. University of California; San Francisco, CA: 2006.
63. Brown GM, Levy HA. Acta Crystallogr Sect B: Struct Sci. 1979; 35:656.
64. Korolevich M. J Mol Struct. 1994; 306:261.
65. Korolevich M, Sivchik V, Zhbankov R, Marchenko G, Zabelin L. J Appl Spectrosc. 1992; 56:217.
66. Korolevich M, Zhbankov R, Sivchik V. J Mol Struct. 1990; 220:301.
67. Dauchez M, Derremaus P, Vergoten G. J Comput Chem. 1992; 14:263.
68. Kuttel M, Brady JW, Naidoo KJ. J Comput Chem. 2002; 23:1236. [PubMed: 12210149]
69. Huvenne J, Vergoten G, Fleury G, Legrand P. J Mol Struct. 1981; 74:169.
70. Gregurick S, Kafafi S. J Carbohydr Chem. 1999; 18:867.
71. Wells H, Atalla R. J Mol Struct. 1990; 224:385.
72. Prima A, Zhbankov R, Marupov R. J Struct Chem. 1967; 5:783.
73. Jorgensen WL, Tiradorives J. J Am Chem Soc. 1988; 110:1657.
74. Woods RJ, Khalil M, Pell W, Moffat SH, Smith VH Jr. J Comput Chem. 1990; 11:297.
75. Breneman CM, Wiberg KB. J Comput Chem. 1990; 11:361.
76. Mahoney MW, Jorgensen WL. J Chem Phys. 2000; 112:8910.
77. Francl MM, Carey C, Chirlian LE. J Comput Chem. 1996; 17:367.
78. Juaristi, E., editor. Conformational Behavior of Six-Membered Rings: Analysis, Dynamics and Stereochemical Effects. New York: VCH; 1995.
79. Appell M, Strati G, Willett JL, Momany FA. Carbohydr Res. 2004; 339:537. [PubMed: 15013391]
80. Jeffrey GA, Pople JA, Radom L. Carbohydr Res. 1972; 25:117.
81. Jeffrey GA, Pople JA, Binkley JS, Vishveshwara S. J Am Chem Soc. 1978; 100:373.
82. Schleifer L, Senderowitz H, Aped P, Tartakovsky E, Fuchs B. Carbohydr Res. 1990; 206:21.
83. Dinur, U.; Hagler, AT. Reviews in Computational Chemistry. VCH; New York: 1991. Chapter 4
84. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. J Comput Chem. 1983; 4:187.
85. Hermans J, Berendsen HJC, van Gunsteren WF, Postma PM. Biopolymers. 1984; 23:1513.
86. MacKerell AD, 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, Wiorkiewicz-Kuczera J, Yin D, Karplus M. J Phys Chem B. 1998; 102:3586. [PubMed: 24889800]
87. Mackerell AD Jr. J Comput Chem. 2004; 25:1584. [PubMed: 15264253]
88. Brisson JR, Baumann H, Imberty A, Perez S, Jennings H. J Biochemistry. 1992; 31:4996.
89. Potenzone JR, Hopfinger AJ. Carbohydr Res. 1976; 46:67. [PubMed: 129281]
90. Arnott S, Scott WE. J Chem Soc Perkin Trans. 2(1972):324.
91. Wang J, Cieplak P, Kollman PA. J Comput Chem. 2000; 21:1049.
92. Compton DAC, Montero S, Murphy WF. J Phys Chem. 1980; 84:3587.
93. Heenan RK, Bartell LS. J Chem Phys. 1983; 78:1270.
94. Rasanen M, Bondybey VE. J Chem Phys. 1985; 82:4718.
95. Durig JR, Bucy WE, Wurrey CJ, Carreira LA. J Phys Chem. 1975; 79:988.
96. van Alsenoy C, Scarsdale JN, Williams JO, Schafer L. J Mol Struct. 1982; 86:365.
97. Schafer L, van Alsenoy C, Scarsdale JN. J Mol Struct. 1982; 86:349.
98. Barrow GM. J Chem Phys. 1952; 20:1739.
99. Sasada Y. J Mol Struct. 1988; 190:93.
100. Weldon AJ, Vickrey TL, Tschumper GS. J Phys Chem A. 2005; 109:11073. [PubMed: 16331953]
101. Houk KN, Eksterowicz JE, Wu Y-D, Fuglesang CD, Mitchell DB. J Am Chem Soc. 1993; 115:4170.
102. Brown E, Brey W, Weltner W. Biochim Biophys Acta. 1975; 399:124. [PubMed: 1148272]
103. Michon F, Brisson JR, Jennings HJ. Biochemistry. 1987; 26:8399. [PubMed: 2450562]
104. Poppe L, Stuike-Prill S, Meyer B, van Halbeek H. J Biomol NMR. 1992; 2:109. [PubMed: 1422148]
105. Flippen JL. Acta Crystallogr Sect B: Struct Sci. 1973; 29:1881.
106. Kooijman H, Kroonbatenburg LMJ, Kroon J, Breg JN, Deboer JL. Acta Crystallogr Sect C Cryst Struct Commun. 1990; 46:407
107. Alagona G, Ghio C. J Mol Struct (THEOCHEM). 1992; 254:287.
108. Bako I, Grosz T, Palinkas G. J Chem Phys. 2003; 118:3215.
109. Bastiansen O. Acta Chem Scand. 1949; 3:415.
110. Chidichimo G, Imbardelli D, Longeri M, Saupe A. Mol Phys. 1988; 65:1143.
111. Cramer CJ, Truhlar DG. J Am Chem Soc. 1994; 116:3892.
112. Erdem SS, Varnali T, Aviyente V. J Phys Org Chem. 1997; 10:196.
113. Gubskaya AV, Kusalik PG. J Phys Chem A. 2004; 108:7165.
114. Hommel EL, John MK, Ma G, Hadad CM, Allen HC. J Phys Chem B. 2005; 109:811. [PubMed: 16866446]
115. Hooft RWW, van Eijck BP, Kroon J. J Chem Phys. 1992; 97:3639.
116. Nagy PI, Dunn WJ III, Alagona G, Ghio C. J Am Chem Soc. 1992; 113:6719.
117. Pachler KGR, Wessels PL. J Mol Struct. 1970; 6:471.
118. Saiz L, Padro JA, Guardia E. J Chem Phys. 2001; 114:3187.
119. Fateley WG, Miller FA. Spectrochim Acta. 1962; 18:977.
120. Bedrov D, Borodin O, Smith GD. J Phys Chem B. 1998; 102:5683.
121. Liu H, Muller-Plathe F, van Gunsteren WF. J Chem Phys. 1995; 1042:1722.
122. Smith GD, Jaffe RL, Yoon DY. J Am Chem Soc. 1995; 117:530.
123. Begum R, Matsuura H. J Chem Soc Faraday Trans. 1997; 93:3839.
124. Goutev N, Ohno K, Matsuura H. J Phys Chem A. 2000; 104:9226.
125. Howarth OW, Michael M. Polymer. 1997; 38:355.
126. Senderowitz H, Parish C, Still WC. J Am Chem Soc. 1996; 118:2078.
127. Lemieux, RU.; Koto, S.; Voisin, D. Anomeric Effect Origin Consequences. Szarek, WA.; Horton, D., editors. American Chemical Society; Washington, DC: 1979. Symposium Series 87Chapter 2
128. Fogarasi G, Szalay PG. J Phys Chem A. 1997; 101:1400.
129. Langley CH, Allinger NL. J Phys Chem A. 2002; 106:5638.
130. Ross BD, True NS, Matson GB. J Phys Chem. 1984; 88:2675.
131. Shen Y, Kohla G, Lrhorfi AL, Sipos B, Kalthoff H, Gerwig GJ, Kamerling JP, Schauer R, Tiralongo J. Eur J Biochem. 2004; 271:281. [PubMed: 14717696]
132. Sheridan J, Bossert W, Bauder A. J Mol Spectrosc. 1980; 80:1.
133. Blom CE, Gunthard HH. Chem Phys Lett. 1981; 84:267.
134. O’Connell AM. Acta Crystallogr Sect B: Struct Sci. 1973; 29:2320.
135. Davis JT, Hirani S, Bartlett C, Reid BR. J Biol Chem. 1994; 269:3331. [PubMed: 8106372]
136. Mobli M, Almond A. Org Biomol Chem. 2007; 4:2243. [PubMed: 17609755]
137. Smondyrev AM, Berkowitz ML. J Comput Chem. 1999; 25:531.
138. Breg J, Kroon-Batenburg MJL, Strecker G, Montreuil J, Vliegenthart JFG. Eur J Biochem. 1989; 178:727. [PubMed: 2912732]
139. Momany FA, Willet JL. J Comput Chem. 2000; 21:1204.
140. Stenutz R, Carmichael I, Widmalm G, Serianni AS. J Org Chem. 2002; 67:949. [PubMed: 11856043]
141. Haasnoot CAG, Deleeuw F, Deleeuw HPM, Altona C. Biopolymers. 1981; 20:1211.
142. Bock K, Duus JO. J Carbohydr Chem. 1994; 13:513.
143. Nishida Y, Hori H, Ohrui H, Meguro H. J Carbohydr Chem. 1988; 7:239.
144. Nishida Y, Ohrui H, Meguro H. Tetrahedron Lett. 1984; 25:1575.
145. Cremer D, Pople JA. J Am Chem Soc. 1975; 97:1354.
146. Jeffrey GA, McMullan RK, Takagi S. Acta Crystallogr Sect B: Struct Sci. 1977; 33:728.
147. Takagi S, Jeffrey GA. Acta Crystallogr Sect B: Struct Sci. 1979; 35:902.
148. Ha SN, Giammona A, Field M, Brady JW. Carbohydr Res. 1988; 180:207. [PubMed: 3203342]
149. Kouwijzer MLCE, Vaneijck BP, Kroes SJ, Kroon J. J Comput Chem. 1993; 14:1281.
150. Weiner JS, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta SJ, Weiner P. J Am Chem Soc. 1984; 106:765.


Figure 1.
Tetrahydro-2-methoxy-5-(tetrahydro-2H-pyran-2-yloxy)-2H-pyran (A), methyl $\beta$-D-glucopyra-nosyl-( $1 \rightarrow 4$ )- $\beta$-D-glucopyranoside $(\mathrm{B})$, and methyl $a$-D-glucopyranosyl-( $1 \rightarrow 4$ )-$\alpha$-D-glucopyranoside (C).


Figure 2.
B3LYP/6-31++ G(2d,2p) ( $\mathbf{\Delta})$ and GLYCAM06 ( $)$ rotational energy curves about the exocyclic CG-CG bond in hydroxymethylcyclohexane.


Figure 3.
Key exocyclic torsion angles in $\alpha$-neuraminic acid.


Figure 4.
Differences in average molecular mechanical energies from MD simulation data ( 1 ns ) for the ${ }^{4} \mathrm{C}_{1}$ and ${ }^{1} \mathrm{C}_{4}$ chair forms as a function of the $V_{1}$ coefficient for the $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ atomic sequence in $a$-D-mannopyranose (upper), and $a$-D-idopyranose (lower).



Figure 5.
PMF-WHAM analysis for the $\omega_{8}$ (O7-C7-C8-O8) torsion in $\alpha-5-N$-acetyl-neuraminic acid (A). $\omega_{8}$ population sampling over a 20-ns MD trajectory ( $86 \%$ trans and $14 \%$-gauche) $V_{1}=-0.1 \mathrm{kcal} / \mathrm{mol}(\mathrm{B})$.




Figure 6.
Rotational energy curves for 2,3-butanediol, with internal H-bonds (A), without internal Hbonds (B) and (C) 2,3-dimethoxybutane in which internal H-bonds are prohibited by the substitution of hydroxyl protons with methyl groups. B3LYP/6-31++ G(2d,2p) ( $\mathbf{\Delta}$ ); GLYCAM06 (•).



Figure 7.
B3LYP/6-31++ G(2d,2p) ( $\mathbf{\Delta})$, GLYCAM06 ( $\boldsymbol{\bullet}$ ) and GLYCAM_93 ( $\boldsymbol{*})$ rotational energy curves about the CG-OS bond in tetrahydro-2-methoxy-2H-pyran (axial), (A), and tetrahydro-2-methoxy-2H-pyran (equatorial), (B), for the united-atom (GLYCAM06) and all-atom (GLYCAM_93) charge models.




Figure 8.
B3LYP/6-31++G(2d,2p) ( $\mathbf{(})$ and GLYCAM06 ( $)$ ) rotational energy curves about the $\mathrm{N}-\mathrm{C}$ bond in (A) acetamide, (B) N,N-dimethylacetamide, and (C) $N$-methylacetamide.


Figure 9.
Rotational energy curves about the N - CG bond in N -(tetrahydro- 2 H -pyran-3-yl) acetamide (A), B3LYP/6-31++G(2d,2p) ( $\mathbf{(})$, GLYCAM06 ( $)$, H1—CG—N—H ( $\left.V_{1}=0.0 \mathrm{kcal} / \mathrm{mol}\right)$ and $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{C}\left(V_{1}=-0.17 \mathrm{kcal} / \mathrm{mol}\right)$, in vacuum. PMF-WHAM analysis for the CG $-\mathrm{CG}-\mathrm{N}-\mathrm{C}$ sequence in N -(tetrahydro-2H-py-ran-3-yl) acetamide, $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{H}\left(V_{1}\right.$ $=1.0 \mathrm{kcal} / \mathrm{mol})$ and $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{C}\left(V_{1}=-1.0 \mathrm{kcal} / \mathrm{mol}\right)(\mathrm{O})$ and $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{H}\left(V_{1}=\right.$ $0.0 \mathrm{kcal} / \mathrm{mol})$ and $\mathrm{H} 1-\mathrm{CG}-\mathrm{N}-\mathrm{C}\left(V_{1}=-0.17 \mathrm{kcal} / \mathrm{mol}\right)(\boldsymbol{O})$, (B), in explicit solvent. Rotational energy curves about the OS-CG bond in tetrahydro- $2 H$-pyran-3-yl acetate, B3LYP/6-31++G(2d,2p) ( $\mathbf{\Delta})$, GLYCAM06 ( ${ }^{(\boldsymbol{O}), ~(C) . ~}$



Figure 10.
B3LYP/6-31++ G(2d,2p) ( $\mathbf{\Delta})$, and GLYCAM06 ( $\boldsymbol{(})$ rotational energy curves in (A) tetrahydro-2-methoxy-2H-pyran-3-yl acetate and (B) (tetrahydro-2H-pyran-2-yl)methyl acetate.


Figure 11.
QM and GLYCAM06 rotational energy curves about the C-CG bond in 2methylpropanoate MP2/cc-pVDZ ( $\boldsymbol{\bullet}$ ), B3LYP/6-3111G(2d,2p) ( $\mathbf{(})$,GLYCAM06 ( $\boldsymbol{\bullet}$ ).



Figure 12.
B3LYP/6-31++G(2d,2p)( $\mathbf{\Delta})$ and GLYCAM06 ( $\boldsymbol{\bullet}$ ) rotational energy curves about the CGCG bond in 2-methoxybutan-3-ol with internal hydrogen bonds (A) and without internal hydrogen bonds (B).


Figure 13.
B3LYP/6-31++G(2d,2p) ( $\mathbf{\Delta})$ and GLYCAM06 ( $)$ rotational energy curves about the CGCG bond in $N$-ethanolacetamide.


Figure 14.
Population distributions from a 2-ns MD simulation trajectory $V_{1}=3.5 \mathrm{kcal} / \mathrm{mol}$, (A).
Potential of mean force (PMF) and statistical weighted-histogram analysis method (WHAM) for $\phi\left(\mathrm{C}_{x}-\mathrm{C}_{x}-\mathrm{O}_{x-1}-\mathrm{C}_{x-1}\right)$ in methyl $\alpha-5-N$-acetylneuraminic acid, (B).


Figure 15.
The rotational energy curves about the CG-OS bond in (R)-tetrahydro-2-methoxy-2H-pyran-2-carboxylic acid computed with GLYCAM06 ( ) employing the ionized species, and at the B3LYP/6-31++G(2d,2p) level for both the ionized $(\mathbf{\Delta})$ and neutral $(\triangle)$ species.


Figure 16.
Comparison with experimental results of frequencies calculated by NMODE/GLYCAM for a monomer and a 64-unit crystal. The diameter of the circles for the calculated values corresponds to the relative kinetic energy associated with the motion used to determine the assignment, and not to photoabsorption intensity. For the monomer, the kinetic energy is relative to only the one molecule; for the crystal, it corresponds to a motion in a particular molecule, but relative to all 64 molecules. Some frequencies included here might not be observable by IR and/or Raman spectroscopy. Calculated crystal frequencies lower than 40 $\mathrm{cm}^{-1}$ are not included in the figure. In the list of assignment classes, "Exo. O" refers to exocyclic oxygens, oxygen atoms connected to the main ring but that are not part of the ring. $\left({ }^{*}\right)$ These frequencies occurred in a single molecule at the face of the crystal and are not expected to occur frequently in nature. Citations: (A) Wells and Atalla ${ }^{71}$; (B) Dauchez et al. ${ }^{67}$ (C), Gregurick and Kafafi ${ }^{70}$; (D) Kuttel et al. ${ }^{68}$ Please see the text for further discussion of the figure. The details of the individual assignments, as well as the frequencies, eigenvectors, etc., can be obtained by contacting the authors.


Figure 17.
Range of frequencies associated with $\mathrm{C}-\mathrm{O}$ and $\mathrm{C}-\mathrm{C}$ stretches as calculated by NMODE/ GLYCAM for the 64-unit monomer. The figure includes all $\mathrm{C}-\mathrm{O}$ and $\mathrm{C}-\mathrm{C}$ stretches in each molecule, regardless of relative importance to the mode. Circle diameters correspond to the relative kinetic energy associated with the stretching motion within the mode (see also discussion of Fig. 16). The experimental range is taken from Wells and Atalla, ${ }^{71}$ and corresponds to "heavy atom stretches."

## Table 1

Model Compounds Employed in the Development of GLYCAM06.

| Molecular class | Parameter | Training set | Test set |
| :---: | :---: | :---: | :---: |
| Hydrocarbons | HC-CG-CG-HC | Ethane |  |
|  | CG-CG-CG-CG | Butane | 2,3-dimethylbutane, 2-methylbutane, and ethylcyclohexane |
|  | HC-CG-CG-CG | Propane | 2-methylpropane, 2,2-dimethylpropane, and methyl cyclohexar |
| Alcohols | HO-OH-CG-H1 | Methanol | Propan-2-ol |
|  | HO-OH-CG-CG | Ethanol and propan-2-ol |  |
|  | $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{HC}$ | Ethanol and 2-methylpropan-2-ol |  |
|  | CG-CG-CG-OH | Propanol | 1,1,2,2-tetramethylpropanol, 2,2-dimethylpropanol, hydroxym (equatorial) and butanol |
|  | $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OH}$ | 1,2-ethanediol and 1,2-propanediol | 2-methyl-1,2-propanediol, 2,3-butanediol, 2-methyl-2,3-butane dimethyl-2,3-butanediol |
| Ethers | H1-CG-OS-CG | Methoxymethane | Methoxyethane, 2-methoxypropane, and 2-methoxy-2-methylp |
|  | CG-CG-OS-CG | Methoxyethane and 2-methoxypropane | 2-methoxy-2-methylpropane, 2-methoxypropane, 2-tert-butox methoxycyclohexane (axial and equatorial), and 2-isopropoxyp |
|  | CG-CG-CG-OS | Methoxypropane and 1-methoxy-2-methylpropane | 1-methoxy-2,2-dimethylpropane |
|  | OS-CG-CG-OS | 1,2-dimethoxyethane | 1,2-dimethoxypropane, 1,2-dimethoxy-2-methylpropane, 2,3-d dimethoxy-2-methylbutane, and 2,3-dimethoxy-2,3-dimethylb |
|  | OS-CG-OS-CG | 1,1-dimethoxyethane, 2,2-dimethoxypropane and Tetrahydro-2-methoxy- 2 H -pyran (axial and equatorial) | Tetrahydro-2-methoxy-2H-pyran-3-ol (axial and equatorial) an (tetrahydro-2H-pyran-2-yloxy)-2H-pyran |
| Amides | N-CG-CG-CG | N -(2,33-trimethylbutan-2-yl) acetamide and N -propylacetamide | N -tert-pentylacetamide and N -sec-butylacetamide |
|  | $\mathrm{H}-\mathrm{N}-\mathrm{C}-\mathrm{O}$, | Acetamide |  |
|  | HC-CG-C-N |  |  |
|  | H1-CG-N-C/CG | $\mathrm{N}, \mathrm{N}$-dimethylacetamide, and N -isopropylacetamide |  |
|  | HC-CG-CG-N | $N$-tert-butylacetamide |  |
|  | CG-N-C-O, | $\mathrm{N}, \mathrm{N}$-dimethylacetamide | $N$-methylacetamide, $N$-ethylacetamide, |
|  | CG-N-C-CG |  | N -isopropylacetamide and N -tert-butylacetamide |
|  | CG-CG-N-C | N -tert-butylacetamide and N -isopropylacetamide, N -(tetrahydro-2 H -pyran-3-yl) acetamide | N -ethylacetamide and N -(tetrahydro-2-methoxy- 2 H -pyran-3-yl |
| Esters | H1-CG-OS-C | Methyl acetate |  |
|  | OS-C-CG-HC | Methyl acetate |  |
|  | CG-OS-C-O, | Methyl acetate |  |
|  | CG-C-OS-CG |  |  |
|  | CG-CG-OS-C | Ethyl acetate and tetrahydro-2H-pyran-3-yl acetate | tetrahydro-2-methoxy-2H-pyran-3-yl acetate and (tetrahydro-2 acetate |
| Carboxylates | O2-C-CG-CG | 2-methyl propanoate |  |
|  | OH-CG-CG-C | 3-hydroxy propanoate | 2-methyl-3-hydroxy propanoate |
|  | $\mathrm{HC}-\mathrm{CG}-\mathrm{C}-\mathrm{O} 2$ | Acetate |  |
| Ether alcohols | $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OS}$ | Methoxyethanol and 2-methoxypropanol | 2-methoxybutan-3-ol, 2-methoxy-2-methylpropanol, 2-methox and 2-methoxy-2,3-dimethylbutan-3-ol |
| Ether amides | OS-CG-CG-N | $N$-(1,1-dimethyl-2-methoxyethyl) acetamide and N -(1-methoxypropan-2-yl)acetamide | N -(1-methoxy-2-methylpropan-2-yl)acetamide and N -(1-metho |
|  | CG-OS-CG-N | N -(2-methoxyethyl)acetamide and N -(2-methoxypropyl)acetamide |  |
| Molecular class | Parameter | Training set | Test set |


| Molecular class | Parameter | Training set | Test set |
| :--- | :--- | :--- | :--- |
| Alcohol amides | OH-CG-CG-N | $N$-ethanolacetamide | $N$-(2-hydroxypropyl)acetamide, $N$-(2-hydroxy-2-methylpropyl) <br> hydroxy-2,3-dimethylbutan-2-yl)acetamide |
| Ether carboxylates | O2-C-CG-OS | 2-methoxyacetate, 2-methoxypropanoate, 2,2-dimethoxyacetate, <br> 2-methoxy-2-methylpropanoate and $(R)$-tetrahydro-2- <br> methoxy-2H-pyran-2-carboxylic acid (ionized) |  |
|  | OS-CG-CG-C | 3-methoxypropanoate | 2-methyl-3-methoxypropanoate |

## Id!ıssnuew ıOपłn*

Author Manuscrip
Ensemble-Averaged Partial Charge ${ }^{a}$ Sets for the Methyl Glycosides of D-Glcp, D-Manp, and D-Galp.

| Atom | D-GlcpoMe |  | D-ManpOMe |  | D-GalpOMe |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $a$ | $\beta$ | $a$ | $\beta$ | $a$ | $\beta$ |
| CMe | $0.259 \pm 0.01$ | $0.267 \pm 0.01$ | $0.275 \pm 0.01$ | $0.262 \pm 0.01$ | $0.262 \pm 0.01$ | $0.264 \pm 0.01$ |
| OMe | $-0.470 \pm 0.03$ | $-0.454 \pm 0.02$ | $-0.532 \pm 0.03$ | $-0.406 \pm 0.03$ | $-0.478 \pm 0.03$ | $-0.455 \pm 0.03$ |
| C1 | $0.527 \pm 0.08$ | $0.378 \pm 0.08$ | $0.496 \pm 0.07$ | $0.262 \pm 0.05$ | $0.488 \pm 0.09$ | $0.372 \pm 0.08$ |
| C2 | $0.246 \pm 0.09$ | $0.310 \pm 0.10$ | $0.245 \pm 0.07$ | $0.241 \pm 0.09$ | $0.339 \pm 0.10$ | $0.357 \pm 0.10$ |
| O 2 | $-0.713 \pm 0.03$ | $-0.718 \pm 0.02$ | $-0.716 \pm 0.03$ | $-0.664 \pm 0.03$ | $-0.724 \pm 0.02$ | $-0.727 \pm 0.03$ |
| H2O | $0.437 \pm 0.03$ | $0.437 \pm 0.02$ | $0.446 \pm 0.02$ | $0.431 \pm 0.02$ | $0.431 \pm 0.02$ | $0.432 \pm 0.02$ |
| C3 | $0.286 \pm 0.10$ | $0.284 \pm 0.11$ | $0.251 \pm 0.11$ | $0.263 \pm 0.13$ | $0.253 \pm 0.09$ | $0.259 \pm 0.11$ |
| O3 | $-0.699 \pm 0.03$ | $-0.709 \pm 0.03$ | $-0.696 \pm 0.03$ | $-0.692 \pm 0.04$ | $-0.701 \pm 0.02$ | $-0.695 \pm 0.02$ |
| H3O | $0.427 \pm 0.02$ | $0.432 \pm 0.01$ | $0.431 \pm 0.02$ | $0.430 \pm 0.02$ | $0.433 \pm 0.02$ | $0.433 \pm 0.02$ |
| C4 | $0.254 \pm 0.12$ | $0.276 \pm 0.13$ | $0.326 \pm 0.16$ | $0.349 \pm 0.15$ | $0.204 \pm 0.08$ | $0.203 \pm 0.10$ |
| O4 | $-0.710 \pm 0.03$ | $-0.714 \pm 0.03$ | $-0.732 \pm 0.04$ | $-0.746 \pm 0.03$ | $-0.673 \pm 0.04$ | $-0.664 \pm 0.04$ |
| H4O | $0.436 \pm 0.02$ | $0.440 \pm 0.02$ | $0.438 \pm 0.02$ | $0.443 \pm 0.02$ | $0.436 \pm 0.03$ | $0.436 \pm 0.02$ |
| C5 | $0.283 \pm 0.11$ | $0.225 \pm 0.12$ | $0.276 \pm 0.15$ | $0.206 \pm 0.13$ | $0.216 \pm 0.09$ | $0.140 \pm 0.09$ |
| 05 | $-0.574 \pm 0.07$ | $-0.470 \pm 0.06$ | $-0.507 \pm 0.07$ | $-0.393 \pm 0.06$ | $-0.527 \pm 0.07$ | $-0.402 \pm 0.07$ |
| C6 | $0.276 \pm 0.04$ | $0.282 \pm 0.044$ | $0.266 \pm 0.05$ | $0.273 \pm 0.05$ | $0.308 \pm 0.05$ | $0.319 \pm 0.05$ |
| O6 | $-0.682 \pm 0.02$ | $-0.688 \pm 0.02$ | $-0.688 \pm 0.02$ | $-0.680 \pm 0.02$ | $-0.684 \pm 0.03$ | $-0.692 \pm 0.02$ |
| H6O | $0.418 \pm 0.02$ | $0.424 \pm 0.02$ | $0.422 \pm 0.02$ | $0.419 \pm 0.02$ | $0.418 \pm 0.02$ | $0.419 \pm 0.02$ |

${ }^{a}$ From 200 snapshots evenly extracted from a 50 -ns MD simulation.
łd!ıכsnuew roułn*




Author Manuscript

Author Manuscript

Author Manuscript


[^1]Author Manuscript

Table 4

| Bond | $K_{r}{ }^{\text {b }}$ | $r_{\text {eq }}{ }^{\text {c }}$ | Bond | $K_{\mathrm{r}}$ | $r_{\text {eq }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| C-OS ${ }^{\text {a }}$ | 450.0 | 1.323 | C-HC ${ }^{\text {a }}$ | 331.0 | 1.090 |
| CG-CG | 310.0 | 1.520 | C-H1 | 410.0 | 1.092 |
| CG-HC ${ }^{\text {a }}$ | 340.0 | 1.090 | OH-HO ${ }^{a}$ | 553.0 | 0.960 |
| CG-H1 ${ }^{a}$ | 340.0 | 1.090 | $\mathrm{N}-\mathrm{H}^{a}$ | 434.0 | 1.010 |
| CG-H2 ${ }^{\text {a }}$ | 340.0 | 1.090 | $\mathrm{CY}-\mathrm{OH}{ }^{\text {a }}$ | 320.0 | 1.410 |
| CG-OH | 320.0 | 1.430 | CG-OS | 285.0 | 1.460 |
| CG-C | 220.0 | 1.530 | CG-OY | 320.0 | 1.410 |
| CG-N | 337.0 | 1.450 | CY-CG | 310.0 | 1.520 |
| $\mathrm{C}-\mathrm{O}^{\text {a }}$ | 570.0 | 1.229 | $\mathrm{CY}-\mathrm{OS}{ }^{a}$ | 320.0 | 1.410 |
| $\mathrm{C-N}{ }^{\text {a }}$ | 490.0 | 1.335 | CY-OY ${ }^{\text {a }}$ | 320.0 | 1.410 |
| $\mathrm{C-O} 2^{\text {a }}$ | 656.0 | 1.250 | CY-C | 220.0 | 1.530 |
| Angle | $K_{\theta}{ }^{\text {d }}$ | $\theta_{\text {cq }}{ }^{\text {e }}$ | Angle | $K_{\theta}$ | $\theta_{\text {eq }}$ |
| CG-OS-C | 60.0 | 117.0 | OS-CG-N | 106.9 | 107.9 |
| OS-C-O ${ }^{\text {a }}$ | 80.0 | 125.0 | CG-CG-N ${ }^{a}$ | 80.0 | 109.7 |
| OS-C-CG | 95.0 | 110.8 | H1-CG-N ${ }^{\text {a }}$ | 50.0 | 109.5 |
| HC-CG-HC | 40.0 | 109.5 | H2-CG-N ${ }^{\text {a }}$ | 50.0 | 109.5 |
| H1-CG-H1 | 45.0 | 109.5 | C-CG-CG ${ }^{a}$ | 63.0 | 111.1 |
| H2-CG-H2 | 45.0 | 109.5 | C-CG-HC ${ }^{\text {a }}$ | 50.0 | 109.5 |
| CG-CG-HC | 45.0 | 112.0 | C-CG-H1 ${ }^{\text {a }}$ | 50.0 | 109.5 |
| CG-CG-H1 | 45.0 | 111.0 | C-CG-H2 ${ }^{\text {a }}$ | 50.0 | 109.5 |
| CG-CG-H2 | 45.0 | 111.0 | $\mathrm{C-N-CG}{ }^{\text {a }}$ | 50.0 | 120.0 |
| CG-CG-CG | 45.0 | 113.5 | CG-N-CG ${ }^{\text {a }}$ | 50.0 | 118.0 |
| OH-CG-CG | 70.0 | 107.5 | H-N-CG ${ }^{\text {a }}$ | 30.0 | 118.0 |
| H1-CG-OH | 60.0 | 110.0 | C-N-H ${ }^{\text {a }}$ | 30.0 | 122.0 |
| H2-CG-OH | 60.0 | 110.0 | $\mathrm{H}-\mathrm{N}-\mathrm{H}^{\text {a }}$ | 35.0 | 120.0 |

Author Manuscript
Author Manuscript

| nuew douın* |  |  | łd!ıosnuew ıOuın* |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Bond | $K_{r}{ }^{\text {b }}$ | $r_{\text {eq }}{ }^{\text {c }}$ | Bond | $K_{\text {r }}$ | $r_{\text {eq }}$ |
| CG-OH-HO | 55.0 | 109.5 | H2-C-N | 55.0 | 112.4 |
| Су-ОН-Но | 55.0 | 109.5 | H2-C-O | 55.0 | 112.2 |
| CG-OS-CG | 50.0 | 111.6 | OS-CG-C | 63.0 | 112.36 |
| CT-OS-CG | 50.0 | 111.6 | $\mathrm{O}^{2}-\mathrm{C}-\mathrm{OH}^{\text {a }}$ | 80.0 | 126.0 |
| os-CG-CG | 70.0 | 108.5 | OS-CY-CG | 70.0 | 108.5 |
| OY-CG-CG | 70.0 | 108.5 | OH-CY-CG | 70.0 | 108.5 |
| H1-CG-OS | 60.0 | 110.0 | OY-CG-CG | 70.0 | 108.5 |
| H1-CG-OY | 60.0 | 110.0 | OH-CY-C | ${ }_{63.0}$ | 112.4 |
| H2-CG-OS | 60.0 | 110.0 | OS-CY-C | ${ }^{63.0}$ | 112.4 |
| H1-C-O | 60.0 | 110.0 | OY-CY-C | 63.0 | 112.4 |
| H1-C-N | 55.0 | 112.4 | OY-CY-OH | 100.0 | 112.0 |
| OS-CG-OS | 100.0 | 112.0 | OY-CY-OS | 100.0 | 112.0 |
| OH-CG-OS | 100.0 | 112.0 | CG-CY-C ${ }^{\text {a }}$ | 63.0 | 111.1 |
| O2-C-02 | 80.0 | 126.0 | CG-OS-CY | 50.0 | 111.6 |
| O2-C-CG | 70.0 | 115.0 | CG-OY-CY | 50.0 | 111.6 |
| CG-C-N ${ }^{\text {a }}$ | 70.0 | 116.6 | O2-C-CY | 70.0 | 115.0 |
| CG-C-O ${ }^{\text {a }}$ | 80.0 | 120.4 | HC-CG-CY | 45.0 | 112.6 |
| $\mathrm{N}-\mathrm{C-} \mathrm{O}^{\text {a }}$ | 80.0 | 122.9 | CG-CG-CY | 45.0 | 113.5 |
| N-CG-HC ${ }^{\text {a }}$ | 50.0 | 109.5 |  |  |  |
| Torsion | $\mathrm{V}_{n} \mathrm{f}_{2}$ | $r^{g, n^{h}}$ | Torsion | $\mathrm{v}_{n} / 2$ | $\gamma, n$ |
| OS-CG-CG-C | -0.10 | 0.0, 3 | Os-CG-OS-CT | 0.37 | 0.0, 3 |
|  | 0.10 | 0.0, 2 |  | 1.27 | 0.0,2 |
|  | -1.00 | 0.0, 1 |  | 0.30 | 0.0, 1 |
| OH-CG-CG-C | 0.10 | 0.0, 3 | OH-CG-OS-CG | 0.37 | 0.0, 3 |
|  | 0.20 | 0.0, 2 |  | 1.27 | 0.0,2 |
|  | $-2.50$ | 0.0, 1 |  | 0.30 | 0.0, 1 |
| CG-CG-OS-C | -0.04 | 0.0, 3 | H2-CG-OS-CG | 0.10 | 0.0, 3 |
|  | 0.47 | 0.0, 1 |  | 0.60 | 0.0,2 |
| CG-OS-C-O | -3.20 | 0.0, 2 | H2-CG-OS-CT | 0.10 | 0.0, 3 |
| CG-C-OS-CC | 3.00 | 0.0, 2 |  | 0.60 |  |

Author Manuscript

| UueW dOYłn |  |  | ıd!ıssnuew ı0чın* |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Bond | $K_{\mathrm{r}}{ }^{b}$ | $r_{\mathrm{eq}}{ }^{c}$ | Bond | $K_{\mathrm{r}}$ | $r_{\text {eq }}$ |
| H1-CG-OS-C | 0.00 | 0.0, 3 | HC-CG-CG-OS | 0.05 | 0.0, 3 |
| OS-C-CG-HC | 0.00 | 0.0, 3 | HC-CG-CG-OY | 0.05 | 0.0, 3 |
| HC-CG-CG-HC | 0.13 | 0.0, 3 | H1-CG-CG-OS | 0.05 | 0.0, 3 |
| H1-CG-CG-HC | 0.17 | 0.0, 3 | H2-CG-CG-OS | 0.05 | 0.0, 3 |
| H2-CG-CG-HC | 0.17 | 0.0, 3 | $\mathrm{H}-\mathrm{N}-\mathrm{C}-\mathrm{O}$ | -2.50 | 0.0, 2 |
| H1-CG-CG-H1 | 0.17 | 0.0, 3 |  | 2.00 | 0.0, 1 |
| H1-CG-CG-H2 | 0.17 | 0.0, 3 | $\mathrm{H}-\mathrm{N}-\mathrm{C}-\mathrm{H} 1$ | $-2.50$ | 0.0, 2 |
| HC-CG-CG-CG | 0.10 | 0.0, 3 | H-N-C-CG | 0.0 | 0.0, 1 |
| H1-CG-CG-CG | 0.15 | 0.0, 3 | HC-CG-C-O | 0.0 | 0.0, 3 |
| H2-CG-CG-CG | 0.15 | 0.0, 3 | HC-CG-C-N | 0.0 | 0.0, 3 |
| HO-OH-CG-H1 | 0.18 | 0.0, 3 | CG-OS-CG-N | -0.90 | 0.0, 1 |
| HO-OH-CG-H2 | 0.18 | 0.0, 3 | OS-CG-CG-N | -1.30 | 0.0, 1 |
| $\mathrm{HO}-\mathrm{OH}-\mathrm{CY}-\mathrm{OY}$ | 0.18 | 0.0, 3 | OY-CG-CG-N | -1.30 | 0.0, 1 |
| HO-OH-CY-C | 0.18 | 0.0, 3 | OH-CG-CG-N | -1.50 | 0.0, 1 |
| $\mathrm{HO}-\mathrm{OH}-\mathrm{CY}-\mathrm{CG}$ | 0.18 | 0.0, 3 | CG-CG-CG-N | 0.40 | 0.0, 1 |
| $\mathrm{HO}-\mathrm{OH}-\mathrm{CG}-\mathrm{CG}$ | 0.18 | 0.0, 3 | HC-CG-CG-N | 0.10 | 0.0, 3 |
| CG-CG-CG-CG | 0.45 | 0.0, 1 | H1-CG-CG-N | 0.10 | 0.0, 3 |
| $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{HC}$ | 0.05 | 0.0, 3 | H2-CG-CG-N | 0.10 | 0.0, 3 |
| $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{H} 1$ | 0.05 | 0.0, 3 | N-CG-CG-N | 0.00 | 0.0, 1 |
| OH-CG-CG-H2 | 0.05 | 0.0, 3 | HC-CG-CG-C | 0.10 | 0.0, 3 |
| $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{CG}$ | 0.10 | 0.0, 3 | H1-CG-CG-C | 0.10 | 0.0, 3 |
| OS-CG-CG-CG | -0.27 | 0.0, 1 | CG-CG-CG-C | 0.45 | 0.0, 1 |
| OS-CG-CG-CY | -0.27 | 0.0, 1 | CG-CG-N-C | 0.00 | 0.0, 1 |
| OY-CG-CG-CG | -0.27 | 0.0, 1 | OS-CG-N-C | 2.03 | 0.0, 2 |
| H1-CG-OS-CG | 0.27 | 0.0, 3 |  | 1.00 | 0.0, 1 |
| CG-OS-CG-CG | 0.16 | 0.0, 3 | OS-CG-N-H | -0.43 | 0.0, 2 |
| CT-OS-CG-CG | 0.16 | 0.0, 3 |  | 1.52 | 0.0, 1 |
| OH-CG-CG-OH | 0.55 | 0.0, 3 | CG-CG-N-H | 0.10 | 0.0, 3 |
|  | 0.95 | 0.0,2 | H1-CG-N-CG | 0.00 | 0.0, 3 |
|  | -0.10 | 0.0, 1 | H2-CG-N-CG | 0.00 | 0.0, 3 |
| $\mathrm{OH}-\mathrm{CG}-\mathrm{CG}-\mathrm{OS}$ | 0.25 | 0.0, 2 | H1-CG-N-H | 1.00 | 0.0, 1 |

Author Manuscript
Author Manuscript

| UeW douln |  |  | łd!ı0snuew ıOપın* |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Bond | $K_{r}{ }^{\text {b }}$ | $r_{\text {eq }}{ }^{\text {c }}$ | Bond | $K_{\text {r }}$ | $r_{\text {eq }}$ |
|  | ${ }^{-1.10}$ | 0.0, 1 | H2-CG-N-H | 1.00 | 0.0, 1 |
| OH-CG-CG-OY | 0.25 | 0.0,2 | H1-CG-N-C | $-1.00$ | 0.0, 1 |
|  | -1.10 | $0.0,1$ | H2-CG-N-C | ${ }^{-1.00}$ | 0.0, 1 |
| H1-CG-CG-OY | 0.05 | 0.0,3 | CG-N-C-O | $-2.80$ | 0.0,2 |
| OS-CG-CG-OS | 0.82 | 0.0,2 | CG-N-C-CG | $-2.70$ | 0.0,2 |
| OS-CG-CG-OY | 0.82 | 0.0,2 | HC-CG-C-O2 | 0.00 | 0.0, 1 |
| OS-CG-OH-HO | 0.18 | 0.0,3 | CG-CG-C-O2 | $-1.50$ | 0.0,2 |
| OS-CG-OS-CG | 0.37 | 0.0,3 | H1-CG-C-O2 | 0.00 | 0.0, 1 |
|  | 1.27 | 0.0,2 | H2-CG-C-O2 | 0.00 | 0.0, 1 |
|  | 0.30 | $0.0,1$ | OS-CG-C-O2 | $-2.25$ | 0.0,2 |
|  |  |  | CG-OS-CG-C | 0.00 | 0.0,2 |
| Torsion | $\mathrm{v}_{1} / 2$ | $\gamma, n$ | Torsion | $\mathrm{v}_{n} / 2$ | \%,n |
| H1-CT-OS-CG | 0.27 | 0.0,3 | OS-CY-CG-HC | 0.05 | 0.0, 3 |
| CG-OS-CT-CT | 0.16 | 0.0,3 | OY-CY-CG-HC | 0.05 | 0.0, 3 |
| CT-os-CG-CG | 0.16 | 0.0,3 | CG-OS-CY-CG | 0.16 | 0.0, 3 |
| OS-CG-OS-CT | 0.37 | 0.0,3 | CG-OY-CY-CG | 0.16 | 0.0, 3 |
|  | 1.27 | 0.0,2 | $\mathrm{OH}-\mathrm{CY}-\mathrm{C}-\mathrm{O} 2$ | -2.25 | 0.0,2 |
|  | 0.30 | 0.0, 1 | OS-CY-C-O2 | $-2.25$ | 0.0, 2 |
| CG-N-C-CT | 0.00 | 0.0,2 | OY-CY-C-O2 | $-2.25$ | 0.0,2 |
| OY-CY-OS-CG | 0.37 | 0.0,3 | CG-CY-C-O2 | -1.50 | 0.0,2 |
|  | 1.27 | 0.0,2 | CG-OS-CY-C | 3.50 | 0.0, 1 |
|  | 0.30 | 0.0, 1 | CG-OY-CY-C | 0.00 | 0.0, 1 |
| OH-CY-OY-CG | 0.37 | 0.0,3 | HC-CG-CY-C | 0.10 | 0.0, 3 |
|  | 1.27 | 0.0,2 | CG-CG-CY-C | 0.45 | 0.0, 3 |
|  | 0.30 | $0.0,1$ | CY-OS-CG-CG | 0.16 | 0.0, 3 |
| OS-CY-OY-CG | 0.37 | 0.0,3 | CY-OY-CG-CG | 0.16 | 0.0,3 |
|  | 1.27 | 0.0,2 | CY-OS-CG-H1 | 0.27 | 0.0, 3 |
|  | 0.30 | $0.0,1$ | CY-OY-CG-H1 | 0.27 | 0.0, 3 |
| OH-CY-CG-CG | -0.27 | $0.0,1$ | CY-CG-CG-CG | 0.45 | 0.0, 1 |
| OS-CY-CG-CG | -0.27 | $0.0,1$ | CY-CG-CG-HC | 0.10 | 0.0, 3 |
| OY-CY-CG-CG | -0.27 | 0.0, 1 | CY-CG-CG-H1 | 0.15 | 0.0, |


łd!ıОsnuew ıOપłn $\forall$

| MueW 1OYł |  |  | ıd!̣ısnuew doułn ${ }^{\text {a }}$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Bond | $K_{\mathrm{r}}{ }^{\text {b }}$ | $r_{\text {eq }}{ }^{\text {c }}$ | Bond | $K_{\text {r }}$ | $r_{\text {eq }}$ |
| OH-CY-CG-HC | 0.05 | 0.0, 3 | CY-CG-CG-OH | 0.10 | 0.0, 3 |
| Improper torsions |  |  |  |  |  |
| X-X-N-H | 1.0 | 180.0,2 |  |  |  |
| $\mathrm{X}-\mathrm{X}-\mathrm{C}-\mathrm{O}$ | 10.5 | 180.0,2 |  |  |  |
| $\mathrm{X}-\mathrm{X}-\mathrm{N}-\mathrm{CG}$ | 1.5 | 180.0,2 |  |  |  |
| $\mathrm{X}-\mathrm{O} 2-\mathrm{C}-\mathrm{O} 2$ | 10.5 | 180.0,2 |  |  |  |
| $\mathrm{X}-\mathrm{O} 2-\mathrm{C}-\mathrm{OH}$ | 10.5 | 180.0,2 |  |  |  |
| van der Waals |  |  |  |  |  |
|  | $R^{* i}$ | $8^{j}$ |  |  |  |
| H | 0.6000 | 0.0157 |  |  |  |
| но | 0.0000 | 0.0000 |  |  |  |
| HC | 1.4870 | 0.0157 |  |  |  |
| H1 | 1.3870 | 0.0157 |  |  |  |
| H2 | 1.2870 | 0.0157 |  |  |  |
| O | 1.6612 | 0.2100 |  |  |  |
| O 2 | 1.6612 | 0.2100 |  |  |  |
| OH | 1.7210 | 0.2104 |  |  |  |
| OS | 1.6837 | 0.1700 |  |  |  |
| OY | 1.6837 | 0.1700 |  |  |  |
| CG | 1.9080 | 0.1094 |  |  |  |
| CY | 1.9080 | 0.1094 |  |  |  |
| C | 1.9080 | 0.0860 |  |  |  |

[^2]d!ıəsnuew ıOપłn $\forall$
Average ${ }^{a}$ Overall Errors in the GLYCAM06 Torsional Energies Relative to B3LYP/6-31++ G(2d,2p)//HF/6-31G* Values for the Molecules and Specific Atomic Sequences in the Training Set.

| Atomic sequence | A | B | Relative error |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | $\left\langle\text { Error }{ }^{b}\right\rangle_{\text {curve }}$ | \% of maximum barrier ${ }^{c}$ | $\left\langle\text { Error }^{d}\right\rangle_{\text {minima }}$ |
| Hydrocarbons (Overall) |  |  | 0.15 | 2.0 | 0.11 |
| A-CG-CG-B | HC, CG | HC, CG | 0.15 | 2.0 |  |
| Alcohols (Overall) |  |  | 0.45 | 5.0 | 0.28 |
| A-CG-CG-B | OH, H1 | OH, H1, HC, CG | 0.50 | 6.0 |  |
| A-OH-CG-B | HO | H1, CG | 0.09 | 6.0 |  |
| Ethers (Overall) |  |  | 0.56 | 4.0 | 0.35 |
| A-CG-CG-B | OS, H1 | OS, H1, HC, CG | 0.30 | 2.0 |  |
| A-OS/OY-CG/CY-B | CG | OS, H1, H2 | 0.65 | 6.0 |  |
| Amides (Overall) |  |  | 0.64 | 4.0 | 0.38 |
| A-N-C-B | H, CG | CG, O | 1.14 | 7.0 |  |
| A-CG-N-B | H1, CG | H, C | 0.18 | 4.0 |  |
| A-CG-C-B | HC, CG | $\mathrm{O}, \mathrm{N}$ | 0.63 | 7.0 |  |
| A-CG-CG-B | N, H1 | HC, H1, CG, N | 0.41 | 4.0 |  |
| Esters (Overall) |  |  | 0.44 | 3.0 | 0.23 |
| A-CG-C-B | HC | OS | 0.02 | 8.0 |  |
| A-CG-OS-B | H1,CG | C | 0.46 | 4.0 |  |
| A-OS-C-B | CG | CG, O | 0.40 | 3.0 |  |
| Carboxylates (Overall) |  |  | 0.72 | 38.0 | 0.60 |
| $\mathrm{A}-\mathrm{CG} / \mathrm{CY}^{e}-\mathrm{C}-\mathrm{B}$ | HC, CG | O2 | 0.14 | 7.0 |  |
| Ether alcohols (Overall) |  |  | 0.61 | 6.0 | 0.59 |
| A-CG-CG-B | OS, H1 | OH, H1 | 0.61 | 6.0 |  |
| Ether amides (Overall) |  |  | 1.12 | 9.0 | 1.10 |
| A-CG-N-B | OS, H2 | H, C | 1.93 | 15.0 |  |
| A-OS-CG-B | CG | N, H2 | 0.64 | 9.0 |  |
| A-CG-CG-B | N, H1 | OS, H1 | 0.37 | 6.0 |  |
| Alcohol amides (Overall) |  |  | 0.49 | 5.0 | 0.41 |

əd!ıכsnuew ıOчłn $\forall$


Table 6
$\omega_{8}$ (O7-C7-C8-O8) Trans and Gauche Rotamer Populations From a 10-ns Explicit Solvent MD Simulation of $\alpha$-Neu5Ac as a Function of the $V_{1}$ Coefficient.

| $\boldsymbol{V}_{\mathbf{1}}(\mathbf{k c a l} / \mathbf{m o l})$ | Trans | Gauche |
| :---: | :---: | :---: |
| -1.0 | 30 | 70 |
| -0.8 | 59 | 41 |
| -0.6 | 32 | 68 |
| -0.4 | 41 | 59 |
| -0.2 | 88 | 12 |
| 0.0 | 84 | 16 |
| 0.2 | 67 | 33 |
| 0.4 | 83 | 17 |
| 0.6 | 86 | 14 |
| 0.8 | 88 | 12 |
| 1.0 | 98 | 2 |

Scalar ${ }^{3} J$-Couplings (Hz) and Exocyclic Hydroxymethyl Rotamer Populations Computed from MD Simulation (50 ns) Versus Experimental Data for Methyl $a$-D-Glc $p$ and $a$-D-Galp.

|  | ${ }^{3} J_{\mathrm{H} 5, \mathrm{H} G \mathrm{R}}$ | ${ }^{3} \mathrm{~J}_{\mathrm{H} 5, \mathrm{HGS}}$ | $\mathrm{gg}(\%)$ | $\mathrm{gt}(\%)$ | $\operatorname{tg}(\%)$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| Methyl $a$-D-Glc $p$ |  |  |  |  |  |


| Experiment $^{a}$ | 5.49 | 2.39 | 53 | 47 | 0 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Experiment $^{b}$ | 5.4 | 2.3 | 57 | 38 | 5 |
| GLYCAM06 | $5.4 \pm 1.7$ | $2.9 \pm 2.0$ | 62 | 36 | 2 |
| Methyl $a$-D-Galp |  |  |  |  |  |
| Experiment $^{a}$ | 7.8 | 6.0 | 14 | 47 | 39 |
| Experiment $^{b}$ | 8.6 | 3.7 | 16 | 75 | 9 |
| Experiment $^{c}$ | 8.3 | 4.0 | 21 | 61 | $18^{d}$ |
|  |  |  | 13 | 70 | $17^{e}$ |
|  |  |  | 15 | 69 | $16^{f}$ |
| GLYCAM06 | $7.9 \pm 1.6$ | $3.7 \pm 1.8$ | 8 | 75 | 18 |


| GLYCAM06 | $.6 \quad 1.8 \quad-8 \quad 1$ |
| :--- | :--- | :--- | :--- |
| $a_{\text {Ref } 142}$ |  |

${ }^{\text {Ref. } 142 .}$
$b_{\text {Ref. } 144 .}$
${ }^{c}$ Ref. 143.
${ }^{d}$ Employing equations A (Table 3), with the ${ }^{3} J_{\mathrm{HH}}$ values derived from a Karplus equation that includes the effects of the electronegativities, and orientations of $a$-substituents).
${ }^{e}$ Employing equations B1 (Table 3), with the ${ }^{3} J_{\mathrm{HH}}$ values derived from a Karplus equation that includes the effects of the electronegativities, and orientations of $a$-substituents).
$f_{\text {Employing equations }}$ (Table 3), with the ${ }^{3} J_{\mathrm{HH}}$ values derived from a Karplus equation that includes the effects of the electronegativities, and orientations of $a$-substituents).
дd!ısnuew roułn*
Distances and Dihedral Angles Between the Ring Hydrogen Atoms Computed From MD Simulation ( 50 ns ) Versus Experimental Data for Methyl $a$-DGal $p, a$-D-Glc $p$ and $a$-D-Man $p$.

|  | $a$-D-GalpOMe |  | $a$ - D-GlcpoMe |  | $a$ - D-ManpOMe |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Experiment ${ }^{\boldsymbol{a}}$ | GLYCAM06 | Experiment ${ }^{\text {b }}$ | GLYCAM06 | Experiment ${ }^{b}$ | GLYCAM06 |
| H1-H2 | 2.396 | $2.40 \pm 0.1$ | 2.473 | $2.41 \pm 0.1$ | 2.562 | $2.52 \pm 0.1$ |
| H1-H3 | 3.827 | $3.83 \pm 0.1$ | 3.795 | $3.86 \pm 0.1$ | 3.817 | $3.79 \pm 0.1$ |
| H1-H4 | 4.874 | $4.95 \pm 0.1$ | 4.065 | $4.23 \pm 0.1$ | 4.109 | $4.21 \pm 0.1$ |
| H1-H5 | 3.605 | $3.75 \pm 0.1$ | 3.634 | $3.74 \pm 0.1$ | 3.639 | $3.78 \pm 0.1$ |
| H2-H3 | 3.053 | $3.05 \pm 0.1$ | 3.046 | $3.05 \pm 0.1$ | 2.440 | $2.42 \pm 0.1$ |
| H2-H4 | 3.792 | $3.83 \pm 0.1$ | 2.500 | $2.78 \pm 0.2$ | 3.829 | $3.86 \pm 0.1$ |
| H2-H5 | 3.907 | $3.99 \pm 0.1$ | 3.919 | 4.0060 .1 | 4.057 | 4.1760 .1 |
| H3-H4 | 2.436 | 2.4560 .1 | 3.033 | $3.05 \pm 0.1$ | 3.053 | $3.05 \pm 0.1$ |
| H3-H5 | 2.485 | $2.53 \pm 0.2$ | 2.607 | $2.76 \pm 0.2$ | 2.592 | $2.66 \pm 0.2$ |
| H4-H5 | 2.457 | $2.45 \pm 0.1$ | 3.050 | $3.04 \pm 0.1$ | 3.070 | $3.05 \pm 0.1$ |
| H1C1C2H2 | 50.49 | $51.07 \pm 7.5$ | 56.86 | $54.02 \pm 7.7$ | -69.03 | $-67.21 \pm 6.9$ |
| H 2 C 2 C 3 H 3 | -176.08 | $-175.74 \pm 7.3$ | -178.44 | $-171.18 \pm 7.9$ | -53.69 | $-54.98 \pm 6.9$ |
| H3C3C4H4 | 53.97 | $55.99 \pm 6.9$ | 174.53 | $166.95 \pm 8.4$ | 172.41 | $172.99 \pm 7.7$ |
| H4C4C5H5 | -59.23 | $-58.12 \pm 6.8$ | -175.73 | $-171.72 \pm 8.4$ | -173.52 | $-171.59 \pm 8.3$ |

[^3]Id!̣こsnuew ıOपłn $\forall$
Author Manuscript
Author Manuscript
Author Manuscript

| Spins | $\alpha$-D-GlcpoMe | $\beta$-D-Glcp OMe | $a$-D-GalpOMe | $\beta$-D-GalpOMe | $\alpha$-D-ManpOMe | $\beta$-D-ManpOMe |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| ${ }^{3} J_{\mathrm{HIH} 2}$ | $3.8 / 3.5 \pm 0.8$ | $8.0 / 9.9 \pm 0.5$ | $4.0 / 3.8 \pm 0.8$ | 7.9/9.8 $\pm 0.9$ | $1.8 / 2.2 \pm 0.5$ | $0.9 / 3.0 \pm 0.7$ |
| ${ }^{3} J_{\mathrm{HIH} 2}$ | $9.8 / 9.9 \pm 0.4$ | $9.4 / 9.7 \pm 0.6$ | $10.3 / 10.1 \pm 0.2$ | $9.9 / 10.0 \pm 0.4$ | $3.5 / 3.4 \pm 0.8$ | $3.2 / 3.2 \pm 0.8$ |
| ${ }^{33} J_{\mathrm{H} 1 \mathrm{H} 2}$ | $9.1 / 9.7 \pm 0.6$ | $9.2 / 9.7 \pm 0.6$ | $3.4 / 3.3 \pm 0.7$ | $3.4 / 3.2 \pm 0.8$ | $9.5 / 10.0 \pm 0.3$ | $9.6 / 10.0 \pm 0.4$ |
| ${ }^{3} J_{\mathrm{HIH} 2}$ | 10.1/9.9 $\pm 0.4$ | $9.7 / 10.0 \pm 0.4$ | $1.0 / 3.0 \pm 0.7$ | $1.1 / 2.9 \pm 0.7$ | - | $9.7 / 9.9 \pm 0.6$ |

Table 10
Primary Assignments for Monomer Frequencies Calculated by NMODE/GLYCAM.

| Frequency, $\mathrm{cm}^{-1}$ | Primary assignment |
| :---: | :---: |
| 66 | T[O6-C6-H61(1.000)] |
| 91 | T[O6-C6-C5(1.000)] |
| 107 | $\mathrm{U}(\mathrm{ETO})[\mathrm{O} 3(1.064), \mathrm{O} 4(0.904), \mathrm{O} 2(0.568),[\& \mathrm{~B}, \mathrm{O} 3-\mathrm{C} 3-\mathrm{H} 3(1.000)]]$ |
| 149 | $\mathrm{U}(\mathrm{ETO})[\mathrm{O} 1(0.976),[\& T, \mathrm{O} 1-\mathrm{C} 1-\mathrm{C} 2(1.000)]]$ |
| 214 | T[H4O-O4-C4(1.000)] |
| 231 | T[H4O-O4-C4(1.000), H6O-O6-C6(0.498),C6-O6-H6O(0.427), C6-C5-C4(0.422)] |
| 253 | $\mathrm{U}(\mathrm{ECO})[\mathrm{O} 1(1.027),[\& \mathrm{~S}, \mathrm{O} 1-\mathrm{HO} 1(1.000)][\& \mathrm{~B}, \mathrm{O} 1-\mathrm{C} 1-\mathrm{C} 2(1.000)]]$ |
| 272 | T[H6O-O6-C6(1.000)] |
| 282 | T[H2O-O2-C2(1.000)] |
| 298 | $\mathrm{U}(\mathrm{ECO})[\mathrm{O} 3(0.950), \mathrm{O} 4(0.907), \mathrm{O} 2(0.377),[\& \mathrm{~B}, \mathrm{O} 4-\mathrm{C} 4-\mathrm{C} 3(1.000)]]$ |
| 303 | T[HO1-O1-C1(1.000)] |
| 313 | T[HO1-O1-C1(1.000)] |
| 353 | $\mathrm{U}(\mathrm{ERBO})[\mathrm{O} 3(1.076),[\& \mathrm{~S}, \mathrm{O} 3-\mathrm{C} 3(1.000)]]$ |
| 369 | $\mathrm{U}(\mathrm{ERBO})[\mathrm{O} 2(0.982),[\& \mathrm{~B}, \mathrm{O} 1-\mathrm{C} 1-\mathrm{H1}(1.000)]]$ |
| 382 | T[H2O-O2-C2(1.000)] |
| 410 | T[H3O-O3-C3(1.000)] |
| 422 | T[H3O-O3-C3(1.000), $\mathrm{HO1} 1-\mathrm{O} 1-\mathrm{C} 1(0.742)]$ |
| 458 | T[H3O-O3-C3(1.000)] |
| 481 | $\mathrm{U}(\mathrm{RB})[\mathrm{O} 5(0.833),[\& \mathrm{~B}, \mathrm{O} 5-\mathrm{C} 1-\mathrm{C} 2(1.000)]]$ |
| 521 | U(ROPD)[O5(0.735),C4(0.668),[\&T,O5-C5-C4(1.000)]] |
| 570 | $\mathrm{U}(\mathrm{RC})[\mathrm{O}(0.657), \mathrm{C} 3(0.505),[\& \mathrm{~S}, \mathrm{O}-\mathrm{C} 1(1.000)]]$ |
| 584 | $\mathrm{U}(\mathrm{RC})[\mathrm{C} 4(0.974), \mathrm{C} 5(0.734),[\& B, \mathrm{C} 4-\mathrm{C} 3-\mathrm{O} 3(1.000)]]$ |
| 661 | $\mathrm{U}(\mathrm{RB})[\mathrm{Cl}(0.929), \mathrm{O} 5(0.386),[\& \mathrm{~T}, \mathrm{C} 1-\mathrm{O} 1-\mathrm{HO} 1(1.000)]]$ |
| 835 | ST[H61(1.000),H62(0.620)] |
| 881 | $\mathrm{U}(\mathrm{RB})[\mathrm{O} 5(0.335),[\& \mathrm{~S}, \mathrm{C} 1-\mathrm{H} 1(1.000)]]$ |
| 951 | T[C6-C5-O5(1.000)] |
| 977 | U(ROPD)[C4(0.870),[\&T,C4-C3-O3(1.000)]] |
| 983 | $\mathrm{U}(\mathrm{RB})[\mathrm{C} 1(0.616), \mathrm{C} 4(0.356),[\& T, \mathrm{C} 3-\mathrm{C} 4-\mathrm{O} 4(1.000)]]$ |
| 995 | $\mathrm{U}(\mathrm{RC})[\mathrm{Cl}(0.770), \mathrm{O5}(0.755),[\& \mathrm{~S}, \mathrm{C} 1-\mathrm{O}(1.000)]]$ |
| 1028 | U(ROPD)[C5(1.005),C2(0.388),[\&S,C5-H5(1.000)]] |
| 1054 | $\mathrm{U}(\mathrm{RC})[\mathrm{C} 5(0.316),[\& T, \mathrm{C} 3-\mathrm{C} 4-\mathrm{H} 4(1.000)]]$ |
| 1073 | $\mathrm{U}(\mathrm{ECH})[\mathrm{H} 3(0.989),[\& T, \mathrm{H} 3-\mathrm{C} 3-\mathrm{O} 3(1.000)]]$ |
| 1088 | U(RB)[C2(0.768),C4(0.667),[\&S,C2-O2(1.000)]] |
| 1103 | $\mathrm{U}(\mathrm{ECH})[\mathrm{H} 4(0.953), \mathrm{H} 5(0.877),[\& T, H 4-\mathrm{C} 4-\mathrm{O} 4(1.000)]]$ |
| 1130 | U(ETH)[H5(0.837),H4(0.529),[\&T,H5-C5-C4(1.000)]] |
| 1135 | $\mathrm{U}(\mathrm{ECH})[\mathrm{H} 2(0.964),[\& \mathrm{~T}, \mathrm{H} 2-\mathrm{C} 2-\mathrm{O} 2(1.000)]]$ |
| 1236 | $\mathrm{B}[\mathrm{HO1-O1-C1(1.000)]}$ |
| 1274 | B[H3O-O3-C3(1.000)] |
| 1285 | B[H2O-O2-C2(1.000)] |


| Frequency, $\mathrm{cm}^{-1}$ | Primary assignment |
| :---: | :---: |
| 1302 | B[H6O-O6-C6(1.000)] |
| 1303 | B[H4O-O4-C4(1.000), $\mathrm{H} 6 \mathrm{O}-\mathrm{O} 6-\mathrm{C} 6(0.588)$ ] |
| 1315 | ST[H61(1.000),H62(0.900)] |
| 1344 | $\mathrm{U}(\mathrm{ECH})[\mathrm{H} 5(0.821), \mathrm{H} 4(0.802),[\& T, \mathrm{H} 5-\mathrm{C} 5-\mathrm{O} 5(1.000)]]$ |
| 1367 | $\mathrm{U}(\mathrm{ECH})[\mathrm{H} 2(0.985),[\& T, \mathrm{H} 2-\mathrm{C} 2-\mathrm{O} 2(1.000)]]$ |
| 1393 | U(ETH)[H4(1.005),H2(0.725),[\&B,H3-C3-C4(1.000)]] |
| 1433 | U(ETH)[H2(0.984),H3(0.896),[\&T,H2-C2-C1(1.000)]] |
| 1466 | U(ETH)[H5(0.923),H4(0.630),[\&T,H5-C5-C4(1.000)]] |
| 1513 | U(ETH)[H1(0.999),[\&T,H1-C1-C2(1.000)]] |
| 1537 | U(ECH)[H5(0.857),[\&T,H5-C5-C6(1.000)]] |
| 1545 | $\mathrm{U}(\mathrm{RC})[\mathrm{C} 4(0.921), \mathrm{C} 1(0.456),[\& T, \mathrm{C} 3-\mathrm{C} 2-\mathrm{H} 2(1.000)]]$ |
| 1571 | $\mathrm{U}(\mathrm{RB})[\mathrm{C} 3(0.580),[\& \mathrm{~S}, \mathrm{C} 2-\mathrm{C} 1(1.000)]]$ |
| 1599 | $\mathrm{U}(\mathrm{RB})[\mathrm{C} 2(0.799), \mathrm{C} 4(0.558),[\& T, \mathrm{C} 2-\mathrm{C} 1-\mathrm{O} 1(1.000)]]$ |
| 1613 | AT[H62(1.000),H61(0.999)] |
| 1635 | S[C6-O6(1.000),C6-C5(0.868)] |
| 2912 | SS[H61(1.000),H62(0.975)] |
| 2944 | U(EOPSH)[H3(0.993), $\mathrm{H} 4(0.873), \mathrm{H} 2(0.434),[\& S, \mathrm{H} 3-\mathrm{C} 3(1.000)]]$ |
| 2947 | $\mathrm{U}(\mathrm{ERBH})[\mathrm{H} 1(0.894),[\& S, \mathrm{H} 1-\mathrm{C} 1(1.000)]]$ |
| 2947 | $\mathrm{U}(\mathrm{ERBH})[\mathrm{H} 1(0.893),[\& \mathrm{~S}, \mathrm{H} 1-\mathrm{C} 1(1.000)]]$ |
| 2951 | U(EOPSH)[H5(1.000),H2(0.947),H3(0.422),[\&S,H5-C5(1.000)]] |
| 2956 | U(EOPSH)[H4(0.997),H3(0.980),[\&S,H4-C4(1.000)]] |
| 2980 | AS[H62(1.000),H61(0.954)] |
| 3709 | S[H2O-O2(1.000)] |
| 3711 | S[H6O-O6(1.000)] |
| 3711 | $\mathrm{S}[\mathrm{H} 4 \mathrm{O}-\mathrm{O} 4(1.000), \mathrm{H6O}-\mathrm{O6}(0.518)]$ |
| 3714 | S[H3O-O3(1.000), $\mathrm{H} 4 \mathrm{O}-\mathrm{O} 4(0.420)]$ |
| 3720 | S[HO1-O1(1.000)] |

Assignment notation details: Assignments begin with a class abbreviation (e.g., S, AT, ETH, etc.) and are followed by a list of atom sets significantly involved in that sort of motion. In any atom set, the first atom is the one whose motion is being considered. Stretches and bends are referenced to 1 and 2 other valence-connected atoms, respectively. Torsions are referenced to only two other atoms (and not three) because torsional motion is considered as being about a given bond. For toroidal and circumferential (ring) motions, each set includes only the atom(s) moving most significantly with respect to the ring. The weights in parentheses are a general indication of the significance of the motion. Weights for simple (non-ring) motions are all scaled so that the maximum significance is 1 . Weights for ring-related motions are scaled accordingly, but after the simple motions are scaled, so their significances might be greater than one. Assignment abbreviations: Simple motions are noted without prefix or suffix. They are: $\mathrm{S}=$ stretch; $\mathrm{B}=$ bend; $\mathrm{T}=$ torsion. Complex motions typically have a prefix, a suffix, or both, in addition to an indication of motion type. The motion types, in addition to those employed in simple modes are: OPS = out of plane stretch (axial atoms); OPD = out of plane distortion; $\mathrm{RB}=$ ring "breathing" (expansion/contraction) motion; $\mathrm{RC}=$ ring circumferential motion; $\mathrm{RT}=$ ring toroidal motion. The prefix E indicates that the motion is of exocyclic atoms $(\mathrm{O}$ or H$)$. In the combinations ET and EC , the T stands for toroidal and the C , for circumferential. The prefix S or A represents "symmetric" or "asymmetric." For example, SS indicates symmetric stretch of two equivalent atoms. An "O" or "H" suffix indicates oxygen or hydrogen. A lack of a suffix indicates motion of the main ring. Concerted ring-related motions that do not fall easily into one of the above classes are prefixed with "U."

Table 11
Comparison of Vibrational Analyses for the $a$-D-Glucopyranose Monomer Using Geometries Computed at the Molecular Mechanics and Quantum Mechanics Levels.

| NMODE GLYCAM | B3LYP /6-31++g(2d,2p)//6-31++g(2d,2p) |  | HF/6-31G*//6-31G* |  |
| :---: | :---: | :---: | :---: | :---: |
| Frequency $\mathrm{cm}^{-1}$ | Frequency $\mathrm{cm}^{-1}$ | IR Intensity ( $\mathbf{k m m o l}^{-1}$ ) | Frequency $\mathrm{cm}^{-1}$ | IR Intensity ( $\mathrm{kmmol}^{-1}$ ) |
| H-O Stretches |  |  |  |  |
| 3720 | 3857 | 39.1 | 4132 | 63.8 |
| 3714 | 3851 | 40.3 | 4119 | 56.3 |
| 3711 | 3826 | 45.9 | 4099 | 59 |
| 3711 | 3825 | 52 | 4095 | 95.4 |
| 3709 | 3811 | 49.4 | 4089 | 56.1 |
| H-C Stretches |  |  |  |  |
| 2980 | 3063 | 43.3 | 3283 | 78.6 |
| 2956 | 3057 | 15.7 | 3270 | 32.8 |
| 2951 | 3040 | 19.7 | 3266 | 34.3 |
| 2947 | 3031 | 17 | 3224 | 56.9 |
| 2947 | 3012 | 42.4 | 3207 | 58.6 |
| 2944 | 2992 | 17.9 | 3195 | 2.8 |
| 2912 | 2988 | 58.4 | 3189 | 49.8 |
| -C6- Distortions |  |  |  |  |
| 1635 | 1458 | 12.2 | 1635 | 0.6 |
|  | 1065 | 52.3 | 1206 | 37.4 |
|  | 421 | 3.2 | 463 | 42.7 |
| Other H-C (Bends, Distortions, Toroidal and Circumferential Motions) |  |  |  |  |
| 1613 | 1498 | 6.9 | 1667 | 7 |
| 1537 | 1441 | 19.1 | 1620 | 11 |
| 1513 | 1416 | 13.3 | 1593 | 6 |
| 1466 | 1410 | 3.8 | 1583 | 25.1 |
| 1433 | 1387 | 1.9 | 1567 | 27.5 |
| 1393 | 1382 | 12.6 | 1552 | 47.9 |
| 1367 | 1366 | 8.1 | 1536 | 30.5 |
| 1344 | 1357 | 4.8 | 1515 | 23.2 |
| 1315 | 1351 | 38.3 | 1502 | 17.7 |
| 1135 | 1290 | 7 | 1486 | 26 |
| 1130 | 1281 | 13.5 | 1416 | 6.4 |
| 1103 | 1265 | 14.7 | 1402 | 42.4 |
| 1073 | 1232 | 20.9 | 1368 | 91.3 |
| 835 | 1222 | 22.1 | 1300 | 102.5 |
|  | 1194 | 25.7 | 944 | 24.5 |
|  | 850 | 12.8 |  |  |
|  | H-O-C Bends |  |  |  |
| 1303 | 1276 | 62.3 | 1437 | 72.5 |


| $\frac{\text { NMODE GLYCAM }}{\text { Frequency cm }^{-1}}$ | B3LYP /6-31 $++\mathrm{g}(2 \mathrm{~d}, 2 \mathrm{p}) / / 6-31++\mathrm{g}(2 \mathrm{~d}, 2 \mathrm{p})$ |  | HF/6-31G*//6-31G* |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Frequency $\mathrm{cm}^{-1}$ | IR Intensity $\left(\mathrm{kmmol}^{-1}\right)$ | Frequency $\mathrm{cm}^{-1}$ | IR Intensity ( $\mathrm{kmmol}^{-1}$ ) |
| 1302 |  |  | 1339 | 51.6 |
| 1285 |  |  | 1321 | 57.5 |
| 1274 |  |  |  |  |
| 1236 |  |  |  |  |
| H-O-C Torsions |  |  |  |  |
| 458 | 384 | 65.1 | 444 | 84.8 |
| 422 | 366 | 58.7 | 424 | 51.9 |
| 410 | 360 | 44.9 | 419 | 105.2 |
| 382 | 332 | 93 | 386 | 98.9 |
| 313 | 226 | 79.7 | 306 | 20.1 |
| 303 | 221 | 50.8 | 262 | 123.2 |
| 282 | 171 | 94 | 252 | 55.4 |
| 272 |  |  | 207 | 90.3 |
| $231$ |  |  |  |  |
| 214 |  |  |  |  |
| Other Exocyclic (Non-Ring) O Motions |  |  |  |  |
| 881 | 1102 | 72.2 | 1298 | 69.2 |
| 570 | 1042 | 287.4 | 1274 | 190.4 |
| 521 | 637 | 28 | 696 | 25.6 |
| 481 | 555 | 23.7 | 602 | 18.3 |
| 369 | 534 | 18.9 | 581 | 51.4 |
| 353 | 436 | 25.7 | 486 | 65.9 |
| 298 | 400 | 38.5 | 399 | 17.2 |
| 253 | 269 | 13.6 | 293 | 2.8 |
| 149 | 251 | 8.6 | 274 | 2.9 |
| $107$ | 247 | 9.9 | 155 | 1.8 |
|  | 143 | 3.9 |  |  |
|  | 113 | $3$ |  |  |
| Motions of the Ring |  |  |  |  |
| 1599 | 1464 | 12 | 1258 | 142.8 |
| 1571 | 1148 | 53.8 | 1255 | 29.1 |
| 1545 | 1129 | 13.3 | 1228 | 25.6 |
| 1088 | 1117 | 95.9 | 1215 | 207.9 |
| 1054 | 1082 | 71.9 | 1152 | 16.6 |
| 1028 | 1075 | 42.2 | 1126 | 32 |
| 995 | 1049 | 112.1 | 1099 | 16.5 |
| 983 | 1012 | 43 | 979 | 20.5 |
| 977 | 995 | 14.3 | 851 | 49.3 |
| 661 | 908 | 12.2 | 653 | 43.7 |
| $584$ | 770 | 38.4 |  |  |
|  | 595 | 22.1 |  |  |


| $\begin{aligned} & \text { NMODE GLYCAM } \\ & \text { Frequency } \mathrm{cm}^{-1} \end{aligned}$ | B3LYP /6-31++g(2d,2p)//6-31++g(2d,2p) |  | HF/6-31G*//6-31G* |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Frequency $\mathrm{cm}^{-1}$ | IR Intensity $\left(\mathrm{kmmol}^{-1}\right)$ | Frequency $\mathrm{cm}^{-1}$ | IR Intensity ( $\mathrm{kmmol}^{-1}$ ) |
| Other Motions of C6 and O6 |  |  |  |  |
| 951 | 306 | 14.3 | 1193 | 60.8 |
| 91 | 94 | 6.4 | 334 | 9.3 |
| $66$ | 65 | 0.9 | 123 | 2.4 |
|  |  |  | 106 | 4.7 |
|  |  |  | 75 | 2.9 |


[^0]:    © 2007 Wiley Periodicals, Inc.
    Correspondence to: R. J. Woods; rwoods@ccrc.uga.edu.
    The first two authors contributed equally to this work
    Current address for Karl N. Kirschner: Center for Molecular Design, Hamilton College, Clinton, New York 13323
    Current address for Jorge González-Outeiriño: Information Technology and Services, Syracuse University, 250 Machinery Hall Syracuse, NY 13244
    This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/0192-8651/ suppmat

[^1]:    $a_{\text {Relative energies taken from Ref. 61. Dotted lines indicate hydrogen bonds. }}$
    ${ }^{b}$ Charges computed by averaging the EA charges of: the $\alpha$ - and $\beta$-anomers of all glycosides.
    ${ }^{c}$ Charges computed by averaging the EA charges of: the $\alpha$ - and $\beta$-anomers of each glycoside.

[^2]:    ${ }^{a}$ Valence terms incorporated from the AMBER PARM94 parameter sets (CG corresponds to CT). $b_{\text {Bond stretching force constant }}\left(\mathrm{kcal} / \mathrm{mol} \AA^{2}\right)$. ${ }^{c}$ Bond length $(\AA)$.
    ${ }^{d}$ Angle-bending force constant $\left(\mathrm{kcal} / \mathrm{mol} \mathrm{rad}^{2}\right)$.
    ${ }^{e}$ Angle (degrees).
    Indicates relative barrier to rotation $(\mathrm{kcal} / \mathrm{mol})$.
    $g_{\text {Phase factor (degrees). }}$

[^3]:    ${ }^{a}$ Ref. 147.
    $b_{\text {Ref. } 146}$

