Published in final edited form as:

J Chem Theory Comput. 2011 October 11; 7(10): 3162–3180. doi:10.1021/ct200328p.

CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling

Olgun Guvench[ø], **Sairam S. Mallajosyula**[†], **E. Prabhu Raman**[†], **Elizabeth Hatcher**[†], **Kenno Vanommeslaeghe**[†], **Theresa J. Foster**[ø], **Francis W. Jamison II**[ø], and **Alexander D. MacKerell Jr.**[†,*]

[ø]Department of Pharmaceutical Sciences, University of New England College of Pharmacy, Portland, Maine 04103

[†]Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, 20 Penn St., HSF II-629, Baltimore, MD 21201

Abstract

Monosaccharide derivatives such as xylose, fucose, *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GlaNAc), glucuronic acid, iduronic acid, and *N*-acetylneuraminic acid (Neu5Ac) are important components of eukaryotic glycans. The present work details development of force-field parameters for these monosaccharides and their covalent connections to proteins via O-linkages to serine or threonine sidechains and via N-linkages to asparagine sidechains. The force field development protocol was designed to explicitly yield parameters that are compatible with the existing CHARMM additive force field for proteins, nucleic acids, lipids, carbohydrates, and small molecules. Therefore, when combined with previously developed parameters for pyranose and furanose monosaccharides, for glycosidic linkages between monosaccharides, and for proteins, the present set of parameters enables the molecular simulation of a wide variety of biologically-important molecules such as complex carbohydrates and glycoproteins. Parametrization included fitting to quantum mechanical (QM) geometries and conformational energies of model compounds, as well as to QM pair interaction energies and distances of model compounds with water. Parameters were validated in the context of crystals of relevant monosaccharides, as well NMR and/or x-ray crystallographic data on larger systems including oligomeric hyaluronan, sialyl Lewis X, O- and N-linked glycopeptides, and a lectin:sucrose complex. As the validated parameters are an extension of the CHARMM all-atom additive biomolecular force field, they further broaden the types of heterogeneous systems accessible with a consistently-developed force-field model.

Keywords

xylose; fucose; GlcNAc; GalNAc; glucuronic acid; sialic acid; Neu5Ac; O-linked; N-linked; glycoprotein; glycopeptide; glycan; peptidoglycan; glycosaminoglycan; force field; molecular dynamics; molecular mechanics; CHARMM

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

Supporting Information Available: Supporting Information includes model compound and crystalline intramolecular geometries, descriptions of O- and N-linked systems, model compound vibrational frequencies, model compound:water pair interaction geometries, dihedral potential energy scans for O-linked model compounds, and RMSD analysis of N-linked glycoprotein MD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

Introduction

Monosaccharides having the canonical formula $C_n(H_2O)_n$ are essential biomolecular components of life. Examples such as glucose are central to bioenergetics, and their polymers serve both structural and energy-storage functions, with prominent examples including cellulose, starch, and glycogen. However, the role of carbohydrates extends beyond this realm to include biomolecular functions such as molecular recognition. For example, the quality-control mechanism for protein folding,[1] the differences between blood group antigens,[2] and the ability of viruses to infect host cells[3,4] all have carbohydrates as a critical components. A common theme among the monosaccharides involved in such biomolecular functions is that their atomic compositions differ from the canonical formula. In particular, they are often deoxy, oxidized, or *N*-methylamine derivatives of $C_n(H_2O)_n$ monosaccharides, and/or are covalently liked to other biomolecules such as proteins and lipids via bonds involving oxygen or nitrogen atoms.

Classical force field development efforts aimed at enabling accurate modeling of carbohydrates and carbohydrate-containing biomolecular systems have been ongoing for over a decade.[5–21] While increased availability of computing resources has allowed for extensive use of quantum mechanical (QM) target data in an effort to capture the conformational energetics of carbohydrates, much of the focus has been on glucose and its diastereomers. Further limiting the scope of force-field based carbohydrate modeling is the fact that much of the parameter development work has not been done in the wider context of biomolecular force fields, such that attempts to model heterogeneous biomolecular systems containing proteins, lipids, and/or nucleic acids with carbohydrates may be hampered by differences in force field parametrization protocols and/or functional forms. It is of note that a recent parametrization of hexopyranoses (such as glucose) and their polymers was explicitly made to be compatible with the GROMOS family of biomolecular force fields,[20,22] and also that the most recent iteration of the GLYCAM force field, GLYCAM06,[21] contains parametrization for carbohydrate derivatives that can form the foundation for a generalizable biomolecular force field.[23]

Toward developing a comprehensive additive all-atom carbohydrate force field, we have developed and validated parameter sets for pyranose[24] and furanose[25] monosaccharides, as well as aldose and ketose linear carbohydrates and their reduced counterparts, the sugar alcohols.[26] Parameter sets have also been developed for glycosidic linkages involving both pyranoses[27,28] and furanoses,[28] with the force field shown to reproduce NMR elucidated solution conformational properties of the disaccharides of maltoside and cellobioside.[29] Combined, these parameter sets yield a force field that covers most carbohydrates that serve bioenergetic, structural, and energy-storage functions. The present work extends the parameter set to deoxy, oxidized, or *N*-methylamine monosaccharide derivatives as well as covalent linkages to proteins, thereby allowing the simulation of carbohydrates that are important in biomolecular function and molecular recognition. As with the stated previous efforts, the present parameter development was done explicitly in a fashion to make these new models compatible with the CHARMM additive all-atom biomolecular force field for proteins,[30,31] nucleic acids,[32,33] lipids,[34–38] and drug-like small molecules,[39] with the intention of creating a widely-applicable and robust force field for the modeling of biomolecular systems consisting of any combination of proteins, nucleic acids, lipids, carbohydrates, and/or small molecules.

Methods

Molecular mechanics (MM) calculations for parameter development were performed with the CHARMM software.[40,41] The force field potential energy function $U(r)$ was the same as that for the CHARMM protein,[30,31,42] nucleic acid,[32,33,43] lipid,[34–37,44,45], carbohydrate,[24–28] and small molecule all-atom additive force fields,[39]

$$\begin{aligned} U(r)= & \sum_{\text{bonds } b} K_b(b-b_0)^2 \\ & + \sum_{\text{valence angles } \theta} K_\theta(\theta-\theta_0)^2 \\ & + \sum_{\text{Urey-Bradley angles } S} K_S(S-S_0)^2 \\ & + \sum_{\text{dihedrals } \chi} K_\chi(1+\cos(n\chi \\ & -\delta)) \\ & + \sum_{\text{impropers } \varphi} K_\varphi(\varphi-\varphi_0)^2 \\ & + \sum_{\text{nonbonded pairs } ij} \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right] \\ & + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}. \end{aligned} \qquad (1)$$

The first five sums in Eqn. 1 account for bonded interactions. In these sums, K_b, K_θ, K_S, K_χ, K_φ and are bond, valence angle, Urey-Bradley angle, dihedral angle, and improper dihedral angle force constant parameters, respectively. b, θ, S, χ and φ are the bond distance, valence angle, Urey-Bradley angle 1,3-distance, dihedral angle, and improper dihedral angle values. The subscript 0 indicates an equilibrium value parameter. Additionally, for the dihedral term, n is the multiplicity and δ is the phase angle as in a cosine series. The sum over nonbonded pairs ij includes a Lennard-Jones (LJ) 6–12 term to account for dispersion and Pauli exclusion and a Coulomb term to account for electrostatic interactions. ε_{ij} is the LJ well depth, $R_{\min,ij}$ is the interatomic distance at the LJ energy minimum, q_i and q_j are the partial atomic charges, and r_{ij} is the distance between atoms i and j. The Lorentz-Berthelot combining rules are used to determine LJ parameters between different atom types.[46] There is no separate term for hydrogen bonding interactions, as these are accounted for in the parametrization through a combination of LJ and Coulomb energies

A modified version of the rigid three-site TIP3P model was used to represent water,[47,48] and the SHAKE algorithm[49] was applied to keep water molecules rigid and to constrain covalent bonds between hydrogens and their covalently bound heavy atoms to their equilibrium values. Gas-phase molecular mechanics energies were calculated using infinite nonbonded cutoffs. Aqueous and crystal simulations employed periodic boundary conditions[46] to minimize boundary artifacts and to simulate the infinite crystal environment. A force-switched (aqueous) or energy-switched (crystal) smoothing function[50] was applied to LJ interactions in the range of c-2 to c, where c is the cutoff distance in Å. Long-range Coulomb interactions were handled using particle mesh Ewald[51] with a real-space cutoff of c. The equations of motion were integrated with the “leapfrog” integrator[52] and a timestep of dt. In the molecular dynamics (MD) simulations, the isothermal-isobaric ensemble was generated via Nosé-Hoover thermostating,[53,54] Langevin piston barostating,[55] and a long-range correction to the pressure to account for LJ interactions beyond the cutoff distance c.[46] Condensed-phase simulations were done at experimental temperature and pressure, which was 298 K and 1 atm for all simulations. Simulations of crystals were based on availability of relevant systems in the Cambridge Structural Database[56] (CSD), and employed the appropriate experimental unit cell geometries with crystallographic water molecules and/or crystallographic counter-ions; aqueous simulations employed a truncated octahedron as the periodic system. For aqueous simulations, the cell length dimensions were varied isotropically to maintain the target pressure during simulation, whereas unit cell edge lengths in crystal simulations were allowed to vary independently. Angular crystal cell parameters of 90° were constrained to this value while those not 90° were allowed to vary independently. Table 1 lists the simulation *c* and *dt* values, along with simulation lengths, MD snapshot frequency, and the number of times each system was simulated. In cases where a system was simulated once, error estimates for data were generated by treating each MD snapshot as an independent sample and using the expression $t_{\mathrm{critical}}*s/(n^{0.5})$, where *n* is the number of snapshots, *s* is the sample standard deviation, and $t_{\mathrm{critical}} = 1.960$, which is the value for a 95% confidence level for a *t* distribution with infinite degrees of freedom. In cases where a system was simulated more than once, different trajectories were generated by random assignment of initial velocities, and each trajectory was treated as an independent sample to generate error estimates for data using the expression $t_{\mathrm{critical}}*s/(n^{0.5})$, where *n* is the number of trajectories, *s* is the standard deviation of the average values calculated for each trajectory, and t_{critical} is the value for a 95% confidence level for a *t* distribution with *n* −1 degrees of freedom. Additional simulation details such as system construction are mentioned in the Results and Discussion section for each system.

The Gaussian 03 program[57] was used for all QM calculations. For small model compounds (Figure 1 **M4**, **M6a**, **M6b**, **M8**), geometry optimization was done at the MP2/cc-pVTZ level;[58,59] otherwise geometry optimization was done at the MP2/6-31G(d) level[60] followed by a MP2/cc-pVTZ single point energy calculation (MP2/cc-pVTZ//MP2/6-31G(d)). The cc-pVTZ basis set was used for evaluating all conformational energies as it demonstrates a favorable combination of efficiency and accuracy on carbohydrate systems.[24,61,62] Vibrational calculations were performed using the MP2/6-31G(d) model chemistry, with tight convergence tolerances applied; the geometries from these unconstrained optimizations were also used as reference geometries for gas-phase minimized MM model compound geometries. A scale factor of 0.9434 was applied to QM vibrational frequencies, as required to account for limitations in the level of theory and reproduce experimental frequencies.[63] Potential energy decomposition analysis was performed using the MOLVIB utility in CHARMM using the internal coordinate convention of Pulay et al.[64] All potential energy scans were performed with only the scanned dihedral angles constrained. MM energies were fit to QM potential energies scans using the freely-available Monte Carlo simulated annealing (MCSA) dihedral parameter fitting program “fit_dihedral.py”[27,65] (available for download at http://mackerell.umaryland.edu). For each dihedral being fit, three multiplicities *n* of 1, 2, and 3 were included and the corresponding K_χ values (Equation 1) were optimized to minimize the root mean square error *RMSE* between the MM and QM energies as defined by

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

where the sum is over all conformations *i* of the molecule in the scan, w_i is a weight factor for conformation *i*, E_i^{QM} is the QM energy of conformation *i*, E_i^{MM} is the total MM energy, including the energy of the dihedrals for which the parameters are being optimized (Equation 1), and *c* is a constant that vertically aligns the data as the optimization proceeds to minimize the *RMSE* and is defined by

$$\frac{\partial RMSE}{\partial c}=0. \tag{3}$$

w_i values can be empirically chosen to, for example, favor more accurate fitting of low-energy conformations while sacrificing the fit of high-energy ones. In fitting the dihedral parameters, K_χ values were constrained to be no more than 3 kcal/mol, and phase angles δ were limited to 0 and 180° to maintain symmetry of the dihedral potentials about $\chi = 0°$ and thereby ensure applicability of dihedral parameters to both enantiomers of a chiral compound.

To test the nonbonded parametrization for charged or hydrogen bond-forming moieties in model compounds, QM calculations were done to determine interaction energies for model compound:water-molecule pairs. To ensure consistency across the CHARMM additive force field, these calculations followed a standard procedure.[66] First, the solute:water interaction distance was optimized at the HF/6-31G(d) level, with constraints on all other degrees of freedom. Here, the water intramolecular geometry in both QM and MM calculations of pair interaction data was that of the TIP3P water model,[47] and the model compound geometry was one that was previously gas-phase optimized in the QM or CHARMM representation, respectively. Second, following optimization, HF/6-31G(d) interaction energy target data were calculated as $s^*(E_{\mathrm{pair}} - E_{\mathrm{solute}} - E_{\mathrm{water}})$, with no basis-set superposition-error correction and the empirical scaling factor of s introduced to yield parameters appropriate for a condensed phase force field,[30,67] where s=1.00 in the case of solutes having moieties with non-zero formal charge and 1.16 otherwise. The interaction distance target data were calculated as the QM-optimized distance minus 0.2 Å, again to yield parameters appropriate for a condensed phase force field.

Results and Discussion

I. Parameter development

Truncated derivatives—The monosaccharides xylose and fucose can be viewed as truncated derivatives of $C_6H_{12}O_6$ hexopyranoses like glucose (Figure 2, compounds **2**, **3**, and **1** respectively). Relative to these hexopyranoses, xylose is missing the entire hydroxymethyl group while fucose, a deoxyhexopyranose, lacks the hydroxyl on the hydroxymethyl group. From a parametrization standpoint, this suggested the immediate transfer of existing parameters. In particular, existing hexopyranose parameters,[24] combined with existing alkane[34–38,68] and linear ether parameters[69] applied to the methyl group on fucose, provided coverage for all atoms and connectivities in these molecules. Crystal simulations of xylose, fucose, and rhamnose (a diastereomer of fucose, with inverted chiralities at C2 and C4) demonstrated the suitability of these transferred parameters. With the exception of the C5-O5 bond length in fucose, all average bond, angle, and dihedral values were consistent with the experimental crystallographic values (Table S1 of the supporting information). Further optimization of the C5-O5 bond parameters, which were originally developed for use in hexopyranoses,[24] and subsequently used in other hexopyranose derivatives (see below), would have required the creation of a new atom type. However, in the interest of balancing accuracy with simplicity and generality of the force field parameters, a new atom type was not introduced as the extent of disagreement with the crystal data was deemed acceptable. Unit cell geometries were consistent with the experimental values and, in line with prior results,[24–27] unit cell volumes were systematically overestimated by several percent (Table 2).

N-acetylamines—One of the most common modifications to hexopyranoses is the replacement of the C_2 hydroxyl with an *N*-acetylamine group, resulting in monosaccharides like *N*-acetylglucosamine (GlcNAc; Figure 2, **4**) and *N*-acetylgalactosamine (GalNAc; Figure 2, **5**). In eukaryotes, both GlcNAc and GalNAc are important components of the oligosaccharides that are post-translationally attached to proteins to create glycoproteins.[70] Using a fragment-based approach, isopropylacetamide (Figure 1, **M4**) was used to develop parameters for these types of sugars. Initial parameters for **M4** were transferred from existing parameters previously developed for *N*-methylacetamide in the context of proteins,[30] and with carbon Lennard-Jones parameters taken from an improved set of alkane parameters.[68,69] The 3-fold dihedral term for rotation of the isopropyl group was fit to the QM relaxed potential energy scan (PES), and the transferred C-CT equilibrium length was increased by 0.030 Å. These bonded parameters yielded near-ideal agreement between the QM and the MM conformational energies (Figure 3a), as well as good agreement with bond lengths, valence angles and dihedral angles (Table S2, Supporting Information). Vibrational frequencies also showed good agreement with the exception of those that, in the QM representation, involved wagging or deformation of atoms in the amide bond (Figure S1, frequency #'s 10–13, Supporting Information). While these frequencies were overestimated in the MM representation compared to the gas-phase QM, this is in fact appropriate behavior for a condensed-phase force field as the relevant frequencies tend to increase in going from the gas phase to an aqueous environment.[30] Finally, using the transferred Lennard-Jones and partial charge nonbonded parameters, water pair interaction energies with the amide CO and NH groups faithfully reproduced the target data (Table 3; Figure S2, Supporting Information).

Transfer of the parameters allowed for immediate creation of models for GlcNAc and GalNAc. Only crystals of the α-anomers of these two sugars were available through the Cambridge Structural Database;[56] MD simulations of infinite crystals using the lowest *R*-value structures as starting conformations (ACGLUA11: two molecules of α-GlcNAc in unit cell; AGALAM10: two molecules of α-GalNAc in unit cell) showed that the bonds, valence angles, and dihedral angles were all well-represented by the force field model in that average values from the simulations corresponded to those in the reference experimental crystals (Table S3, Supporting Information). The availability of a crystal structure of *N*-acetyl-β-mannosamine (ManNAc) monohydrate (NACMAN10: four molecules of β-ManNAc + four water molecules in unit cell) allowed the testing of a β-anomer using the same parameters and gave similarly good results (Table S3, Supporting Information), and as with the truncated derivatives, crystal volumes for all the *N*-acetylamines were overestimated by several percent (Table 2).

Carboxylates

Glucuronate and iduronate: Oxidation of the C6 alcohol to form a carboxylic acid yields hexopyranose derivatives such as glucuronic acid and iduronic acid, which ionize at physiological pH to yield glucuronate (Figure 2, **6**) and iduronate (Figure 2, **7**). Among other functions, these compounds are important as components of glycosaminoglycans[70] and for metabolic conjugation with drugs by the liver.[71] Continuing with a model compound-based approach, **M6a** and **M6b** (Figure 1) were used to develop parameters for glucuronate and iduronate. As detailed below, bonded and nonbonded parameters were transferred from analogous parameters, and missing angle and dihedral parameters were fit to QM geometries and conformational energies.

M6a (propanoate) parameters were previously developed for the CHARMM protein force field,[30] allowing for immediate extension to **M6b** (α-methoxy-propanoate), which mimics the C6 carboxylate in the context of the C4, C5, O5, and C1 atoms of the hexopyranose ring.

Lennard-Jones parameters were updated based on recent work on alkanes.[68,69] Additional parameters required upon introduction of the methoxy group were transferred by analogy, leaving only the O5-C5-C6-O dihedrals to be fit. **M6b** was constructed with the C1-O5-C5-C6 analogous dihedral in the *trans* conformation. An optimized MP2/cc-pVTZ scan was done on the OCCO torsion with no other constraints on the system, and the C1-O5-C5-C6 dihedral stayed *trans* for the entire scan. Self-consistent optimization of the O61-C6-O62 angle (equilibrium angle increased by 8 degrees), the C5-C6-O61 and -O62 angles (equilibrium angle decreased by 4 degrees), the O5-C5-C6 angle (equilibrium angle decreased by 8.5 degrees relative to linear ethers[69]), and the OCCO torsions yielded good conformational energies (Figure 3b), minimum-energy geometries (Table S4, Supporting Information), vibrational frequencies (Figure S3, Supporting Information), and water interaction energies (Table 3 and Figure S4, Supporting Information) as compared to the QM target data.

Relevant crystals included β-glucuronate (NABDGC) and α-galacturonate (CANAGLC10). In the case of glucuronate, the full monoclinic unit cell consisted of 2 monosaccharides, 2 water molecules, and 2 sodium ions, and in the case of galacturonate the full hexagonal unit cell consisted of 6 monosaccharides, 12 water molecules, 2 sodium ions, and 2 calcium ions. After initial simulations, the C5-C6 equilibrium bond length was reduced by 0.042 Å, following which the bonds, valence angles, and dihedral angles had average values from MD simulations that were consistent with the experimental geometries (Table S5, Supporting Information). An interesting exception was bonds and angles involving oxygens in the carboxyl groups. In the crystals, there is a 2–3% difference between equivalent C6-O bonds and between equivalent C5-C6-O angles, presumably due to differences in the local chemical environments of the carboxyl oxygens. In the MM representation, there is no difference in the parameters for these equivalent oxygens, and these large asymmetries in bonds and angles are not reproduced, pointing to some limitations of the pairwise-additive functional form of the force field where the bonds and angles are treated with harmonic terms. Also of note is that the O5-C5-C6 angles in these crystals have accurate values in the simulations relative to the experiments, demonstrating the transferability of the corresponding parameter optimized on **M6b**. Finally, both crystals had average unit cell parameters from the simulations that were largely consistent with the experimental reference values (Table 2).

Sialic acid: Like glucuronic acid and iduronic acid, *N*-acetylneuraminic acid (Neu5Ac), or “sialic acid” as it is commonly called, has a carboxylic acid moiety that is deprotonated at physiological pH (Figure 2, **8**). Neu5Ac is important not only as a common component of glycosyl groups added as post-translational modifications to proteins, but also as a critical participant in molecular recognition resulting in viral infection of human cells, in particular influenza virus infection.[4,72] In addition to the carboxyl group, Neu5Ac contains an *N*-acetylamine group, like GlcNAc and GalNAc, as well as a linear polyalcohol group, like linear carbohydrates and sugar alcohols.[26] While parameters for the *N*-acetylamine group can readily be transferred from those developed for GlcNAc and GalNAc, and parameters for the linear polyalcohol group from previously-developed linear polyalcohol parameters,[26] parametrization of the carboxyl group is complicated by the presence of not only an ether moiety connected to the same carbon atom C2, but also a hydroxyl group.

M8 (α-methoxy-lactate), which is **M6b** with the addition of a hydroxyl group, was used to develop parameters involving the C2 anomeric carbon in Neu5Ac. After transferring analogous parameters from **M6b**, both bonded and nonbonded parameters were optimized to reproduce target QM geometries, vibrational frequencies, conformational energies, and water pair interaction energies. In particular, to reproduce angle geometries, equilibrium angle values for the C1-C2-C3, O2-C2-C3, and C3-C2-O6 angles were adjusted, resulting in good agreement with the target data (Table S6, Supporting Information). Angle parameters optimized using **M6b** proved transferable, and as with **M6b**, asymmetries in bond and angle geometries involving the equivalent carboxyl oxygen atoms seen in the QM representation were not captured in the MM representation. The MM vibrational frequencies were consistent with those from QM calculations, with the exception of the hydroxyl OH stretch, which has the highest frequency in the empirical model (Figure S5, Supporting Information). In the force field, this OH stretch frequency is ~700 cm^{-1} greater than the next-highest set of frequencies, which are due to methyl CH stretches, and is similar to calculated QM or experimental infrared OH stretching frequencies for simple alcohols.[24] In contrast, in the QM representation, the OH stretch has a frequency similar to the CH stretches, and therefore much lower than OH stretching frequencies for simple alcohols. The reason for this is quite clear from the minimum-energy geometry used to compute the vibrational frequencies, in which the hydroxyl group is oriented to form an intramolecular hydrogen bond with the carboxyl group. This strong hydrogen bond leads to weakening of the OH bond as evidenced by an increase of 0.04 Å in the OH bond length for conformations having this intramolecular interaction relative to conformations without the interaction, as observed in a relaxed QM scan of hydroxyl rotation (described below). This phenomenon cannot be captured using molecular mechanics; however, since covalent bonds involving hydrogens are typically constrained to their equilibrium values using SHAKE[49] or a related algorithm, this limitation is not a significant concern for intended applications of the model.

QM target conformational energies for **M8** were from geometry-optimized MP2/cc-pVTZ scans of carboxyl and of hydroxyl rotation in increments of 15°. The only constrained degree of freedom during the scans was the dihedral being scanned. Therefore, during dihedral scanning of the carboxyl group, the hydroxyl group underwent rotation due to nonbonded interactions with the carboxyl group, and vice versa. In contrast, the C1-C2-O6-C6 dihedral was built in the *trans* geometry and stayed in this local minimum throughout the dihedral scanning. Dihedral parameters for $O_{carboxyl}$-C1-C2-O2 and C1-C2-O2-HO2 were simultaneously fit to the 50 QM conformational energies/geometries,[65] with harmonic restraining potentials applied to the $O_{carboxyl}$-C1-C2-O2, C1-C2-O2-HO2, and C1-C2-O6-C6 dihedrals in the MM representation to ensure a match between the MM and QM conformations during the fitting process. The resulting optimized parameters yielded good agreement with the target data (Figure 3c, d). Finally, water interaction energies with the carboxyl group as a hydrogen bond acceptor (analogous to Figure S4), the ether as a hydrogen bond acceptor (analogous to Figure S4), and the hydroxyl as both an acceptor and a donor (Figure S6) showed systematically too favorable interaction energies with the carboxyl group using the partial charges transferred from **M6a/b**, which, as described above, were themselves directly transferred from previous work and seen to be suitable in the contexts of **M6a/b**. The partial charges on the carboxyl oxygens were therefore adjusted from −0.76 to −0.60 e and the partial charge on the carboxyl carbon decreased from 0.62 to 0.30 e. While this may appear to be a large change, it is important to note that the net charge of −1 e on the carboxylate group remains unaltered, and that this net charge is the main determinant for the strength of electrostatic interactions with this moiety. Prior to the charge redistribution, QM interactions of the carboxyl oxygens with water were too favorable by ~2 kcal/mol. Additionally, interactions with the ether oxygen were too unfavorable by ~2 kcal/mol owing to electrostatic repulsion of water by the adjacent carboxyl carbon. In contrast, with the new partial charge set, good agreement was achieved with the target data (Table 3). Bonded and nonbonded parameter optimization was done self-consistently, and all presented data are from the final parameter set. To create a force field model for Neu5Ac **8**, parameters from **M8** were combined with those from hexopyranoses,[24] polyalcohols,[26] and **M4**. MD simulation of the single example of Neu5Ac in the deprotonated form in the Cambridge Structural Database[56] (KEMYAC; 4 molecules of monosaccharide in the β anomeric form + 4 sodium ions + 12 water molecules, in the complete tetragonal unit cell) pointed to additional parameter optimization. As with the analogous parameter in crystals for β-glucoronic acid and α-galacturonic acid, the C1-C2 equilibrium bond length was reduced by 0.042 Å; additionally, the C2-C3 equilibrium bond length was increased by 0.035 Å, the O2-C2-O6 equilibrium angle value was reduced by 3.5 degrees, and the C3-C2-O6 equilibrium angle value was increased by 2.5 degrees in order to achieve good agreement of bonds, angles, and geometries at and near the carboxylate moiety (Table S7). Finally, geometric parameters for the rectangular unit cell were in very close agreement with the experimental values (Table 2).

O-glycans—The O-glycosidic carbohydrate-protein bond, between the anomeric carbon of a carbohydrate and the sidechain alcohol of either the amino acid serine (Ser) or threonine (Thr),[70,73] is an important biological linkage present, for example, in mucin glycoproteins[74] and as a post-translational modification for cytosolic proteins in signaling pathways.[75] To develop force field parameters for O-glycosidic linkages, four dipeptide derivatives of the hexopyranose analog tetrahydropyran were chosen as model compounds. These model compounds correspond to the α and β anomers of Ser- (Figure 1, **MO1** and **MO2**) and of Thr-linked (Figure 1, **MO3** and **MO4**) carbohydrates. Initial values for bonded and nonbonded parameters were transferred from ethers,[69] carbohydrates,[24] and proteins,[30] leaving as targets for parametrization the glycosidic dihedrals about the C1-O1 bond (O5-C1-O1-Cβ and C2-C1-O1-Cβ), O1-Cβ bond (C1-O1-Cβ-Cα and additionally C1-O1-Cβ-Cγ in the Thr-linked analogs) and the Cβ-Cα bond (O1-Cβ-Cα-N and O1-Cβ-Cα-C).

To parameterize glycosidic dihedral rotation about the C1-O1 bond and the O1-Cβ bond, 2D MP2/cc-pVTZ//MP2/6-31G(d) scans were performed on the O5-C1-O1-Cβ (ϕ_S) / C1-O1-Cβ-Cα (ψ_S) surfaces for all four model compounds (Figure 4). Global minima for the α anomers were located at ϕ_S/ψ_S = −105°/60° and −90°/45° for the Ser and Thr dipeptides (**MO1** and **MO3**) respectively. For the β anomers, global minima corresponded to ϕ_S/ψ_S values of −75°/75° and −75°/45° for the Ser and Thr dipeptides (**MO2** and **MO4**), respectively. During the optimized dihedral scans the only constraints were on the dihedrals being scanned, thereby allowing full relaxation of all other degrees of freedom. For example, the peptide backbone geometry relaxed to various parts of the extended region of the protein ϕ/ψ Ramachandran surface for each of the four global-minimum structures, with QM-optimized ϕ/ψ values for the Ser α,β anomers **MO1** and **MO2** being −156°/168°, −157°/161°, respectively, and for the Thr α,β anomers **MO3** and **MO4** being −153°/170°, −153°/172°, respectively.

For the Ser analogs **MO1** and **MO2**, dihedral parameters for O5-C1-O1-Cβ, C2-C1-O1-Cβ (both involving rotation about the same bond C1-O1), and C1-O1-Cβ-Cα were simultaneously fit to both QM potential energy scans. Similarly, for the Thr analogs **MO3** and **MO4**, parameters for the dihedrals O5-C1-O1-Cβ, C2-C1-O1-Cβ (both involving rotation about the same bond C1-O1), and C1-O1-Cβ-Cα, C1-O1-Cβ-Cγ (both involving rotation about the same bond O1-Cβ), were simultaneously fit to both QM scans.

To parameterize the dihedral rotations O1-Cβ-Cα-N and O1-Cβ-Cα-C about the Cβ-Cα bond, 1D MP2/cc-pVTZ//MP2/6-31G(d) scans were performed on the O1-Cβ-Cα-N dihedral for all four model compounds (Figure S7). Global minima for the α anomers were located at −60° and 75° for the Ser and Thr dipeptides (**MO1** and **MO3**) respectively. For the β anomers, global minima corresponded to dihedral values of −165° and 45° for the Ser and Thr dipeptides (**MO2** and **MO4**), respectively. For both the Ser and Thr analogs, **MO1** to **MO4**, dihedral parameters for O1-Cβ-Cα-N and O1-Cβ-Cα-C were simultaneously fit to both the anomeric QM potential energy scans. To ensure faithful reproduction of conformational energies near the QM minima, during the MCSA fitting the global minimum conformations were given weight factors w_i (Equation 2) of 3, conformations with energies above 14 kcal/mol weight factors of 0 and all other conformations weights of 1.

A single relevant crystal structure, namely that of a Thr α-anomer (CSD code, R factor, compound name: COSHEX, 3.8, O-α-D-Mannopyranosyl-(1-3)-L-threonine), was found through a CSD search, and this structure along with QM-optimized structures of **MO1-4**, were used to guide additional parametrization of bonded terms. MD simulations of the crystal with two monomers per unit cell showed that a few transferred equilibrium bond lengths and valence angles had to be modified to better match the experimental intramolecular geometries. Therefore, the O1-Cβ equilibrium bond length was increased by 0.01 Å, the equilibrium valence angles for O1-Cβ-Cα and O1-Cβ-Cγ were decreased by 4.5° and 1.5°, and the equilibrium valence angle for Cβ-O1-C1 was increased by 2.0°. These additional optimizations yielded good reproduction of both QM and crystal geometries (Table S8, Supporting Information), and crystal unit cell dimensions consistent with other crystals in the present study (Table 2).

Of note, the O-glycan parametrization was also able to reproduce the correct anomeric configurational preference when compared with QM calculations. Defining the anomeric ΔE as $E_\alpha - E_\beta$, for the Ser O-linkage, the MM ΔE value of 1.72 kcal/mol compares reasonably well with the QM value of 3.17 kcal/mol. Similar agreement is seen for the Thr O-linkage, where the MM ΔE value is −3.28 kcal/mol and the QM value −5.27 kcal/mol. Thus we note that the force field is able to predict the correct configurational preferences for both the Ser and Thr linkages with the Ser linkage favoring the α configuration and the Thr linkage favoring the β configuration. The final potential energy surfaces using the final optimized parameters are presented in Figure 4a to 4d and Figure S7, with *RMSE* values with respect to the QM conformational energies ranging from 0.9 to 2.0 kcal/mol across all the dihedral scans.

N-glycans—Post-translational protein modification by N-linked glycosylation consists of the addition of oligosaccharides to the sidechains of asparagine (Asn) residues. This covalent modification, which occurs in the endoplasmic reticulum, plays a critical role in cell surface expression and is often required for protein stability and biological function.[70,76] It has been found that N-linked glycosylation generally occurs at the sequence Asn-X-Ser/Thr, where X is any amino acid except proline.[77,78] This type of glycosylation is found in nearly all eukaryotes,[70,76] and in most cases the linkage occurs between Asn and *N*-acetylglucosamine (GlcNAc), replacing the alcohol moiety of GlcNAc C1 with an amide linkage to the Asn side chain.

To develop the force field parameters for this linkage, tetrahydropyran with *N*-acetylamine substituted at C1 was chosen as the model compound, with both α and β anomer analogs used for the parametrization process (Figure 1, **MN1** and **MN2**). Most of the initial bond, angle, and dihedral parameters were readily transferred from the *N*-acetylamine substitution at the C2 position, as developed for GlcNAc and GalNAc. Additionally, the O5-C1-N angle parameter was transferred from O5-C1-O by analogy, and the dihedral parameters C5-O5-C1-N and O5-C1-N-C were transferred from those for C5-O5-C1-O and C1-C2-N-C.

To test the transferred parameters QM MP2/cc-pVTZ//MP2/6-31G(d) scans were performed for the O5-C1-N-C dihedral in the two model compounds (Figure 5a,b). These scans were followed by two additional scans of the C5-O5-C1-N dihedral with the O5-C1-N-C dihedral constrained to its corresponding QM global minimum (Figure 5c,d). In the case of the C5-O5-C1-N scans, which correspond to ring deformation, the transferred parameters (Figure 5c,d “MM Trsfd”) adequately reproduced the target data and could not be further improved by additional fitting (Figure 5c,d “MM Fit”). In contrast, the transferred parameters for O5-

C1-N-C, which determine the energetics of rotation of the *N*-acetylamine group, gave incorrect locations for the minimum energies as well as barriers to rotation that were too high (Figure 5a,b "MM Trsfd"). Using the QM O5-C1-N-C scans for both model compounds as target data, dihedral force constants were developed for the O5-C1-N-C dihedral using the MCSA fitting procedure, with the same weighting protocol described previously for the O-glycan model compounds. After fitting, *RMSE* values for the **MN1** and **MN2** O5-C1-N-C scans were 1.36 and 0.58 kcal/mol, respectively, reflecting a much closer match to the target QM surfaces (Figure 5a,c "MM Fit") than with the transferred parameters.

To test the transferred and optimized parameters, geometrical descriptors of the QM and MM minimized geometries were compared along with crystalline intramolecular geometries and unit cell parameters for C1 mono-substituted monosaccharides. A CSD survey yielded four mono-substituted *N*-acetylamine crystals (CSD ref. code, R factor, compound name: AVUVES, 4.02, β-1-*N*-acetamido-ᴅ-mannopyranose monohydrate, AVUVIW, 3.82, β-1-*N*-acetamido-ᴅ-galactopyranose, AVUVOC, 3.86, β-1-*N*-acetamido-ᴅ-xylopyranose, RESJEE, 3.28, β-1-*N*-acetamido-ᴅ-glucopyranose). The results of the comparison between the QM and MM intramolecular geometries and the MD simulations of the infinite crystals are tabulated in Tables S9 and S10 of the Supporting Information, and demonstrate good agreement of the MM data with regard to both QM-optimized model compounds geometries and crystal data for bonds and angles. MM dihedral angles are also consistent with target QM and crystal data with the exception of those involving rotation around the C1-N bond in the crystal. However, for these O5-C1-N-C and C2-C1-N-C dihedrals, discrepancies between the crystallographic and MD average values can be explained as resulting from a flat energy profile in the region of the global minima. In particular, for the β anomer analog **MN2**, the energy cost for going from a O5-C1-N-C dihedral value of −90° to one of −120° is only 0.5 kcal/mol (Figure 5b). And finally, as with other crystal simulations, a systematic overestimation of the crystal volumes was observed (Table 2).

Since N-glycosylation commonly involves linkage of *N*-acetylglucosamine (GlcNAc) to the sidechain of Asn, parameters are required for the dihedral angle between the nitrogens of the *N*-acetlyamine groups at position C1 (anomeric carbon) and C2 of GlcNAc involved in such a linkage. To parameterize this dihedral, tetrahydropyran with *N*-acetlyamine substitutions at both the C1 and C2 positions was chosen as the model compound, and QM conformational energies were collected for all four possible diastereomers (Figure 1, **MN3**, **MN4**, **MN5**, **MN6**). Initial parameters for N2-C2-C1-N1 were transferred from the analogous OCCO dihedral of the hexopyranose monosaccharide force field, which were developed using ethylene glycol as a model compound and validated based on crystallographic ring pucker geometries.[24] These transferred parameters (Figure 6, "MM Trsfd") reproduce the QM energy scans (Figure 6, "QM") as well as do the parameters explicitly fit to the target data (Figure 6, "MM Fit"), including good reproduction of the locations and shapes of global energy minima. Reflecting the appropriateness of the transferred dihedral parameters are the similarities in the *RMSE* values of the transferred vs. fit parameters for **MN3**, **MN4**, **MN5** and **MN6**: 0.88 kcal/mol vs. 0.70 kcal/mol, 1.86 kcal/mol vs. 1.73 kcal/mol, 0.43 kcal/mol vs. 0.44 kcal/mol, and 3.12 kcal/mol vs. 3.41 kcal/mol. Thus, the transferred parameters were retained as the final parameters, and were subsequently used to compare the geometries of the QM global minimum and the MM minimized geometries (Table S11, Supporting Information). The MM model, which uses a single set of parameters for all four model compounds, faithfully captures bond lengths and angles, thereby requiring no further adjustment of the equilibrium bond lengths and angles, as well as most dihedrals. The exception is for the dihedrals O5-C1-N-C and C2-C1-N-C dihedrals, where average errors are 20.9° and 18.0°, in part due to the flat potential profile associated with this dihedral as discussed above. Furthermore, these average errors are heavily influenced by one outlier, namely **MN6**, because the MM optimized geometry of **MN6** favors an intramolecular hydrogen bond between the two acetylamine units thereby locking the O5-C1-N-C dihedral angle at 153° compared to the QM value of 83°. Excluding this compound from the analysis yields average errors of 4.8° and 0.3° for these two dihedrals. A CSD survey yielded only one di-substituted *N*-acetylamine crystal (CSD ref. code, R factor, compound name: CAKFAV, 2.50, N'-(2-Acetamido-2-deoxy-β-D-glucopyranosyl) acetamide monohydrate). Based on an MD simulation of the crystal, all bond lengths and valence angles are well reproduced by the transferred force field parameters (Table S12, Supporting Information), with average errors for bond lengths of 0.008 Å and ranging from −1.1° to +2.1° for valence angles. Furthermore, all dihedral angles, including those for rotation about both the C1-N and C2-N bonds, are well reproduced by the transferred parameters. Finally, percentage errors for the unit cell parameters *A*, *B*, *C* and β were calculated to be −0.1%, 0.8%, 4.7% and −19.6%, respectively, and the error in the crystal volume was 1.0% (Table 2). The large change in β is due to a slight shift of the two monomers in the crystal with respect to each other; however, this change does not lead to a significant change in the unit cell volume.

To test the applicability of the above parametrization for amino acid – carbohydrate conjugates, the optimized parameters were applied to crystalline N-linked monosaccharides. A CSD survey yielded three crystal structures of N-linked monosaccharides, all with the N-linkage in the β conformation (CSD ref. code, R factor, compound name: ASGPRS, 6.00, 2-Acetamido-1-N-(L-aspart-4-oyl)-2-deoxy-β-D-glucopyranosylamine hydrate, BEHPIN. 5.40, 4-N-(2-Acetamido-2-deoxy-β-D-glucopyranosyl)-L-asparagine trihydrate, BEHPOT, 7.20, 4-N-(β-D glucopyranosyl)-L-asparagine monohydrate). In all cases, the MD simulations reproduced crystallographic bond lengths, valence angles, and dihedral angles to within acceptable errors (Table S13, Supporting Information), with the exception of the C1-C2-N2-C and C3-C2-N2-C dihedrals. These latter dihedrals correspond to rotation about the N2-C2 bond, and the observed error can be rationalized from the potential energy surface for the model compound **M4**, where the conversion from one minimum to another in the crystal simulation corresponds to conversion across the 0.5 kcal/mol barrier from one of the two global minima to the other in **M4**. And finally, all unit cell parameters as calculated by averaging across the MD trajectories were close to the corresponding experimental crystallographic values (Table 2).

II. Application to example systems

Toward demonstrating the utility of the new parameter set, MD simulations were done on relevant carbohydrates alone and covalently conjugated or non-covalently bound to proteins. Carbohydrate-only systems consisted of aqueous simulations of monomeric GlcNAc, the linear glycosaminoglycan polymer hyaluronan, and the branched glycan sialyl Lewis X. Covalent carbohydrate-protein conjugates consisted of 2 glycoproteins containing only N-linked glycans and 2 glycoproteins having both N-linked and O-linked glycans. Finally, an MD simulation was performed on sucrose non-covalently bound to a lectin.

Conformational properties of the GlcNAc acetamido group—NMR studies on GlcNAc allow for the ability of the force field to reproduce conformational sampling of this sugar in solution. Based on the potential energy scan for **M4** (Figure 3), the acetamido group of GlcNAc is anticipated to have three stable conformations. Using previously-developed Karplus equations for 3-bond *J*-coupling $^{3}J(H^{N}H^{2})$ between the protons on C2 and the amide nitrogen,[79] it is possible to compare the conformational properties of this moiety in aqueous MD simulations with NMR data. The relevant H-N-C2-H dihedral value θ can be related to the **M4** CT-N-C2-C3 dihedral scan data by the relationship $\theta_{\text{H-N-C-H}} = \theta_{\text{C-N-C-C}}$ - 60°, and yields three minima for $\theta_{\text{H-N-C-H}}$: one at 0° (*cis*) and two energetically-equivalent minima on either side of 180° (*trans*$^{-}$ and *trans*$^{+}$), with the *trans* minima 1.7 kcal/mol more stable than the *cis* minimum in the MM representation (Figure 3). $\theta_{\text{H-N-C-H}}$ values from 3 10-ns trajectories each of the α- and β-anomers of GlcNAc show very different behavior for the two anomers. In the case of the α-anomer, the population of sampled states is *trans*$^{-} \gg$ *trans*$^{+} >$ *cis*, where overall *trans* > *cis* as anticipated from the vacuum potential energy surface of the model compound, but with a clear energetic asymmetry introduced between *trans*$^{-}$ and *trans*$^{+}$ (Figure 7a). The deviation from the vacuum potential energy surface of **M4** is even more striking in the case of the β-anomer, in which the population of sampled states is *cis* ≫ *trans* (Figure 7b); it is worth emphasizing that all force field parameters are exactly the same for the two anomers. Interestingly, in cases where *trans* is undersampled relative to the **M4** surface, namely *trans*$^{+}$ for the α-anomer and both *trans*$^{-}$ and *trans*$^{+}$ for the β-anomer, the sampled *trans* conformations deviate from the **M4** ideal values of *trans*$^{-}$/*trans*$^{+}$ = +135°/−135° (dashed lines in Figure 7a, b).

Visualization of the trajectories did not point to any obvious stabilizing or destabilizing interactions as causing the difference between the acetamido *cis*/*trans* conformational preferences between the two anomers. However, the populations may be rationalized in the context of the rotational profiles of the acetamido group in the absence of electrostatic interactions (Figure 7c). Without electrostatic interactions, there is no possibility of intramolecular hydrogen bonding or electrostatic repulsion between the acetamido group on C2 and the hydroxyl groups attached to C1 and C3, and to a very rough approximation, this mimics the electrostatic shielding in aqueous solution. The energy of the β-anomer as a function of HN-N-C2-H2 has a global minimum in the *cis* conformation, with a broad local minimum of +1 kcal/mol at the *trans* conformation (Figure 7c). This energy surface explains the preference for the *cis* conformer along with sampling of the ideal *trans* conformation instead of the *trans*$^{+}$ and *trans*$^{-}$ conformations (Figure 7b). Likewise, for the α-anomer, the vacuum potential energy surface in the absence of electrostatic interactions mirrors the acetamido conformational sampling for aqueous α-GlcNAc (Figure 7a, c). In particular, the surface is no longer symmetric about HN-N-C2-H2 = 0°, because, unlike β-GlcNAc which has both C1 and C3 hydroxyls in equatorial configurations, the C1 hydroxyl is axial. As a result, the combination of bonded and LJ force field terms yields a global minimum at *trans*$^{-}$, which is the conformational state preferentially sampled by β-GlcNAc. Finally, to highlight the importance of electrostatics, it is worth noting that in the absence of electrostatic interactions, the *cis* H-N-C-H conformation for **M4** becomes the global minimum by 0.4 kcal/mol relative to the the *trans*$^{+/-}$ local minima, whereas using the full force field representation, the *trans*$^{+/-}$ conformations are 1.7 kcal/mol more stable than the *cis* (Figure 7d).

Mobli and Almond recently developed Karplus equations specifically for the α- and β-anomers of GlcNAc,[79] where $^{3}J(\text{H}^{\text{N}}\text{H}^{2})$ for the α-anomer is described by

$$J=9.56\cos^{2}(\theta)-1.62\cos(\theta)+0.69 \tag{4}$$

and for the β-anomer by

$$J=9.45\cos^{2}(\theta)-2.08\cos(\theta)+0.63 \tag{5}$$

Using the above equations to calculate ensemble average coupling values $<^{3}J>$ from the MD trajectories yields values of 7.0±0.1 and 7.7±0.4 Hz (where errors are 95% confidence intervals using the $<^{3}J>$ from each trajectory as an independent sample) for the α- and β-anomers respectively. In comparison, the values from NMR experiments are 8.9 Hz and 9.1 Hz.[79] Equation 4 for the α-anomer has maxima of 11.9 Hz at ±180° and 8.6 Hz at 0° and a minimum of 0.7 Hz at ±90°. Similarly, equation 5 for the β-anomer has maxima of 12.2 Hz at ±180° and 8.0 Hz at 0° and minima of 0.6 Hz at ±90°. Assuming the above QM-derived Karplus equations are appropriate, interpretation of the MD data is complicated by the fact that both the *cis* and *trans* conformations correspond to maxima in Equations 4 and 5. Thus, the underestimation of the ^3J values from the MD simulations relative to NMR may arise from local structural deviations from idealized *cis* or *trans* geometries, from inaccurate sampling of *cis* vs. *trans* conformational populations, or from a combination of these factors. Using a similar MD protocol with a different force field, the developers of the above equations computed $<^3J>$ values of 8.9 and 10.4 Hz for the two anomers.[79] While their computed value for the α-anomer exactly reproduces their NMR value, the computed β-anomer value is overestimated by nearly the same amount as it is underestimated in the present study. More importantly, in the previous work, the α-anomer exclusively sampled the *trans* conformation and the β-anomer preferentially sampled the *trans* conformation vs. *cis* by a factor of 9 to 1. This is in contrast to the present work, where the α-anomer does demonstrate some sampling of the *cis* conformation and where the β-anomer samples the *cis* conformation almost exclusively (Figure 7). With regard to the significant preference of the β-anomer for the *cis* conformation observed here, it is unlikely that it is due to kinetic trapping as all three trajectories were minimized, heated, and equilibrated with positional restraints on the monosaccharide atoms such that the acetamido group maintained a $trans^+$ geometry during these initial phases of the simulation. Thus, it remains an open question as to whether or not the β-GlcNAc acetamido group prefers the *cis* or *trans* conformation in aqueous solution, or some combination of the two, as all of these possibilities are consistent with the experimental ^3J value.

Conformational properties of oligomeric hyaluronan—The linear glycosaminolygcan hyaluronan, composed of GlcNAc and GlcUA residues, is an important component of the extracellular matrix, and plays structural as well as molecular-recognition roles in biology.[80,81] The component monosaccharides of hyaluronan are linked together in the repeating motif, …GlcUA-β(1→3)-GlcNAc-β(1→4)-GlcUA… (Figure 8a), and the linear polymer can reach molecular weights of over one million Daltons.[80,81] Recently, the aqueous structure of hyaluronan oligomers has been deduced via NMR spectroscopy as being close to a contracted left-handed 4-fold helix,[82] and the ϕ/ψ dihedral angles of this repeating structure are maintained when hyaluronan oligomers form complexes with the hyaluronan-binding domain (HABD) of the cell-surface protein CD44.[83] Using the hyaluronan 8-mer (HA8; Figure 8, *n*=4) coordinates from the A-form HABD:HA8 complex [PDB ID 2JCQ[83]], five 50-ns simulations of aqueous HA8 were performed. The crystallographically-unresolved coordinates of residue GlcUA8 were generated using force field default internal geometries for the monosaccharide, in which the hexopyranose ring is in the energetically-favored $^4C_1$ chair conformation, and GlcUA8-β(1→3)-GlcNAc7 ϕ/ψ dihedral angles geometries were constructed in accord with the NMR/crystallographic conformations. The HA8 molecule was centered in a truncated octahedron with sufficient water molecules so that it was at least 10 Å from the nearest edge of the system, overlapping water molecules were deleted, and 4 sodium ions were added at random positions to achieve a system of net neutral charge. The system was briefly minimized, heated, and equilibrated with positional restraints on HA8 atomic positions, and then the system was simulated with a timestep *dt* of 0.002 ps and a cutoff value *c* of 10 Å with only a harmonic restraining potential on the HA8 center-of-mass, with other MD details per “Methods.” Five independent trajectories were achieved by random assignment of velocities to the same system at the start of the five separate simulations.

HA8 contains 4 GlcNAc residues, allowing for the calculation of four separate ensemble-average $^3J(H^NH^2)$ values $<^3J>$, one for each residue, using the Karplus relationship in equation 5 for β-GlcNAc. Except for the reducing end residue, for which $<^3J>$ is underestimated by 1.6 Hz, the values computed using Equation 5 and the MD conformations are in excellent agreement with values for NMR experiments on hyaluronan oligomers (Table 4). The conformational properties of the acetamido group are seen to be sensitive to the local environment, as was the case for the α- vs. β-anomers of GlcNAc. In particular, the reducing-end GlcNAc acetamido group, which has its C1 hydroxyl in the β-anomeric configuration, primarily samples the *cis* conformation, whereas the other three GlcNAc acetamido residues preferentially sample the *trans* conformation, and this conformational difference is reflected in the slightly lower value of $<^3J>$ for GlcNAc1 (Table 4). The key difference between the local environments of the acetamido group for GlcNAc1 vs. the three other GlcNAc residues is the presence of the carboxyl group on the preceding GlcUA residue. As GlcNAc1 has no such neighboring residue, its acetamido group cannot act as a hydrogen bond donor to the neighboring carboxyl group, which when formed acts to stabilize the *trans* conformation.[82] Lacking this stabilization, the GlcNAc1 acetamido group shows conformational behavior similar to that described above for the β-anomer of GlcNAc monosaccharides.

The HA8 ϕ/ψ dihedral distributions from these simulations are consistent with NMR[82] and X-ray crystallographic[83] structures of oligomeric hyaluronan. There is one well-pronounced global free-energy minimum in each of the distributions for the GlcUA-β(1→3)-GlcNAc and GlcNAc-β(1→4)-GlcUA linkages, and for both linkage types, the location of the MD global free-energy minimum coincides with the experimental ϕ/ψ values (Figure 8b). While there are two additional local free-energy minima in the case of GlcUA-β(1→3)-GlcNAc and one additional minimum in the case of GlcUA-β(1→4)-GlcNAc, these are 2–3 kcal/mol higher in free energy relative to the global minimum and as such correspond to <5% of the sampled conformations. Therefore, the glycosidic linkage parameters that were previously developed using model compound hexopyranose disaccharide analogs lacking hydroxyl or hydroxymethyl groups and validated in the context of hexopyranoses[27] do demonstrate transferability to a polymer composed of hexopyranose derivatives.

Conformational properties of sialyl Lewis X—Sialyl Lewis X (sLe^X) is a tetrasaccharide carbohydrate moiety of particular importance in molecular recognition, and has roles in normal cell function such as leukocyte homing[84,85] as well as in disease states such as cancer[86] and chronic inflammatory conditions.[87] The sLe^X tetrasaccharide consists of Neu5Acα(2→3)Galβ(1→4)[Fucα(1→3)]GlcNAcβO-R, where "-R" indicates linkage of the reducing-end GlcNAc to another moiety. Given the three glycosidic linkages connecting the four component monosaccharides, a significant degree of conformational heterogeneity is, in principle, possible, and pioneering work combined carbohydrate synthesis, NMR spectroscopy, and computational studies to elucidate the structural and conformational properties of sLe^X.[88–92] Here, using glycosidic dihedral angles from one of these studies,[89] sLe^X was built, solvated in a truncated octahedron of water molecules extending at least 10 Å in all directions from the solute molecule and with a single neutralizing sodium counterion, briefly minimized and heated with harmonic restraining potentials on sLe^x heavy atom positions, and then, after removal of the positional restraints, simulated for 25 ns with a cutoff value c = 10 Å and a timestep dt = 0.002 ps. Five such simulations were done, with random assignment of initial velocities to generate different MD trajectories.

While two of the three sets of glycosidic linkage dihedral angles retained their initial values for all 25 ns in all five trajectories, the third set, namely Neu5Acα(2→3)Galβ, was not stable in any of the trajectories. Rather, the ϕ dihedral angle, defined as C1-C2-O$_{link}$-C3, rapidly relaxed in all cases from the initial value of +163° to one of −70° (Figure 9a). Likewise the ψ dihedral, defined as C2-O$_{link}$-C3-H3, rapidly relaxed in all cases from the initial value of −61° to one of 0° (Figure 9b). Interestingly, this relaxation was to the region of glycosidic ϕ/ψ space corresponding to that sampled in a previous combined NMR/MD study[92] that represents one of the pioneering studies on the conformational properties of sLe^X. Importantly, in that study, a 5 ns MD simulation was done *in vacuo* with a dielectric constant of 80 to account for solvent screening of electrostatic effects; this was in contrast to the reference study,[89] in which *in vacuo* molecular mechanics minimization was used for model refinement.

In the time since the initial work on isolated sLe^x, crystal structures of protein:sLe^x complexes have become available. Searches of the PDB using first the search term “slex” and second the search term “lewis x” yielded six such complexes as of January 2011 (PDB IDs 1G1R, 1G1T, 2KMB, 2R61, 2RDG, 2Z8L), all of which are noncovalent complexes of sLe^x with sLe^x-binding proteins. For comparison with the reference NMR ϕ/ψ angles, missing hydrogen atoms were assigned to the crystal sLe^x structures using force field geometries, and then each complete crystallographic sLe^x molecule was minimized with harmonic dihedral restraining potentials on heavy-atom ϕ/ψ dihedral angles. Using these optimized crystallographic models, ϕ/ψ angles were noted to be in the same region as the global free-energy minimum in the present MD study for the Neu5Acα(2→3)Galβ linkage, as well as for the other two glycosidic linkages in sLe^x (Figure 10). Thus, in accord with early NMR/MD work[92] and later crystallographic work, each of the three ϕ/ψ angles in sLe^x has a prominent global free-energy minimum, which is preserved in going from aqueous solution to a protein-bound state. Additionally, prior MD and NMR studies[92,93] have noted that the Neu5Acα(2→3)Galβ linkage in unbound sLe^x is not confined to the global minimum, consistent with the present results, whereas upon protein binding this glycosidic linkage does become conformationally constrained,[93] consistent with the later crystal structures of sLe^x:protein complexes.

Glycoprotein Systems—Four crystal structures obtained from the PDB database were used to study O- and N- linkages in a protein environment. These structures were chosen as they contain multiple glycosylations sites and have been solved at a high resolution (Table S14). The Reduce software[94] was used to place missing hydrogen positions and to choose optimal Asn and Gln sidechain amide and His sidechain ring orientations. Patch residues were used to incorporate disulfide bonds and the O- and N- glycosidic linkages between Ser/Thr or Asn residues and the relevant sugar units. Crystallographic water molecules, counter ions, and heteroatoms were included while building the crystal structure for simulations. Scripts obtained from CHARMM-GUI[95] and modified accordingly were used to set up the simulations, and the CHARMM software was used to solvate each system in a box of dimensions chosen so as to have 10 Å between the protein extremities and the edge of the solvent box (Table S14). Systems were neutralized by adding the appropriate number of counter ions and then energy minimized. MD equilibration involved a 100 ps NVT simulation in which harmonic restraints were applied on the protein and the carbohydrate moieties followed by a 200 ps NPT simulation in which all the restraints were removed. The equilibrated structures were then used for 16 ns production simulations that were performed using NAMD version 2.7b1.[96] the last 10 ns of which were used for subsequent analysis of ensemble properties.

Analysis of the trajectories revealed a common theme, namely that the glycan portions of the glycoprotein systems exhibit greater conformational variability than the protein portions. For all systems studied, the overall RMSDs of the complete glycoproteins remain lower than 3 Å for the entire simulation lengths (Figure S8a). On decomposing the overall RMSD into carbohydrate (Fig S8b) and protein components (Fig S8c), the carbohydrate regions demonstrate high flexibility with deviations as large as 8 Å in some cases, while the underlying protein regions remain very stable with RMSD always lower than 2 Å. The high RMSD for the carbohydrate regions is consistent with the high flexibility of carbohydrates, as observed in both NMR and crystallographic studies. In fact the high conformational variability of carbohydrate regions, combined with the variable glycosylation of identical sites in a sample of a given protein ("microheterogeneity"), is known to hinder crystallographic studies of glycoproteins, posing a barrier to progress for the accumulation of structural data on glycoproteins.[97,98]

Pooled data from the last 10 ns of the simulation trajectories were used to assess the flexibility of Asn sidechains conjugated to glycans as well as the flexibility of the associated glycosidic linkages. The key observations from this pooled data are that (1) the sampled conformations strongly overlap with crystallographically-observed conformations, suggesting correct placement of the molecular mechanics free-energy minima, and (2) there is greater flexibility associated with dihedral atoms exclusively in the Asn sidechain as compared to dihedral atoms involved in the glycosidic linkage. With regard to the glycosidic linkage O5-C1-Nδ-Cγ/C1-Nδ-Cγ-Cβ dihedrals, two well-defined minima are sampled in the simulations, and both minima are populated in the crystals (Figure 11a). Moving from the glycan toward the protein backbone, the C1-Nδ-Cγ-Cβ/Nδ-Cγ-Cβ-Cα dihedrals sample a narrow distribution in the C1-Nδ-Cγ-Cβ coordinate and a broad distribution in the Nδ-Cγ-Cβ-Cα Asn sidechain atoms, similar to that seen in crystal structures (Figure 11b). Finally, looking at sampling of the Nδ-Cγ-Cβ-Cα/Cγ-Cβ-Cα-N, wherein all 8 atoms belong to the Asn sidechain, a great deal of flexibility is seen, both in the simulations and in the crystal structures. Here, the Nδ-Cγ-Cβ-Cα dihedral confers high flexibility to the N-linkage. In addition to being consistent with the crystal structures considered here, the varying degrees of conformational flexibility in the simulation data are consistent with a survey of over 500 N-linked glycans in the PDB.[99] Of particular note is the Cγ-Cβ-Cα-N dihedral, which adopts three well-defined conformations with values of 60°, 180° and 300° corresponding to the *g*+, *anti*, and *g*− conformational states. The simulation probabilities are *g*− (23%), *anti* (64%), and *g*+ (13%), which compares favorably to probabilities from the latter survey of *g*− (18%), *anti* (50%) and *g*+ (32%).

Pooled data for O-linkages – namely O5-C1-O1-Cβ/C1-O1-Cβ-Cα and C1-O1-Cβ-Cα/O1-Cβ-Cα-N dihedral distributions – are presented in Figure 12. The type of O-linkage affected the flexibility of the C1-O1-Cβ-Cα dihedral, which was found to be flexible in Ser linkages but more restricted in Thr linkages, with the dihedral sampling conformations around +120° for α- Thr linkages and around +150° for β-Thr linkages. This latter pattern of sampling has also been observed in a combined NMR/MD study of model (α/β)Thr-O-GalNAc diamides.[100] The O1-Cβ-Cα-N dihedral is found to adopt three well defined conformations with values of −60°, ±180°, and +60° which correspond to the *g*−, *anti*, and *g*+ conformational states, which in turn influences the folding-back of the carbohydrate moiety onto the peptide backbone. Additional analysis of the individual linkages revealed O5-C1-O1-Cβ dihedral angle sampling in the region of +60° for the α anomers and −60° for the β anomers, consistent with the exo-anomeric effect seen in sugars.[101]

Lectin-sucrose non-covalent interactions—The designed chimeric cyanovirin-N homolog protein[102] is composed of two domains (A and B), each of which binds one sucrose molecule in sites well separated from each other. Since both X-ray and NMR structures of the complex have been solved,[102] this protein was chosen as a test case for non-covalent protein-carbohydrate interactions. Chain A was chosen out of the two very similar molecules resolved in the crystal asymmetric unit cell of PDB entry 3HP8.[102] The system preparation consisted of adding the 3 missing N-terminal residues to the protein using the MODELLER package,[103–106] followed by applying the Reduce software[94] to choose optimal Asn and Gln sidechain amide and His sidechain ring orientations and the CHARMM software[107] to add missing hydrogens and solvate the system in a rectangular box with dimensions 74 Å × 53.8 Å × 52 Å, chosen to have 10 Å between the protein extremities and the edge of the solvent box; the net system charge was made neutral by replacing four randomly chosen water molecules with sodium ions. The system was minimized, heated by periodic reassignment of velocities, and equilibrated for 50 ps, all with harmonic restraints on protein and sucrose atoms, after which the system was simulated for 21 ns without restraints, the last 20 ns of which was used to collect data for analysis.

During the 20ns simulation the sucrose molecule associated with the A-domain (SucA) and the sucrose molecule associated with the B-domain (SucB) remained bound to the shallow, surface exposed sites in domains A and B, respectively. Over the course of the simulation, both molecules sampled only a narrow range of glycosidic ϕ-dihedral values (Figure 13a,b). However, the much broader range of glycosidic ψ-dihedral values spanned by SucA points to greater flexibility of the sucrose molecule bound to domain A. In addition to the conformational region around $\psi = -60°$, which is populated by SucB, SucA also visits regions near $\psi=+60°$. These observations, particularly the alternate $\psi=+60°$ conformational basin populated by SucA, are consistent with the conformational behavior of sucrose in solution as studied previously.[108] This greater flexibility is mirrored in the higher RMSD of SucA compared to SucB (Figure 13c), despite both molecules remaining bound to their respective pockets. The higher flexibility of SucA observed in the simulation is consistent with experimental data for the system: more NMR resonances are affected in domain A as a result of sucrose binding than in domain B; the average of the SucA atoms crystallographic B-factors is 33.7 $Å^2$ vs. 18.4 $Å^2$ for SucB; and, though both are only weakly bound, the experimentally measured apparent binding affinity of SucA is lower than that of SucB (K_d = 15.2 and 7.3 mM, respectively).[102]

To better understand the differences between SucA and B, the probability of protein-sucrose hydrogen bonds (H-bonds) was analyzed. The presence of a hydrogen bond was based on a distance cutoff of 2.4 Å for the acceptor to hydrogen distance. The SucA glucose moiety preserves the H-bonds observed in the crystal structure between the C3 hydroxyl group and the N99 backbone amide nitrogen and Q98 carbonyl oxygen with 95% probability (Residue naming is per Ref [102]). H-bonds between the C4 hydroxyl group and S2 backbone carbonyl and S6 side chain hydroxyl are preserved with 100% probability. A third H-bond between the C2 hydroxyl and carbonyl oxygen of N99 is preserved with 95% probability. In contrast, the H-bonds formed by the fructose moiety are not as highly preserved as the ones by the glucose moiety, and this, combined with flexibility about the ψ-dihedral, accounts for the flexibility of bound SucA. In particular, the H-bond between the C3-hydroxyl and the N99 backbone carbonyl oxygen and N101 amide is preserved with 60% probability, whereas the one between the C4-hydroxyl group and R24 backbone carbonyl oxygen is preserved with only 40% occupancy.

As with SucA, H-bonds for the glucose moiety of SucB tend to be stable during the course of the simulation. In particular, H-bonds between the glucose C4-hydroxyl group and the N43 backbone amide and Q53 carbonyl oxygen are preserved with > 90% probability, and the C3-hydroxyl oxygen of the glucose moiety preserves a water mediated H-bond with Q53 backbone amide with 60% probability. And while the water mediated hydrogen bond between the C3-hydroxyl hydrogen and N54 side-chain carbonyl observed in the crystal is preserved with only 10% occupancy, the glucose ring nonetheless remains firmly bound in its crystallographic conformation. Additionally, unlike SucA, H-bonds observed in the crystal between the protein and the SucB fructose moiety are maintained. These include preservation of the hydrogen bonding involving C3-hydroxyl group and N43 carbonyl oxygen and D45 backbone amide with > 80% probability. The H-bond between C4 hydroxyl group and carbonyl oxygen of R81 is preserved with 100% probability. The preservation of the H-bond network between the fructose moiety of SucB and the protein is consistent with the lesser flexibility about the ψ dihedral for SucB (Figure 13b).

Based on these simulations, while binding of the glucose moiety is preserved for sucrose in both binding pockets, subtle structural differences in the binding pockets in the two domains yield a sucrose molecule bound to domain A with a higher degree of flexibility, consistent with NMR, crystallographic, and binding affinity data.[102] These results suggest that the interactions between the carbohydrate and protein aspects of the force field, as well as competition with solvent, are properly balanced, an outcome of the consistent approach used for the optimization of the nonbond parameters in the comprehensive CHARMM additive force field for biomolecules.

Conclusions

The present set of parameters is an important addition to the existing CHARMM carbohydrate force field as it enables the modeling of common eukaryotic glycans, including glycoproteins. The parametrization has in fact already shown its utility in studying such systems. In one case, simulations were undertaken on a series of compounds containing 5 different sugars and the dipeptides of Ser and Thr, yielding 14 molecules when the different anomers are taken into account (S. Mallajosyula and A.D. MacKerell, Jr. submitted for publication). For 8 of these molecules NMR experimental *J*-coupling and NOE solution data are available,[100,109,110] and there was overall excellent agreement between the experimental NMR observables and those calculated from simulations using the present force field. In another case, simulations were undertaken on the glycosaminoglycan polymer hyaluronan in a non-covalent complex with the hyaluronan-binding domain of the Type I transmembrane protein CD44, resulting in the description of two key monosaccharides in the polymer important for binding as well as a key residue in the protein involved in conformational switching of the hyaluronan-binding site.[111] Additional future directions of interest include evaluation of ring conformational equilibria including the complicated behavior of iduronate,[112,113] glycosidic conformational transitions that can occur on timescales longer than tens of nanoseconds,[114] and the force field description of sulfated and phosphorylated carbohydrates. With regard to this latter direction, work is underway both with regard to parametrization and application.

One consistent trend in the present work is the overestimation of crystal volumes for neutral compounds; this trend is not unexpected given similar results in CHARMM force field models for hexopyranose and furanose monosaccharides,[24,25] linear sugars and sugar alcohols,[26] and disaccharides.[27,28] One possible explanation is that the highly directional hydrogen bonding in the crystal environment is at odds with the parametrization protocol for hydroxyl groups, which targeted the molecular volumes and heats of vaporization of neat alcohols and therefore is, in a sense, a mean-field approach to developing transferable additive force field parameters. Current work on introducing electronic polarizability into the molecular mechanics framework may help to alleviate this limitation, ideally by yielding a force field where a single set of parameters can yield quantitative results in the gas phase, the crystalline environment, and both aqueous and organic solutions.[115]

Finally, it is worth noting that much of the present work is transferable to glycans linked to lipids,[116] which represent another major class of biomolecules – in addition to proteins – having covalent linkages to carbohydrates. The completion of work presently underway toward this aim will result in an optimized CHARMM additive force field capable of describing the vast majority of heterogeneous biomolecular systems known in eukaryotic biology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support from the NIH (GM070855, ADM) and University of New England College of Pharmacy start-up funds (OG) is acknowledged as is computational support from the Department of Defense and NPACI Alliance, and we thank Drs. Rich Pastor, Richard Venable and John Brady for helpful discussions.

References

1. Helenius A, Aebi M. Annu. Rev. Biochem. 2004; 73:1019. [PubMed: 15189166]
2. Hakomori S. Biochim. Biophys. Acta. 1999; 1473:247. [PubMed: 10580143]
3. Lue J, Hsu M, Yang D, Marx P, Chen Z, Cheng-Mayer C. J Virol. 2002; 76:10299. [PubMed: 12239306]
4. Viswanathan K, Chandrasekaran A, Srinivasan A, Raman R, Sasisekharan V, Sasisekharan R. Glycoconj J. 2010; 27:561. [PubMed: 20734133]
5. Glennon TM, Zheng YJ, Legrand SM, Shutzberg BA, Merz KM. J. Comput. Chem. 1994; 15:1019.
6. Woods RJ, Dwek RA, Edge CJ, Fraserreid B. J. Phys. Chem. 1995; 99:3832.
7. Ott KH, Meyer B. J. Comput. Chem. 1996; 17:1068.
8. Senderowitz H, Parish C, Still WC. J. Am. Chem. Soc. 1996; 118:2078.
9. Reiling S, Schlenkrich M, Brickmann J. J. Comput. Chem. 1996; 17:450.
10. Senderowitz H, Still WC. J Org Chem. 1997; 62:1427.
11. Durier V, Tristram F, Vergoten G. J. Mol. Struct.: THEOCHEM. 1997; 395:81.
12. Damm W, Frontera A, Tirado-Rives J, Jorgensen WL. J. Comput. Chem. 1997; 18:1955.
13. Momany FA, Willett JL. Carbohydr. Res. 2000; 326:194. [PubMed: 10903029]
14. Momany FA, Willett JL. Carbohydr. Res. 2000; 326:210. [PubMed: 10903030]
15. Basma M, Sundara S, Calgan D, Vernali T, Woods RJ. J. Comput. Chem. 2001; 22:1125. [PubMed: 17882310]
16. Kirschner KN, Woods RJ. Proc. Natl. Acad. Sci. U. S. A. 2001; 98:10541. [PubMed: 11526221]
17. Kuttel M, Brady JW, Naidoo KJ. J. Comput. Chem. 2002; 23:1236. [PubMed: 12210149]
18. Kony D, Damm W, Stoll S, van Gunsteren WF. J. Comput. Chem. 2002; 23:1416. [PubMed: 12370944]
19. Lii JH, Chen KH, Allinger NL. J. Comput. Chem. 2003; 24:1504. [PubMed: 12868113]
20. Lins RD, Hunenberger PH. J. Comput. Chem. 2005; 26:1400. [PubMed: 16035088]
21. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. J. Comput. Chem. 2008; 29:622. [PubMed: 17849372]
22. Hansen HS, Hunenberger PH. J. Comput. Chem. 2011; 32:998. [PubMed: 21387332]
23. Tessier MB, DeMarco ML, Yongye AB, Woods RJ. Mol. Simul. 2008; 34:349.
24. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, MacKerell AD Jr. J. Comput. Chem. 2008; 29:2543. [PubMed: 18470966]
25. Hatcher E, Guvench O, MacKerell AD Jr. J. Phys. Chem. B. 2009; 113:12466. [PubMed: 19694450]
26. Hatcher ER, Guvench O, MacKerell AD Jr. J. Chem. Theory. Comput. 2009; 5:1315. [PubMed: 20160980]
27. Guvench O, Hatcher E, Venable RM, Pastor RW, MacKerell AD Jr. J. Chem. Theory. Comput. 2009; 5:2353. [PubMed: 20161005]
28. Raman EP, Guvench O, MacKerell AD Jr. J. Phys. Chem. B. 2010; 114:12981. [PubMed: 20845956]
29. Hatcher E, Säwén E, Widmalm G, MacKerell AD Jr. J. Phys. Chem. B. 2011; 115:597. [PubMed: 21158455]
30. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. J. Phys. Chem. B. 1998; 102:3586.

31. MacKerell AD Jr, Feig M, Brooks CL III. J. Comput. Chem. 2004; 25:1400. [PubMed: 15185334]

32. Foloppe N, MacKerell AD Jr. J. Comput. Chem. 2000; 21:86.

33. MacKerell AD Jr, Banavali NK. J. Comput. Chem. 2000; 21:105.

34. Schlenkrich, M.; Brinkman, J.; MacKerell, AD., Jr; Karplus, M. An empirical potential energy function for phospholipids: criteria for parameter optimization and applications. In: Merz, KM.; Roux, B., editors. Membrane Structure and Dynamics. Boston: Birkhauser; 1996. p. 31

35. Feller SE, Gawrisch K, MacKerell AD Jr. J. Am. Chem. Soc. 2002; 124:318. [PubMed: 11782184]

36. Yin DX, MacKerell AD Jr. J. Comput. Chem. 1998; 19:334.

37. Feller SE, MacKerell AD Jr. J. Phys. Chem. B. 2000; 104:7510.

38. Klauda JB, Brooks BR, MacKerell AD Jr, Venable RM, Pastor RW. J. Phys. Chem. B. 2005; 109:5300. [PubMed: 16863197]

39. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr. J. Comput. Chem. 2010; 31:671. [PubMed: 19575467]

40. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. J. Comput. Chem. 1983; 4:187.

41. MacKerell, AD., Jr; Brooks, B.; Brooks, CL., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The energy function and its paramerization with an overview of the program. In: Schleyer, PvR; Allinger, NL.; Clark, T.; Gasteiger, J.; Kollman, PA.; Schaefer, HF., III; Schreiner, PR., editors. Encyclopedia of Computational Chemistry. Vol. Vol. 1. Chichester: John Wiley & Sons; 1998. p. 271

42. Guvench, O.; MacKerell, AD, Jr. Comparison of protein force fields for molecular dynamics simulations. In: Kukol, A., editor. Molecular Modeling of Proteins. New Jersey: Humana Press, Inc.; 2008. p. 63

43. MacKerell AD Jr, Wiórkiewicz-Kuczera J, Karplus M. J. Am. Chem. Soc. 1995; 117:11946.

44. Feller SE, Yin DX, Pastor RW, MacKerell AD Jr. Biophys. J. 1997; 73:2269. [PubMed: 9370424]

45. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I, MacKerell AD Jr, Pastor RW. J. Phys. Chem. B. 2010; 114:7830. [PubMed: 20496934]

46. Allen, MP.; Tildesley, DJ. Computer Simulation of Liquids. Oxford: Oxford University Press; 1987.

47. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J. Chem. Phys. 1983; 79:926.

48. Durell SR, Brooks BR, Ben-Naim A. J. Phys. Chem. 1994; 98:2198.

49. Ryckaert JP, Ciccotti G, Berendsen HJC. J. Comput. Phys. 1977; 23:327.

50. Steinbach PJ, Brooks BR. J. Comput. Chem. 1994; 15:667.

51. Darden T, York D, Pedersen L. J. Chem. Phys. 1993; 98:10089.

52. Hockney, RW. The potential calculation and some applications. In: Alder, B.; Fernbach, S.; Rotenberg, M., editors. Methods in Computational Physics. Vol. Vol. 9. New York: Academic Press; 1970. p. 136

53. Nosé S. Mol. Phys. 1984; 52:255.

54. Hoover WG. Phys. Rev. A. 1985; 31:1695. [PubMed: 9895674]

55. Feller SE, Zhang YH, Pastor RW, Brooks BR. J. Chem. Phys. 1995; 103:4613.

56. Allen FH. Acta Crystallogr. Sect. B-Struct. Sci. 2002; 58:380.

57. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA.; Vreven, T., Jr; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, K.; Kitao, O.; Nakai, H.; Klene, M.; Li, TW.; Knox, JE.; Hratchian, HP.; Cross, JB.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.;

Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Gaussian 03; Revision B.04 ed. Pittsburgh, PA: Gaussian, Inc.; 2003.

58. Møller C, Plesset MS. Phys. Rev. 1934; 46:618.

59. Dunning TH. J. Chem. Phys. 1989; 90:1007.

60. Hariharan PC, Pople JA. Theor. Chim. Acta. 1973; 28:213.

61. Guvench O, MacKerell AD Jr. J. Phys. Chem. A. 2006; 110:9934. [PubMed: 16898697]

62. Woodcock HL, Moran D, Pastor RW, MacKerell AD Jr, Brooks BR. Biophys. J. 2007; 93:1. [PubMed: 17554075]

63. Scott AP, Radom L. J. Phys. Chem. 1996; 100:16502.

64. Pulay P, Fogarasi G, Pang F, Boggs JE. J. Am. Chem. Soc. 1979; 101:2550.

65. Guvench O, MacKerell AD Jr. J. Mol. Model. 2008; 14:667. [PubMed: 18458967]

66. MacKerell AD Jr. J. Comput. Chem. 2004; 25:1584. [PubMed: 15264253]

67. MacKerell AD Jr, Karplus M. J. Phys. Chem. 1991; 95:10559.

68. Vorobyov IV, Anisimov VM, MacKerell AD Jr. J. Phys. Chem. B. 2005; 109:18988. [PubMed: 16853445]

69. Vorobyov I, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, MacKerell AD Jr. J. Chem. Theory. Comput. 2007; 3:1120.

70. Varki, A.; Cummings, RD.; Esko, JD.; Freeze, HH.; Stanley, P.; Bertozzi, CR.; Hart, GW.; Marilynn, EE., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

71. Lemke, TL.; Williams, DA., editors. Foye's Principles of Medicinal Chemistry. 6th ed.. Baltimore (MD): Lippincott, Williams, and Wilkins; 2008.

72. Varki A. Nature. 2007; 446:1023. [PubMed: 17460663]

73. Spiro RG. Glycobiology. 2002; 12:43R.

74. Strous GJ, Dekker J. Crit Rev Biochem Mol Biol. 1992; 27:57. [PubMed: 1727693]

75. Zachara NE, Hart GW. Chem. Rev. 2002; 102:431. [PubMed: 11841249]

76. Dwek RA. Chem. Rev. 1996; 96:683. [PubMed: 11848770]

77. Hart GW, Brew K, Grant GA, Bradshaw RA, Lennarz WJ. J. Biol. Chem. 1979; 254:9747. [PubMed: 489565]

78. Bause E. Biochem J. 1983; 209:331. [PubMed: 6847620]

79. Mobli M, Almond A. Org. Biomol. Chem. 2007; 5:2243. [PubMed: 17609755]

80. Toole BP. Nat. Rev. Cancer. 2004; 4:528. [PubMed: 15229478]

81. Almond A. Cell. Mol. Life Sci. 2007; 64:1591. [PubMed: 17502996]

82. Almond A, DeAngelis PL, Blundell CD. J. Mol. Biol. 2006; 358:1256. [PubMed: 16584748]

83. Banerji S, Wright AJ, Noble M, Mahoney DJ, Campbell ID, Day AJ, Jackson DG. Nat. Struct. Mol. Biol. 2007; 14:234. [PubMed: 17293874]

84. Kannagi R. Curr. Opin. Struct. Biol. 2002; 12:599. [PubMed: 12464311]

85. Sperandio M. FEBS J. 2006; 273:4377. [PubMed: 16956372]

86. Cazet A, Julien S, Bobowski M, Krzewinski-Recchi MA, Harduin-Lepers A, Groux-Degroote S, Delannoy P. Carbohydr. Res. 2010; 345:1377. [PubMed: 20231016]

87. Romano SJ. Treat Respir Med. 2005; 4:85. [PubMed: 15813660]

88. Berg EL, Robinson MK, Mansson O, Butcher EC, Magnani JL. J. Biol. Chem. 1991; 266:14869. [PubMed: 1714447]

89. Lin YC, Hummel CW, Huang DH, Ichikawa Y, Nicolaou KC, Wong CH. J. Am. Chem. Soc. 1992; 114:5452.

90. Ball GE, O'Neill RA, Schultz JE, Lowe JB, Weston BW, Nagy JO, Brown EG, Hobbs CJ, Bednarski MD. J. Am. Chem. Soc. 1992; 114:5449.

91. Ichikawa Y, Lin YC, Dumas DP, Shen GJ, Garcia-Junceda E, Williams MA, Bayer R, Ketcham C, Walker LE. J. Am. Chem. Soc. 1992; 114:9283.

92. Rutherford TJ, Spackman DG, Simpson PJ, Homans SW. Glycobiology. 1994; 4:59. [PubMed: 8186551]

93. Cooke RM, Hale RS, Lister SG, Shah G, Weir MP. Biochemistry. 1994; 33:10591. [PubMed: 7521209]

94. Word JM, Lovell SC, Richardson JS, Richardson DC. J. Mol. Biol. 1999; 285:1735. [PubMed: 9917408]

95. Jo S, Kim T, Iyer VG, Im W. J. Comput. Chem. 2008; 29:1859. [PubMed: 18351591]

96. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. J. Comput. Chem. 2005; 26:1781. [PubMed: 16222654]

97. Lutteke T, Frank M, von der Lieth CW. Carbohydr. Res. 2004; 339:1015. [PubMed: 15010309]

98. Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, Yu C, Boles KS, Evans EJ, Stuart DI, Dwek RA, Jones EY, Owens RJ, Davis SJ. Structure. 2007; 15:267. [PubMed: 17355862]

99. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Glycobiology. 2004; 14:103. [PubMed: 14514716]

100. Corzana F, Busto JH, Jimenez-Oses G, Garcia de Luis M, Asensio JL, Jimenez-Barbero J, Peregrina JM, Avenoza A. J. Am. Chem. Soc. 2007; 129:9458. [PubMed: 17616194]

101. Rao, VSR.; Qasba, PK.; Balaji, PV.; Chandrasekaran, R. Conformation of Carbohydrates. Amsterdam: Harwood Academic Publishers; 1998.

102. Koharudin LM, Furey W, Gronenborn AM. Proteins. 2009; 77:904. [PubMed: 19639634]

103. Sali A, Blundell TL. J. Mol. Biol. 1993; 234:779. [PubMed: 8254673]

104. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Annu Rev Biophys Biomol Struct. 2000; 29:291. [PubMed: 10940251]

105. Fiser A, Do RKG, Sali A. Protein Sci. 2000; 9:1753. [PubMed: 11045621]

106. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Chapter 5. Curr Protoc Bioinformatics. 2006 Unit 5 6.

107. Brooks BR, Brooks CL III, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. J. Comput. Chem. 2009; 30:1545. [PubMed: 19444816]

108. Raman EP, Guvench O, MacKerell AD Jr. J Phys Chem B. 114:12981. [PubMed: 20845956]

109. Corzana F, Busto JH, Engelsen SB, Jimenez-Barbero J, Asensio JL, Peregrina JM, Avenoza A. Chemistry. 2006; 12:7864. [PubMed: 16850514]

110. Fernandez-Tejada A, Corzana F, Busto JH, Jimenez-Oses G, Jimenez- Barbero J, Avenoza A, Peregrina JM. Chemistry. 2009; 15:7297. [PubMed: 19544521]

111. Jamison FW 2nd, Foster TJ, Barker JA, Hills RD Jr, Guvench O. J. Mol. Biol. 2011; 406:631. [PubMed: 21216252]

112. Babin V, Sagui C. J. Chem. Phys. 2010; 132:104108. [PubMed: 20232948]

113. Sattelle BM, Hansen SU, Gardiner J, Almond A. J. Am. Chem. Soc. 2010; 132:13132. [PubMed: 20809637]

114. Peric-Hassler L, Hansen HS, Baron R, Hunenberger PH. Carbohydr. Res. 2010; 345:1781. [PubMed: 20576257]

115. Lopes PE, Roux B, Mackerell AD Jr. Theor. Chem. Acc. 2009; 124:11. [PubMed: 20577578]

116. Abel S, Dupradeau FY, Raman EP, MacKerell AD Jr, Marchi M. J. Phys. Chem. B. 2011; 115:487. [PubMed: 21192681]

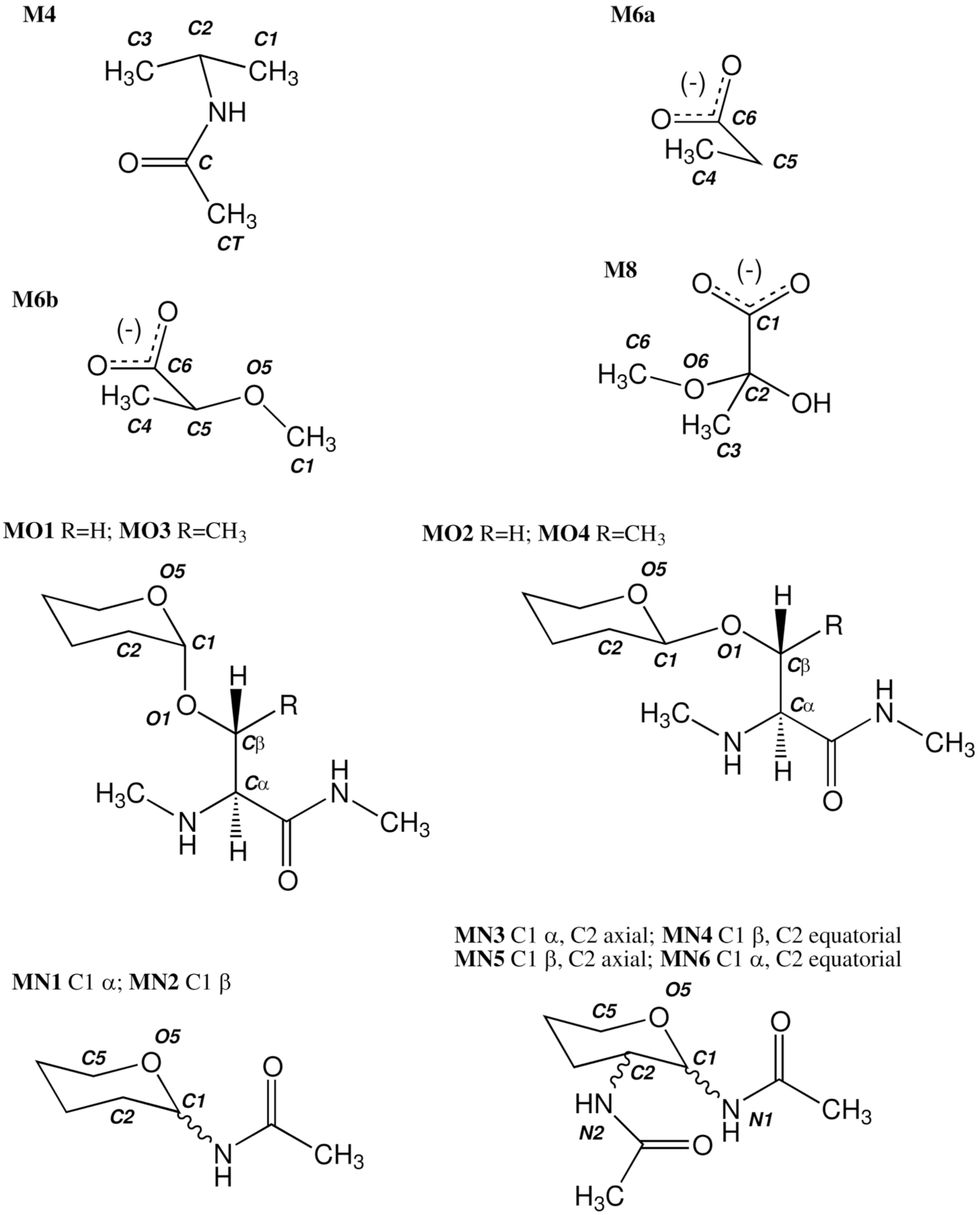

Figure 1.

Model compounds used to develop parameters for glucopyranose derivatives GlcNAc and GalNAc (**M4**), glucuronate and iduronate (**M6a** and **M6b**), and sialic acid (**M8**); to develop O-glycosidic linkage parameters involving Ser (**MO1** and **MO2**) and Thr (**MO3** and **MO4**) amino acid sidechains; and to parameterize N-linked glycosylation (**MN1-6**). Atom labels used in the text are in italics.

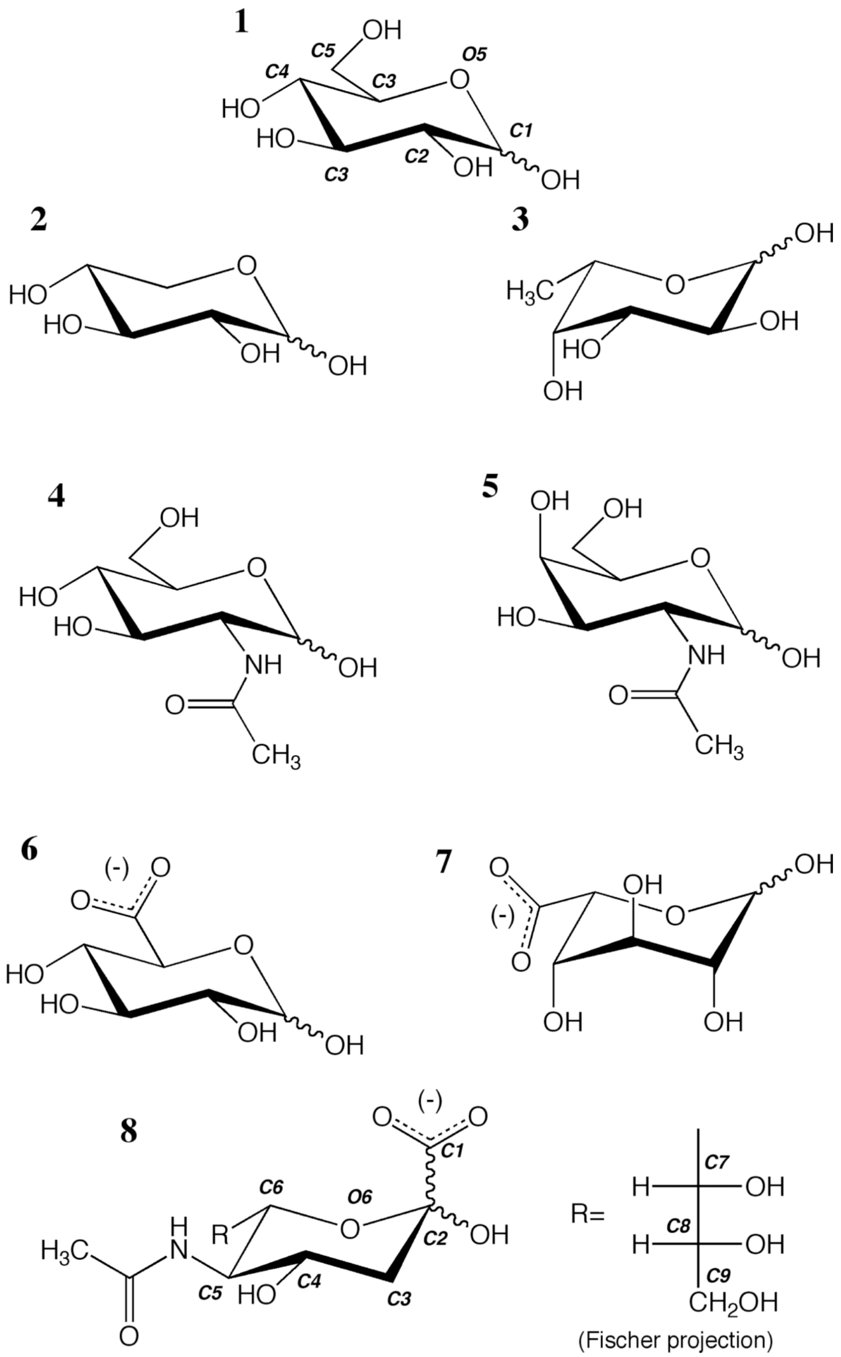

Figure 2.

Chemical structures of the pyranose form of glucose, D-glucopyranose (**1**) and related molecules D-xylose (**2**), L-fucose (**3**), *N*-acetyl-D-glucosamine (GlcNAc) (**4**), *N*-acetyl-D-galactosamine (GalNAc) (**5**), D-glucuronate (**6**), L-iduronate (**7**), *N*-acetyl-D-neuraminic acid (sialic acid) (**8**). Carbon atoms and the ring ether oxygen are labeled in glucose **1**; atom numbering is analogous for **2**–**7**. Sialic acid **8** carbon numbering begins at the carboxyl group. Hydroxyl and carboxyl oxygen atoms derive their numbering from the carbon atom to which they are attached.

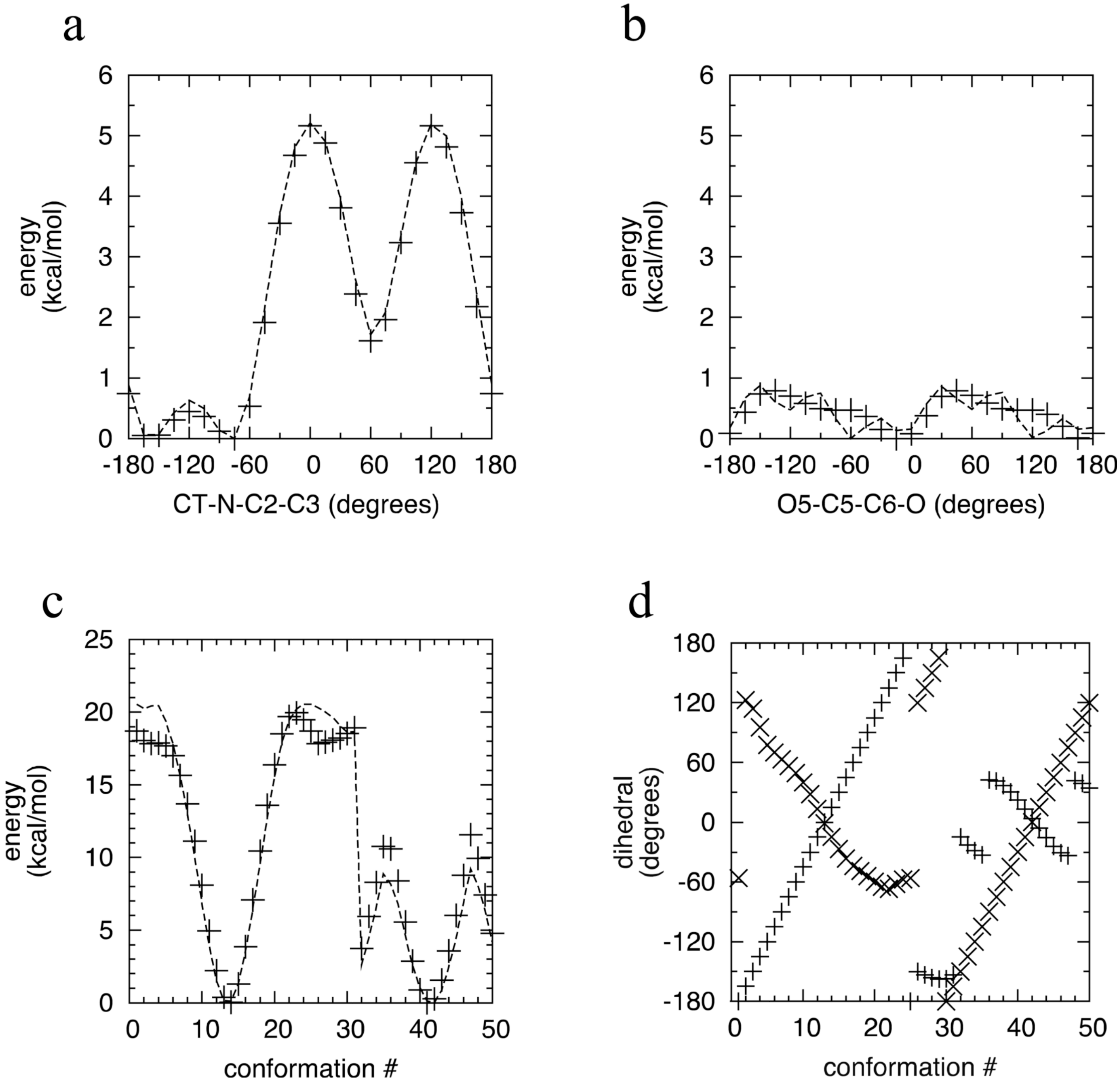

Figure 3.

Relaxed potential energy scans of (a) the isopropyl group in compound **M4** and (b) the carboxyl group in compound **M6b,** (c) conformational energies and (d) hydroxyl and carboxyl dihedral values for compound **M8**.

(a–c): MP2/cc-pVTZ energies are represented as crosses and MM energies as a dashed line; (d): hydroxyl dihedral values are represented as crosses and carboxyl dihedral values as x’s.

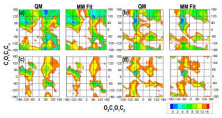

Figure 4.

2D-Dihedral potential energy scans on the O5-C1-O1-Cβ (ϕ_s) and C1-O1-Cβ-Cα (ψ_s) dihedrals of the model compounds (a) **MO1**, (b) **MO2**, (c) **MO3**, and (d) **MO4** representative of the O-glycan linkages. Energies are in kcal/mol, with contours every 2 kcal/mol. Only energies below 14 kcal/mol have been plotted for the sake of clarity.

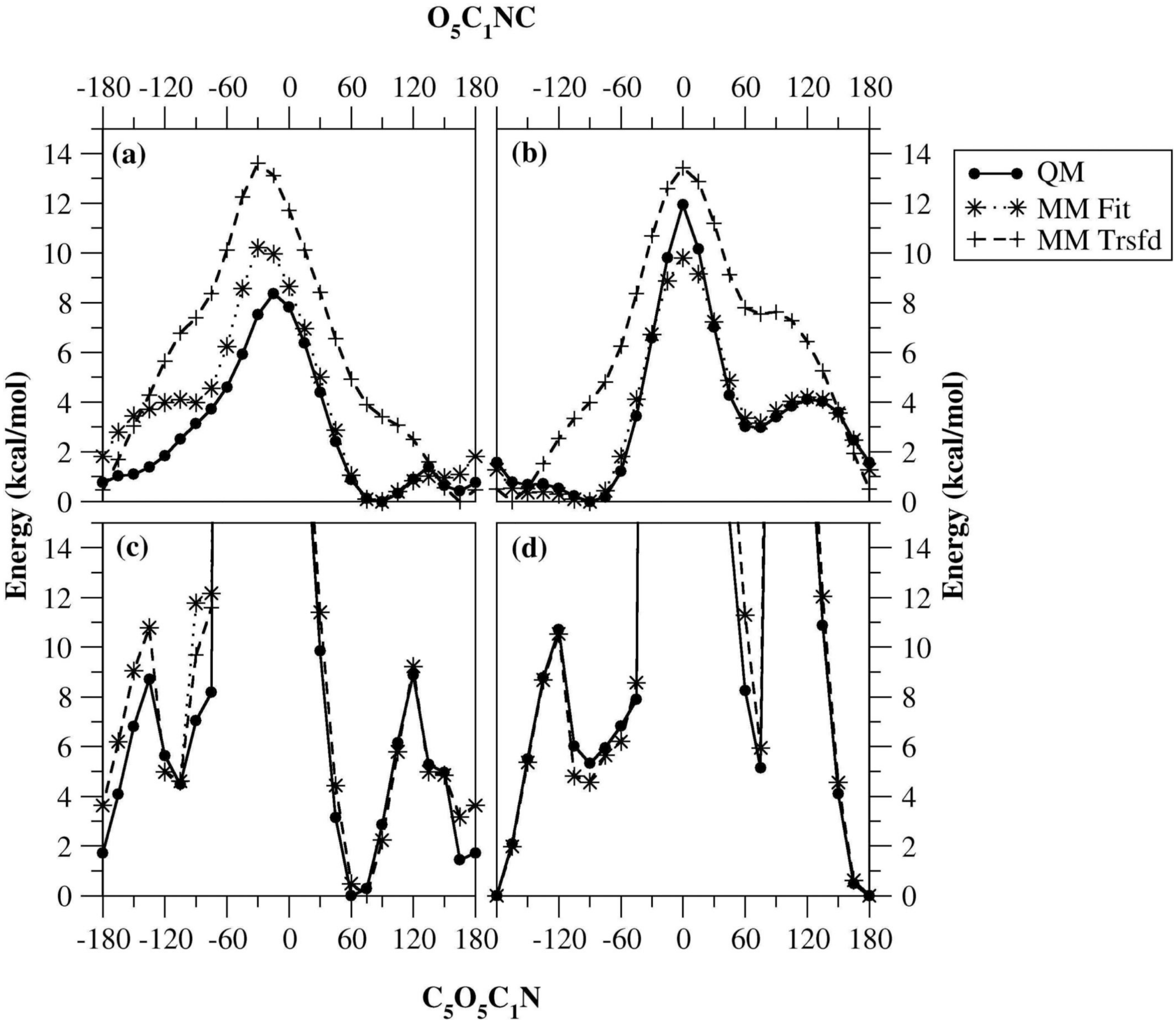

Figure 5.

Dihedral potential energy scans for model compounds **MN1** and **MN2** representative of the N-glycan linkages. (a) **MN1** O5-C1-N-C, (b) **MN2** O5-C1-N-C, (c) **MN1** C5-O5-C1-N, (d) **MN2** C5-O5-C1-N.

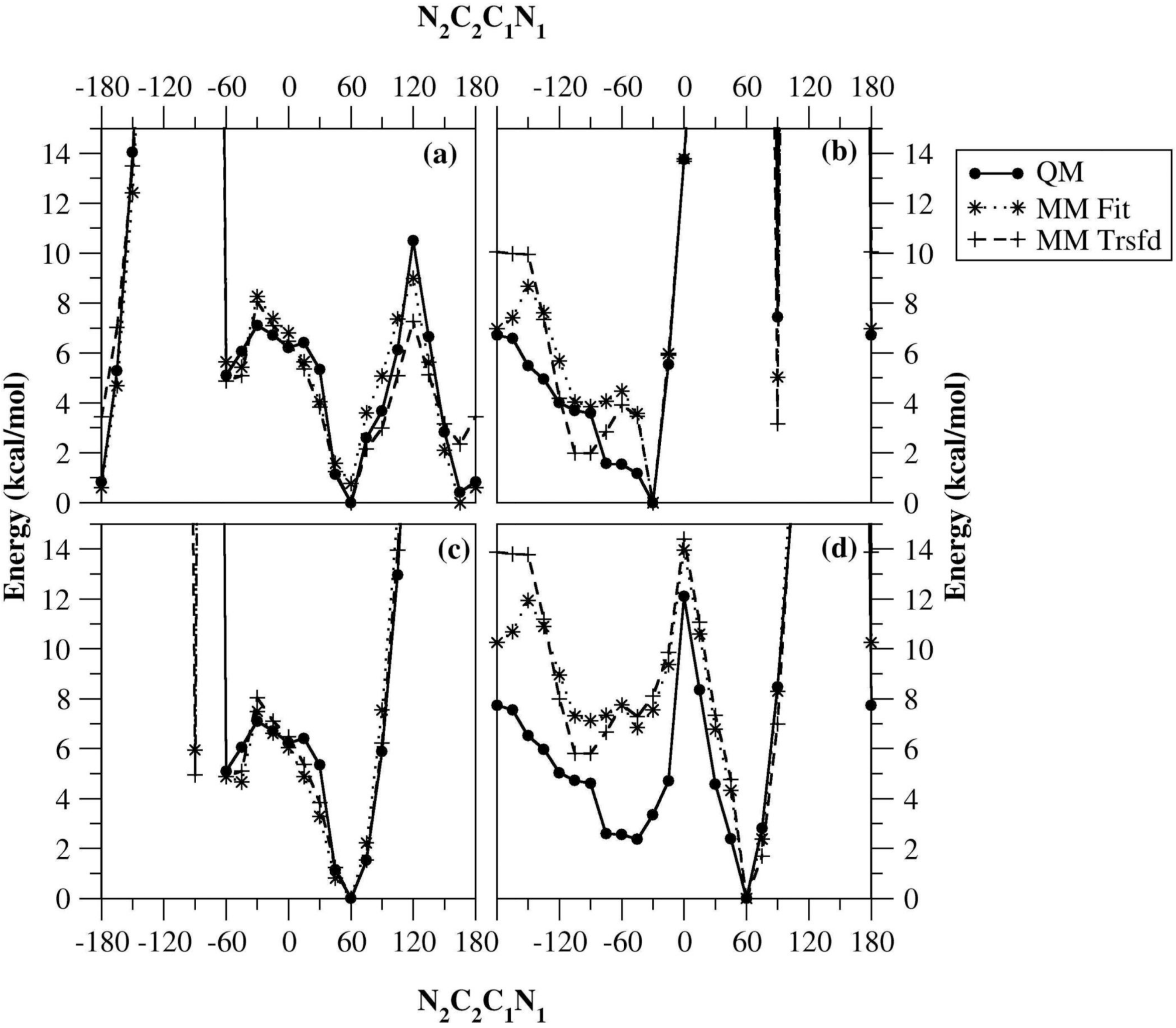

Figure 6.

Dihedral potential energy scans about the N2-C2-C1-N1 dihedral for model compounds (a) **MN3**, (b) **MN4**, (c) **MN5**, and (d) **MN6**.

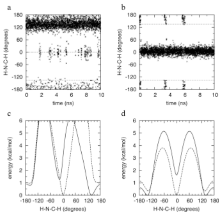

Figure 7.

Conformational properties of the GlcNAc acetamido group. $\theta_{\text{H-N-C-H}}$ vs. time is shown for the α- (a) and β-anomers (b) in three independent aqueous MD simulation trajectories (+, x, *) for each anomer, along with dashed lines at −135°/0°/+135°. Also shown are (c) the gas-phase potential energy surfaces for $\theta_{\text{H-N-C-H}}$ in α- (solid line) and β-GlcNAc (dashed line) in the absence of electrostatic interactions, and (d) the gas-phase potential energy surfaces for $\theta_{\text{H-N-C-H}}$ for **M4** with (solid line) and without (dashed line) electrostatic interactions.

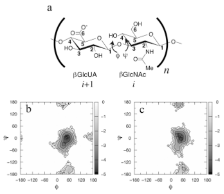

Figure 8.

Conformational properties of the glycosidic linkages in hyaluronan. The chemical structure of hyaluronan is shown (a), demonstrating the GlcUA-β(1→3)-GlcNAc linkage; to either side of this linkage are GlcNAc-β(1→4)-GlcUA linkages, and monosaccharide residues are numbered starting at the reducing end of the polymer. Boltzmann-inverted hyaluronan 8-mer ϕ/ψ probability distributions for GlcUA-β(1→3)-GlcNAc (b) and GlcNAc-β(1→4)-GlcUA glycosidic linkages (c) are shown (in kcal/mol; contour lines are every 1 kcal/mol). ϕ = H1-C1-O_{link}-Cx and ψ = C1-O_{link}-Cx-Hx, and data have been aggregated across all linkages of the same type in the 8-mer and across five independent 50-ns simulations. ϕ/ψ angles from NMR[81] and X-ray crystallographic[83] structures of hyaluronan are shown as x's and open squares, respectively.

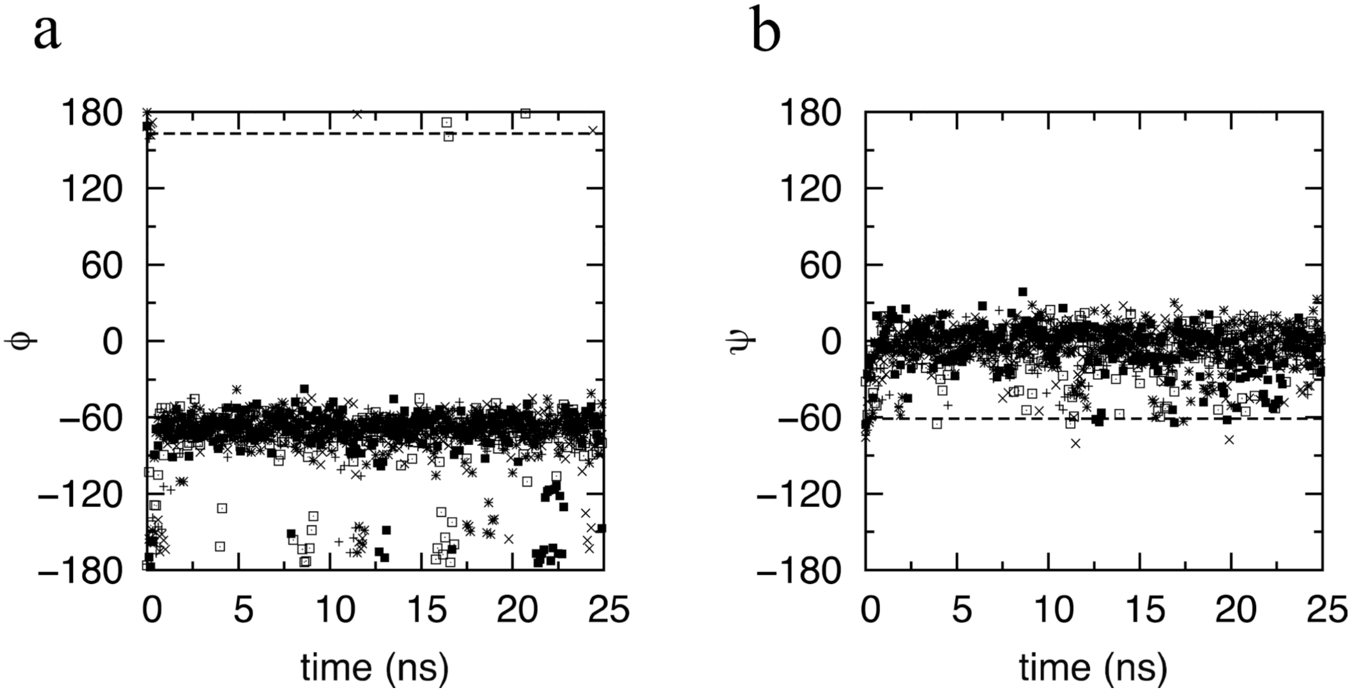

Figure 9.

Time-dependent conformational properties of the Neu5Acα(2→3)Galβ glycosidic linkage in sLe^{X}. Data are shown for the ϕ dihedral angle (a; defined as C1-C2-O$_{link}$-C3) and the ψ dihedral (b; as C2-O$_{link}$-C3-H3) for all five 25-ns simulations. Dashed lines indicate values from the reference NMR structure used to seed the simulations.

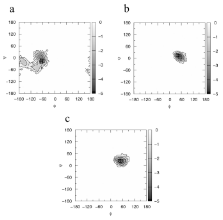

Figure 10.

Conformational properties of the sLe^X glycosidic linkages Neu5Acα(2→3)Galβ (a; ϕ/ψ = C1-C2-O$_{link}$-C3/C2-Olink-C3-H3), Galβ(1→4)GlcNAc (b; ϕ/ψ = H1-C1-Olink-C4/C1-Olink-C4-H4), and Fucα (1→3)GlcNAc (c; ϕ/ψ = H1-C1-O$_{link}$-C3/C1-O$_{link}$-C3-H3). MD data have been aggregated from the 5 ns – 25 ns intervals of five separate simulations and Boltzmann-inverted, with contours every 1 kcal/mol. The 0 ns – 5ns interval from all simulations was excluded to minimize sampling artifacts arising from time-dependent relaxation of the starting conformation. X's indicate values from the reference NMR structure used to seed the simulations, and squares are values from sLe^X:protein noncovalent complexes from the PDB (see text).

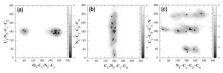

Figure 11.
Boltzmann-inverted glycosidic dihedral angle distributions associated with the N-glycan linkage. (a) O5-C1-Nδ-Cγ/C1-Nδ-Cγ-Cβ distribution. (b) C1-Nδ-Cγ-Cβ/Nδ-Cγ-Cβ-Cα distribution and (c) Nδ-Cγ-Cβ-Cα/Cγ-Cβ-Cα-N distribution. Squares indicate the values observed in the crystallographic structure.

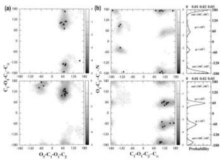

Figure 12.

Boltzmann-inverted glycosidic dihedral angle distributions associated with O-glycan linkages. (a) O5-C1-O1-Cβ/C1-O1-Cβ-Cα distributions and (b) C1-O1-Cβ-Cα/O1–Cβ-Cα-N distributions are shown, with data collected from the Ser O-linkages in the top panel and from Thr O-linkages in the lower panel respectively. Squares indicate the values observed for the Ser and Thr O-linkages in the crystallographic structures respectively. The side panel of (b) contains probability distributions associated with the O1-Cβ-Cα-N dihedral angle.

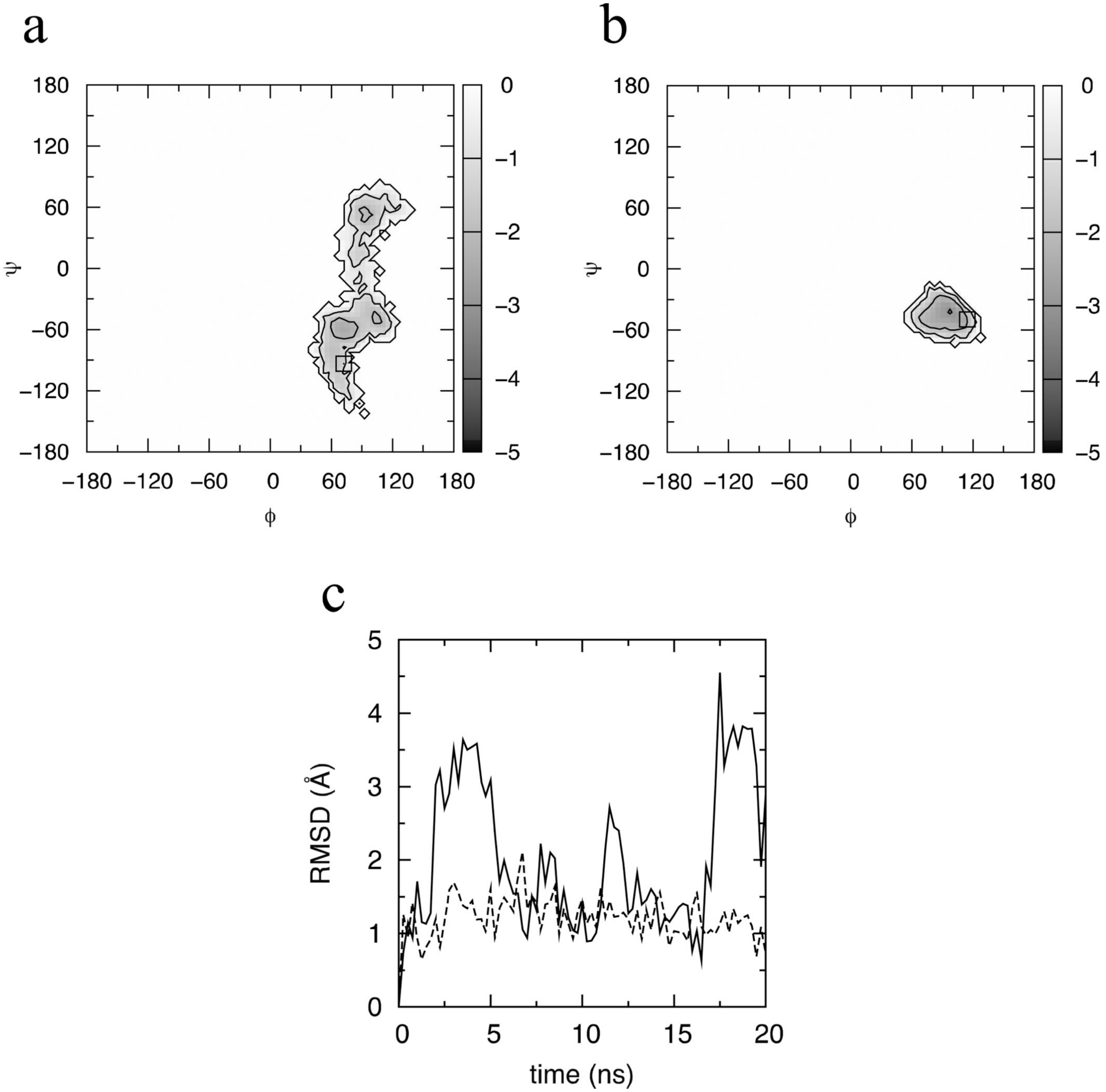

Figure 13.

Glycosidic ϕ/ψ ($^gO_{ring}$-$^gC1$-O_{link}-$^fC2$/$^gC1$-O_{link}-$^fC2$-$^fO_{ring}$, where the superscripts “g” and “f” indicate the glucose and the fructose groups, respectively) dihedral angle distributions and RMSD values for sucrose bound noncovalently to the designed chimeric cyanovirin-N homolog protein. Boltzmann-inverted ϕ/ψ distributions are shown for the sucrose molecule bound to the A-domain (a) and the sucrose molecule bound to the B-domain (b) (contours every 1 kcal/mol), as well as heavy-atom RMSD values for the A-domain and B-domain sucrose molecules with respect to the crystallographic coordinates (c; solid line and dashed line, respectively). Squares indicate the values observed in the crystallographic structure.

Table 1

MD simulation details

System	c (Å)	dt (fs)	length (ns)	snapshot frequency (ps^{-1})	# of simulations	Simulation software
Carbohydrate crystals[a]	12	1	4	1	1	CHARMM[40,107]
Aqueous systems						
α-GlcNAc	12	1	10	10	3	CHARMM
β-GlcNAc	12	1	10	10	3	CHARMM
oligomeric hyaluronan	10	2	50	10	5	CHARMM
sialyl Lewis X	10	2	25	10	5	CHARMM
glycoproteins[b]	12	2	16	5	1	NAMD96
lectin:sucrose[c]	12	2	20	5	1	NAMD96

[a] Full listing in Table 2.

[b] Full listing in Table S14.

[c] Langevin thermostating[46] was used instead of Nosé-Hoover thermostating, and the system was a rectangular prism.

Table 2

Crystalline unit cell geometries and volumes.[a]

CSD code / molecule name	*A* (Å)			*B* (Å)			*C* (Å)			β (degrees)[b]			volume (Å^3)		
	expt	MD	% error[c]	expt	MD	% error	expt	MD	% error	expt	MD	% error	expt	MD	% error
truncated derivatives															
XYLOSE / xylose	9.25	9.25	0.0	12.67	13.29	4.9	5.64	5.55	−1.6	90	90		660.0	681.0	3.2
ALFUCO / fucose	14.48	14.63	1.1	7.60	7.95	4.6	6.68	6.43	−3.6	90	90		733.8	747.6	1.9
RHAMAH01 / rhamnose	7.91	7.97	0.8	7.92	8.16	3.0	6.67	6.76	1.3	95.6	96.2	0.6	415.9	436.5	5.0
N-acetylamines															
ACGLUA11 / GlcNAc	11.57	11.75	1.5	4.85	4.84	−0.2	9.74	10.05	3.2	116.7	117.3	0.5	488.2	507.3	3.9
AGALAM10 / GalNAc	9.16	9.79	6.9	6.32	5.84	−7.6	9.21	9.79	6.4	107.9	109.1	1.2	507.3	527.9	4.1
NACMAN10 / ManNAc	7.56	7.65	1.2	7.73	7.76	0.3	18.61	19.10	2.6	90	90		1088.1	1133.2	4.1
Acids															
NABDGC / glucuronate	9.21	9.80	6.5	7.01	7.02	0.2	7.38	7.06	−4.3	96.8	97.7	1.4	472.5	479.3	1.4
CANAGL10 / galacturonate	13.50	13.19	−2.3	13.50	13.19	−2.3	9.66	9.88	2.3	90	90		1523.7	1489.5	−2.2
KEMYAC / Neu5Ac	7.50	7.50	0.0	7.50	7.50	0.0	29.36	29.08	−1.0	90	90		1652.1	1635.0	−1.0
O-glycans[d]															
COSHEX	7.73	7.79	0.7	8.63	8.61	−0.2	9.94	10.13	1.9	112.5	111.0	−1.3	612.8	634.0	3.5
N-glycans[d]															
AVUVES	6.63	6.90	4.0	19.49	20.63	5.9	8.46	8.02	−5.2	90	90		1093.6	1141.5	4.4
AVUVIW	7.47	7.53	0.8	8.70	8.90	2.2	14.98	15.35	2.5	90	90		973.6	1028.0	5.6
AVUVOC	6.64	6.73	1.4	8.62	8.56	−0.7	15.81	16.28	3.0	90	90		904.3	936.8	3.6
RESJEE	7.86	8.05	2.4	9.42	9.43	0.1	14.01	14.01	0.0	90	90		1038.1	1063.1	2.4
CAKFAV	4.94	4.93	−0.1	7.88	7.94	0.8	17.67	18.50	4.7	91.4	73.5	−19.6	687.8	694.6	1.0
ASGPRS	4.93	4.86	−1.5	24.22	24.46	1.0	7.79	7.76	−0.3	97.7	100.4	2.7	921.8	906.1	−1.7
BEHPIN	4.94	4.92	−0.3	24.26	23.88	−1.6	7.77	7.76	−0.2	97.7	98.1	0.4	922.8	901.9	−2.3
BEHPOT	4.94	4.82	−2.4	16.68	16.31	−2.2	8.08	8.59	6.3	96.1	97.3	1.2	662.0	670.0	1.2

[a] MD values are 4-ns averages; 95% confidence intervals for *A*, *B*, and *C* are < 0.02 Å, for β are < 0.2 degrees, and for volumes are < 0.5 Å^3.

[b] Constrained to 90° in the simulation if equal to 90° in the experimental crystal, otherwise allowed to vary independently during the simulation.

[c] (MD-expt)/expt * 100%.

[d] Please refer to text for full compound names.

Table 3

Solute:water pair interaction energies and distances for model compounds **M4**, **M6B**, and **M8**.

water orientation[a]	**energy (kcal/mol)**			**distance (Å)**		
	QM[b]	**MM**	**MM-QM**	**QM-0.20**[b]	**MM**	**MM-QM**
M4						
a	−5.39	−5.86	−0.48	1.97	1.95	−0.02
b	−5.92	−6.05	−0.13	1.94	1.94	0.00
c	−6.74	−6.84	−0.10	1.81	1.76	−0.05
d	−8.44	−7.23	1.21	1.78	1.77	−0.01
average			0.13			−0.02
standard deviation			0.74			0.02
M6B						
a	−14.67	−15.77	−1.09	1.61	1.64	0.03
b	−15.18	−15.01	0.17	2.27	2.22	−0.05
c	−6.08	−2.46	3.61	2.19	2.39	0.19
d	−15.89	−17.64	−1.75	1.85	1.72	−0.13
e	−10.48	−11.05	−0.57	1.77	1.76	−0.01
f	−7.4	−6.38	1.02	1.87	1.84	−0.03
average			0.23			0.00
standard deviation			1.92			0.10
M8						
a	−12.86	−13.23	−0.37	1.65	1.69	0.03
b	−13.47	−13.56	−0.10	2.30	2.23	−0.08
c	−5.27	−2.93	2.34	1.97	2.01	0.04
e	−8.92	−9.1	−0.18	1.86	1.85	−0.02
f	−6.59	−5.7	0.90	1.97	1.93	−0.04
g	−9.78	−9.56	0.22	1.73	1.82	0.10
h	0.08	−0.7	−0.78	2.10	2.29	0.19
i	0.23	−0.74	−0.97	2.16	2.28	0.12

water orientation[a]	energy (kcal/mol)			distance (Å)		
	QM[b]	MM	MM-QM	QM-0.20[b]	MM	MM-QM
average			0.13			0.04
standard deviation			1.06			0.09

[a] Molecular geometries are as illustrated in Figure S2, S4, and S6 of the supporting information.

[b] HF/6-31G(d) target energies have been scaled by 1.16 for the neutral compounds (**M4** and **M6b**) but not for the charged compound (**M8**) and distances have been shortened by 0.20 Å.

Table 4

Comparison of hyaluronan oligomer acetamido $^{3}J(\mathrm{H}^{\mathrm{N}}\mathrm{H}^{2})$ values (Hz) from simulations with NMR experiments.

	experiment[82]	**experiment**[82]	**MD**[a]
	HA4	**HA6**	**HA8**
reducing end	9.8	9.7	8.1±0.1 (85%±3%)
core[b]	n/a	9.8	9.9±1.3 (17%±46%)
	n/a	n/	10.3±0.1 (1%±3%)
non-reducing end	9.7	9.7	10.1±0.7 (10%±29%)

[a] MD $^{3}J(\mathrm{H}^{\mathrm{N}}\mathrm{H}^{2})$ values are calculated as the average $<^{3}J>$, where $<^{3}J>$ is the ensemble-average $^{3}J(\mathrm{H}^{\mathrm{N}}\mathrm{H}^{2})$ in one 50-ns MD simulation and there are five $<^{3}J>$'s, one from each simulation. Values in parentheses are the average % of conformations sampled that are *cis*. Error values are 95% confidence intervals calculated as 2.78*(average($<^{3}J>$ or % cis))/sqrt(5). % *trans* = 1 - % *cis* and confidence intervals for % *trans* are identical to those for % *cis*.

[b] Refers to GlcNAc residues that are neither at the reducing end (Figure 15, residue i=1) nor at the non-reducing end (Figure 15, residue $i=2n-1$). In the case of HA8, these are listed in order from the reducing end to the non-reducing end.