Published in final edited form as:

J Chem Theory Comput. 2018 June 12; 14(6): 3132–3143. doi:10.1021/acs.jctc.8b00175.

CHARMM Drude Polarizable Force Field for Glycosidic Linkages Involving Pyranoses and Furanoses

Asaminew H. Aytenfisu^{1,§}, Mingjun Yang^{1,2,§}, and Alexander D. MacKerell Jr.^{1,*}

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

²XtalPi Inc., Shennan Road 6025, Futian District, Shenzhen, China

Abstract

We present an extension of the CHARMM Drude polarizable force field to enable modeling of polysaccharides containing pyranose and furanose monosaccharides. The new force field parameters encompass 1 $\leftrightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4 and 1 $\rightarrow$ 6 pyranose-furanose linkages, 2 $\rightarrow$ 1 and 2 $\rightarrow$ 6 furanose-furanose linkages, 2 $\rightarrow$ 2, 2 $\rightarrow$ 3 and 2 $\rightarrow$ 4 furanose-pyranose and 1 $\leftrightarrow$ 1, 1 $\rightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4 and 1 $\rightarrow$ 6 pyranose-pyranose linkages. For the glycosidic linkages both simple model compounds and the full disaccharides with methylation at the reducing end were used for parameter optimization. The model compounds were chosen to be monomers or glycosidic-linked dimers of tetrahydropyran (THP) and tetrahydrofuran (THF). Target data for optimization included one- and two-dimensional potential energy scans of ω and the Φ/Ψ glycosidic dihedral angles in the model compounds and full disaccharides computed by quantum mechanical (QM) RIMP2/cc-pVQZ single point energies on MP2/6-31G(d) optimized structures. Also included in the target data are extensive sets of QM gas phase monohydrate water-saccharide interactions, dipole moments, and molecular polarizabilities for both model compounds and full disaccharides. The resulting polarizable model is shown to be in good agreement with a range of QM data, offering a significant improvement over the additive CHARMM36 carbohydrate force field, as well as experimental data including crystal structures and conformational properties of disaccharides and a trisaccharide in aqueous solution.

Graphical Abstract

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

§The authors contributed equally to this work

Author Contributions: ADM conceived the work. MY and AA performed the calculations. MY, AA and ADM performed the analysis. MY, AA, and ADM wrote and revised the manuscript.

Competing Financial Interests Statement: ADM Jr. is cofounder and CSO of SilcsBio LLC.

Supporting Information

Supporting information includes details of the parametrization of the model compound along with the following tables and figures. Tables include the list of crystals used for parameter validation (Table S1), list of model compounds and associated dihedrals (Table S2–3); summary of optimized QM and MM geometries (Table S4), summary of target data used in dihedral fitting of pyranose-furanose and pyranose-pyranose full disaccharides (Table S5–6), comparison of crystalline internal geometries (Table S7A–B), summary of crystal unit cell parameters (Table S8), and RMSE of relative conformational energy between QM and MM (Table S9–10). Figures include examples of water-disaccharide interaction orientations (Figure S1), images of small model compounds (Figure S2), selected dihedral conformational relative potential energy plots (Figure S3–4), Φ/Ψ 2D conformational relative QM and MM potential energy surfaces for the model compounds included in Figures 1 and 2 (Figure S5A–F), and 2D furanose pucker relative potential energy surfaces (Figure S6). This material is available free of charge via the Internet at <http://pubs.acs.org>.

Keywords

CHARMM; MD simulations; carbohydrates; molecular mechanics; disaccharide; polysaccharide; classical Drude oscillator; electronic polarization

Introduction

Carbohydrates play an important role in biology due to their central role in a range of biological functions including as fuels, in energy storage, as metabolic intermediates, in structural roles and participating in molecular recognition.^{1–7} However, the structural characterization of carbohydrates via experimental method is challenging and progress lags behind such studies of proteins and nucleic acids. To supplement experimental methods theoretical approaches based on both quantum mechanics and empirical force fields offer great potential. For example, NMR methods in conjunction with molecular dynamics simulations have yielded insights into the structural properties of di-, tri- and polysaccharides, including in vaccine epitopes^{8–10} and the HIV Envelope.^{11, 12} A central observation from these studies is the diverse conformational heterogeneity accessible to polysaccharides, the property that leads to challenges for experimental structure determination methods while being well suited for force field methods.

While the utility of empirical force fields for structural studies of carbohydrates is increasing, limitations in force fields, the most notable being the additive or fixed charge approximation, potentially limit their overall accuracy.^{13–27} An important step toward overcoming this are efforts to move beyond the fixed charge force approximation by extending force fields to include explicit treatment of electronic polarization. While progress on polarizable force fields for biological molecules has been made on several fronts,^{28–35} with respect to carbohydrates progress has been largely limited to the CHEQ fluctuating charge model³⁶ and the Drude polarizable force field, with models for polyalcohols, hexopyranose and furanose monosaccharides, and acyclic monosaccharides being presented for the latter.^{37–39}

The classical Drude oscillator model accounts for electronic polarization by introducing a virtual point charge (the Drude oscillator or particle) that is linked to non-hydrogen atoms via a harmonic spring, in addition to a number of other extensions as recently presented.⁴⁰ Notable is the ability of the polarizable Drude carbohydrate model to treat cooperative intramolecular hydrogen bonding between three or more vicinal hydroxyl groups as is often present in carbohydrates.³⁹

The present study focuses on extending the Drude polarizable force field to polysaccharides containing furanoses and pyranoses. These encompass a wide range of carbohydrates from sucrose to glycans occurring on cell surfaces and glycoproteins. The optimized polarizable force field parameters encompass 1 $\leftrightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4 and 1 $\rightarrow$ 6 pyranose-furanose linkages, 2 $\rightarrow$ 2, 2 $\rightarrow$ 3 and 2 $\rightarrow$ 4 furanose-pyranose, 2 $\rightarrow$ 1 and 2 $\rightarrow$ 6 furanose-furanose linkages, 1 $\leftrightarrow$ 1, 1 $\rightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4 and 1 $\rightarrow$ 6 pyranose-pyranose linkages and 1 $\leftrightarrow$ 1, 1 $\rightarrow$ 2, 1 $\rightarrow$ 3 and 1 $\rightarrow$ 5 pyranose-aldopentose linkages using a parametrization strategy that is a continuation of the recent Drude polarizable force field protocol.^{37, 38} This involves the use of high-level QM data both on simple model compounds and full disaccharides as the target data for the optimization of parameters associated with the glycosidic linkages. Two-dimensional (2D) adiabatic potential energy scans (PES) at the MP2/6-31G(d) level for the full range of the Φ/Ψ dihedral angles about the glycosidic linkages were performed followed by RIMP2/cc-pVQZ single point energies. In total, over 31000 conformational energies that encompass all types and chiralities of linkages are used as target data in the parametrization process. Also included in the target data are minimum interaction energies and distances with water at the RIMP2/cc-pVQZ//MP2/6-31G(d) model chemistry used for optimization of the glycosidic linkage oxygen and adjacent carbon atom electrostatic parameters. In addition, QM data on dipole moments and molecular polarizabilities on model compounds and on full disaccharides are included in the optimization. The final parameters are applied to full di-, tri-, and tetrasaccharides in condensed phase simulations in both crystalline and aqueous environments allowing for validation based on their ability to reproduce experimental observables.

Methods

All molecular mechanics calculations were carried out using the CHARMM program version c41a2 and with the polarizable Drude force field,^{28, 37, 38} unless otherwise stated. The water models used in this study are the polarizable SWM4-NDP⁴¹ and CHARMM TIP3P^{42, 43} water models with the latter used in conjunction with the additive CHARMM36 carbohydrate force field.^{13, 14, 17, 44, 45} All MD simulations used a 1 fs time step and were performed at 298 K, unless otherwise stated. All minimizations and molecular dynamics (MD) equilibrations were performed following a previously presented protocol.^{28, 37, 38}

Quantum mechanical calculation were carried out using the Gaussian 03,⁴⁶ QChem,⁴⁷ and Psi4⁴⁸ programs. Adiabatic PES were performed by geometry optimization at the MP2/6-31G(d) level with the target dihedral constrained followed by RIMP2/cc-pVQZ single point energies (i.e. RIMP2/cc-pVQZ//MP2/6-31G(d) QM model chemistry). PES was performed in 15° increments. QM gas phase calculations of full sugar interactions with water were performed via distance scans along defined orientations in 0.1 Å increments with the RIMP2/cc-pVQZ model chemistry with basis set superposition error (BSSE) correction.⁴⁹ In this procedure, only the solute-water interaction distance was allowed to vary in the calculation with all other degrees of freedom constrained with calculations performed on individual monohydrates (see Figure S1). The SWM4-NDP⁴¹ water model internal geometry was used for water⁴² and the solute was constrained to the gas phase optimized geometry in the QM or CHARMM representation.

Crystal MD simulations were initiated from coordinates obtained from the Cambridge Structural Database (CSD).⁵⁰ Hydrogens were rebuilt and minimized with the CHARMM36^{13, 14, 17, 44, 45} additive force field. Supercells were built using the experimental cell dimensions and crystal simulations were performed using the Langevin thermostat with the reference temperature at the experimental temperature, as described previously.^{28, 37, 38} A 20 ns simulation was performed on each crystal including any crystal water if present. Crystals included in the present study are list in Table S1 of the supporting information.

MD simulations to calculate NMR coupling constants consisting of four disaccharides in 1100 molecules of water, corresponding to a concentration of 200 mM, were performed at a temperature of 298K. Simulations were run for a total of 120 ns. Further details can be found in the previously published additive force field papers by Raman¹⁶, Guvench⁵¹ and Patel.¹⁷ NMR heteronuclear three-bond proton-carbon coupling constants (in Hz) for the glycosidic angle rotations J_{COCH} were computed from the simulations using the modified Karplus equation developed by Tvaroska et al.⁵²

$${}^3J_{COCH} = 5.7\cos^2\phi - 0.6\cos\phi + 0.5 \quad 1$$

as well as that by Cloran et al.⁵³

$${}^3J_{COCH} = 7.49\cos^2\phi - 0.96\cos\phi + 0.15 \quad 2$$

${}^3J_{HCCH}$ associated with the fructofuranose ring pucker geometry were also calculated using the equations established by Penhoat et al.^{54, 55}

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

$${}^3J_{HCCH} = 3.97\cos^2\phi - 0.91\cos\phi - 0.10\sin^2\phi + 4.67 \quad 4$$

To compare J coupling between published experimental data and ensemble averages from the MD simulations, the value of the dihedral angle of interest was first calculated for each snapshot of each disaccharide in the simulation for which the corresponding J coupling value was calculated using the above equations. The resulting J coupling values for the dihedral angle of interest were then averaged for all snapshots from the 120 ns MD simulations (coordinates saved every 100 ps) yielding the presented ensemble average.

To facilitate presentation of the results we introduce the following nomenclature to describe the various glycosidic linkages. The type of monosaccharides is indicated by p, f or a for pyranose, furanose or aldopentose, respectively. The atoms involved in the linkages are the standard atoms in the monosaccharides. Accordingly, p1→2a is a 1→2 glycosidic linkage between a pyranose and an aldopentose while, p1→4p is a 1→4 glycosidic linkage between two pyranoses. Use of these descriptors allows for all the disaccharides containing a specific type of linkage to be referred to in a single term.

Results and Discussion

Parameters available for the Drude polarizable force field for protein,^{30, 40, 56} lipids,^{57, 58} nucleic acids,^{29, 59–61} polyalcohols,³⁹ hexopyranose,^{37, 62} and aldopentofuranoses³⁸ provide the foundation for parametrization of carbohydrate glycosidic linkages involving furanoses and pyranoses. The pyranose parameters are an update of the original Drude hexopyranose monosaccharide parameters.^{37, 62} The glycosidic linkages include pyranose - pyranose, pyranose - furanose and furanose - furanose (viz. p1↔1p, p1→2p, p1→3p, p1→4p, p1→6p, p1↔2f, p1→3f, p1→4f, p1→6f, f2→1f, f2→6f), where the numbers in the linkage type notation represents the IUPAC carbon numbering used in glucopyranose and fructofuranose. Linkages involving furanoses other than fructofuranose (namely aldopentose) were included with the numbering in the furanose changed accordingly. The general procedure in this work was adopted from that used for the additive carbohydrate optimization of linkages involving furanose and hexopyranose.^{16, 17, 51} The results presented below are based on the final set of optimized parameters.

Glycosidic linkage electrostatic parameters

The initial electrostatic parameters, including the partial atomic charges, atomic polarizabilities (alphas) and atomic Thole scaling factors, were transferred directly from the full monosaccharides to both the model compounds (Figures 1 and 2) and full sugar disaccharides (Table 1). With the glycosidic oxygen, the charges are located on the two lone pairs (LP) on each oxygen with the polarizability and the Lennard-Jones parameters centered on the oxygen. Several sets of charges ranging between -0.232 (-0.116/LP, based on acyclic ethers) to -0.304 (-0.152/LP, based on the O5 charge in pyranose monosaccharides) were tested both on the model compounds and the full disaccharides. The alphas and Thole scaling factors were not subjected to additional optimization. The overall neutral charge was maintained by modifying the charges on the adjacent carbons. For each set of charges the dihedral parameters were fit targeting the respective QM dihedral PES, following which water interactions, dipole moments and molecular polarizabilities were calculated and compared to QM data. Water interaction energies, based on the interaction orientations shown in Figure S1 (as an example with full list in Table 2) for the full disaccharide, were in better agreement with the QM data with higher charges while the agreement of dipole moments was better with lower charges. Considering both types of target data, charges between -0.240 (-0.120/LP) and -0.280 (-0.140/LP) reasonably reproduce the QM data (Table 1). Based on this level of agreement the O charge was assigned -0.244 (-0.122/LP), as this value is similar to that previously used in the O-methyl group on C1 in aldopentose from the furanose monosaccharide study.⁶³

Validation of the electrostatic parameters in full disaccharides

To validate the quality of transferred atomic charges, polarizabilities and Thole scale factors, the parameters were compared to the total QM dipole moments of low-energy disaccharide conformations and QM disaccharide-water interaction energies on full disaccharides not included in the training set (Table 2). The dipole moments were computed at the MP2/cc-pVQZ level for disaccharide conformations with a relative conformational energy lower than 3.0 kcal/mol selected from the dihedral scans around the corresponding glycosidic linkage. For each type of glycosidic linkage, representative disaccharide conformations were chosen to construct the disaccharide-water interaction models. For each disaccharide model, the water molecule was placed around the linkage oxygen (O_{linkage}) atom with four different orientations, two along the lone pair direction and the other two along the bisector of the covalent angle $C_1-O_{\text{linkage}}-C'_n$ with the two planes (the water plane H-O-H and linkage plane $C_1-O_{\text{linkage}}-C'_n$) perpendicular to each other, analogous to that performed with the furanose-pyranose disaccharide (Figure S1). The disaccharide-water interaction pair was then scanned along the $O_{\text{linkage}}\cdots H_{\text{water}}$ vector at MP2/6-31G(d) level to identify the lowest interaction energy distance followed by single-point energy calculation at MP2/cc-pVQZ level with BSSE correction. The SWM4-NDP water was used to compute the interaction energy in the Drude model at the QM optimized disaccharide-water structure geometry. The transferred parameters result in a RMSE value of 0.52 Debye for dipole moments and 1.17 kcal/mol for disaccharide-water interaction energies, respectively (Table 2). We also optimized the atomic charges, polarizabilities and Thole scale factors of the atoms in the glycosidic linkage ($C_1-O_{\text{linkage}}-C'_n$) targeting the QM total dipole moment and disaccharide-water interaction energy. The resulting parameters are slightly better in the reproduction of the total dipole moment with a RMSE value of 0.45 Debye but a bit worse for the disaccharide-water interaction energies with a RMSE value of 1.27 kcal/mol. In light of the limited improvement of the optimized parameters versus those transferred directly from the monosaccharides, we chose the transferred parameters to maintain transferability across the different glycosidic linkages.

Internal parameter optimization

Internal parameters including bond and valence angle equilibrium values and force constants, and dihedral angle force constants were optimized targeting QM geometries and energies. Crystal geometries⁵⁰ were used to validate the bond and angle parameters as performed previously for the additive CHARMM force field.^{16, 17, 51} Force constants for the dihedral angles about the glycosidic linkages were further adjusted targeting QM potential energy scans (PES) at the RIMP2/cc-pVQZ//MP2/6-31G(d) model chemistry. Analogous adiabatic MM PES were then computed with the dihedral force constant parameters of the dihedral(s) being optimized set to zero. The energy difference between MM and QM (MM-QM) PES was then targeted for the fitting of the dihedral parameters using the least-square fitting program developed in this laboratory.⁶⁴ During fitting four multiplicities n of 1, 2, 3, and 6 were included and the phase angle, δ , was limited to 0 and 180° as required to insure applicability of the parameters to stereoisomers about a chiral center.

Optimization of the internal parameters for the glycosidic linkages initially targeted the model compounds based on THP and THF (Figures 1, 2 and S2: supporting information). The optimized parameters were then transferred to the full disaccharides for the furanose-furanose, pyranose-furanose and pyranose-pyranose species, with additional optimization of selected dihedral parameters undertaken to maximize agreement with the full disaccharide QM target data. The final set of Φ/Ψ and ω dihedral parameters were applied to both the model compounds and full sugars. This leads to some degradation of the quality of the agreement with the target data (see below) for the model compounds, but this was deemed acceptable as the goal of the force field is the study of full polysaccharides in condensed phase environments.

For all compounds, the potential energies for all conformations of the model compounds or full disaccharides with the same number of atoms, including the different linkages and anomers, were offset with respect to the global minimum. For example, model compounds MC1, MC3 and MC4 all have the same number of atoms such that their energies may be offset to the same global minimum. This assures that the relative gas phase energies of the different glycosidic linkages and anomers are correct. We note that with the additive force field this offset was only performed for all the conformations associated with a specific linkage type. The furanose or THF ring conformations in all of the model compounds or full sugar, except where noted, were constrained in the north or south conformations to account for the flexibility of the five member ring, as presented in the additive force field paper.⁴⁵ This was performed by constraining the ${}^5C_5-{}^5O_5-{}^5C_2-{}^5C_3$ dihedral for the north conformation and the ${}^5C_4-{}^5C_5-{}^5O_5-{}^5C_2$ dihedral for the south conformation to zero, as originally performed in studies of nucleic acids.⁶⁵ In all cases parametrization was done in a self-consistent fashion such that whenever one parameter was changed, all target properties were recomputed and parameters re-optimized as required.

Parametrization of the pyranose-pyranose, pyranose-furanose and furanose-furanose glycosidic linkages based on the model compounds

Initial parameters for full disaccharides were optimized based on the model compounds listed in Figures 1 and 2. Missing parameters beyond those associated with the ω and Φ/Ψ dihedrals were adjusted based on the model compounds listed in Table S2 as described in the Supporting Information (Section S1, Table S3 and Figure S2). After transferring the missing parameters to the model compounds, 2D QM dihedral PES about the Φ/Ψ , and 1D QM dihedral PES about ω were computed for all possible anomer combinations of the model compounds (Figures 1 and 2). During the scans the THF ring puckers were constrained to either south or north conformations. Conformations with relative QM energies of 12 kcal/mol or less were used as the target data to fit the Φ/Ψ and ω dihedral parameters. Target dihedrals being optimized are listed in Table S3, columns 5, 6 and 7. The final RMSE values are shown in Figure 3 for the individual model compounds for the fitted parameter along with those where the dihedral force constants were set to zero. More details can be found in Section S.2 of the Supporting Information. Figure 4 and Figure S5A-F show that the optimized MM parameters capture the variation of the QM relative energies for the Φ/Ψ 2D PES to a high degree of accuracy, as well as for the relative energies of the different anomers, consistent with the RMSE values.

Analysis of selected intramolecular geometric terms involving non-hydrogen atoms is shown in Table S4 with respect to the QM MP2/6-31G(d) optimized geometries for all studied model compound anomers. The MM bond lengths show an RMSE of 0.01 Å with respect to the QM target data, while an RMSE of 1.50° is obtained for the valence angles. Notable is the level of agreement for the analyzed torsions, with an RMSE of 6.31°. This agreement indicates that fitting the dihedral parameters to the PES yields parameters that also reproduce the location of the local energy minima on the dihedral surfaces.

Parametrization of the pyranose-furanose and furanose-furanose glycosidic linkages based on the full disaccharides

For optimization of the dihedral parameters the full disaccharides were O-methylated on the reducing end and in the compounds with furanoses those rings were restrained to either the south or north conformation. This process first involved adjusting the aldopentose dihedral parameters involving the O-methyl to achieve improved agreement with QM sugar pucker PES (Figure S6), extending the recent parametrized aldopentose monosaccharide model.³⁸ Listed in Table S5 are the disaccharides included in the training set and the number of conformations of each species. The table also identifies the CHARMM patches used to implement the glycosidic linkages between the individual monosaccharides and to place the O-methyl moiety on the reducing end of the sugars. Each disaccharide conformation was initially generated using the CHARMM36 additive force field by generating conformations on the 2D linkage surfaces in 15° increments and energy minimizing the conformers. Subsequent QM geometry optimizations and single point energy calculations were then performed only on conformations with a relative energy of the 15 kcal/mol or less. Final QM conformations used for fitting were those with relative QM energies of 12 kcal/mol.

Dihedral parameter fitting involved two steps. Parameter optimization was initially performed on the individual disaccharides for the target dihedrals listed in Table S3. Subsequently, global fitting was performed on the disaccharides with common torsions. This leads to the disaccharides containing p1→2a, p1→3a, a1→2p, a1→p3 and a1→p4 glycosidic linkages being fit globally as they share same central atom, while p1↔1a containing disaccharides were fit separately. The final RMSE values are shown in Figure 5. In general the initial parameters obtained from optimization targeting the model compounds yield improved agreement with the QM data versus the additive force field. The additional optimization of the dihedrals unique to the full disaccharides yielded further improvements in the majority of cases. In the case of the a1→2p and a1→3p disaccharides, the Drude model is in poorer agreement with experiment as compared to the additive force field. However, given the overall consistent level of agreement of the Drude with the QM data, the poorer agreement in two cases versus the additive model was deemed acceptable.

Parametrization of the pyranose-pyranose glycosidic linkages based on the full disaccharides

Parametrization of the pyranose-pyranose glycosidic linkages followed the protocol presented above. We note that parameters for the atomic charges, polarizabilities, Thole scale factors, bonds and angles were transferred from the corresponding furanose containing disaccharides to maintain the parameter transferability for atoms with the same local

chemical environment. The full disaccharides, with O-methylation of the reducing end, were constructed for the $p1 \leftrightarrow 1p$, $p1 \rightarrow 2p$, $p1 \rightarrow 3p$ and $p1 \rightarrow 4p$ glycosidic linkage types (see Table 2 and Table S6). For each linkage type there are a total of 256 possible diastereomers associated with the 8 possible diastereomers for each monosaccharide times the α and β anomers for each monosaccharide ($8 \times 8 \times 2 \times 2 = 256$). Since the number of possible diastereomer for each linkage type is quite large, specific diastereomers of each disaccharide for use in the present study were determined according to their frequency of occurring in crystal structures. To this end, a survey of the crystal structure database of glycans constructed by Im and co-workers (<http://www.glycanstructure.org/fragment-db>)^{66, 67} was performed from which the disaccharide diastereomers were selected for use in this study (Table S6). To reduce the computational cost, QM scans for $p1 \leftrightarrow 1p$, $p1 \rightarrow 2p$, $p1 \rightarrow 3p$ and $p1 \rightarrow 4p$ linked disaccharides were performed according to the energy landscape of the THP-THP model compounds used in the additive force field parametrization. Conformations of the THP-THP model from the 2D ϕ/ψ dihedral PES with relative energies lower than 8.0 kcal/mol were used to construct the full disaccharides. For $p1 \rightarrow 6p$ linked disaccharides, in addition to the THP-THP QM scan, QM data were also generated by scanning the ψ/ω and ϕ glycosidic dihedrals every 15° . All generated conformers were then subjected to QM minimization and single point energy calculations to generate the final target data for fitting. The force constants of the dihedrals (Table S3) being fit were set to zero and optimized using the least-square fit program targeting the QM conformational energies. For $p1 \leftrightarrow 1p$ and $p1 \rightarrow 6p$ glycosidic linkages, the parameters were directly fitted targeting the respective disaccharides. Parameters for the $p1 \rightarrow 2p$, $p1 \rightarrow 3p$ and $p1 \rightarrow 4p$ linked disaccharides were optimized globally since they share the same central atoms and associated dihedral parameters for the glycosidic linkages. The results show a remarkable improvement compared to the additive force field (Figure 5 and Table S10), which is evident in Figure 6 that shows the Φ/Ψ energy surface for the $p1 \rightarrow 4p$ of both the Drude and additive force fields along with the QM energy surfaces. The initial parameters for the Drude model based on the model compounds (MC10 and MC11 in Figure 2, MC0 and MC7 In Figure S2) generally show a significant improvement over the additive, with that improvement increasing with the additional optimization in the context of the full disaccharides. As both force fields were optimized targeting the same QM data using similar approaches, the overall improvement in the Drude model indicates the importance of the explicit treatment of polarization in accurately modeling the intrinsic energetics of the disaccharides.

Validation against crystal and aqueous solution experimental data

A total of 25 crystals containing pyranose-furanose or furanose-furanose and 16 pyranose-pyranose full sugar crystal structures were obtained from the Cambridge Structural Database (CSD).⁶⁸ For each system, a supercell was constructed of $2 \times 2 \times 2$ unit cells. Table S1 lists all crystals, the linkages they cover and the numbers of sugar and water molecules in the supercells. Simulations were then performed on each supercell with both additive and the new Drude force fields under the NPT ensemble for 20 ns. In the majority of cases the Drude force field produces smaller deviations of the crystal lattice parameters compared to additive force fields though exceptions occur, as may be seen in Table S8. The average percentile error of the crystal lattice A, B, C and unit cell volumes compared to the

experimental values are 1.35(0.59) Å, 0.85(0.66) Å, 1.69(0.91) Å, and 3.52(0.44) Å³ for the Drude force field versus 1.00(0.31) Å, 1.80(0.36) Å, 2.13(0.36) Å, and 5.08(0.39) Å³ for the additive force field, where the values in parantheses are the standard errors. Overall, the Drude model yields improvement over the additive force field in the description of the solid structures, with the Drude performance better for the B, C and volume terms. In specific cases where the additive force field performance is better, the level of performance of the two models is similar, with the exception of the structure YOXFOG, which is a trihydrate, where significant discrepancies in the C unit cell parameter occur with the Drude model, leading a significant increase in the volume of that unit cell, in contrast to the general decreased volumes with the Drude force field.

Comparison of the MD-averaged intramolecular geometric terms provides a very precise way to judge the accuracy of this aspect of the force field. For each crystal simulated, the values of each bond, angle and dihedral were averaged over all independent molecules and over the simulation time of 20 ns. The difference of these values versus the experimental crystal structures yields the average difference and standard deviation. Table S7A-B shows the results for bonds, angles and torsions associated with the glycosidic linkages. Average differences in bond lengths, angles and torsions are less than 0.03 Å, 3.89° and 13.3°, respectively. Overall, the optimized parameters capture well the intramolecular geometries of the studied di/poly-saccharides in their crystal environments

Additional validation involved solution simulations of selected disaccharides and a trisaccharide from which the conformational properties were compared with experimental NMR data. Karplus-type equations have previously been developed for ¹³C-¹H spin couplings allowing for comparison with the experimentally measured ³J coupling constants to the glycosidic linkage dihedral angles in the trajectories. Saccharides used for the NMR calculations are listed in Tables 3 and 4. Previous studies on these sugars have been performed in the context of the additive force field.^{16, 37, 44}

Experimental ³J coupling constants for trehalose, maltose, cellobiose, melibiose, and gentiobiose are shown in Table 3 with corresponding values from the simulations. Simulation ³J coupling constants were calculated using either equation 1 or 2, which in some cases lead to significantly different estimates of the MD ³J coupling values. Table 4 includes the NMR and MD ³J coupling values for sucrose and kestose. For these sugars the Karplus equation was that presented by Haasnoot-Altona⁵⁵ that does not include furanose. Additive C36 and Drude result are shown in both tables with a summary of the comparison of the MD-calculated and experimental data shown in Table 5. For the pyranose containing polysaccharides in Table 3 the C36 and Drude results are similar using both equations, with the Drude results slightly worse than those from the additive model. With the furanose containing polysaccharides in Table 4 the agreement with experiment is better with the Drude model. Overall, the level of agreement for the saccharides in solution with the Drude model is quite satisfactory, given that the model is based on optimization of the model based entirely on gas phase QM disaccharide properties. In the case of C36, the 1–6 linkage parameters were subjected to additional optimization specifically targeting NMR J-coupling constants,¹⁷ potentially leading to better agreement with experiment over the Drude model for some of the 1–6 linkages (Table 3).

Conclusions

In this work, Drude polarizable force field parameters for glycosidic linkages involving pyranose and furanose sugars were optimized. Linkages for which parameters were optimized includes $p1 \leftrightarrow 2f$, $p1 \rightarrow 3f$, $p1 \rightarrow 4f$ and $p1 \rightarrow 6f$ pyranose-furanose linkages, $f2 \rightarrow 2p$, $f2 \rightarrow 3p$ and $f2 \rightarrow 4p$ furanose-pyranose, $f2 \rightarrow 1f$ and $f2 \rightarrow 6f$ furanose-furanose linkages, $p1 \leftrightarrow 1p$, $p1 \rightarrow 2p$, $p1 \rightarrow 3p$, $p1 \rightarrow 4p$ and $p1 \rightarrow 6p$ pyranose-pyranose linkages and $p1 \leftrightarrow 1a$, $p1 \rightarrow 2a$, $p1 \rightarrow 3a$ and $p1 \rightarrow 5a$ pyranose-aldopentose linkages. Parameters were built on those available for monosaccharides of aldopentoses and hexopyranoses. Target data for torsional parameters associated with the glycosidic linkages included high level QM 1D and 2D PES of the ω and Φ/Ψ glycosidic dihedral angles in both model compounds and full disaccharides. The parameters were then initially fit based on model compounds followed by optimization targeting the full disaccharides. Final results showed a systematic improvement over the additive CHARMM36 carbohydrate force field with respect to the QM data, which, given that similar target data was used for optimization of both models, indicates the importance of the contribution of electronic polarization to treating the conformational energies of the carbohydrates.

As validation both condensed phase aqueous solution properties and crystal unit cell parameters and intramolecular geometries were used. Results with the Drude parameters showed good agreement with the available NMR J coupling data, which was not specifically targeted in the present study. In the case of the di and polysaccharide crystals average differences of 1.35%, 0.85%, 1.69% and 3.52% with respect to experimental crystal unit cell parameters A, B, C and volumes, respectively, were obtained with the Drude model. These results are generally in better agreement with experiment versus the additive force field, though in specific cases the additive force field showed better performance. This general improvement in the polarizable Drude model over the additive force field with respect to experimental data further indicates the importance of the inclusion of electronic polarization in accurately modeling complex carbohydrates.

In summary, the presented Drude polarizable force field can be used to model linear, branched, and cyclic pyranose/furanose/aldopentose glycosides both alone and in heterogeneous systems including proteins, nucleic acids and/or lipids when combined with remainder of the Drude polarizable force field. This offers the capability of studying these systems with a computationally accessible (eg. ~4 fold computational cost over the additive C36 force field), fully polarizable empirical force field not previously available to the scientific community, including the ability to perform MD simulations with a variety of programs including CHARMM^{70, 71}, NAMD^{72, 73}, GROMACS^{74, 75} and OpenMM⁷⁶. To further this capability ongoing efforts are targeting glycosidic linkages involving proteins and lipids as well as a wider range of chemical functional groups that occur in carbohydrates.

The final parameters are available in the CHARMM force field release at (http://mackerell.umaryland.edu/charmm_ff.shtml).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support from the NIH (GM070855) and computational support from the University of Maryland Computer-Aided Drug Design Center and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant Number OCI-1053575 are acknowledged.

References

1. Alonso DM, Wettstein SG, Dumesic JA. Bimetallic catalysts for upgrading of biomass to fuels and chemicals. *Chem Soc Rev.* 2012; 41:8075–8098. [PubMed: 22872312]
2. Dwek RA. Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev.* 1996; 96:683–720. [PubMed: 11848770]
3. Markou G, Angelidaki I, Georgakakis D. Microalgal carbohydrates: an overview of the factors influencing carbohydrates production, and of main bioconversion technologies for production of biofuels. *Appl Microbiol Biotechnol.* 2012; 96:631–645. [PubMed: 22996277]
4. Dwek RA, Butters TD. Introduction: Glycobiology Understanding the Language and Meaning of Carbohydrates. *Chem Rev.* 2002; 102:283–284.
5. Slaney AM, Wright VA, Meloncelli PJ, Harris KD, West LJ, Lowary TL, Buriak JM. Biocompatible Carbohydrate-Functionalized Stainless Steel Surfaces: A New Method For Passivating Biomedical Implants. *ACS Appl Mater Interfaces.* 2011; 3:1601–1612. [PubMed: 21438637]
6. El Kadib A, Bousmina M. Chitosan Bio-Based Organic–Inorganic Hybrid Aerogel Microspheres. *Chem Eur J.* 2012; 18:8264–8277. [PubMed: 22689451]
7. Kaczmarek P, Keay SK, Tocci GM, Koch KR, Zhang CO, Barchi JJ Jr, Grkovic D, Guo L, Michejda CJ. Structure-activity relationship studies for the peptide portion of the bladder epithelial cell antiproliferative factor from interstitial cystitis patients. *J Med Chem.* 2008; 51:5974–5983. [PubMed: 18788730]
8. González-Outeiriño J, Kadirvelraj R, Woods RJ. Structural elucidation of type III group B Streptococcus capsular polysaccharide using molecular dynamics simulations: the role of sialic acid. *Carbohydr Res.* 2005; 340:1007–1018. [PubMed: 15780265]
9. Kuttel MM, Timol Z, Ravenscroft N. Cross-protection in *Neisseria meningitidis* serogroups Y and W polysaccharides: A comparative conformational analysis. *Carbohydr Res.* 2017; 446–447:40–47.
10. Yang M, Simon R, MacKerell AD. Conformational Preference of Serogroup B Salmonella O Polysaccharide in Presence and Absence of the Monoclonal Antibody Se155–4. *J Phys Chem B.* 2017; 121:3412–3423. [PubMed: 28423910]
11. Stewart-Jones GB, Soto C, Lemmin T, Chuang GY, Druz A, Kong R, Thomas PV, Wagh K, Zhou T, Behrens AJ, Bylund T, Choi CW, Davison JR, Georgiev IS, Joyce MG, Kwon YD, Pancera M, Taft J, Yang Y, Zhang B, Shivatare SS, Shivatare VS, Lee CC, Wu CY, Bewley CA, Burton DR, Koff WC, Connors M, Crispin M, Baxa U, Korber BT, Wong CH, Mascola JR, Kwong PD. Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell.* 2016; 165:813–826. [PubMed: 27114034]
12. Yang M, Huang J, Simon R, Wang LX, MacKerell AD Jr. Conformational Heterogeneity of the HIV Envelope Glycan Shield. *Sci Rep.* 2017; 7:4435. [PubMed: 28667249]
13. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, Mackerell AD Jr. Additive empirical force field for hexopyranose monosaccharides. *J Comput Chem.* 2008; 29:2543–2564. [PubMed: 18470966]
14. Hatcher E, Guvench O, MacKerell AD Jr. CHARMM additive all-atom force field for aldopentofuranoses, methyl-aldopentofuranosides, and fructofuranose. *J Phys Chem B.* 2009; 113:12466–12476. [PubMed: 19694450]

15. Hatcher E, Guvench O, MacKerell AD Jr. CHARMM Additive All-Atom Force Field for Acyclic Polyalcohols, Acyclic Carbohydrates and Inositol. *J Chem Theory Comput.* 2009; 5:1315–1327. [PubMed: 20160980]
16. Raman EP, Guvench O, MacKerell AD. CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses. *J Phys Chem B.* 2010; 114:12981–12994. [PubMed: 20845956]
17. Patel DS, Pendrill R, Mallajosyula SS, Widmalm G, MacKerell AD Jr. Conformational properties of alpha- or beta-(1→6)-linked oligosaccharides: Hamiltonian replica exchange MD simulations and NMR experiments. *J Phys Chem B.* 2014; 118:2851–2871. [PubMed: 24552401]
18. Glennon TM, Zheng YJ, Le Grand SM, Shutzberg BA, Merz KM. A force field for monosaccharides and (1 → 4) linked polysaccharides. *J Comput Chem.* 1994; 15:1019–1040.
19. Ott KH, Meyer B. Parametrization of GROMOS force field for oligosaccharides and assessment of efficiency of molecular dynamics simulations. *J Comput Chem.* 1996; 17:1068–1084.
20. Senderowitz H, Still WC. A Quantum Mechanically Derived All-Atom Force Field for Pyranose Oligosaccharides. AMBER* Parameters and Free Energy Simulations. *J Org Chem.* 1997; 62:1427–1438.
21. Momany FA, Willett JL. Computational studies on carbohydrates: in vacuo studies using a revised AMBER force field, AMB99C, designed for α-(1→4) linkages. *Carbohydr Res.* 2000; 326:194–209. [PubMed: 10903029]
22. Eklund R, Widmalm G. Molecular dynamics simulations of an oligosaccharide using a force field modified for carbohydrates. *Carbohydr Res.* 2003; 338:393–398. [PubMed: 12559740]
23. Lii JH, Chen KH, Allinger NL. Alcohols, ethers, carbohydrates, and related compounds. IV carbohydrates. *J Comput Chem.* 2003; 24:1504–1513. [PubMed: 12868113]
24. Vergoten G, Mazur I, Lagant P, Michalski JC, Zanetta JP. The SPASIBA force field as an essential tool for studying the structure and dynamics of saccharides. *Biochimie.* 2003; 85:65–73. [PubMed: 12765776]
25. Singh A, Tessier MB, Pederson K, Wang X, Venot AP, Boons GJ, Prestegard JH, Woods RJ. Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. *Can J Chem.* 2016; 94:927–935. [PubMed: 28603292]
26. Pol-Fachin L, Rusu VH, Verli H, Lins RD. GROMOS 53A6GLYC, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J Chem Theory Comput.* 2012; 8:4681–4690. [PubMed: 26605624]
27. Plazinski W, Lonardi A, Hunenberger PH. Revision of the GROMOS 56A6(CARBO) force field: Improving the description of ring-conformational equilibria in hexopyranose-based carbohydrates chains. *J Comput Chem.* 2016; 37:354–365. [PubMed: 26525424]
28. Small MC, Aytenfisu AH, Lin F-Y, He X, MacKerell AD. Drude polarizable force field for aliphatic ketones and aldehydes, and their associated acyclic carbohydrates. *J Comput-Aided Mol Des.* 2017:1–15.
29. Lemkul JA, MacKerell AD Jr. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: I. Refinement Using Quantum Mechanical Base Stacking and Conformational Energetics. *J Chem Theory Comput.* 2017; 13:2053–2071. [PubMed: 28399366]
30. Lopes PE, Huang J, Shim J, Luo Y, Li H, Roux B, MacKerell AD Jr. Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *J Chem Theory Comput.* 2013; 9:5430–5449. [PubMed: 24459460]
31. Chowdhary J, Harder E, Lopes PE, Huang L, MacKerell AD Jr, Roux B. A polarizable force field of dipalmitoylphosphatidylcholine based on the classical Drude model for molecular dynamics simulations of lipids. *J Phys Chem B.* 2013; 117:9142–9160. [PubMed: 23841725]
32. Lucas TR, Bauer BA, Patel S. Charge equilibration force fields for molecular dynamics simulations of lipids, bilayers, and integral membrane protein systems. *Biochimica et Biophysica Acta (BBA) - Biomembranes.* 2012; 1818:318–329. [PubMed: 21967961]
33. Kumar Sinha S, Mehta M, Patel S. A charge equilibration formalism for treating charge transfer effects in MD simulations: Application to water clusters. *J Comput Chem.* 2017; 38:1389–1409. [PubMed: 28447346]

34. Lucas TR, Bauer BA, Patel S. Charge equilibration force fields for molecular dynamics simulations of lipids, bilayers, and integral membrane protein systems. *Biochim Biophys Acta*. 2012; 1818:318–329. [PubMed: 21967961]
35. Zhong Y, Patel S. Nonadditive empirical force fields for short-chain linear alcohols: methanol to butanol. Hydration free energetics and Kirkwood-Buff analysis using charge equilibration models. *J Phys Chem B*. 2010; 114:11076–11092. [PubMed: 20687517]
36. Zhong Y, Bauer BA, Patel S. Solvation properties of N-acetyl-beta-glucosamine: molecular dynamics study incorporating electrostatic polarization. *J Comput Chem*. 2011; 32:3339–3353. [PubMed: 21898464]
37. Patel DS, He X, MacKerell AD Jr. Polarizable empirical force field for hexopyranose monosaccharides based on the classical drude oscillator. *J Phys Chem B*. 2015; 119:637–652. [PubMed: 24564643]
38. Jana M, MacKerell AD Jr. CHARMM Drude Polarizable Force Field for Aldopentofuranoses and Methyl-aldopentofuranosides. *J Phys Chem B*. 2015; 119:7846–7859. [PubMed: 26018564]
39. He X, Lopes PE, Mackerell AD Jr. Polarizable empirical force field for acyclic polyalcohols based on the classical Drude oscillator. *Biopolymers*. 2013; 99:724–738. [PubMed: 23703219]
40. Lemkul JA, Huang J, Roux B, MacKerell AD Jr. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem Rev*. 2016; 116:4983–5013. [PubMed: 26815602]
41. Lamoureux G, Harder E, Vorobyov I, Roux B, MacKerell AD Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem Phys Lett*. 2006; 418:245–249.
42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983; 79:926–936.
43. Beglov D, Roux B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J Chem Phys*. 1994; 100:9050–9063.
44. Guvench O, Hatcher ER, Venable RM, Pastor RW, MacKerell AD Jr. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J Chem Theory Comput*. 2009; 5:2353–2370. [PubMed: 20161005]
45. Raman EP, Guvench O, MacKerell AD Jr. CHARMM additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses. *J Phys Chem B*. 2010; 114:12981–12994. [PubMed: 20845956]
46. 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 B04 ed. Gaussian, Inc; Pittsburgh, PA: 2003.
47. Shao Y, Gan Z, Epifanovsky E, Gilbert ATB, Wormit M, Kussmann J, Lange AW, Behn A, Deng J, Feng X, Ghosh D, Goldey M, Horn PR, Jacobson LD, Kaliman I, Khaliullin RZ, Ku T, Landau A, Liu J, Proynov EI, Rhee YM, Richard RM, Rohrdanz MA, Steele RP, Sundstrom EJ, Woodcock HL, Zimmerman PM, Zuev D, Albrecht B, Alguire E, Austin B, Beran GJO, Bernard YA, Berquist E, Brandhorst K, Bravaya KB, Brown ST, Casanova D, Chang CM, Chen Y, Chien SH, Closser KD, Crittenden DL, Diedenhofen M, DiStasio RA, Do H, Dutoi AD, Edgar RG, Fatehi S, Fusti-Molnar L, Ghysels A, Golubeva-Zadorozhnaya A, Gomes J, Hanson-Heine MWD, Harbach PHP, Hauser AW, Hohenstein EG, Holden ZC, Jagau TC, Ji H, Kaduk B, Khistyayev K, Kim J, Kim J, King RA, Klunzinger P, Kosenkov D, Kowalczyk T, Krauter CM, Lao KU, Laurent AD, Lawler KV, Levchenko SV, Lin CY, Liu F, Livshits E, Lochan RC, Luenser A, Manohar P, Manzer SF, Mao SP, Mardirossian N, Marenich AV, Maurer SA, Mayhall NJ, Neuscamman E, Oana CM, Olivares-Amaya R, O'Neill DP, Parkhill JA, Perrine TM, Peverati R, Prociuk A, Rehn DR, Rosta

- E, Russ NJ, Sharada SM, Sharma S, Small DW, Sodt A, Stein T, Stück D, Su YC, Thom AJW, Tsuchimochi T, Vanovschi V, Vogt L, Vydrov O, Wang T, Watson MA, Wenzel J, White A, Williams CF, Yang J, Yeganeh S, Yost SR, You ZQ, Zhang IY, Zhang X, Zhao Y, Brooks BR, Chan GKL, Chipman DM, Cramer CJ, Goddard WA, Gordon MS, Hehre WJ, Klamt A, Schaefer HF, Schmidt MW, Sherrill CD, Truhlar DG, Warshel A, Xu X, Aspuru-Guzik A, Baer R, Bell AT, Besley NA, Chai JD, Dreuw A, Dunietz BD, Furlani TR, Gwaltney SR, Hsu CP, Jung Y, Kong J, Lambrecht DS, Liang W, Ochsenfeld C, Rassolov VA, Slipchenko LV, Subotnik JE, Van Voorhis T, Herbert JM, Krylov AI, Gill PMW, Head-Gordon M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol Phys.* 2015; 113:184–215.
48. Parrish RM, Burns LA, Smith DGA, Simonnett AC, DePrince AE 3rd, Hohenstein EG, Bozkaya U, Sokolov AY, Di Remigio R, Richard RM, Gonthier JF, James AM, McAlexander HR, Kumar A, Saitow M, Wang X, Pritchard BP, Verma P, Schaefer HF 3rd, Patkowski K, King RA, Valeev EF, Evangelista FA, Turney JM, Crawford TD, Sherrill CD. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J Chem Theory Comput.* 2017; 13:3185–3197. [PubMed: 28489372]
49. Boys SF, Bernardi F. The Calculation of Small Molecular Interaction by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol Phys.* 1970; 19:553–566.
50. Allen FH, Bellard S, Brice MD, Cartwright BA, Doubleday A, Higgs H, Hummelink T, Hummelink-Peters BG, Kennard O, Motherwell WDS, Rodgers JR, Watson DG. The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information. *Acta Cryst.* 1979; B35:2331–2339.
51. Guvench O, Hatcher E, Venable RM, Pastor RW, MacKerell AD. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J Chem Theory Comput.* 2009; 5:2353–2370. [PubMed: 20161005]
52. Tvaroška I, Hricovíni M, Petráková E. An attempt to derive a new Karplus-type equation of vicinal proton-carbon coupling constants for C-O-C-H segments of bonded atoms. *Carbohydr Res.* 1989; 189:359–362.
53. Cloran F, Carmichael I, Serianni AS. Density Functional Calculations on Disaccharide Mimics: Studies of Molecular Geometries and Trans-O-glycosidic ³JCOCH and ³JCOCC Spin-Couplings. *J Am Chem Soc.* 1999; 121:9843–9851.
54. Herve du Penhoat C, Imbert A, Roques N, Michon V, Mentech J, Descotes G, Perez S. Conformational behavior of sucrose and its deoxy analog in water as determined by NMR and molecular modeling. *J Am Chem Soc.* 1991; 113:3720–3727.
55. Haasnoot CAG, de Leeuw FAAM, Altona C. The relationship between proton-proton NMR coupling constants and substituent electronegativities—I: An empirical generalization of the karplus equation. *Tetrahedron.* 1980; 36:2783–2792.
56. Huang J, MacKerell AD Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem.* 2013; 34:2135–2145. [PubMed: 23832629]
57. Chowdary J, Harder E, Lopes PEM, Huang L, MacKerell AD Jr, Roux B. A polarizable force field of dipalmitoylphosphatidylcholine based on the classical drude model for molecular dynamics simulations of lipids. *J Phys Chem B.* 2013; 117:9142–9160. [PubMed: 23841725]
58. Li H, Chowdhary J, Huang L, He X, MacKerell AD Jr, Roux B. Drude Polarizable Force Field for Molecular Dynamics Simulations of Saturated and Unsaturated Zwitterionic Lipids. *J Chem Theory Comput.* 2017; 13:4535–4552. [PubMed: 28731702]
59. Savelyev A, MacKerell AD Jr. All-atom polarizable force field for DNA based on the classical drude oscillator model. *J Comput Chem.* 2014; 10:1652–1664.
60. Lemkul JA, MacKerell AD Jr. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA. *J Chem Theory Comput.* 2017; 13:2072–2085. [PubMed: 28398748]
61. Baker CM, Anisimov VM, MacKerell AD Jr. Development of CHARMM polarizable force field for nucleic acid bases based on the classical Drude oscillator model. *J Phys Chem B.* 2011; 115:580–596. [PubMed: 21166469]
62. Yang M, Aytenfisu AH, MacKerell AD Jr. Proper balance of solvent-solute and solute-solute interactions in the treatment of the diffusion of glucose using the Drude polarizable force field. *Carbohydr Res.* 2018; 457:41–50. [PubMed: 29422120]

63. Jana M, MacKerell AD Jr. CHARMM Drude polarizable force field for aldehydofuranoses and methyl-aldehydofuranosides. *J Phys Chem B*. 2015; 119(25):7846–7859. [PubMed: 26018564]
64. Vanommeslaeghe K, Yang M, MacKerell AD Jr. Robustness in the fitting of molecular mechanics parameters. *J Comp Chem*. 2015; 36:1083–1101. [PubMed: 25826578]
65. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, MacKerell AD. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J Chem Theory Comput*. 2012; 8:348–362. [PubMed: 22368531]
66. Jo S, Song KC, Desaire H, MacKerell AD Jr, Im W. Glycan Reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J Comput Chem*. 2011; 32:3135–3141. [PubMed: 21815173]
67. Park SJ, Lee J, Patel DS, Ma H, Lee HS, Jo S, Im W. Glycan Reader is Improved to Recognize Most Sugar Types and Chemical Modifications in the Protein Data Bank. *Bioinformatics*. 2017
68. Allen FH. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr B*. 2002; 58:380–388. [PubMed: 12037359]
69. Waterhouse AL, Calub TM, French AD. Conformational analysis of 1-kestose by molecular mechanics and by n.m.r spectroscopy. *Carbohydr Res*. 1991; 217:29–42. [PubMed: 1797404]
70. Brooks BR, Brooks CL, Mackerell AD, 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, Hodosek 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. CHARMM: The biomolecular simulation program. *J Comput Chem*. 2009; 30:1545–1614. [PubMed: 19444816]
71. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983; 4:187–217.
72. Jiang W, Hardy DJ, Phillips JC, Mackerell AD, Schulten K, Roux B. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J Phys Chem Lett*. 2011; 2:87–92. [PubMed: 21572567]
73. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005; 26:1781–1802. [PubMed: 16222654]
74. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1–2:19–25.
75. Lemkul JA, Roux B, van der Spoel D, MacKerell AD Jr. Implementation of extended Lagrangian dynamics in GROMACS for polarizable simulations using the classical Drude oscillator model. *J Comput Chem*. 2015; 36:1473–1479. [PubMed: 25962472]
76. Huang J, Lemkul JA, Eastman PK, MacKerell AD Jr. Molecular Dynamics Simulations of Explicitly Solvated Drude Polarizable Systems on GPUs: Implementation, Validation, and Benchmark. *J Comput Chem*. accepted.

Figure 1.

Model compounds MC1, MC3, MC4, MC5 and MC6 that were used to parametrize the glycosidic linkage in pyranose-furanose and furanose-furanose disaccharides (MC1 for $p1 \rightarrow 2f$, MC3 for $p1 \rightarrow 3/4f$, MC4 for $f2 \rightarrow 2/3/4p$, MC5 for $f2 \rightarrow 6f$, and MC6 for $p1 \rightarrow 6f$).

Figure 2.

Model compounds MC10 and MC11 that were used to parametrize the glycosidic linkages in the pyranose-pyranose disaccharides. (MC10 for $p1 \rightarrow 1p$ and MC11 for $p1 \rightarrow 2/3/4p$)

Figure 3. Model compound root mean square errors (RMSE) of relative conformational energies between the QM and MM results. Energies in kcal/mol. Fitted indicates RMSE using final optimized parameter of the model compounds and Zero indicates RMSE using the target dihedral parameters set to zero. Data presented in Supplement Table S9.

Figure 4.

Φ/Ψ Potential energy surfaces in the QM and MM representations for the 4 anomers ($\alpha\alpha$, $\alpha\beta$, $\beta\alpha$ and $\beta\beta$) of pyranose-furanose model MC1 used to optimize dihedral parameters associated with 1 $\rightarrow$ 2 glycosidic linkages. Results are presented for the five membered (THF) ring constrained to either south (left two columns) or north (right two columns) conformations and the four anomeric types are shown in the difference rows. Energies are in kcal/mol with contours every 1 kcal/mol.

Figure 5.

Full disaccharide root mean square error (RMSE) of relative conformational energies between the QM and MM results for the pyrans. Energies in kcal/mol. Data presented in supporting information Table S10. Drude (full) and (model) indicates parameters optimized based on full disaccharides and the THP/THF model compound QM target data, respectively. The dashed lines demarcates the pyranose-furanose/aldopentose disaccharides to the left and the pyranose-pyranose disaccharides to the right.

Figure 6. Φ/Ψ potential energy surfaces in the QM and MM (fit) representations for the $p1 \rightarrow 4p$ anomers ($\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, $\beta\beta$) of the full sugar. Results are presented for QM, Drude and additive force field across rows. Energies are in kcal/mol with contours every 1 kcal/mol. In the additive $\beta\alpha$ panel, the energies are all > 12 kcal/mol, such that the panel is completely blank.

Table 1

Root mean square errors (RMSE) associated with the different charges applied to the glycosidic linkages in the full disaccharides with respect to QM dipole moments and water-model compound interaction energies using selected pyranose-pyranose, pyranose-furanose and furanose-furanose full disaccharides.^a RMSE values for “Vs. Water and dipole moment” were calculated with equal weighting on the two terms. Optimized indicates the charge set optimized to reproduce the target data and the bold, underlined set are the final charges used in the glycosidic linkages between the disaccharides.

Charge	Vs. Water	Vs. Dipole moment	Vs. Water and dipole moment
Optimized	0.615	0.263	0.550
-0.072	1.236	0.371	1.089
-0.082	1.103	0.353	0.974
-0.092	0.977	0.340	0.865
-0.102	0.863	0.331	0.767
-0.112	0.763	0.327	0.682
<u>-0.122</u>	<u>0.687</u>	<u>0.328</u>	<u>0.618</u>
-0.132	0.642	0.335	0.581
-0.142	0.635	0.347	0.577
-0.152	0.667	0.365	0.607

^aFull disaccharides and data in this table include α -D-Arbf(1 $\rightarrow$ 5)- α -D-Arbf(32 dipole moment and 85 water interaction), α -D-Glcp-(1 $\rightarrow$ 6)- β -D-Fruf(14 dipole moment and 36 water interaction), β -D-Galp-(1 $\rightarrow$ 4)- β -D-Fruf(33 dipole moment and 45 water interaction) and α -D-Glcp-(1 $\rightarrow$ 2)- β -D-Fruf(27 dipole moment and 48 water interaction) in total 106 dipole moment and 214 water interaction pairs.

Table 2

Dipole moments (Debye) and full disaccharide-water interaction energies^a (kcal/mol) computed from Drude force field with transferred and MCSA optimized electrostatic parameters.

disaccharide model	No. of conformations	Dipole moment RMSE		no. of conformations for water interaction ^b	Water interaction RMSE	
		Transferred	Optimized		Transferred	Optimized
α -D-Glcp-(1 $\rightarrow$ 1)- α -D-Glcp	23	0.38	0.45	3	0.78	0.86
α -D-Glcp-(1 $\rightarrow$ 1)- β -D-Glcp	15			3		
β -D-Glcp-(1 $\rightarrow$ 1)- β -D-Glcp				4		
α -D-Manp-(1 $\rightarrow$ 2)- α -D-Manp-OMe	11	0.31	0.54	4	1.55	1.66
α -D-Galp-(1 $\rightarrow$ 2)- β -D-Galp-OMe	19			4		
β -D-Manp-(1 $\rightarrow$ 2)- α -D-Manp-OMe	14			4		
β -D-Glcp-(1 $\rightarrow$ 2)- β -D-Glcp-OMe	5			4		
α -D-Manp-(1 $\rightarrow$ 3)- β -D-Manp-OMe	13	0.46	0.35	4	1.26	1.30
α -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-OMe	7			4		
β -D-Glcp-(1 $\rightarrow$ 3)- β -D-Glcp-OMe	3			4		
β -D-Manp-(1 $\rightarrow$ 3)- β -D-Manp-OMe	6			4		
α -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-OMe	2	0.62	0.55	3	0.72	0.93
α -D-Glcp-(1 $\rightarrow$ 4)- α -D-Glcp-OMe	27			4		
β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-OMe	42			4		
β -D-Glcp-(1 $\rightarrow$ 4)- β -D-Glcp-OMe	11			2		
α -D-Manp-(1 $\rightarrow$ 6)- β -D-Manp-OMe	30	0.65	0.38	4	1.13	1.22
α -D-Glcp-(1 $\rightarrow$ 6)- α -D-Glcp-OMe	20					
β -D-Manp-(1 $\rightarrow$ 6)- β -D-Manp-OMe	6			(3+4) ^a		
β -D-Glcp-(1 $\rightarrow$ 6)- β -D-Glcp-OMe	28			3		
Total	282	0.52	0.45	69	1.17	1.27

^aTwo conformations of the β -D-Manp-(1 $\rightarrow$ 6)- β -D-Manp-OMe disaccharides were used to model the disaccharide-water interactions.

^bOnly successfully completed calculations in QM were listed in this table.

Table 3

J-coupling constants computed from MD simulations of pyranose-pyranose disaccharides and the experimental values. Experimental errors are in parentheses. Two data are presented for the two proton on C₆.

$^3J(\text{H}_1\text{C}_1\text{O}_{\text{link}}\text{C}_n')$				
Disaccharide	Linkage	Exp	MD	
			(eq 1)	(eq 2)
TREHALOSE	$\alpha(1->1)\alpha$	C36	2.55	2.73
		Drude	2.76	3.01
MALTOSE	$\alpha(1->4)$	C36	3.93	4.52
		Drude	2.69	2.92
CELLOBIOSE	$\beta(1->4)$	C36	2.81	3.07
		Drude	3.04	3.36
MELIBIOSE	$\alpha(1->6)$	C36	2.33	2.45
		Drude	2.81	3.07
GENTIOBIOSE	$\beta(1->6)$	C36	2.62	2.82
		Drude	3.35	3.81
$^3J(\text{C}_1\text{O}_{\text{link}}\text{C}_n'\text{H}_n')$				
MALTOSE	$\alpha(1->4)$	C36	4.49	5.27
		Drude	4.81	5.66
CELLOBIOSE	$\beta(1->4)$	C36	4.99	5.90
		Drude	5.01	5.93
MELIBIOSE	$\alpha(1->6)$	C36	2.15/2.22	2.36/2.55
		Drude	2.30/2.44	2.62/2.88
GENTIOBIOSE	$\beta(1->6)$	C36	2.18/2.28	2.34/2.52
		Drude	2.58/2.83	2.98/3.34

J-coupling constants computed from the MD simulations of Sucrose and Kestose and the experimental values.⁶⁹

Table 4

SUCROSE EXP		MD (eq 1, 3 &4)		KESTOSE EXP		MD (eq 1, 3 &4)	
$^6\text{H}_1-^6\text{C}_1-\text{O}_{\text{link}}-^5\text{C}_2$	4.20	C36	4.26	$^6\text{H}_1-^6\text{C}_1-\text{O}_{\text{link}}-^5\text{C}_2$	4.30	C36	4.22
		Drude	4.53			Drude	4.38
$^5\text{H}_3-^5\text{C}_3-^5\text{C}_4-^5\text{H}_4$	8.80	C36	9.77	$^5\text{H}_3-^5\text{C}_3-^5\text{C}_4-^5\text{H}_4$	8.00	C36	9.79
		Drude	9.20			Drude	9.24
$^5\text{H}_4-^5\text{C}_4-^5\text{C}_5-^5\text{H}_5$	8.30	C36	9.12	$^5\text{H}_4-^5\text{C}_4-^5\text{C}_5-^5\text{H}_5$	8.50	C36	9.20
		Drude	7.78			Drude	7.83
				$^5\text{H}_3-^5\text{C}_3-^5\text{C}_4-^5\text{H}_4$	8.0	C36	9.81
						Drude	9.33
				$^5\text{H}_4-^5\text{C}_4-^5\text{C}_5-^5\text{H}_5$	7.5	C36	8.95
						Drude	7.37

Table 5

Summary of the comparison of the MD-calculated and experimental NMR J-coupling constants.

	Ave. Diff		RMS Diff		Abs. Ave. Diff	
	C36	Drude	C36	Drude	C36	Drude
Table 3 data						
Eq. 1	-0.42	-0.23	0.59	0.66	0.46	0.55
Eq. 2	-0.05	0.22	0.60	0.80	0.50	0.67
Table 4 data						
	0.94	0.26	1.16	0.73	0.96	0.59