

CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates

Sairam S. Mallajosyula,[†] Olgun Guvench,[‡] Elizabeth Hatcher,[†] and Alexander D. MacKerell, Jr.*[†]

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

[‡]Department of Pharmaceutical Sciences, University of New England College of Pharmacy, Portland, Maine 04103, United States

Supporting Information

ABSTRACT: Presented is an extension of the CHARMM additive all-atom carbohydrate force field to enable the modeling of phosphate and sulfate linked to carbohydrates. The parameters are developed in a hierarchical fashion using model compounds containing the key atoms in the full carbohydrates. Target data for parameter optimization included full two-dimensional energy surfaces defined by the glycosidic dihedral angle pairs in the phosphate/sulfate model compound analogs of hexopyranose monosaccharide phosphates and sulfates, as determined by quantum mechanical (QM) MP2/cc-pVTZ single point energies on MP2/6-31+G(d) optimized structures. In order to achieve balanced, transferable dihedral parameters for the dihedral angles, surfaces for all possible anomeric and conformational states were included during the parametrization process. In addition, to model physiologically relevant systems, both the mono- and dianionic charged states were studied for the phosphates. This resulted in over 7000 MP2/cc-pVTZ//MP2/6-31G+(d) model compound conformational energies which, supplemented with QM geometries, were the main target data for the parametrization. Parameters were validated against crystals of relevant monosaccharide derivatives obtained from the Cambridge Structural Database (CSD) and larger systems, namely inositol-(tri/tetra/penta) phosphates noncovalently bound to the pleckstrin homology (PH) domain and oligomeric chondroitin sulfate in solution and in complex with cathepsin K protein.

■ INTRODUCTION

Phosphate and sulfate moieties linked to carbohydrates are essential components of biochemical processes. Two important classes of phosphate-bearing carbohydrates are sugar nucleotides and inositol phosphates. The glycosyl esters of nucleoside di- or monophosphates, generally referred to as "sugar nucleotides", are of primary importance in carbohydrate metabolism.^{1–4} These activated sugars serve as the donor during the biosynthesis of oligosaccharides, polysaccharides, and other glycoconjugates and interact with enzymes involved in carbohydrate synthesis, namely epimerases,⁵ transporters,⁶ and glycosyltransferases.⁷ Until recently, very little was known about the three-dimensional (3D) structure and conformational behavior of nucleotide sugars.^{8–10} However, in the past several years, a number of glycosyltransferase crystal structures have been solved that capture various reaction pathway stages, thereby enabling theoretical modeling.^{8–10} Whereas sugar nucleotides have hexopyranose moieties whose core is a six-membered cyclic ether, the inositol phosphates (IP_x , where $x = 1, 2, \dots, 6$) lack the ether functional group and have at their core cyclohexane. Inositol phosphates are involved, for example, in the inositol signaling pathway, in which IP_3 is liberated from the inner leaflet of the cytoplasmic membrane by hydrolysis and acts as an intracellular second messenger.^{11,12}

Sulfate-bearing carbohydrates are important components of glycosaminoglycans (GAGs), which are unbranched carbohydrate polymers composed of repeating disaccharide units. Sulfated GAGs are linked to protein cores to form proteoglycans (PGs), which are found in vertebrate digestive, endocrine, nervous, muscular, skeletal, circulatory, immune,

respiratory, urinary, and reproductive tissues.^{13–16} Sulfation patterns define distinct types of GAGs. One example is chondroitin sulfate (CS), which is found in several PGs. Its linear polymer structure is characterized by sulfated disaccharide units consisting of the hexopyranose monosaccharides β -D-glucuronic acid (GlcUA) and 2-acetamido-2-deoxy- β -D-acetylgalactose (*N*-acetyl-D-galactosamine, GalNAc) joined by $\beta(1 \rightarrow 4)$ and $\beta(1 \rightarrow 3)$ linkages. In CS, sulfation is known to occur at a variety of positions on both of the component monosaccharides. For example, sulfation at the C4 or C6 positions of GalNAc yields chondroitin-4-sulfate (C4S) and chondroitin-6-sulfate (C6S), respectively.^{17,18} CS proteoglycans (CSPGs) are found to be critical for correct neuronal development and cellular signaling,¹⁹ and structural modifications to CSPGs have also been implicated in atherosclerosis and cancer.²⁰

Understanding the 3D conformational properties of carbohydrate-containing molecules such as those described above and relating them to their physical properties and biological function is of prime importance. However, structural studies have been hindered due to the inherent flexibility²¹ of carbohydrates, which hinders their crystallization, leading to, for example, significant errors in crystal structures for molecules deposited in the PDB (Protein Data Bank).²² Structural characterization by NMR has also proven challenging because of the characteristic overlap of NMR resonances in the spectra of oligosaccharides.²³

Received: November 6, 2011

Published: December 26, 2011



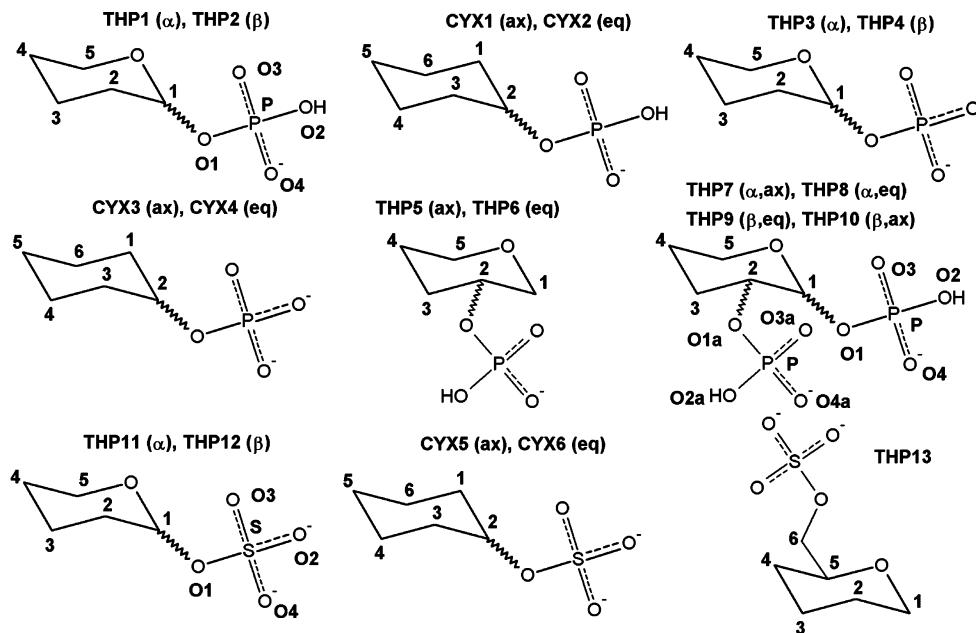


Figure 1. Model compounds used to develop parameters for phosphates and sulfates linked to carbohydrates. Tetrahydropyran derivatives are identified as THP, while cyclohexane derivatives are identified as CYX. Atom labels and numberings are presented in bold font, and for THP, compounds are analogous to D-glucopyranose.

To this end, molecular modeling and dynamics (MD) studies using accurate force fields can provide atomic-level-of-detail properties such as structure, dynamics, and thermodynamics. Classical force field development efforts aimed at enabling accurate modeling of carbohydrates and carbohydrate-containing biomolecular systems have been ongoing for over a decade.^{24–33} However, their scope has been limited by the fact that, in part, much of the parameter development work was not done in the wider context of biomolecular force fields. This complicates attempts to model heterogeneous biomolecular systems containing proteins, lipids, and/or nucleic acids with carbohydrates due to differences in the force field parametrization protocols. Recently, efforts have been made to revise and make the resulting oligosaccharide parameters compatible with the related family of force fields, with two examples being the GROMOS and GLYCAM force fields.^{26,27,34,35}

Toward developing a comprehensive additive all-atom CHARMM carbohydrate force field, we have optimized and validated parameters for pyranose³⁶ and furanose³⁷ monosaccharides, as well as aldose and ketose linear carbohydrates and their reduced counterparts, the sugar alcohols.³⁸ Parameters have also been developed for glycosidic linkages involving both pyranoses^{39,40} and furanoses.⁴⁰ Recently, the parameters have also been extended to cover deoxy, oxidized, or N-methylamine monosaccharide derivatives as well as covalent N and O linkages to proteins.^{41,42} The development protocol for the CHARMM carbohydrate force field is consistent with the other components of the CHARMM additive all-atom biomolecular force field, which includes proteins,^{43,44} nucleic acids,^{45,46} lipids,^{47–50} and drug-like small molecules,⁵¹ with the intention of creating a widely applicable and robust force field for the modeling of heterogeneous biomolecular systems. To date, the CHARMM carbohydrate force field has been validated and applied in the context of oligosaccharides, noncovalent carbohydrate–protein interactions, and covalent O- and N-linkages to proteins.^{41,42,52} The present work extends those

parameters to enable the modeling of phosphate and sulfate groups linked to carbohydrates.

METHODS

Molecular mechanics (MM) calculations for parameter development were done with the CHARMM software.⁵³ The force field potential energy function was same as in the previous studies,^{36–41} as were the details of the molecular mechanics and dynamics calculations. Briefly, a modified version of the TIP3P model was used to represent water.^{54,55} The SHAKE algorithm⁵⁶ was applied to keep water molecules rigid and to constrain covalent bonds between hydrogens and their covalently bound atoms to their equilibrium values. Gas-phase molecular mechanics energies were calculated using infinite nonbonded cutoffs. Aqueous and crystal simulations employed periodic boundary conditions to minimize boundary artifacts and to simulate the infinite crystal environment. Force-switched (aqueous) or energy-switched (crystal) smoothing⁵⁷ was applied to LJ interactions in the range of $c - 2$ to c , where c is the cutoff distance in Ångstroms, and long-range Coulomb interactions were handled using particle mesh Ewald⁵⁸ with a real-space cutoff of c . The value of c was 12 Å for all simulations in the present study. Equations of motion were integrated with the “leapfrog” integrator,⁵⁹ with the integration time step dt of 1 fs for crystal simulations and 2 fs for aqueous simulations. In MD simulations using CHARMM, the isothermal–isobaric ensemble was generated via Nosé–Hoover thermostating,^{60,61} Langevin piston barostating,⁶² and a long-range correction to the pressure to account for LJ interactions beyond the cutoff distance c .⁶³ Condensed-phase simulations were done at 298 K and 1 atm for all simulations, consistent with experimental conditions. Crystal simulations were performed on relevant systems from the Cambridge Structural Database⁶⁴ (CSD) using the appropriate experimental unit cell geometries with crystallographic water molecules and/or crystallographic counterions; aqueous simulations employed a cubic box as the periodic system. Cell length dimensions were varied

isotropically to maintain the target pressure in aqueous simulations. Unit cell edge lengths were allowed to vary independently in crystal simulations. Angular crystal cell parameters of 90° were constrained, while those not 90° were allowed to vary independently. Error estimates for data were generated by treating each MD snapshot as an independent sample and using the expression $t_{\text{critical}}/(n^{0.5})$, where n is the number of snapshots, s is the sample standard deviation, and $t_{\text{critical}} = 1.960$, which is the value for a 95% confidence level for a t distribution with infinite degrees of freedom. Additional simulation details for relevant systems are mentioned in the Results and Discussion section for each system.

The Gaussian 03 program⁶⁵ was used for all QM calculations. For all of the model compounds (Figure 1), geometry optimization was done at the MP2/6-31+G(d) level^{66,67} followed by a MP2/cc-pVTZ single point energy calculation (MP2/cc-pVTZ//MP2/6-31+G). The cc-pVTZ basis set was used for evaluating all conformational energies based on efficiency and accuracy on carbohydrate systems.^{36–41} All potential energy scans were performed with only the scanned dihedral angles constrained.

MM energies were fit to QM potential energies using the freely available Monte Carlo simulated annealing (MCSA) dihedral parameter fitting program “fit_dihedral.py” (available for download at <http://mackerell.umaryland.edu>).⁶⁸ From the expression for the MM dihedral energy:

$$U_{\text{dihedral}} = \sum_{\text{dihedrals}\chi} K_\chi (1 + \cos(n\chi - \delta)) \quad (1)$$

multiplicities n of 1, 2, and 3 were included for each dihedral being fit, and the corresponding K_χ values were optimized to minimize the root-mean-square error RMSE between the MM and QM energies as defined by

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

In eq 2, 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 , and E_i^{MM} is the total MM energy, including the energy of the dihedrals for which the parameters are being optimized (eq 1). The constant c vertically aligns the data to minimize the RMSE as defined by

$$\frac{\partial \text{RMSE}}{\partial c} = 0 \quad (3)$$

w_i values were empirically chosen to favor more accurate fitting of low-energy conformations (see below). K_χ values were constrained to be no more than 3 kcal/mol, and phase angles δ were limited to 0 and 180°. The latter constraint ensures applicability of dihedral parameters to both diastereomers about a chiral center.

Protein:carbohydrate Simulations. Five crystal structures were chosen to validate the parameters. For phosphate linkages, the pleckstrin homology domain noncovalently bound to tri-, tetra-, and penta-inositol phosphates was chosen (PDB id: 1U29,⁶⁹ 1UNQ,⁷⁰ and 1FHW⁷¹). For sulfate linkages, oligomeric chondroitin sulfate in solution (PDB id: 1C4S¹⁸) and in complex with cathepsin K protein (PDB id: 3C9E⁷²) was studied. These structures (protein:carbohydrate) were chosen as they have been solved at a high resolution or involve

molecules relevant for testing the force field (Table 1). Scripts obtained from CHARMM-GUI and modified accordingly to

Table 1. MD Simulation Details

system (PDB id)	resolution (Å)	simulation box (Å ³)	counter ions
1U29	1.80	60 × 60 × 60	3 K ⁺
1UNQ	0.98	70 × 70 × 70	7 K ⁺
1FHW	1.90	87 × 87 × 87	9 K ⁺
1C4S	3.00	56 × 56 × 56	6 Na ⁺
3C9E	1.80	75 × 75 × 75	4 Ca ²⁺

include the linkages were used to set up the simulations.^{73,74} Coordinates of the crystal structures were retrieved from the PDB (Protein Data Bank).⁷⁵ The Reduce software was applied to protein coordinates to place missing hydrogen positions and to choose optimal Asn and Gln side chain amide and His side chain ring orientations.⁷⁶ Patch residues were used to incorporate disulfide bonds and the relevant linkages to monosaccharide components. Crystallographic water molecules, counterions, and heteroatoms were included in the simulation systems. These systems were then immersed in a pre-equilibrated cubic water box that extended at least 10 Å beyond the non-hydrogen atoms of the system. Water molecules having oxygen within a distance of 2.8 Å of non-hydrogen solute atoms were deleted. On the basis of the overall charge, counterions were added to neutralize the solvated systems. Details of each simulation system are summarized in Table 1. Using the CHARMM software, water molecules were allowed to rearrange around the fixed solute atoms by a short minimization of 50 steepest descent (SD) steps followed by 50 adopted basis Newton–Raphson (ABNR) steps.^{77,78} Next, with a mass-weighted harmonic restraint (atomic mass(Δr)² kcal/mol/Å²/AMU) on the non-hydrogen atoms of the solute, all atoms in the system were subjected to a 50-step SD minimization followed by a 50-step ABNR minimization cycle. This was followed by a 100 ps simulation in the NVT ensemble with the same harmonic restraints to equilibrate the solvent molecules around the glycoprotein. An unrestrained 200 ps NPT simulation at 1 atm and 298 K was then performed to complete the equilibration, and NAMD version 2.7b1 was subsequently used for a 20-ns unrestrained production simulation.⁷⁹ In the production simulations, Langevin thermostating was applied to all atoms (coupling coefficient of 1 ps⁻¹, temperature bath of 298 K), and Langevin piston barostating (piston oscillation period of 200 fs, barostat damping time scale of 100 fs, piston pressure of 1 atm, and temperature of 298 K) was used to maintain the system pressure. The last 14 ns of the production runs were used for the analyses.

RESULTS AND DISCUSSION

The focus of this work was to develop parameters to extend the existing carbohydrate force field to include phosphorylated and sulfated sugars. Special care was taken to develop and validate parameters that could be used in tandem with other components of the CHARMM biomolecular force field, such as existing parametrizations for proteins^{43,44} and lipids.^{47–50} Keeping this in mind, judicious parameter transfers were made from these prior efforts. For example, partial charges on phosphate and sulfate groups were transferred from methylphosphates and methylsulfate, since water interaction energies with these groups had previously been studied in detail.^{45–50} A comparison between the charges on methylphosphates and

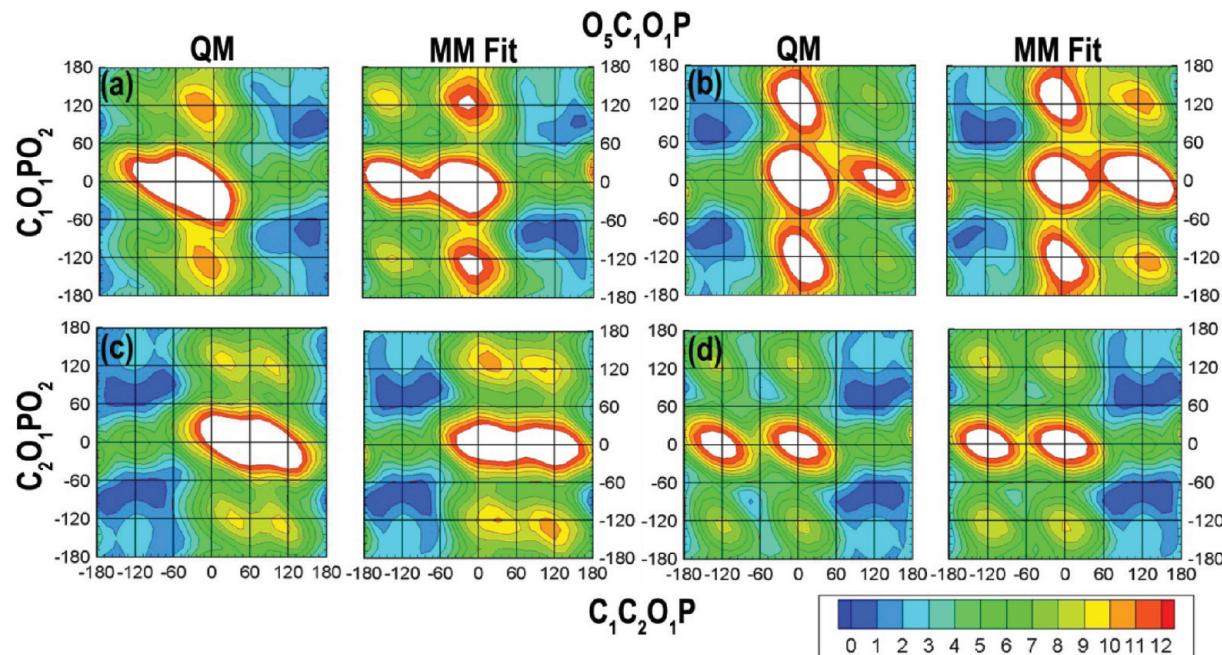


Figure 2. 2D-dihedral potential energy scans about the O₅—C₁—O₁—P/C₁—O₁—P—O₂ dihedrals for model compounds (a) THP1 and (b) THP2 and about the C₁—C₂—O₁—P/C₂—O₁—P—O₂ dihedrals for model compounds (c) CYX1 and (d) CYX2. Energies are in kcal/mol, with contours every 1 kcal/mol. Only energies below 12 kcal/mol have been plotted for the sake of clarity. In each case, the left panel depicts the QM scan and the right panel depicts the MM scan.

methylsulfate and the corresponding moieties in the carbohydrate force field is presented in the Supporting Information (SI). This approach prevents charge mismatch in heterogeneous systems (e.g., carbohydrate/protein or carbohydrate/lipid) that could lead to simulation artifacts. Below, we describe the model compounds used in the parametrization process and the various issues involved.

Model Compounds. Two sets of model compounds were used to parametrize the targeted linkages (Figure 1). The first set consisted of phosphate/sulfate groups attached to the anomeric carbon (C1) of the hexopyranose analog tetrahydropyran in both the α and β conformations (model compounds labeled THP, Figure 1). Parameters developed for these model compounds were then transferred to the second set of model compounds, in which these moieties were attached to the C2 carbon of cyclohexane, with the phosphate/sulfate groups in both the axial and equatorial geometries (model compounds labeled CYX, Figure 1). To account for physiologically relevant systems, both monoanionic ($-HPO_4^{2-}$) (THP1, THP2, CYX1, CYX2) and dianionic ($-PO_4^{2-}$) (THP3, THP4, CYX3, and CYX4) charged states for phosphates were included. Several additional model compounds were chosen to parametrize certain specific dihedral parameters, as follows. Model compounds THP5 and THP6 were chosen to parametrize the attachment of phosphate groups to the C2 carbon of tetrahydropyran. Model compounds THP7–THP10 were used to parametrize the O_{phos}—C—C—O_{phos} dihedral angle, which is present when phosphate groups are attached to adjacent carbons on a carbohydrate ring. Model compounds INI1–INI4 (Figure S1, Supporting Information (SI)) were used to parametrize the dihedral angle between the phosphate ether oxygen and the hydroxyl oxygen (HO₃PO—C—C—OH) as in inositol phosphates.^{11,12} Model compounds THP11, THP12, CYX5 and CYX6 were chosen to parameterize the sulfate attachment and

THP13 was used to parametrize the attachment of sulfate to the C6 carbon of a hexopyranose to enable the modeling of chondroitin-6-sulfates (C6S).^{17,18}

Parameter Optimization Overview. Following the transfer of available bonded and nonbonded parameters for phosphate and sulfate groups from prior work,^{45–50} a number of dihedral parameters spanning these groups and the carbohydrate-analog portion were not available and needed to be parametrized in the present study. A hierarchical parametrization scheme was applied to obtain transferable parameters. First, dihedral parameters were determined for monoanionic phosphates in the context of tetrahydropyran. This involved simultaneous fitting of dihedrals for O₅—C₁—O₁—P, C₂—C₁—O₁—P, C₁—O₁—P—O₂, and C₁—O₁—P—(O₃/O₄) (see Figure 1 for atom naming). The resulting parameters were then used to study monoanionic cyclohexane derivatives, including transferability of the evaluated C₂—C₁—O₁—P dihedral parameters, which also describe the C₁—C₂—O₁—P and C₃—C₂—O₁—P dihedrals. After optimizing parameters for monoanionic phosphates in the context of both THP and CYX model compounds, dianionic derivatives were targeted. Monoanionic C₁—O₁—P—(O₃/O₄) dihedral parameters were tested for transferability and compared to results from independent fitting of C₁—O₁—P—(O₂/O₃/O₄). A set of parameters was judiciously chosen that could reproduce both the monoanionic and dianionic target QM data in the THP compounds, and this set of parameters was also evaluated in the context of dianionic CYX phosphates. For sulfate-containing model compounds, a similar strategy was employed wherein the dihedral parameters were first determined for THP sulfates. These parameters were then used to study CYX sulfates. Parametrization was performed in a self-consistent fashion such that whenever one parameter was changed, properties were recomputed and additional parameters were reoptimized, if necessary.

Table 2. Bond Lengths, Valence Angles, and Dihedral Angles for Model Compounds THP1 and THP2 and CYX1 and CYX2

	THP1			THP2			
	QM	MM	MM-QM	QM	MM	MM-QM	avg err
bonds							
C1–O1	1.419	1.413	−0.006	1.396	1.412	0.016	0.005
O1–P	1.687	1.633	−0.054	1.686	1.636	−0.051	−0.052
P–O2*	1.686	1.592	−0.095	1.684	1.591	−0.093	−0.094
P–O3	1.508	1.519	0.011	1.522	1.518	−0.003	0.004
P–O4	1.509	1.516	0.007	1.500	1.516	0.016	0.012
angles							
C2–C1–O1	111.3	112.5	1.2	111.1	105.6	−5.5	−2.2
O5–C1–P	107.5	108.2	0.7	105.6	107.3	1.7	1.2
C1–O1–P	118.4	117.1	−1.4	117.0	115.4	−1.6	−1.5
O1–P–O2*	100.1	97.4	−2.6	99.5	103.5	4.0	0.7
O1–P–O3	107.4	110.9	3.5	108.0	110.0	2.0	2.8
O1–P–O4	105.9	109.1	3.2	105.6	110.6	5.1	4.1
O2*–P–O3	107.2	110.7	3.5	105.3	107.3	2.0	2.7
O2*–P–O4	106.6	107.2	0.6	110.2	107.1	−3.1	−1.2
O3–P–O4	126.6	119.3	−7.3	125.4	117.3	−8.0	−7.7
dihedrals							
C3–C2–C1–O1	−61.6	−74.3	−12.7	174.4	170.8	−3.6	−8.1
C5–O5–C1–O1	60.1	71.5	11.4	177.2	−174.4	8.3	9.8
C2–C1–O1–P	−75.9	−74.7	1.2	97.5	99.7	2.2	1.7
O5–C1–O1–P	163.2	160.1	−3.1	−143.0	−141.9	1.0	−1.1
C1–O1–P–O2*	88.9	100.2	11.3	71.4	79.5	8.1	9.7
C1–O1–P–O3	−22.9	−15.4	7.5	−38.2	−34.9	3.3	5.4
C1–O1–P–O4	−160.5	−148.7	11.8	−174.4	−166.1	8.3	10.1
CYX1				CYX2			
	QM	MM	MM-QM	QM	MM	MM-QM	avg err
bonds							
C2–O1	1.437	1.417	−0.021	1.430	1.415	−0.015	−0.018
O1–P	1.682	1.642	−0.040	1.674	1.643	−0.031	−0.035
P–O2*	1.686	1.592	−0.094	1.692	1.592	−0.099	−0.097
P–O3	1.511	1.517	0.006	1.522	1.519	−0.003	0.002
P–O4	1.510	1.519	0.010	1.503	1.518	0.015	0.012
angles							
C1–C2–O1	109.6	109.3	−0.2	110.0	107.8	−2.2	−1.2
C3–C2–O1	106.4	106.7	0.3	107.8	105.0	−2.8	−1.3
C2–O1–P	116.8	119.1	2.3	117.1	118.6	1.6	2.0
O1–P–O2*	99.9	96.2	−3.6	98.9	96.5	−2.5	−3.0
O1–P–O3	105.3	109.2	3.9	109.1	109.7	0.6	2.2
O1–P–O4	108.5	111.2	2.7	105.9	110.6	4.7	3.7
O2*–P–O3	107.7	107.5	−0.1	104.9	107.8	2.9	1.4
O2*–P–O4	106.4	110.8	4.4	109.9	110.7	0.7	2.6
O3–P–O4	126.1	119.3	−6.7	125.0	119.2	−5.8	−6.3
dihedrals							
C4–C3–C2–O1	63.4	75.9	12.4	−176.7	−171.6	5.1	8.8
C6–C1–C2–O1	−61.9	−74.4	−12.6	175.5	169.3	−6.3	−9.4
C1–C2–O1–P	−93.8	−90.5	3.2	98.4	92.7	−5.7	−1.2
C3–C2–O1–P	145.7	146.5	0.8	−140.0	−149.4	−9.5	−4.3
C2–O1–P–O2*	−70.2	−84.4	−14.2	69.0	86.9	17.9	1.9
C2–O1–P–O3	178.2	164.6	−13.6	−40.3	−24.7	15.6	1.0
C2–O1–P–O4	40.9	30.9	−10.0	−177.2	−158.2	19.0	4.5

After the initial self-consistent parametrization targeting QM data, the resulting parameters were tested in simulations of small monosaccharide crystals. Special attention was paid to the experimental C1–O1, O1–P/S, and P/S–O2/O3/O4 distances, as studies have shown that the MP2 level of theory tends to overestimate the O1–P/S bond length and underestimate the C1–O1 bond length.^{80–82} The present results were consistent with these prior observations, with the P/S–

O2/O3/O4 distances being slightly overestimated in the MP2 calculations as compared to the crystal geometries. The extent of deviation of QM data from crystal data was dependent on the charge state of the phosphate group, with monoanionic phosphates showing larger deviations in P–O2 distances and dianionic phosphates in O1–P distances. Table S1 of the Supporting Information lists average bond lengths and angles from MP2-optimized geometries of model compounds THP1–

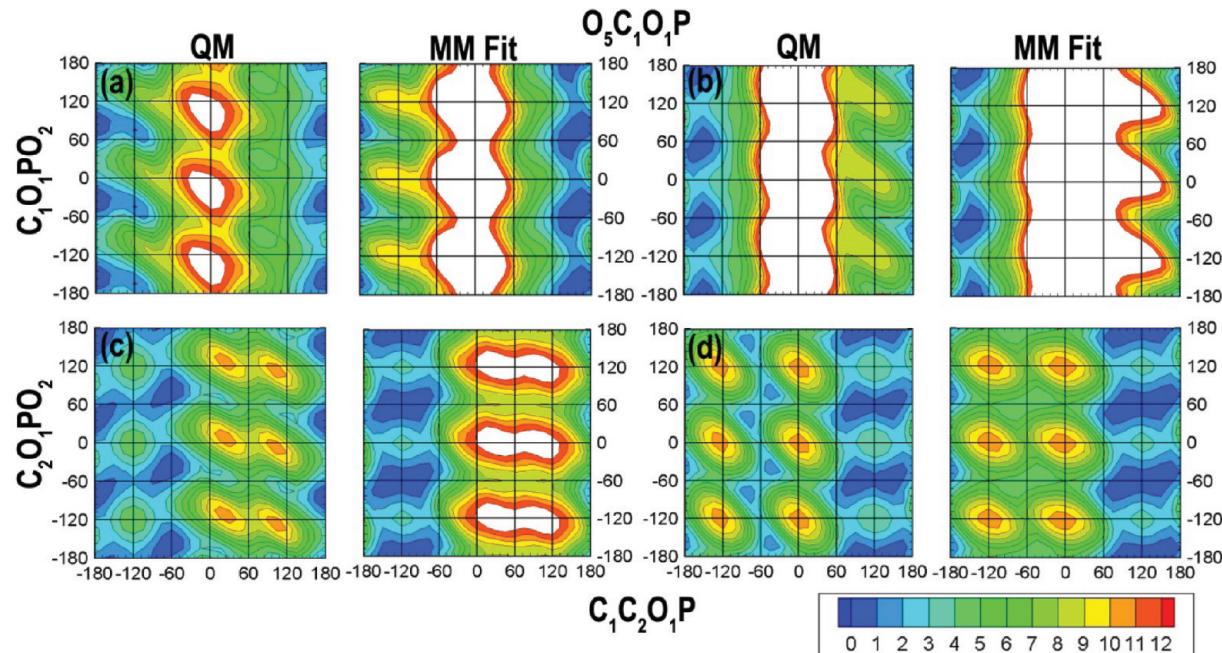


Figure 3. 2D-dihedral potential energy scans about the O₅—C₁—O₁—P/C₁—O₁—P—O₂ dihedrals for model compounds (a) THP3 and (b) THP4 and about the C₁—C₂—O₁—P/C₂—O₁—P—O₂ dihedrals for model compounds (c) CYX3 and (d) CYX4. Energies are in kcal/mol, with contours every 1 kcal/mol. Only energies below 12 kcal/mol have been plotted for the sake of clarity. In each case, the left panel depicts the QM scan and the right panel depicts the MM scan.

THP4 and CYX1–CYX4 and from relevant crystal structures to illustrate these systematic deviations. The final set of parameters was optimized to reproduce, to the best possible extent, both the crystal and QM internal geometries. All presented data are from the final set of self-consistently optimized parameters. These parameters are available as “toppar_carb_sep11.tgz” for download at <http://mackerell.umaryland.edu>.

Monoanionic Phosphates. To parametrize dihedral rotation about the C₁—O₁ and O₁—P bonds, 2D MP2/cc-pVTZ//MP2/6-31+G(d) scans were performed on the O₅—C₁—O₁—P/C₁—O₁—P—O₂ surfaces for model compounds THP1 and THP2 (Figure 2a and b) in 15° increments yielding $2 \times 24 \times 24 = 1152$ conformations. During the QM scans, the O₁—P—O₂—H dihedral was fixed in an *anti* conformation (180°). This was done to avoid contamination of the QM energy surface by rotation of the —OH group in response to energy minimization, which could in turn lead to intramolecular hydrogen bonding between the phosphate group and the ring oxygen. No other constraints were applied so as to allow full relaxation of all other degrees of freedom, including ring deformations. The MCSA fitting procedure was used to simultaneously fit the dihedral parameters for O₅—C₁—O₁—P, C₂—C₁—O₁—P, C₁—O₁—P—O₂, and C₁—O₁—P—(O₃/O₄). To ensure faithful reproduction of conformational energies, conformations with energies above 14 kcal/mol were given weights w_i (eq 2) of 0 and all other conformations weights of 1. After fitting, the low-energy regions observed in the QM surfaces are well-represented by the force field, and the behavior of the conformational energies is consistent with previous *ab initio* studies on THP1 and THP2,⁸⁰ wherein the investigators performed a 1D scan about the O₅—C₁—O₁—P dihedral, which corresponds to a 1D slice of the 2D surfaces presented here.

Minimum-energy QM geometries from the 2D QM scans were subjected to full optimization without restraints to arrive at the respective global minima. These QM-optimized geometries were then used to compare bond lengths, valence angles, and dihedral angles to those of MM-optimized geometries (Table 2). Importantly, the MM data were generated using the final set of parameters, which incorporated crystal data, thereby leading to underestimation of the O₁—P and P—O₂ QM bond lengths with respect to the MP2 data. All other bonds lengths are well reproduced by the final parameter set. Among the bond angles, the O₃—P—O₄ angle has the largest deviation when compared to QM data. The angle between the charged phosphate oxygens is well reproduced for all of the dianionic QM structures and all of the crystal structures studied, as discussed subsequently (for example, for the monoanionic crystal structure, the difference between the MD-averaged bond angle and the crystallographic bond angle is 0.1°). It was found that trying to tune the equilibrium value corresponding to this bond angle to reproduce the monoanionic QM data caused significant deviations in the dianionic phosphates and thus was not performed. The final parameters reproduce the dihedral angles, with a few exceptions. These exceptions, observed in the α -anomer for phosphate rotation (C₁—O₁—P—O₂, +11.3°) and ring distortion (C₃—C₂—C₁—O₁, -12.7°), are due to the flat profile of the local energy landscapes, as noted by the fact that the energy differences corresponding to these dihedrals’ rotations are 0.39 and 0.88 kcal/mol, respectively.

2D MP2/cc-pVTZ//MP2/6-31+G(d) scans for the CYX1 and CYX2 C₁—C₂—O₁—P/C₂—O₁—P—O₂ dihedrals reveal two minimum-energy basins for each compound (Figure 2c and d). Both the depths and breadths of these basins are captured by the parametrized MM model, as are the locations and heights of local maxima. The effect of the shallow minima is seen on comparing the QM- vs MM-minimized global

Table 3. Bond Lengths, Valence Angles, and Dihedral Angles for Model Compounds THP3 and THP4 and CYX3 and CYX4

	THP3			THP4			
	QM	MM	MM-QM	QM	MM	MM-QM	avg err
bonds							
C1–O1	1.377	1.416	0.039	1.360	1.415	0.056	0.047
O1–P	1.821	1.652	-0.170	1.836	1.658	-0.179	-0.174
P–O2	1.549	1.526	-0.023	1.554	1.526	-0.029	-0.026
P–O3	1.538	1.525	-0.014	1.537	1.525	-0.012	-0.013
P–O4	1.550	1.526	-0.025	1.547	1.526	-0.021	-0.023
angles							
C2–C1–O1	113.3	112.7	-0.6	112.6	108.8	-3.8	-2.2
O5–C1–P	108.2	109.7	1.5	107.9	109.0	1.1	1.3
C1–O1–P	117.1	119.4	2.2	112.7	116.9	4.2	3.2
O1–P–O2	101.2	105.8	4.6	100.8	105.7	4.9	4.8
O1–P–O3	100.6	104.9	4.3	100.4	105.4	4.9	4.6
O1–P–O4	101.9	105.7	3.8	102.2	105.4	3.2	3.5
O2–P–O3	117.6	113.3	-4.3	117.0	113.1	-3.8	-4.0
O2–P–O4	114.9	113.2	-1.7	114.5	112.8	-1.7	-1.7
O3–P–O4	116.4	112.9	-3.5	117.6	113.5	-4.1	-3.8
dihedrals							
C3–C2–C1–O1	-58.3	-72.6	-14.3	177.3	176.5	-0.8	-7.6
C5–O5–C1–O1	56.6	68.1	11.5	174.5	176.3	1.8	6.6
C2–C1–O1–P	-62.8	-72.5	-9.8	95.5	93.7	-1.8	-5.8
O5–C1–O1–P	178.1	162.9	-15.2	-145.7	-148.2	-2.5	-8.8
C1–O1–P–O2	-34.9	-39.8	-4.8	-55.9	-64.2	-8.3	-6.6
C1–O1–P–O3	-156.1	-159.9	-3.8	-176.2	175.8	-8.0	-5.9
C1–O1–P–O4	83.8	80.5	-3.2	62.4	55.5	-6.9	-5.0
CYX3							
	QM	MM	MM-QM	QM	MM	MM-QM	avg err
bonds							
C2–O1	1.408	1.420	0.012	1.405	1.419	0.014	0.013
O1–P	1.799	1.660	-0.138	1.798	1.661	-0.137	-0.138
P–O2	1.553	1.527	-0.027	1.542	1.526	-0.016	-0.021
P–O3	1.554	1.527	-0.028	1.553	1.526	-0.027	-0.027
P–O4	1.541	1.526	-0.016	1.557	1.526	-0.030	-0.023
angles							
C1–C2–O1	106.4	107.6	1.2	110.9	108.4	-2.5	-0.6
C3–C2–O1	111.9	109.4	-2.5	109.2	105.8	-3.4	-3.0
C2–O1–P	116.5	120.4	3.9	112.7	119.6	6.9	5.4
O1–P–O2	101.8	105.3	3.5	101.0	104.9	3.9	3.7
O1–P–O3	102.7	106.0	3.3	102.3	105.4	3.1	3.2
O1–P–O4	100.8	105.1	4.4	101.8	106.1	4.3	4.3
O2–P–O3	114.5	113.1	-1.4	116.9	113.2	-3.7	-2.6
O2–P–O4	117.1	113.2	-3.9	116.9	113.1	-3.8	-3.8
O3–P–O4	116.2	113.2	-3.1	114.2	113.2	-1.0	-2.0
dihedrals							
C4–C3–C2–O1	58.9	75.6	16.7	-177.6	-172.3	5.3	11.0
C6–C1–C2–O1	-62.1	-76.4	-14.3	177.2	170.2	-7.0	-10.6
C1–C2–O1–P	-170.6	-140.1	30.5	98.3	93.6	-4.7	12.9
C3–C2–O1–P	69.7	96.1	26.5	-139.9	-147.7	-7.8	9.3
C2–O1–P–O2	38.3	59.2	20.9	-176.4	176.8	6.8	13.8
C2–O1–P–O3	-80.5	-60.9	19.6	62.7	57.1	-5.5	7.0
C2–O1–P–O4	159.2	179.0	19.8	-55.7	-63.2	-7.6	6.1

geometries (Table 2). Significant deviations in the dihedral rotations about the O1–P bond occur for both the axial (-14.2°) and equatorial (19.0°) species. However, the energy differences corresponding to these dihedral deviations turn out to be 0.38 and 0.24 kcal/mol, respectively. Thus, as with phosphorylated THP, the dihedral deviations in CYX1 and CYX2 are due to the broad minima in the 2D surfaces. Finally, as observed for THP1 and THP2, bond lengths and angles are

reproduced in the MM minimized geometries with the exception of the O1–P and P–O2 bond lengths and the O3–P–O4 angle (Table 2).

Dianionic Phosphates. In the case of the dianionic phosphate model compounds (THP3, THP4, CYX3, and CYX4), all of the terminal oxygens on the phosphate moiety are treated as being chemically equivalent, that is, having the same atom type, unlike the monoanionic phosphate model

compounds in which one of these oxygens is protonated (Figure 1). Comparing 2D MP2/cc-pVTZ//MP2/6-31+G(d) scans for phosphate rotation in the dianionic phosphates to the parametrized MM scans shows that the optimized parameters are able to correctly locate the local minima for the β -anomer analog **THP4** and the equatorial phosphate **CYX4** (Figure 3c and d). However, for both the α -anomer analog **THP3** and the axial phosphate **CYX3**, the local minima are shifted by 30° in the horizontal axis, which track rotation about the C–O bond. Furthermore, the high-energy regions for both the α - and β -anomers are overestimated by the new parameters.

Table 3 presents a comparison between the QM- and MM-minimized geometries for **THP3**, **THP4**, **CYX3**, and **CYX4** based on unconstrained minimization of minimum-energy conformations taken from the 2D scans. Other than the C1–O1 and O1–P bonds, bonds and angles are well represented by the new parameters. With regard to dihedrals, the errors in the 2D energy surfaces for **THP3** and **CYX3** stated above manifest as a O5–C1–O1–P dihedral value underestimated by -15.2° for **THP3** and a C1–C2–O1–P dihedral value that is overestimated by 30.5° for **CYX3**. The C2–C1–O1–P/C1–C2–O1–P dihedral parameters are shared between all four model compounds since they share the same atom types. These parameters solely control the rotation about the C1–O1 bond for the cyclohexane derivatives, while in the tetrahydropyran (**THP1**–**THP4**) derivatives the O5–C1–O1–P dihedral also contributes to rotation about the C1–O1 bond. The discrepancy between the QM and MM data stems from the fact that the energy minima for **CYX3** (axial) and **THP3** (α) are significantly shifted when compared to the other model compounds, as seen in 1D QM energy profiles for the O5–C1–O1–P, C2–C1–O1–P, and C1–C2–O1–P dihedrals running through the 2D minima (Supporting Information Figure S2), and therefore could not be simultaneously fit. However, the dihedral parameters do correctly describe the conformational characteristics for six of the eight phosphates **THP1**–**THP4** and **CYX1**–**CYX4**. Also, it is shown below that the balanced dihedral parameters reproduce the O5–C1–O1–P dihedral distributions for all of the dianionic phosphate crystals studied, further justifying the parametrization approach.

Bis-phosphates. Phosphorylation at multiple positions is commonly observed in inositol phosphates, with inositol tri-, penta-, and hexa-phosphates playing diverse roles in biology, including cell growth, apoptosis, cell migration, and endocytosis.^{11,12} To develop parameters for such multiphosphorylated compounds, the tetrahydropyran derivatives **THP5**–**THP10** were chosen since they serve the two-fold purpose of developing parameters for both multiphosphorylated inositols and pyranoses. For a phosphate group at the C2/C4 position in pyranoses, parameters are required for the O5–(C1/C5)–(C2/C4)–O1 dihedral between the pyranose ring oxygen and the phosphate oxygen. To parametrize this dihedral, phosphate substitutions at C2 were chosen, and QM conformational energies were collected for both the axial (**THP5**) and equatorial positions (**THP6**). The MCSA procedure was used to fit the QM energy profile, and the newly developed parameters were used to obtain the MM energy profile. These newly developed dihedral parameters capture the QM energy profiles and correctly predict the QM minima for both the axial and equatorial substitutions (Figure 4a and b).

To parametrize the O1–C1–C2–O1a dihedral (Figure 1), model compounds with phosphate substitutions at both C1 and C2 were chosen. All four possible diasteromers **THP7** (α ,

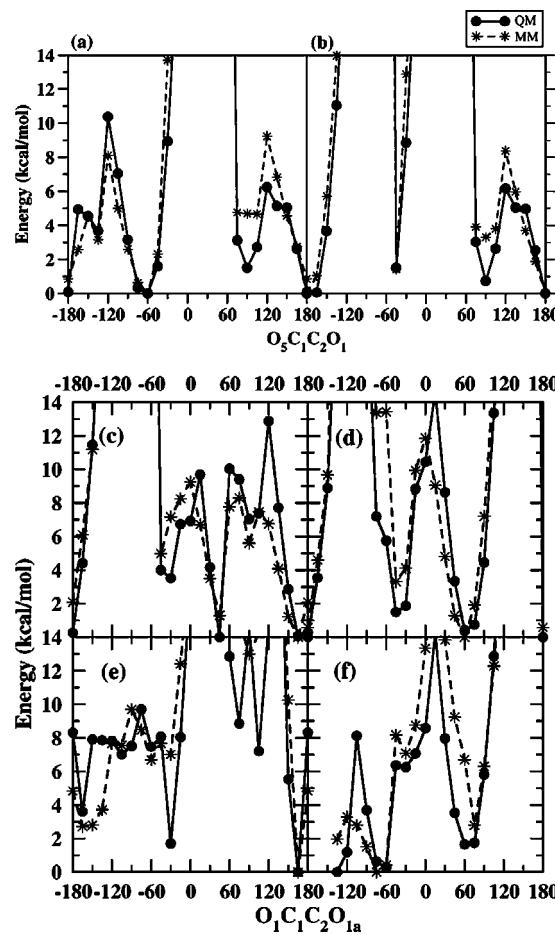


Figure 4. 1D-dihedral potential energy scans about the O5–C1–C2–O1 dihedral for (a) **THP5** and (b) **THP6** and about the O1–C1–C2–O1a dihedral for the diphosphates (c) **THP7**, (d) **THP8**, (e) **THP9**, and (f) **THP10**.

α , equatorial), **THP8** (α , equatorial), **THP9** (β , equatorial), and **THP10** (β , axial) were used for the 1D conformational scans presented in Figure 4c–f. Following MCSA fitting, a single set of O1–C1–C2–O1a dihedral parameters was able to capture the minima for all four diasteromers. Additionally, for the β , equatorial (**THP9**) diastereomer a large discrepancy of 5.3 kcal/mol is present between the QM and MM energies at the O1–C1–C2–O1a dihedral value of -30° . An inspection of the QM and MM geometries revealed the presence of two intramolecular hydrogen bonds in the QM geometry compared to none in the MM geometry. We note that this is due to the longer O1–P (O1a–P) bond in the QM geometry (~ 1.70 Å) when compared to the same in the MM geometry (~ 1.64 Å). This allows the phosphate groups to rotate about the O1–P (O1a–P) bond in the QM geometry and form the intramolecular hydrogen bond. On comparing the QM minimum energies of the four diastereomers, it was found that the α , equatorial (**THP8**) conformation was the most stable conformation followed by the α , axial (**THP7**); β , axial (**THP10**); and β , equatorial (**THP9**) conformations. The energy differences between the four diastereomers relative to the α , equatorial (**THP8**) isomer were **THP8** (0 kcal/mol) $<$ **THP7** (2.50 kcal/mol) $<$ **THP10** (6.72 kcal/mol) $<$ **THP9** (13.34 kcal/mol). Thus, the **THP8** QM minimum energy conformation was used for the comparison of QM and MM geometries (Supporting Information Table S2). The QM

Table 4. Crystalline Intramolecular Geometries for Di-Anionic Phosphate Crystals^a

	CIMDUX			JEYDAS			KGLUCP02			avg error
	CRY5	MD ^b	MD-C	CRY5	MD	MD-C	CRY5	MD	MD-C	
C1–OP1	1.419	1.416	−0.002	1.422	1.426	0.003	1.408	1.418	0.010	0.004
OP1–P1	1.645	1.620	−0.025	1.630	1.631	0.001	1.635	1.616	−0.019	−0.014
P1–OP2	1.515	1.516	0.001	1.496	1.513	0.017	1.518	1.510	−0.007	0.004
P1–OP3	1.512	1.516	0.004	1.523	1.516	−0.008	1.519	1.510	−0.009	−0.004
P1–OP4	1.522	1.517	−0.005	1.532	1.516	−0.016	1.516	1.513	−0.003	−0.008
C2–C1–OP1	109.7	111.5	1.8	111.5	114.6	3.0	109.5	110.4	0.9	1.9
O5–C1–OP1	111.1	111.0	−0.1	110.0	108.3	−1.8	111.8	112.2	0.4	−0.5
C1–OP1–P1	119.1	118.2	−0.9	118.1	117.1	−1.0	122.7	120.4	−2.3	−1.4
OP1–P1–OP2	106.3	106.9	0.6	103.1	105.3	2.2	103.1	106.0	2.9	1.9
OP1–P1–OP3	101.1	103.0	1.9	106.4	105.9	−0.5	106.0	108.0	1.9	1.1
OP1–P1–OP4	107.2	107.2	0.0	107.2	106.7	−0.5	108.6	107.1	−1.6	−0.7
OP2–P1–OP3	114.0	111.6	−2.3	114.2	113.5	−0.7	111.9	111.9	0.0	−1.0
OP2–P1–OP4	112.9	114.2	1.3	113.7	111.6	−2.2	113.3	111.6	−1.7	−0.9
OP3–P1–OP4	114.0	112.8	−1.3	111.4	113.0	1.6	113.1	111.8	−1.3	−0.3
C3–C2–C1–OP1	−65.4	−66.1	−0.7	73.1	74.1	0.9	−63.8	−61.8	2.0	0.7
C5–O5–C1–OP1	58.9	59.7	0.9	−65.7	−69.3	−3.6	62.3	55.9	−6.4	−3.0
C2–C1–OP1–P1	−150.4	−143.7	6.7	132.6	129.9	−2.7	−151.6	−149.6	2.0	2.0
O5–C1–OP1–P1	89.1	95.9	6.8	−103.6	−105.9	−2.3	87.2	91.7	4.6	3.0
C1–OP1–P1–OP2	69.6	63.3	−6.3	−175.9	−169.3	6.6	135.3	137.3 ^c	2.0	0.7
C1–OP1–P1–OP3	−171.1	−178.9	−7.8	−55.4	−48.7	6.7	−106.9	−103.8 ^d	3.1	0.7
C1–OP1–P1–OP4	−51.5	−59.7	−8.3	63.8	71.9	8.1	14.9	17.8 ^e	2.9	0.9
O2–C2–C1–OP1	56.4	61.2	4.8	−47.0	−53.9	−6.9	57.7	64.8	7.1	1.6

^aSee Supporting Information Figure S5 for atom labeling. ^b4-ns MD average computed across 4000 snapshots and averaged over all *n* monosaccharides in the simulated complete unit cell (CIMDUX *n* = 4, JEYDAS, KGLUCP02 *n* = 2; 95% confidence intervals, calculated as 1.96(RMS fluctuation)/sqrt(4000*n*), were < 0.001 Å for bonds, < 0.1° for angles, and <0.4° for dihedrals. The error for C1–OP1–P–OP2/OP3/OP4 (footnotes c–e) was ~2.0°. ^cC1–OP1–P–OP2 sampled four minima: 137.3°, −175.4°, −104.3°, −56.3°. ^dC1–OP1–P–OP3 sampled four minima: −103.8°, −56.0°, 17.6°, 65.1°. ^eC1–OP1–P–OP4 sampled four minima: 17.8°, 65.2°, 136.5°, −175.7°.

geometries are well reproduced by the MM parameters, with the exception of rotation of the phosphate group about the C2–O1a dihedral (Supporting Information Figure S3). Upon restraining this dihedral 5.0° away from the QM value, all of the dihedrals fall within acceptable deviations from the QM geometry, with the largest deviation being 10.8°. The energy cost of restraining this dihedral is 0.68 kcal/mol, which is comparable to thermal fluctuations at room temperature (*RT* = 0.59 kcal/mol). This deviation is due to the broad energy minima associated with the C–C–O–P dihedral, as observed in the monophosphates for which the energy cost of going from, e.g., 90° to 160° in CYHX2 is less than 1 kcal/mol (Figure 2d). The energy cost of restraining the dihedral to the observed QM value is found to be 0.84 kcal/mol, with the largest deviation being −11.8° (Supporting Information Table S2).

Hydroxyl–Phosphate Dihedral. In most phosphate-substituted carbohydrate systems, the carbon atoms adjacent to the phosphate-substituted carbon are substituted with a hydroxyl. To parametrize the dihedral angle involving the phosphate ether oxygen and the hydroxyl oxygen (HO₃PO–C–C–OH), myo-inositol and related substituted monoionic phosphates were used as model compounds (Supporting Information Figure S1, INI1–INI4). Similar to the bisphosphates, conformational scans about the O₃PO–C–C–OH dihedral were performed for the four possible diastereomers: INI1 (ax, ax), INI2 (ax, eq), INI3 (eq, eq), and INI4 (eq, ax). The MM model obtained from MCSA fitting, which uses a single set of parameters for all four model compounds, faithfully captures the minima for two of the conformational scans (ax, eq: Supporting Information Figure

S4b and eq, eq: Figure S4c). For the ax, ax scan (Figure S4a), the ordering of the local minimum by energy was shifted relative to the QM values and could not be adjusted by altering the weighting in the MCSA fit. Last, for the eq, ax scan (Figure S4d) at the $\phi = -60^\circ$ point, the MM minimum is shifted from the QM at -75° , and this was shown to be due to formation of an intramolecular hydrogen bond. On analyzing the MM energy surface in the absence of this point, the MM parametrization correctly captures the QM behavior (Figure S4d). As discussed below, the HO₃PO–C–C–OH dihedral is well reproduced for all monosaccharide crystals. An inspection of the eq, eq (INI3) scan (Figure S4c) revealed an underestimation of the MM energies at dihedral values of 135°, 150°, and 165°. A closer analysis of these conformations revealed that the inositol ring adopted a boat conformation with the phosphate group being present in the axial conformation at the flagpole atom (end atoms), leading to substantial steric repulsion with the flagpole hydrogen atom on the other end of the molecule and thus the high energy in the QM profile. We note that this is an unfavorable conformation in the transition path (chair-boat transition) as the ring generally distorts by pushing two hydrogen atoms into the flagpole position to reduce the steric interactions.⁸³ Interestingly, these conformations with two hydrogen atoms in the flagpole position were well predicted by the MM scans (Figure S4a: -30° , -45° , -60° ; Figure S4b: -60° , -45°). Thus, the discrepancy in the QM and MM scans (at 135°, 150°, and 165°) was considered acceptable as it would not adversely affect the results of an MD simulation. However, sampling of conformations akin to 135°, 150°, and 165° could lead to an unfavorable stabilization of these states.

Table 5. Crystalline Intramolecular Geometries for Monoanionic and Neutral Phosphate Crystals^a

	JUGTAG			MINOSP ^c		
	CRYST	MD ^b	MD-C	CRYST	MD	MD-C
C1–OP1	1.423	1.413	−0.009	C2–OP1	1.458	1.436
OP1–P1	1.595	1.618	0.022	OP1–P1	1.586	1.632
P1–OP2*	1.575	1.576	0.001	P1–OP2*	1.548	1.573
P1–OP3	1.480	1.519	0.039	P1–OP3*	1.553	1.579
P1–OP4	1.500	1.516	0.015	P1–OP4	1.47	1.519
C2–C1–OP1	106.5	110.9	4.5	C1–C2–OP1	109.9	108.2
O5–C1–OP1	111.6	111.2	−0.3	C3–C2–OP1	108.1	108.8
C1–OP1–P1	124.8	120.5	−4.3	C2–OP1–P1	121.1	121
OP1–P1–OP2*	102.1	104.8	2.7	OP1–P1–OP2*	107.3	107.5
OP1–P1–OP3	111.6	109.9	−1.7	OP1–P1–OP3*	108.5	108.9
OP1–P1–OP4	104.5	106.6	2.1	OP1–P1–OP4	108	111.6
OP2*–P1–OP3	111.4	109.8	−1.5	OP2*–P1–OP3*	104.4	102.8
OP2*–P1–OP4	110.3	108.8	−1.5	OP2*–P1–OP4	115.7	112.5
OP3–P1–OP4	115.9	116.0	0.1	OP3*–P1–OP4	112.6	112.8
C3–C2–C1–OP1	−71.9	−71.0	1.0	C4–C3–C2–OP1	68.5	76.5
C5–O5–C1–OP1	60.8	61.4	0.7	C6–C1–C2–OP1	−64	−70.6
C2–C1–OP1–P1	−174.7	−174.2	0.5	C1–C2–OP1–P1	−123.6	−126.2
O5–C1–OP1–P1	63.7	64.4	0.6	C3–C2–OP1–P1	115.3	111.2
C1–OP1–P1–OP2*	39.5	43.5	4.0	C2–OP1–P1–OP2*	59.2	59.6
C1–OP1–P1–OP3	−79.5	−74.5	5.1	C2–OP1–P1–OP3*	−53.1	−51.1
C1–OP1–P1–OP4	154.4	158.8	4.4	C2–OP1–P1–OP4	−175.5	−176.4
O2–C2–C1–OP1	53.9	56.8	2.9	O1–C1–C2–OP1	62.7	60.9
				C4–C3–C2–OP1	−68.5	−76.4
				C6–C1–C2–OP1	63.9	70.5
				C1–C2–OP1–P1	123.6	126
				C3–C2–OP1–P1	−115.3	−111.5
				C2–OP1–P1–OP2*	−59.2	−59.4
				C2–OP1–P1–OP3*	53.1	51.3
				C2–OP1–P1–OP4	175.5	176.6
				O1–C1–C2–OP1	−62.7	−60.9
						1.744

^aSee Supporting Information Figure S5 for atom labeling. ^b4-ns MD average computed across 4000 snapshots and averaged over all *n* monosaccharides in the simulated complete unit cell (JUGTAG, MINOSP *n* = 4; 95% confidence intervals, calculated as 1.96(RMS fluctuation)/sqrt(4000*n*), were < 0.001 Å for bonds, < 0.1° for angles, and <0.4° for dihedrals. ^cUnit cell contained monosaccharides with two different orientations of the phosphate groups in the molecules of carbohydrate in the unit cell.

Table 6. Crystalline Unit Cell Geometries and Volumes

	a			b			c		
	exptl	MD ^a	% err ^b	exptl	MD	% err	exptl	MD	% err
CIMDUX	8.446	8.287	−1.88	10.203	10.101	−1.00	16.594	16.365	−1.38
JEYDAS	6.228	6.175	−0.85	14.600	14.127	−3.24	8.982	8.951	−0.34
KGLUCP02	10.458	10.201	−2.46	9.027	9.291	2.93	7.532	7.202	−4.38
JUGTAG	7.351	7.327	−0.33	9.666	9.648	−0.19	15.230	15.331	0.67
MINOSP	6.810	6.997	2.75	16.548	17.225	4.09	12.536	13.003	3.72
avg			−0.55			0.52			−0.34
	β^c			vol					
	exptl	MD	% err	exptl	MD	% err	exptl	MD	% err
CIMDUX	99.2	96.3	−2.90	1411.6	1361.4	−3.56			
JEYDAS	102.0	101.2	−0.84	798.8	765.8	−4.14			
KGLUCP02	110.4	111.8	1.24	666.5	633.6	−4.94			
JUGTAG	90.0	90.0	0.00	1082.2	1083.3	0.11			
MINOSP	133.3	135.5	1.67	1028.8	1097.9	6.72			
avg			−0.17						−1.16

^aMD values are 4 ns averages. The 95% confidence intervals for A, B, and C are <0.02 Å; for β are <0.2°; and for volumes are <0.5 Å³. ^b(MD – exptl)/exptl × 100%. ^cConstrained to 90° in the simulation if equal to 90° in the experimental crystal, otherwise allowed to vary independently during the simulation.

Crystal Simulations. A CSD survey yielded five phosphate monosaccharide crystals, which were used to test the optimized

parameters. The phosphates were found to be present in all three possible charged states. Three crystals had phosphates in

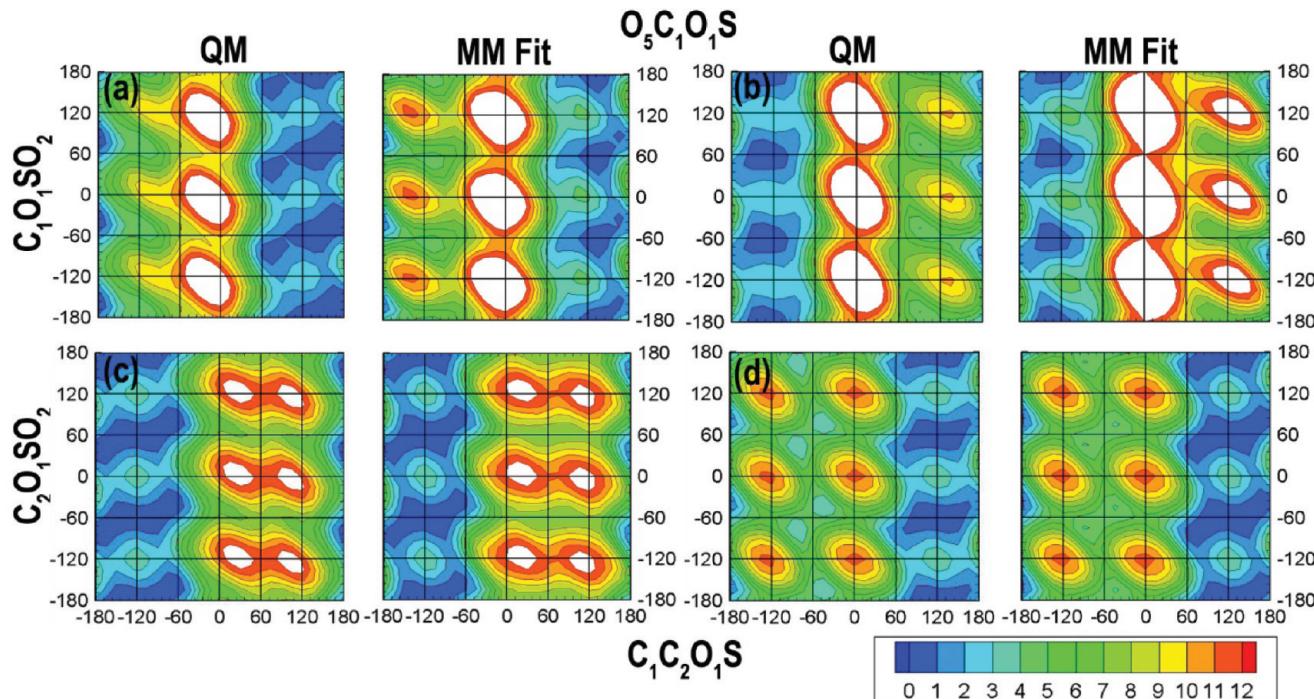


Figure 5. 2D-dihedral potential energy scans about the O5–C1–O1–S/C1–O1–S–O2 dihedrals for model compounds (a) THP11 and (b) THP12 and about the C1–C2–O1–S/C2–O1–S–O2 dihedrals for model compounds (c) CYX5 and (d) CYX6. Energies are in kcal/mol, with contours every 1 kcal/mol. Only energies below 12 kcal/mol have been plotted for the sake of clarity. In each case, the left panel depicts the QM scan and the right panel the MM scan.

the dianionic state (Supporting Information Figure S5; CSD ref code, R factor, compound name: CIMDUX, 1.73, disodium α -D-glucose-1-phosphate hydrate; JEYDAS, 1.77, dipotassium galactose 1-phosphate pentahydrate; KGLUCP02, 1.86, dipotassium glucose-1-phosphate dehydrate). One had the phosphate in a monoanionic state (JUGTAG, 1.83, potassium α -D-glucopyranosyl-hydrogenphosphate), and another had a neutral charged state (MINOSP, 1.80, myo-inositol-2-phosphate monohydrate). The presence of two different counterions in the different crystals, sodium and potassium, allowed testing of ionic interactions. MD simulations of the infinite crystals were performed on these systems, with the MD-averaged intramolecular geometries compared to the experimental values in Tables 4 and 5 and the unit cell parameters summarized in Table 6.

All bond lengths, bond angles, and dihedrals are well represented by the force field. Notably, the problematic bond lengths with regard to QM geometries (i.e., C1–O1 and O1–P) in all systems, as well as the P–O2* bond lengths in the monoanionic systems, are well represented in the crystal simulations. Even the C1–O1–P bond angle associated with the phosphate linkage is well represented in the crystal structure simulations over all the charged states. The dihedrals present the most dramatic improvement when compared to the QM-minimized geometries. None of the phosphate linkage dihedrals show a deviation greater than $\pm 8.3^\circ$ from the starting crystal structure values during the simulations. For the MINOSP crystal structure simulations (Table 5), the MD simulations are able to maintain the different orientations of the phosphate groups in the two molecules of carbohydrate in the unit cell. For one of the dianionic crystals (KGLUCP02), the phosphate group rotated about the O1–P bond and accessed four rotational states during the simulation (Table 4). Taking

into account that all of the terminal oxygen atoms in the dianionic phosphates are equivalent, these four rotational states in fact represent only two unique orientations of the phosphate group, with one of the two corresponding to the starting crystal structure.

All unit cell parameters from the MD trajectories were close to the corresponding crystallographic values (Table 6). The error for the unit cell length and angle parameters averaged over the five systems was -0.55% (a), 0.52% (b), 0.34% (c), and -0.17% (β). The average error in the molecular volume was found to be -1.16% . The present model overestimates the molecular volume for the neutral species (MINOSP, 6.72%), a trend that is consistent with the results in the CHARMM force field for hexopyranose and furanose monosaccharides, linear sugars and sugar alcohols, and disaccharides.^{36–41}

Monoanionic Sulfates. To parametrize dihedral rotation about the C1–O1 and O1–S bonds in sulfated carbohydrates, 2D MP2/cc-pVTZ//MP2/6-31+G(d) scans were performed on the O5–C1–O1–S/C1–O1–S–O2 surfaces for model compounds THP11 and THP12 (Figure 5a and b). As described for the phosphates, the MCSA fitting procedure was used to simultaneously fit the dihedral parameters for O5–C1–O1–S, C2–C1–O1–S, and C1–O1–S–(O2/O3/O4). During fitting, conformations with energies above 14 kcal/mol were given weights w_i (eq 2) of 0 and all other conformations weights of 1. Analysis of Figure 5a and b shows the low-energy basins observed in the QM energy surfaces to be well represented by the MM energy surfaces.

MM internal geometries were compared to those from the unconstrained QM optimized geometries (Table 7). The O1–S bond lengths present the major disagreement, with the MM parameters underestimating the O1–S bond length by -0.141 \AA averaged over THP11 and THP12. The valence angles and

Table 7. Bond Lengths, Valence Angles, and Dihedral Angles for Model Compounds THP11-12 and CYX5-6

	THP11			THP12			
	QM	MM	MM-QM	QM	MM	MM-QM	avg err
bonds							
C1–O1	1.426	1.412	-0.014	1.401	1.415	0.014	0.000
O1–S	1.718	1.580	-0.138	1.727	1.583	-0.144	-0.141
S–O2	1.486	1.452	-0.033	1.483	1.452	-0.031	-0.032
S–O3	1.486	1.453	-0.033	1.487	1.453	-0.034	-0.034
S–O4	1.474	1.452	-0.023	1.474	1.451	-0.023	-0.023
angles							
C2–C1–O1	112.3	115.1	2.7	110.5	106.6	-3.8	-0.6
O5–C1–S	106.3	108.3	2.0	106.0	107.7	1.7	1.8
C1–O1–S	116.6	116.1	-0.4	113.3	113.3	-0.1	-0.3
O1–S–O2	104.2	104.8	0.6	104.3	105.6	1.3	0.9
O1–S–O3	103.4	107.4	4.0	103.0	105.4	2.4	3.2
O1–S–O4	100.8	103.0	2.3	100.4	103.6	3.2	2.7
O2–S–O3	114.1	113.9	-0.2	113.9	113.8	-0.1	-0.2
O2–S–O4	115.3	113.1	-2.2	116.3	113.7	-2.6	-2.4
O3–S–O4	116.3	113.4	-3.0	116.0	113.5	-2.6	-2.8
dihedrals							
C3–C2–C1–O1	-60.9	-73.2	-12.3	174.8	173.4	-1.4	-6.8
C5–O5–C1–O1	61.2	72.9	11.7	177.3	-178.2	4.5	8.1
C2–C1–O1–S	-68.2	-65.5	2.6	108.7	105.6	-3.1	-0.3
O5–C1–O1–S	170.8	168.6	-2.2	-132.0	-136.9	-4.9	-3.5
C1–O1–S–O2	83.2	88.0	4.7	62.0	62.6	0.6	2.7
C1–O1–S–O3	-36.4	-33.6	2.8	-57.2	-58.2	-0.9	0.9
C1–O1–S–O4	-157.0	-153.5	3.5	-177.2	-177.6	-0.4	1.5
CYX5				CYX6			
	QM	MM	MM-QM	QM	MM	MM-QM	avg err
bonds							
C2–O1	1.441	1.419	-0.022	1.437	1.419	-0.018	-0.020
O1–S	1.708	1.590	-0.118	1.707	1.592	-0.116	-0.117
S–O2	1.486	1.453	-0.033	1.488	1.452	-0.036	-0.035
S–O3	1.488	1.453	-0.035	1.487	1.453	-0.034	-0.035
S–O4	1.477	1.452	-0.024	1.476	1.452	-0.024	-0.024
angles							
C1–C2–O1	106.4	107.1	0.6	107.9	105.4	-2.5	-0.9
C3–C2–O1	109.0	109.3	0.2	109.5	108.3	-1.2	-0.5
C2–O1–S	114.0	116.8	2.8	113.5	115.7	2.2	2.5
O1–S–O2	104.1	105.0	0.9	103.9	106.3	2.4	1.6
O1–S–O3	104.0	106.2	2.3	104.1	105.2	1.2	1.7
O1–S–O4	100.6	103.4	2.8	100.8	103.2	2.4	2.6
O2–S–O3	113.8	114.0	0.1	113.7	114.1	0.4	0.2
O2–S–O4	115.8	113.4	-2.4	115.9	113.4	-2.5	-2.4
O3–S–O4	115.9	113.5	-2.3	115.8	113.4	-2.4	-2.4
dihedrals							
C4–C3–C2–O1	62.4	73.4	11.0	-175.6	-170.5	5.1	8.1
C6–C1–C2–O1	-63.6	-74.6	-11.0	176.3	172.7	-3.5	-7.3
C1–C2–O1–S	-137.2	-137.5	-0.3	131.1	142.3	11.2	5.5
C3–C2–O1–S	102.1	99.7	-2.4	-107.3	-99.7	7.6	2.6
C2–O1–S–O2	62.0	63.4	1.4	57.0	59.3	2.2	1.8
C2–O1–S–O3	-57.5	-57.6	-0.2	-62.3	-62.1	0.2	0.0
C2–O1–S–O4	-177.7	-177.4	0.3	177.4	178.9	1.5	0.9

dihedral angles are well represented for both the α and β configurations (THP11 and THP12), with the maximum average error for the valence angles and dihedral angles being 3.2° and 8.1° , respectively.

Figure 5c and d show 2D MP2/cc-pVTZ//MP2/6-31+G(d) scans performed on the C1–C2–O1–S/C2–O1–S–O2 surfaces for the cyclohexane derivatives CYX5 and CYX6 along with the final MM surfaces. The parameters developed

using THP11 and THP12 accurately reproduce the low energy regions of these surfaces. In Table 7, the QM and MM minimized geometries are compared. As in THP11 and THP12, the O1–S bond length averaged over CYX5 and CYX6 is underestimated by -0.117 \AA . The valence angles and dihedral angles are again well represented with the maximum average error being 2.6° and 8.1° , respectively. Notably, the average error for the O1–S bond averaged over three crystal

Table 8. Crystalline Intramolecular Geometries for Sulfate Crystals^a

	SOJHAA			POCSOP			HAHZEV ^b			avg err ^c		
	CRYs	MD ^b	MD-C	CRYs	MD	MD-C	CRYs	MD	MD-C			
C4–OS1	1.466	1.422	-0.045	C3–OS1	1.458	1.431	-0.028	C2–OS1	1.471	1.430	-0.041	-0.038
OS1–S1	1.572	1.562	-0.010	OS1–S1	1.599	1.581	-0.018	OS1–S1	1.604	1.581	-0.023	-0.017
S1–OS2	1.443	1.453	0.011	S1–OS2	1.416	1.449	0.033	S1–OS2	1.448	1.451	0.003	0.016
S1–OS3	1.455	1.448	-0.007	S1–OS3	1.450	1.451	0.001	S1–OS3	1.454	1.454	0.000	-0.002
S1–OS4	1.450	1.449	-0.001	S1–OS4	1.442	1.448	0.005	S1–OS4	1.438	1.452	0.013	0.006
C3–C4–OS1	108.8	110.7	1.8	C2–C3–OS1	104.2	110.0	5.7	C1–C2–OS1	109.3	110.7	1.5	3.0
C5–C4–OS1	106.1	107.8	0.8	C4–C3–OS1	111.8	112.2	0.4	C3–C2–OS1	105.6	109.4	3.8	1.7
C4–OS1–S1	121.3	122.0	0.7	C3–OS1–S1	119.0	117.6	-1.3	C2–OS1–S1	119.0	118.3	-0.6	-0.4
OS1–S1–OS2	101.8	102.5	-2.1	OS1–S1–OS2	101.4	103.0	1.6	OS1–S1–OS2	100.4	101.6	1.2	0.2
OS1–S1–OS3	108.0	105.9	-1.2	OS1–S1–OS3	106.0	106.6	0.6	OS1–S1–OS3	106.1	105.6	-0.5	-0.4
OS1–S1–OS4	108.3	107.1	0.4	OS1–S1–OS4	107.1	106.1	-1.0	OS1–S1–OS4	107.3	106.3	-1.0	-0.6
OS2–S1–OS3	113.1	113.5	-0.5	OS2–S1–OS3	114.1	113.3	-0.8	OS2–S1–OS3	113.3	113.9	0.7	-0.2
OS2–S1–OS4	114.0	113.5	2.2	OS2–S1–OS4	114.6	112.9	-1.8	OS2–S1–OS4	115.2	113.3	-1.8	-0.5
OS3–S1–OS4	111.0	113.2	-4.6	OS3–S1–OS4	112.2	113.7	1.4	OS3–S1–OS4	113.2	114.3	1.2	-0.7
C2–C3–C4–OS1	-56.3	-60.9	3.3	C1–C2–C3–OS1	-172.9	-175.0	-2.1	C4–C3–C2–OS1	-170.0	-177.5	-7.4	-2.1
				C5–C4–C3–OS1	175.6	-178.4	6.0					
C3–C4–OS1–S1	-122.5	-119.2	0.1	C2–C3–OS1–S1	-145.2	-149.3	-4.1	C1–C2–OS1–S1	97.0	94.0	-3.0	-2.4
C5–C4–OS1–S1	122.1	122.5	0.1	C4–C3–OS1–S1	93.9	89.9	-4.0	C3–C2–OS1–S1	-141.2	-144.6	-3.4	-2.4
O5–C5–C4–OS1	52.1	57.6	0.5					O5–C1–C2–OS1	171.5	178.3	6.8	
C4–OS1–S1–OS2	-160.9	-160.8	-0.7	C3–OS1–S1–OS2	177.4	174.5	-2.9	C2–OS1–S1–OS2	-165.6	-167.0	-1.4	-1.7
C4–OS1–S1–OS3	79.8	79.9	1.5	C3–OS1–S1–OS3	57.9	54.9	-3.0	C2–OS1–S1–OS3	76.3	73.8	-2.6	-1.4
C4–OS1–S1–OS4	-40.4	-41.1	1.5	C3–OS1–S1–OS4	-62.2	-66.7	-4.5	C2–OS1–S1–OS4	-44.9	-48.1	-3.2	-2.1
O3–C3–C4–OS1	65.3	66.8	1.5	O2–C2–C3–OS1	65.4	60.0	-5.4	O3–C3–C2–OS1	70.4	58.3	-12.1	-5.3
				O4–C4–C3–OS1	55.3	59.7	4.4					

^aSee Supporting Information Figure S7 for atom labeling. ^b4-ns MD average computed across 4000 snapshots and averaged over all n monosaccharides in the simulated complete unit cell (SOJHAA, POCSOP $n = 4$; HAHZEV $n = 2$; 95% confidence intervals, calculated as $1.96(\text{RMS fluctuation})/\sqrt{4000n}$, were $< 0.001 \text{ \AA}$ for bonds, $< 0.1^\circ$ for angles, and $< 0.4^\circ$ for dihedrals. ^cAverage error for the relevant dihedral is evaluated only if the dihedral is shared in all of the crystals. ^bIntramolecular geometries for the C6-sulfate substitution are presented in Table S4.

structure simulations (discussed below) is -0.017 \AA , which is much lower than the average errors of -0.141 \AA and -0.117 \AA for **THP11** and **THP12** and **CYX5** and **CYX6**,

C-6 Sulfates. Since sulfates occur at the C6 position of the monosaccharides in chondroitin (chondroitin-6-sulfate (C6S)),^{17,18} the parametrization was extended to include this substitution. A tetrahydropyran ring with a $-\text{CH}_2\text{OSO}_3$ group attached to C5 in an equatorial configuration was used to model the attachment of the sulfate group to C6 (**THP13**, Figure 1). After transferring parameters for bonds, angles, and dihedrals, only new dihedral parameters for rotation about the C5–C6 and C6–O1 bonds were required. To parametrize these terms, 2D MP2/cc-pVTZ//MP2/6-31+G(d) scans were performed to generate a C4–C5–C6–O1/C5–C6–O1–S1 energy surface for the model compound (Supporting Information Figure S6a). Following MCSA optimization, the MM energy surface (Supporting Information Figure S6b)

captures most of the QM energy surface features, although there is a slight shift in the local minima in the upper left quadrant of the surface, with the relative energies of the minimum energy regions (i.e., $-60^\circ/135^\circ$ and $-105^\circ/75^\circ$) being reversed. However, the MM model faithfully captures the global minimum located at approximately $-45^\circ/-75^\circ$. Supporting Information Table S3 includes the geometric descriptors of the QM- and MM-minimized geometries for **THP13**, showing the MM model to accurately reproduce the QM-minimized geometry.

Crystal Simulations. A CSD survey yielded three sulfate monosaccharide crystals which were used to test the optimized parameters (Supporting Information Figure S7; CSD ref code, R factor, compound name: SOJHAA, 1.69, methyl α -D-galactopyranoside 4-(sodium sulfate) dehydrate; POCSOP, 1.72, sodium methyl- α -D-galactopyranoside-3-sulfonate mono-hydrate; HAHZEV, 1.79, methyl α -D-galactopyranoside 2,6-

Table 9. Crystalline Unit Cell Geometries and Volumes

	a			b			c		
	exptl	MD ^a	% err ^c	exptl	MD	% err	exptl	MD	% err
SOJHAA	28.618	29.156	1.88	5.800	5.821	0.37	7.983	7.869	-1.42
POCSOP	6.896	6.590	-4.44	12.916	13.122	1.59	13.654	14.164	3.74
HAHZEV	10.904	11.758	7.83	5.511	5.19	-5.83	13.443	13.025	-3.11
			1.76			-1.29			-0.27
	β^b			vol					
	exptl	MD	% err	exptl	MD	% err	exptl	MD	% err
SOJHAA	99.4	101	1.62	1307.2	1310.9	0.28			
POCSOP	90	90	0	1216.1	1224.4	0.68			
HAHZEV	93.1	95.7	2.75	806.6	789.4	-2.14			
			1.46						-0.39

^aMD values are 4 ns averages. The 95% confidence intervals for A, B, and C are <0.02 Å; for β are <0.2°; and for volumes are <0.5 Å³. ^bConstrained to 90° in the simulation if equal to 90° in the experimental crystal, otherwise allowed to vary independently during the simulation. ^c(MD – exptl)/exptl × 100%.

bis(sodium sulfate) dihydrate). The three crystals covered four different sulfate attachment points on the monosaccharide at C2, C3, C4, and C6, such that they act as a good test in preparation for calculations on more complex sulfated oligosaccharides like chondroitin-4-sulfate (C4S) and chondroitin-6-sulfate (C6S). The results comparing the intramolecular geometrical descriptors from MD simulations for the three crystals are summarized in Table 8.

The bond lengths, valence angles, and dihedrals are well represented by the final set of parameters. As mentioned above, the average error for the O1–S bond is found to be -0.017 Å, which is considerably lower than the average errors of -0.141 Å and -0.117 Å for the QM-minimized structures of **THP11** and **THP12** and **CYX5** and **CYX6**. The highest average errors for the valence angles and dihedral angles are 3.0° and -5.3°, indicating satisfactory agreement of the MD data with the experimental values. Supporting Information Table S4 includes the results for the sulfate attachment at the C6 position in the HAHZEV crystal. The bond lengths and valence angles are reproduced satisfactorily by the MM model, with the largest average errors being 0.027 Å and 2.9°, respectively. Analyzing the dihedrals shows the sulfate group at C6 samples two distinct conformations associated with the C5–C6 and C6–OS5 bonds (See Supporting Information Figure S7 for atom naming and numbering), while the sulfate group rotates freely about the OS5–S2 bond. Since the three sulfate oxygens are identical in the MM representation, pooled MD data for the three dihedrals (C6–OS5–S2–OS6/7/8) were used to analyze the dihedral distribution. This analysis shows that the dihedral samples three distinct conformational bins separated by ~60°, which is in agreement with the crystallographic values of 179.2°, -60.7°, and 60.7° (Supporting Information Table S4). Thus, the MM parameters are able to describe the conformational characteristics of a C6 sulfate attachment, thereby enabling the study of chondroitin-6-sulfate (C6S). Additionally, the unit cell parameters averaged over the MD trajectories are in agreement with experimental crystallographic values (Table 9). Errors for unit cell parameters averaged over the three crystal structure simulations are 1.76% (a), -1.29% (b), -0.27% (c), and 1.46% (β), and for the unit cell volume, the average error is -0.39%.

Protein:Carbohydrate Systems: Phosphates. Inositol phosphates (phosphoinositides) play critical roles in the localization and assembly of protein complexes involved in

signal transduction, membrane trafficking, and cytoskeletal dynamics.^{84–86} Phosphoinositides bind to pleckstrin homology (PH) domains, which are protein modules of 120 amino acids found in many proteins.^{86,87} Despite high sequence variability, PH domains contain a common core structure consisting of a partly open barrel capped at one end by a C-terminal α -helix.⁸⁸ A subset of PH domains has been shown to bind phosphoinositide head groups with affinity and specificity.^{89,90} To validate the present parameters in the context of interactions with proteins, calculations were performed on tri-, tetra-, and penta-inositol phosphates interacting with the PH domains of 3G ARNO (the triglycine variant of the Arf nucleotide binding site opener; PDB id: 1U29),⁶⁹ protein kinase B (PKB/Akt; PDB id: 1UNQ),⁷⁰ and Grp1 (a general receptor for phosphoinositides; PDB id: 1FHW),⁷¹ respectively.

For all systems studied, the overall RMSD of the complete system remains lower than 4 Å for the entire simulation length (Figure 6). The high RMSD values of 3 Å and 4 Å for protein

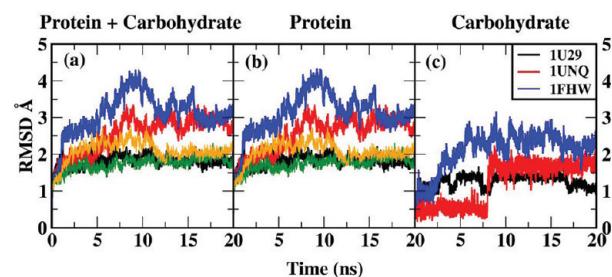


Figure 6. RMS difference (RMSD) analysis of inositol phosphate/pleckstrin homology (PH) domain glycoprotein systems. (a) RMSD for all carbohydrate–protein heavy atoms, (b) RMSD for the protein heavy atoms only, and (c) RMSD for carbohydrate heavy atoms only. RMSDs were calculated following the alignment of all non-hydrogen atoms of protein and carbohydrate for a, all non-hydrogen atoms of protein for b, and all non-hydrogen atoms of the carbohydrate for c. In a and b, the RMSD excluding the flexible terminal regions for 1UNQ (green line) and 1FHW (orange line) is also shown.

kinase B (PKB/Akt) and Grp1 PH domains are due to the flexible terminal regions in these proteins that do not interact with the rest of the protein or the carbohydrates. The overall RMSD excluding these regions (the first five and last three amino acids for protein kinase B (PKB/Akt) and the last six

Table 10. Significant Phosphate–Protein Contact Occupancies from MD Simulations of the PH Homology Domains^a

1U29			1UNQ			1FHW (domain A) ^b		
Ins(1,4,5)P ₃			Ins(1,3,4,5)P ₄			Ins(1,3,4,5,6)P ₅		
amino acid	phosphate	occu	amino acid	phosphate	occu	amino acid	phosphate	occu
phosphate-1								
	Ile 19 (N)			OP4 (2.76)	0.997			
	Arg 23 (NH ₂)			OP3 (2.80)	0.975			
	Tyr 18 (N)			OP4 (2.85)	0.929			
	Glu 17 (N)			OP1 (3.37)	0.889			
	Arg 23 (NH ₂)			OP3 (3.04)	0.783			
	Arg 23 (NH ₂)			OP4 (3.10)	0.709			
	Arg 23 (C _ε)			OP3 (3.38)	0.659			
	Glu 17 (N)			OP4 (5.33)	0.623			
	Ile 19 (O)			OP4 (3.56)	0.569			
phosphate-3								
	Arg 25 (N _ε)			OP4 (2.93)	0.989	Arg 284 (N _ε)	OP4 (2.84)	0.999
	Arg 25 (NH ₂)			OP3 (2.98)	0.954	Arg 284 (NH ₂)	OP2 (2.84)	0.973
	Lys 14 (N _ε)			OP1 (2.82)	0.927	Lys 273 (C _ε)	OP4 (3.73)	0.780
						Lys 273 (N _ε)	OP1 (3.31)	0.721
						Lys 273 (N _ε)	OP4 (3.01)	0.593
						Arg 305 (NH ₁)	OP3 (2.94)	0.549
phosphate-4								
Lys 273 (N _ε)	OP3 (2.80)	0.994				Lys 273 (N _ε)	OP3 (2.83)	1.000
Arg 285 (NH ₂)	OP4 (2.83)	0.986				Hse 355 (N _{ε2})	OP3 (3.50)	0.801
Arg 285 (N _ε)	OP4 (2.86)	0.895				Lys 343 (N _ε)	OP4 (4.77)	0.690
Arg 285 (C _ε)	OP4 (3.47)	0.864				Asn 354 (N _{δ2})	OP2 (4.63)	0.604
Tyr 296 (OH)	OP2 (2.67)	0.804						
Lys 273 (C _ε)	OP3 (3.25)	0.788						
phosphate-5								
His 356 (N _{ε2})	OP4 (2.55)	0.908	Arg 86 (NH ₂)	OP4 (4.38)	0.925	Gly 276 (N)	OP4 (3.76)	0.990
Gly 276 (N)	OP3 (2.78)	0.900	Arg 86 (NH ₁)	OP4 (6.64)	0.752			
Asn 355 (O _{δ1})	OP2 (3.14)	0.585	Arg 86 (C _ε)	OP4 (5.55)	0.738			
phosphate-6								
						Arg 277 (NH ₂)	OP4 (7.17)	0.528

^aA cutoff distance of 3.50 Å was used to detect the close contacts. Only occupancies greater than 0.5 are presented for the sake of clarity. Contact distances in the crystal structure are presented in parentheses. Distances are in Å. ^bOnly domain A is selected for the analyses.

amino acids for Grp1) is less than 2.5 Å (shown in Figure 6a and b; green line, protein kinase B (PKB/Akt; PDB id: 1UNQ); orange line, Grp1 (PDB id: 1FHW)). Structural changes in the phosphoinositides were relatively small, being less than 3 Å in all cases, with visual inspection of the trajectories showing them to remain bound to their respective proteins.

To analyze phosphoinositide–protein interactions, contact probabilities between the phosphates in the phosphoinositides and the non-hydrogen atoms in the PH domains were calculated over the final 14 ns of the simulations. Contacts were considered significant if the occupancy was found to be greater than 0.5, with results for all systems tabulated in Table 10. For 3G ARNO, the phosphates at positions 4 and 5 in inositol(1,4,5)P₃ are in close contact with the PH domain. The phosphate group at position 4 was found to be in contact with the side chain nitrogen atoms of Lys 273 and Arg 285, which belong to the β 1 and β 2 strands of the PH domain; the phosphate group also interacted with the phenolic side chain of Tyr 296.^{69,88} The position 5 phosphate was found to be in contact with His 356 and the amide nitrogen of Gly 276. These occupancy data agree with the mutation studies for the 3G ARNO PH domain, wherein six mutations—K273A, R285A, Y296F, R306A, K344A, and H356A—resulted in the loss of binding.⁶⁹ Simulation data identify four of the experimental

mutations points, K273, R285, Y296, and H356. The occupancy data for K344 are found to be 0.423, and it interacts with the phosphate at position 5. Thus, the MM model is able to also capture the mutation at K344. For the PKB/Akt PH domain, the inositol ring in inositol(1,3,4,5)P₄ rotates during the simulation, as seen by a shift in carbohydrate RMSD (Figure 6c), and the contacts between the phosphate group at position 4 and the PH domain are lost. However, even with this change in the conformation, the force field is still able to identify all the close contact interactions as observed in the crystal structure (contact distances in the crystal structure are presented in parentheses in Table 10).⁷⁰ Close contact occupancies for domain A of the Grp1-PH/inositol(1,3,4,5,6)P₅ complex were also used to test the force field, and the amino acids involved in binding this phosphoinositide are similar to those in 3G ARNO PH (PDB id: 1U29), which shares close to 90% sequence similarity with Grp1.^{69–71} As with the previous system, Lys 273 and Arg 284 are again involved in the binding (Table 10; note that Arg 284 in Grp1 corresponds to Arg 285 in the 3G ARNO PH domain). Thus, the force field maintains stable noncovalent interactions between various inositol-phosphates and proteins on the 20 ns time scale, in agreement with experimental structures and mutation data, helping to validate the force field with respect to the treatment of such heterogeneous systems.

Protein:Carbohydrate System: Sulfates. Cathepsin K is a major collagenolytic enzyme produced by bone-resorbing osteoclasts, which are bone cells that removes bone tissue by removing its mineralized matrix and breaking up the organic bone.⁹¹ The lack and excess of cathepsin K activity have been implicated in defects such as pycnodysostosis⁹² and osteoporosis,⁹³ respectively. Recently, it has been demonstrated that the collagenase activity of cathepsin K requires the formation of a molecular complex with glycosaminoglycans (GAGs) such as chondroitin 4-sulfate (C4S).^{94,95} Those initial studies were followed by the determination of the crystal structure of the complex of cathepsin K:C4S where both direct and water mediated interactions between C4S and cathepsin K were identified.⁷²

C4S is a repeating copolymer of alternating β -D-glucuronic acid (GCU) and 2'-deoxy-2'-acetamido- β -D-galactose-4-sulfate (ASG). The GCU units are linked to ASG via a $\beta(1\rightarrow 3)$ linkage, while the ASG units are linked to GCU by a $\beta(1\rightarrow 4)$ linkage. Twenty-nanosecond MD trajectories of the Cathepsin K/C4S complex (PDB id: 3C9E) as well as C4S alone in solution (PDB id: 1C4S) were performed. The carbohydrate–protein non-hydrogen atom RMSD in cathepsin K/C4S remains lower than 2 Å for the entire simulation, as do the protein and C4S alone (Supporting Information Figure S8). The increased flexibility of C4S alone in solution is evident from the large fluctuations in RMSD (Supporting Information Figure S8c), indicating the stabilizing effect of the protein on the carbohydrate structure.

To better understand the structural changes occurring upon C4S complexation to cathepsin K, the conformational space of C4S both in the presence and in the absence of cathepsin K was analyzed. The conformational space of C4S is defined by the ϕ/ψ dihedrals for the $\beta(1\rightarrow 3)$ and $\beta(1\rightarrow 4)$ glycosidic linkages. In Figure 7a and b, the Boltzmann-inverted ϕ/ψ dihedral distributions for the two linkages are presented. The conformational space of all the linkages, other than those between sugars at positions 1 and 2, are affected upon the binding of 1C4S with the protein (cathepsin K) due to the noncovalent contacts between the carbohydrate and the protein. To identify these contacts, contact probabilities between the non-hydrogen atoms of 1C4S and cathepsin K were evaluated for the final 14 ns of the 1C4S:cathepsin K simulation, with a cutoff distance of 3.50 Å. Contacts with occupancy greater than 0.5 were considered to be significant and are tabulated in Supporting Information Table S5. The sugars at positions 3, 4, and 6 interact with amino acids Lys9, Gln172, Ile171, and Asp6. The simulation results, though limited to 20 ns, agree with the initial contacts observed in the crystal structure, which have been discussed by Li et al.,⁷² reinforcing confidence in the validity of the newly developed parameters for modeling carbohydrate:protein interactions. It is to be noted that the sulfate groups of ASG are involved in very weak interactions with the protein backbone. The highest occupancy of a sulfate–protein noncovalent contact was found to be 0.396 between the sulfate group of ASG6 and the N_{ζ} atom of Lys 10. The longer contact distance between these atoms in the crystal structure (3.35 Å) is consistent with the low occupancy in the simulation.

CONCLUSIONS

The newly developed parameters are an extension of the existing CHARMM carbohydrate force field to enable the modeling of phosphates and sulfates linked to carbohydrates, as occurs in biological systems such as sugar nucleotides, inositol

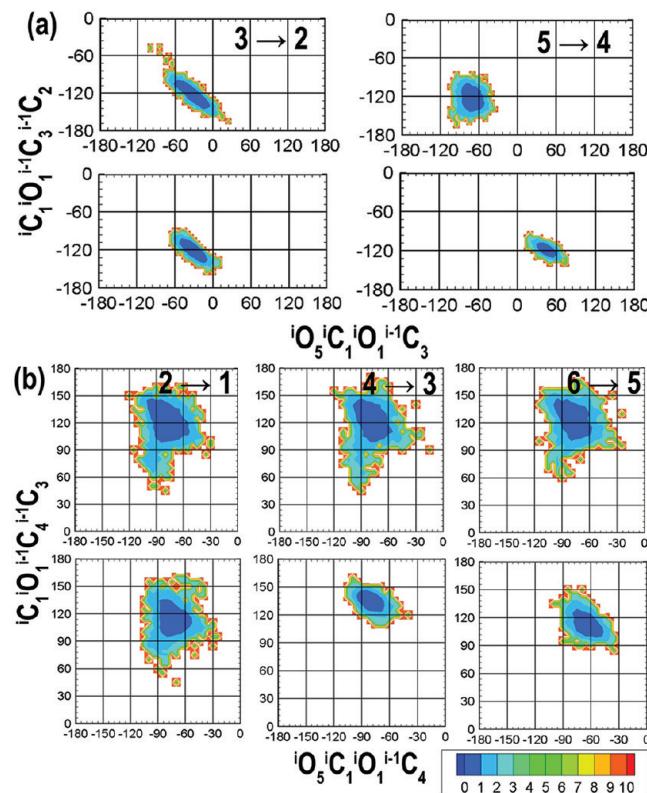


Figure 7. Boltzmann-inverted ϕ/ψ dihedral angle distributions. (a) $\beta(1\rightarrow 3)$ ϕ/ψ dihedral distribution ($\phi = {}^1\text{O}_5-{}^1\text{C}_1-{}^1\text{O}_1-{}^{i-1}\text{C}_3$, $\psi = {}^1\text{C}_1-{}^1\text{O}_1-{}^{i-1}\text{C}_3-{}^{i-1}\text{C}_2$), (b) $\beta(1\rightarrow 4)$ ϕ/ψ dihedral distribution ($\phi = {}^1\text{O}_5-{}^1\text{C}_1-{}^1\text{O}_1-{}^{i-1}\text{C}_4$, $\psi = {}^1\text{C}_1-{}^1\text{O}_1-{}^{i-1}\text{C}_4-{}^{i-1}\text{C}_3$). Contours are mapped every 1 kcal/mol. In each case, the ϕ/ψ distribution from the unbound C4S simulation is presented in the top panel, while that from the complex is in the bottom panel.

phosphates, and glycosaminoglycans. The new parameters are compatible with the existing CHARMM protein force field^{43,44} both by design and as demonstrated by simulations of proteins complexed with monosaccharides (bearing phosphate) and polysaccharides (bearing sulfate). These studies help to validate the parameter development methodology, which included transfer of charge and nonbonded parameters from existing CHARMM phosphate and sulfate parametrizations to maintain internal consistency of the expanding CHARMM biomolecular force field and allow its application to the study of complex heterogeneous biological systems. Excellent agreement was found between crystallography-identified contacts and contact probabilities from the simulations, thereby demonstrating the force field's ability to maintain key crystal contacts.

With regard to the parametrization methodology, an overestimation of the crystal volumes was observed for the neutral compounds, a trend which is consistent with the results in the CHARMM force field for models of hexopyranose and furanose monosaccharides, linear sugars and sugar alcohols, and disaccharides.^{36–41} As mentioned in previous studies, the introduction of electronic polarizability into the molecular mechanics framework may help to alleviate this limitation, and work in this direction is currently underway. Finally, we note that the present work enables the study of phospholipids, which represent another major class of biomolecules—in addition to proteins—having covalent linkages to carbohydrates. Simulations are being undertaken to validate the newly developed parameters in the context of phospholipids.

■ ASSOCIATED CONTENT

§ Supporting Information

Comparison between the charges on methyphosphates and methylsulfate and the corresponding moieties in the carbohydrate force field. Comparison between the intramolecular QM minimized geometries and crystal structure geometries. Model compound intramolecular geometries for THP8 and THP13. Crystalline intramolecular geometries for the C6 sulfate attachment in HAHZEV crystal. 1D energy scans for THP1-4, CYX1-4, and INI1-4. 2D energy scans for THP13. Monosaccharide unit geometries in the phosphate and sulfate crystals. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: alex@outerbanks.umaryland.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support from the NIH (GM070855, A.D.M.) and University of New England College of Pharmacy startup funds (O.G.) is acknowledged, as is computational support from the Department of Defense and NPACI Alliance, and we thank Drs. E. Prabhu Raman and Kenno Vanommeslaeghe for helpful discussions.

■ REFERENCES

- (1) Ginsburg, V. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1964**, *26*, 35.
- (2) Kochetkov, N. K.; Shibaev, V. N.; Tipson, R. S.; Derek, H. *Adv. Carbohydr. Chem. Biochem.* **1973**, *28*, 307.
- (3) Neufeld, E. F.; Hassid, W. Z. *Adv. Carbohydr. Chem.* **1963**, *18*, 309.
- (4) Shibaev, V. N. *Adv. Carbohydr. Chem. Biochem.* **1986**, *44*, 277.
- (5) Thoden, J. B.; Frey, P. A.; Holden, H. M. *Biochemistry* **1996**, *35*, 5137.
- (6) Kawakita, M.; Ishida, N.; Miura, N.; Sun-Wada, G. H.; Yoshioka, S. *J. Biochem.* **1998**, *123*, 777.
- (7) Breton, C.; Snajdrova, L.; Jeanneau, C.; Koca, J.; Imbert, A. *Glycobiology* **2006**, *16*, 29R.
- (8) Fritz, T. A.; Raman, J.; Tabak, L. A. *J. Biol. Chem.* **2006**, *281*, 8613.
- (9) Kubota, T.; Shiba, T.; Sugioka, S.; Furukawa, S.; Sawaki, H.; Kato, R.; Wakatsuki, S.; Narimatsu, H. *J. Mol. Biol.* **2006**, *359*, 708.
- (10) Fritz, T. A.; Hurley, J. H.; Trinh, L. B.; Shiloach, J.; Tabak, L. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15307.
- (11) Stephen, B. S. *Cell Signal* **2001**, *13*, 151.
- (12) Berridge, M. J. *Nature* **1993**, *361*, 315.
- (13) Haltiwanger, R. S.; Lowe, J. B. *Annu. Rev. Biochem.* **2004**, *73*, 491.
- (14) Bulow, H. E.; Hobert, O. *Annu. Rev. Cell Dev. Biol.* **2006**, *22*, 375.
- (15) Hacker, U.; Nybakken, K.; Perrimon, N. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 530.
- (16) Bishop, J. R.; Schuksz, M.; Esko, J. D. *Nature* **2007**, *446*, 1030.
- (17) Sattelle, B. M.; Shakeri, J.; Roberts, I. S.; Almond, A. *Carbohydr. Res.* **2010**, *345*, 291.
- (18) Winter, W. T.; Arnott, S.; Isaac, D. H.; Atkins, E. D. *J. Mol. Biol.* **1978**, *125*, 1.
- (19) Sugahara, K.; Mikami, T.; Uyama, T.; Mizuguchi, S.; Nomura, K.; Kitagawa, H. *Curr. Opin. Struct. Biol.* **2003**, *13*, 612.
- (20) Malavaki, C.; Mizumoto, S.; Karamanos, N.; Sugahara, K. *Connect. Tissue Res.* **2008**, *49*, 133.
- (21) Wormald, M. R.; Petrescu, A. J.; Pao, Y. L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371.
- (22) Crispin, M.; Stuart, D. I.; Jones, E. Y. *Nat. Struct. Mol. Biol.* **2007**, *14*, 354; discussion 354.
- (23) Duus, J. A.; Gotfredsen, C. H.; Bock, K. *Chem. Rev.* **2000**, *100*, 4589.
- (24) Momany, F. A.; Willett, J. L. *Carbohydr. Res.* **2000**, *326*, 210.
- (25) Momany, F. A.; Willett, J. L. *Carbohydr. Res.* **2000**, *326*, 194.
- (26) Kirschner, K. N.; Yongye, A. B.; Tschaepel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622.
- (27) Lins, R. D.; Hunenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400.
- (28) Basma, M.; Sundara, S.; Çalgan, D.; Vernali, T.; Woods, R. J. *J. Comput. Chem.* **2001**, *22*, 1125.
- (29) Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955.
- (30) Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W. F. *J. Comput. Chem.* **2002**, *23*, 1416.
- (31) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832.
- (32) Senderowitz, H.; Parish, C.; Still, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 2078.
- (33) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10541.
- (34) Hansen, H. S.; Hünenberger, P. H. *J. Comput. Chem.* **2011**, *32*, 998.
- (35) Foley, B. L.; Tessier, M. B.; Woods, R. J. *Wiley Interdisciplinary Reviews: Computational Molecular Science*; Wiley: New York, 2011.
- (36) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. Jr. *J. Comput. Chem.* **2008**, *29*, 2543.
- (37) Hatcher, E.; Guvench, O.; MacKerell, A. D. Jr. *J. Phys. Chem. B* **2009**, *113*, 12466.
- (38) Hatcher, E.; Guvench, O.; MacKerell, A. D. Jr. *J. Chem. Theory Comput.* **2009**, *5*, 1315.
- (39) Guvench, O.; Hatcher, E. R.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. Jr. *J. Chem. Theory Comput.* **2009**, *5*, 2353.
- (40) Raman, E. P.; Guvench, O.; MacKerell, A. D. Jr. *J. Phys. Chem. B* **2010**, *114*, 12981.
- (41) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. Jr. *J. Chem. Theory Comput.* **2011**, *7*, 3162.
- (42) Mallajosyula, S. S.; MacKerell, A. D. Jr. *J. Phys. Chem. B* **2011**, *115*, 11215.
- (43) MacKerell, A. D. Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (44) MacKerell, A. D. Jr.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400.
- (45) Foloppe, N.; MacKerell, A. D. Jr. *J. Comput. Chem.* **2000**, *21*, 86.
- (46) MacKerell, A. D.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105.
- (47) Klauda, J. B.; Brooks, B. R.; MacKerell, A. D.; Venable, R. M.; Pastor, R. W. *J. Phys. Chem. B* **2005**, *109*, 5300.
- (48) Feller, S. E.; Gawrisch, K.; MacKerell, A. D. *J. Am. Chem. Soc.* **2001**, *124*, 318.
- (49) Feller, S. E.; MacKerell, A. D. *J. Phys. Chem. B* **2000**, *104*, 7510.
- (50) Yin, D.; MacKerell, A. D. *J. Comput. Chem.* **1998**, *19*, 334.
- (51) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671.
- (52) Jamison II, F. W.; Foster, T. J.; Barker, J. A.; Hills, R. D., Jr.; Guvench, O. *J. Mol. Biol.* **2011**, *406*, 631.

- (53) Brooks, B. R.; Brooks, C. L. 3rd; MacKerell, A. D. Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
- (54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (55) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198.
- (56) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (57) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667.
- (58) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (59) Hockney, R. W. *Methods Comput. Phys.* **1970**.
- (60) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.
- (61) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (62) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613.
- (63) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U. K., 1989.
- (64) Allen, F. *Acta Crystallogr.* **2002**, *B58*, 380.
- (65) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Laham, A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian Inc.: Wallingford, CT, 2003.
- (66) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
- (67) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
- (68) Guvench, O.; MacKerell, A. D. Jr. *J. Mol. Model.* **2008**, *14*, 667.
- (69) Cronin, T. C.; DiNitto, J. P.; Czech, M. P.; Lambright, D. G. *EMBO J.* **2004**, *23*, 3711.
- (70) Milburn, C. C.; Deak, M.; Kelly, S. M.; Price, N. C.; Alessi, D. R.; Van Aalten, D. M. *Biochem. J.* **2003**, *375*, 531.
- (71) Ferguson, K. M.; Kavran, J. M.; Sankaran, V. G.; Fournier, E.; Isakoff, S. J.; Skolnik, E. Y.; Lemmon, M. A. *Mol. Cell* **2000**, *6*, 373.
- (72) Li, Z.; Kienetz, M.; Cherney, M. M.; James, M. N.; Bromme, D. *J. Mol. Biol.* **2008**, *383*, 78.
- (73) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J. Comput. Chem.* **2008**, *29*, 1859.
- (74) Jo, S.; Song, K. C.; Desaire, H.; MacKerell, A. D. Jr.; Im, W. *J. Comput. Chem.* **2011**, *32*, 3135.
- (75) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
- (76) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735.
- (77) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; John Wiley & Sons: New York, 1988.
- (78) Becker, A. D.; Roux, B.; Watanabe, M. *Computational Biochemistry and Biophysics*; Marcel-Dekker, Inc.: New York, 2001.
- (79) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulter, K. *J. Comput. Phys.* **1999**, *151*, 283.
- (80) Petrova, P.; Koca, J.; Imbert, A. *J. Am. Chem. Soc.* **1999**, *121*, 5535.
- (81) Andre, I.; Tvaroska, I.; Carver, J. P. *J. Phys. Chem. A* **2000**, *104*, 4609.
- (82) Kona, J.; Tvaroska, I. *Chem. Papers* **2009**, *63*, 598.
- (83) Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*; John Wiley and Sons: New York, 1994.
- (84) Corvera, S.; Czech, M. P. *Trends Cell Biol.* **1998**, *8*, 442.
- (85) Fruman, D. A.; Meyers, R. E.; Cantley, L. C. *Annu. Rev. Biochem.* **1998**, *67*, 481.
- (86) Lemmon, M. A.; Ferguson, K. M.; Abrams, C. S. *FEBS Lett.* **2002**, *513*, 71.
- (87) Rebecchi, M. J.; Scarlata, S. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 503.
- (88) DiNitto, J. P.; Cronin, T. C.; Lambright, D. G. *Sci. STKE* **2003**, *2003*, re16.
- (89) Isakoff, S. J.; Cardozo, T.; Andreev, J.; Li, Z.; Ferguson, K. M.; Abagyan, R.; Lemmon, M. A.; Aronheim, A.; Skolnik, E. Y. *EMBO J.* **1998**, *17*, 5374.
- (90) Rameh, L. E.; Arvidsson, A.; Caraway, K. L. 3rd; Couvillon, A. D.; Rathbun, G.; Crompton, A.; VanRenterghem, B.; Czech, M. P.; Ravichandran, K. S.; Burakoff, S. J.; Wang, D. S.; Chen, C. S.; Cantley, L. C. *J. Biol. Chem.* **1997**, *272*, 22059.
- (91) Drake, F. H.; Dodds, R. A.; James, I. E.; Connor, J. R.; Debouck, C.; Richardson, S.; Lee-Ryckaszewski, E.; Coleman, L.; Rieman, D.; Barthlow, R.; Hastings, G.; Gowen, M. *J. Biol. Chem.* **1996**, *271*, 12511.
- (92) Gelb, B. D.; Shi, G.-P.; Chapman, H. A.; Desnick, R. J. *Science* **1996**, *273*, 1236.
- (93) Zaidi, M.; Troen, B.; Moonga, B. S.; Abe, E. *J. Bone Miner. Res.* **2001**, *16*, 1747.
- (94) Li, Z.; Hou, W.-S.; Escalante-Torres, C. R.; Gelb, B. D.; Bromme, D. *J. Biol. Chem.* **2002**, *277*, 28669.
- (95) Li, Z.; Yasuda, Y.; Li, W.; Bogyo, M.; Katz, N.; Gordon, R. E.; Fields, G. B.; Bromme, D. *J. Biol. Chem.* **2004**, *279*, 5470.