Supporting Information

for

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

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

Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, 20 Penn St., HSF II-629, Baltimore, MD 21201; Department of Pharmaceutical Sciences, University of New England College of Pharmacy, Portland, Maine 04103

---

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

[†] University of Maryland

[ø] University of New England

Charges on methylphosphates (anionic and dianionic) and methylsulfate from the CHARMM nucleic acid and lipid force fileds used to model the phosphate and sulfate linkage to carbohydrates.
45-50

```
! Phosphates from the nucleic acid force field.
RESI MP_1        -1.00   ! Methylphosphate, anionic
GROUP                    !
! atom order for molvib
ATOM C1   CN9    -0.170  !
ATOM O1   ON2    -0.620  !                 H11
ATOM P1   P       1.500  !                  |
ATOM O2   ON4    -0.680  !            H13--C1--H12
ATOM O3   ON3    -0.820  !                  |
ATOM O4   ON3    -0.820  !                 O1
ATOM H11  HN9     0.090  !                  |
ATOM H12  HN9     0.090  !              O4==P1==O3 (-)
ATOM H13  HN9     0.090  !                  |
ATOM H2   HN4     0.340  !                 O2
                         !                   \
                         !                    H2

! Transfer of charge to the carbohydrate force field.
PRES PH_A          -1.00 ! The phosphate is added to the axial group
dele atom H1A            !  O-PO3H is added to C1; apply to THP2
GROUP
ATOM C1   CC311D    0.21 !
ATOM H1B  HCA1      0.09 !
ATOM O1   OC30P    -0.62 !
ATOM P1   PC        1.50 !
ATOM O2   OC312    -0.68 !
ATOM O3   OC2DP    -0.82 !
ATOM O4   OC2DP    -0.82 !
ATOM H2   HCP1      0.34 !
ATOM C5   CC321C    0.02 !
ATOM H5A  HCA2      0.09 !
ATOM H5B  HCA2      0.09 !
ATOM O5   OC3C61   -0.40 !
```

```
! Phosphates from the nucleic acid force field.
RESI MP_2        -2.00   ! Methylphosphate, dianionic
GROUP                    !
ATOM P1   P       1.100  !
ATOM O1   ON2    -0.400  !                 H11
ATOM O2   ON3    -0.900  !                  |
ATOM O3   ON3    -0.900  !            H13--C1--H12
ATOM O4   ON3    -0.900  !                  |
GROUP                    !                 O1
ATOM C1   CN9    -0.270  !                  |
ATOM H11  HN9     0.090  !        (-)  O4==P1==O3  (-)
ATOM H12  HN9     0.090  !                 ||
ATOM H13  HN9     0.090  !                 O2

! Transfer of charge to the carbohydrate force field.
PRES DPH_A         -2.00 ! The phosphate is added to the axial group
dele atom H1A            !  O-PO3 is added to C1; apply to THP2
GROUP
ATOM C1   CC311D    0.11 !
ATOM H1B  HCA1      0.09 !
ATOM O1   OC30P    -0.40 !
ATOM P1   PC        1.10 !
ATOM O2   OC2DP    -0.90 !
ATOM O3   OC2DP    -0.90 !
ATOM O4   OC2DP    -0.90 !
ATOM C5   CC321C    0.02 !
ATOM H5A  HCA2      0.09 !
ATOM H5B  HCA2      0.09 !
ATOM O5   OC3C61   -0.40 !
```

```
! Sulfate from the lipid force field.
RESI MSO4          -1.00 ! CH3O4S Methylsulfate
GROUP                    !
ATOM S    SG3O1     1.33 !           OS2(-1)
ATOM OS1  OG303    -0.28 !            |
ATOM OS2  OG2P1    -0.65 ! (-1) OS2--S(+2)--OS4 (-1)
ATOM OS3  OG2P1    -0.65 !            |
ATOM OS4  OG2P1    -0.65 !           OS1
ATOM C1   CG331    -0.37 !             \
ATOM H11  HGA3      0.09 !         H11-C1-H13
ATOM H12  HGA3      0.09 !              |
ATOM H13  HGA3      0.09 !             H12

! Transfer of charge to the carbohydrate force field.
PRES SH_A         -1.00  !
dele atom H1A            !
GROUP
ATOM C1   CC311D    0.01 !
ATOM H1B  HCA1      0.09 !
ATOM O1   OC30P    -0.28 !
ATOM S1   SC        1.33 !
ATOM O2   OC2DP    -0.65 !
ATOM O3   OC2DP    -0.65 !
ATOM O4   OC2DP    -0.65 !
ATOM C5   CC321C    0.02 !
ATOM H5A  HCA2      0.09 !
ATOM H5B  HCA2      0.09 !
ATOM O5   OC3C61   -0.40 !
```

Table S1. Bond lengths and valence angles averaged over QM-optimized geometries of model compounds and over relevant crystal geometries.

Phosphate

	Mono-anion			Di-anion		
	QM	Crys	QM-C	QM	Crys	QM-C
C1-O1/C2-O1	1.420	1.423	-0.002	1.387	1.416	-0.029
O1-P	1.682	1.595	0.087	1.814	1.637	0.177
P-O2*	1.687	1.575	0.112			
P-O2				1.550	1.510	0.040
P-O3	1.516	1.480	0.035	1.546	1.518	0.027
P-O4	1.505	1.500	0.005	1.549	1.523	0.026
C2-C1-O1/C1-C2-O1	110.5	106.5	4.016	110.8	110.3	0.543
C3-C2-O1	107.1	111.6	-4.482	110.6		
O5-C1-P	106.6			108.1	111.0	-2.904
C1-O1-P/C2-O1-P	117.3	124.8	-7.506	114.7	120.0	-5.215
O1-P-O2*	99.6	102.1	-2.524			
O1-P-O2				101.2	104.2	-2.998
O2*-P-O3	106.3	111.4	-5.082			
O2*-P-O4	108.3	110.3	-2.001			
O2-P-O3				116.5	113.4	3.145
O2-P-O4				115.9	113.3	2.542
O1-P-O3	107.5	111.6	-4.135	101.5	104.5	-2.977
O1-P-O4	106.5	104.5	1.949	101.7	107.7	-6.001
O3-P-O4	125.8	115.9	9.845	116.1	112.8	3.281

Table S1 continued:

Sulfate

	QM	Crys	QM-C
			-
C1-O1/C2-O1	1.426	1.465	0.039
O1-S	1.715	1.592	0.123
S-O2	1.486	1.435	0.051
S-O3	1.487	1.453	0.034
S-O4	1.475	1.444	0.032
C2-C1-O1/C1-C2-O1	109.3	107.4	1.8
C3-C2-O1	109.3	107.8	1.4
O5-C1-S	106.1		
C1-O1-S/C2-O1-S	114.3	119.7	-5.4
O1-S-O2	104.1	101.2	2.9
O1-S-O3	103.6	106.7	-3.1
O1-S-O4	100.7	107.6	-6.9
O2-S-O3	113.9	113.5	0.4
O2-S-O4	115.8	114.6	1.2
O3-S-O4	116.0	112.1	3.9

O2* is the protonated oxygen in mono-anionic phosphates.

Table S2. Bond lengths, valence angles and dihedral angles for model compound **THP8**.

	QM	MM	MM-QM	MM	MM-QM	MM	MM-QM
		THP8		**THP8[(a)]**		**THP8[(b)]**	
C1-O1	1.409	1.421	0.012	1.422	0.014	1.422	-0.013
O1-P	1.694	1.629	-0.065	1.626	-0.069	1.624	0.070
P-O2*	1.678	1.586	-0.092	1.587	-0.091	1.587	0.091
P-O3	1.511	1.521	0.010	1.52	0.009	1.519	-0.008
P-O4	1.511	1.522	0.011	1.521	0.010	1.521	-0.010
C2-O1a	1.428	1.431	0.003	1.431	0.003	1.431	-0.003
O1a-P	1.696	1.646	-0.05	1.643	-0.052	1.645	0.051
P-O2*a	1.665	1.590	-0.074	1.589	-0.076	1.588	0.076
P-O3a	1.513	1.518	0.005	1.519	0.006	1.519	-0.006
P-O4a	1.515	1.521	0.006	1.522	0.006	1.522	-0.006
C2-C1-O1	109.2	115.1	5.8	111.9	2.6	111.5	-2.3
O5-C1-P	110.0	107.0	-2.9	108.0	-2.0	108.0	1.9
C1-O1-P	119.3	119.3	0.0	118.6	-0.7	118.9	0.4
O1-P-O2*	98.2	97.8	-0.4	94.2	-4.0	93.8	4.4
O1-P-O3	110.6	110.2	-0.4	110.8	0.2	110.9	-0.3
O1-P-O4	104.1	109.5	5.3	109.4	5.3	109.4	-5.3
O2*-P-O3	105.7	109.8	4.1	110.9	5.2	111.0	-5.3
O2*-P-O4	111.1	110.0	-1.1	110.2	-0.9	110.2	0.9
O3-P-O4	124.1	117.7	-6.4	118.6	-5.5	118.6	5.5
C1-C2-O1a	111.1	115.5	4.4	116.2	5.1	116.6	-5.5
C3-C2-O1a	107.6	108.0	0.4	106.1	-1.5	106.1	1.5
C2-O1a-P	118.0	118.5	0.6	119.4	1.5	119.3	-1.4
O1a-P-O2*A	99.2	94.9	-4.3	98.0	-1.2	98.9	0.3
O1a-P-O3a	103.7	109.9	6.3	109.0	5.4	108.9	-5.2
O1a-P-O4a	109.5	111.4	1.9	110.3	0.8	110.1	-0.6
O2*a-P-O3a	111.0	108.8	-2.2	108.9	-2.1	108.8	2.2
O2*a-P-O4a	106.7	110.7	4.0	110.9	4.2	110.7	-4.1
O3a-P-O4a	124.0	118.6	-5.4	117.9	-6.1	117.8	6.2
C3-C2-C1-O1	-61.6	-68.8	-7.2	-68.2	-6.6	-67.2	5.6
C5-O5-C1-O1	56.2	69.2	13.0	66.0	9.7	65.0	-8.7
C2-C1-O1-P	-120.3	-99.8	20.4	-117.2	3.1	-118.2	-2.1
O5-C1-O1-P	120.5	136.0	15.5	120.0	-0.4	119.5	1.0
C1-O1-P-O2*	62.4	86.5	24.1	73.2	10.8	74.3	-11.8
C1-O1-P-O3	-47.9	-28.0	19.8	-41.0	6.9	-40.0	-7.9
C1-O1-P-O4	176.7	-159.1	24.2	-173.6	9.7	-172.7	10.6
C4-C3-C2-O1a	-175.4	-178.9	-3.4	-178.2	-2.8	-179.1	3.7
O5-C1-C2-O1a	177.5	176.6	-0.9	174.2	-3.2	175.1	2.3
C3-C2-O1a-P	-138.0	-92.8	45.2	-129.3	8.7	-134.1	-3.9
C1-C2-O1a-P	**101.4**	**142.1**	**40.7**	**106.4**	**5.0**	**101.4**	**0.0**
C2-O1a-P-O2*a	-67.6	-84.1	-16.5	-75.4	-7.7	-73.6	5.9
C2-O1a-P-O3a	178.0	163.9	-14.1	171.4	-6.6	173.0	5.0

C2-O1a-P-O4a	43.8	30.4	-13.4	40.4	-3.4	42.4	1.4
O1-C1-C2-O1a	57.5	54.7	-2.8	53.4	-4.1	54.6	2.9

O2* is the protonated phosphate oxygen. [(a)] C1-C2-O1a-P dihedral restrained 5° from the QM geometry. [(b)] C1-C2-O1a-P dihedral restrained to the QM geometry.

Table S3. Bond lengths, valence angles and dihedral angles for model compound THP13.[(a)]

	QM	MM	MM-QM
	THP13		
Bonds			
C6-O1	1.430	1.448	0.018
O1-S	1.714	1.596	-0.118
S-O2	1.475	1.452	-0.023
S-O3	1.489	1.453	-0.036
S-O4	1.486	1.452	-0.034
Angles			
C5-C6-O1	112.9	112.1	-0.8
C6-O1-S	115.3	116.9	1.6
O1-S-O2	100.9	103.0	2.1
O1-S-O3	104.1	105.6	1.6
O1-S-O4	103.9	106.3	2.4
O2-S-O3	115.2	113.2	-2.1
O2-S-O4	116.3	113.4	-2.9
O3-S-O4	113.9	114.1	0.3
Dihedrals			
C4-C5-C6-O1	-51.9	-51.1	0.7
O5-C5-C6-O1	-172.5	-170.8	1.7
C5-C6-O1-S	-70.0	-77.2	-7.2
C6-O1-S-O2	-160.3	-166.7	-6.4
C6-O1-S-O3	80.0	74.4	-5.6
C6-O1-S-O4	-39.5	-47.2	-7.8

[(a)]See Figure 1 for atom naming and numbering.

Table S4. Crystalline intramolecular geometry for the C6 sulfate in the HAHZEV crystal.[a]

	CRYS	MD[b]	MD-CRYS
	HAHZEV		
Bonds			
C6-OS5	1.445	1.446	0.002
OS5-S2	1.579	1.581	0.002
S2-OS6	1.452	1.453	0.001
S2-OS7	1.439	1.448	0.010
S2-OS8	1.425	1.452	0.027
Angles			
C5-C6-OS5	106.2	107.6	1.4
C6-OS5-S2	118.2	115.9	-2.3
OS5-S2-OS6	101.7	104.6	2.9
OS5-S2-OS7	106.8	107.8	1.0
OS5-S2-OS8	107.3	105.4	-1.8
OS6-S2-OS7	112.9	113.3	0.5
OS6-S2-OS8	114.1	111.9	-2.2
OS7-S2-OS8	113.0	112.7	-0.2
Dihedrals			
C4-C5-C6-OS5[c]	-166.4	-166.5, -93.6	-0.2
O5-C5-C6-OS5[c]	71.9	73.3, 143.0	1.4
C5-C6-OS5-S2[c]	-175.9	175.7, -149.9	-8.4
C6-OS5-S2-OS6[d]	179.2	173.6	-5.7
C6-OS5-S2-OS7[d]	60.7	54.0	-6.8
C6-OS5-S2-OS8[d]	-60.7	-68.2	-7.6

[a] See Supporting Information Figure S7 for atom numbering.

[b] 4-ns MD average computed across 4000 snapshots and averaged over all n monosaccharides in the simulated complete unit cell (HAHZEV n=2; 95% confidence intervals, calculated as 1.96*(RMS fluctuation)/sqrt(4000*n), were < 0.001 Å for bonds, < 0.1 degrees for angles, and < 0.4 degrees for dihedrals .

[c] Dihedrals span two conformational bins. The MD averages are presented over the two distinct conformational regions adopted by the dihedrals. The difference is calculated for the conformational bin closest to the initial value.

[d] Due to the equivalencing of the atoms types OS6, OS7 and OS8, the dihedral distributions for these three dihedrals are pooled together for analysis. The combined C6-OS5-S2-OS6/7/8 dihedral distribution is found to populate three conformational bins with means at 173.6˚, 54.0˚ and -68.2˚.

Table S5. Significant carbohydrate-protein contact occupancies from MD simulations of the 1C4S: Cathepsin K system.[(a)]

Amino Acid	Carbohydrate	Occu	Crystallographic Distance
GCU1	---	---	---
ASG2	---	---	---
GCU3			
Lys 9 (N_ζ)	O6a	0.943	2.85
Lys 9 (N_ζ)	O6b	0.906	3.87[(d)]
Gln 172 ($N_{\varepsilon 2}$)	O2	0.812	2.80
ASG4			
Ile 171 (N)	O6	0.819	2.80
Ile 171 (O)	O6	0.716	3.05
GCU5	---	---	---
ASG6			
Asp 6 (N)	O6	0.964	2.76

[(a)]A cutoff distance of 3.50 Å was used to detect the close contacts. Occupancies greater than 0.5 are presented for the sake of clarity. Contact distances in the crystal structure are also summarized. Distances are in Å.

[(b)] ASG: 2′-deoxy-2′-acetamido-β-D-galactose-4-sulfate.

[(c)] GCU: β-D-glucuronic acid

[(d)]O6a and O6b correspond to the carboxylate oxygen's in β-D-glucuronic acid. During the simulation both the oxygen's are in contact with N_ζ of Lys9, giving rise to the high occupancy for both.

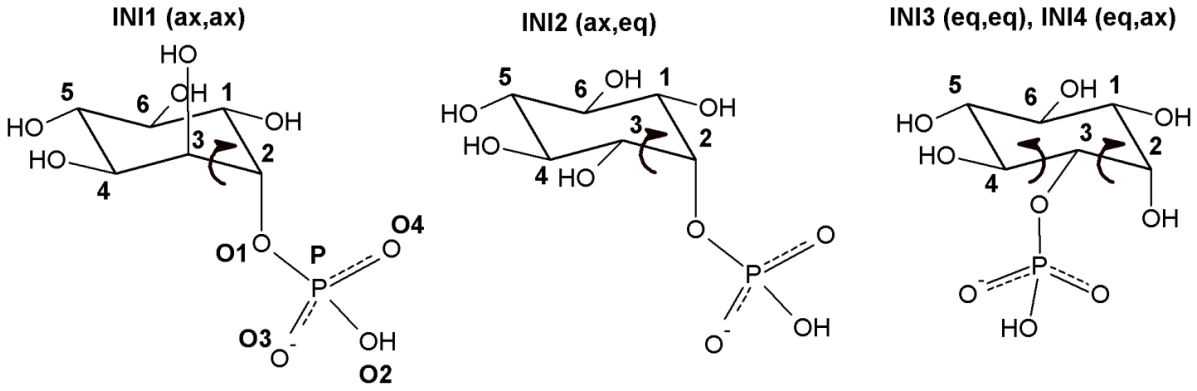

Figure S1. Model compounds used to develop parameters for HO_3PO-C-C-OH dihedrals. Arrows indicate the scanned dihedral for each compound. *Naming convention:* The configuration at the phosphate-substituted carbon is described first followed by the configuration at the hydroxyl-substituted carbon ("eq" = equatorial, "ax" = axial). For **INI3** (eq, eq) the scanned dihedral is O-C3-C4-O, while the scanned dihedral for **INI4** (eq, ax) is O-C3-C2-O. Atom labels are in bold.

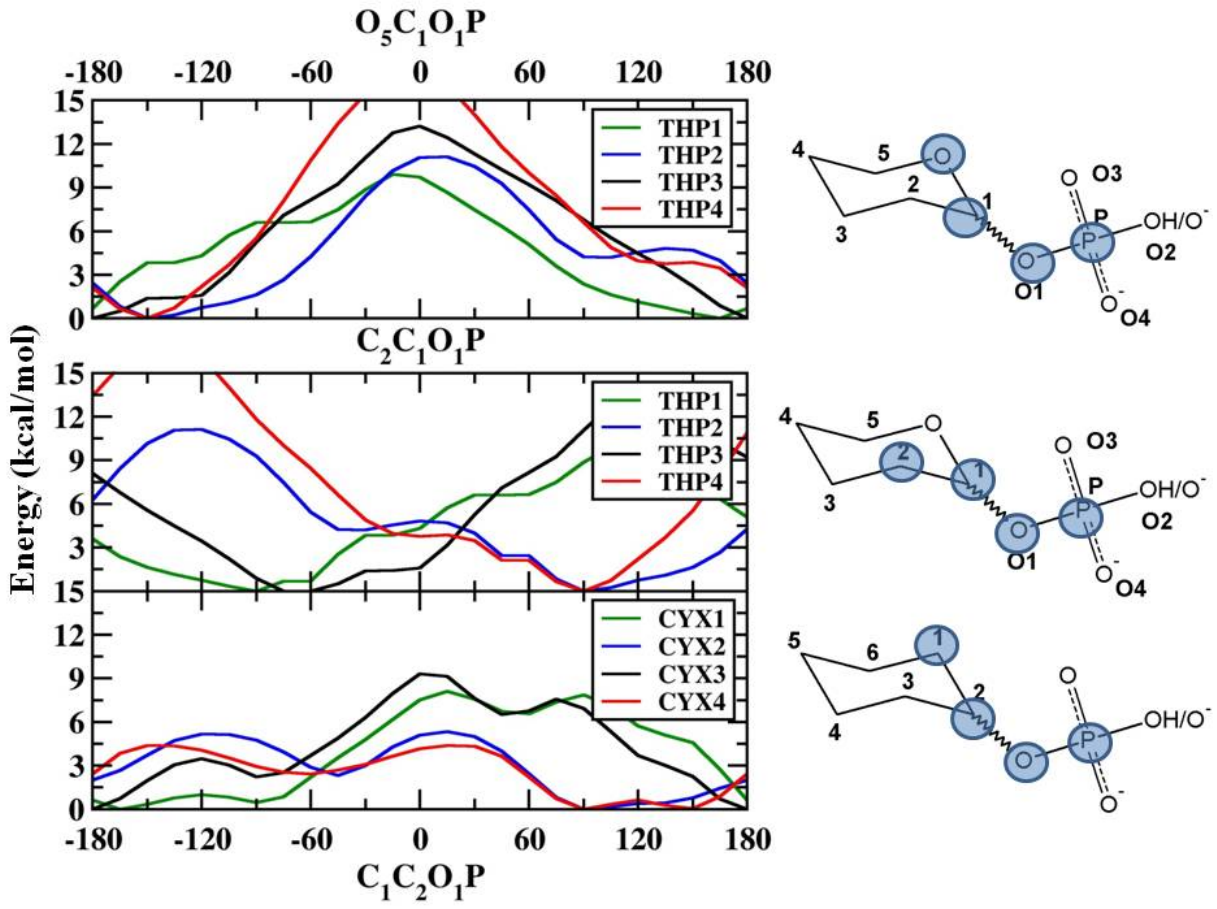

Figure S2. 1D QM energy profiles about the O5-C1-O1-P, C2-C1-O1-P, and C1-C2-O1-P dihedrals corresponding to the local minima of the 2D QM energy surfaces for **THP1-4** and **CYX1-4**. Atoms involved in the dihedrals are highlighted by shaded circles.

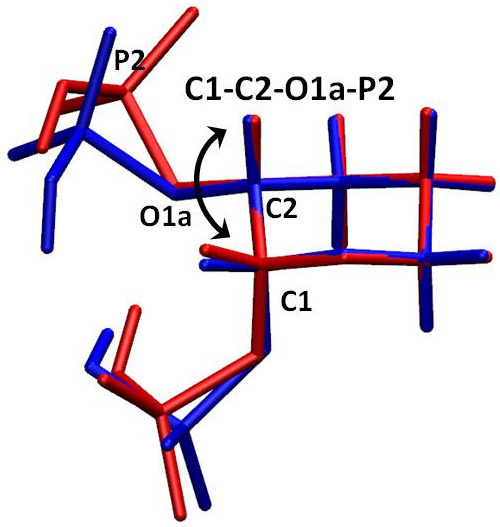

Figure S3. Overlay of the QM (blue) and MM (red) minimized geometries of **THP8**. Arrow indicates the concerned C1-C2-O1a-P dihedral.

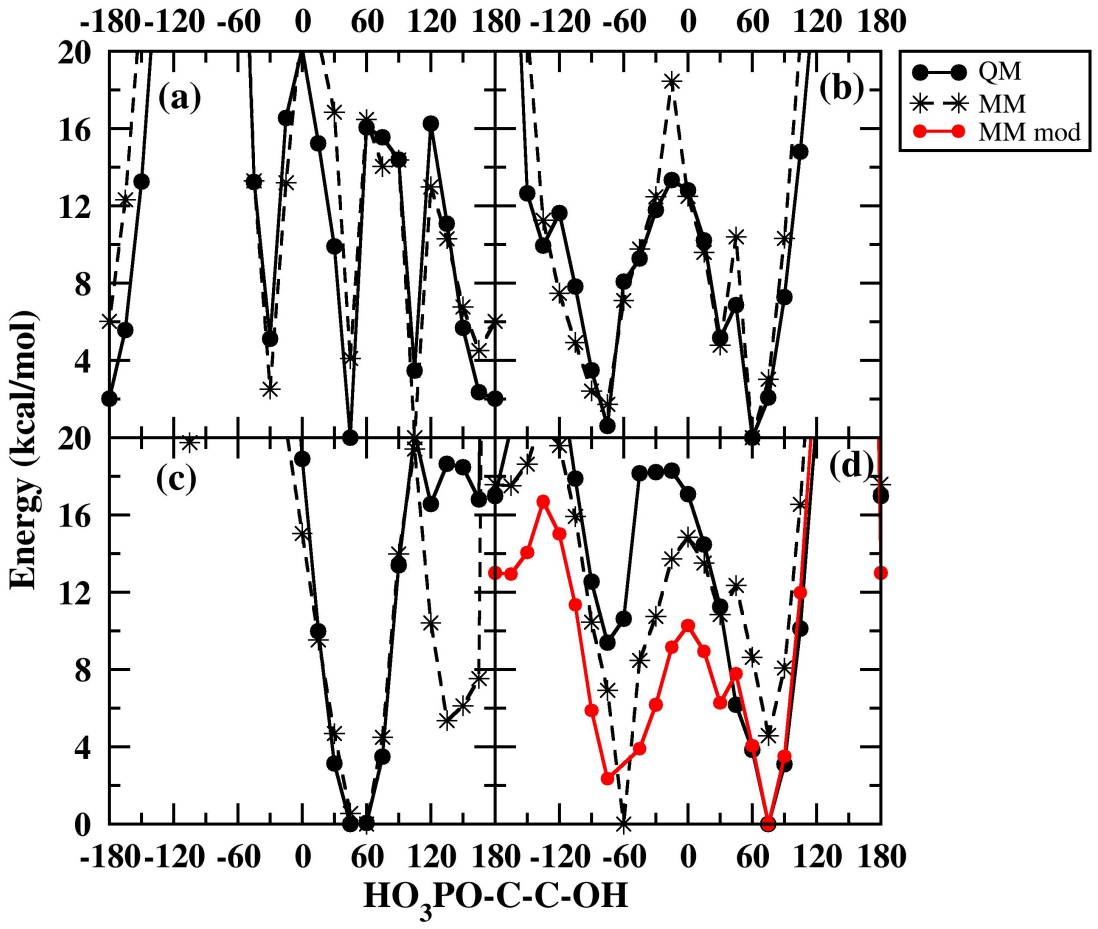

Figure S4. 1D-Dihedral potential energy scans about the HO_3PO-C-C-OH dihedral for model compounds (a) **INI1**, (b) **INI2**, (c) **INI3**, and (b) **INI4**. Legend: solid line with filled circles- QM scan, dashed line with stars- MM scan. For **INI4** the MM data are shown twice: first including all energies (dashed line with stars), and then again excluding the -60˚ point and vertically re-aligned such that the +75˚ conformation has an energy of 0 (red line with filled circles).

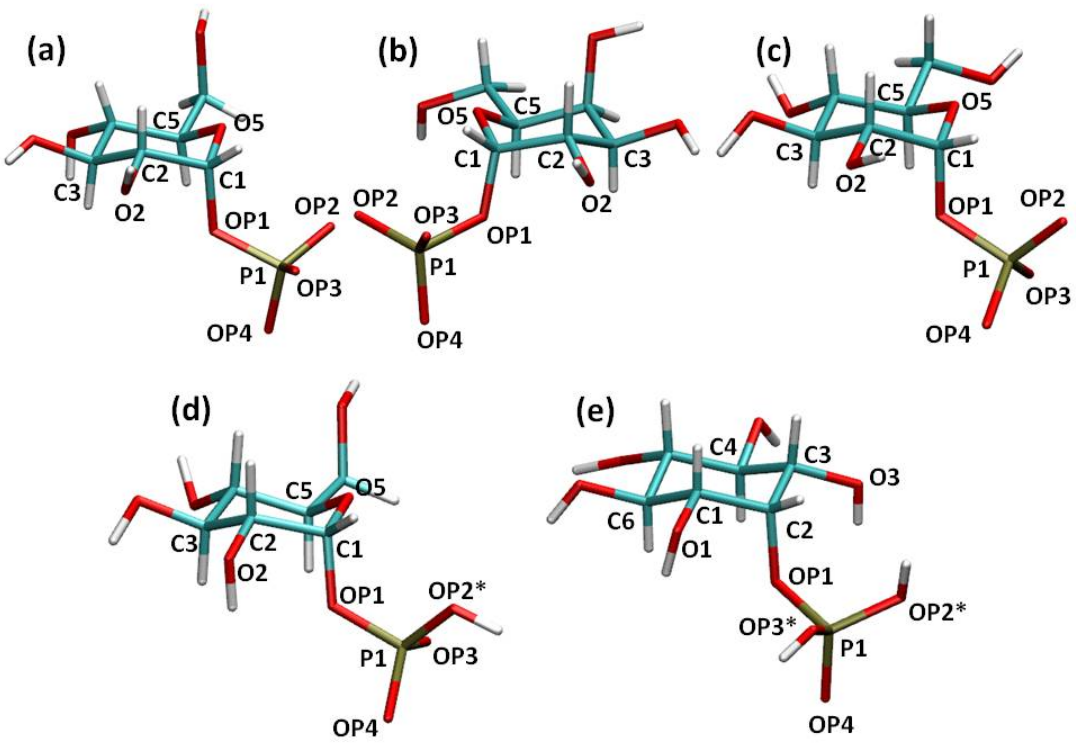

Figure S5. Monosaccharide phosphate crystals from the Cambridge Structural Database. (a) CIMDUX, (b) JEYDAS, (c) KGLUCP02, (d) JUGTAG and (e) MINOSP. Atom labels and numberings used to describe bond lengths, valence angles, and dihedral angles in Table 4 and Table 5 are in bold.

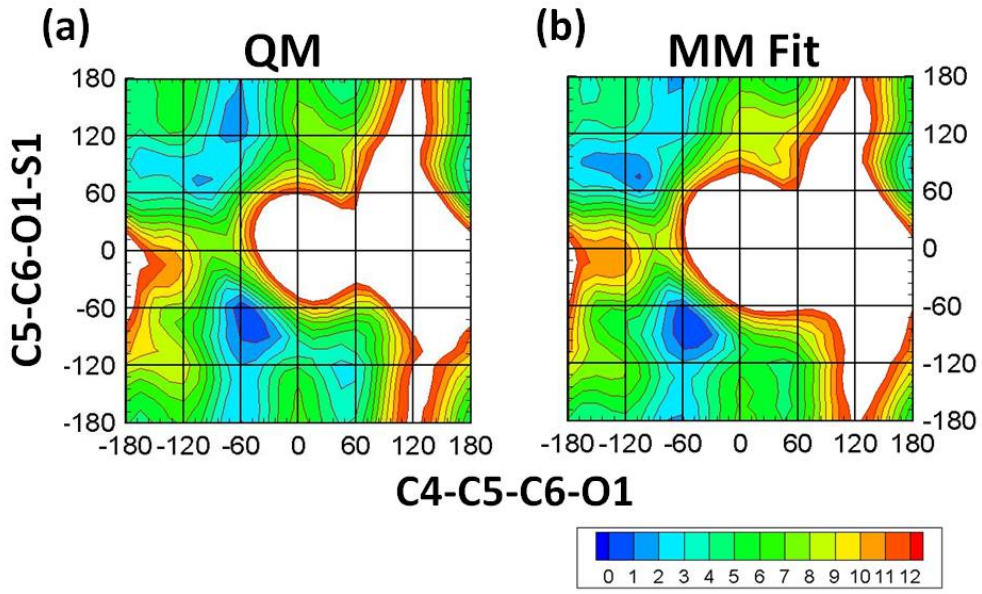

Figure S6. 2D-Dihedral potential energy scans about the C4-C5-C6-O1/C5-C6-O1-S1 dihedrals for model compound **THP13**. Energies are in kcal/mol, with contours every 1 kcal/mol. Only energies below 12 kcal/mol have been plotted for the sake of clarity. The left panel depicts the QM scan and the right panel the MM scan.

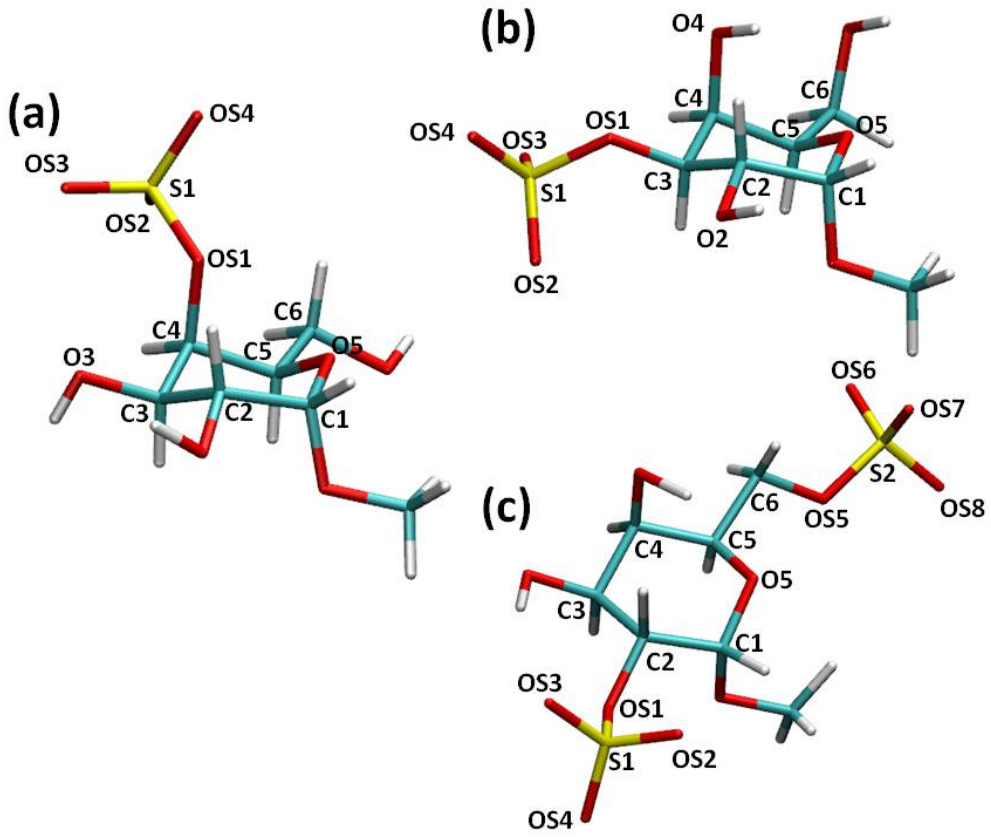

Figure S7. Monosaccharide sulfate crystals from the Cambridge Structural Database. (a) SOJHAA, (b) POCSOP and (c) HAHZEV. Atom labels and numberings used to describe bond lengths, valence angles, and dihedral angles in Table 8 and Supporting Information Table S4 are in bold.

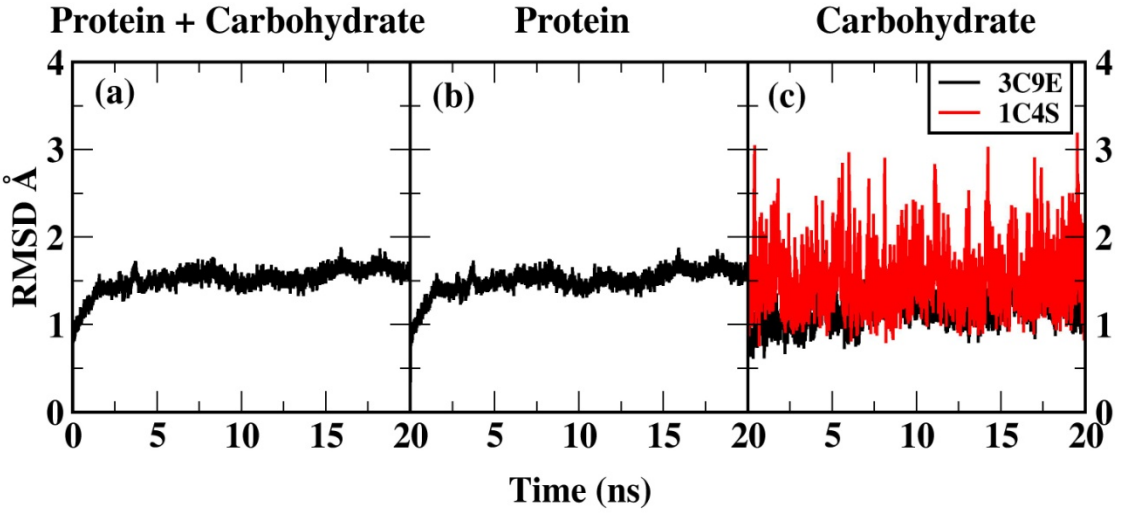

Figure S8. RMS difference (RMSD) analysis of Cathepsin K/C4S complex and C4S in solution. (a) RMSD for all carbohydrate-protein heavy atoms, (b) RMSD for the protein heavy atoms only, and (c) RMSD for carbohydrate heavy atoms only. RMSD in Å and time in ns. RMSDs were calculated following the alignment 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).