ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

Check for updates

Polarizable AMOEBA force field predicts thin and dense hydration layer around monosaccharides†

Luke A. Newman, [ID] [ab] Mackenzie G. Patton,[c] Breyanna A. Rodriguez,[d]
Ethan W. Sumner[ab] and Valerie Vaissier Welborn [ID] *[ab]

Polarizable force fields crucially enhance the modeling of macromolecules in polar media. Here, we present new parameters to model six common monosaccharides with the polarizable AMOEBA force field. These parameters yield a thinner, but denser, hydration layer than that previously reported. This denser hydration layer results in eliminating non-physical aggregation of glucose in water—an issue that has plagued molecular dynamics simulations of carbohydrates for decades.

Ubiquitous in Nature, monosaccharides have become building blocks for novel molecules and materials.[1–3] A vast set of monosaccharides with differing numbers and types of functional groups exist. Further, these building blocks form a vast array of polysaccharides, glycoproteins, glycolipids or glycomolecules. This modularity makes carbohydrates relevant to many research fields. However, it also hinders our ability to fully realize, predict and control the properties of natural and synthetic molecules containing carbohydrates. Characterizing these molecules with solution nuclear magnetic resonance, X-ray and neutron spectroscopies provides microscopic, ensemble properties, which can be hard to interpret at the molecular level.[4–7] Modeling can inform these properties, providing that the level of theory used accounts for the intermolecular interactions that govern carbohydrate properties. In practice, carbohydrate modeling gravitates towards classical molecular dynamics (MD), whose accuracy depends on that of the underlying force field.[1,2,8,9] Over the past decade, multiple force fields have been tested, or developed specifically, for carbohydrate modeling, including, but not limited to, GROMOS, CHARMM, Drude (polarizable), GLYCAM, and CSFF.[8–11]

Previous MD studies highlight the importance of polarization for the modeling of monosaccharides in water.[8,9] Indeed, explicit polarization shows an increase in conformational flexibility, expansion of monosaccharides' solvation shell, and changes in distance and orientation of proximal waters.[9] Meanwhile, implicit water polarization (*e.g.*, TIP5P) better balances the van der Waals attraction term between and among carbohydrates. More specifically, force fields that use the 12–6 Lennard-Jones potential overestimate the attraction between monosaccharides, yielding to non-physical aggregation of β-D-glucose at low concentration.[8,12] Although they demonstrated that the issue can be mitigated by improving the electrostatic term with polarization and re-optimizing the Lennard-Jones parameters,[8,13] non-physical aggregation remains a challenge to predictive models of carbohydrates in diverse environments.

Here, we assessed the AMOEBA polarizable force field[14] for modeling carbohydrates in solution. AMOEBA includes explicit polarization for water and carbohydrates *via* induced atomic dipoles, as well as a buffered 14–7 Halgren potential for van der Waals interactions, which addresses the limitations of other force fields listed above. We used poltype 2[15] as is (*i.e.*, without refining the generic van der Waals or torsion parameters), to compute new AMOEBA parameters for the α and β anomers of D-glucose (Glc), D-glucosamine (GlcN), *N*-acetyl-D-glucosamine (GlcNAc), D-galactose (Gal), D-galactosamine (GalN), and *N*-acetyl-D-galactosamine (GalNAc), shown in Fig. 1.

Using the new poltype 2 AMOEBA parameters, we performed an energy minimization, followed by 150 ns MD simulation of each monosaccharide in water with constant number of particles, pressure and temperature (NPT) in Tinker 9.[16] The first 5 ns of the MD were used as equilibration and the remaining 145 ns as production run to compute the properties described below. Each simulation was run twice, once with mutual polarization, converging induced dipoles at each step within 10^{-5} D, and once without the dipole polarization term (see input files in ESI†).

Average bond lengths, angles, and ring dihedrals (ESI†) agree with experimental data for all twelve anomers. Comparing these

[a] Department of Chemistry, Virginia Tech, Blacksburg, VA 24061, USA.
  E-mail: vwelborn@vt.edu
[b] Macromolecules Innovation Institute, Virginia Tech, Blacksburg, VA 24061, USA
[c] Department of Chemistry and Biochemistry, Bridgewater College, Bridgewater, VA 22812, USA
[d] Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, VA 23529, USA

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4cc04415k
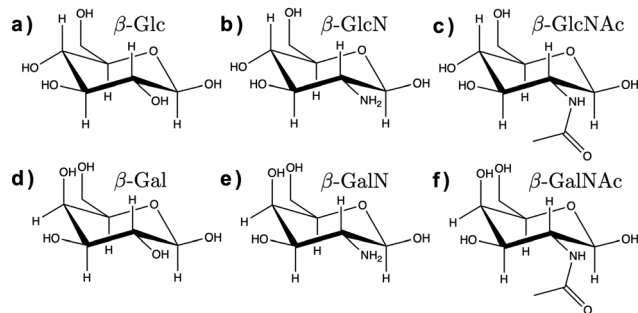
Fig. 1 Chemical structure of the β anomers of the six monosaccharides studied here. (a) Glc, (b) GlcN, (c) GlcNAc, (d) Gal, (e) GalN, (f) GalNAc. The chemical structure of the α anomers is provided in ESI† (Fig. S1).

geometric variables with and without mutual polarization, we see that polarization does not change the structure of the monosaccharides, nor does it yield enhanced flexibility. This suggests that force field accuracy is mainly revealed in the solution properties of the modeled carbohydrates (*i.e.*, the structure and density of the hydration layer, diffusivity, *etc.*), which will be the focus of this paper. Nevertheless, we note two observations about the conformation of these monosaccharides, as predicted with AMOEBA.

First, we record ring puckering for a majority of the α anomers. This contrasts with the β anomers that all remain in a stable chair conformation throughout the duration of the simulation. In Fig. 2, we show the evolution of the ring dihedral $C_1-C_2-C_3-C_4$ for α-Glc. We see that α-Glc spends a majority of the simulation in the chair conformation ($C_1-C_2-C_3-C_4 \sim -55°$) but also temporarily visits alternative ring puckered conformations. Overall, α-Glc, α-GlcN, α-GlcNAc, α-Gal, α-GalN and α-GalNAc spends 2.9, 7.4, 0.09, 0.07, 2.6 and 0.05% of the 145 ns MD trajectory, respectively, in a conformation other than chair. Extensions of the simulations of α- and β-Glc, GlcN, Gal and GalN to 300 ns, and replicate simulations (ESI†), confirm this trend. Although this effect has been reported before,[17,18] it is non trivial to validate against experiments and it remains possible that these conformational switches are non-physical. However, we obtain similar results when using the β AMOEBA
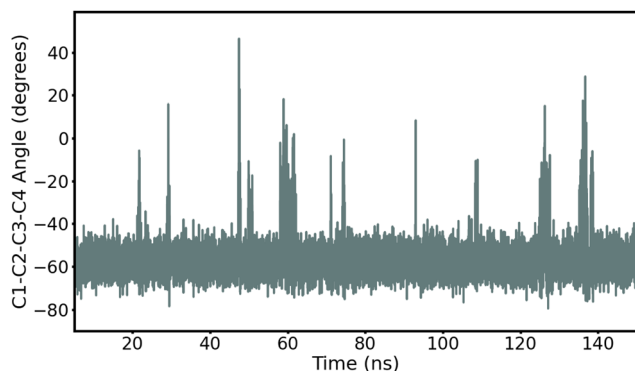
parameters to model the α anomers (Fig. S4–S7, ESI†), suggesting that this effect is due to the structure and not the force field parameters.

Second, the dihedral $O_6-C_6-C_5-O_5$ reveals a ratio for the three gg, gt, and tg populations that do not agree with experimental data. For example, we find that β-Glc is 70(39)% gg, 28(22)% gt and 2(39)% tg in our simulations with mutual (no) polarization, compared with 31, 61 and 8% in experiments.[4,5] This likely indicates that this degree of freedom equilibrates on timescales longer than is simulated here. A solution would be to separate the frames that characterize each rotamer, compute their properties independently and reweigh them according to the experimental ratio. However, we do not anticipate the different rotamers to have significantly different solution properties. Therefore, for the remainder of this paper, the data are computed from the unmodified MD trajectories.

In Fig. 3, we present the radial distribution function of the hydroxyl oxygens to the water oxygens for the β anomers (see Fig. S2 for the α anomers plot, ESI†). In Table 1, we show the position of the first peak, position of the first minima and coordination number at the first minima, for β-Glc, as reference.

The six monosaccharides exhibit a very similar hydration pattern with water oxygens, most likely, 2.85 Å away from the hydroxyl oxygens (Table 1). Looking at the radial distribution functions from our MD simulations performed without polarization (dotted lines in Fig. 3), we see shorter and broader peaks. These results indicate that polarization enhances the structure of proximal waters, but does not result in an expanded hydration shell as reported elsewhere.[9] Rather, we compute an hydration shell of 3.5 Å for β-Glc, compared to 4.05 Å without polarization (Tables 1), and 3.8 Å as predicted by the polarizable Drude force field.[9] In our simulations, we see a tighter, more ordered, hydration layer with higher coordination number, corresponding to a significantly denser hydration shell than has been previously observed with other force fields.[9]

Fig. 2 Time evolution of the torsion angle $C_1-C_2-C_3-C_4$ in α-Glc during the 145 ns production MD with AMOEBA (mutual polarization). $C_1-C_2-C_3-C_4$ is −55° when α-Glc is in chair conformation.
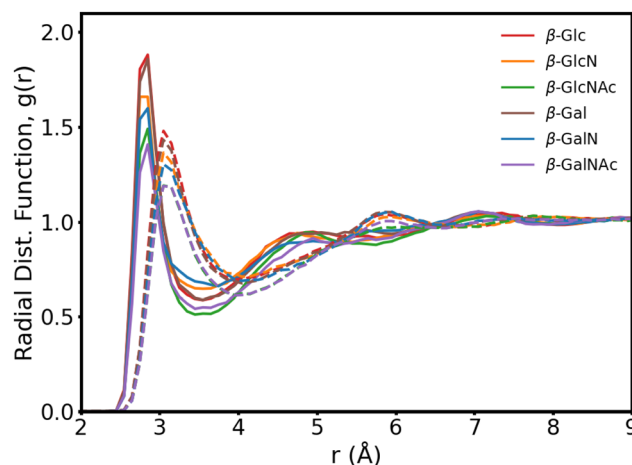

Fig. 3 Carbohydrate–water radial distribution function for the six β anomers, with (solid lines) and without (dotted lines) mutual polarization. Each curve is an average of the O_1-O_w, O_2/N_2-O_w, O_3-O_w and O_4-O_w distribution functions, where O_w is the water oxygen.

Table 1 First peak position (r_0), first minimum position (r_{min}) and coordination number (CN) for hydroxyl–water radial distribution function of β-Glc with and without mutual polarization

Oxygen	Mutual pol.			No pol.		
	r_0 (Å)	r_{min} (Å)	CN	r_0 (Å)	r_{min} (Å)	CN
O_1	2.85	3.55	4.16	3.05	4.25	6.18
O_2	2.85	3.45	3.59	3.05	4.05	4.94
O_3	2.85	3.65	4.22	3.05	3.95	4.48
O_4	2.85	3.45	3.37	3.05	3.95	4.21
Average	2.85	3.55	3.90	3.05	4.05	4.92
O_6	2.85	3.35	3.31	3.05	3.85	4.28

In Table 2, we list the hydration number of each monomer, computed with and without mutual polarization, as well as the number of acceptor and donor hydrogen bonds (H-bonds) to the hydroxyls and nitrogen groups. We find that Glc is surrounded by 25 ± 8 water molecules, which is consistent with the 21 hydration number found with dielectric spectroscopy.[19] Further, we observe a relatively large standard deviation of the hydration number for all monosaccharides, regardless of the anomeric form. This suggests a dynamic hydration layer where the water molecules exchange position on a timescale of tens of picoseconds. Interestingly, we see that polarization increases the hydration number by about 5 in all cases and the standard deviation by 2. Therefore, polarization induces a thin, ordered, hydration layer that is also very dynamic. Finally, we see that the hydroxyl groups in all monosaccharides are better hydrogen bond acceptors than they are donors, consistent with previous work.[11]

To further test the performance of these new AMOEBA force field parameters for monosaccharides, we ran MD simulations

Table 2 Hydration number (number of water molecules within 4 Å of each monosaccharide) and number of acceptor and donor hydrogen bonds to the hydroxyl groups

System	Hydration number	Acceptor H-bonds	Donor H-bonds
α-Glc (mutual pol.)	24.7 ± 7.3	5.4 ± 1.3	3.6 ± 1.0
α-Glc (no pol.)	19.4 ± 5.9	2.9 ± 1.3	1.7 ± 1.0
β-Glc (mutual pol.)	24.6 ± 7.6	5.5 ± 1.4	3.5 ± 1.0
β-Glc (no pol.)	19.4 ± 5.9	2.7 ± 1.3	1.6 ± 1.0
α-GlcN (mutual pol.)	24.5 ± 7.3	5.3 ± 1.2	3.2 ± 1.0
α-GlcN (no pol.)	19.0 ± 5.8	3.0 ± 1.3	1.5 ± 1.0
β-GlcN (mutual pol.)	24.3 ± 7.4	5.4 ± 1.3	3.4 ± 1.1
β-GlcN (no pol.)	19.5 ± 5.9	3.0 ± 1.3	1.6 ± 1.0
α-GlcNAc (mutual pol.)	28.6 ± 8.3	4.5 ± 1.2	3.2 ± 1.0
α-GalNAc (no pol.)	22.3 ± 7.0	2.2 ± 1.2	1.7 ± 1.1
β-GlcNAc (mutual pol.)	28.0 ± 8.8	4.6 ± 1.3	3.6 ± 1.1
β-GlcNAc (no pol.)	22.5 ± 6.8	2.3 ± 1.2	1.8 ± 1.0
α-Gal (mutual pol.)	25.3 ± 7.3	5.2 ± 1.3	3.5 ± 1.0
α-Gal (no pol.)	19.6 ± 5.9	2.8 ± 1.3	1.6 ± 1.0
β-Gal (mutual pol.)	24.4 ± 7.4	5.3 ± 1.3	3.4 ± 1.0
β-Gal (no pol.)	19.6 ± 5.9	2.8 ± 1.3	1.6 ± 1.0
α-GalN (mutual pol.)	24.0 ± 7.4	5.0 ± 1.2	3.1 ± 1.0
α-GalN (no pol.)	19.3 ± 6.0	3.0 ± 1.3	1.6 ± 1.0
β-GalN (mutual pol.)	25.2 ± 7.3	5.3 ± 1.3	3.3 ± 1.1
β-GalN (no pol.)	19.6 ± 5.9	2.9 ± 1.3	1.6 ± 1.0
α-GalNAc (mutual pol.)	28.2 ± 8.7	4.3 ± 1.2	3.3 ± 1.0
α-GalNAc (no pol.)	22.6 ± 6.9	2.3 ± 1.2	1.6 ± 1.0
β-GalNAc (mutual pol.)	27.6 ± 8.7	4.4 ± 1.2	3.4 ± 1.0
β-GalNAc (no pol.)	22.7 ± 6.9	2.4 ± 1.2	1.7 ± 1.0

(150 ns in the NPT ensemble) of multiple β-Glc in water. Following previous work,[8] we simulated six different concentrations, all below the concentration at which β-Glc aggregates: 0.25, 0.5, 1, 2, 3 and 4 mol kg^{-1}. In each case, we computed the radial distribution function of the center-of-mass (COM) of β-Glc, as well as their mean-square displacement. We fitted the mean-square displacement as a function of time with a linear function (Fig. S3, ESI†) to extract the diffusion coefficient for each concentration:

$$D = \lim_{t \to \infty} \frac{\langle r^2 \rangle}{6t} \tag{1}$$

where D is the diffusion coefficient, $\langle r^2 \rangle$ the mean-square displacement of β-Glc (the mean is carried over the number of β-Glc in the simulation) and t the time. Fig. 4 presents the results. In previous MD studies, non-physical β-Glc aggregation manifested within a few ns, characterized by a sharp peak in the radial distribution function at about 5.5 Å. In our simulations, we do not see this feature (Fig. 4c). Rather, we see a much smaller and broader peak at 7 Å, whose integration (Fig. 4d) scales with the concentration of β-Glc. This suggests that non-physical aggregation of the monosaccharides in water does not equate to the broad peak in our radial distribution function. Further, the diffusion coefficients computed from these simulations agree well with experimental data (Fig. 4b), where AMOEBA only slightly underestimates the diffusion of β-Glc in water, outperforming other force fields.[8] Running MD for the highest concentration (c = 4 mol kg^{-1}) without mutual polarization results in a similar, although slightly broader, COM radial distribution function (light green curve in Fig. 4c and d). This result emphasizes van der Waals interactions as governing the aggregation of monosaccharides in solution, not polarization.

In summary, the polarizable AMOEBA force field with its robust multipole expansion, atomic induced dipoles and buffered 14–7 Halgren potential outperforms other force fields when modeling monosaccharides in solution. As AMOEBA has also been proven accurate for water properties and protein modeling, it provides the opportunity for a unified set of parameters that can described saccharides in diverse hydrated environments.

Data availability

The data supporting this article have been included as part of the ESI.† Input files, including Poltype 2 and Tinker 9 input
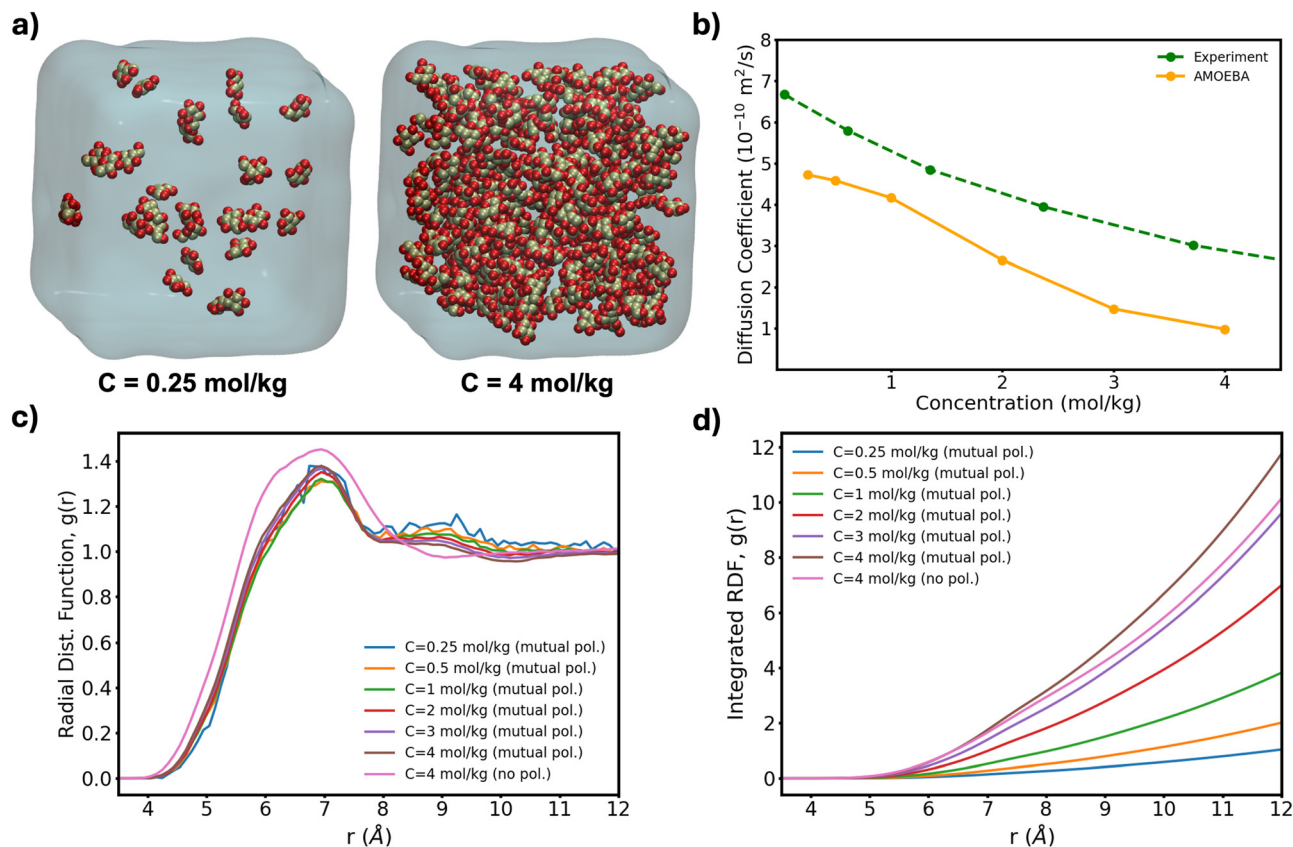
Fig. 4 (a) Lowest and highest Glc concentration modeled here. (b) Diffusion coefficient computed from eqn (1) as a function of concentration for our box sizes of ~60^3 Å^3 (~20 k atoms). Extrapolation of MD-calculated diffusion coefficients to bulk conditions can yield a 10% increase[20] although no correction has been applied here. (c) Radial distribution function of the COM of β-Glc and (d) corresponding integration data.

with parameters files are available at **https://github.com/WelbornGroup**.

Conflicts of interest

There are no conflicts to declare.

References

1 A. Plazinska and W. Plazinski, *J. Chem. Theory Comput.*, 2021, **17**, 2575–2585.

2 X. Xiong, Z. Chen, B. P. Cossins, Z. Xu, Q. Shao, K. Ding, W. Zhu and J. Shi, *Carbohydr. Res.*, 2015, **401**, 73–81.

3 C. Molteni and M. Parrinello, *J. Am. Chem. Soc.*, 1998, **120**, 2168–2171.

4 R. Stenutz, I. Carmichael, G. Widmalm and A. S. Serianni, *J. Org. Chem.*, 2002, **67**, 949–958.

5 C. Thibaudeau, R. Stenutz, B. Hertz, T. Klepach, S. Zhao, Q. Wu, I. Carmichael and A. S. Serianni, *J. Am. Chem. Soc.*, 2004, **126**, 15668–15685.

6 J. Ø. Duus, C. H. Gotfredsen and K. Bock, *Chem. Rev.*, 2000, **100**, 4589–4614.

7 J. S. Stevens and S. L. Schroeder, *Surf. Interface Anal.*, 2009, **41**, 453–462.

8 J. Sauter and A. Grafmüller, *J. Chem. Theory Comput.*, 2015, **11**, 1765–1774.

9 P. Pandey and S. S. Mallajosyula, *J. Phys. Chem. B*, 2016, **120**, 6621–6633.

10 D. S. Patel, X. He and A. D. MacKerell Jr, *J. Phys. Chem. B*, 2015, **119**, 637–652.

11 P. Mason, G. Neilson, J. Enderby, M.-L. Saboungi and J. Brady, *J. Phys. Chem. B*, 2005, **109**, 13104–13111.

12 P. S. Schmalhorst, F. Deluweit, R. Scherrers, C.-P. Heisenberg and M. Sikora, *J. Chem. Theory Comput.*, 2017, **13**, 5039–5053.

13 C. R. Rupakheti, A. D. MacKerell Jr and B. Roux, *J. Chem. Theory Comput.*, 2021, **17**, 7085–7095.

14 Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder and P. Ren, *J. Chem. Theory Comput.*, 2013, **9**, 4046–4063.

15 B. Walker, C. Liu, E. Wait and P. Ren, *J. Comput. Chem.*, 2022, **43**, 1530–1542.

16 M. Harger, D. Li, Z. Wang, K. Dalby, L. Lagardère, J.-P. Piquemal, J. Ponder and P. Ren, *J. Comput. Chem.*, 2017, **38**, 2047–2055.

17 I. Alibay and R. A. Bryce, *J. Chem. Inf. Model.*, 2019, **59**, 4729–4741.

18 V. Lutsyk and W. Plazinski, *J. Phys. Chem. B*, 2024, **128**, 472–491.

19 K. Shiraga, T. Suzuki, N. Kondo, T. Tajima, M. Nakamura, H. Togo, A. Hirata, K. Ajito and Y. Ogawa, *J. Chem. Phys.*, 2015, **142**, 234504.

20 I.-C. Yeh and G. Hummer, *J. Phys. Chem. B*, 2004, **108**, 15873–15879.