International Journal of
Molecular Sciences

MDPI

Article

Pyranose Ring Puckering Thermodynamics for Glycan Monosaccharides Associated with Vertebrate Proteins

Olgun Guvench [1,2,*], Devon Martin [1,2] and Megan Greene [1]

[1] Department of Pharmaceutical Sciences and Administration, School of Pharmacy, University of New England, 716 Stevens Avenue, Portland, ME 04103, USA; dmartin11@une.edu (D.M.); mgreene3@une.edu (M.G.)
[2] Graduate School of Biomedical Science and Engineering, University of Maine, 5775 Stodder Hall, Orono, ME 04469, USA
* Correspondence: oguvench@une.edu; Tel.: +1-207-221-4171

Abstract: The conformational properties of carbohydrates can contribute to protein structure directly through covalent conjugation in the cases of glycoproteins and proteoglycans and indirectly in the case of transmembrane proteins embedded in glycolipid-containing bilayers. However, there continue to be significant challenges associated with experimental structural biology of such carbohydrate-containing systems. All-atom explicit-solvent molecular dynamics simulations provide a direct atomic resolution view of biomolecular dynamics and thermodynamics, but the accuracy of the results depends on the quality of the force field parametrization used in the simulations. A key determinant of the conformational properties of carbohydrates is ring puckering. Here, we applied extended system adaptive biasing force (eABF) all-atom explicit-solvent molecular dynamics simulations to characterize the ring puckering thermodynamics of the ten common pyranose monosaccharides found in vertebrate biology (as represented by the CHARMM carbohydrate force field). The results, along with those for idose, demonstrate that the CHARMM force field reliably models ring puckering across this diverse set of molecules, including accurately capturing the subtle balance between $^{4}C_{1}$ and $^{1}C_{4}$ chair conformations in the cases of iduronate and of idose. This suggests the broad applicability of the force field for accurate modeling of carbohydrate-containing vertebrate biomolecules such as glycoproteins, proteoglycans, and glycolipids.

Keywords: glucose; GlcNAc; galactose; GalNAc; mannose; xylose; fucose; Neu5Ac; glucuronate; iduronate; tetrahydropyran

check for updates

Citation: Guvench, O.; Martin, D.; Greene, M. Pyranose Ring Puckering Thermodynamics for Glycan Monosaccharides Associated with Vertebrate Proteins. *Int. J. Mol. Sci.* **2022**, *23*, 473. https://doi.org/10.3390/ijms23010473

Academic Editors: Istvan Simon and Csaba Magyar

Received: 6 December 2021
Accepted: 28 December 2021
Published: 31 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Glycosylation is a common and important post-translational modification to proteins in eukaryotic biology. Additionally, carbohydrates are key components of eukaryotic lipids that make up the bilayers in which transmembrane proteins are embedded [1]. The carbohydrate portions of glycosylated proteins and glycolipids are called glycans. Naturally occurring glycans in vertebrates, including in humans, are composed of the monosaccharides D-glucose (Glc), *N*-acetyl-D-glucosamine (GlcNAc), D-galactose (Gal), *N*-acetyl-D-galactosamine (GalNAc), D-mannose (Man), D-xylose (Xyl), L-fucose (Fuc), *N*-acetyl-D-neuraminic acid (Neu5Ac), D-glucuronic acid (GlcA), and L-iduronic acid (IdoA), all in their pyranose forms [2] (Figure 1). As Neu5Ac, GlcA, and IdoA are expected to be deprotonated under typical physiological conditions, Figure 1 shows their conjugate base forms, *N*-acetyl-D-neuraminate, D-glucuronate, and L-iduronate, and it is these forms that are exclusively considered in what follows. Examples of glycans as components of glycosylated proteins are the *N*-glycans [3] and O-glycans [4] attached to glycoproteins and the glycosaminoglycans attached to proteoglycans [5]. Experimental atomic-resolution structural biology on glycosylated proteins is complicated by the non-template based synthesis of the attached glycans [6], which precludes a convenient source of homogeneous

sample from biological sources, the intrinsic flexibility of glycans, which hinders conformational analysis by X-ray crystallography and NMR spectroscopy [7], and the covalent linkage of proteins with glycans, which can affect the structural properties of both the glycan and protein components [8–10]. In the context of membrane proteins, experimental atomic-resolution structural biology using X-ray crystallography entails extracting the membrane protein from its native lipid environment in order to create protein crystals [11], which means the effects of natural glycolipids in the native bilayer are not included in the structure determination. Therefore, the impact of glycans on protein structure continues to be at the frontiers of protein structure research.

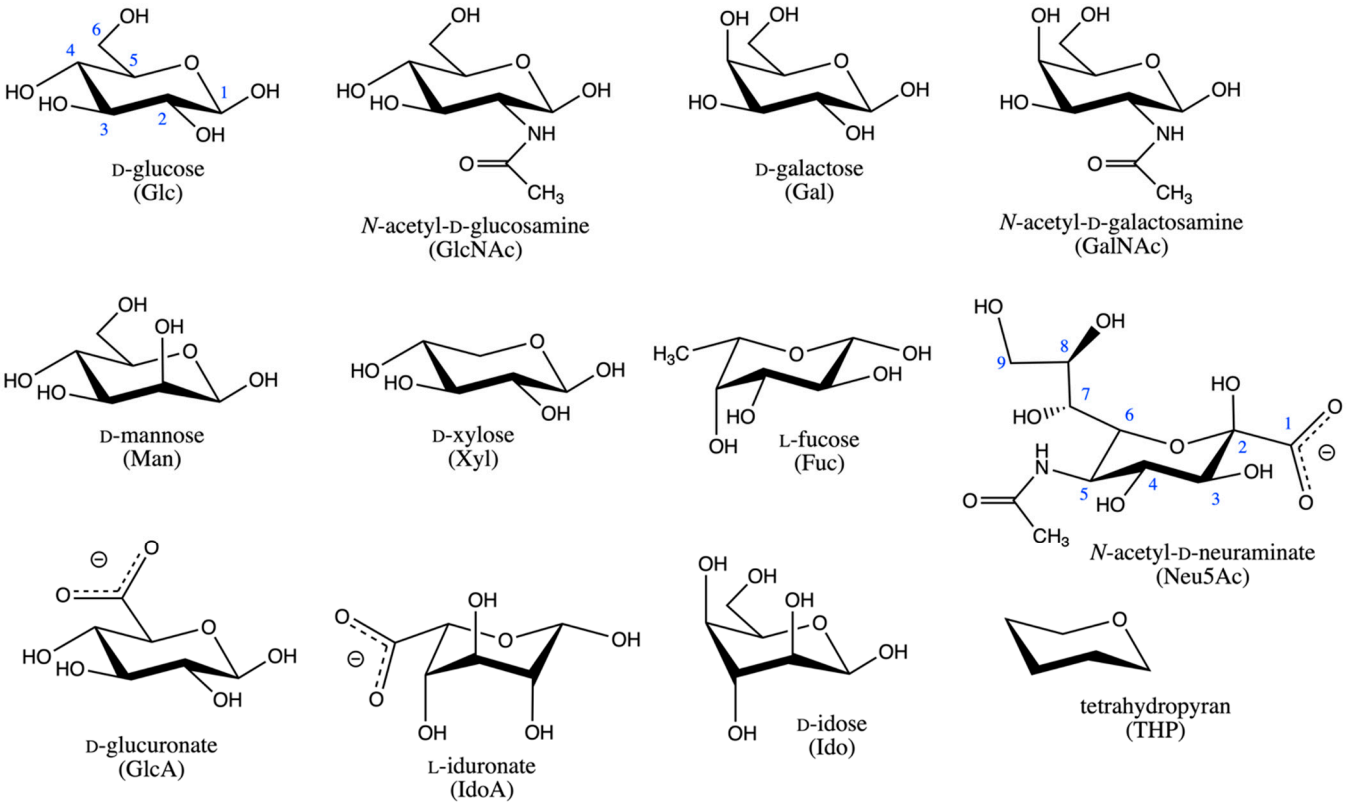

Figure 1. Compounds considered in the current study. Glc carbon atoms are numbered in blue. All other monosaccharides follow the same numbering scheme, except for Neu5Ac, which is numbered as pictured. All monosaccharides are drawn as the β anomer. The α anomer is created by inversion of the configuration at carbon 2 for Neu5Ac and at carbon 1 for all other monosaccharides. Both anomers for each monosaccharide as well as the corresponding O-methyl glycosides, formed by methylation at the anomeric carbon hydroxyl, were studied, for a total of 45 compounds (44 monosaccharides + THP).

Computational approaches for three-dimensional modeling of the atomic-resolution conformational properties of glycans have been developed and applied to help bridge the gaps in experimental methods [12–26]. Widely used among these computational approaches are explicit solvent molecular dynamics simulations employing atomistic force fields such as GLYCAM06 [27,28], GROMOS 53A6GLYC [29,30], GROMOS 56a6CARBO/CARBO_R [31–33], OPLS-AA [34,35], and CHARMM [36–39]. The quality of the results from such molecular dynamics simulations depends upon the quality of the force field parametrization. The conformational properties of glycans are determined principally by flexibility in the rings of the constituent monosaccharides and in the glycosidic linkages connecting them (Figure 2) [12,40], and thus it is important for force field parametrizations to accurately capture the physics of these sources of flexibility in order to ensure reliable modeling results.

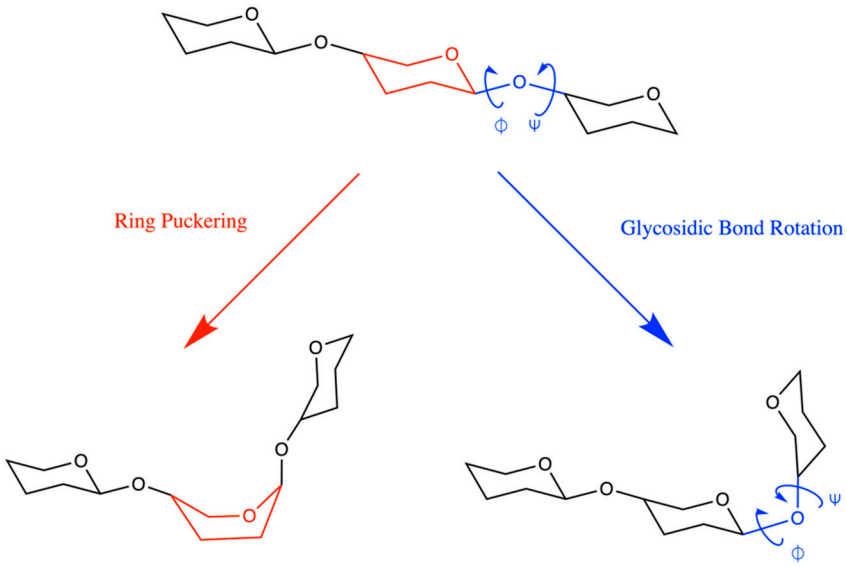

Figure 2. Pyranose ring puckering (red) and glycosidic bond rotation (blue) are the major sources of polymer flexibility in vertebrate glycans.

Since pyranose ring puckering occurs at the microsecond and beyond timescale [40–42], which is near the upper limit of typical present-day all-atom explicit-solvent molecular dynamics simulations, limitations in force field accuracy may not be readily apparent simply based on analysis of such simulation results. Here, we systematically determine the ring puckering thermodynamics of all compounds in Figure 2, including both the α and the β anomers and their corresponding O-methyl glycosides for the ten monosaccharides (i.e., 45 systems total), with the widely-used CHARMM force field. Extended System Adaptive Biasing Force (eABF) [43,44] is applied to achieve well-converged equilibrium statistics for ring puckering probabilities, with error estimates from triplicate 200-ns simulations for each system. The ring puckering thermodynamics from these simulations are in line with expected behavior, including for the highly flexible IdoA, and imply that the CHARMM force field can be used with confidence to correctly capture pyranose ring puckering contributions to glycan conformational heterogeneity in the context of the vertebrate glycans such as N-glycans, O-glycans, glycosaminoglycans, and glycolipids.

We additionally consider idose in its pyranose form, since, like IdoA, idose has a close balance between $^4C_1$ and $^1C_4$ chair probabilities, which makes it a useful test of force field accuracy. In agreement with prior computational results [45] and recent NMR data [46], the CHARMM carbohydrate force field performs very well in capturing the close balance for idose as well as for IdoA. Finally, for completeness, we include tetrahydropyran, which is the basic six-membered ring scaffold common to all of the monosaccharides considered here (Figure 2). As expected, there is an exact 50:50 balance for chair-chair interconversion for THP.

It is possible to tune ring puckering thermodynamics by selectively refining specific force field parameters and by using ring puckering thermodynamics as target data in the parametrization process. In the case of the GROMOS force field, such an approach was taken as a force field revision [31,32,47], and has yielded excellent results for ring puckering across a wide variety of pyranoses [32,33,45,48]. In the case of CHARMM, ring puckering thermodynamics in solution were not used as target data for CHARMM parametrization, and both bonded and nonbonded force field parameters, which built upon quantum mechanical gas phase puckering energetics for tetrahydropyran [36,49], are conserved across all of the different monosaccharides considered here. This demonstrates it is possible to correctly account for pyranose monosaccharide ring puckering thermodynamics in solution with a general transferable bonded and nonbonded force field parameter set. In the case of CHARMM, combining this parameter set with CHARMM force field parameters

for proteins [50–52] can enable accurate modeling of glycoproteins and proteoglycans, and combining these parameters set with CHARMM force field parameters for lipids can do the same for glycolipids [53,54], which in turn can enable accurate modeling of transmembrane proteins embedded in complex bilayers composed of natural lipids.

2. Results and Discussion

2.1. Reaction Coordinate and Sampling Approach

Ring puckering for pyranose monosaccharides is commonly described using the Cremer-Pople (C-P) parameters (Q, θ, ϕ) [55], which provide a convenient quantitative means to identify both the extent and the nature of the puckering using spherical coordinates. The puckering amplitude Q describes the extent or magnitude of the puckering, while the angular values $0° \leq \theta \leq 180°$ and $0° \leq \phi < 360°$ describe the nature of the puckering. "Polar" values of θ near 0° and 180° correspond to $^4C_1$ and $^1C_4$ chair conformations, respectively, while "equatorial" values of θ near 90° correspond to boat and skew boat conformations, with the ϕ value indicating the specific boat or skew boat (e.g., ^2S_O). Intermediate or "tropical" values of θ, which are between the poles and the equator, correspond to envelope and half-envelope conformations, with the ϕ value indicating the specific envelope or half envelope [56].

Due to the long timescale for interconversion between $^4C_1$ and $^1C_4$, it is impractical to precisely determine the balance of probabilities and, hence, free energy difference, ΔG, between these conformations for pyranose monosaccharides on a routine basis using standard all-atom explicit-solvent molecular dynamics simulations. This is true for pyranoses where $\Delta G \approx 0$ due to the energy barrier separating the conformations [41,42], and the situation is even more difficult in cases where ΔG is substantially different than zero due to the difficulty in achieving equilibrium sampling of the unfavored conformation.

A logical means to address this issue is to apply a bias to θ during a simulation and either to reweight the sampling distribution to get unbiased conformational probabilities or to directly compute ΔG from the bias. Such an approach employing metadynamics [57,58] has enabled a number of studies to this end [29,32,33,45,48,59]. As demonstrated in these studies, this approach allows one to obtain a good estimate for ΔG with much less computation time than through standard (non-biased) molecular dynamics. There are two potential downsides to using a bias on θ. The first is the need to develop specialized computer code for the bias since the C-P θ is not a standard cartesian or internal coordinate. The second is that, while the single parameter θ can differentiate the two chair conformations from each other and also non-chair conformations from chair conformations, it cannot differentiate one non-chair conformation from another non-chair conformation. This second potential downside can be addressed by introducing a second simultaneous bias on ϕ but at the expense of further complicating the first downside.

For these reasons, direct use of dihedral angles is an attractive alternative. For example, Pickett and Strauss (P-S) defined three out-of-plane dihedrals constructed as various combinations of atoms in the pyranose ring [60], and it has been shown that simultaneous biases on all three of these angles can be effectively used to sample pyranose ring puckering [61]. In fact, the P-S and C-P approaches are mathematically equivalent [62]. However, there is an important practical difference with regard to applying biases on C-P parameters versus P-S out-of-plane dihedrals: only the two angular C-P parameters are required to uniquely identify the pucker nature (as opposed to magnitude) of a particular conformation whereas all three P-S out-of-plane dihedrals are required to do the same [59,63].

Babin and Sagui (B-S) have also proposed using dihedral angles for biased sampling of pyranose ring pucker [64]. In contrast to the P-S approach, only two dihedral angles are used in their scheme, $\alpha_1 \equiv$ O5–C1–C2–C3 and $\alpha_2 \equiv$ C3–C4–C5–O5, and the dihedral angles are real dihedrals determined by sequentially bonded atoms. Babin and Sagui have shown that biased sampling of (α_1, α_2) is an effective approach for sampling IdoA and GlcA puckering, and Alibay and Bryce have extended on these two monosaccharides to sulfated variants, as well as to non-sulfated and sulfated variants of GlcNAc, Gal, and

GalNAc [65]. In what follows, we demonstrate that major minima in $\Delta G(\alpha_1, \alpha_2)$ are populated by unique conformations. As such, it is possible to do direct integration of regions of $\Delta G(\alpha_1, \alpha_2)$ to determine ΔG not only between the $^4C_1$ and $^1C_4$ chairs, but also between specific boat/skew-boat conformations.

2.2. Extended System Adaptive Biasing Force (eABF) Sampling of the B-S (α_1, α_2) Reaction Coordinate

Methyl α-L-idopyranosiduronic acid (MeαIdoA) (Figure 2 "IdoA" with a methylated axial C1 hydroxyl) serves as a good test system to demonstrate the efficacy of eABF sampling of (α_1, α_2) owing to a small (<1 kcal/mol [46]) ΔG for conversion between the $^4C_1$ and $^1C_4$ chair conformations and a large energy barrier (~10 kcal/mol from the present work based on transition path saddle points in Figure 3), and therefore, slow kinetics, for this transition. Triplicate 200-ns eABF simulations with simultaneous biases on α_1 and α_2 and seeded with different randomized initial velocities yield essentially identical results across the entire $\Delta G(\alpha_1, \alpha_2)$ surface (Figure 3). Not only are the thermodynamic minima equal in both value and location, but so are the saddle regions and even the maxima, which demonstrates the excellent convergence properties of eABF for this system. $\Delta G(\alpha_1, \alpha_2)$ data are similarly well-converged for all 45 systems in this study (four different anomerization/methylation states for each of the 11 monosaccharides in Figure 2 plus tetrahydropyran; Supplementary Material Figures S1–S12).

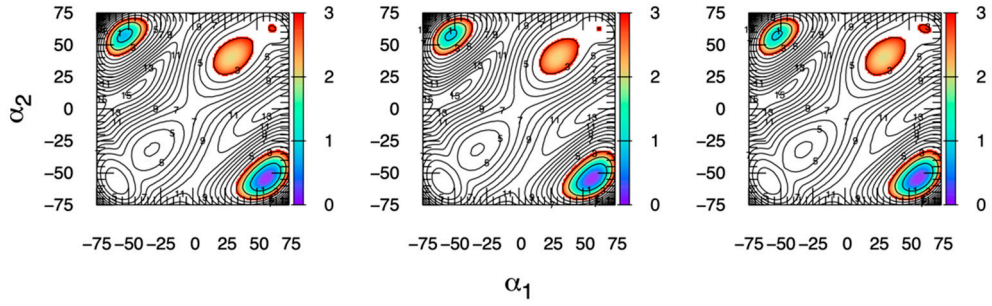

Figure 3. MeαIdoA $\Delta G(\alpha_1, \alpha_2)$ from eABF simulation. Each panel is from a separate 200-ns simulation seeded with different initial random velocities. α_1 and α_2 are in degrees. $\Delta G(\alpha_1, \alpha_2)$ is in kcal/mol, with contours drawn every 1 kcal/mol, colored from 0–3 kcal/mol, and labeled every 2 kcal/mol.

Additionally, each major thermodynamic minimum, that is, where $\Delta G(\alpha_1, \alpha_2) < 3$ kcal/mol, is populated by a single type of ring puckering conformation (Figure 4). We have chosen 3 kcal/mol as a cutoff value for the definition of major thermodynamic minimum since, at the simulation temperature of 298 K, values greater than 3 kcal/mol correspond to small probabilities, specifically, less than 0.64%. This association between a single ring puckering conformation and each major thermodynamic minimum in $\Delta G(\alpha_1, \alpha_2)$ holds for all 44 monosaccharides in this study, which illustrates the practical utility of the B-S reaction coordinate for characterizing pyranose ring puckering not only for chair conformations but also for specific non-chair conformations.

Kinetic data from the simulations clearly show the efficacy of eABF combined with the B-S (α_1, α_2) reaction coordinate for effectively sampling pyranose ring pucker, which is not surprising given the excellent convergence properties of $\Delta G(\alpha_1, \alpha_2)$ with eABF as discussed previously. Serving as a negative control, standard (non-biased) triplicate simulations of MeαIdoA starting from the $^1C_4$ chair undergo at most one transition in C-P θ during 200 ns (Figure 5a). Specifically, two of the simulations maintain $\theta \cong 180°$, indicating they are trapped in the initial conformation, while the third transitions at 25 ns to $\theta \cong 90°$ and remains there, indicating it is stuck in the equatorial boat/skew-boat region of puckering space. Therefore, standard sub-microsecond explicit solvent molecular dynamics simulation is inadequate for the task of sampling puckering conformations for pyranoses modeled with the CHARMM force field.

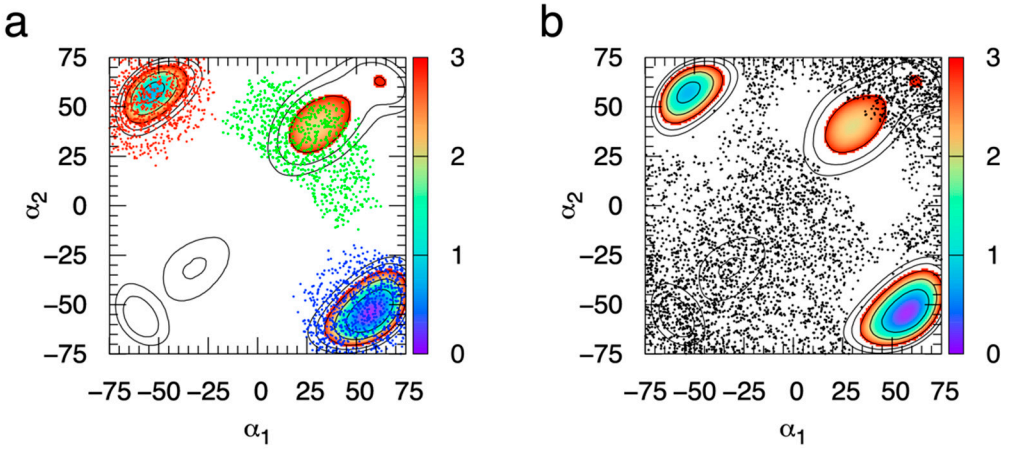

Figure 4. Sampling of specific MeαIdoA ring puckering conformations during eABF simulation with the Babin-Sagui (α_1, α_2) reaction coordinate. Sampled (α_1, α_2) values are separated into those for $^4C_1$, $^1C_4$, and ^2S_O (blue, red, and green dots, respectively, in panel "a") and for all other (black dots, panel "b") puckering conformations. α_1 and α_2 are in degrees. $\Delta G(\alpha_1, \alpha_2)$ is in kcal/mol, with contours drawn every 1 kcal/mol from 0–5 kcal/mol and colored from 0–3 kcal/mol. Puckering data have been aggregated across the triplicate simulations, and $\Delta G(\alpha_1, \alpha_2)$ is from the first simulation in the triplicate.

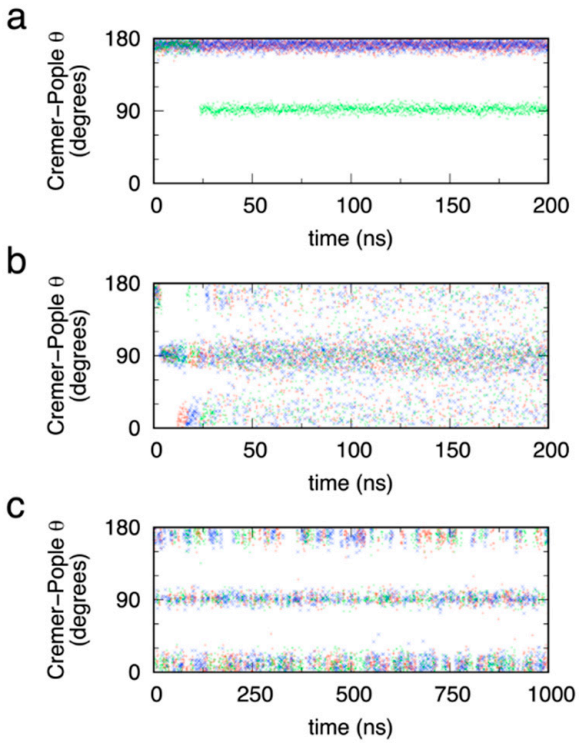

Figure 5. MeαIdoA conformational transitions in standard (non-biased) (**a**), eABF (**b**), and CMAP-biased (**c**) molecular dynamics simulations. eABF and CMAP biased simulations have biasing on the Babin-Sagui (α_1, α_2) reaction coordinate. Data in each panel are from triplicate simulations (blue, red, and green) seeded with different random initial velocities.

In contrast, with eABF sampling, during the first 25 ns, as the time-dependent biasing force becomes a progressively better estimate of the thermodynamic force along (α_1, α_2), transitions in θ start to become induced (Figure 5b). Beyond t = 25 ns, there is rapid transitioning on the nanosecond timescale from the $^4C_1$ chair ($\theta \cong 0°$), through boat/skew-boat conformations ($\theta \cong 90°$), to the $^1C_4$ chair ($\theta \cong 90°$) and back again, indicating sufficient

sampling of (α_1, α_2) by eABF to provide an accurate estimate of the thermodynamic force along (α_1, α_2). As a technical point, the eABF approach applies a bias not to (α_1, α_2) directly but to extended degrees of freedom attached to (α_1, α_2), and the thermodynamic force on (α_1, α_2) is recovered from the biasing force applied to these extended degrees of freedom [43,44]. Standard ABF is not possible for sampling (α_1, α_2) since α_1 and α_2 do not meet the required orthogonality condition for standard ABF [66–68] owing to the sharing of atoms carbon 1 and oxygen 5 in both of the dihedral angle definitions. For additional information on this point, we refer interested readers to the cyclohexane data in Figure 2 of reference [44] and the associated discussion therein, which vividly demonstrates errors in estimation of cyclohexane puckering free energy with standard ABF that are corrected with eABF.

As a positive control, and similar to the approach of Babin and Sagui [64], we ran an additional set of simulations that employed CMAP-biased sampling [51,69]. In these simulations, the potential energy was defined by $U_{\text{non-biased}} + U_{\text{CMAP}}$, where $U_{\text{non-biased}}$ is the same CHARMM additive force field function used in the non-biased simulations here and U_{CMAP} is $U_{\text{CMAP}}(\alpha_1, \alpha_2) \cong -0.5 \times \Delta G(\alpha_1, \alpha_2)$. Unlike in the eABF simulations, the bias, in this case from the CMAP term, is fixed. We note that $U_{\text{CMAP}}(\alpha_1, \alpha_2)$ is only approximately equal to $-0.5 \times \Delta G(\alpha_1, \alpha_2)$ since, while $\Delta G(\alpha_1, \alpha_2)$ was computed on a square grid with a grid spacing of 1°, the grid spacing for the CMAP potential is 15°. As expected, there is excellent sampling of C-P θ from the very beginning of the triplicate simulations (Figure 5c). While there is rapid barrier crossing with this approach, there is less uniform sampling across all values of θ as compared to eABF sampling, with a strong tendency to favor sampling of polar and equatorial values of θ as compared to tropical values (Figure 5b vs. Figure 5c). This resulted from the factor of 0.5 used in the definition of $U_{\text{CMAP}}(\alpha_1, \alpha_2)$, and was done to maximize importance sampling of thermodynamically favored regions of (α_1, α_2) space while still lowering barriers sufficiently to achieve ergodic sampling of (α_1, α_2) on the 200-ns time scale of the simulations. As expected, thermodynamically unfavored regions of (α_1, α_2) correspond to tropical values, which in turn are envelope and half-envelope conformations with high degrees of ring strain.

Plotting C-P (θ, ϕ) values sampled during the eABF and the CMAP-biased simulations further validates the degree to which these two biasing methods applied to the (α_1, α_2) reaction coordinate enable sampling of pyranose puckering space. In addition to excellent coverage of the two chair conformations $^4C_1$ and $^1C_4$ located in the polar regions, there is good coverage of the equatorial region for $75° < \phi < 270°$, which includes $^5S_1$, $^{2,5}B$, ^2S_O, $B_{3,O}$, $^1S_3$, $^{1,4}B$, and $^1S_5$, in order of increasing ϕ (Figure 6). That said, there is very limited sampling of equatorial regions outside this range of ϕ values, resulting from the fact that the two-dimensional B-S (α_1, α_2) reaction coordinate is not a perfect replacement for biased sampling of the two-dimensional C-P (θ, ϕ) reaction coordinate. Nonetheless, it is reasonable to assume conformations not sampled are very high in free energy and that the thermodynamically relevant conformations have all been sampled. This latter point is emphasized by comparing these sampling data for eABF versus CMAP biasing. In the case of eABF, as time increases, sampling approaches that for a distribution biased by $-\Delta G(\alpha_1, \alpha_2)$, whereas for CMAP biasing, sampling is that for a distribution biased by $-0.5 \times \Delta G(\alpha_1, \alpha_2)$, as discussed above. As such, eABF provides more complete coverage of (θ, ϕ) space (Figure 6a) as compared to CMAP-biased sampling (Figure 6b).

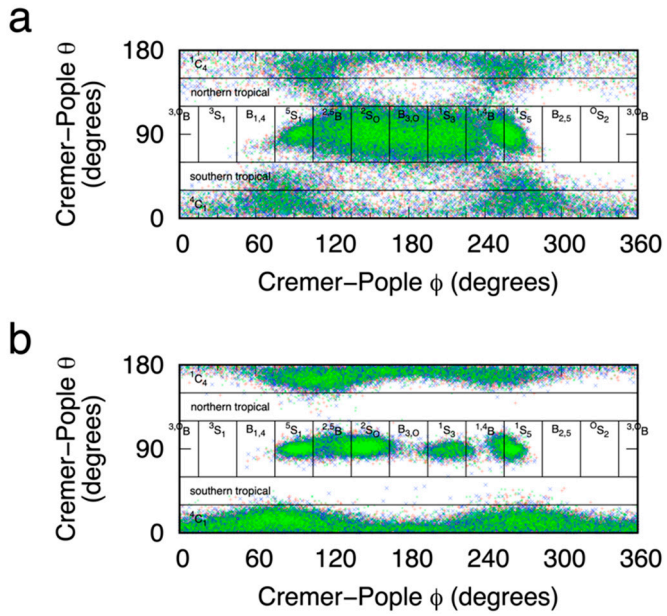

Figure 6. MeαIdoA Cremer-Pople (θ, ϕ) values sampled during eABF (**a**) and CMAP-biased (**b**) molecular dynamics simulations. Pyranose ring puckering regions [56] ("$^4C_1$", "northern tropical", "^2S_O", etc.) are labeled as defined in the Materials and Methods section. Biasing was applied to the Babin-Sagui (α_1, α_2) reaction coordinate. Data in each panel are from triplicate simulations (blue, red, and green) seeded with different random initial velocities. eABF simulations were 200 ns and CMAP-biased simulations were 1000 ns.

2.3. Using eABF-Computed $\Delta G(\alpha_1, \alpha_2)$ to Calculate Specific Ring Puckering Conformation Probabilities

Given that each major thermodynamic minimum for MeαIdoA is populated by a single type of puckering conformation, as shown above, it is possible simply to integrate the probabilities associated with each minimum to determine relative probabilities for specific ring puckering conformations. We operationalized this by converting $\Delta G(\alpha_1, \alpha_2)$ values from eABF simulations to probabilities $p(\alpha_1, \alpha_2)$ using the Boltzmann relationship $p(\alpha_1, \alpha_2) = \exp(\Delta G(\alpha_1, \alpha_2)/-RT)$, where R is the universal gas constant and T is the temperature. We then separated the data based on the (α_1, α_2) quadrant, and summed up all values of p for each (α_1, α_2) bin having an associated value $\Delta G(\alpha_1, \alpha_2) < 3$ kcal/mol within a 20° degree radius of the most favorable thermodynamic minimum in that quadrant. This yields at most one summed probability, P, per quadrant of the (α_1, α_2) coordinate. In the case of MeαIdoA, there are three such values, $P_{+,+}$, $P_{-,+}$, and $P_{+,-}$; the subscript here indicates the quadrant, for example, the quadrant defined by ($\alpha_1 < 0°$, $\alpha_2 > 0°$) for "−, +". As discussed above, for MeαIdoA, the "+, −" minimum corresponds uniquely to the $^4C_1$ ring pucker conformation, "−, +" to $^1C_4$, and "+, +" to ^2S_O (Figure 4), which allows for the assignment of probability values to specific ring pucker conformations based on eABF $\Delta G(\alpha_1, \alpha_2)$ results.

2.4. Ring Puckering Probabilities: Idose and Iduronate

Among the molecules considered in this study (Figure 2), the Ido and IdoA compounds are well known to exhibit significant conformational flexibility with regard to ring pucker. There are recent high-quality experimental results quantifying this, but with variable agreement with prior molecular dynamics simulation studies [46]. Comparison of $^4C_1$:$^1C_4$ ring puckering probability ratios shows good agreement between the present simulation results and these available experimental data (Table 1). In addition to probabilities from the eABF $\Delta G(\alpha_1, \alpha_2)$ results, we have included probabilities computed from the CMAP-biased simulations. These were determined by collecting all $^4C_1$ conformations from a CMAP-biased simulation, assigning a probability $p = \exp(U_{CMAP}/-RT)$ to each conformation to

account for the effect of the CMAP bias, and then summing up the p values to get the total probability for the $^4C_1$ pucker. This was likewise carried out for the $^1C_4$ pucker, and the two total probabilities were normalized to sum to 100% (Table 1, "CMAP-biased simulations").

Table 1. $^4C_1$:$^1C_4$ ring puckering probability ratios in idose (Ido) and iduronate (IdoA) compounds.

Compound	eABF Simulations [1]	CMAP-Biased Simulations [1]	Experimental [46]
αIdo	17.6:82.4 (1.8)	15.1:84.9 (1.9)	18:82
MeαIdo	16.1:83.9 (0.7)	18.1:81.9 (2.1)	42:58
βIdo	97.1:2.9 (0.7)	90.7:9.3 (1.7)	82:18
MeβIdo	82.8:17.2 (2.2)	76.6:23.4 (1.5)	74:26
MeαIdoA	82.9:17.1 (1.4)	77.2:22.8 (1.0)	61:39

[1] Data are averages from triplicate simulations with standard error of the mean values in parentheses.

Converting the $^4C_1$:$^1C_4$ ring puckering probability ratios r to free energies using the relationship $\Delta G = -RT\ln(r)$ and plotting these ΔG values further illustrates how well the force field approach treats the close balance between $^4C_1$ and $^1C_4$ ring conformations. These ΔG values for the $^4C_1$ to $^1C_4$ equilibrium from the eABF and from the CMAP-biased simulations are typically within 0.5 kcal/mol of the experimental values (Figure 7a and Figure 7b, respectively). This very small degree of error is excellent for a force field model, and is not much different than what is seen when comparing the results from the two different simulation approaches using the same force field (Figure 7c).

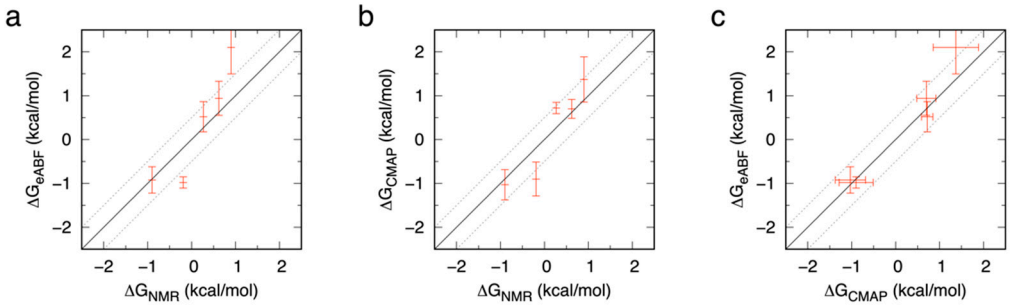

Figure 7. Comparison of ΔG values for the $^4C_1$ to $^1C_4$ equilibrium in Ido and IdoA compounds from eABF simulations, CMAP-biased simulations, and NMR experiments. Data are presented as eABF vs. NMR (**a**), CMAP-biased vs. NMR (**b**), and eABF vs. CMAP-biased (**c**). The specific compounds and the experimental data from NMR experiments are as detailed in Table 1. Simulation data points are averages from triplicate simulations, with error bars representing 95% confidence intervals. The solid diagonal is the line $y = x$, and the dotted diagonal lines are ±0.5 kcal/mol.

2.5. Ring Puckering: $\Delta G(\alpha_1, \alpha_2)$ Minima for All Compounds

Quantitative calculation of pyranose ring puckering probabilities is valuable for comparison to high-quality experimental data for pyranoses with multiple thermodynamically accessible puckering conformations, as in the case of IdoA and Ido. However, such calculation by either integration around eABF $\Delta G(\alpha_1, \alpha_2)$ minima or summing of re-weighted probabilities for individual snapshots from CMAP-biased simulations entails substantial post-simulation effort following the initial computation of $\Delta G(\alpha_1, \alpha_2)$ with eABF. Unlike IdoA and Ido, most of the pyranose monosaccharides considered here are expected to have a single thermodynamically important $\Delta G(\alpha_1, \alpha_2)$ minimum that corresponds to either the $^4C_1$ or $^1C_4$ chair pucker conformation. As such, tabulation of ΔG minima values in the four quadrants of (α_1, α_2) space provides a convenient semi-quantitative means to evaluate the behavior of the force field model for those compounds.

Table 2 lists the ΔG minimum value in each of the four quadrants of (α_1, α_2) space for each of the 45 compounds studied. It also correlates each thermodynamically important minimum (i.e., having a value of <3 kcal/mol) with the puckering conformation associated with the value of (α_1, α_2) for that ΔG minimum. This correlation was carried out using computed Cremer-Pople parameters (detailed in "Materials and Methods: Definition of $^4C_1$, $^1C_4$, ^2S_O, $^OS_2$, and other ring puckering conformations") for trajectory snapshots with (α_1, α_2) values within a 10° radius of the location of the ΔG minimum.

Table 2. Minimum $\Delta G(\alpha_1, \alpha_2)$ values in each of the four quadrants of the (α_1, α_2) reaction coordinate, and the corresponding major ring puckering conformation(s).

Compound	$\Delta G_{+,-}$ [1]	$\Delta G_{-,+}$ [1]	$\Delta G_{-,-}$ [1]	$\Delta G_{+,+}$ [1]	Major Pucker Conformation(s) [2]
αGlc	0	5.47 (0.04)	6.05 (0.04)	8.46 (0.05)	$^4C_1$
MeαGlc	0	6.83 (0.06)	7.06 (0.05)	9.39 (0.06)	$^4C_1$
βGlc	0	8.43 (0.19)	5.44 (0.04)	7.03 (0.01)	$^4C_1$
MeβGlc	0	8.27 (0.08)	5.38 (0.03)	6.91 (0.02)	$^4C_1$
αGlcNAc	0	5.01 (0.05)	6.11 (0.07)	7.19 (0.05)	$^4C_1$
MeαGlcNAc	0	6.19 (0.11)	7.08 (0.04)	8.20 (0.04)	$^4C_1$
βGlcNAc	0	4.95 (0.09)	4.01 (0.02)	6.60 (0.05)	$^4C_1$
MeβGlcNAc	0	4.73 (0.06)	2.83 (0.09)	6.85 (0.04)	$^4C_1 > {}^1S_5$
αGal	0	4.25 (0.06)	6.11 (0.04)	8.73 (0.06)	$^4C_1$
MeαGal	0	5.62 (0.01)	7.35 (0.04)	8.74 (0.03)	$^4C_1$
βGal	0	6.56 (0.09)	5.80 (0.01)	8.35 (0.04)	$^4C_1$
MeβGal	0	7.10 (0.06)	6.63 (0.04)	8.29 (0.04)	$^4C_1$
αGalNAc	0	3.09 (0.09)	7.00 (0.07)	7.72 (0.09)	$^4C_1$
MeαGalNAc	0	4.33 (0.06)	8.21 (0.05)	7.77 (0.06)	$^4C_1$
βGalNAc	0	2.47 (0.05)	3.66 (0.07)	7.03 (0.04)	$^4C_1 > {}^1C_4$
MeβGalNAc	0	2.90 (0.04)	3.58 (0.01)	6.87 (0.05)	$^4C_1 > {}^1C_4$
αMan	0	5.26 (0.6)	6.83 (0.02)	9.99 (0.06)	$^4C_1$
MeαMan	0	5.82 (0.03)	7.54 (0.06)	10.74 (0.03)	$^4C_1$
βMan	0	6.89 (0.05)	7.27 (0.03)	8.98 (0.04)	$^4C_1$
MeβMan	0	6.20 (0.05)	6.24 (0.05)	8.14 (0.01)	$^4C_1$
αXyl	0	2.17 (0.01)	6.03 (0.02)	6.00 (0.05)	$^4C_1 > {}^1C_4$
MeαXyl	0	3.60 (0.02)	6.95 (0.02)	6.73 (0.02)	$^4C_1$
βXyl	0	3.77 (0.05)	5.25 (0.01)	3.24 0.01)	$^4C_1$
MeβXyl	0	4.19 (0.00)	4.91 (0.02)	3.51 (0.01)	$^4C_1$
αFuc	3.87 (0.06)	0	8.11 (0.06)	5.97 (0.02)	$^1C_4$
MeαFuc	5.14 (0.03)	0	8.15 (0.01)	7.10 (0.02)	$^1C_4$
βFuc	6.48 (0.03)	0	8.10 (0.03)	5.73 (0.01)	$^1C_4$
MeβFuc	7.01 (0.04)	0	7.86 (0.02)	6.54 (0.04)	$^1C_4$

Table 2. *Cont.*

Compound	$\Delta G_{+,-}$ [1]	$\Delta G_{-,+}$ [1]	$\Delta G_{-,-}$ [1]	$\Delta G_{+,+}$ [1]	Major Pucker Conformation(s) [2]
αNeu5Ac	2.71 (0.09)	0	2.79 (0.02)	1.42 (0.07)	$^{2}C_{5} > {}^{3}S_{O} > {}^{5}C_{2} \cong {}^{4,O}B$
MeαNeu5Ac	4.89 (0.29)	0	6.37 (0.10)	2.71 (0.24)	$^{2}C_{5} > {}^{3}S_{O}$
βNeu5Ac	6.79 (0.09)	0	7.18 (0.10)	4.01 (0.03)	$^{1}C_{4}$
MeβNeu5Ac	8.76 (0.08)	0	8.88 (0.10)	5.93 (0.11)	$^{1}C_{4}$
αGlcA	0	4.53 (0.12)	5.64 (0.04)	6.28 (0.05)	$^{4}C_{1}$
MeαGlcA	0	5.80 (0.04)	6.75 (0.03)	7.20 (0.03)	$^{4}C_{1}$
βGlcA	0	5.96 (0.11)	5.69 (0.01)	4.31 (0.06)	$^{4}C_{1}$
MeβGlcA	0	8.30 (0.09)	6.49 (0.02)	6.22 (0.04)	$^{4}C_{1}$
αIdoA	0	0.31 (0.09)	3.84 (0.04)	1.74 (0.02)	$^{4}C_{1} \cong {}^{1}C_{4} > {}^{2}S_{O}$
MeαIdoA	0	0.77 (0.05)	3.23 (0.01)	2.04 (0.02)	$^{4}C_{1} > {}^{1}C_{4} > {}^{2}S_{O}$
βIdoA	2.29 (0.03)	0	4.47 (0.08)	3.81 (0.03)	$^{1}C_{4} > {}^{4}C_{1}$
MeβIdoA	3.29 (0.06)	0	4.31 (0.05)	3.53 (0.06)	$^{1}C_{4}$
αIdo	0.73 (0.08)	0	0.88 (0.08)	3.20 (0.06)	$^{1}C_{4} > {}^{4}C_{1} \cong {}^{O}S_{2}$
MeαIdo	0.82 (0.04)	0	1.00 (0.02)	2.81 (0.03)	$^{1}C_{4} > {}^{4}C_{1} \cong {}^{O}S_{2} > {}^{3}S_{1}$
βIdo	0	2.15 (0.10)	4.17 (0.08)	5.30 (0.09)	$^{4}C_{1} > {}^{1}C_{4}$
MeβIdo	0	1.12 (0.09)	3.49 (0.04)	5.00 (0.09)	$^{4}C_{1} > {}^{1}C_{4}$
THP	0	0.03 (0.01)	5.14 (0.01)	5.13 (0.00)	$^{4}C_{1} = {}^{1}C_{4}$

[1] Data in kcal/mol are averages from triplicate simulations for the minimum $\Delta G(\alpha_1, \alpha_2)$ in that quadrant. For example, "−,+" indicates the quadrant defined by ($\alpha_1 < 0°$, $\alpha_2 > 0°$). Standard error of the mean values are in parentheses. [2] Conformations are listed in order of most likely to least likely. Only conformations corresponding to $\Delta G_{+,-}$, $\Delta G_{-,+}$, $\Delta G_{+,+}$, and/or $\Delta G_{-,-} < 3$ kcal/mol are listed.

As expected, most of the pyranose monosaccharides have a single major pucker conformation: the $^{4}C_{1}$ or the $^{1}C_{4}$ chair. Aside from IdoA and Ido, which were discussed in the previous section, exceptions to this are MeβGlcNAc, βGalNAc, MeβGalNAc, αXyl, αNeu5Ac, and MeαNeu5Ac.

MeβGlcNAc, βGalNAc, and MeβGalNAc all have their $\Delta G(\alpha_1, \alpha_2)$ global minimum corresponding to the $^{4}C_{1}$ chair conformation, as expected. They each also have a secondary minimum, but in all three cases the associated $\Delta G(\alpha_1, \alpha_2)$ is no less than 2.5 kcal/mol, which translates to a probability of no more than 1.5%. For MeβGlcNAc, the secondary minimum arises from skew-boat puckering, whereas for βGalNAc and MeβGalNAc, the secondary minimum is the $^{1}C_{4}$ chair conformation.

αXyl has the $^{4}C_{1}$ chair conformation as its global minimum and a secondary $\Delta G(\alpha_1, \alpha_2)$ minimum corresponding to the $^{1}C_{4}$ chair and with a value of 2.17 kcal/mol. This compares favorably to the value of 1.65 kcal/mol computed with the GROMOS 56a6CARBO force field (Table 1 in [48]), which is also exactly the value from Angyal's scheme for determining ring puckering free energies [70]. We note that data from Angyal's scheme have been used for quantitative comparison in other force field evaluations for a wide variety of pyranoses. It is worth emphasizing here that the Angyal data, though based in experiment, are indirect and were deemed by Angyal to be "calculated interaction energies". Concerning his "calculated interaction energies", Angyal writes, "an approximate calculation serves as a useful guide and can be readily carried out by adding the values of all of the non-bonded interactions occurring in each conformer, plus the value of the anomeric effect [70]". That is, those Angyal data for the $^{4}C_{1}$ to $^{1}C_{4}$ equilibrium in D-aldopyranoses listed in Table 1 of [70] are calculated as a simple sum of experimental values from model compounds, in

contrast to being directly measured for each monosaccharide, for example, through NMR experiments [46].

Neu5Ac is discussed at length in Appendix A, below. Part of that discussion involves comparison to structures from PDB crystal structures. On the one hand, all simulation data here are for isolated Neu5Ac monosaccharides in liquid water, whereas the PDB data are from crystal environments and typically involve Neu5Ac having non-covalent interactions with other biomolecules and/or being covalently attached to other monosaccharides. On the other hand, there is substantial congruence between the aqueous simulation data and the experimental crystal data for Neu5Ac (Figure A1b,d,f,h in Appendix A). Indeed, a computational study of Neu5Ac ring puckering in vacuum and in explicit water noted that the structure of Neu5Ac bound in influenza neuraminidase belonged to conformations preferentially sampled in the aqueous simulations [71]. And, an analysis of high-resolution PDB data for MeβGlcNAc noted that while nearly 97% of structures in the data set were in the $^4C_1$ chair conformation, 2.6% were boats or skew boats [72], which correlates closely with data from the present work. Therefore, in addition to NMR data from directly analogous experimental systems of monosaccharides in liquid water, PDB data may be useful as benchmarks for the type of force field-based simulations described here.

On a final note, control eABF simulations for THP yield a $\Delta G(\alpha_1, \alpha_2)$ plot that is symmetric about both $\alpha_1 = \alpha_2$ and about $\alpha_1 = -\alpha_2$ (Supplementary Material Figure S12), as expected. There are two equivalent global minima at $^4C_1 = {}^1C_4$, and boat/skew-boat conformations are over 5 kcal/mol higher in free energy. Thus, the exocyclic functional groups in the pyranose monosaccharides considered here can be thought of as introducing two types of perturbations to the THP $\Delta G(\alpha_1, \alpha_2)$: breaking of the symmetry, and altering the balance of chair vs. boat/skew-boat energetics.

3. Materials and Methods

3.1. Force Field

All systems were modeled using the CHARMM all-atom additive force field for carbohydrates [36,38] and the CHARMM-modified TIP3P water parameters [73,74] as contained in the "jul20" release of the CHARMM force field available as "toppar_c36_jul20.tgz" from http://mackerell.umaryland.edu/charmm_ff.shtml (accessed on 3 March 2021). Systems with a carboxylate functional group additionally used sodium ion parameters [75,76] included in the same release. During the course of the present work, we discovered a set of typos in the jul20 parameter file that affect Neu5Ac puckering energetics. Full details are provided in Appendix A. The data presented in this manuscript and the associated Supplementary Material reflect the correct parameters as developed in [38].

3.2. System Construction

Solvated systems were constructed for each monosaccharide in Figure 2 using either the α or the β anomer or one of the corresponding O-methyl glycosides, resulting in four unique systems for each monosaccharide. Monosaccharide coordinates were constructed from default force field internal geometries. The solvent consisted of a cubic box of water molecules at the experimental density of water and having an edge length of the longest dimension of the monosaccharide plus 30 Å; water molecules within 3 Å of the monosaccharide were deleted. In systems with a carboxylate group, a single sodium ion replaced a water molecule randomly selected and at least 6 Å from the monosaccharide. All system construction was carried out using the CHARMM program, v. c45b1 [77]. A single system containing tetrahydropyran (THP) was similarly constructed.

3.3. Molecular Dynamics Simulations

Each system was simulated in triplicate under periodic boundary conditions. Each replicate within the triplicate was assigned random initial velocities using a unique random seed to generate a unique trajectory. Simulations were carried out using the NAMD software, v. 2.13 [78]. Electrostatic and Lennard-Jones interactions employed a 10-Å spherical

cutoff. Lennard-Jones interaction energies were smoothly switched to zero in the interval 8–10 Å [79], an isotropic correction was applied for Lennard-Jones interactions beyond the cutoff [80], and the particle-mesh Ewald method with a 1 Å grid spacing accounted for electrostatic interactions beyond the cutoff [81]. After 1000 steps of energy minimization, each system was heated through re-initializing velocities to the target temperature of 298 K every 1000 molecular dynamics steps across 20,000 total steps with an integration timestep of 0.5 fs and positional restraints on solute non-hydrogen atoms. The SHAKE [82] and SETTLE algorithms [83] were respectively used to constrain all bonds involving hydrogen atoms and water geometries to their equilibrium values, and a temperature of 298 K and a pressure of 1 atm were maintained using Langevin thermostatting [84] and Nosé-Hoover Langevin barostatting [85,86]. Following heating, positional restraints were removed and data were collected from 200-ns production simulations (100×10^6 timesteps with an integration timestep of 2.0 fs).

The Extended-System Adaptive Biasing Force (eABF) methodology [43,44] was used to determine the free energy of pyranose ring puckering, $\Delta G(\alpha_1, \alpha_2)$, using reaction coordinates proposed by Babin and Sagui [64], where α_1 is the dihedral angle defined by the atoms O5-C1-C2-C3 and α_2 is the dihedral defined by the atoms C3-C4-C5-O5, except in the case of sialic acid in which these dihedrals are defined by O6-C2-C3-C4 and C4-C5-C6-O6, respectively. $\Delta G(\alpha_1, \alpha_2)$ was computed from the CZAR gradient estimate [43] using a Poisson equation formalism [87] implemented within NAMD via the Colvars software module [88]. eABF parameters included a fictitious particle spring constant of $k_B T$/degree/degree and sampling with a $1^\circ \times 1^\circ$ bin size and restrained with half-harmonic potentials to the range $-75^\circ < \alpha_{1,2} < 75^\circ$. Application of the biasing force in a given bin was scaled by 0 for the first 100 samples and then linearly scaled from 0 to 100% between 100 and 200 samples. Non-biased control simulations followed the same protocol but with no eABF sampling.

Additional CMAP-biased simulations were carried out for iduronate and for idose by applying a fixed bias equal to $-0.5 \times \Delta G(\alpha_1, \alpha_2)$ through the CHARMM force field CMAP term [69]. The representation of this bias using CMAP is not exact relative to the reference values computed by eABF simulation, as CMAP uses a square grid with 15° intervals between grid points and bicubic interpolation approximate $-0.5 \times \Delta G(\alpha_1, \alpha_2)$ for off-grid values of (α_1, α_2). CMAP-biased simulations were run using the OpenMM software, v. 7.5.1 [89] and a molecular dynamics protocol similar to that used for non-biased control NAMD simulations.

3.4. Molecular Dynamics Trajectory Analysis

Molecular dynamics trajectories were analyzed with the CHARMM software, including for the computation of Cremer-Pople ring puckering parameters [55]. VMD [90] was used for visualization and the creation of molecular graphics.

3.5. Definition of $^4C_1$, $^1C_4$, ^2S_O, $^OS_2$, and Other Ring Puckering Conformations

C-P parameters (θ, ϕ) were used to define ring puckering conformations as follows (note: analogous puckers for Neu5Ac compounds have all superscripted/subscripted numbers in puckering conformations incremented by 1 to reflect the different atom numbering in Neu5Ac, as shown in Figure 2):

- $^4C_1$: $0^\circ \leq \theta < 30^\circ$, ϕ = any
- Southern tropical: $30^\circ \leq \theta < 60^\circ$, ϕ = any
- Equatorial: $60^\circ \leq \theta < 120^\circ$, with specific conformations defined by,
  - $^{3,O}B$: $0^\circ \leq \phi < 15^\circ$ or $345^\circ \leq \phi < 360^\circ$
  - $^3S_1$: $15^\circ \leq \phi < 45^\circ$
  - $B_{1,4}$: $45^\circ \leq \phi < 75^\circ$
  - $^5S_1$: $75^\circ \leq \phi < 105^\circ$
  - $^{2,5}B$: $105^\circ \leq \phi < 135^\circ$
  - ^2SO: $135^\circ \leq \phi < 165^\circ$

- $B_{3,O}$: $165° \leq \phi < 195°$
- $^{1}S_{3}$: $195° \leq \phi < 225°$
- $^{1,4}B$: $225° \leq \phi < 255°$
- $^{1}S_{5}$: $255° \leq \phi < 285°$
- $B_{2,5}$: $285° \leq \phi < 315°$
- $^{O}S_{2}$: $315° \leq \phi < 345°$

- Northern tropical: $120° \leq \theta < 150°$, ϕ = any
- $^{4}C_{1}$: $150° \leq \theta \leq 180°$, ϕ = any

4. Conclusions

The data presented here provide a thorough accounting of the ring puckering free energies for the ten common vertebrate monosaccharides and idose, as represented by the CHARMM force field. In addition to demonstrating that the CHARMM force field reliably models ring puckering across this set diverse of molecules, the results show that doing so is possible with a single set of self-consistent force-field parameters developed using a standardized force field parametrization protocol [36,38]. This, in combination with examples of CHARMM force field studies on glycosidic linkages [91–96], lends confidence to the application of these parameters in the modeling of carbohydrate-containing protein systems, such as glycoproteins and proteoglycans as well as transmembrane proteins in glycolipid-containing bilayers. Accurate simulations for these types of systems can help expand the frontiers of protein structural biology by bridging gaps in experimental approaches for characterizing carbohydrate-containing protein systems.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms23010473/s1.

Author Contributions: Conceptualization, O.G.; methodology, O.G. and D.M.; software, O.G. and D.M.; validation, O.G. and D.M.; formal analysis, O.G. and D.M.; resources, O.G.; data curation, O.G.; writing—original draft preparation, O.G.; writing—review and editing, O.G., D.M. and M.G.; visualization, O.G. and M.G.; supervision, O.G.; project administration, O.G.; funding acquisition, O.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation, grant number MCB-1453529 to O.G. The APC was funded by the University of New England School of Pharmacy.

Acknowledgments: O.G. gratefully acknowledges communications with Jérôme Hénin and Anthony Serianni.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

During the course of the present work, we discovered a set of typos in the jul20 parameter file that affect Neu5Ac puckering energetics. These typos affect only Neu5Ac in the present work and will be corrected in a future official update to the CHARMM force field (A. D. MacKerell, Jr., personal communication). For the time being, the jul20 "par_all36_carb.prm" CHARMM parameter file can be manually corrected by adding the following lines to that file and deleting all other lines that refer to these same parameters:

NC2D1	CC3161	CC3161	CC3261	0.20	3	0.0
CC312	CC3163	CC3161	NC2D1	0.20	3	0.0
OC3C61	CC3163	CC3161	NC2D1	0.20	3	0.0

The typos affect two dihedrals in the Neu5Ac ring, with the first parameter affecting rotation about the C4-C5 bond and the second two rotation about the C5-C6 bond. The above three lines revert the parameters to the original values in the publication describing

parametrization for Neu5Ac [38]. Figure A1 demonstrates the large qualitative difference between the eABF $\Delta G(\alpha_1, \alpha_2)$ results using the incorrect force field dihedral parameters resulting from the typos and the correct force field dihedral parameters that are the original values from that publication. In Figure A1, (α_1, α_2) values from all instances of Neu5Ac in the PDB are overlaid on top of the $\Delta G(\alpha_1, \alpha_2)$ contour plots, and clearly show the superiority of the correct force field parameters as judged by the overlap of the PDB data with the global minima in the $\Delta G(\alpha_1, \alpha_2)$ contour plots (data were extracted from the PDB on 30 July 2021 by searching with the SMILES string "CC(=O)NC1C(CC(OC1C(C(CO)O)O)(C(=O)O)O)O" and separating hits into either α anomers or β anomers, of which there were 439 and 52, respectively found across a total of 170 PDB entries). In the case of the incorrect parameters, there is poor overlap, while with the correct parameters there is excellent overlap.

For αNeu5Ac and MeαNeu5Ac simulated using the incorrect parameters (Figure A1a and Figure A1c, respectively), the global minimum is in a boat/skew-boat region of (α_1, α_2) space whereas the vast majority of crystallographic structures in the α anomeric form are in the $^2C_5$ chair pucker conformation. However, with the correct force field parameters, the global minimum is in the $^2C_5$ region of (α_1, α_2) space for both αNeu5Ac (Figure A1b) and MeαNeu5Ac (Figure A1d), and the small proportion of α anomeric crystallographic structures outside of this region are located in or near a secondary minimum with favorable free energy (i.e., < 3 kcal/mol).

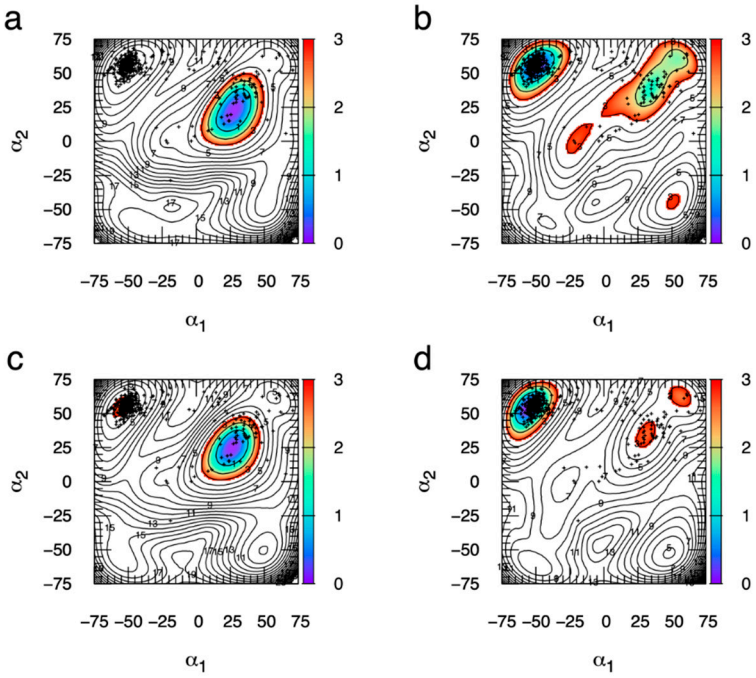

Figure A1. *Cont.*

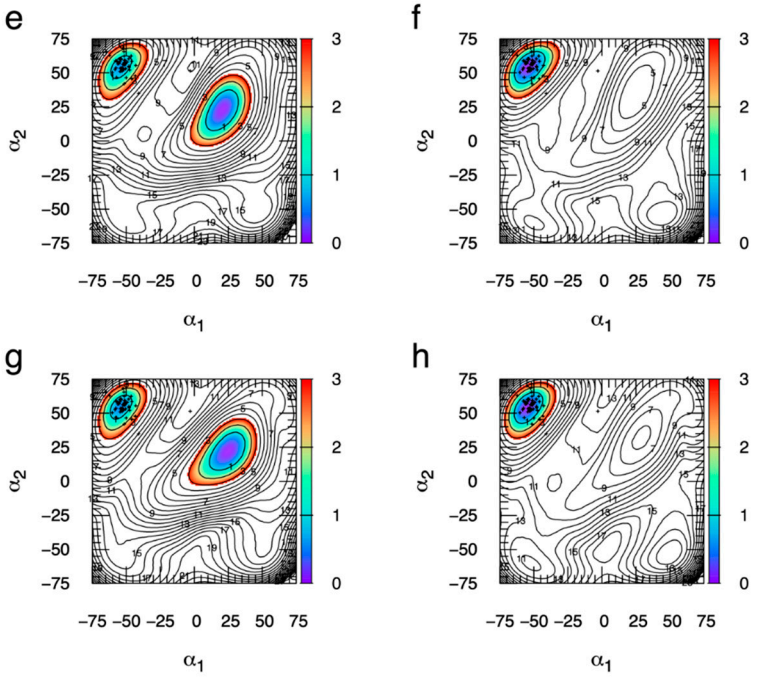

Figure A1. $\Delta G(\alpha_1, \alpha_2)$ from eABF simulations using incorrect versus correct dihedral force field parameters for Neu5Ac along with (α_1, α_2) data from all Neu5Ac structures in the PDB (searched 30 July 2021). $\Delta G(\alpha_1, \alpha_2)$ contour data are for αNeu5Ac with incorrect parameters (**a**), αNeu5Ac with correct parameters (**b**), MeαNeu5Ac with incorrect parameters (**c**), MeαNeu5Ac with correct parameters (**d**), βNeu5Ac with incorrect parameters (**e**), βNeu5Ac with correct parameters (**f**), MeβNeu5Ac with incorrect parameters (**g**), and MeβNeu5Ac with correct parameters (**h**). α_1 and α_2 are in degrees. $\Delta G(\alpha_1, \alpha_2)$ is from the first simulation in the triplicate and is in kcal/mol, with contours drawn every 1 kcal/mol and colored from 0–3 kcal/mol. PDB data were divided into two groups: those from α anomers and those from β anomers. Crystallographic data from the α anomers are displayed as small +'s in (**a**–**d**) and crystallographic data from the β anomers are displayed as small +'s in (**e**–**h**).

For the β anomers simulated using the incorrect parameters (βNeu5Ac (Figure A1e) and MeβNeu5Ac (Figure A1g)), there are no crystallographic Neu5Ac structures in the β anomeric form that coincide with the global minimum. In contrast, with the correct parameters, nearly all of these crystallographic structures in the β anomeric form, which are in the $^2C_5$ chair pucker conformation, coincide with the global $\Delta G(\alpha_1, \alpha_2)$ minimum for both βNeu5Ac (Figure A1f) and MeβNeu5Ac (Figure A1h).

References

1. Schnaar, R.L.; Kinoshita, T. Glycosphingolipids. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015; pp. 125–135. [CrossRef]
2. Seeberger, P.H. Monosaccharide Diversity. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2017; pp. 19–30. [CrossRef]
3. Stanley, P.; Taniguchi, N.; Aebi, M. N-Glycans. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015; pp. 99–111. [CrossRef]
4. Brockhausen, I.; Stanley, P. O-GalNAc Glycans. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015; pp. 113–123. [CrossRef]
5. Lindahl, U.; Couchman, J.; Kimata, K.; Esko, J.D. Proteoglycans and Sulfated Glycosaminoglycans. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015; pp. 207–221. [CrossRef]

6. Varki, A.; Kornfeld, S. Historical Background and Overview. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015; pp. 1–18. [CrossRef]
7. Imberty, A.; Prestegard, J.H. Structural Biology of Glycan Recognition. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015; pp. 387–400. [CrossRef]
8. Kumar, A.; Narayanan, V.; Sekhar, A. Characterizing Post-Translational Modifications and Their Effects on Protein Conformation Using NMR Spectroscopy. *Biochemistry* **2020**, *59*, 57–73. [CrossRef] [PubMed]
9. Xin, F.; Radivojac, P. Post-translational modifications induce significant yet not extreme changes to protein structure. *Bioinformatics* **2012**, *28*, 2905–2913. [CrossRef] [PubMed]
10. Craveur, P.; Narwani, T.J.; Rebehmed, J.; de Brevern, A.G. Investigation of the impact of PTMs on the protein backbone conformation. *Amino Acids* **2019**, *51*, 1065–1079. [CrossRef]
11. Kermani, A.A. A guide to membrane protein X-ray crystallography. *FEBS J.* **2021**, *288*, 5788–5804. [CrossRef]
12. Woods, R.J. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem. Rev.* **2018**, *118*, 8005–8024. [CrossRef] [PubMed]
13. Samsonov, S.A.; Pisabarro, M.T. Computational analysis of interactions in structurally available protein-glycosaminoglycan complexes. *Glycobiology* **2016**, *26*, 850–861. [CrossRef] [PubMed]
14. Whitmore, E.K.; Vesenka, G.; Sihler, H.; Guvench, O. Efficient Construction of Atomic-Resolution Models of Non-Sulfated Chondroitin Glycosaminoglycan Using Molecular Dynamics Data. *Biomolecules* **2020**, *10*, 537. [CrossRef]
15. Whitmore, E.K.; Martin, D.; Guvench, O. Constructing 3-Dimensional Atomic-Resolution Models of Nonsulfated Glycosaminoglycans with Arbitrary Lengths Using Conformations from Molecular Dynamics. *Int. J. Mol. Sci.* **2020**, *21*, 7699. [CrossRef]
16. Bohne-Lang, A.; von der Lieth, C.-W. GlyProt: In silico glycosylation of proteins. *Nucleic Acids Res.* **2005**, *33*, W214–W219. [CrossRef]
17. Singh, A.; Montgomery, D.; Xue, X.; Foley, B.L.; Woods, R.J. GAG Builder: A web-tool for modeling 3D structures of glycosaminoglycans. *Glycobiology* **2019**, *29*, 515–518. [CrossRef]
18. Engelsen, S.B.; Hansen, P.I.; Perez, S. POLYS 2.0: An open source software package for building three-dimensional structures of polysaccharides. *Biopolymers* **2014**, *101*, 733–743. [CrossRef] [PubMed]
19. Kuttel, M.M.; Ståhle, J.; Widmalm, G. CarbBuilder: Software for building molecular models of complex oligo- and polysaccharide structures. *J. Comput. Chem.* **2016**, *37*, 2098–2105. [CrossRef]
20. Clerc, O.; Deniaud, M.; Vallet, S.D.; Naba, A.; Rivet, A.; Perez, S.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB: Integration of new data with a focus on glycosaminoglycan interactions. *Nucleic Acids Res.* **2019**, *47*, D376–D381. [CrossRef]
21. Clerc, O.; Mariethoz, J.; Rivet, A.; Lisacek, F.; Pérez, S.; Ricard-Blum, S. A pipeline to translate glycosaminoglycan sequences into 3D models. Application to the exploration of glycosaminoglycan conformational space. *Glycobiology* **2019**, *29*, 36–44. [CrossRef] [PubMed]
22. Park, S.-J.; Lee, J.; Qi, Y.; Kern, N.R.; Lee, H.S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **2019**, *29*, 320–331. [CrossRef]
23. Almond, A. Multiscale modeling of glycosaminoglycan structure and dynamics: Current methods and challenges. *Curr. Opin. Struct. Biol.* **2018**, *50*, 58–64. [CrossRef]
24. Sattelle, B.M.; Shakeri, J.; Cliff, M.J.; Almond, A. Proteoglycans and their heterogeneous glycosaminoglycans at the atomic scale. *Biomacromolecules* **2015**, *16*, 951–961. [CrossRef]
25. Frank, M. Conformational analysis of oligosaccharides and polysaccharides using molecular dynamics simulations. *Methods Mol. Biol.* **2015**, *1273*, 359–377. [CrossRef] [PubMed]
26. Widmalm, G. A perspective on the primary and three-dimensional structures of carbohydrates. *Carbohydr. Res.* **2013**, *378*, 123–132. [CrossRef]
27. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **2008**, *29*, 622–655. [CrossRef]
28. Singh, A.; Tessier, M.B.; Pederson, K.; Wang, X.; Venot, A.P.; Boons, G.-J.; Prestegard, J.H.; Woods, R.J. Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. *Can. J. Chem.* **2016**, *94*, 927–935. [CrossRef]
29. Pol-Fachin, L.; Rusu, V.H.; Verli, H.; Lins, R.D. GROMOS 53A6GLYC, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theory Comput.* **2012**, *8*, 4681–4690. [CrossRef] [PubMed]
30. Pol-Fachin, L.; Verli, H.; Lins, R.D. Extension and validation of the GROMOS 53A6$_{GLYC}$ parameter set for glycoproteins. *J. Comput. Chem.* **2014**, *35*, 2087–2095. [CrossRef]
31. Hansen, H.S.; Hünenberger, P.H. A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers. *J. Comput. Chem.* **2011**, *32*, 998–1032. [CrossRef]
32. Plazinski, W.; Lonardi, A.; Hünenberger, P.H. Revision of the GROMOS 56A6$_{CARBO}$ force field: Improving the description of ring-conformational equilibria in hexopyranose-based carbohydrates chains. *J. Comput. Chem.* **2016**, *37*, 354–365. [CrossRef] [PubMed]

33. Panczyk, K.; Gaweda, K.; Drach, M.; Plazinski, W. Extension of the GROMOS 56a6$_{CARBO/CARBO_R}$ Force Field for Charged, Protonated, and Esterified Uronates. *J. Phys. Chem. B* **2018**, *122*, 3696–3710. [CrossRef]
34. Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W.L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **1997**, *18*, 1955–1970. [CrossRef]
35. Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W.F. An improved OPLS-AA force field for carbohydrates. *J. Comput. Chem.* **2002**, *23*, 1416–1429. [CrossRef] [PubMed]
36. Guvench, O.; Greene, S.N.; Kamath, G.; Brady, J.W.; Venable, R.M.; Pastor, R.W.; Mackerell, A.D., Jr. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543–2564. [CrossRef]
37. Guvench, O.; Hatcher, E.R.; Venable, R.M.; Pastor, R.W.; MacKerell, A.D., Jr. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353–2370. [CrossRef]
38. Guvench, O.; Mallajosyula, S.S.; Raman, E.P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T.J.; Jamison, F.W., II; Mackerell, A.D., Jr. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput.* **2011**, *7*, 3162–3180. [CrossRef] [PubMed]
39. Mallajosyula, S.S.; Guvench, O.; Hatcher, E.; MacKerell, A.D., Jr. CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *J. Chem. Theory Comput.* **2012**, *8*, 759–776. [CrossRef]
40. Sattelle, B.M.; Shakeri, J.; Almond, A. Does Microsecond Sugar Ring Flexing Encode 3D-Shape and Bioactivity in the Heparanome? *Biomacromolecules* **2013**, *14*, 1149–1159. [CrossRef]
41. Sattelle, B.M.; Hansen, S.U.; Gardiner, J.; Almond, A. Free energy landscapes of iduronic acid and related monosaccharides. *J. Am. Chem. Soc.* **2010**, *132*, 13132–13134. [CrossRef] [PubMed]
42. Sattelle, B.M.; Bose-Basu, B.; Tessier, M.; Woods, R.J.; Serianni, A.S.; Almond, A. Dependence of pyranose ring puckering on anomeric configuration: Methyl idopyranosides. *J. Phys. Chem. B* **2012**, *116*, 6380–6386. [CrossRef]
43. Lesage, A.; Lelièvre, T.; Stoltz, G.; Hénin, J. Smoothed Biasing Forces Yield Unbiased Free Energies with the Extended-System Adaptive Biasing Force Method. *J. Phys. Chem. B* **2017**, *121*, 3676–3685. [CrossRef] [PubMed]
44. Fu, H.; Shao, X.; Chipot, C.; Cai, W. Extended Adaptive Biasing Force Algorithm. An On-the-Fly Implementation for Accurate Free-Energy Calculations. *J. Chem. Theory Comput.* **2016**, *12*, 3506–3513. [CrossRef]
45. Plazinski, W.; Plazinska, A. Molecular dynamics simulations of hexopyranose ring distortion in different force fields. *Pure Appl. Chemistry. Chim. Pure Appl.* **2017**, *89*, 1283–1294. [CrossRef]
46. Bose-Basu, B.; Zhang, W.; Kennedy, J.L.W.; Hadad, M.J.; Carmichael, I.; Serianni, A.S. ^{13}C-Labeled Idohexopyranosyl Rings: Effects of Methyl Glycosidation and C6 Oxidation on Ring Conformational Equilibria. *J. Org. Chem.* **2017**, *82*, 1356–1370. [CrossRef]
47. Lins, R.D.; Hünenberger, P.H. A new GROMOS force field for hexopyranose-based carbohydrates. *J. Comput. Chem.* **2005**, *26*, 1400–1412. [CrossRef]
48. Panczyk, K.; Plazinski, W. Pyranose ring puckering in aldopentoses, ketohexoses and deoxyaldohexoses. A molecular dynamics study. *Carbohydr. Res.* **2018**, *455*, 62–70. [CrossRef]
49. Guvench, O.; MacKerell, A.D., Jr. Automated conformational energy fitting for force-field development. *J. Mol. Model.* **2008**, *14*, 667–679. [CrossRef]
50. 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.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616. [CrossRef]
51. MacKerell, A.D., Jr.; Feig, M.; Brooks, C.L., III. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–699. [CrossRef] [PubMed]
52. Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.; Mittal, J.; Feig, M.; Mackerell, A.D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273. [CrossRef]
53. Klauda, J.B.; Venable, R.M.; Freites, J.A.; O'Connor, J.W.; Tobias, D.J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A.D., Jr.; Pastor, R.W. Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843. [CrossRef] [PubMed]
54. Klauda, J.B.; Monje, V.; Kim, T.; Im, W. Improving the CHARMM Force Field for Polyunsaturated Fatty Acid Chains. *J. Phys. Chem. B* **2012**, *116*, 9424–9431. [CrossRef]
55. Cremer, D.; Pople, J.A. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358. [CrossRef]
56. Dowd, M.K.; French, A.D.; Reilly, P.J. Modeling of aldopyranosyl ring puckering with MM3 (92). *Carbohydr. Res.* **1994**, *264*, 1–19. [CrossRef]
57. Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603. [CrossRef]
58. Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826–843. [CrossRef]
59. Autieri, E.; Sega, M.; Pederiva, F.; Guella, G. Puckering free energy of pyranoses: A NMR and metadynamics-umbrella sampling investigation. *J. Chem. Phys.* **2010**, *133*, 095104. [CrossRef]
60. Pickett, H.M.; Strauss, H.L. Conformational structure, energy, and inversion rates of cyclohexane and some related oxanes. *J. Am. Chem. Soc.* **1970**, *92*, 7281–7290. [CrossRef]

61. Hansen, H.S.; Hünenberger, P.H. Using the local elevation method to construct optimized umbrella sampling potentials: Calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water. *J. Comput. Chem.* **2010**, *31*, 1–23. [CrossRef]
62. Boeyens, J.C.A.; Evans, D.G. Group theory of ring pucker. *Acta Crystallogr. Sect. B* **1989**, *45*, 577–581. [CrossRef]
63. Sega, M.; Autieri, E.; Pederiva, F. Pickett angles and Cremer–Pople coordinates as collective variables for the enhanced sampling of six-membered ring conformations. *Mol. Phys.* **2011**, *109*, 141–148. [CrossRef]
64. Babin, V.; Sagui, C. Conformational free energies of methyl-alpha-L-iduronic and methyl-beta-D-glucuronic acids in water. *J. Chem. Phys.* **2010**, *132*, 104108. [CrossRef]
65. Alibay, I.; Bryce, R.A. Ring Puckering Landscapes of Glycosaminoglycan-Related Monosaccharides from Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2019**, *59*, 4729–4741. [CrossRef] [PubMed]
66. Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008**, *128*, 144120. [CrossRef]
67. Hénin, J.; Chipot, C. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* **2004**, *121*, 2904–2914. [CrossRef]
68. Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M.L. Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables. *J. Chem. Theory Comput.* **2010**, *6*, 35–47. [CrossRef]
69. MacKerell, A.D., Jr.; Feig, M.; Brooks, C.L., III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415. [CrossRef] [PubMed]
70. Angyal, S.J. The Composition and Conformation of Sugars in Solution. *Angew. Chem. Int. Ed. Engl.* **1969**, *8*, 157–166. [CrossRef]
71. Spiwok, V.; Tvaroška, I. Conformational free energy surface of alpha-N-acetylneuraminic acid: An interplay between hydrogen bonding and solvation. *J. Phys. Chem. B* **2009**, *113*, 9589–9594. [CrossRef]
72. Sattelle, B.M.; Almond, A. Is N-acetyl-D-glucosamine a rigid 4C1 chair? *Glycobiology* **2011**, *21*, 1651–1662. [CrossRef]
73. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
74. Durell, S.R.; Brooks, B.R.; Ben-Naim, A. Solvent-induced forces between two hydrophilic groups. *J. Phys. Chem.* **1994**, *98*, 2198–2202. [CrossRef]
75. Beglov, D.; Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050–9063. [CrossRef]
76. Venable, R.M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R.W. Simulations of anionic lipid membranes: Development of interaction-specific ion parameters and validation using NMR data. *J. Phys. Chem. B* **2013**, *117*, 10183–10192. [CrossRef] [PubMed]
77. Brooks, B.R.; Brooks, C.L., III; MacKerell, A.D., Jr.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [CrossRef]
78. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [CrossRef]
79. Steinbach, P.J.; Brooks, B.R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **1994**, *15*, 667–683. [CrossRef]
80. Shirts, M.R.; Mobley, D.L.; Chodera, J.D.; Pande, V.S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J. Phys. Chem. B* **2007**, *111*, 13052–13063. [CrossRef] [PubMed]
81. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An *N* log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]
82. Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]
83. Miyamoto, S.; Kollman, P.A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962. [CrossRef]
84. Kubo, R.; Toda, M.; Hashitume, N. *Statistical Physics II: Nonequilibrium Statistical Mechanics*, 2nd ed.; Springer: New York, NY, USA, 1991.
85. Martyna, G.J.; Tobias, D.J.; Klein, M.L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189. [CrossRef]
86. Feller, S.E.; Zhang, Y.H.; Pastor, R.W.; Brooks, B.R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621. [CrossRef]
87. Hénin, J. Fast and accurate multidimensional free energy integration. *J. Chem. Theory Comput.* **2021**, *17*, 6789–6798. [CrossRef] [PubMed]
88. Fiorin, G.; Klein, M.L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345–3362. [CrossRef]
89. Eastman, P.; Swails, J.; Chodera, J.D.; McGibbon, R.T.; Zhao, Y.; Beauchamp, K.A.; Wang, L.P.; Simmonett, A.C.; Harrigan, M.P.; Stern, C.D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659. [CrossRef] [PubMed]
90. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

91. Mallajosyula, S.S.; MacKerell, A.D. Influence of Solvent and Intramolecular Hydrogen Bonding on the Conformational Properties of O-Linked Glycopeptides. *J. Phys. Chem. B* **2011**, *115*, 11215–11229. [CrossRef]
92. Faller, C.E.; Guvench, O. Sulfation and cation effects on the conformational properties of the glycan backbone of chondroitin sulfate disaccharides. *J. Phys. Chem. B* **2015**, *119*, 6063–6073. [CrossRef] [PubMed]
93. Yang, M.; Angles d'Ortoli, T.; Säwén, E.; Jana, M.; Widmalm, G.; MacKerell, A.D. Delineating the conformational flexibility of trisaccharides from NMR spectroscopy experiments and computer simulations. *Phys. Chem. Chem. Phys.* **2016**, *18*, 18776–18794. [CrossRef]
94. Ng, C.; Premnath, P.N.; Guvench, O. Rigidity and flexibility in the tetrasaccharide linker of proteoglycans from atomic-resolution molecular simulation. *J. Comput. Chem.* **2017**, *38*, 1438–1446. [CrossRef] [PubMed]
95. Rönnols, J.; Engström, O.; Schnupf, U.; Säwén, E.; Brady, J.W.; Widmalm, G. Inter-residual Hydrogen Bonding in Carbohydrates Unraveled by NMR Spectroscopy and Molecular Dynamics Simulations. *Chem. Bio. Chem.* **2019**, *20*, 2519–2528. [CrossRef]
96. Lutsyk, V.; Plazinski, W. Conformational Properties of Glycosaminoglycan Disaccharides: A Molecular Dynamics Study. *J. Phys. Chem. B* **2021**, *125*, 10900–10916. [CrossRef]