JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

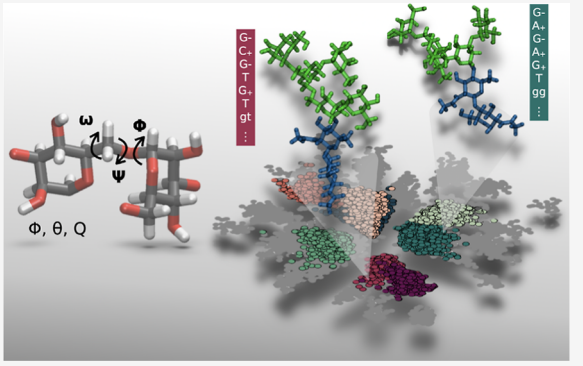
Exploration, Representation, and Rationalization of the Conformational Phase Space of N-Glycans

Isabell Louise Grothaus, Giovanni Bussi, and Lucio Colombi Ciacchi*

Cite This: J. Chem. Inf. Model. 2022, 62, 4992−5008 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Despite their fundamental biological relevance, structure−property relationships in *N*-glycans are fundamentally lacking, and their highly multidimensional compositional and conformational phase spaces remain largely unexplored. The torsional flexibility of the glycosidic linkages and the ring dynamics result in wide, rugged free-energy landscapes that are difficult to sample in molecular dynamics simulations. We show that a novel enhanced-sampling scheme combining replica exchange with solute and collective-variable tempering, enabling transitions over all relevant energy barriers, delivers converged distributions of solvated *N*-glycan conformers. Several dimensionality-reduction algorithms are compared and employed to generate conformational free-energy maps in two dimensions. Together with an originally developed conformation-based nomenclature scheme that uniquely identifies glycan conformers, our modeling procedure is applied to reveal the effect of chemical substitutions on the conformational ensemble of selected high-mannose-type and complex glycans. Moreover, the structure-prediction capabilities of two commonly used glycan force fields are assessed via the theoretical prediction of experimentally available nuclear magnetic resonance J-coupling constants. The results especially confirm the key role of ω and ψ torsion angles in discriminating between different conformational states and suggest an intriguing correlation between the torsional and ring-puckering degrees of freedom that may be biologically relevant.



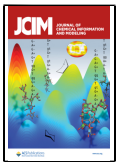
INTRODUCTION

Biologically relevant polysaccharides, or glycans, present the largest compositional and architectural variability among all other biomolecular classes and thus harbor a rich and complex coding capability, complementing that of polypeptides and polynucleotides.[1] In fact, they do not only provide mechanical support in biological systems (e.g., as cellulose or chitin) but also act as fine-tuned encoders of information in biomolecular recognition processes. The enzyme-regulated attachment of glycans to polypeptide chains via covalent tethering to the terminal NH_2 group of asparagine (*N*-glycosylation) is the most common post-translational modification. *N*-glycosylation broadens the range of functionality of the underlying protein (i) by regulating its folding, stability and function (ii) by providing target structures for carbohydrate-binding proteins (lectins) and specific antibodies and (iii) by mediating cell−matrix interactions as well as cell−cell recognition.[2,3] This emphasizes the need to identify and rationalize the rules underlying what has been named as the sugar code, namely the transmission of information mediated by glycans resulting in a specific biochemical function or signal.

Unraveling the sugar code is strongly impaired both by the lack of standard structural descriptors (such as α-helices and β-sheets in polypeptides) and by the high dynamical flexibility of glycan chains, reminiscent of the conformational variability of disordered peptides. Moreover, the non-linear, branched architecture of glycan chains and the variability of the type of linkages between the sugar monomers prevent the description of their structure in terms of few conformational variables, as done in proteins via the two-dimensional representation of all their torsional angle values in a Ramachandran plot.[4] As a first step toward a fundamental study of structure−property relationships in *N*-glycan systems, in this work we employ enhanced-sampling molecular dynamics (MD) simulations and advanced dimensionality-reduction techniques to explore the high-dimensional free-energy landscape of a set of biologically relevant *N*-glycan models.

Built up of monosaccharides linked via O-glycosidic linkages, glycans harbor two main structural degrees of freedom: torsion angles and ring distortions. The ϕ, ψ, and, in the case of 1−6 linkages, the ω torsion angles define the relative positions of

Received: August 17, 2022
Published: September 30, 2022



https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

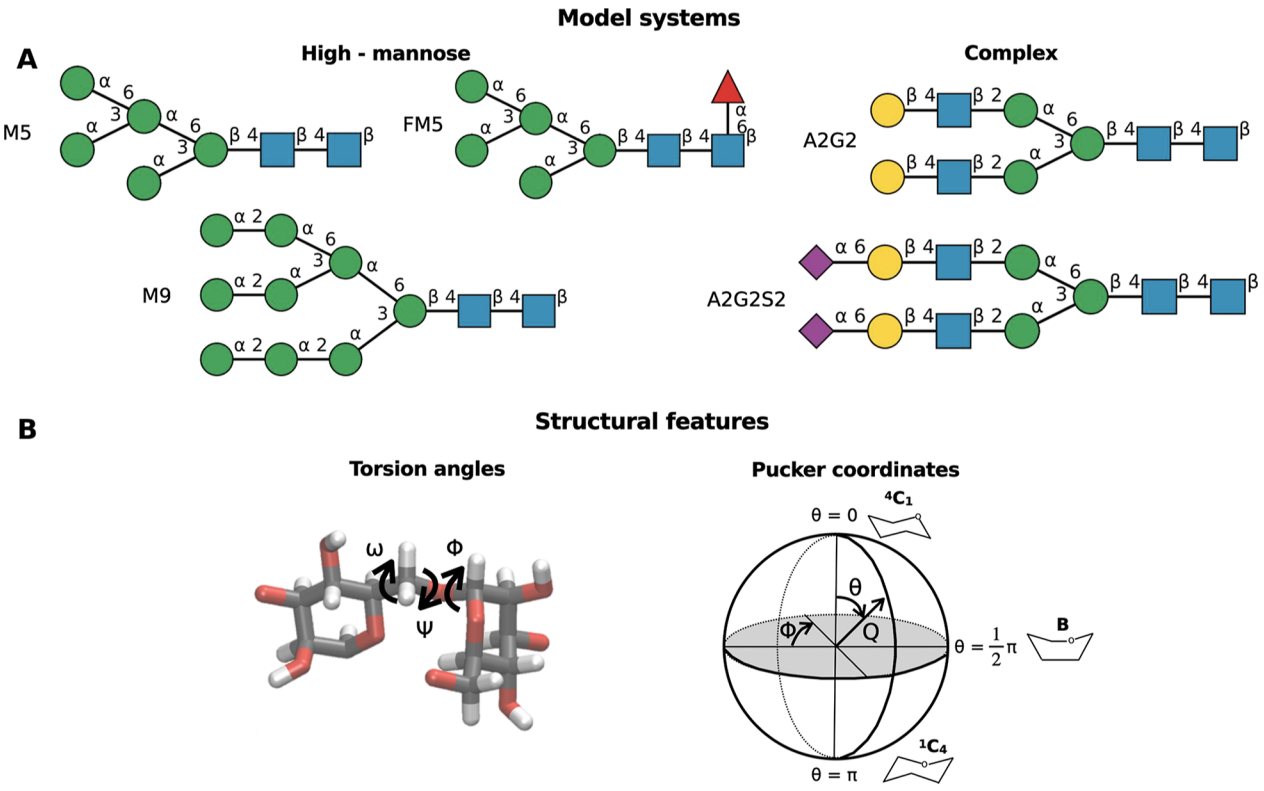


Figure 1. Top: model systems employed in this study, namely three high-mannose type *N*-glycans (M5, FM5, and M9) and two complex *N*-glycans (A2G2 and A2G2S2). Bottom: representative atomistic details are described for a torsion angle along a 1−6 O-glycosidic linkage as well as for the pucker coordinates according to the Cremer−Pople representation.[5] Monosaccharide symbols are in congruence with the nomenclature of the Consortium for functional glycomics,[6] using the Oxford notation for *N*-glycan names. The models were drawn using DrawGlycan.[7]

the individual saccharide monomers (Figure 1) within each possible conformer, which are stabilized by hydrogen bonds between the hydroxyl groups of the monomers.[8,9] The distortions, or puckering, of the six-membered saccharide rings are classified as chair (C), half-chair (H), enveloped (E), skew (S), or boat (B).[10] These distortions can be unambiguously mapped using the spherical pucker coordinates ϕ, θ, and Q, introduced by Cremer and Pople.[5] Information about the average structures adopted by *N*-glycans can be obtained by nuclear magnetic resonance (NMR) or, in some cases, by X-ray diffraction.[11] However, only atomic-scale simulations are, in principle, able to deliver full details of the probability distribution of all possible conformers in an *N*-glycan population and of their dynamical behavior.[12−14] In the case of classical MD, crucial in this regard are the accuracy of the employed force fields on one side[15,16] and the ergodicity of the used method on the other side. In particular, the slow transitions between different conformational (rotameric) states prevent efficient phase-space sampling and convergence of conformer distributions in plain MD simulations.[17,18]

For this reason, various enhanced-sampling MD techniques have been used to facilitate the crossing of the relevant energy barriers and accelerate the transition probabilities. These include replica-exchange MD (REMD),[19] Hamiltonian replica-exchange MD (H-REMD) with solute scaling (REST2),[20] well-tempered metadynamics (WT-MetaD),[21] Umbrella Sampling (US),[22] and their different combinations. On the one hand, methods based on bias potentials applied to specific collective variables (CVs), such as WT-MetaD, have so far focused on only few specific torsion angles (e.g., ω),[18] not giving justice to the structural complexity of *N*-glycans with multiple branches.[18,23−25] On the other hand, CV-independent methods such as REMD do not guarantee complete phase-space sampling[18,23,26] and require elaborate pre-calculations when used together with additional bias potentials.[24,27] To overcome these difficulties, in this study, we combine REST2 with the recently introduced replica exchange with collective-variable tempering (RECT) algorithm[28] (Figure 2). RECT enhances the transition probability between conformers separated by high-energy barriers by means of one-dimensional bias potentials applied to a large number of selected degrees of freedom, for example all torsion angles, in a MetaD framework. Simultaneously, REST2 samples all other degrees of freedom via a solute-scaling approach, which requires only a low number of replicas.

Due to the large number of CVs used to define the conformational states of an *N*-glycan, structural analyses require the rationalization of a high-dimensional vector space. Dimensionality-reduction techniques are thus required for an efficient graphical representation of the glycan conformational space and of its associated free-energy landscape. Such reduction should be performed in a way that best differentiates among the most probable conformers, possibly revealing mutual functional dependencies and hidden correlations among the many CVs. Diverse strategies have been developed in recent years to map a high-dimensional feature matrix, **X**, onto a low-dimensional latent-space matrix, **T**, and their applicability mainly depends on the underlying data structure.[29] For instance, principal component analysis (PCA)[30] projects the data onto the linear eigenvector space defined by the largest eigenvalues k obtained by diagonalization of the covariance matrix of **X**. Therefore, the top k

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

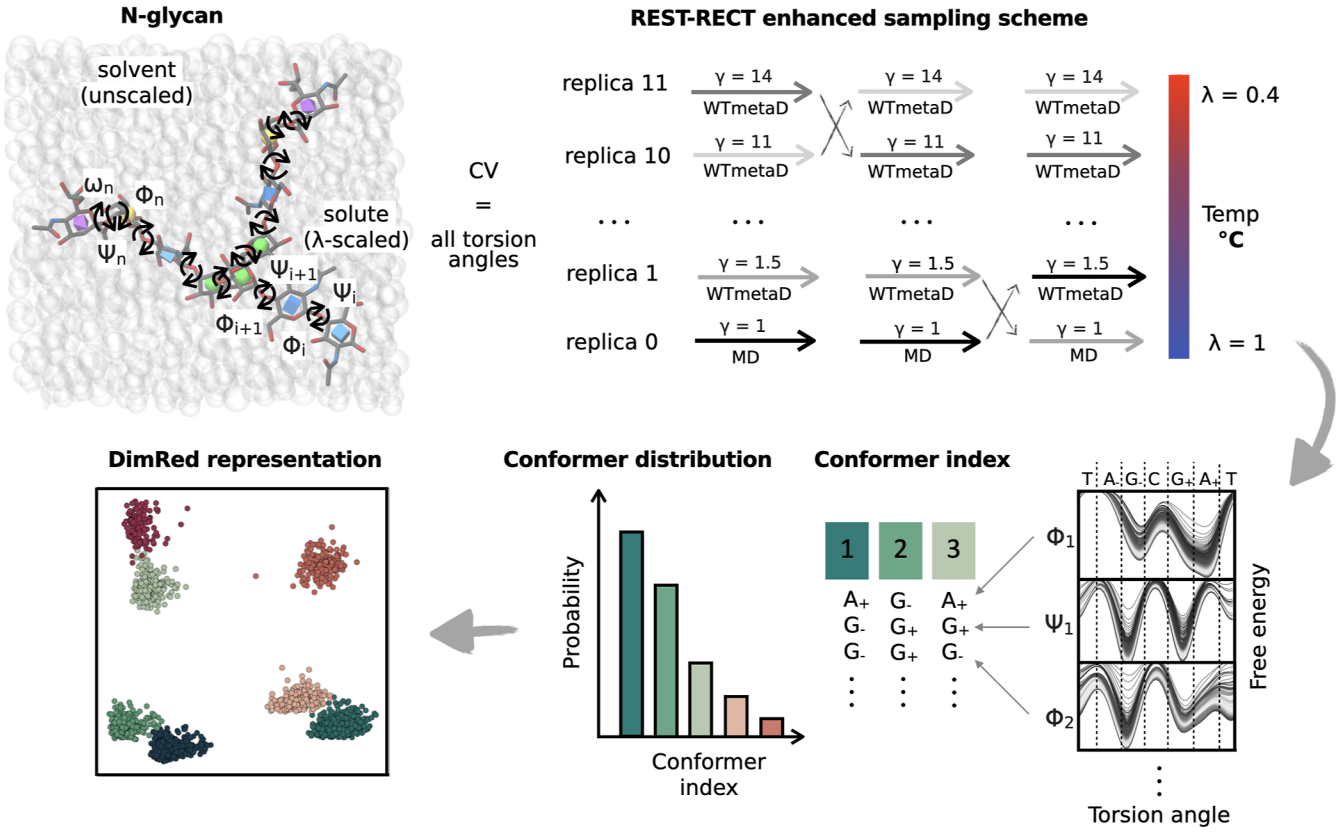


Figure 2. Schematic overview of the approach followed in this study. Free *N*-glycans in solution are simulated by employing the enhanced-sampling method REST-RECT to accelerate transitions over barriers for all torsion angles and pucker coordinates. Conformer strings are then constructed based on the free-energy landscape of each torsion angle. Individual conformers are grouped together according to these strings, and conformer distributions are constructed from simulated trajectories. Low-dimensional representations of the conformer clusters are finally generated using dimensionality-reduction methods.

eigenvectors (the principal components) represent the maximum data variance and are used as the axes of two-dimensional or three-dimensional graphs representing the entire data space. In contrast, diffusion map[31] is a non-linear method, where the connectivity (or diffusion distance) between individual data points (in our case, individual *N*-glycan conformers) is quantified by the likelihood of transitioning from one to the other, as expressed with the help of a diffusion kernel function. Data points of **X** are projected onto a two-dimensional matrix **T** in such a way that the diffusion distances in the high-dimensional feature space can be approximated by Euclidean distances between points in the reduced space, ensuring the preservation of the local vector-space structure (so-called isometric embedding). The non-linear sketch map algorithm[32,33] also ensures isometric embedding but is based on a different approach named multidimensional scaling.[34] The mutual distance of data points in **X** is conserved in **T** by the application of a sigmoid function focusing on the reproduction of intermediate distances rather than far-away distances, which are dominated by the topology of the high-dimensional space. Finally, besides these unsupervised techniques, there are also supervised algorithms, which make use of the information included in an additional property matrix **Y**. A recent example is the kernel principal covariates regression model (kPcovR),[35] which combines the robust methods of linear regression and PCA.

In this study, we investigate the ability of all these four dimensionality-reduction techniques to effectively measure the similarity among *N*-glycan conformers and partition them into meaningful subgroups (clusters) for further analysis. The proposed simulation and analysis scheme (Figure 2) is applied to different *N*-glycans (Figure 1) that are dominant in human plasma samples and therefore represent valuable model systems.[36] Our aim is to reveal how differences in the chemical composition of *N*-glycans impact the probability distribution of their conformers. In this way, the *N*-glycan heterogeneity can be better correlated with their biological functions. Additionally, we assess the performance of two force fields, namely CHARMM36[37] and GLYCAM06j,[38] in predicting the correct torsion angle and pucker distributions by comparison of our computational results with experimental NMR data.

METHODS

Simulation Setup. Five different *N*-glycans were investigated in this study, namely M5, FM5, M9, A2G2S2, and A2G2 (Figure 1). Their three-dimensional structures were constructed using the CHARMM-GUI glycan modeler based on averaged structures from the glycan fragment database.[39−42] The *N*-glycans were solvated with a 15 Å thick water layer in a cubic box. Two sodium counter-ions were added to compensate for the two net negative charges of A2G2S2.

MD Simulations. MD simulations were performed with the GROMACS code, version 2018.4,[43] patched with the PLUMED package, version 2.6.[44] Either the CHARMM36[37,45,46] or the GLYCAM06j[38] force field was used for the *N*-glycans molecules in combination with the

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

CHARMM-modified TIP3P water model (mTIP3P)[47] or the standard version (sTIP3P),[48] respectively. The GLYCAM06j force field parameters in a GROMACS format were obtained from the CHARMM-GUI glycan modeler while constructing the glycan structure and simulation box. This Amber force field parameter input generation for GROMACS has been available since version 3.6. In the case of the CHARMM36 force field, the recent correction to a previous faulty implementation, affecting, in particular, the ring-inversion properties of Neu5Ac, has been applied in all simulations.[49] The leap-frog algorithm was used as an integrator with a 2 fs time step, and the LINCS algorithm[50] was employed to constrain bonds connected to hydrogen atoms. Temperature control was realized via velocity rescaling[51] using a time constant of 0.1 ps, setting a reference temperature of 310.15 K. The pressure was set to 1 bar with a compressibility of 4.5×10^{-5} bar^{-1} and kept constant via the Parrinello-Rahman barostat with a time constant of 5 ps. The Verlet list scheme[52] was employed with a neighbor list updated every 80 steps. The calculation of electrostatic interactions was done with the particle mesh Ewald[53] method using a cut-off distance of 1.2 nm for the real-space contribution. The following steps were performed to equilibrate the systems. First, an energy minimization of water and ions was performed using the steepest-descent algorithm with a tolerance of 1000 kJ mol^{-1} nm^{-1}, restraining the *N*-glycan atoms. Then, the solvent was equilibrated in one *NVT* and one *NPT* run, each lasting 1 ns, with restrained *N*-glycans. After that, a second energy minimization of all atoms was performed with no constraints, with the same parameters as before. Finally, unrestrained *NVT* and *NPT* equilibration runs were performed, lasting 1 and 100 ns, respectively. For each simulated system, two different starting conformers, named s1 and s2, were generated by setting the ω angle of the **6**− branch (see below for the definition of the branch labeling) either in a gt (for s1) or a gg (for s2) conformation. Separate simulations starting with the two initial conformations s1 and s2 were performed to validate the convergence of the conformer distributions predicted by the different simulation techniques employed. For plain MD simulations, each production run lasted 6 μs, recording frames every 8 ps.

Enhanced-Sampling Simulations. Enhanced-sampling MD simulations were performed with a combination of the REST2 replica exchange method[20,54] and the RECT[28] method based on WT-MetaD[21] (henceforth referred to as REST-RECT). In each simulation, the whole *N*-glycan was defined as the REST2 solute, whose temperature was scaled in different replicas with index i by means of scaling factors λ_i acting on the long-range electrostatics, the Lennard−Jones interactions, as well as the dihedral angles. For non-neutral systems, a neutralizing background was automatically added via GROMACS. We used 12 replicas and a geometric progression of λ_i values equal to 1, 1, 0.92, 0.84, 0.77, 0.71, 0.65, 0.60, 0.55, 0.50, 0.46, and 0.42, spanning an effective temperature ladder from 310.15 to 800.00 K. Note that both the ground replica ($i = 0$) and the first replica ($i = 1$) were at the same ground temperature of $T_0 = 310.15$ K for convenience of the RECT implementation and analysis (see below). Water and ions were always kept at the ground temperature. Replica exchanges were attempted every 400 steps, following a Metropolis-Hastings acceptance criterion. In the RECT part, all n torsion angles, as listed below, of the simulated glycan were defined as CVs and biased simultaneously by n one-dimensional potentials in each replica i. n amounted to 14, 17, 24, 17, and 23 in the M5, FM5, M9, A2G2, and A2G2S2 *N*-glycans, respectively. Torsion angles were defined as ϕ = O5′−C1′−Ox−Cx, ψ = C1′−Ox−Cx−C(x−1), and ω = O6−C6−C5−O5, with x being the carbon number of the linkage at the non-reducing end. An exception is the 2−6 angles between Gal and Neu5Ac, which are defined as ϕ = O6′−C2′−O6−C6, ψ = C2′−O6−C6−C5, and ω = O6−C6−C5−O5. The replicas were biased with bias factors γ_i following a geometric progression of values equal to 1, 1.2, 1.46, 1.82, 2.3, 2.94, 3.78, 4.89, 6.34, 8.23, 10.7, and 14. Gaussian hills were deposited at time intervals of τ_G = 1 ps, with a width of 0.35 rad and a height corresponding to $h_i = (k_B \Delta T_i/\tau) \times \tau_G$, where k_B is the Boltzmann constant, $\Delta T_i = T_0(\gamma_i - 1)$ is the boosting temperature, and τ = 4 ps is the characteristic time for the bias evolution in the RECT part. The geometric progressions of λ_i and γ_i ensured sufficient overlaps of the potential energy distributions at all temperatures, resulting in uniform round trip times for different replicas (Figures S3 and S5). We note that the ground replica was fully unbiased ($\lambda_0 = 1$, $\gamma_0 = 1$), allowing for an unbiased statistical distribution of frames, which could be trivially used in all subsequent analyses. The first replica was biased but kept at the ground temperature ($\lambda_1 = 1$, $\gamma_1 = 1.2$) to ensure sufficient overlaps between the first two replicas. The inclusion of higher-order replicas in the analyses requires the application of the weighted-histogram analysis method.[55] However, including replicas up to $i = 4$ in the analyses did not result in a more effective sampling. Higher replicas should not be entered into the analyses because of the expected low ensemble overlap with the reference replica.

Conformer String Generation. *N*-glycans are multi-branched structures characterized by the specific linkages between saccharide monomers. Each glycosidic linkage gives rise to at least two torsion angles (ϕ and ψ), where 1−6 and 2−6 linkages harbor an additional torsion angle, ω. Based on these structural characteristics, we constructed an unambiguous labeling scheme to distinguish different conformers of the same *N*-glycan (see Figure 3 for example, related to M5 and M9 glycans). The scheme is also applicable to other glycans, independently of their size, number or type of branches, and amount of substitutions such as fucosylation. Each conformer is identified by a digit string of length n equal to the number of torsion angles in the glycan. The string begins at the free-reducing end of the glycan, which can be linked to the asparagine residue of proteins by oligosaccharyltransferase enzymes. *N*-glycans always start with a β 1 − 4-linked GlcNAc dimer, followed by a mannose residue (compare Figure S1). For each linkage, the string reports digits assigned to ϕ, ψ, and ω (if applicable) in this order. In correspondence of a junction (leading e.g., to an α 1−6 and an α 1−3 branch after the first mannose), a string separator is introduced, labeled according to the C atom at the branch origin (e.g., **6**− for 1−6 linkages). The string continues first along the branch of the higher C atom (6 in our case) until reaching the terminal residue, prior to returning to the last junction and following the branch of the next-lower C atom (3 in our case). Additional modifications like the attachment of fucose residues or bisecting GlcNAc residues are included after all other branches are assigned. Since our models have only one or two junctions, the separators of primary branches are labeled with bold numbers (**6**− or **3**−) for clarity. The string digits indicate the intervals of values in which the torsion angle lies, following the IUPAC nomenclature for dihedrals.[56] Namely, the digits for ϕ and ψ and the corresponding interval of radian values are C =

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

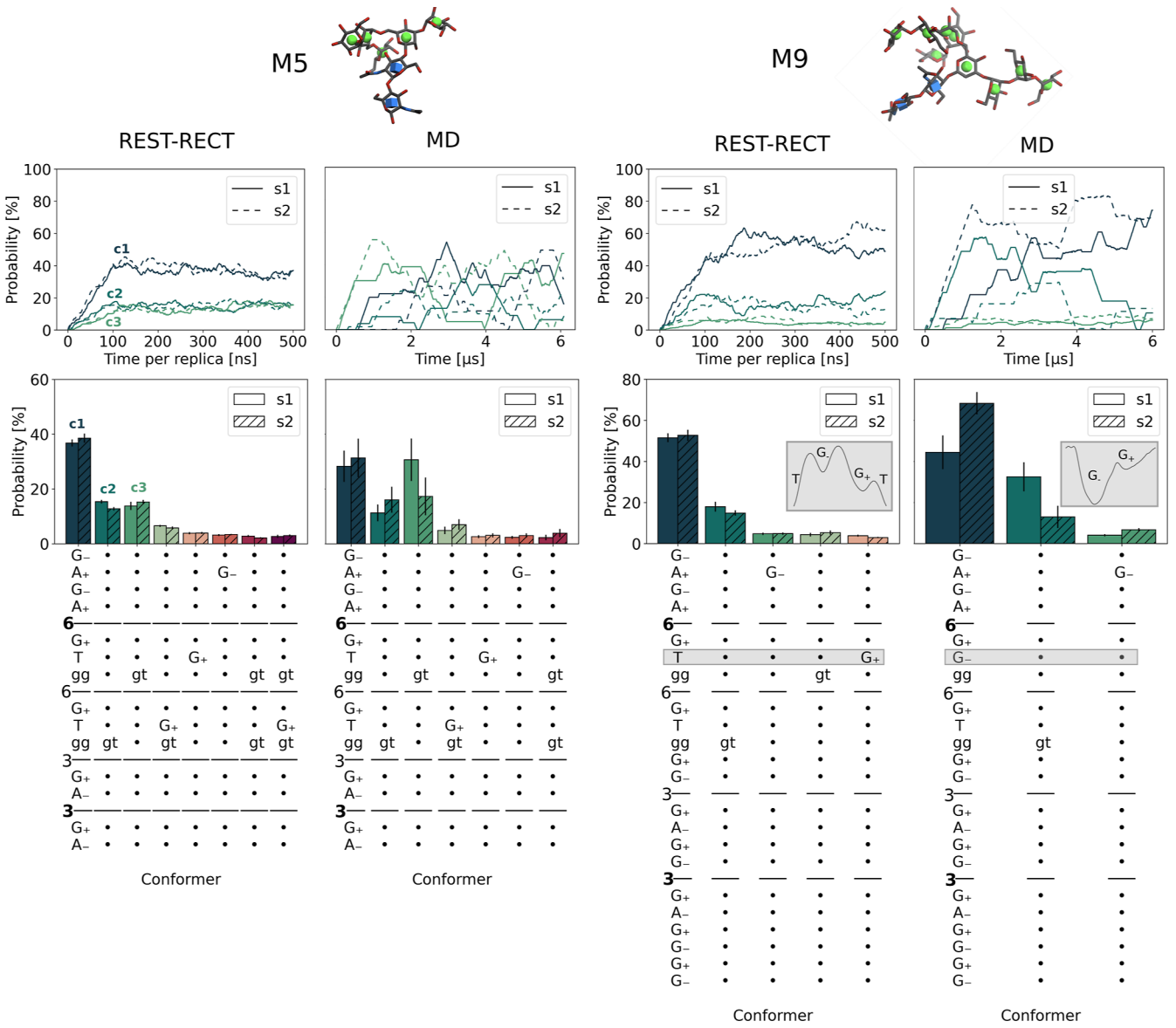


Figure 3. Comparison of REST-RECT with plain MD simulations for systems M5 and M9 using the CHARMM36 force field. The upper panel includes atomistic structures of each glycan, visualized with the 3D-SNFG tool of VMD.[65] The middle panel shows the moving average for the three most-populated conformer clusters using a window size of 100 ns (REST-RECT) and 1.2 μs (MD), corresponding to the same sampling time. Two separate simulations were performed with differing initial starting configurations (s1 and s2, see Methods). The lower panel reports the resulting conformer distributions. The conformer string is given on the *x*-axis, where each digit stands for a torsion angle, the letter representing the occupied free-energy minima (see Methods). Dots are used instead of letters when no change could be observed in comparison with the most-populated conformer cluster. The gray boxes highlight a key conformational difference between REST-RECT and MD simulations, with a depicted free-energy landscape of the corresponding angle (inserts). Only conformers with a probability higher than 2.5% are plotted.

$[-0.52, +0.52)$, $G_+ = [+0.52, +1.57)$, $A_+ = [+1.57, +2.62)$, and $T = [+2.62, \pi)$, or $[-\pi, -2.62)$, $A_- = [-2.62, -1.57)$, and $G_- = [-1.57, -0.52)$. The digits for ω are gg = $[-2.62, 0)$, gt = $[0, 2.62)$, and tg = $[2.62, \pi)$ or $[-\pi, -2.62)$. The assignment of each torsion angle to a given interval is performed in the following way. First, the free-energy profile associated with rotation along the torsion angle is calculated from all frames of MD trajectories (either plain MD or REST-RECT MD). The position of the free-energy minima is then labeled according to the nomenclature above. All angles belonging to the same free-energy basin (around a minimum between the two neighboring maxima) are finally labeled equally to the minimum of their basin.

Probability Distributions. In all simulations, torsion angle values were recorded for each frame and converted into conformer strings. Histograms were constructed using the individual conformers as bins. In order to assess statistics, block averaging was performed, separating the data set in evenly distributed blocks. The average of all blocks $\overline{X} = \frac{1}{N}\sum_{j=1}^{N} X_j$ was calculated over $N = 10$ blocks, where X_j is the average calculated within each jth block. Error bars were calculated as standard deviations of the sampling distributions (standard error of the mean): $\mathrm{std}(\overline{X}) = \sqrt{\frac{\mathrm{var}(\overline{X})}{N}}$, with the variance of the sampling distributions being $\mathrm{var}(\overline{X}) = \left(\frac{N}{N-1}\right)\left[\frac{1}{N}\sum_{j=1}^{N} X_j^2 - \left(\frac{1}{N}\sum_{j=1}^{N} X_j\right)^2\right]$.

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

Dimensionality Reduction. Four different dimensionality-reduction techniques were compared in this study. We always included 31 250 data points ($n_{samples}$), with the different torsion angles of each *N*-glycan defined as features ($n_{features}$), resulting in a feature matrix $\mathbf{X}$ with the shape $n_{samples} \times n_{features}$. In order to account for the periodicity of torsion angles, their sin and cos values were used in the feature matrix $\mathbf{X}$ instead of the torsion angles, in all cases except for the sketch map.

Principal Component Analysis. The PCA calculations were performed with the scikit-learn package.[57] Whenever two different feature matrices $\mathbf{X}$ were compared to each other, for example, stemming from simulations performed with two different force fields, the corresponding data sets were concatenated prior to PCA calculation. Free-energy differences (ΔG) along the principal components 1 and 2 defining the low-dimensional latent-space matrix $\mathbf{T}$ were calculated by constructing two-dimensional histograms with 35 bins and converting the histogram probabilities P according to $\Delta G = -k_B T \ln(P)$.

Diffusion Maps. Diffusion maps were computed based on the same algorithm as described by Bottaro et al.[58] using the Gaussian kernel $K_{ij} = \exp\left(\frac{-d_{ij}^2}{2\sigma^2}\right)$, where d_{ij} represents the pairwise distance matrix between sin and cos of the torsion angles values of conformers x_i and x_j. The parameter σ, defining the size of the neighborhood including similar conformational structures, was set equal to 1.7. We note that the used algorithm is different from the original implementation[31] in that the transition matrix is simultaneously normalized and made symmetric by means of an iterative procedure.[58] This yields results that are equivalent to those of the recently introduced bi-stochastic kernel method.[59]

Sketch Map. Sketch maps were calculated with the DimRed module of PLUMED (version 2.6). The matrix of dissimilarities between the frames in the feature space was calculated using the Euclidean distance measure. 500 landmark points were obtained from the farthest point sampling and subjected to minimization of the stress function. The switching distance was chosen to be equal to 2.5 for all *N*-glycans, which lies roughly in the middle of the range of distances characterized by Gaussian fluctuations (Figure S6). We set $A = B = 4$ in the high-dimensional space and $a = b = 2$ in the low-dimensional space, as defined by Ceriotti and coworkers.[33] The tolerance for the conjugate gradient minimization was set to 10^{-3} using 20 grid points in each direction and 200 grid points for interpolation. Five annealing steps were used, and the remaining trajectory data were projected on the constructed sketch map.

Kernel Principal Covariates Regression Model. The recently developed kPcovR algorithm was employed as described in the tutorials at https://github.com/lab-cosmo/kernel-tutorials, using the scikit-cosmo implementation. Besides the feature matrix $\mathbf{X}$, the conformer strings assigned to each frame were employed as properties in the property matrix $\mathbf{Y}$. Prior to fitting, the input was centered and standardized by removing the mean and scaling the data to obtain a unit variance. We note that our data set was used as a whole and not split into separate training and testing data sets. A Gaussian kernel with $\gamma = 1$ was used, and mixing parameters α were calculated for each simulated glycan on a subset of 1000 frames for two dimensions, leading to values of 0.1 for M5, 0.5 for FM5, 0.9 for M9, and 0.1 for A2G2S2. The regularization parameter λ was set to 10^{-4} for the linear regression part. The error of the linear regression part was assessed by comparing the true properties $\mathbf{Y}$ and the predicted properties $\hat{\mathbf{Y}}$.

J-Coupling Calculations. Validation of the glycan structures obtained from MD simulations was carried out by comparing theoretically calculated with experimentally measured scalar $^3J_{H,H}$ NMR coupling constants. The comparison is meaningful only for ω torsion angles in α 1 − 6 linkages since ϕ and ψ lack the necessary proton pair, whereas the $J_{H5,H6}$ and $J_{H5,H6'}$ constants can be both computed and measured[60] (see Figure 7 below for the atom nomenclature). We note that caution must be taken when comparing the results of different force fields because of the inconsistencies in atom labeling conventions. In particular, the H6 (H6S) and H6′ (H6R) hydrogens are named H61 and H62 in CHARMM36, respectively, while the opposite names (H62 and H61) are used in GLYCAM06j. The theoretical calculations of the coupling constants were performed using three different sets of modified Karplus equations,[61] namely:

1) The equation of Altona and Haasnoot[62]

$$^3J_{H,H} = P_1\cos^2\omega + P_2\cos\omega + P_3 + \sum_{i=1}^{4}\Delta_{\chi_i}\{P_4 + P_5\cos^2(\zeta_i\omega + P_6|\Delta_{\chi_i}|)\}$$

where the sum runs over the different substituents (in our case, H, C, and two O), the P parameters are taken from the original data set, the electronegativity values Δ_{χ_i} are equal to 0 for H, 0.4 for C, and 1.3 for O, and the substituent orientations ζ_i are either −1 or 1. The equation is applied to the torsion angles ω = H5−C5−C6−H6 or ω = H5−C5−C6−H6′. For example, the former has electronegativity values of 0.4 for i = 1, 1.3 for i = 2 and 3, and 0 for i = 4, with $\zeta_{1,2} = 1$ and $\zeta_{3,4} = -1$.

2) The equations of Stenutz[63]

$$^3J_{H5,H6'} = 5.08 + 0.47\cos\omega + 0.90\sin\omega - 0.12\cos 2\omega + 4.86\sin 2\omega$$

$$^3J_{H5,H6} = 4.92 - 1.29\cos\omega + 0.05\sin\omega + 4.58\cos 2\omega + 0.07\sin 2\omega$$

with ω = O5 − C5 − C6 − O6.

3) The equations of Tafazzoli[64]

$$^3J_{H5,H6'} = 5.06 + 0.45\cos\omega - 0.90\cos 2\omega + 0.80\sin\omega + 4.65\sin 2\omega$$

$$^3J_{H5,H6} = 4.86 - 1.22\cos\omega + 4.32\cos 2\omega + 0.04\sin\omega + 0.07\sin 2\omega$$

with ω = O5 − C5 − C6 − O6.

The general equation of Altona and Haasnoot can be adapted to different kinds of linkages due to the flexible choice of substituents. In contrast, the equations of Stenutz and Tafazzoli were derived specifically from *J*-coupling constants computed with density functional theory for a model aldopyranosyl ring and D-glucose/D-galactose. The computed coupling constants were averaged over all 62 500 frames in each REST-RECT simulation of the *N*-glycans (considering only the starting conformation s1). Block averaging was used

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

to compute error bars, as described above for probability distributions.

Puckering. The general Cremer−Pople pucker coordinates θ and ϕ for six-membered rings were used to evaluate the ring distortions (Figure 1, bottom) using the Puckering module of PLUMED (version 2.6). Two-dimensional histograms with 200 × 200 bins were constructed from the frames of REST-RECT simulations (only for the s1 starting conformation) and converted to free-energy surfaces as described above. Two-dimensional pucker plots along ϕ and θ were plotted using the Mollweide projection, also termed homolographic or elliptical projection. This pseudocylindrical map projection is equal-area, meaning that areas, densities, and, thus, free-energy values are preserved.

RESULTS

In this section, we first applied the REST-RECT method to different *N*-glycans and benchmarked its ergodicity against plain MD employing the CHARMM36 force field. In the second step, we compared the ability of different dimensionality-reduction techniques, namely PCA, diffusion map, sketch map, and kPcovR, to effectively cluster distinct *N*-glycan conformations. The sampling and clustering protocol was then employed to compare the free-energy landscapes of different *N*-glycan systems, to assess the quality of different force fields, to validate the results against measured NMR J-coupling parameters, and to investigate the dependence of ring puckering on glycan conformation.

Comparison of REST-RECT and MD. Enhanced-sampling REST-RECT and plain MD simulations were performed to explore the conformational phase space spanned by the torsional angles of the *N*-glycan models shown in Figure 1A. A fair comparison of the two methods was ensured by using the same total simulation time, amounting to 6 μs for plain MD and to 500 ns for the REST-RECT simulations, which included 12 system replicas. The distribution of replica temperatures led to uniform and adequate exchange probabilities, as indicated by the achievement of at least 10 round-trips per replica in all cases (Figure S3). The obtained probability distributions of the conformer populations are shown in Figure 3 for the representative cases of the M5 and M9 glycans. Labeling of conformers, in line with the official IUPAC nomenclature for dihedral angles,[56] was performed as described in detail in the Methods section (see also Figure S1). The first notable result is that stable and consistent conformer population distributions were already obtained after about 100 ns of REST-RECT simulation, irrespective of the starting conformation s1 or s2 (see Methods), as shown in the upper panels of Figure 3. In contrast, plain MD simulations displayed large fluctuations, poor convergence, and significant dependency upon the starting conformation for individual conformers. This resulted in much larger error bars associated with the conformer distribution histograms (Figure 3 lower panel) in comparison with the enhanced-sampling simulations, especially in the case of M9 glycan. Furthermore, plain MD simulations predicted different conformer populations than REST-RECT. For M5, the differences were not dramatic, and in particular, the most-populated clusters corresponded to the same conformer. For M9, however, the ψ angle in the main 1−6 linkage between two mannose residues (gray box in Figure 3, lower panel) remained stuck in a $G\_$ free-energy minimum and did not reach the global-minimum T conformation predicted by REST-RECT. Analogous conclusions were drawn for the other *N*-glycans FM5, A2G2S2, and A2G2 (Figure S2), with FM5 showing the largest improvements associated with proper REST-RECT sampling.

The analysis revealed interesting common patterns of torsion-angle conformations in certain structural elements across the investigated models. For instance, the sequence $G\_A_+G\_A_+$ was predicted as the global minimum of the chitobiose core for all glycans, and the G_+Tgg sequence characterizes the 1−6 linkages in most cases. Moreover, there was an evident preference for a *gg* conformation of the ω angle, which originates from the *gauche* effect.

Dimensionality Reduction of *N*-Glycans. The set of strings associated with the most-populated conformer clusters of a given *N*-glycan (as shown in Figure 2) is a reduced representation of the free-energy minima in the high-dimensional conformational phase space spanned by all torsion angles. However, such representation becomes cumbersome when comparing different *N*-glycan systems and does not give a measure of the structural differences among the different conformers. We therefore investigated the ability of several dimensionality-reduction techniques to deliver two-dimensional representations of the conformer clusters in an efficient and physically meaningful manner using all torsion angles as input features. As stated in the Introduction, we concentrated on three unsupervised learning algorithms, namely PCA, diffusion maps, and sketch maps, and the supervised-learning algorithm kPcovR.

PCA and diffusion maps generated almost identical two-dimensional representations (Figures 4 and S7) and gave similar eigenvalue progressions along the PCA/diffusion components (Figure S6). For all *N*-glycans, there was an obvious gap between the first one to three eigenvalues (two for the case of M5 shown in Figure 4) and the remaining ones, indicating that a few corresponding structural features are more important than others in differentiating the glycan conformers. Having a closer look at the PCA of M5, *PC1* differentiates conformers along their ω torsion angles in the side branch 6− and *PC2* along the main branch **6**−, corresponding to 25.0 and 19.5% of variance in the underlying data, respectively. Knowing that the highest variance is included in the rotations around ω torsion angles gave us an a-posteriori justification for the two selected initialization states s1 and s2, differing in the states of ω in **6**−, situated in very different energetically conformers. This differentiation of conformers from each other by two ω angles gave rise to four main groups of conformer clusters (Figure 4). The overlap between the conformer clusters in each of those groups originated from the fact that these conformers only differ in ψ angles, which can not be resolved in this two-dimensional representation, revealing the limitations of the PCA and diffusion-map algorithms (Figure 3).

In fact, it is interesting to note that the number of highest, well-separated PCA or diffusion-map eigenvalues is equal to the number of ω angles present in the *N*-glycan structures M5 (two), FM5 (three), and A2G2S2 (one) (Figure S6). In this respect, M9 is an exception, presenting only one well-separated eigenvalue (corresponding to the 6− branch) but two ω angles. Correspondingly, in the two-dimensional maps, there is a clear separation of cluster conformers along the *PC1* component but some overlap along the *PC2* component (Figure S7).

The sketch-map algorithm clustered conformers in a similar way to PCA and diffusion map, differing in the overall spatial arrangement but with only marginally better separation of the

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

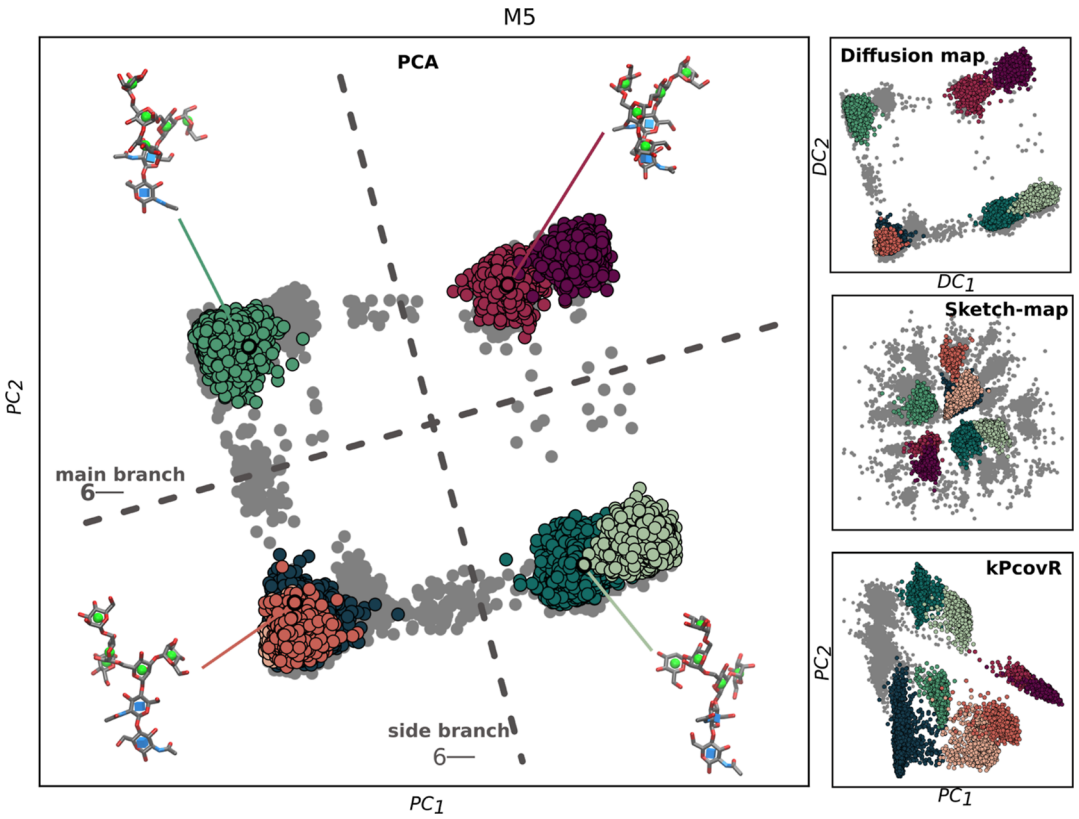


Figure 4. Comparison of four different dimensionality-reduction algorithms to cluster the distinct conformers of *N*-glycan M5. PCA, diffusion map, and sketch map employ all M5 torsion angles as features, whereas kernel principle covariates regression (kPcovR) additionally uses the conformer strings as an additional property. Each gray point corresponds to one frame, and colored points correspond to the respective conformers given in Figure 3. Sampling was performed from REST-RECT simulations (only s1).

clusters. The sketch-map analysis did not allow for a ranking of the components and thus for an unbiased identification of the most important structural features of the system. The kPcovR algorithm separated the most-occupied conformers in the most effective way, resulting in cluster clouds with only little overlap to neighboring ones (Figure 4). However, the algorithm did not allow for a meaningful interpretation of how conformers are separated or clustered together since no characteristic feature could be assigned to the kPcovR principal components. It rather seems that the clusters were ranked according to their population probabilities along *PC1*, as suggested by the progression of colors from left to right in the two-dimensional map and in the histograms of conformer probabilities (Figures 3 and 4). The calculated losses for the regressions in kPcovR can only be interpreted as relative values, as $\mathbf{Y}$ consists of discrete conformers, but the linearity between $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ along the target (red line) is very clear (Figure S6).

From this analysis, we conclude that, for the investigated systems, PCA, diffusion map, and sketch map can be used almost interchangeably with respect to their physical meaning, while kPcovR may be useful whenever a two-dimensional representation with well-separated conformer clusters is sought for. In this study, we employed PCA in the following applications of the method, as it is the most straightforward and computationally effective algorithm.

Comparison of *N*-Glycan Structures. As a first application, we compared the conformational landscape of high-mannose-type and complex *N*-glycans. For this purpose, we used PCA as a method to represent the free-energy maps associated with the conformational ensembles of our five glycan models in two dimensions. We focused on the influence that various chemical modifications (fucosylation, sialylation, and variation of branch length) might have on the resulting structures. In doing so, we included in the analysis only the structural features common to all compared structures, highlighting the effect of the mentioned modifications on the free-energy maps (Figure 5). In general, the analysis showed that the core structure of the high-mannose-type glycans was only marginally affected by the addition of further residues. Fucosylation of M5 (FM5) had no effect at all on the conformational free-energy landscape, whereas elongation of the shorter branch by two units (M9) led to a slight stabilization of the main conformer and destabilization of the secondary minima at larger *PC1* values (Figure 5, upper row). Instead, sialylation of A2G2 (A2G2S2) with additional Neu5Ac units on both branches slightly deepened all three secondary energy minima at the expense of the most-populated region of the conformational phase space in the bottom-right corner of the map (Figure 5, lower row).

Comparison of Force-Field Predictions. Besides enabling comparisons of different *N*-glycan structures, the developed methodology also allowed us to compare very accurately the structural prediction capability of different force fields. Here, we compared the two most widely used force fields for protein and carbohydrate systems, namely CHARMM36 and GLYCAM06j. The results for A2G2S2 are shown exemplarily in Figure 6 and for the other glycans (except FM5 because of its similarity to M5) in Figure S8. The comparison revealed very substantial differences in terms of both conformer distributions and free-energy landscapes, and

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

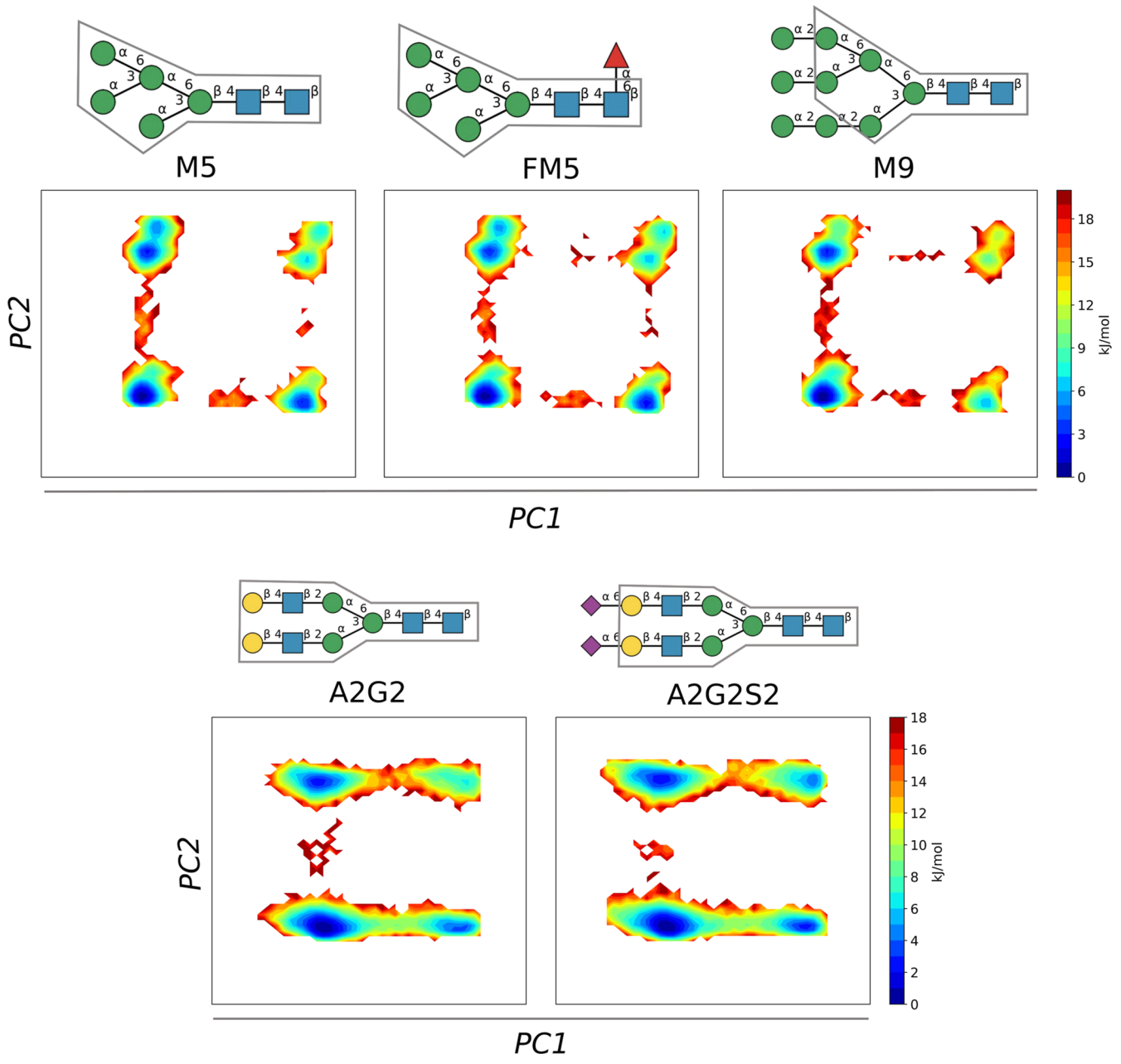


Figure 5. PCA free-energy maps of different *N*-glycans. The upper panels compare the three high-mannose-type *N*-glycans M5, FM5, and M9, whereas the lower panels compare A2G2 against its sialylated variant A2G2S2. Only the torsion angles common to all structures (boxes around the schematic glycan models) were used as features in the analysis. A common PCA was performed by concatenating the data sets of M5, FM5, and M9 and those of A2G2 and A2G2S2, respectively. Sampling of the phase space was performed with REST-RECT simulations, starting from the s1 conformation.

even the global-minimum structures were different. CHARMM36 predicted that the majority of conformers (and thus the global free-energy minimum) cluster in the right-bottom region of the PCA map, in contrast with GLYCAM06j, which predicted a global minimum in the left-bottom region. The position of the secondary minima was also different in the two cases. Comparison of the predicted conformer strings indicated that the major differences arise from the ψ angle of the main branch **6**− as well as the two ω angles of the terminal 1−6 linkages between Gal and Neu5Ac (Figure 6, lower panels). In the global-minimum structure predicted by CHARMM36, ψ was in a *T* conformation and ω in a *gt* conformation. These conformations changed to G_+ and *tg* in the global-minimum structure predicted by GLYCAM06j, respectively. Similar considerations hold for M5, M9, and A2G2 (Figure S8), although the conformer distributions were dramatically less different than in the case of AG2S2.

By looking at the one-dimensional profiles along selected torsion angles, it becomes evident that the discrepancies arise from only subtle differences in the force field parametrizations. For instance, the free-energy differences between the *T* and G_+ conformations of the ψ angle or between the *gt* and *tg* conformations of the ω angle were less than 5 kJ/mol. However, such small differences have a profound effect on the resulting multidimensional free-energy landscape and lead to rather distant global minima, as observed above. In the next

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

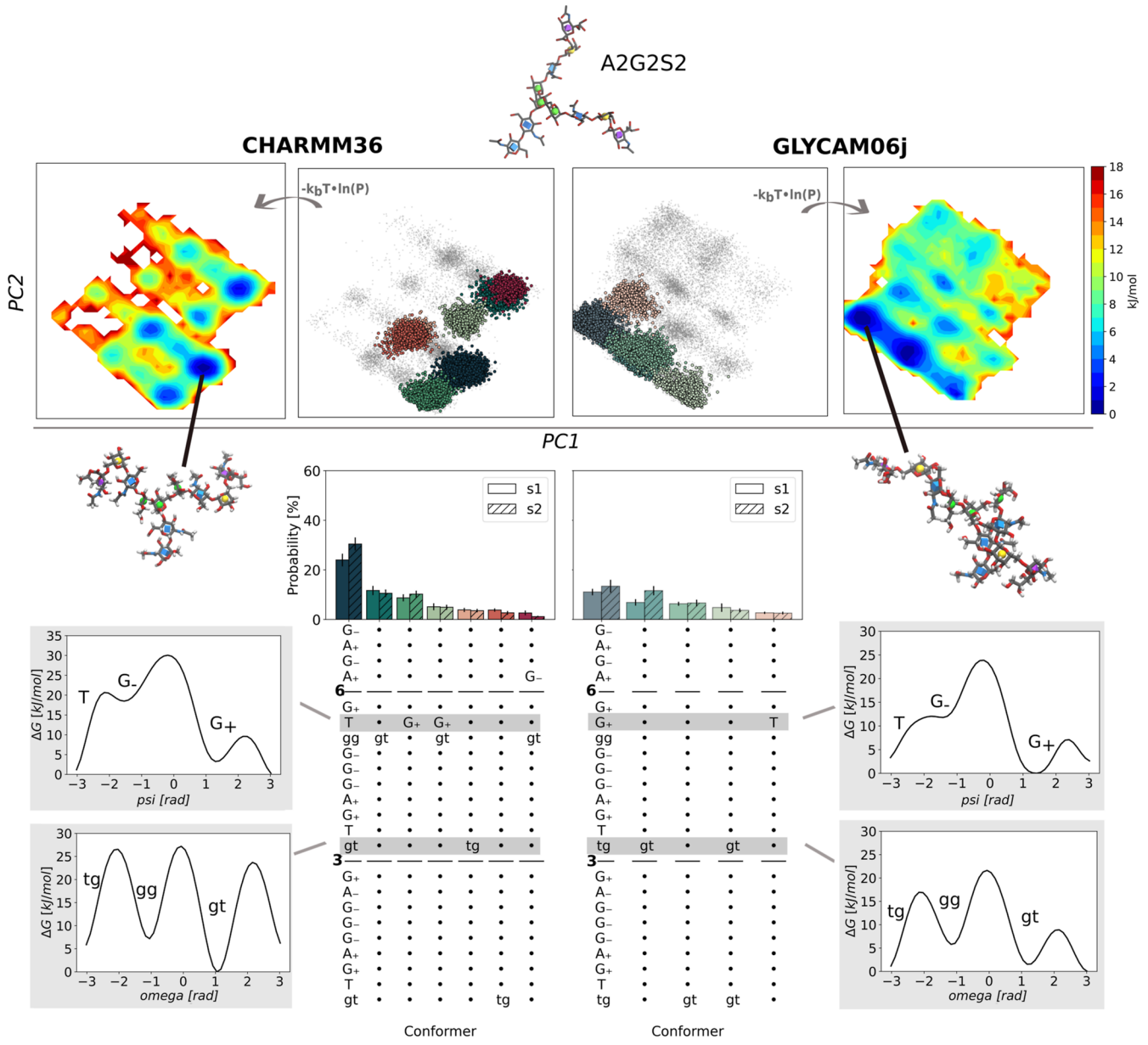


Figure 6. Comparison of the conformational ensembles of A2G2S2 predicted by REST-RECT simulations with either the CHARMM36 or the GLYCAM06j force field. The upper panels show the PCA maps of conformer clusters and the corresponding free-energy landscape. The clusters are colored in accordance with the conformer distributions shown in the lower panels. Free-energy profiles along selected torsion angles (indicated by the gray rectangles) are represented beside the conformer strings and labeled with the conformations of the free-energy minima.

section, we will show that these differences led to markedly different predictions of NMR spectroscopic fingerprints of the glycan populations by different force fields.

Force-Field Validation by Calculation of NMR Parameters. A comparison of experimental NMR data with the corresponding observables predicted theoretically by the two force fields was made to ascertain which one can be considered more accurate in terms of torsion angle description. We stress that this assessment is only reliable when the ergodicity of the simulations is fulfilled, which requires complete phase-space sampling by means of converged REST-RECT simulations.

As shown above, the largest variability among the different conformers of *N*-glycans originated from the ω torsion angles around the 1−6 O-glycosidic linkages. The three protons H5, H6, and H6′ harbored by these linkages (see Figure 7) gave rise to well-defined NMR J-coupling constants, whose values depended on the relative distances between the H nuclear spins and thus on the conformation of the ω angles. Therefore, a comparison between measured and theoretically computed $J_{H5,H6}$ and $J_{H5,H6'}$ frequencies enabled a clear validation of the predicted glycan structures.[60] The calculations were performed by ensemble averages of the coupling constants computed for all conformers sampled by the REST-RECT simulations using three different parametrizations of the empirical Karplus equation, as described in the Methods section (Figure 7).

For the main branch of M5, GLYCAM06j led to better agreement between experimental and theoretical $J_{H5,H6}$ and $J_{H5,H6'}$ frequencies, whereas CHARMM36 performed better for the side branch. Regarding M9, CHARMM36 performed better than GLYCAM06j for both ω angles, although both force fields overestimated the $J_{H5,H6'}$ frequency by over 2 Hz.

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

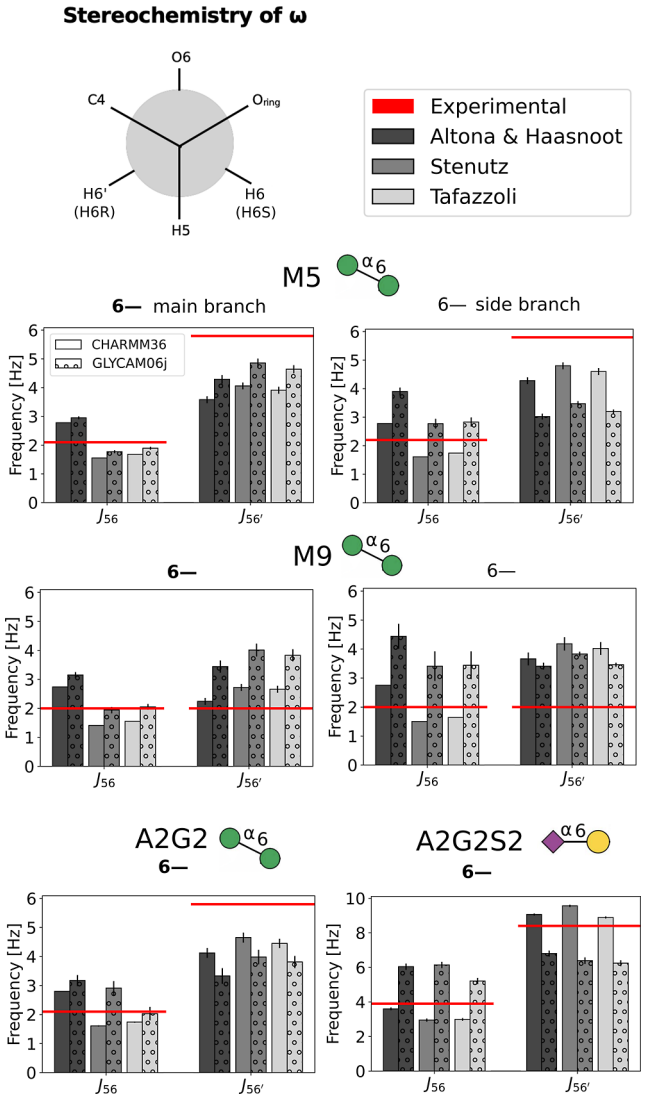


Figure 7. Validation of the ω angle populations by comparison of computed and experimental NMR J-coupling constants for M5, M9, A2G2,[60] and A2G2S2.[27,66] The upper-left panel shows the stereochemistry of an ω angle in a *gg* conformation along its C5 and C6 atoms, with labeled protons. The legend on the right reports the color code of the plots below, referring to three different parametrizations of the Karplus equation used to compute the J-coupling constants. All plotted values are also reported in Table S1.

We would like to note that the experimental J-couplings of M9 were only reported as approximations in the original paper[60] but are used here due to the lack of other data sources. For the only ω torsion angle in A2G2, CHARMM36 predicted slightly better frequencies than GLYCAM06j. For the sialylated variant A2G2S2, no experimental parameter of the ω torsion angle between two Man residues was available; therefore, the comparison was made for the J-coupling constants of the 1−6 linkage between Gal and Neu5Ac. In this case, the experimental values were collected for a different glycan, namely trisaccharide sialyl-α-(2−6)-lactose,[27,66] but could be used here as an approximation because this glycan carries the same terminal branches as those of A2G2S2. The predicted CHARMM36 values of both $J_{H5,H6}$ and $J_{H5,H6'}$ were in very good agreement with the experimental ones.

Overall, neither force field reproduced all experimental J-coupling constants with great accuracy (i.e., within the intrinsic error bars of the theoretical method), but CHARMM36 seemed to deliver better structural predictions than GLYCAM06j, especially in the cases of M9 and A2G2S2.

Analysis of the Sugar Puckering Conformations. So far, we have only focused on the torsion angles of the *N*-glycan molecules, which were considered as explicit RECT CVs in our simulations. However, the combination with the REST2 method also allowed good sampling of other structural degrees of freedom, in particular of the puckering conformations of the individual monosaccharide units. In this section, we investigate whether the puckering free-energy landscapes spanned by the Cremer−Pople parameters θ and ϕ were dependent on the specific conformation of the glycan, as determined by its torsion angle.

For this aim, we constructed and represented two-dimensional polar free-energy maps in a way that conserves the area defined by intervals of the θ and ϕ pucker coordinates. As usual, the free energy was computed from the histograms of conformer population probability in the ground replica of the REST-RECT simulations. In Figure 8 we show the puckering maps of the GlcNAc1 and GlcNAc2 units of the most-populated and the second-most-populated torsion angle conformer clusters (with respect to the distribution reported in Figures 3 and S2).

For all considered glycans, the CHARMM36 force field predicted a $^4C_1$ chair conformation as the global minimum of the monosaccharide units in both conformer states. Minor differences could be observed in the relative position (or even appearance) of the secondary minima, corresponding to boat conformations (at the equator of the maps, for $\theta = \pi/2$) and mostly for monosaccharide units situated in the core of the *N*-glycans. The third-most and fourth-most conformers have also been analyzed but not explicitly shown as they resemble the same occupation of minima as seen in the most-populated and second-most-populated clusters. Whether larger shifts of the local minima or even stabilization of boat conformations could result from constraining the glycans within conformer states far from their natural global minima (e.g., upon binding of the molecule into tight protein pockets) is a question that should be investigated in future studies.

The analysis of the pucker landscapes along the branches of the A2G2S2 glycan, which is composed of very diverse monosaccharide units, revealed again a strong propensity for the chair conformation $^4C_1$, except for the terminal Neu5Ac units that were in a $^1C_4$ conformation (Figure S10). However, the relative boat propensities were quite different for different units, being very low or absent for Man and Gal units, more evident for GlcNAc units, and the strongest for Neu5Ac units. We note that the GLYCAM06j, in comparison with CHARMM36, generally predicted a broader exploration of the pucker phase space, resulting in an increased appearance of local minima and smaller energy differences between different regions of the maps (Figure S10). Only the terminal Neu5Ac units presented very similar maps for both force fields, with the same distribution of minima and the same degree of phase-space exploration.

DISCUSSION

Glycosylation is a ubiquitous modification of biomolecular systems, involving proteins, lipids, and, as more recently discovered, RNA.[67] As a prominent example, glycans tethered to the SARS-CoV-2 spike protein modulate its enzymatic function,[68,69] providing but a small glimpse of the importance

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

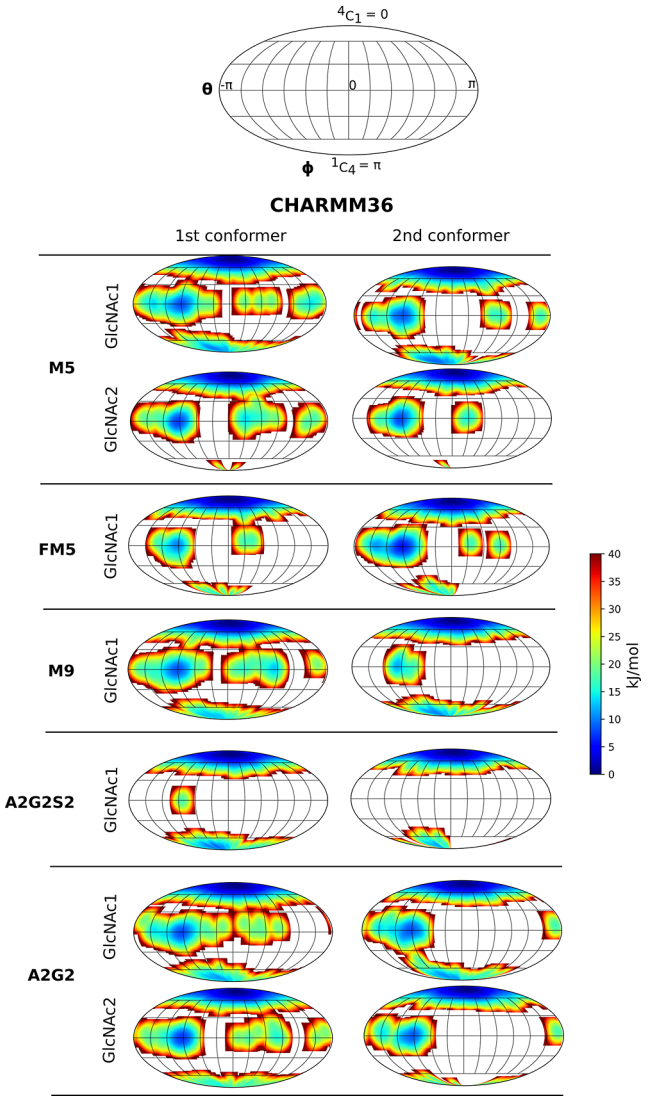


Figure 8. Free-energy surface along the Cremer−Pople puckering coordinates θ and ϕ for GlcNAc residues of all N-glycans compared for their most and second-most occurring conformers. Collective variables were computed from REST-RECT simulations by histogram construction and conversion to free energies.

of glycosylation in life science.[3] However, the tremendous diversity of *N*-glycan structures promoted by all possible cellular glycosylation pathways hampers a rational understanding of clear structure−function relationships at the basis of a still mysterious sugar code.[70] Our claim is that a correct and comprehensive description of the whole ensemble of conformers is crucial for the prediction of the biological function of glycan systems, motivating a fundamental exploration of the conformational phase space of representative *N*-glycans by means of enhanced-sampling molecular dynamics simulations (Figure 2).

We were able to show that REST-RECT simulations provide converged conformer distributions with a complete sampling of all torsion and puckering angles within a few hundred nanoseconds of cumulative time, with better accuracy and using less computational time than long plain MD simulations. Sufficiently short round trip times reveal the strength of the RECT method, capable of biasing up to 22 CVs simultaneously while still ensuring adequate diffusion in the replica space. This behavior originates from an adjusted ergodicity by scaled bias factors over the replica ladder, ensuring a proper compensation of free-energy barriers by a self-consistent addition of one-dimensional bias potentials.[28] Alternative methods such as temperature REMD[19] are computationally too demanding in solvated systems.[71] Bias-exchange metadynamics[72] would have required more replicas for the same number of biased CVs and would have only enabled biasing them one at a time.[28] It may be argued that some of the chosen CVs are in fact redundant; in particular, axial ϕ torsion angles in α-linkages occupy only the gauche conformation (G_+) due to the so-called exoanomeric effect. This is due to the favorable overlap of one oxygen lone-electron pair with the antibonding orbital of the adjacent C−O bond,[14,73,74] an effect that needs to be appropriately mapped by torsion, Lennard−Jones, and Coulomb terms in force-field potentials. However, there is no computational advantage in excluding those CVs from the biased scheme, and it is indeed reassuring to see that the results do confirm such background-knowledge details.

Although the computational method ensures ergodicity of the performed simulations, their accuracy in predicting experimental observables remains limited by the functional form and parametrization of the employed force fields. Various carbohydrate force fields have been developed and refined for selected systems and cases, such as the stability of protein−carbohydrate complexes, the conformational behavior of linear polysaccharides, or the ring distortions of monosaccharide units.[15,16,75] Generally speaking, the main families of biomolecular force fields, namely CHARMM, GLYCAM (AMBER), and GROMOS, have been shown to have good performance in reproducing the behavior and predicting experimental data of polysaccharide systems, with few exceptions.[14,16] However, depending on the saccharide size, the property under investigation, and the required levels of detail, differences among the force field families do emerge, which can be traced back to how well the steric, electrostatic, and torsional energy terms represent the physical reality and mimic the actual glycan behavior.[14]

We have performed an in-depth analysis of the *N*-glycan conformer distributions predicted by the CHARMM36 and GLYCAM06j force fields, focusing on converged free-energy profiles of torsion angles that shape the three-dimensional glycan structure and its flexibility. The observed different phase-space distributions for A2G2S2 and A2G2, as well as the different conformer distributions for M5 and M9, originate from different free-energy profiles around the ψ and ω torsion angles in 1−6 linkages. Especially, the ψ angle of branch **6**− in A2G2 and A2G2S2 is a critical feature, which differentiates between two main conformers, previously named "backfold" and "extended".[26] CHARMM36 consistently produced conformer distributions with only a few highly populated states, whereas GLYCAM06j produced broader distributions and flatter associated free-energy landscapes. The frequent revisions of the force-field parametrization of torsional terms up to the present day indeed show that a correct description of rotational barriers in glycan systems is not at all straightforward.[76] In particular, the contribution of torsional energy and electrostatic interactions needs to be balanced with great care.[76] The two force fields compared here differ especially in the latter term: CHARMM36 adjusts partial atomic charges to fit solute−water interactions of carbohydrate fragments computed with quantum mechanical methods, whereas the partial charges of GLYCAM06s are derived from the restrained

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

electrostatic potential (RESP) method.[76] Our comparison of computed J-coupling constant values to the sparsely available data from NMR experiments indicated an overall better performance of the CHARMM36 force field. The three tested parametrizations of the Karplus equation yielded consistent results, although they are all based on empirical parameters so that some discrepancies are both expected and unavoidable. While all of our calculations were performed with the TIP3P water model, more complex water models like TIP5P have been shown to impact the predicted carbohydrate aggregation and protein−carbohydrate interactions.[77−79] A recent review about the modeling of complex carbohydrates summarizes the difficulties that might arise from the usage of different water models and general issues of overestimating carbohydrate−environment interactions.[80] Therefore, future studies should also focus on a critical assessment of how different solvent models impact the performance of glycan force fields in predicting the correct conformer cluster distributions.

The here-examined *N*-glycans have been the subject of several previous investigations. The comparison of conformer distributions upon chain elongations like the addition of core fucosylation or sialylation (Figure 5) has also been shown in previous studies to leave the equilibrium distribution almost unaffected,[81] although the here-used PCA representation visualizes results much more comprehensively. When analyzing glycan structures on a detailed torsion-angle-based level, the focus lies on the more flexible ω angles, preferably found in their gauche conformation as outlined for different *N*-glycans.[9] Both experimental and computational studies of M9 suggested that it is mainly confined in a gauche conformer, meaning that the ω torsion angle in the **6−** branch should be in a *gg* conformation, which is in agreement with our findings for both employed force fields.[82−84] The observed stabilization of the global minimum of M9 after elongation of one M5 branch by two 1−2-linked mannose units (see Figure 5) is probably due to an increased number of inter-branch hydrogen bonds.[85] A2G2S2 has been studied by Yang and co-workers[24] using REST2 in combination with Hamiltonian bias potentials, employing the CHARMM36 force field. This approach is similar to REST-RECT, the only difference being that Yang and co-workers used a biasing profile on the torsional angles as obtained in preliminary umbrella sampling simulations, whereas in our RECT scheme, the compensating profiles are computed on the fly. In fact, the reported free-energy profiles for individual torsion angles in ref 24 are overall in agreement with our CHARMM36 simulations (Figure S9), demonstrating converged phase-space sampling in both studies.

A2G2 was previously investigated by REMD using the GLYCAM06g force field, and the predicted relative populations of the ω angle (O6−C6−C5−C4) in branch **6−** amounted to 71 and 28% for the *gg* and *gt* conformers, respectively.[26] Our simulations with the GLYCAM06j force field, however, gave average values of 80% for *gg*, 11% for *gt* and 9% for *tg*. While the force field versions g and j only differ in the atom labeling for consistency with other AMBER force fields or in the addition of parameters for the protein−carbohydrate linkage, the simulations by Nishima and co-workers[26] differ from ours with respect to the type of sampling method. We believe that our REST-RECT simulations provide a more complete phase-space sampling, as demonstrated by the very good convergence (Figure S4) and the clear independence from the chosen initial configurations. Other earlier investigations of A2G2 using the CHARMM36 force field revealed a distribution of 52% *gg* versus 48% *gt* conformations in the ω torsion angle (O6−C6−C5−O5).[18] Our values of 71% for *gg* and 29% for *gt* computed with CHARMM36, however, are closer to the experimentally estimated values of 65 and 35%, respectively.[18,86] As identical force field parameters were used in both studies, the associated differences can have multiple reasons, namely (i) the use of the sTIP3P water model in contrast to the mTIP3P model used here, again pointing toward the need to better assess the performance of different solvent models; (ii) incomplete phase-space sampling in the earlier simulations; and (iii) the fact that, in their simulations, Galvelis and co-workers prevented ring inversion for all monosaccharide rings, although there are hints about a possible influence of puckering states on the glycan linkage conformations.[87,88]

The last point raises a further argument worth discussing. Ring flipping (inversion) or extended puckering distortion is a process that happens on the μs time scale and therefore requires very long simulation times or enhanced sampling.[14,89] However, for single monosaccharides, it can be occasionally observed in classical MD simulations. For instance, isolated GlcNAc and Neu5Ac readily undergo ring flips,[90,91] which could also be observed in our *N*-glycan systems when they are part of polysaccharide chains (Figures 8 and S10). Puckering was long considered irrelevant for most monosaccharides in polysaccharide chains, and therefore, the degrees of freedom associated with ring distortion were often ignored, reducing the complexity of MD simulations.[18] However, as mentioned above, several computational studies revealed the importance of ring-flipping events in determining the polysaccharide conformer distribution[75,88] where the degree of puckering flexibility is influenced by the molecular context and size of the *N*-glycan.[12,89] Another interesting feature is that the puckering of saccharide units connected via 1−4 linkages within glycan chains has been suggested to influence the ϕ and ψ distributions of the adjacent linkages. While we do observe that puckering free-energy landscapes of GlcNAc residues located in the chitobiose core differ among different conformers, we cannot confirm a direct effect on the adjacent linkages of puckering-prone units. In line with our observations, previous studies of glycan monomers and trimers did notice differences in the puckering landscapes predicted by different force fields and, in particular, pointed toward a better performance of the CHARMM36 force field.[75] Accurate prediction of ring-inversion free energies is expected to be very important for strongly constrained systems, such as glycan chains bound in protein pockets and subjected to enzymatic reactions, where ring inversion is often a key step of substrate activation before, for example, hydrolysis of the adjacent linkages.[92]

Importantly, we report a systematic comparison of dimensionality-reduction techniques applied to *N*-glycans. Up to now, this has been limited to PCA using atom coordinates as input features.[84,93] Nevertheless, clustering of glycan conformers has already been performed in the past, using, for example, end-to-end distances between glycan branches to describe their flexibility and identify the main conformers.[23] Additionally, the usage of spherical coordinates was introduced to describe the dynamical behavior of the **6−** branch, albeit the procedure was applied to the description of only one single branch.[26] Our successful application of various dimensionality-reduction techniques to represent the high-dimensional phase space of *N*-glycans highlights their enormous potential in

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

delivering a consistent analysis of the most-populated conformer clusters while simultaneously providing meaningful information about the most important structural features behind the used descriptive variables. This becomes very important for glycans composed of diverse monosaccharide units arranged in complex branched chains, presenting additional chemical modifications (fucosylation and sialylation) and including more than two ω torsion angles, making structural relationships not as intuitive as for small glycans like M5.

Complementing the clustering and dimensionality-reduction analysis, the here-introduced conformer strings provide a solid and IUPAC nomenclature-compliant way of labeling different conformers. Comparison of the strings easily reveals differences and similarities, which are immediately traceable to specific linkages and torsion angles. Previous classifications of identified conformers were performed with less clear nomenclature rules. For instance, the groups "backfold", "half backfold", "tight backfold", "extended-a", and "extended-b" were defined according to their ψ and ω torsion values of the first 1−6 linkage. These groups can still be differentiated using a low-dimensional representation constructed by PCA (see e.g., Figure S11 for A2G2). In line with a previous study, we found that "half backfold" conformers are in fact part of the "backfold" cluster.[18] However, this classification system is again limited to the description of only one single branch.

Finally, we note that our developed analysis framework is not limited to the investigation of *N*-glycans but can be extended to other glycan classes with other linkage types, such as O-glycans, GPI-anchors, or glycosphingolipids. Furthermore, the analysis can be extended to glycans bound to protein systems, although care must be taken in ensuring complete phase-space exploration despite the constraints dictated by amino-acid/glycan interactions. Adaptation of the REST-RECT algorithm to the study of protein-bound glycans will be the aim of future investigations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c01049.

> Guide with rules to construct the conformer string; probability distributions for various *N*-glycans (M5, FM5, M9, A2G2, and A2G2S2) simulated with the CHARMM36 or GLYCAM06j; round-trip times of all replica for REST-RECT simulations; performance plots of employed dimensionality-reduction techniques with representation for all N-glycans; force-field comparison for N-glycans M5, M9, and A2G2; representative free-energy profiles of torsion angles and monosaccharide puckering for glycan A2G2S2; and table including all computed and experimentally reported J-coupling constants (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Lucio Colombi Ciacchi − *Hybrid Materials Interfaces Group, Bremen Center for Computational Materials Science and MAPEX Center for Materials and Processes, University of Bremen, 28359 Bremen, Germany;* orcid.org/0000-0003-1444-9733; Email: colombi@hmi.uni-bremen.de

Authors

Isabell Louise Grothaus − *Hybrid Materials Interfaces Group, Bremen Center for Computational Materials Science and MAPEX Center for Materials and Processes, University of Bremen, 28359 Bremen, Germany;* orcid.org/0000-0002-0441-3860

Giovanni Bussi − *Scuola Internazionale Superiore di Studi Avanzati (SISSA), 34136 Trieste, Italy;* orcid.org/0000-0001-9216-5782

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c01049

Notes

The authors declare no competing financial interest.

The generation of conformer strings and plotting of conformer string distribution plots were carried out using GlyCONFORMER, a Python-based package deposited on GitHub under https://github.com/IsabellGrothaus/GlyCONFORMER. An example data set is given, demonstrating the analysis workflow in a Jupyter notebook. Additionally, the repository includes exemplary Jupyter notebooks offering help in performing dimensionality-reduction algorithms. Employed PLUMED input files for REST-RECT simulations were deposited in the PLUMED-NEST repository[94] under PlumID:22.028. Structure and trajectory files for all *N*-glycans simulated with the two employed force fields, CHARMM36 and GLYCAM06j, can be accessed from https://doi.org/10.5281/zenodo.6542267.

■ ACKNOWLEDGMENTS

The authors thank Georg Bossenz for the support implementing a Python script to calculate J-coupling values from MD simulations. Computational resources were provided by the North German Supercomputing Alliance (HLRN), project hbb00001. Additionally, computational resources were provided by the supercomputing centre CINECA in the framework of the HPC-Europa3 Transnational Access programme under application HPC177SW8P. This work has been supported by the Deutsche Forschungsgemeinschaft through the Research Training Group 2247, Quantum Mechanical Material Modeling-QM3.

■ REFERENCES

(1) Gabius, H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* **2011**, *36*, 298−313.

(2) Skropeta, D. The effect of individual N-glycans on enzyme activity. *Bioorg. Med. Chem.* **2009**, *17*, 2645−2653.

(3) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **2019**, *15*, 346−366.

(4) Ramachandran, G.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **1963**, *7*, 95−99.

(5) Cremer, D.; Pople, J. A. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354−1358.

(6) Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Marth, J. D.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E. Symbol nomenclature for glycan representation. *Proteomics* **2009**, *9*, 5398−5399.

(7) Cheng, K.; Zhou, Y.; Neelamegham, S. DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology* **2017**, *27*, 200−205.

(8) Battistel, M. D.; Azurmendi, H. F.; Freedberg, D. I.*NMR in Glycoscience and Glycotechnology*; New Developments in NMR, 2017; pp 1−19.
(9) Kirschner, K. N.; Woods, R. J. Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541−10545.
(10) Schwarz, J. C. P. Rules for conformation nomenclature for five- and six-membered rings in monosaccharides and their derivatives. *J. Chem. Soc., Chem. Commun.* **1973**, *0*, 505−508.
(11) Prestegard, J. H. A perspective on the PDB's impact on the field of glycobiology. *J. Biol. Chem.* **2021**, *296*, 100556.
(12) Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T. How Sugars Pucker: Electronic Structure Calculations Map the Kinetic Landscape of Five Biologically Paramount Monosaccharides and Their Implications for Enzymatic Catalysis. *J. Am. Chem. Soc.* **2014**, *136*, 1008−1022.
(13) Re, S.; Watabe, S.; Nishima, W.; Muneyuki, E.; Yamaguchi, Y.; MacKerell, A. D.; Sugita, Y. Characterization of Conformational Ensembles of Protonated N-glycans in the Gas-Phase. *Sci. Rep.* **2018**, *8*, 1644.
(14) Woods, R. J. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem. Rev.* **2018**, *118*, 8005−8024.
(15) Plazinska, A.; Plazinski, W. Comparison of Carbohydrate Force Fields in Molecular Dynamics Simulations of Protein-Carbohydrate Complexes. *J. Chem. Theory Comput.* **2021**, *17*, 2575−2585.
(16) Lazar, R. D.; Akher, F. B.; Ravenscroft, N.; Kuttel, M. M. Carbohydrate Force Fields: The Role of Small Partial Atomic Charges in Preventing Conformational Collapse. *J. Chem. Theory Comput.* **2022**, *18*, 1156−1172.
(17) Re, S.; Nishima, W.; Miyashita, N.; Sugita, Y. Conformational flexibility of N-glycans in solution studied by REMD simulations. *Biophys. Rev.* **2012**, *4*, 179−187.
(18) Galvelis, R.; Re, S.; Sugita, Y. Enhanced Conformational Sampling of N-Glycans in Solution with Replica State Exchange Metadynamics. *J. Chem. Theory Comput.* **2017**, *13*, 1934−1942.
(19) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(20) Wang, L.; Friesner, R. A.; Berne, B. J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(21) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2007**, *100*, 020603.
(22) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(23) Re, S.; Miyashita, N.; Yamaguchi, Y.; Sugita, Y. Structural diversity and changes in conformational equilibria of biantennary complex-type N-glycans in water revealed by replica-exchange molecular dynamics simulation. *Biophys. J.* **2011**, *101*, L44−L46.
(24) Yang, M.; Huang, J.; MacKerell, A. D. Enhanced conformational sampling using replica exchange with concurrent solute scaling and hamiltonian biasing realized in one dimension. *J. Chem. Theory Comput.* **2015**, *11*, 2855−2867.
(25) Oshima, H.; Re, S.; Sugita, Y. Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. *J. Chem. Theory Comput.* **2019**, *15*, 5199−5208.
(26) Nishima, W.; Miyashita, N.; Yamaguchi, Y.; Sugita, Y.; Re, S. Effect of bisecting GlcNAc and core fucosylation on conformational properties of biantennary complex-type N-glycans in solution. *J. Phys. Chem. B* **2012**, *116*, 8504−8512.
(27) Patel, D. S.; Pendrill, R.; Mallajosyula, S. S.; Widmalm, G.; MacKerell, A. D. Conformational Properties of α- or β-(1→6)-Linked Oligosaccharides: Hamiltonian Replica Exchange MD Simulations and NMR Experiments. *J. Phys. Chem. B* **2014**, *118*, 2851−2871.
(28) Gil-Ley, A.; Bussi, G. Enhanced Conformational Sampling using Replica Exchange with Collective-Variable Tempering. *J. Chem. Theory Comput.* **2015**, *11*, 1077−1085.
(29) Tribello, G. A.; Gasparotto, P. Using Dimensionality Reduction to Analyze Protein Trajectories. *Front. Mol. Biosci.* **2019**, *6*, 46.
(30) Jolliffe, I.*Principal Component Analysis*; Springer Verlag, 1986.
(31) Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7426−7431.
(32) Tribello, G. A.; Ceriotti, M.; Parrinello, M. Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 5196−5201.
(33) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13023−13028.
(34) Torgerson, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika* **1952**, *17*, 401−419.
(35) Helfrecht, B. A.; Cersonsky, R. K.; Fraux, G.; Ceriotti, M. Structure−property maps with Kernel Principal Covariates Regression. *Mach. Learn.* **2020**, *1*, 045021.
(36) Lageveen-Kammeijer, G. S. M.; de Haan, N. d.; Mohaupt, P.; Wagt, S.; Filius, M.; Nouta, J.; Falck, D.; Wuhrer, M. Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. *Nat. Commun.* **2019**, *10*, 2137.
(37) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353−2370.
(38) 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.* **2007**, *29*, 622−655.
(39) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(40) Jo, S.; Song, K. C.; Desaire, H.; MacKerell, A. D.; Im, W. Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J. Comput. Chem.* **2011**, *32*, 3135−3141.
(41) Park, S.-J.; Lee, J.; Patel, D. S.; Ma, H.; Lee, H. S.; Jo, S.; Im, W. Glycan Reader is improved to recognize most sugar types and chemical modifications in the Protein Data Bank. *Bioinformatics* **2017**, *33*, 3051−3057.
(42) Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUIGlycan Modelerfor modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **2019**, *29*, 320−331.
(43) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.
(44) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(45) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L. d.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.
(46) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. 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.
(47) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.;

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(48) 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.
(49) Guvench, O.; Martin, D.; Greene, M. Pyranose Ring Puckering Thermodynamics for Glycan Monosaccharides Associated with Vertebrate Proteins. *Int. J. Mol. Sci.* **2021**, *23*, 473.
(50) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(51) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(52) Páll, S.; Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **2013**, *184*, 2641−2650.
(53) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(54) Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol. Phys.* **2013**, *112*, 379−384.
(55) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(56) McNaught, A. D.; Wilkinson, A.IUPAC Compendium of Chemical Terminology − The Gold Book. http://goldbook.iupac.org/ (2009), accessed March 10, 2022.
(57) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(58) Bottaro, S.; Gil-Ley, A.; Bussi, G. RNA folding pathways in stop motion. *Nucleic Acids Res.* **2016**, *44*, 5883−5891.
(59) Marshall, N. F.; Coifman, R. R. Manifold learning with bi-stochastic kernels. *J. Inst. Math. Its Appl.* **2019**, *84*, 455−482.
(60) Homans, S. W.; Dwek, R. A.; Boyd, J.; Mahmoudian, M.; Richards, W. G.; Rademacher, T. W. Conformational transitions in N-linked oligosaccharides. *Biochemistry* **1986**, *25*, 6342−6350.
(61) Karplus, M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* **1959**, *30*, 11−15.
(62) Haasnoot, C.; de Leeuw, F. d.; Altona, C. The relationship between proton-proton NMR coupling constants and substituent electronegativities-I. *Tetrahedron* **1980**, *36*, 2783−2792.
(63) Stenutz, R.; Carmichael, I.; Widmalm, G.; Serianni, A. S. Hydroxymethyl Group Conformation in Saccharides: Structural Dependencies of 2JHH, 3JHH, and 1JCH Spin−Spin Coupling Constants. *J. Org. Chem.* **2002**, *67*, 949−958.
(64) Tafazzoli, M.; Ghiasi, M. New Karplus equations for 2JHH, 3JHH, 2JCH, 3JCH, 3JCOCH, 3JCSCH, and 3JCCCH in some aldohexopyranoside derivatives as determined using NMR spectroscopy and density functional theory calculations. *Carbohydr. Res.* **2007**, *342*, 2086−2096.
(65) Thieker, D. F.; Hadden, J. A.; Schulten, K.; Woods, R. J. 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology* **2016**, *26*, 786−787.
(66) Poppe, L.; Stuike-Prill, R.; Meyer, B.; van Halbeek, H. v. The solution conformation of sialyl-α(2→6)-lactose studied by modern NMR techniques and Monte Carlo simulations. *J. Biomol. NMR* **1992**, *2*, 109−136.
(67) Flynn, R. A.; Pedram, K.; Malaker, S. A.; Batista, P. J.; Smith, B. A.; Johnson, A. G.; George, B. M.; Majzoub, K.; Villalta, P. W.; Carette, J. E.; Bertozzi, C. R. Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell* **2021**, *184*, 3109−3124.
(68) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6*, 1722−1734.
(69) Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.* **2021**, *13*, 963−968.
(70) Schjoldager, K. T.; Narimatsu, Y.; Joshi, H. J.; Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 729−749.
(71) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604−1608.
(72) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(73) Thøgersen, H.; Lemieux, R. U.; Bock, K.; Meyer, B. Further justification for the exo-anomeric effect. Conformational analysis based on nuclear magnetic resonance spectroscopy of oligosaccharides. *Can. J. Chem.* **1982**, *60*, 44−57.
(74) Wolfe, S.; Myung-Hwan, W.; Mitchell, D. J. On the magnitudes and origins of the ″anomeric effects″, ″exo-anomeric effects″, ″reverse anomeric effects″, and C—X and C—Y bond -lengths in XCH2YH molecules. *Carbohydr. Res.* **1979**, *69*, 1−26.
(75) Plazinski, W.; Plazinska, A. Molecular dynamics simulations of hexopyranose ring distortion in different force fields. *Pure Appl. Chem.* **2017**, *89*, 1283−1294.
(76) Fadda, E.; Woods, R. J. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discovery Today* **2010**, *15*, 596−609.
(77) Lay, W. K.; Miller, M. S.; Elcock, A. H. Reparameterization of Solute-Solute Interactions for Amino Acid-Sugar Systems Using Isopiestic Osmotic Pressure Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 1874−1882.
(78) Lay, W. K.; Miller, M. S.; Elcock, A. H. Optimizing Solute-Solute Interactions in the GLYCAM06 and CHARMM36 Carbohydrate Force Fields Using Osmotic Pressure Measurements. *J. Chem. Theory Comput.* **2016**, *12*, 1401−1407.
(79) Sauter, J.; Grafmüller, A. Solution Properties of Hemicellulose Polysaccharides with Four Common Carbohydrate Force Fields. *J. Chem. Theory Comput.* **2015**, *11*, 1765−1774.
(80) Fadda, E. Molecular simulations of complex carbohydrates and glycoconjugates. *Curr. Opin. Chem. Biol.* **2022**, *69*, 102175.
(81) Harbison, A. M.; Brosnan, L. P.; Fenlon, K.; Fadda, E. Sequence-to-structure dependence of isolated IgG Fc complex biantennaryN-glycans: a molecular dynamics study. *Glycobiology* **2018**, *29*, 94−103.
(82) Wooten, E. W.; Bazzo, R.; Edge, C. J.; Zamze, S.; Dwek, R. A.; Rademacher, T. W. Primary sequence dependence of conformation in oligomannose oligosaccharides. *Eur. Biophys. J.* **1989**, *18*, 139−148.
(83) Woods, R. J.; Pathiaseril, A.; Wormald, M. R.; Edge, C. J.; Dwek, R. A. The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. *Eur. J. Biochem.* **1998**, *258*, 372−386.
(84) Shahzad-ul-Hussan, S.; Sastry, M.; Lemmin, T.; Soto, C.; Loesgen, S.; Scott, D. A.; Davison, J. R.; Lohith, K.; O'Connor, R.; Kwong, P. D.; Bewley, C. A. Insights from NMR Spectroscopy into the Conformational Properties of Man-9 and Its Recognition by Two HIV Binding Proteins. *ChemBioChem* **2017**, *18*, 764−771.
(85) Fogarty, C. A.; Fadda, E. Oligomannose N-Glycans 3D Architecture and Its Response to the FcγRIIIa Structural Landscape. *J. Phys. Chem. B* **2021**, *125*, 2607−2616.
(86) Canales, A.; Mallagaray, A.; Pérez-Castells, J.; Boos, I.; Unverzagt, C.; André, S.; Gabius, H.; Cañada, F. J.; Jiménez-Barbero, J. Breaking Pseudo-Symmetry in Multiantennary Complex N-Glycans Using Lanthanide-Binding Tags and NMR Pseudo-Contact Shifts. *Angew. Chem., Int. Ed.* **2013**, *52*, 13789−13793.
(87) Plazinski, W.; Drach, M. The influence of the hexopyranose ring geometry on the conformation of glycosidic linkages investigated

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008

using molecular dynamics simulations. *Carbohydr. Res.* **2015**, *415*, 17−27.
(88) Plazinski, W.; Drach, M.; Plazinska, A. Ring inversion properties of 1→2, 1→3 and 1→6-linked hexopyranoses and their correlation with the conformation of glycosidic linkages. *Carbohydr. Res.* **2016**, *423*, 43−48.
(89) Sattelle, B. M.; Almond, A. Shaping up for structural glycomics: a predictive protocol for oligosaccharide conformational analysis applied to N-linked glycans. *Carbohydr. Res.* **2014**, *383*, 34−42.
(90) Sattelle, B. M.; Almond, A. Is N-acetyl-d-glucosamine a rigid 4C1 chair? *Glycobiology* **2011**, *21*, 1651−1662.
(91) Spiwok, V.; Tvaroška, I. Conformational Free Energy Surface of α-N-Acetylneuraminic Acid: An Interplay Between Hydrogen Bonding and Solvation. *J. Phys. Chem. B* **2009**, *113*, 9589−9594.
(92) Ardèvol, A.; Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights fromab InitioQuantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* **2015**, *137*, 7528−7547.
(93) Roy, R.; Poddar, S.; Sk, M. F.; Kar, P. Conformational preferences of triantennary and tetraantennary hybrid N-glycans in aqueous solution: Insights from 20 μs long atomistic molecular dynamic simulations. *J. Biomol. Struct. Dyn.* **2022**, 1−16.
(94) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; Pierre, M. D. L.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **2019**, *16*, 670−673.

https://doi.org/10.1021/acs.jcim.2c01049
J. Chem. Inf. Model. 2022, 62, 4992−5008