Mechanical Compressibility of the Glycosylphosphatidylinositol (GPI) Anchor Backbone Governed by Independent Glycosidic Linkages
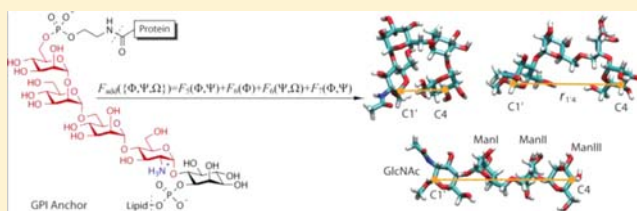
Marko Wehle,[†] Ivan Vilotijevic,[†] Reinhard Lipowsky,[†] Peter H. Seeberger,[†,‡] Daniel Varon Silva,[†] and Mark Santer*,[†]

[†]Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany
[‡]Department of Chemistry and Biochemistry, Free University of Berlin, Arnimallee 22, 14195 Berlin, Germany

Ⓢ *Supporting Information*

ABSTRACT: About 1% of the human proteome is anchored to the outer leaflet of cell membranes via a class of glycolipids called GPI anchors. In spite of their ubiquity, experimental information about the conformational dynamics of these glycolipids is rather limited. Here, we use a variety of computer simulation techniques to elucidate the conformational flexibility of the Man-$\alpha(1{\rightarrow}2)$-Man-$\alpha(1{\rightarrow}6)$-Man-$\alpha(1{\rightarrow}4)$-GlcNAc-α-OMe tetrasaccharide backbone **2** that is an essential

and invariant part of all GPI-anchors. In addition to the complete tetrasaccharide structure, all disaccharide and trisaccharide subunits of the GPI backbone have been studied as independent moieties. The extended free energy landscape as a function of the corresponding dihedral angles has been determined for each glycosidic linkage relevant for the conformational preferences of the tetrasaccharide backbone (Man-$\alpha(1{\rightarrow}2)$-Man, Man-$\alpha(1{\rightarrow}6)$Man and Man-$\alpha(1{\rightarrow}4)$-GlcNAc). We compared the free energy landscapes obtained for the same glycosidic linkage within different oligosaccharides. This comparison reveals that the conformational properties of a linkage are primarily determined by its two connecting carbohydrate moieties, just as in the corresponding disaccharide. Furthermore, we can show that the torsions of the different glycosidic linkages within the GPI tetrasaccharide can be considered as statistically independent degrees of freedom. Using this insight, we are able to map the atomistic description to an effective, reduced model and study the response of the tetrasaccharide **2** to external forces. Even though the backbone assumes essentially a single, extended conformation in the absence of mechanical stress, it can be easily bent by forces of physiological magnitude.

■ INTRODUCTION

Glycosylphosphatidylinositols (GPIs) are complex glycolipids present in eukaryotic cells, typically covalently bound to the C-terminus of proteins via the phosphoethanolamine unit of the GPI molecule (Figure 1).[1] Although they primarily serve to anchor biomolecules to the outer leaflet of the plasma membrane, the complex Man-$\alpha(1{\rightarrow}2)$-Man-$\alpha(1{\rightarrow}6)$-Man-$\alpha(1{\rightarrow}4)$-GlcN-$\alpha(1{\rightarrow}6)$-*myo*-inositol pseudopentasaccharide core and a variety of possible side chains and branches suggests that additional biological functions could be associated with these structures. GPIs are involved in localization of the anchored proteins in membrane microdomains, commonly referred to as lipid rafts.[2] GPI-anchoring brings proteins into close proximity with other raft-associated species and enables their interactions, which underline and determine diverse processes such as signal transduction, cell adhesion, protein trafficking, and antigen presentation.[3−5] These interactions are affected by the properties of the GPI anchor itself, namely its conformational flexibility and dynamics. High lateral mobility and diffusivity of the GPI anchored proteins has been attributed to the rigidity of the GPI anchor and its ability to prevent intermittent interactions with the lipid bilayer.[6] Flexibility of the anchor, on the other hand, is thought to secure access to multiple spatial orientations of the anchored protein and

facilitate its interactions with other membrane-associated molecules.[7]

There is a general consensus that assigning the biological function of an oligosaccharide requires not only a concise structural characterization, but also the overall *dynamic* picture of the carbohydrate topology.[8] Due to a high degree of internal flexibility, conformational dynamics blurs the appearance of a structure for an experimental investigation, making direct characterization of oligosaccharides challenging.[9] Computational studies with atomistic models have been a valuable tool for elucidating some important aspects on glycan dynamics, for example, for interpreting NMR data[10] or exploring possible mechanisms of carbohydrate recognition and binding.[11−13] Nonetheless, numerical characterization of oligosaccharides is burdened by the requirements to sample a large conformational space and to employ explicit aqueous solvent models,[14,15] both of which significantly increase the computational overhead and slow down the global dynamics. This makes extended case studies on larger systems prohibitively expensive.

Only a few numerical studies dealing explicitly with the GPI molecule have been described, but they concisely illustrate the
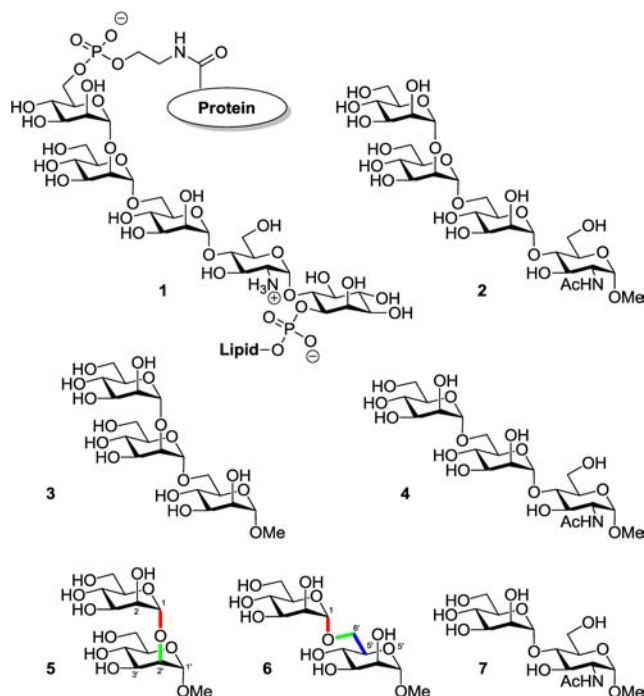
Figure 1. Structure of a simple GPI-anchored protein **1** depicting the GPI core pseudopentasaccharide that connects a protein to the lipid. Methyl glycoside fragments **3**−**7** of the GPI tetrasaccharide backbone **2**. The red and the blue bonds in **5** and **6** define the three glycosidic dihedral angles Φ, Ψ, and Ω, as explained in the text.

difficulties associated with atomistic modeling. A complete GPI anchor of the parasite *Trypanosoma brucei*, including a galactose side chain connected to the pseudopentasaccharide core, was the first structure of this kind to be modeled.[16] Owing to the size of the system, the model which emerged from this study is necessarily an average, static space filling representation of a GPI anchor. A different approach has been taken by Nieto and co-workers who combined experimental results from nuclear magnetic resonance (NMR) studies and molecular dynamics (MD) simulations data for a series of GPI-related compounds to propose a motional model for the GPI core, in the context of adaptive molecular recognition, characterizing it as a hinge.[17,18] Although good agreement with experimental data was achieved, with average simulation times of only a few nanoseconds per compound, the final picture inferred from such simulations has to be considered as preliminary.

Here, we demonstrate that if the general strategy of dividing the glycan topology of interest into a set of tractable substructures is pursued systematically and in an iterative fashion, a reliable and powerful qualitative picture can be derived. Focusing on the tetrasaccharide backbone **2** of the GPI core, we considered the complete set of its di- and trisaccharide fragments (compounds **3**−**7**, Figure 1). Upon an initial cross-compilation of efficient, plain MD runs, we obtain an instructive survey of how conformational preferences of glycosidic linkages can be expected to change across different molecules, and of whether the numerical sampling is sufficiently exhaustive. A selective set of biased MD simulations, acquiring accurate and extensive free energy landscapes for glycosidic angles, is then needed to determine at which point, for example, sampling of the backbone **2** from trajectories becomes reasonably complete.

Only after such an investigation is it possible to conclude rather than to assume that the tetrasaccharide has relatively simple conformational properties which derive directly from its disaccharide constituents and furthermore support the use of a reduced numerical model in which the oligosaccharide is treated as a sequence of independent glycosidic linkages. Exploiting this model, we are able to reconcile the different pictures quoted above. The GPI backbone essentially takes on a single extended conformation in equilibrium, but it is susceptible to forces of physiological magnitude.

■ METHODS

To reduce the computational burden, we have excluded the inositol moiety in **1** from the current study and consequently eliminated the charge associated with protonated amine of GlcN in the remaining tetrasaccharide by addition of an acetyl group. Comparative simulations we carried out on molecules containing amine, protonated amine and acetamide show minimal differences in their conformational dynamics.

Some of the di- and trisaccharide fragments of the GPI are themselves of biological importance and have been the subject of previous structural studies. For example, the trimannose fragment **3** may be present in *N*-linked glycans attached to various extra-cellular proteins;[19,20] the dimannoside **5** is recognized by the lectin concanavalin A[13] and has been investigated by X-ray diffraction[21] and NMR techniques.[22,23] The body of published data will be shown to sufficiently support our simulation results.

Carbohydrate conformations in this study are characterized by the glycosidic dihedral angles within Ramachandran-type plots. Structure **5** in Figure 1 exemplifies the definition of glycosidic angles of an $\alpha(1\rightarrow 2)$ linkage, extending from C1 to C2′; we adhere to the convention as described by Chandrasekaran in which the atoms in the reducing end monosaccharide are labeled with primed numbers.[24] Φ is defined by the sequence (C2,C1,O,C2′) represented in red, the angle around the O−C2′ bond in **5** (green bond in Figure 1) is termed Ψ and defined by (C1,O,C2′,C1′). Compound **6** represents an $\alpha(1\rightarrow 6)$-linkage extending from C1 to C5′, where the additional glycosidic dihedral angle Ω is given by the sequence (O,C6′,C5′,O5′) highlighted in blue. The above definitions have been chosen to enhance the graphical representation of Ramachandran plots. Values for Φ, Ψ and Ω according to strict IUPAC conventions for dihedral torsions in polysaccharides[25,26] are provided as a reference in Figure S7 of the Supporting Information. The three staggered conformations related to Ω are further distinguished as *trans−gauche* (tg), *gauche−trans* (gt) and *gauche−gauche* (gg), where the *gg*-conformation corresponds to Ω = 300°.[27]

We do not *a priori* employ the frequently invoked procedure of dividing an oligo- or polysaccharide into its constituting disaccharide fragments. Our computational investigation, instead, reveals *to what extent* this can be done. The discussion starts with the analysis of disaccharides **5**−**7** followed by the analysis of how the conformational behavior of the three linkages changes, for example, along the series **6**→**3**→**4**→**2**. We employ regular MD simulations at ambient conditions in the NPT-ensemble, biased molecular dynamics and Monte Carlo (MC) simulations. The details about computational methods and the choice of numerical algorithms are provided in the Supporting Information (section A1) along with a comment on the development of force fields for carbohydrates[28−31] (section A2), in particular the GLYCAM[31] force field that is employed here with the simulation engines AMBER[32] and GROMACS.[33−36]

■ RESULTS AND DISCUSSION

Disaccharide Fragments 5 and 7: $\alpha(1\rightarrow 2)$ and $\alpha(1\rightarrow 4)$ Linkages. We illustrate the compilation of Ramachandran-type plots for displaying free energies along with analyzing the $\alpha(1\rightarrow 2)$ linkage of compound **5**. For example, a probability
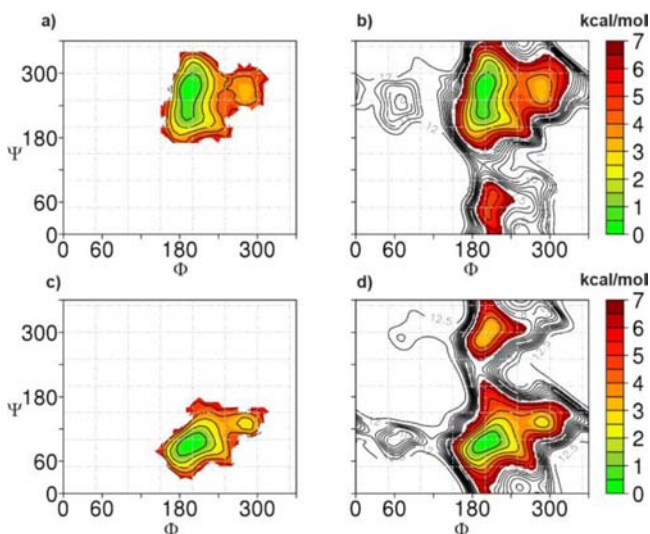
distribution p in the glycosidic dihedral angles (Φ,Ψ) may be acquired from a long MD trajectory as

$$p(\Phi, \Psi) = \langle \delta(\Phi - \sigma_\Phi(\mathbf{r}_1, ..., \mathbf{r}_N))\delta(\Psi - \sigma_\Psi(\mathbf{r}_1, ..., \mathbf{r}_N))\rangle \tag{1}$$

where angular brackets denote the ensemble average and σ expresses the dihedral angle as a function of Cartesian atomic coordinates. This distribution is formally represented as a free energy

$$F(\Phi, \Psi) = -k_B T \log p(\Phi, \Psi) \tag{2}$$

a contour plot of which is depicted in Figure 2a (in the following, all values of glycosidic angles will be reported in units

Figure 2. Free energy landscapes as functions of the glycosidic dihedral angles Φ and Ψ for the $\alpha(1\rightarrow2)$ linked dimannose **5** (a and b, and for the $\alpha(1\rightarrow4)$ linkage in fragment **7** (c and d). (a and c) Estimates of the free energy landscape from a histogram analysis according to eqs 1 and 2, assembled from 200 ns long MD-trajectories with 200.000 data points each. The points are collected into bins of size 10° by 10°; contour lines are obtained by smooth interpolation between data points of the histogram. (b and d) Free energy landscapes from flooding simulations with the flooding time scale τ_F set to 50 ps (see Supporting Information, section A3), the total duration of each ABMD run was 400 ns. The free energy is presented on a grid of 144 × 144 points. The step size of contour lines is 0.5 kcal/mol. Free energy values up to 7 kcal/mol are color coded.

of degrees). It indicates a typical level of 5−7 kcal/mol above the global minimum in F that is reached within a simulation time of 200 ns at 300 K. In order to explore the conformational space more thoroughly, we selectively perform biased MD simulations, based on the adaptive biasing molecular dynamics (ABMD) scheme.[37] The strategy is to find a biasing potential acting on the glycosidic angles in order to achieve their uniform sampling (Supporting Information, section A3). Pictorially, the free energy landscape is "flooded". Figure 2b shows the result of such a simulation. As expected, the two maps agree closely up to the level where the trajectory sampling yields statistically significant contributions to eq 2. Flooding may, in principle, proceed until sampling across the whole angular space becomes uniform although strong steric clashes can lead to steep gradients in the free energy profile and hinder convergence. For the cases considered here, we find it sufficient to conduct flooding out of the main energy wells up to a level of maximally

12−13 kcal/mol, usually covering more than half of the accessible space of glycosidic angles with high accuracy. To facilitate comparison with free energy maps acquired from trajectory sampling, we keep color coding up to 7 kcal/mol and use only contour lines beyond this level.

The behavior of the dimannoside **5** has been described as carrying out fast transitions in Ψ between two low-lying minima that are connected by a low-energy pathway.[38] Plots (a) and (b) in Figure 2 are consistent with this picture showing a minimum in Ψ as an elongated island. In contrast, Φ is more restrained with essentially a single preference at 200°. The force field here correctly reflects the expected *exo*-anomeric effect.[39] Note the secondary minimum along Ψ in plot (b); although accessible in principle, it is usually not visited during a typical simulation run of 200 ns (as depicted in a) because of the large barrier of ∼10 kcal/mol and the topography of the free energy profile, exhibiting a narrow saddle in (b).

The disaccharide **7**, a mannose connected to a glucose via an $\alpha(1\rightarrow4)$ glycosidic linkage, behaves in a qualitatively similar manner. The conformational freedom is also restricted to essentially one major minimum at $(\Phi,\Psi) = (200°,100°)$ shown in plots (c) and (d), Figure 2. The flooding simulation reveals an additional shallow minimum $(\Phi,\Psi) = (200°,300°)$, connected to the main minimum only through a narrow saddle with a barrier between 7.5 and 8.0 kcal/mol. We note that our results for the free energy maps, Figure 2, are very similar to those obtained by Nieto and co-workers,[18] arrived at with different techniques. They find overall agreement with their extensive compilation of NMR studies on saccharides containing these linkages. This is also true for the (Φ,Ψ) map of the $\alpha(1\rightarrow6)$ linkage, to be discussed further.

Fragment 6: $\alpha(1\rightarrow6)$ Linkage. In line with the discussion of conformational behavior of fragments **5** and **7**, the Φ angle of the $\alpha(1\rightarrow6)$ linkage in dimannoside **6** should essentially be restricted to one minimum. Since the additional bond defining the Ω dihedral angle puts more distance between the two pyranose rings, one might expect the flexibility of Ψ to be increased. As can be inferred from Figure 3, Ψ can access two minima with a sufficiently low barrier to allow for frequent transitions, reflected by Ψ as function of time in panel c. The landscapes (a) and (b) closely resemble that obtained by Dowd[40] or Petkowicz[41] by grid searches in vacuum. Up to a level of 4 kcal/mol, the agreement between trajectory data and the flooding simulation is nearly quantitative. In contrast, more pronounced variations for F are found across repetitive MD runs, especially with respect to the infrequently sampled region of $\Omega < 200°$ in the (Ψ, Ω) free energy landscape (Figure 4a, left chart). Statistical fluctuations for the $\alpha(1\rightarrow6)$ linked fragment **6** among different runs are strongly influenced by the dynamics of Ω. Dwell times between the rotamer states *gg*, *gt* and *tg* can be fairly long, up to several tenths of nanoseconds. In Figure 4 we compare $\Omega(t)$ of two independent runs of 200 ns length, starting in the *gg* (b) and the *tg* (c) rotamer (for corresponding free energy landscapes for each starting conformation see Figure S2 in the Supporting Information). The estimate for the limiting rotamer populations is, however, quite robust and deserves some discussion. It may be expected that the conformational behavior of Ω within a $\alpha(1\rightarrow6)$ linkage is derived from that of a free exocyclic hydroxymethyl group. Populations of the exocyclic Ω of around 60%/40% for *gg*/*gt* are reported if the orientation of the hydroxyl group at C4 is equatorial (as in mannose or glucose); for an axial orientation (as in galactose) the *gt* rotamer is favored and the *gg*
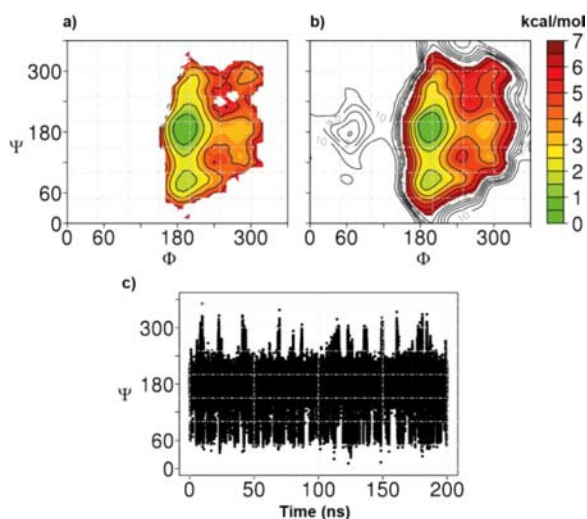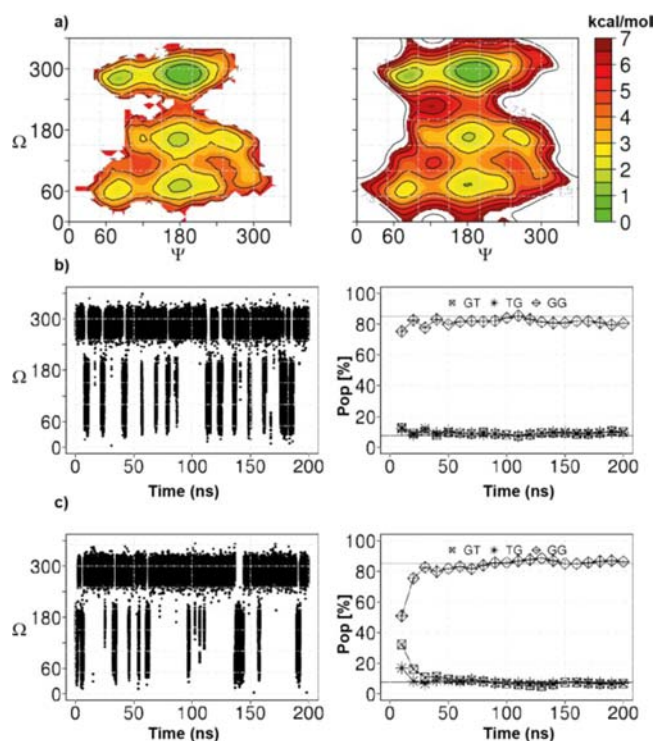
Figure 3. Conformational behavior of the dihedral angles (Φ,Ψ) within the $\alpha(1\rightarrow6)$ linkage of the disaccharide fragment **6**. (a) Free energy map assembled from a 200 ns MD-trajectory (200.000 data points), with the rotamer *gg* as the starting conformation for the third dihedral angle Ω. (b) Corresponding flooding simulation with $\tau_F = 50$ ps and total runtime of 400 ns. (c) Trajectory of Ψ showing frequent transitions between the two major free energy minima at $(\Phi,\Psi) = (180°,180°)$ and $(180°,90°)$.

Figure 4. (a) (Ψ,Ω) free energy landscapes of the $\alpha(1\rightarrow6)$ linked dimannose **6**, acquired from a 200 ns trajectory starting in the *gg* conformation (left); from a flooding simulation with $\tau_F = 50$ ps and 400 ns total runtime (right). We can compute populations of the three rotamer states of Ω from the data presented by inverting eq 2 taking the values of the angle Ω from 240° to 360° as *gg*, from 0° to 120° as *gt* and from 120° to 240° *tg*. We obtain 85% for the *gg* rotamer at 290°, and 7.5% for *gt* and *tg*, respectively (using the ABMD data). (b and c) Trajectories of Ω for the disaccharide **6** from a 200 ns MD run (left) and the cumulative rotamer populations binned into segments of 10 ns length (right), with (b) *gg* or (c) *tg* as the start conformation. Horizontal lines depict the populations computed from (a).

conformation is suppressed.[15,42,43] Within a glycosidic linkage, however, the situation is less clear. Although we can confirm all expected rotamer distributions on exocyclic Ω angles, within the glycosidic linkage this dihedral angle behaves differently. The *gg* rotamer prevails, whereas *tg* and *gt* are equally but only weakly populated.

In *N*-linked glycans containing the $\alpha(1\rightarrow6)$ linkage of the dimannoside **6**, the distribution among the Ω rotamers depends on steric interactions with other parts of the glycan[44] or with the polypeptide/protein to which it is attached.[45] A comprehensive experimental investigation of **6** by a combined MD and rotating frame nuclear Overhauser enhancement (ROE) spectroscopy approach, however, found that the *gg* conformation should indeed be favored by more than 90%.[46] In recent computational[38] and experimental[53] studies, the authors arrived at varying conclusions.

It cannot be said, however, whether mannoses are in any respect special for the GPI backbone or not. Perić-Hassler et al.[47] have compiled the conformational characteristics of a series of glucose based disaccharides, using a specially adapted version of the GROMOS force field.[29] They also make extensive use of biasing techniques to enhance sampling of glycosidic dihedrals. There are indeed some close similarities. For instance, the free energy landscape of our $(1\rightarrow4)$ linkage is in almost quantitative agreement with its close relative maltose; kojibiose is very similar to our $(1\rightarrow2)$ linkage, but the flexibility in Ψ is somewhat reduced. Isomaltose qualitatively agrees with the $(1\rightarrow6)$ linkage in (Φ,Ψ) space, but not in Ω. There *gg* is dominating as well, but the *gt* conformer is also significantly populated.[47]

Global Conformation of the Tetrasaccharide Backbone. A natural and frequently invoked simplification which gives rough predictions of global glycan conformations is to assume that the glycan is composed of independent glycosidic linkages, the conformational characteristics of which are inferred from corresponding disaccharides. For the special case at hand, interactions between distant parts of the oligosaccharide may suspend the validity of this *additivity hypothesis* (for a brief discussion on this issue, see Supporting Information, section B). Indeed, for the tetrasaccharide **2** this possibility cannot be excluded (Figure 5). A rigorous conformational analysis by acquiring multidimensional free energy maps as employed in the previous section becomes increasingly complicated if many variables are considered simultaneously, although it is in principle possible with the algorithm employed here.

For assembling free energy landscapes with multidimensional histograms or grids, sampling quickly becomes poor and might require excessive runtimes. Visualization of data also becomes difficult.[48] Even a restricted projection on all seven glycosidic angles of **2** is clearly prohibitive. To analyze its conformational behavior, we therefore make use of simulation data compiled from the collection of fragments **3** to **7** in Figure 1.

We first performed a series of regular MD runs on the remaining compounds **3**, **4** and **2**, each 200 ns in length, starting from each of three rotamer states of Ω as initial configuration. By inspection of the combined data set, one can infer that coarse conformational characteristics of disaccharides **5**-**7** survive within the compounds **2**−**4**. In particular, we see a stably weak population of the Ω *tg* and *gt* rotamer states. With respect to conformational details and sampling accuracy, we
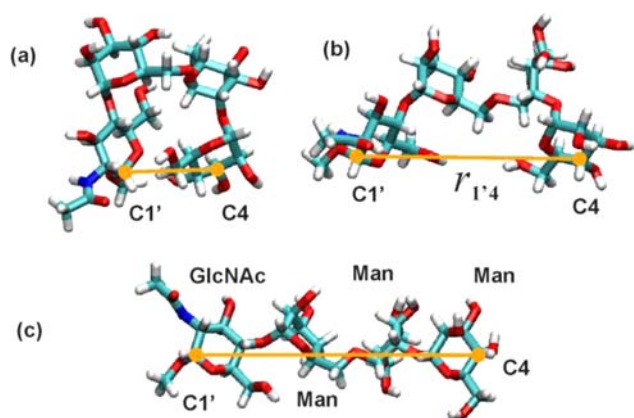
Figure 5. Three configurations of the tetrasaccharide backbone **2**. Frequently visited conformations are similar to the C-like shape in (b), for which each glycosidic dihedral angle pair is close to its global minimum within the corresponding disaccharide fragment. In this snapshot, the distance $r_{1'4}$ equals 12.7 Å. During a typical 200 ns long trajectory, one observes substantial excursions from this conformation, as shown in (a) and (c). The distance $r_{1'4}$ between carbon atoms C4 and C1′ (signified by the solid orange lines, all parallel to the plane of the paper) can be as short as 5 Å in (a) and as long as 16 Å in (c).

Figure 6. Conformational behavior of the glycosidic dihedral angles Φ, Ψ and Ω of the tetrasaccharide backbone **2**. (a) Time evolution of dihedral angle Ω recorded over 1 μs. (b) Free energy landscape for the pair (Ψ,Ω) compiled from a trajectory of 1 μs with 10^6 data points.

clearly recognize a distinction between the linkages $\alpha(1\rightarrow2)$ of **5** and $\alpha(1\rightarrow4)$ of **7** on one side and the $\alpha(1\rightarrow6)$ linkage of **6** on the other. The behavior of $\alpha(1\rightarrow2)$ and $\alpha(1\rightarrow4)$ suffers only minimal changes (see Supporting Information, Figure S3). This indicates that disaccharide units **5** and **7** as part of the tetrasaccharide **2** can be represented as two independent entities kept apart by the $\alpha(1\rightarrow6)$ linkage, with low mutual interaction. This seems plausible as conformational freedom of both **5** and **7** is restricted to a relatively small region in (Φ,Ψ) space. In contrast, the $\alpha(1\rightarrow6)$ linkage shows significant scatter in the free energy, particularly with respect to the (Ψ,Ω) pair. This is consistent with the observed increase in dwell times for Ω in each of the three rotamer states as we move from the disaccharide **6** to the tetrasaccharide **2**. Ψ and Ω can probe a relatively wide range of conformational space and rotations around these glycosidic angles involve a considerable change in orientation or position of the saccharide units they connect, leading to viscous damping. The slowing in dynamics of these angles indicates their insufficient sampling, and changes noticed in conformational behavior must be confirmed otherwise. For instance, by evaluating the simulation data along the series **6**→**3**→**4**→**2**, a slight increase in population for the *tg* rotamer in Ω and the secondary minimum in Ψ is suggested. By carrying out a selected, numerically expensive flooding simulation on the (Ψ,Ω) pair of **2**, we can estimate at which point a long MD trajectory will provide sufficient statistics, see Figure 6 (here starting from the *gg* conformation). We find that from ~1 μs onward, the sampling can be considered reasonably complete: additional MD runs of 1 μs duration, starting from the *gt* and *tg* rotamer, respectively, lead to equivalent results. The slight shifts in populations of Ω and Ψ can now be confirmed, for each starting configuration the distribution in these angles is reproduced almost identically (see Supporting Information, Figure S4). Note that dwell times in different rotamer states can range from several 100 ps up to 100 ns (in *gg*).

It is useful to summarize what has been achieved by the spreadsheet procedure outlined above: (i) a series of relatively long but manageable MD runs on the full tetrasaccharide **2** and the ensemble of substructures **3−7** characterize **2** as composed of two disaccharide building blocks, (ii) the connecting $\alpha(1\rightarrow6)$ linkage is seen to be most sensitive to the overall composition of the saccharide containing it, especially with respect to sampling efficiency, (iii) for the slow degrees of freedom, the sampling efficiency from trajectories of **2** could be evaluated based on only one selected biased MD simulation. That is, the data sets from the 1 μs long MD runs may now be considered to represent all of the relevant phase space/conformational space. This is important, as the flooding mechanism projects information only on the selected set of variables, such as the pair (Ψ,Ω).
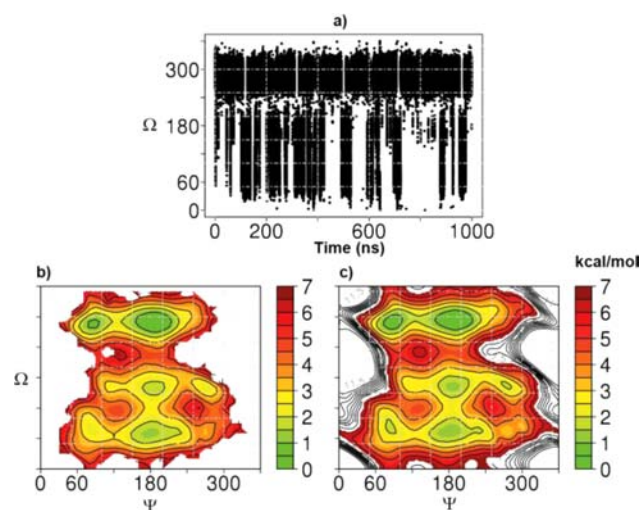
Having verified that conformational preferences of **2** per linkage are essentially inherited from disaccharide fragments, we can finally turn to the question to what extent different linkages are also independent. For this, we estimated probability distributions p of the desired observables by assembling normalized histograms from data of the μs MD trajectories. For instance, we find that glycosidic angles Θ_1 and Θ_2 that belong to different linkages are independent: their joint distribution $p(\Theta_1, \Theta_2)$ can, to a good approximation, be represented as a simple product $p(\Theta_1, \Theta_2) \approx p(\Theta_1)p(\Theta_2)$. Note that a one-point probability distribution such as, for example, $p(\Theta_1)$, can always be obtained exactly by integrating the two-point distribution $p(\Theta_1, \Theta_2)$ over Θ_2; but this in general implies a loss of information: $p(\Theta_1)p(\Theta_2)$ might not resemble $p(\Theta_1, \Theta_2)$ at all if Θ_1 and Θ_2 are strongly correlated.

A product decomposition usually fails for pairs of dihedral angles of the same glycosidic linkage due to the interaction of the neighboring carbohydrate rings. For a $\alpha(1\rightarrow6)$ linkage with three consecutive glycosidic angles one would, therefore, expect a genuine three-point function $p(\Phi,\Psi,\Omega)$ to be required, but this is surprisingly not so. In fact, by computing one- and two-point functions from ABMD data of the disaccharide **6** (e.g., from the data in Figures 3b and 4a and solving eq 2 for p), we find that a product of the form $p(\Phi)p(\Psi,\Omega)$ approximates $p(\Phi,\Psi,\Omega)$ quite well up to a level of ~5 kcal/mol, as can be inferred by inspecting the explicitly three-dimensional free energy landscapes assembled from a μs-trajectory of both, the disaccharide **6** and the tetrasaccharide **2**, compared to a map

corresponding to the above product decomposition (Supporting Information, Figure S5).

The particular product decomposition was motivated by exploring the free energy landscape of the pair Ω, Φ. Figure 7
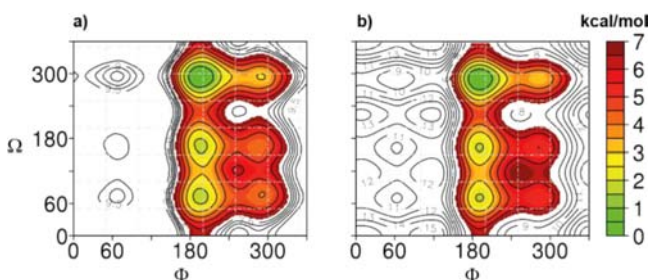
Figure 7. Free energy landscapes of the glycosidic dihedral angles (Φ,Ω) for the disaccharide **6** as obtained from (a) a flooding simulation providing $F(\Phi,\Omega)$ directly and (b) approximated by the product $p(\Phi)p(\Omega)$ and applying eq 2: $p(\Phi)$ and $p(\Omega)$ have been obtained by integrating $p(\Phi,\Omega) \approx \exp(-F(\Phi,\Omega)/k_BT)$ over Ω and Φ, respectively. The Ω-rotamer populations are ($gg/gt/tg$) = (85.4/6.7/7.9%) for (a) and (85.5/7.1/7.4%) for (b), respectively. The bounds defining the three rotamers gg, gt and tg are as in Figure 4.

illustrates that the product $p(\Phi)p(\Omega)$, which can be obtained from $p(\Phi)p(\Psi,\Omega)$ by integrating over Ψ, is in almost perfect agreement with the corresponding free energy landscape for this pair acquired from ABMD data. In summary, we can now with confidence represent the free energy of the tetrasaccharide core **2**, as a function of all glycosidic angles $\{\Phi,\Psi,\Omega\}$, as an additive function $F_{add}(\{\Phi,\Psi,\Omega\})$ as follows:

$$F_{add}(\{\Phi, \Psi, \Omega\}) = F_5(\Phi, \Psi) + F_6(\Phi) + F_6(\Psi, \Omega)$$
$$+ F_7(\Phi, \Psi) \qquad (3)$$

where the subscripts indicate the disaccharide, the simulation data (flooding or plain MD) of which is employed to represent the corresponding F. The possibility to decompose the free energy landscape $F_6(\Phi,\Psi,\Omega)$ into the two terms in eq 3 is remarkable, but also of practical importance: the acquisition of free energies as a function of only one or two variables is still relatively fast, and quite extensive landscapes can be compiled. This will prove convenient for exploiting a reduced model for the tetrasaccharide backbone, as suggested below.

We should note that the approximation of a product decomposition as described above (both, independent disaccharide linkages and the special case of the $1\rightarrow6$ linkage) cannot be automatically expected to be a rule. However, carbohydrate moieties are usually well hydrated, and frequent intramolecular hydrogen bridges (that would render a product decomposition useless) are likely to be screened away by the solvent. Oligosaccharides are rather known not to fold into compact structures in general (as polypeptides do). For this reason, additivity can be considered a useful starting point, even if the global structure of an oligosaccharide is seen to deviate strongly from a set of independent glycosidic linkages. It would rather help identifying unusual interactions between distant glycan moieties. Relating to the latter statement, the results arrived at in this section justify the exclusion of the inositol moiety from the present study. The inositol would, as depicted in Figure 1, be appended via an equatorial hydroxyl group and be distal to the tetrasaccharide part. The conformational preferences of this trailing glycosidic linkage should be similar

to our ($1\rightarrow2$) or ($1\rightarrow4$) linkages, with essentially one major combination of dihedral angles. This was also found by Nieto and co-workers.[18] Preliminary results (MD trajectories, 160 ns) on a version of the complete structure **1** that we have parametrized (including the phosphoinositol) indeed sustain the prediction that the inositol does not change the overall conformational characteristics. As expected, we observe a further slowing down of the kinetics upon appending this additional residue to the tetrasaccharide.

Reduced Backbone Model and Response to External Forces. Relying on the results of the previous section, it is now straightforward to reduce the topology of the tetrasaccharide backbone such that an efficient conformational analysis can be accomplished. We first select a suitable small subset of atoms defining a sequence of bonds, angles and dihedral angles (Figure 8a). In this sparse geometry, the carbohydrate rings are
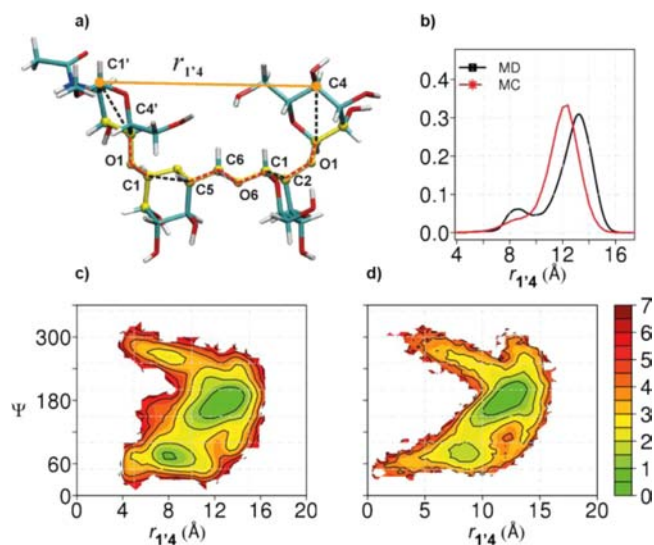
Figure 8. (a) Translation of the all-atom tetrasaccharide backbone **2** into a reduced topology, defined by the labeled atoms along the dashed path. Atom names denote elements together with their ring position (1−6) within the monosaccharide unit. Dihedral torsion is allowed only around red path segments. Dashed black path segments represent sugar rings as effective bonds between two consecutive glycosidic linkages. All atoms kept in yellow provide sufficient information to define orientation and position of the carbohydrate rings. (b) Distribution $p(r_{1'4})$ for the end-to-end distance from MD data (black) and the MC result (red). (c) Free energy landscape for the pair $r_{1'4}$ and Ψ created from the MD data set as in (a). (d) Corresponding free energy landscape obtained from an MC-simulation.

now represented by bonds around which no torsion is allowed.[49] The only remaining degrees of freedom are torsions around glycosidic bonds, that is, the degrees of freedom dominating the global flexibility of the molecule. We further retain 4 additional ring atoms in order to define these dihedral angles and to define the orientation of the sugar rings. F_{add} of eq 3 defines the (free) energy landscape of the reduced model. F_5, F_7 and $F_6(\Psi,\Omega)$ are directly taken from the ABMD data displayed in Figure 2b, 2d and Figure 4a (right panel), respectively. $F_6(\Phi)$ is computed from the data displayed in Figure 3b, as described above.

F_{add} can be sampled by, for example, stochastic dynamics or a Monte Carlo (MC) procedure. We choose the latter approach, as we find that MC sampling most efficiently covers the

18969

dx.doi.org/10.1021/ja302803r | *J. Am. Chem. Soc.* 2012, 134, 18964−18972

relevant space of all glycosidic angles. Trial moves are carried out successively and independently according to each term in eq 3. They consist of random increments of the constituting dihedral torsions with a maximum step size of 30°, and a corresponding change ΔF_{add} in free energy. Note that the $\alpha(1\rightarrow6)$ linkage can now be treated with two independent trial moves: one for a rotation around Φ and one in the two-dimensional space of (Ψ,Ω). The acceptance probability is given by the Metropolis function min$[1, \exp(-\Delta F_{add}/k_B T)]$, resulting in an average acceptance rate of roughly 30%. Moves leading beyond 360° or 0° are treated by applying periodic boundary conditions. By repeatedly carrying out the MC procedure with the same initial geometry, we find that 10^6 MC trials are sufficient to reach convergence, and this number has been adopted in all subsequent simulations. For the initial geometry, we selected a configuration of **2** during a solvated MD run with Ω in the gg conformation, and all other dihedral angles close to their respective free energy minima. This configuration was then relaxed (energy minimized) with the method of steepest descent for an additional 500 steps, and adopted the initial geometry from which positions for all atoms involved in the reduced model were taken. In the sampling procedure described, atomic coordinates change only through dihedral rotations; bond lengths and angles remain fixed.

To test the performance of the reduced model simulated with the MC method compared to a full MD simulation, we consider the distribution of the distance $r_{1'4}$ between atoms C1′ and C4 (Figure 8b). Both approaches yield essentially one major conformation with $r_{1'4}$ at roughly 13 Å. The shift of the MC result toward shorter distances is mainly due to the relative rigidity of the reduced model. It varies with the particular initial geometry chosen and deviations of up to 1 Å appear to be typical. In contrast, the width of the major maximum is well reproduced by the MC results, simply reflecting that the glycosidic dihedral angles dominate the overall shape of the molecule. MC and MD also agree in the prediction of the minor maximum at ∼8 Å. The presence of this maximum can be related to the variation in Ψ of the $\alpha(1\rightarrow6)$ linkage. Plots (c) and (d) in Figure 8 show free energy landscapes of the pair of variables $r_{1'4}$ and Ψ. Note that the second maximum in the MC data is less pronounced and almost appears as a shoulder. This can be explained by the difference in the population of Ψ, the secondary maximum of which is enhanced in the tetrasaccharide compared to the disaccharide **6** entering the reduced model (Supporting Information, Figure S4). We note that for a remote pair of atoms such as C1′, C4, the distance $r_{1'4}$ with the particular topology of the GPI tetrasaccharide backbone depends only weakly on Ω, and hence only weakly on the distribution of the $gg/gt/tg$ rotamers (Supporting Information, Figure S6). Ω mainly contributes to the broadening of the main maximum in Figure 8b. Thus we may conclude that the GPI backbone assumes essentially one extended conformation, with a residual flexibility given by the variations in the Ψ glycosidic angle of the $\alpha(1\rightarrow6)$ linkage.

The latter conclusion can be made more precise. The free energy landscapes acquired with the ABMD method cover a broader region in the space of glycosidic angles, and with higher overall accuracy than a typical MD simulation. We may exploit this in order to probe the response of the reduced model to the action of weak externally applied forces that would shift the probability distributions of glycosidic angles away from their equilibrium positions. Such forces may arise, for instance, during cell−cell contact or association of the full GPI anchor

with rafts or ligands. We can expect that glycosidic linkages dominantly contribute to shape changes of the molecule at small loads, because bond stretching and angle bending are both usually governed by much higher force constants than torsions. To emulate mechanical action on the molecule, we apply a pair of oppositely pointing forces at C1′ and C4. If, for simplicity, we assume that the forces are directed along the line connecting these atoms, they give rise to a tension f (with units of force) of equal magnitude. For instance, a typical value of 5 pN corresponds to $1.2k_B T$/nm. The sign of f distinguishes whether the molecule is stretched or compressed: when adding a term of the form $F_{tens} = f\,r_{1'4}$ to F_{add} in eq 3, a negative value of f will lead to a decrease of the total free energy if a change $\Delta r_{1'4}$ during a MC trial is positive, and $r_{1'4}$ tends to increase. Likewise, a positive f will lead to compression.

The results are summarized in Figure 9, where f is chosen to be within the physiologically relevant range -15 pN $\le f \le 15$
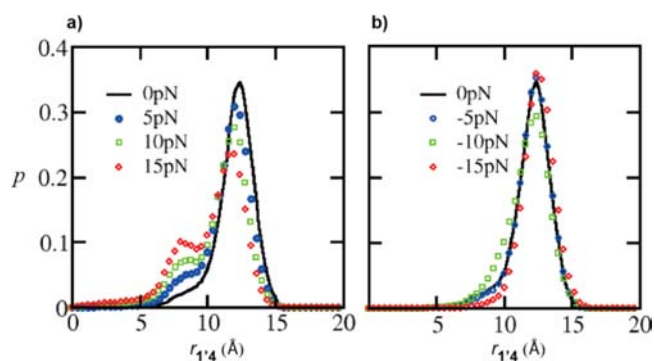
Figure 9. Probability distribution $p(r_{1'4})$ for the tetrasaccharide under (a) compression and (b) tension. Corresponding values for f are indicated in the two panels.

pN. The tetrasaccharide is seen to be more susceptible to compression than to stretching. This is consistent with Figure 8, suggesting that a reduction in length should correspond to a shift in population between the two main free energy minima in Ψ. Under tension, the main minimum at $\Psi \approx 180°$ gradually becomes the only one populated and up to -15 pN, there is no significant average change in length.

It is important to reiterate that we did not initially rely on the additivity hypothesis, which *a posteriori* proved to work so well. Conformational preferences of glycosidic linkages can change significantly when they are studied in different topologies. Landström and Widmalm have recently shown that this may happen even when going from a di- to a trisaccahride.[50] The authors clearly emphasize the need for careful sampling in an MD study when comparing to experimental results. The important point in our work is that along with the computational strategy chosen, such nontrivial behavior would have been detected automatically. A corresponding reduced model could then be derived incrementally, with additive glycosidic free energy landscapes as the initial approximation, improvements upon which may include steric interactions between whole carbohydrate rings,[51] or inter-actions between parts of the molecule that can be grouped into effective force centers ("beads"), as in systematic coarse-grain strategies that are commonly applied to more homogeneous systems.[52]

CONCLUSIONS

We have performed a thorough conformational analysis and a concise characterization of the tetrasaccharide backbone **2** of the GPI core. A spreadsheet compilation of data from regular MD trajectories covering the complete collection of substructures provides a survey of the overall conformational character of the target structure **2**. Even for this coarse analysis, we find that run times of at least 200 ns must be employed to sufficiently account for degrees of freedom that are rather difficult to sample. We have also performed biased MD simulations on a selected set of substructures. The biased dynamics permit us to explore free energy landscapes of glycosidic angles far more extensively than with plain MD runs. With biased MD, we reach levels of 12−13 kcal/mol, and additional, high energy metastable minima in solution for the $\alpha(1{\rightarrow}2)$ linked dimannose **5** and for the $\alpha(1{\rightarrow}4)$ linked fragment **7** could be identified. The landscapes generated from histograms over MD trajectories agree quantitatively with the biased dynamics up to 4−5 kcal/mol.

The spreadsheet analysis clearly identifies the $\alpha(1{\rightarrow}6)$ linkage as critical with respect to sampling efficiency and accuracy. For the tetrasaccharide, a comparison with an ABMD simulation reveals that only run times of 1 μs and above will provide a sufficient sampling. Corresponding data sets from regular MD runs were then used, in combination with the previous results, to verify that the tetrasaccharide can indeed be viewed as a sequence of independent glycosidic linkages, the conformational preferences of which are essentially inherited from disaccharide substructures. This finally permits mapping to a reduced model, in which extensive free energy landscapes from biased MD runs were employed to probe the response of the molecule to external loads. We emphasize the simplicity of the obtained picture: the tetrasaccharide backbone consists of two rigid disaccharide units that are connected by a 1−6 linkage, together forming a rather stiff rod with a residual bending elasticity conveyed by the central Ψ glycosidic dihedral angle. Thus, a reconciliation of the two opposing intuitive pictures of the GPI anchor backbone raised in the introduction is indeed possible by characterizing the backbone as mechanically compressible.

The computational effort expended, although it may appear discouraging at first sight, is fully justified. Our differential approach provides insights into the nature of glycan structural dynamics and we reveal a simplicity of the behavior of the GPI tetrasaccharide backbone **2** that might be difficult to obtain otherwise. The simple reduced model obtained here illustrates how one can devise a reliable mapping strategy based on fully atomistic simulations.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional figures that further support conclusions from the main text and the details of numerical algorithms and used software are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

mark.santer@mpikg.mpg.de

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Paulick, M. G.; Bertozzi, C. R. *Biochemistry* **2008**, *47*, 6991.

(2) Simons, K.; Toomre, D. *Nat. Rev. Mol. Cell Biol.* **2000**, *1*, 31.

(3) Low, M. G.; Saltiel, A. R. *Science* **1988**, *239*, 268.

(4) Ferguson, M. A. J. *J. Cell Sci.* **1999**, *112*, 2799.

(5) Maeda, Y.; Kinoshita, T. *Prog. Lipid Res.* **2011**, *50*, 411.

(6) Paulick, M. G.; Forstner, M. B.; Groves, J. T.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20332.

(7) Paulick, M. G.; Wise, A. R.; Forstner, M. B.; Groves, J. T.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2007**, *129*, 11543.

(8) DeMarco, M. L.; Woods, R. J. *Glycobiology* **2008**, *18*, 426.

(9) Johannes, J. F. G. In *NMR Spectroscopy and Computer Modeling of Carbohydrates*, Vol. 930; American Chemical Society: Washington, D.C., 2006; pp 1−19.

(10) Engelsen, S. B.; du Penhoat, C. H.; Pérez, S. *J. Phys. Chem.* **1995**, *99*, 13334.

(11) Brás, N. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. *J. Int. J. Quantum Chem.* **2008**, *108*, 2030.

(12) Feinberg, H.; Mitchell, D. A.; Drickamer, K.; Weis, W. I. *Science* **2001**, *294*, 2163.

(13) Moothoo, D. N.; Canan, B.; Field, R. A.; Naismith, J. H. *Glycobiology* **1999**, *9*, 539.

(14) Almond, A.; Sheehan, J. K. *Glycobiology* **2003**, *13*, 255.

(15) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541.

(16) Homans, S. W.; Edge, C. J.; Ferguson, M. A. J.; Dwek, R. A.; Rademacher, T. W. *Biochemistry* **1989**, *28*, 2881.

(17) Chevalier, F.; Lopez-Prados, J.; Groves, P.; Pérez, S.; Martín-Lomas, M.; Nieto, P. M. *Glycobiology* **2006**, *16*, 969.

(18) Chevalier, F.; Lopez-Prados, J.; Pérez, S.; Martín-Lomas, M.; Nieto, P. M. *Eur. J. Org. Chem.* **2005**, 3489.

(19) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832.

(20) Woods, R. J.; Pathiaseril, A.; Wormald, M. R.; Edge, C. J.; Dwek, R. A. *Eur. J. Biochem.* **1998**, *258*, 372.

(21) Srikrishnan, T.; Chowdhary, M. S.; Matta, K. L. *Carbohydr. Res.* **1989**, *186*, 167.

(22) Peters, T. *Liebigs Ann. Chem.* **1991**, 135.

(23) Helander, A.; Kenne, L.; Oscarson, S.; Peters, T.; Brisson, J. R. *Carbohydr. Res.* **1992**, *230*, 299.

(24) Rao, V. S. R.; Qasba, P. K.; Balaji, P. V.; Chandrasekaran, R. *Conformation of Carbohydrates*; Harwood Academic: Chur, Switzerland, 1998.

(25) Moss, G. P. *Pure. Appl. Chem.* **1996**, *68*, 2193.

(26) Eur., J. *Biochem.* **1983**, *131*, 5.

(27) The letter combinations *tg/gt/gg* codify the angles the bonds C5−O5 and C5−C4 make with the C6−O6 bond in a staggered projection, when looking in the direction of the C5−C6 bond. The first letter signifies whether O5 is gauche or trans with respect to O6, and the second refers to C4.

(28) Kräutler, V.; Müller, M.; Hünenberger, P. H. *Carbohydr. Res.* **2007**, *342*, 2097.

(29) Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400.

(30) MacKerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584.

(31) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622.

(32) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L., III; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER*; University of California: San Francisco, 2010.

(33) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.

(34) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.

(35) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.

(36) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(37) Babin, V.; Roland, C.; Sagui, C. *J. Chem. Phys.* **2008**, *128*, 134101 .

(38) Bernardi, A.; Colombo, A.; Sánchez-Medina, I. *Carbohydr. Res.* **2004**, *339*, 967.

(39) Allinger, N. L. In *Molecular Structure*; John Wiley & Sons, Inc.: New York, 2010; pp 189−211.

(40) Dowd, M. K.; French, A. D.; Reilly, P. J. *J. Carbohydr. Chem.* **1995**, *14*, 589.

(41) Petkowicz, C. L. O.; Reicher, F.; Mazeau, K. *Carbohydr. Polym.* **1998**, *37*, 25.

(42) González-Outeiriño, J.; Kirschner, K. N.; Thobhani, S.; Woods, R. J. *Can. J. Chem.* **2006**, *84*, 569.

(43) Marchessault, R. H.; Pérez, S. *Biopolymers* **1979**, *18*, 2369.

(44) Wooten, E. W.; Bazzo, R.; Edge, C. J.; Zamze, S.; Dwek, R. A.; Rademacher, T. W. *Eur. Biophys. J.* **1990**, *18*, 139.

(45) Brisson, J. R.; Carver, J. P. *Biochemistry* **1983**, *22*, 3680.

(46) Spronk, B. A.; Rivera-Sagredo, A.; Kamerling, J. P.; Vliegenthart, J. F. G. *Carbohyd. Res.* **1995**, *273*, 11.

(47) Perić-Hassler, L.; Hansen, H. S.; Baron, R.; Hünenberger, P. H. *Carbohydr. Res.* **2010**, *345*, 1781.

(48) Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. *J. Chem. Theory Comput.* **2010**, *6*, 35.

(49) Perico, A.; Mormino, M.; Urbani, R.; Cesaro, A.; Tylianakis, E.; Dais, P.; Brant, D. A. *J. Phys. Chem. B* **1999**, *103*, 8162.

(50) Landström, J.; Widmalm, G. *Carbohydr. Res.* **2010**, *345*, 330.

(51) Bathe, M.; Rutledge, G. C.; Grodzinsky, A. J.; Tidor, B. *Biophys. J.* **2005**, *88*, 3870.

(52) Hynninen, A. P.; Matthews, J. F.; Beckham, G. T.; Crowley, M. F.; Nimlos, M. R. *J. Chem. Theory Comput.* **2011**, *7*, 2137.

(53) Olsson, U.; Säwén, E.; Stenutz, R.; Widmalm, G. *Chem.—Eur. J.* **2009**, *15*, 8886.

18972

dx.doi.org/10.1021/ja302803r | *J. Am. Chem. Soc.* 2012, 134, 18964−18972