

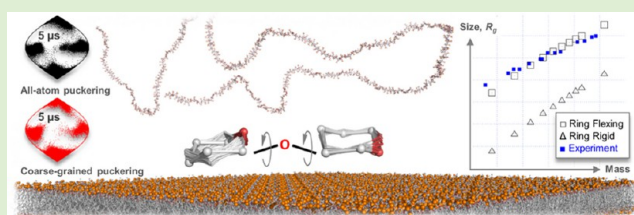
# Does Microsecond Sugar Ring Flexing Encode 3D-Shape and Bioactivity in the Heparanome?

Benedict M. Sattelle, Javad Shakeri, and Andrew Almond\*

Faculty of Life Sciences, University of Manchester, Manchester Institute of Biotechnology, 131 Princess Street, Manchester, M1 7DN, United Kingdom

## S Supporting Information

**ABSTRACT:** The biological information encoded in carbohydrate sequences dwarfs that of proteins and nucleic acids. Deciphering structure–function relationships in heparin and heparan sulfate (the heparanome) is further compounded by extreme sequence diversity, experimental difficulties, and the computational cost of rigorous modeling. Here we perform unbiased microsecond dynamics simulations of 11 heparanome oligosaccharides (55 microseconds total) to investigate the effect of sequence on 3D-structure and to underpin a coarse-grained model that is consistent with long-time scale experimentally validated atomic motions in water. Pyranose ring flexing (puckering) in 2-*O*-sulfo- $\alpha$ -L-iduronic acid, which underlies heparin-mediated anticoagulation, was modulated by polymerization (chain position and adjacent residues), which is supported by previous experiments. Furthermore, in coarse-grained simulations, inclusion of puckering was essential to predict macroscopic hydrodynamic properties of heparan sulfate chains containing hundreds of monosaccharaides. Our structural findings and model enable rational molecular design, and we propose that, in the heparanome, puckering, polymer 3D-shape, and bioactivity are inextricably linked.



## INTRODUCTION

Complex carbohydrates perform key roles in organismal gene expression, development, defense and cellular communication.<sup>1–7</sup> In contrast to proteins and nucleic acids, the branching and extensive chemical decoration of carbohydrates represent a vast and enigmatic code that underlies their bioactivities.<sup>8</sup> Mammalian heparan sulfate and structurally related heparin (Figure 1) serve to exemplify how sequence variability in carbohydrate chains has evolved to encode complex biological information and discrete functions. Although initially synthesized as a common repeating disaccharide, heparin and heparan sulfate are highly modified in the Golgi by epimerases and sulfotransferases to produce numerous glycoforms (collectively the heparanome)<sup>9</sup> that are implicated in diverse processes including cancer, coagulation, nervous system development, and inflammation.<sup>7</sup> To kick-start development of novel therapeutics,<sup>10</sup> there is a pressing need for improved understanding of how these sequences give rise to bioactivity, in particular, whether glycoforms impart their function via encoded 3D-structural information.

A case in point is the heparin pentasaccharide that initiates anticoagulation through binding to the protein antithrombin-III (AT-III), Figure 1C. This sequence is present in low molecular weight heparin (LMWH), one of the most widely clinically used natural products, but such sequences are problematic to isolate from heterogeneous polymer mixtures, and structurally related contaminants have caused patient deaths.<sup>11</sup> While some drugs (e.g., Fondaparinux and Idraparinux) mimic the AT-III binding sequence,<sup>12</sup> such synthetic alternatives remain difficult

to develop. Additionally, carbohydrate chains containing greater than a few residues are not amenable to solution-phase or solid-state structure determination due to inherent 3D flexibility and severe resonance overlap in NMR spectra, further driving the need for accurate computational approaches to decipher their enigmatic 3D structure–function relationships.

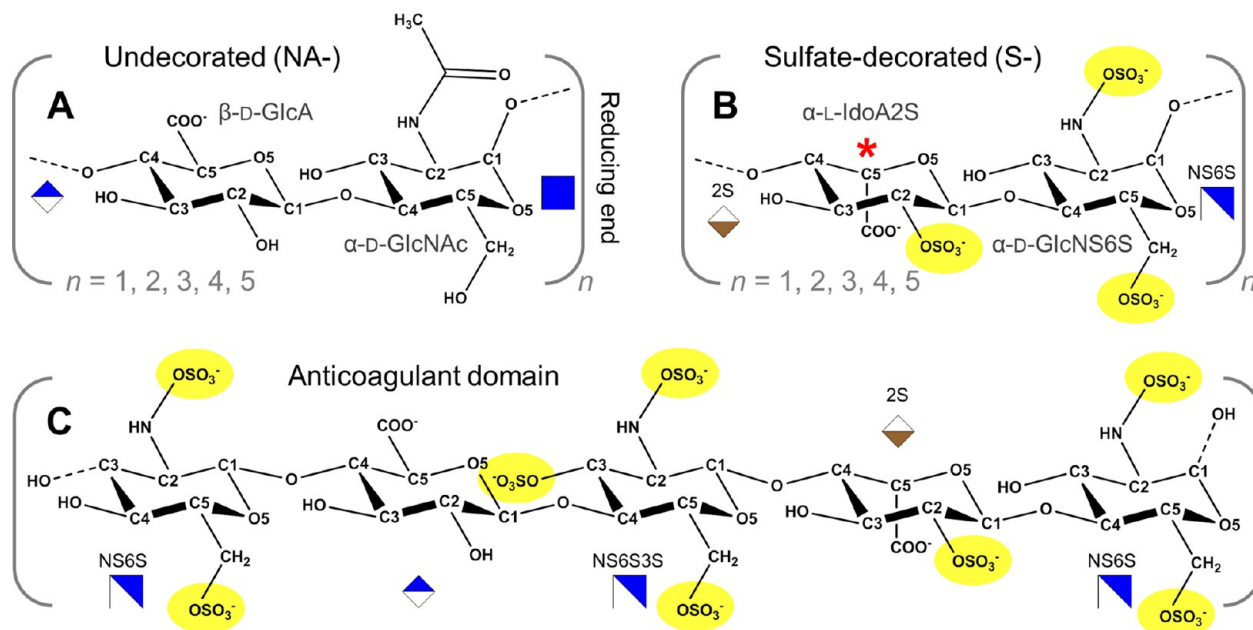
Computational modeling provides a novel route to rationalizing biological function in the heparanome. However, heparin and heparan sulfate present significant challenges to this field, not least because their sequences in vivo are typically large chains of 50–200 disaccharides in length.<sup>13</sup> Furthermore, copolymer heparan sulfate is a complex linear repeat of flexible<sup>14–16</sup> undecorated regions (or NA-domains, Figure 1A) that are sandwiched between shorter protein-binding sulfate-decorated sequences (S-domains, Figure 1B). While theoretical methods offer the promise of new structure–function relationships, they are largely untested on heparanomic sequences and they must be validated by comparing rigorous simulations of model oligosaccharides with experimental data.

The ability to accurately model heparin, heparan sulfate and other carbohydrate sequences at length scales ranging from bioactive oligosaccharides (e.g., Figure 1C) to large polysaccharides is also complicated by ring flexing (or puckering) motions that occur in individual monosaccharides (see Figure 2 and Supporting Information, Figure S1). Puckering of six-

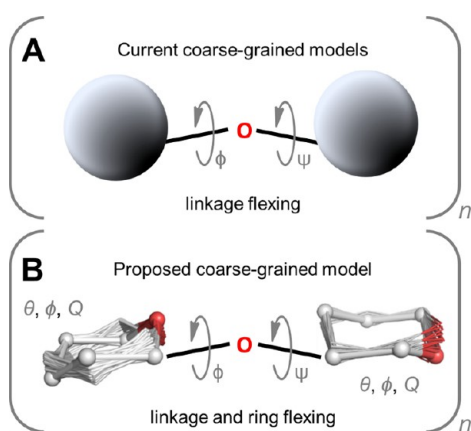
Received: January 15, 2013

Revised: February 22, 2013

Published: February 25, 2013



**Figure 1.** Model oligosaccharides from heparin and heparan sulfate. Chemical modification of the repeating heparan sulfate disaccharide ( $n = 1$ ), i.e., sulfate group attachment and C5 epimerization (\*), causes different pyranose ring flexing (or puckering), diverse sequences, and bioactivities. Unbiased explicit solvent simulations were extended to 5  $\mu$ s for five sequences (2–10 residues) of (A) undecorated or NA-domains, (B) sulfate-decorated or S-domains, and (C) the anticoagulant heparin pentasaccharide.



**Figure 2.** (A) Current coarse-grained carbohydrate models group pyranose ring atoms into clusters (or “beads”) and only allow linkage motion, thereby reducing the computational burden of modeling large chains. (B) In contrast, the coarse-grained approach reported herein uniquely includes both glycosidic linkage and pyranose ring flexibility, allowing accurate microsecond simulations of carbohydrate biopolymers (e.g.,  $n = 400$ ) within hours using standard computer hardware. Parameterized degrees of freedom are shown and ring substituents are hidden.

membered pyranose rings is proposed to be a defining microsecond-time-scale, 3D signature of numerous chemically distinct monosaccharides<sup>17–20</sup> and is mechanistically important for molecular recognition and bioactivity (e.g., in heparin-mediated anticoagulation).<sup>21,22</sup> However, microsecond puckering is difficult to quantify experimentally and no computational study to date has performed rigorous microsecond simulations of oligosaccharides to investigate this phenomenon in biologically relevant carbohydrate sequences.

Modeling such long-time scale dynamic structural degrees of freedom in water requires extensive conformational sampling and, therefore, significant computational resources, to reach

structural convergence. Unbiased nanosecond simulations are insufficient to capture equilibrium microsecond events and artificial enhanced sampling methods can distort computed 3D kinetics.<sup>23</sup> Even allowing for projected advances in hardware acceleration technology, explicit simulation of microsecond atomic motions in micrometer solvated biopolymers will remain unfeasible for the foreseeable future. Current approximate methods for modeling large carbohydrate chains include abstractions of linkage dynamics from unrealistic nanosecond simulations and do not consider biologically important pyranose ring motions (Figure 2A).<sup>24,25</sup>

Here we use hardware acceleration to perform unbiased aqueous dynamics (55  $\mu$ s in total) of 11 biologically important oligosaccharides from heparan sulfate and heparin: undecorated NA-domains and sulfate-decorated S-domains of length 2–10 residues and the anticoagulant AT-III binding pentasaccharide (see Figure 1 for structures). These detailed long-time scale simulations were used to understand local 3D structure in heparin and heparan sulfate protein-binding and linker regions. Development and parametrization of a coarse-grained model (Figure 2B), based on the performed molecular dynamics, enabled investigation of the effect of microsecond puckering on 3D shape in heparan sulfate polymers containing hundreds of residues.

We observed that polymerization biases the puckering equilibrium of unstable 2-*O*-sulfo- $\alpha$ -L-iduronic acid, in agreement with previous experimental observations. This prediction may underlie interactions with proteins and has ramifications for structure-based design of new anticoagulant drugs. We also deduced that in large heparan sulfate chains inclusion of microsecond pyranose ring flexing was required for accurate prediction of macroscopic experimental observables. This insight highlights the need for explicit ring flexing in coarse-grained models of carbohydrates (and other biomolecules) and suggests that microsecond pyranose puckering is a missing 3D-

structural link between sequence and function in the heparanome.

## MATERIAL AND METHODS

**Explicit Aqueous Simulations.** Unbiased all-atom explicit solvent molecular dynamics were performed as described previously<sup>17–20</sup> using NVIDIA graphics processing units and ACEMD software (Acellera Ltd.). A total of 11 oligosaccharides containing 2, 4, 5, 6, 8, and 10 residues (Figure 1A–C) were modeled with the GLYCAM11<sup>26</sup> force field and solvated in cubic TIP3P<sup>27</sup> water boxes of equal side length: 28, 32, 48, 48, 48, and 56 Å, respectively. Sulfate group partial charge parameters were derived from quantum mechanical calculations, reported previously,<sup>18,19</sup> and neutrality was achieved by adding the appropriate number of Na<sup>+</sup> counterions. Following 1000 steps of initial conjugate-gradient energy minimization, each assembly was heated from 0 to 298 K and the box sizes were equilibrated for 20 ns in the NPT ensemble prior to 5.25 μs NVT production simulations at 298 K. The first 250 ns were discarded and data were recorded at 10 ps intervals. The velocity-Verlet integration algorithm and a hydrogen mass repartitioning scheme allowed a 4 fs time step without affecting the equilibrium distribution.<sup>28</sup> Hydrogen atoms were constrained using the M-SHAKE algorithm<sup>29</sup> and electrostatic interactions were calculated via the PME method with a grid spacing of less than or equal to 1.0 Å (in the X, Y, and Z dimensions). Electrostatic and van der Waals interactions were truncated at 9 Å, and a scaling factor of 1.0 was employed for carbohydrate 1–4 interactions.<sup>26</sup>

**Structural, Equilibrium, and Rate Calculations.** Molecular properties were calculated from 5 μs of explicit aqueous or coarse-grained simulations (500000 frames). Glycosidic linkage  $\phi$  and  $\psi$  torsions (O5–C1–O4–C4 and C1–O4–C4–C3, respectively) and pyranose ring puckers were defined and quantified as described previously;<sup>17–20</sup> see Supporting Information, Figure S1 for illustrations and atom numbering. Cremer–Pople<sup>30</sup> parameters  $\theta$ ,  $\phi$ , and  $Q$  were used to quantify pyranose puckers and the angle  $\theta$  was defined as O5–C1–C2–C3–C4–C5. Ring <sup>1</sup>H–<sup>1</sup>H three-bond vicinal couplings (<sup>3</sup>J<sub>H,H</sub> values) were computed using substituent adjusted Karplus equations.<sup>31,32</sup> Average end lengths and radii of gyration were computed using standard equations after appending appropriate chemical decorations to provide the correct mass distribution. Structural equilibration was assessed by monitoring convergence of the simulation average pucker ( $\theta$ , as previously)<sup>17–20</sup> and linkage ( $\phi$  and  $\psi$ ). Puckering rates were calculated as previously<sup>17–20</sup> by measuring the average frequency of transitions. Trajectories were clustered hierarchically using the Amber12<sup>33</sup> tool ptraj with a sampling frequency of 100 frames (5000 in total).

**Extracting Free Energies from Explicit Aqueous Simulations.** Linkages and puckers and their relative free energies ( $\Delta G$ , kcal mol<sup>-1</sup>) were computed at equilibrium from binned conformer populations by applying the Boltzmann eqs 1 and 2 and assuming  $\Delta G \approx \Delta E$ , where  $k_B$  is the Boltzmann constant,  $T$  is temperature (298 K), and  $p_i$  is the probability of bin (state)  $i$ .

$$E_i = -k_B T \ln(p_i) \quad (1)$$

$$\Delta E_{ij} = E_j - E_i = k_B T \ln(p_j/p_i) \quad (2)$$

Linkages were placed in square bins of equal area spanning 30° × 30°. Puckers were placed in bins of equal area on the Cremer–Pople<sup>30</sup> sphere (unequal spacing in the azimuthal angle  $\theta$ , such that there were a total of 12 bins, and equal 30° spacing in the meridional angle  $\phi$ ). To achieve this, the bin sizes were adjusted to normalize for the area on the sphere (the circumference of a spherical cross-section at azimuthal angle  $\theta$  is  $2\pi r \sin \theta$  on a sphere of radius  $r$ ).

**Coarse-Grained Model.** The coarse-grained model comprised two empirical functions, one for the linkage and one for the pucker; both were derived from the set of functions shown in eq 3.

$$\begin{aligned} f_1 &= 1; & f_2 &= \sin x; & f_3 &= \cos x; & f_4 &= \sin 2x; & f_5 &= \cos 2x; \\ & & f_6 &= \sin 3x; & f_7 &= \cos 3x \end{aligned} \quad (3)$$

For linkages, the two-dimensional function shown in eq 4 was used, which satisfies the torsion angle conditions that  $\phi(k) = \phi(k + 2\pi)$  and  $\psi(k) = \psi(k + 2\pi)$  and is differentiable in both  $\phi$  and  $\psi$  (required for performing iterative calculations).

$$V(\phi, \psi) = \sum_{n=1,7;m=1,7} a_{nm} f_n(\phi) f_m(\psi) \quad (4)$$

For Cremer–Pople parameters,<sup>30</sup> the condition  $\phi(k) = \phi(k + 2\pi)$  needed to be satisfied for the meridional angle. To make  $\theta$  parametrize correctly on the surface of a sphere and be differentiable an extra sine term (as found in spherical harmonics) was added and  $\theta$  was doubled to take account of  $0 \leq \theta \leq \pi$ , eq 5. The averages and standard deviations of ring  $Q$  values were measured from explicit aqueous simulations and used to produce the distribution of  $Q$  values in the mesoscale model.

$$V(\theta, \phi) = c_1 + c_2 \theta + c_3 \theta^2 + \sum_{n=1,7;m=1,7} b_{nm} f_n(2\theta) f_m(\phi) \sin \theta \quad (5)$$

Energy surfaces were calculated for linkages and ring puckers by first calculating their probability from microsecond explicit aqueous simulations in water and converting this to energy surfaces using eqs 1 and 2. Each ring and linkage was parametrized in the coarse-grained representation by varying the coefficients in eqs 4 and 5 until the optimum least-squares fit for  $V$  was obtained. Example linkage and ring parameters for an undecorated heparan sulfate disaccharide are reported in Supporting Information, Figure S9. For polymers, suitable coefficients were used for each linkage and ring. Linkages and rings were integrated independently using the Leapfrog algorithm (with a time step of 2.5 fs) and the Langevin equation of motion, eq 6, where  $L$  is stochastic noise from a Gaussian distribution (for coarse-grained simulations performed here the mean was set to 1.0),  $V$  is the potential energy function, and  $x$  represents a degree of freedom (two for each linkage and two for each pucker). For the coarse-grained simulations performed here,  $m$  and  $\gamma$  were set to 200 and 2.0, respectively.

$$m\ddot{x} + \gamma\dot{x} = L(t) - \frac{\delta V}{\delta x} \quad (6)$$

Polymers were reconstructed into three-dimensional space using standard torsion and bond geometries<sup>26</sup> and puckers were built from the Cremer–Pople parameters using standard geometries,<sup>26</sup> with methods described previously<sup>30</sup> but in reverse. Simulated parameters ( $\theta$ ,  $\phi$ , and  $Q$ ) were transformed as described in the original work,<sup>30</sup> eq 7. The six  $z$ -coordinates were calculated by using the standard Cremer–Pople equation for a six-membered ring, eq 8.

$$q_2 \stackrel{\text{def}}{=} Q \sin \theta; \quad \phi_2 \stackrel{\text{def}}{=} \phi; \quad q_3 \stackrel{\text{def}}{=} Q \cos \theta \quad (7)$$

$$z_j = \sqrt{\left(\frac{1}{3}\right)} q_2 \cos \left[ \phi_2 + \frac{2\pi}{3}(j-1) \right] + \sqrt{\left(\frac{1}{6}\right)} q_3 (-1)^{(j-1)} \quad (8)$$

The  $x$ - and  $y$ -coordinates were recovered using geometrical arguments. First, the projections of the bonds in the  $xy$ -plane were calculated from the bond lengths ( $r_i$ ) and the  $z_i$  values using eq 9, where  $z_0 \equiv z_6$  for a six-membered ring. The radius of the ring and an angle in the  $xy$ -plane was then approximated, eqs 10, and finally the  $x$ - and  $y$ -coordinates recovered, eq 11. The rings and linkages were recombined into a chain using fixed geometry using standard bonds and angles.<sup>26</sup>

$$d_j = 3 \sin^{-1} \left[ \frac{1}{3} \sqrt{r_j^2 - (z_j - z_{j-1})^2} \right] \quad (9)$$

$$\sigma = \frac{1}{2\pi} \sum_{j=1}^6 d_j; \quad \theta_j = \frac{1}{\sigma} \sum_{i=1}^j d_i \quad (10)$$

$$x_j = \sigma \sin \theta_j; \quad y_j = \sigma \cos \theta_j \quad (11)$$

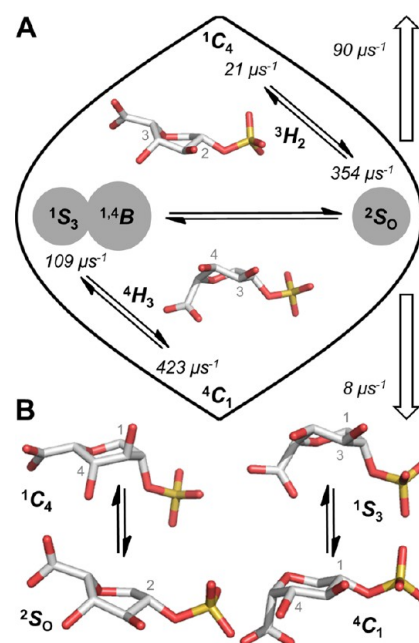


**Polysaccharide Production and Light Scattering.** The heparan sulfate polymer backbone (without sulfate decoration) was produced from a single batch of *Escherichia coli* O10/KS/H4 and purified as described previously.<sup>16</sup> For multiangle laser light scattering analysis, 4 mg were dissolved in 0.5 mL of ammonium bicarbonate (100 mM) and the samples were centrifuged for 10 min (at 12000 g). The supernatant was collected and purified polymers were applied to a Superose 12 (10/300GL) gel filtration column (GE Healthcare) with a flow rate of 0.5 mL/min in ammonium bicarbonate (100 mM) buffer. The eluent was passed into a multiangle laser light (658 nm) scattering detector (DAWN HELEOS-II, Wyatt) and subsequently through a refractive index detector (Wyatt). Data were analyzed using ASTRA software (Wyatt) using the method of Zimm and a suitable specific refractive increment ( $dn/dc$ ) of 0.153.

## RESULTS AND DISCUSSION

**Convergence in Microsecond Simulations of S-Domains, NA-Domains, and the AT-III Binding Pentasaccharide.** Unbiased explicit aqueous simulations of 11 model oligosaccharides from heparan sulfate and heparin, namely, NA- and S-domains of variable length and the AT-III binding pentasaccharide (see Figure 1 for structures), were extended to 5  $\mu$ s each. Dynamic exchange between different pyranose ring puckers converged to equilibrium (a prerequisite for accurate prediction of experimental measurements) after at most 3  $\mu$ s in all 65 rings. This finding is consistent with previous work<sup>17–20</sup> and is exemplified in Supporting Information, Figure S2, which shows puckering convergence plots for each ring in the model AT-III binding pentasaccharide (Figure 1C). Glycosidic linkage equilibrium was assessed by inspection of dihedral angle time series plots for the 108  $\phi$  and  $\psi$  torsions from the 54 linkages, data for the eight glycosidic dihedrals in the model pentasaccharide are shown in Supporting Information, Figure S3. The torsions underwent multiple transitions between expected conformations, indicating that all low-energy aqueous glycosidic linkage conformers had been explored. While unsampled high-energy linkage conformations may remain and microsecond time scale exchange was predicted (see Supporting Information, Figure S3), our results intimate that these are unlikely to represent a significant population in the aqueous glycosidic linkage equilibrium ensembles. Therefore, it was assumed that the 5  $\mu$ s oligosaccharide simulations had sampled all accessible aqueous conformers (i.e., reached equilibrium), permitting prediction of conformer populations, free energies, and pucker exchange kinetics.

**Effect of Polymerization on Uronic Acid Pucker.** The 2-*O*-sulfo- $\alpha$ -L-iduronic acid (IdoA2S) residue (an unstable  ${}^1C_4$  chair in water)<sup>18</sup> from the AT-III binding pentasaccharide simulation underwent forward ( ${}^1C_4 \rightarrow {}^4C_1$ ) and backward ( ${}^4C_1 \rightarrow {}^1C_4$ ) chair–chair conformational exchange at computed rates of 8 and 90  $\mu$ s<sup>-1</sup>, respectively (Figure 3). The occupancy of chairs was 5% ( ${}^1C_4$ ) and 79% ( ${}^4C_1$ ) and the free energy separating them was predicted to be  $\Delta G \approx 1.7$  kcal mol<sup>-1</sup> ( ${}^4C_1$  lower in energy). All but 2 of the 36 canonical nonchair puckers (the  ${}^1C_4$  hemisphere half-chair  ${}^5H_0$  and envelope  ${}^5E$ ) were sampled and the boats (B) and skew-boats (S)  ${}^{1,4}B$  (5%),  ${}^1S_3$  (3%),  ${}^2S_0$  (3%),  ${}^1S_5$  (1%), and  ${}^{2,5}B$  (1%) were the predominant nonchair puckers (all predicted to be less than 3.0 kcal mol<sup>-1</sup> above the lowest energy  ${}^4C_1$  chair). Similar observations were made for the internal IdoA2S residues of the S-domain oligosaccharides comprising 4–10 residues (Figure 1B). For comparison, the IdoA2S monosaccharide in water was predicted to have equivalent forward and backward transition

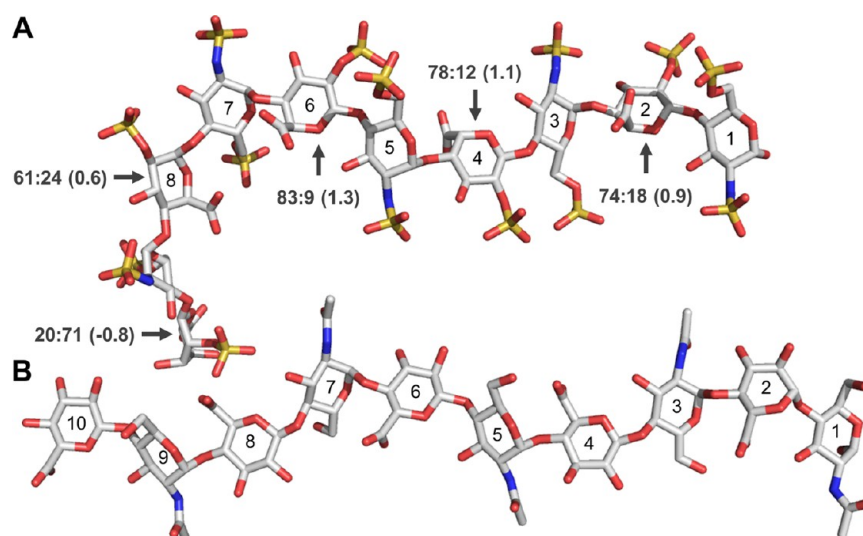


**Figure 3.** (A) Schematic representation of IdoA2S puckering in a 5  $\mu$ s simulation of the AT-III binding heparin pentasaccharide in water (Figure 1C). The predominant equatorial boat (B) and skew-boat (S) puckers are indicated within the spherical projection (the diameters of the gray circles represent their respective populations) and their rates of kinetic exchange with chair puckers are shown (double harpoons). The computed overall exchange rates between chairs ( ${}^1C_4$ ,  ${}^4C_1$ ) are indicated by vertical arrows (far right). Representative 3D-conformers of the most populated half-chair (H) intermediates between the equator and the polar chair puckers are also shown. (B) Representative IdoA2S chair and skew-boat puckers extracted from the AT-III binding heparin pentasaccharide 5  $\mu$ s simulation. Pyranose ring atoms are numbered and hydrogen atoms are hidden for clarity.

rates of 2 and 176  $\mu$ s<sup>-1</sup>.<sup>18</sup> In the monosaccharide, the occupancy of chair puckers was 92% ( ${}^1C_4$ ) and 1% ( ${}^4C_1$ ), with  ${}^2S_0$  (3%),  ${}^3S_1$  (1%), and  $B_{30}$  (1%) being the predominant nonchair conformations, and the free energy separating the two chair puckers was computed to be  $\Delta G \approx 2.7$  kcal mol<sup>-1</sup> ( ${}^1C_4$  lower in energy).

Our prediction also differs from the computed metastable  ${}^1C_4$  chair in nanosecond metadynamics<sup>34</sup> and the IdoA2S residue has been found, by NMR and X-ray crystallography, to possess different chair/skew-boat puckering equilibria (predominant  ${}^1C_4$ ) when in complex with various proteins (e.g., AT-III and Fibroblast Growth Factor Receptor 2).<sup>21,35</sup> This suggests that the low free energy barrier to exchange in water (1.7 kcal mol<sup>-1</sup> in our heparin simulation) may allow for differential selectivity. Figure 3 illustrates the IdoA2S puckering itinerary predicted in our 5  $\mu$ s<sup>-1</sup> AT-III binding pentasaccharide simulation. Conformational exchange between the  ${}^1C_4$  chair and the equatorial  ${}^2S_0$  skew-boat occurred via the intermediate half-chair  ${}^3H_2$  at computed rates of 354  $\mu$ s<sup>-1</sup> ( ${}^1C_4 \rightarrow {}^2S_0$ ) and 21  $\mu$ s<sup>-1</sup> ( ${}^2S_0 \rightarrow {}^1C_4$ ). The IdoA2S residue underwent pseudorotation (during which all equatorial boats and skew-boats were sampled) and exchange between the  ${}^1S_3$  skew-boat and the  ${}^4C_1$  chair occurred at computed rates of 423  $\mu$ s<sup>-1</sup> ( ${}^1S_3 \rightarrow {}^4C_1$ ) and 109  $\mu$ s<sup>-1</sup> ( ${}^4C_1 \rightarrow {}^1S_3$ ), via the intermediate half-chair  ${}^4H_3$ .

Thus, based on the most extensive simulations to date, we predict that polymerization shifts the aqueous equilibrium



**Figure 4.** Predominant 3D conformers of heparan sulfate S-domain (A) and NA-domain (B) deca-saccharides determined by hierarchical 3D clustering of 5  $\mu$ s unbiased explicit solvent simulations. The coarse-grained dynamics reported herein is consistent with the extended 3D shape of the NA-domain oligosaccharide (cf. the contracted S-domain). Percentage ratios of S-domain IdoA2S residue chair puckers ( ${}^4C_1/{}^1C_4$ ) and their computed relative free energies (in parentheses) are shown. Pyranose rings are numbered from the reducing end and relative free energies were calculated as the free energy of  ${}^1C_4$  minus the free energy of  ${}^4C_1$ .

distribution of internal IdoA2S residue puckers, from the  ${}^1C_4$  chair (in the monosaccharide) toward the  ${}^4C_1$  chair (as found here in oligosaccharides), with a computed relative free energy change of  $\Delta\Delta G \approx 4.4$  kcal mol $^{-1}$ . The IdoA2S residue  ${}^2S_0$  pucker, which is present in the AT-III-heparin pentasaccharide cocrystal structure,<sup>21</sup> was equally populated in the monosaccharide and the AT-III binding pentasaccharide simulations (3%). Interestingly, in the S-domain deca-saccharide simulations, the IdoA2S residues at positions 2, 4, 6, 8, and 10 (relative to the reducing end) all exhibited relatively greater populations of  ${}^2S_0$ ; these were computed to be 7, 8, 6, 13, and 6%, respectively.

When IdoA2S was adjacent to two GlcNS6S residues (cf. GlcNS6S3S and GlcNS6S in the AT-III binding pentasaccharide) the average chair–chair free energy difference of 10 internal residues (from all of the simulated S-domain oligosaccharides) reduced to  $\Delta G \approx 0.8$  kcal mol $^{-1}$  ( ${}^4C_1$  lower in energy). In all four terminal (nonreducing end) IdoA2S residues from the S-domain oligosaccharides (comprising 4–10 residues), the  ${}^1C_4$  chair was lower in energy, with the average chair–chair free energy difference being  $\Delta G \approx 1.2$  kcal mol $^{-1}$ . Furthermore, IdoA2S puckering was predicted to be increasingly  ${}^4C_1$ -like toward the middle of an oligosaccharide chain. For example, in the S-domain deca-saccharide the IdoA2S residues at positions 2, 4, 6, and 8 were predicted to favor the  ${}^4C_1$  chair (cf.  ${}^1C_4$ ) by 0.9, 1.1, 1.3, and 0.6 kcal mol $^{-1}$ , respectively. Conversely, the nonreducing end terminal IdoA2S residue (at position 10) favored the  ${}^1C_4$  chair (by 0.8 kcal mol $^{-1}$ ), as found in the monosaccharide. This prediction is illustrated on the most populated S-domain 3D conformer, determined by clustering the 5  $\mu$ s trajectory, in Figure 4A.

The simulations performed here therefore suggest that longer S-domains contain higher proportions of  ${}^4C_1$  chairs and allow us to hypothesize that puckering is a mechanism for sequence-specific protein recognition. If so, mimetic drugs would have to replicate these complex dynamics and our findings may explain why the production of potent AT-III binding anticoagulants has been difficult in the absence of

complex carbohydrate chemistry resembling the AT-III binding heparin pentasaccharide. Further work is needed to elucidate the detailed molecular mechanisms resulting in these pucker population differences and, in particular, the positional dependence of puckering equilibria in oligosaccharides. Possible causes include the increased strain forces present within longer chains and also the effect of solvent water.

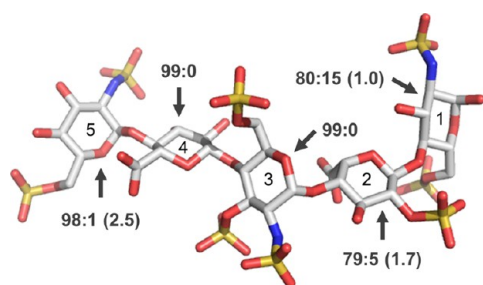
The pyranose ring in the  $\beta$ -D-glucuronic acid (GlcA) monosaccharide was not predicted to undergo  ${}^4C_1 \rightarrow {}^1C_4$  transitions in previous unbiased microsecond simulations.<sup>18</sup> This was also the case for internal and terminal (nonreducing end) GlcA residues in the current 5  $\mu$ s simulations of the AT-III binding pentasaccharide and the heparan sulfate NA-domain oligosaccharides (of length 4–10 residues). The predominant NA-domain deca-saccharide 3D-conformer is illustrated in Figure 4B, which is notably more extended than the most populated S-domain of equivalent degree of polymerization (Figure 4A). Computed GlcA ring puckering was extremely similar in all model systems, regardless of chain length, chain position and adjacent residues. While the role of IdoA 2-O-sulfation remains uncertain in heparan sulfate mediated processes, such as development,<sup>3</sup> on the basis of these simulations we can speculate that the diverse bioactivity in the heparanome is enabled by the presence of variably flexible IdoA and IdoA2S residues.

#### Effect of Polymerization on Hexosamine Pucker.

Considering the unsulfated, di-, and trisulfated D-glucosamines GlcNAc, GlcNS6S and GlcNS6S3S, only the terminal residues underwent chair–chair exchange in simulations performed here. Specifically these were: reducing end GlcNAc and GlcNS6S in the NA- and S-domain models, respectively, and both terminal GlcNS6S residues in the AT-III binding pentasaccharide. Reducing-end GlcNAc residues in the undecorated oligosaccharides transitioned from  ${}^4C_1$  to  ${}^1C_4$  at a faster rate than predicted for the free monosaccharide ( $\approx 150$   $\mu$ s $^{-1}$  vs  $\approx 1$   $\mu$ s $^{-1}$  previously).<sup>19</sup> Interestingly, this faster GlcNAc  ${}^4C_1 \rightarrow {}^1C_4$  exchange rate occurred in NA-domain oligosaccharides containing 4–10 residues, while in the disaccharide

simulation the reducing-end GlcNAc underwent this  ${}^4C_1 \rightarrow {}^1C_4$  transition at a slower rate ( $\approx 20 \mu s^{-1}$ ) resembling monosaccharide chair–chair exchange kinetics. This finding suggests an effect of chain length on the reducing-end GlcNAc ring puckering equilibrium.

In the S-domain oligosaccharides, the reducing-end GlcNS6S residue was the only hexosamine to undergo chair–chair transition. This occurred only in the di- and hexasaccharide simulations in which the computed  ${}^4C_1 \rightarrow {}^1C_4$  exchange rates were  $\approx 1 \mu s^{-1}$ . In the AT-III binding pentasaccharide both terminal GlcNS6S residues puckered between chair forms. Although there were insufficient transitions in the nonreducing end terminal residue to compute reliable exchange rates, relative chair–chair free energies for the reducing and nonreducing end residues were estimated to be 1.0 and 2.5 kcal mol $^{-1}$ , respectively, indicating greater flexibility of the GlcNS6S residue adjacent to the flexible IdoA2S (cf. stable GlcA). This predicted effect is illustrated on the most populated AT-III binding pentasaccharide 3D conformer in Figure 5. These results concur with our previous<sup>19</sup> microsecond



**Figure 5.** Predominant 3D conformer of the anticoagulant AT-III binding heparin pentasaccharide determined by hierarchical 3D clustering of the 5  $\mu s$  unbiased explicit solvent simulation. Percentage ratios of chair pucker ( ${}^4C_1/{}^1C_4$ ) in each residue and their computed relative free energies (in parentheses) are shown. Relative free energies were calculated as the free energy of  ${}^1C_4$  minus the free energy of  ${}^4C_1$ .

dynamics studies of both GlcNS6S and GlcNS6S3S monosaccharides, wherein chair–chair puckering was predicted only in GlcNS6S and only after several microseconds. Therefore, while hexosamine puckering may contribute to bioactivity in the heparanome, in these residues it is likely that charge interactions involving sulfate groups are the dominant factor in binding.

Our long-time scale simulations and comparisons herein intimate that the equilibrium of aqueous microsecond pyranose puckering in unstable and metastable rings (which undergo chair–chair exchange as free monosaccharides) is mediated by the effects of polymerization: chain length, chain position and the chemical identity of neighboring residues. In particular, for unstable IdoA2S, a key determinant of the AT-III binding pentasaccharide pharmacophore, the heparan sulfate S-domain terminal and internal IdoA2S residues were predicted to have significantly different puckering signatures. While the sulfation pattern of adjacent hexosamines did influence puckering (comparing nonterminal IdoA2S in S-domains and the AT-III binding pentasaccharide), the 3D-shape of the relatively rigid sulfated hexosamines themselves was only significantly different at the reducing end termini.

**Comparison of Oligosaccharide Microsecond Dynamics to Experiments.** The equilibrium predictions for glycosidic linkage geometries were validated by comparison

with experimental data from protein cocrystallography, solution scattering, and nuclear magnetic resonance spectroscopy, Table 1. Computed average torsion values in the undecorated and

**Table 1.** Calculated and Experimental Glycosidic Linkage Geometries in Heparan Sulfate Domains<sup>a</sup>

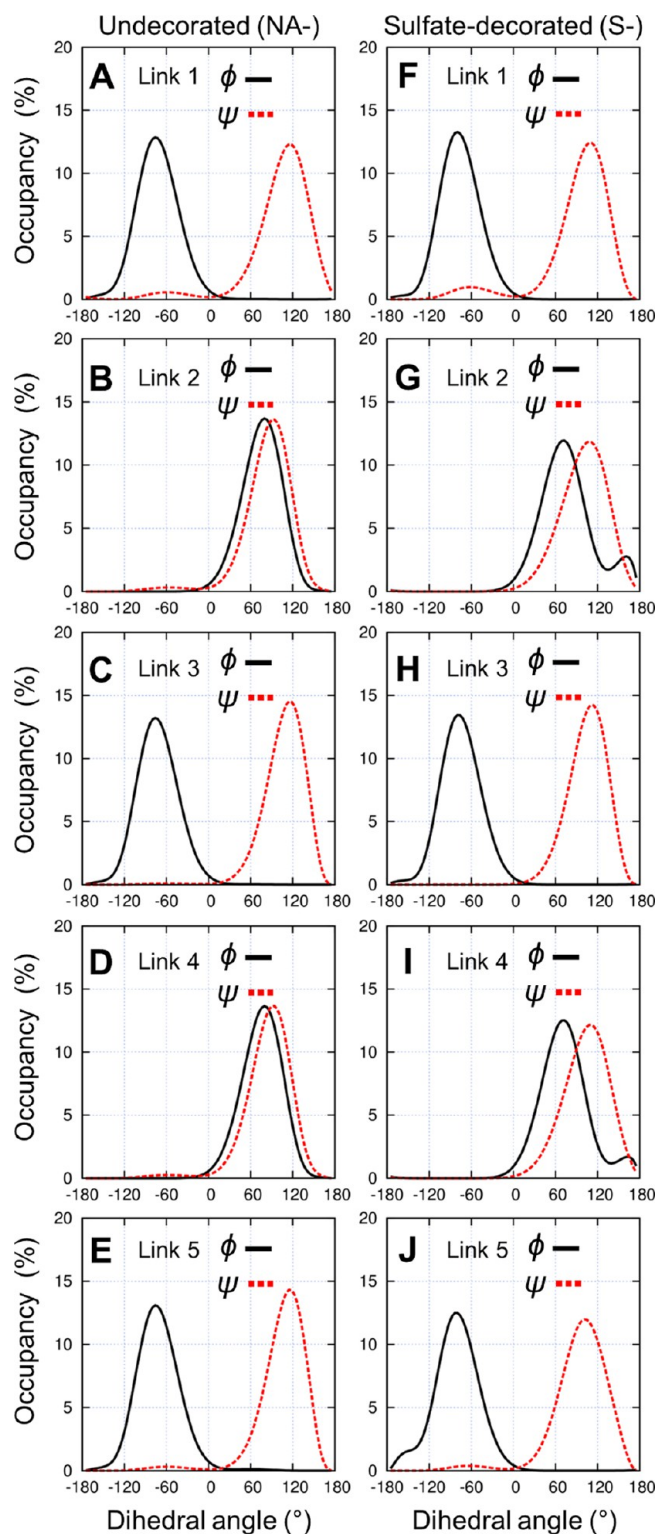
A				
method	GlcNAc-GlcA (NA-)		GlcA-GlcNAc (NA-)	
	$\phi$ (deg)	$\psi$ (deg)	$\phi$ (deg)	$\psi$ (deg)
CAL (24) <sup>b</sup>	80 $\pm$ 13	91 $\pm$ 27	-77 $\pm$ 18	115 $\pm$ 27
XTL (2) <sup>c</sup>	83 $\pm$ 3	94 $\pm$ 1	-91 $\pm$ 1	127 $\pm$ 4
XS (27) <sup>d</sup>	86 $\pm$ 22	91 $\pm$ 25	-87 $\pm$ 27	129 $\pm$ 26
NMR (1) <sup>c</sup>	90 $\pm$ 30	90 $\pm$ 30	-60 $\pm$ 30	120 $\pm$ 30
B				
method	GlcNS6S-IdoA2S (S-)		IdoA2S-GlcNS6S (S-)	
	$\phi$ (deg)	$\psi$ (deg)	$\phi$ (deg)	$\psi$ (deg)
CAL (24) <sup>b</sup>	76 $\pm$ 26	107 $\pm$ 24	-82 $\pm$ 16	108 $\pm$ 31
XTL (19) <sup>c</sup>	84 $\pm$ 22	100 $\pm$ 19	-79 $\pm$ 20	132 $\pm$ 19
XS (50) <sup>c</sup>	98 $\pm$ 25	86 $\pm$ 22	-61 $\pm$ 21	132 $\pm$ 22
NMR/C (1) <sup>c</sup>	79 $\pm$ 30	150 $\pm$ 30	-77 $\pm$ 30	110 $\pm$ 30
NMR/S (1) <sup>c</sup>	109 $\pm$ 30	158 $\pm$ 30	-55 $\pm$ 30	107 $\pm$ 30

<sup>a</sup>The number of observations for each method is reported in parentheses. <sup>b</sup>Calculated (CAL) torsion values are averages ( $\pm$ standard deviations) computed from 5  $\mu s$  simulations performed here. The hexosamine  $\rightarrow$  uronic acid and uronic acid  $\rightarrow$  hexosamine values were calculated from 10 and 14 measurements, respectively, and from oligosaccharides containing 4, 6, 8, and 10 residues. Structures and torsion definitions are detailed in Figure 1, Methods, and Supporting Information, Figure S1. <sup>c</sup>Experimental data from protein cocrystallography (XTL), X-ray scattering (XS),<sup>14,15</sup> and solution NMR spectroscopy<sup>36,37</sup> (NMR) in  ${}^1C_4$  chair (C) and  ${}^2S_0$  (S) puckers. Torsions derived from experimental NMR data were converted from their standard definitions and errors were estimated. <sup>d</sup>Perkins, S., personal communication (correction to an error in published data<sup>15</sup>).

sulfate-decorated oligosaccharides, Table 1A and B, respectively, were all within the experimental error of structural inferences from this large body of solid-state and solution phase data.<sup>14,15,36,37</sup> Furthermore, computed glycosidic linkage dihedrals from NA- and S-domains were almost identical, exemplified in Figure 6, which compares the computed glycosidic torsion distributions in each model hexasaccharide (NA- and S-domains). By inspection of pyranose ring flexing and glycosidic torsion time series plots in the all-atom simulations, there was no apparent correlation between puckering and linkage conformation. These observations suggest that sulfate groups have little effect on linkage geometry, as found experimentally for heparan and chondroitin sulfates.<sup>41,42</sup>

As previously,<sup>17–20</sup> the computed aqueous equilibrium dynamics were used to predict nuclear magnetic resonance pyranose ring vicinal couplings ( ${}^3J_{H,H}$ ). Calculated  ${}^3J_{H,H}$  values from microsecond-time-scale monosaccharide simulations, and employing the force field used here, have been shown<sup>17–20</sup> to achieve good trend agreements with experimental data and here they were validated against equivalent measurements from undecorated and sulfate-decorated oligosaccharides in solution,<sup>36,39,40,43</sup> which are difficult to measure accurately due to resonance overlap. Comparisons for simulated tetrasaccharides (undecorated and sulfate-decorated) are shown in Table 2. Data for the other model systems, which trended similarly and were also consistent with previous work,<sup>17–20</sup> are shown in





**Figure 6.** Comparisons of glycosidic linkage dynamics in 5  $\mu$ s unbiased explicit solvent simulations of undecorated (NA-domain) and sulfate-decorated (S-domain) hexasaccharides from heparan sulfate (see Figure 1A,B). Linkages are numbered from the reducing end, dihedral angles were binned in  $10^\circ$  increments, and the plots are Bezier-smoothed.

Supporting Information, Figure S4. Computed and experimental  $^3J_{\text{H,H}}$  values show good agreement (Table 2), 18 of the 24 compared values are within 2 Hz. In the undecorated sugar, the average absolute deviation of predictions from experimental

values for internal GlcNAc and GlcA residues were 1.2 and 1.6 Hz, respectively. Predictions for the terminal (reducing end) GlcNAc deviated from experiment by 2.7 Hz, on average. Computed internal and terminal (nonreducing end) GlcA  $^3J_{\text{H,H}}$  values were effectively identical ( $\pm \approx 1$  Hz) and are also indicative of stable  $^4C_1$  chairs. For the sulfate-decorated tetrasaccharide, computed  $^3J_{\text{H,H}}$  values for internal and reducing-end GlcNS6S deviated on average from experiment by just 0.2 Hz.

The maximum deviation of calculated and experimental values was seen for internal IdoA2S residues. As discussed previously, only a small (energetic) change in force field parameters would be needed to alter the finely balanced equilibrium of idopyranose puckers.<sup>18</sup> In Table 2 the  $^3J_{\text{H,H}}$  predictions for this residue are on average 3.0 Hz greater than the corresponding experimental values. A similar magnitude of discrepancy in this sugar and the difficulty of predicting its ring vicinal  $^1\text{H}$ – $^1\text{H}$  couplings have been reported and discussed previously.<sup>18</sup> However, while experiment is perhaps suggestive of smaller shifts toward the  $^4C_1$  chair in these residues, the trend for a bias toward the  $^4C_1$  chair in a polymer (indicated by the larger  $^3J_{\text{H,H}}$  values) is consistent. Ring  $^3J_{\text{H,H}}$  values for the free IdoA2S monosaccharide were measured previously using 900 MHz NMR to be 1.8 ( $^3J_{1,2}$ ), 3.3 ( $^3J_{2,3}$ ), 3.5 ( $^3J_{3,4}$ ), and 2.2 ( $^3J_{4,5}$ ) Hz<sup>18</sup> and the measurements and calculations for the internal IdoA2S residue in Table 2 are all greater than these prior observations. Additionally, ring vicinal  $^1\text{H}$ – $^1\text{H}$  couplings for internal IdoA2S residues from S-domain tri-, tetra-, and pentasaccharides have been reported previously to be greater still at 5.2 ( $^3J_{1,2}$ ), 4.3 ( $^3J_{2,3}$ ), 4.2 ( $^3J_{3,4}$ ), and 4.2 ( $^3J_{4,5}$ ) Hz,<sup>43</sup> further supporting our computational prediction that a greater percentage of  $^4C_1$  chair is present in the IdoA2S equilibrium of puckers when present in a polymer compared to a free monosaccharide.

Our rigorous all-atom simulations provide evidence that the equilibrium 3D-kinetics and conformer free energies of unstable pyranoses, which is modulated by the extent of chemical decoration, is also dependent on polymerization. These findings are consistent with previous experiments and long-time scale simulations,<sup>17</sup> they may also be central to molecular recognition and could provide ideas for structure-based design of new drugs and bioactive materials. All puckering degrees of freedom identified in previous<sup>17–20</sup> microsecond calculations were sampled in the explicit solvent microsecond simulations (see Supporting Information, Figures S5 and S6) and the data were considered an excellent basis for studying the effect of puckering in longer physiologically relevant polymers.

**Impact of Puckering in Large Biopolymers.** As discussed, explicitly simulating large solvated biopolymers on micrometer scales is computationally unfeasible and alternative approaches are needed. In this case, equilibrium glycosidic linkage and pyranose ring motions from the aqueous microsecond simulations were converted to free energy surfaces and used to parametrize a coarse-grained model by least-squares fitting to empirical energy functions (see Methods). Using these energy functions, coarse-grained simulations were performed for all 11 model oligosaccharides (Figure 1), which executed more than 3 orders of magnitude faster than the equivalent all-atom simulations, even at this length scale and without hardware acceleration.

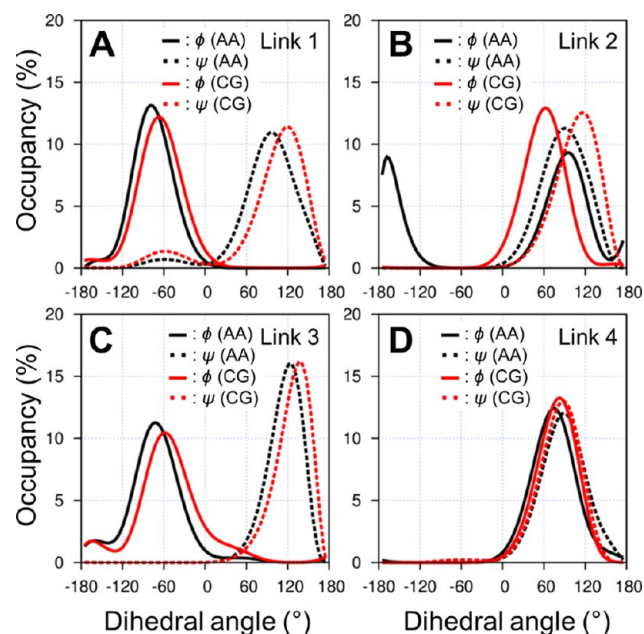
The precision and accuracy of the coarse-grained model in terms of structural, thermodynamic, and 3D-kinetic properties is exemplified here by comparing linkage and ring flexing in 5

**Table 2.** Calculated and Experimental Pyranose Ring Three-Bond Vicinal  $^1\text{H}$ – $^1\text{H}$  NMR Couplings for Undecorated and Sulfate-Decorated Heparan Sulfate Tetrasaccharides<sup>a</sup>

		undecorated (NA-domain)				sulfate-decorated (S-domain)			
		GlcA	GlcNAc	GlcA	GlcNAc	IdoA2S	GlcNS6S	IdoA2S	GlcNS6S
CAL	$J_{1,2}$	6.7 (0.0)	2.9 (0.0)	6.5 (0.0)	2.8 (0.0)	2.0 (0.4)	2.5 (0.0)	7.0 (1.1)	2.7 (0.0)
CAL	$J_{2,3}$	7.5 (0.0)	8.7 (0.1)	7.3 (0.0)	6.0 (0.2)	2.2 (0.4)	10.2 (0.0)	7.2 (1.0)	10.2 (0.0)
CAL	$J_{3,4}$	7.5 (0.0)	7.8 (0.0)	7.3 (0.0)	6.4 (0.1)	2.4 (0.4)	9.9 (0.0)	7.2 (1.1)	10.0 (0.0)
CAL	$J_{4,5}$	8.3 (0.0)	8.2 (0.0)	8.2 (0.0)	6.6 (0.2)	3.4 (0.1)	10.0 (0.0)	3.9 (0.1)	10.0 (0.0)
EXP <sup>b</sup>	$J_{1,2}$	-	3.8 (0.1)	8.0 (0.1)	3.4 (0.1)	-	3.7 (0.1)	2.4 (0.1)	3.5 (0.1)
EXP <sup>b</sup>	$J_{2,3}$	-	10.5 (0.1)	9.3 (0.1)	10.7 (0.1)	-	10.6 (0.1)	4.8 (0.1)	10.2 (0.1)
EXP <sup>b</sup>	$J_{3,4}$	-	8.5 (0.1)	8.9 (0.1)	8.6 (0.1)	-	9.0 (0.1)	3.4 (0.1)	8.8 (0.1)
EXP <sup>b</sup>	$J_{4,5}$	-	9.7 (0.1)	9.6 (0.1)	9.9 (0.1)	-	10.1 (0.1)	2.5 (0.1)	9.8 (0.1)

<sup>a</sup> $J_{\text{H,H}}$  values are reported in units of Hz. Calculated (CAL) averages (and standard deviations) are from 5  $\mu\text{s}$  all-atom simulations, computed using the substituent adjusted Karplus equations of Altona & Haasnoot<sup>38</sup> (NA-domain) and Hricovini et al.<sup>32</sup> (S-domain). Experimental (EXP) measurements (and estimated errors) are from solution NMR spectroscopy.<sup>b</sup>Blundell et al.,<sup>36</sup> Sue et al.,<sup>39</sup> and Jin et al.<sup>40</sup> See Figure 1 for structures. Experimental values at nonreducing termini (“-”) were not compared due to the presence of unsaturated C4–C5 bonds

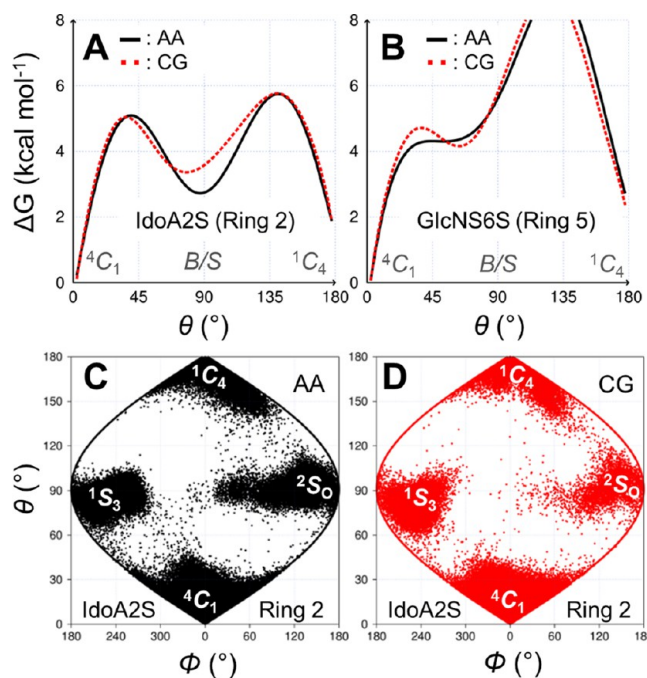
$\mu\text{s}$  all-atom and coarse-grained simulations of the AT-III binding pentasaccharide. The all-atom dynamics of all four glycosidic linkages were faithfully reproduced by the coarse-grained model (Figure 7A–D) and the pyranose ring atomic



**Figure 7.** Comparisons of glycosidic linkage dynamics in 5  $\mu\text{s}$  unbiased explicit solvent all-atom (AA, black) and coarse-grained simulations (CG, red) of the AT-III binding heparin pentasaccharide (see Figure 1C). Linkages are numbered from the reducing end, dihedral angles were binned in 10° increments, and the plots are Bezier-smoothed.

motions were also in precise agreement, typified by comparisons of computed all-atom and coarse-grained puckering free energy profiles in Figure 8A,B and sampled individual puckers in unstable IdoA2S, Figure 8C,D.

In residues that underwent significant exchange between the two pyranose chair puckers ( $^4\text{C}_1$  and  $^1\text{C}_4$ ), the relative chair conformer free energies in the all-atom and coarse-grained simulations were all within 0.3 kcal mol<sup>-1</sup>. Furthermore, the computed all-atom and coarse-grained chair–chair puckering rates trended in agreement (Supporting Information, Figure S7). These findings and trends were evident in the other 10 model oligosaccharides (Supporting Information, Figures S5

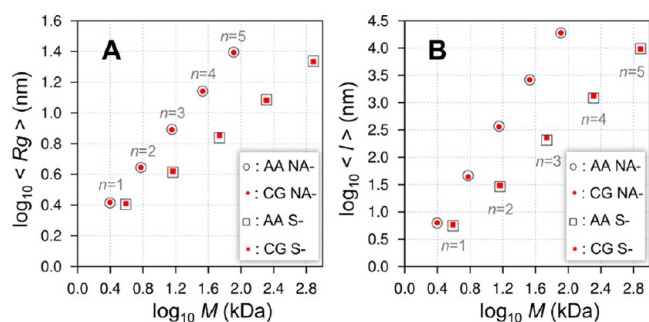


**Figure 8.** Comparisons of calculated pyranose ring puckering in 5  $\mu\text{s}$  all-atom (AA) and coarse-grained (CG) simulations of the AT-III binding heparin pentasaccharide. Bezier-smoothed one-dimensional puckering free energy profiles in (A) IdoA2S and (B) terminal GlcNS6S residues are shown with chair ( $^4\text{C}_1$ ,  $^1\text{C}_4$ ) and boat/skew-boat (B/S) conformers indicated. Panels (C) and (D) compare individual IdoA2S puckers sampled in the all-atom and coarse-grained simulations on sinusoidal projections of puckering space (which faithfully represent the populations in spherical phase-space). Ring numbering corresponds to Figure 5 and the angles  $\theta$  and  $\phi$  describe puckering in six-membered rings (see Methods).

and S6). The computed radii of gyration and end-to-end lengths for the 10 model NA- and S-domains from heparan sulfate were identical in the 5  $\mu\text{s}$  all-atom and coarse-grained simulations, Figure 9A,B.

In the all-atom simulations, other than at the oligosaccharide termini, the conformations of chemically distinct glycosidic linkages and pyranose rings were similar. This allowed repetition of linkage and ring parameters derived from nonterminal residues to simulate long heparan sulfate polymers. In the coarse-grained models, parameters from internal residues





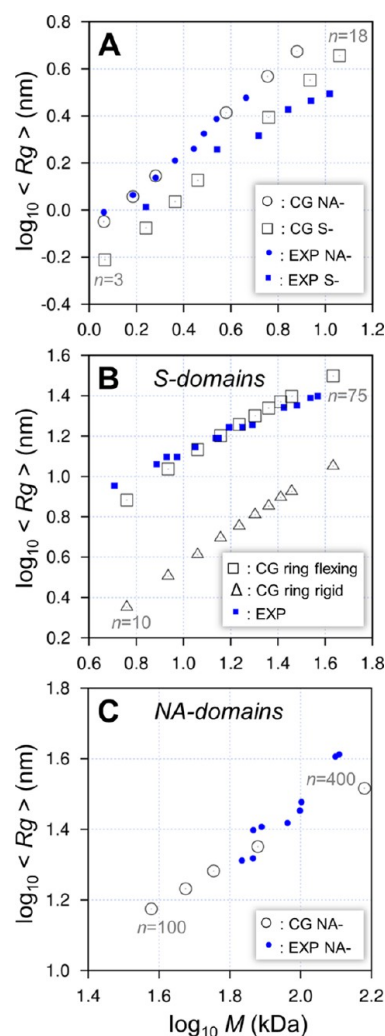
**Figure 9.** Precise 3D-properties from rapid coarse-grained simulations of heparan sulfate oligosaccharides. Undecorated (NA-) and sulfate-decorated (S-) domains, containing 2–10 residues, were modeled using 5  $\mu$ s explicit solvent all-atom (AA) and coarse-grained (CG) dynamics (open and closed symbols, respectively). Computed average radii of gyration,  $R_g$  (A) and end-to-end lengths,  $l$ , (B) are plotted against molecular mass ( $M$ ). Calculated standard deviations from the all-atom and coarse-grained simulations were identical ( $\pm 0.01$  nm),  $n$  is one disaccharide (see Figure 1 for structures).

of explicitly modeled decasaccharides were used, and for completeness the termini were modeled using parameters for equivalent residues from the all-atom simulations. Coarse-grained dynamics of large heparan sulfate polymers (undecorated and fully sulfate-decorated carbohydrates in the range 6–800 residues) were extended to 5  $\mu$ s each using the coarse-grained model and derived parameters.

The accuracy of structural predictions for these large polymers were confirmed by comparing calculated radii of gyration with X-ray scattering measurements for heparan sulfate polymers (undecorated and sulfate-decorated regions) containing 6–128 monosaccharides (Figure 10A,B)<sup>14,15,44</sup> and our observations from light scattering for undecorated domains in the range 400–800 residues (Figure 10C). The coarse-grained simulations predicted the NA-domains to be more extended than S-domains of equivalent molecular weight, which concurs with the previous experimental observations (Figure 10A).

Exclusion of the ring puckering degree of freedom, wherein the monosaccharides were fixed in their most populated chair pucker, had a small effect on predicted radii of gyration for NA-domains containing metastable pyranose rings (Supporting Information, Figure S8). Here, removing ring motions in the coarse-grained simulations caused only a slight increase in predicted radii of gyration. However, this was dramatically not the case for large protein-binding S-domains, which instead contain the highly flexible pyranose IdoA2S. In this case, the absence of puckering in the coarse-grained simulations resulted in significantly reduced computed radii of gyration values, which are clearly not in agreement with experiment, Figure 10B.

Therefore, while the effect of pyranose puckering on 3D-shape in short heparin and heparan sulfate sequences (e.g. tetrasaccharides)<sup>40</sup> has been proposed to be negligible, consideration of ring flexing in longer chains (and by implication other carbohydrate polymers containing flexible monosaccharides) is essential to correctly predict their macroscopic 3D-properties. Our computational data suggest that microsecond pyranose puckering is responsible for the hitherto unexplained difference between experimentally observed radii of gyration measurements for NA- and S-domains. This emphasizes that accurate simulations of carbohydrate

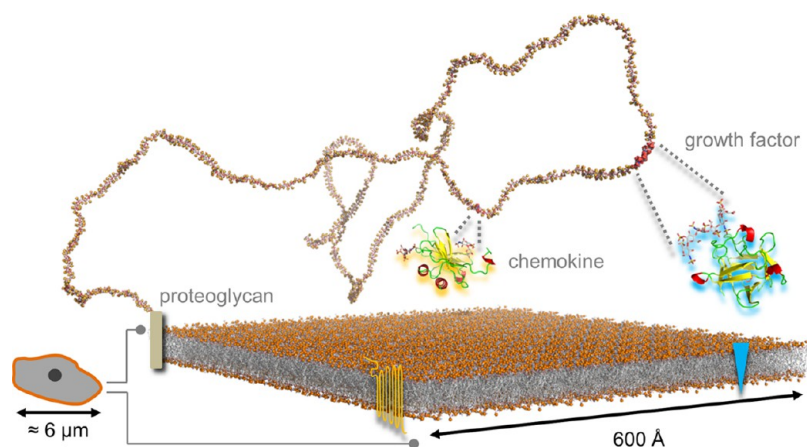


**Figure 10.** Fast, accurate 3D-predictions from 5  $\mu$ s coarse-grained (CG) dynamics of heparan sulfate biopolymers. Average calculated (CG) and experimental (EXP) radii of gyration ( $R_g$ ) values (open and closed symbols, respectively) are plotted against molecular mass ( $M$ ) for (A) S-domains (S-) and NA-domains (NA-) in the range 6–36 residues,<sup>14,15</sup> (B) S-domains in the range 20–150 residues,<sup>44</sup> where open triangles show data from coarse-grained calculations excluding ring flexing (puckers were fixed as rigid chairs), and (C) NA-domains of 200–800 residues in length;  $n$  is one disaccharide.

biopolymers requires consideration of both glycosidic linkage and pyranose ring flexibility.

The heparanome and the wider glycome contain too many sequence permutations to practically characterize them using biophysical experiments and rigorous simulations (i.e., all-atom, explicit solvent, and long-time scale molecular dynamics). While currently available mesoscale approaches based on nanosecond simulations and excluding sugar ring puckering have provided useful insights,<sup>24,25</sup> the simulations and experiments performed here clearly demonstrate that more realistic computational methods (which include the effect of pyranose ring puckering) are necessary for studying 3D structure–function relationships in the heparanome, and potentially in other carbohydrates.

Results herein further intimate that approximate representations of noncarbohydrate biopolymers, for example, proteins and nucleic acids, must also encapsulate equilibrium atomic motions of complex structural degrees of freedom that define



**Figure 11.** Schematic mammalian extracellular environment including an example heparan sulfate chain containing 800 residues and extracted from a 5  $\mu$ s coarse-grained simulation. Mammalian cell surfaces exhibit large sulfated carbohydrates (e.g., heparan, chondroitin, dermatan, and keratan sulfates), which stem from transmembrane proteoglycans and interact with numerous proteins via their sulfate-decorated domains. A lipid bilayer and heparan sulfate binding proteins, a chemokine and basic fibroblast growth factor, are shown (not to scale). The physical dimensions of a typical mammalian cell are illustrated (left).

specific molecular interactions such as drug–protein binding,<sup>21</sup> high-order 3D structure (e.g., in DNA and RNA)<sup>45</sup> and ultimately cell physiology. With recent access to microsecond simulations, current hypotheses based on nanosecond predictions may need reassessment.

## CONCLUSIONS

Unbiased microsecond dynamics simulations of explicitly solvated oligosaccharides from the mammalian heparanome predicted that the effect of polymerization was to alter the dynamic equilibrium of ring puckers in IdoA2S residues. This trend was already present (but overlooked) in previously published experimental 3D-data. Furthermore, both chain position and the adjacent monosaccharide chemistry (e.g., sulfate groups) were predicted to modulate IdoA2S puckering equilibria, which are possible mechanisms for sequence-specific protein recognition of heparan sulfate S-domains and could contribute to the vast functional diversity of the heparanome. Thus, development of potent heparin mimetics may require accurate 3D emulation of this behavior. Microsecond glycosidic linkage torsion distributions were predicted to be unaffected by sequence and hexosamine residues had relatively stable puckers (cf. IdoA2S) in the modeled oligosaccharides, suggesting that sulfate charge is the dominant physical property underlying their bioactivity.

The computed long-time scale atomic motions in heparan sulfate NA- and S-domains (glycoside linkage and pyranose ring flexing) were used to underpin a coarse-grained model that is uniquely capable of accurately simulating macromolecular sugars, such as those found on cell surfaces (Figure 11), within hours on standard computer hardware while simultaneously including atomic detail of the key 3D dynamic carbohydrate degrees of freedom. Because inclusion of ring flexing in our coarse-grained model was required to predict experimental 3D data for large heparan sulfate chains, we propose that puckering plays a role in encoding the numerous discrete biological functions associated with the heparanome, which ultimately manifest via molecular interactions and 3D shape. Replicating such predictive dynamical structural models, including long-time scale atomic motions, will be essential for accurate computational representations of other cellular components

(e.g., proteins and nucleic acids) and structure-based design of therapeutics and biotechnological devices.

## ASSOCIATED CONTENT

### Supporting Information

Carbohydrate 3D definitions (S1), detailed analyses of all-atom and coarse-grained simulations (S2–S8), and example coarse-grained parameters (s9) are provided. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Tel.: +44 (0)161 306 4199. Fax: +44 (0)161 306 5201. E-mail: [andrew.almond@manchester.ac.uk](mailto:andrew.almond@manchester.ac.uk)

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Financial support was provided by the U.K. Biotechnology and Biological Sciences Research Council (Grant BB/J00040X/1).

## REFERENCES

- (1) Gama, C. I.; Tully, S. E.; Sotogaku, N.; Clark, P. M.; Rawat, M.; Vaidehi, N.; Goddard, W. A., 3rd; Nishi, A.; Hsieh-Wilson, L. C. *Nat. Chem. Biol.* **2006**, *2*, 467.
- (2) Liu, J.; Pedersen, L. C. *Appl. Microbiol. Biotechnol.* **2007**, *74*, 263.
- (3) Gorski, B.; Stringer, S. E. *Trends Cell. Biol.* **2007**, *17*, 173.
- (4) Parish, C. R. *Nat. Rev. Immunol.* **2006**, *6*, 633.
- (5) Sasisekharan, R.; Shriver, Z.; Venkataraman, G.; Narayanasami, U. *Nat. Rev. Cancer* **2002**, *2*, 521.
- (6) Smits, N. C.; Kurup, S.; Rops, A. L.; ten Dam, G. B.; Massuger, L. F.; Hafmans, T.; Turnbull, J. E.; Spillmann, D.; Li, J. P.; Kennel, S. J.; Wall, J. S.; Shworak, N. W.; Dekhuijzen, P. N.; van der Vlag, J.; van Kuppevelt, T. H. *J. Biol. Chem.* **2010**, *285*, 41143.
- (7) Bishop, J. R.; Schuksz, M.; Esko, J. D. *Nature* **2007**, *446*, 1030.
- (8) Werz, D. B.; Ranzinger, R.; Herget, S.; Adibekian, A.; von der Lieth, C. W.; Seeberger, P. H. *ACS Chem. Biol.* **2007**, *2*, 685.
- (9) Lamanna, W. C.; Kalus, I.; Padva, M.; Baldwin, R. J.; Merry, C. L. R.; Dierks, T. *J. Biotechnol.* **2007**, *129*, 290.
- (10) Gandhi, N. S.; Mancera, R. L. *Drug Discovery Today* **2010**, *15*, 1058.
- (11) Liu, H.; Zhang, Z.; Linhardt, R. J. *Nat. Prod. Rep.* **2009**, *26*, 313.

- (12) de Kort, M.; Buijsman, R. C.; van Boeckel, C. A. *Drug Discovery Today* **2005**, *10*, 769.
- (13) Rabenstein, D. L. *Nat. Prod. Rep.* **2002**, *19*, 312.
- (14) Khan, S.; Gor, J.; Mulloy, B.; Perkins, S. J. *J. Mol. Biol.* **2010**, *395*, 504.
- (15) Khan, S.; Rodriguez, E.; Patel, R.; Gor, J.; Mulloy, B.; Perkins, S. *J. J. Biol. Chem.* **2011**, *286*, 24842.
- (16) Mobli, M.; Nilsson, M.; Almond, A. *Glycoconjugate J.* **2008**, *25*, 401.
- (17) Sattelle, B. M.; Almond, A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 5843.
- (18) Sattelle, B. M.; Hansen, S. U.; Gardiner, J.; Almond, A. *J. Am. Chem. Soc.* **2010**, *132*, 13132.
- (19) Sattelle, B. M.; Almond, A. *Glycobiology* **2011**, *21*, 1651.
- (20) Sattelle, B. M.; Bose-Basu, B.; Tessier, M.; Woods, R. J.; Serianni, A. S.; Almond, A. *J. Phys. Chem. B* **2012**, *116*, 6380.
- (21) Hricovini, M.; Guerrini, M.; Bisio, A.; Torri, G.; Petitou, M.; Casu, B. *Biochem. J.* **2001**, *359*, 265.
- (22) Das, S. K.; Mallet, J. M.; Esnault, J.; Driguez, P. A.; Duchaussoy, P.; Sizun, P.; Herault, J. P.; Herbert, J. M.; Petitou, M.; Sinay, P. *Angew. Chem., Int. Ed.* **2001**, *40*, 1670.
- (23) Schlick, T. *Biol. Reprod.* **2009**, *1*, 51.
- (24) Bathe, M.; Rutledge, G. C.; Grodzinsky, A. J.; Tidor, B. *Biophys. J.* **2005**, *88*, 3870.
- (25) Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195.
- (26) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622.
- (27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (28) Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; De Fabritiis, G. *J. Chem. Inf. Model.* **2010**, *50*, 397.
- (29) Krautler, V.; Van Gunsteren, W. F.; Hunenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501.
- (30) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354.
- (31) Haasnoot, C. A. G.; Deleeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783.
- (32) Hricovini, M.; Bizik, F. *Carbohydr. Res.* **2007**, *342*, 779.
- (33) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. J. Comput. Chem.* **2005**, *26*, 1668.
- (34) Oborsky, P.; Tvaroska, I.; Kralova, B.; Spiwok, V. *J. Phys. Chem. B* **2013**, *117*, 1003.
- (35) Nieto, L.; Canales, A.; Gimenez-Gallego, G.; Nieto, P. M.; Jimenez-Barbero, J. *Chemistry* **2011**, *17*, 11204.
- (36) Blundell, C. D.; Roberts, I. S.; Sheehan, J. K.; Almond, A. *J. Mol. Microbiol. Biotechnol.* **2009**, *17*, 71.
- (37) Mulloy, B.; Forster, M. J.; Jones, C.; Davies, D. B. *Biochem. J.* **1993**, *293* (Pt3), 849.
- (38) Altona, C.; Haasnoot, C. A. G. *Org. Magn. Reson.* **1980**, *13*, 417.
- (39) Sue, S. C.; Brisson, J. R.; Chang, S. C.; Huang, W. N.; Lee, S. C.; Jarrell, H. C.; Wu, W. G. *Biochemistry* **2001**, *40*, 10436.
- (40) Jin, L.; Hricovini, M.; Deakin, J. A.; Lyon, M.; Uhrin, D. *Glycobiology* **2009**, *19*, 1185.
- (41) Mulloy, B.; Forster, M. J. *Glycobiology* **2000**, *10*, 1147.
- (42) Sattelle, B. M.; Shakeri, J.; Roberts, I. S.; Almond, A. *Carbohydr. Res.* **2010**, *345*, 291.
- (43) Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Torri, G.; Casu, B.; Gatti, G.; Jacquinet, J. C.; Sinay, P.; Petitou, M.; Choay, J. *J. Am. Chem. Soc.* **1986**, *108*, 6773.
- (44) Pavlov, G.; Finet, S.; Tatarenko, K.; Korneeva, E.; Ebel, C. *Eur. Biophys. J.* **2003**, *32*, 437.
- (45) Rich, A. *Nat. Struct. Biol.* **2003**, *10*, 247.