

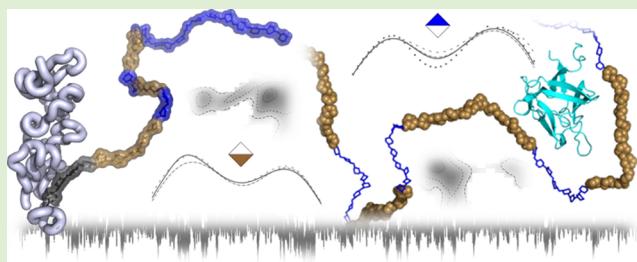
# Proteoglycans and Their Heterogeneous Glycosaminoglycans at the Atomic Scale

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

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

## Supporting Information

**ABSTRACT:** Proteoglycan spatiotemporal organization underpins extracellular matrix biology, but atomic scale glimpses of this microarchitecture are obscured by glycosaminoglycan size and complexity. To overcome this, multi-microsecond aqueous simulations of chondroitin and dermatan sulfates were abstracted into a prior coarse-grained model, which was extended to heterogeneous glycosaminoglycans and small leucine-rich proteoglycans. Exploration of relationships between sequence and shape led to hypotheses that proteoglycan size is dependent on glycosaminoglycan unit composition but independent of sequence permutation. Uronic acid conformational equilibria were modulated by adjacent hexosamine sulfonation and iduronic acid increased glycosaminoglycan chain volume and rigidity, while glucuronic acid imparted chain plasticity. Consequently, block copolymeric glycosaminoglycans contained microarchitectures capable of multivalent binding to growth factors and collagen, with potential for interactional synergy at greater chain number. The described atomic scale views of proteoglycans and heterogeneous glycosaminoglycans provide structural routes to understanding their fundamental signaling and mechanical biological roles and development of new biomaterials.



## INTRODUCTION

Proteoglycans (PGs) are vital constituents of cellular membranes and extracellular matrices, where they confer tissue mechanical properties, facilitate constitutive cell–cell interactions, and are implicated in neurodegeneration, arthritis, and cancer.<sup>1–4</sup> Human PGs, such as bikunin and decorin,<sup>5,6</sup> comprise a core protein plus one or more covalently linked glycosaminoglycan (GAG) polymers, which possess functionally important patterns of postsynthetic modification. In particular, domains rich in sulfate and iduronic acid maintain connective tissue integrity and mediate signal transduction via multivalent interactions with collagen and growth factors.<sup>7–10</sup> However, the molecular basis for such bioactivity remains enigmatic, in part, because current representations of PGs and attendant GAGs lack atomic scale conformational detail.

Obtaining high resolution 3D data for PGs is hampered by GAG sequence heterogeneity, conformational flexibility, and molecular size. Electron microscopy and small-angle X-ray scattering have provided low resolution insights.<sup>11,12</sup> However, atomic resolution crystallographic studies have been limited to either the protein core (e.g., for small leucine-rich repeat PGs<sup>13–15</sup>) or isolated GAG components (e.g., fiber X-ray scattering data for chondroitin-4-sulfate<sup>16</sup>). Modeling offers an avenue to progress, but accurate fine-grained simulations (i.e., including all atoms and water solvent) have been restricted to oligosaccharides and short sub- $\mu$ s time scales (notwithstanding approaches that distort biologically important kinetics<sup>17</sup>). This limitation has, to some extent, been recently overcome by the

use of application-specific integrated circuits and graphics processing units, which have enabled unbiased multi- $\mu$ s computational studies.<sup>18,19</sup> However, application of fine-grained simulations to even small PGs will remain out of reach for the foreseeable future.

Less computationally demanding coarse-grained (CG) carbohydrate models have been successfully developed,<sup>20,21</sup> but in these, only the polymer glycosidic linkage degrees of freedom remain and functionally important pyranose ring flexing (or puckering) is neglected. Recently, we developed a CG approach that includes both carbohydrate linkage and ring motions.<sup>22</sup> In this method, linkage torsional and ring puckering free energy surfaces are calculated from microsecond time scale fine-grained simulations (containing implicit information on sugar substituents) and used to parametrize a simplified CG representation, which can be used to rapidly predict the motions of polymers using numerical integration. Although promising, the method's ability to model heterogeneous GAG sequences or glycoconjugates, such as PGs, has not been established. Therefore, at present, neither experimental nor theoretical techniques have been shown capable of characterizing heterogeneous GAGs or PGs at high resolution. Consequently, the relationships between GAG postsynthetic modification, domain organization, and PG dynamic shape, and

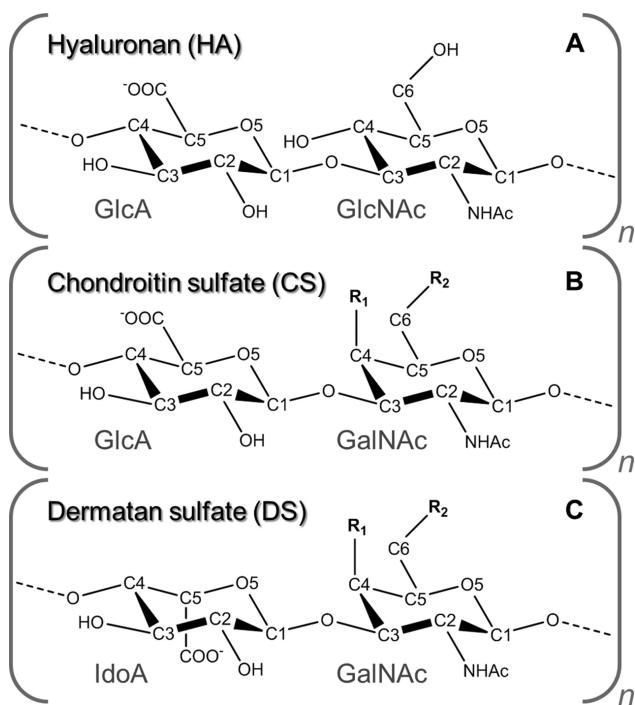
Received: December 18, 2014

Revised: January 30, 2015

Published: February 2, 2015

their role in multivalent interactions with proteins, remain enigmatic at the atomic scale.

Here, we overcome this inability by broadening the applicability of our CG approach to characterize heterogeneous GAGs and PGs both accurately and at the atomic scale. Unbiased microsecond fine-grained simulations of hyaluronan and six key sequence units from chondroitin and dermatan sulfates (Figure 1) were performed to investigate the effect of



**Figure 1.** Chemical structures of hyaluronan, chondroitin, and dermatan repeating disaccharide units. All-atom explicit solvent (or fine-grained) 10  $\mu$ s simulations were performed for model decasaccharides ( $n = 5$ ) of (A) hyaluronan, (B) the chondroitin O-unit ( $R_1 = R_2 = \text{OH}$ ), A-unit ( $R_1 = \text{OSO}_3^-$ ,  $R_2 = \text{OH}$ ), and C-unit ( $R_1 = \text{OH}$ ,  $R_2 = \text{OSO}_3^-$ ), and (C) the dermatan iO-unit ( $R_1 = R_2 = \text{OH}$ ), iA-unit ( $R_1 = \text{OSO}_3^-$ ,  $R_2 = \text{OH}$ ), and iC-unit ( $R_1 = \text{OH}$ ,  $R_2 = \text{OSO}_3^-$ ).

postsynthetic modifications on molecular motions. Using these data to further parametrize our CG model enabled simulations of GAGs to be performed on complex heterogeneous sequences, in general, and physiologically relevant block copolymeric structures, in particular. In further key steps, the CG approach was extended to PGs, and by exploiting GAG sequencing information and X-ray diffraction data, realistic atomic models of small leucine-rich repeat PGs were constructed. Where possible, the derived models were validated by comparison with hydrodynamic measurements. The results represent the first accurate atomistic views of heterogeneous GAGs and PGs, allowing relationships between sequence, atomic scale shape, and dynamics to be explored in these complex extracellular molecules. The reported models transcend simple schematic representations of PGs and advance understanding of their assembly and protein interactions toward clarifying the roles of PGs in health and disease and furthering their use as therapeutics.

## MATERIALS AND METHODS

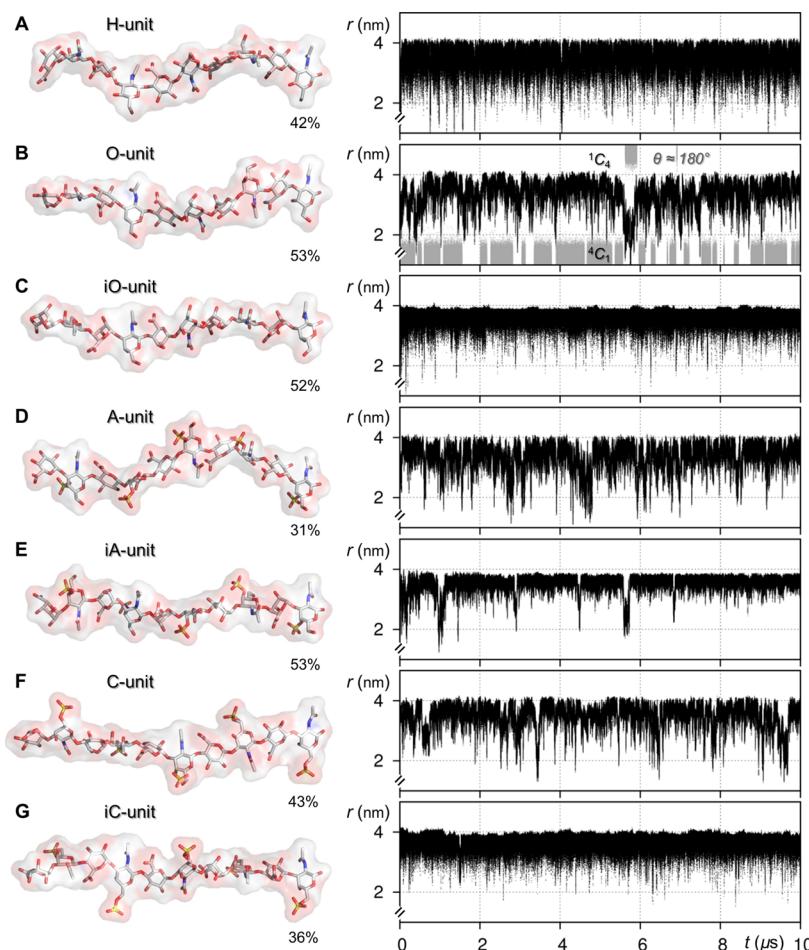
**Materials.** Medical grade hyaluronan was obtained from Genzyme Corp. (Boston, U.S.A.). Shark cartilage chondroitin sulfate (cat. no.

C4384), bovine trachea chondroitin sulfate (cat. no. 27042), and porcine intestinal mucosa dermatan sulfate (cat. no. C3788) were obtained from Sigma-Aldrich (Dorset, U.K.). Two disaccharide samples were produced by bacterial chondroitinase digestion of chondroitin and dermatan sulfate and isolated using high resolution gel filtration and ion exchange chromatography (Iduron Ltd., Manchester, U.K.) to use as reference compounds for NMR assignment of polysaccharides. Each disaccharide sample contained a mixture of unsulfonated uronic acids (UA) with unsaturated UA C4–C5 bonds ( $\Delta\text{UA}$ ) and a single  $O$ -sulfate at either C4 (4S) or C6 (6S) on the galactosamine (GalNAc); chemically they were  $\Delta\text{UA}\beta(1 \rightarrow 3)\text{GalNAc-4S}$  (cat. no. CD002) and  $\Delta\text{UA}\beta(1 \rightarrow 3)\text{GalNAc-6S}$  (cat. no. CD003).

**NMR Spectroscopy.** For NMR analyses, sulfated polysaccharide and disaccharide samples (see Materials) were prepared from lyophilized material reconstituted in 595  $\mu$ L of D<sub>2</sub>O (CK Gas Products Ltd., Ibstock, U.K.) and 5  $\mu$ L of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). All samples were centrifuged at 12000g for 10 min before transfer to 5 mm NMR tubes (part no. 542-PP-7, Wilmad-LabGlass, Vineland, U.S.A.). Each was adjusted to pD 7.3–7.4, and final concentrations were in the range 0.5–2.0 mM. The [<sup>1</sup>H]-1D, [<sup>1</sup>H,<sup>13</sup>C]-HSQC, [<sup>1</sup>H]-COSY, and [<sup>1</sup>H,<sup>13</sup>C]-HSQC-TOCSY spectra were recorded at 25 °C with a <sup>1</sup>H frequency of 800 MHz (Avance III spectrometer with TCI <sup>1</sup>H-<sup>13</sup>C/<sup>15</sup>N-<sup>1</sup>H cryoprobe equipped with z-gradients, Bruker Corp., Billerica, U.S.A.). Heteronuclear experiments were gradient selected and sensitivity enhanced; chemical shifts were referenced relative to internal DSS. Spectra were processed using Topspin software (Bruker Corp., Billerica, U.S.A.); 1D experiments were apodized by an exponential function (0.3 Hz decay constant) and zero-filled to twice the acquired points; 2D spectra were linear predicted in the indirect dimension to twice the acquired points and apodized with a 90° sine-bell function in both dimensions. Assignments for two sulfated disaccharides (see Materials) were performed with reference to prior data for similar compounds.<sup>23–27</sup> Glycosaminoglycan polysaccharide <sup>1</sup>H and <sup>13</sup>C assignments were aided by comparison with the spectra and assignments from these disaccharides. In polysaccharide [<sup>1</sup>H,<sup>13</sup>C]-HSQC spectra, overlap of GalNAc C4 and C6 peaks hindered integration and precluded estimation of 6-O-sulfate; in this case, C5 peak integrals were used to the estimate relative levels of sulfate.

**Laser Light Scattering.** Glycosaminoglycan polysaccharides (4 mg) were dissolved in 0.5 mL of ammonium bicarbonate (100 mM) for 24 h at 4 °C and then centrifuged at 12000g for 10 min prior to application to a Superose 6 10/300GL gel filtration column (GE Healthcare, Little Chalfont, U.K.) running at a flow rate of 0.75 mL/min in the same eluent. Samples eluting from the column at room temperature passed through an in-line DAWN HELEOS-II laser photometer with a 658 nm laser (Wyatt Technology Corp., Santa Barbara, U.S.A.), WyattQELS detector model number Q-03 and an Optilab rEX refractometer; a refractive index increment ( $dn/dc$ ) of 0.16 mL g<sup>-1</sup> was used throughout.<sup>28</sup> Light scattering and concentration during elution were recorded using ASTRA software (v5.3.4.13). Multiangle laser light scattering was used to determine weight-averaged molecular masses ( $M$ ) and radius of gyration ( $R_g$ ) using analysis modules within the ASTRA software. For molecules too small to accurately determine  $R_g$  quasi-elastic light scattering was used to measure hydrodynamic radii ( $R_h$ ), again using analysis modules within the ASTRA software. Conversion of calculated  $R_g$  to  $R_h$  was achieved using the formula  $R_g = 1.5R_h$ , derived previously for a random coil.<sup>29</sup>

**Fine-Grained Simulations.** Fine-grained molecular modeling of seven GAG decasaccharides and the GAG-PG linker trisaccharide, Gal $\beta(1 \rightarrow 3)\text{Gal}\beta(1 \rightarrow 4)\text{Xyl}$ , were performed using the GLYCAMS06 force-field topologies and parameters for carbohydrates.<sup>30</sup> Each was built using the AMBER12<sup>31</sup> tool “tleap” and then added to cubic boxes of TIP3P water,<sup>32</sup> resulting in side lengths in the range 6.2–7.2 nm ( $\approx$ 12000 water molecules). Solute atoms were positioned at least 1.2 nm from the solvent box edge and charge neutral assemblies were achieved by adding the appropriate number of explicit Na<sup>+</sup> ions. Following initial conjugate-gradient energy minimization (1000 steps),

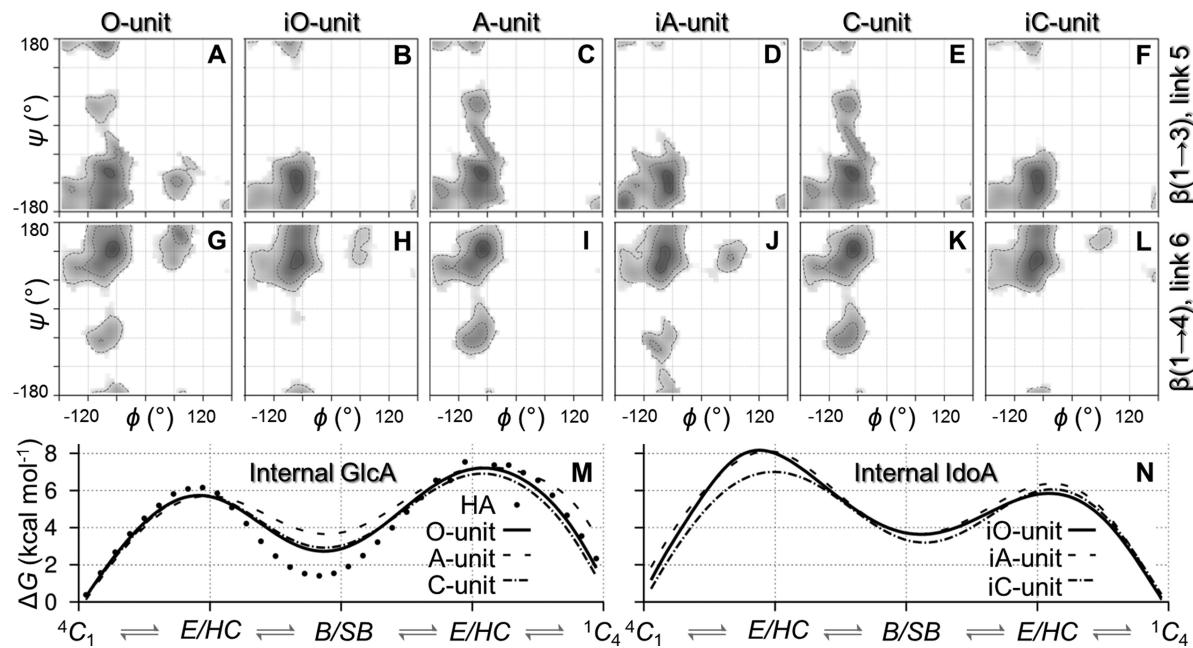


**Figure 2.** (A–G) Predominant conformers (left, with their percentage occupancy) and microsecond conformational dynamics in 10  $\mu$ s fine-grained simulations of GAG decasaccharides, measured using the distance ( $r$ , right) between the two endmost glycosidic oxygen atoms. The structures comprise homogeneous repeats of (A) hyaluronan (H-unit), (B, D, F) chondroitin sulfate containing glucuronic acid (O-, A-, and C-units), and (C, E, G) dermatan sulfate containing iduronic acid (iO-, iA-, and iC-units). The correlation between  $r$  and pyranose pucker is exemplified in gray (B) by the Cremer-Pople azimuthal angle ( $\theta$ ) for an internal glucuronic acid ring (number four counting from the reducing end).

each assembly was heated to 298 K and equilibrated in the NPT ensemble for 20 ns prior to 10.3  $\mu$ s unbiased NVT production molecular dynamics, performed using ACEMD software<sup>33</sup> and GTX TITAN graphics processing unit hardware (Nvidia Corp., Santa Clara, U.S.A.), as described previously.<sup>19,34</sup> The initial 300 ns was discarded and for the remaining 10  $\mu$ s, atomic coordinate data were saved at 10 ps intervals for analyses. Molecular properties and axial helical rises (per disaccharide) were computed from complete 10  $\mu$ s trajectories (i.e., 1000000 data sets) using standard equations described previously.<sup>35</sup> The distance  $r$  (graphed in Figure 2) was calculated over the four internal disaccharides within decasaccharides (i.e., between reducing-end and non-reducing-end glycosidic linkage oxygen atoms). Predominant decasaccharide conformers (illustrated in Figure 2) were derived using the AMBER12<sup>31</sup> tool “ptraj” and a hierarchical 3D-clustering algorithm with a sampling frequency of 100 frames (10000 in total). Glycosidic torsion angles  $\phi$  and  $\psi$  in (1  $\rightarrow$   $n$ ) linkages (where  $n = 3$  or 4) were computed using the IUPAC definitions:  $\phi$  defined as O<sub>5</sub>-C<sub>1</sub>-O-C <sub>$n$</sub>  and  $\psi$  defined as C<sub>1</sub>-O-C <sub>$n$</sub> -C<sub>( $n$ -1)</sub>. A charge neutral 1  $\mu$ s simulation of bikunin with a GAG-PG linker trisaccharide attached at serine residue 10 (Ser-10) was also performed using the methods detailed above; in this case both the GLYCAMS06 and AMBER FF03<sup>36</sup> force-fields were used and the proteoglycan was solvated in a cubic box of TIP3P water with a side length of 7.3 nm.

**Coarse-Grained Simulations.** The coarse-grained (CG) model, utilizing empirical energy functions for glycosidic linkage and ring pucker, has been described previously.<sup>22</sup> Two-dimensional free energy surfaces for glycosidic linkage torsions and ring azimuthal and

meridian angles were derived from fine-grained 10  $\mu$ s simulations and parametrization of the CG empirical energy functions was achieved by fitting to these surfaces, as described previously.<sup>22</sup> For validation purposes only, all linkages and all rings present in decasaccharides were parametrized from their respective fine-grained simulation equivalent. For simulations of polymers,  $\beta(1 \rightarrow 3)$ -linked disaccharide units were parametrized from data for the central two linkages and two rings of respective fine-grained decasaccharides. These units were stitched together via their  $\beta(1 \rightarrow 4)$  linkages to produce model chains, that is, the constituent disaccharide [ $\beta(1 \rightarrow 4)$ U $\beta(1 \rightarrow 3)$ N] is a representative unit with two linkages and two rings parametrized from a single fine-grained simulation (here, U and N represent uronic acid and hexosamine, respectively). If the ends of the chain did not naturally fit within this disaccharide register, then individual linkages and rings (with identical parameters) were appended (N to the nonreducing end and  $\beta(1 \rightarrow 4)$ U to the reducing end). This allowed the construction of heterogeneous polymers of arbitrary length and sequence. All CG simulations were performed for 10  $\mu$ s, accomplished by integrating the Langevin equation of motion independently for the linkages and rings, using the Leapfrog algorithm and a time-step of 2.5 fs, described previously.<sup>22</sup> During the simulations, only the positions of the glycosidic oxygen atom and pyranose ring atoms were calculated (i.e., ring substituents were not explicitly modeled) and coordinates were saved at regular intervals for analyses. No steric interactions were calculated during CG simulations (other than those implicit in the parameters). Therefore, when it was deemed necessary, and after completion of each



**Figure 3.** Glycosidic linkage and pyranose ring motions in 10  $\mu$ s unbiased fine-grained simulations of model homogeneous chondroitin sulfate and dermatan sulfate decasaccharide subunits. Two-dimensional ( $\phi, \psi$ ) surfaces of free energy (contours at 2, 4 and 6 kcal mol $^{-1}$ ) for (A–F) the internal  $\beta(1 \rightarrow 3)$  linkage 5 and (G–L) the adjacent  $\beta(1 \rightarrow 4)$  linkage 6 (counting from the reducing end). (M, N) Puckering free energy graphs for uronic acid ring 6 (counting from the reducing end), with ring conformational space labeled as C, chair; E, envelope; HC, half-chair; B, boat; and SB, skew-boat.

simulation, coordinate sets with steric clashes were identified by detecting those in which CG atoms had approached within 1 nm (excluding those in the same or neighboring residues) and then removed from the simulated ensemble. Radius of gyration ( $R_g$ ) calculations on polymers were performed without adding any substituents to the carbohydrate rings (test calculations on polymers showed this had a negligible effect on the outcome).

**Proteoglycan Simulations.** Models were constructed using Cartesian coordinates for small leucine-rich repeat PG core proteins (bikunin, decorin and biglycan); these were obtained from the Protein Data Bank (identification codes: 1BIK,<sup>13</sup> 1XKU,<sup>15</sup> and 2FT3,<sup>14</sup> respectively). The decorin dimer was built by applying a transformation matrix (the PDB file BIOMT record). Terminal amino acids missing from these experimental data (residues 1–24 in bikunin and biglycan, 1–21 in decorin) were constructed using a knowledge-based algorithm with default settings in the “Loop Modeler” feature of MOE software (Chemical Computing Group Inc., Montreal, Canada). In each case the lowest energy loop conformer was selected. In the absence of experimental coordinates, this approach provides an energetically favorable N-terminal loop conformer that is based on experimental 3D-structures from the Protein Data Bank. Subsequently, all experimentally refined nonprotein atoms (i.e., ions, waters, carbohydrate glycosylation fragments) and all amino acid substituents were removed, leaving only Cartesian coordinates of the protein backbone atoms.

Polymer chains representing GAGs with specific sequences (as described above) were attached to a linker trisaccharide and onto previously identified Ser residues (within the terminal loop regions) to form the final modeled sequence:  $[\beta(1 \rightarrow 4)U\beta(1 \rightarrow 3)N]_n\beta(1 \rightarrow 4)U\beta(1 \rightarrow 3)Gal\beta(1 \rightarrow 3)Gal\beta(1 \rightarrow 4)-Xyl-Ser$ . Parameters for the CG model of the linker were derived from a fine grained simulation of the trisaccharide (see above) and the parameters for the CG model of the GAG portion were as described above. The glycoprotein linkages between Xyl and Ser were fixed using the following torsional angles ( $C-C_{\alpha}-C_{\beta}-O_G$ :  $\pm 60^\circ$ ,  $C_{\alpha}-C_{\beta}-O_G-C_1$ :  $+100^\circ$ ,  $C_{\beta}-O_G-C_1-O_S$ :  $-60^\circ$ ), which are consistent with previous NMR measurements<sup>37</sup> and a 1  $\mu$ s fine-grained simulation of bikunin with an attached linker trisaccharide. The PG core protein was also kept in a fixed geometry while 10  $\mu$ s CG simulations were performed on the carbohydrate chains. Radii of

gyration ( $R_g$ ) calculations on PGs were performed without adding substituents to the GAG pyranose rings or side chains to the PG core protein amino acids. Additional post-translational modifications beyond the GAG chains, such as N-linked glycans, were assumed to be a minor contributor to  $R_g$  and were not considered in CG simulations for practical reasons. Conversion of calculated  $R_g$  to  $R_h$  was achieved using the formula  $R_g = 0.77R_h$  for a constant density sphere.

## RESULTS AND DISCUSSION

**Compositional Analyses and Nomenclature.** The sulfate, D-glucuronic acid (GlcA), and L-iduronic acid (IdoA) content of shark cartilage and bovine trachea chondroitin sulfate (CS) and porcine intestinal dermatan sulfate (DS) were assayed using  $^{13}\text{C}$  NMR spectra, Figure S1. The GAG compositional nomenclature in this work follows prior descriptions,<sup>38</sup> wherein CS  $\beta(1 \rightarrow 3)$  linked disaccharide units comprising unsulfonated, 4-O-sulfo and 6-O-sulfo N-acetyl-D-galactosamine (GalNAc) are designated O, A, and C, respectively. In DS these are prefixed with “i” to indicate the presence of IdoA (cf. GlcA in CS). For shark CS, the proportions of O-, A-, and C-units were estimated to be 20, 20, and 60%, respectively. Similarly, in bovine CS, NMR analyses indicated approximate levels of 20, 60, and 20% for these units. Porcine DS was estimated to contain a mix of GalNAc 4-O-sulfo (A/iA) and 6-O-sulfo (C/iC) units at 80 and 20%, respectively; the GlcA/IdoA ratio was approximated at 1:4. It was concluded that neither 2-O-sulfo IdoA nor 4,6-O-sulfo GalNAc were present in substantial quantity (greater than 10%) in shark, bovine, or porcine GAG samples. These findings concur with previous chromatographic studies of similar but not identical biological materials,<sup>39</sup> and for the purposes of this work, shark (CS), bovine (CS), and porcine (DS) materials were deemed to be repeating blocks of [A][C]<sub>3</sub>[O], [C]-[A]<sub>3</sub>[O], and [A][iA]<sub>4</sub>, respectively. Medical grade hyaluronan (HA), with N-acetyl-D-glucosamine (GlcNAc), was assumed to

be a homogeneous (100%) repeating disaccharide, referred to here (unconventionally) as the H-unit.

### Dynamics of Subunits from Heterogeneous GAGs.

Decasaccharides comprising five repeats of subunits from HA, CS, and DS (H, O, A, C, iO, iA, and iC) were subjected to unbiased 10  $\mu$ s explicit molecular dynamics simulations, including all atoms (GAG, water solvent and counterions), referred to here as fine-grained. For each of the simulated molecules, the distance  $r$  between the two endmost glycosidic oxygen atoms (across the internal four disaccharides and thereby excluding possible end effects<sup>40</sup>) was computed. Based on time series plots of  $r$  (Figure 2), each decasaccharide had a characteristic dynamic range and frequency of motion. In all molecules, transitions occurred between numerous conformations, but those containing iduronic acid were observed to spend more time in extended geometries than their glucuronic acid equivalents (predominant shapes are illustrated in Figure 2). Further inspection and analysis revealed extremes of maximally extended chains and compact turns, with exchange being correlated to  $\mu$ s time scale linkage and ring flexing (exemplified in Figure 2B), as found previously in amylose.<sup>34</sup> It was confirmed that, in every simulation, arithmetic mean  $r$ -values for 1  $\mu$ s subsets had a standard deviation of less than 0.1 nm from the mean for the entire 10  $\mu$ s data set, substantiating that motions in the model GAG decasaccharides had approached equilibrium.

In the H-unit decasaccharide simulation, 67% of conformers were 2- to 4-fold left-handed helices and 81% had an axial rise per disaccharide ( $h$ ) of >0.82 nm. Prior X-ray fiber diffraction observations for HA include 2-, 3-, and 4-fold left-handed helices with  $h$  in the range 0.84–0.98 nm.<sup>41</sup> In the A-unit simulation, 43% of conformers were 2- to 3-fold left-handed helices, with 42% having  $h$  > 0.94 nm. Experimentally, chondroitin-4-sulfate has been observed as 2- to 3-fold left-handed helices with  $h$  in the range 0.96–0.98 nm.<sup>42</sup> The C-unit decasaccharide simulation contained 65% of conformers between right-handed 8-fold helices and left-handed 3-fold helices, with 52% having  $h$  > 0.94 nm, consistent with previously described chondroitin-6-sulfate extended left-handed 3-fold and right-handed 8-fold helices with  $h$  in the range 0.96–0.98 nm.<sup>43</sup> The iA-unit decasaccharide simulation had 58% of conformers with  $h$  in the range 0.90–0.96 nm, 44% were between right-handed 8-fold helices and left-handed 3-fold helices. Dermatan-4-sulfate has been observed to form 2-, 3-, and 8-fold helices with  $h$  in the range 0.92–0.94 nm.<sup>44</sup> Considering that these prior experiments were performed in the solid state and, hence, only a fraction of solution phase conformers could satisfy the packing constraints, it was concluded that the predicted conformer populations had a good overlap with published X-ray fiber diffraction refinements.

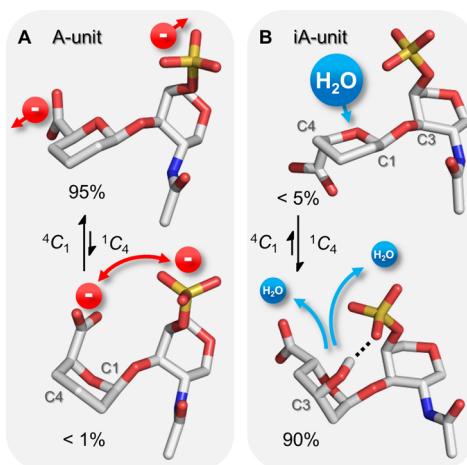
Computed glycosidic linkage free energy surfaces (Figure S2) showed all  $\beta(1 \rightarrow 3)$  linkages to be similar and  $(\phi, \psi)$  for the major minima were centered on  $\approx(-60^\circ, -120^\circ)$ , with minor conformers at  $\approx(-60^\circ, +60^\circ)$  and  $\approx(+60^\circ, -120^\circ)$ , Figure 3A–F. A slight difference in these was the relatively restricted range of the  $\beta(1 \rightarrow 3)$   $\psi$  torsion in DS decasaccharides. Furthermore, the illustrated free energy surfaces were derived from central linkages, and it should be noted that other minor differences were observed within each chain and at their ends. Constituent  $\beta(1 \rightarrow 4)$  linkages, Figure 3G–L, were almost indistinguishable and  $(\phi, \psi)$  for the major minima were centered on  $\approx(-60^\circ, 120^\circ)$ , with minor populations at  $(-120^\circ, -60^\circ)$  and  $(+60^\circ, +120^\circ)$ ; the latter was not sampled in decasaccharides of

A- or C-units. Prior NMR studies of homogeneous unsulfonated oligosaccharides have refined the H-unit  $\beta(1 \rightarrow 3)$  and  $\beta(1 \rightarrow 4)$  linkages to be  $(-68^\circ, -111^\circ)$  and  $(-71^\circ, 124^\circ)$ , respectively.<sup>45</sup> Analogous refinements for the O-unit resulted in  $(-72^\circ, -132^\circ)$  and  $(-73^\circ, 122^\circ)$ , which was concluded to be virtually identical to the A-unit.<sup>46</sup> Taken together, the computed linkage free energy surfaces for each of the decasaccharides are predicted to have surprising similarities, irrespective of their sulfate, hexosamine, and uronic acid composition, and are consistent with available solution phase experimental data.

**Molecular Mechanism for Encoding GAG Shape via Sequence.** Puckering occurred in all decasaccharide pyranose rings and converged to equilibria over 2–5  $\mu$ s (Table S3), consistent with previous simulations.<sup>19</sup> Based on an average of the four internal hexosamines in each model, all were principally  $^4C_1$  chairs (>97%) and motions were invariant of C4-position stereochemistry (GlcNAc or GalNAc) and O-sulfonation (at the C4- or C6-position). While all GlcA pyranoses were predominantly  $^4C_1$ -chairs, they also underwent  $^4C_1 \leftrightarrow ^1C_4$  exchange (e.g., Figure 3M). This behavior is qualitatively similar to prior experimentally consistent GlcA monosaccharide simulations,<sup>47</sup> but with average  $^4C_1$  populations of 64% (H-unit), 65% (O-unit), 95% (A-unit), and 75% (C-unit). The predicted ring conformational equilibria in the H- and O-unit decasaccharides were consistent with prior NMR measurements of pyranose  $^1H$ – $^1H$  vicinal couplings ( $^3J_{H,H}$ ), Table S4.<sup>46,48</sup>

The DS decasaccharide IdoA pockers were predicted to be predominantly  $^1C_4$  chairs (e.g., Figure 3N) with populations of 82% (iO-unit), 90% (iA-unit), and 56% (iC-unit). All IdoA rings underwent  $^4C_1 \leftrightarrow ^1C_4$  exchange and the mix of IdoA ring conformers produced a qualitative agreement with measured low  $^3J_{H,H}$  values (<5 Hz) for iA regions of porcine DS.<sup>49</sup> The deviation between predicted and experimentally measured  $^3J_{H,H}$  values were as described and discussed previously.<sup>47</sup> Although  $^2S_0$  was the predominant nonchair conformer, it was predicted to be a minor ensemble component (<7%) in the iO-, iA-, and iC-unit decasaccharides. This skew-boat is reportedly key for protein binding<sup>50,51</sup> but has been difficult to observe in free crystals<sup>52</sup> and was infrequently populated in prior IdoA monosaccharide simulations (<10%).<sup>47</sup>

The fine-grained simulations performed here represent the most extensive for GAG oligosaccharides to date and suggest that CS and DS uronic acid ring puckering equilibria are influenced across  $\beta(1 \rightarrow 3)$  glycosidic linkages by neighboring hexosamine sulfate groups. In particular, GalNAc 4-O-sulfonation biases GlcA and IdoA toward  $^4C_1$  and  $^1C_4$ , respectively. To investigate, conformers with both uronic acid chair pockers were extracted from A- and iA-unit decasaccharide simulations. Inspection of the A-unit at the chain center suggested that repulsion of the GalNAc 4-O-sulfate and GlcA carboxyl groups disfavored the  $^1C_4$  chair, which was not present in the favored  $^4C_1$  pucker (Figure 4A). For the iA-unit, the IdoA  $^1C_4$  pucker was stabilized by interaction of its C3-position hydroxyl substituent with the adjacent GalNAc 4-O-sulfo group (Figure 4B). This allowed water to be excluded from the IdoA ring face, which was not the case when IdoA was in the less favored  $^4C_1$  chair conformation. It is hypothesized that, within the fundamental  $\beta(1 \rightarrow 3)$ -linked GAG unit, networks of through-space interactions between uronic acid, adjacent hexosamine sulfate groups, and water establish characteristic uronic acid puckering equilibria that contribute to polymer



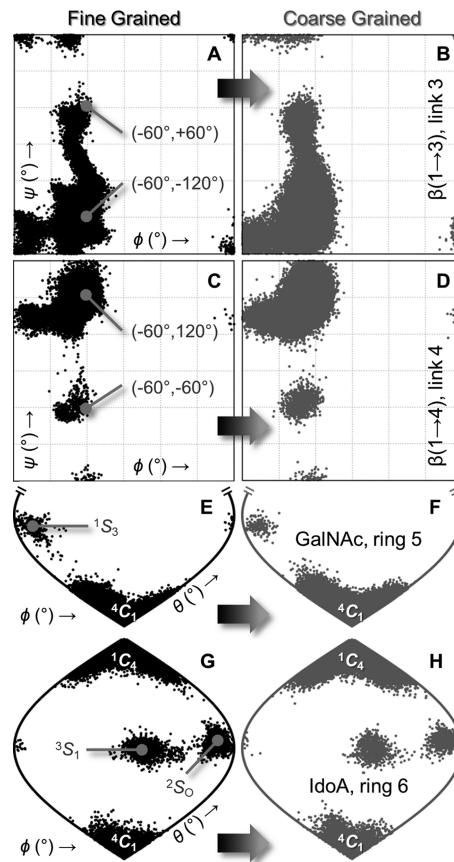
**Figure 4.** Representative conformers with the most and least favored uronic acid chair puckerers from 10  $\mu$ s unbiased fine-grained simulations of homogeneous decasaccharides comprising A-units (A) and iA-units (B). Geometries of disaccharides from the chain centers are illustrated, with putative molecular mechanisms for conformational preferences highlighted: intermolecular repulsion (red spheres), attraction (dashed line), and interactions with water solvent (blue spheres). Ring substituents are hidden for clarity, except at the C3-position in the iA-unit IdoA <sup>1</sup>C<sub>4</sub> conformer (B), highlighting a predicted through-space interaction with the adjacent hexosamine sulfate group.

hydrodynamic properties. This represents a mechanism for encoding shape via sequence and may be a key factor in both specific GAG-protein interactions and PG organization.

**Validating Simulations of Homogeneous and Heterogeneous GAG Polymers.** Computed glycosidic linkage and pyranose ring free energy surfaces were used to parametrize a coarse-grained (CG) model, described previously.<sup>22</sup> For validation purposes, the decasaccharides studied above were resimulated for 10  $\mu$ s using the CG model. Comparing CG with reference fine-grained simulations, equivalent calculated mean *r* values (see above) were all within 0.3 nm and linkage and ring motions were faithfully reproduced (Figure 5).

Radius of gyration ( $R_g$ ) and molecular mass ( $M$ ) were measured simultaneously, using multiangle laser light scattering (MALLS), on a sample of medical grade HA. Model polymers in the range [H]<sub>100</sub> to [H]<sub>400</sub> (reflecting the range of measured molecular weights) were simulated for 10  $\mu$ s using the CG method. Experimental data and simulated mean  $R_g$  are graphed in Figure 6A (see “CG all”). Within these simulations, unrealistic atomic overlap occurred due to the absence of steric interactions in the CG model.<sup>22</sup> Correcting for this by removing those computed structures containing steric clashes and recalculating  $R_g$  achieved excellent agreement with experiment (see Figure 6A “CG no clash”). Finally, recalculation of  $R_g$  (after removing steric clashes) from identical CG simulations but with all constituent monosaccharides fixed in rigid <sup>4</sup>C<sub>1</sub> chair puckerers resulted in predictions that disagreed with experimental measurements (see Figure 6A “CG no clash, rigid rings”).

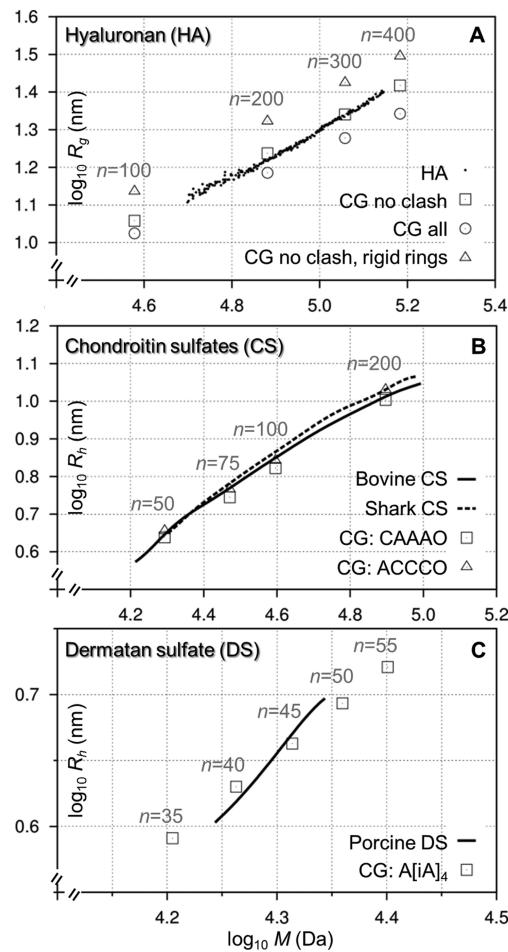
In the cases of CS and DS materials from shark, bovine, and porcine sources, the small polymer sizes ( $\approx$ 10 nm) precluded reliable  $R_g$  measurements using MALLS. Therefore, hydrodynamic radii ( $R_h$ ) at defined  $M$  were derived from quasi-elastic light scattering and MALLS (Figure 6B,C and Figure S5). Model CG polymers of these materials were built from blocks of five disaccharides (with compositions reflecting NMR



**Figure 5.** Homogeneous decasaccharide dynamics from fine-grained simulations (left, black) were faithfully reproduced by the coarse-grained model (right, gray) in equivalent 10  $\mu$ s simulations (A–D) and (E–H). The dynamic range of the model A-unit  $\beta(1 \rightarrow 3)$  and  $\beta(1 \rightarrow 4)$  glycosidic linkages (numbers 3 and 4 counting from the reducing end) and (E–H) iA-unit GalNAc and IdoA pyranose rings (numbers 5 and 6 counting from the reducing end) are shown, using glycosidic ( $\phi, \psi$ ) and ring puckering ( $\theta, \phi$ ) parameters, respectively.

analyses reported above) and simulations were performed within the experimentally observed mass ranges (again following removal of steric clashes). Simulated  $R_g$  values for representative shark and bovine CS polymers (in the range 50–200 disaccharides) concurred with the measurements after conversion to equivalent  $R_h$  (Figure 6B). Here, the experimental trend for shark CS to have slightly larger  $R_h$  than bovine CS at equivalent mass was reproduced in the CG simulations. For model porcine DS polymers of 35–55 disaccharides in length, simulated  $R_g$  values (converted to  $R_h$ ) were also close to measured equivalents (Figure 6C). Marginal deviations in experimental and calculated data for sulfonated polymers may be attributed to sequence differences in the biological materials and models, the assumed refractive index increment value (determined from bovine nasal cartilage PG fractions<sup>28</sup>) or the conformationally dependent  $R_g$  to  $R_h$  conversion factor.

This validation demonstrates the applicability of the CG approach to all human GAGs except keratan sulfate (the model's predictive ability for heparan sulfate was shown previously<sup>22</sup>). It also reiterates that inclusion of both glycosidic linkage and pyranose ring  $\mu$ s time scale motions are necessary to produce accurate GAG polymer models. Furthermore, the CG method enables rapid and accurate predictions using



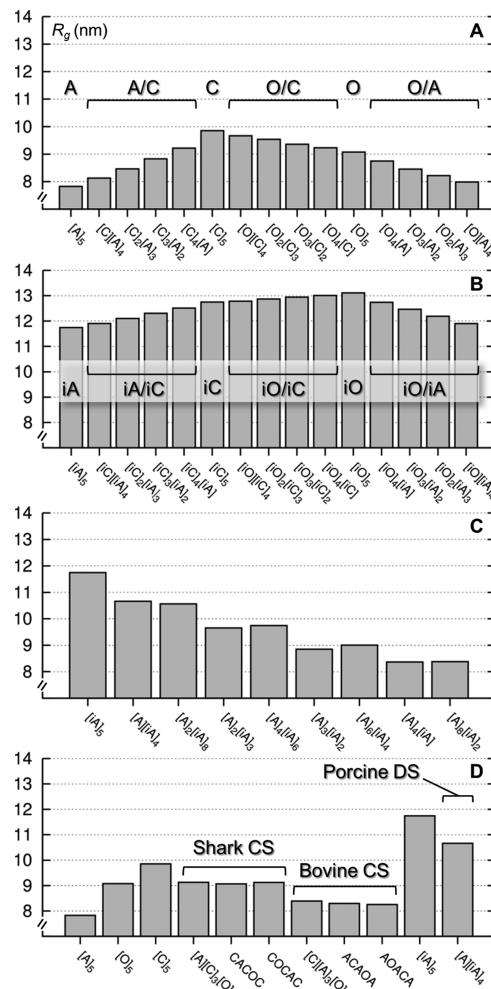
**Figure 6.** Accurate, rapid prediction of hydrodynamic size for polymers of (A) hyaluronan, (B) chondroitin sulfates, and (C) dermatan sulfate from biological sources. Predictions from  $10\ \mu\text{s}$  coarse-grained (CG) polymer simulations (open symbols) and experimental radius of gyration ( $R_g$ ) or hydrodynamic size ( $R_h$ ) are plotted against molecular mass ( $M$ ) on logarithmic axes. Note:  $n$  is the number of disaccharides in the model polymers and analyzed experimental data points are shown as dots in (A), while in (B, C) the experimental data are depicted as lines.

standard computer hardware; for example, these  $10\ \mu\text{s}$  simulations for polymers of 100 disaccharides in length were each performed in less than 1 h.

#### Effects of Combining and Permuting GAG Chain Units.

The CG method is particularly suited to studying the relationship between chain dynamics and GAG sequence. To effectuate a study of this kind, polymer chains with 100 disaccharide units were constructed by repeating blocks of either 5 or 10 units (idiosyncratic patterns were chosen to explore the effect of both hexosamine sulfonation and uronic acid type). Performing 10 simulations of  $[A]_{100}$  resulted in a calculated  $R_g$  standard deviation of 0.03 nm, confirming the utility of mean  $R_g$  as a sensitive indicator of chain conformation. Chains were simulated for  $10\ \mu\text{s}$  using the CG approach; steric clashes occurred infrequently at this degree of polymerization and removal was not necessary for accurate predictions. Computed  $R_g$  values for each polymer are shown in Figure 7.

Of any CS sequence considered, chains with  $[A]_{100}$  and  $[C]_{100}$  resulted in the smallest and largest  $R_g$  values, respectively (Figure 7A). Successive substitution of A- with C-units (or vice versa) resulted in a systematic increase (or decrease) in

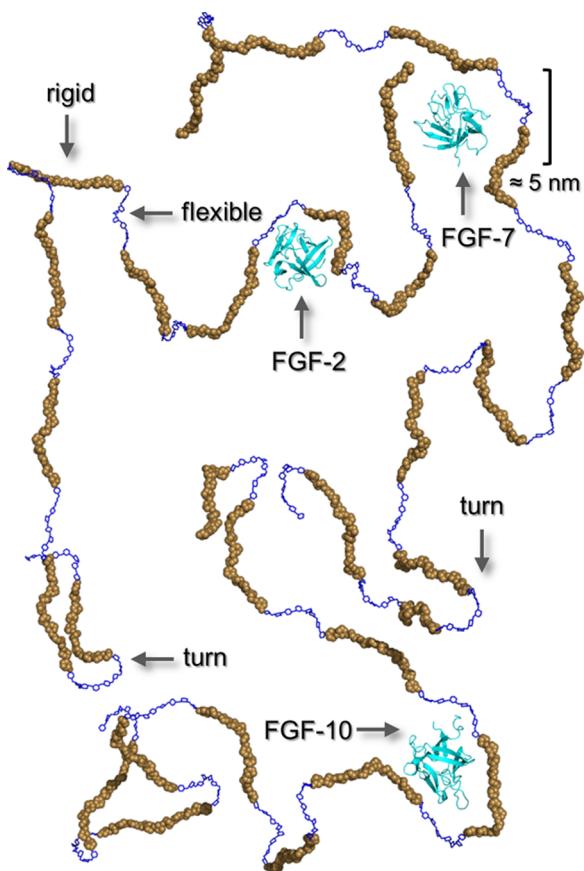


**Figure 7.** Prediction of the effect of sequence and domain organization on radius of gyration ( $R_g$ ) in model homogeneous and heterogeneous polymers of 100 disaccharide units in length: (A) chondroitin sulfate polymers, (B) dermatan sulfate polymers, (C) hybrid chondroitin/dermatan sulfate polymers, and (D) models of biological materials studied experimentally here. All  $R_g$  values were derived from  $10\ \mu\text{s}$  coarse-grained simulations.

polymer size. Chains with only O-units were significantly longer than those with only A-units, but shorter than polymers comprising only C-units. Thus, inclusion of O-units increased the average hydrodynamic size of A-unit polymers and decreased that of C-unit polymers. In model DS chains, equivalent sulfonation patterns (to CS) resulted in consistently larger  $R_g$  values (Figure 7B), a trend that reflected observations from fine-grained simulations (Figure 2). Here, the 4-O-sulfo  $[iA]_{100}$  chain had a smaller calculated  $R_g$  than either the  $[iO]_{100}$  or  $[iC]_{100}$  polymers. In contrast to CS, successive inclusion of iO- into iC-unit polymers marginally increased calculated radii. Polymers comprising exclusively IdoA-units had a relatively restricted range of  $R_g$  values; all were within 1.4 nm of each other (cf. 2.0 nm in GlcA-unit polymers). Progressive addition of A-units into an iA-based polymer decreased the overall size, and maintaining the A/iA stoichiometry while increasing the iA-block size had a negligible effect on calculated hydrodynamic radii (Figure 7C). Therefore, it was hypothesized that permutation of units (while maintaining the combination) has little effect on  $R_g$  and the combinations of uronic acid type and hexosamine sulfonation contribute to the final size of  $R_g$  in

the following order (smallest to largest): A < O < C < iA < iO < iC. This suggests that combining these units in heterogeneous polymers provides a mechanism for tuning the hydrodynamic properties of GAG chains.

**Block Copolymer Microarchitectures as a Basis for Multivalency.** Supporting these hypotheses, permutations of a sequence combination mimicking shark CS were predicted to have identical  $R_g$  values, and the same was true for bovine CS (Figure 7D). Shark CS polymers were predicted to be slightly larger than their bovine equivalents, but the porcine DS models were, as expected, the most extended model biological sequences considered. As a consequence of this it would be expected that PGs comprising DS would have a larger hydrodynamic domain than their CS equivalents. However, in DS block copolymers, overall polymer  $R_g$  did not provide the full picture. For example, inspection of conformational snapshots from a simulation of the  $\{[A]_4[iA]_6\}_{10}$  polymer revealed flexibility in the GlcA-containing regions, resulting in tight turns reminiscent of protein  $\beta$ -hairpins, and comparative rigidity in the IdoA-containing domains (Figure 8). These predicted microarchitectures could be a molecular basis for multivalent interactions with proteins, such as fibroblast growth factors.<sup>7,53</sup> The differential dynamics of these copolymeric



**Figure 8.** Three representative conformers of a heterogeneous dermatan sulfate polymer with the sequence  $\{[A]_4[iA]_6\}_{10}$  (a 2D projection of 3D data). Constituent A- and iA-units are blue and sand colored, respectively. Emergent rigid, flexible, and turn conformational features in the  $10 \mu s$  coarse-grained simulation are labeled. These generate microarchitectures hypothesized to enable multivalent interactions with fibroblast growth factors; structures of FGF-2, FGF-7, and FGF-10 (Protein Data Bank entries: 1FQ9, 1QQK, and 1NUN) are shown to scale and colored cyan.

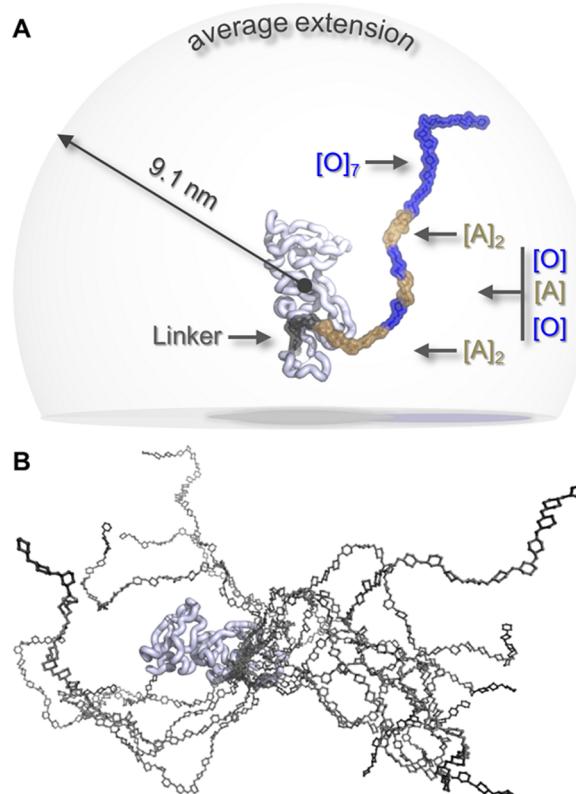
sequences can thus be expected to have significance in signal transduction.

Given the documented ability of IdoA to be an equilibrium of chair conformations, prior work had logically assumed IdoA to impart GAG chain flexibility and GlcA to confer rigidity.<sup>24</sup> Our current studies, based on  $10 \mu s$  fine- and coarse-grained simulations, have come to an alternative set of conclusions: (a) pyranose rings of both IdoA and GlcA are flexible, (b) their conformational equilibria are driven by neighboring hexosamine 4-O-sulfonation toward either  $^1C_4$  or  $^4C_1$ , (c) equivalent chains with IdoA are more extended than those with GlcA (irrespective of sulfonation), and (d) IdoA does not substantially populate nonchair boat or skew-boat pucksers.

#### Accurate Proteoglycan Models: Bikunin and Decorin.

In a  $10 \mu s$  fine-grained simulation of the GAG-PG interface trisaccharide, Gal $\beta(1 \rightarrow 3)$ Gal- $\beta(1 \rightarrow 4)$ Xyl, both galactose (Gal) pyranose rings were predicted to be stable  $^4C_1$  pucksers (>99%), while xylose (Xyl) was an equilibrium comprising mostly  $^4C_1$  chairs (70%) with a minor  $^1C_4$  component (26%). The Gal-Gal and Gal-Xyl glycosidic linkages were centered on  $(\phi,\psi) \approx (-60^\circ, -120^\circ)$  and  $\approx (-60^\circ, +120^\circ)$ , respectively, which are consistent with prior NMR<sup>37</sup> and a  $1 \mu s$  fine-grained simulation of bikunin glycosylated with the GAG-PG trisaccharide. Comparing linkage and ring dynamics in these two simulations (Figure S6) suggested that motions in this region do not differ substantially when free in water and when linked to the core protein. Models of PGs were constructed by parametrizing a fully dynamic CG model of the linker from these fine grained simulations and elaborating with specific GAG chain sequences; these polymers were then attached to previously identified PG core protein serine (Ser) residues. The PG core proteins were kept rigid while the dynamics of GAG chains were predicted in  $10 \mu s$  CG simulations. Conformers that contained steric clashes were removed from resulting ensembles prior to calculation of  $R_g$  (as above).

A human bikunin PG model was built from an X-ray refinement of its core protein, the linker (attached to Ser-10), and a single CS chain. Several human bikunin GAG glycoforms have been sequenced, with 27–43 monosaccharides.<sup>5</sup> Initially, an intermediate length chain (33 monosaccharides) was modeled: GalNAc-[O]<sub>7</sub>[A]<sub>2</sub>[O][A][A][O][A]<sub>2</sub>-GlcA-. In this simulation, the GAG folded and transiently enveloped the core protein. The distance between the core protein and terminal monosaccharide geometric centers was on average 9.1 nm and thus significantly larger than the protein core (Figure 9A). This distance could range between 2.0 and 14.8 nm, highlighting the extent of sampled GAG conformations. Representative 3D structures from this human bikunin simulation are illustrated in Figure 9B and an  $R_g$  of 3.4 nm was predicted (repeated calculations of  $R_g$  had a standard deviation below the number of significant figures reported, see Table S7). Furthermore, performing this simulation with an alternative extended N-terminal loop conformer, derived from a  $1 \mu s$  fine-grained simulation of the bikunin core protein (see Figure S8), resulted in an indistinguishable  $R_g$  value, suggesting that the N-terminal loop conformation has very little effect on intact PG hydrodynamic properties. Prefixing the sequence described above (at its nonreducing end) with [O]<sub>5</sub> resulted in the longest CS chain identified previously.<sup>5</sup> In an ancillary simulation of this glycoform, the computed  $R_g$  increased by almost a quarter, to a value of 4.1 nm. For comparison, an experimental  $R_h$  value for rat bikunin (thought to have similar hydrodynamic properties to human bikunin) was measured

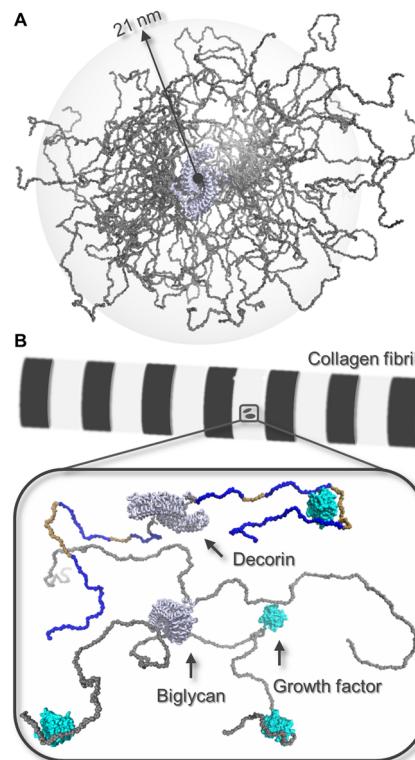


**Figure 9.** (A) Average conformer of human bikunin derived from a 10  $\mu$ s coarse-grained simulation. The model comprises a heterogeneous chondroitin sulfate sequence of intermediate length (see text) and constituent O- and A-units are blue and sand colored, respectively. The mean chain extension from the core protein geometric center is depicted as a sphere (see Supporting Information for the model coordinates). (B) The 15 representative conformations from the same simulation illustrate the dynamic range of chondroitin sulfate conformational sampling within human bikunin.

previously at  $3.5 \pm 0.1$  nm.<sup>54</sup> This is within the predicted range of  $R_g$  values for the two human bikunin glycoforms considered here.

Crystallographic data for the bovine skin decorin core protein (for modeling purposes assumed here to be a homomeric dimer) was used as the basis for a second PG model. In the absence of sequence data for bovine decorin GAG chains, they were initially based on recent sequence analysis of porcine decorin GAGs<sup>55</sup> and the average molecular weight of bovine decorin GAG chains ( $\approx 24$  kDa).<sup>56</sup> A linker and a single DS chain with the sequence  $\{[A][iA]_4\}_{10}$ -GlcA- was appended to each subunit (at Ser-5). The calculated  $R_g$  for this dynamic PG model was 8.2 nm. Bovine decorin can be assumed to exhibit GAG glycoform variability, and thus, two further simulations were performed with variations in the IdoA content and chain length. Increasing the IdoA block length in both GAG chains with the sequence  $\{[A][iA]_5\}_9$ -GlcA- ( $\approx 25$  kDa) resulted in an  $R_g$  of 8.7 nm. A further PG simulation where both chains were extended with the sequence  $\{[A][iA]_4\}_{12}$ -GlcA- ( $\approx 27$  kDa) resulted in a concomitant increase in  $R_g$  to 9.4 nm. These predictions are consistent with an experimentally measured  $R_g$  of  $9.3 \pm 0.1$  for intact bovine decorin,<sup>56</sup> underscoring the CG method's ability to reproduce hydrodynamic data for intact PGs. The predictions also provide evidence for a dependence of PG size on the length and sequence combinations of the attendant GAGs.

In the absence of an X-ray diffraction refinement for human decorin core protein, the homologous bovine core protein was again used as a start point for building a human decorin PG model. Recent sequencing work has resulted in a human decorin GAG glycoform map, comprising multiple domains of variable length, sulfonation, and IdoA content.<sup>6</sup> Based on this, a presumed consensus sequence was constructed:  $[C]_6[A]_8[O]_6$ - $[iA]_4[iC]_3[iO]_3$ - $[A]_3[C]_4[O]_3$ - $[iC][iA][iO]$ - $[O]_3[iO][O]_4$ -GlcA-. Two of these chains, one per subunit, were appended to the bovine decorin PG model at Ser-5 (a homomeric dimer was assumed). Representative conformations from the subsequent CG simulation are illustrated in Figure 10A. The predicted



**Figure 10.** (A) Representative conformers derived from a 10  $\mu$ s coarse-grained simulation of decorin with a single consensus human dermatan sulfate polymer attached to each core protein subunit (see text for the sequence). The 30 conformations reveal the dynamic range of polymer motion and their significant contribution to PG size; mean chain extension from the protein geometric center is depicted as a sphere of radius 21 nm. (B) Conformers of decorin and biglycan (two chains per subunit) with consensus human dermatan sulfate polymers (see text for sequences and Supporting Information for model coordinates). In decorin, GlcA and IdoA-containing blocks are blue and sand colored, respectively. A schematic collagen fibril with D-spacing of  $\approx 70$  nm and growth factor proteins (Protein Data Bank entries: 1FQ9, 1QQK, and 1NUN) are drawn to scale, these (colored cyan) are positioned to illustrate the potential for multivalent binding.

overall  $R_g$  value was 8.0 nm, and thus, compared to a bovine skin decorin model, the hydrodynamic size of human skin decorin is predicted to be smaller, in part due to reduced IdoA composition. The dynamics of the two GAG chains featured bending and wrapping motions, with the average distance between the  $[iA]_4$  regions of each chain predicted to be 23 nm. As alluded to above, multivalent interactions are theorized to be possible between a protein and a single GAG chain (see Figure 8). In this PG model,  $[iA]_4$  regions of separate chains

approached with a proximity of less than 5 nm, the approximate size of an FGF protein, confirming the feasibility of increased multivalent binding in a PG system with two GAG chains (Figure 10B).

**Substantiating Functional Roles for GAG Chain Length and Density: Biglycan.** Two human biglycan models were constructed using crystallographic data for the bovine core protein (again assumed to be a homomeric dimer). In each, both subunits were appended (via linkers at Ser-4 and Ser-11) with two GAG chains containing an average number of units found in human mitral valve tissue associated with either tensile ( $\{[A][iA]_4\}_{12}\text{-GlcA}\}$ ) or compressive ( $\{[A][C]_2[iA]_2\}_{17}\text{-GlcA}\}$ ) regions.<sup>57</sup> In these dynamic models, the respective predicted  $R_g$  values were 13.0 and 15.4 nm. The larger hydrodynamic radii of PGs in compressive tissue, as reproduced here, are thought to be important in resisting stress.<sup>57</sup> Based on the biglycan simulations, it is simply concluded that the GAG chains with longer sequences found in compressive tissue are primarily responsible for the enhanced hydrodynamic radii. While the exact sequences have some effect, this is predicted to be a secondary factor. Indeed, the simulations imply that at equivalent GAG length the tension associated sequence, with a higher IdoA content, would be larger in size than the sequence found in compressive mitral valve tissue. If so, the specific sequences of IdoA and sulfonation are likely to be responsible for other functions such as protein recruitment. In the tensile tissue biglycan model, the average approach distances of the  $[iA]_4$  blocks at the chain centers was 16 nm, significantly closer than in the human decorin model (23 nm), implying that the synergistic effect of four GAG chains increased the possibility of multivalent interactions with other proteins (Figure 10B), as recently evidenced.<sup>8</sup> These simulations also highlight the vast contribution of GAG chains to PG size. The computed  $R_g$  for the homodimeric biglycan core protein alone was around 3 nm, much less than the 13–15 nm found here for fully glycoconjugated biglycan dimer models (with a total mass of  $\approx 200$  kDa). Moreover, based on standard bovine serum albumin, a folded globular protein would require a mass of approximately 3 MDa to achieve this physical size.

## CONCLUSIONS

It has been shown that accurate spatiotemporal PG models, including complex heterogeneous GAG sequences, can be constructed by abstracting glycosidic linkage and pyranose ring atomic motions from multi- $\mu$ s fine-grained simulations. The current CG method implementation was successful for the cases of small PGs and isolated GAGs, resulting in predictions that supersede schematic “bottle brush” PG representations. These atomic scale dynamic perspectives are essential to uncover how PGs collectively perform constitutive signal transduction and mechanical functions, and extension to GAG sequences not considered here should be straightforward.

The described CG technique can address PG assemblies comprising millions of atoms, which remain beyond the capability of even hardware accelerated fine-grained modeling. Therefore, emerging GAG sequence and core protein 3D data, in concert with methodological developments enabling molecular interactions, will allow larger cell-surface and aggregating PGs, plus assemblies thereof, to be studied routinely at the atomic scale. In time, the approach could inform how GAG synthetic enzyme deficiencies lead to connective tissue diseases,<sup>58</sup> provide mode-of-action insights

for biopharmaceutical PGs, and facilitate the development of new biologically inspired medicines and materials.

## ASSOCIATED CONTENT

### S Supporting Information

(S1) 800 MHz NMR spectra for glycosaminoglycan polymers from biological sources, (S2) decasaccharide linkage dynamics in fine- and coarse-grained 10  $\mu$ s simulations, (S3) decasaccharide ring dynamics in fine- and coarse-grained 10  $\mu$ s simulations, (S4) pyranose ring  $^1\text{H}$ – $^1\text{H}$  three-bond vicinal spin-couplings, (S5) light scattering data for glycosaminoglycan polymers from biological sources, (S6) protein-GAG linker dynamics during free and protein-attached  $\mu$ s fine-grained simulations, (S7) statistical analyses of radii of gyration from coarse-grained 10  $\mu$ s simulations, (S8) representative bikunin conformers from a 1  $\mu$ s fine-grained simulation, (S9) ring puckering definitions, and (S10) decasaccharide  $r$  values from fine- and coarse-grained 10  $\mu$ s simulations. Three-dimensional models from 10  $\mu$ s coarse-grained simulations of bikunin, decorin, and biglycan (as shown in Figures 9A and 10B) are also 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.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

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

## REFERENCES

- (1) Ariga, T.; Miyatake, T.; Yu, R. K. *J. Neurosci. Res.* **2010**, *88*, 2303–15.
- (2) Li, L.; Ly, M.; Linhardt, R. *J. Mol. Biosyst.* **2012**, *8*, 1613–25.
- (3) Kim, S. H.; Turnbull, J.; Guimond, S. *J. Endocrinol.* **2011**, *209*, 139–51.
- (4) Sugahara, K.; Mikami, T.; Uyama, T.; Mizuguchi, S.; Nomura, K.; Kitagawa, H. *Curr. Opin. Struct. Biol.* **2003**, *13*, 612–20.
- (5) Ly, M.; Leach, F. E.; Laremore, T. N.; Toida, T.; Amster, I. J.; Linhardt, R. *J. Nat. Chem. Biol.* **2011**, *7*, 827–833.
- (6) Laremore, T. N.; Ly, M.; Zhang, Z.; Solakyildirim, K.; McCallum, S. A.; Owens, R. T.; Linhardt, R. *J. Biochem. J.* **2010**, *431*, 199–205.
- (7) Radek, K. A.; Taylor, K. R.; Gallo, R. L. *Wound Repair Regen.* **2009**, *17*, 118–26.
- (8) Nguyen, T. K.; Tran, V. M.; Sorna, V.; Eriksson, I.; Kojima, A.; Koketsu, M.; Loganathan, D.; Kjellen, L.; Dorsky, R. I.; Chien, C. B.; Kubera, B. *ACS Chem. Biol.* **2013**, *8*, 939–48.
- (9) Fransson, L. A.; Coster, L.; Havasmark, B.; Malmstrom, A.; Sjoberg, I. *Biochem. J.* **1974**, *143*, 379–89.
- (10) Seidler, D. G.; Faiyaz-Ul-Haque, M.; Hansen, U.; Yip, G. W.; Zaidi, S. H.; Teebi, A. S.; Kiesel, L.; Gotte, M. *J. Mol. Med.* **2006**, *84*, 583–94.
- (11) Buckwalter, J. A.; Rosenberg, L. C. *J. Biol. Chem.* **1982**, *257*, 9830–9.
- (12) Watanabe, Y.; Inoko, Y. *J. Chromatogr. A* **2013**, *1303*, 100–4.
- (13) Xu, Y.; Carr, P. D.; Guss, J. M.; Ollis, D. L. *J. Mol. Biol.* **1998**, *276*, 955–66.
- (14) Scott, P. G.; Dodd, C. M.; Bergmann, E. M.; Sheehan, J. K.; Bishop, P. N. *J. Biol. Chem.* **2006**, *281*, 13324–32.
- (15) Scott, P. G.; McEwan, P. A.; Dodd, C. M.; Bergmann, E. M.; Bishop, P. N.; Bella, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15633–8.

- (16) Cael, J. J.; Winter, W. T.; Arnott, S. *J. Mol. Biol.* **1978**, *125*, 21–42.
- (17) Schlick, T. *F1000 Biol. Rep.* **2009**, *1*, 51.
- (18) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341–6.
- (19) Sattelle, B. M.; Almond, A. *Carbohydr. Res.* **2014**, *383*, 34–42.
- (20) Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.
- (21) Ingolfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. *WIREs Comput. Mol. Sci.* **2014**, *4*, 225–248.
- (22) Sattelle, B. M.; Shakeri, J.; Almond, A. *Biomacromolecules* **2013**, *14*, 1149–59.
- (23) Bossennec, V.; Petitou, M.; Perly, B. *Biochem. J.* **1990**, *267*, 625–30.
- (24) Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Casu, B.; Torri, G.; Bossennec, V.; Perly, B.; Sinay, P.; Petitou, M.; Choay, J. *Carbohydr. Res.* **1990**, *195*, 157–67.
- (25) Silipo, A.; Zhang, Z.; Canada, F. J.; Molinaro, A.; Linhardt, R. J.; Jimenez-Barbero, J. *ChemBioChem* **2008**, *9*, 240–52.
- (26) Yamada, S.; Yamane, Y.; Sakamoto, K.; Tsuda, H.; Sugahara, K. *Eur. J. Biochem.* **1998**, *258*, 775–83.
- (27) Yang, H. O.; Gunay, N. S.; Toida, T.; Kuberan, B.; Yu, G. L.; Kim, Y. S.; Linhardt, R. J. *Glycobiology* **2000**, *10*, 1033–1040.
- (28) Kitchen, R. G.; Cleland, R. L. *Biopolymers* **1978**, *17*, 759–83.
- (29) Cohen, D. E.; Thurston, G. M.; Chamberlin, R. A.; Benedek, G. B.; Carey, M. C. *Biochemistry* **1998**, *37*, 14798–814.
- (30) Kirschner, K. N.; Yonge, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–55.
- (31) 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*, 1668–88.
- (32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (33) Harvey, M. J.; Giupponi, G.; De Fabritiis, G. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.
- (34) Sattelle, B. M.; Almond, A. *Phys. Chem. Chem. Phys.* **2014**, *16*, 8119–26.
- (35) Almond, A.; Brass, A.; Sheehan, J. K. *J. Mol. Biol.* **1998**, *284*, 1425–37.
- (36) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (37) Agrawal, P. K.; Jacquinet, J. C.; Krishna, N. R. *Glycobiology* **1999**, *9*, 669–77.
- (38) Malmstrom, A.; Bartolini, B.; Thelin, M. A.; Pacheco, B.; MacCarana, M. *J. Histochem. Cytochem.* **2012**, *60*, 916–25.
- (39) Volpi, N. *J. Pharm. Pharmacol.* **2009**, *61*, 1271–80.
- (40) Blundell, C. D.; DeAngelis, P. L.; Day, A. J.; Almond, A. *Glycobiology* **2004**, *14*, 999–1009.
- (41) Winter, W. T.; Smith, P. J.; Arnott, S. *J. Mol. Biol.* **1975**, *99*, 219–35.
- (42) Millane, R. P.; Mitra, A. K.; Arnott, S. *J. Mol. Biol.* **1983**, *169*, 903–20.
- (43) Arnott, S.; Guss, J. M.; Hukins, D. W.; Matthews, M. B. *Biochem. Biophys. Res. Commun.* **1973**, *54*, 1377–83.
- (44) Mitra, A. K.; Arnott, S.; Atkins, E. D.; Isaac, D. H. *J. Mol. Biol.* **1983**, *169*, 873–901.
- (45) Almond, A.; Deangelis, P. L.; Blundell, C. D. *J. Mol. Biol.* **2006**, *358*, 1256–69.
- (46) Sattelle, B. M.; Shakeri, J.; Roberts, I. S.; Almond, A. *Carbohydr. Res.* **2010**, *345*, 291–302.
- (47) Sattelle, B. M.; Hansen, S. U.; Gardiner, J.; Almond, A. *J. Am. Chem. Soc.* **2010**, *132*, 13132–4.
- (48) Blundell, C. D.; Roberts, I. S.; Sheehan, J. K.; Almond, A. *J. Mol. Microbiol. Biotechnol.* **2009**, *17*, 71–82.
- (49) Gatti, G.; Casu, B.; Torri, G.; Vercellotti, J. R. *Carbohydr. Res.* **1979**, *68*, C3–C7.
- (50) Pellegrini, L.; Burke, D. F.; von Delft, F.; Mulloy, B.; Blundell, T. L. *Nature* **2000**, *407*, 1029–34.
- (51) Hricovini, M.; Guerrini, M.; Bisio, A.; Torri, G.; Petitou, M.; Casu, B. *Biochem. J.* **2001**, *359*, 265–272.
- (52) Ochsenbein, P.; Bonin, M.; Schenk-Joss, K.; El-Hajji, M. *Angew. Chem., Int. Ed.* **2011**, *50*, 11637–9.
- (53) Taylor, K. R.; Rudisill, J. A.; Gallo, R. L. *J. Biol. Chem.* **2005**, *280*, S300–6.
- (54) Lindstrom, K. E.; Blom, A.; Johnsson, E.; Haraldsson, B.; Fries, E. *Kidney Int.* **1997**, *51*, 1053–8.
- (55) Zhao, X.; Yang, B.; Solakyildirim, K.; Joo, E. J.; Toida, T.; Higashi, K.; Linhardt, R. J.; Li, L. *J. Biol. Chem.* **2013**, *288*, 9226–37.
- (56) Scott, P. G.; Grossmann, J. G.; Dodd, C. M.; Sheehan, J. K.; Bishop, P. N. *J. Biol. Chem.* **2003**, *278*, 18353–18359.
- (57) Grande-Allen, K. J.; Calabro, A.; Gupta, V.; Wight, T. N.; Hascall, V. C.; Vesely, I. *Glycobiology* **2004**, *14*, 621–33.
- (58) Miyake, N.; Kosho, T.; Mizumoto, S.; Furuchi, T.; Hatamochi, A.; Nagashima, Y.; Arai, E.; Takahashi, K.; Kawamura, R.; Wakui, K.; Takahashi, J.; Kato, H.; Yasui, H.; Ishida, T.; Ohashi, H.; Nishimura, G.; Shiina, M.; Saitsu, H.; Tsurusaki, Y.; Doi, H.; Fukushima, Y.; Ikegawa, S.; Yamada, S.; Sugahara, K.; Matsumoto, N. *Hum. Mutat.* **2010**, *31*, 966–74.