Computational and Structural Biotechnology Journal 20 (2022) 3884–3898

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

3-*O*-Sulfation induces sequence-specific compact topologies in heparan sulfate that encode a dynamic sulfation code

Samuel G. Holmes [a,b,1,2], Balaji Nagarajan [a,b,1,3], Umesh R. Desai [a,b,*,4]

[a] Department of Medicinal Chemistry, School of Pharmacy, Virginia Commonwealth University, Richmond, VA 23298, USA
[b] Institute for Structural Biology, Drug Discovery and Development, Virginia Commonwealth University, Richmond, VA 23219, USA

ARTICLE INFO

Article history:
Received 17 May 2022
Received in revised form 5 July 2022
Accepted 5 July 2022
Available online 18 July 2022

Keywords:
3-*O*-Sulfation
Heparan sulfate
Molecular dynamics
Conformational sampling
Conformational changes
3OSTs

ABSTRACT

Heparan sulfate (HS) is arguably the most diverse linear biopolymer that is known to modulate hundreds of proteins. Whereas the configurational and conformational diversity of HS is well established in terms of varying sulfation patterns and iduronic acid (IdoA) puckers, a linear helical topology resembling a cylindrical rod is the only topology thought to be occupied by the biopolymer. We reasoned that 3-*O*-sulfation, a rare modification in natural HS, may induce novel topologies that contribute to selective recognition of proteins. In this work, we studied a library of 24 distinct HS hexasaccharides using molecular dynamics (MD). We discovered novel compact (C) topologies that are populated significantly by a unique group of 3-*O*-sulfated sequences containing IdoA residues. 3-*O*-sulfated sequences containing glucuronic acid (GlcA) residue and sequences devoid of 3-*O*-sulfate groups did not exhibit high levels of the C topology and primarily exhibited only the canonical linear (L) form. The C topology arises under dynamical conditions due to rotation around an IdoA → GlcN glycosidic linkage, especially in psi (Ψ) torsion. At an atomistic level, the L → C transformation is a multi-factorial phenomenon engineered to reduce like-charge repulsion, release one or more HS-bound water molecules, and organize a bi-dentate "IdoA-cation-IdoA" interaction. These forces also drive an L → C transformation in a 3-*O*-sulfated octasaccharide, which has shown evidence of the unique C topology in the co-crystallized state. The 3-*O*-sulfate-based generation of unique, sequence-specific, compact topologies indicate that natural HS encodes a dynamic sulfation code that could be exploited for selective recognition of target proteins.

1. Introduction

Heparin/heparan sulfate (Hp/HS), two members of the superfamily of glycosaminoglycans (GAGs), play important roles in numerous molecular and cellular processes. Classically, Hp/HS have been studied extensively as modulators of coagulation; however, growing evidence indicates that these biopolymers may possess major clinical value as anti-infectives and anti-cancer agents [1–4]. Hp/HS have also been implicated as major regulators in diverse processes such as axonal growth [5], neurofibrillary tangles [6], and native immune response [7,8].

The diverse biological activities of Hp/HS are mediated by different proteins including those that are secreted or membrane-bound, either within or outside cells. The recent release of the first draft of GAG interactome suggests more than 500 protein partners of these highly anionic polymers [9]. Among these, a number of proteins preferentially bind to 3-*O*-sulfate (3OS)-containing sequence(s) in Hp/HS including antithrombin [10], fibroblast growth factor receptor [11], glycoprotein D [12], cyclophilin B [13], neuropilin-1 [5], heparin cofactor II [14], tau glycoprotein [6] and spike glycoprotein [15]. Each of these sequences are distinct owing to the structural pattern flanking the 3OS group, which suggests the possibility of a "sulfation code". It is expected that more proteins will be identified in the future as preferred targets with their own 3OS-based sulfation code.

The 3OS group is introduced in the final step of Hp/HS biosynthesis and is generally rare in most biological tissues [16]. For

Abbreviations: GAGs, Glycosaminoglycans; Hp/HS, Heparin/Heparan Sulfate; 3OS, 3-O-sulfate; MD, Molecular Dynamics; EED, End-to-End Distance.

* Corresponding author at: Institute for Structural Biology, Drug Discovery, and Development, 800 E. Leigh Street, Suite 212, Richmond, VA 23219, USA.
E-mail address: urdesai@vcu.edu (U.R. Desai).
[1] Equal contributing authors.
[2] ORCID: 0000-0001-8589-1241.
[3] ORCID: 0000-0003-3892-6178.
[4] ORCID: 0000-0002-1976-6597.

https://doi.org/10.1016/j.csbj.2022.07.013

example, clinically used Hp has only one in three chains carrying the antithrombin-binding 3OS sequence, which approximates to one 3OS-containing residue in nearly 150 residues. On the other hand, human follicular fluid contains ~6 % GlcNS3S residues [17]. Interestingly, nature has designed seven isoforms of 3-*O*-sulfotransferase (HS3ST-1, -2, -3_A, -3_B, 4, -5, and -6) to introduce just one 3OS group in an HS chain. In comparison, only one and three isoforms are known to date for other sulfotransferases (STs), which introduce *O*-sulfates at the 2- and 6-positions, respectively. Further, a high level of regulation is apparent in the biosynthesis of Hp/HS despite being a fully template-independent process. For example, the expressions of different STs are tissue- and time- dependent [18–21]. Such a regulation appears to help generate a fairly consistent phenotype, as noted in the expression of STs in a specific tissue.

The seven isoforms of HS3ST display significant differences in substrate specificities. Whereas HS3ST-1 preferentially generates the antithrombin-binding HS sequence, HS3ST-3_B generates binding motifs for herpes simplex virus-1 glycoprotein D [16]. The specificities of other isoforms have been presented in terms of generation of either antithrombin- or glycoprotein D- binding motifs. Yet, obvious differences in specificities among the seven isoforms have also been clearly noted [22–25]. Alternatively, the HS3ST family of enzymes appears to encompass some overlap of specificity, while also exhibiting important differences.

The rarity of 3OS groups in nature coupled with an abundance of HS3ST isoforms that present an overlapping, and yet varying substrate specificity, make this functional group an enigmatic structural modification. A burning question is why has nature engineered 3-*O*-sulfation? Growing evidence points to its role in generating unique "sulfation codes" that are recognized by different proteins in a sequence-specific manner. We reasoned that 3-*O*-sulfation may induce unique conformational properties that endow an Hp/HS sequence with novel protein recognition properties. More specifically, we reasoned whether incorporation of a 3OS group in some Hp/HS sequences, but not all, generate novel conformational states?

In this work, we present molecular dynamics-based evidence that the presence of 3OS group induces a unique compact ('C') form distinct from the well-established linear, helical form known for decades. We studied a library of 24 Hp/HS hexasaccharides with varying distribution of sulfate groups to comprehensively understand their conformational states. We find that when a 3-*O*-sulfated GlcN residue is flanked by two iduronic acid (IdoA) residues, a unique 'C' shaped topology is significantly populated. The C topology arises from inversion of a preferred glycosidic torsion, which facilitates reduction in electrostatic crowding, release of one or more bound waters, and bi-dentate sequestration of a cation by iduronic acids. MD studies of a 3OS-containing octasaccharide also generate a compact form, similar to that observed for hexasaccharides. Interestingly, literature-reported crystal structure of the 3-*O*-sulfated octasaccharide shows the occupation of this compact topology in the bound state confirming the existence of distinct, sequence-dependent topologies, different from the cylindrical rod form known so far. Overall, the 3OS-based generation of unique, sequence-specific, compact topologies indicate that natural HS encodes a dynamic sulfation code that could be exploited for selective recognition of target proteins.

2. Methods

2.1. Molecular dynamics (MD) simulations

The library of 24 variably sulfated hexasaccharides was built using GLYCAM web-server (https://glycam.org) and files for performing MD simulations were downloaded from the server. The structure of each sequence was first confirmed, which included the order of residues, the bond lengths and angles, and initial torsional angles. For MD simulations, ring puckering of IdoA was restrained using weak torsional restraints and puckering occupancies were ascertained before and after simulations using the BFMP method [26,27]. MD simulations were performed using AMBER18 with GLYCAM06J force field, essentially as described earlier [28,29]. Each sequence was neutralized with an appropriate number of Na^+ counterions and then solvated using TIP3P water with periodic boundary conditions such that a minimal distance of 12 Å was maintained between solute atoms and the nearest box edge. A two-step minimization was carried out for each hexasaccharide, wherein the solute was restrained with a harmonic potential of 100 kcal/(mol $Å^2$) in the first step. Here, water molecules were "relaxed" for 500 iterations of steepest descent and 1500 cycles of conjugate gradient minimization. In the second step, conjugate gradient minimization was performed for 2500 iterations without any restraint. In the next step, the entire system (Hp/HS sequence in TIP3P water box) was equilibrated by implementing a Berendsen thermostat to achieve a constant temperature of 300 K first and then a pressure of 1 atm. Finally, the system was equilibrated without any restraints. All minimization and equilibration steps utilized a 2 fs timestep for a total of 1 ns. MD production runs were performed for 1 μs during which the coordinates were collected and stored every 10 ps. Covalently bonded hydrogen atoms were constrained using the SHAKE algorithm, as described in AMBER [28,29].

2.2. Analysis of MD production runs

Torsional (Φ,Ψ) space, solvation, hydrogen bonding, end-to-end distance (EED), sodium cation (Na^+) binding, and potential energy calculations were performed using the CPPTRAJ package with AmberTools14 [30]. IUPAC definitions for HS were used for glycosidic torsional angles, moving from the non-reducing end to reducing end: Φ = O5-C1-O1-C4 and Ψ = C1-O1-C4-C5 [31]. Torsional density maps were prepared using the 'displot' function in Seaborn for bivariate frequency distributions with the default parameters. Selection of Φ,Ψ data points corresponding to each minima was enabled by the findpeaks function in SciPy. The proportion of the secondary minimum at each linkage was calculated by generating a 2D histogram using the histogram2d function from the NumPy library and dividing by the total number of frames (100,000). EED was calculated as mentioned previously to approximate the length of the GAG, respectively [32]. Na^+ binding was modeled by establishing distance criteria, where a Na^+ was observed within 3.5 Å of C6 of Residues 2 and 4. Both distances were measured in parallel using the *DISTANCE* command in CPPTRAJ throughout the trajectory. Na^+ binding was measured as a binary function, that is, 1 if present at the time the trajectory was captured (every 0.01 ns) and 0 if not. To evaluate changes in like-charge repulsions between carboxylates and sulfates, the coulombic potential energy term was isolated from the force field for the functional groups of interest using the energy command in CPPTRAJ. Default parameters were used in calculating electrostatics. Only terminal oxygens of residue III 3-*O*-sulfate, residue IV carboxylate, and residue V 6-*O*-sulfate were included in the calculations. Number of hydrogen bonds (H-bonds) per functional group were calculated by including only hydrogen bond acceptors and donors, as shown in Table S2. Water loss during conformational changes relied on changes in these H-bonds along the MD trajectory. An H-bond was considered formed if the distance between donor (D) and acceptor (A) atoms was less than 3.5 Å and the D-H····A angle was 180 ± 60°.

2.3. MD simulations of the 3-O-sulfated octasaccharide

Octasaccharide GlcNS6S(1 → 4)IdoA2S(^2S_O) (1 → 4) GlcNS6S3S (1 → 4) IdoA2S($^1C_4$) (1 → 4) GlcNS6S (1 → 4) IdoA2S ($^1C_4$) (1 → 4) GlcNS6S (1 → 4) IdoA2S($^1C_4$) was built using the GLYCAM web server following the process for hexasaccharides as described above. Residue puckers were restricted to those observed in the co-crystal structure (PDB:3INA). Puckering occupancies were assessed using the BFMP method [26]. The minimization, equilibration, and production MD protocols for this octasaccharide were identifical to those as described for the hexasaccharide library. The analysis of MD production runs for EED, Φ/Ψ, water loss, Na^+ binding analyses, etc. was performed in a manner identical to that of the hexasaccharides. A binary function allowed for the monitoring of three separate Na^+ binding interactions, which included residue III to I, residue V to III, and residue V to I.

3. Results

3.1. Rationale behind the design of the hexasaccharide library

To understand whether the presence of a 3OS group endows a Hp/HS chain with unique conformational states, we studied a library of 24 Hp/HS hexasaccharides. Theoretically, four uronic acid (UA) and twelve glucosamine (GlcN) monomers are possible owing to variations at the 2-, 3-, and/or 6-positions,[a] which can generate 48 unique Hp/HS disaccharide building blocks, or 110,592 theoretically possible unique hexasaccharides of $Acid_{NRE}$–$Amine_{RE}$ class (NRE = non-reducing end; RE = reducing end). An identical number of hexasaccharide sequences are possible for the $Amine_{NRE}$–$Acid_{RE}$ class. This represents an astounding configurational diversity for such a small sequence. This diversity is further exacerbated by the different puckers possible for IdoA, of which $^1C_4$ and ^2S_O forms are the most populated [31].

It is now well recognized that most protein-binding Hp/HS sequences are tetra- to octa- saccharide long [12,14,15,33–35]. Hence, we chose hexameric sequences, which allow placing the 3OS group in the middle of the chain. A single 3OS group was positioned on the internal GlcN residue III, which had variations at 2- (acetyl or sulfate) and 6- (sulfate or unsubstituted) positions (Fig. 1). Three other variations were introduced. First, residues II and IV were either GlcA or IdoA with the former in $^4C_1$ form, while the latter in either $^1C_4$ or ^2S_O form. We chose to include IdoA because it has been shown to exhibit a nearly equal proportion of the two puckers [36], which could offer a better glimpse into interesting conformational dynamics. Second, residue V was either GlcNS6S or GlcNAc6S. Third, the 3-position of residue III could be either *O*-sulfated or —OH; the latter forming the six control sequences. The residues at the two termini, i.e., residues VI and I, were maintained invariant as GlcA and GlcNS6S, respectively. Thus, the library could be thought as six sub-libraries, each containing one non-3OS-containing control sequence and three 3OS-containing sequences with variations in the internal tetrasaccharide (Fig. 1). Of the 3OS-hexasaccharides, 12 contained IdoA (either $^1C_4$ or ^2S_O), while 6 had GlcA ($^4C_1$). We reasoned that this library design afforded significant variation in configurational and conformational space. It is important to note that some of these 24 sequences are not known to exist in mammals as yet. The known substrate specificity of the C5 epimerase requires a GlcNS on the non-reducing side of the acceptor GlcA, which disfavors generation of the GlcNAc → IdoA disaccharide units [37]. However, these sequences are likely to be synthetically accessible and could serve as interesting chemical biology and/or drug discovery leads.

3.2. MD simulations indicate occupancy of a novel compact topology in a sequence-specific manner

Recently, a number of MD studies on GAGs in water have been performed using three major forcefields, including GLYCAM, CHARMM and GROMOS [29]. Considerable insights into glycosidic torsional angles, ring puckers, inter-atomic interactions, intermolecular interactions and global shapes of GAG chains have been derived from these studies [38]. In fact, recent studies on large number of GAG disaccharides have further confirmed the classic works by Mulloy and co-workers [31] that glycosidic torsions Φ and Ψ vary within a relatively small range irrespective of sulfation pattern [39,40]. This gives rise to the canonical 'helical' topology of Hp/HS chains, which is used as the primary Hp/HS topology in docking [41–43], although pucker and glycosidic bond flexibility is typically allowed in these studies [44,45]. Yet, Hp/HS chains are thought to be linear cylinders, which implies that "sulfation code" is wholly configurational with only a marginal room for IdoA conformations.

To quickly understand whether 3-*O*-sulfation impacts overall Hp/HS topology, we analyzed a parameter of global 'shape' called the end-to-end distance (EED, see definition in Fig. S1A) [32]. The helical form of the common Hp/HS hexasaccharide (i.e., $(IdoA2S–GlcNS6S)_3$) exhibits an EED of 26.4 Å in its static state [46], which is similar to that derived from its MD simulations [32]. We calculated the EEDs at every 1 ns, i.e., for 1000 frames over 1 μs, for each of the 24 sequences. The results indicated that the majority of sequences displayed EEDs of ∼27 Å. This implies that the majority of sequences occupy the helical, linear topology (Fig. S1B). However, several sequences exhibited major reduction in EEDs. Specifically, major reductions were observed for sequences 2, 3, 12, and 14. Interestingly, each of these sequences contain IdoA residues at positions II and IV, either in the $^1C_4$ (Seq. 2, 3 and 12) or ^2S_O form (Seq. 14).

To understand this better, we analyzed the EEDs of MD frames as a function of time. While the EEDs were essentially invariant (∼27 Å) for all control non-3OS sequences, the profiles for sequences 2, 3, 12 and 14 were significantly different (Figs. 2 and S2). Sequences 2, 3 and 12 displayed a transition to an alternate conformational form of ∼16 Å EED, while sequence 14 preferred a ∼13 Å form (Fig. 2A). Sequence 2 also displayed an additional more compact form with an EED of ∼8 Å. This suggested that this group of four sequences exhibit a high tendency to occupy non-linear, perhaps non-helical, topologies. A comprehensive survey of EED dynamics over the MD runs showed that sequences 1, 4–6 and 13 also sample non-linear conformational forms (Fig. S2); however, the extent of their compaction is not as high as sequences 2, 3, 12 and 14.

To further parse the differences, we calculated the frequencies of MD frames with similar EEDs (every 0.2 Å). Whereas each of the six control sequences displayed a primarily unimodal EED profile with maximal length of ∼27 Å, 3OS containing sequences displayed a non-unimodal distribution (Fig. S3). The most interesting among these were sequences 2, 3 and 12, which exhibited an unusual bimodal or trimodal distribution with EEDs of ∼27, ∼16 and ∼8 Å (Fig. 2B). Interestingly, sequence 14 was observed to be in its compact form for most of its MD simulation time with an EED of ∼13 Å. Other 3OS containing sequences, i.e., 1, 4–6 and 13, also displayed a bimodal distribution, *albeit* with a different distribution of EEDs (Fig. S3). In contrast, 3OS containing sequences 7 → 9 and 16 → 18, which have GlcA residues flanking 3OS residue, display a primarily unimodal distribution similar to their corresponding controls 3 and 6 (Fig. S3).

[a] Four uronic acids include GlcA, GlcA2S, IdoA, and IdoA2S. Twelve glucosamine residues include $GlcNH_2$, $GlcNH_23S$, $GlcNH_26S$, $GlcNH_23S6S$, GlcNS, GlcNS3S, GlcNS6S, GlcNS3S6S, GlcNAcS, GlcNAc3S, GlcNAc6S, and GlcNAc3S6S.

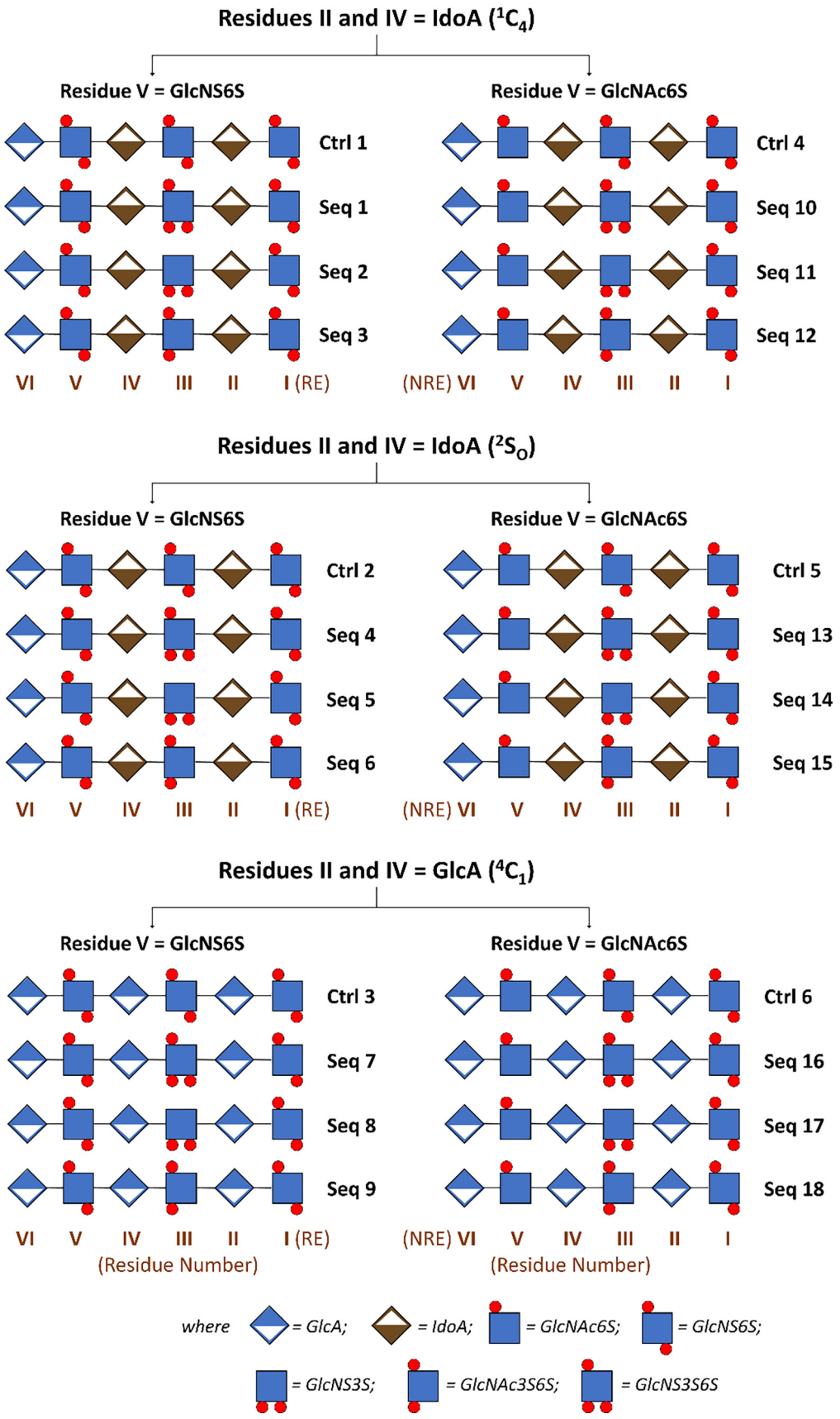

Fig. 1. The library of Hp/HS hexasacchrides. The library can be thought of in terms of 6 subsets derived from variations in the sulfation pattern and uronic acid composition (IdoA ($^1C_4$ or ^2S_O) or Glc($^4C_1$)) of the four middle residues (Residues II through V). The terminal (reducing end (RE) and non-reducing end (NRE)) residues were maintained invariant as GlcNS6S and GlcA, respectively. Residue III was either 3-*O*-sulfated or unmodified at 3-position, which served as a control (Ctrl) for each subset. Symbols for each monosaccharide follow the Symbol Nomenclature for Glycans (SNFG) system, while presence of sulfate groups is represented by a red circle (see legend at the bottom). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

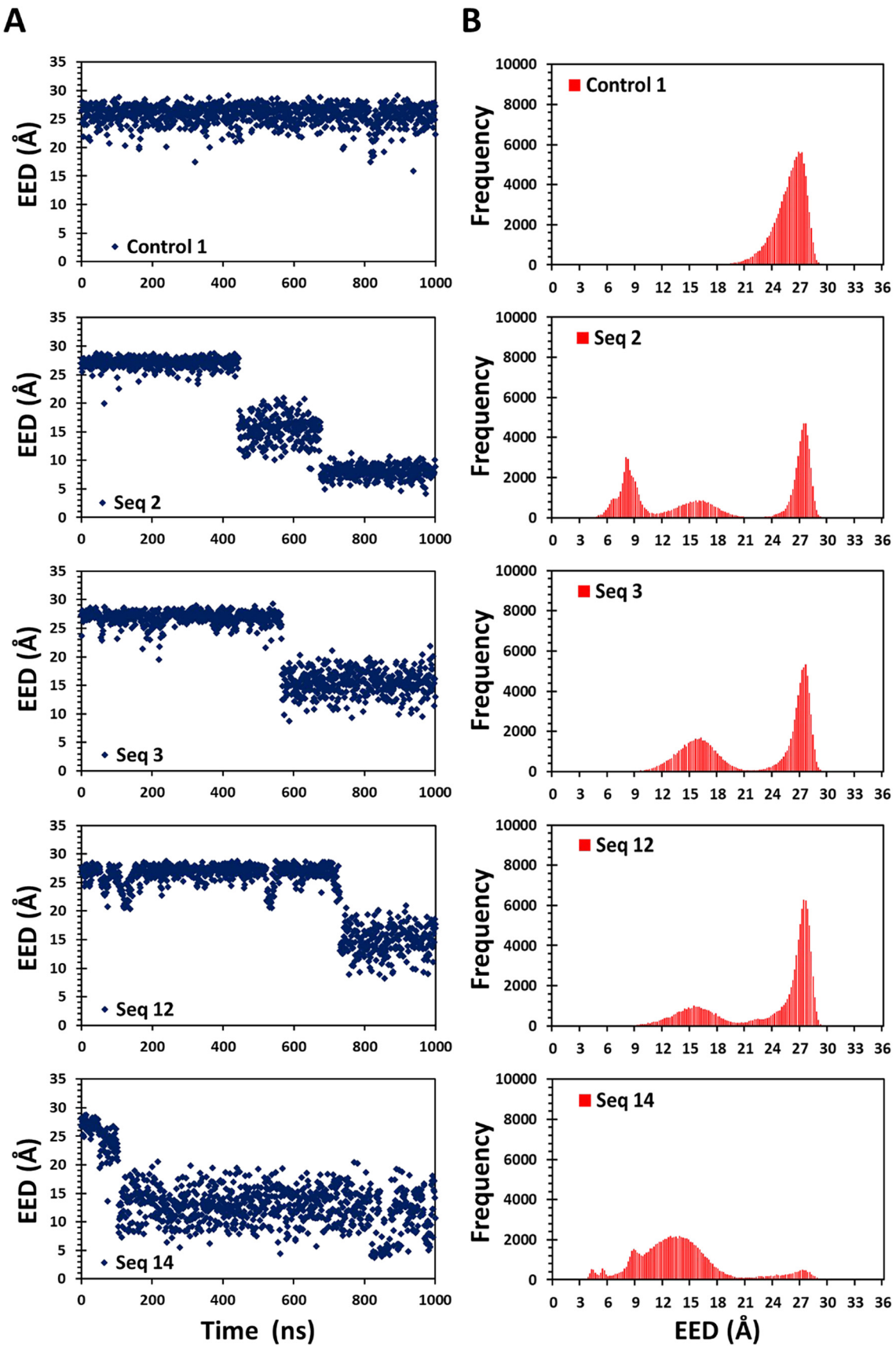

Fig. 2. Profiles of end-to-end distances (EEDs). EED as a function of MD simulation time (A) and frequency of occurrence (B) are shown for control 1 and sequences 2, 3, 12 and 14. Corresponding profiles for control sequences 2 → 6 and sequences 4 → 11, 13, and 15 → 18 are shown in supplementary Figs. S2 and S3. EEDs were calculated every 1 ns for the 1 μs MD simulations (i.e., 1000 frames). For frequency distributions, the bin size was set to every 0.2 Å.

The above results indicate that when a 3OS group is present, a hexasaccharide samples multiple topologies depending on the structure of the sequence. Hexasaccharides with 3OS group and IdoA residues considerably populate non-linear, compact topologies. In fact, differing levels of compaction are observed for different types of sequences. Further, hexasaccharides with 3OS group and GlcA residues do not exhibit any tendency to populate one or more C form(s). Thus, the results indicate that the precise sequence of a 3-*O*-sulfated hexasaccharide determines the tendency and extent of compaction. This was the first evidence that some 3OS groups may induce an unusual compact form in Hp/HS sequences. Such unique, three-dimensional, structural motifs likely form the "sulfation code".

3.3. 3-O-sulfation selectively modifies a Psi torsional angle

To identify the site of origin for compaction, we studied torsions Φ and Ψ for each glycosidic linkage of the 24 sequences (Fig. S4). The IdoA/GlcA → GlcN and GlcN → IdoA/GlcA linkages of all control sequences C1 through C6 exhibited Φ/Ψ preferences that were identical to that observed by NMR of heparin dodecasaccaride having the common sequence $(IdoA2S–GlcNS6S)_6$ [46]. This preference was also noted for 3OS sequences (see Table S1 and Fig. S4). However, sequences 2, 3, 12 and 14 displayed additional Φ/Ψ preferences, especially at the internal IdoA → GlcN linkage (i.e., IV → III residues). For example, sequence 2 displayed a second Φ/Ψ preference at −80.8°/66.4°, which represents a massive change in Ψ of nearly 165° from its linear torsional preference (-80.8°/-99.2°) (Fig. 3A). Likewise, sequences 3, 12 and 14 also displayed similar torsional preferences (Fig. 3A, Table S1). Extraction of geometries corresponding to these Φ/Ψ preferences from MD simulations shows that this major change in preference transforms the helical, linear form (EED ~27 Å) to the compact form with EEDs of ~16 Å in sequences 2, 3 and 12, and 13 Å in seq 14 (see models in Fig. 3A). Although the exact topologies of these compact forms are structurally unique, we refer to each of these as 'C' topologies, which are radically different from the linear 'L' form.

To gain some quantitative insights, we calculated the torsional occupancies using EED histograms, which corresponded to the occupancy of the compact form (Φ/Ψ = ~-80°/~66°) (Fig. 3B). The population of these Φ/Ψ torsions was found to be small, or non-existent, in the large majority of sequences, except for sequences 2, 3, 12 and 14. Whereas the C topology was populated nearly 30–70 % in sequences 2, 3, 12 and 14, it was 0–10 % in others. GlcA flanking hexasaccharides were essentially devoid of this conformer (Fig. 3B and S4). Thus, the results convey the role of HS structural pattern in populating the novel compact topology.

3.4. Atomistic origin of the L → C transformation

We sought to reveal the foundational factors that drive the topological switch from helical L shape to the C form. At the thermodynamic level, multiple enthalpic and/or entropic factor(s) may be responsible. We hypothesized that negative enthalpic factors, such as electrostatic repulsion due to sulfate group crowding, may initiate the process. This could then be followed by favorable entropic factors, such as release of Hp/HS-bound water molecules, which may synergize to drive the L → C topological transformation.

To test the above conjecture, we analyzed the L topologies of sequences 2, 3, 12 and 14 and their corresponding controls C1, C4 and C5. Installation of a 3OS group appeared to increase electrostatic crowding due to close proximity of 3-*O*-sulfate, carboxylate, and 6-*O*-sulfate of residues III, IV and V, respectively (Fig. 4A). In contrast, these groups appeared 'isolated' in the C form. To gain better insight, we plotted the repulsive electrostatic force calculated by GLYCAM using a CPPTRAJ script between these anionic groups across the entire MD trajectory [30] (Fig. 4B). The plots show that rotation in Ψ at the IV → III linkage induces a significant drop in Coulombic energy of ≥20 kcal/mol for sequences 2, 3, 12 and 14. In fact, the timing of these drops correlated well with the observed topological transformation from the L to C shape. Further, such decreases in electrostatic energy were not noted for the corresponding control sequences 1, 4 and 5 as well as other 3OS sequences (Fig. S5). Thus, like-charge repulsion may be one of the factors initiating topological transition in the distinct group of 3OS sequences.

3.5. Release of HS-bound water molecules favors compaction

The release of protein-bound water molecules during protein folding has been documented extensively [47,48]. Solvent water molecules provide many roles in protein folding, one of which is to enhance entropic energy resulting in stabilization of the folded state. For GAGs, the role of such water molecules, if any, remains unknown. To assess whether water molecules are involved in L → C topological transformation, we first calculated the number of intermolecular hydrogen bonds (H-bonds) at the individual residue level as a function of MD simulation time [9,30,32,40] (Fig. S6). Residue II of sequences 2, 3, 12 and 14 displayed a significant decrease in the number of intermolecular H-bonds that coincided with the topological transition. This phenomenon was absent in control sequence 1, 4 and 5 as well as in other 3OS sequences (Fig. S6). Interestingly in addition to residue II, residue IV of sequence 14 also showed a loss in number of intermolecular H-bonds.

To understand the atomistic origins of water loss, we repeated the analysis with focus on the Hp/HS partners involved in intermolecular H-bonds. Although several polar atoms/groups were involved, the carboxylates and *O*-sulfates of residues II, III and IV contributed most H-bonds, which remained consistent over the MD time period. However, a decrease of ≥2 intermolecular H-bonds was observed for sequences 2, 3, 12 and 14 upon formation of the compact form, specifically at IdoA carboxylates (Figs. 5 and S7). This loss in bound waters was not observed for their corresponding control sequences as well as for other 3OS sequences (Figs. 5 and S7). The results support our conjecture that the novel compact topology formation manifest from both enthalpic and entropic contributions. Interestingly, a careful review shows that the H-bonding profile of the 3OS groups of sequences 2, 3, and 12 increases by ~1H-bonds. This is absent in other 3OS sequences, such as 9 and 18 (Fig. S8).

3.6. Sodium cation binding stabilizes the compact topology

Since the early 1990s, the binding of metal cations to GAGs has been the subject of many studies [49–52]. A common observation has been that divalent cations, e.g., Ca^{2+}, preferentially bind to carboxylate groups of GAGs. A more recent MD-based investigation of chondroitin sulfate disaccharides concluded that Ca^{2+} prefer to engage GlcA carboxylates rather than *O*-sulfate groups [53].

We hypothesized that additional favorable enthalpic contributions may arise from metal ion binding to contribute to the L → C transformation. To assess this possibility, we evaluated the MD trajectory of sequence 3 for the presence of metal ion, i.e., Na^+, within the coordination sphere (≤3.5 Å) of each carboxylate and sulfate group. Interestingly, one Na^+ ion was found to be engaged by the carboxylates of IdoA residues II and IV in a series of MD frames (Fig. 6A). This represents a bidentate IdoA–Na^+–IdoA interaction, which arises only if the two carboxylates are brought in close proximity due to compaction. Thus, the corresponding control sequence 1 did not exhibit such a phenomenon.

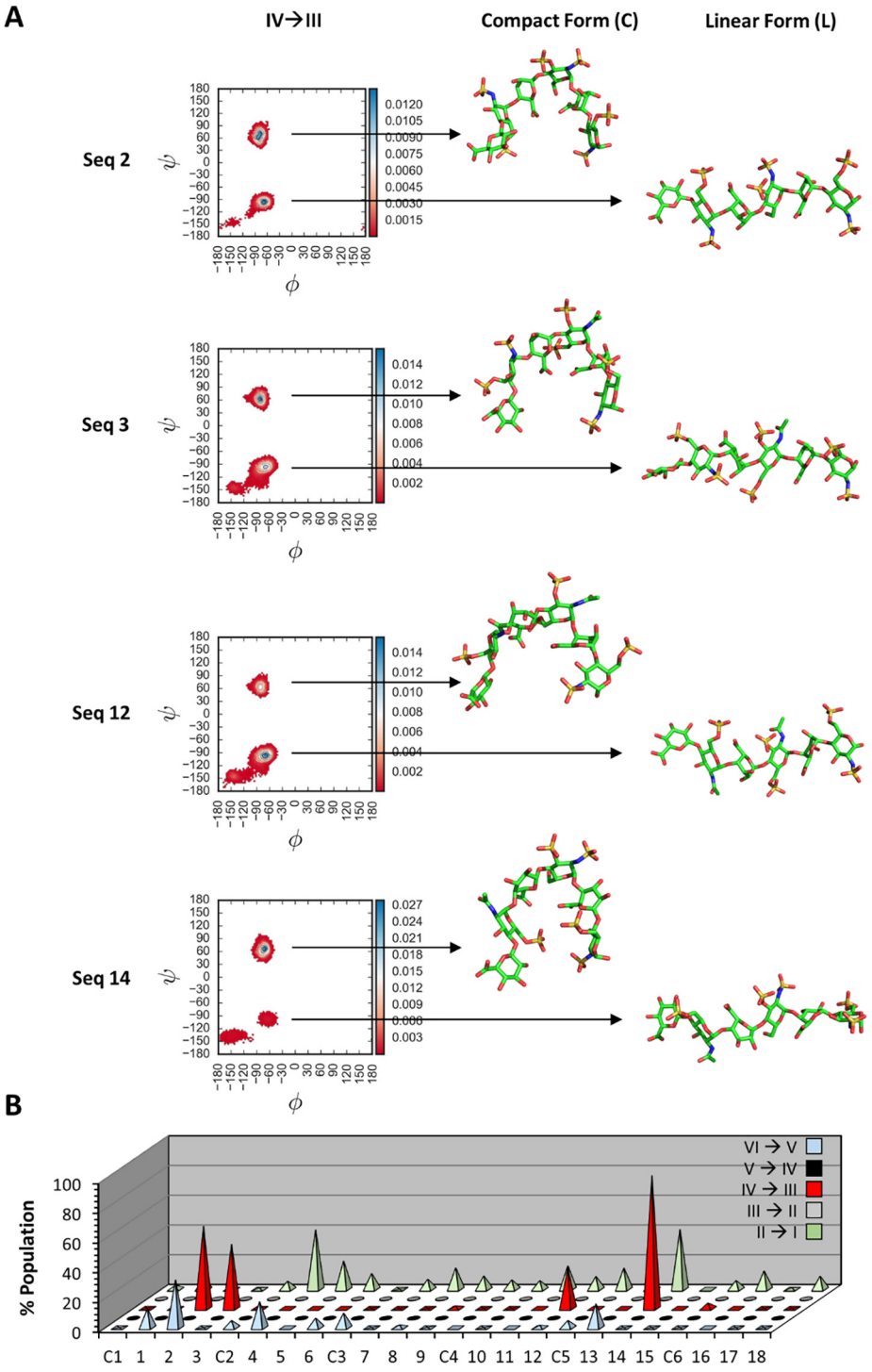

Fig. 3. Maps of glycosidic torsions pinpoint the origin of the compact (C) topology. (A) Ramachandran plot (Φ and Ψ ranges) for the IV → III glycosidic linkage of sequences 2, 3, 12 and 14 (see Fig. S4 of all 24 sequences). IV → III linkage corresponds to the middle IdoA → GlcN linkage of Hp/HS hexasaccharides. Blue regions refer to highly preferred Φ and Ψ. Sequences 2, 3, 12 and 14 exhibit an unusual Φ/Ψ of ∼-80°/∼66°, which induces a unique, compact (C) topology (shown on the right). Also shown is the linear, helical topology corresponding to the Φ/Ψ of ∼-80°/∼-105°. (B) Populations (in %) at the Φ/Ψ = ∼-80°/∼66° for all 24 hexasaccharides calculated using NumPy. Non-reducing terminus residue is labeled VI and reducing end residue is labeled I. Molecular models were generated in PyMOL. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To assess whether this phenomenon is generally applicable, we developed a CPPTRAJ script for identifying the bidentate COO^{-}–Na^{+}–COO^{-} interaction for all 24 Hp/HS sequences. Fig. 6B profiles the results of this analysis. Although several MD trajectories did show the presence of such a bidentate interaction, only those corresponding to sequences 2, 3, 12, and 14 displayed a highly consistent phenomenon that lasted in a prolonged fashion. In fact, the timing of the formation of the bidentate IdoA–Na^{+}–IdoA interaction correlated well with the L → C transition (see Fig. 5).

It is important to note that the tendency to occupy the C form-related Φ/Ψ space at the IV → III glycosidic linkage is theoretically accessible for all Acid → Amine linkages (see Fig. S4, Table S1). This

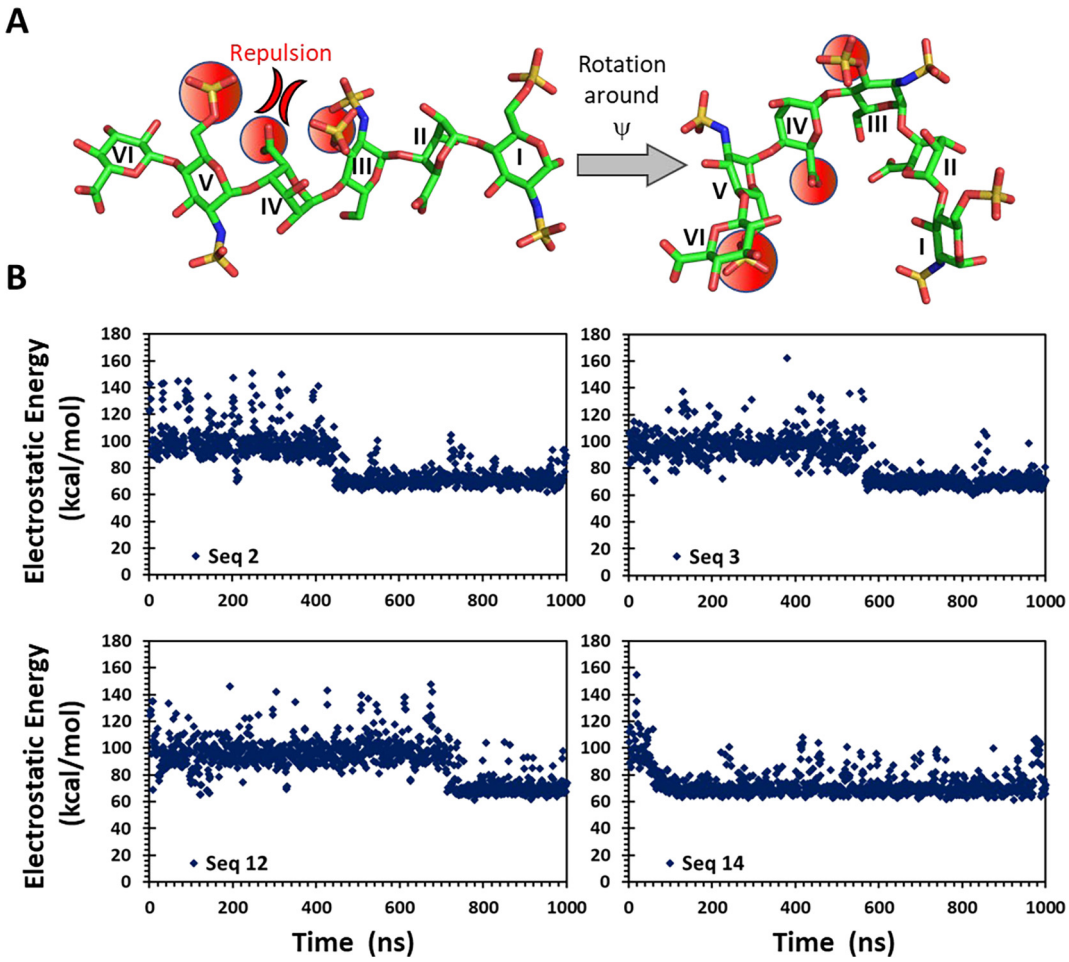

Fig. 4. Bad electrostatics contributes to the induction of the compact form. (A) Sequence 2 exemplifies this phenomenon. The 3-O-sulfate, 5-carboxylate, and 6-O-sulfate groups of residues III, IV and V, respectively (red spheres) in sequence 2 generate electrostatic crowding, which is relieved by rotation around Ψ of the IV → III glycosidic linkage. (B) Profiles of changes in electrostatic energy in the above dense anionic cluster as a function of the MD simulation time. The decrease in electrostatic energy coincides with the observed change in EED as a function of time (see Fig. 2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

implies that longer oligosaccharide may present opportunities for the formation of additional bidentate IdoA–Na^+–IdoA interactions, which may or may not compete with each other. Overall, the results support the hypothesis that the C form is likely to be stabilized by IdoA–Na^+–IdoA interaction, just as loss of bound waters favor the L → C transition.

3.7. Kinetics of the L → C transformation suggests a concerted process

A key question arises whether the three different contributors to the L → C transformation, including electrostatic repulsion arising from multi-anion crowding, intermolecular H-bonding between IdoA carboxylates and water, and simultaneous engagement of Na^+ to two IdoA residues, work independently or in a concerted manner? To assess this, we identified the MD time frames that display the maximal changes in the two descriptors of the L → C transformation, i.e., EED (Fig. 7A) and Ψ of IV → III linkage (Fig. 7B). Whereas sequence 2 exhibits start of topological transformation at about 444 ns, the start times for sequences 3, 12 and 14 were ~566, ~729 and ~104 ns. Interestingly, these time frames for completion of L → C transformation for each of these sequences, as evidenced by EED and Ψ (IV → III), were very short and in the range of 2 to 5 ns. More importantly, the changes in intermolecular H-bonding (Fig. 7D) and bidentate Na^+ complexation (Fig. 7E) occur over nearly the same time frame. In fact, the bidentate Na^+ binding phenomenon appears to arise at the end of the transformation phase following which the C topology remains stable for an extended period of time. A closer review of changes in the electrostatic repulsion between 3OS of residue III, carboxylate of residue IV, and 6OS of residue V indicates a more variable and subtle process (Fig. 7C) that occurs over a longer time period and is initiated much before the rapid L → C glycosidic reorientation. This suggests that bad electrostatic crowding may be a contributor but not essential for L → C transformation. Overall, the kinetics of L → C transformation suggests a concerted process.

3.8. MD of an HS octasaccharide supports the principles enunciated from HS hexasaccharides

To assess whether the novel compact topology identified for 3OS HS hexasaccharides can be found in nature, we reviewed 3D structures of proteins in complex with oligosaccharides in the protein data bank (rcsb.org). A pentasaccharide containing one 3OS group has remained the major ligand studied to date. This includes fondaparinux in complex with adeno-associated virus-DJ [54], bacterial polysialyltransferase [55], platelet factor 4 [56], staphylococcus TarM [57], thrombospondin-1 [58] and antithrombin [34]. Unfortunately, fondaparinux has a GlcA-GlcN-IdoA motif that is different from the motifs studied in this work. Crystal structures of several longer oligosaccharides are reported; however, none of

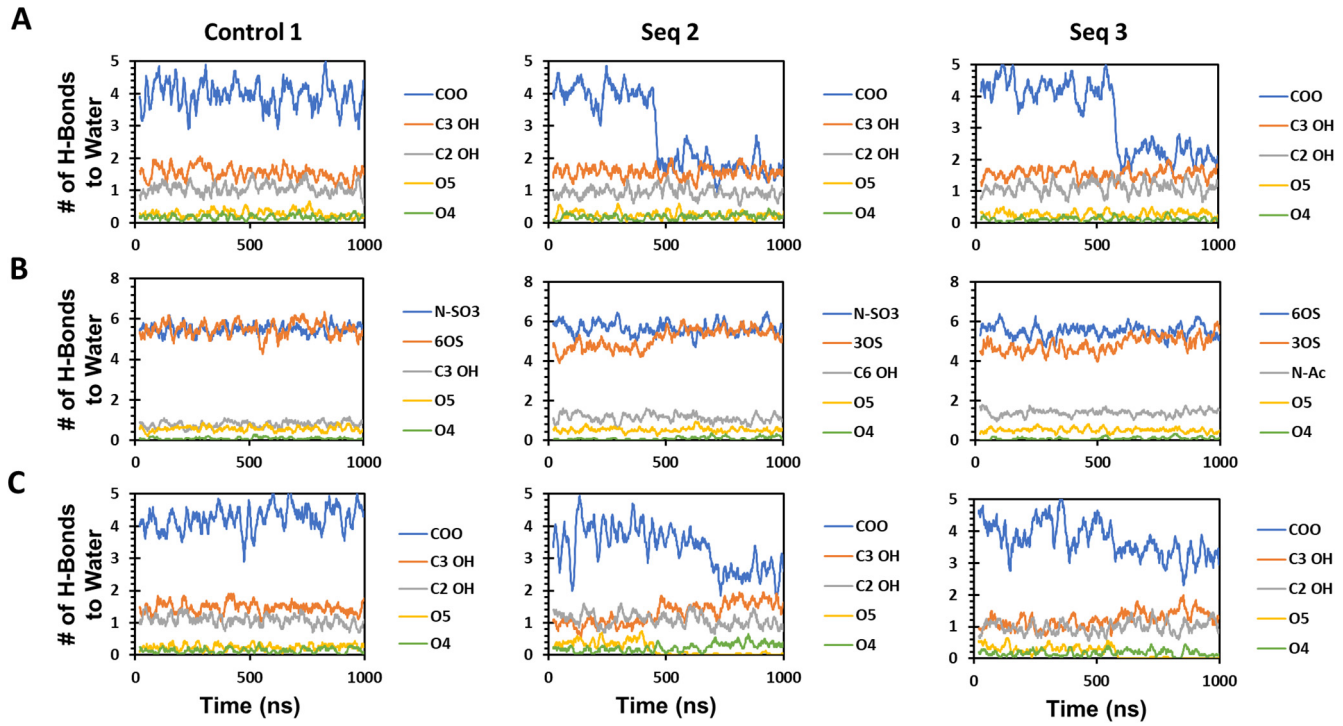

Fig. 5. Intermolecular H-bonds between solvent water and specific groups of HS residues. H-bonds formed over the 1 μs MD trajectory are shown for residues II (A), III (B) and IV (C) of control 1 and 3OS sequences 2 and 14 (see Fig. S7 for sequences 3 and 12). Note residues II and IV are IdoA, while residue III is 3-*O*-sulfated GlcN. Intermolecular H-bonds were calculated using CPPTRAJ.

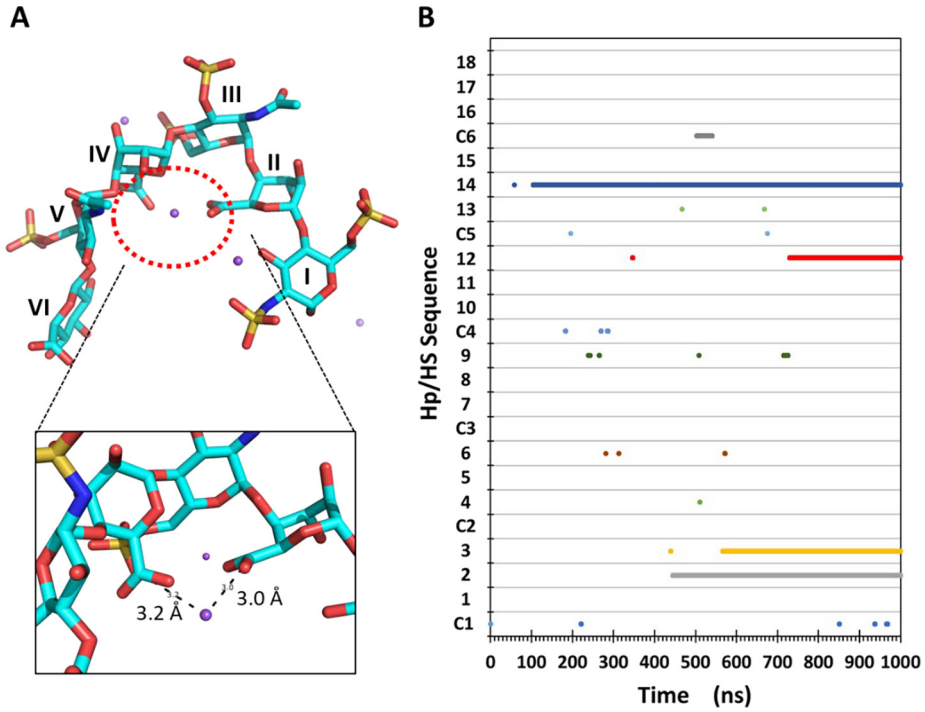

Fig. 6. Stabilization of compact form arising from formation of bidentate IdoA-Na^+-IdoA interaction. (A) The carboxylates of IdoA residues II and IV of sequence 3 generate a favorable site for Na^+ resulting in stabilization of the compact (C) form. (B) Of the 24 Hp/HS sequences, a stable, consistent bidentate IdoA-Na^+-IdoA interaction occurs only for sequences 2, 3, 12 and 14.

these contain the key 3OS group. Only one crystal structure of a heparin lyase in complex with an octasaccharide with 3OS group (PDB: 3INA) has been reported to date [59]. Although the structure of this 3INA octasaccharide (GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S-IdoA2S-GlcNS6S-IdoA2S) is different from the library of our hexasaccharides, it displays a characteristic compact topology shown by the four 3OS sequences in our MD simulations (Fig. 8A). Additionally, the compact topology was stabilized by a key biden-

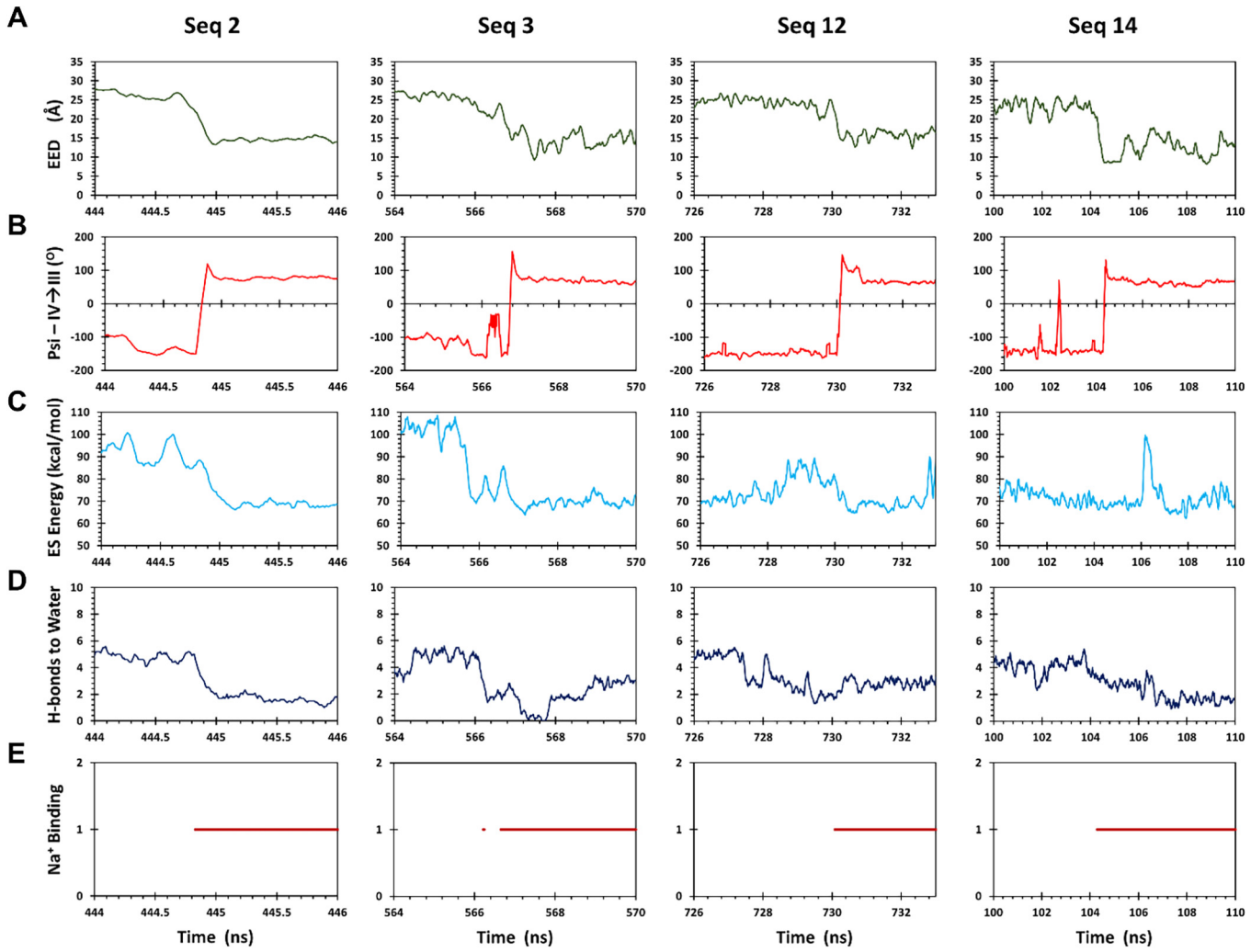

Fig. 7. The L–>C transition is concerted. Each descriptor of the L → C transition including EED (in Å), Ψ of IV → III linkage (in degrees), electrostatic energy between anions of residues III and IV (in kcal/mol), number of intermolecular H-bonds between carboxylates of IdoA and water, and number of Na^+ bound simultaneously to IdoA residues II and V (bi-dentate interaction) as a function of MD simulation time is shown for sequences 2 (A), 3 (B), 12 (C) and 14 (D). Only the brief window of the L → C transition, wherein significant changes in each of these descriptors, is shown for clarity. Profiles shown are 10-point moving averages calculated for 100,000 frames over 1 μs simulations.

tate IdoA-Cation-IdoA interaction in which an arginine residue served as a cation (Fig. 8B). This provides strong support for the existence of unique C topologies in nature.

To assess whether the octasaccharide exhibits similar behavior in MD, we built the starting linear topology using GLYCAM-Web and subjected it to 1 μs MD simulation, as described for the hexasaccharide library. Visual inspection of the 100,000 MD frames of the octasaccharide showed that ~100 ns into the simulation, the linear topology transformed into a compact form. This L → C transformation was accompanied with an EED decrease from ~33 Å to ~9 Å (Fig. 8C). Frequency distribution of the different conformers formed during the MD run indicated that the compact topology was the most abundant form (Fig. 8D). Closer inspection of the different glycosidic torsions revealed that the compact topology was generated by two Ψ reorientations at the III → II and V → IV linkages (Fig. 8E). Of these, the crystal structure presents only the latter Ψ reorientation. In fact, the V(IdoA2S) → IV(GlcNS6S) linkage exhibits a Ψ of 63.5°, which is remarkably close to the value presented by 3OS-containing hexasaccharide sequences at ~63° (Fig. 8E and Table S1).

To further address the similarity of atomistic mechanisms behind the topological transformation in 3INA octasaccharide to 3OS hexasaccharides, we analyzed the two key components that help stabilize the C topology, i.e., loss of intermolecularly H-bonded water molecules and gain of bidentate IdoA-Na^+-IdoA interaction. Similar to the 3OS-containing hesasaccharides, the results show that IdoA2S residues I and III of the 3INA octasaccharide lose about 2 to 3 water molecules as a results of the topological transformation around 100 ns (Fig. 9A). IdoA2S residues V and VII do not exhibit this phenomenon (Fig. S10), which correlates well with no noticeable changes in glycosidic torsions near the non-reducing terminus of the 3INA octasaccharide (not shown). Likewise, the MD profile shows a robust IdoA-Na^+-IdoA bidentate interaction for the 3INA octasaccharide in which three IdoA carboxylates take part (Fig. 9B). Of these, residues I and III form more consistent interaction between them than residues I and V. Overall, atomistic foundations behind the MD profile of the octasaccharide are very similar to that of sequences 2, 3, 12 and 14, thereby supporting the mechanics and generality of the L → C transformation.

4. Discussion

In this study, we discovered that a small group of 3OS-containing hexasaccharides significantly occupied one or more compact topologies (C forms) in the dynamical state, while also populating the linear, helical topology (H form) to some extent.

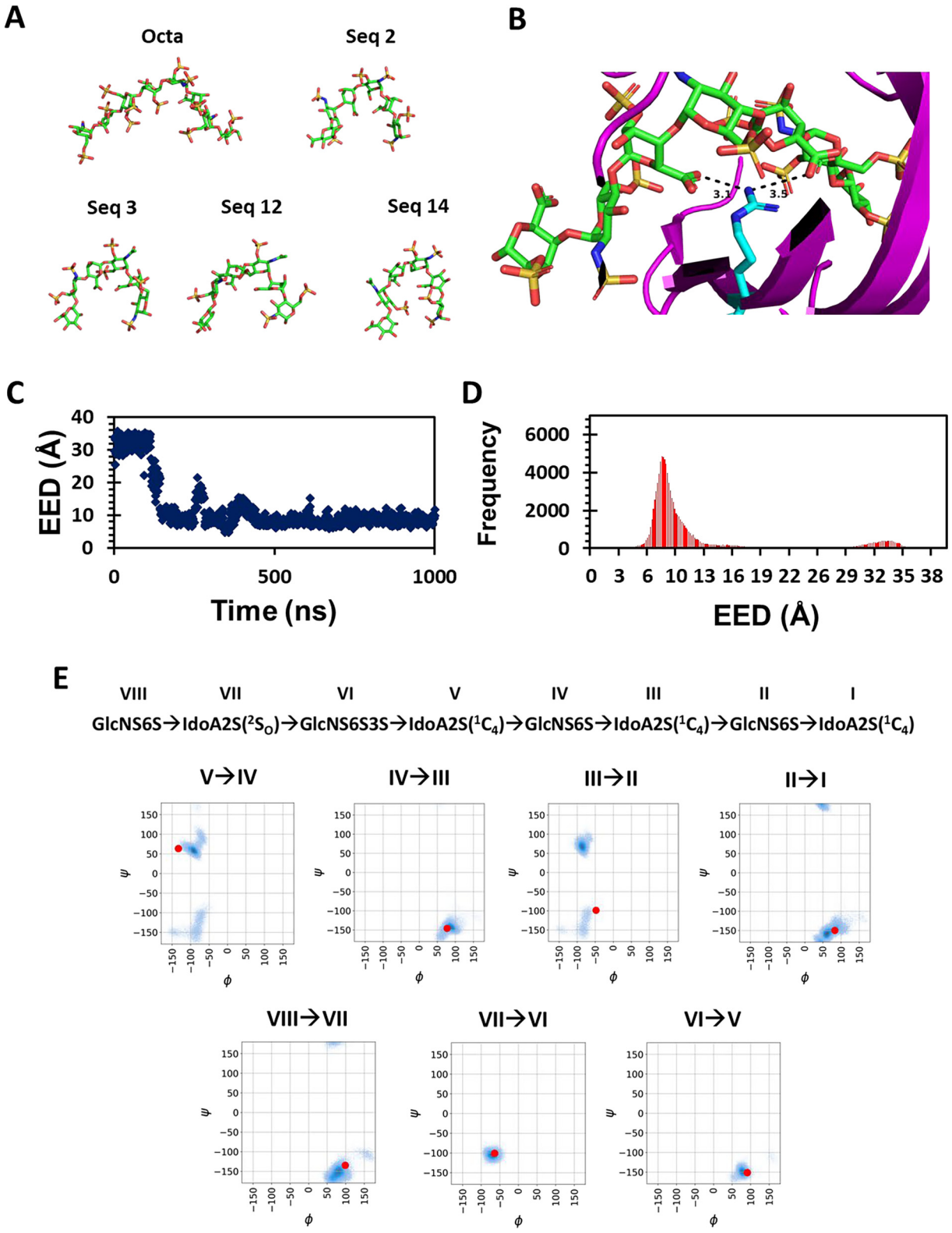

Fig. 8. MD simulations of a 3-*O*-sulfated octasaccharide. A crystallized co-complex (PDB: 3INA) of 3OS octasaccharide in complex with a heparin lyase shows the characteristic compact topology observed for the four 3OS-containing HS hexasaccharides. (A) PyMOL-genereated molecular models of the 3INA octasaccharide from the co-crystal structure and MD-derived compact topologies of Seq 2, 3, 12 and 14. (B) Arg83 forms a key salt-bridging interaction with adjacent IdoA COO^- groups, stabilizing the compact form of the 3INA octasaccharide. (C and D) MD studies on the 3INA octasaccharide showing EED as a function of time (C) and frequency distribution (D). EEDs (shown for every 1 ns) and frequency distributions (over 1 μs) were calculated for all 100,000 frames using a bin size of 0.16 Å. (E) The sequence of the 3INA octasaccharide and the Φ/Ψ space for each of its linkages derived from MD run (Res I is RE; VIII is NRE). Blue regions correspond to higher probability of occurrence, while red dots correspond to crystallographically observed Φ/Ψ values for the 3INA octasaccharide. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

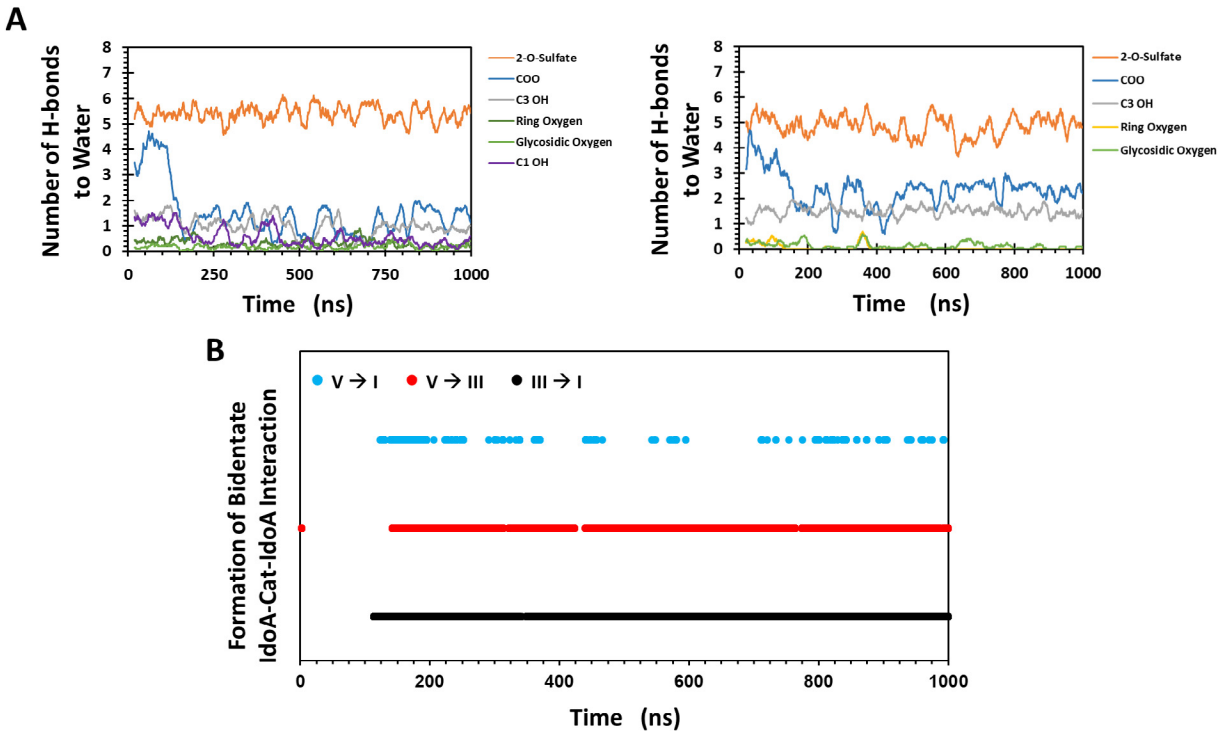

Fig. 9. MD simulations of a 3INA octasaccharide supports principles enunciated from the study of HS hexasaccharides. The 3INA octasaccharide displayed a loss of 2–3 water molecules (A) and gain of bidentate IdoA2S-Na^+-IdoA2S interaction (B) as the L → C transformation sets in around 100 ns. Intermolecular H-bonds were calculated using CPPTRAJ for all frames of 1 μs MD trajectories. Shown here are residues I and III, while other residues are shown in Fig. S9.

Whereas the C form(s) were observed for sequences containing two IdoA residues, not all 3OS hexasaccharides with two IdoA residues sampled the C form to a significant extent. Also, none of the GlcA-containing 3OS hexasaccharides populated the C form. Likewise, none of the control sequences, which were devoid of a 3OS group, populated any of the C forms. These results were also found to hold parallelly for a naturally occurring octasaccharide containing one 3OS group. Our discovery challenges the field's long-held understanding that irrespective of the sulfation pattern, HS exhibits a linear helical form with the possibility of small "kink(s)" around IdoA residues [27,60–67]. Our work shows for the first time that the conformational diversity of HS is much higher and detailed studies with configurationally diverse 3OS-containing oligosaccharides are critically needed to truly appreciate the static and dynamic nature of the HS sulfation code.

Within the class of 3OS sequences populating the compact form, certain sulfation pattern, e.g., sequences 3 and 12, may present one C topology, whereas others, e.g., sequences 2 and 14, may exhibit more (Figs. 2 and 3). Alternatively, compact topologies are induced in a sequence-specific manner. This observation was also supported in MD studies of the 3INA octasaccharide, which contained one 3OS group and two 2-*O*-sulfated IdoA residues. Thus, 3INA octasaccharide and hexasaccharides 2, 3, 12 and 14 are structurally different. Yet, the 3INA octasaccharide exhibited a compact topology, which corresponded to the same change in Ψ as the hexasaccharides (Fig. 8). The observation that the 3INA octasaccharide populates a compact topology in the co-crystallized state with a protein lends experimental foundation to the concept of sequence-specific induction of compact topology.

One can think of natural HS as a repository of the most diverse "sulfation codes" that engineer selectivity of protein recognition. Yet, the sulfation code is not biosynthesized in the classic template-driven manner, as demonstrated for nucleic acids and proteins. To explain its origin, it has been suggested that the sulfation code could arise from the controlled spatiotemporal expression of the superfamily of *N*- or *O*-sulfotransferases [21,68–69]. Of these, the group of 3-*O*-sulfotransferases are most probably the most important. Not only is the 3OS group rare in HS [17,70–71] and has been implicated in mediating selective recognition of proteins [14–15,34,6], nature utilizes seven different isoforms to engineer just one structural modification. Thus, the 3OS group has been thought of as the key driver of the sulfation code. Yet, how it engineers a unique sulfation code has remained elusive.

A priori, one may theorize that an HS sulfation code is generated by either 1) the unique sequence of the oligosaccharide and manifested in the linear helical form and/or 2) conformational effects arising from occupation of different ring puckers, e.g., $^1C_4$, ^2S_O, etc., in the linear helical form. Available co-crystal structural data support the first hypothesis wherein HS sequences exhibit linear, helical form with a complete turn every-four residues [60–66]. Many NMR-based studies in solution support the second hypothesis, wherein neighboring sulfation patterns impact ring puckering equilibria to engineer a sulfation code within the linear, helical form [27,67]. Yet, none of these clearly identify the criticality of a 3OS group.

Interestingly, some clues on how a 3OS group could contribute to this phenomenon were already available in co-crystal structures of fondaparinux, the five-residue sequence with one 3OS group in the center of the chain. This pentasaccharide displays a "kink" in multiple co-crystal structures with antithrombin and platelet factor 4 [34,56], which indicates a slight deviation from the linear, helical form. This departure from the linear form may not necessarily originate during protein binding because recent MD studies with 3OS-containing di- and penta- saccharides in the free state demonstrate the sampling of the kinked forms [39,72]. With this work on longer oligosaccharides, it is clear that sequence-specific generation of compact topologies, i.e., dynamical sulfation codes, requires the presence of certain 3OS groups. It is important to note

that 3OS group may not be the only contributor, especially when considering binding to proteins. For example, the ability of IdoA to switch ring puckers upon binding to antithrombin has been well-documented and likely to be an additional mechanism to explain selectivity of binding [73]. It is also important to not discard the possible contributions of IdoA2S, which are known to be important [74–76] and may or may not be similar to IdoA.

The compact topology-based sulfation code requires the formation of one or more bi-dentate IdoA-Cation-IdoA interaction, which stabilizes L → C transformation through a reorientation of Ψ at an appropriate IdoA → GlcN linkage (Fig. 6). Our observation that reorientation occurs at an IdoA → GlcN linkage, and not alternate GlcN → IdoA/GlcA linkages, agrees with earlier literature highlighting the flexibility of these linkages in HS [67,77]. Interestingly, sequences 2, 3 and 12 carried IdoA in $^{1}C_{4}$ form, while sequence 14 had IdoA in $^{2}S_{O}$ form (Fig. 1), which further conveys the importance of conformational bias in guiding overall topology.

At an atomistic level, our work resulted in delineating the underlying forces that direct the L → C transformation. These forces include 1) like-charge repulsion arising from sulfate group crowding (Fig. 4), 2) loss of HS-bound water molecules upon compaction (Fig. 5), and 3) favorable interactions of the bidentate IdoA-Cation-IdoA type following L → C transition (Fig. 6). Of these, electrostatic crowding while being clearly synchronous for sequences 2 and 3, was less synchronous for sequences 12 and 14 (Fig. 7). This suggests that the latter two forces, i.e., energetic gain from water loss and bidentate interactions, are likely to be more dominant contributors.

Compaction of proteins, which is associated with loss of bound water molecules, has been known to provide entropic benefits [47,48]. In fact, it is known that protein folding relies on formation of a hydrophobic core that necessitates expulsion of water molecules. In fact, in a two-state protein folding model, stabilization of the 'native' state occurs following expulsion of "interstitial water and consolidation of the [hydrophobic] core" [78]. Our work demonstrates a rather similar phenomenon for HS sequences that undergo compaction. Expulsion of two or more water molecules occurs essentially simultaneously with bidentate IdoA-Cation-IdoA interaction, which could be thought of as consolidation of the core. To the best of our knowledge, this is the first report on the expulsion of bound water molecules during a topological transformation. In contrast, the cation binding property of HS has been well documented [49–52]. Yet, the bidentate IdoA-Cation-IdoA "code" is novel.

A key question originates from this study on oligosaccharides whether larger HS chains exhibit compaction. Compaction or bending has been observed in HS and other glycosaminoglycans [79–81]. Yet, the extent of compaction is difficult to study because of polydispersity and heterogeneity. In fact, details on points of compaction or bending remain unknown. This work pinpoints atomistic origins and mediators of compactions, which could now be assessed at the polymeric level.

This work also conveys the importance of studying other members of the hexasaccharide library. Our library was designed to include IdoA and GlcA residues in a uniform manner. Natural and synthetic libraries could include a mix of IdoA, IdoA2S, GlcA and GlcA2S. It is very likely that these expand the concept of "dynamic sulfation code" including variations around the L → C transition.

Finally, our work does not attempt to offer one-to-one correlation between the substrate specificity of an HS3ST and a specific dynamic sulfation code. With the understanding gained so far, it is difficult to derive such as one-to-one correlation, even if it exists. Our work provides a reasonable foundation for supporting such a possibility and validating such as hypothesis will require much advanced work.

Overall, this study proposes a fundamental change in the concept of "sulfation code". Rather than the sulfation code being a static configurational entity, it proposes a more dynamic conformational entity. Although we have studied 24 hexasaccharides and one octasaccharide, it is important to note that this is still a small number of HS sequences. Our work conveys the importance of understanding the topological terrain of 3OS oligosaccharides. Equally important would be to identify unique HS sequences that manifest the dynamic sulfation code and employ it in developing therapeutics or diagnostics against specific protein targets.

Author contributions

SGH performed analysis of MD experiments and contributed to paper writing; BN set up all MD experiments, reviewed and contributed to paper writing; URD supervised and guided research, revised and finalized the paper, and acquired funding. All authors have read and agreed to the publication of this work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by NIH grants K12 HL141954, U01 CA241951, P01 HL151333 and R61 HL161813 to URD. We also thank the computing resources made available through the S10RR027411 grant from the National Center for Research Resources to VCU.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.07.013.

References

[1] Critical Care 2020 24 454 van Haren FMP, Page C, Laffey JG, Artigas A, Camprubi-Rimblas M, Nunes Q, Smith R, Shute J, Carroll M, Tree J, Carroll M, Singh D, Wilkinson T, Dixon B. Nebulised heparin as a treatment for COVID-19: scientific rationale and a call for randomised evidence. Crit Care. 2020 Jul 22;24(1):454.
[2] Huang Y, Song Y, Li J, Lv C, Chen ZS, Liu Z. Receptors and ligands for herpes simplex viruses: Novel insights for drug targeting. Drug Discov Today 2022 Jan;27(1):185–95.
[3] Kovacsovics TJ, Mims A, Salama ME, Pantin J, Rao N, Kosak KM, et al. Combination of the low anticoagulant heparin CX-01 with chemotherapy for the treatment of acute myeloid leukemia. Blood Adv 2018;2(4):381–9.
[4] Patel NJ, Sharon C, Baranwal S, Boothello RS, Desai UR, Patel BB. Heparan sulfate hexasaccharide selectively inhibits cancer stem cells self-renewal by activating p38 MAP kinase. Oncotarget 2016;7(51):84608–22.
[5] Thacker BE, Seamen E, Lawrence R, Parker MW, Xu Y, Liu J, et al. Expanding the 3-O-sulfate proteome-enhanced binding of neuropilin-1 to 3-O-sulfated heparan sulfate modulates its activity. ACS Chem Biol 2016;11(4):971–80.
[6] Zhao J, Zhu Y, Song X, Xiao Y, Su G, Liu X, et al. 3-O-Sulfation of heparan sulfate enhances tau interaction and cellular uptake. Angew Chem Int Ed Engl 2020 Jan 27;59(5):1818–27.
[7] Farrugia BL, Lord MS, Melrose J, Whitelock JM. The role of heparan sulfate in inflammation, and the development of biomimetics as anti-inflammatory strategies. J Histochem Cytochem 2018;66(4):321–36.
[8] Rajarathnam K, Desai UR. Structural insights into how proteoglycans determine chemokine-CXCR1/CXCR2 interactions: progress and challenges. Front Immunol 2020;24(11):660.
[9] Vallet SD, Clerc O, Ricard-Blum S. Glycosaminoglycan-protein interactions: the first draft of the glycosaminoglycan interactome. J Histochem Cytochem 2021;69(2):93–104.
[10] Raghuraman A, Mosier PD, Desai UR. Finding a needle in a haystack: development of a combinatorial virtual screening approach for identifying high specificity heparin/heparan sulfate sequence(s). J Med Chem 2006;49 (12):3553–62.

[11] McKeehan WL, Wu X, Kan M. Requirement for anticoagulant heparan sulfate in the fibroblast growth factor receptor complex. J Biol Chem 1999;274 (31):21511–4.
[12] Liu J, Shriver Z, Pope RM, Thorp SC, Duncan MB, Copeland RJ, et al. Characterization of a heparan sulfate octasaccharide that binds to herpes simplex virus type 1 glycoprotein D. J Biol Chem 2002;277(36):33456–67.
[13] Deligny A, Denys A, Marcant A, Melchior A, Mazurier J, van Kuppevelt TH, et al. Synthesis of heparan sulfate with cyclophilin B-binding properties is determined by cell type-specific expression of sulfotransferases. J Biol Chem 2010;285(3):1701–15.
[14] Sankarayanarayanan NV, Strebel TR, Boothello RS, Sheerin K, Raghuraman A, Sallas F, et al. A hexasaccharide containing rare 2-O-sulfate-glucuronic acid residues selectively activates heparin cofactor II. Angew Chem Int Ed Engl 2017;56(9):2312–7.
[15] Chittum JE, Sankaranarayanan NV, O'Hara CP, Desai UR. On the selectivity of heparan sulfate recognition by SARS-CoV-2 spike glycoprotein. ACS Med Chem Lett 2021;12(11):1710–7.
[16] Thacker BE, Xu D, Lawrence R, Esko JD. Heparan sulfate 3-O-sulfation: a rare modification in search of a function. Matrix Biol 2014;35:60–72.
[17] de Agostini AI, Dong JC, de Vantéry AC, Ramus MA, Dentand-Quadri I, Thalmann S, et al. Human follicular fluid heparan sulfate contains abundant 3-O-sulfated chains with anticoagulant activity. J Biol Chem 2008;283 (42):28115–24.
[18] Mochizuki H, Yoshida K, Shibata Y, Kimata K. Tetrasulfated disaccharide unit in heparan sulfate: enzymatic formation and tissue distribution. J Biol Chem 2008;283(45):31237–45.
[19] Yabe T, Hata T, He J, Maeda N. Developmental and regional expression of heparan sulfate sulfotransferase genes in the mouse brain. Glycobiology 2005;15(10):982–93.
[20] Turnbull J, Powell A, Guimond S. Heparan sulfate: decoding a dynamic multifunctional cell regulator. Trends Cell Biol 2001;11(2):75–82.
[21] Borjigin J, Deng J, Sun X, De Jesus M, Liu T, Wang MM. Diurnal pineal 3-O-sulphotransferase 2 expression controlled by beta-adrenergic repression. J Biol Chem 2003;278(18):16315–9.
[22] Xia G, Chen J, Tiwari V, Ju W, Li JP, Malmstrom A, et al. Heparan sulfate 3-O-sulfotransferase isoform 5 generates both an antithrombin-binding site and an entry receptor for herpes simplex virus, type 1. J Biol Chem 2002;277 (40):37912–9.
[23] Liu J, Shworak NW, Sinaÿ P, Schwartz JJ, Zhang L, Fritze LM, et al. Expression of heparan sulfate D-glucosaminyl 3-O-sulfotransferase isoforms reveals novel substrate specificities. J Biol Chem 1999;274(8):5185–92.
[24] Li J, Su G, Xu Y, Arnold K, Pagadala V, Wang C, et al. Synthesis of 3-*O*-Sulfated Heparan Sulfate Oligosaccharides Using 3-*O*-Sulfotransferase Isoform 4. ACS Chem Biol 2021;16(10):2026–35.
[25] Liu J, Pedersen LC. Emerging chemical and biochemical tools for studying 3-O-sulfated heparan sulfate. Am J Physiol Cell Physiol 2022 (in press) https://journals.physiology.org/doi/epdf/10.1152/ajpcell.00110.2022.
[26] Makeneni S, Foley BL, Woods RJ. BFMP: a method for discretizing and visualizing pyranose conformations. J Chem Inf Model 2014 Oct 27;54 (10):2744–50.
[27] Hsieh PH, Thieker DF, Guerrini M, Woods RJ, Liu J. Uncovering the relationship between sulphation patterns and conformation of iduronic acid in heparan sulphate. Sci Rep 2016;14(6):29602.
[28] Nagarajan B, Sankaranarayanan NV, Desai UR. Rigorous analysis of free solution glycosaminoglycan dynamics using simple, new tools. Glycobiology 2020;30(8):516–27.
[29] Nagarajan B, Sankaranarayanan NV, Desai UR. Perspective on computational simulations of glycosaminoglycans. Wiley Interdiscip Rev Comput Mol Sci 2019;9(2):e1388.
[30] Roe DR, Cheatham 3rd TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 2013;9 (7):3084–95.
[31] Mulloy B, Forster MJ. Conformation and dynamics of heparin and heparan sulfate. Glycobiology 2000;10(11):1147–56.
[32] Nagarajan B, Sankaranarayanan NV, Patel BB, Desai UR. A molecular dynamics-based algorithm for evaluating the glycosaminoglycan mimicking potential of synthetic, homogenous, sulfated small molecules. PLoS ONE 2017;12(2): e0171619.
[33] Chopra P, Joshi A, Wu J, Lu W, Yadavalli T, Wolfert MA, Shukla D, Zaia J, Boons GJ. The 3-*O*-sulfation of heparan sulfate modulates protein binding and lyase degradation. Proc Natl Acad Sci U S A. 2021 Jan 19;118(3):e2012935118.
[34] Jin L, Abrahams JP, Skinner R, Petitou M, Pike RN, Carrell RW. The anticoagulant activation of antithrombin by heparin. Proc Natl Acad Sci U S A 1997;94 (26):14683–8.
[35] Schlessinger J, Plotnikov AN, Ibrahimi OA, Eliseenkova AV, Yeh BK, Yayon A, et al. Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. Mol Cell 2000;6 (3):743–50.
[36] Haasnoot CAG, de Gelder R, Kooijman H, Kellenbach ER. The conformation of the idopyranose ring revisited: How subtle O-substituent induced changes can be deduced from vicinal 1H-NMR coupling constants. Carbohydr Res 2020;496:108052.
[37] Debarnot C, Monneau YR, Roig-Zamboni V, Delauzun V, Le Narvor C, Richard E, et al. Substrate binding mode and catalytic mechanism of human heparan sulfate d-glucuronyl C5 epimerase. Proc Natl Acad Sci U S A 2019;116 (14):6760–5.
[38] Nagarajan B, Holmes SG, Sankaranarayanan NV, Desai UR. Molecular dynamics simulations to understand glycosaminoglycan interactions in the free- and protein-bound states. Curr Opin Struct Biol 2022;17(74):102356.
[39] Lutsyk V, Plazinski W. Conformational properties of glycosaminoglycan disaccharides: a molecular dynamics study. J Phys Chem B 2021;125 (39):10900–16.
[40] Nagarajan B, Sankaranarayanan NV, Desai UR. In-depth molecular dynamics study of all possible chondroitin sulfate disaccharides reveals key insight into structural heterogeneity and dynamism. Biomolecules 2022;12(1):77.
[41] Mottarella SE, Beglov D, Beglova N, Nugent MA, Kozakov D, Vajda S. Docking server for the identification of heparin binding sites on proteins. J Chem Inf Model 2014;54(7):2068–78.
[42] Poluri KM, Joseph PRB, Sawant KV, Rajarathnam K. Molecular basis of glycosaminoglycan heparin binding to the chemokine CXCL1 dimer. J Biol Chem 2013;288(35):25143–53.
[43] Joseph PR, Mosier PD, Desai UR, Rajarathnam K. Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: structural plasticity mediates differential binding interactions. Biochem J 2015;472(1):121–33.
[44] Marcisz M, Zacharias M, Samsonov SA. Modeling protein-glycosaminoglycan complexes: does the size matter? J Chem Inf Model 2021;61(9):4475–85.
[45] Sankaranarayanan NV, Desai UR. Toward a robust computational screening strategy for identifying glycosaminoglycan sequences that display high specificity for target proteins. Glycobiology 2014;24(12):1323–33.
[46] Mulloy B, Forster MJ, Jones C, Davies DB. N.m.r. and molecular-modelling studies of the solution conformation of heparin. Biochem J. 1993 Aug 1;293 (Pt 3)(Pt 3):849-58.
[47] Levy Y, Onuchic JN. Water mediation in protein folding and molecular recognition. Annu Rev Biophys Biomol Struct 2006;35:389–415.
[48] Mattos C, Clark AC. Minimizing frustration by folding in an aqueous environment. Arch Biochem Biophys 2008;469(1):118–31.
[49] Whitfield DM, Sarkar B. Metal binding to heparin monosaccharides: D-glucosamine-6-sulphate, D-glucuronic acid, and L-iduronic acid. J Inorg Biochem 1991;41(3):157–70.
[50] Whitfield DM, Choay J, Sarkar B. Heavy metal binding to heparin disaccharides. I. Iduronic acid is the main binding site. Biopolymers 1992;32(6):585–96.
[51] Furth G, Knierim R, Buss V, Mayer C. Binding of bivalent cations by hyaluronate in aqueous solution. Int J Biol Macromol 2008;42(1):33–40.
[52] Stevic I, Parmar N, Paredes N, Berry LR, Chan AK. Binding of heparin to metals. Cell Biochem Biophys 2011;59(3):171–8.
[53] Faller CE, Guvench O. Sulfation and cation effects on the conformational properties of the glycan backbone of chondroitin sulfate disaccharides. J Phys Chem B 2015;119(20):6063–73.
[54] Xie Q, Spear JM, Noble AJ, Sousa DR, Meyer NL, Davulcu O, et al. The 2.8 Å electron microscopy structure of adeno-associated virus-DJ bound by a heparinoid pentasaccharide. Mol Ther Methods Clin Dev 2017;8(5):1–12.
[55] Lizak C, Worrall LJ, Baumann L, Pfleiderer MM, Volkers G, Sun T, et al. X-ray crystallographic structure of a bacterial polysialyltransferase provides insight into the biosynthesis of capsular polysialic acid. Sci Rep 2017;7(1):5842.
[56] Cai Z, Yarovoi SV, Zhu Z, Rauova L, Hayes V, Lebedeva T, et al. Atomic description of the immune complex involved in heparin-induced thrombocytopenia. Nat Commun 2015;22(6):8277.
[57] Sobhanifar S, Worrall LJ, Gruninger RJ, Wasney GA, Blaukopf M, Baumann L, et al. Structure and mechanism of Staphylococcus aureus TarM, the wall teichoic acid α-glycosyltransferase. Proc Natl Acad Sci U S A 2015;112(6): E576–85.
[58] Tan K, Duquette M, Liu JH, Zhang R, Joachimiak A, Wang JH, et al. The structures of the thrombospondin-1 N-terminal domain and its complex with a synthetic pentameric heparin. Structure 2006;14(1):33–42.
[59] Han YH, Garron ML, Kim HY, Kim WS, Zhang Z, Ryu KS, et al. Structural snapshots of heparin depolymerization by heparin lyase I. J Biol Chem 2009;284(49):34019–27.
[60] Faham S, Hileman RE, Fromm JR, Linhardt RJ, Rees DC. Heparin structure and interactions with basic fibroblast growth factor. Science 1996;271 (5252):1116–20.
[61] Pellegrini L, Burke DF, von Delft F, Mulloy B, Blundell TL. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. Nature 2000;407(6807):1029–34.
[62] Lietha D, Chirgadze DY, Mulloy B, Blundell TL, Gherardi E. Crystal structures of NK1-heparin complexes reveal the basis for NK1 activity and enable engineering of potent agonists of the MET receptor. EMBO J 2001;20 (20):5543–55.
[63] Carter WJ, Cama E, Huntington JA. Crystal structure of thrombin bound to heparin. J Biol Chem 2005;280(4):2745–9.
[64] Lee SC, Guan HH, Wang CH, Huang WN, Tjong SC, Chen CJ, et al. Structural basis of citrate-dependent and heparan sulfate-mediated cell surface retention of cobra cardiotoxin A3. J Biol Chem 2005;280(10):9567–77.
[65] Shao C, Zhang F, Kemp MM, Linhardt RJ, Waisman DM, Head JF, et al. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. J Biol Chem 2006;281(42):31689–95.
[66] Li W, Huntington JA. The heparin binding site of protein C inhibitor is protease-dependent. J Biol Chem 2008;283(51):36039–45.
[67] Muñoz-García JC, López-Prados J, Angulo J, Díaz-Contreras I, Reichardt N, de Paz JL, et al. Effect of the substituents of the neighboring ring in the conformational equilibrium of iduronate in heparin-like trisaccharides. Chemistry 2012;18(51):16319–31.

[68] Kuberan B, Lech M, Borjigin J, Rosenberg RD. Light-induced 3-O-sulfotransferase expression alters pineal heparan sulfate fine structure. A surprising link to circadian rhythm. J Biol Chem 2004;279(7):5053–4.
[69] Cadwallader AB, Yost HJ. Combinatorial expression patterns of heparan sulfate sulfotransferases in zebrafish: I. The 3-O-sulfotransferase family. Dev Dyn 2006;235(12):3423–31.
[70] Marcum JA, Atha DH, Fritze LM, Nawroth P, Stern D, Rosenberg RD. Cloned bovine aortic endothelial cells synthesize anticoagulantly active heparan sulfate proteoglycan. J Biol Chem 1986;261(16):7507–17.
[71] Pejler G, Bäckström G, Lindahl U, Paulsson M, Dziadek M, Fujiwara S, et al. Structure and affinity for antithrombin of heparan sulfate chains derived from basement membrane proteoglycans. J Biol Chem 1987;262(11):5036–43.
[72] Janke JJ, Yu Y, Pomin VH, Zhao J, Wang C, Linhardt RJ, et al. Characterization of heparin's conformational ensemble by molecular dynamics simulations and nuclear magnetic resonance spectroscopy. J Chem Theory Comput 2022;18 (3):1894–904.
[73] Ferro D, Provasoli A, Ragazzi M, Torri G, Casu B, Gatti G, et al. J Am Chem Soc 1986;108(21):6773–8.
[74] Guerrini M, Guglieri S, Beccati D, Torri G, Viskov C, Mourier P. Conformational transitions induced in heparin octasaccharides by binding with antithrombin III. Biochem J 2006;399(2):191–8.
[75] Turnbull JE, Fernig DG, Ke Y, Wilkinson MC, Gallagher JT. Identification of the basic fibroblast growth factor binding sequence in fibroblast heparan sulfate. J Biol Chem 1992;267(15):10337–41. PMID: 1587820.
[76] Jemth P, Kreuger J, Kusche-Gullberg M, Sturiale L, Giménez-Gallego G, Lindahl U. Biosynthetic oligosaccharide libraries for identification of protein-binding heparan sulfate motifs. Exploring the structural diversity by screening for fibroblast growth factor (FGF)1 and FGF2 binding. J Biol Chem 2002;277 (34):30567–73.
[77] Muñoz-García JC, Corzana F, de Paz JL, Angulo J, Nieto PM. Conformations of the iduronate ring in short heparin fragments described by time-averaged distance restrained molecular dynamics. Glycobiology 2013;23(11):1220–9.
[78] Akmal A, Muñoz V. The nature of the free energy barriers to two-state folding. Proteins 2004;57(1):142–52.
[79] Sattelle BM, Shakeri J, Cliff MJ, Almond A. Proteoglycans and their heterogeneous glycosaminoglycans at the atomic scale. Biomacromolecules 2015;16(3):951–61.
[80] Ingr M, Kutálková E, Hrnčiřík J. Hyaluronan random coils in electrolyte solutions-a molecular dynamics study. Carbohydr Polym 2017;15(170):289–95.
[81] Khan S, Fung KW, Rodriguez E, Patel R, Gor J, Mulloy B, et al. The solution structure of heparan sulfate differs from that of heparin: implications for function. J Biol Chem 2013;288(39):27737–51.