HHS Public Access

Author manuscript

Wiley Interdiscip Rev Comput Mol Sci. Author manuscript; available in PMC 2020 March 01.

Published in final edited form as:

Wiley Interdiscip Rev Comput Mol Sci. 2019 ; 9(2): . doi:10.1002/wcms.1388.

Perspective on computational simulations of glycosaminoglycans

Balaji Nagarajan[#1,2], **Nehru Viji Sankaranarayanan**[#1,2], and **Umesh R. Desai**[1,2,*]

[1]Institute for Structural Biology, Drug Discovery and Development, Virginia Commonwealth University, Richmond, VA 23298, USA

[2]Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298, USA

[#] These authors contributed equally to this work.

Abstract

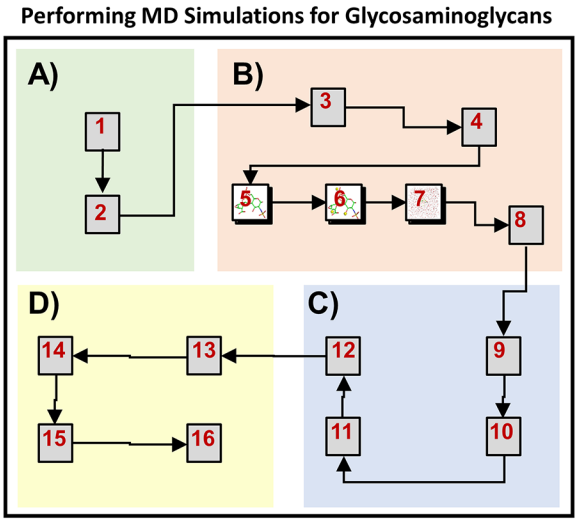
Glycosaminoglycans (GAGs) represent a formidable frontier for chemists, biochemists, biologists, medicinal chemists and drug delivery specialists because of massive structural complexity. GAGs are arguably the most complex, natural linear biopolymers with theoretical diversity orders of magnitude higher than proteins and nucleic acids. Yet, this diversity remains generally untapped. Computational approaches offer major routes to understand GAG structure and dynamics so as to enable novel applications of these biopolymers. In fact, computational algorithms, softwares, online tools and techniques have reached a level of sophistication that help understand atomistic details of conformational variation and protein recognition of individual GAG sequences. This review describes current approaches and challenges in computational study of GAGs. It presents a history of major findings since the earliest mention of GAGs (the 1960s), the development of parameters and force fields specific for GAGs, and the application of these tools in understanding GAG structure–function relationship. This review also presents a section on how to perform simulation of GAGs, which is directed toward researchers interested in entering this promising field with potential to impact therapy.

Abstract

[*]Address for correspondence: Dr. Umesh R. Desai, 800 E. Leigh Street, Suite 212, Richmond, VA 23219. Ph (804) 828-7328; Fax (804) 827-3664; urdesai@vcu.edu.

FURTHER READING /RESOURCES

All the needed additional resources are given in the supplementary material.



Keywords

Molecular dynamics; Force fields; Glycosaminoglycans; Solvent effects; hydrogen bonds; Torsion; Free energy

INTRODUCTION

Despite decades of research, it appears that glycosaminoglycans (GAGs) still represent a frontier that is just beginning to be appreciated by chemists, biochemists, biologists, medicinal chemists and drug delivery specialists. From novice beginnings as 'messy' components of mucosal exudates that retain water and ensure lubricability of joints and cartilage, GAGs are now being accepted as important components of cellular signal transduction. Science is now advancing more in the direction of GAGs as informational molecules with some key sequences exhibiting promising biological activity (Hu et al., 2011; Xu et al., 2011; DeAngelis, Liu, & Linhardt, 2013; Griffin & Hsieh-Wilson, 2013; Schworer, Zubkova, Turnbull, & Tyler, 2013; Patel et al., 2016; Sankaranarayanan et al., 2017; Maza, Gandia-Aguado, de Paz, & Nieto, 2018; Lu et al., 2018; Sankaranaryanan, Nagarajan, & Desai, 2018; Xu, Arnold, & Liu, 2018). The next frontier to be tackled is converting this information into drugs. This frontier presents major opportunities for computational chemist to make an impact, especially because GAGs are arguably the most diverse, natural linear biopolymers made from very few foundational saccharide units.

Structurally, GAGs are linear polysaccharides consisting of repeating disaccharide units made up of hexuronic acid and hexosamine linked by glycosidic bonds (Figure 1). The type of foundational saccharide units and the nature of their inter-residue connectivity differentiates different GAGs. For example, heparin (HP) and heparan sulfate (HS) consist of alternating α(1→4)- and β(1→4)-linked glucosamine (GlcN) and uronic acid (UA) residues, which can either be glucuronic acid (GlcA) or iduronic acid (IdoA), that carry varying levels of sulfate and acetyl groups. On the other hand, chondroitin sulfate (CS) is an alternate co-polymer of repeating *N*-acetylgalactosamine (GalNAc) and GlcA units con-

joined by β(1→4) and β(1→3) linkages, as shown in Figure 1. Dermatan sulfate (DS) is a form of CS, in which the UA residues are more of the IdoA type. Finally, hyaluronic acid (HA), or hyaluronan, is a non-sulfated GAG consisting of *N*-acetylglucosamine (GlcNAc) and GlcA linked by alternating β(1→4) and β(1→3) inter-glycosidic bonds. Besides HP/HS, CS, DS and HA, several other GAGs exist in nature, e.g., keratan sulfate, acharan sulfate. These GAGs, although important in terms of biology, are not discussed in this review and the reader is encouraged to refer to specialized reviews elsewhere (Funderburgh, 2000; Pomin, 2015; Vieira et al., 2004).

The theoretical diversity of GAGs is massive. A back-of-the-envelope comparison of the number of unique, repeating 6-mers of GAGs with 6-mers of proteins and nucleic acids reveals astounding possibilities. Whereas there can be 4.096×10^3 distinct hexa-nucleotides ($=4^6$) from the four common nucleic acid bases, a hexa-peptide made up of the 20 common amino acids could any of the 6.4×10^6 sequences ($=20^6$). In contrast, a HP/HS repeating 6-mer can be any of the 12.2×10^9 sequences originating from a theoretical possibility of 48 unique, repeating disaccharide units. Although not all of the 48 disaccharide units have been identified in nature and such a possibility is also low considering the current knowledge about substrate specificity of GAG biosynthetic enzymes (Esko & Selleck, 2002), the projected chemical space is enticing to both the synthetic chemist as well as the computational biologist interested in advancing the frontier of GAGs as drugs.

In nature, GAGs are found mostly as covalently bound to cell membrane proteins in the form of proteoglycans (PGs), except for HA, which is extruded from cells into the extracellular matrix. Work accomplished in the last decade has presented a large number of interesting roles for GAGs. GAGs bind to various proteins and contribute to fundamental cellular processes, such as growth, differentiation, morphogenesis, and adhesion (Xu & Esko, 2014), while also modulating physiologic and pathologic processes, such as hemostasis, inflammation, tumor growth, microbial infection, immune response, and others (Varki, 2017; Xu & Esko, 2014). At a biochemical level, GAGs can serve as storehouses of proteins and small molecules (Ricard-Blum & Lisacek, 2017; Varki, 2017), modulators of enzymatic activity (Huntington, 2013; Li et al., 2008; Li & Huntington, 2012; O'Keeffe et al., 2004) and protein folding (Hong et al., 2001; Iannuzzi, Irace, & Sirangelo, 2015), and mediators of cartilage biomechanics (Fox, Bedi, & Rodeo, 2009) among others.

Understanding structure and dynamics of GAGs is essential for gaining insight into their myriad biological roles. Although highly sought, it has been difficult to study structure of every GAG sequence through solution-based experiments because GAG synthesis and/or preparation has been challenging (Mende et al., 2016; Pomin & Wang, 2018). Further, conformational and dynamical properties of GAGs is sometimes not possible to elucidate because appropriate high-resolution biophysical approaches are still to be developed. To address these gaps, researchers have developed computational approaches to understand GAG conformation, dynamics and structure–function relationships.

In fact, computational algorithms, software, online tools and techniques have reached a level of sophistication to enables one to understand atomistic details of conformational variation and protein recognition of individual GAG sequences. Although computational studies of

GAGs, especially longer sequences, from the ab-initio perspective continues to be challenging, approaches have been developed to simulate the behavior of GAGs in solution. Computational algorithms, parameters and force fields have been developed to study nearly all saccharide components of GAG as well as their conformation, configuration, and inter-glycosidic linkages with and without solvation.

This review describes current approaches and challenges in computational study of GAGs. It presents a history of major findings since the earliest mention of GAGs (the 1960s), the development of parameters and force fields specific for GAGs, and the application of these tools in understanding GAG structure–function relationship. This review also presents a section on how to perform simulation of GAGs, which is directed toward researchers interested in entering this promising field with potential to impact therapy. The authors do not purport to convey that this review is exhaustive but special attempt has been made to cover the large number of useful papers. Any misgivings are authors' inadvertent mistakes and should not be taken as referring to less important works.

History and Development

The development of computational methods for studying GAGs has relied on tools and protocols developed earlier for neutral carbohydrates and glycans. In fact, the development of the Ramachandran plot (Ramachandran, Ramakrishnan, & Sasisekharan, 1963) helped inquiry into other biomolecules including carbohydrates. For example, preferred torsions in seaweed polysaccharides and their allowed/disallowed regions were studied based on close contacts (Rees, 1969). The study also identified left- or right-handed helical folds. In another study, the conformations of an ionic polysaccharide were explored using the chain model, which took into account dipole-dipole interactions in potential energy calculations (Cleland, 1971). Simultaneously, efforts to understand conformational preferences of individual saccharides in solution, e.g., aldohexopyranose and aldopentofuranose, were also made by using MO-LCAO methods (Del Re, Pullman, & Yonezawa, 1963). It was established that these residues may take multiple conformational forms, also referred to as ring puckers, in solution (Vijayalakshmi, Yathindra, & Rao, 1973).

By the early 1970s, simple computational investigations on structures related to GAGs were being reported. In 1975, studies on charge distribution, torsional potential and a steric energy mapping for 1→4-linked disaccharides of chondroitin were performed using CNDO/2 quantum mechanical treatment (Potenzone & Hopfinger, 1975). Likewise, conformational analysis of hyaluronic acid (HA) in either charged, uncharged or neutral states was carried out (Potenzone Jr & Hopfinger, 1978). This study enhanced the understanding of surface energy and chain dimensions, which agreed fairly well with the crystal structure of HA (Atkins & Sheehan, 1971). A year later, conformational analysis of heparin was performed by taking into account close contacts (Nagarajan & Rao, 1979), which updated the model of heparin from a stereo-chemical perspective.

While these studies were being put forward, Jeffrey and Taylor modeled 13 pyranose structures using molecular mechanics programs in existence at that time (Allinger & Chung, 1976), which correlated with their crystalline states (Jeffrey & Taylor, 1980). They developed MM1-CARB parameters that satisfactorily reproduced anomeric and exo-

anomeric effects. Yet, challenges in modeling possibly coupled with sub-optimal parameters led to a discrepancy between modeled β-maltose conformation in different solvents and the crystal structure (Tvaroška, 1982).

Beginning 1980s, computational modeling of GAGs received much boost owing to the development of molecular mechanics (MM) force fields, which advanced significant understanding of saccharide conformations. Many groups utilized hard-sphere exo-anomeric (HSEA) calculations by treating rings as rigid moieties and energy as a function of rotatable torsion (Bock et al., 1984; Lemieux, Bock, Delbaere, Koto, & Rao, 1980) and compared results to conformer populations derived from ^{1}H NMR spectroscopy. For example, Ragazzi et al. used MM methods to investigate uronic acid conformation and identified $^{1}C_{4}$,$^{2}S_{O}$, and $^{4}C_{1}$ forms of nearly equivalent energies (Ragazzi, Ferro, & Provasoli, 1986). Using the same technique, conformational analysis of seven different heparin mono- to penta- saccharides, including the antithrombin (AT)-binding sequence, was studied (Ferro et al., 1986). The ^{3}J coupling constants simulated using molecular geometries obtained from a MM method helped elucidate equilibrium between $^{1}C_{4}$ and $^{2}S_{O}$ forms of sulfated iduronic acid (IdoA) in heparin. These studies raised a key question whether the relative conformer populations, and the corresponding sulfate group orientations, could be correlated with specific biological activity. In 1987, a conformational model of a synthetic pentasaccharide, based on force-field calculations and NMR studies, was put together, which showed four major overall topologies of the high-affinity antithrombin-binding heparin pentasaccharide in solution. Advanced MM further refined the proposed model and elucidated the role of IdoA conformations (Ferro, Provasoli, & Ragazzi, 1992; Ferro et al., 1990; Ragazzi et al., 1990).

In 1993, Mulloy et al. studied the effect of different conformations of IdoA through molecular dynamic simulations (Forster & Mulloy, 1993). AMBER4 MM parameters were used; however, the studies did not involve sulfated sequences because parameters were not available then. No explicit conformational transition was observed for $^{1}C_{4}$ and $^{4}C_{1}$ puckers, whereas the $^{2}S_{O}$ form displayed chain-length dependent rate of transition from boat to twist-boat. Correlation with solution experiments showed the presence of 1:1 mixture of chair and boat forms in trisaccharides (Forster & Mulloy, 1993). This was used in the prediction of solution conformation of heparin using molecular modeling, monomer crystal structure and NMR. In this prediction a modified MM2 force field, developed earlier (Ragazzi et al., 1986), was used to derive conformational energy map for IdoA. The work led to deposition of the structure of heparin dodecasaccharide in two conformationally homogeneous forms ($^{1}C_{4}$ and $^{2}S_{O}$) (PDB ID:1HPN) (Mulloy, Forster, Jones, & Davies, 1993), which correlates fairly well with an earlier X-ray-diffraction structure (Nieduszynski & Atkins, 1973). Later, various chemically modified HP/HS sequences were studied and found to possess small variation in the inter-glycosidic conformations irrespective of substituent pattern (Barbara Mulloy et al., 1994). Likewise, the flexibility of pyranose ring in dermatan sulfate (DS) was also studied by imposing non-bonded contact and other restraints (Venkataraman, Sasisekharan, Cooney, Langer, & Sasisekharan, 1994). CHARMM minimization was used to fix the positions of endocyclic and exocyclic atoms belonging to different disaccharide repeats containing IdoA in various forms ($^{4}C_{1}$, $^{2}T_{O}$, $^{O}T_{2}$). Interestingly for CS, $^{O}T_{2}$ form of IdoA was found to possess the same non-bonded energy as $^{4}C_{1}$ and $^{2}T_{O}$ forms.

Development of sulfate (—OSO_3^-) and sulfamate (—$NHSO_3^-$) parameters

The earliest work on studying sulfated species is by Kaliannan et al., who studied the *N*-methyl sulfamate as a representative of the component of heparin through an *ab-initio* method (Kaliannan, Vishveshwara, & Rao, 1983). This work identified that the energy barrier around S—O bond was negligible, which implied nearly free rotation. Aside from this work, computational modeling of GAGs or GAG residues through the mid-1990s was performed without, or with inadequate, parameters for *O*-sulfonate and *N*-sulfamate groups. The first group of parameters came from quantum mechanical (QM) *ab initio* self-consistent Hartree-Fock calculations performed by several groups. Huige and Altona derived the potential energy function for the two key groups (*O*-sulfonate and *N*-sulfamate), located at 2-, 3-, and 6- positions of the glucosamine (GlcN) ring and incorporated them into AMBER and CHARMM force field packages (Huige & Altona, 1995). Gaussian 80 and Gamess-UK were used at 6–31G* basis set to optimize the structures of the two functional groups. The results showed that the two force fields reasonably reproduced *ab initio*-derived conformers and energy levels.

These studies further led to computations using advanced basis sets such as 6–31+G** on *N*-methyl sulfamate in ionic and neutral forms. The atomic charges, computed by the Merz-Kollman method for the NSO_3^- and OSO_3^- groups, were inserted in MM2-derived force-field (Ferro, Pumilia, Cassinari, & Ragazzi, 1995; Ragazzi & Ferro, 1997). This eventually led to the parameters for the two key groups for conformational analysis of sulfated polysaccharides, which correlated closely with X-ray structure (deviation of 0.21 Å). These studies accounted for explicit solvation and counterions (Ragazzi & Ferro, 1997).

Limitations of earlier modeling work on GAGs

The early force fields such as MM1 to MM3 (Allinger, 1976; Allinger, 1977; Allinger, Yuh, & Lii, 1989), modified MM (Jeffrey & Taylor, 1980), HSEA (Lemieux et al., 1980), and TRIPOS (Clark, Cramer, & Van Opdenbosch, 1989) were developed to understand the lowest energy conformers, for which parameters were derived from experimentally available data. Later, QM approaches were implemented in developing force field parameters (Rees, 1969; Vijayalakshmi et al., 1973; Zhdanov, Minkin, Minjaev, Zacharov, & Alexeev, 1973). It is instructive to note that pioneering work of several groups provided experimental data to better predict and refine theoretical models (Chakrabarti, 1977; Chakrabarti, Park, & Stevens, 1980; Cifonelli & King, 1973; Gatti, Casu, Hamer, & Perlin, 1979; Gatti, Casu, & Perlin, 1978; Heatley, Scott, Jeanloz, & Walker-Nasir, 1982). Yet, the limitations of these early approaches arose from a focus on certain types molecules and lack of computational resources to perform all-atom, exhaustive orbital calculations. Also, the experimental data available for appropriate treatment of explicit solvent and cations were still limited. Finally, most of the theoretical studies were validated through experimental test cases involving mono-, di- or tri-saccharides, and conformer populations of longer oligosaccharides was generally beyond reach. Finally, computational resources for *ab initio* calculations on longer GAG chains or sequences so as to study their interaction with proteins in the presence of solvent was difficult, time consuming and expensive.

Development of classical all-atom, force fields

For an efficient molecular dynamics (MD) simulation that surveys all possible conformations experienced by GAGs in aqueous conditions, a comprehensive force field is critically needed. Ideally, the force field should comprise of parameters for studying all possible building blocks of GAGs such as HP, HS, CS, DS, HA and others. Availability of such a force field would enable prediction of GAG sequences that have unique properties such as an unusual fold, specific recognition of a protein binding site, or interesting conformational dynamism with biological consequences.

Major classical force fields such as AMBER (Cornell et al., 1995; Weiner et al., 1984), CHARMM (Brooks et al., 1983), GROMOS (Schuler, Daura, & van Gunsteren, 2001) and OPLS-AA (Jorgensen, Maxwell, & Tirado-Rives, 1996) have been available to simulate proteins, DNA and small organic molecules for some time. Each of these force fields have their own method, solvent characteristics, and parameters. Over time these force fields have continued to refine parameters to more precisely reproduce experimental results (Monticelli & Tieleman, 2013).

Figure 2 shows the timeline of key advances made in the direction of GAG force field/parameter development and application. The earliest application of force fields for GAGs focused on elucidating conformational preferences of unsulfated GAGs including HA and unsulfated CS/DS in solution (Heatley, Brass, & Sheehan, 1998; Almond, Brass, & Sheehan, 1998; Furlan, La Penna, Perico, & Cesàro, 2005; Holmbeck, Petillo, & Lerner, 1994; Letardi, La Penna, Chiessi, Perico, & Cesàro, 2002) followed by validation of results using NMR or diffraction techniques. Since the parameters for sulfation were not developed, most computational studies involved smaller oligosaccharides (mono, di or tri saccharides) (Allinger et al., 1989; Brooks et al., 1983; Cornell et al., 1995). Following this *ab initio* methods enhanced the parameterization of force fields, which afforded elucidation of conformations of higher order oligosaccharides (Bayraktar, Akal, Sarper, & Varnali, 2004; Verli & Guimarães, 2004). As *O*-sulfonate and *N*-sulfamate parameters became available, conformational preferences of GAGs of varying substitution pattern were studied (Guvench et al., 2011; Kirschner et al., 2008; Pol-Fachin, Rusu, Verli, & Lins, 2012). These studies enabled advanced understanding of the torsional space, inter- and intra- molecular interactions that govern stability/flexibility of a GAG sequence, anomeric and exo- anomeric effect of IdoA residues (Pol-Fachin & Verli, 2008; Verli & Guimarães, 2004), while also enabling description of the conformational impact on neighboring saccharide units (Muñoz-García et al., 2012). Further recent development of force fields has enabled microsecond simulations of GAG sequences (Sattelle, Hansen, Gardiner, & Almond, 2010; Sattelle, Shakeri, & Almond, 2013).

Three force fields are most often used in GAG studies today. These include GLYCAM, CHARMM, and GROMOS (Guvench, Hatcher, Venable, Pastor, & MacKerell, 2009; Kirschner et al., 2008; Lins & Hunenberger, 2005). Much information about these force field parameters can be obtained from their webpages (see Supplementary Information). The reader is also directed to a recent review on carbohydrate force field development for additional information (Foley, Tessier, & Woods, 2012). Of the three, GLYCAM supports a

vast library of saccharide sequences, which includes reducing-end unsaturated GAG chains (Singh et al., 2016).

The GROMOS force field

GROMOS was proposed for hexopyranose-based carbohydrates in 2005 (Lins & Hunenberger, 2005). This parameter set is referred to as 45A4 and it was developed based on a new set of charge and torsions for hexopyranoses from QM calculations at 6–31G* level. The atomic partial charges were derived using restrained electrostatic potential fit (RESP). A series of MD simulations using GROMOS96 program for mono- and di- saccharides were performed using the newly derived parameters with simple point charge (SPC) water model. Restrained and unrestrained simulations showed $^4C_1$ as the stable ring conformer, which was known to be the dominant form in solution (Lins & Hunenberger, 2005).

Experiments with the early 45A4 parameters led to findings of limitations in conformational puckering (Autieri, Sega, Pederiva, & Guella, 2010). In 2012, the 45A4 potentials of GROMOS96 were revised with a new set of charges, atom types, torsional potentials, and nonstandard van der Waals scaling to deduce GROMOS $53A6_{GLYC}$ (Pol-Fachin et al., 2012), which was validated for 16 aldopyranose monomers and 8 disaccharides. Metadynamics and unbiased MD for these test cases predicted $^1C_4$ pucker for IdoA and $^4C_1$ for all other aldohexopyranose monosaccharides. These metadynamics ensembles were comparable with the NMR-based experimentally preferred conformations. Unrestrained MD studies revealed that the $^4C_1$ and $^1C_4$ puckers were stable, in appropriate equilibrium ratios, and devoid of any other puckers. Recently GROMOS $56A6_{CARBO}$ was developed (Plazinski, Lonardi, & Hünenberger, 2016).

The CHARMM force field

In 2009, CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses was proposed (Guvench et al., 2009). CHARMM incorporated all possible combination of chiral centers for pyranoses, including both α and β anomers. The QM calculations were performed using Gaussian 03 in MP2/6–31G(d) basis set (Guvench et al., 2009). Using the calculated parameters, MM calculations were performed for test cases of *O*-methyl-tetrahydropyran and inter-glycosidic–linked dimers of two tetrahydropyrans in transferable intermolecular potential with 3 points (TIP3P) water box. The optimized parameter sets gave results that were in good agreement with the experimental NMR observations. In 2010, the CHARMM parameter set was extended to furanose residues using a procedure similar to that used for hexopyranose parameter development (Raman, Guvench, & MacKerell, 2010). Once again, all possible, natural inter-glycosidic linkages of furanose residues were taken into account to optimize the parameter set. The CHARMM force field derived conformational samplings for model compounds were found to be in good agreement with the experimental results. In 2011, parameters to include deoxy, oxidized and *N*-methylamine monosaccharides in CHARMM (Guvench et al., 2011) were derived using Gaussian 03 QM calculations. For modeling carbohydrates in TIP3P water box, MM calculations utilized a potential energy function that was identical to the one used for other biomolecular applications including proteins, nucleic acid, lipids and small molecules.

Interestingly, the simulations showed a good overall agreement with J-coupling and NOE solution data.

Although *N*-methyamine monosaccharides were addressable by CHARMM, parameters for *N*-sulfamate group were not available. This parameter set has been recently been incorporated in CHARMM36 parameter files (Sarkar, Yu, Desai, MacKerell, & Mosier, 2016). The model compound with the *N*-sulfamate group was simulated through QM calculations using Gaussian 03 with geometry optimization. Vibrational calculations were carried out at MP2/6–31+G(d) level and followed by single point energy calculations. These force field parameters were successfully tested with heparin bound to antithrombin.

The GLYCAM force field/parameters

The first GLYCAM force field parameter set, named GLYCAM_93, was developed to perform wider biomolecular simulations, e.g., studying protein–glycan complexes (Woods, Dwek, Edge, & Fraser-Reid, 1995). A consistent carbohydrate parameter set, derived using geometry optimization by Gaussian 90 and 92 at Hartree-Fock (RHF) with 6–31 G* basis set, was made compatible to AMBER all-atom force field to enable simultaneous MM calculations on proteins and glycans. The results generated from conformational sampling of hexopyranoses predicted rotational properties of inter-glycosidic linkages belonging to residues carrying normal and epimerized C2 substituents. Additionally, GLYCAM_93 offered a platform to create and study new monosaccharides from the existing parameter set (Woods et al., 1995). Later, validation of ω-angle preferences in 1→6 *D*-gluco- and *D*-galactopyranosyl linkages were performed using QM and the new parameters, called GLYCAM2000, were deployed in AMBER MM force field (Kirschner & Woods, 2001). This paremeter set reproduced both rotamer population of conformers in solution and QM-based gas phase conformers.

It is important to note that GLYCAM2000 identified an important role of intermolecular water interactions in equilibrium population of different rotamers. Yet, the earlier parameters did not reproduce diffusion rates and radial pair distribution between hydroxyl group and TIP3P water in MD. The first and second hydration shells were not appropriately reproduced when compared with crystal structures (Jeffrey, 1994). To address this, GLYCAM06 was developed, which enabled transfer of parameters to all possible saccharide rings and glycan size (Kirschner et al., 2008). Additionally, no specific atom-types for alpha- or beta-anomers were necessary and the parameters could be extended to other biomolecules. GLYCAM06, derived using QM calculations on B3LYP/6–3111G(2d, 2p)//HF/6–31G*, were compatible to AMBER protein parameters (Kirschner et al., 2008). The partial charge derivations were similar to earlier versions but did not involve aliphatic hydrogen atoms. GLYCAM06's MM calculations utilized AMBER7/8 set with TIP3P water box. The results were better than that achieved earlier with conformational preferences and primary/secondary hydration shells similar to experimental observations. Further, the GLYCAM06 parameter set had well-established parameters for GAGs, except for inability to address unsaturated uronic acid monosaccharides (ΔUA). Recently, GLYCAM06 was updated with a more generalized parameter set to model *N*- and *O*-sulfation, ΔUA residues, neutral and protonated GlcN, IdoA and GlcA residues (Singh et al., 2016). MM calculations were

carried out using AMBER11 and TIP3P water box, and simulation predictions were comparable with experimental NMR scalar coupling and NOE measurements. Interestingly, conformer populations of ΔUA rings changed with sulfation. Likewise, sulfation of a tetrasaccharide sequence appeared to favor certain conformations with respect to *N*- and non-*N*- sulfated GlcN residues. These parameters are given in Table 1, while useful links are listed in Supplementary Information.

Applications of the three major force fields

Although the three carbohydrate force fields/parameter sets continue to be developed, several research groups have performed simulations using the current versions to address some important questions. These simulations have addressed varying lengths of glycans from the monosaccharides to the polysaccharides as well as helped design new sequences. Table 2 describes these studies in brief. Below we describe some key studies utilizing all-atom force field parameters to generate initial input structures. A comprehensive presentation of results follows this section.

Simulations involving GROMOS force field in GROMACS

In early 1998, Kaufmann and co-workers simulated HA dimer and trimer in water using GROMOS to understand the effect of hydration. A cubic solvation box with SPC water molecules was used. The initial coordinates for the GAG were taken from crystal structures and optimized in GROMOS to generate input parameter/topology and coordinates for MD simulation (Kaufmann, Möhle, Hofmann, & Arnold, 1998; Winter, Smith, & Arnott, 1975). Dynamic reorientation of water molecules around polar groups of HA and their average lifetimes were reported.

While it was easier to study HA because of lack of sulfate groups, sulfated GAGs presented challenges for a long time because of absence of direct parameters in GROMOS. Verli and co-workers first attempted to simulate heparin decasaccharide using this force field in a rather inexpensive manner (Verli & Guimarães, 2004). They derived the crude GAG topology with the PRODRG server using solution ^{1}H NMR structure (PDB ID: 1HPN) as input (Mulloy et al., 1993; Schuttelkopf & van Aalten, 2004). The crude topology was then modified to reference values of the S—N bond present in the sulfonamide group of GlcN residues. The atomic charges were then calculated using *ab initio* QM calculations using GAMESS at 3–21G level to derive partial charges including those for the inter-glycosidic linkage. Once the geometries were optimized, single point energy calculations at 6–31G** were performed to derive Lowdin atomic charges (Schmidt et al., 1993). The optimized parameters for heparin sequences gave average torsions of inter-glycosidic linkages in good agreement with experimental observations. In following years, the group studied the effect of different atomic charge calculations (e.g., Mulliken, Lowdin or electrostatic potential derived charges) and forces dictating IdoA2S conformational preferences ($^{2}S_{O}/^{1}C_{4}$) in heparin (Becker, Guimarães, & Verli, 2005; Pol-Fachin & Verli, 2008).

Gandhi and Mancera carried out a study to reproduce ring puckering of IdoA2S residue (Gandhi & Mancera, 2010). They compared GLYCAM and GROMOS parameters. To perform MD using GROMOS, they derived the $^{2}S_{O}$ conformation from the NMR structure,

generated an initial topology using the PRODRG server, and then optimized its geometry using HF/6–31G* basis set to compute partial charges. These parameters helped to perform MD in explicit water molecule (SPC/E) using GROMACS simulation software. The combination of these parameters showed transition from skew boat to chair in presence of water with no observation of chair to boat transition.

Cipla and co-workers used Accelrys Discovery studio 2.1 (Accelrys Discovery Studio 2.1. Accelrys: San Diego & http://www.accelrys.com.) to build the initial configuration of chondroitin sulfate C (CS-C, chondroitin-6-sulfate) dodecasaccharide. They adopted the methodology discussed above including initial generation of glycan topology on the PRODRG server using the all-atom GROMOS96 force field for two disaccharide building blocks (GlcA1→3GalNAc6S, GalNAc6S1→4GlcA) from which the polysaccharide topology was built manually. The resulting total charge was a non-integer value, which was set to the expected integer value. Preliminary simulations with these parameters resulted in a nonphysical behavior of sulfate groups including deformation of geometry and disorientation of sulfate groups. Following refinement of charges using QM, the electrostatic potentials improved and enabled simulation of the CS-C chain in SPC water using GROMOS96 (Cilpa, Hyvönen, Koivuniemi, & Riekkola, 2010; Schuler & Van Gunsteren, 2000).

Simulations involving the CHARMM force field in CHARMM

The CHARMM force field was extensively used in understanding the conformational preference of GAGs. In the early 1990s, Scoot et al., used CHARMM along with QUANTA template charges to perform MD simulations in vacuum and in water for HA and CS, respectively, to deduce secondary and tertiary structural characteristics (Brooks et al., 1983; John E. Scott, Chen, & Brass, 1992; Scott, Cummings, Brass, & Chen, 1991). Almond et al. have reported conformational preferences of HA using an advanced version 22 of CHARMM (Almond et al., 1998; Almond et al., 1998; Almond, Sheehan, & Brass, 1997). They compared conformational sampling of 150 ns for two CHARMM force fields and performed MD studies in TIP3P water box. The same methodology was also utilized to understand conformational preferences of two alternate HA tetrasaccharides, which carried the carboxyl groups in un-protonated states during the explicit water simulation. In a separate study, the group compared simulation of the HA tetrasaccharide to x-ray fiber diffraction data (Almond et al., 1998).

The x-ray fiber diffraction structure of CS-C led Weigel and Arnold to perform a MD study to understand interaction with water molecules. The initial coordinates of CS-C tetrasaccharide, taken from the structure reported earlier (Winter, Arnott, Isaac, & Atkins, 1978), were optimized using the CHARMM force field. QUANTA/CHARMM charges were assigned to atoms and input topologies were generated to perform a 4 ns MD in TIP3P water box (Wiegel, Kaufmann, & Arnold, 1999). Kaufmann et al. have also performed a 4 ns MD simulation of a CS-A tetrasaccharide (chondroitin-4-sulfate, (Kaufmann, Möhle, Hofmann, & Arnold, 1999), which included initial structure re-optimization using the sulfate group charges from Lamba et al. (Lamba et al., 1994).

To understand the solvent conformational preferences of CS and DS unsulfated tetrasaccharides, an explicit water MD study has been performed (Almond & Sheehan, 2000). This study analyzed long-lived intra-molecular hydrogen bonds (H-bonds) as well as handedness of the helical arrangement. Initial structures were derived from the PDB and CHARMM25 force field parameters with extension to carbohydrates, including explicit consideration of the exo-anomeric effect, were used to generate the input topology files. Partial charges were assigned in Gaussian 98 using the HF/6–31G* basis set for *N*-acetylated sugars and HF/6–31G** for carboxylated sugars. For DS simulations, all possible IdoA ring puckers were considered with a specific set of partial charges.

In recent years, the CHARMM36 carbohydrate force field has incorporated patches for sulfate groups at varying positions on the sugar backbone (see Supporting Information) (Guvench et al., 2009; Guvench et al., 2011; Mallajosyula, Guvench, Hatcher, & MacKerell, 2012; Raman et al., 2010). Based on these parameters, an adaptive biasing force (ABF) sampling of CS disaccharides was carried out using nanoscale molecular dynamics (NAMD) (Faller & Guvench, 2015). Since the earlier CHARMM force fields did not carry *N*-sulfamate parameters, this enhancement is expected to enable more effective applications of the recent versions of CHARMM (Sarkar et al., 2016). In fact, a recent application attempts to understand the flexibility of the core tetrasaccharide linker sequence connecting the GAG to the protein of the proteoglycan using CHARMM C36 force field in CHARMM (Ng, Nandha Premnath, & Guvench, 2017).

GLYCAM force field/parameter set using AMBER

More computational studies on sulfated GAGs have been carried out using the GLYCAM parameters in AMBER than any other force field. One of the reasons for this is the earlier widespread use of AMBER in MD simulations of peptides, proteins and small molecules (Salomon-Ferrer, Case, & Walker, 2013). For example, the AMBER biomolecular force field was used for non-sulfated versions of heparin to understand IdoA puckering (Forster & Mulloy, 1993). Recently, this was expanded to study conformational free energies of methyl α-*L*-IdoA and methyl β-*D*-GlcA using AMBER10 and GLYCAM06. The AMBER software modules link, edit and parm (LEaP) program was used to prepare the initial structure with the appropriate topology and parameters of GLYCAM06 and solvated in TIP3P water molecules (Babin & Sagui, 2010). As of now, structures of GAGs can be built using the GLYCAM web tool (http://glycam.org).

To understand the conformational influence of different water models (TIP3P, TIP4P, TIP4PEW and TIP5P) on chondroitin 4-sulfate (CS-A) octasaccharide, Nematu et al. used GLYCAM parameter set. They included new torsional parameters and partial charge for the —SO_3^- group using Restricted Electrostatic Potential (RESP) method and CHELPG algorithm up on monosaccharide 1-*O*-Me-β-GalNAc4S, as described in GLYCAM parameterization toolkit. The results showed that conformational preferences of β(1→3) linkage were different for different water models (A. Neamtu, B. Tamba, & X. Patras, 2013).

GLYCAM has been used to simulate the MD behavior of a library of heparin hexasaccharides by Munoz and co-workers. The topology and coordinates files were built with LEaP module of AMBER11, while GLYCAM_06h parameters were used to model the

heparin molecule including *O*-sulfonate and *N*-sulfamate groups. Here, the partial charge re-modification of *O*- and *N*- atoms bound to the —SO_3^- group was performed using the GLYCAM procedure so as to ensure an appropriate integral charge for the oligosaccharide, while parameters for TIP3P water and counter ions were used from AMBERff 12SB parameters (Muñoz-García, Corzana, de Paz, Angulo, & Nieto, 2013; Munoz-Garcia et al., 2013)

Another study has investigated the effect of ring conformational equilibrium of IdoA in a library of heparin-like trisaccharides carrying variation in sulfation positions (α*D*-GlcNS-(1→4)-α*L*-IdoA2S-(1→4)-α*D*-GlcN). The starting geometries of these sequences were developed from the 1HPN structure and the topologies were built using AMBER5 by employing a residual set of partial charges from a published work (Rodríguez-Carvajal, Imberty, & Pérez, 2003). AMBER 91 and GLYCAM 93 were used to generate the trisaccharides and Altona parameters of sulfates were appended for sulfate atoms (Huige & Altona, 1995). They performed independent simulations for IdoA2S both in $^1C_4$ and ^2S_O conformations in TIP3P water molecule (Muñoz-García et al., 2012).

Samsonov et al. (2014) analyzed GAG monosaccharides using three different techniques, namely QM, NMR and MD, to compare and contrast results. For MD simulations, they generated initial configurations using the (Molecular Operating Environment (MOE)) (https://www.chemcomp.com) and GLYCAM06 parameters implemented in AMBER11 package. The atomic charges for sulfate groups were derived using RESP calculations at 631(d)G levels for methylsulfate. Simulations for a hexasaccharide sequence were also performed and compared with monosaccharide results (Samsonov, Theisgen, Riemer, Huster, & Pisabarro, 2014; Vanquelef et al., 2011). Finally, the validation of new GLYCAM06 parameter set (discussed above) included variably sulfated GAG di- and tri-saccharides containing terminal ΔUA residue, in which the initial structures were built using the LEaP module of AMBER12 and TIP3P water molecule for solvation (Singh et al., 2016). These newly developed parameters have been employed for conformational study of eight HS hexasaccharides containing variable sulfated GlcN residues (2-*O*-, 3-*O*- and 6-*O*-substitution) to elucidate influence on neighboring IdoA and IdoA2S residues (Hsieh, Thieker, Guerrini, Woods, & Liu, 2016).

Simulations performed in other ways

A variety of open source MD software tools available for academic users such as NAMD, GROMACS and ACEMD (Harvey, Giupponi, & Fabritiis, 2009; B. Hess, Kutzner, van der Spoel, & Lindahl, 2008; Phillips et al., 2005) could also be used to perform simulations of GAGs. For example, GLYCAM06-generated topology of α-*L*-IdoA2S-OMe, provided by the Sattelle group, was converted into the GROMOS format using amb2gmx.pl script for simulations in GROMACS 4.0.7 (Oborský, Tvaroška, Králová, & Spiwok, 2013; Sattelle et al., 2010; Sorin & Pande, 2005). Similarly, parameters of β-*D*-GlcN, obtained from the GLYCAM website, were converted to the GROMOS format using the above approach (Spiwok, Králová et al. 2010). Here we highlight some studies that employed similar approaches.

Settelle and Almond (Sattelle & Almond, 2011) describe the preparation of unsulfated and sulfated variants of GlcN residue (with and without *N*-acetylation) using LEaP tool in AMBER followed by use of in-house developed coordinates for $^4C_1/^1C_4$ puckers in GLYCAM06g force field (Kirschner et al., 2008), which was implemented in ACEMD software to perform MD simulations (Harvey et al., 2009). It is important to note that at the time of this study, parameters for the O—S—O angle as well as C2—O2—S—O torsional angle of the sulfated residues were not available. These parameters were adopted from the general AMBER force field (GAFF, see http://ambermd.org/tutorials/basic/tutorial4b/), whereas partial charges for the sulfonated residues were computed using Gaussian 03 with consistent basis set (Frisch MJ; Wang, Wolf, Caldwell, Kollman, & Case, 2004). The same group used similar protocol to study the free energy landscape of IdoA and related monosaccharides in explicit water (Sattelle et al., 2010).

Murphy et al. (2008) generated initial parameter and topology files for IdoA and IdoA2S in the $^1C_4$ forms based on the geometries reported in the PDB using the antechamber module of AMBER tools. AM1-BCC charges and GAFF atom types were assigned and coordinates for the ^2S_O and $^4C_1$ forms were derived from a short 5 ns Generalized Born (GB) implicit solvent simulation of $^1C_4$ conformation in water (Murphy, McLay, & Pye, 2008).

Simulations of interactions between oligosaccharides and water molecules were performed by Almond (2005). These simulations used the GLYCAM93 parameters in CHARMM MD module on multiple HA and CS di- and oligo- saccharides. Likewise, Arixtra, a clinically used heparin pentasaccharide, was simulated using structure from the PDB (2GD4), initial coordinates from xLEaP module of AMBER Tools 1.5 and GLYCAM06g parameters. The GlcNS residue was derived from the standard GlcNAc residue of GLYCAM by removing the acetyl group and adjusting the partial charge on the nitrogen atom to give a formal charge of −1 to the *N*-sulfamate group. Following addition of Na^+ ions using AMBERff99SB force field and solvation using TIP3P box (Hornak et al., 2006; Kirschner et al., 2008), MD simulation was performed using NAMD 2.8 software (Phillips et al., 2005). In this study, the van der Waals and electrostatic 1–4 scaling factors (SCNB and SCEE, respectively) were set to 1 to be consistent with GLYCAM06 parameterization (Langeslay et al., 2012). Similar protocol was also utilized by Beecher et al. (2014) in the study of this heparin pentasaccharide(Beecher, Young, Langeslay, Mueller, & Larive, 2014). Finally, the Boltzmann jump method derived input CS tetrasaccharides to perform MD in Cerius 2 package using the Dreiding force field for sulfate groups, which were optimized using the Jaguar QM treatment (Bochevarov et al., 2013; Gama et al., 2006; Greeley et al., 1994; Kian-Tat Lim et al., 1997).

How to Simulate a GAG – An overview

The tools developed so far for simulating GAGs enable beginners and non-computational experts to perform all-atom MD simulations with relative ease. A generic flow of an effective simulation protocol is depicted in Figure 3. Once a desired GAG of defined chain length is selected, steps A) to D) can be undertaken to elucidate its structural and dynamical characteristics. These steps are not restricted to a particular force field or MD program, although a few suggestions are included.

A) Generation of the initial GAG structure—A three-dimensional (3D) structure of the chosen GAG sequence is needed to perform an MD simulation (Figure 3A). Several coordinates are available in the structural databases such as the Protein Data Bank (https://www.rcsb.org), the Cambridge Structural Database (https://www.ccdc.cam.ac.uk) and others (Table 3). If the GAG sequence is not available in a database, structure-building tools are available (see Table 3), which can help build a sequence from scratch. Some of the tools, e.g., GLYCAMWeb, CHARMMGUI, and PRODRG server, also provide the initial structure in a form compatible to commonly used force fields. Finally, the Desai laboratory has developed an extensive library of GAG sequences (HP, HS, and CS) ranging in length from disaccharide to hexasaccharide using SYBYL scripts for studying their interaction with target proteins using a dual-filter screening algorithm (Sankaranarayanan Nehru et al., 2017; Sankaranarayanan & Desai, 2014). These GAG sequences are also available upon request.

A few checks and balances have to be performed to ensure that the initial structures are appropriate for further studies. The initial 3D structure should be checked for missing atoms, ring conformation, sulfate position, hydrogens, charges and energy minimized to remove steric hindrances. It is important to note here that the GLYCAM web server includes all possible combinations of sulfation and puckering for monomers for each GAG type. A user can generate the desired heterogeneity and output the structural coordinates following minimization performed in presence of counter ions and water (e.g., TIP3P model) using the AMBER–GLYCAM06 force field, which offers special advantages for studying GAGs.

B) Generating input parameters and coordinates—The next step is to generate the parameter/topology and coordinate files to perform an MD simulation using a force field. The minimized structure file of a GAG sequence from step A) should consist of atom name and type matching that of the force field to be used for MD simulation. The format of this information required by the force field can be reviewed from its documentation pages (See ‘Understanding Parameter Topology Files’ section in Table 3). An example of non-sulfated residues of IdoA and GlcNAc from the GLYCAM06 library is shown in Supporting Information (see Figure S1). Following steps 3 and 4 in Figure 3B, the input initial structures are called upon by the respective programs (either Leap, X-leap, tleap, VMD, PRODRG, ATB, ACPYPE, CGenFF, Topolbuild, TopolGen, or pdb2gmx). An example of heparin disaccharide (GlcNS6S($^4C_1$)-1→4-IdoA2S(^2S_O)) built using X-leap with GLYCAM06 library parameters is shown in step 5 of Figure 3B.

The next step 6 is to check the total charge of the sequence, which should be an integer and balanced by a counter ion (e.g., Na^+) (Figure 3B). This is then followed by solvation of the GAG sequence in water of desired shape (e.g., cubic, rectangle, truncated octahedral, etc.) The solvation water models used in the literature are listed in Table 3. An example of a TIP3P water box solvating a heparin disaccharide sequence generated using GLYCAM06–AMBER14 is shown in step 7 of Figure 3B. The 3D water box is made to fully immerse the GAG sequence and counter ions in water by imposing the periodic boundary condition. It is important to note that the concentration of Na^+ in the water box could be made to reach any desired value, e.g., 0 to 145 mM, by adjusting the size of the water box or the number of water molecules. Literature reports that counter ions could be added manually, closest to the negatively charged groups, by removing water molecules in their vicinity (Verli &

Guimarães, 2004). Save the initial coordinate and parameter file to perform further steps as shown in step 8 of Figure3B.

C) Minimization, equilibration and MD simulation—The initial structure generated above is further minimized to remove close contacts (step 9 in Figure 3C), which performed in minimally two steps. The solute molecules, i.e., the GAG and counter ions, are fully constrained first and the water molecules relaxed to the desired number of iterations (e.g., 2000), followed by relaxation of the entire system without constraints to reach a minimum energy state. Literature reports a three-step minimization protocol too, which includes a step of restraining sugar ring atoms (Hsieh et al., 2016).

The minimization and equilibration steps are fraught with inconsistencies if performed in an inappropriate manner. Equilibration is performed while weakly restraining the solute to avoid unwanted fluctuations in the system. The system is heated in small steps from 0 K to the desired temperature using a thermostat (e.g., Berendsen/Langevin) at constant volume (NVT) over several picoseconds (e.g., 60 ps) (step 10 of Figure 3C), which is followed by equilibration at a constant pressure (usually 1 atm (or NPT)) using a barostat for ~1 ns (step 11 of Figure 3C). The literature reports several algorithms, e.g., SHAKE, LINCS, SETTLE, or RATTLE, have been used to constrain the hydrogen-containing bonds (Andersen, 1983; Berk Hess, Bekker, Berendsen, & Fraaije, 1997; Miyamoto & Kollman, 1992; van Gunsteren & Berendsen, 1977). Periodic boundary conditions have been employed throughout these simulations. Long-range electrostatic interactions have been treated using the Particle-Mesh Ewald algorithm (Darden, York, & Pedersen, 1993), whereas cutoffs for non-bonded interaction have been varied from 8 to 12 Å.

Following completion of equilibration, the system is taken up for a MD “production” run, which implies data generation at a desired step size (e.g., 1 to 4 fs/step) for the desired length of time (e.g., few ns to ~μs, step 12 of Figure 3C). The evolution trajectories (i.e., data) are collected at a constant step length (e.g., every 1 to 10 ps) for further analysis (step 13 −16 of Figure 3D). It is important to note that during the MD simulation it may be necessary to constrain a ring to a specific pucker, e.g., IdoA to either $^{1}C_{4}$ or $^{2}S_{O}$, so as to avoid a non-sampled pucker, e.g., $^{4}C_{1}$ for IdoA. Such restraints are imposed using a flat-well parabolic energy potential (Hsieh et al., 2016; B. Nagarajan, Sankaranarayanan, Patel, & Desai, 2017; Verli & Guimarães, 2004). Links to the procedures described above are included in the Supporting Information file. In addition, the best practices to perform MD simulations are also included in Supporting Information.

Analysis and Results of MD trajectories

Deductions on inter-glycosidic torsions, torsional space and flexibility

General—: The primary focus of studying GAGs in MD is to understand the overall solution conformational behavior including local structure and dynamical properties. The overall conformational behavior is a sum of the forces acting upon individual residues/ groups, constraints imposed by the saccharide ring (or lack thereof), and the torsions sampled by each inter-glycosidic bond present in the sequence. The overall conformational motions of a GAG arise from dynamism of the inter-glycosidic torsional angles, as shown in

Figures 4 and 5. Therefore, the easiest parameters that can be followed in an MD simulation are the inter-glycosidic torsion angles and their correlation with experimentally derived values from X-ray or NMR techniques. An advantage with MD is that it offers all torsions, i.e., Φ & Ψ for each inter-glycosidic bond, from an equilibrated simulation.

It is important to note that authors have used two definitions of Φ & Ψ in the literature, either an IUPAC or a non-IUPAC definition. The IUPAC definition relies on using heavy atoms, i.e., Φ=O5—C1—O1—C3 and Ψ=C1—O1—C3—C4, while the non-IUPAC definition is NMR-user friendly and describes Φ as H1—C1—O1—C3 and Ψ as C1—O1—C3—H3 (see Supplementary Figure S2). The literature has presented many ways to display this torsional space. For example, linear plots of time-dependent fluctuations in Φ & Ψ (Verli & Guimarães, 2004), or 2D contour plots (Almond & Sheehan, 2000; Almond et al., 1997; Cros, Petitou, Sizun, Pérez, & Imberty, 1997), probability distributions (Muñoz-García et al., 2013; Muñoz-García et al., 2012), free energy plots (Neamtu et al., 2013), adiabatic maps (Blanchard et al., 2007; Rodríguez-Carvajal et al., 2003), or heat maps (Hsieh et al., 2016).

In general, the Φ & Ψ values are extracted from each frame of the recorded trajectory, using the above definitions. The occurrence of Φ & Ψ pairs is calculated based on specific intervals, say by dividing them in 6° interval bins. Once the simulations have enough equilibrium conformers, the probability of the conformers in particular Φ & Ψ bins is calculated using the total number of frames. If represented in this format, it is called probability distribution. The energy of each (Φ, Ψ) geometry is calculated by relating its probability to Boltzmann energy (Mandal, 1988). Using such energy values, a contour map of Φ,Ψ is reconstructed to locate the different energy regions.

Application to HA sequences—: Analysis of Φ & Ψ have offered insight into many aspects of GAG structure and biophysical properties. One of the earliest applications of MD in this direction related to studying HA and torsional space of its Φ & Ψ angles (Almond et al., 1997; Holmbeck et al., 1994). Following MD, Φ & Ψ (non-IUPAC definition) were extracted from each frame, binned every 6°, and converted into bin probabilities using the total population of torsional angles, which were then used to calculate Boltzmann energies of each bins. These energies were plotted as contour maps for the ensample of conformers across the MD run. Figure S3 shows one such map for the β1→4 linkage of HA in vacuum and in water. The map reveals much lower conformational spreading in vacuum as compared to that in water, which highlights the role of solvent in narrowing the energy funnel. Analysis of longer sequences showed that an internal disaccharide exhibits reduced flexibility with Φ & Ψ of 50° & 0°, respectively, which compared favorably to values of 60° and 0° from another study (Almond et al., 1998; Almond et al., 1998). In fact, the terminal residues showed multiple allowed regions in the contour map. These results also compared favorably with Φ, Ψ data from an x-ray fiber diffraction study (Figures 4A & 4B).

Application to CS sequences—: A conformational analysis study for CS showed a behavior similar to HA, especially for the β1→4-linked disaccharide (Φ=50°; Ψ=0°) (Almond & Sheehan, 2000). It is important to note that CS possesses two types of inter-glycosidic linkages as compared to HA, which is uniformly β1→4-linked. Thus, the β1→3

linkage of CS was expected to deviate from that of HA and displayed Φ & Ψ of 50° and −30°, respectively (Figures 4C & 4D). In comparison, DS displayed conformational spread of torsions similar to HA and CS but positioned at entirely different regions for both β1→4 (Φ=50°, Ψ=45°) and β1→3 (Φ=45°, Ψ=−45°). This arises from the presence of IdoA as well as GlcA in DS, while HA and CS predominantly carry the GlcA residue in the $^4C_1$ form. Studies have also explored the conformational space with different forms ($^1C_4$, ^2S_O, etc.) of IdoA in solution and shown the ^2S_O form DS displays properties similar to HA, except for a deviation in β1→4 Ψ of 30° (Almond & Sheehan, 2000).

Another earlier study on conformational sampling of CS-A explored the torsional map based on IUPAC definition of torsions (Kaufmann et al., 1999). Interestingly, this study captured two energy minima for β1→3 linkage (Φ,Ψ = −10°,−85° and Φ,Ψ = 80°,90°) in water; however, the latter minima was found to be short lived (200 ps) and rapidly switched to the more stable former minima. In contrast, the β1→4 linkage was stable with a broad minimum centered around Φ,Ψ = −10°,−70°. The torsional flexibility observed in this study did not correlate with the X-ray results (Winter et al., 1978) but matched with earlier computational study (Zsiška & Meyer, 1993). A recent study has shown that conformational preferences of the β1→4 linkage of CS-A does not depend on the types of water models using in simulation, which is an interesting conclusion different from studies on other GAGs. In this study, β1→4 linkage exhibited a unique basin with mean values of Φ and Ψ at 289±2.5° and 242±5.5°, respectively. These observations were in agreement with NMR-based data (Blanchard et al., 2007; Sattelle & Almond, 2010; Yu, Wolff, Amster, & Prestegard, 2007) but deviated from X-ray fiber diffraction data (J. J. Cael, Winter, & Arnott, 1978; Winter et al., 1978), perhaps due to crystal packing forces. In contrast, the β1→3 linkage also showed a major, specific preference for Φ and Ψ (288±1.5° and 137±4°). Here, Φ is comparable with that derived from NMR while Ψ showed a deviation of 30°, but also displayed two minor conformational basins at Φ and Ψ of 209° and 92°, and 56° and 116°, respectively, with TIP3P water model.

In 2010, a comparative study on multiple CS saccharides using MD and semi-empirical methods showed that GLYCAM06 simulations showed better correlation with experimental results in predicting backbone torsional space (Sattelle & Almond, 2010). Use of an improved electrostatic potential function on CS-C resulted in a torsional probability density profile resembling a single uniform Gaussian distribution for the β1→4 linkage with Φ and Ψ centered at −71° and −109°, respectively. For the β1→3 linkage, the distribution was much broader with Φ and Ψ center around −85° and 110°, respectively (Cilpa et al., 2010). These results were in agreement with the earlier experimental observations (J. J. Cael et al., 1978; Michel et al., 2004; Yu et al., 2007). Researchers have also showed similar observations using GRID search method using MM3 FF and obtained adiabatic maps (Blanchard et al., 2007; Rodríguez-Carvajal et al., 2003).

Application to HP/HS sequences—: MD simulations of a heparin decasaccharide ([IdoA2S-GlcNS6S]$_5$), which represents the most common heparin sequence, have shown time dependent torsional angles to be relatively stable around their average values (Φ,Ψ=96.6±14.1, −125.0±15.1 (β1→4) and −75.0±10.7, −116±10.6 (α1→4), respectively (Verli & Guimarães, 2004). Comparison of the dynamical behavior of the IdoA-GlcN

torsional angles with those of the GlcN-IdoA showed reduced flexibility of the former. The mean torsions observed in MD runs were comparable to NMR-derived values (Mulloy et al., 1993).

Another report studied two heparin trisaccharides carrying variation in internal IdoA2S residue, so as to impact their $^{1}C_{4}$ and $^{2}S_{O}$ populations, which suggested considerable flexibility around $^{2}S_{O}$ energy minimum in comparison to that for the $^{1}C_{4}$ dominant conformer (Figures 4F & 4E)(Pol-Fachin & Verli, 2008). It is possible that this inherent broad minima in the Φ,Ψ space affords the $^{2}S_{O}$ form to fit into many protein binding pockets and induce biological activity. The observation also showed that the IdoA2S ($^{2}S_{O}$)–GlcNS6S torsional space occupied more space when compared to IdoA2S in $^{1}C_{4}$ and also possessed four distinct representative conformers (Figure 4H). On evaluation of different combination of disaccharide pairs with the trisaccharide showed that there was an inter-conversion path of torsion angle pairs. The disaccharide pair GlcNAc-IdoA2S in $^{1}C_{4}$ showed a much narrower region of torsional space (Figure 4G).

Muñoz-García et al. carried out an analysis of global conformations for a library of trisaccharides and plotted the distribution frequencies of their inter-glycosidic torsions (Φ and Ψ) using a bin size of 10° (Muñoz-García et al., 2013; Muñoz-García et al., 2012). Figures 5A & 5B shows their distribution with respect to GlcN-IdoA2S and IdoA2S-GlcN linkages for all eight sequences. For the GlcN-IdoA2S linkages Φ and Ψ were centered at −45° and −30°, respectively, which Ψ distribution being wider in comparison to Φ. For IdoA2S-GlcN linkages, a sharp narrow region was observed for Φ centered at 40°. A broad distribution with two peaks of maximum value at −30° and 15° was observed for Ψ (Figure 6B). These observations were similar to an earlier unrestrained MD with some small variations.

Validation studies performed with new GLYCAM06 parameter set, with three ΔUA-containing disaccharides carrying variation in sulfation, showed a high density of structures for Φ and Ψ at 50° and 0°, respectively, irrespective of sulfation of GlcN residue (Figures 5C & 5D). The work also showed an anti Ψ state near Φ = 50° and Ψ = 180°. The same report presented analysis of two synthetic sequences of tetrasaccharides using heat maps based on the MD trajectories of each disaccharide pair containing IdoA2S (Figures 5H and 5I). The results indicated that inter-glycosidic torsions were not significantly dependent on IdoA puckers. Likewise, there was only a small effect of *N*–sulfation on the inter-glycosidic preferences irrespective of puckering (Singh et al., 2016).

Deductions on the intra-molecular hydrogen bond interactions

General—: Hydrogen bonding occurs between and within GAG residues generating secondary structures such as two-fold and three-fold helices as well as chiral handedness of chains (Almond & Sheehan, 2000). Deduction of a H-bond is typically carried out by measuring the distance between the donor (Dn) and acceptor (Ac) atoms, which should be less than 3.5 Å, and the angle between Dn—H•••Ac atoms, which should be 180±60° (Supplementary Figure S4). The presence of such a bond is deduced in an automated manner for every structural frame of a trajectory. Generally, the lifetime and/or overall occupancy percentage (frames with H-bond/total number of MD frames) of a hydrogen bond is

calculated for comparative analysis. An overall occupancy of >50% is referred to as a persistent H-bond. At the same time, it is important to recognize a H-bond may be cooperative (one donor, multiple simultaneous acceptors) or mutually exclusive (one donor, one acceptor). Tools and links for such an analysis are given in the Supplementary Information section. Also Supplementary Table S1 lists the observed H-bond in MD studies of HA, CS and DS sequences.

Application to HA structures—: The Almond group has analyzed intra-molecular H-bond formation for a range of HA di- and tetra- saccharides (Almond et al., 1998a; Almond et al., 1998b; Almond & Sheehan, 2000; Almond et al., 1997). In tetrasaccharides, >30 distinct H-bonds were identified in a short MD simulation (500 ps), of which about 10 intra-molecular H-bonds persisted more 50% of the total time (Almond et al., 1998). The authors noted that H-bond formation was more favored in solvent than in vacuum (Figures 6A & 6B). For HA, the strongest H-bonds were formed between oxygen atoms of adjacent residues, e.g., O5 of GlcNAc to O3 of GlcA or O5 of GlcA to O4 of GalNAc (Figure 6C).

Application to CS structures—: Almond and co-workers also studied unsulfated chondroitin tetrasaccharides, which showed results similar to HA oligosaccharides as far as β(1→4) linkage is concerned, while differences were noted for H-bond interactions for the β(1→3)-linked residues (Figures 6D through 6F). As expected, the variation in H-bonding impacted flexibility or rigidity of torsional angles. Further, the persistence of H-bonds, or lack thereof, impacts the formation of secondary helical folds, which supports the multiple helical folds and secondary structures observed in crystals (Almond et al., 1998; Almond & Sheehan, 2000).

An earlier study on CS-A identified a very strong intra-molecular H-bond (>80% occupancy) between acetamido hydrogen (NHAc) and oxygen atom (O6) of the carboxylate group (Figure 6G), (Kaufmann et al., 1999). A later study of H-bonding has shown that the conformation of β(1→4) linkage is stabilized by two intra-molecular H-bonds (E.D. Atkins, Meader, & Scott, 1980; J. J. Cael et al., 1978). A H-bond of particular interest – between HO3 of GlcA and the ring oxygen O5 of GalNAc4S – because of its high lifetime was found to exhibit dependence on the model of water used in simulations (e.g., TIP4P, TIP4P-EW or TIP5P). Yet, the authors conclude that the choice of different water models influence the β(1→4) linkage much less than the β(1→3) linkage (Andrei Neamtu, Bogdan Tamba, & Xenia Patras, 2013).

The two types of inter-glycosidic linkages present in CS offer an interesting opportunity to compare local dynamics and inter-molecular forces. In the minimum energy conformation, β(1→3) linkage of CS-A displays a H-bond between the HO2 group of GlcA to carbonyl oxygen (O2) of the GalNAc4S residue (Figure 6H), which when absent reduces the energy contour basin. To be specific, the other direct H-bond between HO2 of GlcA and SO_3^- group of GalNAc is the wholly responsible for the structures to fall in the least energy basin.

Application to DS structures—: MD results for DS sequences carrying IdoA residues in three conformations ($^1C_4$, ^2S_O, and $^4C_1$) have also been analyzed for H-bond patterns (Almond & Sheehan, 2000). Unlike HA and CS structures in β(1–4), the $^4C_1$ form of IdoA

prevented establishment of simultaneous H-bonds between NHAc and COO^- group and between OH3 of IdoA to ring oxygen of GalNAc residue at the same time. The only strong interaction observed was from the carboxyl group to acetamido group of the neighboring residues. When the IdoA switches to ^2S_O conformer for the β(1→4) linkage, three H-bonds are invoked including i) IdoA OH3 and GlcNAc O5 (ring oxygen), ii) GlcNHAc acetamido to IdoA carboxyl (O6) and iii) CH_2OH of GalNAc and OH3 of IdoA. Of these, however the H-bond iii) is preferred over ii) (Figures 6K and 6L, (Almond & Sheehan, 2000).

A comparative study of HA, chondroitin and CS-A disaccharides using GLYCAM06 and other two semi-empirical methods showed similar intra-molecular H-bond interactions, as discussed above (Sattelle & Almond, 2010). A study by Cilpa et al. in understanding the atomistic insights of CS-C showed the persistence of intra-molecular H-bonds as a function of time during the simulation. These results were comparable to earlier results with each β(1→4) and β(1→3) having two main intra-molecular H-bonds (Figures 6I and 6J). This study identified the variation in strengths of H-bonds across different GAGs. In fact, the H-bond between HO6 of GalNAc and O3 of GlcA was found to be disfavored owing to the epimerization in HA, while it is favored for chondroitin (Almond & Sheehan, 2000; Almond et al., 1997). An interesting observation from this study was the identification of an additional H-bond between sulfate of $6SO_3$ – GalNAc and O3 of GlcA, which not identified in earlier QM studies but arises from side chain flexibility (Cilpa et al., 2010).

Application to HP/HS structures—: A distinctive feature of the HP/HS structures is the complexity arising from the presence of IdoA, which exhibits considerable flexibility. Pol-Fachin and Verli carried out a 0.2 μs simulation to identify the interactions responsible for IdoA2S conformational equilibrium in solution (Pol-Fachin & Verli, 2008). They selected two trisaccharide sequences including GlcNS6S-IdoA2S-GlcNS6S and GlcNAc6S-IdoA2S-GlcNS carrying the IdoAS residue in both $^1C_4$ and ^2S_O forms. The analysis revealed a range of two to four H-bonds with varying strengths, of which the H-bond involving the inter-glycosidic oxygen of the α and β(1–4) linkage was common for both forms. For the GlcNS6S-IdoA2S linkage, the ^2S_O or $^1C_4$ geometry of IdoA residue determined H-bond formation (Figures 7A & 7B). Between the two forms of IdoA, a reduction of ~8% in intra-molecular H-bond interactions was observed for the ^2S_O form. Interestingly, a H-bond common to both forms was observed for the IdoA2S-GlcNS6S linkage between HO3 of GlcN and ring oxygen (O5) of IdoA. Yet, MD studies revealed that the strength of even consistent H-bonds may be different for as judged by distributions of angle and distance (Pol-Fachin & Verli, 2008). A comparison of H-bond interactions for sequence GlcNAc6S-IdoA2S-GlcNS to its fully sulfated counterpart showed that acetylation did not affect the H-bond interaction between O2 of IdoA2S and HN2 of GlcNAc for the $^1C_4$ form. But acetylation impacted the ^2S_O form by forming H-bond between HO3 and HH2 of GlcNAc (Figures 7C & 7D, (Pol-Fachin & Verli, 2008).

The first simulation on a heparin decasaccharide by Verli and group showed variations in Φ and Ψ, which were well correlated with the intra-molecular distances (Figures 7E & 7F). For IdoA2S in ^2S_O conformation, but not in $^1C_4$ conformation, the distance between OH of GlcN and COO^- of IdoA correlated well with results from NMR (Verli & Guimarães, 2004).

The intra-molecular H-bonding pattern of Arixtra, the heparin pentasaccharide that binds to antithrombin with high affinity, is of particular interest because its therapeutic value. Two explicit MD studies have been performed on this pentasaccharide (Beecher et al., 2014; Langeslay et al., 2012). Figures 7G and 7H show representative structures from these simulations with IdoA2S in $^{2}S_{O}$ and $^{1}C_{4}$, respectively. The results showed an interesting intra-residue H-bond for the central GlcN residue between NH and 3-*O*-sulfonate groups. The occupancy of this intra-residue H-bond was 89% when IdoA2S residue was in the $^{2}S_{O}$ conformation but only 64% for the $^{1}C_{4}$ conformer (Remko & von der Lieth, 2006). In fact, this characteristic H-bond in $^{2}S_{O}$ conformer significantly stabilized the overall Arixtra conformation so as to better fit into its site of binding on AT (HricovÍNi et al., 2001; M et al., 2008). In addition to this H-bond, analysis of the MD runs also showed other stabilizing inter-residue H-bonds between residues D and E, residues F and G, and residues G and H (see Figures 7G and 7H).

Studies on a library of 16 heparin trisaccharides reveal similar insights (Muñoz-García et al., 2013; Muñoz-García et al., 2012). For these sequences, a H-bond between IdoA2S's ring oxygen with HO3 group of adjacent GlcN occurred approximately 57% of the time and possessed an average life time of 2.6 ps. Further, this particular H-bond is not altered by conformational variation of IdoA2S residues as well as by variation in sulfation patterns between the interacting residues. These results are in line with the general conclusions of several studies that H-bonds are formed between adjacent residues and these interactions help stabilize the overall 3D geometry of GAGs. A recent study also derived conclusions similar to these (Hsieh et al., 2016). This study relied on MD studies with eight hexasaccharides variation in sulfation of GlcN and IdoA residue. Interestingly, an intra-residue H-bond between 2-*O*-sulfonate and HO3 of IdoA2S, which stabilizes the skew boat form, was identified in this study. The $^{1}C_{4}$ form did not display this intra-residue H-bond (see Figures 7I – 7K; https://www.youtube.com/watch?v=aQYLXtKSIvs).

Understanding GAG and solvent interactions—The hydrophilic and charged nature of GAGs encourages extensive interactions with water molecules. GAG–water interactions are closely associated with viscoelastic properties of these natural polymers, which play important role in human biology (Lujan, Underwood, Jacobs, & Weiss, 2009). These arise from the conformational flexibility of GAG sequences and the caging of GAG's polar atoms (e.g., acetamido, carboxyl, hydroxyl and sulfate groups) with water molecules (Isaac & Atkins, 1973). In fact, the inter-molecular GAG–water interactions help stabilize certain secondary structures (e.g., two-fold helices), especially when bridging neighboring monomers (Yuan, & Andy, 2005; Millane, Mitra, & Arnott, 1983). Studies using X-ray, NMR and other techniques have shown a strong influence of water on their GAG structure (Atkins, Phelps, & Sheehan, 1972; Cael et al., 1976; F. Heatley & Scott, 1988; Isaac & Atkins, 1973; Joshi & Topp, 1992; Scott et al., 1991).

Recent computational simulations have shown that bridging waters may alter H-bonding network between neighboring residues (Sattelle, Shakeri, Cliff, & Almond, 2015). Solvent interaction appears to be different for different sequences and this may play a role in protein recognition (Muñoz-García et al., 2012; Verli & Guimarães, 2004). In fact, a number of computational studies have predicted the importance of water molecules in GAG recognition

of basic binding sites on proteins, which have been supported by experimental results (Jana & Bandyopadhyay, 2012; Samsonov, Teyra, & Pisabarro, 2011; Sepuru, Nagarajan, Desai, & Rajarathnam, 2016). These and other studies suggest that inter-molecular GAG–water interactions are important to gain insight into GAG recognition of proteins.

Pair distribution function—: For a GAG sequence at thermodynamic equilibrium, the average distribution of water molecules around its constituent atoms is called a pair distribution function. Figure 8A shows a distribution of water molecules around a heparin disaccharide. The distribution shows some water molecules closest (within 5Å) to GAG sequence (black dotted circles enclosing blue spheres), which are often called 'structured' water molecules. Water molecules not close to sequence form the bulk water (>5Å). Figures 8B and 8C show distribution of water molecules around three groups of atoms (—OH, —COO^- and —SO_3^-) in IdoA. Quantitative information on the nature of hydration around any atom can be derived using the pair distribution parameter g_{WA} (Equation 1). In this equation, $N_{WA}(r)$ is the average number of water oxygen atoms within a sphere of radius r around the atom of interest and ρ_w is the density of water in the system. This can be calculated using different tools (Refer to Supporting Information).

$$g_{WA}(r) = \frac{1}{4\pi\rho_W r^2}\frac{dN_{WA}(r)}{dr} \quad (1)$$

Pair distribution function and CS studies—: Wiegel et al. analyzed the pair distribution function for —OH, —OSO_3^- and —CH_3 atoms towards the oxygen atom of water molecules in MD simulations of CS-C (Figures 8D–8H, Wiegel et al., 1999). The results showed a characteristic sharp peak at 2.7 Å for oxygens of water around the OH group of CS-C corresponding to the inter-molecular O–HO_W H-bond (Figure 8F). In contrast, the repulsive Lennard-Jones potential presented a broad distribution for water molecules around the —CH_3 group at 3.8 Å suggesting an a-polar interaction (Figure 8G). At the same time, the pair distribution function of water molecules around the —OSO_3^- group was similar to water–water or water–OH distribution (Figure 8H).

The number of water molecules surrounding a solute group, e.g., OH, is determined by integrating the pair distribution function, while the strength of inter-molecular interactions is calculated from the H-bond energy and/or geometry features. For energy, −12 kcal/mol or lower is regarded as a strong H–bond (Rossky & Karplus, 1979). For CS-C, the interaction energy showed an asymmetry in distribution of solute water interaction energies, which implies presence of structured as well as bulk water. Likewise, a distance in the range of 2.5 Å and 3.9 Å between the donor and acceptor atoms was also used for CS-C (Figure 8I, (McDonald & Thornton, 1994). Both calculations showed ~20 water molecules for a disaccharide unit of CS-C with most structured H-bonds formed by —OH and —OSO_3^- groups. Interestingly, the orientation of CH group(s) may be the reason for forming a hydrophobic sub-domain for C6S chain, which helps form helical secondary structures (Scott, 1995; Wiegel et al., 1999).

MD simulations of CS-A also showed similar results using the pair distribution function and interaction energies (Kaufmann et al., 1999). The average number and lifetimes of resident water molecules around the polar atoms of CS-A were measured (Figures 8J-L). The measured lifetimes of bulk water H-bonds were observed to be in the range of 0.1 – 1.4 ps, which indicated a high dynamical behavior involving rapid bond formation and break down (Cavatorta, Deriu, Cola, & Middendorf, 1994). In contrast, the lifetimes were larger for sulfate groups than bulk water (~1.4 ps, Figure 9L). On an average, 19.7 H-bonds per disaccharide unit of CS-A were formed in hydrated environment. Although no direct comparison with experimental studies was possible, non-structured water molecules estimated using differential scanning calorimetry were in agreement with MD studied (~20/ disaccharide unit, (Joshi & Topp, 1992; Jouon, Rinaudo, Milas, & Desbrières, 1995). Interestingly, this count was higher than that measured for other carbohydrates, which ranged from 6 to 8 (Kawai, Sakurai, Inoue, Chûjô, & Kobayashi, 1992).

To understand the impact of different solvent models (e.g., TIP3P, TIP4P, TIP4P-EW and TIP5P) independent MD simulations were performed on CS-A sequences (Neamtu et al., 2013). The radial distribution function (RDF, which is essentially pair distribution function) of hydroxyl O2 of GlcUA and O6 of GalNAc4S to water oxygen showed significant participation in inter-molecular hydrogen bond interactions (Zhong, Bauer, & Patel, 2011). A major first peak for structured water was observed at 2.8 Å for the OH group, which was organized in all water models, except for TIP5P. In contrast, the second layer of hydration was better represented by the TIP5P water model. RDF for sulfate groups showed a well-defined first layer of hydration for both three-point and four-point water models. Also, the results showed that conformational preferences of the β(1→4) linkage are not affected by differences in hydration water when compared to the β(1→3) linkage conformations. Overall, computational performance was good when TIP3P water model was used, which conveys the idea that proper selection of water model is needed to conduct good and efficient MD simulations.

Pair distribution function and HP/HS studies—: Pol-Fachin and Verli have performed two MD studies on trisaccharides of HP/HS containing variations in sulfation level around a central IdoA residue, which was simulated in $^{2}S_{O}$ and $^{1}C_{4}$ forms (Pol-Fachin & Verli, 2008). The water molecules tend to position the inter-glycosidic bonds to maintain higher flexibility when conformational switching occurs, which can be observed from the deviations in H-bond angles and distances as the MD run progresses. No differences were observed in H-bond groups of GlcN upon solvation but changes in hydration shell for IdoA2S was apparent upon transition between $^{2}S_{O}$ and $^{1}C_{4}$ conformations for both sequences. Here, the $^{2}S_{O}$ form appeared to be more solvated and was carrying more tightly bound water molecules in both sequences.

Studies of a library of HP-like trisaccharides carrying a central IdoA2S residue showed more generalized solvation effect around the IdoA2S–GlcN linkage (Muñoz-García et al., 2013; Muñoz-García et al., 2012). These studies identified two solvation shells around IdoA2S in $^{2}S_{O}$ conformation 2 and 3 Å around the inter-glycosidic oxygen, which were not apparent for IdoA2S in $^{1}C_{4}$ form (Figure 8M).

Pair distribution function and HA studies—: One fairly old simulation of HA in water has shown a rather similar pair distribution function for the polar (OH) as well as a-polar (CH_2) groups with the first minimum within 3.4 Å, an average of 3.2 water molecules and an average of 10–15 H-bonds per disaccharide unit (Kaufmann et al., 1998).

Bridging water molecules—: In late 1990s, Almond et al. carried out a study to understand the nature of water interactions (Almond et al., 1997), which led to identification of water molecules that bridge two contiguous residues in GAGs. This study was on HA (GlcA-GlcNAc-GlcA-GlcNAc) and unsulfated CS (GlcA-GalNAc-GlcA-GalNAc) sequences (Almond, 2005). Two types of water-mediated interactions were found – a) a single water molecule and b) two water molecules – that form a bridge between residues of the same sequence (Figures 8N and 8O). For the β(1→3) linkage of HA, bridged water was found between the O7 of GlcNAc and O2 of GlcA and between the O4 of GlcNAc and the O6 of GlcA. Of which, the latter displayed high probability of a dimeric water bridge. Likewise, the β(1→4) linkage showed water bridges between the N2 of GlcNAc and O6 of GlcA and between the O3 of GlcA and O5 of GlcNAc, of which the former was identified as the most consistent across all MD frames.

Epimerization at the 2-position in HA to give GalNAc in chondroitin alters the orientation of water molecule around the β(1→3) linkages such that a bridging water between GlcNAc O4 and O6 of GlcA in HA is found to be lacking in chondroitin (Figure 8O). Meanwhile the β(1→4) linkage of chondroitin possessed water bridges similar to that present in HA. Such precise information is very useful in understanding the overall shape of a GAG sequence in biological recognition. Finally, one of the earliest MD study on sulfated CS-A sequences has shown that a water molecule bridges non-contiguous residues almost 50% of the time (Kaufmann et al., 1999).

Understanding ring conformations in GAGs

Introduction to Cremer-Pople analysis and vicinal scalar coupling constants: To understand ring puckering, Cremer-Pople established three generic parameters Q, φ and θ, which correspond to the radius, meridian angle and azimuthal, respectively (Cremer & Pople, 1975). For a hexapyranose ring, the Cremer-Pople parameters can help uniquely define any ring pucker. Q is the puckering magnitude, which reflects the deviation from a flat ring. φ and θ are puckering parameters and uniquely correspond individual forms such as a $^{1}C_4$ chair (φ,θ = 0°,180°), $^{4}C_1$ chair (φ,θ = 0°,0°), $^{2}S_O$ sofa (φ,θ = 150°,90°), $^{2,5}B$ boat (φ,θ = 120°,90°), etc. (Supplementary Figure S5). Automated tools have been developed to capture these ring parameters to understand the ring flexibility and dynamics throughout an MD run e.g., g-puckering in GROMACS, pucker in cpptraj, getpucker in VMD, etc. Figure 9A shows the schematic representation of these puckers from a 20 ns test simulation starting with IdoA in an arbitrary conformation with ring puckers on the right. Apart from the conformations shown, the ring could also scan intermediate conformers (Supplementary Figure S5). Even a small deviation in φ and θ can be deduced using automated tools (an example and links provided in Supporting Information) (Makeneni, Foley, & Woods, 2014).

Vicinal coupling constant ($^3J_{H,H}$) arises from interaction between scalar (through-bond) coupling of nuclear spins separated three bonds apart. $^3J_{H,H}$ is sensitive to ring conformational flexibility and changes as a function of orientation of scalar spins. ^{1}H NMR is the experimental technique that reports on this parameter. Yet, the observations from NMR are too fast and only provide an average information, while leaving out fine details on the flexibility of the sequences. MD computational simulations are the means to capture these details. Equation 2, called Karplus equation, is used for this purpose(Haasnoot, de Leeuw, & Altona, 1980). In this equation, a three-bond dihedral angle (θ) is related to the corresponding scalar vicinal coupling ($^3J_{H,H}$). Parameter ψ reflects a phase shift between the cosine expansion and the measured torsion angle, while the coefficients A, B and C are empirical constants derived from experimental results on related systems. Analysis of the above two descriptors using MD simulations have offered significant insight into GAG conformational dynamics and structure.

$$^3J_{H,H} = \cos^2(\theta + \psi) + B\cos(\theta + \psi) + C \quad \text{Equation 2}$$

Conformational properties of HP/HS—: The conformational behavior of IdoA residues in a HP/HS chain has been considerably debated in the literature (Casu, 1985; Linhardt, 2003). C5 epimerization in *L*-IdoA induces significant ring flexibility in comparison to *D*-GlcA, which exhibits a stable $^4C_1$ ring conformer in solution. In fact as described above, IdoA and its analogs can adopt many conformations ($^1C_4$, $^4C_1$, ^2S_O, etc.) and experimentally two forms have been the major ones found to exist at equilibrium. Importantly, the conformational equilibrium of the IdoA analogs could be modulated either by sulfation of the neighboring residues or by its location at the chain termini (reducing or non-reducing end). This conformational plasticity has been proposed as a key factor responsible for its various biological functions, which arise from binding to proteins.

Puckering parameters and vicinal coupling constants, discussed above, have been used in many studies to follow ring dynamics and stability in MD runs. The parameters also help compare results of MD simulations to NMR-based experimental results. One of the earliest results using this approach stated that IdoA tends to be in equilibrium between the two chairs, namely $^1C_4$ or $^4C_1$. Later, the discovery of an oversulfated pentasaccharide, which bound better to antithrombin and displayed skew boat ^2S_O form as a key conformer (Casu et al., 1986; Das Sanjoy et al., 2001; Rees, Morris, Stoddart, & Stevens, 1985), kindled theoreticians to explore conformational sampling of GAG residues in more detail. These then shed light on other canonical puckers and helped identify transition intermediates such as half-chair and envelop forms.

The prediction of conformational behavior of a heparin-like structure in water using Particle Mesh Ewald (PME) (Darden et al., 1993)), electrostatic treatment has shown that explicit solvent environment clarifies IdoA behavior better (Angulo, Nieto, & Martin-Lomas, 2003). This method increased the accuracy of IdoA equilibria and their equivalence with experimental coupling constants. These predictions were the first to suggest that sulfated IdoA exhibits an equilibrium of ^2S_O and $^1C_4$ forms. Later, microsecond simulation of IdoA

analogs showed that equilibrium was reached at about 3 μs (Sattelle et al., 2010), which is a fairly long time on an MD timescale. Analysis of the MD runs using Cremer-Pople parameters showed 7:1:2 population of $^{1}C_{4}$:$^{4}C_{1}$:$^{2}S_{O}$ forms for unsulfated IdoA at equilibrium. Upon sulfation at 2-position, the ring was found to be biased towards the $^{1}C_{4}$ chair form with only 10% of the $^{2}S_{O}$ form (Sattelle et al., 2010). In contrast, irrespective of the starting conformation, GlcA was found to be always stable in the $^{4}C_{1}$ chair form. Interestingly, further detailed analysis of MD results has shown that two quasi-stable states – $^{2}S_{O}$ and $^{3}S_{1}$ – for IdoA are populated, irrespective of sulfation (Sattelle et al., 2010).

Puckering analysis of MD runs with both anomers of methyl idopyranosides has shown significant variation as a function of the solvent environment (Sattelle et al., 2012). Two chair conformations were populated for both anomers with $^{1}C_{4}$ being the most favored (84% α- and 99% β-anomer). Theoretically, the transition between the two chair forms ($^{1}C_{4}$ and $^{4}C_{1}$) occurs through half-chair, boat, skew boat and/or envelop forms; yet the β-anomer of methyl idopyranoside did not exhibit half chair and envelops forms (Sattelle et al., 2012). In these studies, the $^{3}J_{H,H}$ vicinal coupling constants were in agreement with the NMR-derived values. Likewise, a report on a library of heparin-like trisaccharides showed that the presence of 6-*O*-sulfation at the reducing end enhances the proportion of $^{2}S_{O}$ form of central IdoA2S (45%) (Muñoz-García et al., 2012). In contrast, when this 6-*O* sulfation is absent, there was a significant drop in $^{2}S_{O}$ population. A later study by the group using tar-MD method showed that there was an increase of population of equatorial conformers (θ=90) of IdoA2S, including $^{2}S_{O}$, $^{2,5}B$, $B_{3,O}$, $^{3}S_{1}$, etc., when 6-*O*-sulfation is present at the reducing end GlcN (Muñoz-García et al., 2013).

A systematic investigation of the role of sulfation of the neighboring residues on the nature of conformational preferences of IdoA and IdoA2S residues has been recently carried out (Hsieh et al., 2016). The study focused on understanding effects of 2-*O*- and 3-*O*-sulfations. Eight hexasaccharides were synthesized containing four each of IdoA and IdoA2S residues located in the middle disaccharide of the chain. MD simulations showed that both IdoA and IdoA2S interconvert between different conformations including $^{1}C_{4}$, $^{2}S_{O}$, and $^{4}C_{1}$ with a distinct set of vicinal coupling constant $^{3}J_{H,H}$. Figure 9G shows the population analysis of ring conformation as a function of hexasaccharide sequence. The lower sulfated sequences preferred IdoA2S in $^{1}C_{4}$ conformation, whereas higher sulfation in the neighboring residues shifted the ring towards the $^{2}S_{O}$ to a maximum of 75% (Hsieh et al., 2016). This correlated with reports on the heparin pentasaccharide (Hricovíni, 2015). When the IdoA is not sulfated at position 2, there was re-adjustment of ring puckering. In lower sulfated sequences, IdoA preferred the skew boat conformers (~47%), whereas higher adjacent sulfation lowered this population (29 to 19%). The GLYCAM06 analysis of MD simulations performed on variably sulfated tetrasaccharides containing ΔUA residue confirmed the effect of neighboring sulfation (GlcNS/GlcNAc) on IdoA2S conformer population. Sequences carrying neighboring GlcNAc favored $^{1}C_{4}$ for IdoA2S, while two adjacent GlcNS increased the population of $^{2}S_{O}$ by 37% (Singh et al., 2016).

The use of MD simulations and conformational analysis to develop insight into protein recognition was presented in a study with a group of heparin hexasaccharide sequences, which were not active towards FGF-1 (Munoz-Garcia et al., 2013). MD-based theoretical

inter-proton distances were in good agreement with results from ROESY experiments, which supported conformational analysis showing IdoA populations changing along the heparin chain. The inactive sequence, IdoA2S-GlcNS6S-IdoA-GlcNAc6S-IdoA2S-GlcNAc6S, showed that the IdoA2S at the non-reducing end is depleted in $^{2}S_{O}$ form, while internal IdoA residues displayed substantial $^{2}S_{O}$. In contrast, the active sequence with higher sulfation (IdoA2S-GlcNS6S-IdoA2S-GlcNS6S-IdoA2S-GlcNS6S) exhibits altered orientation of sulfate groups on account of appropriate conformational preferences, which induce binding to FGF-1 (Munoz-Garcia et al., 2013).

Although majority of long MD simulations (μs) on ring conformations sampled by GAGs have been with mono- or di- saccharide sequences, a couple of studies have attempted to elucidate the effect of chain length. The IdoA residues in oligosaccharides with different chain lengths and variable sulfation levels were especially investigated. It is known that HS carries sulfated and unsulfated domains, called NS and NA domains, respectively (Sattelle et al., 2013), which may range in size from 6 to 10 residues long. Analysis of MD simulations for IdoA2S belonging to the AT-binding heparin pentasaccharide showed back and forth transitions from $^{1}C_{4}$ to $^{4}C_{1}$. The overall pucker occupancy favored $^{4}C_{1}$ (79%) with the presence of half-chair, boat and skew-boat conformers with varying percentages from 1 to 5%. Similar observations were also made for internal IdoA2S residues present in oligosaccharides of the NS-type (Sattelle et al., 2013). It is important to note that observations are different from the earlier monosaccharide simulations in which IdoA2S was found to mostly favor $^{1}C_{4}$ form. A possible explanation is the lower free energy barrier for internal IdoA2S to undergo transitions in water, aided by chain length-specific interactions. Interestingly for longer NS domains, IdoA2S at even numbered positions exhibited a higher population of $^{2}S_{O}$ ranging from 6 to 13%. Overall, IdoA2S located in the middle of a long chain favored $^{4}C_{1}$, whereas that at the terminus tended to prefer $^{1}C_{4}$ form. This is extremely important if identifying protein binding sequences have to be deduced. For GlcA residue, $^{4}C_{1}$ form was nearly uniformly populated irrespective of chain length, position and sulfation level of adjacent residues (Sattelle et al., 2013).

Researchers have also queried whether *D*-GlcNAc is a rigid chair. Extended microsecond MD studies on *N*- and *O*-sulfonated variant monosaccharides (GlcNAc, GlcNS, GlcNS6S, GlcNS6S3S and GlcNS3S) free α-anomeric and methylglycoside forms have helped drive variation in equilibrium populations as a function of substituents (Sattelle & Almond, 2011). For GlcNAc, equilibrium was achieved following 3–5 μs of experimentation. The $^{4}C_{1}$ form was found to be the dominant conformer (99.6%), which has been experimentally validated ($^{3}J_{H,H}$). Yet, a small proportion of $^{1}C_{4}$ was possible to detect during simulations, especially at a transient level. The pathways between these two chairs, $^{4}C_{1}$ and $^{1}C_{4}$, were achieved through canonical non-chair and skew boat puckers (Figure 9B). 1-*O*-methylation in GlcNAc also showed transitions between two chair forms, *albeit* with a lower frequency (Supplementary Figures S6 A and B).

Increasing sulfation level of GlcN, as in GlcNS, resulted in 100% of $^{4}C_{1}$ form. It is important to note that although the simulation was started with $^{1}C_{4}$ pucker, it switched to the $^{4}C_{1}$ chair almost instantly and stayed stable throughout the MD run (Sattelle & Almond, 2011). Up on 6-sulfation, GlcNS6S showed a single change from $^{4}C_{1}$ to $^{1}C_{4}$ with slower

exchange rate (Supplementary Figure S6 C). Further increase in sulfation at the 3-position, as in GlcNS3S and GlcNS3S6S, did not impact much change to the stability of its $^4C_1$ conformer. In fact, the miniscule proportions of other forms became even more inaccessible. The MD results were in good agreement with the crystallographic data on GlcN structures (Ikemizu et al., 1999; Maveyraud et al., 2009; Xu, McBride, Paulson, Basler, & Wilson, 2010). Also, the MD simulations agreed with QM and NMR vicinal coupling constant results (Sattelle & Almond, 2011).

Long microsecond MD simulations of HP/HS chains that are 2 to 10 residues in length and carrying NS and NA type domains haven given insight into the nature of GlcNAc and GlcNS6S too. GlcN residues irrespective of sulfation/acetylation state, i.e., GlcNAc, GlcNS6S and GlcNS6S3S, at the reducing termini showed transitions between the two chair forms. The rate of transition was higher for NA-containing species than for NS-containing oligosaccharides (Sattelle et al., 2013). Interestingly, this was slower for disaccharides in comparison to tetra- to deca- saccharides.

Conformational properties of other GAGs—: In the manner of MD studies on HP/HS sequences, conformational preferences of unsulfated chondroitin, CS-A and HA disaccharides have also been studied (Sattelle & Almond, 2010). MD studies were performed using GLYCAM06 and analyzed for monosaccharide geometries. The results showed that the structures deviated least from the $^4C_1$ form for hexosamine in all test cases with only a 1% proportion of the ^2S_O form. Likewise, GlcA exhibited minimal deviation from the $^4C_1$ geometry but displayed a significantly higher ^2S_O population of 9, 4 and 11% for chondroitin, CS-A and HA, respectively. It is important to note that $^1C_4$ chair form was not identified at all in this analysis. The MD results were supported by experimental measurements of average vicinal coupling constants $^3J_{H,H}$ for each hexopyranose ring.

Overall, very few MD studies on CS and DS sequences have been performed perhaps arising from the results obtained at disaccharide levels showing an overwhelming abundance of $^4C_1$ form. Yet, longer sequences may exhibit differences, which should be studied in detail.

Deducing changes in free energy—As discussed above, the Cremer-Pople parameters derived from MD simulations can be transformed into populations of different conformers at equilibrium, which in turn can be converted into population probabilities (p) for the occurrence of a conformer. This affords deduction of difference in free energy (ΔG) between two conformers (e.g., 1 and 2) at equilibrium using equation 3. In this equation, k_B is the Boltzmann constant and T is the absolute temperature, whereas p_1 and p_2 are the probabilities of occurrence of two conformers 1 and 2. For the simulations of IdoA described above, the $^4C_1$ chair and ^2S_O skew-boat forms displayed higher energies than the $^1C_4$ chair form by 0.9 and 2.6 kcal/mol, respectively. Using the same equation, $^1C_4$ and ^2S_O energies were calculated to be equivalent irrespective of sulfation of IdoA; while the energy of the $^4C_1$ conformer was significantly higher, which ensured a slow conformational exchange between the two chair forms (Sattelle et al., 2010).

$$\Delta G = -k_B T \ln \frac{p_1}{p_2} \quad \text{Equation 3}$$

A later extended MD study (microseconds) on α and β methyl idopyranosides using different models of water showed a free energy difference of 1.7 kcal/mol between the two chair forms for the former anomer (Sattelle et al., 2012). The key driving factor impacting these differences were the distribution of water molecules around the anomeric configurations. These solvent geometries impacted stability of puckers and their conformational kinetics. Likewise, microsecond simulations of GlcN analogs (acetyl and sulfonated variants) showed that the $^4C_1$ form was more stable than the $^1C_4$ form by about 3.5 kcal/mol (Sattelle & Almond, 2011).

An identical free energy analysis can also be performed for the bi-dimensional inter-glycosidic torsional space involving Φ and ψ. The densities of different identifiable states (e.g., '*i,j*' with probability $P_{i,j}$) in the torsional space with maximal probability P_{max} (corresponding to the entire torsional space for a given inter-glycosidic bond) can be transformed using the Boltzmann formulation 3 described above. Application of such as calculation in CS-A did not show noticeable differences in the β1→4 linkage as a function of different water models, as expected; however, three distinct minima were observed for the β1→3 linkage (Neamtu et al., 2013). A major basin of high probability corresponding to a free energy of −3.8 kcal/mol arose from a unique H-bond between the —OH group at position 2 of GlcA and oxygen of GalNAc's acetamido group. Two minor basins observed in the bi-dimensional space had energies of −0.93 kcal/mol and −1.3 kcal/mol and could be rationalized by the presence of some specific H-bonds (Neamtu et al., 2013). It is important to note that minor basins were not observed when higher-order water models (TIP4P, TIP5P) were used in simulations.

Finally, some groups have used advanced sampling techniques and more in-depth calculations to map the free energy landscape such as metadynamics, replica exchange dynamics, and adaptively biased molecular dynamics (Babin & Sagui, 2010; Oborský et al., 2013; Spiwok, Králová, & Tvaroska, 2010; Faller & Guvench, 2015). Generally, these studies have been performed on mono- or di- saccharides belonging to HP/HS and CS. Overall, the results were in good agreement with unbiased MD simulations as well as experimental observations.

Coarse grain simulations—Experimental and computational structural analysis of GAGs has given much insights into their free and protein-bound states in solution (Casu, Naggi, & Torri, 2015; Kjellén & Lindahl, 2018; Pomin & Mulloy, 2018; Sankaranarayanan, Nagarajan, & Desai, 2018). Generally, computational studies have been performed for sequences that are mono- to dodeca- saccharide long. Yet, GAGs are much longer. A typical GAG chain may carry 50 to 200 residues. Considering the configurational and conformational complexity of constituent saccharides, study of a GAG chain's conformational properties is extremely difficult. Even though the approach involving all-atom computational classical MD simulations (discussed above) laid a path to understanding

GAG solution behavior, it is more likely a time and resource consuming on increasing chain length.

For longer HP/HS sequences carrying multiple NA and NS domains, which may or may not bind a protein in a specific manner, it is difficult to extract detailed structural information within relevant time using all-atom simulations is computationally challenging. An alternative approach that is gaining strength is the coarse grain (CG) simulation. This approaches reduces the number of degrees of freedom involved in computational calculations by assigning pseudo-atoms (Ingólfsson et al., 2014). Many CG models, including MB3, Martini, Srinivas, Markutsya, etc. have been proposed to perform carbohydrate simulations (López et al., 2009; Molinero & Goddard, 2004; Srinivas, Cheng, & Smith, 2011).

Bathe et al. performed the first successful CG approach specific to GAGs and analyzed chondroitin and HA characteristics as a function of ionic strength and pH (Bathe, Rutledge, Grodzinsky, & Tidor, 2005). Their CG topology was built using the standard degrees of freedom (DOF) and included inter-glycosidic linkage torsion angles. The DOFs were reduced by considering three CG atoms for each monosaccharide and two additional virtual pseudo-CG atoms including the center of charge (V_Q) and the center of geometry (V_G). During simulations all bonds were treated as rigid and only flexible inter-glycosidic torsions were allowed to move. Since the internal motions (DOFs) of each residue reflects a constant Cremer-Pople ring puckering, CG parameters were fine-tuned using results from the classical all-atom disaccharide simulations. Non-bonded interactions (electrostatic and steric) between the adjacent residues were taken into consideration using appropriate potentials. Predictions based on the CG model were comparable to experimental observations; however, it is important to note that these experiments were much more efficient.

As discussed above, ring conformations of UA are important, which implies that it is not enough to take only inter-glycosidic torsions and CG beads, which simulate residues. To gain insight into conformational dynamics, it is essential to include pyranose ring motions in CG model(s). Satelle et al. included ring puckering in their CG model in addition to inter-glycosidic torsions by using the Cremer–Pople parameters as DOFs. They employed two separate energy functions for inter-glycosidic linkage and puckering motions. The overall hydrodynamic motions not only reproduced orientations at individual residue level but also explored conformational energy landscape of sugar puckering conformers (Sattelle et al., 2013; Sattelle et al., 2015).

Samsonov et al. have developed an alternating approach of coining 28 pseudo-atoms corresponding to functional groups of 17 different GAG residues and a CG parameter set that is compatible to the AMBER force field (Salomon-Ferrer et al., 2013). An analysis of PDB-deposited protein–GAG complexes was performed using three key interactions of functional groups to proteins. CG parameters such as bonds, virtual bonds, bond and torsional angles of pseudo-atoms were extracted from all-atom force fields and restraints were used to maintain equilibrium values. The non-bonded interactions were taken into account by assigning integer charges to sulfate (−1) and carboxyl group (−1), zero charges to

all other atoms, and implementing Lenard-Jones parameters using potential of mean force approach (PMF). Overall, a good agreement on global (RMSD, length and R gyration) and local characteristics (bond, angle and dihedrals) of GAGs was derived between CG and all-atom approaches (Samsonov, Bichmann, & Pisabarro, 2015).

Recently, Kolesnikov et al. developed a field-theoretic approach in CG modeling to describe the solution properties of GAGs in presence of added salt concentrations (Kolesnikov, Budkov, & Nogovitsyn, 2014). This work was based on the data from experimental osmotic pressure in calculating the static structure factor, degree of dissociation, persistence length and radius gyration for semi flexible polyelectrolyte chains in presence of salt. This simplistic theoretical approach by means of adjustable parameters showed the agreement between experimental and theoretical approach.

Conclusions and Future Directions

This review was directed toward beginners and experimental GAG biologists to help understand development and application of computational techniques for GAG. Understanding free solution behavior of GAGs and sequences of interest is key to performing complex simulations for answering biological questions. Rather than synthesizing or isolating GAG sequences, computational simulation are relatively inexpensive. In fact, the technology has developed to a level that meets experimental results in many cases.

The all-atom computational conformational studies of GAGs have reached a state of maturity that enables an interested non-computational specialist with some basic understanding to perform MD simulations and derive meaningful conclusions. Many tools are available and parameters for many atoms and groups have been deduced. Insight on many different aspects of GAG structure can be derived including overall shape(s) of GAGs in solution, flexibility of chains of different lengths, populations of different conformers likely to exist in solution, rate of conversion of different conformer populations, dynamical variations in surface electrostatics, strength and persistence of H-bonds, impact of GAG sequence on ground state conformations, and role of water molecules in preservation of GAG structure and conformation. These insights are extremely useful for identifying and explaining protein recognition by certain GAG sequences.

Unfortunately, most biology-based researchers have not availed of these advances to pinpoint GAG sequences of interest. The traditional discovery of promising GAG sequences has sought to use solution-based experimental techniques as the first step; although computational methods of identifying GAG sequences of interest may save much time and resources. Here it is important to reiterate that computational methods of simulation have been able to put forward predictions for GAG–protein systems in solution (Samsonov, Gehrcke, & Pisabarro, 2014; Sapay, Cabannes, Petitou, & Imberty, 2011; Sepuru et al., 2016; Verli & Guimarães, 2005) and computational study of GAGs is no longer an esoteric technique.

The subject of computational studies of GAGs is poised for major leaps in many directions. Considering that simulation of behavior of oligosaccharides is now feasible, the next step is

going to be designing GAG sequences that bind to a specific site on target proteins. Efforts in this direction have already yielded fruits suggesting that most GAG-binding proteins should be targetable. A key aspect of this effort will be to predict the exact overall fold and electrostatics of sequences with high level of accuracy, especially for longer chains. Also it would be important to elucidate elements of GAG sequences that engineer specificity of interaction with proteins. In this connection, a key direction may be to rely on residues that are not present in nature, e.g., IdoA2S3S and GlcA2S3S, which could offer a route to engineering specificity. Finally, another important direction would be to design GAG-like chains that carry non-natural, synthetic variations of saccharide substituents.

A primary goal of enabling computational technology for understanding GAG structure and recognition is to realize more clinically useful GAGs. The current state of DruGAGome (drugs that are GAGs) is poor. The sole member of DruGAGome is heparin pentasaccharide, in its various forms including fondaparinux, enoxaparin and heparin. Identification of few other sequences that bind to proteins, e.g., heparin cofactor II and glycoprotein D, suggests that membership of DruGAGome may increase in the future. But major efforts are needed to enhance the probability of realizing this possibility.

Yet, there are challenges. In addition to the fundamental scientific question whether highly specific GAG sequences are possible at the level of proteins and nucleic acid, technological challenges abound. Understanding all possible topologies for a given chain length is not easy and needs massive computational power. There is a chain-length dependent proportional increase in computational complexity and resource needs. Method such as united atom model and coarse grain have started to make some inroads but at the cost of fine information at atomistic level. However, it is likely that models of binary and ternary complexes of GAGs with proteins may be realized through these approaches.

Overall, advances in simulation of GAGs made over the past decade or so have positioned interested researchers, especially the experimental biologist, to undertake fairly sophisticated experiments on how a given GAG sequence may behave in solution and interact with desired targets as well as un-intended targets. The level of sophistication has reached to the point of being able to designing GAG sequences with beneficial properties. Computational approaches should continue to gain strength and help expand the DruGAGome in future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the NIH including HL107152, HL090586 and HL128639.

REFERENCES

Accelrys Discovery Studio 2.1. Accelrys: San Diego, C., & http://www.accelrys.com.

Allinger NL (1976). Calculation of Molecular Structure and Energy by Force-Field Methods In Gold V & Bethell D (Eds.), Advances in Physical Organic Chemistry (Vol. 13, pp. 1–82): Academic Press.

Allinger NL (1977). Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. Journal of the American Chemical Society, 99(25), 8127–8134. doi:10.1021/ja00467a001

Allinger NL, & Chung DY (1976). Conformational analysis. 118. Application of the molecular-mechanics method to alcohols and ethers. Journal of the American Chemical Society, 98(22), 6798–6803. doi:10.1021/ja00438a004

Allinger NL, Yuh YH, & Lii JH (1989). Molecular mechanics. The MM3 force field for hydrocarbons. 1. Journal of the American Chemical Society, 111(23), 8551–8566. doi:10.1021/ja00205a001

Almond A (2005). Towards understanding the interaction between oligosaccharides and water molecules. Carbohydrate Research, 340(5), 907–920. doi:10.1016/j.carres.2005.01.014 [PubMed: 15780256]

Almond A, & Sheehan JK (2000). Glycosaminoglycan conformation: do aqueous molecular dynamics simulations agree with x-ray fiber diffraction? Glycobiology, 10(3), 329–338. doi:10.1093/glycob/10.3.329 [PubMed: 10704532]

Almond A, Brass A, & Sheehan JK (1998a). Deducing polymeric structure from aqueous molecular dynamics simulations of oligosaccharides: predictions from simulations of hyaluronan tetrasaccharides compared with hydrodynamic and X-ray fibre diffraction data11Edited by R. Huber. Journal of Molecular Biology, 284(5), 1425–1437. doi:10.1006/jmbi.1998.2245 [PubMed: 9878361]

Almond A, Brass A, & Sheehan JK (1998b). Dynamic exchange between stabilized conformations predicted for hyaluronan tetrasaccharides: Comparison of molecular dynamics simulations with available NMR data. Glycobiology, 8(10), 973–980. doi:10.1093/glycob/8.10.973 [PubMed: 9719678]

Almond A, Sheehan JK, & Brass A (1997). Molecular dynamics simulations of the two disaccharides of hyaluronan in aqueous solution. Glycobiology, 7(5), 597–604. [PubMed: 9254042]

Andersen HC (1983). Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. Journal of Computational Physics, 52(1), 24–34. doi:10.1016/0021-9991(83)90014-1

Angulo J, Nieto PM, & Martin-Lomas M (2003). A molecular dynamics description of the conformational flexibility of the l-iduronate ring in glycosaminoglycans. Chemical Communications. (13), 1512–1513. doi:10.1039/B303386B [PubMed: 12868729]

Atkins EDT, Phelps CF, & Sheehan JK (1972). The conformation of the mucopolysaccharides. Hyaluronates. Biochemical Journal, 128(5), 1255. [PubMed: 4643702]

Atkins ED, & Sheehan JK (1971). The molecular structure of hyaluronic acid. Biochemical Journal, 125(4), 92P.

Atkins ED, Meader D, & Scott JE (1980). Model for hyaluronic acid incorporating four intramoledular hydrogen bonds. International Journal of Biological Macromolecules, 2(1), 318–319. doi:10.1016/0141-8130(81)90029-5

Autieri E, Sega M, Pederiva F, & Guella G (2010). Puckering free energy of pyranoses: A NMR and metadynamics-umbrella sampling investigation. The Journal of Chemical Physics, 133(9), 095104. doi:10.1063/1.3476466 [PubMed: 20831339]

Babin V, & Sagui C (2010). Conformational free energies of methyl-α-L-iduronic and methyl-β-D-glucuronic acids in water. The Journal of Chemical Physics, 132(10), 104108. doi:10.1063/1.3355621 [PubMed: 20232948]

Bathe M, Rutledge GC, Grodzinsky AJ, & Tidor B (2005). A coarse-grained molecular model for glycosaminoglycans: Application to chondroitin, chondroitin sulfate, and hyaluronic Acid. Biophysical Journal, 88(6), 3870–3887. doi:10.1529/biophysj.104.058800 [PubMed: 15805173]

Bayraktar H, Akal E, Sarper O, & Varnali T (2004). Modeling glycosaminoglycans—hyaluronan, chondroitin, chondroitin sulfate A, chondroitin sulfate C and keratan sulfate. Journal of Molecular Structure: THEOCHEM, 683(1), 121–132. doi:10.1016/j.theochem.2004.07.001

Becker CF, Guimarães JA, & Verli H (2005). Molecular dynamics and atomic charge calculations in the study of heparin conformation in aqueous solution. Carbohydrate Research, 340(8), 1499–1507. doi:10.1016/j.carres.2005.03.018 [PubMed: 15882850]

Beecher CN, Young RP, Langeslay DJ, Mueller LJ, & Larive CK (2014). Hydroxyl-proton hydrogen bonding in the heparin oligosaccharide Arixtra in aqueous solution. The Journal of Physical Chemistry B, 118(2), 482–491. doi:10.1021/jp410540d [PubMed: 24354321]

Blanchard V, Chevalier F, Imberty A, Leeflang BR, Basappa, Sugahara K, & Kamerling JP (2007). Conformational studies on five octasaccharides isolated from chondroitin sulfate using NMR spectroscopy and molecular modeling. Biochemistry, 46(5), 1167–1175. doi:10.1021/bi061971f [PubMed: 17260946]

Bochevarov AD, Harder E, Hughes TF, Greenwood JR, Braden DA, Philipp DM, Friesner RA (2013). Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. International Journal of Quantum Chemistry, 113(18), 2110–2142. doi: 10.1002/qua.24481

Bock K, Meldal M, Bundle DR, Iversen T, Garegg PJ, Norberg T, Svenson SB (1984). The conformation of Salmonella O-antigenic polysaccharide chains of serogroups A, B, and D1 predicted by semi-empirical, hard-sphere (HSEA) calculations. Carbohydrate Research, 130(Supplement C), 23–34. doi:10.1016/0008-6215(84)85267-2 [PubMed: 6478459]

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, & Karplus M (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry, 4(2), 187–217. doi:10.1002/jcc.540040211

Brooks BR, Brooks CL 3rd., Mackerell AD Jr., Nilsson L, Petrella RJ, Roux B, Karplus M (2009). CHARMM: The Biomolecular simulation program. Journal of Computational Chemistry, 30*(*10): 1545–1614. doi: 10.1002/jcc.21287 [PubMed: 19444816]

Cael JJ, Isaac DH, Blackwell J, Koenig JL, Atkins EDT, & Sheehan JK (1976). Polarized infrared spectra of crystalline glycosaminoglycans. Carbohydrate Research, 50(2), 169–179. doi:10.1016/S0008-6215(00)83848-3 [PubMed: 991159]

Cael JJ, Winter WT, & Arnott S (1978). Calcium chondroitin 4-sulfate: Molecular conformation and organization of polysaccharide chains in a proteoglycan. Journal of Molecular Biology, 125(1), 21–42. doi:10.1016/0022-2836(78)90252-8 [PubMed: 712856]

Casu B (1985). Structure and Biological Activity of Heparin In Tipson RS & Horton D (Eds.), Advances in Carbohydrate Chemistry and Biochemistry (Vol. 43, pp. 51–134): Academic Press. [PubMed: 3913287]

Casu B, Choay J, Ferro DR, Gatti G, Jacquinet JC, Petitou M, Torri G (1986). Controversial glycosaminoglycan conformations. Nature, 322, 215. doi:10.1038/322215b0 [PubMed: 3736652]

Casu B, Naggi A, & Torri G (2015). Re-visiting the structure of heparin. Carbohydrate Research, 403, 60–68. doi:10.1016/j.carres.2014.06.023 [PubMed: 25088334]

Cavatorta F, Deriu A, Cola DD, & Middendorf HD (1994). Diffusive properties of water studied by incoherent quasi-elastic neutron scattering. Journal of Physics: Condensed Matter, 6(23A), A113.

Chakrabarti B (1977). Effect of counterions on the conformation of hyaluronic acid. Archives of Biochemistry and Biophysics, 180(1), 146–150. doi:10.1016/0003-9861(77)90018-2 [PubMed: 16566]

Chakrabarti B, Park JW, & Stevens ES (1980). Glycosaminoglycans: Structure and Interactio. Critical Reviews in Biochemistry, 8(3), 225–313. doi:10.3109/10409238009102572 [PubMed: 6774852]

Cifonelli JA, & King J (1973). Structural studies on heparins with unusually high N-acetylglucosamine contents. Biochimica et Biophysica Acta (BBA) - General Subjects, 320(2), 331–340. doi: 10.1016/0304-4165(73)90313-9 [PubMed: 4750749]

Cilpa G, Hyvönen MT, Koivuniemi A, & Riekkola ML (2010). Atomistic insight into chondroitin-6-sulfate glycosaminoglycan chain through quantum mechanics calculations and molecular dynamics simulation. Journal of Computational Chemistry, 31(8), 1670–1680. doi:10.1002/jcc.21453 [PubMed: 20087899]

Clark M, Cramer RD, & Van Opdenbosch N (1989). Validation of the general purpose tripos 5.2 force field. Journal of Computational Chemistry, 10(8), 982–1012. doi:10.1002/jcc.540100804

Cleland RL (1971). Ionic polysaccharides. V. Conformational studies of hyaluronic acid, cellulose, and laminaran. Biopolymers, 10(10), 1925–1948. doi:10.1002/bip.360101012

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Kollman PA (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society, 117(19), 5179–5197. doi:10.1021/ja00124a002

Cremer D, & Pople JA (1975). General definition of ring puckering coordinates. Journal of the American Chemical Society, 97(6), 1354–1358. doi:10.1021/ja00839a011

Cros S, Petitou M, Sizun P, Pérez S, & Imberty A (1997). Combined NMR and molecular modeling study of an iduronic acid-containing trisaccharide related to antithrombotic heparin fragments. *Bioorganic &* Medicinal Chemistry, 5(7), 1301–1309. doi:10.1016/S0968-0896(97)00087-4

Darden T, York D, & Pedersen L (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics, 98(12), 10089–10092. doi:10.1063/1.464397

Das Sanjoy K, Mallet JM, Esnault J, Driguez PA, Duchaussoy P, Sizun P, Sinaÿ P (2001). Synthesis of conformationally locked carbohydrates: A skew-boat conformation of L-iduronic acid governs the antithrombotic activity of heparin. Angewandte Chemie International Edition, 40(9), 1670–1673. doi:10.1002/1521-3773(20010504)40:9<1670::AID-ANIE16700>3.0.CO;2-Q [PubMed: 11353474]

DeAngelis PL, Liu J, & Linhardt RJ (2013). Chemoenzymatic synthesis of glycosaminoglycans: re-creating, re-modeling and re-designing nature's longest or most complex carbohydrate chains. Glycobiology, 23(7), 764–777. doi:10.1093/glycob/cwt016 [PubMed: 23481097]

Del Re G, Pullman B, & Yonezawa T (1963). Electronic structure of the α-amino acids of proteins: I. Charge distributions and proton chemical shifts. Biochimica et Biophysica Acta, 75, 153–182. doi:10.1016/0006-3002(63)90595-X [PubMed: 14083899]

E SJ, Yuan C, & Andy B (2005). Secondary and tertiary structures involving chondroitin and chondroitin sulphates in solution, investigated by rotary shadowing/electron microscopy and computer simulation. European Journal of Biochemistry, 209(2), 675–680. doi:10.1111/j.1432-1033.1992.tb17335.x

Esko JD, & Selleck SB (2002). Order out of chaos: assembly of ligand binding sites in heparan sulfate. Annual Reviews of Biochemistry, 71, 435–471. doi:10.1146/annurev.biochem.71.110601.135458

Faller CE, & Guvench O (2015). Sulfation and cation effects on the conformational properties of the glycan backbone of chondroitin sulfate disaccharides. The Journal of Physical Chemistry B, 119(20), 6063–6073. doi:10.1021/jp511431q [PubMed: 25906376]

Ferro DR, Provasoli A, & Ragazzi M (1992). Driving the pyranoid ring conformation in molecular mechanics calculations. Carbohydrate Research, 228(2), 439–443. doi:10.1016/0008-6215(92)84137-H

Ferro DR, Provasoli A, Ragazzi M, Casu B, Torri G, Bossennec V, Choay J (1990). Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences. Carbohydrate Research, 195(2), 157–167. doi:10.1016/0008-6215(90)84164-P [PubMed: 2331699]

Ferro DR, Provasoli A, Ragazzi M, Torri G, Casu B, Gatti G, Choay J (1986). Evidence for conformational equilibrium of the sulfated L-iduronate residue in heparin and in synthetic heparin mono- and oligo-saccharides: NMR and force-field studies. Journal of the American Chemical Society, 108(21), 6773–6778. doi:10.1021/ja00281a052

Ferro DR, Pumilia P, Cassinari A, & Ragazzi M (1995). Treatment of ionic species in force-field calculations: Sulfate and carboxylate groups in carbohydrates. International Journal of Biological Macromolecules, 17(3), 131–136. doi:10.1016/0141-8130(95)92679-K [PubMed: 7577811]

Foley BL, Tessier MB, & Woods RJ (2012). Carbohydrate force fields. *Wiley interdisciplinary reviews*. Computational molecular science, 2(4), 652–697. doi:10.1002/wcms.89 [PubMed: 25530813]

Forster MJ, & Mulloy B (1993). Molecular dynamics study of iduronate ring conformation. Biopolymers, 33(4), 575–588. doi:10.1002/bip.360330407

Frisch MJ TG, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JJA, Vreven T, Kudin KN, Burant JC, et al. Gaussian 03, Revision C.02. 2004.

Funderburgh JL (2000). MINI REVIEW Keratan sulfate: structure, biosynthesis, and function. Glycobiology, 10(10), 951–958. doi:10.1093/glycob/10.10.951 [PubMed: 11030741]

Furlan S, La Penna G, Perico A, & Cesàro A (2005). Hyaluronan chain conformation and dynamics. Carbohydrate Research, 340(5), 959–970. doi:10.1016/j.carres.2005.01.030 [PubMed: 15780260]

Gama CI, Tully SE, Sotogaku N, Clark PM, Rawat M, Vaidehi N, Hsieh-Wilson LC (2006). Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. Nature Chemical Biology, 2, 467. doi:10.1038/nchembio810 [PubMed: 16878128]

Gandhi NS, & Mancera RL (2010). Can current force fields reproduce ring puckering in 2-O-sulfo-α-l-iduronic acid? A molecular dynamics simulation study. Carbohydrate Research, 345(5), 689–695. doi:10.1016/j.carres.2009.12.020 [PubMed: 20097328]

Gatti G, Casu B, & Perlin AS (1978). Conformations of the major residues in heparin. 1H-NMR spectroscopic studies. Biochemical and Biophysical Research Communications, 85(1), 14–20. doi:10.1016/S0006-291X(78)80004-7 [PubMed: 743267]

Gatti G, Casu B, Hamer GK, & Perlin AS (1979). Studies on the conformation of heparin by ^{1}H and ^{13}C NMR spectroscopy. Macromolecules, 12(5), 1001–1007. doi:10.1021/ma60071a044

Golabek AA, Walus M, Wisniewski KE, & Kida E (2005). Glycosaminoglycans modulate activation, activity, and stability of tripeptidyl-peptidase I in vitro and in vivo. Journal of Biological Chemistry, 280(9), 7550–7561. [PubMed: 15582991]

Greeley BH, Russo TV, Mainz DT, Friesner RA, Langlois JM, Goddard WA, Ringnalda MN (1994). New pseudospectral algorithms for electronic structure calculations: Length scale separation and analytical two-electron integral corrections. The Journal of Chemical Physics, 101(5), 4028–4041. doi:10.1063/1.467520

Griffin ME, & Hsieh-Wilson LC (2013). Synthetic probes of glycosaminoglycan function. Curr Opin Chem Biol, 17(6), 1014–1022. doi:10.1016/j.cbpa.2013.09.015 [PubMed: 24148269]

Guvench O, Hatcher E, Venable RM, Pastor RW, & MacKerell AD (2009). CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. Journal of Chemical Theory and Computation, 5(9), 2353–2370. doi:10.1021/ct900242e [PubMed: 20161005]

Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, MacKerell AD (2011). CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. Journal of Chemical Theory and Computation, 7(10), 3162–3180. doi:10.1021/ct200328p [PubMed: 22125473]

Haasnoot CAG, de Leeuw FAAM, & Altona C (1980). The relationship between proton-proton NMR coupling constants and substituent electronegativities—I: An empirical generalization of the karplus equation. Tetrahedron, 36(19), 2783–2792. doi:10.1016/0040-4020(80)80155-4

Harvey MJ, Giupponi G, & Fabritiis GD (2009). ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. Journal of Chemical Theory and Computation, 5(6), 1632–1639. doi:10.1021/ct9000685 [PubMed: 26609855]

Heatley F, & Scott JE (1988). A water molecule participates in the secondary structure of hyaluronan. Biochemical Journal, 254(2), 489. [PubMed: 2845953]

Heatley F, Scott JE, Jeanloz RW, & Walker-Nasir E (1982). Secondary structure in glycosaminoglycuronans: N.m.r. spectra in dimethyl sulphoxide of disaccharides related to hyaluronic acid and chondroitin sulphate. Carbohydrate Research, 99(1), 1–11. doi:10.1016/S0008-6215(00)80969-6 [PubMed: 6799200]

Hess B, Bekker H, Berendsen HJC, & Fraaije JGEM (1997). LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry, 18(12), 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hess B, Kutzner C, van der Spoel D, & Lindahl E (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation, 4(3), 435–447. doi:10.1021/ct700301q [PubMed: 26620784]

Holmbeck SMA, Petillo PA, & Lerner LE (1994). The solution conformation of hyaluronan: A combined NMR and molecular dynamics study. Biochemistry, 33(47), 14246–14255. doi:10.1021/bi00251a037 [PubMed: 7947836]

Hong S, Choi G, Park S, Chung A-S, Hunter E, & Rhee SS (2001). Type D retrovirus gag polyprotein interacts with the cytosolic chaperonin TRiC. Journal of Virology, 75(6), 2526–2534. doi:10.1128/JVI.75.6.2526-2534.2001 [PubMed: 11222675]

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, & Simmerling C (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and* Bioinformatics, 65(3), 712–725. doi:10.1002/prot.21123

Hricovíni M (2015). Solution structure of heparin pentasaccharide: NMR and DFT Analysis. The Journal of Physical Chemistry B, 119(38), 12397–12409. doi:10.1021/acs.jpcb.5b07046 [PubMed: 26340667]

Hricovíni M, Guerrini M, Bisio A, Torri G, Petitou M, & Casu B (2001). Conformation of heparin pentasaccharide bound to antithrombin III. Biochemical Journal, 359(2), 265. [PubMed: 11583572]

Hsieh P-H, Thieker DF, Guerrini M, Woods RJ, & Liu J (2016). Uncovering the relationship between sulphation patterns and conformation of iduronic acid in heparan sulphate. Scientific Reports, 6, 29602. doi:10.1038/srep29602 [PubMed: 27412370]

Hu YP, Lin SY, Huang CY, Zulueta MM, Liu JY, Chang W, Hung SC (2011). Synthesis of 3-O-sulfonated heparan sulfate octasaccharides that inhibit the herpes simplex virus type 1 host-cell interaction. Nature Chemistry 3(7), 557–563. doi: 10.1038/nchem.1073

Huige CJM, & Altona C (1995). Force field parameters for sulfates and sulfamates based on ab initio calculations: Extensions of AMBER and CHARMm fields. Journal of Computational Chemistry, 16(1), 56–79. doi:10.1002/jcc.540160106

Huntington JA (2013). Thrombin inhibition by the serpins. Journal of Thrombosis and Haemostasis. 11 Suppl 1, 254–264. doi: 10.1111/jth.12252. [PubMed: 23809129]

Iannuzzi C, Irace G, & Sirangelo I (2015). The effect of glycosaminoglycans (GAGs) on amyloid aggregation and toxicity. Molecules, 20(2). doi:10.3390/molecules20022510

Ikemizu S, Sparks LM, van der Merwe PA, Harlos K, Stuart DI, Jones EY, & Davis SJ (1999). Crystal structure of the CD2-binding domain of CD58 (lymphocyte function-associated antigen 3) at 1.8-Å resolution. Proceedings of the National Academy of Sciences, 96(8), 4289.

Ingólfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, & Marrink SJ (2014). The power of coarse graining in biomolecular simulations. *Wiley interdisciplinary reviews*. Computational molecular science, 4(3), 225–248. doi:10.1002/wcms.1169 [PubMed: 25309628]

Isaac DH, & Atkins EDT (1973). Molecular conformations of chondroitin-4-sulphate. Nature New Biology, 244, 252. doi:10.1038/newbio244252a0 [PubMed: 4516605]

Jana M, & Bandyopadhyay S (2012). Conformational flexibility of a protein-carbohydrate complex and the structure and ordering of surrounding water. Physical Chemistry Chemical Physics, 14(18), 6628–6638. doi:10.1039/C2CP24104H [PubMed: 22460826]

Jeffrey GA (1994). The role of the hydrogen bond and water in biological processes. Journal of Molecular Structure, 322, 21–25. doi:10.1016/0022-2860(94)87017-9

Jeffrey GA, & Taylor R (1980). The application of molecular mechanics to the structures of carbohydrates. Journal of Computational Chemistry, 1(1), 99–109. doi:10.1002/jcc.540010114

Jorgensen WL, Maxwell DS, & Tirado-Rives J (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. Journal of the American Chemical Society, 118(45), 11225–11236. doi:10.1021/ja9621760

Joshi HN, & Topp EM (1992). Hydration in hyaluronic acid and its esters using differential scanning calorimetry. International Journal of Pharmaceutics, 80(1), 213–225. doi: 10.1016/0378-5173(92)90279-B

Jouon N, Rinaudo M, Milas M, & Desbrières J (1995). Hydration of hyaluronic acid as a function of the counterion type and relative humidity. Carbohydrate Polymers, 26(1), 69–73. doi: 10.1016/0144-8617(95)98837-7

Kaliannan P, Vishveshwara S, & Rao VSR (1983). Ab initio SCF—MO study of the molecular structures of aminomethanol, aminesulfonic acid and N-methyl-sulfamate. Journal of Molecular Structure: THEOCHEM, 105(3), 359–374. doi:10.1016/0166-1280(83)80214-0

Kaufmann J, Möhle K, Hofmann H-J, & Arnold K (1998). Molecular dynamics study of hyaluronic acid in water. Journal of Molecular Structure: THEOCHEM, 422(1), 109–121. doi:10.1016/S0166-1280(97)00084-5

Kaufmann J, Möhle K, Hofmann H-J, & Arnold K (1999). Molecular dynamics of a tetrasaccharide subunit of chondroitin 4-sulfate in water. Carbohydrate Research, 318(1), 1–9. doi:10.1016/S0008-6215(99)00091-9 [PubMed: 10515047]

Kawai H, Sakurai M, Inoue Y, Chûjô R, & Kobayashi S (1992). Hydration of oligosaccharides: Anomalous hydration ability of trehalose. Cryobiology, 29(5), 599–606. doi: 10.1016/0011-2240(92)90064-9 [PubMed: 1424716]

Lim Kian-Tat, Brunett Sharon, Iotov Mihail, McClurg Richard B., Vaidehi Nagarajan, Dasgupta Siddharth, … G. WA III (1997). Molecular dynamics for very large systems on massively parallel computers: The MPSim program. Journal of Computational Chemistry, 18, 501–521.

Kirschner KN, & Woods RJ (2001). Solvent interactions determine carbohydrate conformation. Proceedings of the National Academy of Sciences, 98(19), 10541–10545.

Kirschner KN, Yongye AB, Tschampel SM, GonzÁLez-OuteiriÑO J, Daniels CR, Foley BL, & Woods RJ (2008). GLYCAM06: A generalizable biomolecular force dield. Carbohydrates. Journal of Computational Chemistry, 29(4), 622–655. doi:10.1002/jcc.20820 [PubMed: 17849372]

Kjellén L, & Lindahl U (2018). Specificity of glycosaminoglycan–protein interactions. Current Opinion in Structural Biology, 50, 101–108. doi:10.1016/j.sbi.2017.12.011 [PubMed: 29455055]

Kolesnikov AL, Budkov YA, & Nogovitsyn EA (2014). Coarse-grained model of glycosaminoglycans in aqueous salt solutions. A field-theoretical approach. The Journal of Physical Chemistry B, 118(46), 13037–13049. doi:10.1021/jp503749a [PubMed: 25319727]

Lamba D, Glover S, Mackie W, Rashid A, Sheldrick B, & Pérez S (1994). Insights into stereochemical features of sulphated carbohydrates: X-ray crystallographic and modelling investigations. Glycobiology, 4(2), 151–163. doi:10.1093/glycob/4.2.151 [PubMed: 8054715]

Langeslay DJ, Young RP, Beni S, Beecher CN, Mueller LJ, & Larive CK (2012). Sulfamate proton solvent exchange in heparin oligosaccharides: evidence for a persistent hydrogen bond in the antithrombin-binding pentasaccharide Arixtra. Glycobiology, 22(9), 1173 −1182. [PubMed: 22593556]

Lemieux RU, Bock K, Delbaere LTJ, Koto S, & Rao VS (1980). The conformations of oligosaccharides related to the ABH and Lewis human blood group determinants. Canadian Journal of Chemistry, 58(6), 631–653. doi:10.1139/v80-098

Letardi S, La Penna G, Chiessi E, Perico A, & Cesàro A (2002). Conformational dynamics of hyaluronan in solution. 2. Mode-coupling diffusion approach to oligomers. Macromolecules, 35(1), 286–300. doi:10.1021/ma0112220

Li W, & Huntington JA (2012). Crystal structures of protease nexin-1 in complex with heparin and thrombin suggest a 2-step recognition mechanism. Blood. 120*(2)*, 459–67. doi: 10.1182/blood-2012-03-415869. [PubMed: 22618708]

Li W, Adams TE, Nangalia J, Esmon CT, & Huntington JA (2008). Molecular basis of thrombin recognition by protein C inhibitor revealed by the 1.6-A structure of the heparin-bridged complex. Proceedings of National Academy of Sciences USA. 105*(*12*)*, 4661–4666. doi: 10.1073/pnas.0711055105.

Linhardt RJ (2003). 2003 Claude S. Hudson award address in carbohydrate chemistry. Heparin: structure and activity. Journal of Medicinal Chemistry, 46(13), 2551–2564. doi:10.1021/jm030176m [PubMed: 12801218]

Lins RD, & Hunenberger PH (2005). A new GROMOS force field for hexopyranose-based carbohydrates. Journal of Computational Chemistry, 26(13), 1400–1412. doi:10.1002/jcc.20275 [PubMed: 16035088]

López CA, Rzepiela AJ, de Vries AH, Dijkhuizen L, Hünenberger PH, & Marrink SJ (2009). Martini coarse-grained force field: Extension to carbohydrates. Journal of Chemical Theory and Computation, 5(12), 3195–3210. doi:10.1021/ct900313w [PubMed: 26602504]

Lu W, Zong C, Chopra P, Pepi Lauren E, Xu Y, Amster IJ, Boons GJ (2018). Controlled chemoenzymatic synthesis of heparan sulfate oligosaccharides. Angewandte Chemie International Edition, 57(19), 5340–5344. doi:10.1002/anie.201800387 [PubMed: 29512241]

Lujan TJ, Underwood CJ, Jacobs NT, & Weiss JA (2009). Contribution of glycosaminoglycans to viscoelastic tensile behavior of human ligament. Journal of Applied Physiology, 106(2), 423–431. doi:10.1152/japplphysiol.90748.2008 [PubMed: 19074575]

Guerrini M, Guglieri S, Casu B, Torri G, Mourier P, Boundier C, & Viskov C (2008). Antithrombin-binding octasaccharides and role of extensions of the active pentasaccharide sequence in the specificity and strength of interaction. Journal of Biological Chemistry, 282, 2662–26675.

Makeneni S, Foley BL, & Woods RJ (2014). BFMP: A method for discretizing and visualizing pyranose conformations. Journal of Chemical Information and Modeling, 54(10), 2744–2750. doi:10.1021/ci500325b [PubMed: 25289680]

Mallajosyula SS, Guvench O, Hatcher E, & MacKerell AD (2012). CHARMM additive all-atom force field for phosphate and sulfate linked to carbohydrates. Journal of Chemical Theory and Computation, 8(2), 759–776. doi:10.1021/ct200792v [PubMed: 22685386]

Mandal F (1988). Statistical Physics, Second edition, Wiley, Chichester.

Maveyraud L, Niwa H, Guillet V, Svergun Dmitri I, Konarev Peter V, Palmer Rex A, Mourey L (2009). Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from Sambucus nigra. Proteins: Structure, Function, and Bioinformatics, 75(1), 89–103. doi: 10.1002/prot.22222

Maza S, Gandia-Aguado N, de Paz JL, Nieto PM, (2018). Fluorous-tag assisted synthesis of a glycosaminoglycan mimetic tetrasaccharide as a high-affinity FGF-2 and midkine ligand. *Bioorganic &* Medicinal Chemistry, 26(5), 1076–1085. doi: 10.1016/j.bmc.2018.01.022.

McDonald IK, & Thornton JM (1994). Satisfying hydrogen bonding potential in proteins. Journal of Molecular Biology, 238(5), 777–793. doi:10.1006/jmbi.1994.1334 [PubMed: 8182748]

Mende M, Bednarek C, Wawryszyn M, Sauter P, Biskup MB, Schepers U, & Bräse S (2016). Chemical synthesis of glycosaminoglycans. Chemical Reviews, 116(14), 8193–8255. doi:10.1021/acs.chemrev.6b00010 [PubMed: 27410264]

Michel G, Pojasek K, Li Y, Sulea T, Linhardt RJ, Raman R, Cygler M (2004). The structure of chondroitin B lyase complexed with glycosaminoglycan oligosaccharides unravels a calcium-dependent catalytic machinery. The Journal of Biological Chemistry, 279(31), 32882–32896. doi: 10.1074/jbc.M403421200 [PubMed: 15155751]

Millane RP, Mitra AK, & Arnott S (1983). Chondroitin 4-sulfate: Comparison of the structures of the potassium and sodium salts. Journal of Molecular Biology, 169(4), 903–920. doi:10.1016/S0022-2836(83)80142-9 [PubMed: 6415288]

Miyamoto S, & Kollman PA (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. Journal of Computational Chemistry, 13(8), 952–962. doi: 10.1002/jcc.540130805

Molecular Operating Environment (MOE), C. C. G. U., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018.

Molinero V, & Goddard WA (2004). M3B: A coarse grain force field for molecular simulations of malto-oligosaccharides and their water mixtures. The Journal of Physical Chemistry B, 108(4), 1414–1427. doi:10.1021/jp0354752

Monticelli L, & Tieleman DP (2013). Force Fields for Classical Molecular Dynamics In Monticelli L & Salonen E (Eds.), Biomolecular Simulations: Methods and Protocols (pp. 197–213). Totowa, NJ: Humana Press.

Mulloy B, Forster MJ, Jones C, & Davies DB (1993). N.m.r. and molecular-modelling studies of the solution conformation of heparin. Biochemical Journal, 293(3), 849. [PubMed: 8352752]

Mulloy B, Forster MJ, Jones C, Drake AF, Johnson EA, & Davies DB (1994). The effect of variation of substitution on the solution conformation of heparin: a spectroscopic and molecular modelling study. Carbohydrate Research, 255(Supplement C), 1–26. doi:10.1016/S0008-6215(00)90968-6 [PubMed: 8181000]

Muñoz-García JC, Corzana F, de Paz JL, Angulo J, & Nieto PM (2013). Conformations of the iduronate ring in short heparin fragments described by time-averaged distance restrained molecular dynamics. Glycobiology, 23(11), 1220–1229. doi:10.1093/glycob/cwt058 [PubMed: 23903025]

Muñoz-García JC, López-Prados J, Angulo J, Díaz-Contreras I, Reichardt N, de Paz JL, Nieto PM (2012). Effect of the substituents of the neighboring ring in the conformational equilibrium of iduronate in heparin-like trisaccharides. Chemistry – A European Journal, 18(51), 16319–16331. doi:10.1002/chem.201202770

Munoz-Garcia JC, Solera C, Carrero P, de Paz JL, Angulo J, & Nieto PM (2013). 3D structure of a heparin mimetic analogue of a FGF-1 activator. A NMR and molecular modelling study. Organic

& Biomolecular Chemistry, 11(47), 8269–8275. doi:10.1039/C3OB41789A [PubMed: 24178304]

Murphy KJ, McLay N, & Pye DA (2008). Structural studies of heparan sulfate hexasaccharides: New insights into iduronate conformational behavior. Journal of the American Chemical Society, 130(37), 12435–12444. doi:10.1021/ja802863p [PubMed: 18722428]

Nagarajan B, Sankaranarayanan NV, Patel BB, & Desai UR (2017). A molecular dynamics-based algorithm for evaluating the glycosaminoglycan mimicking potential of synthetic, homogenous, sulfated small molecules. PLoS ONE, 12(2), e0171619. doi:10.1371/journal.pone.0171619 [PubMed: 28182755]

Nagarajan M, & Rao VSR (1979). Conformational studies of proteoglycans: Theoretical studies on the conformation of heparin. Biopolymers, 18(6), 1407–1420. doi:10.1002/bip.1979.360180607 [PubMed: 465649]

Neamtu A, Tamba B, & Patras X (2013). Molecular dynamics simulations of chondroitin sulfate in explicit solvent: Point charge water models compared. Cellulose Chemistry and Technology., 47, 191−202.

Ng C, Nandha Premnath P, & Guvench O (2017). Rigidity and flexibility in the tetrasaccharide linker of proteoglycans from atomic-resolution molecular simulation. Journal of Computational Chemistry, 38(16), 1438–1446. doi:10.1002/jcc.24738 [PubMed: 28101951]

Nieduszynski IA, & Atkins EDT (1973). Conformation of the mucopolysaccharides. X-ray fibre diffraction of heparin. Biochemical Journal, 135(4), 729–733. [PubMed: 4273186]

Oborský P, Tvaroška I, Králová B, & Spiwok V (2013). Toward an accurate conformational modeling of iduronic acid. The Journal of Physical Chemistry B, 117(4), 1003–1009. doi:10.1021/jp3100552 [PubMed: 23286518]

O'Keeffe D, Olson ST, Gasiunas N, Gallagher J, Baglin TP, & Huntington JA (2004). The heparin binding properties of heparin cofactor II suggest an antithrombin-like activation mechanism. Journal of Biological Chemistry. 279*(48)*, 50267–2073. [PubMed: 15371417]

Patel NJ, Sharon C, Baranwal S, Boothello RS, Desai UR, Patel BB (2016). Heparan sulfate hexasaccharide selectively inhibits cancer stem cells self-renewal by activating p38 MAP kinase. Oncotarget 7(51), 84608–84622. doi: 10.18632/oncotarget.12358 [PubMed: 27705927]

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Schulten K (2005). Scalable molecular dynamics with NAMD. Journal of Computational Chemistry, 26(16), 1781–1802. doi: 10.1002/jcc.20289 [PubMed: 16222654]

Plazinski W, Lonardi A, & Hünenberger PH (2016). Revision of the GROMOS 56A6CARBO force field: Improving the description of ring-conformational equilibria in hexopyranose-based carbohydrates chains. Journal of Computational Chemistry, 37(3), 354–365. doi:10.1002/jcc.24229 [PubMed: 26525424]

Pol-Fachin L, & Verli H (2008). Depiction of the forces participating in the 2-O-sulfo-α-l-iduronic acid conformational preference in heparin sequences in aqueous solutions. Carbohydrate Research, 343(9), 1435–1445. doi:10.1016/j.carres.2008.04.016 [PubMed: 18452898]

Pol-Fachin L, Rusu VH, Verli H, & Lins RD (2012). GROMOS 53A6GLYC, an improved GROMOS force field for hexopyranose-based carbohydrates. Journal of Chemical Theory and Computation, 8(11), 4681–4690. doi:10.1021/ct300479h [PubMed: 26605624]

Pomin HV, & Mulloy B (2018). Glycosaminoglycans and Proteoglycans. *Pharmaceuticals*, 11(1). doi: 10.3390/ph11010027

Pomin VH (2015). Keratan sulfate: an up-to-date review. International Journal of Biological Macromolecules, 72, 282–289. doi:10.1016/j.ijbiomac.2014.08.029 [PubMed: 25179279]

Pomin VH, & Wang X (2018) Synthetic oligosaccharide libraries and microarray technology: A powerful combination for the success of current glycosaminoglycan interactomics. ChemMedChem. 13*(7)*, 648–661. [PubMed: 29160016]

Potenzone R Jr, & Hopfinger AJ (1978). Conformational analysis of glycosaminoglycans. III. Conformational properties of hyaluronic acid and sodium hyaluronate. Polymer Journal, 10, 181. doi:10.1295/polymj.10.181

Potenzone R, & Hopfinger AJ (1975). Conformational analysis of glycosaminoglycans. I. Charge distributions, torsional potentials, and steric maps. Carbohydrate Research, 40(2), 323–336. doi:10.1016/S0008-6215(00)82613-0

Ragazzi M, & Ferro DR (1997). Toward a realistic force field for the treatment of ionic sugars. Journal of Molecular Structure: THEOCHEM, 395–396(Supplement C), 107–122. doi:10.1016/S0166-1280(96)04955-X

Ragazzi M, Ferro DR, & Provasoli A (1986). A force-field study of the conformational characteristics of the iduronate ring. Journal of Computational Chemistry, 7(2), 105–112. doi:10.1002/jcc.540070203 [PubMed: 29160575]

Ragazzi M, Ferro DR, Perly B, Sinaÿ P, Petitou M, & Choay J (1990). Conformation of the pentasaccharide corresponding to the binding site of heparin for antithrombin III. Carbohydrate Research, 195(2), 169–185. doi:10.1016/0008-6215(90)84165-Q [PubMed: 2331700]

Ramachandran GN, Ramakrishnan C, & Sasisekharan V (1963). Stereochemistry of polypeptide chain configurations. Journal of Molecular Biology, 7(1), 95–99. doi:10.1016/S0022-2836(63)80023-6 [PubMed: 13990617]

Raman EP, Guvench O, & MacKerell AD (2010). CHARMM additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses. The Journal of Physical Chemistry B, 114(40), 12981–12994. doi:10.1021/jp105758h [PubMed: 20845956]

Rees DA (1969). Conformational analysis of polysaccharides. Part II. Alternating copolymers of the agar-carrageenan-chondroitin type by model building in the computer with calculation of helical parameters. Journal of the Chemical Society B: Physical Organic(0), 217–226. doi:10.1039/J29690000217

Rees DA, Morris ER, Stoddart JF, & Stevens ES (1985). Controversial glycosaminoglycan conformations. Nature, 317, 480. doi:10.1038/317480a0 [PubMed: 2995834]

Remko M, & von der Lieth C-W (2006). Gas-phase and solution conformations of selected dimeric structural units of heparin. Journal of Chemical Information and Modeling, 46(4), 1687–1694. doi:10.1021/ci060060+ [PubMed: 16859300]

Ricard-Blum S, & Lisacek F (2017). Glycosaminoglycanomics: where we are. Glycoconjugate Journal. 34*(3)*, 339–349. doi: 10.1007/s10719-016-9747-2. [PubMed: 27900575]

Rodríguez-Carvajal MA, Imberty A, & Pérez S (2003). Conformational behavior of chondroitin and chondroitin sulfate in relation to their physical properties as inferred by molecular modeling. Biopolymers, 69(1), 15–28. doi:10.1002/bip.10304 [PubMed: 12717719]

Rossky PJ, & Karplus M (1979). Solvation. A molecular dynamics study of a dipeptide in water. Journal of the American Chemical Society, 101(8), 1913–1937. doi:10.1021/ja00502a001

Salomon-Ferrer R, Case DA, & Walker RC (2013). An overview of the amber biomolecular simulation package. Wiley Interdisciplinary Reviews: Computational Molecular Sciences, 3(2), 198–210. doi:10.1002/wcms.1121

Samsonov SA, Bichmann L, & Pisabarro MT (2015). Coarse-grained model of glycosaminoglycans. Journal of Chemical Information and Modeling, 55(1), 114–124. doi:10.1021/ci500669w [PubMed: 25490039]

Samsonov SA, Gehrcke J-P, & Pisabarro MT (2014). Flexibility and explicit solvent in molecular-dynamics-based docking of protein–glycosaminoglycan systems. Journal of Chemical Information and Modeling, 54(2), 582–592. doi:10.1021/ci4006047 [PubMed: 24479827]

Samsonov SA, Teyra J, & Pisabarro MT (2011). Docking glycosaminoglycans to proteins: analysis of solvent inclusion. Journal of Computer-Aided Molecular Design, 25(5), 477–489. doi:10.1007/s10822-011-9433-1 [PubMed: 21597992]

Samsonov SA, Theisgen S, Riemer T, Huster D, & Pisabarro MT (2014). Glycosaminoglycan Monosaccharide Blocks Analysis by Quantum Mechanics, Molecular Dynamics, and Nuclear Magnetic Resonance (Vol. 2014).

Sankaranarayanan Nehru V, Strebel Tamara R, Boothello Rio S, Sheerin K, Raghuraman A, Sallas F, … Desai Umesh R (2017). A hexasaccharide containing rare 2-O-sulfate-glucuronic acid residues selectively activates heparin cofactor II. Angewandte Chemie International Edition, 56(9), 2312–2317. doi:10.1002/anie.201609541 [PubMed: 28124818]

Sankaranarayanan NV, & Desai UR (2014). Toward a robust computational screening strategy for identifying glycosaminoglycan sequences that display high specificity for target proteins. Glycobiology, 24(12), 1323–1333. doi:10.1093/glycob/cwu077 [PubMed: 25049239]

Sankaranarayanan NV, Nagarajan B, & Desai UR (2018). So you think computational approaches to understanding glycosaminoglycan–protein interactions are too dry and too rigid? Think again! Current Opinion in Structural Biology, 50, 91–100. doi:10.1016/j.sbi.2017.12.004 [PubMed: 29328962]

Sapay N, Cabannes E, Petitou M, & Imberty A (2011). Molecular modeling of the interaction between heparan sulfate and cellular growth factors: Bringing pieces together. Glycobiology, 21(9), 1181–1193. doi:10.1093/glycob/cwr052 [PubMed: 21572110]

Sarkar A, Yu W, Desai UR, MacKerell AD, & Mosier PD (2016). Estimating glycosaminoglycan–protein interaction affinity: water dominates the specific antithrombin–heparin interaction. Glycobiology, 26(10), 1041–1047. doi:10.1093/glycob/cww073 [PubMed: 27496757]

Sattelle BM, & Almond A (2010). Less is more when simulating unsulfated glycosaminoglycan 3D-structure: Comparison of GLYCAM06/TIP3P, PM3-CARB1/TIP3P, and SCC-DFTB-D/TIP3P predictions with experiment. Journal of Computational Chemistry, 31(16), 2932–2947. doi:10.1002/jcc.21589 [PubMed: 20564659]

Sattelle BM, & Almond A (2011). Is N-acetyl-d-glucosamine a rigid $^4C_1$ chair? Glycobiology, 21(12), 1651–1662. doi:10.1093/glycob/cwr101 [PubMed: 21807769]

Sattelle BM, Bose-Basu B, Tessier M, Woods RJ, Serianni AS, & Almond A (2012). Dependence of pyranose ring puckering on anomeric configuration: Methyl idopyranosides. The Journal of Physical Chemistry B, 116(22), 6380–6386. doi:10.1021/jp303183y [PubMed: 22577942]

Sattelle BM, Hansen SU, Gardiner J, & Almond A (2010). Free energy landscapes of iduronic acid and related monosaccharides. Journal of the American Chemical Society, 132(38), 13132–13134. doi:10.1021/ja1054143 [PubMed: 20809637]

Sattelle BM, Shakeri J, & Almond A (2013). Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? Biomacromolecules, 14(4), 1149–1159. doi:10.1021/bm400067g [PubMed: 23439078]

Sattelle BM, Shakeri J, Cliff MJ, & Almond A (2015). Proteoglycans and their heterogeneous glycosaminoglycans at the atomic scale. Biomacromolecules, 16(3), 951–961. doi:10.1021/bm5018386 [PubMed: 25645947]

Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Montgomery JA (1993). General atomic and molecular electronic structure system. Journal of Computational Chemistry, 14(11), 1347–1363. doi:10.1002/jcc.540141112

Schuler LD, & Van Gunsteren WF (2000). On the choice of dihedral angle potential energy functions for n-alkanes. Molecular Simulation, 25(5), 301–319. doi:10.1080/08927020008024504

Schuler LD, Daura X, & van Gunsteren WF (2001). An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. Journal of Computational Chemistry, 22(11), 1205–1218. doi:10.1002/jcc.1078

Schuttelkopf AW, & van Aalten DMF (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallographica Section D, 60(8), 1355–1363. doi:10.1107/S0907444904011679

Schworer R, Zubkova OV, Turnbull JE, Tyler PC (2013). Synthesis of a targeted library of heparan sulfate hexa- to dodecasaccharides as inhibitors of beta-secretase: potential therapeutics for Alzheimer's disease. Chemistry 19(21), 6817–6823. doi: 10.1002/chem.201204519 [PubMed: 23553710]

Scott JE (1995). Extracellular matrix, supramolecular organisation and shape. Journal of Anatomy, 187(Pt 2), 259–269. [PubMed: 7591990]

Scott JE, Chen Y, & Brass A (1992). Secondary and tertiary structures involving chondroitin and chondroitin sulphates in solution, investigated by rotary shadowing/electron microscopy and computer simulation. European Journal of Biochemistry, 209(2), 675–680. doi:10.1111/j.1432-1033.1992.tb17335.x [PubMed: 1425674]

Scott JE, Cummings C, Brass A, & Chen Y (1991). Secondary and tertiary structures of hyaluronan in aqueous solution, investigated by rotary shadowing-electron microscopy and computer

simulation. Hyaluronan is a very efficient network-forming polymer. Biochemical Journal, 274(3), 699. [PubMed: 2012600]

Sepuru KM, Nagarajan B, Desai UR, & Rajarathnam K (2016). Molecular basis of chemokine CXCL5-glycosaminoglycan interactions. Journal of Biological Chemistry. doi: 10.1074/jbc.M116.745265

Singh A, Tessier MB, Pederson K, Wang X, Venot AP, Boons G-J, Woods RJ (2016). Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. Canadian Journal of Chemistry, 94(11), 927–935. doi:10.1139/cjc-2015-0606 [PubMed: 28603292]

Sophia Fox AJ, Bedi A, & Rodeo SA (2009). The basic science of articular cartilage: Structure, composition, and function. Sports Health, 1(6), 461–468. doi:10.1177/1941738109350438 [PubMed: 23015907]

Sorin EJ, & Pande VS (2005). Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophysical Journal, 88(4), 2472–2493. doi:10.1529/biophysj.104.051938 [PubMed: 15665128]

Spiwok V, Králová B, & Tvaroska I (2010). Modelling of beta-D-glucopyranose ring distortion in different force fields: a metadynamics study. Carbohydrate Research, 345(4), 530–537. doi: 10.1016/j.carres.2009.12.011 [PubMed: 20053394]

Srinivas G, Cheng X, & Smith JC (2011). A solvent-free coarse grain model for crystalline and amorphous cellulose fibrils. Journal of Chemical Theory and Computation, 7(8), 2539–2548. doi: 10.1021/ct200181t [PubMed: 26606627]

Tvaroška I (1982). Theoretical studies on the conformation of saccharides. IV. Solvent effect on the stability of β-maltose conformers. Biopolymers, 21(9), 1887–1897. doi:10.1002/bip.360210915

van Gunsteren WF, & Berendsen HJC (1977). Algorithms for macromolecular dynamics and constraint dynamics. Molecular Physics, 34(5), 1311–1327. doi:10.1080/00268977700102571

Vanquelef E, Simon S, Marquant G, Garcia E, Klimerak G, Delepine JC, Dupradeau F-Y (2011). R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. Nucleic Acids Research, 39(Web Server issue), W511–W517. doi:10.1093/nar/gkr288 [PubMed: 21609950]

Varki A (2017). Biological roles of glycans. Glycobiology, 27(1), 3–49. doi:10.1093/glycob/cww086 [PubMed: 27558841]

Venkataraman G, Sasisekharan V, Cooney CL, Langer R, & Sasisekharan R (1994). A stereochemical approach to pyranose ring flexibility: its implications for the conformation of dermatan sulfate. Proceedings of the National Academy of Sciences, 91(13), 6171–6175.

Verli H, & Guimarães JA (2004). Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution. Carbohydrate Research, 339(2), 281–290. doi:10.1016/j.carres.2003.09.026 [PubMed: 14698886]

Verli H, & Guimarães JA (2005). Insights into the induced fit mechanism in antithrombin–heparin interaction using molecular dynamics simulations. Journal of Molecular Graphics and Modelling, 24(3), 203–212. doi:10.1016/j.jmgm.2005.07.002 [PubMed: 16146701]

Vieira TC, Costa-Filho A, Salgado NC, Allodi S, Valente AP, Nasciutti LE, & Silva LC (2004). Acharan sulfate, the new glycosaminoglycan from Achatina fulica Bowdich 1822. Structural heterogeneity, metabolic labeling and localization in the body, mucus and the organic shell matrix. Eur J Biochem, 271(4), 845–854. [PubMed: 14764101]

Vijayalakshmi KS, Yathindra N, & Rao VSR (1973). Theoretical studies on the conformations of methyl glycopyranosides. Carbohydrate Research, 31(2), 173–181. doi:10.1016/S0008-6215(00)86182-0

Wang J, Wolf RM, Caldwell JW, Kollman PA, & Case DA (2004). Development and testing of a general amber force field. Journal of Computational Chemistry, 25(9), 1157–1174. doi:10.1002/jcc.20035 [PubMed: 15116359]

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Weiner P (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. Journal of the American Chemical Society, 106(3), 765–784. doi:10.1021/ja00315a051

Wiegel D, Kaufmann J, & Arnold K (1999). Polar interactions of chondroitinsulfate: Surface free energy and molecular dynamics simulations. Colloids and Surfaces B: Biointerfaces, 13(3), 143–156. doi:10.1016/S0927-7765(98)00115-5

Winter WT, Arnott S, Isaac DH, & Atkins EDT (1978). Chondroitin 4-sulfate: The structure of a sulfated glycosaminoglycan. Journal of Molecular Biology, 125(1), 1–19. doi: 10.1016/0022-2836(78)90251-6 [PubMed: 712855]

Winter WT, Smith PJC, & Arnott S (1975). Hyaluronic acid: Structure of a fully extended 3-fold helical sodium salt and comparison with the less extended 4-fold helical forms. Journal of Molecular Biology, 99(2), 219–235. doi:10.1016/S0022-2836(75)80142-2 [PubMed: 1206703]

Woods RJ, Dwek RA, Edge CJ, & Fraser-Reid B (1995). Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. The Journal of Physical Chemistry, 99(11), 3832–3846. doi:10.1021/j100011a061

Xu D, & Esko JD (2014). Demystifying heparan sulfate–protein interactions. Annual Reviews in Biochemistry, 83(1), 129–157. doi:10.1146/annurev-biochem-060713-035314

Xu D, Arnold K, Liu J (2018). Using structurally defined oligosaccharides to understand the interactions between proteins and heparan sulfate. Current Opinion Structural Biology 50, 155–161. doi: 10.1016/j.sbi.2018.04.003

Xu R, McBride R, Paulson JC, Basler CF, & Wilson IA (2010). Structure, receptor binding, and antigenicity of influenza virus hemagglutinins from the 1957 H2N2 pandemic. Journal of Virology, 84(4), 1715–1721. doi: 10.1128/JVI.02162-09. [PubMed: 20007271]

Xu Y, Masuko S, Takieddin M, Xu H, Liu R, Jing J, Mousa SA, Linhardt RJ, Liu J (2011). Chemoenzymatic synthesis of homogeneous ultralow molecular weight heparins. Science, 334(6055), 498–501. doi: 10.1126/science.1207478 [PubMed: 22034431]

Yu F, Wolff JJ, Amster IJ, & Prestegard JH (2007). Conformational preferences of chondroitin sulfate oligomers using partially oriented NMR spectroscopy of 13c-labeled acetyl groups. Journal of the American Chemical Society, 129(43), 13288–13297. doi:10.1021/ja075272h [PubMed: 17924631]

Zhdanov YA, Minkin VI, Minjaev RM, Zacharov II, & Alexeev YE (1973). Quantum chemistry of carbohydrates: Part III. Calculation of the electronic distribution in some carbohydrate molecules by the CNDO/2 method. Carbohydrate Research, 29(2), 405–411. doi:10.1016/S0008-6215(00)83026-8

Zhong Y, Bauer BA, & Patel S (2011). Solvation properties of N-acetyl-β-glucosamine: A molecular dynamics study incorporating electrostatic polarization. Journal of Computational Chemistry, 32(16), 3339–3353. doi:10.1002/jcc.21873 [PubMed: 21898464]

Zsiška M, & Meyer B (1993). Influence of sulfate and carboxylate groups on the conformation of chondroitin sulfate related disaccharides. Carbohydrate Research, 243(2), 225–258. doi: 10.1016/0008-6215(93)87031-M. [PubMed: 8348541]

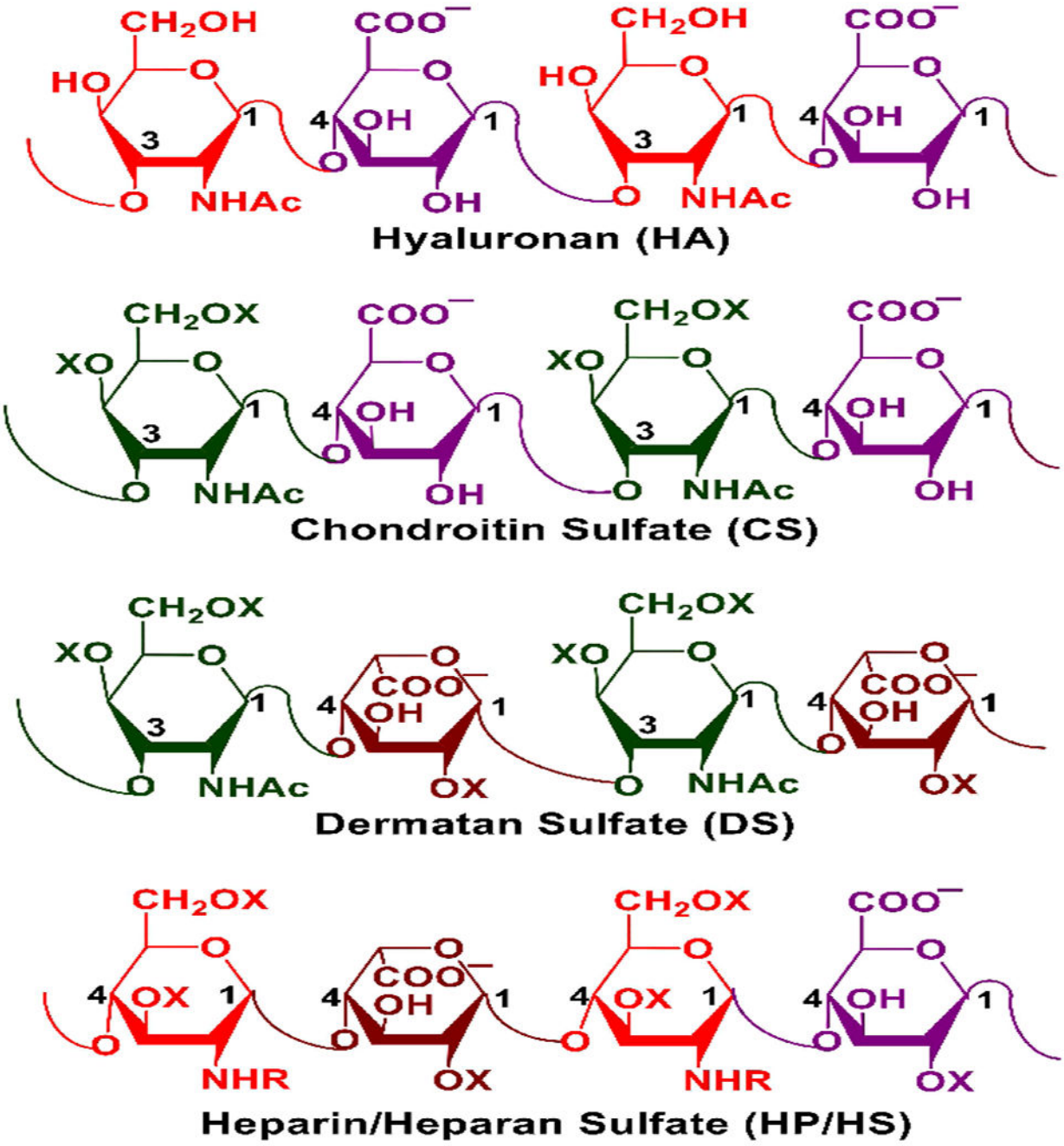


Figure 1: Structures of common glycosaminoglycans (GAGs).

Red - D-glucosamine (GlcN); Green - *N*-acetyl-D-galactosamine (GalNAc); Purple - D-glucuronic acid (GlcA); Brown – L-iduronic acid (IdoA). X = –H or $–SO_3^-$; R = –Ac or $–SO_3^-$ groups. HA has GlcNAc linked to GlcA by β(1→4) inter-glycosidic linkages and GlcA linked to GlcNAc through β(1→3) inter-glycosidic linkages. CS has GalNAc linked to GlcA by β(1→4) linkages and GlcA linked to GalNAc by β(1→3) inter-glycosidic bonds. DS's GalNAc is linked to IdoA by β(1–4) linkages, whereas IdoA is α(1→3) linked to GalNAc. In HP/HS, GlcN is α(1→4) linked to either GlcA or IdoA. In contrast, GlcA

residues are β(1→4) linked to GlcN residues, whereas IdoA is α(1→4) linked to GlcN residues.

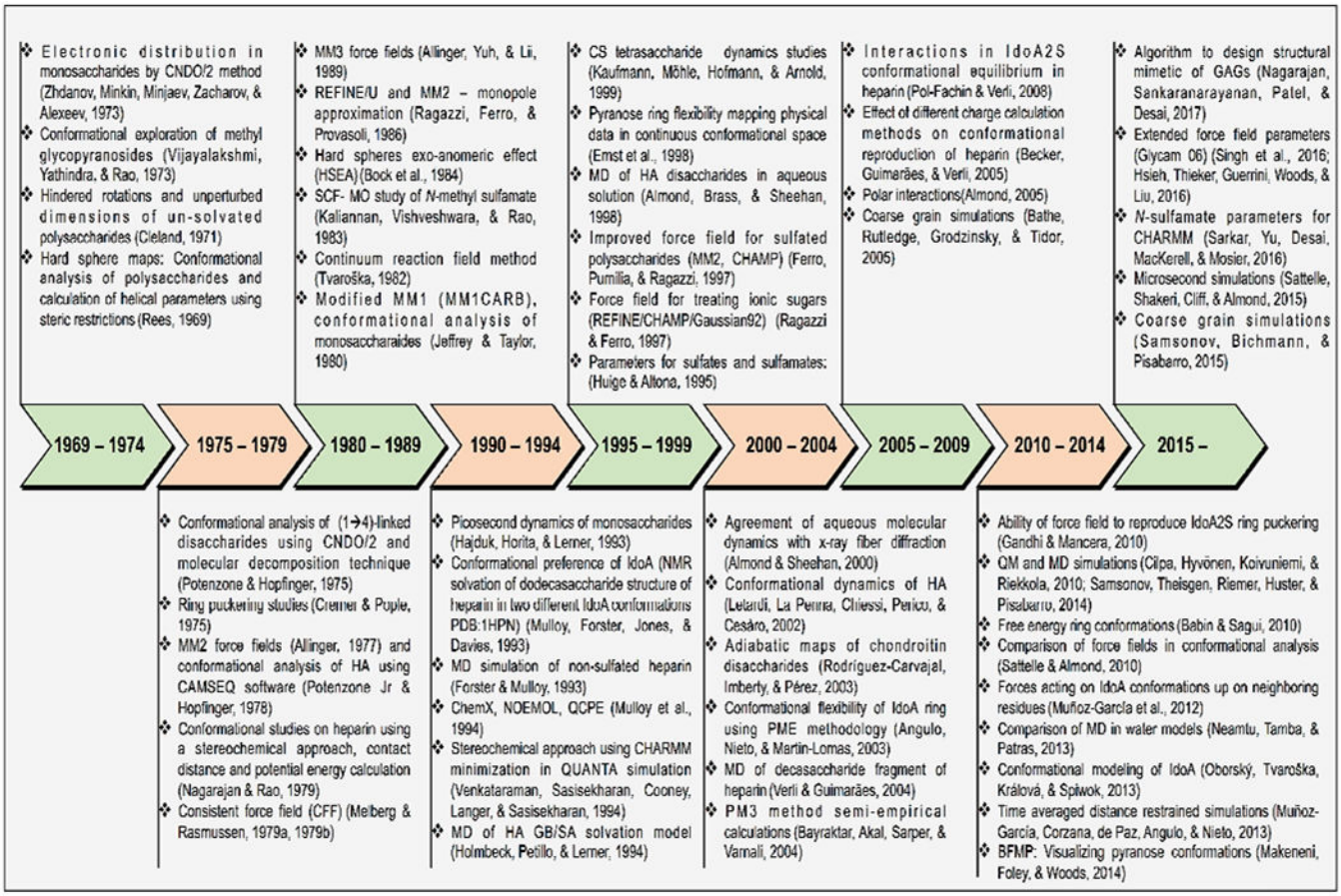


Figure 2: Time line of advances in computational studies of GAGs.
Key findings are listed in chronological order from left to right.

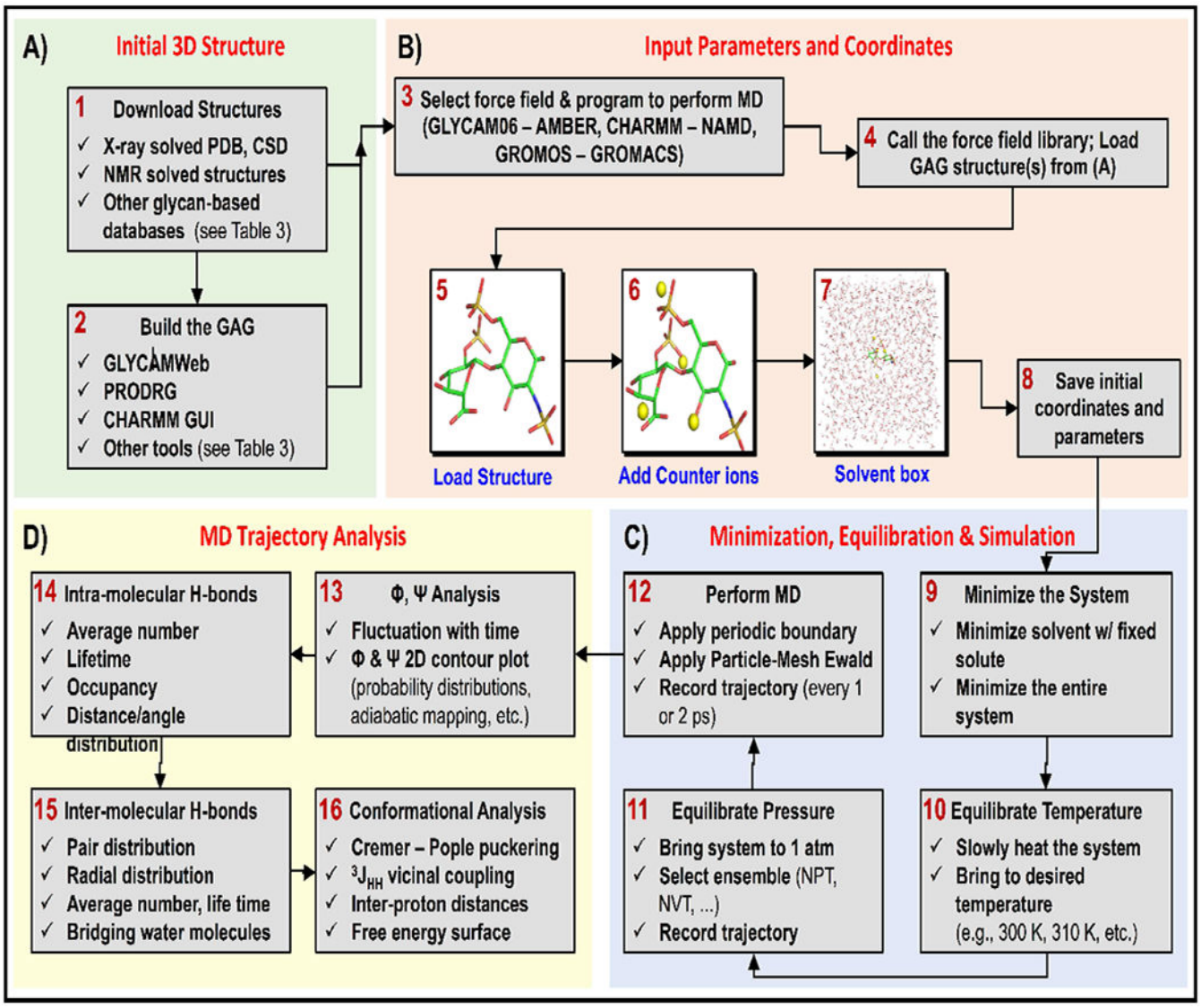


Figure 3: Schematic flow chart for performing MD simulations.

1) The atomic coordinates of GAG should be extracted from the Protein Data Bank (www.rcsb.org) or Cambridge Structural Data base (https://www.ccdc.cam.ac.uk) or from other structural data bases (details are provided in Table3). **2)** If the structure is not available in any of these databases, it can be built using 3-D model building tools (Table3). Once the 3-D structure is built, check for missing atoms and add hydrogens at appropriate places. Appropriate sugar puckering ($^1C_4$, $^4C_1$, ^2S_O) should be included, e.g. IdoA. **3)** After successfully building GAG structure, select the force field and relevant MD program to perform dynamic simulation (GLYCAM06-AMBER, CHARMM-NAMD/CHARMM, or GROMOS/GROMACS). **4**) Using the respective program, call the force field and load the structure obtained from 1 or 2. The loaded structure should be checked for any error and rectified. **5)** An example of a loaded structure is shown (e.g. IdoA2S(^2S_O)-GlcNS6S). **6)** Check the charge of the GAG sequence and add required number of counter ions (using respective programs like addion, genion and autoionize). **7)** Solvate the molecule in a pre-

equilibrated solvent box with defined box size such as cubic 12 Å (e.g., water model: TIP3P, SPC/SPCE). **8)** Once the GAG and ions are solvated save the initial parameter/topology and coordinates to perform further steps. Steps 9–12: A typical way of conducting molecular dynamics (MD) simulations. **9)** Energy minimize the system to remove steric clashes, if any, with defined protocols. **10)** After minimization slowly bring the system to desired temperature using thermostat by restraining the GAG. **11)** Apply constant pressure by coupling the system to a barostat; bring the system to NPT ensemble (different ensembles could be used NVT, NVE) and record outputs at constant time intervals. **12)** Once the system is well equilibrated, perform the final MD production run ranging from nano-seconds (ns) to micro-seconds (μs) of choice; record system trajectory at constant interval of time (e.g., for every 1, 2 or 10 ps). Apply periodic boundary condition and Particle –Meash – Ewald for long range interactions. Steps 13–16: Analyzing the recorded trajectory to extract information. **13)** Extract the inter-glycosidic torsional space and understanding their density/probability/distribution in 2D contour plot by binning the Φ & Ψ. **14)** Trace H-bond interactions within and in-between GAG residues by calculating their occupancy, average number and lifetimes. **15)** Evaluate caging of GAGs in water molecules with respect to time during the simulation using pair distribution function, average number, lifetimes and bridging water molecules. **16)** Analyze the individual puckering states using ring conformational analysis, Cremer-Pople parameters, theoretical and experimental vicinal coupling constants $^{3}J_{HH}$ and inter-proton distance using NOE and MD for adjacent residues. Free energy of conformational transition from one ring to other ring form can also be calculated.

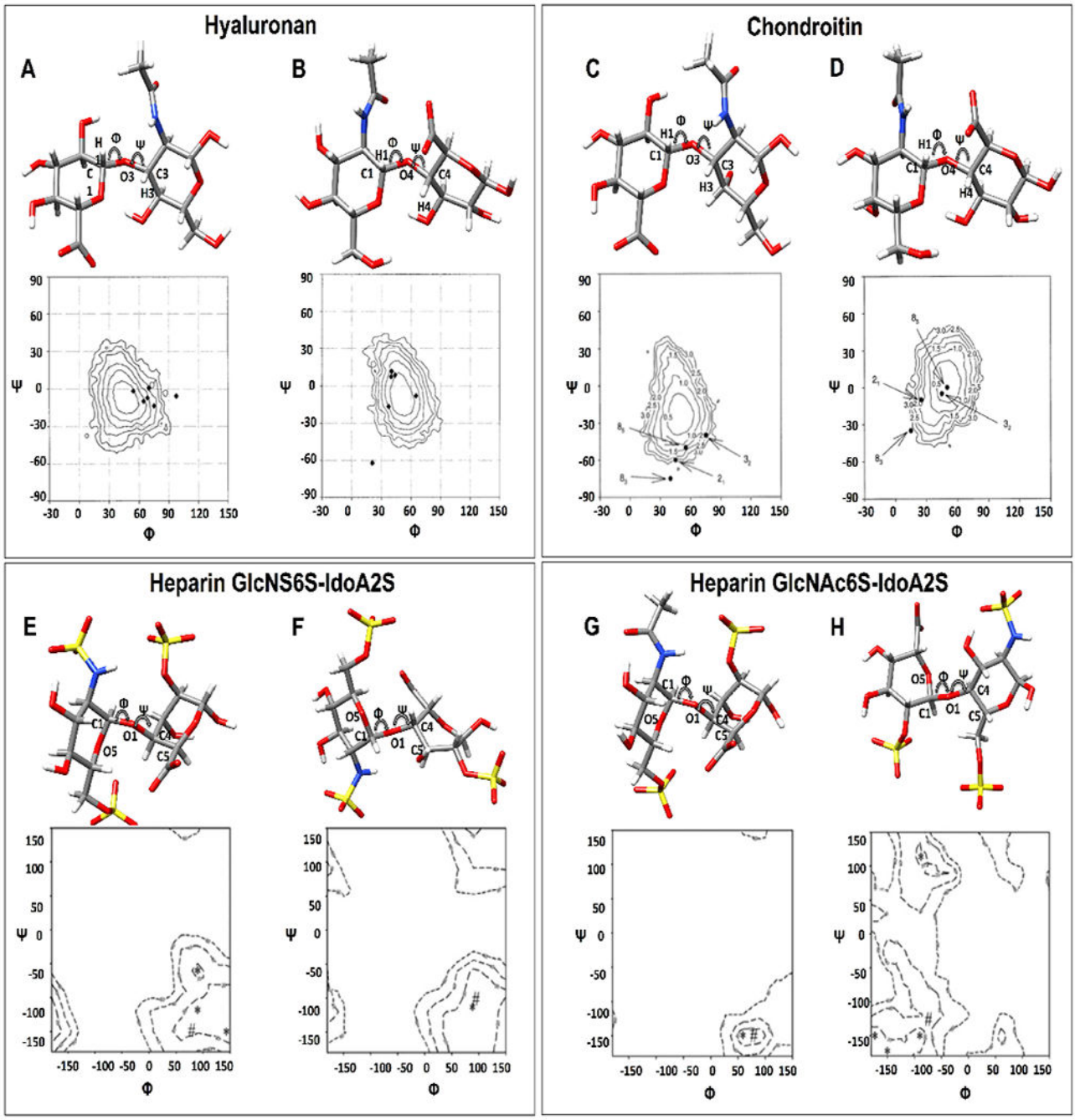


Figure 4: Representation of glycosidic torsion angles (Φ, Ψ).

A) and B) Energy contour exploration of glycosidic angles Φ,Ψ from alternate tetrasaccharide of hyaluronan (HA) shown for middle linkage, extracted from molecular dynamics simulation. The filled diamond boxes represent values of Φ, Ψ extracted from X-ray fiber diffraction refinements (Almond, Brass, & Sheehan, 1998). C) and D) Energy contour exploration of glycosidic angles Φ,Ψ from alternate tetrasaccharide of chondroitin shown for middle linkage, extracted from molecular dynamics simulation. The filled circle represents the value of Φ,Ψ with different helical folds from x-ray-fiber diffraction data

(Almond & Sheehan, 2000). E-H) Contour plots of heparin glycosidic linkages as a function on IdoA conformation. E) GlcNS6S(1→4)IdoA2S linkage shown with IdoA2S in $^{1}C_{4}$, F) with IdoA2S in $^{2}S_{O}$. G) GlcNAc6S(1→4)IdoA2S linkage shown with IdoA2S in $^{1}C_{4}$. H) IdoA2S(1→4)GlcNS6S linkage with IdoA in $^{1}C_{4}$, # represents the final conformation after 200ns and * represents the input minimum-energy conformation for MD refinement (Pol-Fachin & Verli, 2008). ***Figures to be reproduced after permission from respective publishers.***

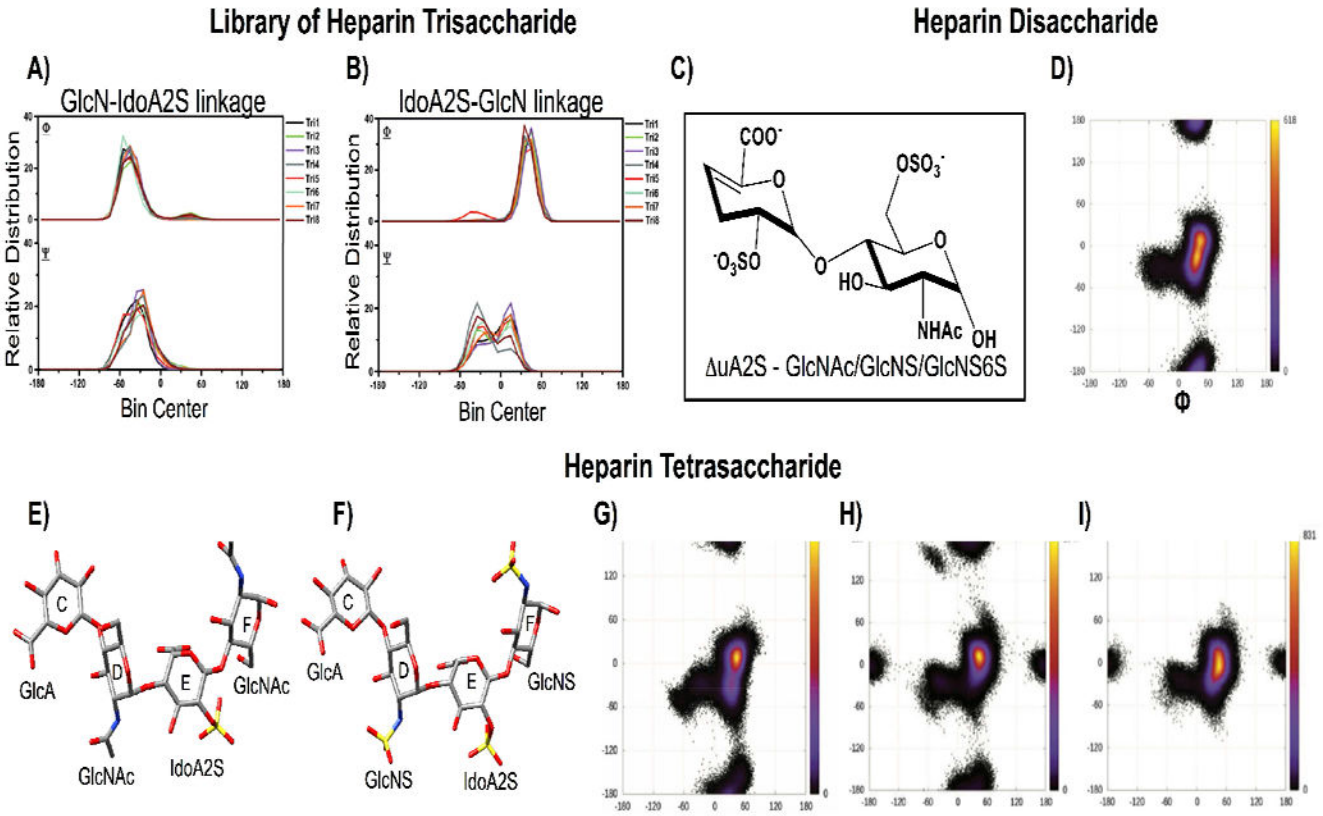


Figure 5: Representation of glycosidic torsion angles Φ,Ψ.

A and B) Glycosidic torsional angle Φ, Ψ distribution with respect to GlcN-IdoA2S and Ido2A-GlcN linkages for a library of 8 sequences (Muñoz-García, Corzana, de Paz, Angulo, & Nieto, 2013). C) Schematic structure of GAG disaccharide with Δ-uronic acid and GlcNAc with varying sulfation of total three sequences (1: ΔUA2S-GlcNAc6S, 2: ΔUA2S-GlcNS, 3: ΔUA2S-GlcNS6S). D) Glycosidic torsional angle Φ, Ψ represented as population distribution heat maps for ΔU2S-GlcNAc6S, with maximum conformers at Φ= 50° , Ψ=0°. E and F) Heparin tetra saccharide (E:GlcA-GlcNAc-IdoA2S-GlcNAc, F:GlcA-GlcNS-IdoA2S-GlcNS) with residue E IdoA2S and with varying sulfation in adjacent GlcNAc in E and F. G-I) Glycosidic torsional angle Φ,Ψ represented as population distribution heat maps for all possible disaccharide pairs were reported, here we show G) distribution of Φ,Ψ for the disaccharide (GlcA-GlcNS) in sequence E. H) Distribution of Φ, Ψ with regard to disaccharide pairs in the last IdoA2S-GlcNAc/GlcNS. I) Distribution of Φ, Ψ with regard to disaccharide pairs in the last IdoA2S-GlcNAc with IdoA2S restrained in $^{2}S_{O}$ form (Singh et al., 2016). ***Figures to be reproduced after permission from respective publishers.***

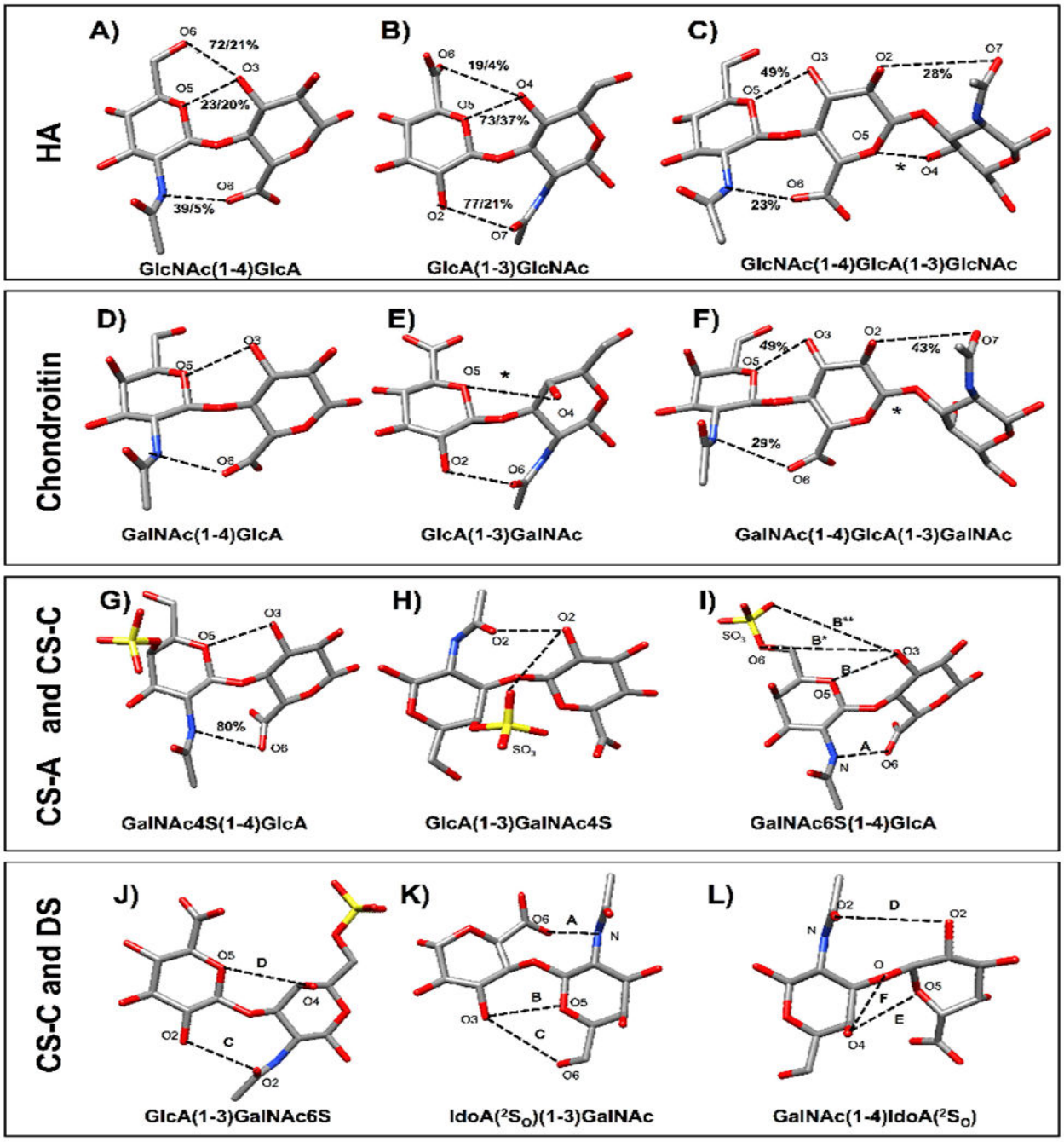


Figure 6: Intra-molecular H-bond interactions observed in HA and CS from various MD simulation studies.

A and B) Represents the H-bond interactions established between the possible donor-acceptor pairs for HA disaccharides. The occupancy in the presence of solvent/vacuum is marked respectively (Almond et al., 1998). C) HA tetrasaccharide – H-bond occupancy in the presence of solvent. D and E) unsulfated CS disaccharides: observed hydrogen bond interactions (Almond & Sheehan, 2000). F) Unsulfated CS tetrasaccharide in the presence of solvent: H-bond occupancy. G and H) Donor-acceptor interaction in CS-A (Kaufmann, Möhle, Hofmann, & Arnold, 1999; Neamtu, Tamba, & Patras, 2013). I and J) Hydrogen

bond interactions from CS-C (Cilpa, Hyvönen, Koivuniemi, & Riekkola, 2010). K and L) H-bond interactions from unsulfated DS with IdoA in $^{2}S_{O}$ puckering (Almond & Sheehan, 2000). Note: For all these the disaccharide pairs are shown for both 1–3 and 1–4 linkages with hydrogen bond donor and acceptor atoms labeled, and existence of hydrogen bonds shown in dashed line. The symbol * shows the difference in occurrence and B*, B** shows the first time appearance. ***Figures to be reproduced after permission from respective publishers.***

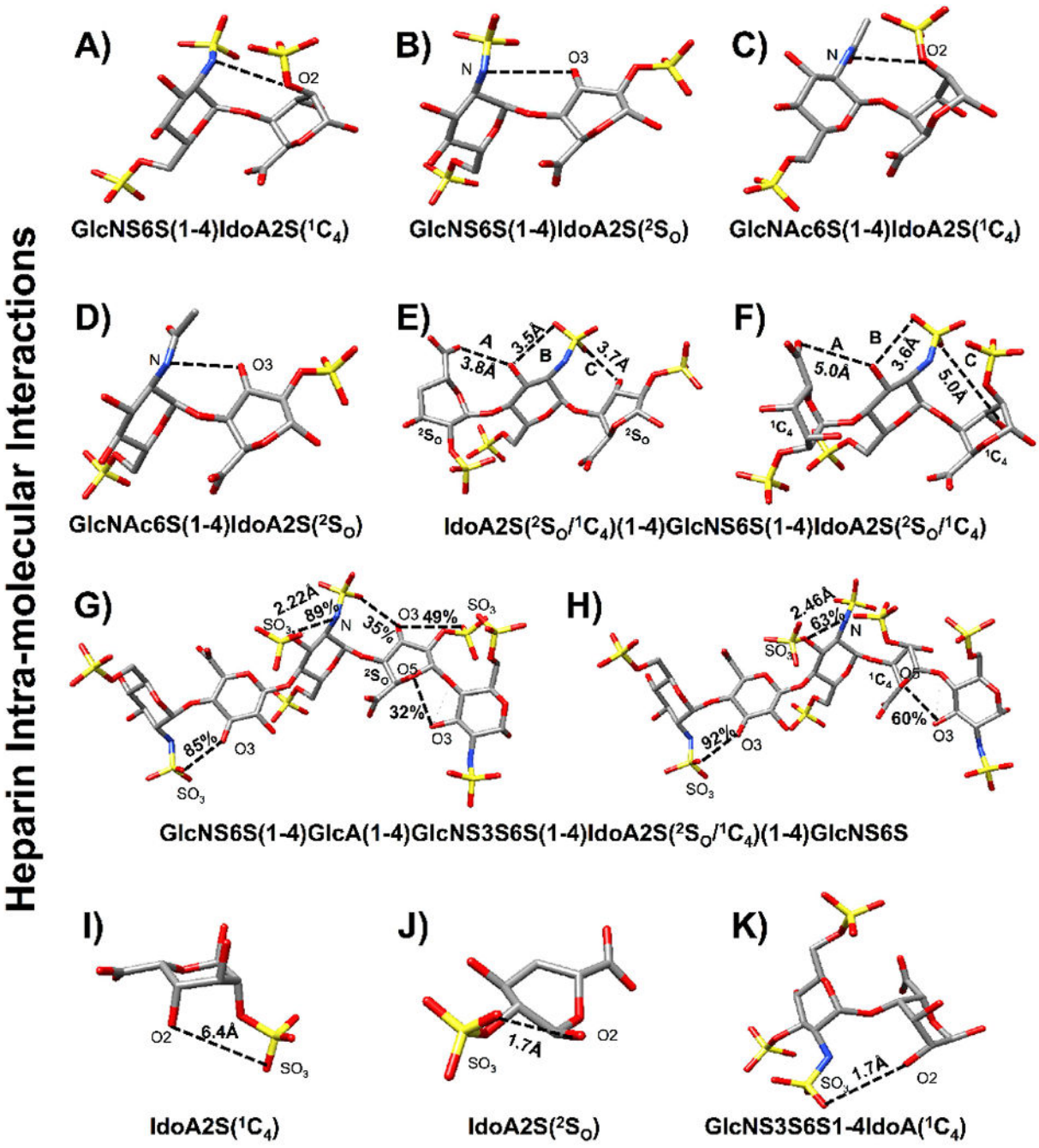


Figure 7: Intra-molecular H-bond interactions observed in HP from various MD simulations.

A and B) Represents the H-bond interactions from heparin disaccharides GlcNS6S-IdoA2S ($^1C_4$) and GlcNS6S-IdoA2S (^2S_O). C and D) The hydrogen bond interactions from heparin disaccharides GlcNAc6S-IdoA2S ($^1C_4$) and GlcNAc6S-IdoA2S(^2S_O) (Pol-Fachin & Verli, 2008). E and F) The hydrogen bond interactions from the sequence $(GlcNS6S\text{-}IdoA2s)_5$ with IdoA2S in both ^2S_O and $^1C_4$ forms (Verli & Guimarães, 2004). G and H) Various observed stabilizing intra molecular hydrogen bond present in pentasaccharide-Arixtra with the participating IdoA2S in both ^2S_O and $^1C_4$ forms (Beecher, Young, Langeslay, Mueller, &

Larive, 2014; Langeslay et al., 2012). I-K) Hydrogen bond distance deviation observed in the individual residue of IdoA in $^{2}S_{O}$ and $^{1}C_{4}$ forms and in between GlcNS3S6S-IdoA ($^{1}C_{4}$). All donor-acceptor atoms and distances are labeled, hydrogen bond interaction shown in dashed line (Hsieh, Thieker, Guerrini, Woods, & Liu, 2016). ***Figures to be reproduced after permission from respective publishers.***

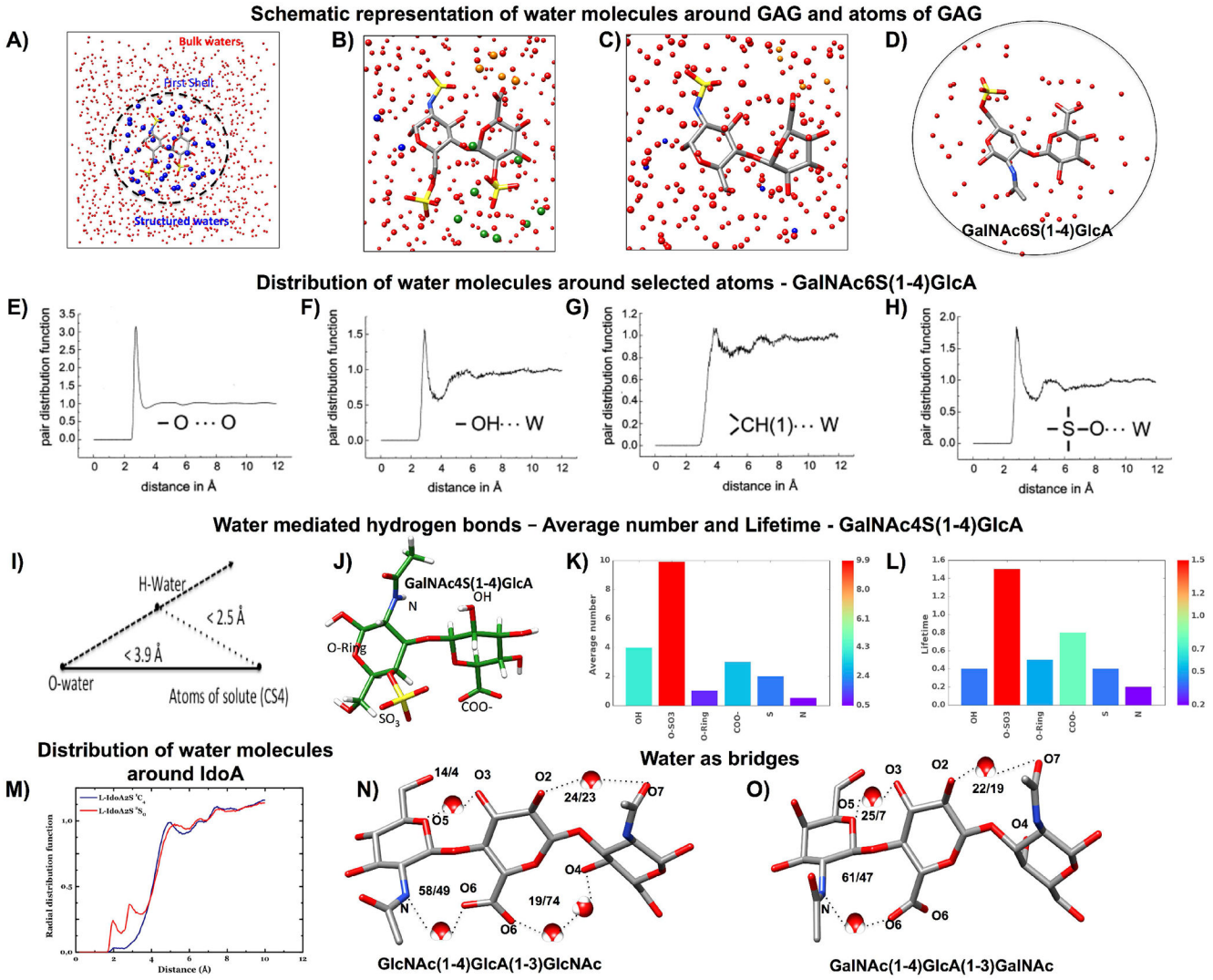


Figure 8: Inter-molecular (water) interactions in GAG.

A) Schematic representation of structured water (water molecules shown in blue spheres) around GAG disaccharide and bulk water molecule, which is not close to GAG (shown in red spheres). B and C) Representation of water molecules close to different polar groups of disaccharide, orange sphere shows water molecules around carboxyl group, Green sphere water molecules shown around sulfate group and blue color water spheres shown around one of the hydroxyl group. D) Shows the structured water molecule around 1–4 linked CS-C (from A-D only oxygen atom of water alone shown for better visibility). E-H) shows the radial distribution of water in CS-C with respect to bulk water-water, hydroxyl group –water, CH –water (Wiegel, Kaufmann, & Arnold, 1999). I) schematic representation of water mediated H-bond interaction. J) Structure of 1–4 CS-A. K) Average number of water mediated H-bonds with respect to each donor and acceptor group in CS-A. L) Life time of established H-bonds between CS-A and water molecules (Kaufmann et al., 1999) (color from blue to red shows the increase in strength in both K and L). M) Radial distribution of water molecules around heparin trisaccharide with I doA2S in $^{2}S_{O}$ in red and $^{1}C_{4}$ in blue

(Muñoz-García et al., 2013). N) Bridging water molecular interaction in HA (one water molecular bridging/two water molecular bridging). O) Bridging water molecular interaction in unsulfated chondroitin (one water molecular bridging/two water molecular bridging) (Almond, 2005). ***Figures to be reproduced after permission from respective publishers.***

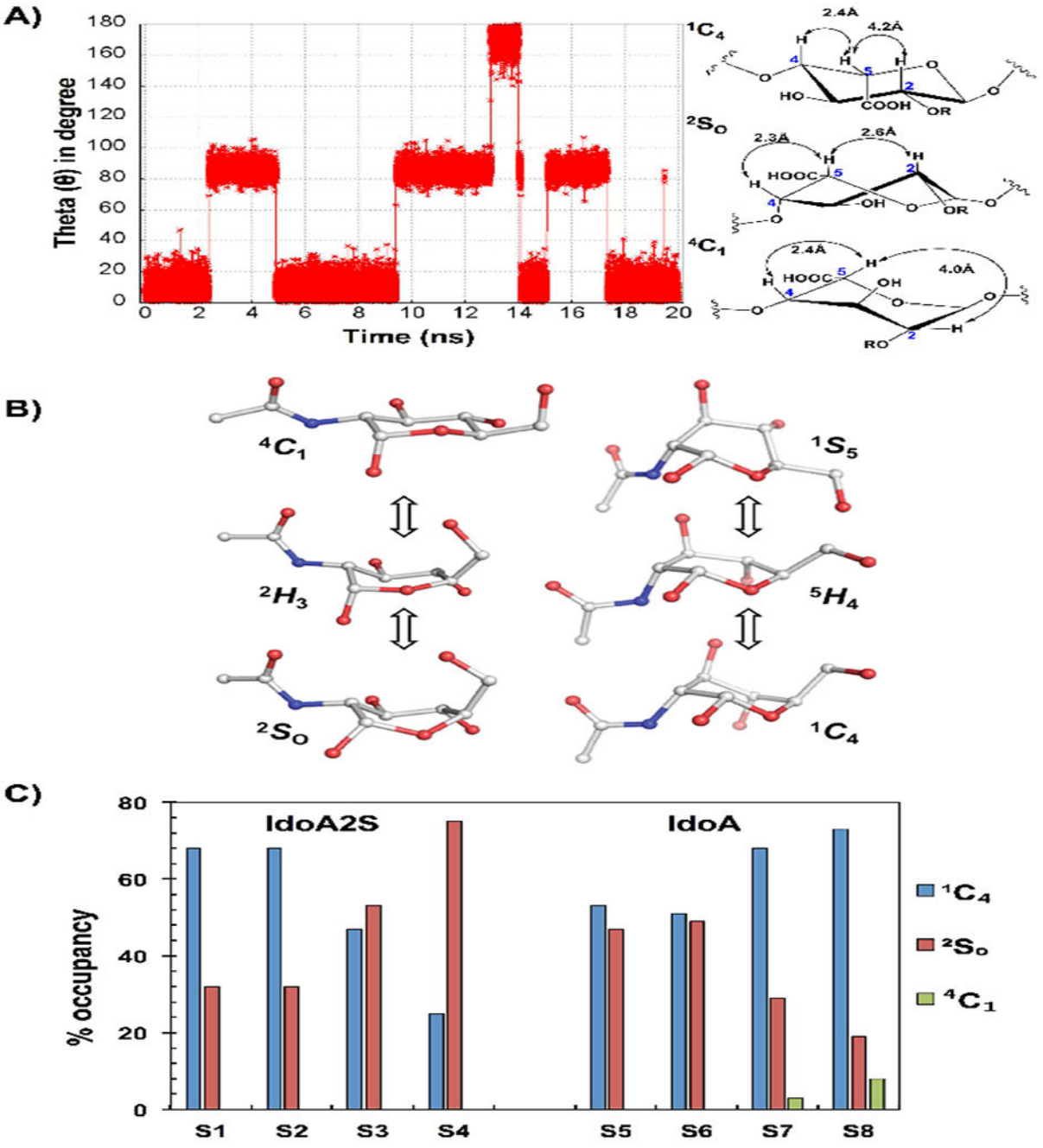


Figure 9: Ring conformations of GAG.

A) Schematic representation of Cremer-Pople puckering parameter theta (θ) for IdoA ring shown for $^4C_1$, θ = 0° to 60°, ^2S_O, θ = 60° to 120°, $^1C_4$, θ = 120° to 180°. The structures are shown in right hand side of A with representations for observed NOE cross peaks from Hsieh et al 2016. From this the vicinal coupling based on four torsion is also shown (e.g. H2-C2-C3-H3). B) The transition path way of ring conformations from chair to skew boat and then to chair observed in microsecond simulation for GlcNAc (Sattelle & Almond, 2011). C) The population of ring conformers obtained for IdoA2S and IdoA in the library of

8 hexasaccharides sequences (S1 to S8) with variation in neighboring residue sulfation are shown. From left to right it shows the increase of sulfation in the sequence (Hsieh et al., 2016). ***Figures to be reproduced after permission from respective publishers.***

Table 1:

Development of MD simulation methodology for glycans and glycosaminoglycans.

Molecule(s)	Force field/Software/Water model	Aim	Method	Reference
Mono- and di- saccharides	GROMOS 45A4 / GROMACS / SPC	Parameterize force field and validate through simulations and comparison with experimental observations	QM geometry optimization, puckering restrained/unrestrained simulations	(Lins and Hunenberger 2005)
16 mono- and 8 di- saccharides	GROMOS53A6$_{GLYC}$ / GROMACS / SPC	Overcome biased stability and improve the parameters over GROMOS 45A4	Metadynamics and unrestrained dynamics	(Pol-Fachin, Rusu et al. 2012)
Pyranoses and collection of inter-glycosidic linkages	CHARMM additive all-atom force field / CHARMM / TIP3P	Derive parameters for all possible chiralities and alpha / beta anomers	QM geometry optimization and vibrational calculation using Gaussian03 (MP2/6–319G(d)) / MD	(Guvench, Hatcher et al. 2009)
Furanose and collection of inter-glycosidic linkages	CHARMM additive all-atom force field / CHARMM / TIP3P	Derive parameters for pyranose–furanose and furanose–furanose inter-glycosidic linkages	QM geometry optimization and vibrational calculation using Gaussian03 (MP2/6–319G(d)) / MD	(Raman, Guvench et al. 2010)
Pyranoses and furanoses, aldoses and ketoses and inter-glycosidic linkages	CHARMM additive all-atom force field / CHARMM-NAMD / TIP3P	Derive parameters and validate using NMR solution properties	QM geometry optimization and vibrational calculation using Gaussian03 (MP2/6–319G(d)) / MD	(Guvench, Mallajosyula et al. 2011, Sarkar, Yu et al. 2016)
Sulfamate-linked glycans	CHARMM additive all-atom force field / CHARMM / TIP3P	Develop parameters for sulfamate linked to carbohydrates (e.g., GlcNS)	QM geometry optimization and single point energy calculation	(Sarkar, Yu et al. 2016)
Carbohydrates	GLYCAM_93 parameters compatible with AMBER force field / TIP3P	Develop a parameter set consistent with all-atom AMBERR force field for glycoproteins and oligosaccharides	Molecular orbital calculations using Gaussian 90 and 92; geometry optimization by Hartree-Fock with 6–31G* level / MM calculations using AMBER4.0	(Woods, Dwek et al. 1995)
Carbohydrates	GLYCAM_2000 parameters compatible with AMBER force field / TIP3P	Develop parameters for 1→6 linked glycans and validate the ω-angle preferences to experimental data	Molecular orbital calculations using Gaussian 94; geomentry optimization by HF 6–31G(d) level / MM calculations using AMBER5.0	(Kirschner and Woods 2001)
Carbohydrates	GLYCAM_06 parameters compatible with AMBER force field / TIP3P	Develop parameters for all possible carbohydrates based on seven critical factors & rectify errors on hydration shell from previous force fields	QM calculations using Gaussian 98 with HF/6–31G* optimization /MM by AMBER 7/8 force fields	(Kirschner, Yongye et al. 2008)
Carbohydrates	GLYCAM_06 parameters compatible with AMBER force field / TIP3P	Develop parameters to model *N*- and *O*-sulfation, ΔUA residues, neutral and charged GlcN & IdoA	QM calculations using Gaussian 09 with HF/6–31++G** diffuse functions / MM by AMBER 7/8 force fields	(Singh, Tessier et al. 2016)

Table 2.

Force fields and types of GAG sequences studied to date.

Force field	Molecular description
GLYCAM	
Glycam93	Library of 8 heparin trisaccharides carrying variable sulfation in IdoA & GlcNAc residues
	Tetrasaccharides with GlcA, GalNAc, or GlcNAc residues
	Hexasaccharides with IdoA2S, GlcNS6S, IdoA, & GlcNAc residues
	Synthetic hexasaccharide carrying IdoA2S and GalNS6S residues
Glycam06	Library of 8 heparin trisaccharides carrying variable sulfation in IdoA & GlcNAc residues
	Monosaccharides with variable puckering in IdoA, IdoA2S and GlcA residues
	2-*O*-Sulfo-α-*L*-IdoA monosaccharide
	Methyl-α-*L*-IdoA & methyl-β-*D*-Glc monosaccharides
	Disaccharides carrying GlcA, GalNAc, GalNAc4S, or GlcNAc residues
	Hexasaccharide carring IdoA2S and GlcNS6S residues
	β-*D*-Glc monosaccharide
	Arixtra (antithrombin binding heparin pentasaccharide)
	α-*L*-IdoA-OMe and α-*L*-IdoA2S-OMe monosaccharides
	Octasaccharide with GlcA and GalNAc4S residues
	Tetra- and hexa- saccharide containing GlcA, GalNAc, GlcNAc, and GalNAc4S residues
	Sulfated analogs of GlcNAc, GalNAc, IdoUA, and GlcUA
	Di- and tetra- saccharides containing ΔUA and sulfated or non-sulfated variants of uronic acid and glucosamine residues
	8 Hexasaccharides carrying variably sulfated IdoA and GlcNAc residues
Glycam06h	5 Hexasaccharides carrying variably sulfate IdoA and GlcNAc residues
Glycam06g	GlcNAc, GlcNS, GlcNS6S, GlcNS6S3S and GlcNS3S monosaccharides as free α-anomers and methylglycosides
	Arixtra (antithrombin binding heparin pentasaccharide)

Force field	Molecular description
Gycam11	Methyl α- and β- *L*-IdoA monosaccharides
CHARMM21A	Tetrasaccharide containing GlcA, GalNAc4S, and GalNAc6S residues
CHARMM 20	GlcA and GlcNAc monosaccharides
CHARMM22	Two HA disaccharides with GlcA and GlcNAc residues
	Two HA tetrasaccharides with GlcA and GlcNAc residues
	HA tetrasaccharides with GlcA and GlcNAc residues
CHARMM 24b2	Tetrasaccharide with GlcA and GalNAc6S residues
	Tetrasaccharide with GlcA and GalNAc6S residues
CHARMM25	HA, DS, and chondroitin tetrasaccharides
CHARMMC36	CS disaccharides
CHARMM 27a2	HA and CS disaccharides
CHARMM36	Linker tetrasaccharide connecting GAG to protein in proteoglycans
GROMACS	Decasaccharide carrying IdoA2S and GlcNS6S residues
	Di- and tri- saccharides of HA type
	Decasaccharide carrying IdoA2S and GlcNS6S residues
GROMOS96	Trisaccharide carrying IdoA2S, GlcNS6S, and GlcNAc residues
	α-*L*-IdoA2S monosaccharide
	Dodecasaccharide carrying GlcA and GalNAc6S residues
OPLS AA	β-*D*-Glc monosaccharide
	HA oligosaccharides
	β-*D*-Glc monosaccharide
SPASIBA	Mono and disaccharide chondroitin sulfate
Dreiding	Tetrasaccharide containing GlcA, GalNAc, GalNAc4S, GalNAc6S, and GalNAc4S6S residues

Force field	Molecular description
MM3	Methyl-*O*-(4-*O*-methyl-2,3,6-tri-*O*-sulfo-β-D-glucopyranosyl-(1→4)-*O*-(2-*O*-sulfo-α-*L*-idopyranosyluronic acid)-(1→4)-*O*-2,6-di-*O*-sulfo-α-D-glucopyra
MM3	Octasaccharides carrying GlcA, GalNAc, GalNAc4S, and GalNAc6S residues
	Disaccharides carrying GlcA, GalNAc, GalNAc4S, and GalNAc6S residues
MM3*	HA octasaccharide
AMBER:GAFF	Hexasaccharide carrying IdoA2S, GlcNAc, GlcNS, IdoA, ΔUA, and ΔUA2S residues
AMBER FF	HA mono- and di- saccharides

Table 3.

Useful sites, resources and programs for computational simulation of glycosaminoglycans.

A) Databases

Database—web-page	Further reading
PDB site—https://www.rcsb.org	(Berman et al., 2000)
CSD—https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/	(Groom, Bruno, Lightfoot, & Ward, 2016)
Glyco3D—http://glyco3d.cermav.cnrs.fr/search.php?type=gag	(Pérez, Sarkar, Rivet, Breton, & Imberty, 2015)
Glycosciences.de—http://www.glycosciences.de/index.php	(Lütteke et al., 2006)
KEGG Glycan database—http://www.genome.jp/kegg/glycan/	(Hashimoto et al., 2006)
GlycomeDB—http://www.glycome-db.org	(Ranzinger, Herget, von der Lieth, & Frank, 2011)
Japan Consortium for Glycobiology and Glycotechnology—http://jcggdb.jp/index.html	(Maeda et al., 2015)

B) Tools to build a GAG

Database—web-page	Further reading
PRODRG Server—http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg	(Schuttelkopf & van Aalten, 2004)
GLYCAM—http://glycam.org	(Woods Group, 2005–2017)
CHARMM-GUI—http://www.charmm-gui.org	(Jo, Kim, Iyer, & Im, 2008; Lee et al., 2016)
SYBYLX—https://www.certara.com/pressreleases/certara-enhances-sybyl-x-drug-design-and-discovery-software-suite/	(Sybyl-X Molecular Modeling Software Packages)
Discover studio—http://accelrys.com/products/collaborative-science/biovia-discovery-studio/	(Dassault Systèmes BIOVIA)
Molecular Operating Environment—https://www.chemcomp.com/MOE-Molecular_Modeling_and_Simulations.htm	(Molecular Operating Environment (MOE))
VEGA ZZ—http://www.vegazz.net	(Pedretti, Villa, & Vistoli, 2004)
CVLS—Libraries di- to hexa- saccharides (available upon request … urdesai@vcu.edu)	Angew. Chem. Int. Ed. Engl. (2017)
Maestro—https://www.schrodinger.com/maestro	(Schrödinger Release 2017–4: Maestro)

C) Force fields

Force field—web-page	Further reading
GLYCAM—http://glycam.org/docs/forcefield/	(Kirschner et al., 2008)
CHARMM—http://mackerell.umaryland.edu/charmm_ff.shtml	(Mallajosyula, Guvench, Hatcher, & MacKerell, 2012)
GROMOS—http://www.gromacs.org/Downloads/User_contributions/Force_fields	(Schuler, Daura, & van Gunsteren, 2001)
OPLSAA—http://onlinelibrary.wiley.com/doi/10.1002/jcc.10139/abstract	(Damm et al., 1997, Kony et al., 2002)
Dreiding—https://authors.library.caltech.edu/24133/	(Mayo, Olafson, & Goddard, 1990)

D) Parameters and topology files

Program—web-page	Further reading
Glycam—http://glycam.org/docs/forcefield/all-parameters/GLYCAM_06j-1	(Kirschner et al., 2008)
CHARMM—http://mackerell.umaryland.edu/charmm_ff.shtml	(Mallajosyula et al., 2012)
Gromacs—http://www.gromacs.org/Downloads/User_contributions/Force_fields	(Hess, Kutzner, van der Spoel, & Lindahl, 2008)

D) Parameters and topology files

Program—web-page	Further reading
Antechamber tool—http://ambermd.org/tutorials/basic/tutorial4b/	(Wang, Wang, Kollman, & Case, 2006)
Gromacs topology—Topology conversion in Gromacs amb2gmx.pl	(Sorin & Pande, 2005)

E) Programs to perform MD simulations

Program—web-page	Further reading
AMBER—http://ambermd.org	(Salomon-Ferrer, Case, & Walker, 2013)
NAMD—http://www.ks.uiuc.edu/Research/namd/	(Phillips et al., 2005)
GROMACS—http://www.gromacs.org	(Hess et al., 2008)
CHARMM—https://www.charmm.org/charmm/	(Brooks et al., 2009)
ACEMD—https://www.acellera.com	(Harvey, Giupponi, & Fabritiis, 2009)

F) Water models

Model	Description	Further Reading
TIP3P	Three point interaction model (Transferable intermolecular potential 3P)	(Jorgensen et al., 1983, Neria and karplus, 1996, Durell et al., 1994)
SPC	Simple point charge	(Berendsen et al., 1981, Berweger, et al., 1995)
SPC/E	Extended simple point charge	(Berendsen et al., 1987)
TIP4P	Four point interaction model	(Jorgensen et al., 1983)
TIP4P-EW	Four point water model with Ewald technique	(Horn et al., 2004)
TIP5P	Five Point interaction model	(Mahoney & Jorgensen, 2000)

G) Tools for analysis of simulations

Tool—web-page	Further Reading
Amber Tools—http://ambermd.org/#AmberTools	(Roe & Cheatham, 2013)
VMD—http://www.ks.uiuc.edu/Research/vmd/	(Humphrey, Dalke, & Schulten, 1996)
BFMP—http://glycam.org/docs/othertoolsservice/download-docs/publication-materials/bfmp/	(Makeneni, Foley, & Woods, 2014)
Gromacs g_puckering—http://manual.gromacs.org/programs/bytopic.html	(Hess et al., 2008)
Chimera—https://www.cgl.ucsf.edu/chimera/	(Pettersen et al., 2004)