Highlights

- Methods established for modeling oligosaccharides, polysaccharides and protein interactions
- Molecular mechanics force-fields now extended to most glycosaminoglycans
- Realistic dynamic predictions from enhanced sampling and microsecond simulations
- Docking protocols for predicting interactions, binding sites and poses
- Coarse graining an emerging route to macroscopic experimental interpretation
Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges.

Andrew Almond*

*School of Chemistry,
The University of Manchester,
Manchester Institute of Biotechnology,
131 Princess Street,
Manchester,
M1 7DN, UK.
Email: Andrew.Almond@manchester.ac.uk
Tel: +44 161 30 64199
Highlights

- Methods established for modeling oligosaccharides, polysaccharides and protein interactions
- Molecular mechanics force-fields now extended to most glycosaminoglycans
- Realistic dynamic predictions from enhanced sampling and microsecond simulations
- Docking protocols for predicting interactions, binding sites and poses
- Coarse graining an emerging route to macroscopic experimental interpretation

Abstract

Glycosaminoglycans are long unbranched and complex polysaccharides that are an essential component of mammalian extracellular matrices. Characterization of their molecular structure, dynamics and interactions is essential to understand important biological phenomena in health and disease, and will lead to novel therapeutics and medical devices. However, this has proven to be a challenge experimentally and theoretical techniques are needed to develop new hypotheses, and interpret experiments. This review aims to examine the current theoretical (rather than experimental) methods used by researchers to investigate glycosaminoglycan structure, dynamics and interactions, from the monosaccharide to the macromolecular scale. It will consider techniques such as quantum mechanics, molecular mechanics, molecular dynamics, coarse graining and docking.

Introduction

Proteoglycans (PGs) are an enigmatic family of structural macromolecules secreted by virtually all mammalian cells [1]. They comprise a protein core (one of at least 43), decorated with the glycosaminoglycans (GAGs): chondroitin sulfate (CS), dermatan sulfate (DS), heparan sulfate (HS) and keratin sulfate (KS). There are two other functionally-different GAGs: heparin, secreted by mast cells [2] and hyaluronan (HA), synthesized at the plasma membrane without covalent attachment to protein [3]. The GAGs are unbranched polymeric repeats of hexosamine and uronic acid (or galactose in KS) sugars. Not being template driven, numerous glycoforms exist in CS and HS, mainly in terms of sulfation and variability in glucuronic acid (GlcA) and iduronic acid (IdoA) content. Heparin is uniformly sulfated and with a high IdoA content, while HA is exclusively GlcA and no sulfate. Although much of their effect is structural, they also exert their influence via interaction with proteins [4].

Studies of GAG 3D-conformation and dynamics are difficult to perform experimentally, and computer modeling is frequently needed, in part due to their repetitive chemistry and flexibility in solution. In general, theoretical methods used to study carbohydrates are applicable to GAGs, but with additional complications. The GAGs are massive, have significant sequence variability, and several chains can be linked to large proteins. Additionally, IdoA has unstable ring puckering on microsecond timescales and as carbohydrates, the GAGs have a strong interaction with water. Furthermore, they are also highly negatively charged, which requires theoretical consideration of their protonation state and potentially the inclusion of charge-balancing counter-ions. This review will examine current approaches and
challenges to understanding 3D-conformation and dynamics via multiscale modeling at several levels of description (see Figure 1), from oligosaccharides to polysaccharides, PGs and protein-GAG complexes.

**Quantum mechanical studies of oligosaccharides**

Oligosaccharides of HA have been successfully modeled using B3LYP hybrid density functional theory (DFT) and the 6-31G** basis set [5]. Both Hartree-Fock ab initio and B3LYP DFT methods (with bases 6-31G*, 6-31+G* and 6-311+G**) have been applied to study sulfated GAGs, allowing computation of their geometries and energies [6, 7]. These studies have been extended (also using B3LYP DFT) to fully optimize the molecular 3D-structure of a heparin trisaccharide in the presence of explicit solvent [8]. The best agreement with experiment was found using the 6-311+G**, and a subsequent study showed how this basis set could be applied to a longer heparin pentasaccharide [9]. Recent computational advances are allowing these more realistic, but expensive, Pople-type basis sets to be used, which can more correctly model sulfate molecular orbitals and non-bonded interactions. Further increases in computer power will allow application of even more accurate correlation-consistent and polarization-consistent basis sets and post-Hartree-Fock methods to GAGs.

Several semi-empirical methods have been applied to GAGs, such as PM3CARB-1 [10] and SCC-DFTB [11]. A comparison of their ability to reproduce ab initio puckering energies showed that both predicted realistic barrier heights for glucose [12]. However, SCC-DFTB more accurately reproduced the overall energy surface. In a separate study, they were compared with molecular mechanics using the GLYCAM06 force-field (see Figure 2) and experiment, indicating that PM3CARB-1 imparted some quantifiable benefits to monosaccharide puckering, while SCC-DFTB appeared to model the sulfonate moieties more realistically [13]. However, the semi-empirical methods did not outperform molecular mechanics in many respects, and this was particularly the case for unsulfated GAGs.

Despite QM’s enhanced ability to model electronic properties, these approaches have a restricted ability to study molecular dynamics and, based on the currently available computational power, are confined to geometry optimization of relatively small molecules. This will undoubtedly change in the years ahead.

**Molecular mechanics and water interactions**

The molecular mechanics (MM) level of theory permits consideration of large molecular systems, bulk water and molecular dynamics. Force-field parameters for carbohydrates such as GLYCAM [14], have been extended with new residues and patches to allow a wide range of GAGs to be modeled with AMBER [15]. Alternative parameter sets are under development for other force-fields, such as CHARMM [16] and GROMOS [17].

Classical molecular dynamics (MD) simulations in explicit aqueous solution using a MM force-field allow the important interaction of GAGs with water to be investigated [18]. One of the key considerations is the water model, and for this reason a CS 8-mer was simulated with four popular choices [19]. Detectable differences were only
observed at the $\beta(1\rightarrow3)$ linkage, and while the 3-site water model (TIP3P) favored intra-molecular hydrogen bonds, the 4-site and 5-site models disfavored them. Other studies have gone beyond these static charge-models, for example in developing a polarizable carbohydrate force-field for aqueous N-acetylglucosamine (GlcNAc) [20]. The results were reported to be in better agreement with ab initio calculations than non-polarizable model, and this promising approach may be extensible to GAGs.

The conformation of HA oligosaccharides have been investigated using aqueous MD simulations and a MM force-field [21], revealing a local conformation close to a left-handed four-fold helix. A more recent study used methodologically-similar MD simulations of 48-merS, including water and ions, as a basis for modeling long HA random coils [22●]. Encouragingly, predictions of radii of gyration with varying electrolyte concentration were in agreement with experiment. One of the first aqueous MD simulations of a CS 4-mer indicated that sulfated GAGs have a higher hydration ability than their non-sulfated equivalents [23]. More recent aqueous MD simulations used an adaptive biasing force to increase the rate of sampling in five different CS disaccharides [24]. The $\beta(1\rightarrow3)$ linkage was predicted to have one minima, while the $\beta(1\rightarrow4)$ had two, and could also be used to infer that calcium ions bind preferentially to carboxylate rather than sulfate groups.

It has been hypothesized for some time that the IdoA sugar has an unstable pyranose ring pucker and this has been corroborated experimentally. As unbiased microsecond MD simulations became available, both IdoA and GlcA sugar rings were confirmed to undergo conformational exchange (see Figure 3) [25, 26]. Furthermore, even sugars typically considered to be rigid chairs, such as GlcNAc, are predicted to have conformational flexibility on microsecond timescales [27]. This has led to a protocol for conformational analysis of oligosaccharides using microsecond simulations [28●●], which has been applied to HA, CS, DS [29●] and heparin [30]. Other enhanced sampling techniques, such as adaptively biased MD, have been applied to this problem in uronic acids to rapidly compute approximate free energies in explicit solvent [31]. Similarly, and more recently, metadynamics have been applied to IdoA and IdoA2S (the 2-O-sulfate) to predict puckering equilibria [32]. It was found that minor changes to the non-bonded parameters could affect the equilibria and potentially be used as a route to tweak simulations to agree with experimental data. An alternative approach is to use time-averaged distance restraints from experimental data in MD simulations, which, when applied to eight heparin-like trisaccharides, allowed accurate quantification of puckers in IdoA [33].

The sulfated GAGs attach to PG cores via a specific tetrasaccharide linker and MD simulations, combined with adaptive biasing force techniques, have recently been used to investigate their conformation and dynamics [34]. Other MD-based algorithms have also been developed to compare non-carbohydrate-based mimetics with GAGs [35]. The simulations show promise for identification of molecules that can act as GAG mimetics. Yet other computational methods have focused on interpretation of complex experiments, such enforced geometry optimization simulations of GAG structural changes, in order to understand atomic force microscopy measurements [36].
Modeling interactions with proteins

A promising method for producing theoretical protein-GAG complex models involves a two-step process based on molecular docking (using AUTODOCK software), followed by MD simulation refinement. This approach was effective in studying the interaction between HS and fibroblast growth factor [37●●], and similarly within the heparin, anti-thrombin and serine protease system [38]. In a related system, free energy perturbation MD simulations provided a reliable method for estimating changes in heparin binding affinity, and water was found necessary for accurate modeling of protein-GAG complexes [39]. Furthermore, it may also be necessary to allow IdoA and IdoA2S rings to adopt multiple puckers during docking exercises [40]. The applicability of these available theoretical approaches to studying protein-GAG complexes has recently been investigated by comparing several methodologies [41].

A novel and promising protocol uses GAG microarray data in combination with rigid-body docking to identify binding sites [42●●]. A slightly different approach identified GAG binding sites using Poisson-Boltzmann electrostatics and orientated using dynamic molecular docking (DMD), allowing exhaustive sampling of GAG conformational space [43]. In a similar fashion, three non-overlapping binding sites were identified in metalloproteinase-3, this time using aqueous MD and docking (AUTODOCK) of HA, CS and heparin in a fully flexible state [44]. Other authors have confirmed that protein electrostatic potential calculations are useful for proposing protein surface regions as putative GAG binding sites, which can be used for local docking calculations [45]. Another option for studying protein-GAG interactions is the ligand-binding site prediction program FINDSITE, which has been successfully used to identify ligand binding poses for heparin analogs bound to heparanase [46].

Related theoretical methods have also been usefully applied to larger complexes. For example, in an investigation of chemokine CXCL8, AUTODOCK was used to dock a heparin 8-mer to the monomer and this was used as a template for constructing a model of a longer heparin 24-mer in a horseshoe geometry interacting with a protein dimer [47●]. The initial model was then refined with MD simulations using the AMBER force-field. Other studies have extended this approach, by applying docking to map the heparin binding sites on the αvβ3 integrin [48]. The large protein size was overcome by splitting the surface into 46 overlapping areas. In another study the authors used TIGER2 replica-exchange MD to produce a model of collagen with telopeptides and, rather than using docking, this was subjected to aqueous MD in the presence of heparin oligosaccharides [49]. The simulations were used to identify heparin binding sites and understand collagen aggregation.

Coarse graining of polysaccharides and proteoglycans

Replacing the atomic detail found in QM and MM levels of theory with lower resolution, coarse grained (CG), representations opens the way to simulations of larger molecules and on longer timescales [50], which is particularly attractive for GAGs. Such methods typically have more structural detail than generic polymer physics theory, and one of the first CG methods for GAGs was developed to study the dynamics and titration behavior of HA and CS [51●]. The predictions were found to be in quantitative agreement with experiment, while allowing the simulations to be
much more computationally efficient than MM. A second approach developed AMBER-compatible CG parameters for GAGs from MD in explicit solvent using a Boltzmann conversion approach [52]. They concluded that the CG model was suitable for investigating the dynamics of long GAG molecules at long time scales. An alternative methodology also used the Boltzmann conversion approach to parametrize the CG model (based on microsecond MD of HS oligosaccharides), while also including dynamic pyranose ring puckering [30]. This model was capable of predicting macroscopic hydrodynamic properties, but only when puckering was taken into account.

There are also models that have shown promise for modeling glycoconjugates, but appear not to have been applied to GAGs, and are worthy of consideration. For example, parameters for carbohydrates have been developed for the MARTINI coarse-grained force field, albeit for the simulation of crystalline cellulose fibers [53]. A further, and recently reported, method uses Monte Carlo simulation followed by torsion-angle MD simulation to produce dynamics for a range of biological molecules, including glycoproteins [54]. These CG models appear to be particularly suited to systems where both carbohydrate and protein are present (and potentially also lipids). Furthermore, there has been work specifically directed towards modeling whole PGs. One approach used the Poisson-Boltzmann equation to model the interactions between CS chains in aggrecan to study its bottle-brush geometry [55]. More recent research used suitably parametrized beads and MD simulations to study the interaction of neutral and charged bottle-brush macromolecules attached to substrates [56]. In another study, a previously derived CG model for HS was extended to include CS and DS (see Figure 4), and used to simulate the dynamics of heterogeneous GAGs and small leucine-rich PGs [29]. It was hypothesized that GAG size is dependent on unit composition and independent of sequence permutation. It was also predicted that IdoA increased GAG chain volume and rigidity, while GlcA imparted chain plasticity; hence block copolymeric GAGs, such as HS, likely contain microarchitectures capable of multivalent binding to growth factors and collagen.

Conclusions

Recent advances in computational power and techniques have enabled major progress in all areas of GAG modeling. At the quantum mechanical level, density functional theory calculations of sulfated oligosaccharides using accurate basis sets and explicit solvent are possible and this is set to transform our understanding of local conformations and interactions with water. Much progress has also been made on development of fully-featured molecular mechanics force-fields for GAGs, which are being used to accurately model dynamics of chains beyond oligosaccharide in water. Moreover, good sampling of energy landscapes are possible with enhanced sampling techniques and also the kinetics of conformational exchange can be predicted using microsecond timescale unbiased molecular dynamics simulations. An emerging technique is coarse-graining, which promises to offer tantalizing insights into whole proteoglycans and assemblies thereof, and an exciting route to experimental interpretation. Finally, new theoretical protocols are being developed for studying interactions between GAGs and proteins. This is allowing prediction of
interactions, binding sites and poses in a research area where experimental data is sparse. Taken together, modeling is having a disproportionate impact on overcoming the significant research challenges in this important biological field.

**Conflicts of interest:** None.

**Funding:** This review did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Figure Legends**

**Figure 1.** Multiscale modeling of glycosaminoglycans from disaccharide to polysaccharide is necessitated by their size and heterogeneity (not to the same scale). (a) Optimized conformation and highest occupied molecular orbital (HOMO) of a heparin disaccharide using B3LYP/6-311+G** ab initio theory, (b & c) molecular modeling of a hyaluronan decasaccharide in water using the GLYCAM06 force-field in TIP3P water, coarse-grained representation of (d) a heparan sulfate polymer and (e) a biglycan proteoglycan [29●, 30].

**Figure 2.** Overlays from 20ns simulations of a chondroitin sulfate disaccharide in water using both semi-empirical (a & b) and molecular modeling levels of theory (c) [13].

**Figure 3.** Puckering dynamics for key monosaccharides (O-methyl) found in glycosaminoglycans from 5µs molecular dynamics simulations using GLYCAM06 in water, for (b) iduronic acid (IdoA), (c) iduronic acid-2-O-sulfate (IdoA2S), (d) glucuronic acid (GlcA) and (e) N-acetylglucosamine [26, 27]. The dynamics are visualized using the Cremer-Pople azimuthal angle (labeled according to canonical pucker) against time and (a) shows the typical 3D-conformers found in the puckering equilibrium.

**Figure 4.** Representative 3D-conformations from coarse-grained simulations of polysaccharides and proteoglycans [29●] (all at the same scale). (a) Chondroitin-4-sulfate chain (blue), (b) heterogeneous chondroitin-4-sulfate (blue) and dermatan-4-sulfate chain (green), and (c) dermatan-4-sulfate chain (all with 200 monosaccharides), (d) Decorin proteoglycan with human-like chondroitin sulfate chains, and (e) is a human serum albumen protein dimer to same scale (protein data bank entry 1AO6).
References


GLYCAM is probably the most widely used molecular mechanics force-field for modeling carbohydrates and this paper details an important update for glycosaminoglycans.


A protocol is proposed for conformational analysis of oligosaccharides using 10 microsecond molecular dynamics simulations. Shorter simulations did not reach equilibrium, while longer simulations did not appear to provide additional sampling benefits.

Multimicrosecond aqueous molecular dynamics simulations of CS and DS are performed, analyzed, and used to parameterize a previously defined coarse-grained model. This model is used to make conformational predictions for heterogeneous glycosaminoglycans and small leucine-rich proteoglycans, and, in particular, study the relationship between sequence and shape.


A two-step method is developed for predicting the molecular interaction between glycosaminoglycans (in this case HS) and proteins based on molecular docking and molecular dynamics simulation. Molecular docking allows the position of a rigid fragment to be identified on the protein surface, while molecular dynamics is used to refine the docked model.


This is a novel approach that combines carbohydrate microarray data and computer modeling to investigate and identify new glycosaminoglycan-protein interactions. Microarrays are used to rapidly screen proteins and, when an interaction is identified, computer modeling is used to predict the binding site.


An extensive and informative molecular modeling study of a glycosaminoglycan-protein interaction, which uses molecular docking and molecular dynamics to produce a model of a CXCL-8 dimer in complex with a 24-mer heparin fragment.


A very solid piece of work describing a theoretical framework for coarse graining of glycosaminoglycans. In particular, the Debye-Hückel interaction potential for studying ionic strength-dependent phenomena is described and implemented.


Molecular dynamics simulations of aggrecan bottle-brush macromolecules tethered to a substrate are performed using chains composed of charged Lennard-Jones beads. These were used to calculate the disjoining pressure between substrates as the distance between them was varied.
Conflicts of interest: None.