Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges

Andrew Almond

Glycosaminoglycans are long unbranched and complex polysaccharides that are an essential component of mammalian extracellular matrices. Characterization of their molecular structure, dynamics and interactions are 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.

Address
School of Chemistry, The University of Manchester, Manchester Institute of Biotechnology, 131 Princess Street, Manchester M1 7DN, UK

Corresponding author: Almond, Andrew (Andrew.Almond@manchester.ac.uk)

Current Opinion in Structural Biology 2018, **50**:58–64

This review comes from a themed issue on **Carbohydrates**

Edited by **Hugues Lortat-Jacob** and **Ralf Richter**

https://doi.org/10.1016/j.sbi.2017.11.008

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/DS 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].

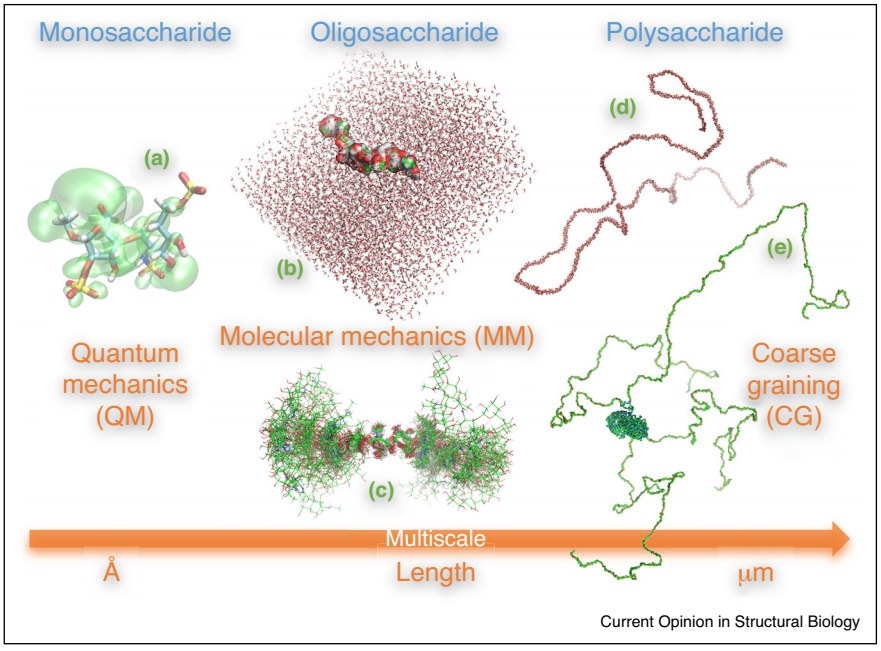
Studies of GAG 3D-conformation and dynamics are difficult to perform experimentally, and computer modeling is frequently needed, in part due to their repetitive chemistry and flexibility in solution. In general, theoretical methods used to study carbohydrates are applicable to GAGs, but with additional 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*

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 representations of **(d)** a heparan sulfate polymer and **(e)** a biglycan proteoglycan [29•,30].

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, this approach has 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 β(1 → 3) 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 models, 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