COMPUTATIONAL MODELING OF GLYCOSAMINOGLYCANS AND THEIR INTERACTIONS WITH PROTEINS

by

ARUNIMA SINGH

(Under the Direction of Robert J. Woods)

ABSTRACT

Glycosaminoglycans (GAGs) are the most abundant hetero-polysaccharides found in mammalian tissues, and serve a variety of important biological roles in the body. They can be found predominantly in the extracellular matrix or cell surface, usually covalently attached to proteins. From providing tensile strength and tissue compressibility, to aiding in cell proliferation and recognition, to being key receptors for viral entry, they offer several avenues of therapeutic interest. GAGs are typically composed of repeating units of hexosamine and uronic acid pairs, that are often variably sulfated, leading to tremendous structural heterogeneity and a high charge density, making structural analysis and elucidation of binding modes very challenging. This presents an opportunity for computational modeling methods to provide insight into the structure and function of these molecules.

The key to effective theoretical modeling of biomolecules is the use of a dependable and validated force field. This work presents a validation of the recent addition of parameters that enable modeling of sulfated GAG sequences to the GLYCAM force field, through a comparison of experimental NMR scalar coupling constants and NOE distances with theoretical data. The analysis demonstrates that the new force field parameters are capable of reproducing NMR

observables for a number of GAG fragments. These parameters have been employed for designing the Glycosaminoglycan Builder, a point-and-click structure modeling utility on GLYCAM-Web, to facilitate 3D structure modeling of GAG fragments. The interface provides separate sets of monosaccharides, unique to each class of GAGs, allowing easy selection of pre-sulfated options for each monosaccharide. We also employ these parameters to study the binding of variably sulfated heparin fragments to chemokine CCL5. The study demonstrated a dependence of the ability of heparin tetrasaccharides to inhibit CCL5-CCR1 binding on the pattern and extent of sulfation. An analysis of the binding mode of a longer heparin fragment to the protein, as well as an examination of the effect of pH on CCL5-heparin binding was also performed.

INDEX WORDS: Glycosaminoglycans, Protein-Glycosaminoglycan interactions, Force field validation, GLYCAM, Molecular dynamics simulations, Automated docking, MM-GBSA, Web-based structure building tools, Theoretical collision cross section calculation

COMPUTATIONAL MODELING OF GLYCOSAMINOGLYCANS AND THEIR INTERACTIONS WITH PROTEINS

by

ARUNIMA SINGH

B.Sc., Gujarat University, India, 2004

M.Sc., Pune University, India, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

COMPUTATIONAL MODELING OF GLYCOSAMINOGLYCANS AND THEIR INTERACTIONS WITH PROTEINS

by

ARUNIMA SINGH

Major Professor:	Robert J. Woods
Committee:	James H. Prestegard
	I. Jonathan Amster
	Natarajan Kannan

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2015

To my parents.

ACKNOWLEDGEMENTS

My sincerest gratitude goes out to my supervisor Prof. Robert J. Woods who helped guide this work and provided continuous encouragement and support. The discussions, disagreements and editorial critique have all shaped the scientist I am today.

I would like to thank my advisory committee members Prof. Prestegard, Prof. Amster and Dr. Kannan for their scientific guidance and suggestions.

A special thanks goes out to all the Woods group members. Spandana, Anita, and Amika, thanks for being there for all the tough times. This wouldn't have been possible without great friends like you. Xiao, David, Mia, Huimin and Yuejie, thanks for all the serious discussions and the fun times. Lachele, thanks for being there for all the math and physics discussions and providing feedback on the research. Mark, of course I don't thank you enough, but without your friendship, help, sense of humor, and the much-needed distractions, these years would have been very dull.

Lastly, I would like to thank my family, my parents and brother, and my loving husband Joy for believing in me all these years.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This dissertation addresses the topic of computational modeling of glycosaminoglycans (GAG) and their interactions with proteins. It comprises the following research projects:

1. Validation of the glycosaminoglycan specific parameter set of the GLYCAM force field, through a comparison of experimental NMR scalar coupling constants and NOE distances with analogous data obtained from theoretical measurements.
2. Design and implementation of the Glycosaminoglycan Builder: a point-and-click structure modeling utility on GLYCAM-Web that enables 3D structure modeling of GAG fragments.
3. Analysis of the binding modes of variably sulfated heparin fragments to CCL5, and the differences in their ability to inhibit the interaction of CCL5 to its receptor CCR1.

These research topics, as well as a review of their respective background information and the computational methods applied to them, are presented as follows:

CHAPTER 2: GLYCOSAMINOGLYCANS

This chapter provides the relevant background information about glycosaminoglycans, their classification, structural features, and interactions. It also presents an overview of the developments in computational modeling of GAGs and protein-GAG interactions.

CHAPTER 3: MOLECULAR MODELING METHODS

Chapter 3 covers the theory behind force fields, molecular dynamics simulations and molecular docking mothodologies.

CHAPTER 4: EXTENSION AND VALIDATION OF THE GLYCAM FORCE FIELD PARAMETERS FOR MODELING GLYCOSAMINOGLYCANS

Chapter 4 is an original research study describing the development of GAG-specific parameters for GLYCAM, and their validation through comparison of theoretical simulation data to NMR. The results of this study will be soon be published.

Contributions to this article include charge development for protonated uronic acids and $\Delta^{4,5}$-unsaturated uronates, validation of new parameters through MD simulations and comparison to NMR, and manuscript preparation.

CHAPTER 5: GLYCOSAMINOGLYCAN BUILDER

Chapter 5 describes the development of the Glycosaminoglycan builder utility of GLYCAM-Web. The design, implementation details, and features will be included as part of the publication describing the structure building features on GLYCAM-Web.

CHAPTER 6: THE INTERACTION OF HEPARIN TETRASACCHARIDES WITH CHEMO-KINE CCL5 IS MODULATED BY SULFATION PATTERN AND pH

Chapter 6 is an original research study undertaken to predict the binding modes of variably sulfated heparin tetrasaccharides to CCL5, in order to understand the structural basis for the difference in the ability of certain fragments to inhibit the binding of CCL5 to its receptor CCR1.

The results of this study were published as a journal article:

Singh, A., Kett, W. C., Severin, I. C., Agyekum, I., Duan, J., Amster, I. J., Proudfoot, A. E. I., Coombe, D. R., and Woods, R. J. (**2015**) *J. Biol. Chem.* 290, 15421–36

Contributions to this chapter include computational modeling, data analysis and manuscript preparation.

CHAPTER 7: THEORETICAL CALCULATION OF COLLISION CROSS SECTION

Chapter 7 provides information about how theoretical collision cross section areas are calculated and provides necessary script for input file preparation.

CHAPTER 8: CONCLUSIONS AND FUTURE PROSPECTS

The last chapter summarizes the major findings and conclusions of this work, and provides comment on possible future directions.

CHAPTER 2

GLYCOSAMINOGLYCANS

Glycosaminoglycans (GAGs) are the most abundant hetero-polysaccharides found in animal tissues. They are long, unbranched polymers with a highly heterogeneous structure (1, 2) and usually a high charge density. They are primarily found on the surface of cells or in the extra cellular matrix (ECM), and sometimes stored in secretory vesicles of some cells to help regulate availability of positively charged molecules in the vesicles (3). Virtually all mammalian-cells produce a variety of GAGs specific to the tissue composition and function (4).

GAGs are composed of repeating disaccharide units made of a hexosamine and a uronic acid or galactose (Gal). The hexosamine may be an *N*-sulfated or an *N*-acetylated glucosamine (GlcNS, GlcNAc) or galactosamine (GalNS, GalNAc), variably *O*-sulfated at the 3, 4 and/or 6 positions. The uronic acid may be a glucuronic acid or an iduronic acid, and may be 2-*O*-sulfated. GAGs can be classified into 5 main groups based on unique disaccharide compositions: hyaluronic acid (HA), heparin/heparan sulfate (HS) (**Figure 2.1**), chondroitin sulfate (CS), dermatan sulfate (DS), and keratan sulfate (KS). Keratan sulfate lacks uronic acids and instead contains variably sulfated galactose residues, and while most GAGs have a heterogeneous pattern of sulfation, hyaluronan is the only unsulfated GAG. All GAGs except HA and heparin are covalently attached to proteins, via a linker sequence connected to their reducing ends, forming a glycosidic linkage to serine/threonine or asparagine residues (5, 6). This forms a class of macromolecules called proteoglycans.

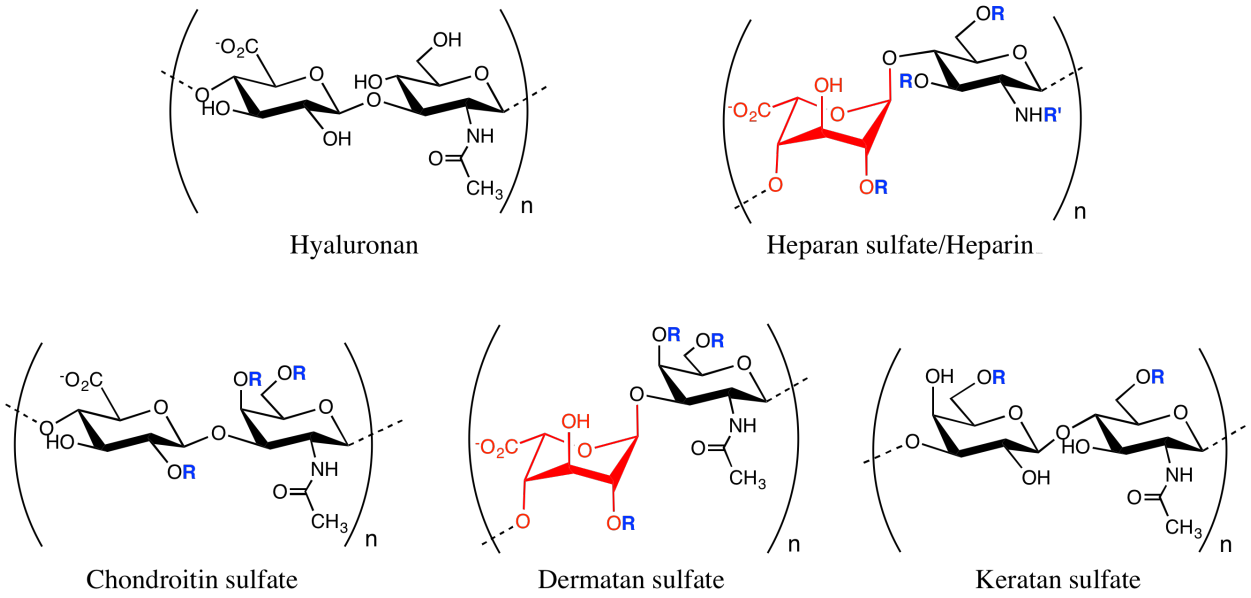

Figure 2.1 Disaccharide structures of different classes of GAGs. R (shown in blue) represents H or SO_3^-, R' represents H, SO_3^- or $COCH_3$. Residues in red (shown as IdoA) may be either GlcA or IdoA.

GAGs play various important biological roles. In the ECM they provide tensile strength, lubrication, elasticity and compressibility to various tissues. Together with the membrane proteoglycans that also interact with the ECM, they regulate cell adhesion, cell growth and proliferation, maintenance of protein concentration gradient in regions of inflammation, immobilization of proteins, immune system invasion, and tumor metastasis (4, 7–10). GAGs are stabilizers, cofactors, and/or coreceptors for growth factors, cytokines, and chemokines (11). In addition, the proteoglycans that are stored in secretory granules help sequester various positively charged proteases and amines through attraction towards the highly negative charge density of the GAGs (8).

Classification of GAGs - Structure, Function And Significance

Hyaluronan

Hyaluronan or hyaluronic acid is one of the largest polysaccharides produced by vertebrates with molecular weight often reaching the millions (12). It is not sulfated and is not covalently attached to proteins, like other GAGs, to form proteoglycans. Instead it is found as part of non-covalent complexes with proteoglycans in the ECM. It is produced by virtually all the cells in the human body, and distributed heavily in connective, neural and epithelial tissues. Unlike other GAGs, it is synthesized in the plasma membrane, and not the Golgi apparatus. Its primary function is to provide a matrix for cell proliferation and migration, tissue organization, and signal transduction processes. Due to its ability to absorb and displace large volume of water, it is highly compressible, and forms an excellent lubricator, and shock absorber in the synovial tissues and joints.

Hyaluronan synthesis takes places on the inner side of the plasma membrane via a class of integral membrane proteins called hyaluronan synthases. As the enzyme adds *N*-acetylglucosamine and glucuronic acid residues alternately to the growing chain, it is extruded out of the cell membrane through the ABC transporter proteins (13). The lengths of the chains are frequently on the order of 10^4 disaccharides.

Hyaluronan interacts with proteoglycans via hyaluronan-binding motifs in a non-covalent manner. It is an important component of the cartilage, where it forms stable ternary complexes with aggrecan, a cartilage proteoglycan, to provide resilience towards compressive load in joints (12). In addition, it also binds to other members of the lectican family of interstitial proteoglycans that include brevican, neurocan, and versican, in addition to aggrecan (4). Hyaluronan is also a component of skin, and is involved in skin and wound healing.

The transmembrane cell surface receptor CD44, which is present on leukocytes and other cells surfaces, also contains a hyaluronan binding link module and its interaction with hyaluronan is implicated in tumor cell survival and metastasis (14, 15). CD44 also associates with RHAMM (Receptor for Hyaluronan Mediated Motility), also a hyaluronan binder, to activate signaling processes involved in cell growth and motility in cells, including cancerous cells (16). Inhibiting these interactions is an area of active research for cancer treatment.

Heparin and Heparan Sulfate

Virtually all the cells in the body produce heparan sulfate, while heparin is synthesized solely by the mast cells and basophils (17). Both are initially synthesized as polymers of *N*-acetyl glucosamine and glucuronic acid in the Golgi bodies, and then modified extensively by enzymes such as *N*-deacetylase, that cleaves the *N*-acetyl group, sulfotransferases, that add sulfates, and epimerase, that converts glucuronic acid to iduronic acid (18). Since these modifications do not occur exhaustively throughout the length of the chains, it leads to tremendous heterogeneity making their sequence analysis very challenging. The uronic acids may be sulfated at the 2-*O* position (IdoA2S or GlcA2S), while the glucosamine may rarely occur as a free amine or more often as an *N*-sulfated (GlcNS) or an *N*-acetylated (GlcNAc) moiety. The GlcNS may also be *O*-sulfated at the 3 or 6 positions, giving rise to GlcNS6S, GlcNS3S, or GlcNS3S6S. In contrast, the GlcNAc residues may be unsulfated or *O*-sulfated only at C6 (GlcNAc6S) (19). The modifications usually occur in clusters along the GAG chain, leading to formation of *N*-sulfated, *N*-acetylated, and mixed domains. The formation of these domains is critical to the interaction of these molecules to proteins such as antithrombin III (ATIII) and fibroblast growth factors (FGFs) that recognize and bind to particular structural motifs along the length of the GAG polymers.

In general, heparin is more heavily sulfated and epimerized than heparan sulfate; in fact, it contains the highest negative charge density of all biological molecules known (20). During synthesis, heparin is covalently attached to the protein serglycin (21), found exclusively in mast cell granules and some other hematopoietic cells. Heparin is released as a peptidoglycan from the mast cells, and immediately cleaved from the peptide as shorter fragments of free heparin.

Heparin is best known for its anticoagulant activity and has been used clinically as an anticoagulant for over 60 years. The anticoagulant activity of heparin is attributed to its interaction with ATIII via specific a pentasaccharide sequence containing a critical 3-*O*-sulfate modification. Binding of heparin to ATIII brings about an inhibition of blood coagulation factors, thrombin and factor Xa. Heparin binds to a myriad of proteins involved in various functions in the body ranging from cell proliferation and differentiation, to inflammation and wound healing, to viral entry and localization, to development of amyloid plaque, to cytokines and chemokines to name a few. The exact physiological functions of heparin remain unclear in spite of identification and analysis of multiple heparin-protein interaction, and its extensive clinical use as an anticoagulant.

HS is present in the form of HS proteoglycans (HSPGs) in all mammalian tissues (**Figure 2.2**). The core protein of the HSPGs generally contain one of more HS chains covalently *O*-linked (22) via a GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xylβ1-*O*-Ser linkage tetrasaccharide sequence. Based on the type of core protein, HSPGs are divided into three classes: transmembrane syndecans, glycosylphosphatidylinositol (GPI) anchored glypicans and extracellular forms that include agrin, perlecan, and collagen XVII (6, 23, 24). The syndecans and the glypicans can be released from the membrane into circulation via proteolysis.

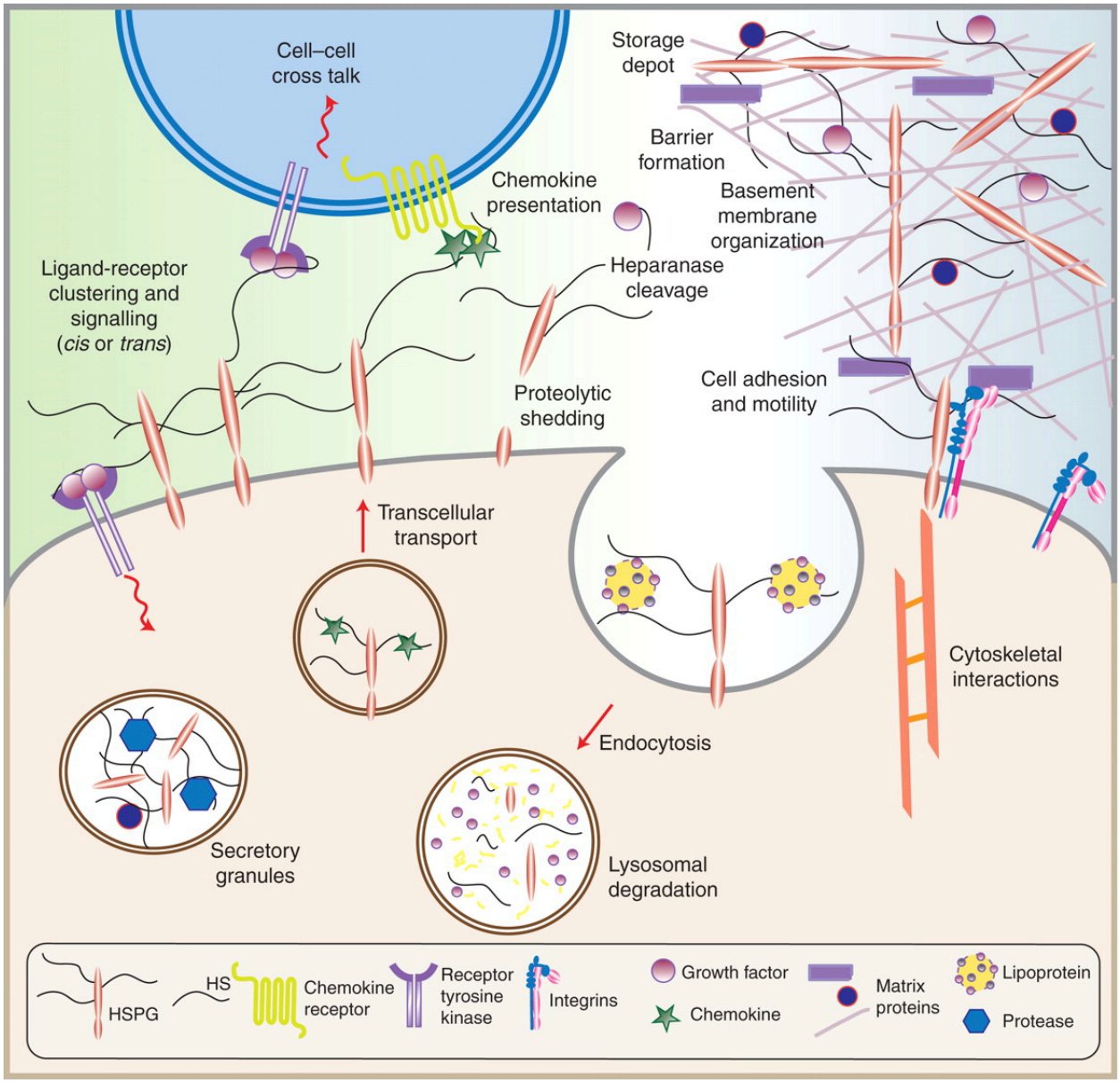

Figure 2.2 Role of HSPGs in cells and tissues. (Adapted from (25); reprinted with permission from Nature Publishing Group © 2007 and Cold Spring Harbor Laboratory Press © 2011)

Syndecans are involved in cell-cell, cell-ECM, and cytoskeletal organization for cell adhesion and signal transduction (26). Glypicans play a role in developmental morphogenesis, being highly expressed in embryonic tissues (27). They also regulate FGFs and bone morphogenic proteins and other cell signaling pathways (28). Agrin, perlecan and collagen XVIII form the secreted class of HSPGs. Agrin is primarily found in kidney and brain, perlecan

is expressed in nearly all basement membranes and connective tissues, and collagen XVIII is found in epithelial and endothelial basement membranes and in cartilage and fibrocartilage (29, 30).

Chondroitin Sulfate

Chondroitin sulfate is one of the first GAGs that was studied as part of "chondromucoid", obtained from cartilage and liver (4). It is found most abundantly in cartilage, bone and heart valves, providing structural integrity and resistance to compression. CS chains are linked via a GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xylβ1-*O*-Ser linker, same as the linker found in HSPGs, to form CS proteoglycans (CSPGs). The CSPGs form a class of proteins called lecticans that are a component of the ECM, and include four proteins: aggrecan, brevican, neurocan, and versican (31). Aggrecan is primarily found in cartilage, providing high compressibility due to the large negative charge density and elongated structure, while versican is expressed in various connective tissues in smooth muscles, epithelial cells and nervous system, aiding in cell adhesion, migration and proliferation. Neurocan and brevican are both restricted to neural tissues, and although the exact functions remain elusive, they are believed to help stabilize brain synapses, and prevent regeneration of damaged nerve endings.

The CS (and DS) synthesis pathways are highly conserved across invertebrates and vertebrates, and involve the activity of the enzyme chondroitin synthase, that is capable of polymerizing GlcNAc and GlcA residues via its dual β1–3-glucuronosyltransferase and β1–4-*N*-acetylgalactosaminyltransferase activities. The chain is consequently acted upon by different sulfotransferases to add 4-O and 6-O sulfates.

Dermatan Sulfate

Dermatan sulfate, as the name suggests, is the predominant glycan present in skin (dermis). It is also present in blood vessels, heart valves, tendons and lungs (32). The synthesis process of DS is similar to that of CS, except for the epimerization of GlcA residues to IdoA. In fact, historically DS was considered to be a variant of CS and was called chondroitin sulfate B, but the presence of IdoA distinguishes it from CS, and is now considered to be functionally similar to HS in some respects (33). The degree of epimerization and sulfation varies along the DS chain, leading to additional level of complexity, which is thought to be controlled via an enzymatic system to encode functional information (32).

DS also attaches to proteins forming DS proteoglycans (DSPGs), and the well-studied DSPGs are biglycan and decorin. DS and DSPGs have been implicated to have a role in cardiovascular disease, tumorigenesis, infection, wound repair, and fibrosis. Accumulation of DS in the mitral valve has been associated with mitral valve prolapse (34).

Keratan Sulfate

Keratan sulfate is found primarily in the cornea, cartilage, and bone. It is made of sulfated poly-N-acetyllactosamine ($[Gal\beta1–4GlcNAc\beta1–3]_n$) that is commonly seen on other glycoproteins and mucins. Depending upon the type of linkage to proteins, KS can be of two types: KS I, the corneal KS, attached as a complex-type *N*-linked branched oligosaccharide to an asparagine residue, and KS II, the skeletal KS, connected via an *N*-acetylgalactosamine residue to serine/threonine similar to the mucin core-2 branched oligosaccharides (35). The concentration of KS in the cornea is 10-fold higher than in cartilage and 2-4 times higher than in other tissues (36). It helps maintain corneal hydration, and defects in sulfation cause distortions in corneal

opacity and macular corneal dystrophy (37, 38). Although the function of KS II in cartilage currently remains unclear, KS is found attached to several core proteins in various tissues, and has been implicated in anti-adhesive functions such as regulation of macrophage adhesion, embryo implantation and motility of corneal endothelial cells (35).

Three Dimensional Structural Details of Heparin, Heparan Sulfate And Interactions

The structure and interaction of heparin and heparan sulfate have attracted more physico-chemical interest than other classes of GAGs (20). This may be in part due to the widespread use of heparin as an anticoagulant, as well as it being commercially available. Heparin also serves as a suitable model for structural studies involving HS, due to the structural similarities between the two, leading to considerable advances in the knowledge about this particular class of GAGs.

Solution Conformation of Heparin

The epimerization of the uronic acid residues in heparin and HS have important conformational effects on these molecules. Conformational analysis of heparin has shown that while the D-glucopyranose residues (GlcA, GlcNAc, and their derivatives) are the most stable in the $^{4}C_{1}$ conformation, the L-iduronate residues show high plasticity in their conformational preference (39). The internal iduronates (IdoA or IdoA2S) in the polymer chains exist in equilibrium between a number of different conformations, the most predominant being the $^{1}C_{4}$ followed by the $^{2}S_{O}$ (40) (**Figure 2.3**). This equilibration depends upon the 2-*O*-sulfate substitution on the residues as well as the substitution of adjacent sugars (41).

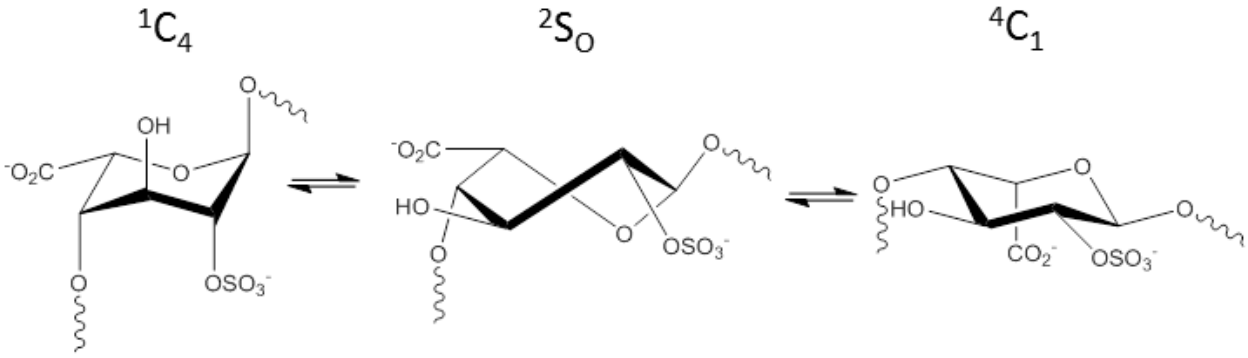

Figure 2.3 Most predominant solution conformations of IdoA.

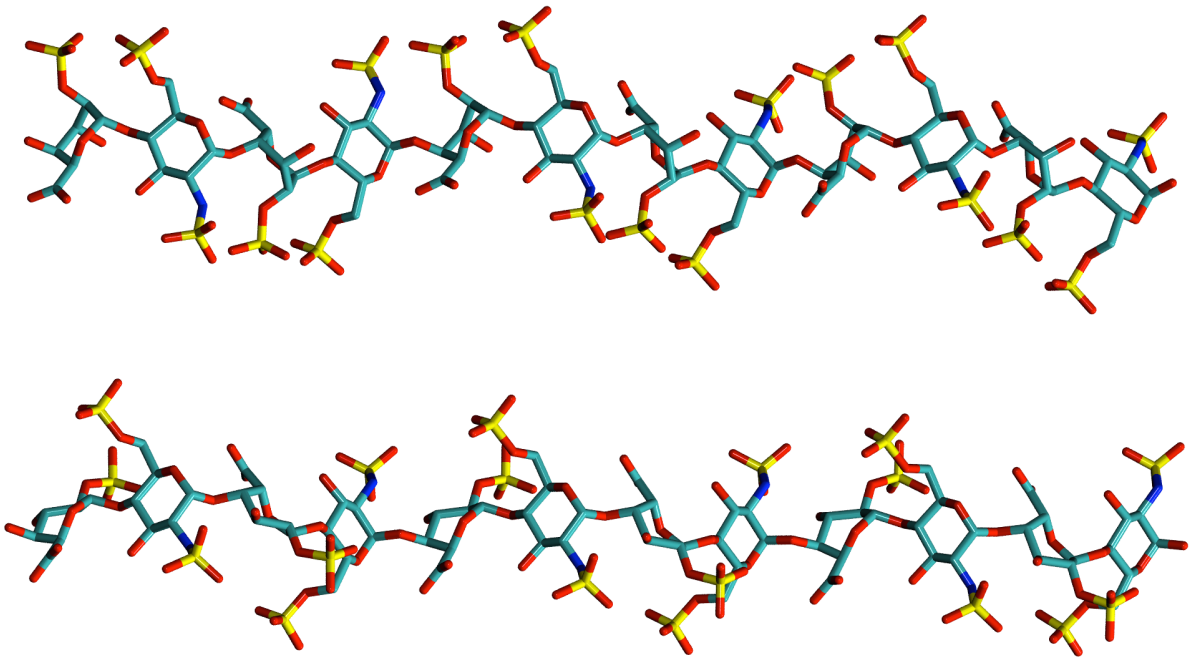

Figure 2.4 Heparin dodecasaccharide with IdoA in $^1C_4$ (top) and ^2S_O (bottom) conformation (PDB ID: 1HPN (42)). The overall length of each dodecasaccharide as well as the number of residues per turn (about 4 in this case), remain the same with change in conformation of IdoA.

The overall conformation of the polymer backbone, however, remains unchanged with the $^1C_4$-^2S_O inter-conversion (**Figure 2.4**) (42). NMR studies have also revealed that in solution,

unbound heparin exhibits similar Ψ and ϕ glycosidic conformations, regardless of the sulfation pattern (43).

Oligosaccharides obtained as a product of lyase cleavage from intact heparin also contain a non-reducing $\Delta^{4,5}$-unsaturated uronic acid (ΔUA) residue, which may or may not be 2-*O*-sulfated. ΔUA exist in equilibrium between the $^{1}H_{2}$ and $^{2}H_{1}$ solution conformations, with $^{1}H_{2}$ being the more predominant form (44).

Binding Properties of Protein-Heparin Complexes

Numerous (>100) protein-GAG interactions have been demonstrated in literature. The importance of these interactions in various biological processes has stimulated further investigation of the details of these interactions. The available co-crystal complexes of heparin fragments and heparin binding proteins elucidate the conformational preference of iduronate residues when bound to proteins. In general, the preferred conformation varies between various proteins in a site-specific manner, with the bound IdoA2S residue adopting the $^{2}S_{O}$ conformation with annexin V (45), $^{1}C_{4}$ conformation for one iduronate and $^{2}S_{O}$ for the other in the heparin pentasaccharide bound to bFGF (46), and a mixed conformation (either $^{1}C_{4}$ or $^{2}S_{O}$), between the several heparin fragments bound to the viral capsid of with the foot and mouth virus (47). This suggests that the binding of heparin to proteins may induce conformational changes in the iduronate residues for an enhanced binding mode and higher affinity. The conformation of the glucosamine residues and its derivatives remained unaltered.

Strong ionic interactions play an important role in GAG binding proteins. The common feature in most GAG binding sites is the presence of basic amino acids, lysine, arginine and to some extent histidine, (48) which facilitate interactions with the negatively charged sulfate and

carboxylate groups on GAGs. These positive residues often form clusters on the surface of the protein. The negatively charged sulfate and carboxylate groups on GAGs interact with basic side chains of arginine and lysine, and histidine (48). Arginine shows higher number of possible hydrogen bonds, through the guanidino group in its side chain, than lysine, and the affinity of binding sites is generally defined by the ratio of these two residues (49). Clusters of basic amino acids on the surface of proteins are either linear motifs on the protein chain, or are brought together by protein folding to create the binding site (50, 51). An analysis of the binding sites of several heparin-binding proteins has shown the sequences XBBXBX or XBBBXXBB, where B is a basic amino acid (Arg or Lys), and X is any other residue, to be the consensus binding sequence (52).

In addition to electrostatic interactions, van der Waals (VDW) forces, hydrogen bonding, and hydrophobic interactions with the carbohydrate backbone also plays a role in protein-GAG interactions (46, 53).

Effect of pH on Protein-Glycosaminoglycan Interactions

Certain protein-GAG interactions can be altered or regulated by changes in pH. This is especially true for proteins that contain histidine residues close to the GAG binding sites. Histidine side chains have a pKa of ~6 (54) and thus get protonated when the pH falls to 6 or lower. Protonation increases the positive charge at the GAG binding interface, augmenting the ionic interactions between the protein and GAG chains (55). The significance of this pH dependent functional regulation has been observed in several instances of GAG binding proteins.

GAGs are involved in the formation of protein aggregates associated with neurodegenerative diseases and Alzheimer's disease (56). β-Amyloid peptide, a 40-43 residue

long peptide derived from the amyloid precursor protein (57), is a component of the senile plaques associated with the etiology of Alzheimer's disease. The peptide associates with various sulfated GAGs: heparin, heparin sulfate, dermatan sulfate and chondroitin sulfate, but does not bind non sulfated GAGs (58). Using heparin affinity chromatography, it has been determined that the maximum binding between the peptide and GAGs occurs below pH 7.0 and very little binding is observed above pH 8.0. To test the specificity of interaction, the heparin-binding motif Val-His-His-Gln-Lys-Leu at residues 12-17, was mutated to Val-Ser-Ser-Gln-Lys-Leu. Mutant peptides did not bind the heparin column at pH 4.0 or pH 8.0, indicating that protonated histidine residues are essential for β-Amyloid peptide and glycosaminoglycan interaction (58).

In contrast to promoting aggregation, some studies have demonstrate that sulfated GAGs can prevent heat-induced aggregation of certain proteins such as Antithrombin III, RNase A, β-Lactoglubulin and bovine serum albumin (BSA) (59–62). The prevention of heat induced aggregation of BSA by dermatan sulfate has been found to be pH sensitive (63). Dextran sulfate suppresses BSA aggregation at acidic pH of 5.1 and 6.2, however aggregation is observed at pH 7.5 (62). The intermediate state of denatured BSA forms a complex with dextran sulfate preventing further oligomerization.

Mast cell proteases are released by the mast cells as part of the innate immune response. These proteases, classified as tryptases, chymases and carboxypeptidases, are stored in biologically active form inside the secretory granules of mast cells at pH of 5.5 (64). Storage of tryptases in the mast cell granules involves interaction of the protein with the negatively charged, heparin-containing, serglycin proteoglycans (65). Human β–tryptase is stabilized by heparin at acidic pH, but dissociation from heparin leads to inactivation of the enzyme at neutral pH (66). Mouse MCP-7 storage also involves interaction of the positively charged side chains of

histidines (67), and the dissociation of the protein from the proteoglycan upon exocytosis from the mast cells into a neutral pH environment indicates that mMCP-7-heparin binding is weakened when histidine side chains are deprotonated (68).

The histidine rich region of the Histidine-rich glycoprotein (HRG or HRGP), is thought to act as a pH sensor and Zn^{2+} detector, helping regulate HRG activity. When the local pH drops, the histidine rich regions become highly positively charged, and in turn bind strongly to cell surface GAGs. This binding helps co-localize other ligands of HRG such as plasminogen and is thought to help displace other GAG-binding proteins such as Antithrombin.

pH-dependence of glycosaminoglycan binding to proteins appears to be a prevalent physiological phenomenon. The importance of the role of pH for these interactions can be understood in terms of the functional role of GAGs. GAGs are principally involved in cell adhesion, immobilization of proteins and maintenance of protein concentration gradient in regions of inflammation. Many signaling molecules such cytokines and chemokines bind to GAGs and initiate a series of signaling events. In conditions such as hypoxia and ischemia, when the pH of the affected tissue drops, this pH dependent binding likely acts as a sensor and initiates recovery measures. Additionally, chemokine-GAG interaction at site of inflammation, may be influenced by changes in pH (69).

Developments in Molecular Modeling of GAGs and Protein-GAG Interactions

Protein-glycosaminoglycan complexes are often refractory to crystallization, and as a consequence limited structural data is available for these interactions. This necessitates a use of alternative approaches for characterization of their 3D complexes, such as computational modeling. However, the presence of high negative charge density, structural heterogeneity, and

high conformational flexibility of GAGs, as well as the lack of well defined binding pockets on the surface of many GAG-binding proteins, make molecular modeling of GAGs and their interactions challenging (55, 70, 71).

A number of computational modeling studies have been attempted, and been reasonably successful, for modeling the binding of GAGs on protein surfaces. Prediction of the GAG binding sites have been made using the GRID algorithm (72), that probes protein surfaces to locate the most favorable binding positions for sulfate groups. This technique helps narrow down the general GAG binding site, and subsequent docking analysis may be performed to find the most optimal GAG binding pose. The technique has been used for studying the binding of heparin to aFGF, bFGF, Antithrombin and IL8 (73).

Molecular docking studies on several heparin binding proteins have been performed using a variety of docking software such as the AutoDock suite of docking programs (74–76), GOLD (77), Glide (78), and DOCK (79). The docking method can first be validated by predicting the heparin binding location on proteins, with known heparin binding sites, and then employed for making predictions for proteins where the binding site is not previously known, as was done in the case of IL8 (73). In the case of bFGF and FGFR1, the docking results were compared to experimental site-directed mutagenesis and biochemical cross-linking data (80). Docking has also been used to predict the binding of heparin to SDF -1α (81), MIP-1α (82), endostatin (83), PECAM-1 (84, 85), and endo-β-glucuronidase (86).

A novel combinatorial virtual screening approach was utilized for identifying high-specificity heparin/HS sequences that bind Antithrombin (87), using GOLD (77). A rigid docking approach for the glycosidic linkages, combined with flexible substituents at the 2-, 3-, and 6- positions, and conformational variability for IdoA residues was utilized. The filtering

strategy identified 10 heparin hexasaccharide sequences, from the 6859 that were screened, as high specificity binders. 9 of these bound in the known heparin pentasaccharide binding orientation, while one bound in a unique geometry and orientation.

Molecular dynamics (MD) simulations of protein-GAG complexes have only recently become more popular. This could be in part due to availability of better parameters for simulating GAGs better computational resources for carrying out simulations at relevant time scales, as well as an overall heightened interest in the field. In the past, MD simulations of heparin oligosaccharides in solution were carried out to determine the conformation of heparin in solution (88), and to explore the flexibility of IdoA residues and glycosidic linkages (89). Recently, several studies have combined molecular docking with MD simulations for studying the dynamics of protein-GAG interactions. These include the study of heparin binding to PECAM-1 and annexin A2 (90), modeling the interaction of HS to CXCL12α (91), binding of hyaluronan, CS, and DS to IL8 (92), and investigation of differences in binding of variably sulfated heparin fragments to CCL5 (69), to name a few. Binding free energy calculations using the MM-GBSA/PBSA methods, to investigate the nature of binding of certain GAG fragments to proteins, have also been reported (69, 90).

CHAPTER 3

MOLECULAR MODELING METHODS

Molecular modeling comprises of a variety of theoretical and computational techniques employed to mimic the behavior of molecules. These techniques are frequently utilized in the fields of computational chemistry, computational biology, and drug design, to study systems ranging in size from small compounds, consisting of a few atoms, to large multimeric biomolecules, consisting of thousands of atoms. This includes prediction of the binding orientations between molecules, their binding affinities, as well as the prediction of molecular motions inherent to their structure-function relationships.

An atomistic level description of molecular systems and the forces between the atoms is a general requirement for most modeling techniques. However, depending upon the motivation, and the application, the level of detail employed for the modeling may vary. While quantum mechanics (QM) explicitly models the electronic environment of each atom with detailed mathematical formulations, the computational expense associated with these calculations limits its applicability to systems with only a small number of atoms. Molecular mechanics (MM), on the other hand, uses classical mechanics to model systems defined in terms of simplified atomic models, therefore permitting the study of larger, more complex systems. In conjunction with basic Newtonian physics, molecular mechanics has been employed for performing molecular dynamics (MD) simulations of biomolecules (93–95), for predicting molecular motions on timescales relevant for elucidating their conformational behavior and generating structure-

function relationships. Molecular docking is another commonly used molecular modeling technique that is employed to predict the preferred binding orientation and affinity between two molecules.

Force field

Classical mechanics force fields are a mathematical formulation that describes the potential energy *V* of a system of atoms, and includes the parameter that define all the predefined values to be used with the formula to quantify the energy of the system.

The force field equation or the potential energy function is usually formulated as a summation of the bonded and non-bonded interactions. Equation 3.1 represents the form adopted by the AMBER/GLYCAM (96, 97) family of force fields.

$$V_{Total} = \sum^{Bonds} 1/2K_r(r - r_0)^2 + \sum^{Angles} 1/2K_\theta(\theta - \theta_0)^2 + \sum^{Dihedrals} \frac{V_n}{2}[1 + \cos(n\emptyset - \gamma]$$

$$+ \sum_{i<j}^{vdW} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] + \sum_{i<j}^{Electrostatics} \frac{1}{4\pi\varepsilon_0}\left[\frac{q_i q_j}{R_{ij}^2}\right] \qquad (3.1)$$

The equation represents a pair-wise summation over all the atomic interactions in the system. The bonds, angles, and dihedral angles, between two, three, and four, covalently attached atoms, represent the bonded interactions. The non-bonded terms are comprised of the vdW and electrostatic interactions.

Bonds and angle terms are modeled using simple harmonic functions, such as Hooke's law. The vdW interactions between two atoms represent the repulsive and attractive forces

between them, and are often modeled using the Lennard-Jones 12-6 potential. Electrostatic effects that represent the interactions between positive and negative charges (or dipoles) are commonly modeled with Coulomb's law using partial charges assigned to the atoms. However, other force fields that implement charge polarizability utilize more complex methodologies (98). The covalent 1-4 interactions represented by the dihedral angle terms, are generally included at the last stages of force field parameterization. This is done because ideally the other terms should be enough to reasonably compute the total potential energy of the system. However, due to the approximations made during the calculations of all the other terms, there may be some non-classical contributions with respect to torsional rotation. Thus, the dihedral angle term represents a quantum correction to the potential energy.

Force field parameterization is often validated by comparing the results from MD simulations to experimentally analogous data to assess the accuracy of the force field. This is an important step, because some parameters may need to be subsequently adjusted in order to reproduce experimental results. This validation is crucial for confirming the applicability of the force field to accurately predict molecular motion.

Molecular Dynamics Simulations

General Theory

The atomic models employed in classical mechanics MD simulations ignore electronic details and instead consider each atom to be a single particle with a fixed point-charge and a van der Waals (vdW) radius. Bonded interactions are modeled as "springs", and Newton's equations of motion are utilized to calculate the motions (trajectory) of a set of atoms, given initial starting

conditions of the system. The motion of the atoms is a time-dependent phenomenon, measured in discrete time intervals or steps (Δt).

A commonly employed numerical method used to integrate Newton's equations of motion over time (that is, to calculate the trajectories of the system under study) is the Verlet algorithm (99), given in Equation 3.2.

$$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + a(t)\,\Delta t^2 \tag{3.2}$$

According to Equation 3.2, an atom's future position, $x(t+\Delta t)$, can be predicted based on its current $x(t)$ and previous $x(t-\Delta t)$ positions and current acceleration, $a(t)$. While, during a simulation, the previous and current positions are known, the current acceleration must be calculated.

Newton's second law (Equation 3.3) relates the motion of an object to any external force acting upon it.

$$F = ma \tag{3.3}$$

Thus, if the force acting upon a particle is known, the acceleration of the particle can be calculated. Force can also be calculated as the gradient of the potential energy (Equation 3.4). The change in potential (∂V) as a function of a change in atomic position (∂x) can be readily generated from the force field, thus allowing the use of potential energy V to calculate the force, and in turn, the acceleration.

$$F = -\partial V/\partial x \tag{3.4}$$

During an MD simulation, the motion is thus propagated through the evaluation of the potential energy of every atom, via the force field equation, and derivation of the acceleration so

as to calculate the position of every atom at each time step Δt using the knowledge of the current and previous positions. Since, at the beginning of a simulation, the system is static, the initial velocities are assigned randomly based on a Maxwell-Boltzmann distribution appropriate to the system temperature.

MD simulation setup and general protocol

The starting 3D structure for an MD simulation may be obtained from an experimentally determined structure, a theoretical model, or a combination of the two. Depending on the information required from the MD simulation, molecules may be simulated *in vacuo* or in the presence of a solvent. Typically, for biomolecular simulations, interactions with the solvent provide a more biologically relevant environment. Solvent effects can be modeled either by employing the implicit solvent formulation that approximates the bulk solvent as a continuum, or employing explicit solvent models that utilize discrete water molecules. Implicit solvent schemes lack the ability to model discrete solute-solvent, and hydrogen bonding interactions that are important to mediate certain inter-molecular interactions. Explicit water models are more accurate at modeling the solvent effect, and are more commonly employed. Some of the popular water models are TIP3P (100), TIP4P (100), and TIP5P (101).

Energy minimization is typically performed before MD simulation, to eliminate large interatomic forces from steric clashes and unrealistic geometries in the starting structure. The energy minimization step usually finds the best nearby conformation, while reducing the net forces on the atoms in the molecule. Steepest descent (102) and conjugate gradient (103) are two popular energy minimization algorithms. While steepest descent minimization reduces energy in the direction of the steepest slope on the potential energy surface, with step sizes proportional to

the magnitude of the slope, conjugate gradient performs calculations in perpendicular directions to ensure better refinement towards the energy minima. To decrease computational expense, conjugate gradient minimization is performed after steepest descent.

Energy minimization is followed by steps that equilibrate the temperature and pressure of the system before data is collected from the MD simulation for analysis. MD simulations of biomolecules are often performed with a time step of 1-2 fs, and the positions, velocities, and energies of the system are collected at regular intervals.

Molecular Docking

Molecular docking is a computational method used to predict the binding orientation of one molecule to another, to form a stable complex. The smaller molecule is referred to as the ligand, and the larger molecule the receptor or target. This technique is frequently used for finding the binding modes of small molecules to proteins, such as in rational drug design (104), but methods to predict binding of two or more larger molecules, such as protein-protein docking have also been developed (105).

Before performing docking, target and ligand molecules need to be selected, their 3D structures obtained from available structure databases or modeled theoretically, and molecular files prepared according to the specifications of the docking software. These are given as input, and the results are analyzed to determine the highest scoring pose. Docking may be performed with the ligand or receptor maintained as a rigid molecule or allowed to be flexible. Rigid docking is faster, but flexible docking allows for a better fit in cases where the provided structure may not be in optimal conformation for binding. Typically, a search space is defined on the receptor surface to limit the potential interaction area, and minimize computation time. This is

usually done based on some experimental knowledge about residues that may be directly involved in binding, or information about putative binding sites based on similar macromolecular interactions. If no information is available about the interaction, "blind" docking may be performed, wherein the entire surface of the receptor may be specified as the potential binding area. Blind docking is slow and may be less accurate at identifying the correct binding mode.

Docking procedures make use of a search algorithm that explores the state variables: position, orientation, and conformation (in case of flexible-ligand docking) of the ligand with respect to the receptor, to find an optimal solution. Search procedures (106) may be systematic, where the search space is sampled at regular intervals in a deterministic fashion, or stochastic, where random changes to the state variables are made till the termination criterion is met. The candidate binding poses for the ligands are scored using a scoring function. The scoring functions may be empirical, force field based, or knowledge based.

CHAPTER 4

EXTENSION AND VALIDATION OF THE GLYCAM FORCE FIELD PARAMETERS FOR MODELING GLYCOSAMINOGLYCANS[1]

[1] Singh, A., Tessier, M.B., Pederson, K., Wang, X., Venot, A., Boons, G.J., Prestegard, J.H., and Woods, R.J. *To be submitted.*

Introduction

Glycosaminoglycans (GAGs) are linear polysaccharide, generally found covalently attached to proteins, forming a protein-class called proteoglycans that are widely present on the plasma membrane, in the extracellular matrix, and in secretory granules of all animal cells (5). GAGs can be classified into 5 main categories based on the unique composition of the polysaccharide: hyaluronan (HA), heparin/heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), and keratan sulfate (KS). The polysaccharides are typically composed of repeating units of a hexosamine-uronic acid disaccharide. The hexosamine may be an N-sulfated or an N-acetylated glucosamine (HS, HS and KS) or galactosamine (CA and DS), variably O-sulfated at the 3, 4 and/or 6 positions. The uronic acid may be a glucuronic acid or an iduronic acid formed as a result of enzymatic epimerization of a glucuronic acid at the C-5 position. These uronic acid moieties may also be 2-O-sulfated. KS lacks uronic acids and instead contains variably sulfated galactose residues and while most GAGs have a heterogeneous pattern of sulfation, hyaluronan is unsulfated GAG (4). GAG-protein interactions are critical in biological processes such as cell adhesion, anticoagulation, regulation of cell growth and proliferation, immobilization of proteins, maintenance of protein concentration gradient in regions of inflammation, viral invasion and tumor metastasis (4, 6–11). Each tissue produces a distinctive repertoire of GAGs that interact with proteins in a tissue specific manner. Most GAG binding proteins interact with heparin (11), because due to structural similarity, it mimics the interaction of these proteins with the widely abundant cell surface HS chains.

GAG sulfation patterns have been demonstrated to modulate biological function, for example in the cases of heparan sulfate in growth factor activation and cellular defense (6, 107), chondroitin sulfate growth factor recognition (108), and synthetic heparin/heparan sulfate in

anticoagulant activity (109). Differences in the sulfation pattern also alter the mode of interaction of heparin oligosaccharides with proteins, such at CCL5, where they have been shown to interact selectively with certain residues depending on the degree and pattern of sulfation (69). This property consequently alters their ability to inhibit the interaction of CCL5 to its receptor CCR1. These sulfation patterns, in addition to altering the charge, also impact the 3-dimensional structure of GAG fragments. Nuclear magnetic resonance spectroscopy (NMR) of GAG fragments has shown that sulfation patterns can alter the ring conformations of non-reducing terminal $\Delta^{4,5}$-unsaturated uronates (ΔUA) (110–112) and IdoA (40, 113, 114). Recently published simulations of HS GAGs have shown that IdoA ring-flipping can have a significant impact on the 3D shape of the GAG polymer (115), however, as noted earlier (43), not all differences in ring puckering lead to an altered overall shape (42).

The variable levels and patterns of sulfation make the structural analysis of GAGs a challenge. This often limits experimental characterization of GAG structures to composition-based analyses of digested fragments of native GAGs. More detailed analysis, such as by NMR or crystallography usually employ short, isolated, or synthetic oligomers, where the sulfation patterns are well controlled. Theoretical methods like molecular dynamics (MD) simulations have been widely used (93–95) to augment experimental methods in studying the conformational and binding properties of biomolecules, and the development of an accurate force field is key to the accuracy of these simulations. Molecular simulations, employing a consistent and validated force field not only provide a basis for interpreting experimental NMR data, but also enable structural analysis of polymers that are either too large for NMR analysis or too complex for routine synthetic preparation. Previously, existing carbohydrate force fields have been augmented in an *ad hoc* manner for examining specific sulfation patterns (116) and only recently

have parameters been developed for transferable sulfate moieties (117). In this work we add two key features to the GLYCAM force field to enable accurate MD simulations of sulfated GAG sequences with AMBER (118). The first addition is the creation of a generalizable sulfate parameter set to model N- and O-sulfation, including new bond, angle, and torsion terms, as well as partial atomic charges, consistent with existing GLYCAM partial atomic charges (97, 119, 120). The second is the development of force field parameters for ΔUA residues, which will permit simulation of the non-reducing terminal residue, typically resulting from use of bacterial heparinase to cleave GAGs. This unsaturated uronate residue is often present in GAG-protein crystal structures and in GAG fragments employed in experimental binding studies. In addition, parameters for neutral (NH_2) and protonated (NH_3^+) glucosamine, and protonated glucuronic and iduronic acids have been included.

To test the performance of the new parameters set, MD simulations were performed on variably sulfated GAG disaccharides containing ΔUA residues, and NMR scalar coupling and NOE measurements were collected for comparison with the theoretical data. MD and NMR data were collected for two synthetic GAG tetrasaccharides, with the aim of confirming the accuracy of the MD simulations and examining any influence of sulfation pattern on GAG conformation. The analysis presented here demonstrates that the new force field parameters reproduce the NMR data for a number of GAG fragments, both with and without terminal ΔUA. The simulations confirm the previous observation (40) that the IdoA ring populates two conformations ($^1C_4$ 63% and ^2S_O 37%), and surprisingly indicate that the terminal GlcA ring does not exclusively adopt the expected $^4C_1$ conformation.

Methods

NMR

NMR spectroscopy was carried out on a spectrometer operating at 18.8 T for disaccharides, and 14.0 T for tetrasaccharides, equipped with a Varian Inova console and a 5 mm cryogenically cooled probe. 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was included as an internal reference in each sample. NMR samples consisted of 0.5 mg of disaccharide in 100% D2O buffer containing 20 mM sodium phosphate and 1 mM DSS, pH 6.5. The sample was shimmed to a DSS linewidth < 1 Hz. Proton resonances were assigned using a standard COSY experiment (Varian ChemPack), processed with NMRpipe (121) and assigned in Sparky (122). 3J-coupling measurements were made from a 1D proton experiment with presaturation to suppress signal due to any residual H2O, collected with a spectral width of 9000 Hz and 32k points, processed and analyzed in MestReNova.

NOEs were measured using a standard 2D NOESY experiment (Varian ChemPack) with a mixing time of 0.4 s, 512 increments and 9000 points for disaccharides, and 0.3 s, 512 increments and 6000 points for tetrasaccharides, processed with NMRpipe. NOE peaks were integrated in NMRViewJ (123) and the distance was calibrated using the distance from the MD simulations between either the $H1^B$ and $H2^B$ or $H2^B$ and $H3^B$ protons on the disaccharide reducing terminal residue (residue B) for **1**-**5**. Tetrasaccharides, **6** and **7**, were calibrated using the distance from the MD simulations between H1 and H5 protons on glucuronate (residue C).

Calculation of Theoretical NMR properties

Theoretical nuclear Overhauser effects (NOEs) were calculated using the Isolated Spin-Pair Approximation (124), in which NOE intensity is assumed to be proportional to $1/R^6$, where R is

the distance between the two spin pairs. Three-bond proton-proton scalar couplings ($^3J_{HH}$) were calculated using Karplus-like equation developed by Haasnoot et al. (125) using the electronegativity values identified by Altona et al. (126) (Equation S4.1 & Table S4.3). Where relevant, experimental $^3J_{HH}$-couplings were decomposed into populations by least-squares fitting of the contributions from theoretical *J*-values computed for each individual state (127).

Molecular mechanics (MM) calculations

The SANDER program from the AMBER11 software package was used to compute the MM energies associated with the parameter development. None of the 1-4 non-bonded interactions were scaled, and torsions were restrained at their desired values with a restraint weight of 5000 kcal/mol•rad^2. A 12Å cut-off for non-bonded interactions was applied.

Molecular Dynamics Simulations

Initial structures for performing the MD simulations of the methyl glycosides for ensemble-averaged charge calculation were obtained from QM optimized models. Solutes were solvated with explicit TIP3P waters (100) with at least 12 Å buffer between the glycan solute and each edge of the solvated cubic box, using the LEaP module of AMER12 (118). Counter-ions were used to neutralize the net charge of each system. Energy minimization was performed under nVT conditions (500 steps steepest descent, followed by 24500 steps of conjugate-gradient minimization). Each system was then heated under nPT conditions for 50 ps, raising the temperature from 0 to 300 K, followed by 100 ps of equilibration while the temperature was maintained at 300 K. All simulations used periodic boundary condition where the pressure was maintained at 1.0 atm, the external dielectric was set to 1.0, and the system compressibility was

set to that of water. The Berendsen thermostat (128) was used for all temperature controls and the SHAKE algorithm (129) was used to constrain bonds with hydrogens, allowing a 2 fs timestep to be used. Non-bonded scaling factors were set to unity, and a 10.0 Å non-bonded cut-off was employed in all steps. Minimization and equilibration were performed using the PMEMD (130) implementation for CPU, in AMBER12. Subsequently, production simulations were performed with the PMEMD-Cuda (130) implementation for GPUs.

The simulations of heparin disaccharides **1**-**5**, and tetrasaccharide **6** and **7** were performed using a similar protocol, except for the minimization steps. The first minimization step was performed in Generalized Born implicit solvent (131) with an infinite non-bonded cutoff, prior to addition of counter-ions and explicit solvent. A second minimization step was performed after each system was explicitly solvated and neutralized.

Quantum mechanics (QM) calculations

All QM calculations were performed using the Gaussian 09 (132) software package.

Parameter development (partial charges)

Partial atomic charges were derived from the restrained electrostatic potential (RESP) charge fitting methodology (133). The ESPs for the small molecules employed in parameter development were computed from the lowest energy conformational state, at the HF/cc-pVTZ level of theory with a RESP weight of 0.0005. For anionic monosaccharides, ESPs were computed with diffuse functions at the HF/6-31++G**//HF/6-31++G** level, whereas for neutral and cationic monosaccharides, calculations were performed at the HF/6-31G*//HF/6-

31G* level, in each case a RESP weight of 0.01 was employed in order to be compatible with GLYCAM06.

Charge models for *N*- and *O*-sulfates; glucosamine; and ΔUA were developed using the standard GLYCAM ensemble-averaged charge method (119). The charges were developed for sulfates using 4-*O*- and 6-*O*-sulfated β-D-GalNAc, and both anomers of *N*-sulfated α- and β-D-glucosamine (D-GlcNS), using initial glycan geometries extracted from co-crystallized protein-sugar complexes. For the ensemble-averaged charge calculation, an initial QM optimized structure was used to derive single point RESP charges, and employed for 10 to 50ns of MD simulations, as required, for adequate sampling of exocyclic rotamers. From the simulations, one hundred evenly-spaced snapshots were extracted, as a representative ensemble of the 3D structures. Each of these geometries was subjected to QM-optimization with all torsion angles frozen in their MD-conformation. RESP charges were calculated for each frame and averaged to get the ensemble-averaged charge set for each particular molecule.

The computed charges for the sulfate moieties (SO_3^-) in both *N*- and *O*-sulfates were within statistical variance of each other, allowing the creation of an interchangeable sulfate residue. Examination of the sulfated sugars revealed similar atomic charges on the sulfated and non-sulfated atoms in GLYCAM06 (97). The most significant deviation between them was associated with O or N atom at the point of sulfate attachment. Consequently, for transferability, the charge on the linking heteroatom was adjusted as necessary to achieve a net integer charge on each sulfated sugar (Table S4.4).

Charges for protonated α- and β-D-glucosamine ($GlcNH_3^+$) were similarly developed, and found to significantly vary from the GLYCAM charges for α- and β-glucose and N-acetyl-glucosamine, particularly for the ring carbon atoms (Table S4.4). This variation suggests that

such analogs require unique charge sets for each monosaccharide, which is not surprising, as the positively charged site is directly adjacent to the sugar ring.

Charges for the ΔUA monomers were obtained from by averaging the charges for each of the low-energy half-chair states, $^{1}H_{2}$ and $^{2}H_{1}$, (**Figure 4.1**a and Table S4.4).

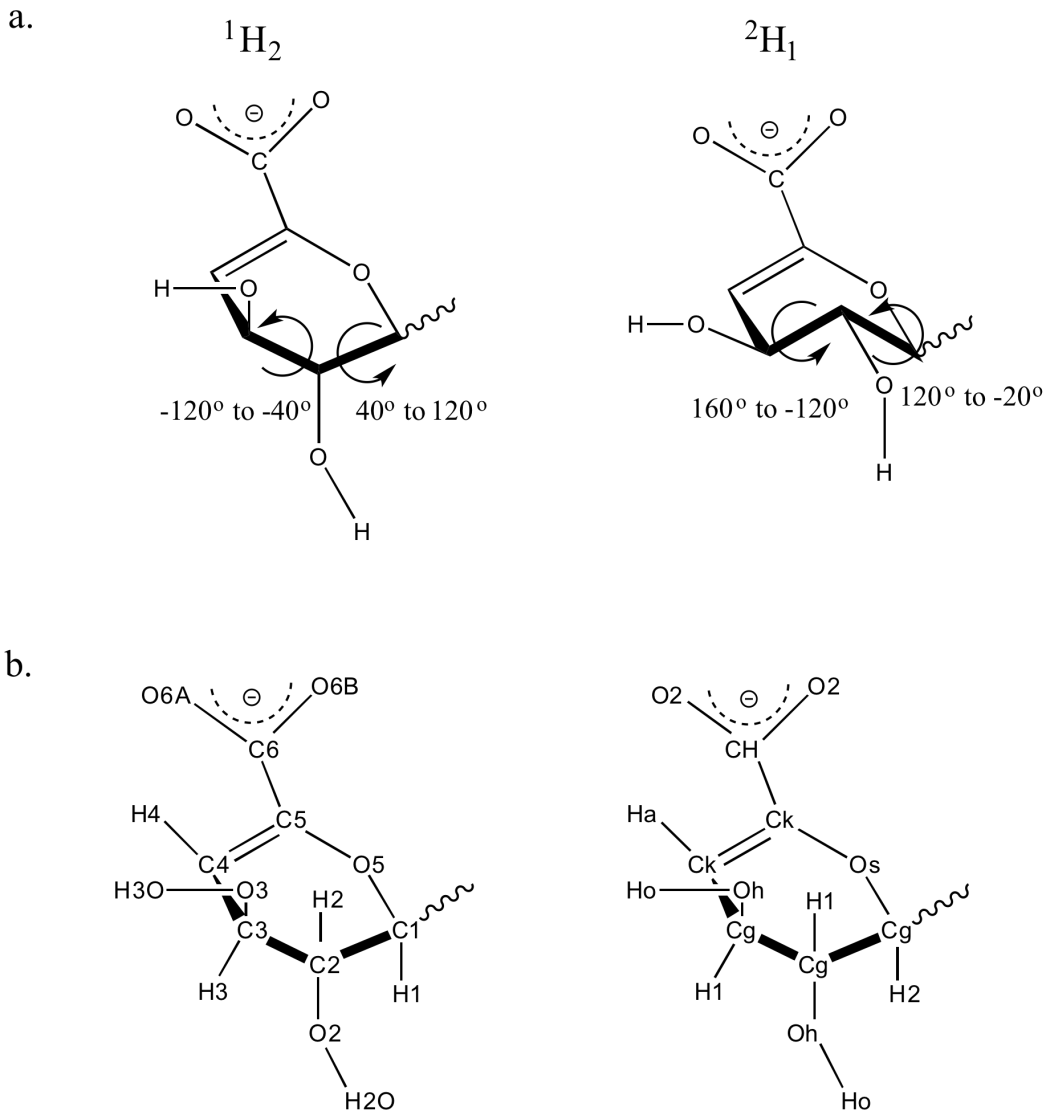

Figure 4.1 a. Δ4,5-unsaturated uronate ring conformations with the torsion angle ranges typically associated with H_1-C_1-C_2-H_2 and H_2-C_2-C_3-H_3 atomic sequences. b. The atom names (left) and atom types (right), employed in GLYCAM for Δ4,5-unsaturated uronates.

Parameter development (atom, bond, angle, and torsion parameters)

The GLYCAM06 force field for carbohydrates (97) and lipids (120) was adapted to include new terms required to model the double bonds found in unsaturated uronic acids. The only new atom type added in this work was the sulfate sulfur (S) atom, for which the van der Waals parameters were transferred from the sulfate atom type (S1) found in parm99 (96). All valence and torsion terms were developed using the hierarchical development procedure outlined in two prior GLYCAM06 publications (97, 120), wherein bond terms are developed first, followed by angle, and torsion terms (Table S4.1 and Table S4.2). Small molecules were selected for parameter development, such that, each contained as few new terms as possible while maintaining an electronic environment relevant to the carbohydrate (97). Equilibrium values for bonds and angles were obtained from the averages of crystal structures found in the Cambridge Structural Database (134) with molecule IDs HEMKEP, KOCOJ, SRHXGU, MIZFUX, GUVFOS, GUVFEI, GUVFAE & ZULPIF (Table S4.1), force constants were derived by fitting to QM data computed at the B3LYP/6-31++g(2d,2p)//HF/6-31++g(2d,2p) level.

Torsion potentials were generated for the relevant bonds in the molecules found in Table S4.2. Rotations were sampled in 30° increments with the exception of terms describing double bond rotations, in which only 0, 90 and 180° orientation were used to characterize the *cis/trans* relative energies and the barrier height between them. All torsion terms were developed without the use of a phase shift adjustment. Exocyclic torsion terms were developed using tetrahydropyran or its unsaturated analog of ΔUA. In the case of *N*-sulfate parameters, planarity of the nitrogen atom was maintained during the QM torsion rotation to reflect solution conformations of an sp^2-hybridized nitrogen.

In the development of the torsion terms associated with the unsaturated bond, a better fit to experimentally observed rotamer preferences required the use of a higher level of QM theory (second order Moller-Plesset, MP2) during geometry optimization. This was the case for rotation about the central Cg-Os bond in the Os-Cg-Os-Ck sequence (**Figure 4.1**b), where the terminal Os-Cg bond did not favor the experimentally observed rotamers at the HF-level. This behavior was corrected by performing geometry optimizations at the MP2 level. The need for a higher QM level that includes electron correlation may reflect the presence of hyperconjugation between the oxygen atoms (Os) and the unsaturated carbon center (Ck). Having observed a dependence of rotamer preference on level of QM theory for this term, all other terms were re-examined, and found not to show any notable dependence on the QM level. The energy contributions to the barrier for *cis/trans* rotation in double bonds was distributed equally between heavy and light atom terms, Cg-Ck-Ck-C and Ha-Ck-Ck-C, avoiding the need for improper torsions (120).

Tuning torsion terms to reproduce solution populations for $\Delta^{4,5}$-unsaturated uronates

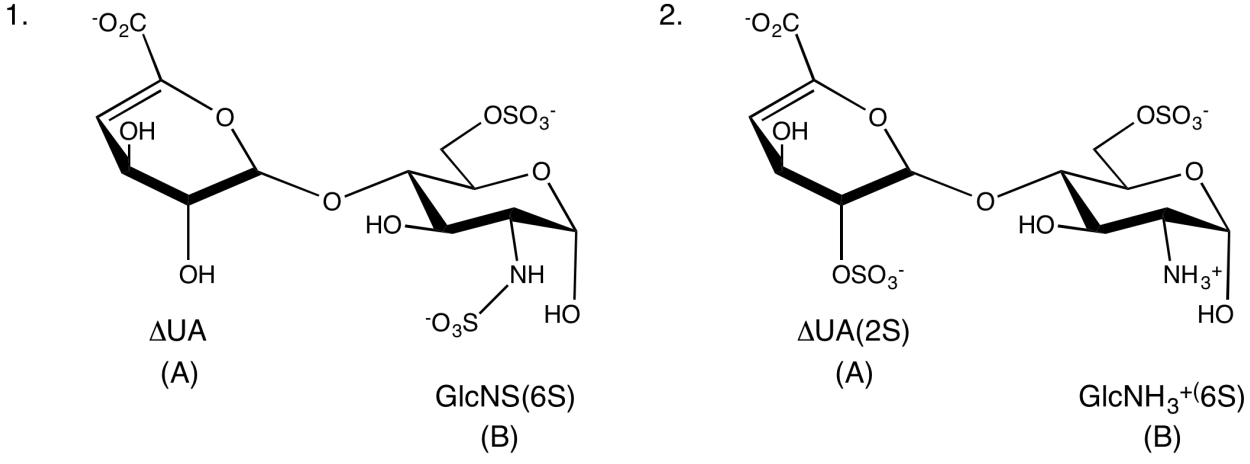

Figure 4.2 Schematic structures of GAG disaccharides (**1** and **2**), ΔUA monosaccharide labeled A, and glucosamine residue B.

MD simulations (100 ns) with the preliminary parameters were collected for disaccharides **1** and **2** (**Figure 4.2**), which contain ΔUA residues. An analysis of the populations of the $^1H_2$ and $^2H_1$ ring states ($^1H_2$:$^2H_1$ = 70:30 and 35:65, respectively) showed poor agreement with the NMR-derived populations, 40:60 and 69:31, respectively. As the partial atomic charges in the ΔUA residue had been derived under the assumption of an equal population of half-chair states, this appeared to be a potential source of error. However, the populations from MD simulations, in which the contribution of the partial charges from each half-chair was varied from 0% $^1H_2$ to 100% $^1H_2$, were relatively insensitive to the atomic charges.

Table 4.1 NMR J-couplings and ring state populations for the ΔUA residue in **1** and **2**.

	1		**2**	
^3J-coupling[a] (Hz)	NMR	Optimized Theoretical	NMR	Optimized Theoretical
H_1-C_1-C_2-H_2	5.5	4.6	3.4	3.4
H_2-C_2-C_3-H_3	4.9	5.0	2.8	3.4
H_3-C_3-C_4-H_4	3.8	3.4	4.4	4.4
Pop. ($^1H_2$: $^2H_1$)	40:60	42:58	69:31	67:33
*NOE's (*Å*)*				
H_1^A-H_3^B	3.0	4.2	3.1	4.4
H_1^A-H_5^B	-	3.9	2.8	3.9

[a] ^3J-couplings for H1-H2, H2-H3, and H3-H4 in the $^1H_2$ ring form are 1.7, 1.2, and 5.8 Hz; whereas for the $^2H_1$ form the couplings are 6.7, 7.8, and 1.6 Hz. The *J*-values were independent of the anomeric configuration (α or β) at the reducing terminus.

Subsequently, the ring torsion terms (Oh-Cg-Cg-Ck, **1;** Os-Cg-Cg-Ck, **2**) were iteratively adjusted so as to obtain optimal agreement with the NMR populations. MD simulations with the optimized torsion terms, yielded average population ratios for **1** and **2** of 42:58 and 67:33 ($^1H_2$:$^2H_1$), respectively, that were then in good agreement with experimental values (**Table 4.1**). All subsequent simulations employed these parameters

Results and Discussion

Conformational analysis of $\Delta^{4,5}$-unsaturated uronate disaccharides (3-5)

GAG disaccharides **3-5** (**Figure 4.3**) were analyzed using NMR and MD simulations, to validate ring conformational populations and glycosidic linkage geometry profiles obtained using the new parameters. Ring conformations and populations were determined from homonuclear $^{3}J_{HH}$-couplings, while NOEs were collected to characterize the global 3D shape of these GAG fragments.

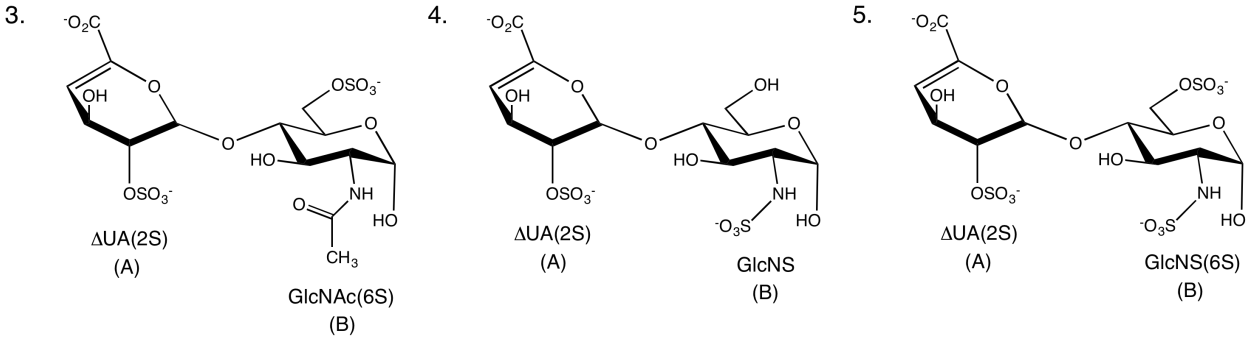

Figure 4.3 Schematic structures of GAG disaccharides (**3-5**).

Ring state populations for the ΔUA residue from MD simulations showed a preference for the $^{1}H_{2}$ state in all cases, consistent with the populations derived from NMR *J*-couplings (**Table 4.2**). Conformational analysis of the ^{3}J-couplings for the GlcNx residues (data not shown) was consistent with the $^{4}C_{1}$ conformation exclusively.

NMR characterization of the glycosidic linkages was provided by an analysis of $H1^{A}$-$H3^{B}$ and $H1^{A}$-$H5^{B}$ proton-proton NOE contacts observed for trans-glycosidic interactions. Comparison of the theoretical and NMR derived NOE distances for these protons shows agreement within 0.6 Å in **4** and **5** (**Table 4.2**).

Table 4.2 Experimental observables for residue A of the ΔUA containing disaccharides **3-5**.

Residue **A**	**3**		**4**		**5**	
^{3}J-coupling[a] (Hz)	NMR	Theoretical	NMR	Theoretical	NMR	Theoretical
H1-H2	3.4	2.7	3.0	2.7	3.7	2.9
H2-H3	-	2.5	2.5	2.5	2.8	2.8
H3-H4	4.7	5.0	4.7	5.0	4.4	4.8
Pop. ($^{1}H_2$: $^{2}H_1$)	70:30	80:20	76:24	80:20	67:33	76:24
NOE's (Å)						
$H1^{A}$-$H3^{B}$	-	3.5	2.7	3.3	2.4	2.9
$H1^{A}$-$H5^{B}$	-	3.5	3.1	3.6	-	3.3

[a] *J*-couplings for H1-H2, H2-H3, and H3-H4 in the $^{1}H_2$ ring form are 1.7, 1.2, and 5.8 Hz; whereas for the $^{2}H_1$ form the couplings are 6.7, 7.8, and 1.6 Hz. The *J*-values were independent of the anomeric configuration (α or β) at the reducing terminus.

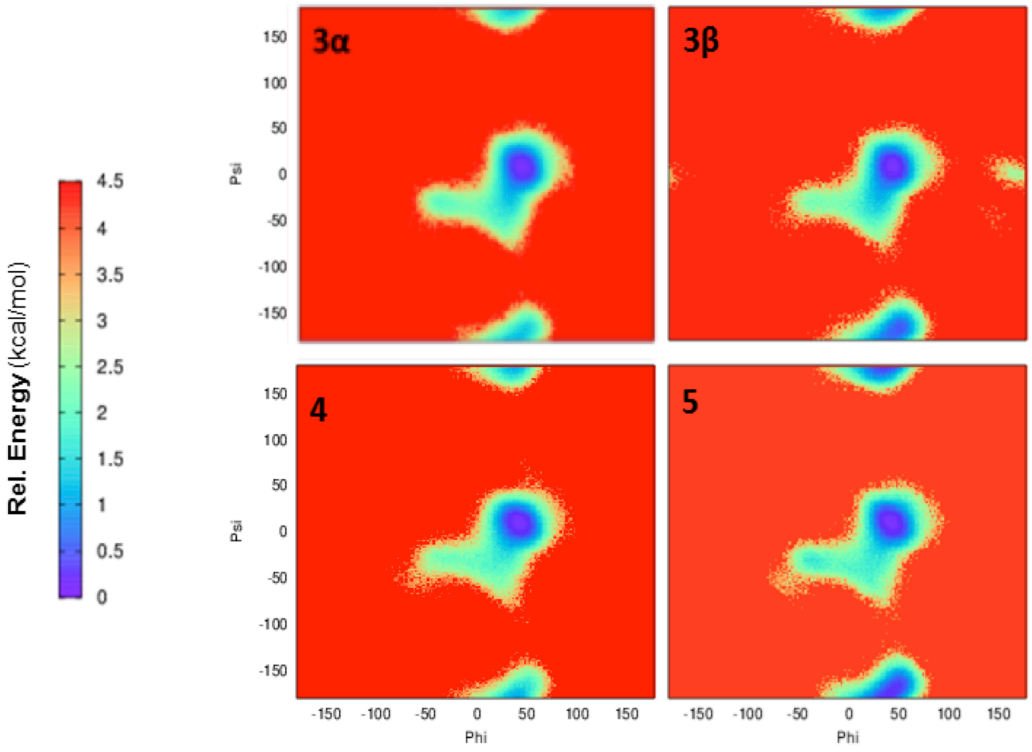

Figure 4.4 MD solution relative free energies for glycosidic rotamers based on Boltzmann weighted populations at 300K. Most regions are not sampled (red colors) and have >4 kcal/mol relative energy. All models show a similar global minimum around 50° and 0° for φ and ψ, respectively.

The φ, ψ population distribution heat maps presented in **Figure 4.4** show the global minimum from the MD simulations around is approximately φ = 50°, ψ = 0° for all the linkages. Each disaccharide also shows a second stable anti-ψ state near φ = 50°, ψ = 180°.

Conformational analysis of heparin tetrasaccharides

The conformational properties of two synthetic GAG tetrasaccharides, differing only in the presence (**7**) or absence (**6**) of N-sulfation (**Figure 4.5**) were characterized by NMR. To deconvolute the NMR data, MD simulations were performed on **6** and **7** for each of the three common IdoA ring conformations ($^{1}C_{4}$, $^{2}S_{O}$ and $^{4}C_{1}$) in each tetrasaccharide.

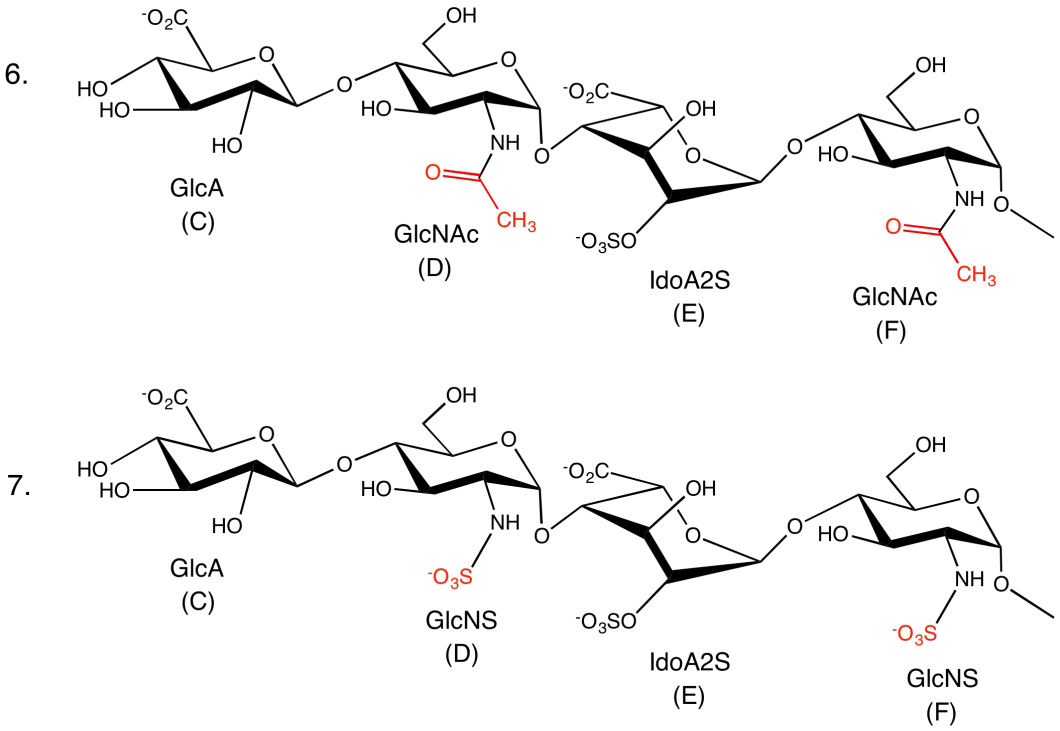

Figure 4.5 Schematic structures of GAG tetrasaccharides (**6** and **7**). Labels C-F are used to identify the monosaccharide residues.

Ring Conformational Analysis

During the 1 μs MD simulations, residues C, D, and F each populated only the $^4C_1$ ring conformation, and back calculation of the *J*-couplings led to agreements with the experimental values within 0.5 Hz for D and F (**Table 4.3**). However, the theoretical H1-H2 *J*-values for the terminal C residue (9.8 Hz) in both **6** and **7** were larger than those observed experimentally by almost 2 Hz. Given the otherwise close agreements, this suggests that the MD simulation may not have detected all of the conformations adopted by the C-rings, despite the relatively long simulation time. In the $^4C_1$ conformation, protons H1 and H2 in the β-anomer of the GlcA ring are *anti* to each other, leading to a large *J*-coupling, whereas in the $^1C_4$ conformation they would be *gauche*, leading to minimal coupling, and a mixture of approximately 80% $^4C_1$ and 20% $^1C_4$ would explain the observed *J*-value in the C residue. However, in the absence of further experimental data, this is not necessarily a unique solution. It is notable that, at least in the case of a fully sulfated GlcA residue, NMR data indicated that the uronate preferred to adopt the $^1C_4$ conformation rather than the $^4C_1$ (135).

Table 4.3 Theoretical and experimental ^3J-couplings for **6** and **7**.

		6				**7**			
Residue	^3J-coupling	Theoretical ($^4C_1$)[a]			NMR	Theoretical ($^4C_1$)[a]			NMR
C	H1-H2	9.8			7.9	9.8			7.9
D	H1-H2	3.5			3.6	3.4			3.6
D	H2-H3	10.1			10.6	10.1			10.2
F	H1-H2	3.5			3.6	3.4			3.5
F	H2-H3	10.0			N/A	10.0			10.2
F	H4-H5	10.0			N/A	10.0			9.5
		$^1C_4$	^2S_O	$^4C_1$	NMR	$^1C_4$	^2S_O	$^4C_1$	NMR
E	H1-H2	1.7	8.0	9.9	0.9	1.7	7.9	9.9	3.2
E	H2-H3	1.8	10.0	9.7	N/A	1.8	10.0	9.7	5.9
E	H3-H4	2.0	6.6	10.0	N/A	2.0	6.6	10.0	3.7
E	H4-H5	3.4	4.5	4.0	2.4	3.4	4.5	4.0	2.7

[a]Only the $^4C_1$ conformation was sampled during the MD simulations

Optimal agreement between the experimental and theoretical *J*-values for the IdoA (residue E) ring structure in **7** was achieved using least square fitting analysis of the contributions from multiple ring forms, resulting in a population distribution of 63:37 ($^1C_4$: ^2S_O) with no contribution from $^4C_1$. The absence of the $^4C_1$ state is supported by NMR data for similar GAGs (40, 42, 43, 89, 136), which indicate this state to be the least populated of the three, if present at all. A search of the Protein Databank (PDB) (137) revealed the $^1C_4$ (73%) and ^2S_O (24%) states to be the dominant forms of IdoA. For **6**, only the $^3J_{H1H2}$ and $^3J_{H4H5}$ couplings were experimentally observed (0.9 Hz and 2.4 Hz respectively), and both were below the theoretical values computed from any of the ring conformations. Nevertheless, the small value of the experimental H1-H2 coupling indicates that there cannot be significant amounts of either the ^2S_O or $^4C_1$ conformations present.

Inter-residue Conformational Analysis

Table 4.4 Theoretical and experimental inter-ring NOE distances for **6** and **7**, measured in Å.

		6			**7**		
C	D	Theoretical		NMR	Theoretical		NMR
H1	H4	2.3		2.8	2.4		2.9
H1	H61/2	2.9		2.8	2.7		2.8
D	E	$^1C_4$	^2S_O		$^1C_4$	^2S_O	
H1	H2	4.6	4.8	4.2	4.9	4.8	N/A
H1	H3	2.4	2.7	2.5	2.5	2.7	2.7
H1	H4	2.5	2.5	3.3	2.3	2.6	2.7
H5	H4	3.1	3.2	2.7	3.4	3.2	N/A
E	F	$^1C_4$	^2S_O		$^1C_4$	^2S_O	
H1	H3	3.2	4.3	N/A	3.1	3.2	2.6
H1	H4	2.3	2.3	2.7	2.3	2.4	2.6
H1	H61/2	3.0	3.0	2.7	2.9	2.9	2.7

The theoretical inter-proton distances (**Table 4.4**) showed agreement to within 0.5 Å of the NMR-derived values for all but the distance between protons H1 and H4 (0.8 Å) in residues D and E of **6**. The theoretical distances were very similar between the $^{1}C_{4}$ and $^{2}S_{O}$ conformations for the NOEs between residues D and E, and E and F indicating that these IdoA ring conformations do not have a significant influence on the overall shape of the tetrasaccharide. This has previously been determined to be the case in NMR structures of heparin octasaccharides (42).

To understand the role of N-sulfation on the glycosidic linkages, heat maps were plotted for the φ versus ψ values for the three glycosidic linkages in each trajectory (**Figure 4.6**). The glycosidic linkages between residues E and F showed very similar distributions for both **6** and **7**. In addition to the observed major conformation (φ ≈ 0° to 60°, ψ ≈ -60° to 60°), this linkage also sampled both the *anti-exo* (φ ≈ -90° to 0°, ψ ≈ -60° to 0°) and the *anti*-ψ (ψ ≈ -150° to +150°) states. Only the simulation of **6** restrained in the $^{2}S_{O}$ conformation did not sample the anti-ψ state for this linkage. The percent distribution of each state is presented in **Table 4.5**.

The linkage between residues C and D in both **6** and **7** also showed very similar distribution, with an additional sparsely populated *anti*-φ state (φ ≈ -150° to +150°) for the $^{1}C_{4}$ conformation. This state was also observed in **7** in the $^{4}C_{1}$ simulation, but not in **6**. Additionally, the anti-ψ orientation was missing for **6** in $^{2}S_{O}$ and anti-φ was missing for **7** in $^{2}S_{O}$. Overall, N-sulfation had little impact on the preferences of the glycosidic linkages, with the possible exception of the E-F linkage in **7**, where regardless of the conformation of the IdoA ring, there appeared to be a modest increase in the population of the anti-ψ conformation (shaded cells in **Table 4.5**).

Table 4.5 Percent distribution of the φ versus ψ states sampled by the glycosidic linkages during of **6** and **7** during the MD simulations performed with each conformation of IdoA.

Tetrasaccharide	Conformation	Linkage	Exo	Anti-exo	Anti-ψ	Anti-φ
6	1C4	C-D	97.6	2.2	0.1	0.1
		D-E	99.0	1.0	0.0	0.0
		E-F	92.4	2.2	5.4	0.0
	2SO	C-D	97.2	1.9	0.0	0.9
		D-E	99.3	0.7	0.1	0.0
		E-F	97.5	2.5	0.0	0.0
	4C1	C-D	95.9	2.5	1.7	0.0
		D-E	99.3	0.7	0.0	0.0
		E-F	94.8	2.1	3.1	0.0
7	1C4	C-D	87.3	4.1	6.3	2.3
		D-E	94.7	5.3	0.0	0.0
		E-F	87.9	3.9	8.2	0.0
	2SO	C-D	95.0	4.5	0.5	0.0
		D-E	92.6	0.4	7.1	0.0
		E-F	88.1	4.5	7.4	0.0
	4C1	C-D	94.8	4.4	0.5	0.2
		D-E	98.9	1.1	0.0	0.0
		E-F	87.9	3.3	8.8	0.0

For the GlcNx (D) and IdoA (E) linkage (**Figure 4.6**b), each ring shape showed a slightly different distribution of glycosidic angles. While the $^{2}S_{O}$ conformation showed the tightest distribution around φ ≈ -50°, ψ ≈ -50°, $^{1}C_{4}$ showed a wider spread of the φ angle and $^{4}C_{1}$ a wider spread of the ψ angle. Each of the three also showed a sparsely populated distribution around φ ≈ 40°, ψ ≈ 0°. For the simulation with $^{2}S_{O}$ conformation, the D-E linkage also sampled the anti-psi state for both 6 and 7, albeit differing in the percentage distribution.

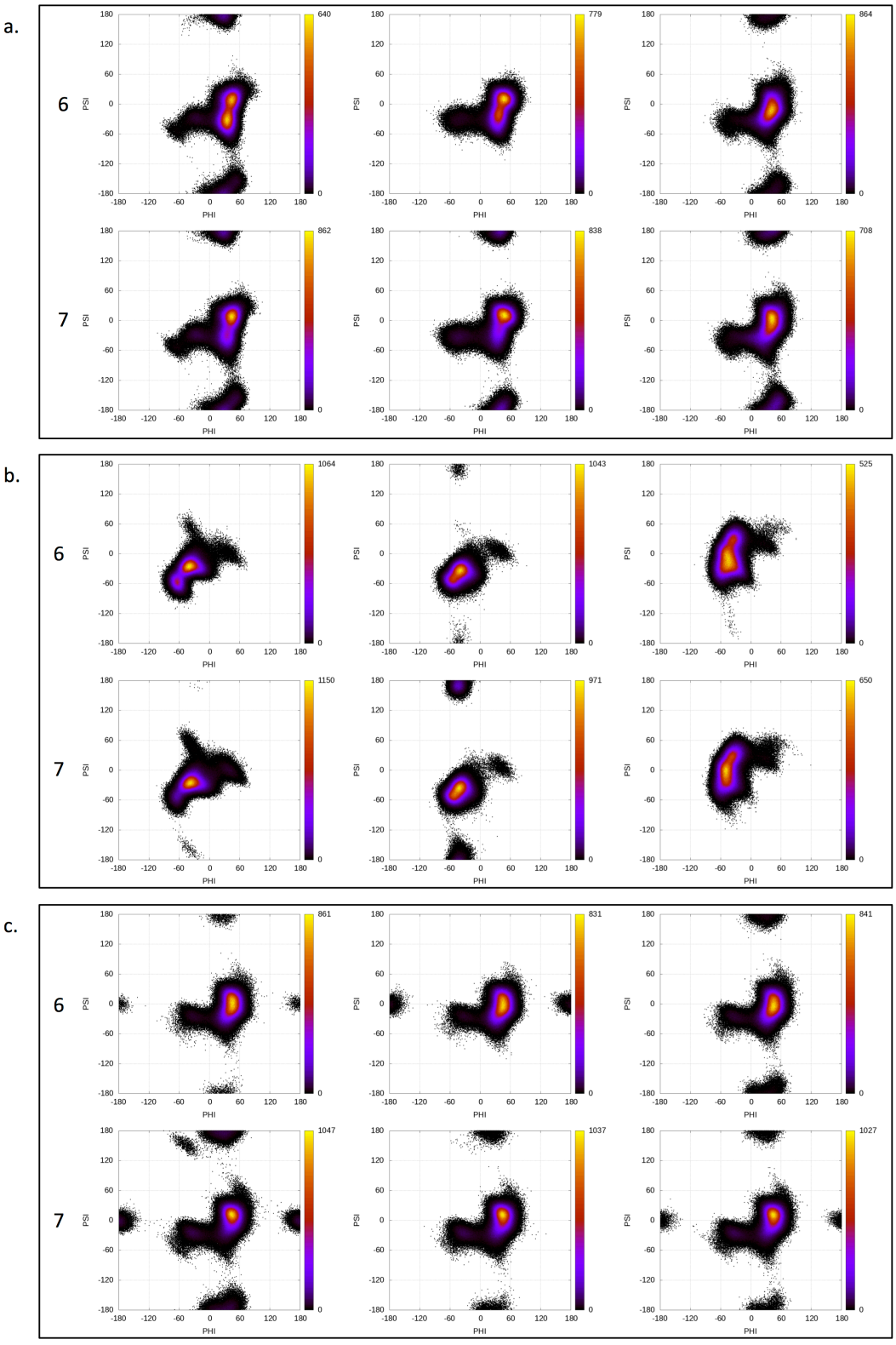

Figure 4.6 Heat maps for φ versus ψ angles for the glycosidic linkages between C-D (a.), D-E (b.) and E-F (c.) for **6** and **7**.

Conclusions

A new parameter set for GAGs containing iduronic acid, Δ4,5-unsaturated uronate, sulfate, as well as protonated glucuronic and iduronic acids has been added to GLYCAM. The development of a transferable sulfate model allows it to be used for multiple attachment points without a need for development of separate charge sets. In addition, development of a single model for ΔUA that reproduces solution conformations permits more accurate modeling of these residues.

The performance of the new parameters set was tested by performing MD simulations on variably sulfated GAG disaccharides containing ΔUA residues, and two synthetic GAG tetrasaccharides. NMR scalar coupling and NOE measurements were collected for comparison with the theoretical data, with the aim of verifying the accuracy of the MD simulations and examining any influence of sulfation pattern on GAG conformation.

Unrestrained simulations of ΔUA on timescales that allowed direct parameterization of the ring populations were performed. Analysis of NMR J_{HH}-couplings showed that the conformation populations of the ΔUA ring is largely insensitive to the adjacent sulfation patterns and the *N*-substituent; however, presence (**2**-**5**) or absence (**1**) of 2-*O*-sulfation on ΔUA altered the favored geometry.

The most notable effect of sulfation, in the case of GAG tetrasaccharides, was observed on the ring geometries for IdoA. Examination of the NMR data showed that tetrasaccharide **6**, which contained 2-*O*-sulfated IdoA but no *N*-sulfated glucosamine residues, exclusively favored the $^{1}C_{4}$ conformation while **7**, which contained two *N*-sulfated glucosamine residues adjacent to the 2-*O*-sulfated IdoA, sampled a substantial $^{2}S_{O}$ population (37%). While the ring flip dynamics were not captured by this work, long timescale simulations of IdoA using GLYCAM have previously shown experimentally-consistent ring populations (138). The $^{3}J_{HH}$-coupling

analysis also suggested that the terminal GlcA ring may not exclusively adopt the expected $^4C_1$ conformation.

These parameters and related structure files are available for download from the GLYCAM website (www.glycam.org) and are included in the new 3D structure building utility, the "Glycosaminoglycan Builder", on GLYCAM-Web (www.glycam.org/gag).

CHAPTER 5

GLYCOSAMINOGLYCAN BUILDER

Introduction

Carbohydrates or glycans play a vital role in biological systems: storing and providing energy, providing structural elements of cell walls, enabling cell-cell recognition through their presence on cell surfaces, providing mechanism for host-pathogen interaction, being markers for diseases, and sequestering cells for tissue repair, growth, and immune response, to name a few (4). Understanding glycan structures, activities, as well as modes of interaction are vital for designing modulators for various biological processes that involve interaction with carbohydrates. Polysaccharide-protein complexes are often refractory to crystallization, and therefore require alternative approaches for the characterization of their 3D complexes. Computational modeling methodologies are widely employed to obtain atomic level understanding of interactions between biomolecules; and with the development of more accurate force fields and availability of better computational resources, modeling techniques have become more effective and performing longer time scale simulations feasible. Several theoretical analyses of protein-carbohydrate interactions have recently been shown helpful in supporting or elucidating experimental findings such as binding specificities of certain antigens (139, 140) and provide possible structural explanation for experimental differences in the inhibitory capabilities of heparin fragments towards the chemokine CCL5 and receptor CCR1 binding (69).

A special class of carbohydrates, glycosaminoglycans (GAGs) are linear polysaccharide molecules, predominantly found covalently attached to proteins, and typically composed of a hexosamine and a uronic acid pair. GAGs play important roles in cell adhesion, hemostasis, anticoagulation, regulation of cell growth and proliferation, immobilization of proteins, maintenance of protein concentration gradient in regions of inflammation, regulation of enzyme activity, lubrication, viral invasion and tumor metastasis (4, 11). Each tissue produces a distinctive repertoire of GAGs that interact with numerous proteins in a tissue specific manner.

Based on the unique disaccharide composition, GAGs can be divided into 5 classes: hyaluronan (HA), heparin/heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), and keratan sulfate (KS). The hexosamine may be an N-sulfated or an N-acetylated glucosamine or galactosamine, variably O-sulfated at the 3, 4 and/or 6 positions. The uronic acid may be a glucuronic acid or an iduronic acid, formed as a result of epimerization at the C-5 position. These uronic acid moieties may be 2-O-sulfated. Keratan sulfate lacks uronic acids and instead contains variably sulfated galactose residues. Most GAGs have a heterogeneous pattern of sulfation except hyaluronan, which is the only unsulfated GAG.

Recently the GLYCAM06 force field has been augmented to include GAG-specific parameters: iduronic acid, O- and N-sulfate, $\Delta^{4,5}$-unsaturated uronate, as well as protonated uronic acids. In order to facilitate ease of specifying and building GAG structures, the carbohydrate modeling tools at GLYCAM-Web (www.glycam.org) have been expanded to provide a separate, specialized user-interface: the Glycosaminoglycan Builder (http://glycam.org/gag). The GAG Builder provides an easy, point-and-click interface for the user to specify sequences of choice and obtaining molecular structure files for visualization, automated docking, and MD simulations. It provides separate sets of monosaccharides, unique

to each class of GAGs (**Table 5.1**), allowing easy selection of pre-sulfated options for each monosaccharide as well as eliminating chances of errors.

Table 5.1 Unique monosaccharides and their sulfated derivatives that constitute each class of GAGs.

GAG class	Hexosamine	Uronic acid
Heparin/Heparan Sulfate	GlcNAc; may be O-sulfated at 3, 6 or both GlcNS; may be O-sulfated at 3, 6 or both	GlcA, GlcA(2S), IdoA, IdoA(2S)
Chondroitin Sulfate	GalNAc; may be O-sulfated at 4, 6 or both	GlcA, GlcA(2S)
Dermatan Sulfate	GalNAc; may O-sulfated at 4, 6 or both	GlcA, GlcA(2S), IdoA, IdoA(2S)
Keratan Sulfate	GlcNAc; may be O-sulfated at 6	Gal, Gal(6S)
Hyaluronan	GlcNAc	GlcA

In addition, $\Delta^{4,5}$-unsaturated uronic acid and its sulfated derivative are provided as options along with the uronic acids for the non-reducing terminus for all GAG classes (but KS). This residue is formed as a result of heparinase digestion, frequently employed to obtain shorter oligosaccharides from long GAGs chains. Many biochemical studies are carried out with oligosaccharides that contain this residue, and it is present in several GAG-protein crystal structures. Thus to model the experimentally observed GAG-protein interactions accurately, it may be important to employ this residue in modeling studies.

The user also has the choice of using an alternate conformation for one or more iduronic acid (IdoA) residues. The default conformation for IdoA residue and its derivatives is the most abundant solution conformation, $^{1}C_{4}$, but the option to model the less predominant $^{2}S_{O}$ and $^{4}C_{1}$ conformations is also provided. The interface, the build-process details, as well as the implementation for the GAG Builder are discussed below.

User Interface and Workflow

The main page of the builder (**Figure 5.1**) provides the user with option to choose from one of the five classes of GAGs. Based on this choice, two sets of monosaccharides, specific for that class of GAGs (**Table 5.1**), are displayed: one for the hexosamines and their sulfated derivatives, the other for uronic acids and their derivatives (galactose and its derivatives in case of KS).

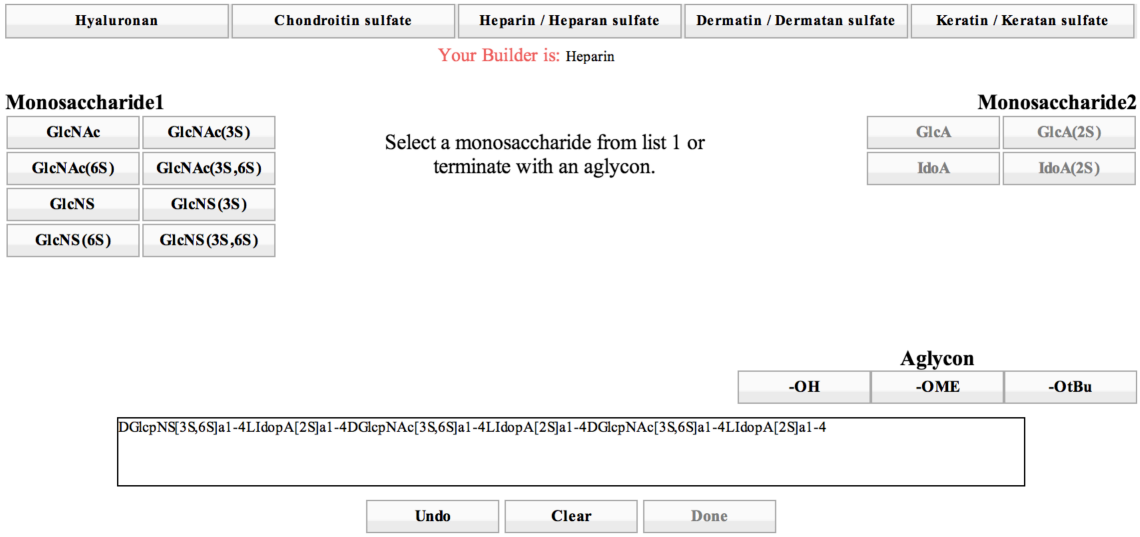

Figure 5.1 Interface of the Glycosaminoglycan Builder on GLYCAM-Web.

The user can choose to build the structure starting from either of these two classes; this will form the non-reducing end of the glycan. Once the first sugar is selected, that set of monosaccharides is disabled from selection and the user chooses one sugar from the other set. The user continues to build the sequence by alternating between the two sets of monosaccharides till they choose to end the sequence using one of the three terminal residues: OH, OMe or OtBu. Users can change the GAG class mid sequence, and opt to include sugars from other classes of GAGs or from the carbohydrate builder to suit their requirements.

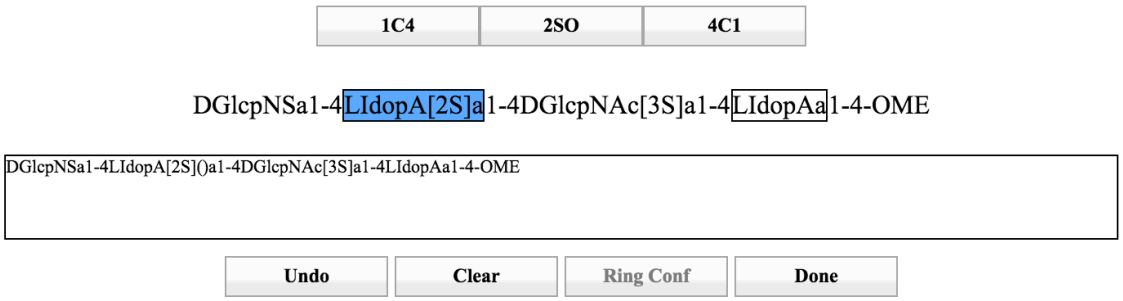

Figure 5.2 View of the GAG Builder showing the option to choose alternate conformations of IdoA residues.

Once the sequence is submitted, the user has the choice of using an alternate conformation for one or more iduronic acid (IdoA) residues, the default being $^{1}C_{4}$, with the option to model the less predominant $^{2}S_{O}$ and $^{4}C_{1}$ conformations (**Figure 5.2**).

Upon completion of the sequence specification process, when the user clicks "done", the process continues to the options page (**Figure 5.3**). This page provides the option of adding ions and/or solvent. The GAG is modeled using the structure prep files library, part of the GLYCAM force field. The structure is energy minimized to remove any bad contacts under implicit solvent conditions with a dielectric constant of 80. If the user indicates to add ions, the GAG is neutralized by the addition of appropriate counter ions (Cl^{-} or Na^{+}). If a solvent box is desired, the system is solvated using a TIP3P water model (100). PDB format files are then made available for download, both with and without the ions/solvent. The entire build process is summarized in **Figure 5.4**.

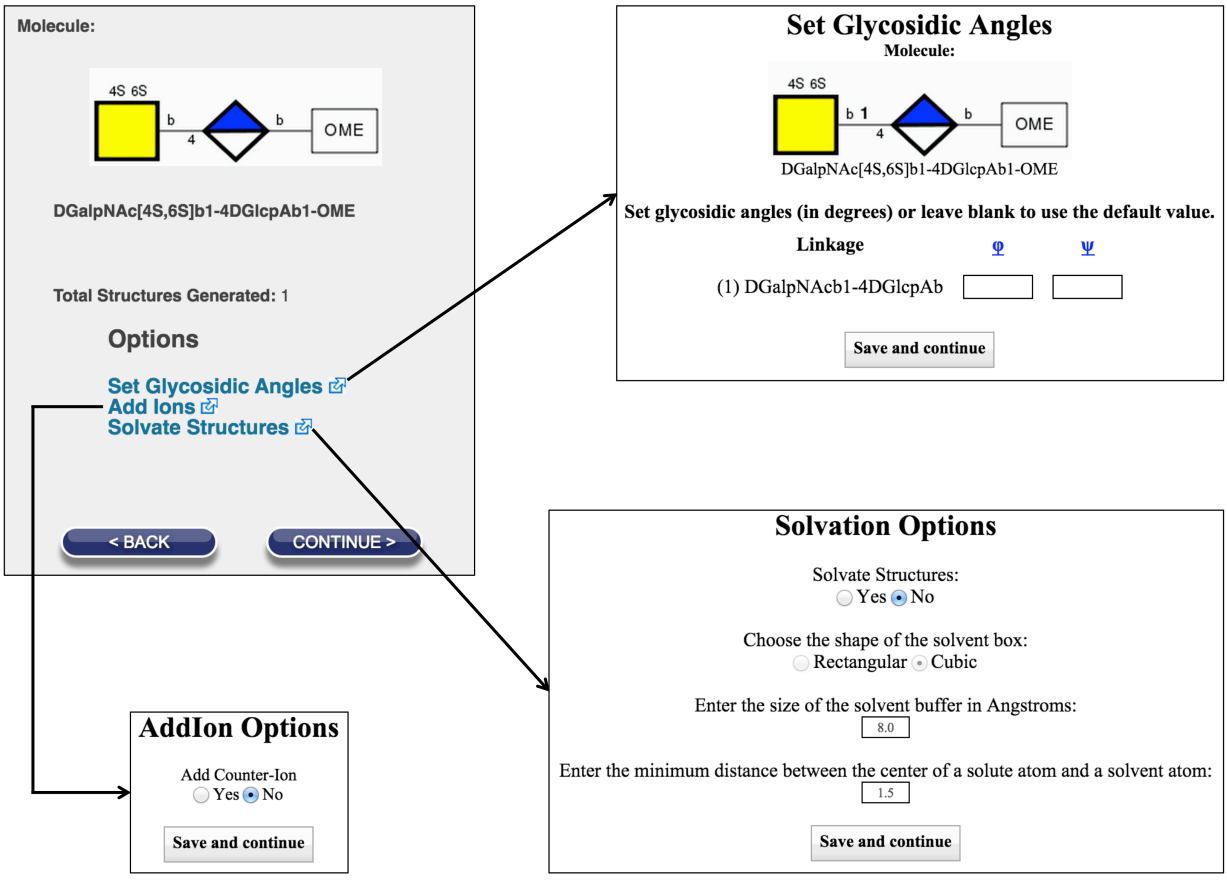

Figure 5.3. Current features available on the Options-page for modeling GAG structures.

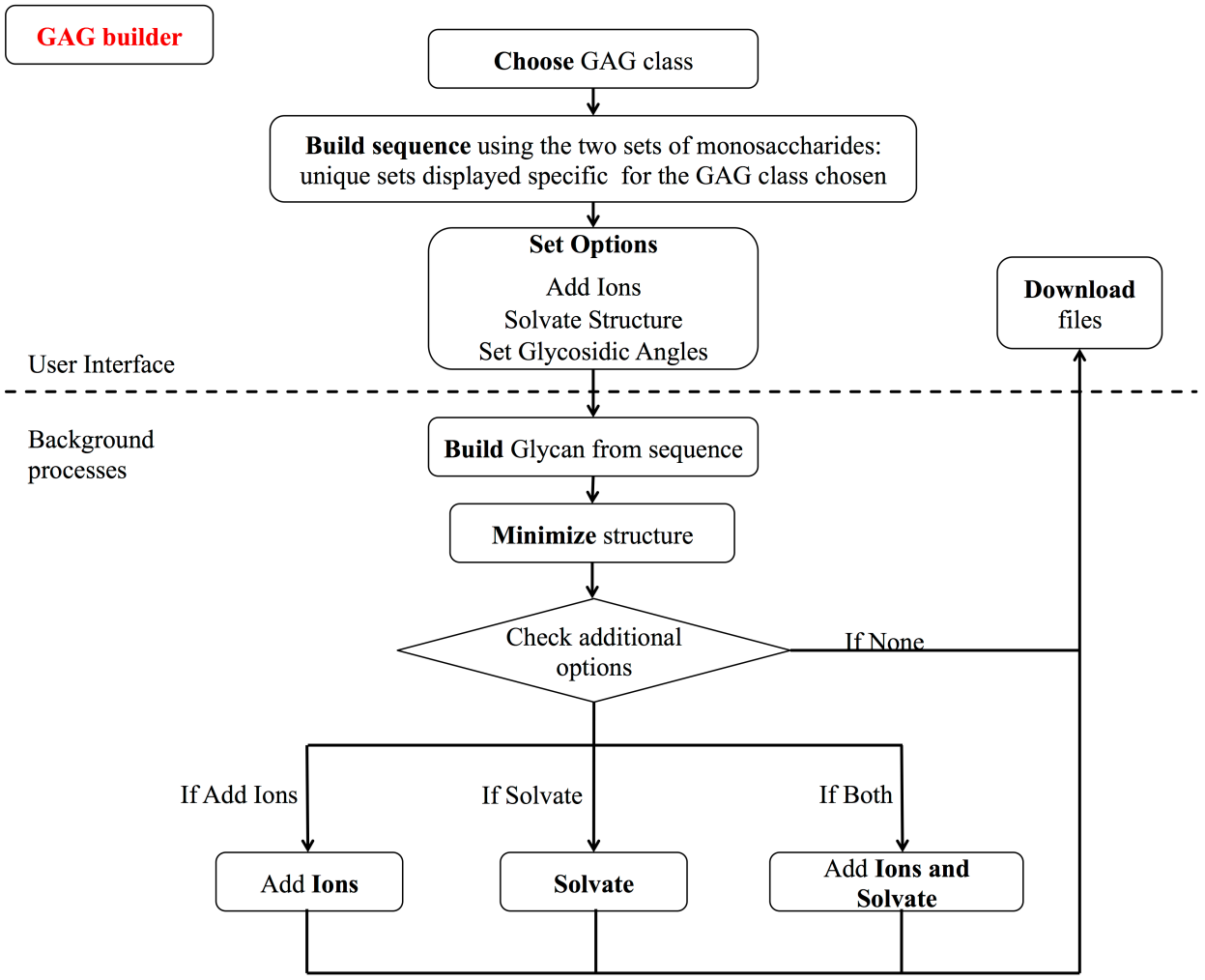

Figure 5.4 Workflow of the GAG Builder showing the processes that occur at the user-interface and in the background for building the structures.

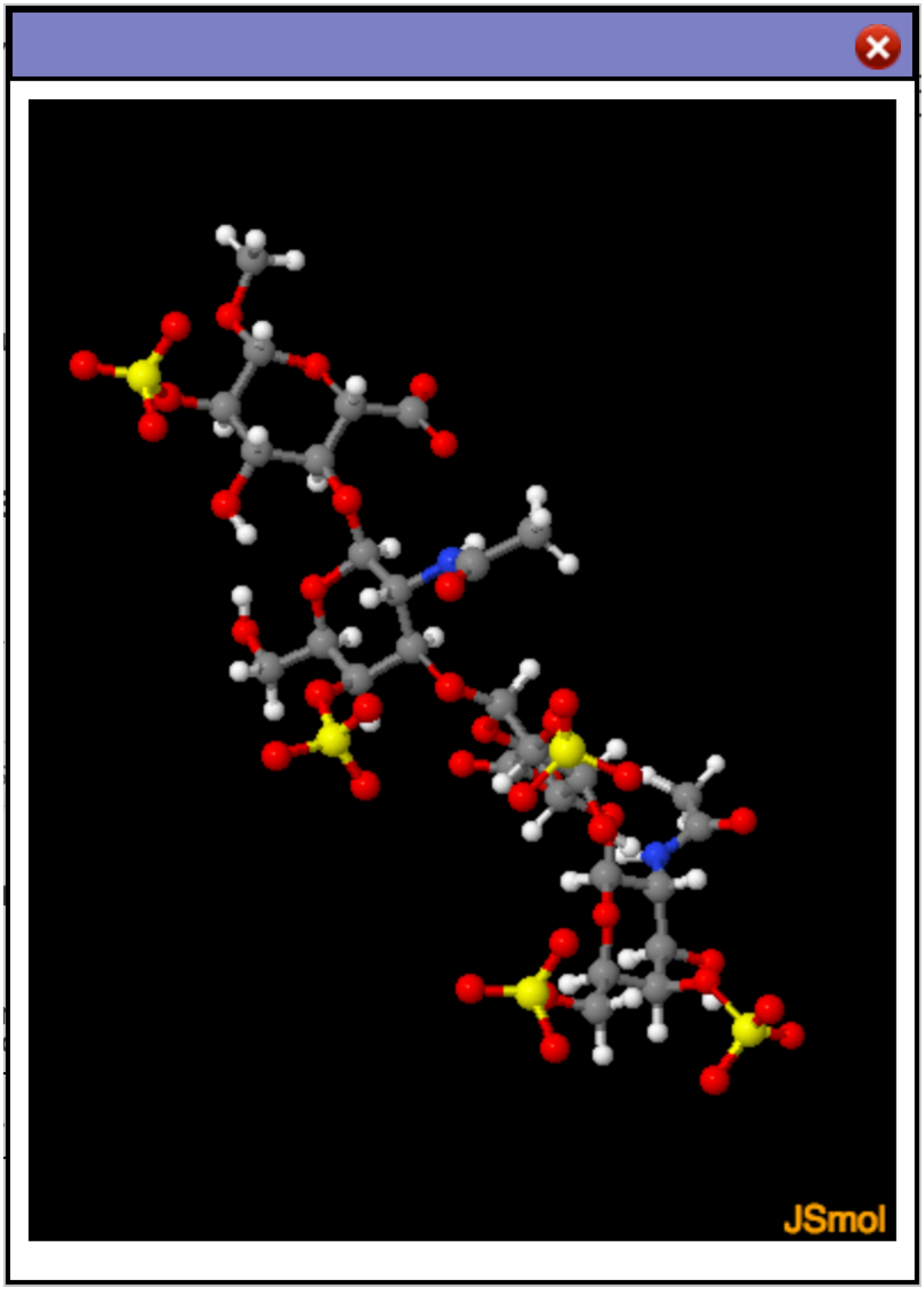

Figure 5.5 JMol applet showing the built GAG structure.

The downloads page can display the built structure using a JMol applet (**Figure 5.5**) and currently provides un-minimized and minimized PDB files as well as topology and restart files compatible with AMBER molecular dynamics package (141).

Implementation

The user interface has been designed with HTML and JavaScript. The pages are deployed using Apache Tomcat and Java. The back end utilizes C++ to read the query sequences and build the structure using a library of monosaccharide prep files.

Future work

The GAG structure builder will be extended to provide protonated uronic acids as well as linker peptides for covalently attaching these GAG structures to proteins.

The output files will also include input files for AutoDock suite of docking software.

CHAPTER 6

THE INTERACTION OF HEPARIN TETRASACCHARIDES WITH CHEMOKINE CCL5 IS MODULATED BY SULFATION PATTERN AND pH[2]

[2] Singh, A., Kett, W. C., Severin, I. C., Agyekum, I., Duan, J., Amster, I. J., Proudfoot, A. E. I., Coom be, D. R., and Woods, R. J. (2015) *J. Biol. Chem.* **290**, 15421–36. Reprinted here with permission of the publisher

Abstract

Interactions between chemokines such as CCL5 and glycosaminoglycans (GAGs) are essential for creating haptotactic gradients to guide the migration of leukocytes into inflammatory sites; GAGs that interact with CCL5 with the highest affinity being heparan sulfates/heparin. The interaction between CCL5 and its receptor on monocytes, CCR1, is mediated through residues R17 and R47 in CCL5, which overlap with the GAG binding ^{44}RKNR^{47}, "BBXB" motifs. Here we report that heparin and tetrasaccharide fragments of heparin are able to inhibit CCL5-CCR1 binding, with IC_{50} values showing strong dependence on the pattern and extent of sulfation. Modeling of the CCL5-tetrasaccharide complexes suggested that interactions between specific sulfate and carboxylate groups of heparin and residues R17 and R47 of the protein are essential for strong inhibition; tetrasaccharides lacking the specific sulfation pattern were found to preferentially bind CCL5 in positions less favorable for inhibition of the interaction with CCR1.

Simulations of a 12-mer heparin fragment bound to CCL5 indicated that the oligosaccharide preferred to interact simultaneously with both ^{44}RKNR^{47} motifs in the CCL5 homodimer, and engaged residues R47 and R17 from both chains. Direct engagement of these residues by the longer heparin oligosaccharide, provides a rationalization for its effectiveness as an inhibitor of CCL5-CCR1 interaction. In this mode, histidine (H23) may contribute to CCL5-GAG interactions when the pH drops just below neutral, as occurs during inflammation.

Additionally, an examination of the contribution of pH to modulating CCL5-heparin interactions suggested a need for careful interpretation of experimental results when they are performed under non-physiological conditions.

Introduction

Chemokines are small proteins (8-10 kDa) that guide the migration of leukocytes to the site of infection during an inflammatory immune response (142). Some chemokines are also involved in the migration of cells into tissues during tissue repair and development. Pro-inflammatory chemokines are immobilized on cell surfaces and extracellular matrices by interactions with glycosaminoglycan (GAG) chains of proteoglycans (143, 144). It is believed that this interaction with GAGs, primarily heparan sulfate (HS), allows the formation of haptotactic chemokine gradients that direct leukocytes into sites of inflammation. Recently endogenous gradients of the chemokine CCL21 in mouse skin were visualized and heparinase treatment disrupted both the CCL21 gradient and dendritic cell migration (145). This and a number of other studies conducted over the last decade have provided convincing evidence that the interaction of chemokines with GAGs is a critical component of directed leukocyte migration. (146). As a result, disruption of chemokine-GAG binding events have been targeted for the development of novel anti-inflammatory agents that are GAG analogs (147, 148).

Chemokines are classified according to the spacing between the first two cysteine (C) residues in their primary sequence into four categories: C, CC, CXC and CX_3C, where X indicates any other residue. These cysteine residues form characteristic intramolecular disulfide bonds that stabilize the tertiary structure, which typically consists of a disordered N-terminus (N-loop), followed by a 3_{10} helix, a three-stranded β-sheet, and a helix at the C-terminus (149). CCL5 (also known as RANTES, for Regulated on Activation Normal T cell ExpreSsed) was chosen for this study because of its well characterized pro-inflammatory activity and because an interaction with GAGs has been reported to be essential for this pro-inflammatory activity (150). CCL5 binds different types of GAGs to varying degrees, but heparin and HS bind with highest

affinity (151). Upon secretion from endothelial cells and activated leukocytes, CCL5 localizes on GAGs at the site of inflammation and triggers the migration of T-cells, monocytes, basophils, eosinophils, natural killer cells, and dendritic cells (152) via engagement with one or more of its receptors: CCR1, CCR3 and CCR5, which are expressed on leukocyte cell surfaces (153).

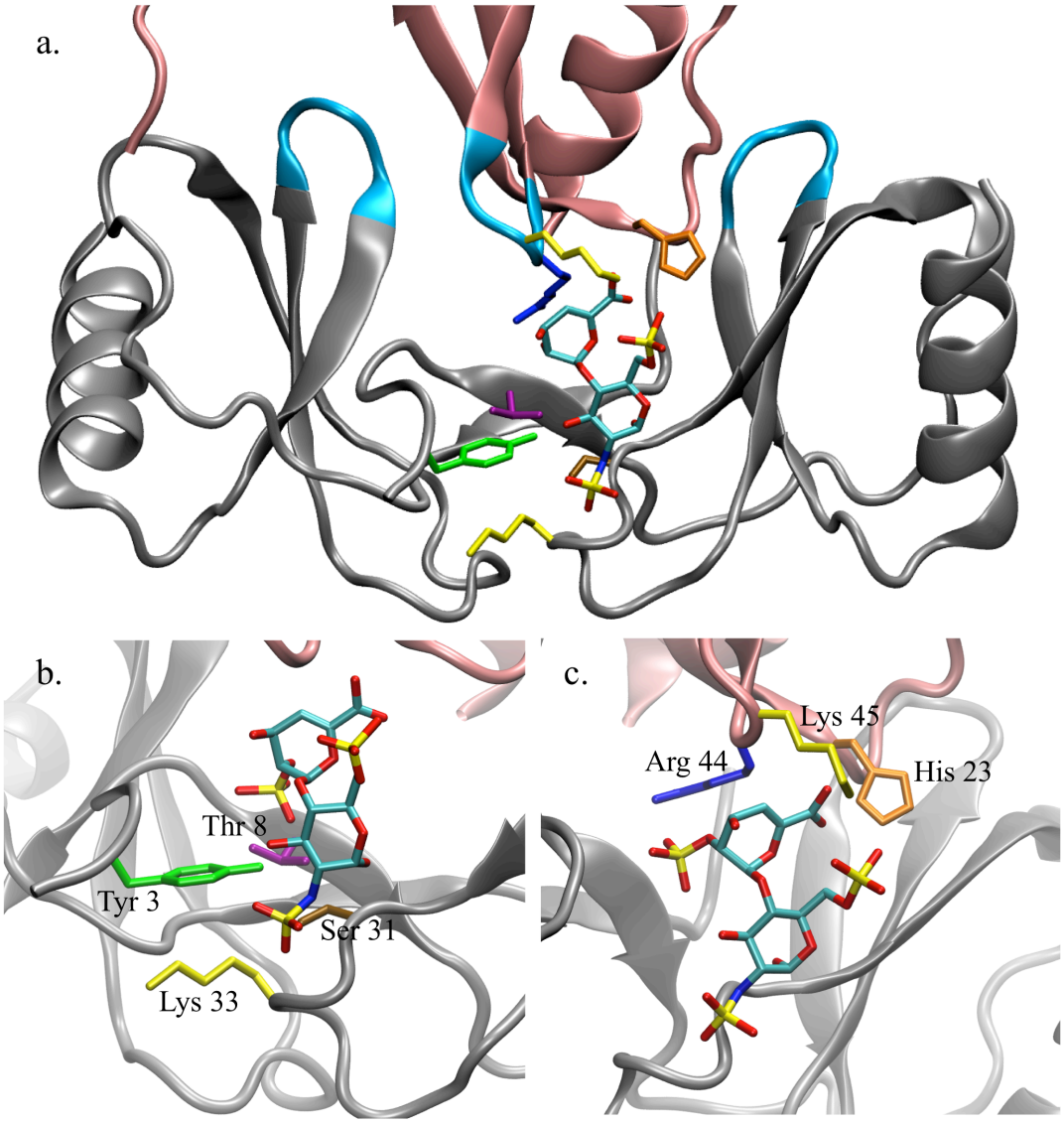

Figure 6.1 Crystal structure of two dimers of CCL5 (PDB ID: 1U4L) showing heparin-related disaccharide (stick representation) interacting with the groove of one CCL5 dimer (grey) and with the lobes from a second dimer (pink), shown as cartoon representation; the ^{44}RKNR^{47} motif is colored light blue (panel "a"). Close-up views of the ligand-protein contacts in the groove-binding and lobe-binding modes shown in panels "b" and "c", respectively. Key side chains are labeled.

The interaction between CCL5 and GAGs has been studied primarily using heparin as a model *in vitro* for the HS structures, which bind CCL5 *in vivo*. CCL5-heparin interactions have been probed by site directed mutagenesis studies, and shown to be mediated primarily through a highly basic ^{44}RKNR^{47} motif on the surface of CCL5 located in a loop termed the "40s loop" (154) (**Figure 6.1**). This is consistent with the fact that protein–GAG interactions generally involve interactions between the anionic sulfate or carboxylate moieties in GAGs and clusters of basic amino acids on the protein. More recent studies indicate that GAG-protein interactions are not only dependent on these linear motifs, but also involve residues that can engage in hydrogen bonding and hydrophobic interactions on the three-dimensional (3D) surface of the protein (50, 51). A further complexity in characterizing GAG-CCL5 interactions is the tendency of these complexes to aggregate at physiological pH or at higher GAG concentrations (155). CCL5 naturally forms high MW oligomers, in which the interaction between E66 and R47 of adjacent monomers plays a pivotal role (156). This role is further substantiated by the observation that the ^{44}AANA^{47}-CCL5 variant does not oligomerize, and furthermore disrupts oligomeric CCL5 to form inactive heterodimers with the WT chemokine (157).

The 3D structure of CCL5 has been determined by crystallography and NMR spectroscopy, in each case at low pH, so as to prevent CCL5 oligomerization (156, 158). The pH that has been employed for the structural studies (pH 3.5-4.8) (156, 158) would be expected to alter the protonation states of histidine residues (pK_A ~6.5), and potentially also of the carboxylates in aspartate (pK_A ~3.6), glutamate (pK_A ~4.2), and IdoA2S (pK_A ~3.1- 3.5) or GlcA (pK_A ~2.8 – 3.2) (159–161) residues. In our crystallographic study (147) (at pH 4.5), employing small molecules identified as CCL5 ligands by library screening performed at pH 3.2, some of the ligands bound to a region outside of the ^{44}RKNR^{47} motif, whereas others bound to a pocket

close to H23. The earlier X-ray crystallography study of CCL5 complexed with heparin disaccharide analogs also detected extensive interactions in a region coined the "30s loop" which is outside of the BBXB motif, although interactions with the BBXB motif of the second monomer in the dimer were also detected (17). This structure is shown in **Figure 6.1**. In contrast, point mutagenesis studies, which identified the importance of the ^{44}RKNR^{47} motif, were carried out under physiological conditions, at approximately neutral pH and did not detect evidence of major contributions from outside of this motif (154). Given the highly ionic nature of heparin, it might be expected that pH could alter the preferred interaction sites between CCL5 and heparin. Accordingly, in the present work we explored the impact of pH on the binding mode.

Heparan sulfate and heparin are highly heterogeneous linear polysaccharides composed of repeating 1,4-linked disaccharide units of β-D-glucuronate (GlcA) or α-L-iduronate (IdoA) and D-glucosamine (GlcN) (162, 163). The uronic acids may be sulfated at the 2-*O* position (IdoA2S or GlcA2S), while the glucosamine may rarely occur as a free amine or more often as an *N*-sulfated (GlcNS) or an *N*-acetylated (GlcNAc) moiety. The GlcNS may also be *O*-sulfated at the 3 or 6 positions, giving rise to GlcNS6S, GlcNS3S, or GlcNS3S6S. In contrast, the GlcNAc residues may be unsulfated or *O*-sulfated only at C6 (GlcNAc6S) (19). Heparin is produced by mast cells and commercial heparin is obtained from tissues with abundant mast cells, such as porcine intestinal mucosa. Structurally, heparin and HS differ in sulfation patterns and the proportion of the various disaccharide units, with there being more IdoA and GlcNS in heparin, as well as more overall sulfation. Common features in heparin are repeating stretches of the tri-sulfated disaccharide structure, IdoA2S-GlcNS6S. In contrast, HS has a more-organized structure with regions of low or no sulfation separating highly sulfated regions that resemble mast cell heparin. Heparan sulfate is produced by almost all cell types and its functions *in vivo*

are primarily to bind and present a range of different growth factors and chemokines to their cell surface receptors (19, 50).

Both CCL5 and CCR1, a major receptor on circulating monocytes, have been proposed as therapeutic targets for cancer related inflammation (164, 165) as well as for infectious diseases (166). Residues R47 and R17 of CCL5 have been shown to play a crucial role in the CCL5-CCR1 binding event (167, 168), and the *N*-terminus of CCL5 is known to be crucial for CCR1 signaling (168, 169). Although it has previously been shown that pools of heparin-derived oligosaccharides can inhibit the binding of CCL5 to its receptor, CCR1 (167), no specific sulfation pattern or motif is known to be optimal for binding to CCL5. GAG heterogeneity makes such evaluations particularly challenging, and presents a role for computational methods to provide theoretical insights. A prerequisite for development of molecules that modify this interaction is the characterization of the dependence of the CCL5-CCR1 binding on the structures and sulfation properties of GAGs and GAG fragments.

Towards the goal of developing small-molecule inhibitors of chemokines, we previously determined that tetrasaccharide fragments from heparinase-digested heparin were able to inhibit both, receptor binding and *in vivo* peritoneal recruitment in an inflammation model (156). In the present study, heparin tetrasaccharides were purified to homogeneity, assayed for their ability to inhibit CCL5-CCR1 binding *in vitro,* and fully structurally characterized by mass spectrometry. The interactions of these tetrasaccharides with CCL5 were then modeled and compared to that of a 12-mer model of intact heparin.

As there are no crystal structures of CCL5 with any of the tetrasaccharides isolated in this work, computational docking was employed to generate initial structures for the 3D complexes. Docking GAGs to proteins is challenging due to the internal flexibility of the GAGs, the high

charge density of the GAGs, and the fact that proteins that bind GAGs often lack well-defined binding pockets (70, 71). The development and testing of protocols that address the issues associated with GAG docking is an area of active research (170–174). Keeping these caveats in mind, the initial CCL5–heparin tetrasaccharide complexes were generated by docking using AutoDock Vina (76). Vina has recently gained popularity in docking carbohydrate ligands (175–177).

To aid in overcoming inaccuracies in the initial structures, as may arise from approximations associated with docking, the complexes were subjected to long (100 ns) fully-solvated molecular dynamics (MD) simulations to refine the docked pose and to assess the characteristics of binding. MD simulations could also be performed under conditions that modeled the effects of low or neutral pH. Docking followed by MD simulation and subsequent binding energy evaluation is an approach that is frequently used to predict the properties of GAG-protein complexes (90, 178, 179).

The results were analyzed in terms of structural stability, preferred binding mode, and computed interaction energies. The theoretical results clearly indicated a dependence of binding site preference and interaction energy on both pH, and the positions of the sulfates within the tetrasaccharides. The heparin fragments were assessed for their ability to inhibit CCL5-CCR1 binding in terms of their ability to engage certain key residues on CCL5, implicated in its interaction with the receptor. The results highlight the important contribution of computational modeling for interpreting biological data and in predicting GAG binding preferences. In addition, these studies offer a cautionary note with regard to the treatment of pH in experimental studies of GAG-protein interactions.

Methods

Preparation of Heparin Oligosaccharides

Heparin was depolymerized according to the procedure described by Chai et al. (180). Briefly, heparin (5 g) and albumin (4 mg) were dissolved in 50 mL of 30 mM CH_3CO_2Na containing 3 mM $CaCl_2$ and adjusted to pH 7 with 0.2 M $NaHCO_3$. Heparinase I (2 IU) or heparinase III (2 IU) (both from Grampain Enzymes, Aberdeen, UK) was added, and the mixture was incubated at 30 °C for 16 h. The mixture was boiled for 3 min, centrifuged and then filtered (0.45 μm). Size-exclusion chromatography was performed on two 90 × 2.5 cm glass columns connected in series. The first column was packed with Bio-Gel P6 fine, and the second column was packed with Bio-Gel P10 fine (both from Bio-Rad Laboratories Inc., Hercules, CA). The columns were eluted with 0.25 M NaCl at a flow rate of 0.5 mL/min using a Gilson HPLC (Middleton, WI), and the effluent was monitored with a refractive index detector. Data were acquired using Gilson Unipoint software. Fractions (1 mL) adjacent to the peak maxima were pooled, lyophilized, and after reconstituting in a minimum of water, desalted on a fast desalting column (10 × 100 mm, GE Healthcare) to give pools of oligosaccharides of a uniform degree of polymerization; see our earlier publication for an example of the separation achieved (181). The desalted fragments were lyophilized, redissolved in water and stored at –20°C. The concentration of each fragment was determined spectrophotometrically at 232 nm in 30 mM HCl using the extinction coefficient of 5500 mol^{-1} cm^{-1}.

Anion-exchange Purification of Tetrasaccharides

Anion exchange chromatography on a C_{18} stationary phase coated with cetyltrimethyl-ammonium ions was performed by adapting the guidelines presented by Mourier and Viskov

(182). A preparative 250 x 21.2 mm, 5 μm Prep C_{18} column (Phenomenex, Torrance, CA) was coated with cetyltrimethylammonium using 1 mM cetyltrimethylammonium bromide dissolved in 32% methanol at a flow rate of 2.5 ml/min at room temperature overnight. For analytical purposes, a 250 x 4.6 mm, 5 μm Luna C_{18} column was prepared in a similar manner, although at a lower flow rate of 1 ml/min for 4 h.

The preparative column was fitted to a preparative Gilson HPLC consisting of 2 model 306 pumps fitted with 25SC pump heads, a model 306 injection pump connected via a t-piece prior to the column, a model 151 UV detector fitted with a short path length cell and a 215 fraction collector, all under the control of Unipoint software. Preparative purifications were achieved in two steps, the first at pH 3 (10 mM phosphoric acid) and the second at pH 7 (10 mM NaH_2PO_4). In each case, elution was effected with a salt gradient formed by addition of 3 M NaCl at the appropriate pH. The column was maintained at room temperature and a flow rate of 20 ml/min was used. The detection wavelength was 235 nm.

Pooled fractions from the preparative column were diluted to a salt concentration of 0.3 to 0.35 M and applied to an anion exchange cartridge (5 ml EconoQ, Bio-Rad Laboratories Inc., Hercules, CA) by gentle vacuum to achieve a flow rate of 2 to 3 ml/min. The cartridge was washed with water (10 ml) and fitted to the above HPLC system. The cartridge was subjected to short step gradients, washing first with 0.6M NaCl, pH 7 for 4 min then with 2.5 M NaCl, pH 7 for 4 min. The flow rate was 2.5 ml/min and the effluent was monitored at 235 nm. Samples destined for re-purification were diluted to yield 0.35 M NaCl and re-purified on the preparative HPLC column described above, and final preparations were desalted as described.

Analytical HPLC

The C_{18} analytical column coated with cetyltrimethylammonium ions was fitted to a Gilson HPLC comprising 2 model 306 pumps fitted with 10SC pump heads, a model 819 injector, a model 119 UV detector and a model 215 liquid handler. The column was maintained at 40°C and a flow rate of 1 ml/min was used with detection at 235 nm. Elution gradients were formed from buffer A (10 mM NaH_2PO_4, pH 7) and buffer B (10 mM NaH_2PO_4 containing 3 M NaCl, pH 7).

Mass Spectrometry Analysis

MALDI: The matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS, abbreviated as MALDI-MS) technique used involved complexing peptides formed from arginine-glycine repeats with heparin tetrasaccharides. Detailed experimental protocols have been presented previously (181). Briefly, the basic peptide $(RG)_{19}R$ was prepared as the trifluoroacetate salt by Auspep (Melbourne, Australia). AG-1 X2 anion exchange resin (20 mg) in the hydroxide form (Biorad, Sydney, Australia) was added to an ice-cold aliquot (100 µl) of 50 µM peptide. The resulting suspension was pelleted and maintained in an ice-bath. An aliquot of peptide (1 µl) was mixed with 10 mg/ml caffeic acid in 50% v/v acetonitrile (8 µl) and 5 to 100 µM sample (1 µl), 1 µl spotted onto a stainless steel sample plate and allowed to dry. MALDI-MS spectra were acquired in the linear mode by using a PerSeptive Biosystems (Applied Biosystems, Melbourne, Australia) Voyager reflectron time-of-flight instrument fitted with a 337 nm nitrogen laser. Delayed extraction was used to increase resolution (22 kV, grid at 93%, guide wire at 0.15%, pulse delay 150 ns, low mass gate at 2000, 50 shots averaged). Mass calibration was achieved by external calibration with the peptide calibration mixture provided by

the manufacturer. The mass of the oligosaccharide was deduced by subtracting the mass of the $(RG)_{19}R$ peptide observed for that sample.

Electron-Detachment Dissociation (EDD) and Collision-Induced Dissociation (CID): Mass spectrometry analyses on the tetrasaccharide fragments were performed with a 9.4T Bruker Apex Ultra Qh-FTICR instrument (Billerica, MA) fitted with an indirectly heated hollow cathode (HeatWave, Watsonville, CA) for electron generation. The sample solutions were infused at concentrations of 2-250 mM in 50:50 methanol:H_2O at a rate of 120 mL/h. Ions were generated by negative-mode electrospray ionization using a metal capillary (Agilent Technologies, Santa Clara, CA, #G2427A). Precursor ions of interest were isolated in the external quadrupole and activated using both collisional induced collision (CID) in the collision cell and electron detachment dissociation (EDD) (183) in the infinity cell. For each spectrum, 512K points were acquired, padded with 1 zero fill and apodized sinebell window. External calibration produced a 5 ppm mass accuracy, and using confidently assigned glycosidic bond cleavage product ions, internal calibration yielded mass accuracy higher than 1 ppm. All MS/MS products are reported using Domon and Costello (184) nomenclature. Structural assignments were determined using in-house software written in MATLAB. In general, the assignment of sulfate groups and N-acetylation was determined by analysis of all possible structures that matched the sulfated tetrasaccharide composition. Details of the assignments for each of the tetrasaccharides are presented in Supplementary Material.

Equilibrium Competition Receptor Binding Assays

The assays were carried out on membranes from CHO cell transfectants expressing CCR1 and the binding of CCL5 was assessed using a Scintillation Proximity Assay (SPA) with [^{125}I]-CCL5

as tracer according to the previously published method (147). Serial dilutions of the heparin tetrasaccharides, covering the range from 0.25 x 10^{-3} M to 1 x 10^{-11}M, were prepared in binding buffer (50 mM Tris/HCl, pH 7.2, containing 1 mM $CaCl_2$, 5 mM $MgCl_2$ and 0.5% BSA). The tetrasaccharide pool and heparin (H 3400; Sigma) were also tested in the same assay. Wheatgerm SPA beads (Amersham) were solubilized in PBS at 50 mg/ml, diluted in binding buffer to a concentration of 10 mg/ml, and membranes of CHO cell transfectants were solubilized at 80 μg/ml in binding buffer. Equal volumes of membrane and SPA bead solutions were mixed before adding them to the assay. The final membrane concentration in the assay was 20μg/ml and the concentration of [^{125}I]-CCL5 was 0.05 nM. The plates (Corning, 96 well, flat and clear bottom) were incubated at room temperature under agitation for 3.5 h. Radioactivity was counted with a beta counter for 1 minute/well and the data analyzed using Graphpad Prism software. Data are expressed as a percent where 100% is the value obtained for CCL5 binding in the absence of either heparin or the tetrasaccharides.

Automated Docking

Automated docking was performed using the molecular docking and virtual screening program, AutoDock Vina (76). To emulate the effects of pH on ionization states in CCL5 and the GAG fragments, the following models were generated. For low pH, the side chain in the only histidine residue (H23) was fully protonated (net charge = 1), as were the carboxylates in all Glu, Asp, GlcA and IdoA residues (net charge = 0). For neutral pH, H23 was assigned zero charge (protonated only at Hε), while the other ionizable residues were appropriately charged

Initial 3D structures for the GAG sequences were built using the tLeap program of the AMBER 12 (118) molecular dynamics package using GLYCAM06 (97) force field, augmented for sulfate groups and unsaturated $\Delta^{4,5}$-uronate (UA) residues (185) . Protein coordinates were obtained from the crystal structure of the human CCL5 dimer in complex with heparin-derived disaccharide (PDB ID: 1U4L) (156). This model of CCL5 contains truncated N-termini (3-68 variant CCL5) naturally found in serum (186). This protein model was employed for automated docking, after removal of the co-crystallized ligand and all water molecules.

For docking, the entire surface of the protein was used as the search space (blind docking). This was done both to avoid biasing the subsequent analysis, and because, although the ^{44}RKNR^{47} motif is known to modulate GAG-binding, the CCL5 crystal structure had the GAG-disaccharide placed in the groove between the CCL5 monomers (156). All glycosidic linkages and exocyclic torsion angles were allowed flexibility during docking. Docking was performed for models of CCL5 corresponding both to low and neutral pH states.

MD simulations

Topology and coordinate files for the CCL5-heparin complexes were generated with the tLeap program, employing the Protein ff99SB (187) and GLYCAM06 (version h) (97) parameters for the protein and GAGs, respectively. The net charge on each system was neutralized with the addition of Na+ or Cl- counter ions, as required. The systems were solvated with TIP3P water (100) in a cubic box extending to at least 12 Å from any atom of the solute.

All MD simulations were performed with the GPU implementation of pmemd, pmemd.cuda_SPDP (130) in Amber12 (118). Energy minimization of the solvent was performed in an NVT ensemble (1000 steps of steepest descent, 24000 steps of conjugate gradient),

followed by a full system energy minimization (1000 steps of steepest descent, 24000 steps of conjugate gradient). The systems were heated from 5 K to 300 K over 60 ps in an NVT ensemble, with a weak positional restraint (10 kcal/mol-$Å^2$) on the atoms in the solute. A Berendsen-type thermostat (128) with a time coupling constant of 1 ps, was utilized for temperature regulation. Equilibration and production was performed at constant pressure (NPT ensemble; 1 atm), with a pressure relaxation time of 1 ps. After the heating step the restraints were removed from the solute atoms, and the entire system was allowed to equilibrate at 300K for 1 ns. All covalent bonds involving hydrogen atoms were constrained using the SHAKE (129) algorithm, allowing a simulation time step of 2 fs. Scaling factors for 1–4 interactions were set to the recommended values of 1.0 and 1.2 for the GAG (97) and protein (187), respectively, and a non-bonded interaction cutoff of 8.0 Å was employed. Long-range electrostatics were computed with the particle mesh Ewald (PME) method. Data were collected for 100 ns for each CCL5-heparin tetrasaccharide (**1-6**) system at neutral, and low pH conditions. For CCL5-heparin dodecasaccharide (**7**), the simulations were performed at slightly acidic (protonated histidine residues) and neutral pH conditions, and data were collected for 200ns each.

Post processing of the MD simulations was performed using ptraj (188) module of Amber and graphical representations of the results were generated with the R-package (189) and VMD (190).

Binding energy calculations

Energetic post processing of the trajectories to obtain binding free energies and per-residue contribution of binding free energies was done with the single-trajectory MM-GBSA method (191, 192) in a continuum solvent model using MMPBSA.py script (193) from the AMBER 12

package. All water molecules were removed from the complexes, and solvation energies were approximated through the modified generalized Born (GB) model (igb=2) (131, 194). Dielectric constant of 4 was used to model the interior of the solute, since a higher internal dielectric constant for proteins with highly charged binding-interfaces has been shown to be appropriate for accurate free energy calculations (195).

For each of the CCL5-heparin tetrasaccharide simulations, energetic convergence was estimated by extracting 100 frames from each of the 40 equal parts (2.5 ns each) of the trajectory and performing binding free energy calculation. This, along with the root-mean square deviation (RMSD) of the tetrasaccharides through the course of the simulations, was used as an indicator of convergence. Consequently, for each of the CCL5-heparin tetrasaccharide simulation, the first 40 ns from the production run were discarded to allow the heparin tetrasaccharide to form stable interactions with the protein. From the last 60 ns of the production run, 6000 frames were extracted and used for more accurate binding energy and per residue decomposition calculations using MM-GBSA (igb=2) with internal solute dielectric of 4.

For the CCL5-heparin dodecasaccharide complexes, first 50 ns of the 200 ns production run were discarded and 10000 frames were employed for the final calculation of binding free energies and per-residue decomposition.

To get interaction energies for the CCL5-disaccharide complex (PDB ID: 1U4L), CCL5 dimer bound to the disaccharide in the groove mode, as well as CCL5 dimer bound to the disaccharide through the ^{44}RKNR^{47} motif from the crystal unit-cell, were subjected to a short implicit solvent minimization, followed by single frame MM-GBSA energy calculation, with settings similar to those described above.

Contact area calculation

Contact areas between the heparin disaccharide and CCL5 dimer in the groove-binding mode, as well as between the disaccharide and the lobes of the opposing CCL5 dimer were calculated using NACCESS (196) using appropriate radii for sugar atoms.

Electrostatic potential calculation

To calculate the electrostatic potential on the surface of CCL5 at neutral and low pH, the input files containing the appropriate atomic charges and radii were prepared using the PD2PQR server (197, 198), which uses PROPKA (199–202) for assigning the protonation states of the residues. The electrostatic potential was calculated using Adaptive Poisson-Boltzmann Solver (APBS) (203).

Results and Discussion

The heparinase-digested heparin fragments were separated into pools of oligosaccharides each with a uniform degree of polymerization, and the tetrasaccharide pool was subjected to anion-exchange chromatography to purify individual structures. An example of the separation profile achieved for the tetrasaccharide pool at pH 3 is presented in **Figure 6.2**a. Additional separation at pH 7 was employed for further resolution (**Figure 6.2**b).

The tetrasaccharides resolved from the pool were tested for their ability to inhibit the binding of CCL5 to its receptor CCR1, expressed on membranes prepared from CHO cells transfected with CCR1 (**Figure 6.3**). Some tetrasaccharides had markedly better inhibitory activity than the tetrasaccharide pool (DP4) from which they were isolated, although none matched the activity of intact heparin (**Figure 6.3**). Calculation of IC_{50} values from the inhibition

data revealed that three tetrasaccharides (**1**, **2**, **3**) were significantly better at inhibiting the binding of CCR1 than the tetrasaccharide pool (DP4 IC_{50}=183 ± 18 μM) (**Table 6.1**). A fourth tetrasaccharide (**4**) was a relatively weak inhibitor, and the remaining two failed to achieve 50% inhibition at the concentration ranges tested (**Figure 6.3** and **Table 6.1**).

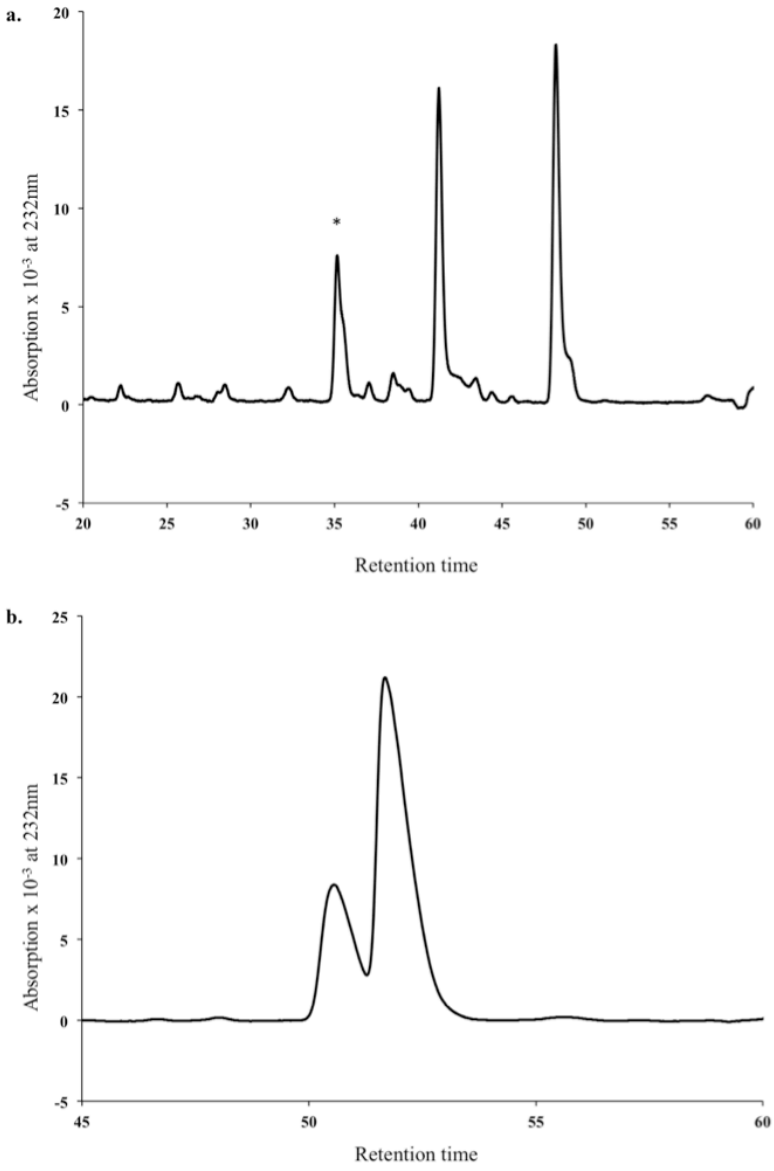

Figure 6.2 Anion exchange chromatography of the tetrasaccharide pool. A C_{18} column coated with cetyltrimethylammonium ions was used. (a) The first preparative separation at pH 3, (b) preparative fractionation of the peak marked with * in A rechromatographed at pH 7. In each instance gradients of increasing NaCl concentration (maximum 3M NaCl) were used.

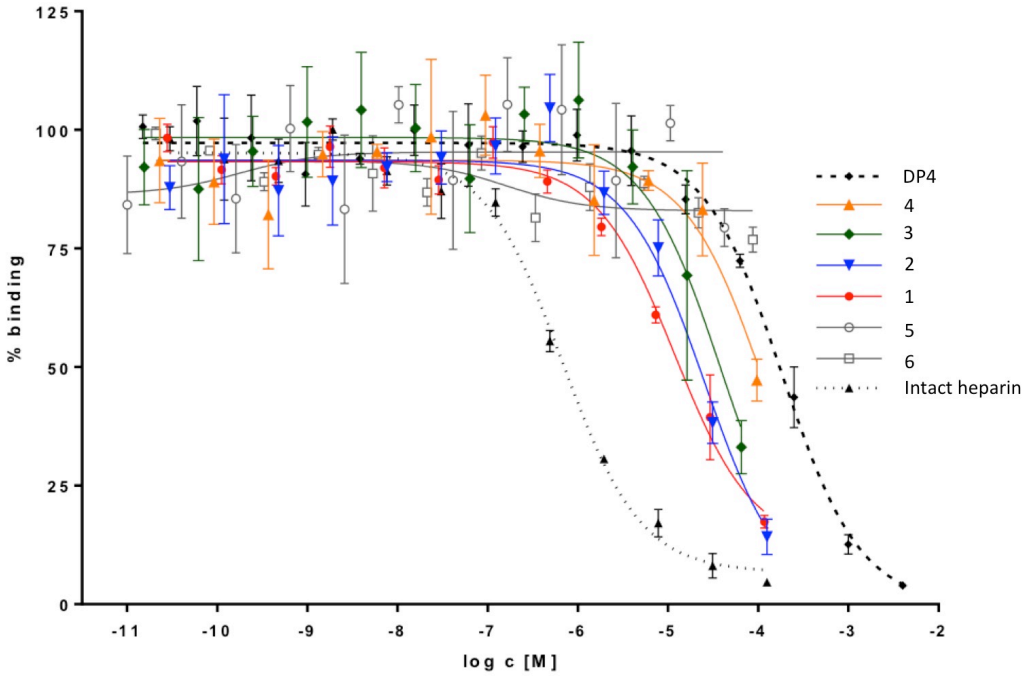

Figure 6.3 Inhibition of CCL5 binding to its receptor CCR1 by heparin tetrasaccharides. Inhibition was determined by competition equilibrium binding assays using membranes from transfected cells expressing CCR1. The six tetrasaccharides, tested to inhibit binding, are shown in comparison to the tetrasaccharide pool (DP4) and intact heparin (H-3400). The mean and standard error are shown for three independent replicates at each data point.

Table 6.1 Ability of heparin tetrasaccharides to inhibit the binding of CCL5 to CCR1

ID	Structure	IC_{50} (μM)	Sulfation Pattern
1	ΔUA2OS-(1-4)-GlcNS6OS-α-(1-4)-IdoA2OS-α-(1-4)-GlcNS	11.8 ± 4.8	3OS, 2NS
2	ΔUA2OS-(1-4)-GlcNS6OS-α-(1-4)-IdoA2OS-α-(1-4)-GlcNS6OS	25.4 ± 2.3	4OS, 2NS
3	ΔUA2OS-(1-4)-GlcNS6OS-α-(1-4)-IdoA-α-(1-4)-GlcNS6OS	39.1 ± 18.5	3OS, 2NS
4	ΔUA2OS-(1-4)-GlcNS6OS-α-(1-4)-IdoA-α-(1-4)-GlcNAc6OS	115 ± 39.2	3OS, 1NS
5	ΔUA2OS-(1-4)-GlcNS-α-(1-4)-IdoA-α-(1-4)-GlcNS6OS	---[a]	2OS, 2NS
6	ΔUA2OS-(1-4)-GlcNS-α-(1-4)-IdoA-α-(1-4)-GlcNAc6OS	---[a]	2OS, 1NS

[a]No measurable inhibition.

For comparison, intact heparin inhibited the binding of CCL5 to CCR1 with an IC_{50} of 0.68 ± 0.1 μM. A 3D model for a 12-mer fragment of heparin (ΔUA2OS-(1-4)-[GlcNS6OS-α-(1-4)-IdoA2OS-α-(1-4)-]$_5$GlcNS6OS) was generated as a trimer of the most highly sulfated tetrasaccharide, **2**.

MALDI-TOF-MS analyses of the tetrasaccharides tested in the biological assays indicated that the sulfate content varied among the fragments (**Figure 6.4**), with inhibitory activity requiring at least four sulfate groups (**Figure 6.4**). However, the data also indicated that the sulfation pattern affected activity, and that solely the presence of four sulfate groups (as in **5**) was not sufficient to achieve inhibition.

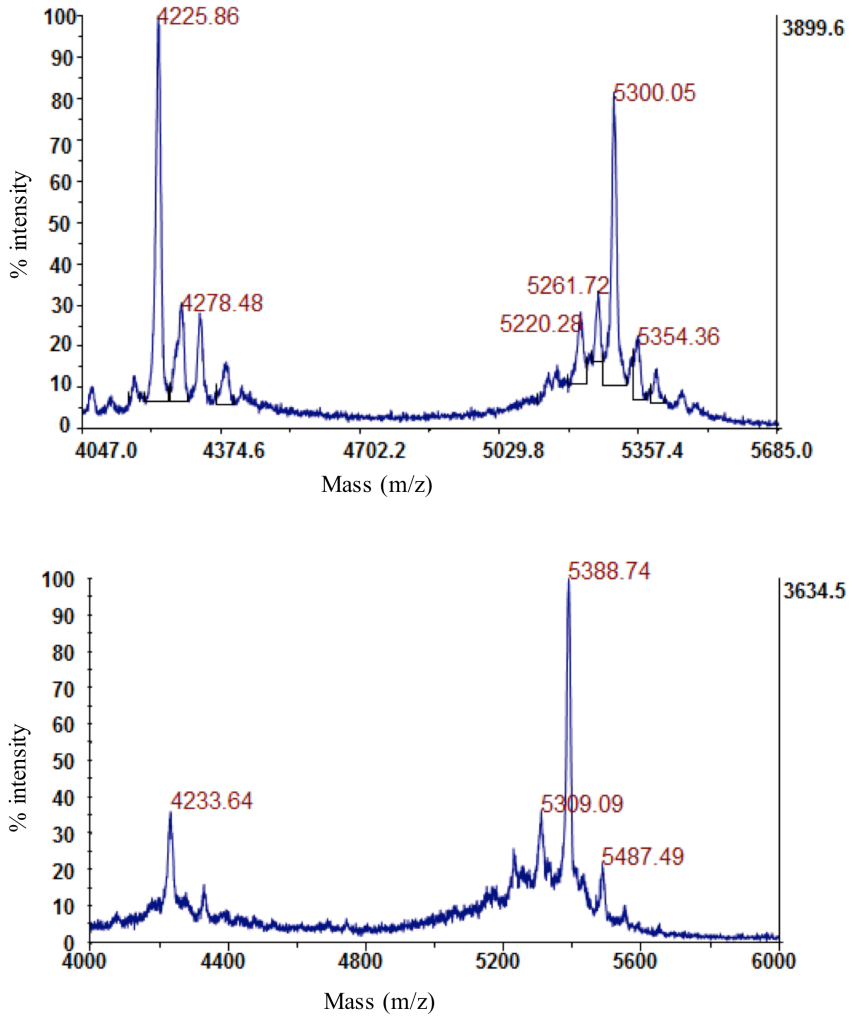

Figure 6.4 MALDI-TOF-MS spectra of two tetrasaccharides, tetrasaccharide **1** (top) and tetrasaccharide **2** (bottom). Masses of the peptide, peptide plus tetrasaccharide and various adducts are shown. A mass indicating loss of a sulfate is also shown for each spectra. Mass of tetrasaccharides = mass of tetrasaccharide peak plus peptide – mass of peptide. Mass of tetrasaccharide **1**: 5300.1- 4225.9 = 1074.2 (theoretical mass for 2N-sulfates, 3O-sulfates = 1074.9); and mass of tetrasaccharide **2**: 5388.74 – 4233.64 = 1155.1 (theoretical mass for 2N-sulfates, 4O-sulfates = 1154.9).

Accurate structural elucidation of highly sulfated GAGs by MS is challenging due to the labile nature of the sulfate half-ester bonds, as they can readily be cleaved during ionization and ion activation stages. The decomposition of sulfate groups is more pronounced in heparin oligomers due to the high sulfate density. Such molecules require careful control of ion activation to retain the sulfate modifications, while providing the necessary fragmentation of glycosidic bonds and pyranose rings. We have shown that, with the aid of Na^+/H^+ exchange coupled with CID (204), it is possible to obtain detailed structural information on even highly complex heparin oligomers like arixtra (205). EDD has also been shown to produce highly informative tandem mass spectra with minimal sulfate loss. Here, assignments of the sulfation positions in the tetrasaccharides (**Table 6.1**) were obtained using CID and EDD MS (see details in Supplementary Data). Cross-ring cleavages and glycosidic bond products were used to assign sites of sulfation.

With the goal of determining how the sulfation pattern of these tetrasaccharides affects their inhibitory activity, 3D structures of the heparin fragments bound to CCL5 were generated using a combination of molecular docking and molecular dynamics simulation. The docking employed crystallographic data for CCL5, obtained at low pH, in complex with a sulfated disaccharide (PDB ID 1U4L). Because the inhibition experiments were carried out at pH 7.2, the effect of pH on the predicted modes of interaction was also examined. In addition to the six tetrasaccharides, a dodecasaccharide with the repeating trisulfated disaccharide pattern of **2**, was modeled to assess the binding of longer heparin fragments.

Molecular docking

Automated docking of **1-6** was performed using AutoDock Vina using the entire surface of the protein as the search space. Docking consistently placed all tetrasaccharides in the groove between the CCL5 monomers (**Figure 6.5**), in agreement with the placement of the disaccharide fragment in the crystal structure. Changing the protonation states of the ionizable side chains to that expected at pH 3.5 did not change the general distributions of the tetrasaccharide poses (**Figure 6.5**).

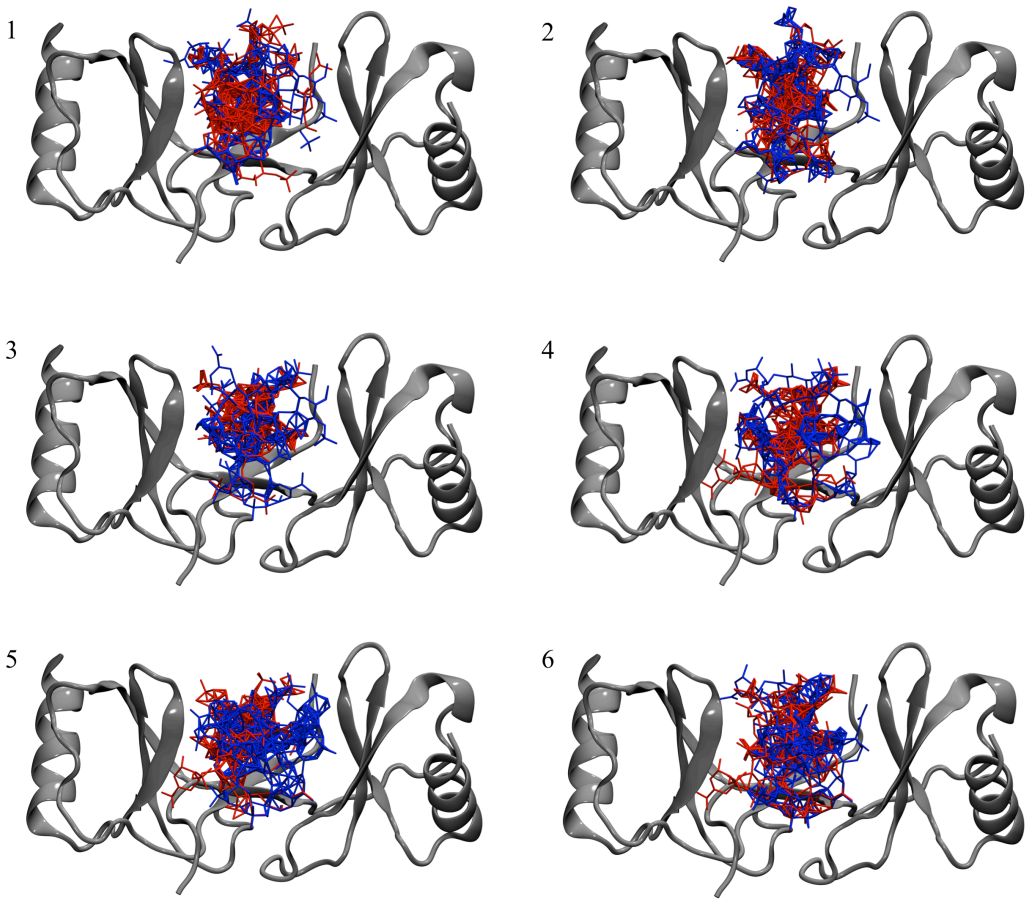

Figure 6.5 Results from automated docking of modeled tetrasaccharides **1-6** to CCL5 dimer (respectively a-f) using AutoDock VINA. Top 5 highest ranked poses are depicted from docking to CCL5 with ionization states corresponding to low pH (red) and neutral (blue).

Table 6.2 Binding energies[a] for the top-ranked pose and average binding affinities for the 10 highest-ranked poses from docking of **1-6** to CCL5 at neutral and low pH conditions.

Neutral pH	**1**	**2**	**3**	**4**	**5**	**6**
Pose 1	-5.8	-6.0	-6.0	-5.9	-5.1	-6.0
Average	-5.4	-5.7	-5.6	-5.4	-4.9	-5.4
Low pH						
Pose 1	-6.3	-6.5	-6.8	-7.3	-6.2	-7.0
Average	-6.1	-6.1	-6.5	-6.6	-5.8	-6.4

[a]kcal/mol

The ranking of the theoretical binding energies from docking was not consistent with the IC_{50} data, although the predicted affinities were significantly enhanced at low pH (**Table 6.2**), as might be expected on the basis of enhanced electrostatic interactions between the anionic ligand and the protein surface (**Figure 6.6**). The predicted binding energies varied by less than 1 kcal/mol for each of **1-6**, whereas the IC_{50} data indicated that tetrasaccharides with the same overall degree of sulfation inhibited the CCL5-CCR1 interaction at neutral pH with varying efficacy (**Table 6.1**). Nor was there a pronounced enhancement of the binding energy for the more highly sulfated ligand **2**. In the groove-based ligand alignment produced by docking, the tetrasaccharides do not interact directly with the known ^{44}RKNR^{47} heparin-binding motif (**Table 6.3**), or with residue R17 reported to be important for the binding of CCL5 and CCR1 (168).

Table 6.3 Average distances[a] between any sulfate group in **1-6** and the RKNR motif as a function of pH

	1	**2**	**3**	**4**	**5**	**6**
Neutral pH	20.4 (12.4, 32.4)	17.1 (9.9, 24.6)	17.0 (5.5, 25.0)	18.8 (10.9, 26.2)	17.5 (10.7, 26.2)	18.1 (10.6, 26.2)
Low pH	16.3 (5.7, 27.6)	21.8 (12.2, 32.5)	19.1 (5.2, 26.8)	17.3 (5.8, 24.5)	17.3 (8.4, 26.1)	17.8 (8.3, 26.1)

[a]In Å, between the sulfate sulfur atom and any of the nitrogen atoms in the side chains of the arginine (R-44a/b or R-47a/b) or lysine (K-45a/b) residues.
The smallest and largest distances are given in brackets.

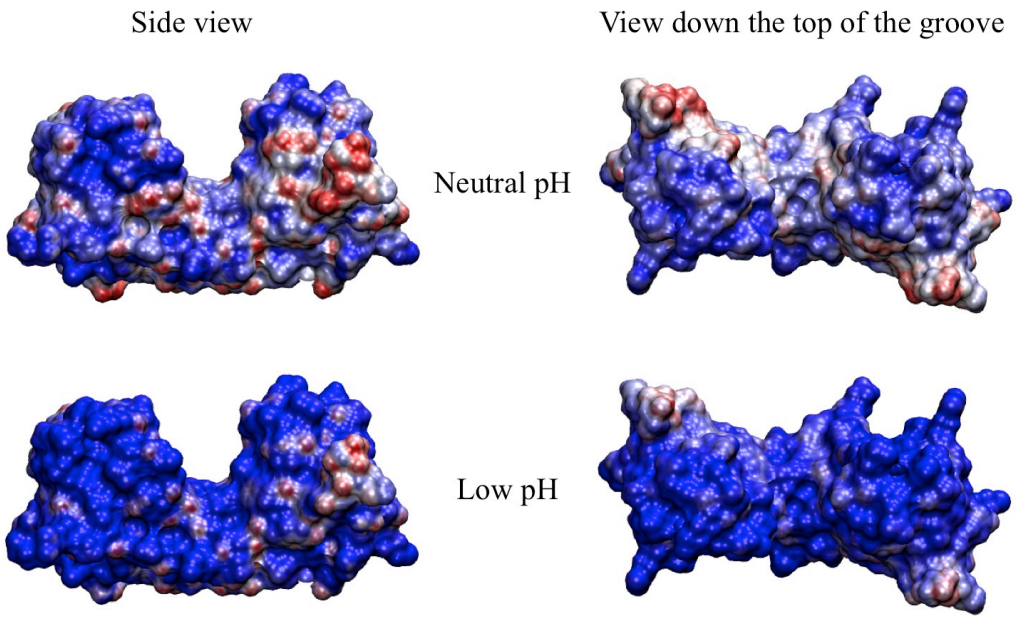

Figure 6.6 Representation of the electrostatic potential surfaces of CCL5 dimer at neutral and low pH. Positive potential is shown in blue and negative potential is in red. At low pH, the aspartic and glutamic acids are neutral and histidine residues are positively charged, expanding the distribution of positive charge on the entire surface of the protein. At neutral pH the region around the ^{44}RKNR^{47} motif has a higher density of positive charge relative to the groove.

As these docking experiments were unable to identify the biologically known binding motif or discriminate between ligands, a representative complex for each CCL5-tetrasaccharide interaction was selected and subjected to MD simulation to more accurately model the effects of molecular dynamics, solvation, and pH.

MD simulation

Within the first 30 ns of simulations performed at neutral pH, tetrasaccharides **1**-**5** diffused out of the groove (**Figure 6.7**) and created new interactions, including with R17 and the ^{44}RKNR^{47} motifs (**Figure 6.8**). Non-inhibiting tetrasaccharide **6**, which has the lowest degree of sulfation, remained bound in the groove throughout the 100ns MD simulation. In contrast, for the

simulations corresponding to low pH, all the heparin tetrasaccharides remained bound in the groove between the CCL5 monomers in the dimer, in similar positions as obtained from docking (**Figure 6.8**).

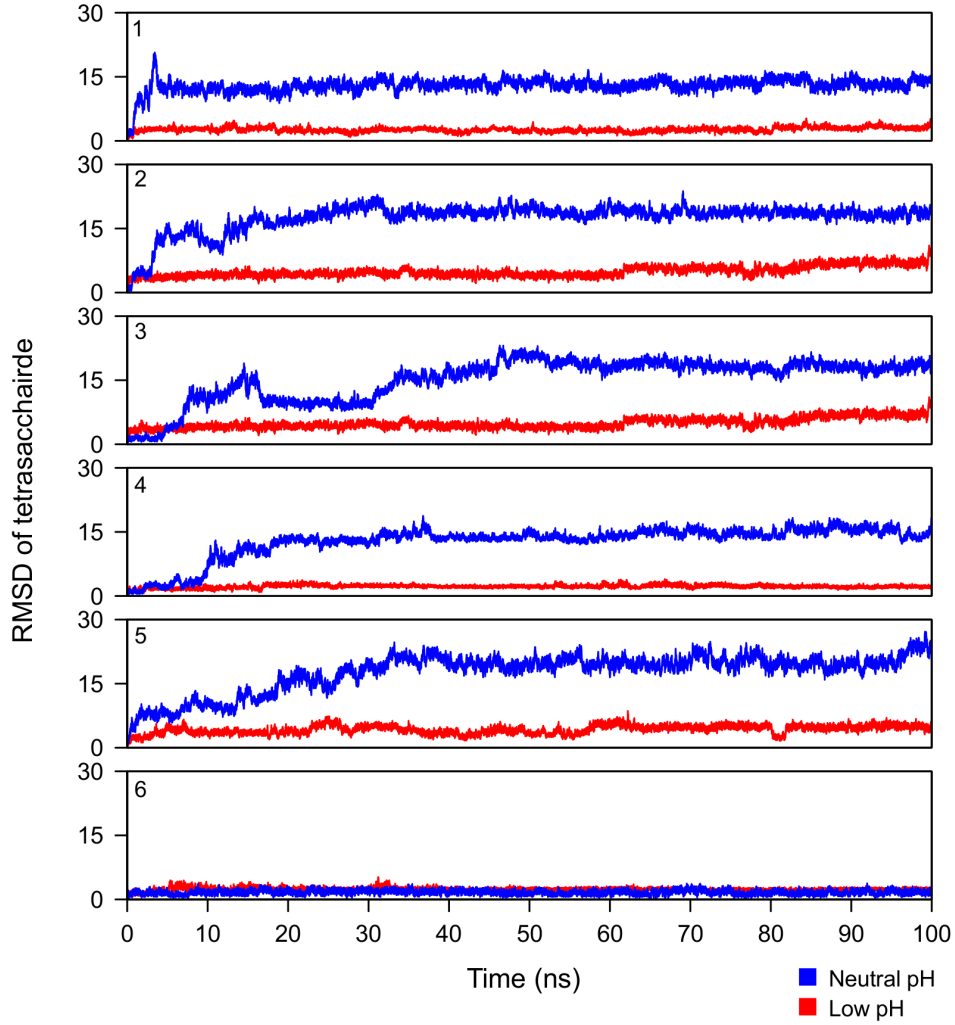

Figure 6.7 RMSD of 1-6 with respect to their starting position over the course of the MD simulations (100 ns) at low (red) and neutral (blue) pH. All tetrasaccharides, but 6, move out of the initial docked site and make interactions with residues outside of the groove of the dimer.

The origin of the difference in the behavior of the complexes at neutral and low pH conditions can be understood in terms of the difference in the distribution of positive potential on the surface of the protein (**Figure 6.6**). At neutral pH, the ^{44}RKNR^{47} motifs have a higher

positive charge density as compared to other regions of CCL5, and consequently the negatively charged heparin fragments are attracted to these regions. The binding of **6** to the groove between the two RANTES monomers at neutral pH may be explained by its low degree of sulfation, and consequently, weaker attraction to the ^{44}RKNR^{47} motifs. Subsequent analyses were performed only on the complexes simulated at neutral pH, unless noted otherwise in the text.

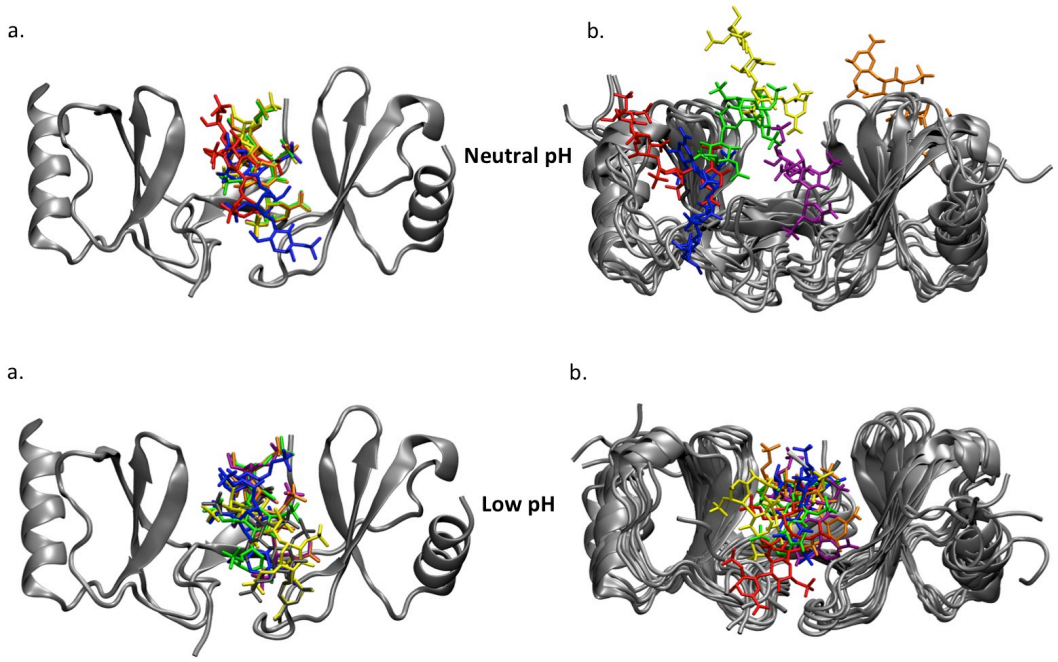

Figure 6.8 Poses of **1-6** at the start (a.) and end (b.) of the MD simulations at neutral (upper) and low (lower) pH (tetrasaccharide colors: **1**–red, **2**-blue, **3**-yellow, **4**-green and **5**-orange, **6**-purple). At low pH the tetrasaccharides show minimal movement within the groove, which at neutral pH they make interactions with residues outside the groove of the CCL5 dimer.

Binding free energy (MM-GBSA) analysis

The tetrasaccharide interaction energies ranged from -71.0 to -36.4 kcal/mol. These values are considerably higher than would be expected experimentally, due in part to the omission of conformational entropic effects in the computational analysis (206). Theoretical entropic contributions were not computed as they were unlikely to have converged over the time scale of

the simulations (195). However, because of the similarity of ligand sizes and compositions, the relative trends in MM-GBSA values are expected to be less sensitive to entropic differences. Per-residue binding energy decomposition identified the protein residues that contributed most significantly to binding of the tetrasaccharides, and revealed differences in the binding modes of each tetrasaccharide (**Table 6.4** and **Figure 6.9**).

Table 6.4 Summary of the per-residue theoretical interaction energies[a] (MM-GBSA) for **1**-**6** at neutral for residues in the ^{44}RKNR^{47} motif and other key[b] residues

GAG	**1**	**2**	**3**	**4**	**5**	**6**
Neutral pH	3OS, 2NS	4OS, 2NS	3OS, 2NS	3OS, 1NS	2OS, 2NS	2OS, 1NS
^{44}RKNR^{47} A	-1.3	-1.3	-8.6	-2.8	-13.3	-1.7
^{44}RKNR^{47} B	-11.6	-14.4	-15.6	-14.1	-1.3	-4.3
N terminus A	-11.4	-4.6	0.0	-7.9	0.0	-5.2
N terminus B	0.0	0.0	-1.7	0.0	-1.2	-7.5
R^{47} A	-0.5	-0.5	-2.1	-1.0	-5.3	-0.7
R^{47} B	-7.2	-7.0	-5.0	-7.2	-0.5	-2.9
R^{17} A	-0.3	-0.3	-0.5	0.0	0.0	0.0
R^{17} B	-8.6	-8.0	-0.6	0.0	0.0	0.0
Chain A	-12.7	-5.9	-10.5	-10.7	-16.8	-16.1
Chain B	-24.3	-25.1	-19.3	-14.1	-2.5	-11.8
Total MM-GBSA[c]	-71.0	-60.1	-52.9	-47.8	-36.4	-61.9

[a]kcal/mol
[b]residues that contribute more than 1 kcal/mol to the interaction energy.
[c]Total from both monomers, conformational entropic effects not included.

As expected from the structural data, the ^{44}RKNR^{47} motifs contributed most strongly to ligand binding. The tetrasaccharides generally interacted preferentially with only one of the motifs in the dimer, with the exceptions of **3**, which appeared to alternate in its interactions between the two domains, and **6**, which remained bound in the inter-domain groove. In addition to the interactions with the ^{44}RKNR^{47} motifs, the tetrasaccharides frequently formed interactions with residues from the *N*-terminus of the opposite domain (shaded cells in **Table 6.4**).

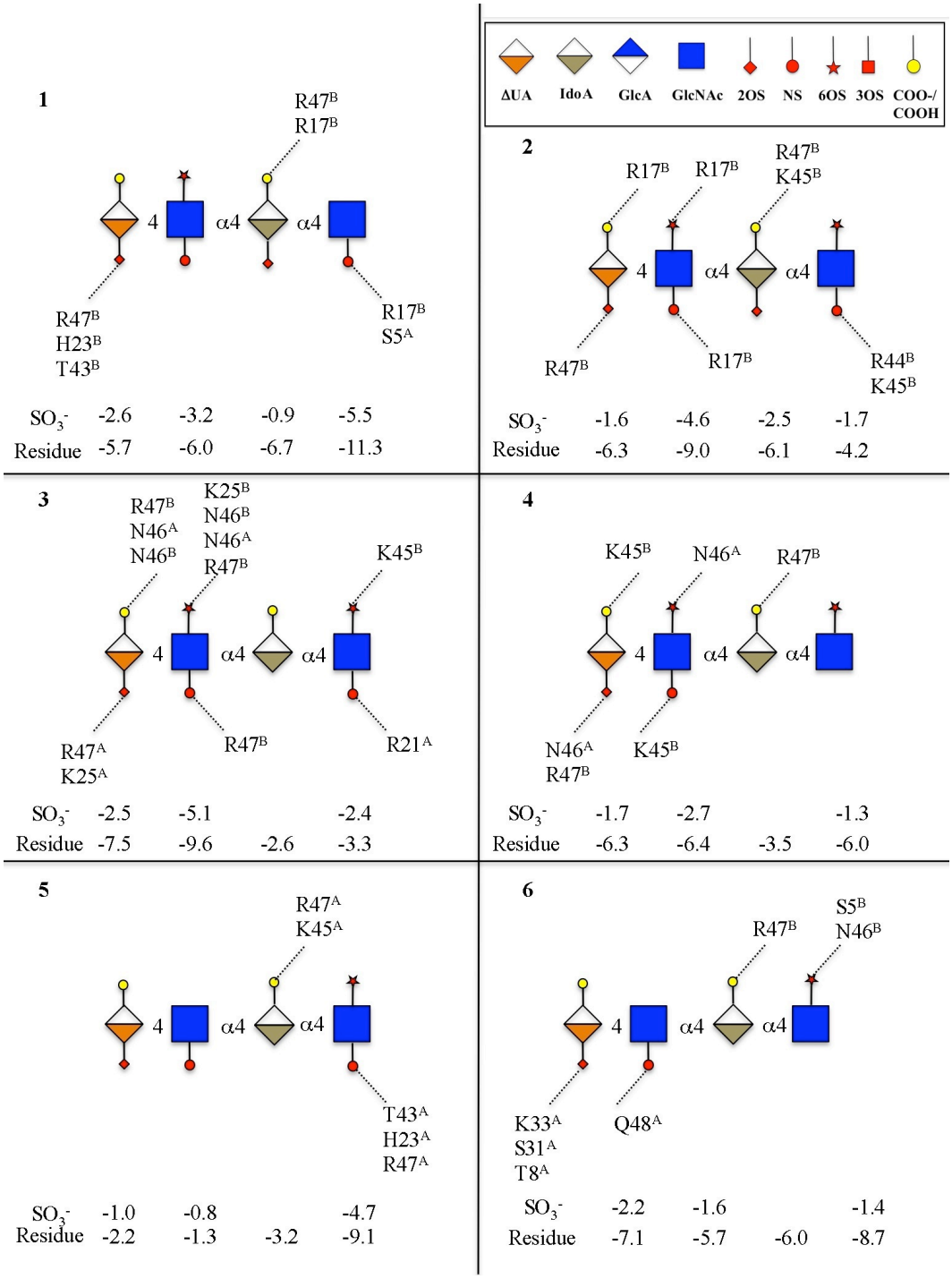

Figure 6.9 Direct interactions between the sulfate and carboxylate groups of **1-6** and key residues from CCL5, at neutral pH. Interactions present for more than 20% of the simulation time, using a 3.5 Å distance cutoff between interacting atoms, during the last 60 ns of each simulation, are presented. Contributions to the interaction energy from the sulfate groups on each residue, as well as the total per-monosaccharide contribution are indicated.

Although contacts with the *N*-termini made only a small contribution to the binding of **2**, **3** and **5**, they played a more significant role for **1, 4** and **6**. Since **6** remained bound in the groove region, the predominant contribution to its binding came from the N-termini.

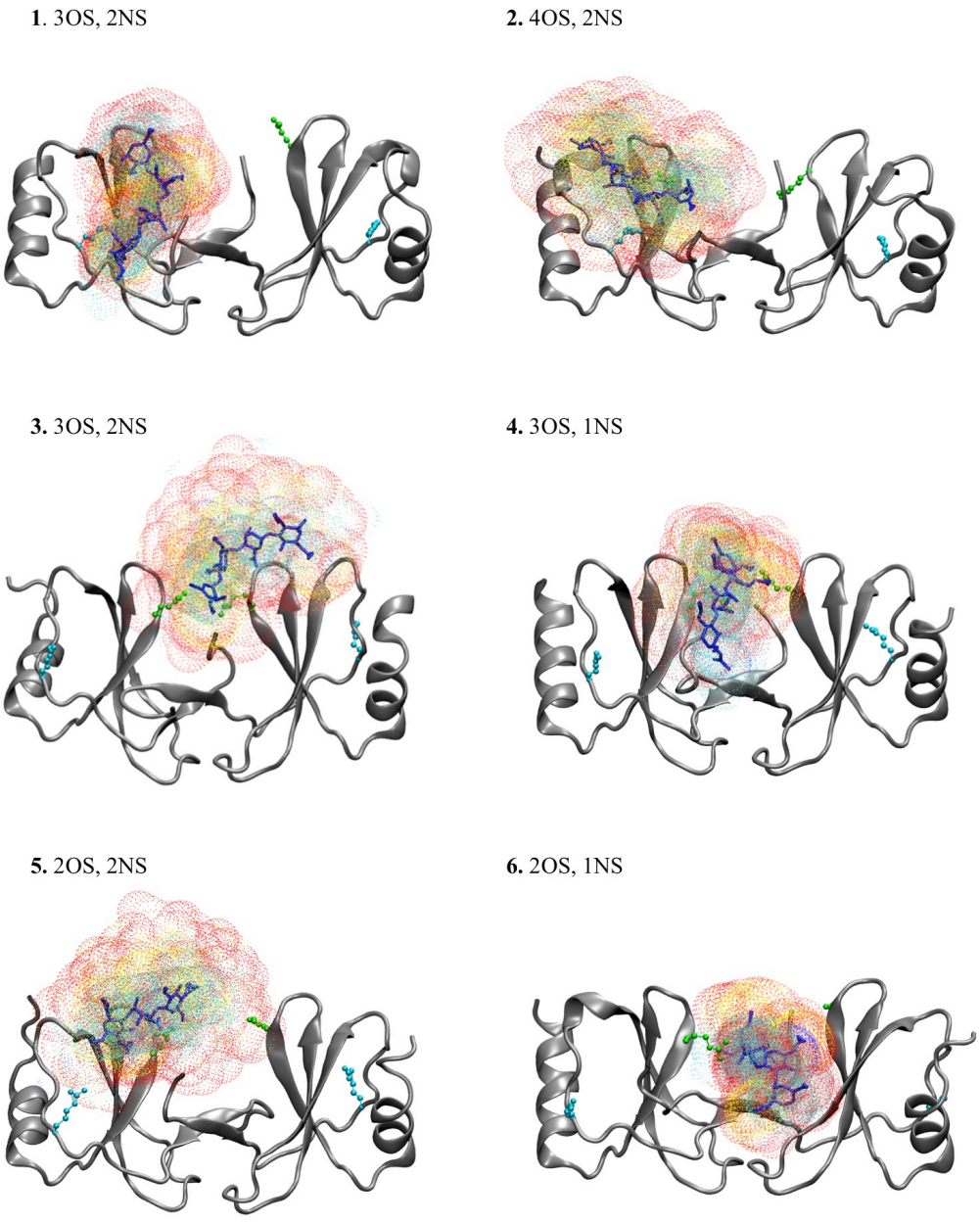

Figure 6.10 Solvent accessible surfaces for 100 poses of the tetrasaccharides extracted at 60 ps intervals from the last 60 ns of the simulations, superimposed to depict their motion and indicate the differences in their contacts with CCL5. Average structure of the tetrasaccharides (blue) shown in stick representation, and average protein structure (silver) shown in ribbon representation. Residue R17 (cyan) and R47 (green) shown in CPK.

Residues R47 and R17 of CCL5 are known to play a crucial role in the CCL5-CCR1 binding event (167, 168), whereas the *N*-terminus of CCL5 is crucial for CCR1 signaling (168, 169). Based on the per-residue MM-GBSA data, each ligand interacts with these residues to varying degrees. Specifically, R47 contributed more than 5 kcal/mol in each case towards binding of tetrasaccharides **1, 2**, **3** and **4** (shaded cells in **Table 6.4**), whereas R17 contributed significantly to binding of the top two inhibitors **1** and **2**.

In terms of the overall interaction energies, the binding of **1**-**5** could be ranked in the same order as their inhibitory capabilities (**Table 6.1**), with **1** being both the tightest binder and most potent inhibitor. However, interaction energy alone was not sufficient to define a strong inhibitor. This is illustrated by the behavior of **6**, which was predicted to bind tightly to CCL5, but because it remained in the inter-domain groove, and did not interact very strongly with either R47 or R17, it could not inhibit CCR1 binding. The binding modes for each tetrasaccharide are presented in **Figure 6.10**, and indicate that different sulfation patterns lead to variations in the preferred CCL5 contact regions, even for tetrasaccharides that contain the same number of sulfate groups.

Broadly, the binding affinity was predicted to increase with increasing number of sulfate groups, with the contribution of specific sulfate groups in defining the interaction being subtle. Specifically, the energetic contributions of individual sulfate groups ranged from -0.8 to -5.5 kcal/mol, and frequently accounted for approximately half of the total energetic contribution of the monosaccharide. Carboxylate groups generally contributed only weakly to binding, relative to the sulfate groups. Frequently, the *O*-sulfate group in the terminal uronic acid, as well as the carboxylate group of the iduronic acid, formed hydrogen bonds with the R47 residues, however there was no clearly discernable pattern between specific hydrogen bonds and affinity.

Docking and MD simulation of a model heparin dodecasaccharide

To examine how a longer heparin fragment would interact with CCL5, and inhibit CCL5-CCR1 interaction, a dodecasaccharide consisting of a repeating sulfation pattern similar to **2** was generated, docked to CCL5, and subjected to MD simulation. This sulfation pattern was chosen because it is the most common structure in heparin. At neutral pH, as in the case of the smaller heparin fragments, docking placed the dodecamer in the groove of the CCL5 dimer. As seen with the tetrasaccharides, during the MD simulation the dodecamer drifted out of the groove to form new interactions with the ^{44}RKNR^{47} motifs. However, the larger heparin fragment interacted with both ^{44}RKNR^{47} motifs simultaneously, spanning the "40s loops". It remained in this orientation for the remainder (approximately 150 ns) of the simulation (**Figure 6.11**). Our findings for the binding mode of a longer heparin sequence match the predictions of Vivès et al. (207) for the binding of heparin oligosaccharide (17 monomers) to CCL5.

Per-residue binding energy decomposition identified that residues R47 and R17 make significant contribution to the binding of the dodecasaccharide to CCL5 (**Table 6.5**). Direct engagement of these residues by a longer heparin oligosaccharide, demonstrates its potential efficacy as an inhibitor of CCL5-CCR1 interactions.

Several GAG binding proteins that contain histidine residues close to the GAG binding sites, have previously demonstrated a heightened GAG binding response under slightly acidic conditions (58, 64, 208, 209). Both CCR1 and CCL5 are up-regulated in many renal inflammatory conditions (210) and blocking CCR1 activity has been shown to reduce inflammation. CCL5 contains one histidine residue in each domain, in the 3_{10}-helix that lies close to the ^{44}RKNR^{47} motif. Because CCL5 binds to GAGs *in vivo* during an inflammatory immune response, which is accompanied by a drop in pH (to around 6 – 6.3) (211–214), we wanted to

examine the binding of heparin to CCL5 under pH conditions slightly below neutral. The CCL5-dodecamer complex was subjected to an MD simulation with protonated histidine residues, to mimic the protonation state at slightly acidic pH conditions. The simulation was initiated from the docked complex and showed that the general binding mode of heparin at slightly acidic inflammatory pH is similar to that of heparin at neutral pH (**Figure 6.11**). The contribution of the histidine residues, as well as the other residues involved in binding, was quantified using MM-GBSA, for the simulations at both pH conditions (**Table 6.5**).

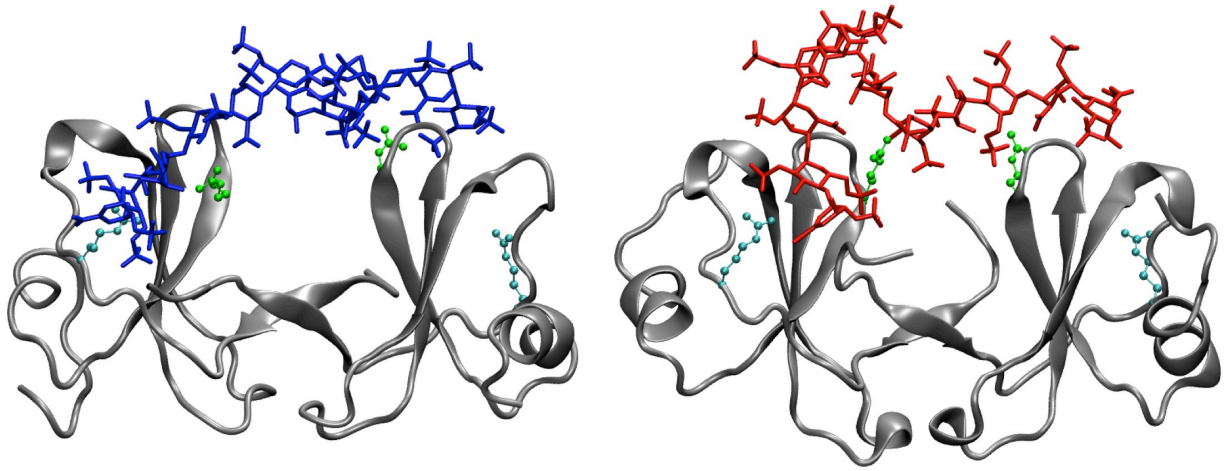

Figure 6.11 Stable pose for dodecamer-CCL5 complex during the MD simulations at neutral (blue) and slightly acidic (red) pH. Residue R17 (cyan) and R47 (green) shown in CPK.

Based on the MM-GBSA interaction energies for the simulation at inflammatory pH, protonation of the histidine residues enabled the heparin chain to be tethered across the two ^{44}RKNR^{47} motifs more strongly, allowing the ^{44}RKNR^{47} motif, particularly from chain B, to make much tighter interactions with the GAG chain. The highest binding contributions still came from the ^{44}RKNR^{47} motifs, but the contribution of one motif was enhanced over the other when the histidine residues were charged. These findings suggest a potential role for H23 as a contributor to CCL5-GAG interactions at sites of inflammation, when the pH is likely to drop

slightly below neutral. The likelihood that a basic charge at position 23 enhances GAG binding has similarly been indicated by a H23K mutant, which was found to bind GAGs with higher affinity than wild-type CCL5 (215).

Table 6.5 Summary of the per-residue theoretical interaction energies[a] (MM-GBSA) for **7** at neutral and slightly acidic pH conditions for residues in the ^{44}RKNR^{47} motif and other key[b] residues

	Neutral pH	Slightly acidic pH
^{44}RKNR^{47} A	-27.9	-27.8
^{44}RKNR^{47} B	-26.6	-40.6
N terminus A	-12.5	-14.8
N terminus B	-2.9	-5.8
R^{47} A	-8.3	-14.0
R^{47} B	-14.0	-16.1
R^{17} A	-1.9	-1.9
R^{17} B	-13.9	-12.4
H^{23}A	-0.9	-5.9
H^{23}B	-2.2	-8.7
Chain A	-52.3	-56.7
Chain B	-50.5	-73.4
Total MM-GBSA[c]	-135.2	-182.2

[a]kcal/mol
[b]residues that contribute more than 1.5 kcal/mol to the interaction energy.
[c]Total from both monomers, conformational entropic effects not included.

Examination of the CCL5-dodecasaccharide complex

Given our data on the contribution of pH to the binding mode of heparin tetrasaccharides we re-examined the crystal structure of the CCL5-disaccharide complex as presented by Shaw et al (17). Interestingly, in the CCL5-heparin disaccharide complex obtained at low pH (156), packing of the protein dimers into the crystallographic unit cell enables the disaccharide to interact simultaneously with the groove region in one dimer, and with the lobes containing the ^{44}RKNR^{47} motif in another (**Figure 6.12**). It is unclear whether in solution the heparin fragment would

prefer to interact with the groove region or with the RKNR motif. The contact area between the disaccharide and the two CCL5 dimers shows that the disaccharide makes a larger contact (222 $Å^2$) with the protein in the groove-binding mode than with the lobes of the opposing dimer (177 $Å^2$). In addition, MM-GBSA binding energy calculations indicate a preference for the groove-interaction mode ($\Delta\Delta G$ = -3.4 kcal/mol). The packing of the protein-disaccharide complexes in the unit cell makes it difficult to say whether the low pH employed in the crystallization protocol is responsible for the placement of the ligand in the groove. However, in the context of our MD simulations of the CCL5-tetrasaccharide complex a pH effect is one possible explanation.

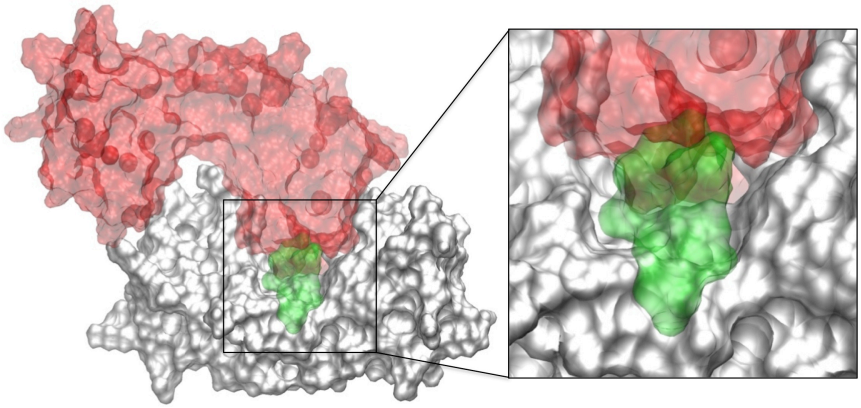

Figure 6.12 Unit cell packing of CCL5 dimers around heparin disaccharide (green). Contact surface with CCL5 groove in dimer 1 (silver) calculated to be 222 $Å^2$, and with the lobes in dimer 2 (red), 177 $Å^2$.

Conclusions

This study integrated multiple experimental and computational methods to elucidate details of the interactions between CCL5 and heparin fragments with varying sulfation pattern and lengths. The modeling data suggested that the positions where heparin fragments bind CCL5 are influenced by the pattern of sulfation as well as the degree of sulfation. The effects of changes in pH on the interaction of heparin tetrasaccharides with CCL5 was monitored by ligand docking

followed by MD simulation. Although AutoDock Vina has been reported to perform reasonably well in carbohydrate docking (175–177), in case of the highly-charged GAG-protein complexes studied here, it was unable to differentiate between changes in the surface charges of CCL5 at the two pH conditions for placement of the GAGs. It is worth noting that the scoring function in AutoDock Vina does not consider partial charges, and instead utilizes an internal method for detecting the potential contributions from hydrogen bonds (76). It is unclear to what extent that approach is able to treat the highly ionic interactions associated with glycosaminoglycans. In addition, the predicted poses from docking were not in agreement with biological expectations, in contrast to the data from MD simulations, where, during the simulations, the change in ionization states markedly altered the positions where most, but not all, of the heparin fragments were found to bind CCL5.

Changes in the surface charge distribution of CCL5, as would result from changes in pH, alter the preferred GAG-binding sites. At low pH the GAG tetramers preferred to interact within the groove between the CCL5 dimer, while at neutral pH most tetrasaccharides preferentially interact with residues outside the groove. Theoretical interaction energy analysis was able to identify the key residues on CCL5 and the key GAG epitopes that are important for binding. These data indicate that variations in the sulfation patterns in GAG fragments that share the same overall level of sulfation affects the nature of their interactions with CCL5, leading to variations in affinity and in the site(s) on the protein most favored for binding. This was also shown to be the case for two tetramers having the same overall sulfation, but a different pattern of sulfation. This suggests that the GAG-CCL5 binding interface differs according to the fine structure of the GAG fragment. The computed interaction energies allowed the tetrasaccharides that demonstrate the best experimental IC_{50} value for inhibition of CCL5-CCR1 binding to be distinguished from

the weak inhibitors. These tetrasaccharides showed the strongest interactions with the residues R47 and R17, both of which are implicated in CCL5-CCR1 binding, further confirming our findings.

The binding of a longer heparin chain to CCL5 was examined with a heparin dodecasaccharide model, which was shown to interact with the ^{44}RKNR^{47} motifs from both chains of CCL5 simultaneously spanning the "40s loops". To model the binding of CCL5 to heparin chains at sites of inflammation *in vivo*, simulations of the CCL5-heparin dodecasaccharide were performed at slightly acidic pH. Residue H23 was identified as a contributor to the CCL5-GAG binding interface at the slightly acidic pH that would be encountered at an inflammatory site. The involvement of this residue during inflammation may be crucial to the heightened CCL5-HS interaction, and to the creation of a concentration gradient for leukocyte migration. The dodecamer also demonstrated its effectiveness as an inhibitor of CCL5-CCR1 interaction through engagement of residues R17 and R47.

Collectively the data presented here cause re-evaluation of the current view that proteins can bind several GAG structures of similar overall charge density (216). To date this discussion has focused on the affinity of the GAG-protein interaction with little or no attention being given to the possibility that different GAG structures of similar overall charge density may bind to different amino acids within the binding face and thereby exhibit different biological activities. In the case of CCL5 this is particularly important as it provides a mechanism for potentially fine tuning the activity of a chemokine which interacts with several receptors, and has thus been described as redundant. The extracellular HS proteoglycan content is known to be modulated under inflammatory conditions, for example chronic exposure to an allergen increases airway HS proteoglycan levels (217), and although HS structural modifications during inflammation have

not yet been defined, they are likely to occur through the activities of the 6-*O*-endosulfatases. Hence, CCL5 could be activated, or inhibited, or directed to bind a receptor that is not influenced by GAG binding depending on the pH and the structure and quantity of HS in that tissue locality. This has significant implications for the development of GAG-related therapeutics.

CHAPTER 7

THEORETICAL CALCULATION OF COLLISION CROSS SECTION

Introduction

Mass spectrometry (MS) has become an important analytical technique for the study of protein-ligand interactions in the past 20 years. Ganem et al. (218) provided the first successful application of mass spectrometry for understanding non-covalent interactions between protein and small molecules when they studied the binding of cytoplasmic receptor FKBP with inhibitors FK506 and rapamycin. Mass spectrometry has since been widely used for characterizing week protein-ligand interactions and has provided specific mass and/or sequence information, binding stoichiometry (219), dissociation constants (220), and information regarding conformational changes associated with binding. Recently, ion mobility spectrometry (IMS), which separates ions based on charge, shape and size, has been used along with mass spectrometry to study protein-ligand complexes. IMS-MS or IMMS, as it is commonly referred to, allows a higher-level separation of complex samples than provided by mass spectrometry alone. The method has gained much popularity since Waters Corporation made Synapt® High Definition Mass Spectrometer (HDMS) commercially available for standalone use in 2006 (221).

A number of IMS techniques have been coupled with MS but each one works on a similar basic principle for ion separation. After ionization of the sample using a suitable ionization technique, a pulse of ions is injected into a cell filled with inert gas at reduced or atmospheric pressure. The ions move through the cell under the influence of an electric field.

The time taken for the ions to travel though the cell depends on the overall shape, size and charge of the ion and is proportional to the rotationally averaged collision cross section (CCS). Different ion mobility separation techniques differ in how the electric field is applied and how the ions travel through the drift cell (221–223). When coupled with mass spectrometry, which separates ions based on mass-to-charge (*m/z*) ratio, the technique provides a two-dimensional separation, based on shape as well as mass. IMS-MS thus allows separation of complex mixtures of structural isomers, chiral compounds and protein/polymer conformers.

Theoretical calculations of CCS using NMR models or X-ray crystal structures are useful for comparing to the differences in experimentally measured CCS values, in order to determine the factors lead to these structural changes. Ligand binding and conformational changes in biomolecules, often account for these differences, and theoretical calculations provide means for elucidating this biological relevance.

Method

The calculation of theoretical CCSs were made using MOBCAL [15, 16]. MOBCAL uses projection approximation (PA) and trajectory method (TM) for calculating the CCS. A static structure may be used for the PA method, but to use the trajectory method an MD simulation needs to be performed. Typically, experimental CCS values are lie between the theoretical PA and TM values.

To use MOBCAL the PDB structure or frames from MD simulation need to be provided in .mfj format, as specified in the user guide. A perl script (provided below) was created that can take MD frames in pdb format to create this input file.

It is also important to note that missing loops and missing residues should ideally be added for more accurate calculation of CCS for proteins. This can be done using MODELLER (226).

Perl script to prepare MOBCAL input file:

```
#Usage Input:
# frames_from_traj(.pdb)   output_file     num_of_atoms
     num_of_frames    job_name
use strict;
if ($#ARGV != 4)
{
     print "Error: Correct usage for this program is:
INPUT_FILE_CONTAINING_FRAMES_FROM_MD OUTPUT_FILE.mfj NUM_OF_ATOMS
NUM_OF_FRAMES JOB_NAME";
     exit();
}

my $input_file  = $ARGV[0];#Maximum 50 frames from MD simulation
trajectory combined into one file
my $output_file = $ARGV[1];      #extension .mfj
my $num_of_atoms = $ARGV[2];     #number of atoms
my $num_of_frames = $ARGV[3];    #number of frames
my $name = $ARGV[4];       #Job name

open (IN,"$input_file") or die "Cannot open input file $!!!\n";
open (OUT,">$output_file") or die "Cannot open output file $!!!\n";

print OUT "$name\n";
print OUT "$num_of_frames\n";
print OUT "$num_of_atoms\n";
print OUT "ang\n";
print OUT "none\n";
print OUT "1.0000\n";

my $i = 0;

while(my $line = <IN>)
{
     if ($i>0 && $line =~ /^MODEL.*\n/)
     {
          print OUT "$num_of_atoms\n";
     }

     if ($line =~ /^ATOM/)
     {
          my @temp_line = split(//,$line);
```

```
            my $count = 0;
            while ($temp_line[$count] ne "")
            {
                  $temp_line[$count] =~ s/\-/ \-/g;
                  $count++;
            }

            $line = join("", @temp_line);

            my @line = split(/\s+/,$line);

            $line[8]=1 if ($line[2] =~ /^H/ || $line[2] =~ /^\dH/);
            $line[8]=12 if ($line[2] =~ /^C/);
            $line[8]=14 if ($line[2] =~ /^N/);
            $line[8]=16 if ($line[2] =~ /^O/);
            $line[8]=32 if ($line[2] =~ /^S/);
            $line[8]=23 if ($line[2] =~ /^Na+/);

            print OUT
"\t$line[5]\t$line[6]\t$line[7]\t$line[8]\t$line[9]\n";
      }

      $i++;
}

print OUT "$num_of_atoms\n";
```

Results

The application of theoretical calculation of collision cross sections of biomolecules to analyze their structural properties is given in appendix “Investigating changes in the Gas-Phase Conformation of Antithrombin III upon binding of Arixtra using Traveling Wave Ion Mobility Mass Spectrometry (TWIMS)”.

CHAPTER 8

CONCLUSIONS AND FUTURE PROSPECTS

The research presented in this thesis aims at computational modeling of GAGs and their interactions with proteins. As outlined in the introduction, modeling GAGs is particularly challenging, and a validated force field parameter set is essential for accurately predicting the conformational properties and binding properties of these molecules. Chapter 4 presents the force field parameters developed in the Woods group for simulating GAG molecules, and their validation through comparison with NMR data. MD simulations were performed on variably sulfated GAG disaccharides containing ΔUA residues, and two synthetic GAG tetrasaccharides. NMR scalar coupling and NOE measurements were collected for comparison with the theoretical data. The new force field parameters were effective at reproducing the NMR data and were able to confirm that the IdoA ring predominantly populates two conformations, $^{1}C_{4}$ and $^{2}S_{O}$. The data also demonstrated that the presence of *N*-sulfation on residues adjacent to sulfated IdoA, increased the solution population of the $^{2}S_{O}$ conformation. The work provides a validated parameter set for the scientific community for further modeling studies of GAG structures.

The Glycosaminoglycan Builder was designed with the aim of facilitating GAG structure modeling. The various classes of GAGs differ in sugar composition and sulfation pattern, and the GAG Builder provides the constituent monosaccharides for each class in their variably sulfated forms for structure building. It also allows flexibility in building alternate conformations of IdoA residues by allowing the user to choose from the three most predominant conformations, default

being $^{1}C_{4}$. This is the first online utility of its kind and we anticipate it being very useful for the community. The interface is intuitive, and may be used by those not experienced in the techniques of molecular modeling, with ease.

Chapter 6 presents the application of these parameters to investigate the binding of heparin fragments to chemokine CCL5. The study integrates experimental and computational methods to elucidate details of the interactions between CCL5 and heparin fragments with varying sulfation pattern and lengths. Six variable sulfated tetrasaccharides were used to demonstrate that the pattern and degree of sulfation, as well as pH, influence the positions adopted by these fragments on the surface of CCL5. Theoretical interaction-energy analysis identified the key residues on CCL5 and the key GAG epitopes that are important for binding. In addition, the tetrasaccharides that demonstrate the best experimental IC_{50} value for inhibition of CCL5-CCR1 binding were distinguished from the weak inhibitors through this analysis.

The binding of a longer heparin chain was used to predict a possible binding mode of HS chains to CCL5 *in* vivo, and offered an explanation for heightened affinity of CCL5 to HS at sites of inflammation. Residue H23 was identified as a contributor to the CCL5-GAG binding interaction at the slightly acidic pH that would be encountered at an inflammatory site, suggesting a role of this residue in binding.

Overall, the data suggests that GAG structures of similar overall charge density bind in specific manner to their interaction partners, making it is important to map these interactions for better design of molecules that may be able to modulate their biological activity.

Future prospects

Interactions of GAGs with chemokines are often thought to be redundant, thereby discouraging GAG-based therapeutic interventions against inflammation. *In vitro* findings suggest that chemokines interact with several GAG fragments with varying affinity, but a lack of specificity is not true, especially *in vivo* (227). Expression of GAGs is a tissue specific phenomenon, and it is believed that fine-tuning of the interactions with chemokines occurs via expression of certain GAGs as a response to certain stimuli (227), and involves specific structural motifs in the GAG chains (107). Thus, a tissue specific identification of precise motifs that engage chemokines to create concentration gradients for sequestering cells is required. Once these motifs are identified therapeutic intervention to modulate their activity would be most feasible. A combination of experimental and modeling techniques will be imperative to gaining better understanding of these systems.

GAGs are also the first adhesion molecules that several viruses recognize and bind to for gaining entry into the host bodies. Many viruses, including the Human immunodeficiency virus (HIV) type 1, Herpes simplex virus (HSV) type 1 and 2, and the Dengue virus are known to utilize this mechanism for enhanced virulence (228–233). Glycoprotein 120 (gp20) on HIV-1 is known to bind HS proteoglycans *in vivo*, and binding affinity experiments show that heparin is the best binder, followed by HS, and that DP16 is the shortest length of heparin oligomer that demonstrates significant binding to gp120 (234). HSV-1 and HSV-2 glycoproteins B and C (gB and gC, respectively) are known to bind to HS chains before the cell fusion event (235). However, these structurally similar proteins on the two serotypes recognize different structural features of heparin; 2,3-*O*-sulfation and 6-*O*-sulfation on heparin have been shown to be critical for binding to HSV-1 but not to HSV-2 (236, 237). An understanding of the exact nature of

these interactions, including the specific structural motifs on GAG chains is essential for designing effective anti-viral strategies.

The availability of the new parameter set and the Glycosaminoglycan builder will allow easy access to molecular structures for modeling, allowing the scientific community to study GAG-protein interactions of interest in detail. An approach that combines experimental-result guided structure modeling and theoretical calculation guided experimental design will be the most useful in this field.

REFERENCES

1. Karamanos, N. K., Syrokou, A., Vanky, P., Nurminen, M., and Hjerpe, A. (1994) Determination of 24 variously sulfated galactosaminoglycan- and hyaluronan-derived disaccharides by high-performance liquid chromatography. *Anal. Biochem.* **221**, 189–99

2. Malavaki, C., Mizumoto, S., Karamanos, N., and Sugahara, K. (2008) Recent advances in the structural study of functional chondroitin sulfate and dermatan sulfate in health and disease. *Connect. Tissue Res.* **49**, 133–9

3. Langer, M. R. (1992) Biosynthesis of glycosaminoglycans in foraminifera: A review. *Mar. Micropaleontol.* **19**, 245–255

4. Esko, J. D., Kimata, K., and Lindahl, U. (2009) Proteoglycans and Sulfated Glycosaminoglycans. in *Essentials of Glycobiology, 2nd edition*, 2nd editio (Varki, A., Cummings, R., and Esko, J. eds), Cold Spring Harbor Laboratory Press

5. Kjellén, L., and Lindahl, U. (1991) Proteoglycans: structures and interactions. *Annu. Rev. Biochem.* **60**, 443–75

6. Bernfield, M., Götte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., and Zako, M. (1999) Functions of cell surface heparan sulfate proteoglycans. *Annu. Rev. Biochem.* **68**, 729–77

7. Antonio, J. D. S., and Iozzo, R. V. (2001) Glycosaminoglycans: structure and biological functions. *eLS*

8. Raman, R., Sasisekharan, V., and Sasisekharan, R. (2005) Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chem. Biol.* **12**, 267–77

9. Rostand, K. S., and Esko, J. D. (1997) Microbial adherence to and invasion through proteoglycans. *Infect. Immun.* **65**, 1–8

10. Schmidtchen, A., Frick, I. M., and Björck, L. (2001) Dermatan sulphate is released by proteinases of common pathogenic bacteria and inactivates antibacterial alpha-defensin. *Mol. Microbiol.* **39**, 708–13

11. Esko, J. D., and Linhardt, R. J. (2009) Proteins that Bind Sulfated Glycosaminoglycans. in *Essentials of Glycobiology, 2nd edition*, 2nd editio (Varki, A., Cummings, R., and Esko, J. eds), Cold Spring Harbor Laboratory Press

12. Fraser, J. R., Laurent, T. C., and Laurent, U. B. (1997) Hyaluronan: its nature, distribution, functions and turnover. *J. Intern. Med.* **242**, 27–33

13. Schulz, T., Schumacher, U., and Prehm, P. (2007) Hyaluronan export by the ABC transporter MRP5 and its modulation by intracellular cGMP. *J. Biol. Chem.* **282**, 20999–1004

14. Bourguignon, L. Y. W. (2009) Hyaluronan-Mediated CD44 Interaction with Receptor and Non-Receptor Kinases Promotes Oncogenic Signaling, Cytoskeleton Activation and Tumor Progression. *Hyaluronan Cancer Biol.* 89-107

15. Banerji, S., Wright, A. J., Noble, M., Mahoney, D. J., Campbell, I. D., Day, A. J., and Jackson, D. G. (2007) Structures of the Cd44-hyaluronan complex provide insight into a fundamental carbohydrate-protein interaction. *Nat. Struct. Mol. Biol.* **14**, 234–9

16. McCarthy, J. B., and Turley, E. A. (2009) RHAMM/HMMR: An Itinerant and Multifunctional Hyaluronan Binding Protein That Modifies CD44 Signaling and Mitotic Spindle Formation, *Hyaluronan Cancer Biol*, 147-168

17. Carlsson, P., and Kjellén, L. (2012) Heparin biosynthesis. *Handb. Exp. Pharmacol.*

18. Lindahl, U., Bäckström, G., Malmström, A., and Fransson, L.-Å. (1972) Biosynthesis of L-iduronic acid in heparin: Epimerization of D-glucuronic acid on the polymer level. *Biochem. Biophys. Res. Commun.* **46**, 985–991

19. Coombe, D. R., and Kett, W. C. (2005) Heparan sulfate-protein interactions: therapeutic potential through structure-function insights. *Cell. Mol. Life Sci.* **62**, 410–424

20. Capila, I., and Linhardt, R. J. (2002) Heparin-protein interactions. *Angew. Chem. Int. Ed. Engl.* **41**, 391–412

21. Kolset, S. O., and Tveit, H. (2008) Serglycin--structure and biology. *Cell. Mol. Life Sci.* **65**, 1073–85

22. Fransson, L. A., Belting, M., Jönsson, M., Mani, K., Moses, J., and Oldberg, A. (2000) Biosynthesis of decorin and glypican. *Matrix Biol.* **19**, 367–76

23. Iozzo, R. V (2005) Basement membrane proteoglycans: from cellar to ceiling. *Nat. Rev. Mol. Cell Biol.* **6**, 646–56

24. Halfter, W., Dong, S., Schurer, B., and Cole, G. J. (1998) Collagen XVIII is a basement membrane heparan sulfate proteoglycan. *J. Biol. Chem.* **273**, 25404–12

25. Bishop, J. R., Schuksz, M., and Esko, J. D. (2007) Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature*. **446**, 1030–7

26. Tkachenko, E., Rhodes, J. M., and Simons, M. (2005) Syndecans: new kids on the signaling block. *Circ. Res.* **96**, 488–500

27. Veugelers, M., De Cat, B., Ceulemans, H., Bruystens, A. M., Coomans, C., Dürr, J., Vermeesch, J., Marynen, P., and David, G. (1999) Glypican-6, a new member of the glypican family of cell surface heparan sulfate proteoglycans. *J. Biol. Chem.* **274**, 26968–77

28. De Cat, B., and David, G. (2001) Developmental roles of the glypicans. *Semin. Cell Dev. Biol.* **12**, 117–25

29. Rehn, M., and Pihlajaniemi, T. (1995) Identification of three N-terminal ends of type XVIII collagen chains and tissue-specific differences in the expression of the corresponding transcripts. The longest form contains a novel motif homologous to rat and Drosophila frizzled proteins. *J. Biol. Chem.* **270**, 4705–11

30. Pufe, T., Petersen, W. J., Miosge, N., Goldring, M. B., Mentlein, R., Varoga, D. J., and Tillmann, B. N. (2004) Endostatin/collagen XVIII--an inhibitor of angiogenesis--is expressed in cartilage and fibrocartilage. *Matrix Biol.* **23**, 267–76

31. Milev, P., Maurel, P., Chiba, A., Mevissen, M., Popp, S., Yamaguchi, Y., Margolis, R. K., and Margolis, R. U. (1998) Differential regulation of expression of hyaluronan-binding proteoglycans in developing brain: aggrecan, versican, neurocan, and brevican. *Biochem. Biophys. Res. Commun.* **247**, 207–12

32. Trowbridge, J. M., and Gallo, R. L. (2002) Dermatan sulfate: new functions from an old glycosaminoglycan. *Glycobiology*. **12**, 117R–125R

33. Arai, T., Parker, A., Busby, W. J., and Clemmons, D. R. (1994) Heparin, heparan sulfate, and dermatan sulfate regulate formation of the insulin-like growth factor-I and insulin-like growth factor-binding protein complexes. *J. Biol. Chem.* **269**, 20388–20393

34. Grande-Allen, K. J., Griffin, B. P., Calabro, A., Ratliff, N. B., Cosgrove, D. M., and Vesely, I. (2001) Myxomatous mitral valve chordae. II: Selective elevation of glycosaminoglycan content. *J. Heart Valve Dis.* **10**, 325–32; discussion 332–3

35. Funderburgh, J. L. (2002) Keratan Sulfate Biosynthesis. *IUBMB Life (International Union Biochem. Mol. Biol. Life)*. **54**, 187–194

36. Funderburgh, J. L., Caterson, B., and Conrad, G. W. (1987) Distribution of proteoglycans antigenically related to corneal keratan sulfate proteoglycan. *J. Biol. Chem.* **262**, 11634–40

37. Edward, D. P., Thonar, E. J., Srinivasan, M., Yue, B. J., and Tso, M. O. (1990) Macular dystrophy of the cornea. A systemic disorder of keratan sulfate metabolism. *Ophthalmology*. **97**, 1194–200

38. Hassell, J. R., Newsome, D. A., Krachmer, J. H., and Rodrigues, M. M. (1980) Macular corneal dystrophy: failure to synthesize a mature keratan sulfate proteoglycan. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 3705–9

39. Mulloy, B., and Forster, M. J. (2000) Conformation and dynamics of heparin and heparan sulfate. *Glycobiology*. **10**, 1147–56

40. Ferro, D. R., Provasoli, A., Ragazzi, M., Casu, B., Torri, G., Bossennec, V., Perly, B., Sinaÿ, P., Petitou, M., and Choay, J. (1990) Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr. Res.* **195**, 157–67

41. van Boeckel, C. A. A., van Aelst, S. F., Wagenaars, G. N., Mellema, J.-R., Paulsen, H., Peters, T., Pollex, A., and Sinnwell, V. (2010) Conformational analysis of synthetic heparin-like oligosaccharides containing α-L-idopyranosyluronic acid. *Recl. des Trav. Chim. des Pays-Bas*. **106**, 19–29

42. Mulloy, B., Forster, M. J., Jones, C., and Davies, D. B. (1993) NMR and molecular-modelling studies of the solution conformation of heparin. *Biochem. J.* **293 (Pt 3)** 849–58

43. Mulloy, B., Forster, M. J., Jones, C., Drake, A. F., Johnson, E. A., and Davies, D. B. (1994) The effect of variation of substitution on the solution conformation of heparin: a spectroscopic and molecular modelling study. *Carbohydr. Res.* **255**, 1–26

44. Ragazzi, M., Ferro, D. R., Provasoli, A., Pumilia, P., Cassinari, A., Torri, G., Guerrini, M., Case, B., Nader, H. B., and Dietrich, C. P. (1993) Conformation of the Unsaturated Uronic Acid Residues of Glycosamtnoglycan Disaccharides. *J. Carbohydr. Chem.***12(4-5)** 523-535

45. Capila, I., Hernáiz, M. J., Mo, Y. D., Mealy, T. R., Campos, B., Dedman, J. R., Linhardt, R. J., and Seaton, B. A. (2001) Annexin V--heparin oligosaccharide complex suggests heparan sulfate--mediated assembly on cell surfaces. *Structure*. **9**, 57–64

46. Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J., and Rees, D. C. (1996) Heparin structure and interactions with basic fibroblast growth factor. *Science*. **271**, 1116–20

47. Fry, E. E., Lea, S. M., Jackson, T., Newman, J. W., Ellard, F. M., Blakemore, W. E., Abu-Ghazaleh, R., Samuel, A., King, A. M., and Stuart, D. I. (1999) The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex. *EMBO J.* **18**, 543–54

48. Fromm, J. R., Hileman, R. E., Caldwell, E. E., Weiler, J. M., and Linhardt, R. J. (1997) Pattern and spacing of basic amino acids in heparin binding sites. *Arch. Biochem. Biophys.* **343**, 92–100

49. Hileman, R. E., Fromm, J. R., Weiler, J. M., and Linhardt, R. J. (1998) Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *Bioessays*. **20**, 156–67

50. Coombe, D. R. (2008) Biological implications of glycosaminoglycan interactions with haemopoietic cytokines. *Immunol. Cell Biol.* **86**, 598–607

51. Lortat-Jacob, H., Grosdidier, A., and Imberty, A. (2002) Structural diversity of heparan sulfate binding domains in chemokines. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1229–1234

52. Cardin, A. D., and Weintraub, H. J. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arterioscler. Thromb. Vasc. Biol.* **9**, 21–32

53. Thompson, L. D., Pantoliano, M. W., and Springer, B. A. (1994) Energetic characterization of the basic fibroblast growth factor-heparin interaction: identification of the heparin binding domain. *Biochemistry*. **33**, 3831–40

54. FAUCHÈRE, J.-L., CHARTON, M., KIER, L. B., VERLOOP, A., and PLISKA, V. (2009) Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int. J. Pept. Protein Res.* **32**, 269–278

55. Gandhi, N. S., and Mancera, R. L. (2008) The structure of glycosaminoglycans and their interactions with proteins. *Chem. Biol. Drug Des.* **72**, 455–482

56. Goedert, M., Jakes, R., Spillantini, M. G., Hasegawa, M., Smith, M. J., and Crowther, R. A. (1996) Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans. *Nature*. **383**, 550–3

57. Mori, H., Takio, K., Ogawarag, M., and Selkoe, D. J. (1992) Mass Spectrometry of Purified Amyloid β Protein in Alzheimer's Disease. *J. Biol. Chem.* **267**, 17082–17086

58. Brunden, K. R., Richter-Cook, N. J., Chaturvedi, N., and Frederickson, R. C. (1993) pH-dependent binding of synthetic beta-amyloid peptides to glycosaminoglycans. *J. Neurochem.* **61**, 2147–2154

59. Busby, T. F., Atha, D. H., and Inghamg, K. C. (1981) Thermal Denaturation of Antithrombin III. *J. Biol. Chem.* **256**, 12140–12147

60. Tsai, a M., van Zanten, J. H., and Betenbaugh, M. J. (1998) II. Electrostatic effect in the aggregation of heat-denatured RNase A and implications for protein additive design. *Biotechnol. Bioeng.* **59**, 281–5

61. Zhang, G., Foegeding, E. A., and Hardin, C. C. (2004) Effect of sulfated polysaccharides on heat-induced structural changes in beta-lactoglobulin. *J. Agric. Food Chem.* **52**, 3975–81

62. Chung, K., Kim, J., Cho, B.-K., Ko, B.-J., Hwang, B.-Y., and Kim, B.-G. (2007) How does dextran sulfate prevent heat induced aggregation of protein? The mechanism and its limitation as aggregation inhibitor. *Biochim. Biophys. Acta*. **1774**, 249–257

63. Militello, V., Casarino, C., Emanuele, A., Giostra, A., Pullara, F., and Leone, M. (2004) Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and light scattering. *Biophys. Chem.* **107**, 175–87

64. Hallgren, J., Bäckström, S., Estrada, S., Thuveson, M., and Pejler, G. (2004) Histidines are critical for heparin-dependent activation of mast cell tryptase. *J. Immunol.* **173**, 1868–1875

65. Humphries, D. E., Wong, G. W., Friend, D. S., Gurish, M. F., Huang, C., Sharpe, A. H., and Stevens, R. L. (1999) Heparin is essential for the storage of specific granule proteases in mast cells. *Nature*. **400**, 769–772

66. Ren, S., Sakai, K., and Schwartz, L. B. (1998) Regulation of human mast cell beta-tryptase: conversion of inactive monomer to active tetramer at acid pH. *J. Immunol.* **160**, 4561–9

67. Matsumoto, R., Sali, A., Ghildyal, N., Karplus, M., and Stevens, R. L. (1995) Packaging of Proteases and Proteoglycans in the Granules of Mast cells and other Hematopoetic cells. *J. Biol. Chem.* **270**, 19524–19531

68. Ghildyal, B. N., Friend, D. S., Stevens, R. L., Austen, K. F., Huang, C., Penrose, J. E., Sali, A., and Gurish, M. E. (1996) Fate of Two Mast Cell Tryptases in V3 Mastocytosis and Normal BALB/c Mice Undergoing Passive Systemic Anaphylaxis: Prolonged Retention of Exocytosed mMCP-6 in Connective Tissues, and Rapid Accumulation of Enzymatically Active mMCP-7 in the Blood. *J. Exp. Med.* **184**, 1061–1073

69. Singh, A., Kett, W. C., Severin, I. C., Agyekum, I., Duan, J., Amster, I. J., Proudfoot, A. E. I., Coombe, D. R., and Woods, R. J. (2015) The Interaction of Heparin Tetrasaccharides with Chemokine CCL5 Is Modulated by Sulfation Pattern and pH. *J. Biol. Chem.* **290**, 15421–36

70. Forster, M., and Mulloy, B. (2006) Computational approaches to the identification of heparin-binding sites on the surfaces of proteins. *Biochem. Soc. Trans.* **34**, 431–434

71. Imberty, A., Lortat-Jacob, H., and Pérez, S. (2007) Structural view of glycosaminoglycan-protein interactions. *Carbohydr. Res.* **342**, 430–439

72. Goodford, P. J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849–57

73. Bitomsky, W., and Wade, R. C. (1999) Docking of Glycosaminoglycans to Heparin-Binding Proteins: Validation for aFGF, bFGF, and Antithrombin and Application to IL-8. *J. Am. Chem. Soc.* **121**, 3004–3013

74. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662

75. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–91

76. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461

77. Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–48

78. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–49

79. Ewing, T. J., Makino, S., Skillman, A. G., and Kuntz, I. D. (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided. Mol. Des.* **15**, 411–28

80. Lam, K., Rao, V. S., and Qasba, P. K. (1998) Molecular modeling studies on binding of bFGF to heparin and its receptor FGFR1. *J. Biomol. Struct. Dyn.* **15**, 1009–27

81. Sadir, R., Baleux, F., Grosdidier, A., Imberty, A., and Lortat-Jacob, H. (2001) Characterization of the stromal cell-derived factor-1alpha-heparin complex. *J. Biol. Chem.* **276**, 8288–96

82. Stringer, S. E., Forster, M. J., Mulloy, B., Bishop, C. R., Graham, G. J., and Gallagher, J. T. (2002) Characterization of the binding site on heparan sulfate for macrophage inflammatory protein 1alpha. *Blood*. **100**, 1543–50

83. Ricard-Blum, S., Féraud, O., Lortat-Jacob, H., Rencurosi, A., Fukai, N., Dkhissi, F., Vittet, D., Imberty, A., Olsen, B. R., and van der Rest, M. (2004) Characterization of endostatin binding to heparin and heparan sulfate by surface plasmon resonance and molecular modeling: role of divalent cations. *J. Biol. Chem.* **279**, 2927–36

84. Gandhi, N. S., Coombe, D. R., and Mancera, R. L. (2008) Platelet endothelial cell adhesion molecule 1 (PECAM-1) and its interactions with glycosaminoglycans: 1. Molecular modeling studies. *Biochemistry*. **47**, 4851–62

85. Coombe, D. R., Stevenson, S. M., Kinnear, B. F., Gandhi, N. S., Mancera, R. L., Osmond, R. I. W., and Kett, W. C. (2008) Platelet endothelial cell adhesion molecule 1 (PECAM-1) and its interactions with glycosaminoglycans: 2. Biochemical analyses. *Biochemistry*. **47**, 4863–75

86. Gandhi, N. S., Freeman, C., Parish, C. R., and Mancera, R. L. (2012) Computational analyses of the catalytic and heparin-binding sites and their interactions with glycosaminoglycans in glycoside hydrolase family 79 endo-β-D-glucuronidase (heparanase). *Glycobiology*. **22**, 35–55

87. Raghuraman, A., Mosier, P. D., and Desai, U. R. (2006) Finding a needle in a haystack: development of a combinatorial virtual screening approach for identifying high specificity heparin/heparan sulfate sequence(s). *J. Med. Chem.* **49**, 3553–62

88. Verli, H., and Guimarães, J. A. (2004) Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution. *Carbohydr. Res.* **339**, 281–290

89. Ferro, D. R., Provasoli, A., Ragazzi, M., Torri, G., Casu, B., Gatti, G., Jacquinet, J. C., Sinay, P., Petitou, M., and 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. *J. Am. Chem. Soc.* **108**, 6773–6778

90. Gandhi, N. S., and Mancera, R. L. (2009) Free energy calculations of glycosaminoglycan-protein interactions. *Glycobiology*. **19**, 1103–1115

91. Sapay, N., Cabannes, E., Petitou, M., and Imberty, A. (2011) Molecular modeling of the interaction between heparan sulfate and cellular growth factors: bringing pieces together. *Glycobiology*. **21**, 1181–93

92. Pichert, A., Samsonov, S. a, Theisgen, S., Thomas, L., Baumann, L., Schiller, J., Beck-Sickinger, A. G., Huster, D., and Pisabarro, M. T. (2012) Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. *Glycobiology*. **22**, 134–145

93. Cheatham, T. E. I. (2004) Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.* **14**, 360–7

94. Fadda, E., and Woods, R. J. (2010) Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today*. **15**, 596–609

95. Feller, S. E. (2000) Molecular dynamics simulations of lipid bilayers. *Curr. Opin. Colloid Interface Sci.* **5**, 217–223

96. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. **78**, 1950–8

97. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655

98. Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J., Haque, I., Mobley, D. L., Lambrecht, D. S., DiStasio, R. A., Head-Gordon, M., Clark, G. N. I., Johnson, M. E., and Head-Gordon, T. (2010) Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B*. **114**, 2549–64

99. Verlet, L. (1967) Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 98–103

100. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935

101. Mahoney, M. W., and Jorgensen, W. L. (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **112**, 8910

102. Arfken, G. (1895) The method of steepest descents. in *Mathematical methods for physicists*, Academic Press Orlando, FL

103. Hestenes, M. R., and Eduard, S. (1952) Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand. (1934).* **49**, 409–436

104. Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–49

105. Smith, G. R., and Sternberg, M. J. E. (2002) Prediction of protein–protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **12**, 28–35

106. Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2006) Protein-ligand docking: current status and future challenges. *Proteins*. **65**, 15–26

107. Kreuger, J., Spillmann, D., Li, J., and Lindahl, U. (2006) Interactions between heparan sulfate and proteins: the concept of specificity. *J. Cell Biol.* **174**, 323–7

108. Gama, C. I., Tully, S. E., Sotogaku, N., Clark, P. M., Rawat, M., Vaidehi, N., Goddard, W. A., Nishi, A., and Hsieh-Wilson, L. C. (2006) Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat. Chem. Biol.* **2**, 467–73

109. van Boeckel, C. A. A., and Petitou, M. (1993) The Unique Antithrombin III Binding Domain of Heparin: A Lead to New Synthetic Antithrombotics. *Angew. Chemie Int. Ed. English*. **32**, 1671–1690

110. Bazin, H. G., Capila, I., and Linhardt, R. J. (1998) Conformational study of synthetic Δ4-uronate monosaccharides and glycosaminoglycan-derived disaccharides. *Carbohydr. Res.* **309**, 135–144

111. Jin, L., Hricovíni, M., Deakin, J. A., Lyon, M., and Uhrín, D. (2009) Residual dipolar coupling investigation of a heparin tetrasaccharide confirms the limited effect of flexibility of the iduronic acid on the molecular shape of heparin. *Glycobiology*. **19**, 1185–96

112. Mikhailov, D., Linhardt, R. J., and Mayo, K. H. (1997) NMR solution conformation of heparin-derived hexasaccharide. *Biochem. J.* **328 (Pt 1)** 51–61

113. Blanchard, V., Chevalier, F., Imberty, A., Leeflang, B. R., Basappa, Sugahara, K., and Kamerling, J. P. (2007) Conformational studies on five octasaccharides isolated from chondroitin sulfate using NMR spectroscopy and molecular modeling. *Biochemistry*. **46**, 1167–75

114. Sattelle, B. M., Shakeri, J., Roberts, I. S., and Almond, A. (2010) A 3D-structural model of unsulfated chondroitin from high-field NMR: 4-sulfation has little effect on backbone conformation. *Carbohydr. Res.* **345**, 291–302

115. Sattelle, B. M., Shakeri, J., and Almond, A. (2013) Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules*. **14**, 1149–59

116. Gandhi, N. S., and Mancera, R. L. (2010) Can current force fields reproduce ring puckering in 2-O-sulfo-alpha-L-iduronic acid? A molecular dynamics simulation study. *Carbohydr. Res.* **345**, 689–95

117. Mallajosyula, S. S., Guvench, O., Hatcher, E., and Mackerell, A. D. (2012) CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *J. Chem. Theory Comput.* **8**, 759–776

118. D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B., X. W. and P. A. K. (2012) AMBER12, University of California, San Francisco

119. Basma, M., Sundara, S., Calgan, D., Vernali, T., and Woods, R. J. (2001) Solvated ensemble averaging in the calculation of partial atomic charges. *J. Comput. Chem.* **22**, 1125–37

120. Tessier, M. B., Demarco, M. L., Yongye, A. B., and Woods, R. J. (2008) Extension of the GLYCAM06 Biomolecular Force Field to Lipids, Lipid Bilayers and Glycolipids. *Mol. Simul.* **34**, 349–363

121. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*. **6**, 277–93

122. Goddard, T. D., and Kneller, D. G. SPARKY 3

123. Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–52

124. MACURA, S., and ERNST, R. R. (1980) Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy. *Mol. Phys.* **100**, 135–147

125. Haasnoot, C. A. G., de Leeuw, F. A. A. M., and Altona, C. (1980) The relationship between proton-proton NMR coupling constants and substituent electronegativities—I. *Tetrahedron*. **36**, 2783–2792

126. Altona, C., Francke, R., de Haan, R., Ippel, J. H., Daalmans, G. J., Hoekzema, A. J. A. W., and van Wijk, J. (1994) Empirical group electronegativities for vicinal NMR proton-proton couplings along a C-C bond: Solvent effects and reparameterization of the Haasnoot equation. *Magn. Reson. Chem.* **32**, 670–678

127. Yongye, A. B., Foley, B. L., and Woods, R. J. (2008) On achieving experimental accuracy from molecular dynamics simulations of flexible molecules: aqueous glycerol. *J. Phys. Chem. A*. **112**, 2634–9

128. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690

129. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. . (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341

130. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8**, 1542–1555

131. Onufriev, A., Bashford, D., and Case, D. A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*. **55**, 383–394

132. Frisch, M. J. . (2009) Gaussian 09, Revision D.01

133. Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollmann, P. A. (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **115**, 9620–9631

134. Allen, F. H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B.* **58**, 380–8

135. Maruyama, T., Toida, T., Imanari, T., Yu, G., and Linhardt, R. J. (1998) Conformational changes and anticoagulant activity of chondroitin sulfate following its O-sulfonation. *Carbohydr. Res.* **306**, 35–43

136. Gatti, G., Casu, B., Hamer, G. K., and Perlin, A. S. (1979) Studies on the Conformation of Heparin by 1 H and 13 C NMR Spectroscopy. *Macromolecules*. **12**, 1001–1007

137. Berman, H. M. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242

138. Sattelle, B. M., Bose-Basu, B., Tessier, M., Woods, R. J., Serianni, A. S., and Almond, A. (2012) Dependence of pyranose ring puckering on anomeric configuration: methyl idopyranosides. *J. Phys. Chem. B*. **116**, 6380–6

139. Makeneni, S., Ji, Y., Watson, D. C., Young, N. M., and Woods, R. J. (2014) Predicting the Origins of Anti-Blood Group Antibody Specificity: A Case Study of the ABO A- and B-Antigens. *Front. Immunol.* **5**, 397

140. Lak, P., Makeneni, S., Woods, R. J., and Lowary, T. L. (2015) Specificity of furanoside-protein recognition through antibody engineering and molecular modeling. *Chemistry*. **21**, 1138–48

141. http://ambermd.org/

142. Tanaka, Y., Adams, D. H., and Shaw, S. (1993) Proteoglycans on endothelial cells present adhesion-inducing cytokines to leukocytes. *Immunol. Today*. **14**, 111–115

143. Rot, A. (1992) Endothelial cell binding of NAP-1/IL-8: role in neutrophil emigration. *Immunol. Today*. **13**, 291–294

144. Springer, T. A. (1994) Traffic signals for lymphocyte recirculation and leukocyte emigration: The multistep paradigm. *Cell*. **76**, 301–314

145. Weber, M., Hauschild, R., Schwarz, J., Moussion, C., de Vries, I., Legler, D. F., Luther, S. A., Bollenbach, T., and Sixt, M. (2013) Interstitial dendritic cell guidance by haptotactic chemokine gradients. *Science*. **339**, 328–332

146. Tanino, Y., Coombe, D. R., Gill, S. E., Kett, W. C., Kajikawa, O., Proudfoot, A. E. I., Wells, T. N. C., Parks, W. C., Wight, T. N., Martin, T. R., and Frevert, C. W. (2010) Kinetics of chemokine-glycosaminoglycan interactions control neutrophil migration into the airspaces of the lungs. *J. Immunol.* **184**, 2677–2685

147. Severin, I. C., Soares, A., Hantson, J., Teixeira, M., Sachs, D., Valognes, D., Scheer, A., Schwarz, M. K., Wells, T. N. C., Proudfoot, A. E. I., and Shaw, J. (2012) Glycosaminoglycan analogs as a novel anti-inflammatory strategy. *Front. Immunol.* **3**, 293

148. Nonaka, M., Bao, X., Matsumura, F., Götze, S., Kandasamy, J., Kononov, A., Broide, D. H., Nakayama, J., Seeberger, P. H., and Fukuda, M. (2014) Synthetic di-sulfated iduronic acid attenuates asthmatic response by blocking T-cell recruitment to inflammatory sites. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8173–8178

149. Allen, S. J., Crown, S. E., and Handel, T. M. (2007) Chemokine: receptor structure, interactions, and antagonism. *Annu. Rev. Immunol.* **25**, 787–820

150. Proudfoot, A. E. I., Handel, T. M., Johnson, Z., Lau, E. K., LiWang, P., Clark-Lewis, I., Borlat, F., Wells, T. N. C., and Kosco-Vilbois, M. H. (2003) Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1885–1990

151. Kuschert, G. S. V, Coulin, F., Power, C. A., Proudfoot, A. E. I., Hubbard, R. E., Hoogewerf, A. J., and Wells, T. N. C. (1999) Glycosaminoglycans Interact Selectively with Chemokines and Modulate Receptor Binding and Cellular Responses. *Biochemistry* **38,** 12959-68

152. Baggiolini, M., Dewald, B., and Moser, B. (1997) Human chemokines: an update. *Annu. Rev. Immunol.* **15**, 675–705

153. Proudfoot, A. E. I. (2002) Chemokine receptors: multifaceted therapeutic targets. *Nat. Rev. Immunol.* **2**, 106–115

154. Proudfoot, A. E. I., Fritchley, S., Shaw, J. P., Vilbois, F., Zwahlen, C., Trkola, A., Marchant, D., Clapham, P. R., and Wells, T. N. C. (2001) The BBXB Motif of RANTES Is the Principal Site for Heparin Binding and Controls Receptor Selectivity. *J. Biol. Chem.* **276**, 10620–10626

155. Wang, X., Watson, C., Sharp, J. S., Handel, T. M., and Prestegard, J. H. (2011) Oligomeric structure of the chemokine CCL5/RANTES from NMR, MS, and SAXS data. *Structure*. **19**, 1138–1148

156. Shaw, J. P., Johnson, Z., Borlat, F., Zwahlen, C., Kungl, A., Roulin, K., Harrenga, A., Wells, T. N. C., and Proudfoot, A. E. I. (2004) The X-ray structure of RANTES: heparin-derived disaccharides allows the rational design of chemokine inhibitors. *Structure*. **12**, 2081–2093

157. Johnson, Z., Kosco-Vilbois, M. H., Herren, S., Cirillo, R., Muzio, V., Zaratin, P., Carbonatto, M., Mack, M., Smailbegovic, A., Rose, M., Lever, R., Page, C., Wells, T. N. C., and Proudfoot, A. E. I. (2004) Interference with heparin binding and oligomerization creates a novel anti-inflammatory strategy targeting the chemokine system. *J. Immunol.* **173**, 5776–5785

158. Chung, C. W., Cooke, R. M., Proudfoot, A. E., and Wells, T. N. (1995) The three-dimensional solution structure of RANTES. *Biochemistry*. **34**, 9307–9314

159. Wood, E. J. (1987) Data for Biochemical Research (third edition). *Biochem. Educ.* **15**, 97

160. Wang, H. M., Loganathan, D., and Linhardt, R. J. (1991) Determination of the pKa of glucuronic acid and the carboxy groups of heparin by 13C-nuclear-magnetic-resonance spectroscopy. *Biochem. J.* **278 (Pt 3)**, 689–695

161. Remko, M., Broer, R., and Van Duijnen, P. T. (2013) How acidic are monomeric structural units of heparin? *Chem. Phys. Lett.* **590**, 187–191

162. Salmivirta, M., Lidholt, K., and Lindahl, U. (1996) Heparan sulfate: a piece of information. *FASEB J.* **10**, 1270–1279

163. Turnbull, J., Powell, A., and Guimond, S. (2001) Heparan sulfate: decoding a dynamic multifunctional cell regulator. *Trends Cell Biol.* **11**, 75–82

164. Lazennec, G., and Richmond, A. (2010) Chemokines and chemokine receptors: new insights into cancer-related inflammation. *Trends Mol. Med.* **16**, 133–144

165. Marques, R. E., Guabiraba, R., Russo, R. C., and Teixeira, M. M. (2013) Targeting CCL5 in inflammation. *Expert Opin. Ther. Targets*. **17**, 1439–1460

166. Aldinucci, D., & Colombatti, A. (2014) The inflammatory chemokine CCL5 and cancer progression. *Mediators Inflamm.*

167. Martin, L., Blanpain, C., Garnier, P., Wittamer, V., Parmentier, M., and Vita, C. (2001) Structural and functional analysis of the RANTES-glycosaminoglycans interactions. *Biochemistry*. **40**, 6303–6318

168. Pakianathan, D. R., Kuta, E. G., Artis, D. R., Skelton, N. J., and Hébert, C. A. (1997) Distinct but overlapping epitopes for the interaction of a CC-chemokine with CCR1, CCR3 and CCR5. *Biochemistry*. **36**, 9642–9648

169. Proudfoot, A. E. I., Power, C. A., Hoogewerf, A. J., Montjovent, M.-O., Borlat, F., Offord, R. E., and Wells, T. N. C. (1996) Extension of Recombinant Human RANTES by the Retention of the Initiating Methionine Produces a Potent Antagonist. *J. Biol. Chem.* **271**, 2599–2603

170. Samsonov, S. A., Gehrcke, J.-P., and Pisabarro, M. T. (2014) Flexibility and explicit solvent in molecular-dynamics-based docking of protein-glycosaminoglycan systems. *J. Chem. Inf. Model.* **54**, 582–92

171. Mottarella, S. E., Beglov, D., Beglova, N., Nugent, M. A., Kozakov, D., and Vajda, S. (2014) Docking server for the identification of heparin binding sites on proteins. *J. Chem. Inf. Model.* **54**, 2068–2078

172. Ballut, L., Sapay, N., Chautard, E., Imberty, A., and Ricard-Blum, S. (2013) Mapping of heparin/heparan sulfate binding sites on αvβ3 integrin by molecular docking. *J. Mol. Recognit.* **26**, 76–85

173. Carpentier, M., Denys, A., Allain, F., and Vergoten, G. (2014) Molecular docking of heparin oligosaccharides with Hep-II heparin-binding domain of fibronectin reveals an interplay between the different positions of sulfate groups. *Glycoconj. J.* **31**, 161–169

174. Samsonov, S. A., Teyra, J., and Pisabarro, M. T. (2011) Docking glycosaminoglycans to proteins: analysis of solvent inclusion. *J. Comput. Aided. Mol. Des.* **25**, 477–489

175. Mishra, S. K., Adam, J., Wimmerová, M., and Koča, J. (2012) In silico mutagenesis and docking study of Ralstonia solanacearum RSL lectin: performance of docking software to predict saccharide binding. *J. Chem. Inf. Model.* **52**, 1250–1261

176. Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B., and Woods, R. J. (2014) Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J. Comput. Chem.* **35**, 526–539

177. Frank, M. (2014) Computational Docking as a Tool for the Rational Design of Carbohydrate-Based Drugs. in *Topics in Medicinal Chemistry*, pp. 1–20, Springer Berlin Heidelberg

178. Atkovska, K., Samsonov, S. A., Paszkowski-Rogacz, M., and Pisabarro, M. T. (2014) Multipose binding in molecular docking. *Int. J. Mol. Sci.* **15**, 2622–2645

179. Valle-Delgado, J. J., Alfonso-Prieto, M., de Groot, N. S., Ventura, S., Samitier, J., Rovira, C., and Fernàndez-Busquets, X. (2010) Modulation of Abeta42 fibrillogenesis by glycosaminoglycan structure. *FASEB J.* **24**, 4250–4261

180. Chai, W., Hounsell, E. F., Bauer, C. J., and Lawson, A. M. (1995) Characterisation by LSI-MS and 1H NMR spectroscopy of tetra-, hexa-, and octa-saccharides of porcine intestinal heparin. *Carbohydr. Res.* **269**, 139–156

181. Kett, W. C., and Coombe, D. R. (2004) A structural analysis of heparin–like glycosaminoglycans using MALDI–TOF mass spectrometry. *Spectroscopy*. **18**, 185–201

182. Mourier, P. A. J., and Viskov, C. (2004) Chromatographic analysis and sequencing approach of heparin oligosaccharides using cetyltrimethylammonium dynamically coated stationary phases. *Anal. Biochem.* **332**, 299–313

183. Wolff, J. J., Amster, I. J., Chi, L., and Linhardt, R. J. (2007) Electron detachment dissociation of glycosaminoglycan tetrasaccharides. *J. Am. Soc. Mass Spectrom.* **18**, 234–244

184. Domon, B., and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5**, 397–409

185. Arunima Singh, Matthew B. Tessier, Kari Pederson, Gordon R. Chalmers, James H. Prestegard, R. J. W. Conformational analysis of glycosaminoglycans by nmr and molecular dynamics simulations using parameter extensions for the GLYCAM force field. *Manuscr. Prep.*

186. Lim, J. K., Burns, J. M., Lu, W., and DeVico, A. L. (2005) Multiple pathways of amino terminal processing produce two truncated variants of RANTES/CCL5. *J. Leukoc. Biol.* **78**, 442–452

187. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. **65**, 712–725

188. Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095

189. R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. *R Found. Stat. Comput. Vienna, Austria.*

190. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38

191. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **33**, 889–897

192. Srinivasan, J., Miller, J., Kollman, P. A., and Case, D. A. (1998) Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* **16**, 671–682

193. Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **8**, 3314–3321

194. Onufriev, A., Bashford, D., and Case, D. A. (2000) Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B*. **104**, 3712–3720

195. Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **51**, 69–82

196. Hubbard, S. J., and Thornton, J. M. (1993) NACCESS, Computer Program, Department of Biochemistry and Molecular Biology, University College London

197. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, 665–667

198. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, 522–525

199. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins*. **61**, 704–721

200. Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins*. **73**, 765–783

201. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **7**, 525–537

202. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. *J. Chem. Theory Comput.* **7**, 2284–2295

203. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041

204. Kailemia, M. J., Li, L., Xu, Y., Liu, J., Linhardt, R. J., and Amster, I. J. (2013) Structurally informative tandem mass spectrometry of highly sulfated natural and chemoenzymatically synthesized heparin and heparan sulfate glycosaminoglycans. *Mol. Cell. Proteomics*. **12**, 979–990

205. Kailemia, M. J., Li, L., Ly, M., Linhardt, R. J., and Amster, I. J. (2012) Complete mass spectral characterization of a synthetic ultralow-molecular-weight heparin using collision-induced dissociation. *Anal. Chem.* **84**, 5475–5478

206. Genheden, S., Kuhn, O., Mikulskis, P., Hoffmann, D., and Ryde, U. (2012) The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant. *J. Chem. Inf. Model.* **52**, 2079–2088

207. Vivès, R. R., Sadir, R., Imberty, A., Rencurosi, A., and Lortat-Jacob, H. (2002) A kinetics and modeling study of RANTES(9-68) binding to heparin reveals a mechanism of cooperative oligomerization. *Biochemistry*. **41**, 14779–14789

208. Jones, A. L., Hulett, M. D., and Parish, C. R. (2005) Histidine-rich glycoprotein: A novel adaptor protein in plasma that modulates the immune, vascular and coagulation systems. *Immunol. Cell Biol.* **83**, 106–118

209. Blundell, C. D., Mahoney, D. J., Cordell, M. R., Almond, A., Kahmann, J. D., Perczel, A., Taylor, J. D., Campbell, I. D., and Day, A. J. (2007) Determining the molecular basis for the pH-dependent interaction between the link module of human TSG-6 and hyaluronan. *J. Biol. Chem.* **282**, 12976–12988

210. Vielhauer, V., Berning, E., Eis, V., Kretzler, M., Segerer, S., Strutz, F., Horuk, R., Gröne, H.-J., Schlöndorff, D., and Anders, H.-J. (2004) CCR1 blockade reduces interstitial inflammation and fibrosis in mice with glomerulosclerosis and nephrotic syndrome. *Kidney Int.* **66**, 2264–2278

211. Häbler, C. (1929) Über den K- und Ca-Gehalt von Eiter und Exsudaten und Seine Beziehungen zum Entzündungsschmerz. *Klin. Wochenschr.* **8**, 1569–1572

212. Revici, E., Stoopen, E., Frenk, E., and Ravich, R. (1949) The painful focus. II. The relation of pain to local physiochemical changes. *Bull Inst Appl Biol*

213. Peer, L. A. (1955) Transplantation Of Tissues, Vol-1

214. Jacobus, W. E., Taylor, G. J., Hollis, D. P., and Nunnally, R. L. (1977) Phosphorus nuclear magnetic resonance of perfused working rat hearts. *Nature*. **265**, 756–758

215. Brandner, B., Rek, A., Diedrichs-Möhring, M., Wildner, G., and Kungl, A. J. (2009) Engineering the glycosaminoglycan-binding affinity, kinetics and oligomerization behavior of RANTES: a tool for generating chemokine-based glycosaminoglycan antagonists. *Protein Eng. Des. Sel.* **22**, 367–373

216. Lindahl, U., and Kjellén, L. (2013) Pathophysiology of heparan sulphate: many diseases, few drugs. *J. Intern. Med.* **273**, 555–571

217. Ge, X. N., Ha, S. G., Rao, A., Greenberg, Y. G., Rushdi, M. N., Esko, J. D., Rao, S. P., and Sriramarao, P. (2014) Endothelial and leukocyte heparan sulfates regulate the development of allergen-induced airway remodeling in a mouse model. *Glycobiology*. **24**, 715–727

218. Ganem, B. (1991) Detection of Noncovalent Receptor-Ligand Complexes by Mass Spectrometry. *J. Am. Chem. Soc.*

219. Annis, D. A., Nazef, N., Chuang, C., Scott, M. P., and Nash, H. M. (2004) A General Technique To Rank Protein-Ligand Binding Affinities and Determine Allosteric versus Direct Binding Site Competition in Compound Mixtures. *J. Am. Chem. Soc.* **126**, 15495–15503

220. Zhang, S., Van Pelt, C. K., and Wilson, D. B. (2003) Quantitative determination of noncovalent binding interactions using automated nanoelectrospray mass spectrometry. *Anal. Chem.* **75**, 3010–8

221. Pacholarz, K. J., Garlish, R. a, Taylor, R. J., and Barran, P. E. (2012) Mass spectrometry based tools to investigate protein-ligand interactions for drug discovery. *Chem. Soc. Rev.* **41**, 4335–55

222. Kanu, A. B., Dwivedi, P., Tam, M., Matz, L., and Hill Jr, H. H. (2008) Ion mobility – mass spectrometry. *J. Mass Spectrom.* 10.1002/jms

223. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K., and Heck, A. J. R. (2010) Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **39**, 1633–55

224. Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C., and Jarrold, M. F. (1996) Structural Information from Ion Mobility Measurements: Effect of the Long-Range Potential. **100**, 16082–16086

225. Shvartsburg, A. A., and Jarrold, M. F. (1996) An exact hard-spheres scattering model for the mobilities of polyatomic ions. **261**, 86–91

226. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815

227. Johnson, Z., Proudfoot, A. E. E., Handel, T. M. M., Damme, J. Van, and Mantovani, A. (2005) Interaction of chemokines and glycosaminoglycans: a new twist in the regulation of chemokine function with opportunities for therapeutic intervention. *Cytokine Growth Factor Rev.* **16**, 625–636

228. Patel, M., Yanagishita, M., Roderiquez, G., Bou-Habib, D. C., Oravecz, T., Hascall, V. C., and Norcross, M. A. (1993) Cell-surface heparan sulfate proteoglycan mediates HIV-1 infection of T-cell lines. *AIDS Res. Hum. Retroviruses*. **9**, 167–74

229. Chen, Y., Maguire, T., Hileman, R. E., Fromm, J. R., Esko, J. D., Linhardt, R. J., and Marks, R. M. (1997) Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat. Med.* **3**, 866–71

230. Krusat, T., and Streckert, H. J. (1997) Heparin-dependent attachment of respiratory syncytial virus (RSV) to host cells. *Arch. Virol.* **142**, 1247–54

231. Summerford, C., and Samulski, R. J. (1998) Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J. Virol.* **72**, 1438–45

232. Dechecchi, M. C., Melotti, P., Bonizzato, A., Santacatterina, M., Chilosi, M., and Cabrini, G. (2001) Heparan sulfate glycosaminoglycans are receptors sufficient to mediate the initial binding of adenovirus types 2 and 5. *J. Virol.* **75**, 8772–80

233. Trybala, E., Roth, A., Johansson, M., Liljeqvist, J.-A., Rekabdar, E., Larm, O., and Bergström, T. (2002) Glycosaminoglycan-binding ability is a feature of wild-type strains of herpes simplex virus type 1. *Virology*. **302**, 413–9

234. Matos, P. M., Andreu, D., Santos, N. C., and Gutiérrez-Gallego, R. (2014) Structural requirements of glycosaminoglycans for their interaction with HIV-1 envelope glycoprotein gp120. *Arch. Virol.* **159**, 555–60

235. Shukla, D., and Spear, P. G. (2001) Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry. *J. Clin. Invest.* **108**, 503–10

236. Herold, B. C., Gerber, S. I., Belval, B. J., Siston, A. M., and Shulman, N. (1996) Differences in the susceptibility of herpes simplex virus types 1 and 2 to modified heparin compounds suggest serotype differences in viral entry. *J. Virol.* **70**, 3461–9

237. Trybala, E., Liljeqvist, J. A., Svennerholm, B., and Bergström, T. (2000) Herpes simplex virus types 1 and 2 differ in their interaction with heparan sulfate. *J. Virol.* **74**, 9106–14

238. Rabenstein, D. L. (2002) Heparin and heparan sulfate: structure and function. *Nat. Prod. Rep.* **19**, 312–331

239. Xu, D., and Esko, J. D. (2014) Demystifying heparan sulfate-protein interactions. *Annu. Rev. Biochem.* **83**, 129–157

240. Koide, T. (2008) Antithrombin and Heparin Cofactor II: Structure and Functions. in *Recent Advances in Thrombosis and Hemostasis 2008*, pp. 177–189, Springer

241. Langdown, J., Belzar, K. J., Savory, W. J., Baglin, T. P., and Huntington, J. A. (2009) The critical role of hinge-region expulsion in the induced-fit heparin binding mechanism of antithrombin. *J. Mol. Biol.* **386**, 1278–1289

242. Gettins, P. G. W., and Olson, S. T. (2009) Exosite determinants of serpin specificity. *J. Biol. Chem.* **284**, 20441–20445

243. Whisstock, J. C., Pike, R. N., Jin, L., Skinner, R., Pei, X. Y., Carrell, R. W., and Lesk, A. M. (2000) Conformational changes in serpins: II. The mechanism of activation of antithrombin by heparin. *J. Mol. Biol.* **301**, 1287–1305

244. Jin, L., Abrahams, J. P., Skinner, R., Petitou, M., Pike, R. N., and Carrell, R. W. (1997) The anticoagulant activation of antithrombin by heparin. *Proc. Natl. Acad. Sci.* **94**, 14683–14688

245. Olson, S. T., Richard, B., Izaguirre, G., Schedin-Weiss, S., and Gettins, P. G. W. (2010) Molecular mechanisms of antithrombin–heparin regulation of blood clotting proteinases. A paradigm for understanding proteinase regulation by serpin family protein proteinase inhibitors. *Biochimie*. **92**, 1587–1596

246. Olson, S. T., Swanson, R., Raub-Segall, E., Bedsted, T., Sadri, M., Petitou, M., Hérault, J.-P., Herbert, J.-M., and Björk, I. (2004) Accelerating ability of synthetic oligosaccharides on antithrombin inhibition of proteinases of the clotting and fibrinolytic systems Comparison with heparin and low-molecular-weight heparin. *Thromb Haemost*. **92**, 929–939

247. Lanucara, F., Holman, S. W., Gray, C. J., and Eyers, C. E. (2014) The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **6**, 281–294

248. Giles, K., Pringle, S. D., Worthington, K. R., Little, D., Wildgoose, J. L., and Bateman, R. H. (2004) Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **18**, 2401–14

249. Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J., and Robinson, C. V (2008) Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **3**, 1139–52

250. Seo, Y., Andaya, A., Bleiholder, C., and Leary, J. A. (2013) Differentiation of CC vs CXC chemokine dimers with GAG octasaccharide binding partners: An ion mobility mass spectrometry approach. *J. Am. Chem. Soc.* **135**, 4325–4332

251. Hyung, S.-J., Robinson, C. V, and Ruotolo, B. T. (2009) Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes. *Chem. Biol.* **16**, 382–390

252. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., and Robinson, C. V (2005) Evidence for macromolecular protein rings in the absence of bulk water. *Science (80-.).* **310**, 1658–61

253. Uetrecht, C., Barbu, I. M., Shoemaker, G. K., van Duijn, E., and Heck, A. J. R. (2011) Interrogating viral capsid assembly with ion mobility–mass spectrometry. *Nat. Chem.* **3**, 126–132

254. Chen, S.-H., Chen, L., and Russell, D. H. (2014) Metal-Induced Conformational Changes of Human Metallothionein-2A: A Combined Theoretical and Experimental Study of Metal-Free and Partially Metalated Intermediates. *J. Am. Chem. Soc.* **136**, 9499–9508

255. Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G., Teplow, D. B., Shea, J.-E., Ruotolo, B. T., and Robinson, C. V (2009) Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **1**, 326–331

256. Williams, D. M., and Pukala, T. L. (2013) Novel insights into protein misfolding diseases revealed by ion mobility-mass spectrometry. *Mass Spectrom. Rev.* **32**, 169–187

257. Petitou, M., and van Boeckel, C. A. A. (2004) A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next? *Angew. Chem. Int. Ed. Engl.* **43**, 3118–33

258. Xu, Y., Cai, C., Chandarajoti, K., Hsieh, P.-H., Li, L., Pham, T. Q., Sparkenbaugh, E. M., Sheng, J., Key, N. S., and Pawlinski, R. (2014) Homogeneous low-molecular-weight heparins with reversible anticoagulant activity. *Nat. Chem. Biol.* **10**, 248–250

259. Xiao, Z., Zhao, W., Yang, B., Zhang, Z., Guan, H., and Linhardt, R. J. (2011) Heparinase 1 selectivity for the 3,6-di-O-sulfo-2-deoxy-2-sulfamido-α-D-glucopyranose (1,4) 2-O-sulfo-α-L-idopyranosyluronic acid (GlcNS3S6S-IdoA2S) linkages. *Glycobiology*. **21**, 13–22

260. Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radforda, S. E., and Ashcrofta, A. E. (2009) Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *cell [Equation (1)]*. **12**, 13

261. Salbo, R., Bush, M. F., Naver, H., Campuzano, I., Robinson, C. V, Pettersson, I., Jørgensen, T. J. D., and Haselmann, K. F. (2012) Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers. *Rapid Commun. Mass Spectrom.* **26**, 1181–1193

262. Jurneczko, E., Kalapothakis, J., Campuzano, I. D. G., Morris, M., and Barran, P. E. (2012) Effects of drift gas on collision cross sections of a protein standard in linear drift tube and traveling wave ion mobility mass spectrometry. *Anal. Chem.* **84**, 8524–8531

263. Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V, and Ruotolo, B. T. (2010) Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82**, 9557–9565

264. http://depts.washington.edu/bushlab/ccsdatabase

265. McCoy, A. J., Pei, X. Y., Skinner, R., Abrahams, J.-P., and Carrell, R. W. (2003) Structure of β-Antithrombin and the Effect of Glycosylation on Antithrombin's Heparin Affinity and Activity. *J. Mol. Biol.* **326**, 823–833

266. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926

267. D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, D. M. Y. and P. A. K. (2015) AMBER 2014

268. Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C., and Jarrold, M. F. (1996) Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **100**, 16082–16086

269. Shvartsburg, A. A., Schatz, G. C., and Jarrold, M. F. (1998) Mobilities of carbon cluster ions: critical importance of the molecular attractive potential. *J. Chem. Phys.* **108**, 2416–2423

270. Chowdhury, S. K., Katta, V., and Chait, B. T. (1990) Probing conformational changes in proteins by mass spectrometry. *J. Am. Chem. Soc.* **112**, 9012–9013

271. Fermas, S., Gonnet, F., Varenne, A., Gareil, P., and Daniel, R. (2007) Frontal analysis capillary electrophoresis hyphenated to electrospray ionization mass spectrometry for the characterization of the antithrombin/heparin pentasaccharide complex. *Anal. Chem.* **79**, 4987–4993

272. Abzalimov, R. R., Dubin, P. L., and Kaltashov, I. A. (2007) Glycosaminoglycans as naturally occurring combinatorial libraries: developing a mass spectrometry-based strategy for characterization of anti-thrombin interaction with low molecular weight heparin and heparin oligomers. *Anal. Chem.* **79**, 6055–6063

273. Lindahl, U., and Li, J. (2009) Interactions between heparan sulfate and proteins—design and functional implications. *Int. Rev. Cell Mol. Biol.* **276**, 105–159

274. Huntington, J. A., Olson, S. T., Fan, B., and Gettins, P. G. W. (1996) Mechanism of heparin activation of antithrombin. Evidence for reactive center loop preinsertion with expulsion upon heparin binding. *Biochemistry*. **35**, 8495–8503

275. Johnson, D. J. D., and Huntington, J. A. (2003) Crystal structure of antithrombin in a heparin-bound intermediate state. *Biochemistry*. **42**, 8712–8719

276. Richard, B., Swanson, R., and Olson, S. T. (2009) The signature 3-O-sulfo group of the anticoagulant heparin sequence is critical for heparin binding to antithrombin but is not required for allosteric activation. *J. Biol. Chem.* **284**, 27054–27064

277. Desai, U. R., Petitou, M., Björk, I., and Olson, S. T. (1998) Mechanism of Heparin Activation of Antithrombin ROLE OF INDIVIDUAL RESIDUES OF THE PENTASACCHARIDE ACTIVATING SEQUENCE IN THE RECOGNITION OF NATIVE AND ACTIVATED STATES OF ANTITHROMBIN. *J. Biol. Chem.* **273**, 7478–7487

278. Hopper, J. T. S., and Oldham, N. J. (2009) Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability. *J. Am. Soc. Mass Spectrom.* **20**, 1851–8

279. Han, L., Hyung, S.-J., Mayers, J. J. S., and Ruotolo, B. T. (2011) Bound anions differentially stabilize multiprotein complexes in the absence of bulk solvent. *J. Am. Chem. Soc.* **133**, 11358–11367

280. Shi, H., Pierson, N. A., Valentine, S. J., and Clemmer, D. E. (2012) Conformation types of ubiquitin [M+ 8H] 8+ ions from water: methanol solutions: evidence for the N and A states in aqueous solution. *J. Phys. Chem. B*. **116**, 3344–3352

281. Lima, M. A., Hughes, A. J., Veraldi, N., Rudd, T. R., Hussain, R., Brito, A. S., Chavante, S. F., Tersariol, I. I., Siligardi, G., and Nader, H. B. (2013) Antithrombin stabilisation by sulfated carbohydrates correlates with anticoagulant activity. *Medchemcomm*. **4**, 870–873

282. Gallagher, J. T. (2001) Heparan sulfate: growth control with a restricted sequence menu. *J. Clin. Invest.* **108**, 357

283. Jurneczko, E., and Barran, P. E. (2011) How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst*. **136**, 20–28

284. Scarff, C. A., Thalassinos, K., Hilton, G. R., and Scrivens, J. H. (2008) Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun. Mass Spectrom.* **22**, 3297–3304

285. Hall, Z., Politis, A., Bush, M. F., Smith, L. J., and Robinson, C. V (2012) Charge-state dependent compaction and dissociation of protein complexes: insights from ion mobility and molecular dynamics. *J. Am. Chem. Soc.* **134**, 3429–3438

APPENDIX

INVESTIGATING CHANGES IN THE GAS-PHASE CONFORMATION OF ANTITHROMBIN III UPON BINDING OF ARIXTRA USING TRAVELING WAVE ION MOBILITY MASS SPECTROMETRY (TWIMS) [3]

[3] Zhao, Y., Singh, A., Li, L., Linhardt, R. J., Xu, Y., Liu, J., Woods, R. J., and Amster I. J., (2015) *Analyst.* **140**(20), 6980-6989. Reprinted here with permission of the publisher

Abstract

We validate the utility of ion mobility to measure protein conformational changes induced by the binding of glycosaminoglycan ligands, using the well characterized system of Antithrombin III (ATIII) and Arixtra, a pharmaceutical agent with heparin (Hp) activity. Heparin has been used as therapeutic anticoagulant drug for several decades through its interaction with ATIII, a serine protease inhibitor that plays a central role in the blood coagulation cascade. This interaction induces conformational changes within ATIII that dramatically enhance the ATIII-mediated inhibition rate. Arixtra is a smallest synthetic Hp containing the specific pentasaccharide sequence required to bind with ATIII. Here we report the first travelling wave ion mobility mass spectrometry (TWIMS) investigation of the conformational changes in ATIII induced by its interaction with Arixtra. Native electrospray ionization mass spectrometry allowed the gentle transfer of the native topology of ATIII and ATIII-Arixtra complex. IM measurements of ATIII and ATIII-Arixtra complex showed a single structure, with well-defined collisional cross section (CCS) values. An average 3.6% increase in CCS of ATIII occurred as a result of its interaction with Arixtra, which agrees closely with theoretical estimation of the change in CCS, using protein crystal structures. Comparing the binding behavior of ATIII under both denaturing and non-denaturing conditions revealed the significance of a folded tertiary structure of ATIII for its biological activity. A Hp oligosaccharide whose structure is similar to Arixtra but missing the 3-O-sulfo group on the central glucosamine residue showed a dramatic decrease in binding affinity towards ATIII, but no change in the mobility behavior of the complex, consistent with prior studies that suggested that 3-O-sulfation affects the equilibrium constant for binding to ATIII, but not the mode of interaction. In contrast, nonspecific binding by a Hp tetrasaccharide showed more complex mobility behavior, suggesting more promiscuous interactions with ATIII. The

effect of collisional activation of ATIII and ATIII-Arixtra complex were also assessed, revealing that the binding of Arixtra provided ATIII with additional stability against unfolding. Overall, our results validate the capability of TWIMS to retain the significant features of the solution structure of a protein-carbohydrate complex so that it can be used to study protein conformational changes induced by the binding of glycosaminoglycan ligands.

Introduction

Heparin (Hp) and heparan sulfate (HS) are highly sulfated, linear polysaccharides, consisting of disaccharide repeat units of 1-4 linked hexuronic acid and N-acetyl-glucosamine, and are members of a class of carbohydrates known as glycosaminoglycans (GAGs) (238). The sequence of Hp and HS features three types of domains: highly sulfated (NS) domains, less or non-sulfated N-acetylated (NA) domains and partially sulfated domains (NA/NS). Some of these domains are selectively recognized by over hundreds of secreted and membrane associated human proteins (8, 49, 55, 239). By regulating the location, stability and activity of these interacting proteins, Hp and HS play crucial role in many important physiological and pathological processes (20).

The Hp/HS induced, allosteric activation of Antithrombin (ATIII) is the most studied and best understood example of a specific GAG-protein interaction. The anticoagulant property of Hp was discovered in 1916, and it has been used for prophylaxis and treatment of venous thrombosis, thrombophlebitis and embolism since 1940s (55). Antithrombin III (ATIII), a 58.2 kDa N-glycosylated mono-chain protein in the serpin (serine protease inhibitors) family of proteins, serves as a principal regulator of blood coagulation serine and cysteine proteinases including factor IXa, factor Xa, and thrombin. The inhibitory activity of ATIII is repressed until it is activated by Hp/HS cofactor, either from therapeutic Hp or endothelial cells surface HS

proteoglycans at the site of a vascular injury (240). The native tertiary structure of ATIII is centered at a five-stranded β-sheet A, in which the N-terminal hinge of a reactive center loop (RCL) is initially buried with an orientation unfavorable to react with target proteinases (241). Recent studies also found that the constraint of the RCL also intensifies the repulsive exosite interactions which counteracts the favorable interaction between proteinases and exosite determinants on strand 3 of sheet C surrounding the RCL (242).

The activation of ATIII occurs via its interaction with Hp, and more specifically, with a unique pentasaccharide sequence with a rare 3-O-sulfo group (55). Upon binding of Hp/HS, a local conformational change at the Hp-binding site (the N-terminal region, the N-terminal end of helix A and all of helix D) is triggered. This further induces conformational changes on the proteinase binding site on ATIII (243, 244). As a result of the allosteric activating structure arrangement, the RCL is released and the equilibrium between favorable and repulsive exosite interactions shifts to the favorable side (245). Therefore, the formation and stabilization of ATIII-proteinase complex are promoted. The inhibitory rate of ATIII can be accelerated up to 150~500-fold by the binding of the specific Hp pentasaccharide domain against factors Xa, IXa and VIIa, and this rate can be accelerated up to 2000~200000-fold by full-length Hp (246).

As shown in the case of the interaction between ATIII and Hp/HS, protein-ligand or protein-protein interaction is often characterized by three dimensional conformational change of protein in response to a specific biological function. Conventional solid-phase method (X-ray crystallography) and solution-phase method (NMR spectroscopy) allow elucidation of structural details of protein and protein complex, and the structures of ATIII and ATIII-Arixtra complex have been characterized in this fashion. However, there are issues including difficult sample

preparation, lengthy data analysis and low sample compatibility, which reduce the widespread applicability of these methods to a wide variety of protein-GAG complexes (247).

Ion mobility mass spectrometry (IMMS) is a rapid, sensitive and high-throughput gas-phase technique combing the advantages of both ESI mass spectrometry and ion mobility separation, and has gained much attention and recognition in the field of structural and dynamical biology (223). IMMS separates gas-phase ions according to their mobility, an intrinsic property determined by size, shape and charge state of ions (222). Travelling wave ion mobility spectrometry (TWIMS) is a commonly-used approach for IMMS, and is a commercially available product. In TWIMS, ion mobility separation takes place in an ion guide filled with a neutral gas. A radially confining potential barrier stops ion diffusion away from the path of the ion beam, while a continuous series of low voltage pulses called travelling waves push ions through the device (248). Protein ions with larger collisional cross section (CCS) have lower mobility, undergo more frequent collisions and fall behind the traveling waves, with the result that they are transmitted more slowly than ions with higher mobility. The drift time of a protein ion can be related to its CCS and conformation though calibration with standards. The biological relevance of the measured CCS can be validated by comparison with theoretical CCSs generated from NMR or X-ray crystal structures (249).

Due to the advantage of TWIMS to experimentally estimate the CCS of a gas phase ion in a rapid and sensitive manner, it has been applied by a number of researchers for studies of protein interactions: Leary and coworker have applied TWIMS to detect differences in the conformations of two classes of chemokines as well as the effect of degree of sulfation of Hp-like oligomers on Hp-chemokine interactions (250); Ruotolo and coworkers have applied TWIMS to study the gas-phase conformational stability of wild-type tetrameric transthyretin and

its disease-associated variants with and without ligand binding (251); Robinson and coworkers have applied TWIMS to measure the quaternary structure of the trp RNA binding protein (TRAP) complex and how the addition of tryptophan or RNA enhances the stability of its ring topology (252); Heck and coworkers have applied TWIMS to measure the CCSs of oligomeric viral capsid assembly of Hepatitis B virus (HBV) and norovirus (253); Russell and coworkers have applied TWIMS to study how the binding of metal ions influences the conformation transition of human metallothionein-2A (MT) protein (254); Bowers and coworkers have used TWIMS to investigate the quaternary structure and self-assembling pathways of amyloid-β protein assemblies (255). Though there are concerns regarding the extent to which the structure of a gas-phase ion matches the native structure of a protein in solution, numerous studies have established a good correspondence between the two, particularly for the short time periods of an IMMS measurement, and under carefully controlled experimental conditions (256).

The ATIII-Hp interaction featured by its well-studied conformational change and high binding specificity represents an excellent model to test the applicability of the TWIMS approach to examining conformational changes associated with GAG-protein interaction. In this study, we applied the method of TWIMS to investigate the nature and extent of conformational change within ATIII induced by the binding of Arixtra, a synthetic analogue of the Hp pentasaccharide sequence known to bind this protein with high specificity (257). TWIMS experiments were performed in order to solve these questions: Can the solution structure of ATIII survive the gas-phase environment in TWIMS and still maintain its activity to bind with Hp? Is TWIMS capable of detecting the small difference in CCS for ATIII that is induced by the binding of Arixtra? Do gas-phase ATIII ions exhibit selective binding with Arixtra compared to other structural-similar compounds? Do the complexes of ATIII and Hp oligosaccharides lacking

some of the features of the specific binding motif show similar or different behaviors in their ion mobility? Collectively, the answers to the following questions provide insight into the utility of an IM approach, and more specifically, a TWIMS approach, for studying conformational changes in proteins as a result of their GAG interactions.

Methods

Reagents

All chemicals and solvents (ammonium acetate, methanol, water and formic acid) were of HPLC grade and purchased from Sigma-Aldrich. AT III was purchased from Aniara/Hyphen Biomed as lyophilized powder (West Chester, OH). Stock solution of ATIII was made by dissolving the lyophilized protein into HPLC-grade water and then stored at -80 °C. Arixtra was purchased from the hospital formulary and desalted on a BioGel P2 column BioRad (Hercules, CA, USA) before use. Modified Arixtra was chemoenzymatic synthesized as previously described (258). The Hp tetrasaccharide was produced from naturally occurring source as previously described (259). Protein calibrants (myoglobin from equine heart, cytochrome c from equine heart, avidin from egg white, concanavalin A from *Canavalia ensiformis* and bovine serum albumin) were purchased from Sigma-Aldrich as lyophilized powder.

Sample preparation

For MS analyses under denaturing conditions, ATIII was diluted in a water/methanol/formic acid solution (49.5:49.5:1, v/v/v) to a final concentration of 3 μM. For MS analyses under non-denaturing conditions, ATIII was diluted in 20 mM ammonium acetate buffer, pH 6.8, to a final concentration of 10 μM. ATIII-Hp complex was obtained by incubating ATIII with Arixtra or

other Hp oligosaccharides at a molar ratio of 1:1 at room temperature. Protein calibrants were diluted in either denaturing solution or non-denaturing solution to a final concentration of 10 μM.

IMMS measurement

NanoESI-IMMS experiments were performed using a quadrupole-TWIM-TOF hybrid mass spectrometer (Synapt G2 HDMS, Waters Corp., Manchester, UK) in positive ionization mode. Protein samples were injected into the nanoESI source through a fused-silica emitter (PicoTip New Objective, Woburn, MA) with a flow rate varying from 0.2-0.5 μl/min. Experimental parameters were carefully tuned to prevent the protein and protein complex from unfolding or losing integrity due to extensive activation while keeping substantial ion transmission. The applied experimental parameters were: capillary voltage, 1.5 kV; sampling cone voltage, 30 V; extraction cone voltage, 5 V; source temperature, 90 °C; flow rate of nitrogen in the IM ion guide, 50 ml/min; flow rate of helium in the helium cell, 180 mL/min; Trap collision energy, 0 V; transfer collision energy, 0 V. Different sets of wave height and corresponding wave velocity were used to optimize the mobility separation. The drift times of the calibrants and ATIII samples were measured when identical experimental conditions were stringently applied. Data analysis was performed with MassLynx 4.1(Waters Corp., Manchester, UK).

CCS calibration

The biggest challenge of IMMS using TWIMS is proper calibration. Mobility is a unique property for a given ion, and depends only on the features of a protein ion and neutral gas (mass, charge, size and shape). The measured drift time of an ion is under the influence of many other

factors, including experimental conditions (magnitude and velocity of travelling wave, pressure of neutral gas, temperature of the source and ion guide). Since the electric field in TWIMS is not uniform due to the existence of travelling wave voltages whose electric potential is changing over time and position within the IM ion guide, drift time acquired using TWIMS cannot be converted directly into a CCS. Instead, the CCS of an analyte needs to be determined based on an empirical relationship between the drift times of protein calibrants and their known CCSs obtained previously by conventional drift time IMS (260).

Experimental CCSs were calibrated, as described in several published protocols (249, 260). Briefly, the CCSs of calibrants were corrected for their charge state and reduced mass with respect to the buffer gas. The drift times were corrected for mass-dependent flight time spent in the transfer ion guide and TOF mass analyzer and mass-independent flight time spent in the transfer ion guide. The natural logarithm of corrected CCSs were plotted against the natural logarithm of corrected drift times and a mathematical formula was derived. The calibration coefficient A was extracted to calculate the effective drift times dt": =.(249). A calibration curve was generated by plotting the literature CCSs as a function of dt". The correlation coefficient R^2 of the calibration curve should be higher than 0.98. The experimental CCS of the analyte ion can be derived from this calibration curve based on the measured drift time.

Another intrinsic problem for this experiment is that the calibrant library uses CCSs, which were measured in helium while the TWIMS measurement uses an IM ion guide filled with nitrogen. However, the absolute error of CCS calibration derived from measuring CCS in different gases can be minimized when appropriate separation parameters are applied and when protein calibrants used to construct calibration curves are carefully selected (261, 262).

Table 1 Calibrants

Calibrants and mass	Charge	CCS (He, Å^2)
Avidin	17	3640
64KDa	18	3640
Concanavalin A	20	5550
102KDa	21	5550
	22	5480
	23	5450
Cytochrome c	11	2303
12.4KDa	12	2335
	13	2391
	14	2473
	15	2579
	16	2679
	17	2723
	18	2766
	19	2800
Myoglobin	15	3230
17.6KDa	16	3313
	17	3384
	18	3489
	19	3570
	20	3682
	21	3792
	22	3815
Bovine Serum Albumin	15	4100
66.5KDa	16	4060
	17	4040

We have noted that CCS estimation was substantially improved by using native protein calibrants in the calibration, compared with using only denatured myoglobin (data not shown). Previous studies showed that CCS calibrated using only denatured calibrants were strongly influenced by separation conditions in TWIMS experiment. Therefore, including native calibrants of similar shape and nature as the analytes in the calibrant set greatly increased the accuracy of calibrated CCS of native ATIII and ATIII-Hp complex (263).

A selected set of native and denatured protein calibrants, with a mass range from 12 kDa to 102 kDa and a CCS range from 2303 $Å^2$ to 5550 $Å^2$ were employed. The charge states and literature CCSs of these protein calibrants used to construct the calibration curve are listed in Table 1, selected from the Collision Cross Section Database, Bush Lab (264). The mass range and CCS range chosen were broad enough to bracket the masses and drift times of ATIII and ATIII-Hp complex ions (Figure 1), so no extrapolation of the calibration curve is necessary.

The separation parameters controlling the ion mobility separation were optimized, so that they fit both the analytes and calibrants. Multiple sets of separation parameters were applied to exclude the effect of electric field on the drift time measurement. The calibration curves constructed at wave height of 13, 15 and 17V are shown in Figure 1. R^2 values of 0.9919, 0.9944 and 0.9966 were observed for each trend line.

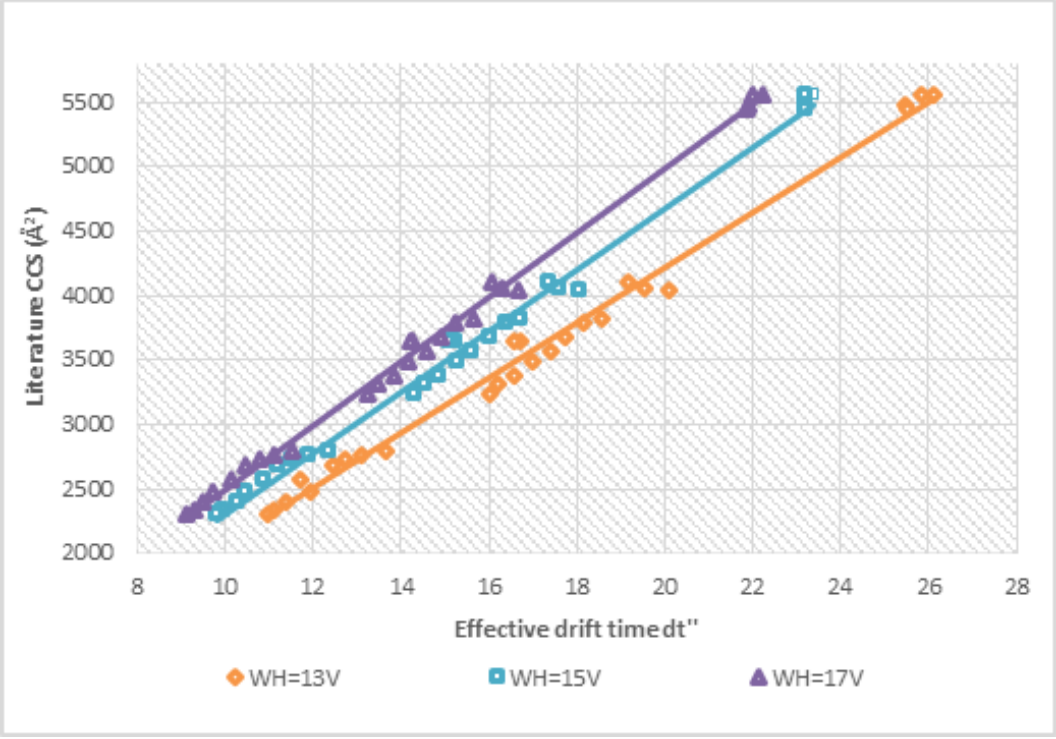

Figure 1 The calibration curves were constructed at three wave heights, combining data from cytochrome c, myoglobin, avidin, BSA and concanavalin A, displayed as literature CCS vs effective drift time for each charge state.

Theoretical calculation of CCSs

The missing residues in the PDB files for the ATIII-Arixtra complex (PDBid: 1E03 (265)) and ATIII (PDBid: 1E05 (265)) were built using MODELLER (226). In the complex, the protein was missing 5 residues from the N-terminus and 8 residues from a disordered portion of the N-terminus. The free protein lacked 2 residues from the N terminus and 12 residues from the disordered region that was missing in the protein in complex with Arixtra. Both forms also lacked one residue at the C-terminus.

Topology and coordinate files for the ATIII-Arixtra complex and free ATIII were generated with the tLeap program, employing the Protein ff99SB (187) and GLYCAM06 (version j) (97) parameters for the protein and GAGs, respectively. The net charge on each system was neutralized with the addition of appropriate number of Na^+ ion. The systems were solvated with TIP3P water (266) in a octahedron box extending to at least 12 Å from any atom of the solute.

All MD simulations were performed with the GPU implementation of pmemd, pmemd.cuda_SPDP (130) in Amber14 (267). Energy minimization of the solvent was performed in an NVT ensemble (1000 steps of steepest descent, 24000 steps of conjugate gradient), followed by a full system energy minimization (1000 steps of steepest descent, 24000 steps of conjugate gradient). The systems were heated from 5 K to 300 K over 60 ps in an NVT ensemble, with a weak positional restraint (10 kcal/mol-Å2) on the atoms in the solute. A Berendsen-type thermostat (128) with a time coupling constant of 1 ps, was utilized for temperature regulation. Equilibration and production was performed at constant pressure (NPT ensemble; 1 atm), with a pressure relaxation time of 1 ps. After the heating step the restraints were removed from the solute atoms, and the entire system was allowed to equilibrate at 300K

for 1 ns. All covalent bonds involving hydrogen atoms were constrained using the SHAKE (129) algorithm, allowing a simulation time step of 2 fs. Scaling factors for 1–4 interactions were set to the recommended values of 1.0 and 1.2 for the GAG (97) and protein (187), respectively, and a non-bonded interaction cutoff of 8.0 Å was employed. Long-range electrostatics were computed with the particle mesh Ewald (PME) method. Data were collected for 20 ns for both the systems. Post processing of the MD simulations was performed using ptraj (188) module of Amber and graphical representations of the results were generated with VMD (190).

Theoretical CCSs of ATIII and ATIII-Arixtra complex were calculated on 30 frames from the last 5 ns of the simulation using MOBCAL (268). Both the projection approximation (PA) (268), and trajectory method (TM) (269) were employed for the calculations.

Results and discussion

Native Mass Spectrometry of unbound ATIII and Arixtra-bound ATIII

ATIII was first analyzed under non-denaturing condition without the addition of any Hp (Figure 2a). The protein was represented by a single charge state envelope, indicating that only protein monomers existed under this experimental condition. Five charge states were observed from +17 to +13 over a range of m/z 3300~4700. Since the charge states detected by using native MS depend on the three dimensional conformation of the protein ion, a narrow distribution of lower charge states indicates a folded and compact conformation with fewer basic sites exposed for protonation (270). The three charge states, +16 to +14, were selected for further investigation since they were the most dominant charge states with relatively higher S/N. An average molecular weight of 57876±6 Da was obtained, in agreement with the literature molecular weight of ATIII (57875 Da) (271).

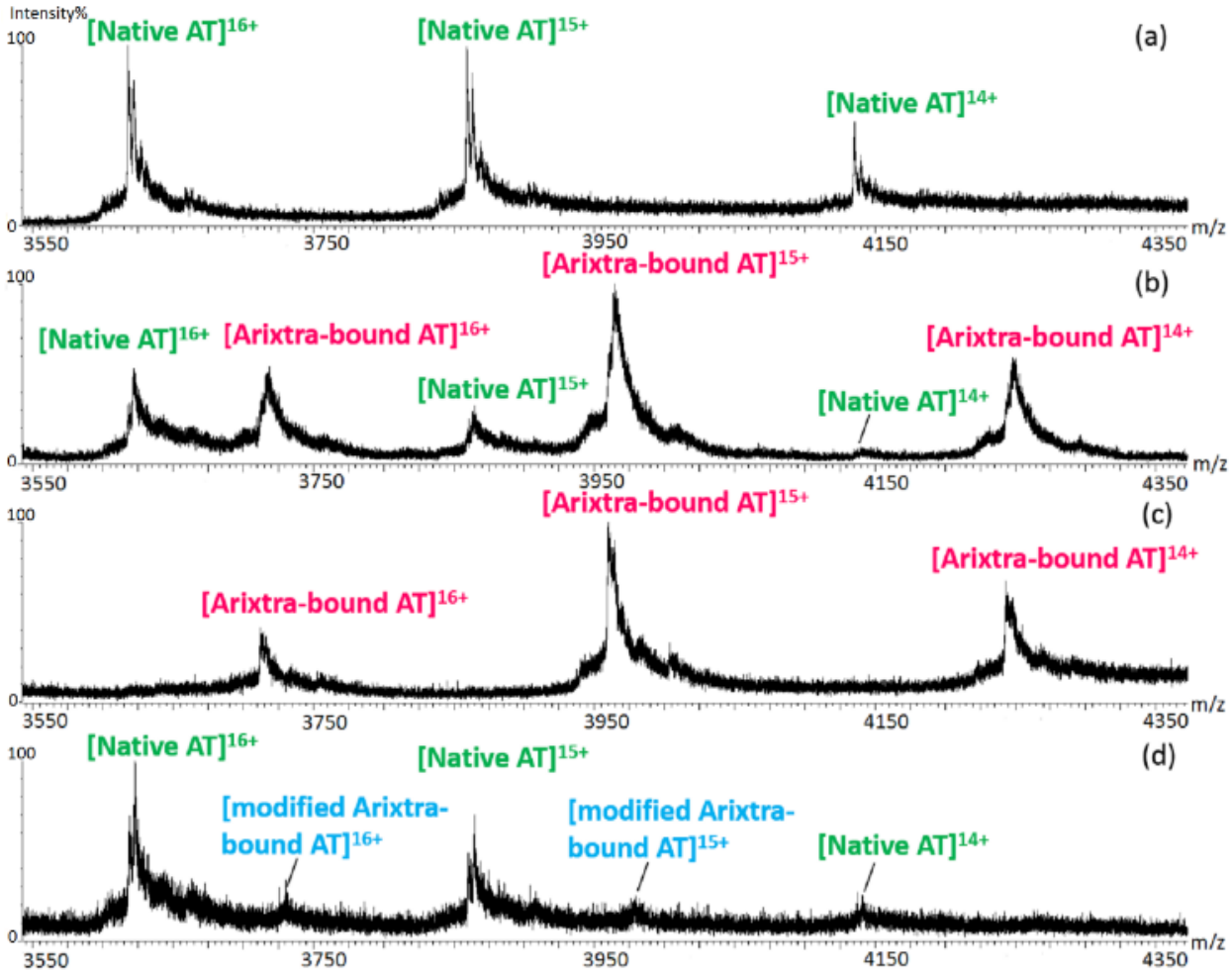

Figure 2 Electrospray ionization mass spectra, obtained under non-denaturing conditions, of (a) ATIII; (b) ATIII incubated with Arixtra for 1h; (c) ATIII incubated with Arixtra for 12h; (b) ATIII incubated with the Arixtra-like hexasaccharide, minus 3-*O*-sulfation, for 12h.

ATIII was next incubated with a small molar excess of Arixtra. Two incubating times were examined to monitor the progress of the binding reaction: 1h (Figure 2b) and 12h (Figure 2c). The binding of Arixtra by ATIII did not change the charge state distribution observed for native protein. An additional peak was observed for each charge state, corresponding to the formation of Arixtra-ATIII complex. An average mass increase of 1512 Da was measured for the

new peaks, in good correspondence to the molecular weight of a single Arixtra molecule (1505 Da). To confirm the binding stoichiometry, titration experiments with different molar ratios were examined, but increasing the amount of Arixtra did not result in the formation of any new peaks. After 12h of incubation, only the peaks corresponding to Arixtra bound to ATIII in a 1:1 ratio were present (Figure 2c). Since longer incubation time allowed the binding reaction to go to completion, this result indicates that the ATIII only presents one binding site for Arixtra, in accordance with the 1:1 binding stoichiometry between ATIII and Arixtra reported previously (272). All of these observations suggest that the solution structure of ATIII survives the ionization process and the environment of the TWIMS instrument.

Interestingly, the intensity ratio between ATIII with and without Arixtra varied with charge state. The ratio for +17, +16, +15 and +14 were 0.69, 1.02, 3.07 and 5.75, respectively. It is evident that the intensity ratio increases with the decrease of charge state, suggesting weakening in the binding affinity with the increase of charge state. One possible explanation is that the lower charge states are more representative of the native structure of ATIII due to a lowering of charge-charge repulsion.

TWIMS of ATIII and its complex with Arixtra

To investigate whether the protein conformational change caused by the interaction between ATIII and Arixtra can be detected by TWIMS, ion mobility measurements were performed on the gas-phase ATIII and Arixtra-bound ATIII ions. The measured drift times of each charge state of ATIII and Arixtra-bound ATIII are shown in Figure 3, at a wave height of 17 V. One narrow drift time distribution was observed for each charge state of ATIII, indicating the presence of a single compact and folded conformation for the ATIII ions. For Arixtra-ATIII complex, a single

narrow peak at higher drift time range was observed for each charge state, implying that binding of Arixtra causes a change in the folded conformation of ATIII, leading to a specific structure with a larger CCS. These observations suggest that the tertiary structures of ATIII and the non-covalent Arixtra-ATIII complex are stable in the TWIMS experiment.

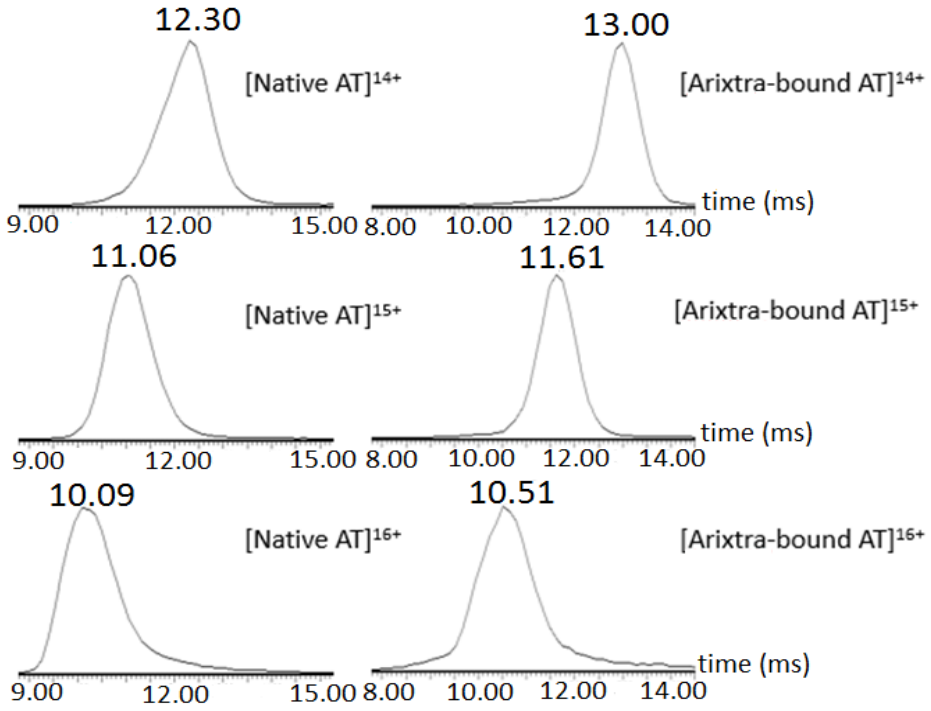

Figure 3 Drift timed distributions (ms) of the dominant charge states (+14 ~ +16) of native ATIII and Arixtra-bound ATIII, at a wave height of 17V.

After calibration, the measured drift times for ATIII and its complex with Arixtra have been converted to CCSs, shown in Table 2. For each charge state, the reported CCS was an average from triplicate experiments, and was found to have a 0.6% relative standard deviation. The binding of Arixtra induces a 3.6% increase in CCS of ATIII, several times larger than the standard deviation in the measurements. The conformational change was highly reproducible, and was observed for each charge state, as well as for a variety of wave heights and wave

velocities. The drift times and peak shape measured for residual bound ATIII, which is present at moderate abundance in the mass spectrum of the 1hour mixture, Figure 2b, were the same as the drift times measurements for the ATIII peaks in the mass spectrum of the pure protein sample, Figure 2a, evidence of the stability of the drift time measurements over different times and different samples.

Table 2 CCSs measured experimentally

Charge State	Free ATIII CCS ($Å^2$)	Arixtra-bound ATIII CCS ($Å^2$)	CCS increase ($Å^2$)	CCS increase%
14	3467.9±14.5	3607.1±16.6	123.5±6.6	3.4
15	3439.8±8.1	3573.1±7.0	133.2±2.3	3.7
16	3462.3±17.6	3591.4±15.9	133.0±8.4	3.7

Specificity and selectivity of ATIII-Arixtra interaction

To test the specificity of the interaction between ATIII and Hp, two control experiments were applied by altering the structure and conformation of both ATIII and its binding partner.

The first control experiment was to test the binding capability of ATIII after being denatured due to the presence of organic solvents under harsh pH condition (49.5 water: 49.5 methanol: 1 formic acid). Denatured ATIII was incubated with Arixtra at a molar ratio of 1:3 for 12h and sprayed in denaturing solution (Figure 4). A wide distribution of higher charge states from + 47 to +23 were observed, in contrast to the narrow distribution of lower charge states under non-denaturing conditions. No peaks corresponding to the formation of ATIII-Arixtra complex were observed for these higher charge states, indicating that the fully denatured ATIII loses its capability to bind with Arixtra. This control eliminates the concern that the complexation of ATIII and Arixtra is non-specific, and results from the columbic attraction of a

positively charged protein with an anionic carbohydrate. A mass scale expansion of the higher mass range of the denatured ATIII saw Arixtra-bound ATIII peaks for charge states lower than +20 (Figure 4, inset). It is possible that these relatively low charge states may represent some partially folded structures of ATIII.

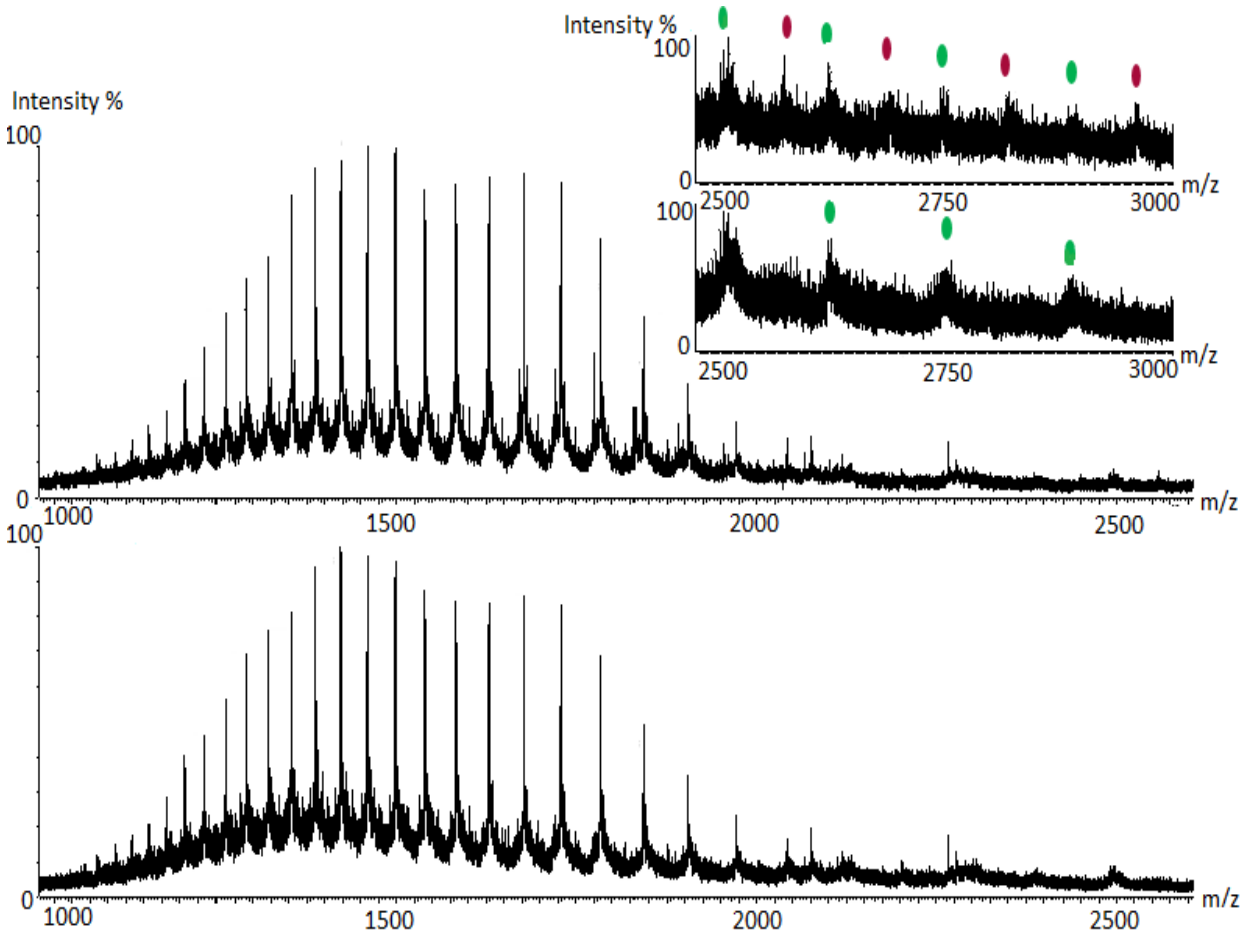

Figure 4 Electrospray ionization mass spectra of denatured ATIII with (upper) and without (lower) addition of Arixtra; The inset shows an expansion of the mass range (2500 m/z-3000 m/z) of the two spectra. Unbound ATIII peaks were labeled with green and Arixtra-bound peaks were labeled with purple.

A broad drift time distribution with multiple features was observed for each charge state of denatured ATIII (data not shown) while a single sharp peak was observed for native ATIII.

This observation indicates that the fully extended or partially unfolded structures of ATIII have higher flexibility than the folded structures, and adopt multiple conformations in the gas phase.

Previous studies have collected evidence that Hp binding sites on proteins are more than clustered basic amino acid residues. Instead, several structural elements (loops, α-helices and β-strands) have been found in Hp binding sites, implying that specific spacing and spatial patterns of the Hp-binding basic residues are essential in Hp-protein interactions (49, 239, 273). As in the case of ATIII-Hp binding, the folded, compact conformation of ATIII allows a spatial alignment of the Hp-binding basic residues which were widely distributed in the sequence (244, 274). This alignment facilitates the formation of a basic patch over several helices and the N-terminal region of ATIII, which provides a binding site for Hp. In this control experiment, denaturing the protein led to a substantial alternation of the topology of the Hp binding sites (8), therefore disrupting the high affinity binding interaction between ATIII and Hp. Our observation revealed the importance of a folded, compact structure of ATIII to maintain its biological activity.

A second control experiment was performed to test the binding specificity between ATIII and Arixtra induced by the pattern of sulfation. Arixtra has eight sites of sulfation, with a sequence of: $GlcNSO_3$, $6SO_3$-GlcA-$GlcNSO_3$, $3SO_3$, $6SO_3$*-IdoA2SO_3-$GlcNSO_3$, $6SO_3$, as shown in Figure 5. The ATIII binding behavior of a Hp hexasaccharide and a Hp tetrasaccharide, with the structures shown in Figure 5, were examined. The sulfation pattern of the Hp hexasaccharide closely resembles that of Arixtra, minus the 3-O-sulfo group on the central glucosamine. The Hp tetrasaccharide lacks two sulfate groups and a monosaccharide on the non-reducing end, compared to the structure of Arixtra.

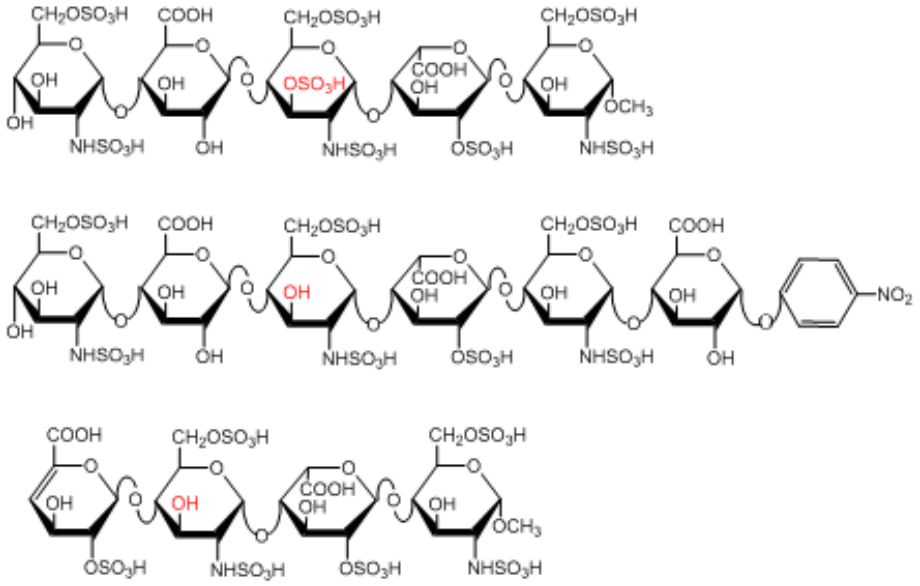

Figure 5 Structures of Arixtra (dp5 with 8 SO_3) (upper), modified Arixtra (dp6 with 7 SO_3) (middle) and the Hp tetrasaccharide (dp4 with 6 SO_3) (below).

After incubating ATIII and the hexasaccharide, peaks corresponding to the hexasaccharide-bound ATIII complex, Figure 2d, were observed, but in much lower abundance compared to Arixtra, for the same incubation time and Hp concentration. Increasing the Hp concentration produced only a small increase in the intensity of the hexasaccharide-bound ATIII complex peaks. This reaction never went to completion, as can be inferred from the presence of unbound ATIII ions which dominated the mass spectra, even after longer incubation time. All of these observations indicate that the binding affinity of the hexasaccharide towards ATIII is much lower than Arixtra, consistent with previous studies (275). Similar results were observed for the binding between ATIII and the tetrasaccharide, with the peaks corresponding to ATIII-tetrasaccharide complex in even lower abundance (results not shown). Our results suggest that the hexasaccharide and tetrasaccharide both have reduced binding affinities for ATIII. The data confirms the significant contribution of the rare 3-O-sulfation in Hp-protein interaction.

The drift time distributions of ATIII-negative control complexes were measured and only the results for +15 charge state were shown in Figure 6. For the complex of ATIII and hexasaccharide, a narrow drift time distribution was observed (Figure 6b), almost identical to that measured for ATIII-Arixtra complex (Figure 6c). The drift time measured for hexasaccharide-bound ATIII was the same as that measured for Arixtra-bound ATIII, which agrees with a previous study that concluded that removing the 3-O-sulfo group affected the equilibrium of native ATIII and activated ATIII but not the conformational change associated with the equilibrium (276). In other words, this control Hp oligosaccharide retains the specificity of binding, but has a reduced binding affinity.

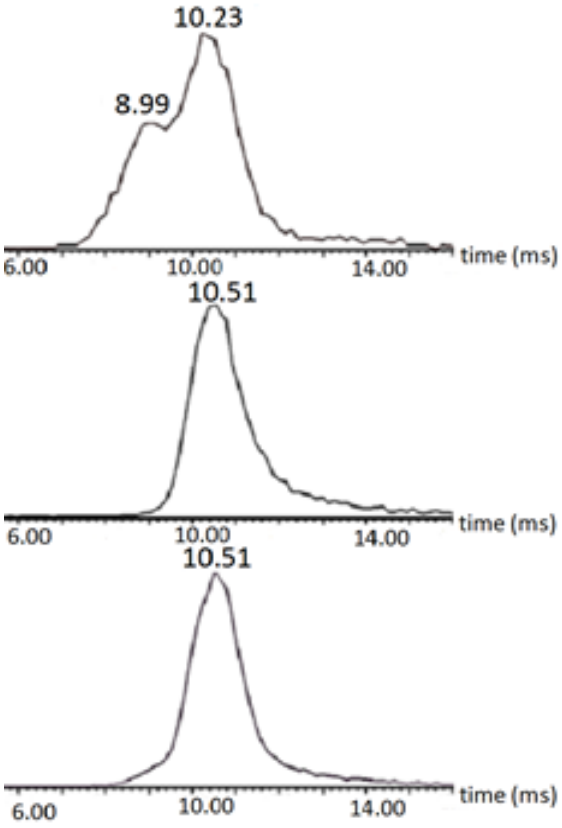

Figure 6 Drift time distributions (ms) of the +15 charge state of ATIII in complex with compounds, shown in Figure 5, the Hp tetrasaccharide (a), modified Arixtra (b) and Arixtra (c), at a wave height of 17V.

In strong contrast, the Hp tetrasaccharide-ATIII complex revealed a very different behavior, exhibiting a broad peak with many features in the drift time spectrum, Figure 6a, suggesting several structures for the complex. We interpret these as reduced specificity in the binding of the Hp ligand. The drift times measured for Hp tetrasaccharide-bound ATIII were smaller than that measured for Arixtra-bound ATIII, suggesting that the binding of the Hp tetrasaccharide may induce less of a conformational change within ATIII, due to the loss of several essential components from the specific pentasaccharide sequence (277).

The results of these control experiments are consistent with the known high-affinity, specific interaction between ATIII and Hp. It requires more than the existence of negatively charged functional groups on the Hp and positively charged residues on the protein. Instead, a well-defined three-dimensional presentation of the negatively charged groups in the Hp and a well-defined tertiary structure of the protein are critical to the specificity of ATIII-Arixtra binding.

Gas phase stability of ATIII and ATIII in complex with Arixtra

Previous studies have shown that the addition of a ligand or counter ions brings in additional conformational stability of protein ions against collisional activation (278, 279). To test whether the specific binding of Arixtra has altered the conformational stability of ATIII, we performed a series of experiments with incremental changes in trapping collisional energy (CE) from 5V to 30V being applied to ATIII with and without Arixtra in the trap ion guide prior to ion mobility separation. This probes the susceptibility to collision-induced unfolding for native ATIII and the ATIII-Arixtra complex.

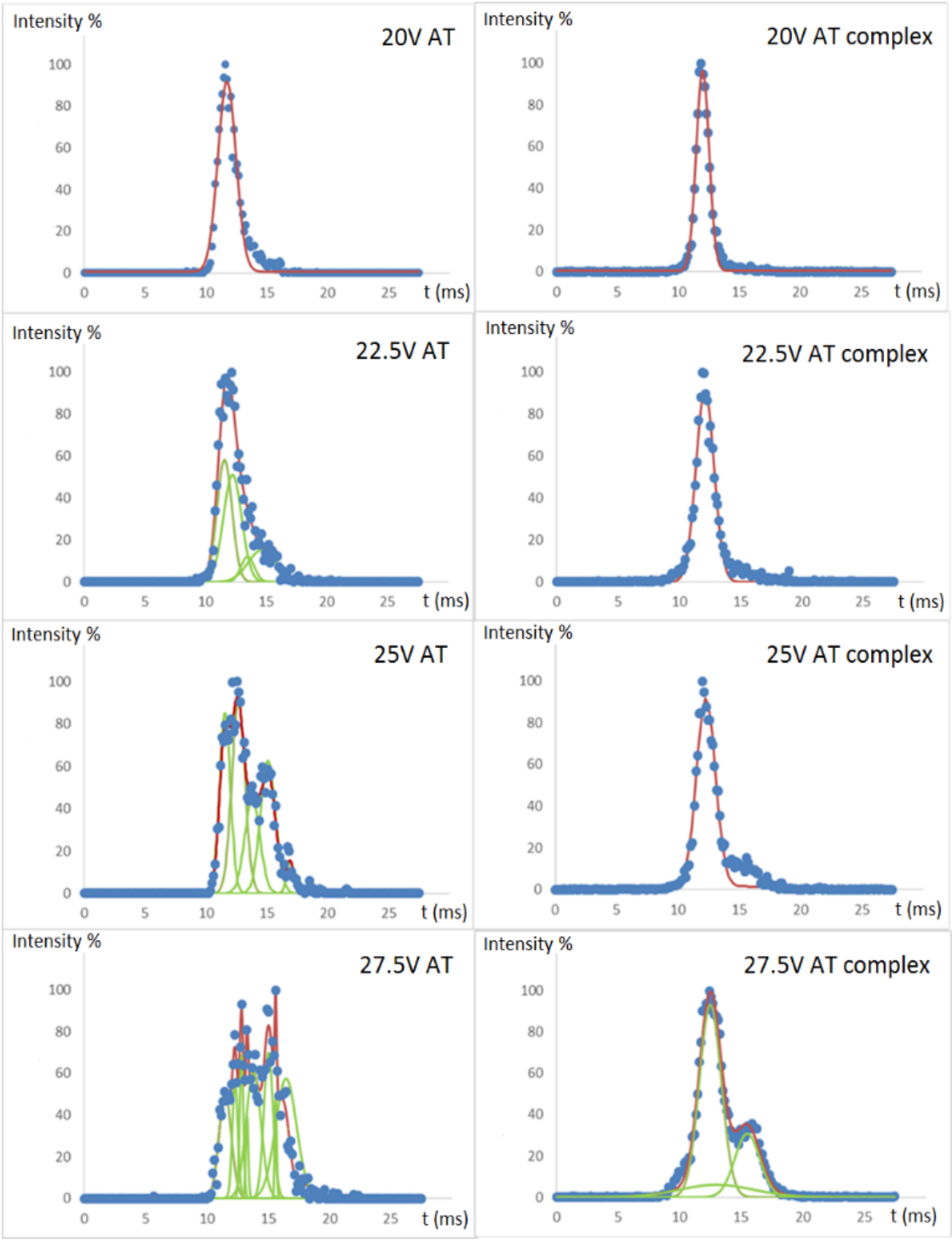

Figure 7 Experimental drift time distributions (ms) of +15 charge state of ATIII (left panels) and ATIII-Arixtra complex (right panels) as a function of trapping CE varying from 20V to 27.5V are shown in 2.5 V increments. The normalized experimental drift time distribution is shown by the blue dots. The fit of the data to a series of Gaussian distributions is shown in green, and the sum of the component Gaussians is shown in red.

The resulting drift time distributions of +15 charge state of both ATIII and ATIII-Arixtra complex are shown in Figure 7. Each drift time distribution was fitted with a minimum numbers of Gaussian distributions according to its shape using the peak analyzer in OriginPro 8.5.0 software (OriginLab Corporation, MA), as described previously (280). Each Gaussian distribution generated as part of the fitness represents a potentially stabilized gas-phase conformation.

At a trapping CE of less than 20V (data not shown) or equal to 20V, a single, narrow peak was observed for both ATIII and ATIII-Arixtra complex, suggesting a single conformation of the protein and its Hp-bound complex.

The behavior of ATIII and ATIII-Arixtra complex started to deviate at elevated activation energy. For ATIII, a shoulder appeared on the high-drift-time side of the original peak when the trapping CE was raised to 22.5V. Four Gaussian distributions fit this drift time distribution, indicating that several partially unfolded or unfolded conformations of gas-phase ATIII ions were present at this collision energy. At a CE of 25 V, the abundance of the shoulder peak clearly increased. Two distinct populations of higher mobility (from 10.37 to 14.10 ms) and lower mobility (from 14.24 to 18.11 ms) conformations were observed, each modeled by two or three Gaussian distributions. At a CE of 27.5 V, a broad drift time distribution resolved by eight Gaussian distributions was observed, indicating the presence of a series of protein conformations of different degree of unfolding.

In contrast, for a range of trapping CE from 0 to 25V, the drift time distributions of ATIII-Arixtra complex remained unaffected, evidenced by a single, narrow peak in the drift time spectrum. A lower mobility peak corresponding to partially unfolded conformation of ATIII complex was not formed until a trapping CE as high as 27.5V was applied, while the higher

mobility peak corresponding to the folded structure was still the dominant peak. Clearly, ATIII, when complexed with Arixtra, is more stable than ATIII itself, also consistent with solution-phase behavior of this system (281).

Previous studies showed that the binding of Hp/HS improves conformational stability of some proteins against heat and proteolysis (282). The additional stability is also related to the activity of proteins.[58] Our results showed that ATIII in complex with Arixtra evidently possesses higher conformational stability than ATIII itself, indicating that the stability of Hp-activated ATIII may be a driving factor for its biological activity, which is responsible for shifting the conformational equilibrium towards greater stability.

Comparison of experimental CCS with theoretical CCS

Theoretical CCSs estimated from X-ray crystal structures of ATIII and Arixtra-bound ATIII were compared with experimentally measured CCSs. The PDB files of ATIII and ATIII-Arixtra complex display two ATIII molecules, one in the inhibitory form and the other in the latent form (265). Only the inhibitory structure was used for the CCS calculation. In addition, N-linked oligosaccharide chains are present in the structures in both PDF files. We chose to leave the oligosaccharides out of the CCS calculation. The N-linked oligosaccharides are not thought to participate in Hp binding, so they are expected to contribute equally to the CCS of ATIII and ATIII-Arixtra complex. As we are interested in the conformation change within the protein rather than the absolute measurement of CCSs, the glycosylation-free estimation of CCS should allow us a direct and simple observation of the conformational change. We added the missing H-atoms to the crystal structures since their effect may not be ignored in the accurate estimation of the subtle conformational change.

Table 3 CCSs calculated using MOBCAL

System	PA method (Å2)	TM method (Å2)
AT III – Arixtra complex	3120.2±3.2	4033.3±14.5
AT III	3010.1±2.7	3900.5±9.7

The calculated CCSs of native ATIII were 3010.1 $Å^2$ estimated using PA and 3900.5 $Å^2$ estimated using TM. The calculated CCSs of Arixtra-bound ATIII were 3120.2 $Å^2$ estimated using PA and 4033.3 $Å^2$ estimated using TM (Table 3). The experimental CCSs for the lowest charge state of native ATIII and ATIII-Arixtra complex were 3467.9 $Å^2$ and 3607.1 $Å^2$, respectively, comparable with those theoretical measurements.

The experimental CCS is approximately 12.1% less than the theoretical CCSs estimated using TM and 15.4 % higher than that estimated using PA. These observations are consistent with previous studies on other proteins, which found that reasonable experimental CCSs should lie in between the theoretical estimates made using PA and TM. The experimental CCSs should be smaller than TM results, due to the collapsed structure of protein ions in the gas-phase in response to desolvation (247, 283, 284). Ignoring the long-range interaction with the neutral gas, collision effect and scattering process, PA results are on average 15% smaller than the experimental CCSs (285), in agreement with our observation.

It is noteworthy that there is close agreement between the relative changes in the CCS measured experimentally versus calculated theoretically. The averaged experimental conformational change was 3.6%, while the calculated conformational change was 3.7% for PA results and 3.4% for TM results, respectively. These results show that the TWIMS experiment generates gas-phase ATIII and Arixtra-bound ATIII ions with a conformation that matches the solution structure close enough, so that the conformational change due to the binding of Arixtra can be accurately measured.

Conclusions

TWIMS was found to provide data that is consistent with the known details of ATIII-Arixtra binding. Our mass spectra showed the formation of a 1:1 complex of ATIII with Arixtra, and the corresponding IM spectra were consistent with a single, folded gas-phase conformation of ATIII in its free and Arixtra-bound form, evidenced by a single narrow drift time distribution in both cases. CCSs derived from these data showed that the binding of Arixtra to ATIII caused a 3.6 % increase of ATIII's CCS. Both the absolute CCSs of ATIII and Arixtra-bound ATIII ions as well as the degree of CCS change were in agreement with theoretical CCSs calculated based on their X-ray crystal structures.

Furthermore, the selectivity and specificity of Hp-ATIII binding known from solution measurements appeared to survive the translation of the ions into the gas-phase, as revealed by the control experiments. Disrupting the folded structure of ATIII caused the loss of its binding affinity towards Arixtra. Removing the 3-O-sulfo group from the known Hp binding sequence resulted in substantial reduction of the abundance of ATIII-Hp complex but did not change the drift time behavior, as expected based on known solution behavior. These two control experiments not only confirmed the biological relevance of the native ATIII and Arixtra-bound ATIII ions that we observed, but also provided evidence to support the existence and significance of the specificity of the ATIII-Hp interaction. A control experiment using a Hp tetrasaccharide with a different pattern of sulfation than the known consensus sequence also showed greatly reduced affinity for ATIII, but also showed a reduction in specificity of binding, evidenced by a broad drift time distribution, consistent with a variety of structures for the complex. Moreover, the TWIMS measurement showed the stabilizing effect of Arixtra binding

on the ATIII folded structure. This result also matched solution studies of ATIII and its Hp complexes.

Collectively, these results highlight the advantages of TWIMS for investigating GAG–protein interactions at the molecular level. Future applications of this approach will be useful for gaining a better understanding of the biological processes mediated by GAG-protein interactions.

Acknowledgements

The work was supported by National Institutions of Health grants R01-GM038060 and P41-GM103390.

YZ gratefully acknowledges training on the use of the Synapt from Dr. Joshua S. Sharp (CCRC, UGA).

SUPPLEMENTARY INFORMATION CHAPTER 4

Equation S4.1

$$^{3}J_{HH} = 13.24\cos^2\theta - 0.91\cos\theta + \sum \Delta\chi_i(0.53 - 2.41\cos^2(\xi_i\theta + 0.27|\Delta\chi_i|))$$

Table S4.1 Bond & Angle terms

Bond	R_o (Å)	k_r (kcal/mol)
C –Ck	1.466	214.0
S -N	1.675	238.0
S -Os[a]	1.589	206.0
S -O2[a]	1.440	620.0
Cg-Oa[b]	1.460	285.0
Cg-Oe[b]	1.460	285.0

Angle	θ_o (°)	k_Θ (kcal/mol)
C -Ck-Ck	126.0	48.0
O2-C-Ck	113.4	49.5
O2-S -N	108.0	84.0
O2-S -O2	113.9	123.0
O2-S -Os	106.9	104.0
OS-Ck-C	109.5	73.0
H1-Cg-N3[c]	109.5	50.0
H -N -S[d]	121.2	37.5
S -N -Cg	110.0	31.0
S -Os-Cg	118.9	50.0
C -Cg-Oa[b]	112.4	63.0
Cg-Cg-Oa[b]	108.5	70.0
Cg-Cg-Oe[b]	108.5	70.0
Cg-Oa-Cg[b]	111.6	50.0
Cg-Oe-Cg[b]	111.6	50.0
Oh-Cg-Oa[b]	112.0	110.0
Os-Cg-Oa[b]	112.0	100.0
Os-Cg-Oe[b]	112.0	100.0

Oa-Cg-Oe[b]	112.0	100.0
H1-Cg-Oa[b]	110.0	60.0
H1-Cg-Oe[b]	110.0	60.0
H2-Cg-Oa[b]	110.0	60.0
H2-Cg-Oe[b]	110.0	60.0

[a]Equilibrium value from Theochem 395/396 (1997) pp 107-122
[b]Copied from GLYCAM06 (97)
[c]Copied from parm99 HP-CT-N3 (96)
[d]From HF/6-31++g(2d,2p) ethylsulfamate

Table S4.2 Torsion terms

Torsion	V_1	V_2	V_3	Source
Cg-Os-Ck-C		-1.40	0.67	2-Methoxyacrylate
Cg-Ck-Ck-C		-11.00		But-2-enoate
Oh-Cg-Cg-Ck	-1.06	-0.60	-0.11	3-Butenol*/Molecule **4**
Os-Cg-Os-Ck	0.75	0.50		2-(Methoxymethoxy) acrylate
Os-Cg-Cg-Ck	-0.27	0.40		4-Methoxybut-1-ene*/Molecule **3**
O2-C -Ck-Ck		-0.80		(2R,3S,4S)-3,4-dihydroxy-2-methoxy-3,4-dihydro-2H-pyran-6-carboxylate**
O2-C -Ck-Os		-0.80		(2R,3S,4S)-3,4-dihydroxy-2-methoxy-3,4-dihydro-2H-pyran-6-carboxylate**
O2-S -N -Cg			0.11	Methylsulfamate
Ha-Ck-Ck-C		-11.00		Acrylate
H1-Cg-N -S	2.00			((2S,3R)-2-methoxytetrahydro-2H-pyran-3-yl)sulfamate***
H1-Cg-N3-H			0.16	X-CT-N3-X[a]
H2-Cg-Os-Ck	1.00	0.50	0.10	2-(1-Methoxyethoxy) acrylate/2-(Methoxymethoxy) acrylate
H2-Cg-Cg-N3			0.10	H2-CT-CT-N[a]
S -N -Cg-Cg	0.50	0.50		((2S,3R)-2-methoxytetrahydro-2H-pyran-3-yl)sulfamate***
Oa-Cg-Cg-Cg	0.19	-0.11	0.14	1,1-dimethoxypropane[b]
Oe-Cg-Cg-Cg	-0.27			Os-Cg-Cg-Cg[c]
Oa-Cg-Cg-Oe		0.40		Os-Cg-Cg-Os[d]
Oe-Cg-Cg-Oe		0.40		Os-Cg-Cg-Os[d]
Oa-Cg-Cg-Os		0.40		Os-Cg-Cg-Os[d]
Oe-Cg-Cg-Os		0.40		Os-Cg-Cg-Os[d]
Oa-Cg-Cg-Oh	-1.10	0.25		Os-Cg-Cg-Oh[c]
Oe-Cg-Cg-Oh	-1.10	0.25		Os-Cg-Cg-Oh[c]
Oa-Cg-Oe-Cg	1.08	1.38	0.96	Os-Cg-Os-Cg[e]
Oa-Cg-Os-Cg	1.08	1.38	0.96	Os-Cg-Os-Cg[e]

Os-Cg-Oa-Cg	1.08	1.38	0.96	Os-Cg-Os-Cg[e]
Oe-Cg-Oa-Cg	-0.25	0.76	1.20	Dimethoxymethane[b]
O2-C -Cg-Oa	0.04	-1.45	0.04	O2-C -Cg-Os[e]
C -Cg-Oa-Cg	-0.60	0.45	0.32	C -Cg-Os-Cg[e]
Cg-Oa-Cg-Cg	-0.70	-0.30	-0.33	2,2-dimethoxypropane[b]
Cg-Oe-Cg-Cg			0.16	Cg-Os-Cg-Cg[c]
N -Cg-Cg-Oa	-1.30			N -Cg-Cg-Os[c]
N -Cg-Cg-Oe	-1.30			N -Cg-Cg-Os[c]
H1-Cg-Cg-Oa			0.05	N -Cg-Cg-Os[c]
H1-Cg-Cg-Oe			0.05	H1-Cg-Cg-Os[c]
H1-Cg-Oa-Cg			0.27	H1-Cg-Os-Cg[c]
H1-Cg-Oe-Cg			0.27	H1-Cg-Os-Cg[c]
H2-Cg-Cg-Oe			0.05	H2-Cg-Cg-Os[c]
H2-Cg-Oa-Cg	-1.20	0.10	0.02	1,1-dimethoxyethane[b]
H2-Cg-Oe-Cg		0.60	0.10	H2-Cg-Os-Cg[c]

Improper Torsion	V_2
Ck-Ha-Ck-Cg	1.0
Os-C -Ck-Ck	1.0
Cg-S -N -H	5.5

*Used for an initial fit
**Fitting equally split between O2-C-Ck-Ck & O2-C-Ck-Os
***Fitting to both 2S and 2R epimers
[a]PARM99 (96)
[b]Published in (138)
[c]Copied from GLYCAM06 (97)
[d]Copied from GLYCAM06g
[e]Copied from GLYCAM06e

Table S4.3 $\Delta\chi i$ Substituents by torsion for Δ4,5-unsaturated uronates for Equation S4.1

Substituent*	ξ_i	H1-C1-C2-H2	H2-C2-C3-H3	H3-C3-C4-H4
S_1	1	1.25 O2 (-OH, -OSO_3^-)	1.25 O3 (-OH)	1.25 O3 (-OH, -OSO_3^-)
S_2	-1	0.35 C3 (-C)	0.35 C4 (-C)	0.35 C2 (-C)
S_3	1	1.26 O1 (-OR)	0.35 C1 (-C)	0.35 C4 (-C)
S_4	-1	1.26 O5 (-OR)	1.25 O2 (-OH)	0.35 C4 (-C)

*Substituent numbers defined in Haasnoot et al. (125)

Table S4.4 Ensemble averaged charges/prep files for new residues.

Transferable N- & O- Sulfate

```
SO3     INT  0
CORRECT     OMIT DU   BEG
-0.8370
 1 DUMM DU   M   0 -1 -2  0.000    0.000     0.000    0.00000
 2 DUMM DU   M   1  0 -1  1.449    0.000     0.000    0.00000
 3 DUMM DU   M   2  1  0  1.522  111.100     0.000    0.00000
 4 O1   O2   M   3  2  1  1.540  111.208   180.000   -0.69400
 5 S1   S    M   4  3  2  1.432   92.882  -123.335    1.24500
 6 O3   O2   E   5  4  3  1.425  115.334   -51.093   -0.69400
 7 O2   O2   M   5  4  3  1.432  113.544    85.155   -0.69400


DONE
```

Δ4,5-unsaturated uronate

```
045  INT 0
CORRECT OMIT DU BEG
-0.8060
 1 DUMM DU   M   0 -1 -2  0.000    0.000     0.00     0.0000
 2 DUMM DU   M   1  0 -1  0.500    0.000     0.00     0.0000
 3 DUMM DU   M   2  1  0  1.296   74.264     0.00     0.0000
 4 C1   Cg   M   3  2  1  4.744  133.011   147.97     0.2500
 5 H1   H2   E   4  3  2  1.085   27.079   135.14     0.0000
 6 O5   Os   S   4  3  2  1.381  117.101  -148.90    -0.2540
 7 C5   Ck   B   6  4  3  1.374  116.893   142.43    -0.0040
 8 C6   C    B   7  6  4  1.552  114.268   153.78     0.9370
 9 O6B  O2   E   8  7  6  1.226  114.187   -23.82    -0.8300
10 O6A  O2   E   8  7  6  1.227  114.136   158.53    -0.8300
11 C4   Ck   B   7  6  4  1.262  123.414   -22.88    -0.1140
12 H4   Ha   E  11  7  6  1.072  117.772  -179.54     0.0000
13 C3   Cg   3  11  7  6  1.506  122.995     1.71     0.3200
14 H3   H1   E  13 11  7  1.084  110.770  -130.68     0.0000
15 O3   Oh   S  13 11  7  1.411  112.274   112.08    -0.7190
16 H3O  Ho   E  15 13 11  0.950  106.837   -56.81     0.4030
17 C2   Cg   B  13 11  7  1.532  109.784   -12.10     0.3570
18 H2   H1   E  17 13 11  1.081  109.778   161.03     0.0000
19 O2   Oh   S  17 13 11  1.405  111.472   -80.76    -0.7130
20 H2O  Ho   E  19 17 13  0.950  108.097    63.37     0.3910

LOOP
C2 C1

DONE
```

α-D-Glucosamine

```
0YNP  INT  0
CORRECT     OMIT DU   BEG
 1.1940
 1 DUMM DU   M   0 -1 -2  0.000     0.0       0.0     0.0000
 2 DUMM DU   M   1  0 -1  1.522     0.0       0.0     0.0000
 3 DUMM DU   M   2  1  0  1.430   109.5       0.0     0.0000
 4 C1   Cg   M   3  2  1  1.390   109.2    -178.5     0.5290
 5 H1   H2   E   4  3  2  1.090    99.8     -58.1     0.0000
 6 C2   Cg   M   4  3  2  1.534   108.4    -170.8     0.1800
 7 H2   H1   E   6  4  3  1.090   107.9     173.1     0.0000
 8 N2   N3   3   6  4  3  1.457   111.1      54.0    -0.0850
 9 H1N  H    E   8  6  4  0.990   117.6     -59.2     0.2310
10 H2N  H    E   8  6  4  1.346   122.2     140.9     0.2310
11 H3N  H    E  10  8  6  1.235   123.1      -9.6     0.2310
12 C3   Cg   M   6  4  3  1.530   110.1     -68.5     0.2120
13 H3   H1   E  12  6  4  1.090   105.9      65.7     0.0000
14 O3   Oh   S  12  6  4  1.430   109.7    -174.9    -0.6410
15 H3O  Ho   E  14 12  6  0.950   109.2    -118.3     0.4360
16 C4   Cg   M  12  6  4  1.521   110.9     -54.6     0.2570
17 H4   H1   E  16 12  6  1.090   108.2     -62.9     0.0000
18 O4   Oh   S  16 12  6  1.435   110.1     175.4    -0.6960
19 H4O  Ho   E  18 16 12  0.949   112.0      43.3     0.4570
20 C5   Cg   M  16 12  6  1.528   108.7      56.8     0.3450
21 H5   H1   E  20 16 12  1.090   110.7      59.8     0.0000
22 O5   Os   E  20 16 12  1.448   108.4     -58.4    -0.5470
23 C6   Cg   3  20 16 12  1.514   114.7    -177.5     0.2940
24 H61  H1   E  23 20 16  1.090   107.4     179.3     0.0000
25 H62  H1   E  23 20 16  1.090   107.3     -62.6     0.0000
26 O6   Oh   S  23 20 16  1.416   110.4      59.4    -0.6710
27 H6O  Ho   E  26 23 20  0.950   103.2    -159.6     0.4310

LOOP
O5 C1

IMPROPER

DONE
```

β-D-Glucosamine

```
0YNp  INT  0
CORRECT     OMIT DU   BEG
 1.1940
 1 DUMM DU   M   0 -1 -2  0.000     0.0       0.0     0.0000
 2 DUMM DU   M   1  0 -1  1.522     0.0       0.0     0.0000
 3 DUMM DU   M   2  1  0  1.422   109.5       0.0     0.0000
 4 C1   Cg   M   3  2  1  1.400   113.3     180.0     0.4580
 5 H1   H2   E   4  3  2  1.102   111.6      56.2     0.0000
 6 O5   Os   M   4  3  2  1.412   112.9     -68.7    -0.4070
 7 C5   Cg   M   6  4  3  1.435   114.3    -179.7     0.1510
 8 H5   H1   E   7  6  4  1.105   109.1     -60.7     0.0000
 9 C6   Cg   3   7  6  4  1.517   106.6    -177.2     0.3300
10 H61  H1   E   9  7  6  1.092   108.3    -178.0     0.0000
11 H62  H1   E   9  7  6  1.093   108.8     -59.0     0.0000
12 O6   Oh   S   9  7  6  1.413   112.7      65.2    -0.6740
13 H6O  Ho   E  12  9  7  0.955   108.1     -58.1     0.4360
14 C4   Cg   M   7  6  4  1.528   110.8      59.1     0.3260
15 H4   H1   E  14  7  6  1.100   107.9      66.4     0.0000
16 O4   Oh   S  14  7  6  1.430   109.8    -174.0    -0.6630
17 H4O  Ho   E  16 14  7  0.958   109.2    -139.6     0.4390
18 C3   Cg   M  14  7  6  1.519   110.1     -54.4     0.1230
19 H3   H1   E  18 14  7  1.101   108.4     -66.7     0.0000
20 O3   Oh   S  18 14  7  1.421   108.7     173.5    -0.6410
21 H3O  Ho   E  20 18 14  0.976   108.5      51.3     0.4480
22 C2   Cg   M  18 14  7  1.529   110.6      52.8     0.2670
23 H2   H1   E  22 18 14  1.105   104.4      63.4     0.0000
24 N2   N3   3  22 18 14  1.415   109.5    -174.5    -0.1400
25 H1N  H    E  24 22 18  1.011   111.0    -180.0     0.2470
26 H2N  H    E  24 22 18  1.011   111.0     -60.0     0.2470
27 H3N  H    E  24 22 18  1.011   111.0      60.0     0.2470

LOOP
C2 C1

IMPROPER

DONE
```

Protonated β-D-Glucuronic acid

```
0ZBP INT 0
CORRECT OMIT DU BEG
0.1940
 1 DUMM DU   M  0 -1 -2  0.000     0.0       0.0     0.0000
 2 DUMM DU   M  1  0 -1  1.522     0.0       0.0     0.0000
 3 DUMM DU   M  2  1  0  1.422   109.5       0.0     0.0000
 4 C1   Cg   M  3  2  1  1.400   113.3     180.0     0.4450
 5 H1   H2   E  4  3  2  1.102   111.6      56.2     0.0000
 6 O5   Os   M  4  3  2  1.412   112.9     -68.7    -0.5210
 7 C5   Cg   M  6  4  3  1.435   114.3    -179.7     0.2150
 8 H5   H1   E  7  6  4  1.105   109.1     -60.7     0.0000
 9 C6   C    B  7  6  4  1.517   106.6    -177.2     0.7840
10 O6A  O2   E  9  7  6  1.230   120.0      59.0    -0.5600
11 O6B  Oh   S  9  7  6  1.230   120.0    -125.0    -0.6470
12 H6O  Ho   E 11  9  7  0.976   108.5     180.0     0.4360
13 C4   Cg   M  7  6  4  1.528   110.8      59.1     0.2730
14 H4   H1   E 13  7  6  1.100   109.8      66.4     0.0000
15 O4   Oh   S 13  7  6  1.430   107.9    -174.0    -0.7330
16 H4O  Ho   E 15 13  7  0.958   109.2    -139.6     0.4520
17 C3   Cg   M 13  7  6  1.519   110.1     -54.4     0.3040
18 H3   H1   E 17 13  7  1.101   108.4     -66.7     0.0000
19 O3   Oh   S 17 13  7  1.421   108.7     173.5    -0.7080
20 H3O  Ho   E 19 17 13  0.976   108.5      51.3     0.4310
21 C2   Cg   M 17 13  7  1.529   110.6      52.8     0.3040
22 H2   H1   E 21 17 13  1.105   104.4      63.4     0.0000
23 O2   Oh   S 21 17 13  1.415   109.5    -174.5    -0.7260
24 H2O  Ho   E 23 21 17  0.957   110.0    -149.4     0.4450

LOOP
C2 C1

IMPROPER
 C5  O6A  C6   O6B

DONE
```

Protonated α-L-Iduronic acid ($^{1}C_{4}$)

```
YuAP  INT 0
CORRECT OMIT DU BEG
-0.1940
 1 DUMM DU   M  0 -1 -2  0.000     0.0       0.0     0.0000
 2 DUMM DU   M  1  0 -1  1.522     0.0       0.0     0.0000
 3 DUMM DU   M  2  1  0  1.422   109.5       0.0     0.0000
 4 C1   Cg   M  3  2  1  1.400   113.3     180.0     0.3600
 5 H1   H2   E  4  3  2  1.102   111.6      56.2     0.0000
 6 O5   Os   M  4  3  2  1.412   112.9     -68.7    -0.4270
 7 C5   Cg   M  6  4  3  1.435   114.3    -179.7     0.2790
 8 H5   H1   E  7  6  4  1.105   109.1     -60.7     0.0000
 9 C6   C    B  7  6  4  1.517   106.6    -177.2     0.7440
10 O6A  O2   E  9  7  6  1.230   120.0      59.0    -0.6180
11 O6B  Oh   S  9  7  6  1.230   120.0    -125.0    -0.6350
12 H6O  Ho   E 11  9  7  0.976   108.5     180.0     0.4350
13 C4   Cg   M  7  6  4  1.528   110.8      59.1     0.2760
14 H4   H1   E 13  7  6  1.100   109.8      66.4     0.0000
15 C3   Cg   3 13  7  6  1.519   110.1     -54.4     0.2640
16 H3   H1   E 15 13  7  1.101   108.4     -66.7     0.0000
17 O3   Oh   S 15 13  7  1.421   108.7     173.5    -0.6940
18 H3O  Ho   E 17 15 13  0.976   108.5      51.3     0.4080
19 C2   Cg   B 15 13  7  1.529   110.6      52.8     0.3360
20 H2   H1   E 19 15 13  1.105   104.4      63.4     0.0000
21 O2   Os   E 19 15 13  1.415   109.5    -174.5    -0.4700
22 O4   Os   M 13  7  6  1.430   107.9    -174.0    -0.4520

LOOP
C2 C1

IMPROPER
 C5  O6A  C6   O6B

DONE
STOP
```

SUPPLEMENTARY INFORMATION CHAPTER 6

Mass spectrometry analysis

Tetrasaccharide 1 : [ΔUA2OS-(1-4)-GlcNS6OS-α-(1-4)-IdoA2OS-α-(1-4)-GlcNS] and

Tetrasaccharide 3 : [ΔUA2OS-(1-4)-GlcNS6OS-α-(1-4)-IdoA-α-(1-4)-GlcNS6OS]

Extensive fragmentation of the precursor ion $[M-5H+Na]^{4-}$ and $[M-6H+2Na]^{2-}$ produced highly informative and structurally relevant ions. From these ions, two structures were determined to have the highest level of agreement with the ions in the MS. All possible glycosidic product ions were observed, and cross-ring products aided in assigning sites of sulfation. Accurate mass measurement of the B_2 ions suggests the positions of the three sulfate groups to be on the non-reducing end disaccharide. The mass difference between cross ring products $^{2,4}A_2$ and $^{0,2}A_2$ positioned a sulfate group on the 6-O position while the mass difference between B_2 and $^{0,2}A_2$ was used to assign the N-sulfation on the middle amino sugar. The two isomeric penta-sulfated tetrasaccharides were differentiated using the glycosidic products B_3, C_3, Z_1, and Y_1. Accurate mass measurement of B_3 and C_3 ions showed that **1** has 1 more sulfate group compared to **3**, while Z_1 and Y_1 showed it had only one sulfo group. Cross-ring products again confirmed the site of sulfation of the reducing end glucosamine. A list of product ions used in structural assignment is provided below.

Tetrasaccharide 2: ΔUA2OS-(1-4)-GlcNS6OS-α-(1-4)-IdoA2OS-α-(1-4)-GlcNS6OS

The addition of an extra sulfate group makes **3** even more prone to sulfate decomposition. To decrease the effect of SO_3 loss, precursor ion $[M-7H+3Na]^{4-}$, was subjected to both EDD and

CID activation. At this charge state, 7 out of the 8 possible ionizable protons were deprotonated, providing an excellent platform to curb sulfate group decomposition. Product ions observed in the MS/MS due to charge reduction span over the mass range of approximately 150-700 m/z. The coverage of cross ring cleavages and glycosidic product ions with their sulfate groups intact, allows for the assignments of all 6 sulfate positions.

Tetrasaccharide 4: [ΔUA2OS-(1-4)-GlcNS6OS-α-(1-4)-IdoA-α-(1-4)-GlcNAc6OS] and Tetrasaccharide 5: ΔUA2OS-(1-4)-GlcNS-α-(1-4)-IdoA-α-(1-4)-GlcNS6OS]

Ionization of all sulfate and carboxylate groups produced sufficient cross-ring cleavages with glyosidic product ions required for sulfate position location. Precursor ions isolated for EDD fragmentation were $[M-5H+2Na]^{3-}$ and $[M-5H+Na]^{4-}$ for **4** and **5** respectively.

Tetrasaccharide 6: ΔUA2OS-(1-4)-GlcNS-α-(1-4)-IdoA-α-(1-4)-GlcNAc6OS

MS/MS analysis on tri-sulfated tetrasaccharide was carried out on the $[M-3H]^{3-}$ precursor ion. Most fragmentation of the precursor ion was observed in the internal uronic acid, with very important cross-ring cleavages on the amino sugars, essential for locating all three sulfates groups. Glycosidic product ions Z_1 and Y_1 established a sulfate group on the reducing end amino sugar, the mass difference between cross-ring products $^{0,2}A_4$ and $^{2,4}X_0$ positioned the sulfate group at the 6-O position, and N-sulfation on the middle amino sugar was un-ambiguously assigned with $^{0,2}A_2$ and B_2 fragments ions.

All uronic acid assignments for the tetrasaccharides were based on diagnostic ratios of synthetically produced standards.

Mass-to-charge/intensity table for tetrasaccharides

Tetrasaccharide 1

m/z	Intensity	Ion
333.4727	2431075	$^{3,5}A_3^{+}Na$
347.0089	2217152	$[^{2,4}X_2\text{-}SO_3]^{2-}$
386.9873	3450791	$^{2,4}X_2^{2-}$
279.496	2229337	$[^{2,4}A_3\text{+Na-}SO_3]^{2-}$
319.4747	3.13E+07	$[^{2,4}A_3\text{+Na}]^{2-}$
560.0003	911022	$^{2,4}A_3\text{+Na-}SO_3$
271.3245	2.72E+07	$[^{2,4}A_4\text{+Na-}SO_3]^{3-}$
296.9922	883224	$^{2,4}A_2$
297.9767	4999005	$[^{2,4}A_4\text{+Na}]^{3-}$
318.974	3405522	$^{2,4}A_2\text{+Na}$
270.5026	8430150	$^{1,5}X_2^{2-}$
563.9959	1877812	$^{1,5}X_2\text{+Na}$
260.3411	4909828	$[^{1,5}X_3\text{-}SO_3]^{3-}$
294.3206	3330858	$[^{1,5}X_3\text{+Na}]^{3-}$
363.483	7013531	$[^{1,5}A_3\text{+Na-}SO_3]^{2-}$
208.9757	1551544	$^{1,5}A_1\text{+Na}$
230.9577	715496	$^{1,5}A_1$
189.4938	596654	$^{0,2}X_1^{2-}$
300.0405	840276	$^{0,2}X_1\text{-}SO_3$
322.0199	1340903	$^{0,2}X_1\text{+Na-}SO_3$
379.9979	3.53E+07	$^{0,2}X_1$
401.9785	3146775	$^{0,2}X_1\text{+Na}$
316.9962	1.38E+07	$^{0,2}X_2^{2-}$
223.9922	1095759	$[^{0,2}X_3\text{+Na-}SO_3]^{4-}$
243.9817	3938072	$[^{0,2}X_3\text{+Na}]^{4-}$
272.3403	1.80E+07	$[^{0,2}X_3\text{+Na-2}SO_3]^{3-}$
291.6661	4.84E+07	$[^{0,2}X_3\text{-SO3}]^{3-}$
298.9926	3.32E+08	$[^{0,2}X_3\text{+Na-}SO_3]^{3-}$
318.3185	8.64E+07	$^{0,2}X_3^{3-}$
325.6456	2965959	$[^{0,2}X_3\text{+Na}]^{3-}$
228.972	2253494	$[^{0,2}A_4\text{-}SO_3]^{3-}$
284.0042	2563157	$[^{0,2}A_4\text{+Na-}SO_3]^{3-}$
291.3316	1.87E+08	$[^{0,2}A_4\text{+Na}]^{3-}$
317.984	3.87E+08	$[^{0,2}A_4\text{+Na}]$
357.0137	4685444	$^{0,2}A\text{-}SO_3$
378.9956	6.90E+07	$[^{0,2}A_2\text{+Na-}SO_3]^{2-}$
437.5017	6509538	$[^{0,2}A_4\text{+Na-}SO_3]^{2-}$

458.953	2628100	$^{0,2}A_2$+Na
477.481	828565	$[^{0,2}A_4$+Na$]^{2-}$
240.018	5.58E+07	Z_1
245.0059	7528803	$[Z_3$-2$SO_3]^{3-}$
247.4997	6376984	Z_2^{2-}
252.333	878741	$[Z_3$+Na-$SO_3]^{3-}$
258.4907	3.64E+07	$[Z_2$+Na$]^{2-}$
271.659	5449872	Z_3^{3-}
328.0345	3998477	$[Z_3$-2$SO_3]^{2-}$
339.0254	3157506	$[Z_3$+Na-2$SO_3]^{2-}$
379.0051	6504145	$[Z_3$+Na-$SO_3]^{2-}$
416.0508	4492925	Z_2-SO_3
438.0334	4132660	Z_2+Na-SO_3
517.9903	1.03E+07	Z_2+Na
170.6672	643264	Y_2^{3-}
216.5264	2832216	$[Y_2$-2$SO_3]^{2-}$
251.0094	2.13E+07	$[Y_3$-$SO_3]^{3-}$
256.505	8.26E+07	Y_2^{2-}
258.0286	7.46E+07	Y_1
267.496	2.97E+08	$[Y_2$+Na$]^{2-}$
277.6618	1758295	Y_3^{3-}
284.9891	6.81E+07	$[Y_3$+Na$]^{2-}$
377.0198	5184818	$[Y_3$-$SO_3]^{2-}$
388.0094	1.87E+07	$[Y_3$+Na-$SO_3]^{2-}$
434.0614	1441788	Y_2-SO_3
456.0441	5538288	Y_2+Na-SO_3
536.0007	4798615	Y_2+Na
247.4997	6376984	$[C_3$-$SO_3]^{4-}$
254.9814	3.72E+07	C_1
257.3209	2788492	$[C_3$+Na-$SO_3]^{3-}$
258.4907	3.64E+07	$[C_2$+Na-$SO_3]$
276.9634	3.62E+07	C_1+Na
283.9733	2984116	$[C_3$+Na$]^{3-}$
298.4692	4.05E+07	$[C_2$+Na$]^{2-}$
346.5072	1.28E+07	$[C_2$+Na-2$SO_3]^{2-}$
386.4858	1.70E+07	$[C_3$+Na-$SO_3]^{2-}$
416.0508	4492925	C_2-2SO_3
438.0334	4132660	C_2+Na-2SO_3
517.9903	1.03E+07	C_2+Na-SO_3
157.0137	931340	B_1-SO_3
198.5159	2107428	$[B_2$-2$SO_3]^{2-}$
236.9708	2.32E+07	B_1
238.4944	4.37E+08	$[B_2$-$SO_3]^{2-}$

243.9897	1888338	$[B_3\text{-}SO_3]^{3-}$
249.4854	3.44E+07	$[B_2\text{+}Na\text{-}SO_3]^{2-}$
251.3174	6.90E+07	$[B_3\text{+}Na\text{-}SO_3]^{3-}$
258.9528	1.21E+07	$B_1\text{+}Na$
277.9698	4.02E+07	$[B_3\text{+}Na]^{3-}$
278.473	9278216	B_2^{2-}
289.4639	2.68E+07	$[B_2\text{+}Na]^{2-}$
337.502	8153033	$[B_3\text{+}Na\text{-}2SO_3]^{2-}$
377.4803	2.86E+07	$[B_3\text{+}Na\text{-}SO_3]^{2-}$
398.0403	2.08E+07	$B_2\text{-}2SO_3$
420.0219	2700897	$B_2\text{+}Na\text{-}2SO_3$
499.9801	4123466	$B_2\text{+}Na\text{-}SO_3$

Tetrasaccharide 2

m/z	Intensity	Ion
344.4625	1204980	$[^{3,5}A_3\text{+}2Na]^{2-}$
364.9564	2711971	$[^{2,5}X_3\text{+}2Na]^{3-}$
408.9686	4266533	$[^{2,4}X_2\text{+}2Na\text{-}SO_3]^{2-}$
437.9552	1182048	$[^{2,4}X_2\text{+}Na]^{2-}$
448.9478	5253160	$[^{2,4}X_2\text{+}2Na]^{2-}$
581.9803	1680743	$^{2,4}A_3\text{+}2Na\text{-}SO_3$
683.9234	1612308	$^{2,4}A_3\text{+}3Na$
429.4725	3.96E+07	$[^{2,4}A_4\text{+}3Na\text{-}SO_3]^{2-}$
312.6308	2054621	$[^{2,4}A_4\text{+}3Na]^{3-}$
469.4512	9.51E+07	$[^{2,4}A_4\text{+}3Na]^{2-}$
318.975	7370662	$^{2,4}A_2\text{+}Na$
387.962	2782092	$^{1,5}X_1\text{+}Na$
665.9349	1453397	$^{1,5}X_2\text{+}2Na$
213.9788	1256837	$[^{1,5}X_2\text{+}Na]^{3-}$
391.0159	6502567	$[^{1,5}X_3\text{-}2SO_3]^{2-}$
463.9682	959878	$[^{1,5}X_3\text{+}3Na\text{-}SO_3]^{2-}$
335.6276	1753207	$[^{1,5}X_3\text{+}3Na]^{3-}$
328.3001	4806872	$[^{1,5}X_3\text{+}2Na]^{3-}$
503.9445	1121187	$[^{1,5}X_3\text{+}3Na]^{2-}$
385.4655	1049666	$[^{1,5}A_3\text{+}3Na\text{-}SO_3]^{2-}$
374.4735	1731915	$[^{1,5}A_3\text{+}2Na\text{-}SO_3]^{2-}$
425.4423	1686556	$[^{1,5}A_3\text{+}3Na]^{2-}$
414.452	1202804	$[^{1,5}A_3\text{+}2Na]^{2-}$
545.9553	3810846	$[^{1,5}A_4\text{+}3Na\text{-}SO_3]^{2-}$
585.9329	1275054	$[^{1,5}A_4\text{+}3Na]^{2-}$
230.9577	1875589	$^{1,5}A_1\text{+}Na$

208.9757	3124121	$^{1,5}A_1$
190.9624	1396998	$^{1,4}X_0$+Na
168.9808	4586108	$^{1,4}X_0$
401.9787	1164212	$^{0,2}X_1$+Na-SO_3
423.9594	1698573	$^{0,2}X_1$+2Na-SO_3
481.9377	5648710	$^{0,2}X_1$+Na
159.9681	3015303	$^{0,2}X_0$+Na
137.9863	3.53E+07	$^{0,2}X_0$
306.3192	944740	$[^{0,2}X_4$+2Na-2$SO_3]^{3-}$
325.6452	1571457	$[^{0,2}X_4$+Na-$SO_3]^{3-}$
340.2989	3.28E+07	$[^{0,2}X_4$+3Na-$SO_3]^{3-}$
332.9723	2.39E+07	$[^{0,2}X_4$+2Na-$SO_3]^{3-}$
274.9621	1488030	$[^{0,2}$X4+3Na$]^{4-}$
352.2988	4312272	$[^{0,2}$X4+Na$]^{3-}$
366.952	3.43E+07	$[^{0,2}X_4$+3Na$]^{3-}$
359.6248	2.66E+08	$[^{0,2}X_4$+2Na$]^{3-}$
344.9716	1250264	$^{0,2}X_4{}^{3-}$
539.9404	2176111	$[^{0,2}X_4$+2Na$]^{2-}$
332.6381	6.63E+07	$[^{0,2}A_4$+3Na-$SO_3]^{3-}$
325.3107	4283711	$[^{0,2}A_4$+2Na-$SO_3]^{3-}$
499.4616	3537453	$[^{0,2}A_4$+3Na-$SO_3]^{2-}$
359.2904	7.83E+08	$[^{0,2}A_4$+3Na^{3-}
351.9628	7155409	$[^{0,2}A_4$+2Na$]^{3-}$
539.4396	6251639	$[^{0,2}A_4$+3Na$]^{2-}$
458.9513	5519062	$^{0,2}A_2$+Na
480.9307	9838277	$^{0,2}A_2$+2Na
240.0178	6298926	Z_1-SO_3
341.9565	1.82E+07	Z_1+Na
363.9403	1018961	Z_1+2Na
319.9762	1573174	Z_1
539.9711	2364589	Z_2+2Na-SO_3
440.9636	4816395	$[Z_3$+3Na-$SO_3]^{2-}$
429.974	1.39E+07	$[Z_3$+2Na-$SO_3]^{2-}$
641.9122	2852458	Z_2+3Na
619.9296	2853514	Z_2+2Na
309.4596	4697566	$[Z_2$+2Na$]^{2-}$
320.2919	2892922	$[Z_3$+3Na$]^{3-}$
312.9651	1148688	$[Z_3$+2Na$]^{3-}$
480.9443	1.03E+07	$[Z_3$+3Na$]^{2-}$
469.9528	3.05E+07	$[Z_3$+2Na$]^{2-}$
258.0282	957228	Y_1-SO_3
337.9875	1452850	Y_1
168.4889	2553235	$Y_1{}^{2-}$

557.9813	1216621	Y_2+2Na-SO_3
449.9685	3406671	$[Y_3$+3Na-$SO_3]^{2-}$
307.4739	1.44E+07	$[Y_2$+Na$]^{2-}$
318.4651	5.28E+07	$[Y_3$+2Na$]^{3-}$
326.2956	4.30E+07	$[Y_3$+3Na$]^{3-}$
318.9678	1.65E+07	$[Y_3$+2Na$]^{3-}$
489.9477	1666493	$[Y_3$+3Na$]^{2-}$
478.9585	1429483	$[Y_3$+2Na$]^{2-}$
539.9711	2364589	C_2+2Na-SO_3
641.9122	2852458	C_2+3Na^{2-}
619.9296	2853514	C_2+2Na
309.4596	4697566	$[C_2$+2Na$]^{2-}$
408.4671	1.35E+07	$[C_3$+3Na-$SO_3]^{2-}$
283.9749	1.21E+07	$[C_3$+Na$]^{3-}$
298.6272	2668033	$[C_3$+3Na$]^{3-}$
426.4643	4251808	$[C_3$+Na$]^{2-}$
448.4453	1.99E+07	$[C_3$+3Na$]^{2-}$
276.963	5462769	C_1+Na
254.9811	6002621	C_1
521.961	2223516	B_2+2Na-SO_3
249.4852	5319634	$[B_2$+Na-$SO_3]^{2-}$
238.494	1150590	B_2-SO_3
623.8996	1024906	B_2+3Na
289.4635	8447862	$[B_2$+Na$]^{2-}$
300.4544	1.08E+07	$[B_2$+2Na$]^{2-}$
399.461	1.66E+07	$[B_3$+3Na-$SO_3]^{2-}$
388.4705	4583231	$[B_3$+2Na-$SO_3]^{2-}$
285.2963	3841507	$[B_3$+2Na$]^{3-}$
439.4399	6.76E+07	$[B_3$+3Na$]^{2-}$
428.448	2808781	$[B_3$+2Na$]^{2-}$
157.0139	4410909	B_1-SO_3
258.9525	1.08E+07	B_1+Na
236.9706	6002360	B_1

Tetrasaccharide 3

m/z	Intensity	Type
333.4714	1360213	$^{3,5}A_3^{3-}$
HS 5	654307	$[^{3,5}A_4$+Na$]^{3-}$
347.0076	1521739	$[^{2,4}X_2$-$SO_3]^{2-}$
386.9851	1833877	$^{2,4}X_2$
279.4956	1819485	$[^{2,4}A_3$+Na-$SO_3]^{2-}$

319.4737	1.11E+07	$[^{2,4}A_3+Na]^{2-}$
217.0358	661325	$^{2,4}A_2-SO_3$
244.6717	664189	$[^{2,4}A_4+Na-SO_3]^{3-}$
271.3241	6805839	$[^{2,4}A_4+Na]^{3-}$
296.9912	862778	$^{2,4}A_2$
286.0232	2404545	$^{1,5}X_1-SO_3$
270.5021	1466935	$^{1,5}X_2{}^{2-}$
563.9954	1314945	$^{1,5}X_2+Na$
260.3404	1253964	$[^{1,5}X_3-SO_3]^{3-}$
294.32	1643931	$[^{1,5}X_3+Na]^{3-}$
363.4814	2987281	$[^{1,5}A_3+Na]^{2-}$
208.9757	2050555	$^{1,5}A_1$
230.9577	672497	$^{1,5}A_1+Na$
168.9809	1045789	$^{1,4}X_0$
190.9627	1104006	$^{1,4}X_0+Na$
300.3253	1086438	$^{1,4}A_4{}^{3-}$
300.0386	860436	$^{0,2}X_1-SO_3$
379.9961	1.68E+07	$^{0,2}X_1$
137.9864	2.89E+07	$^{0,2}X_0$
159.9683	2540262	$^{0,2}X_0+Na$
243.982	711154	$[^{0,2}X_3+Na]^{4-}$
272.3398	6879106	$[^{0,2}X_3+Na-2SO_3]^{3-}$
291.6655	1.42E+07	$[^{0,2}X_3-SO3]^{3-}$
298.9919	9.03E+07	$[^{0,2}X_3+Na-SO3]^{3-}$
318.3175	4.26E+07	$^{0,2}X_3{}^{3-}$
438.0016	2187978	$[^{0,2}X_3-SO_3]^{2-}$
228.9721	1244268	$[^{0,2}A_2+Na]^{2-}$
284.0033	840480	$[^{0,2}A_4-SO_3]^{3-}$
291.331	4.19E+07	$[^{0,2}A_4+Na-SO_3]^{3-}$
317.983	1.39E+08	$[^{0,2}A_4+Na]^{3-}$
357.0126	2646054	$^{0,2}A_2-SO_3$
378.9939	2.97E+07	$^{0,2}A_2+Na-SO_3$
458.9514	1783540	$^{0,2}A_2+Na$
240.0178	2.17E+07	Z_1-SO_3
245.0051	1671386	$[Z_3-SO_3]^{3-}$
247.4992	1946629	$Z_2{}^{2-}$
258.4902	1.08E+07	$[Z_2+Na]^{2-}$
271.6588	2768642	$Z_3{}^{3-}$
278.9857	1094795	$[Z_3+Na]^{3-}$
319.9753	3231493	Z_1-SO_3
328.0328	1466623	$[Z_3-2SO_3]^{2-}$
339.0241	1614568	$[Z_2+Na-SO_3]^{2-}$
341.9561	1.74E+07	Z_1+Na

416.05	2062717	Z_2-SO_3
418.9818	798035	$[Z_3+Na]^{2-}$
438.0306	2401536	Z_2+Na-SO_3
517.988	5032927	Z_2+Na
170.6677	500446	Y_2^{3-}
216.526	1006356	$[Y_2-SO_3]^{2-}$
251.0088	5624073	$[Y_3-SO_3]^{3-}$
256.5045	1.86E+07	Y_2^{2-}
258.0283	1.95E+07	Y_1-SO_3
258.3364	2223404	$[Y_3+Na-SO_3]^{3-}$
267.4955	7.19E+07	$[Y_2+Na]$
284.9884	2.26E+07	$[Y_3+Na]^{3-}$
337.9865	1.47E+07	Y_1
348.0292	837791	$[Y_3+Na-2SO_3]^{2-}$
359.967	1511038	Y_1+Na
377.0181	2940516	$[Y_3-SO_3]^{2-}$
388.008	9819544	$[Y_3+Na-SO_3]^{2-}$
456.0418	2683026	Y_2+Na-SO_3
535.9991	2266567	Y_2+Na
175.0245	2490803	C_1-SO_3
247.4992	1946629	$[C_2-SO_3]^{2-}$
254.9809	1.35E+07	C_1
257.3207	1265177	$[C_3+Na]^{3-}$
258.4902	1.08E+07	$[C_2+Na-SO_3]^{2-}$
276.9628	1.45E+07	C_1+Na
298.4685	1.32E+07	$[C_2+Na]^{2-}$
346.5058	5528174	$[C_3+Na-SO_3]^{2-}$
386.484	8038797	$[C_3+Na]^{2-}$
416.05	2062717	C_2-$2SO_3$
438.0306	2401536	C_2+Na-$2SO_3$
517.988	5032927	C_2+Na-SO_3
157.014	5715306	B_1-SO_3
198.5158	894808	$[B_2-SO_3]^{2-}$
236.9705	7579200	B_1
238.4941	8.62E+07	$[B_2-SO_3]^{2-}$
243.9896	829266	B_3^{3-}
249.4851	1.06E+07	$[B_2+Na-SO_3]^{2-}$
251.3169	1.73E+07	$[B_3+Na]^{3-}$
258.9523	4850051	B_1+Na
278.4726	2722096	$B_2$2-
289.4634	7756332	$[B_2+Na]^{2-}$
337.5008	4551964	$[B_3+Na-SO_3]^{3-}$
377.4788	1.82E+07	$[B_3+Na]^{2-}$

398.0386	9389777	B_2-2SO_3
420.0199	1830413	B_2+Na-2SO_3
499.9785	2481302	B_2+Na-SO_3

Tetrasaccharide 4

m/z	Intensity	Type
344.4627	2794286	$[^{3,5}A_3+2Na]^{2-}$
309.9759	945061	$[^{3,5}A_4+2Na]^{3-}$
358.0443	1.36E+07	$^{2,5}X_1$
220.9733	3655505	$[^{2,5}A_2+2Na]^{2-}$
416.0495	946092	$^{2,4}X_1$
438.0314	1519124	$^{2,4}X_1$+Na
240.0178	4397136	$^{2,4}X_0$
261.9999	1146487	$^{2,4}X_0$+Na
330.4652	2797451	$^{2,4}A_3^{2-}$
581.9835	2471625	$^{2,4}A_3$+2Na-SO_3
458.0607	2.51E+07	$^{2,4}A_3$-2SO_3
480.043	9930864	$^{2,4}A_3$+Na-2SO_3^-
340.9554	2937025	$^{2,4}A_2$+2Na
318.9736	1.12E+07	$^{2,4}A_2$+Na
837.97	6674572	$^{2,4}A_4$+2Na
378.5028	7221381	$[^{2,4}A_4+2Na-SO_3]^{2-}$
758.0107	1459624	$^{2,4}A_4$+2Na-SO_3
678.0557	1958639	$^{2,4}A_4$+2Na-SO_3
328.0342	2443028	$^{1,5}X_1$
526.0481	4235780	$^{1,5}X_2$+Na
434.0022	9569726	$[^{1,5}X_3+2Na]^{2-}$
423.0125	2203005	$[^{1,5}X_3+Na]^{2-}$
374.4731	1229431	$[^{1,5}A_3+Na]^{2-}$
749.9554	1587874	$^{1,5}A_3$+Na
208.9756	2802620	$^{1,5}A_1$
230.9576	1576908	$^{1,5}A_1$+Na
168.9808	5180401	$^{1,5}X_1$
190.9629	1268596	$^{1,5}X_1$+Na
459.0099	3695017	$^{0,2}X_3^{2-}$
320.3231	1615901	$^{0,2}X_3^{2-}$
441.0102	1.78E+07	$[^{0,2}X_3+2Na]^{3-}$
367.4652	1546438	$[^{0,2}A_3+2Na]^{2-}$
327.4868	3223823	$[^{0,2}A_3+2Na-SO_3]^{2-}$
480.9339	1.97E+07	$^{0,2}A_2$+2Na
458.9519	6336015	$^{0,2}A_2$+Na

357.0127	3276534	$^{0,2}A_2$-SO_3
378.9948	1.89E+07	$^{0,2}A_2$+Na-SO_3
488.4702	1.36E+08	$[^{0,2}A_4+2Na]^{2-}$
325.3114	1.76E+08	$[^{0,2}A_4+2Na]^{3-}$
448.4918	1.17E+08	$[^{0,2}A_4+2Na\text{-}SO_3]^{2-}$
298.6585	5.34E+07	$[^{0,2}A_4+2Na\text{-}SO_3]^{2-}$
410.9994	2.62E+07	$[Z_3+2Na]^{2-}$
400.0093	4215256	$[Z_3+Na]^{2-}$
823.0059	3488893	Z_3+2Na
743.0502	1396009	Z_3+2Na-SO_3
458.0607	2.51E+07	Z_2
480.043	9930864	Z_2+Na
378.1032	1741020	Z_2-SO_3
282.0283	3050848	Z_1
304.01	932006	Z_1+Na
420.0048	8.53E+07	$[Y_3+2Na]^{2-}$
409.0148	1.78E+07	$[Y_3+Na]^{2-}$
358.0443	1.36E+07	$[Y_3\text{-}SO_3]^{2-}$
369.0353	3451347	$[Y_3+Na\text{-}SO_3]^{2-}$
476.0717	1.06E+07	Y_2
498.0536	1.95E+07	Y_2+Na
396.1142	4948821	Y_2-SO_3
300.039	4389503	Y_1
254.9811	3339959	C_1
276.963	4692061	C_1+Na
175.0244	4532207	C_1-SO_3
397.4763	2.20E+07	$[C_3+2Na]^{2-}$
386.4844	1132050	$[C_3+Na]^{2-}$
357.4973	2087489	$[C_3+2Na\text{-}SO_3]^{2-}$
309.4598	1.15E+07	$[C_2+2Na]^{2-}$
619.9286	1.73E+07	C_2+2Na
539.971	5241288	C_2+2Na-SO_3
438.0314	1519124	C_2+Na-2SO_3
236.9705	1.26E+07	B_1
258.9525	1.26E+07	B_1+Na
157.0139	6147230	B_1-SO_3
388.4706	8048353	$[B_3+2Na]^{2-}$
377.4793	7874860	$[B_3+Na]^{2-}$
243.9895	1080355	B_3^{3-}
348.4922	972468	$[B_3+2Na\text{-}SO_3]^{2-}$
300.4546	1.35E+07	$[B_2+2Na]^{2-}$
289.4638	2185730	$[B_2+Na]^{2-}$
601.9171	1034700	B_2+2Na

249.4852	4414877	$[B_2+Na-SO_3]^{2-}$
521.9612	3267805	$B_2+Na-2SO_3$

Tetrasaccharide 5

m/z	Intensity	Type
282.5018	506037	$^{3,5}A_3^{2-}$
293.4931	644955	$^{3,5}A_3+Na$
219.9889	444972	$^{2,4}X_0+Na-SO_3$
268.5049	3011025	$^{2,4}A_3^{2-}$
279.4958	3.45E+07	$[^{2,4}A_3+Na]^{2-}$
318.9739	9169035	$^{2,4}A_2+Na$
367.5139	3313673	$[^{2,4}A_4+Na]^{2-}$
736.0312	681139	$^{2,4}A_4+Na$
286.0231	901854	$^{1,5}X_1-SO_3$
260.3409	1793728	$^{1,5}X_3^{3-}$
230.9576	495364	$^{1,5}A_1$
300.04	622537	$^{0,2}X_1-SO_3$
322.0205	666637	$^{0,2}X_1+Na-SO_3$
379.9975	1042287	$^{0,2}X_1$
401.978	874227	$^{0,2}X_1+Na$
137.9864	5367006	$^{0,2}X_0+Na$
159.9683	1280122	$^{0,2}X_0$
223.9923	493462	$[^{0,2}X_3+Na]^{4-}$
291.6661	1.68E+07	$^{0,2}X_3^{3-}$
298.9925	4125438	$[^{0,2}X_3+Na]^{3-}$
284.004	6487862	$^{0,2}A_4^{3-}$
291.3315	6.51E+07	$[^{0,2}A_4+Na]^{3-}$
357.0132	1.68E+07	$[^{0,2}A_2$
437.5007	677397	$[0,2A_4+Na]^{2-}$
240.0178	1232384	Z_1-SO_3
247.4997	1510353	Z_2^{2-}
252.3334	1585878	$[Z_3+Na]^{3-}$
258.4905	1.84E+07	$[Z_2+Na]^{2-}$
319.9769	749118	Z_1
368.0156	619403	Z_3^{2-}
379.0048	1734723	$[Z_3+Na]^{2-}$
416.0509	3919888	Z_2-SO_3
438.0336	1906689	$Z_2+Na-SO_3$
517.9907	2.88E+07	Z_2+Na
168.4889	9497846	Y_1^{2-}
251.0095	7774123	Y_3^{3-}

258.0285	2098118	Y_1-SO_3
267.4956	4639643	$[Y_2+Na]^{2-}$
359.9675	3871987	Y_1+Na
388.009	1293063	$[Y_3+Na]^{2-}$
247.4997	1510353	C_2^{2-}
254.9814	762253	C_1
258.4905	1.84E+07	$[C_2+Na]^{2-}$
276.9633	4250172	C_1+Na
346.5068	1281894	$[C_3+Na]^{2-}$
416.0509	3919888	C_2-SO_3
438.0336	1906689	C_2+Na-SO_3
517.9907	2.88E+07	C_2+Na
157.014	1365344	B_1-SO_3
236.9704	842422	B_1
249.4853	2011587	$[B_2+Na]^{2-}$
258.9523	3649373	B_1+Na
297.5228	1949424	$[B_3+Na$-$SO_3]^{2-}$
326.5102	3576159	B_3^{2-}
337.5012	1.91E+07	$[B_3+Na]^{2-}$
398.04	761988	B_2-SO_3
420.0225	1507521	B_2+Na-SO_3
499.9797	1873528	B_2+Na

Tetrasaccharide 6

m/z	Intensity	Type
282.5022	5954455	$^{3,5}A_3^{2-}$
311.0071	958528	$^{3,5}A_2$
358.0442	2.16E+07	$^{2,5}X_1$
336.0933	1356075	$^{2,4}X_1$-SO_3
416.0502	7228771	$^{2,4}X_1$
240.0178	2.56E+07	$^{2,4}X_0$
268.5041	821863	$^{2,4}A_3$
316.5426	2815127	$[^{2,4}A_4$-$SO_3]^{2-}$
356.5208	3171952	$^{2,4}A_4^{2-}$
424.1096	1883701	$^{1,5}X_2$-SO_3
372.042	3758260	$^{1,5}X_3^{2-}$
312.5128	1120309	$^{1,5}A_3$
208.9755	1692749	$^{1,5}A_1$
168.9807	1121880	$^{1,4}X_1$
298.0232	955730	$^{0,2}X_2^{2-}$
379.0492	2.56E+07	$[^{0,2}X_3$-$SO_3]^{2-}$

277.0561	2.05E+07	$^{0,2}A_2$-SO_3
357.0125	3379999	$^{0,2}A_2$
284.0037	1.42E+07	$^{0,2}A_4^{3-}$
346.553	1178946	$[^{0,2}A_4$-$2SO_3]^{2-}$
386.5312	7.18E+08	$[^{0,2}A_4$-$SO_3]^{2-}$
426.5096	3.83E+08	$^{0,2}A_4^{2-}$
282.0284	1.40E+07	Z_1
349.0389	2979704	Z_3^{2-}
378.1034	3.02E+07	Z_2-SO_3
458.0591	1685640	Z_2
619.1256	1450245	Z_3-SO_3
300.0389	1643163	Y_1
358.0442	2.16E+07	Y_3^{2-}
396.114	1.19E+08	Y_2-SO_3
476.0709	1305532	Y_2
175.0243	1.94E+07	C_1-SO_3
223.341	996944	C_3^{2-}
247.4994	7423082	C_2
254.9811	1.36E+07	C_1
335.5151	9356111	C_3^{2-}
336.0933	1356075	C_2-$2SO_3$
416.0502	7228771	C_2-SO_3
157.0138	1.65E+07	B_1-SO_3
198.5161	924110	$[B_2$-$SO_3]^{2-}$
236.9705	1.09E+07	B_1
238.4941	8.65E+07	B_2^{2-}
286.5317	1442611	$[B_3$-$SO_3]^{2-}$
318.0823	4265602	B_2-SO_3
326.5102	7054038	$B_3$2-
398.0395	2.84E+07	B_2-SO_3
477.9963	7227486	B_2
574.0716	1408488	B_3-SO_3