CHEMICAL DERIVATIZATION COUPLED WITH LIQUID CHROMATOGRAPHY AND TANDEM MASS SPECTROMETRY FOR STRUCTURAL SEQUENCING OF SULFATED GLYCOSAMINOGLYCANS

by

RONGRONG HUANG

(Under the Direction of Joshua S. Sharp)

ABSTRACT

Sulfated glycosaminoglycans (GAGs) are important carbohydrates that participate in essential biological activities by interacting with various proteins in the extracellular matrix. Structural characterization of these biomolecules is of great value to understand their functions, but also challenging due to their highly heterogeneous nature and the lability of the sulfate modifications. The work presented here describes a new method using chemical derivatization coupled with liquid chromatography and tandem mass spectrometry (LC-MS/MS) for structural characterization of two major classes of sulfated GAGs, chondroitin sulfate (CS) and heparan sulfate (HS)/heparin. By replacing the labile and strongly polar sulfate groups with much more stable and hydrophobic acetyl groups, the chemical derivatization method allows isomeric GAG oligosaccharides being separated by standard reversed phase capillary high performance LC and distinguished by their fragmentation patterns generated from conventional collision induced dissociation (CID) MS/MS. The method is successfully used for structural sequencing of CS and HS oligosaccharides, including most of common sulfation positional isomers as well as epimers, and its application in biological relevant cases is also investigated.

INDEX WORDS: Mass Spectrometry, Tandem Mass Spectrometry, High Performance Liquid Chromatography, Reversed Phase Liquid Chromatography, Glycosaminoglycans, Chondroitin Sulfate, Heparan Sulfate, Heparin, Reduction, Permthylation, Desulfation, Peracetylation

CHEMICAL DERIVATIZATION COUPLED WITH LIQUID CHROMATOGRAPHY AND TANDEM MASS SPECTROMETRY FOR STRUCTURAL SEQUENCING OF SULFATED GLYCOSAMINOGLYCANS

by

RONGRONG HUANG

B.S., University of Science and Technology of China, 2008

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

2014

CHEMICAL DERIVATIZATION COUPLED WITH LIQUID CHROMATOGRAPHY AND TANDEM MASS SPECTROMETRY FOR STRUCTURAL SEQUENCING OF SULFATED GLYCOSAMINOGLYCANS

by

RONGRONG HUANG

Major Professor: Joshua S. Sharp

Committee: Ron Orlando
I. Jonathan Amster

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2014

DEDICATION

To my supportive parents, my amazing husband, and my sweet little girl!

ACKNOWLEDGEMENTS

There is no way I can accomplish this work on my own. First and foremost, I would like to thank my advisor Josh Sharp, who always inspires me to think creatively, encourages me to take challenges and provides me valuable career advices. It is his guidance and persistent help that makes this dissertation possible and for that I am very grateful. I also thank my committee members, Ron Orlando and Jon Amster, for their time and advices throughout my study in graduate school.

I would like to acknowledge my collaborators. Thanks to Vitor Pomin, for introducing glycosaminoglycans (GAGs) to me and for the training on various purification and sample preparation techniques involved in GAG analysis. Thanks to Christian Heiss and Zhirui Wang for their advices on chemical derivatizations. Thanks to Jian Liu at the University of North Carolina and Geert-Jan Boons group (Andre Venot and Chengli Zong) for providing synthetic HS standards. Thanks to Barry Boyes (Advanced Materials Technologies, Inc.) for generously providing us with the custom C18 HALO columns and helpful discussions regarding the reversed phase LC separation. Thanks to Kelley Moremen group, Lianchun Wang group (Eduard Condac), Lance Wells group, Parastoo Azadi group and Jon Amster group (John Muchena) for their collaboration and instrumentation on the Robo1 project. Thanks to all my former and current group members and our instrumentation managers, for their assistance, encouragement and company.

Last but not least, a special thank you to my dear family and friends for their continuous love and support.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Glcosaminoglycans

Glycans (also referred to as polysaccharides), either alone or as part of glycoconjungates by attaching to proteins or lipids, are indispensable to all living cells for survival. For the past several decades, increasing research efforts have been made in the area of glycobiology[1-3], where a variety of glycoconjungates have been discovered and studied, including proteoglycans (PGs), glycoproteins, glycolipids and some others as illustrated in **Figure 1.1**. Glycosaminoglycans (GAGs), often attached to proteins as proteoglycans, are a family of negatively charged, linear polysaccharides that can be found in almost all vertebrate and even some invertebrate cells[4,5], involved with various important biological functions[6-8]. Based on the difference of their repeating disaccharide units, GAGs can be divided into four main classes: heparan sulfate (HS)/heparin, chondroitin sulfate (CS)/dermatan sulfate (DS), keratin sulfate (KS) and hyaluronan (HA).

Typically, the biosynthesis of GAGs follows three steps[1,9-11] including initiation, polymerization and chain modification. Unlike proteins or nucleic acid, such process is not template-driven but rather dynamically modulated by a large number of different enzymes that may engage only a subset of potentially available residues leaving others unmodified. Thus, the resulting GAG polysaccharides often present extreme structural complexity with high polydispersity and microheterogeniety. Current knowledge suggests that GAG structure is more cell specific rather than PG type specific[1,10,12], and therefore, the structure may differ in functions regarding to different biological sources, physiological activities or even pathological stages. Each type of GAG has unique structures and properties, and goes through different biosynthesis pathways. Sulfated GAGs (HS/heparin, CS/DS, KS) are synthesized in the Golgi by attaching to core proteins through specific linkage regions (as represented in **Figure 1.1**), whereas the

sulfate-less HA is made as a free glycan via HA synthases (HAS) located in the plasma membrane.

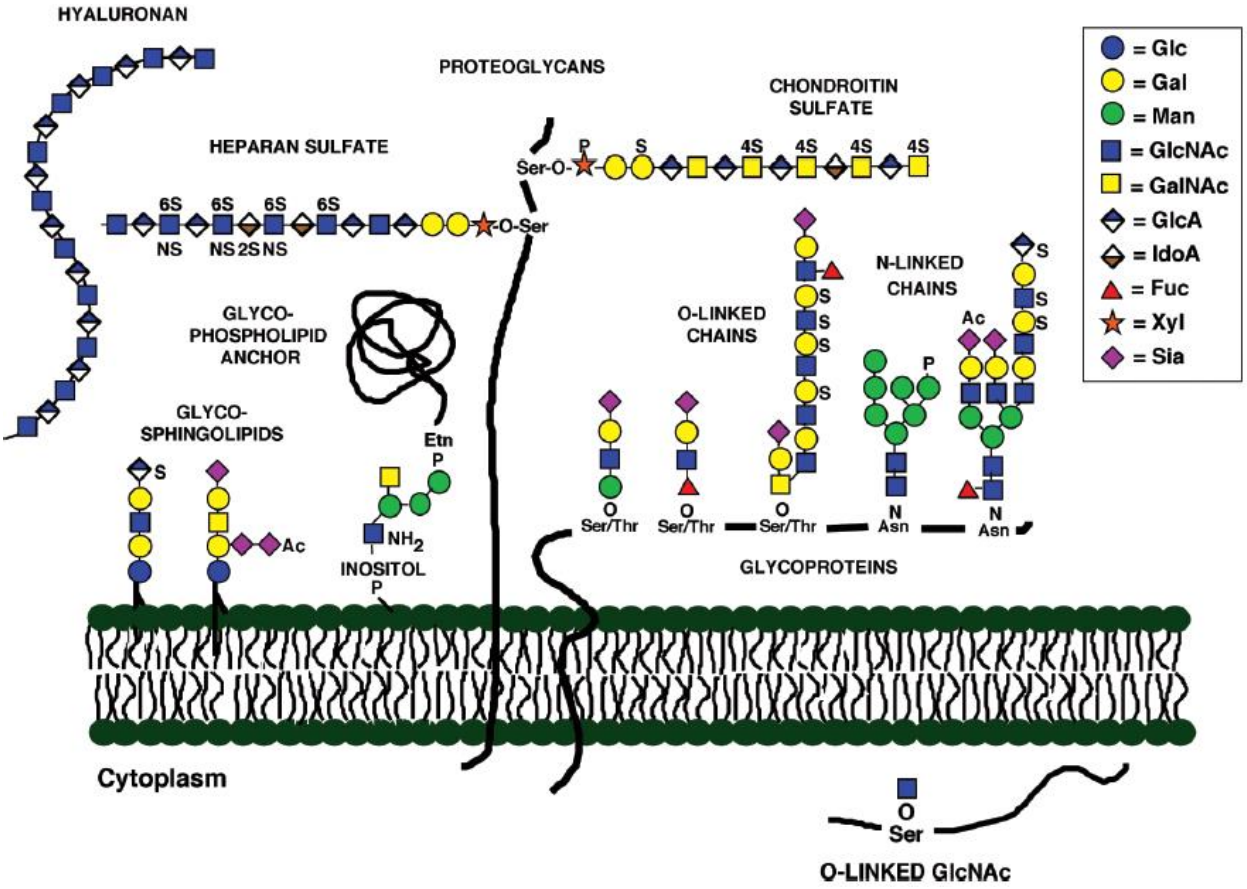


Figure 1.1: Schematic representation of the glycocalyx on vertebrate cells showing the major classes of glycoconjugates. A symbol nomenclature of the various monosaccharides is described in the inset. Glc, glucose; Gal, galactose; Man, mannose; GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine; GlcA, glucuronic acid; IdoA, L-iduronic acid; Fuc. L-fucose; Xyl, xylose; Sia, silic acid; Ac, acetyl; P, phosphate; S, sulfate; NS, N-sulfate; 2S, 3S, 6S indicate the position of O-sulfate groups; Etn, ethanolamine[3].

HS/heparin is O-linked to serine (Ser) residues of core proteins through a tetrasaccharide linkage sequence: glucuronosyl–galactosyl–galactosyl–xylosyl–Ser (GlcAβ1–3Galβ1–3Galβ1–4Xylβ1–Ser). HS/heparin polysaccharides of $(–4GlcAβ1–4GlcNAcα1–)_n$ is generated by

alternately adding N-acetylglucosamine (GlcNAc) and GlcA residues to the linkage region. The subsequence modifications of the precursor chain are implemented by a complex series of reactions including deacetylation, N-/O-sulfation, and epimerization. For GlcNAc residue, the acetylated amine group can be deacetylated into glucosamine (GlcN), or further N-sulfated into N-sulfated glucosamine (GlcNS), whereas the GlcA residue can be converted into iduronic acid (IdoA) differing only by the stereochemistry chemistry at the C5 position (**Figure 1.2**). Additionally, the O-sulfation can happen to GlcA/IdoA residues at the C2 position (2S), and to GlcNAc/GlcN/GlcNS residues at the C6 position (6S) or at, in uncommon but biologically important instances, the C3 position (3S). Heparan sulfate proteoglycans (HSPGs) are made by virtually all cell types, and can be found either expressed on the plasma membrane of cells (e.g. syndecans and glypicans) or secreted in the extracellular matrix (ECM) between cells (e.g. perlecan, agrin and collagen). Heparin PGs, however, are believed only produced by mast cells and stored as serglycin in intracellular secretory granules[13-15].

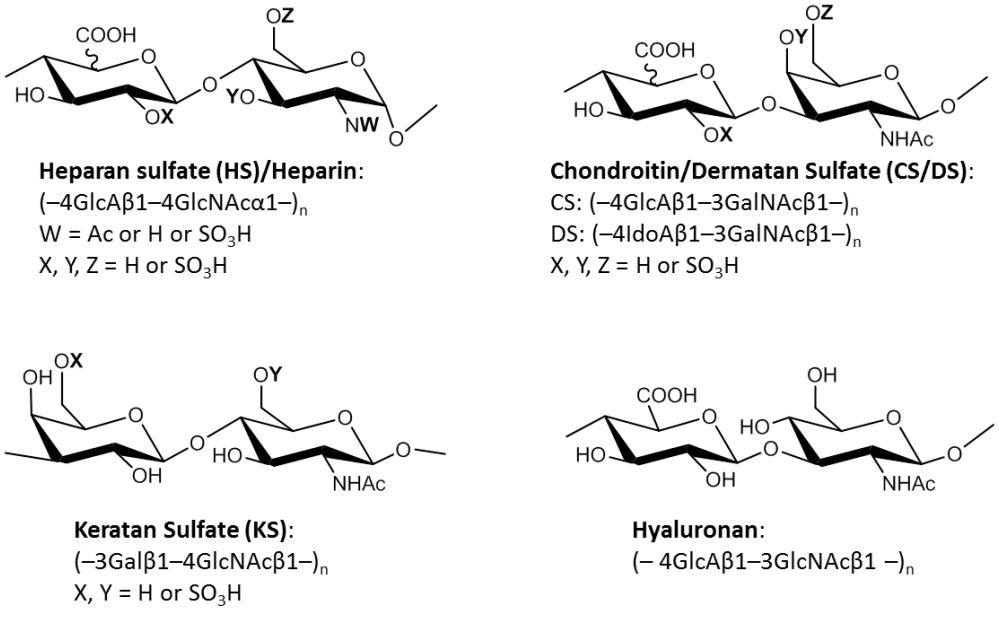


Figure 1.2: Repeating disaccharide structures for the glycosaminoglycans.

CS/DS is synthesized from the identical linkage tetrasaccharide as HS/heparin, by adding N-acetylgalactosamine (GalNAc) (instead of GlcNAc) and GlcA to generate the precursor polysaccharides $(–4GlcA\beta1–3GalNAc\beta1–)_n$. Epimerization on a portion of GlcA residues by converting them into IdoA generates DS, and O-sulfations can occur either at C4/C6 position of GalNAc or C2 position of GlcA/IdoA. CS/DS is generally secreted in the form of proteoglycans (e.g. aggrecan, versican, brevican and neurcan) as major components of the ECM of various tissues including cartilage, skin, heart valve and some others[16-18]. CS/DS is the most abundant type of GAG that can be found in the ECM proteoglycans.

KS can attach to proteins as a branch of either an N-linked glycans through asparagine (Asn) or O-linked glycans through serine/threonine (Ser/Thr) (Figure 1.1). The basic repeating disaccharide unit of KS are composed of galactose (Gal) and GlcNAc (–3Galβ1–4GlcNAcβ1–), where O-sulfation modification can occur on the C6 position of either or both of the monosaccharide residues. KS can be found on limited proteoglycans (e.g. lumican, keratocan, fibromodulin, aggrecan) on the cell surface or in the ECM, with the majority in cornea, but also in some other tissues including cartilage and bone[19,20].

HA is unique among the GAG family, in that it lacks sulfation modifications, is synthesized in plasma membrane rather than in the Golgi and secreted as a free glycan, and has a high molecular weight (MW) of up to 2000kDa, while other GAGs usually fall in the MW range of 5k~70kDa. The repeating disaccharide unit composed of HA is –4GlcAβ1–3GlcNAcβ1–. HA is one of the major components of ECM, and widely distributed in articular cartilage, skin, eye and most body fluid[21,22].

Heparin and heparan sulfate

Heparin and heparan sulfate are the most complex and heterogeneous GAG, as they have the most structural variations of modification patterns made up by a series of human enzymes including four N-deacetylase-N-sulfotransferase (NDST)[23], one C-5 epimerase[24], one 2-O sulfotransferase[25], three 6-O sulfotransferases[26] and seven 3-O sulfotransferases[27]. Mathematically, a total of 48 theoretical disaccharide units could be generated, with the number increasing exponentially as the chains build up, giving enormous potential structural variability, in terms of "sequence", of HS/heparin oligosaccharides. The current data, however, suggest that not all of the modifications are present with equal extent and some of them may only occur when certain adjacent residues are presented[28,29]. For example, epimerization only happens to GlcA residues when they are linked to N-sulfated residue at the non-reducing end[30]. Also, 2-O-sulfation is much more likely to happen on IdoA residues rather than on GlcA resides[25], whereas 3-O-sulfation, happens relatively rarely, and could be found on either GlcN or GlcNS residue, but not on GlcNAc residue[31]. Generally, heparin is more highly sulfated and less heterogeneous along the polysaccharide chain by having primarily (75-90%) disaccharide units of IdoA2S–GlcNS6S, whereas HS is less sulfated and more heterogeneous composed of three different broad types of domains regarding to their N-substitutions: heparin-like N-sulfated (NS) domains, N-acetylated (NA) domains and NS/NA domains[1] (**Figure 1.3**).

HS/heparin can interact with a variety of protein ligands, such as growth factors and their receptors, cytokines and morphogens, enzymes and enzyme inhibitors, and other ECM molecules[32-34]. Through these interactions, HS/heparin has been found involved with numerous physiological processes[35-37], including cell signaling and growth, cell adhesion and migration, anticoagulation, and angiogenesis; as well as a number of diseases, such as, inflammation, viral

infection, tumor development and metastasis, type-2 diabetes and Alzheimer's disease[38-42]. Besides HS/heparin-protein interactions, degradation of HS by human heparanase and endosulfatases in ECM and by exoenzymes in lysosomes is also physiologically important[29]. Products of HS degradation by heparanase have been found directly involved in tumor progression and metastasis[43]. Malfunction of the exoenzymes can lead to mucopolysaccharidoses (MPS), a group of lysosomal storage disorders, resulting in the lysosomal accumulation of partially degraded HS and their excretion into the urine[44].

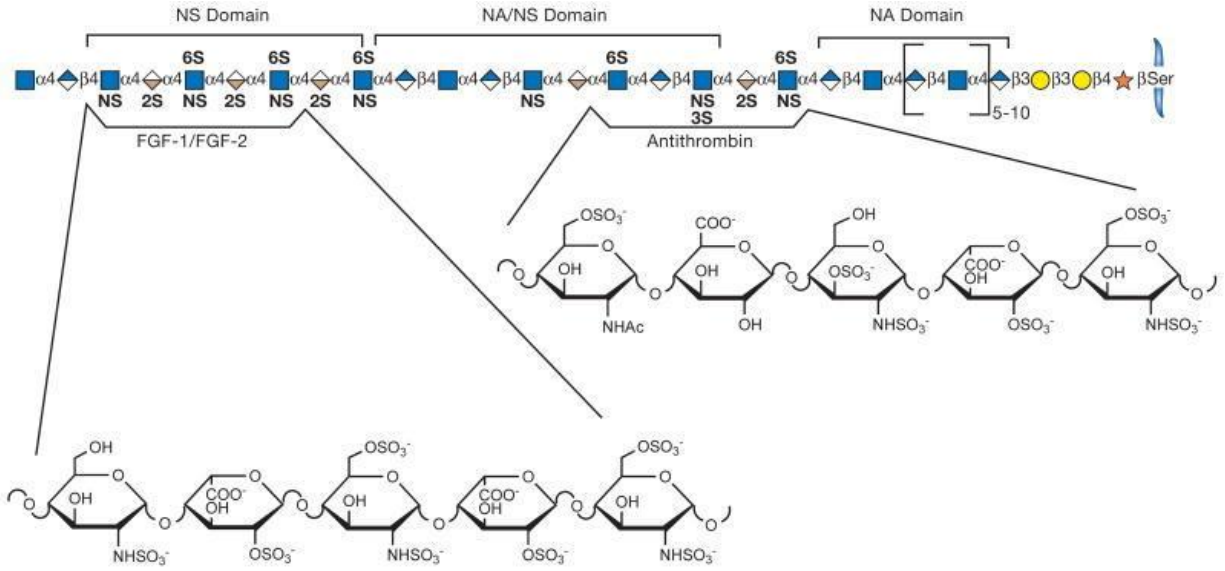


Figure 1.3: Heparan sulfate chain. HS consists of different domains that vary in the extent of modification by sulfation and epimerizaion with binding sites illustrated for fibroblast growth factors (FGFs) and antithrombin[1].

The protein-binding HS domains usually range in size from dp4 (depolymerized tetrasaccharide) to dp12 or even larger[45,46], where specific sequences may be crucial for high-affinity binding and therefore have therapeutic potentials. Heparin is one of the oldest drugs that

have been used as anticoagulants ever since 1930s[47]. Through binding to antithrombin III (AT III), heparin causes a conformational change and activation of AT III, which prevents the formation of clots within blood. A pentasaccharide (**Figure 1.3**) believed to be responsible for the specific binding could be also found in NA/NS domains of certain HS species[48,49], allowing HS to possess anticoagulant activity as well. Two other examples where HS has been found interacting with proteins in selective manners are: 1) a minimum of pentasaccharide sequence of GlcNS–GlcA–GlcNS–IdoA2S–GlcNS required for high affinity binding with fibroblast growth factor 1 (FGF1)[50] and 2) HS domains (around dp10~dp12) containing a sequence of UA±2S–GlcNS–IdoA2S–GlcN3S±6S–UA±2S–GlcANS responsible for interaction with herpes simplex virus glycoprotein gD (HSV gD)[41].

Chondroitin sulfate and dermatan sulfate

CS/DS share similar biosynthesis as HS/heparin. The heterogeneity of CS caused mainly by the variability of O-sulfation positions, where C5-emiperizations generates DS sequences. Several types of CS disaccharide unit have been found in vertebrates, including GlcA–GalNAc, GlcA–GalNAc4S, GlcA–GalNAc6S, GlcA2S–GalNAc6S and GlcA–GalNAc4S6S, whereas DS disaccharide units are usually found as IdoA±2S–GalNAc4S[18]. It is noteworthy that most of CS/DS contains more than one type of disaccharide unit in varying proportions depending on the animal species, the tissue of expression or the (patho)physiological conditions[51]. Structural heterogeneity of CS/DS can be even more diverse in invertebrates, where C3 position of GlcA residue can be sulfated or fucosylated.

CS/DS mediate numerous physiological processes via interactions with key proteins, such as growth factors, cytokines, chemokines, lipoproteins, etc. It has been believed that these interactions are attributed to the particular oligosaccharide domains with specific sulfation

patterns[52], whereas abnormal expression and modification of CS/DS are usually associated with diseases like osteoarthritis, atherosclerosis and even cancer. As an important component of cartilage, the loss of CS from cartilage has been considered a major cause of osteoarthritis. Significant amount increases of the GlcA–GalNAc6S unit in CS chains have been found in early atherosclerotic lesions[53]. Two octasaccharide sequences have been isolated bound to WF6 with high specificity[54], a monoclonal antibody that recognizes CS epitopes expressed in ovarian cancer and some joint diseases. Furthermore, DS oligosaccharides containing IdoA2S–GalNAc4S repeating units have been found responsible for their antithrombotic activity via binding to the thrombin inhibitor heparin cofactor II (HCII)[55].

Structural analysis of GAGs: significance and challenge

Because of their close structure-function relationships, a detailed knowledge of GAG structure is required for an in-depth understanding of the mechanisms by which they mediate so many biological functions. In addition, structural analysis of functional GAG sequences responsible for specific GAG-protein interactions is also critical for screening and discovery of potential therapeutic agents and/or targets. One of the examples is Fondaparinux (Arixtra)[56], a synthetic pentasaccharide anticoagulant with similar structure to the sequence responsible for heparin-antithrombin III binding. It has also been noted that aberrant expressions and modifications or degradations of GAGs are associated with several diseases, where the structural analysis of GAGs could be useful for disease diagnosis or staging, and for indication of disease progression or efficacy of therapy. For example, degraded HS/DS oligosaccharides presented in mucopolysaccharidosis (MPS) patients' urine have been found discriminative from unaffected controls, allowing their use as biomarkers for MPS diseases diagnosis[57]. Furthermore, the health crisis in 2008 resulting from contamination of pharmaceutical heparin with chemically over-

sulfated chondroitin sulfate[58] drew attention to the need of sophisticated analytical methods for securing the quality and safety of GAG-related drugs.

The structural analysis of GAGs are extremely challenging mainly caused by their high density of negative charge, polydispersity and microheterogeneity. Detailed structural characterization requires techniques that are capable of discrimination of small differences between structures, such as sulfation positional isomers and epimers. Nuclear magnetic resonance (NMR) techniques have been demonstrated as a powerful tool to provide accurate structural informations[59,60] including monosaccharide composition, glycosidic linkage, sulfation patterns, and uronic acid types, which also allows for conformational analysis and modeling of GAG oligosaccharides[61]. In most of cases, however, the available material amounts of GAGs from natural souses are often limited and obtained by labor-intensive isolation protocols, where analytical techniques with high sensitivity are more favored, such as mass spectrometry (MS). The major challenges involved in MS-based structural characterization of GAGs include the difficulty to distinguish uronic acid types and to maintain sulfation position information using traditional fragmentation methods[62]. Furthermore, the biologically derived GAG samples usually present as complex mixtures due to their natural heterogeneity, where advanced techniques with high selectivity and resolution are needed for isolating structurally defined GAG sequences. Undoubtedly, the successful combination of separation and spectral techniques can provide critical advantages in structural elucidation of GAGs[63].

Mass spectrometry based structural analysis of GAGs

Owing to their high sensitivity, chromatogram compatibility and rapid progresses of ionization and dissociation techniques for the past decades, mass spectrometry (MS) has become an incredibly valuable tool for structural characterization of GAGs. Research efforts have been

devoted to develop advanced MS-based methods to obtain one or some of the many structural features of GAGs, such as, molecular weight and chain length, disaccharides compositions, as well as glycosidic linkages, sulfation patterns and stereochemistry of uronic acid residues. Most of these methods utilize a bottom-up approach due to GAG's high degree of structural diversity, where chemical or enzymatic degradation of intact GAG polysaccharides chains are applied to generate smaller oligosaccharides for subsequent analysis. GAGs are degraded into their disaccharide building blocks in a typical exhaustive digestion, where compositional analysis is performed to determine disaccharide unit types and their relative abundant. Oligosaccharides larger than disaccharides, which are more biologically relevant, are obtained by partial depolymerizations. The resulted mixtures are then resolved through off- or on-line separations based on their lengths, compositions, or even modification patterns, followed with oligosaccharides profiling by high resolution MS analysis or detailed structural characterization of individual sequence by tandem mass spectrometry (MS/MS).

Enzymatic and chemical depolymerization of GAGs

Enzymes for different type of GAGs are commercially available to generate GAG-derived oligosaccharides for structural analysis. There are two types of enzymes that are mostly used for GAG depolymerizaions[64]: 1) prokaryotic lyases including HA lyases, CS lyases, HS lyases and a relatively new one named K5 lyase[65] and 2) eukaryotic glycosidases (hydrolases) including hyaluronidase, heparanases and keratanase. Most lyases are primarily endolytic enzymes, which β-eliminatively cleave internal hexoamine–uronic acid bonds leaving a free reducing end in hexoamine and a C4-5 double bond in uronic acid respentively. Oligosaccharides generated by lyases usually benefit from their double bond's strong absorptions at 232 nm for sensitive UV detection, however, information about epimerization state is lost for the non-

reducing end uronic acid. Hydrolases, on the other hand, break GAGs by adding water to their glycosidic bond, leaving saturated uronic acid and free reducing end for the derived oligosaccharides. Each type of enzymes has its specificity (**Table 1.1**), regardless its extent, favoring to cleave certain substrates, although many of them are not strict or well-characterized.

Table 1.1: Enzymes used to examine GAG structure[64,65].

GAG digestion enzymes	Substrate	Specificity	Action pattern
Heparin lyase I	Heparin/HS	GlcNS ± 6Sα 1–4 IdoA2Sα	Endolytic
Heparin lyase II	Heparin/HS	Linkages containing 1–4 IdoA ± 2Sα/GlcAβ 1–4	Endolytic
Heparin lyase III	HS	GlcNAcα/GlcNS ± 6Sα 1–4 IdoAα/GlcAβ	Endolytic
Heparanase	HS	At GlcAβ 1–4 GlcNS ± 3S ± 6Sα or GlcAβ 1–4 GlcNα with a GlcA2Sβ (not IdoA2Sα) in proximity	Endolytic
Chondroitin lyase ABC	CS/DS, HA with low efficiency	GalNAc ± 4S ± 6Sβ 1–4 IdoAα/GlcA ± 2Sβ 1–3 or HA	Endolytic
Chondroitin lyase ABC	CS/DS, HA with low efficiency	GalNAc ± 4S ± 6Sβ 1–4 IdoAα/GlcA ± 2Sβ 1–3 or HA	Exolytic
Chondroitinase lyase AC-I	CS/DS/HA	GalNAc ± 4S ± 6Sβ 1–4 GlcA ± 2Sβ 1–3 or HA	Endolytic
Chondroitinase lyase AC-II	CS/DS/HA	GalNAc ± 4S ± 6Sβ 1–4 GlcA ± 2Sβ 1–3 or HA	Exolytic
Chondroitinase lyase B	DS	GalNAc ± 4Sβ 1–4 IdoAα 1–3	Endolytic
Hyaluronan lyase	HA or CS unsulfated regions	GlcNAcβ 1–4 GlcAβ 1–3	Endolytic
Hyaluronidase (mammalian)	HA or CS	GlcNAcβ 1–4 GlcAβ 1–3 and GalNAc ± 4S ± 6Sβ 1–4 GlcA ± 2Sβ 1–3	Endolytic
Keratanase	KS	Galβ 1–4 GlcNAc ± 4S ± 6Sβ 1–3 Nonsulfated linkages are resistant	Endolytic
K5 lyase	HS, NA domain	GlcNAcα 1–4 GlcAα	Endolytic

Chemical depolymerization of GAGs can occur by several ways, including β-elimination, reductive deamination and some others. β-elimination[66] generates oligosaccharides similar to enzymatic degraded ones, which involves chemical reactions including carboxylate ester formation, C5 proton extraction and ester hydrolysis. Reductive deamination[67] is typically

performed with nitrous acid, where cleavages occur at GlcNS residue with pH 1.5 solutions or at GlcNH residue with pH 4 solutions, forming an 2,5-anhydro-D-mannose residue at the reducing end. The nitrous acid degradation is the most commonly used chemical depolymerization methods for GAGs, which, unlike lyase digestions, are able to reserve the epimerization information for non-reducing end uronic acid residues.

Separation and purification of GAGs

Depolymerizations of GAG produce complex mixtures of oligosaccharide varying in compositions and modification patterns, and even in sizes when partial degradation are performed to isolate larger oligosaccharides for study of their protein-binding properties or biological activities. Therefore, it is usually necessary to integrate one or more separation steps into the analysis for on-line separation or off-line purification. Separation of positional and configuration isomers gets harder as oligosaccharide chain lengths increase, due to their increased structural diversity. Various separation methods have been studied for either one or both of off-line and on-line separations of GAG[63,68].

Size-exclusion chromatography (SEC), also referred as gel filtration chromatography, separates molecules based on their sizes and is widely used for resolving GAG-derived oligosaccharides into size-uniform fractions. Large scale SEC is usually used as the first step to prepare oligosaccharide fractions of unique length with amount up to several mg, which can be used for protein-binding studies or further separate to yield simpler mixtures or single-component oligosaccharides for method development studies[69]. Analytical scale SEC has also been used as on-line separation for oligosaccharides profiling, but in a limited use because of their relative low resolution and efficiency comparing to some other methods[70,71].

Strong anion exchange chromatography (SAX) separates molecules based on their negative charge numbers, densities and even distributions, which makes it a great tool for GAG separation of sulfation positional isomers with high resolution. Semi-preparative SAX columns in high performance liquid chromatography (HPLC) system are usually used for further resolving SEC fractions based on different sulfation degrees and/or positions[72,73]. Single-component oligosaccharide could be obtained up to hexasaccharides may be larger depending on the sample complexity, where the resolution, as with other separation techniques, decreases rapidly as oligosaccharide length increases. While it provides highest resolution among other HPLC techniques, the application of SAX is often limited to off-line separation due to the high salt concentration of the mobile phase that prevents its direct interference with MS detection.

Hydrophilic interaction chromatography (HILIC) is a variant of normal phase LC that separate molecules based on their polarity, where more hydrophilic molecules are generally eluted later. HILIC has been shown capable of resolving oligosaccharides with different sizes, acetylation and sulfation contents, but not for isomeric or epimeric ones[74]. Because of their MS compatibility and high efficiency of separating wide MW range, HILIC-MS has been commonly used for oligosaccharides profiling[75,76], where GAG-derived oligosaccharide mixtures could be directly analyzed with or without SEC separation.

Reverse phase ion pairing (RPIP) LC and graphitized carbon chromatography are also reported to be used for on-line LC separation of GAG oligosaccharides. In RPIP-LC, ion pairing reagents are used to increase the retention of GAG on standard reverse phase stationary phase, which has been shown able to resolve all 12 commercial available HS disaccharides making it a useful tool for disaccharide compositional analysis of GAGs[77]. Graphitized carbon LC comes with more hydrophobic stationary phase than standard octadecylsilyl RP-LC, allowing its

retention of GAG oligosaccharides and separation of positional isomers[78]. However, the application of these techniques on separating larger oligosaccharides with isomers and epimers has not been achieved.

Capillary electrophoresis (CE) is an increasingly employed method for separation of GAG oligosaccharides, which resolve molecules based on their ionic mobility. CE has been shown capable of resolving positional and configuration isomeric disaccharides[79] and also been extended to separation of larger oligosaccharides[80] as well as low molecular weight heparin[81] (LMWH). While CE provides high resolution and efficiency as SAX does, the application of CE-MS for GAG analysis is still limited because of the requirement of complicated CE/MS interface optimization and high concentrations of nonvolatile salts for better resolving power.

Ion mobility spectrometry (IMS), which are technologies for gas-phase separation of ions, has also been reported recently for separation of GAG oligosaccharides. IMS has been used for separating two epimeric HS tetrasaccharides[82] as well as certain isomeric heparin octasaccharides containing homogeneous disaccharide units[83], but not for complex oligosaccharide mixtures. While these gas-phase separations have shown selectivity for isomers and epimers, the requirement of subtle instrumental tuning limits their application for robust and high throughput analysis of complex mixtures.

LC separation coupled with partial depolymerization of GAG is usually performed in order to enrich oligosaccharides for analytical method development or biological function studies. SEC is generally employed to separate depolymerized GAG mixtures into fractions with uniform size (dp2, dp4, dp6, dp8, and etc.). When homogeneous samples with defined structures are required for analytical method development, single or multiple steps of SAX is applied to further separate fractions obtained from SEC. GAG oligosaccharides with uniform size, for

example dp8 or dp10 fraction, are usually used to study GAG-protein interactions. In order to define the structure of the oligosaccharides that mediates a particular interaction, affinity purification is employed for isolation of GAG ligands that bind with high affinity to target protein or protein complex of interest. Affinity purification can be performed using either immobilization strategies or trapping assays. For immobilization strategies[84], GAG-binding proteins are generated as tagged forms that will allow direct immobilization to affinity supports pre-conjugated with tag-binding components. The GAG oligosaccharides are loaded on such affinity columns and fractions with low and high affinity will be eluted separately. Affinity purification by trapping assays is also an option, in which protein and GAG oligosaccharides are mixed and then applied to either filtration[85] or hydrophobic[86] trapping assays. The protein and protein-GAG complex are captured by a filter with proper molecular weight cut-off or retained on a hydrophobic stationary phase, where the bound oligosaccharides are released after removing unbound oligosaccharides.

Mass spectrometry of GAGs

Mass spectrometry is an important tool for structural analysis of GAG, offering accurate molecular weight measurement for composition determinations of intact oligosaccharides and several approaches for structural characterization through tandem mass spectrometry. Several ionization techniques have been used for MS analyses of GAGs including fast-atom bombardment (FAB)[87], one of the first techniques, and modern soft methods matrix-assisted laser desorption-ionization (MALDI)[88,89] and electrospray ionization (ESI)[90]. Among these methods, ESI is the most widely used one due to its gentleness for in-source fragmentation minimization and easy chromatography amenability for high-throughput analysis. Multiple MS analyzers have been used with ESI for GAG analysis, including ion trap, quadruple-time of flight

(Q-TOF), and Fourier transform ion cyclotron resonance (FTICR), offering a number of MS/MS techniques to generate structural informative fragments for detailed characterization, such as sulfation patterns and epimerizations.

Collisional induced dissociation (CID) is the most widely employed MS/MS technique that dissociates selected precursor ions by collision with gas atoms in a collision cell. For a typical CID performed in an ion trap, the weakest bonds (usually the glycosidic bonds) rupture to provide most abundant product ions, limiting the cleavage of other bonds required higher vibrational energy (i.e. cross-ring cleavages). Higher energy CID (also referred as HCD), used in a Q-TOF or Orbitrap instrument, provides beam-type collisions for kinetic control of bond rupture generating more abundant cross-ring cleavage useful for determination of linkage and modification positions. In either case, sulfation loss during fragmentations is the major challenge to use CID for structural characterization of highly sulfated GAG oligosaccharides, because the sulfate bonds are much more labile than glycosidic bond resulting in sulfate loss dominated product ions. Although CID multistage tandem mass spectrometry (MS^n) has been successfully used for distinguishing isomeric diasaccharides[91], sulfate loss makes it impossible to obtain accurate sequencing for larger oligosaccharides due to increased structural complexity. Attempts have been made to address this problem, where studies have shown that the sulfate loss can be minimized by lowering the degree of protonation through charge state manipulation and/or proton-sodium exchange[92,93]. However, delicate optimization of buffer and ionization conditions may be required for each oligosaccharide, limiting their application for high-throughput on-line LC-MS/MS analysis of complex mixtures.

Electron based dissociation is an alternation of CID offering milder fragmenting process, which can provide informative glycosidic bond and cross-ring fragmentations while minimizing

the sulfate loss. Both electron detachment dissociation (EDD)[94] and negative electron transfer dissociation (NETD)[95] have been implemented for GAG analysis, where fully structural characterizations were obtained for single-component samples with capability of distinguishing sulfation positions as well as epimerizations. EDD has only been reported to work well for oligosaccharides with relatively low sulfation levels[96], due to the reduced electron/GAG anion collision cross-section, and therefore lower fragmentation efficiency, caused by increased negative charge density. On the other hand, NETD is initiated by electron transfer between GAG anion and reagent radical cation, where charge repulsion between analyte and reagent is not a problem, making NETD a better candidate than EDD for fragmenting oligosaccharides with higher sulfation levels. While current work shows great promise in the application of these dissociation methods for detailed structural characterization, they are generally not amenable to on-line LC separation because of their relative low speed comparing to CID, which limits their application on high-throughput analysis of complex mixtures.

On-line LC-MS/MS of GAGs

Because LC separation and MS/MS analysis are largely involved with purification of GAG oligosaccharides and detailed structural analysis respectively, a successful combination of these two techniques have significant advantages for complex mixtures analysis providing high efficiency and high sensitivity. Research efforts have been made to develop advanced LC-MS based analytical methods to meet the challenges involved in different subjects including disaccharide analysis and oligosaccharide profiling.

Disaccharide analysis is a general and useful analytical method for GAG population, involves exhaustive depolymarizaion of intact heparin/HS into disaccharides and subsequent LC-MS for qualitative and quantitative analysis of the resulting disaccharides. RPIP-LC coupled

with ESI-MS has been successfully served for such purpose[77], whereas its application on larger GAG oligosaccharide has not yet been proved but usually required in order to study specific protein-GAG binding interactions. Oligosaccharide profiling, on the other hand, can provide compositions and abundances of oligosaccharides with moderate lengths by performing partial depolymerization and on-line LC-MS analysis. HILIC LC-MS are largely used for this purpose[74-76], where oligosaccharides can be separated based on their size and sulfation/acetylation degrees, with composition information (such as length, sulfation degree and number of acetyl groups) provided by accurate mass measurement. MS/MS sequencing is usually not involved in HILIC LC-MS of oligosaccharide profiling, due to couple of challenges, as discussed, including low resolution of HILIC for isomers, sulfate loss involved in CID and chromatography-incompatible timescale of current electron based dissociations.

While a number of hyphenated techniques have shown their advantages in addressing one or some of the obstacles, it remains challenging but crucial to develop LC-MS/MS methods that can achieve both on-line separation and accurate detailed identification of oligosaccharides for complex mixture analysis that usually involved with real samples. The work presented in this dissertation provide an alternative strategy, where a chemical derivatization scheme (as illustrated in **Figure 1.4**) was developed that allows the application of the most widely used LC-MS/MS platform for separation and sequencing of GAG oligosaccharides.

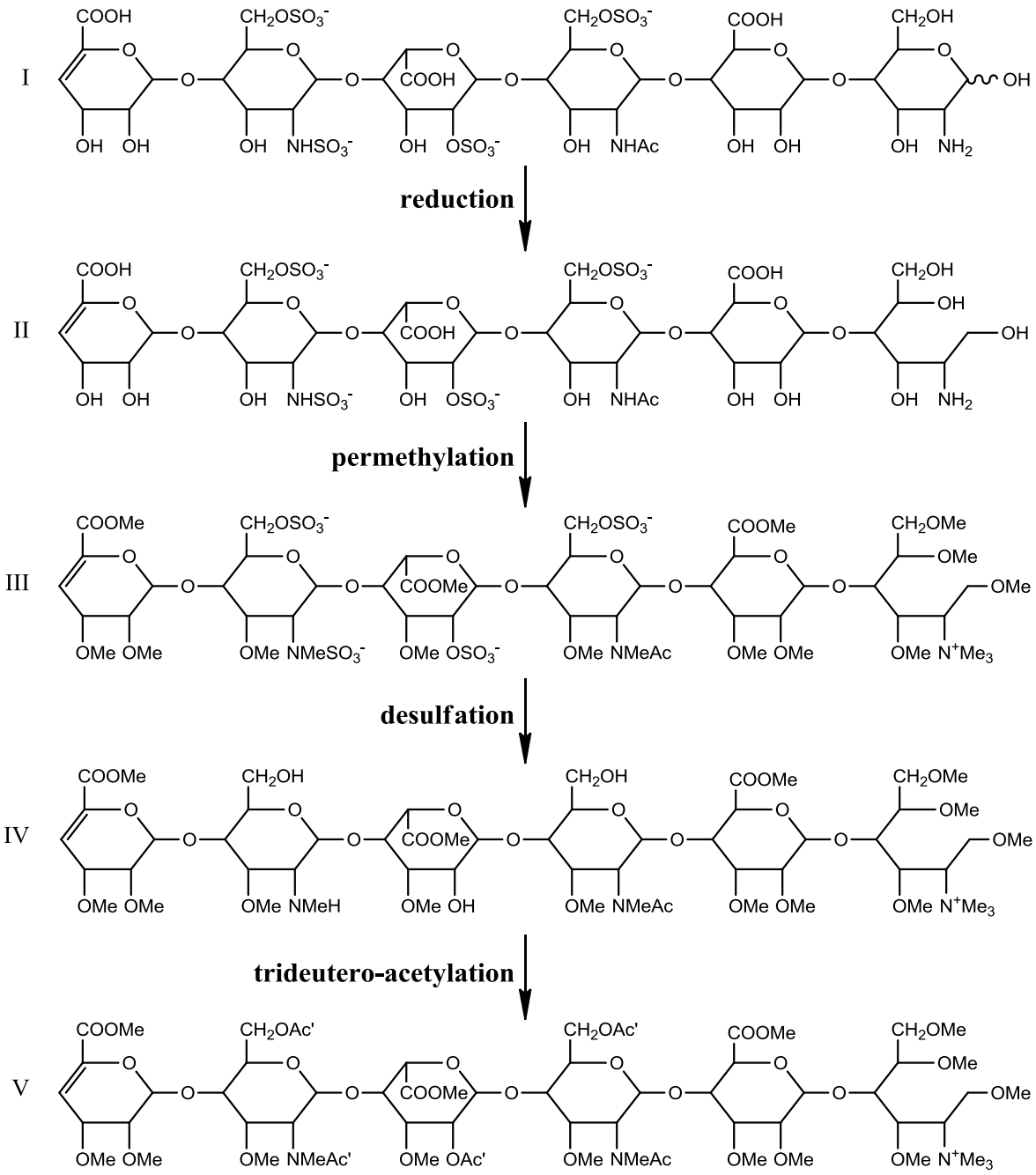


Figure 1.4 Representative schemes of sequential chemical derivatization. Structure of HS oligosaccharides: ΔHexA-GlcNS6S-IdoA2S-GlcNAc6S-GlcA-GlcN and its product after each derivatization step are illustrated. Sequential derivatizations include reduction, permethylation, desulfation and trideutero-acetylation. (Me: -CH_3, Ac: -$COOCH_3$, Ac': -$COOCD_3$)

Chemical derivatization for LC-MS/MS analysis of GAG oligosaccharides

Chemical derivatizations are widely used in structural analysis of glycans to improve ionization and fragmentation in MS analysis, as well as to alternate chromatographic properties[97]. The most commonly used methods involve derivatization of hydroxyl groups (i.e. permethylation and peracetylation) and the reducing terminal (i.e. reduction and reductive amination). While chemical derivatization for structural analysis of GAG oligosaccharides is not as popular as for other carbohydrates, such as N- and O-linked glycans, there have been studies reported on various derivatizations of intact GAG or GAG oligosaccharides, including carboxyl reduction[98], permethylation[99], desulfation[100], peracetylation[101] and reductive amination[77].

Permethylation is one of the most well-established derivatization methods used in MS analysis of N- and O-linked glycans, which greatly increases hydrophobicity of glycans by replacing all hydrophilic hydroxyl groups with methyl groups, and results in much more efficient ionization and more informative fragmentation. Typical permethylation protocols usually use methyl iodide as the methylation reagent together with different choices of strong base for catalysis and organic solvent. Good results have been obtained for permethylation of GAG using the Hakomori methods[102], but the difficulty involved in access and handling of the reagents limited its wide application. On the other hand, the modification of the Ciucanu and Kerek[103] sodium hydroxide method by Anumula and Taylor[104] provides a simple, rapid and effective way for permethylation, which uses dimethyl sulfoxide (DMSO) as solvent, sodium hydroxide DMSO solution as catalyst, methyl iodide as reagent. With this method, the solubility of highly charged GAG oligosaccharides in DMSO became the biggest concern regarding to the effectiveness of the derivatization. Heiss *et al* has reported that converting heparin disaccharides

into their triethylamine salts greatly increases their solubility in DMSO[105], allowing efficient permethylation with the sodium hydroxide method.

During the permethylation of GAGs, not only are the hydroxyl groups are methylated, the hydrogen on the free amine or modified amine group of hexamine residues are also replaced by methyl groups and the carboxyl group of uronic acid residues are converted into its methyl ester (refer to **Figure 1.4**). Sulfate groups are left underivatized during the permethylation. The amounts of sodium hydroxide and methyl iodide are carefully controlled to avoid peeling reactions during the permethylation and saponification of the methyl ester groups after adding water for permethylation quenching[105]. Nevertheless, β–elimination at the uronic acid residues is much more favorable than that for non-acidic residues under basic conditions, like permethylation, giving enzymatic degradation like by-products[99]. In order to overcome this problem, carboxyl reduction may be used to reduce uronic acid into more stable hexose to avoid β–elimination[98].

Desulfation is one of the mostly studied chemical derivatizations of GAG, which are usually used for structural determination or study of the relationship between global sulfation sites (N-/2O-/6O-/3O-) and their biological functions[106]. Various GAG desulfation methods have been reported, among which, acid-catalyzed desulfation and solvolytic desulfation are the two mostly used. While acid-catalyzed desulfation only requires a simple procedure, where sulfate esters are usually methanolyzed by adding methanolic hydrogen chloride and reacting at room temperature[107], it also causes glycosidic bond cleavage due to the close methanolysis rate of glycosidic linkages and sulfate esters. Solvolytic desulfation, on the other hand, does not present obvious sample loss in terms of glycosidic bond cleavage, which makes them the method of choice in most cases. Solvolytic desulfation usually involves with heating pyridinium salts of

GAG in DMSO containing water or methanol (5-10%), with reported reaction rates of N-sulfation>>6-O-sulfation>2-O-sulfation for heparin/HS and 6-O-sulfation>4-O-sulfation>2-O-sulfation for CS/DS[108]. Therefore, selective desulfation may also be obtained by controlling the reaction temperature and/or time[109].

Reduction and peracetylation are also common derivatizations that have been used for glycan analysis. Oligosaccharides with free end are usually reduced by sodium borohydride[105] in order to eliminate anomeric configurations at the reducing end that may cause peak broaden or splitting in LC separation. Peracetylation is an alternative derivatization of permethylation, which replaces hydroxyl groups with acetyl groups rather than methyl groups by reacting oligosaccharides with acetic anhydride[110]. Unlike permethylation, peracetylation of GAG oligosaccharides does not derivatize carboxyl groups or acetylated amine groups, but can acetylate free or mono-methylated amine groups (refer to **Figure 1.4**).

There were studies reported in 1980s using sequential chemical derivatization for structural sequencing of heparin oligosaccharides with fine structure by FAB MS[102], and combined with chemical depolymerization for monosaccharide analysis of heparin polysaccharides by chemical ionization MS coupled with gas-liquid chromatography[101]. In **Chapter 2,** a modified and repurposing sequential chemical derivatization scheme including reduction, permethylation, desulfation and peracetylation is described**,** by which native CS oligosaccharides are modified prior to RPLC-CID MS^n analysis. This methodology consists essentially of replacing the labile sulfate groups with stable acetyl groups while maintaining the positional information of the original sulfation patterns. After derivatization, CS oligosaccharides are not only well retained on standard reversed phase columns without ion pairing, but also efficiently ionized on positive ion mode and fragmented by CID. Therefore, for the first time, on-

line separation and accurate identification of CS oligosaccharide (larger than disaccharides) mixtures containing sulfation positional isomers are successfully achieved on standard LC-MS/MS platform.

After initial success on CS oligosaccharides, we further extend our chemical derivatization scheme to another GAG, heparin/HS, which represent the most structural complexity and therefore the analysis challenge of GAG family. For heparin/HS oligosaccharides, hexadeutero acetyl anhydride is used in the peracetylation step to allow for distinguishing native N-acetylation from peracetylation during the derivatization by mass difference. In **Chapter 3,** we demonstrate that the sequential chemical derivatization strategy allows differentiation between all twelve common HS disaccharides and successful sequencing synthetic HS oligosaccharides up to dodecamers based solely on glycosidic bond cleavages from a single MS/MS experiment.

With the availability of a library of synthetic HS tetrasaccharides varying in sulfation degree, position and uronic acid epimerization, the RPLC selectivity and MS/MS differentiation of sulfation positional isomers and epimers are systematically studied as described in **Chapter 4**. The improved separation, coupled with the CID MS/MS of derivatized oligosaccharides, makes it possible for complete sequencing of a synthetic complex mixture of twenty-one HS tetrasaccharides. Furthermore, it has also been demonstrated that, after chemical derivatization, the combination of molecular weight, chromatography behavior and fragmentation pattern allows for the first complete sequencing of a mixture of HS tetrasaccharides derived from enzymatic depolymerization of native HS.

Roundabout 1 (Robo1), a heparin/HS binding protein, is the receptor for secreted axon guidance molecule Slits, through which heparin/HS mediates the Slit-Robo signaling pathways

that involved in neuronal development and angiogenesis[111]. In order to study the Robo1-binding HS sequences, our derivatization strategy coupled with data-dependent LC-MS/MS are applied to the affinity purified HS octasaccharides as presented in **Chapter 5** and data interpretation are facilitated with the help of our in-house developed program.

References

1. Varki, A.; Cummings, R.; Esko, J. D.; Freeze, H.; Hart, G. W.; Marth, J. *Essentials of Glycobiology*; Cold Spring Harbor Laboratories Press: New York, 1999.

2. Brooks, S. A.; Schumacher, U.; Dweck, M. V. *Functional and Molecular Glycobiology*; Taylor & Francis Ltd. London: London, 2002.

3. Brown, J. R.; Crawford, B. E.; Esko, J. D. *Critical reviews in biochemistry and molecular biology* **2007**, *42*, 481-515.

4. DeAngelis, P. L. *The Anatomical record* **2002**, *268*, 317-326.

5. Yamada, S.; Sugahara, K.; Ozbek, S. *Communicative & integrative biology* **2011**, *4*, 150-158.

6. Bulow, H. E.; Hobert, O. *Annual review of cell and developmental biology* **2006**, *22*, 375-407.

7. Gandhi, N. S.; Mancera, R. L. *Chemical biology & drug design* **2008**, *72*, 455-482.

8. Kresse, H.; Schonherr, E. *Journal of cellular physiology* **2001**, *189*, 266-274.

9. DeAngelis, P. L. *Applied microbiology and biotechnology* **2012**, *94*, 295-305.

10. Victor, X. V.; Nguyen, T. K.; Ethirajan, M.; Tran, V. M.; Nguyen, K. V.; Kuberan, B. *The Journal of biological chemistry* **2009**, *284*, 25842-25853.

11. Zhang, L. *Progress in molecular biology and translational science* **2010**, *93*, 1-17.

12. Suarez, E. R.; Nohara, A. S.; Mataveli, F. D.; de Matos, L. L.; Nader, H. B.; Pinhal, M. A. *Growth Factors* **2007**, *25*, 50-59.

13. Carlsson, P.; Kjellen, L. *Handbook of experimental pharmacology* **2012**, 23-41.

14. Sugahara, K.; Kitagawa, H. *IUBMB life* **2002**, *54*, 163-175.

15. Multhaupt, H. A.; Couchman, J. R. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **2012**, *60*, 908-915.

16. Malmstrom, A.; Bartolini, B.; Thelin, M. A.; Pacheco, B.; Maccarana, M. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **2012**, *60*, 916-925.

17. Mikami, T.; Kitagawa, H. *Biochimica et biophysica acta* **2013**, *1830*, 4719-4733.

18. Silbert, J. E.; Sugumaran, G. *IUBMB life* **2002**, *54*, 177-186.

19. Funderburgh, J. L. *Glycobiology* **2000**, *10*, 951-958.

20. Funderburgh, J. L. *IUBMB life* **2002**, *54*, 187-194.

21. Prehm, P. *The Biochemical journal* **2006**, *398*, 469-473.

22. Viola, M.; Vigetti, D.; Genasetti, A.; Rizzi, M.; Karousou, E.; Moretto, P.; Clerici, M.; Bartolini, B.; Pallotti, F.; De Luca, G.; Passi, A. *Connective tissue research* **2008**, *49*, 111-114.

23. Grobe, K.; Ledin, J.; Ringvall, M.; Holmborn, K.; Forsberg, E.; Esko, J. D.; Kjellen, L. *Biochimica et biophysica acta* **2002**, *1573*, 209-215.

24. Hagner-Mcwhirter, A.; Lindahl, U.; Li, J. *The Biochemical journal* **2000**, *347 Pt 1*, 69-75.

25. Rong, J.; Habuchi, H.; Kimata, K.; Lindahl, U.; Kusche-Gullberg, M. *Biochemistry* **2001**, *40*, 5548-5555.

26. Habuchi, H.; Tanaka, M.; Habuchi, O.; Yoshida, K.; Suzuki, H.; Ban, K.; Kimata, K. *The Journal of biological chemistry* **2000**, *275*, 2859-2868.

27. Liu, J.; Shworak, N. W.; Sinay, P.; Schwartz, J. J.; Zhang, L.; Fritze, L. M.; Rosenberg, R. D. *The Journal of biological chemistry* **1999**, *274*, 5185-5192.

28. Esko, J. D.; Lindahl, U. *The Journal of clinical investigation* **2001**, *108*, 169-173.

29. Kreuger, J.; Kjellen, L. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **2012**, *60*, 898-907.

30. Jacobsson, I.; Lindahl, U.; Jensen, J. W.; Roden, L.; Prihar, H.; Feingold, D. S. *The Journal of biological chemistry* **1984**, *259*, 1056-1063.

31. Xu, D.; Moon, A. F.; Song, D.; Pedersen, L. C.; Liu, J. *Nature chemical biology* **2008**, *4*, 200-202.

32. Rapraeger, A. C. *Methods in cell biology* **2002**, *69*, 83-109.

33. Peysselon, F.; Ricard-Blum, S. *Matrix biology : journal of the International Society for Matrix Biology* **2013**.

34. Lindahl, U.; Li, J. P. *The international journal of biochemistry & cell biology* **2009**, *276*, 105-159.

35. Bishop, J. R.; Schuksz, M.; Esko, J. D. *Nature* **2007**, *446*, 1030-1037.

36. Linhardt, R. J. *Journal of medicinal chemistry* **2003**, *46*, 2551-2564.

37. Rabenstein, D. L. *Natural product reports* **2002**, *19*, 312-331.

38. Sasisekharan, R.; Shriver, Z.; Venkataraman, G.; Narayanasami, U. *Nature reviews. Cancer* **2002**, *2*, 521-528.

39. Li, J. P.; Vlodavsky, I. *Thrombosis and haemostasis* **2009**, *102*, 823-828.

40. Lindahl, B.; Eriksson, L.; Lindahl, U. *Biochem J* **1995**, *306 (Pt 1)*, 177-184.

41. Shukla, D.; Liu, J.; Blaiklock, P.; Shworak, N. W.; Bai, X.; Esko, J. D.; Cohen, G. H.; Eisenberg, R. J.; Rosenberg, R. D.; Spear, P. G. *Cell* **1999**, *99*, 13-22.

42. Lewis, E. J.; Xu, X. *Diabetes care* **2008**, *31 Suppl 2*, S202-207.

43. Vlodavsky, I.; Ilan, N.; Nadir, Y.; Brenner, B.; Katz, B. Z.; Naggi, A.; Torri, G.; Casu, B.; Sasisekharan, R. *Thrombosis research* **2007**, *120 Suppl 2*, S112-120.

44. Tomatsu, S.; Gutierrez, M. A.; Ishimaru, T.; Pena, O. M.; Montano, A. M.; Maeda, H.; Velez-Castrillon, S.; Nishioka, T.; Fachel, A. A.; Cooper, A.; Thornley, M.; Wraith, E.; Barrera, L. A.; Laybauer, L. S.; Giugliani, R.; Schwartz, I. V.; Frenking, G. S.; Beck, M.; Kircher, S. G.; Paschke, E.; Yamaguchi, S.; Ullrich, K.; Isogai, K.; Suzuki, Y.; Orii, T.; Noguchi, A. *Journal of inherited metabolic disease* **2005**, *28*, 743-757.

45. Kreuger, J.; Spillmann, D.; Li, J. P.; Lindahl, U. *Journal of cellular biochemistry* **2006**, *174*, 323-327.

46. Esko, J. D.; Selleck, S. B. *Annual review of biochemistry* **2002**, *71*, 435-471.

47. Petitou, M.; Casu, B.; Lindahl, U. *Biochimie* **2003**, *85*, 83-89.

48. de Agostini, A. I.; Watkins, S. C.; Slayter, H. S.; Youssoufian, H.; Rosenberg, R. D. *The Journal of cell biology* **1990**, *111*, 1293-1304.

49. de Agostini, A. I.; Dong, J. C.; de Vantery Arrighi, C.; Ramus, M. A.; Dentand-Quadri, I.; Thalmann, S.; Ventura, P.; Ibecheole, V.; Monge, F.; Fischer, A. M.; HajMohammadi, S.; Shworak, N. W.; Zhang, L.; Zhang, Z.; Linhardt, R. J. *J Biol Chem* **2008**, *283*, 28115-28124.

50. Raman, R.; Venkataraman, G.; Ernst, S.; Sasisekharan, V.; Sasisekharan, R. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 2357-2362.

51. Malavaki, C.; Mizumoto, S.; Karamanos, N.; Sugahara, K. *Connect Tissue Res* **2008**, *49*, 133-139.

52. Nandini, C. D.; Sugahara, K. *Adv Pharmacol* **2006**, *53*, 253-279.

53. Theocharis, A. D.; Karamanos, N. K. *Atherosclerosis* **2002**, *165*, 221-230.

54. Pothacharoen, P.; Kalayanamitra, K.; Deepa, S. S.; Fukui, S.; Hattori, T.; Fukushima, N.; Hardingham, T.; Kongtawelert, P.; Sugahara, K. *J Biol Chem* **2007**, *282*, 35232-35246.

55. Maimone, M. M.; Tollefsen, D. M. *J Biol Chem* **1990**, *265*, 18263-18271.

56. Petitou, M.; van Boeckel, C. A. *Angew Chem Int Ed Engl* **2004**, *43*, 3118-3133.

57. de Ruijter, J.; de Ru, M. H.; Wagemans, T.; Ijlst, L.; Lund, A. M.; Orchard, P. J.; Schaefer, G. B.; Wijburg, F. A.; van Vlies, N. *Molecular genetics and metabolism* **2012**, *107*, 705-710.

58. Guerrini, M.; Beccati, D.; Shriver, Z.; Naggi, A.; Viswanathan, K.; Bisio, A.; Capila, I.; Lansing, J. C.; Guglieri, S.; Fraser, B.; Al-Hakim, A.; Gunay, N. S.; Zhang, Z.; Robinson, L.; Buhse, L.; Nasr, M.; Woodcock, J.; Langer, R.; Venkataraman, G.; Linhardt, R. J.; Casu, B.; Torri, G.; Sasisekharan, R. *Nature biotechnology* **2008**, *26*, 669-675.

59. Chai, W.; Hounsell, E. F.; Bauer, C. J.; Lawson, A. M. *Carbohydrate research* **1995**, *269*, 139-156.

60. Zhang, Z.; McCallum, S. A.; Xie, J.; Nieto, L.; Corzana, F.; Jimenez-Barbero, J.; Chen, M.; Liu, J.; Linhardt, R. J. *Journal of the American Chemical Society* **2008**, *130*, 12998-13007.

61. Silipo, A.; Zhang, Z.; Canada, F. J.; Molinaro, A.; Linhardt, R. J.; Jimenez-Barbero, J. *Chembiochem : a European journal of chemical biology* **2008**, *9*, 240-252.

62. Zaia, J. *Molecular & cellular proteomics : MCP* **2013**, *12*, 885-892.

63. Yang, B.; Solakyildirim, K.; Chang, Y.; Linhardt, R. J. *Analytical and bioanalytical chemistry* **2011**, *399*, 541-557.

64. Li, L.; Ly, M.; Linhardt, R. J. *Molecular bioSystems* **2012**, *8*, 1613-1625.

65. Murphy, K. J.; Merry, C. L.; Lyon, M.; Thompson, J. E.; Roberts, I. S.; Gallagher, J. T. *J Biol Chem* **2004**, *279*, 27239-27245.

66. Mardiguian, J. US, 1984; Vol. 4440926 A.

67. Shively, J. E.; Conrad, H. E. *Biochemistry* **1976**, *15*, 3932-3942.

68. Zaia, J. *Mass spectrometry reviews* **2009**, *28*, 254-272.

69. Ziegler, A.; Zaia, J. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2006**, *837*, 76-86.

70. Zaia, J.; Costello, C. E. *Analytical chemistry* **2001**, *73*, 233-239.

71. Henriksen, J.; Ringborg, L. H.; Roepstorrf, P. *Journal of mass spectrometry : JMS* **2004**, *39*, 1305-1312.

72. Imanari, T.; Toida, T.; Koshiishi, I.; Toyoda, H. *Journal of chromatography. A* **1996**, *720*, 275-293.

73. Pervin, A.; Gallo, C.; Jandik, K. A.; Han, X. J.; Linhardt, R. J. *Glycobiology* **1995**, *5*, 83-95.

74. Hitchcock, A. M.; Yates, K. E.; Costello, C. E.; Zaia, J. *Proteomics* **2008**, *8*, 1384-1397.

75. Naimy, H.; Buczek-Thomas, J. A.; Nugent, M. A.; Leymarie, N.; Zaia, J. *J Biol Chem* **2011**, *286*, 19311-19319.

76. Staples, G. O.; Bowman, M. J.; Costello, C. E.; Hitchcock, A. M.; Lau, J. M.; Leymarie, N.; Miller, C.; Naimy, H.; Shi, X.; Zaia, J. *Proteomics* **2009**, *9*, 686-695.

77. Yang, B.; Chang, Y.; Weyers, A. M.; Sterner, E.; Linhardt, R. J. *Journal of chromatography. A* **2012**, *1225*, 91-98.

78. Estrella, R. P.; Whitelock, J. M.; Packer, N. H.; Karlsson, N. G. *Analytical chemistry* **2007**, *79*, 3597-3606.

79. Ruiz-Calero, V.; Puignou, L.; Galceran, M. T. *Journal of chromatography. A* **1998**, *828*, 497-508.

80. Desai, U. R.; Wang, H.; Ampofo, S. A.; Linhardt, R. J. *Analytical biochemistry* **1993**, *213*, 120-127.

81. Patel, R. P.; Narkowicz, C.; Hutchinson, J. P.; Hilder, E. F.; Jacobson, G. A. *Journal of pharmaceutical and biomedical analysis* **2008**, *46*, 30-35.

82. Kailemia, M. J.; Park, M.; Kaplan, D. A.; Venot, A.; Boons, G. J.; Li, L.; Linhardt, R. J.; Amster, I. J. *Journal of the American Society for Mass Spectrometry* **2014**, *25*, 258-268.

83. Seo, Y.; Andaya, A.; Bleiholder, C.; Leary, J. A. *J Am Chem Soc* **2013**, *135*, 4325-4332.

84. Loo, B. M.; Kreuger, J.; Jalkanen, M.; Lindahl, U.; Salmivirta, M. *J Biol Chem* **2001**, *276*, 16868-16876.

85. Schenauer, M. R.; Yu, Y.; Sweeney, M. D.; Leary, J. A. *J Biol Chem* **2007**, *282*, 25182-25188.

86. Naimy, H.; Leymarie, N.; Zaia, J. *Biochemistry-Us* **2010**, *49*, 3743-3752.

87. Reinhold, V. N.; Carr, S. A.; Green, B. N.; Petitou, M.; Choay, J.; Sinay, P. *Carbohydrate research* **1987**, *161*, 305-313.

88. Bohme, J.; Anderegg, U.; Nimptsch, A.; Nimptsch, K.; Hacker, M.; Schulz-Siegmund, M.; Huster, D.; Schiller, J. *Analytical biochemistry* **2012**, *421*, 791-793.

89. Laremore, T. N.; Murugesan, S.; Park, T. J.; Avci, F. Y.; Zagorevski, D. V.; Linhardt, R. J. *Analytical chemistry* **2006**, *78*, 1774-1779.

90. Zaia, J. *Mass spectrometry reviews* **2004**, *23*, 161-227.

91. Saad, O. M.; Leary, J. A. *Analytical chemistry* **2003**, *75*, 2985-2995.

92. Kailemia, M. J.; Li, L.; Ly, M.; Linhardt, R. J.; Amster, I. J. *Analytical chemistry* **2012**, *84*, 5475-5478.

93. Zaia, J.; Costello, C. E. *Analytical chemistry* **2003**, *75*, 2445-2455.

94. Leach, F. E., 3rd; Arungundram, S.; Al-Mafraji, K.; Venot, A.; Boons, G. J.; Amster, I. J. *International journal of mass spectrometry* **2012**, *330-332*, 152-159.

95. Leach, F. E., 3rd; Wolff, J. J.; Xiao, Z.; Ly, M.; Laremore, T. N.; Arungundram, S.; Al-Mafraji, K.; Venot, A.; Boons, G. J.; Linhardt, R. J.; Amster, I. J. *Eur J Mass Spectrom (Chichester, Eng)* **2011**, *17*, 167-176.

96. Leach, F. E., 3rd; Xiao, Z.; Laremore, T. N.; Linhardt, R. J.; Amster, I. J. *International journal of mass spectrometry* **2011**, *308*, 253-259.

97. Harvey, D. J. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2011**, *879*, 1196-1225.

98. Taylor, R. L.; Conrad, H. E. *Biochemistry* **1972**, *11*, 1383-1388.

99. Ciucanu, I. *Analytica chimica acta* **2006**, *576*, 147-155.

100. Takano, R. *Trends in Glycoscience and Glycotechnology* **2002**, *14*, 343-351.

101. Barker, S. A.; Hurst, R. E.; Settine, J.; Fish, F. P.; Settine, R. L. *Carbohydr Res* **1984**, *125*, 291-300.

102. Dell, A.; Carman, N. H.; Tiller, P. R.; Thomas-Oates, J. E. *Biomedical & environmental mass spectrometry* **1988**, *16*, 19-24.

103. Ciucanu, I.; Kerek, F. *Carbohydrate research* **1984**, *131*, 209-217.

104. Anumula, K. R.; Taylor, P. B. *Analytical biochemistry* **1992**, *203*, 101-108.

105. Heiss, C.; Wang, Z.; Azadi, P. *Rapid communications in mass spectrometry : RCM* **2011**, *25*, 774-778.

106. Toida, T.; Chaidedgumjorn, A.; Linhardt, R. J. *Trends in Glycoscience and Glycotechnology* **2003**, *15*, 29-46.

107. Taguchi, T.; Iwasaki, M.; Muto, Y.; Kitajima, K.; Inoue, S.; Khoo, K. H.; Morris, H. R.; Dell, A.; Inoue, Y. *European journal of biochemistry / FEBS* **1996**, *238*, 357-367.

108. Nagasawa, K.; Inoue, Y.; Kamata, T. *Carbohydrate research* **1977**, *58*, 47-55.

109. Naggi, A.; Casu, B.; Perez, M.; Torri, G.; Cassinelli, G.; Penco, S.; Pisano, C.; Giannini, G.; Ishai-Michaeli, R.; Vlodavsky, I. *J Biol Chem* **2005**, *280*, 12103-12113.

110. Gunner, S. W.; Jones, J. K. N.; Perry, M. B. *Can. J. Chem.* **1961**, *89*, 1892-1899.

111. Carmeliet, P.; Tessier-Lavigne, M. *Nature* **2005**, *436*, 193-200.

CHAPTER 2

LC-MS^n ANALYSIS OF ISOMERIC CHONDROITIN SULFATE OLIGOSACCHARIDES USING A CHEMICAL DERIVATIZATION STRATEGY[1]

[1] Huang, R., Pomin, V.H., Sharp, J.S., *J. Am. Soc. Mass Sepectrom.*, **2011**, *22*, 1577-1587. Reprinted here with permission of publisher.

Abstract

Improved methods for structural analyses of glycosaminoglycans (GAGs) are required to understand their functional roles in various biological processes. Major challenges in structural characterization of complex GAG oligosaccharides using liquid chromatography-mass spectrometry (LC-MS) include the accurate determination of the patterns of sulfation due to gas-phase losses of the sulfate groups upon collisional activation and inefficient on-line separation of positional sulfation isomers prior to MS/MS analyses. Here, a sequential chemical derivatization procedure including permethylation, desulfation and acetylation was demonstrated to enable both on-line LC separation of isomeric mixtures of chondroitin sulfate (CS) oligosaccharides and accurate determination of sites of sulfation by MS^n. The derivatized oligosaccharides have sulfate groups replaced with acetyl groups, which are sufficiently stable to survive MS^n fragmentation and reflect the original sulfation patterns. A standard reversed-phase LC-MS system with a capillary C18 column was used for separation, and MS^n experiments using collision-induced dissociation (CID) were performed. Our results indicate that the combination of this derivatization strategy and MS^n methodology enables accurate identification of the sulfation isomers of CS hexasaccharides with either saturated or unsaturated non-reducing ends. Moreover, derivatized CS hexasaccharide isomer mixtures become separable by LC-MS method due to different positions of acetyl modifications.

Introduction

Glycosaminoglycans (GAGs) are linear polysaccharides often linked to the protein component of proteoglycans, which can be divided into classes consisting of different repeating disaccharide units: hyaluronan, chondroitin sulfate (CS), dermatan sulfate, heparan sulfate, heparin and keratan sulfate[1]. The repeating disaccharide unit of CS is comprised of glucuronic

acid (GlcA) linked to N-acetylgalactosamine (GalNAc), where sulfations may occur at 4- and/or 6-positions of GalNAc, and/or 2-position of GlcA, yielding five common CS units with different sulfation patterns: [GlcA-GalNAc(4S)] (A-unit), [GlcA-GalNAc(6S)] (C-unit), [GlcA(2S)-GalNAc(6S)] (D-unit), [GlcA-GalNAc(4S6S)] (E-unit) and the non-sulfated unit [GlcA-GalNAc] (O-unit), along with other rare types[2]. The mono-sulfated units, A- and C-units, are commonly found in great proportions in typical CS sources such as chondroitin sulfate-A (CS-A) and chondroitin sulfate-C (CS-C) respectively, with small proportions of other units also present.

Interactions between GAGs and functional proteins are involved in many biological processes including cell signaling, migration, and tissue development[3-7]. It is believed that particular saccharide domains with specific sulfation patterns play crucial rules in these biological functions. For example, studies have shown that significant structural changes in CS chains, caused by alterations in their expressions, lead to functional changes of CS-proteoglycans involved in diseases like atherosclerosis and cancer[8-10]. Therefore, structural investigation of longer biologically active oligosaccharides rather than just disaccharide compositional analyses is quite necessary for further understanding functional domains of GAGs[11,12]. Refined structural characterization of GAG oligosaccharides, especially the identification of their sulfation patterns, is imperative to understand the structure-function relations of GAGs in complexes with target proteins.

The high degree of heterogeneity in GAG polymers caused by the variety of chain lengths and diverse sulfation patterns makes their structural determination a very challenging task. Various techniques have been applied for this purpose including nuclear magnetic resonance (NMR) and mass spectrometry (MS). While NMR techniques have been demonstrated as a powerful tool for detailed structural information[13], their application is typically limited to

highly enriched, relatively abundant samples that are difficult to obtain from many GAG preparations and often exhibits difficulties with spectral complexity. Thus, MS methods are generally more desirable for applications involving low sample amounts and/or heterogeneous samples due to their higher sensitivity and easier compatibility with on-line liquid chromatography separations. The possibility to isolate and structurally interrogate analytes from complex mixtures using tandem MS techniques is an additional important advantage. The major challenges involved in the fine structural characterization of GAG oligosaccharides using MS include the difficulty to obtain accurate identification of the sulfation patterns due to gas-phase losses of the sulfate groups upon collisional activation, and the inefficiency of on-line separation prior to MS/MS analyses of sulfation isomers having identical elemental compositions and saccharide sequences but differing solely in the position of sulfation.

Multiple ionization techniques have been used for structural analyses of GAGs including fast-atom bombardment (FAB)[14,15], matrix-assisted laser desorption-ionization (MALDI)[16,17], as well as electrospray ionization (ESI) coupled with different tandem MS techniques[18-20]. ESI is preferable for its gentle ionization and has been used for structural analysis of GAGs by several groups. Zaia and co-workers have shown that the loss of labile sulfate groups during collision-induced dissociation (CID) can be minimized by deprotonation of these groups, using a combination of charge state manipulation and metal ion adduction. Their tandem mass spectra are often useful in differentiation of sulfation isomers of CS in enriched samples, and even in differentiating DS-like oligosaccharides from CS oligosaccharides with additional MS^3 stage[21,22]. However, unique optimized conditions are required for each different GAG oligosaccharide. Sulfate losses are still quite common, making mixture analysis highly problematic due to the presence of both sulfated and non-sulfated versions of many product ions. Sufficient

fragmentation information for definitive assignment of sulfation sites is not always available, even under optimized conditions, which makes these methods often insufficient for complete sulfation position analysis of GAG oligosaccharide mixtures with high degrees of heterogeneity.

Another way to overcome the limitations of sulfation losses during CID fragmentations in structural analysis of longer oligosaccharides is disaccharide sequencing, in which enzymatic digestions followed by sequential stages of mass spectrometry analyses (MS^n) are employed for disaccharide compositional analysis as well as disaccharide sequencing[19,23]. Sequencing of heparan sulfate hexasaccharides has been successfully achieved by using this methodology coupled with a heparin oligosaccharide sequencing tool for automatic MS^n data interpretation[24]. However, pre-purified oligosaccharides or samples with very few contaminants are still required in order to generate promising structural information. Digestion with multiple enzymes and quantification of the resulting disaccharides mixtures are required prior to this MS-coupled sequencing, increasing the required amount of sample handling.

In addition to CID, electron-detachment dissociation (EDD) has also been applied to the analysis of GAG oligomers. EDD has been demonstrated to produce much higher abundances of cross-ring cleavages and able to distinguish glucuronic acid from iduronic acid based on diagnostic product-ions[25]. However, this dissociation technique also requires relatively high levels of purity of the analytes for accurate structural assignment, and the time-scale for EDD makes coupling with on-line LC separations difficult, limiting the extension of this technique to complex mixtures of oligomers.

On-line LC method coupled with MS has been used to simplify mixtures of GAG oligosaccharides in MS^n analysis while limiting sample losses[26]. Those LC techniques include hydrophilic interaction (HILIC), size-exclusion (SEC), porous graphitized carbon (PGC), and

ion-pairing reversed-phase (IPRP) chromatography. HILIC and SEC have been shown to effectively separate heparin oligosaccharides based on size, and HILIC is also capable of separation based on sulfation and acetylation content. However, both are incapable of separation of sulfation isomers, limiting their effectiveness in analysis of GAG oligosaccharide mixtures[27-29]. IPRP and PGC gives superior retention of sulfated GAGs than using standard C18 reversed-phase chromatography, and have been shown to be capable in separation of isomeric oligosaccharides with small size and less sulfation such as heparan disaccharides, CS disaccharides or mono-sulfated hexasaccharides including the conserved linkage tetrasaccharide where the GAG chain is linked in proteoglycans[30-33]. However, little success has been demonstrated in analyses of larger oligosaccharides with sulfation patterns more complex than one sulfate per disaccharide unit.

While all these techniques have their own advantages in GAG oligosaccharide analyses, there is still no effective methodology that can achieve accurate detailed identification of oligosaccharides in a complex mixture of GAG isomers with different sulfation patterns. Here, we present an approach involving sequential chemical derivatizations to modify native CS oligosaccharides prior to structural MS analysis enabling accurate identification of GAG components in a mixture. This methodology consists essentially of replacing the labile sulfate groups with stable acetyl groups while maintaining the positional information of the original sulfation pattern. This method enables the combination of on-line high-resolution reversed phase LC (RPLC) separation with MS^n fragmentation schemes that allow us to separate sulfation isomers and interrogate them using a step-wise, modular dissociation scheme. In our method, CS hexasaccharides were first permethylated to protect free hydroxyl and carboxyl groups. The sulfate groups were then removed, and the resulting hydroxyl groups were acetylated. The

chemically derivatized isomers were successfully separated by on-line capillary LC using a C18 column, and the original positions of sulfates were determined based on identification of sites of acetyl groups in the MS^n spectra. Purified oligosaccharides thoroughly characterized by 2D NMR were used to verify the accuracy of the derivatization/MS^n technique, and the ability of our protocol to separate and characterize complex mixtures of CS hexamers.

Experimental

Materials

Chondroitin sulfate-A sodium salt from bovine trachea, chondroitin sulfate-C sodium salt from shark cartilage, hyaluronidase from sheep testes type V, chondroitin C lyase from *Flavobacterium heparinum*, Sephadex G-15 resin (Fractionation range of dextrans <1.5 kDa) and Dowex 50WX8-100 ion-exchange resin were purchased from Sigma-Aldrich Inc. (St. Louis, MO). Bio-Gel P-10 gel resin (Fractionation range of dextrans from ~1.5 to ~20 kDa) and the polypropylene chromatographic columns (1.5 x 120 cm for size-exclusion chromatography, and 50 x 1.0 cm for desalting) were obtained from Bio-Rad Laboratories (Hercules, CA). A pre-packed Spherisorb S10 SAX column (10 x 250 mm, 5 μm) was from Waters Corporation (Milford, MA). Chondroitin disaccharide ΔDi-diSE, (sodium salt, with structure of ΔGlcA-GalNAc4S6S) was purchased from Dextra Laboratories Ltd. (Reading, UK). All other regular chemical reagents used in the chemical derivatization, section below, were purchased from Sigma-Aldrich Inc. (St. Louis, MO).

Preparation of CS Hexasaccharides

CS hexasaccharides were prepared as described previously[13]. Briefly, three enzymatically digested samples were applied separately for further chromatography purification including C lyase-digested CS-C, C lyase-digested CS-A, and hyaluronidase digested CS-A. The

depolymerized samples were subjected to size exclusion chromatography on a Bio-Gel P-10 column using 1M NaCl containing 10% ethanol as a mobile phase to obtain oligosaccharides fractions with uniform length. The fraction containing hexasaccharides was concentrated, desalted through Sephadex G-15 column using distilled water mobile phase and then lyophilized. The desalted, dried hexasaccharide fraction was then reduced using sodium borohydride as previously described[13], followed by desalting and lyophilization. The reduced hexasaccharide faction was further separated by strong anion exchange (SAX) chromatography using a linear gradient of NaCl (10 mM/min) in H_2O (pH 5.0). Major fractions were collected, desalted and dried.

Structural Determination by NMR

The purified CS hexasaccharide fractions from SAX chromatography were fully characterized by two-dimensional NMR experiments, $^{1}H/^{1}H$ double quantum filtered correlation spectroscopy (DQF-COSY), $^{1}H/^{1}H$ total correlation spectroscopy (TOCSY), and $^{1}H/^{13}C$ gradient heteronuclear single quantum coherence (gHSQC). The different patterns of sulfation were assigned primarily via $^{1}H/^{13}C$-chemical shifts of cross-peaks in the ^{13}C-gHSQC spectra that were indicative of specific sulfation positions. The ^{13}C-assignment in ^{13}C-HSQC cross-peaks were assigned by the ^{1}H-chemical shifts obtained previously from DQF-COSY or TOCSY experiments. More details about parameters in these NMR experiments had been addressed in previous work[13].

Chemical Derivatization of CS Hexasaccharides

The permethylation of sulfated CS oligosaccharides was a modified from those standard protocols used for unsulfated glycans, because the high degree of sulfation groups and carboxyl groups prevented the dissolution of CS oligosaccharides in DMSO[34]. Reduced CS

hexasaccharides were converted to triethylamine (TEA) salts by a self-packed cation-exchange column with Dowex 50W resin, followed by lyophilization. The dried TEA salts (10-50 μg) were re-suspended in 200 μL DMSO and 200 μL anhydrous suspension of NaOH in DMSO (150 μg/μL) followed by addition of 100 μL iodomethane. After 5 min vortexing and 10 min sonicating, the reaction was stopped by adding 2 mL water and sparged with nitrogen to remove iodomethane, followed by desalting using a C18 Sep-Pak cartridge (Waters Co.). The permethylated products were dried and then made as pyridinium salts using the same type of cation-exchange column and lyophilization. The pyridinium salts (10-50μg) of hexasaccharides were dissolved in 10μL DMSO containing 10% methanol and incubated for 3h at 80°C to remove the sulfate groups[35]. The desulfated product was lyophilized and then re-suspended in 175μL pyridine, 25μL acetic anhydride and incubated at 50°C overnight to acetylate the hydroxyl groups[36]. The solvents were then removed by using a Speed-Vac concentrator, and the samples were re-suspended in water at a concentration of 0.2μg/μl for later analysis.

LC-MS^n Analysis

RPLC was performed on a regular capillary C18 column (Michrom Bioresources, 0.2×50mm, 3μm, 200 Å), using a linear gradient of acetonitrile from 20%- 60% over 60min in 1mM sodium acetate, with a flow rate of 3μL/min and a 10μL injection at a sample concentration of 0.2μg/μL water. Mass spectrometry was performed on a Thermo LTQ-FT instrument using CaptiveSpray ionization with an Advance Ion Source for Thermo MS (Michrom Bioresources Inc.). The spray voltage was set at 1.9kV; capillary temperature was set to 250°C. A full mass scan was acquired under FT mode followed by several directed MS^n scans of precursors with ion trap mode to direct MS^n analysis. The range of the collision energy for all stages of ion activation was set as 35-40V.

Results and Discussion

Homogeneous fractions of CS hexasaccharides were obtained by a combination of SEC and SAX chromatography (**Figure 2.1)**. The structures of these isomers were then determined by 2D-NMR (**Figure 2.2**). Previous studies had already shown that commercial CS sources like CS-A and CS-C are not composed of only A-unit or C-unit but always with substantial amounts of other disaccharide units. The proper validation of any new analytical method requires the use of standards of known structure, therefore the purification and rigorous structural analysis of CS oligomers by LC and NMR spectroscopy, while difficult and laborious, is a key part of the effort to develop analytical methods, especially for heterogeneous polysaccharides such as GAGs.

Two types of enzymes were used to obtain CS oligosaccharides with either unsaturated non-reducing end (by C lyase) or with saturated non-reducing end (by hyaluronidase). Initial work was performed using hexasaccharides generated by lyase digestion, which introduces a double bond at the non-reducing terminus. The asymmetry introduced by the double bond makes interpretation of product ion spectra easier. Four homogeneous CS hexasaccharides were isolated: two trisulfated hexasaccharides [ΔGlcA-GalNAc(4S)-GlcA-GalNAc(4S)-GlcA-GalNAc(4S)-ol] (ΔC4;4;4S-ol), and [ΔGlcA-GalNAc(6S)-GlcA-GalNAc(6S)-GlcA-GalNAc(6S)-ol] (ΔC6;6;6S-ol) from C lyase digested CS-A, one disulfated hexasaccharide [GlcA-GalNAc(6S)-GlcA-GalNAc(4S)-GlcA-GalNAc-ol] (C6;4;0S-ol) from hyaluronidase digested CS-A and one tetra-sulfated hexasacccharide [ΔGlcA-GalNAc(4S)-GlcA(2S)-GalNAc(6S)-GlcA-GalNAc(6S)-ol] (ΔC4;2,6;6S-ol) from C lyase digested CS-C. The symbol Δ indicates 4,5-unsaturation on GlcA at the non-reducing end, the "S" refers to sulfation group, and the "-ol" indicates GalNAc at the reducing end is reduced. A hexasaccharide fraction isolated from hyaluronidase-digested CS-A was collected after SEC separation and without further SAX

separation in order to test the method's ability to separate and characterize structures from mixtures.

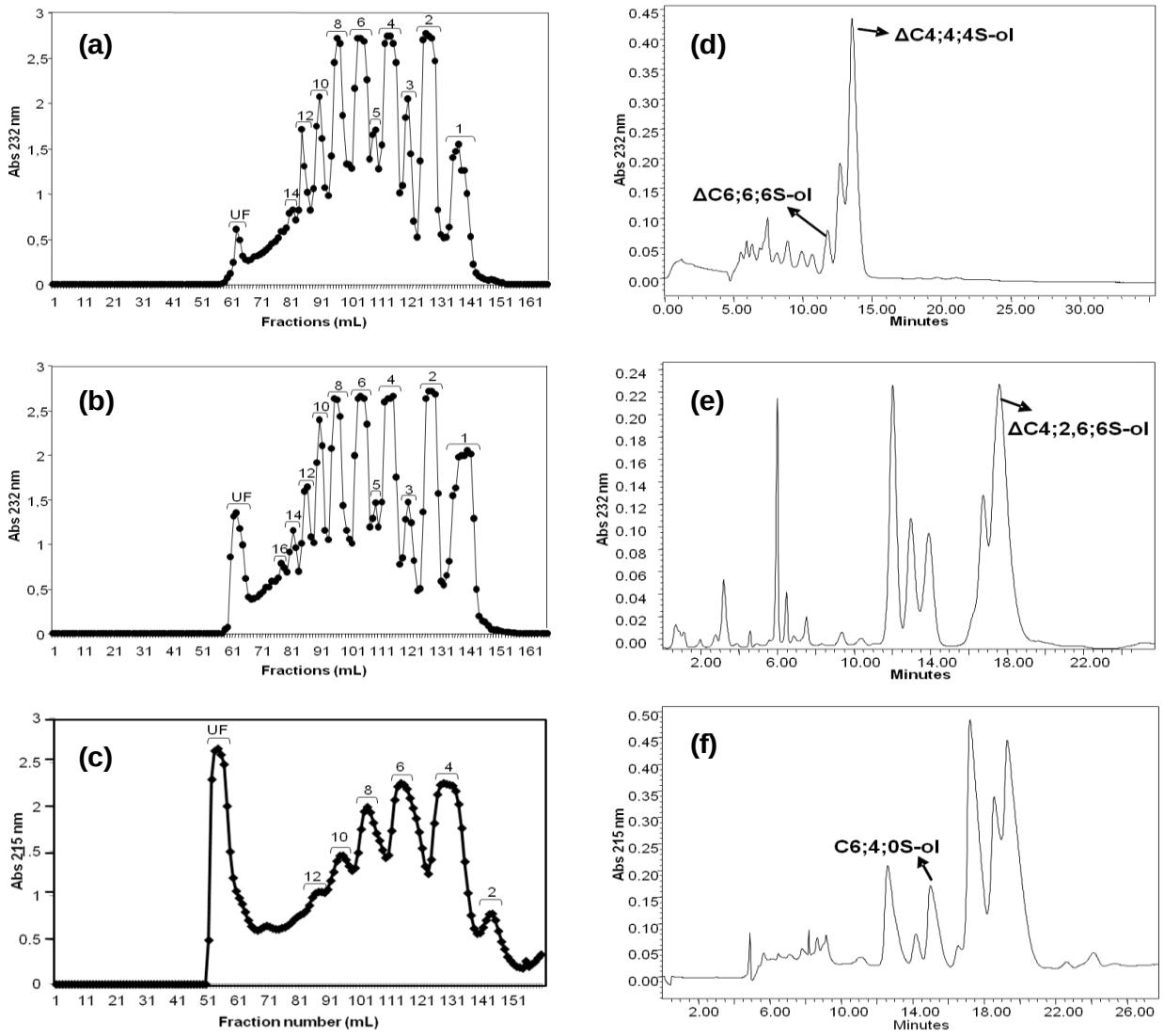


Figure 2.1. Chromatograms for SEC (a-c) and SAX separations (d-f). Three enzymatic digested CS samples subjected to SEC fractionation are showed including C lyase digested CS-A (**a**), C lyase digested CS-C (**b**) and hyaluronidase digested CS-A (**c**). The numbers on the top of each peak correspond to the number of monosaccharide residues. The SAX chromatograms illustrate hexasacchrides with specific sulfation patterns purified from hexasaccharide fractions from each SEC separation: ΔC6;6;6S-ol and ΔC4;4;4S-ol from C lyase digested CS-A (**d**), ΔC4;2,6;6S-ol from C lyase digested CS-C (**e**) and C6;4;0S-ol from hyaluronidase digested CS-A (**f**).

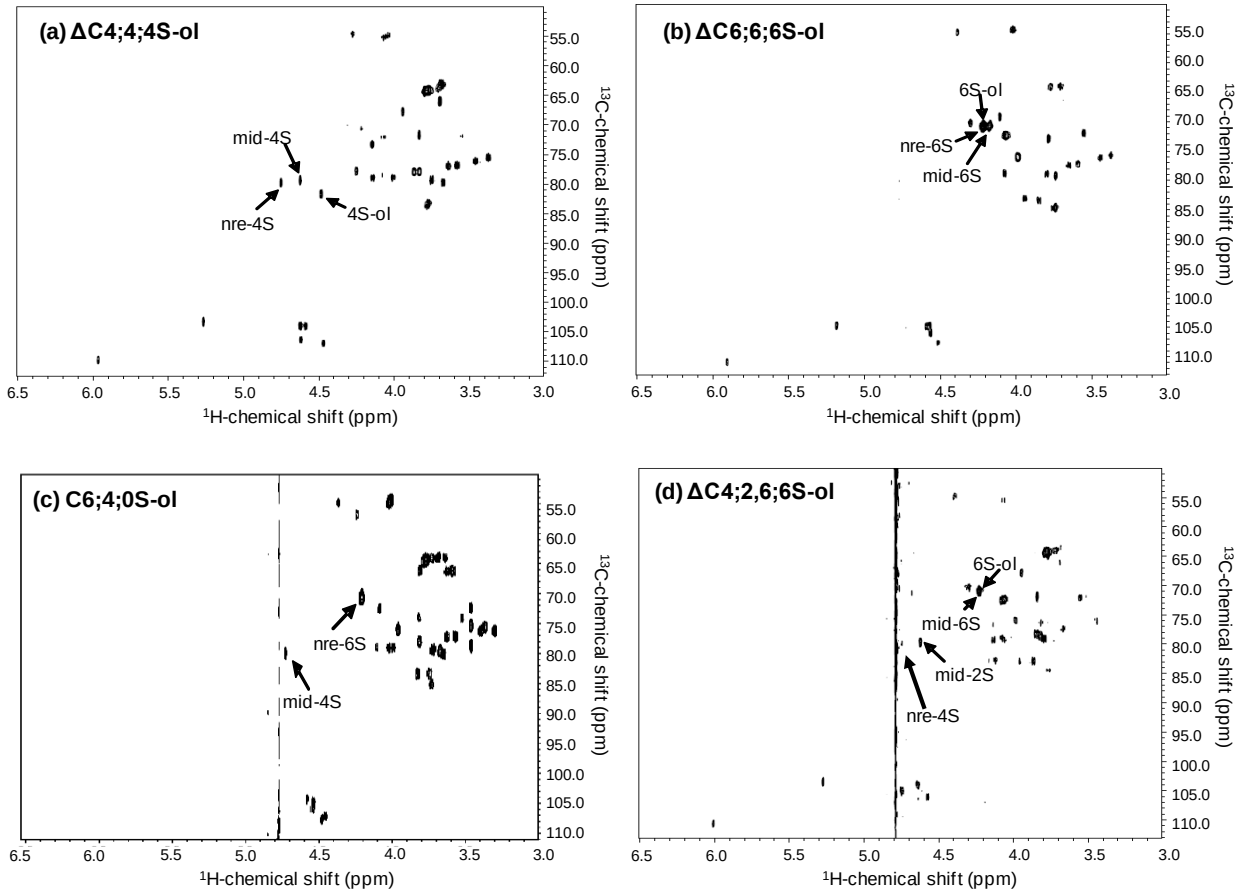


Figure 2.2. ^{13}C-gHSQC NMR analyses of chromatography enriched CS hexasacchrides ΔC4;4;4S-ol (a), ΔC6;6;6S-ol (b), C6;4;0S-ol (c) and ΔC4;2,6;6S-ol (d). The cross-peaks which denote the sulfation sites are indicated respectively in the NMR spectra. The nre-, mid-, and -ol stands for GalNAc units positioned at the non-reducing end, middle, and reducing terminal, respectively. The nre-4S peak in (**d**) has lower intensity due to presaturation effects at peaks nearby the water signal.

Chemical Derivatizations of CS hexasaccharides

The initial part of our methodology is the chemical derivatization of the oligosaccharides, which include sequential permethylation, desulfation and acetylation. The purpose of the chemical derivation strategy is to substitute the labile sulfate groups with a more stable and

distinguishable group that can be differentiated from permethylated groups, both by mass and by LC retention time. This allows the LC separation and structural identification of sulfated GAG oligomers differing only by one or more positions of sulfation. Thus, we substitute the sulfate groups with acetyl groups which are more stable and distinguishable from methyl groups. The position of acetyl groups determined by CID tandem mass spectrometry will reflect the original position of sulfate groups.

The first obstacle encountered for the derivatization scheme was the dissolution of highly sulfated CS hexasaccharides in DMSO, which is the common solvent used in the standard permethylation procedure. Unlike un-sulfated or less densely-sulfated glycans, the high density of the sulfate and carboxyl groups make GAG oligosaccharides highly negative charged and usually exist as sodium salts, which prevents their complete dissolution in DMSO. Therefore, the counter-ions in our CS hexasaccharides were exchanged to form a TEA salt prior to the permethylation reaction. The TEA-CS salts were easily dissolved in DMSO, even for highly sulfated GAG oligomers. The mass of the products showed that the CS hexasaccharides were fully permethylated, including the hydroxyl groups of the carboxyls on the GlcA and the hydrogen of the acetamine group on the GalNAc, leaving the sulfate groups unmodified as shown in **Figure 2.3a** and **2.3b**.

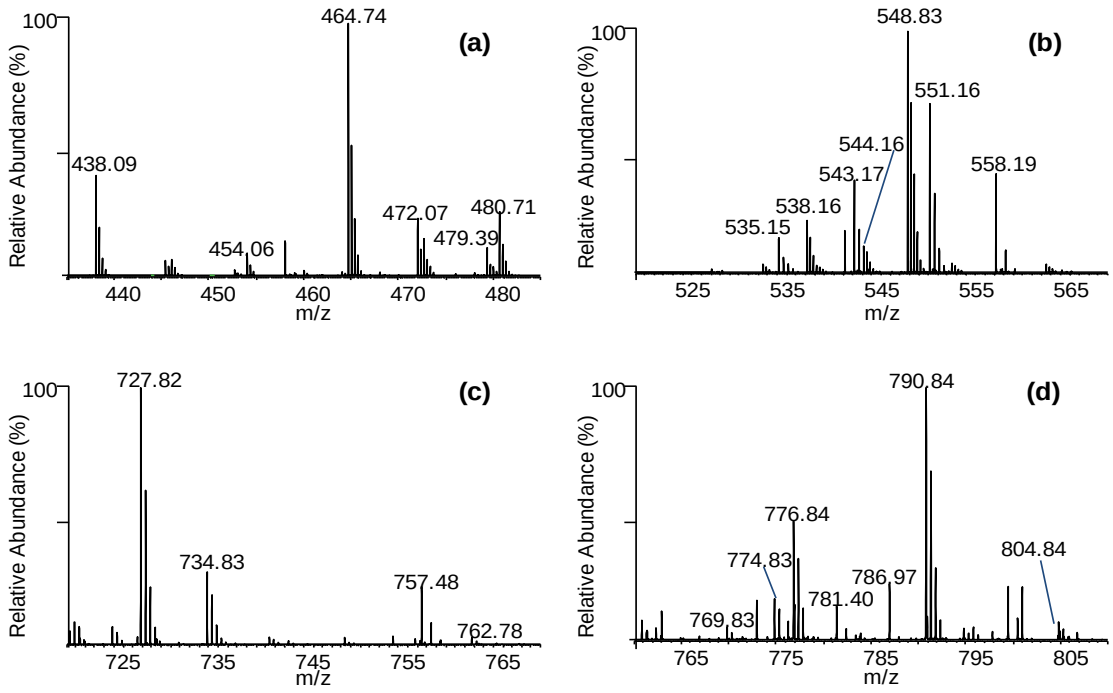


Figure 2.3. MS spectra for reduced saturated trisulfated hexasacchrides with and without derivatizations. Spectra are all full mass scans obtained under FT mode in negative ion mode for underivatized (**a**) and permethylated (**b**) products, and in positive ion mode for permethylated-desulfated (**c**) and permethylated-desulfated-acetylated (**d**) products. (**a**) $[M-3H]^{3-}$ at m/z = 464.74 gives molecular weight of 1397.22. (**b**) $[M-3H]^{3-}$ at m/z = 548.84 gives molecular weight of 1649.52 indicating 18 permethylation sites. Less than 5% of the products are under-permethylated with 14 Da mass off (m/z of 544.16). Other significant peaks are disaccharide (558.19), tetrasaccharide (535.15, 543.17 and 551.16) or hexasaccharide (538.16) by-products caused by the β-elimination reaction during permethylation, which are not selected for MS^n analysis. (**c**) $[M+2Na]^{2+}$ at m/z = 727.82 gives molecular weight of 1409.64 indicating all three sulfate groups are desulfated with less than 10% partial desulfated products (m/z of 767.82). (**d**) $[M+2Na]^{2+}$ at m/z = 790.84 gives molecular weight of 1535.68 and no ion is detected around m/z of 769.83, indicating all three desulfated sites are fully acetylated.

For desulfation reaction, two methods were experimentally compared in order to increase sulfation removal efficiency: methanolic desulfation and the solvolytic desulfation. The former was performed by incubating the permethylated GAG in methanol/HCl, where the sulfate groups are removed by acid methanolysis[37]. However, the acidic conditions seemed to cause the cleavage of some glycosidic bonds concurrently with desulfation, as previously observed with sulfated N-linked glycans[38]. Even though some glycosidic bond cleavages in our products were decreased by optimizing conditions, the amounts of glycolysis were still leading to undesirable levels of sample losses. Adoption of a pyridinium salt-mediated solvolytic desulfation protocol described in the Materials and Methods section reduced the sample loss considerably as determined by direct infusion analysis in both positive and negative ESI-MS (data not shown).

After desulfation of the hexameric mixture, three different derivatization schemes were comparatively tested for determination of the original sites of sulfation. One aliquot was derivatized by acetylation, labeling the original sulfation sites with a group that could be differentiated from permethylated sites by mass and, potentially, by retention on a C18 column. A second aliquot was derivatized by trideuteropermethylation due to the higher efficiency of permethylation compared with peracetylation methods. In this aliquot, the original site of sulfation was labeled with a group that could be easily differentiated from a standard permethylation site by mass, but probably not by retention time. Finally, a third aliquot was kept underivatized at the hydroxyl groups left after desulfation.

All three derivatized versions were retained on a standard C18 column and analyzed in positive ion mode. The subsequent MS^n analyses of the aliquot with underivatized hydroxyl groups at the original sites of sulfation exhibited extensive water loss during the fragmentation procedure. This resulted in fewer informative product ions than the other two derivatized types.

The chromatograms of the unmodified hexasaccharides also gave broader peaks and shorter retention times, making the separation less efficient than for the acetylated product, as shown in **Figure 2.4a**. Trideuteromethylation of former sites of sulfation solved the issues pertaining to water loss during ion activation. The permethylated-desulfated-trideuteromethylated products gave valuable information from the MS^n spectra, and enabled differentiation of isomeric hexasaccharide species based on diagnostic product ions. However, the efficient separation of isomers based on differential position of trideuteromethyl groups in the fully permethylated GAG was not possible, which makes confident identification of specific CS hexamers from an isomeric mixture difficult at best (**Figure 2.4b**). For products where the original sites of sulfation were replaced by acetyl groups, CS sulfation isomers were readily distinguishable through MS^n analyses, as detailed below. Additionally, relatively efficient separations were achieved using standard RPLC based on the different positions of the acetyl groups (**Figure 2.4c**).

After reduction, permethylation, desulfation, and acetylation, the total sample loss is estimated at 50%. The masses of the permethylated-desulfated products and the permethylated-desulfated-acetylated products showed that CS hexasaccharides were properly derivatized (**Figure 2.3c** and **2.3d**). The majority of by-products were tetrasaccharides and disaccharides caused primarily by the β-elimination reaction between the carbon-4 and -5 of the GlcA during the permethylation (**Figure 2.3b**), with a smaller amount of by-products apparently formed during the desulfation procedure. These by-products are notable for not introducing species that would lead to the false identification of sites of sulfation; their major effect is reducing the apparent sensitivity of the technique.

Another problem caused by permethylation is the undermethylation, giving a product (m/z of 544.16 in **Figure 2.3b**) with one free hydroxyl group left. This undermethylation product

can also be acetylated in subsequent acetylation step, giving a final derivatized product with m/z of 804.84 as shown in **Figure 2.3d**. While such product might give a false identification of one additional sulfation site, the amount is very low compare to the fully derivatized product (m/z of 790.84) and the LC chromatogram also gives a separation of 3 min between the two products (data not shown).

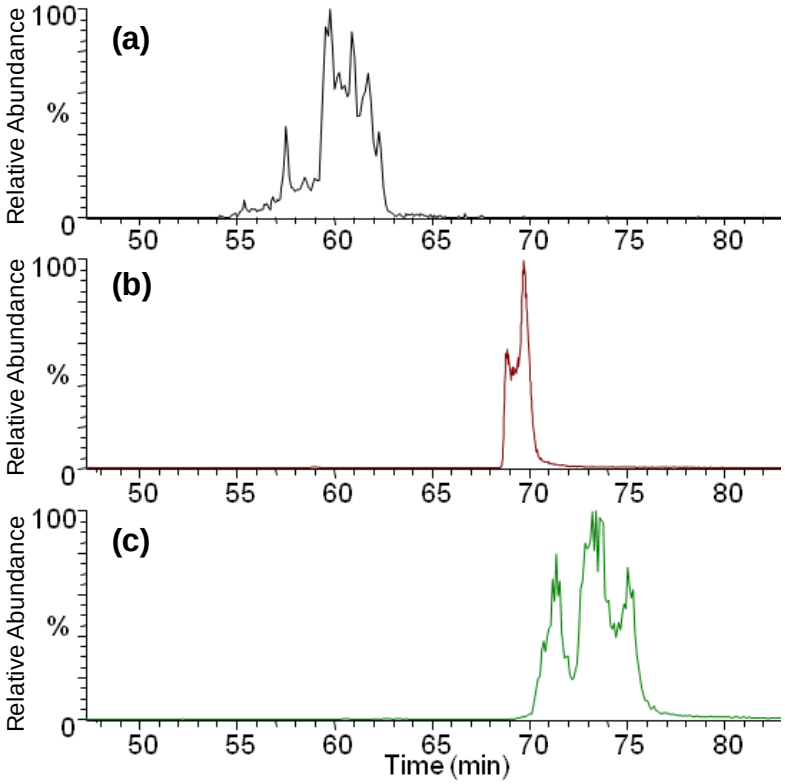


Figure 2.4. RPLC chromatograms for three mixtures of CS trisulfated hexasaccharides from hyaluronidase-digested CS-A. The single ion chromatogram (SIC) for each mixture with different derivatizations is extracted with corresponding parent ions: (**a**) SIC of $[M+2Na]^{2+}$ (m/z 727.8) for the permethylated, desulfated mixture. (**b**) SIC of $[M+2Na]^{2+}$ (m/z 757.8) for the permethylated, desulfated, trideuteromethylated mixture. (**c**) SIC of $[M+2Na]^{2+}$ (m/z 790.8) for the permethylated, desulfated, acetylated mixture. Previous SAX purification and NMR analyses indicate three major trisulfated hexasaccharide products in this mixture.

Structural Differentiation of CS Isomers ΔC4;4;4S-ol and ΔC6;6;6S-ol

As CS molecules vary mainly at 4- and 6-sulfation of the GalNAc, homogeneous tri-sulfated hexasaccharides ΔC4;4;4S-ol and ΔC6;6;6S-ol were selected as initial standards. They allow differentiation during the MS^n analyses between the A-unit and C-unit at each region of the hexamer, including non-reducing end, internal region and reducing end. As illustrated in **Figure 2.5**, all free hydroxyl groups of the oligosaccharides are now methylated after derivatization, and the previous sites of sulfation are labeled by acetyl groups. Fragmentation nomenclature is followed here as described by Domon and Costello[39].

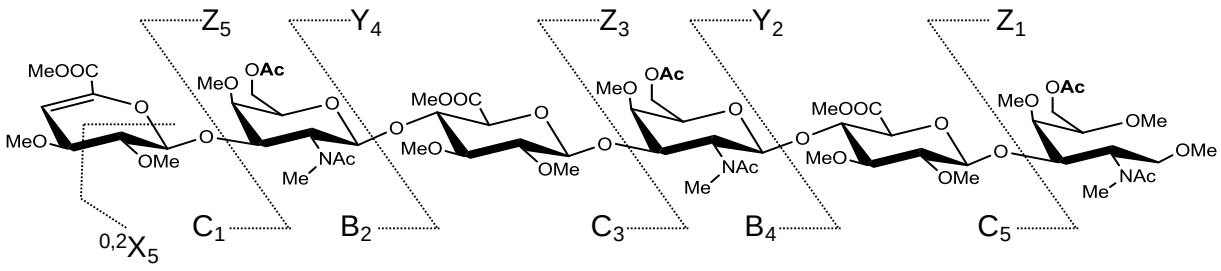


Figure 2.5. Derivatized product of [ΔGlcA-GalNAc(6S)-GlcA-GalNAc(6S)-GlcA-GalNAc(6S)-ol] (ΔC6;6;6S-ol). The symbol Δ indicates 4,5-unsaturation on GlcA at the non-reducing end, the "S" refers to sulfation group, and the "-ol" indicates GalNAc at the reducing end is reduced. The structure shows that there is one acetyl group on the 6 position for each GalNAc indicating the original sulfation sites, while other positions are all methylated.

After all derivatizations, ΔC4;4;4S-ol and ΔC6;6;6S-ol were first analyzed by manual MS^n experiments by direct infusion of solutions of the sample at a concentration of 10μM in 50/50 acetonitrile/water with a flow rate of 3μL/min. Four ions, $[M+H]^+$, $[M+2H]^{2+}$, $[M+Na]^+$ and $[M+2Na]^{2+}$, were detected and analyzed by thorough MS^n up to MS^4 for the most abundant

product ions for each round of fragmentation (data not shown). While it is relatively simple to drive permethylated sugar ions toward sodium adducts by the introduction of sodium salts into the running buffers, it is very difficult to wholly eliminate sodium adduct ions and drive permethylated sugars to protonated ions. Additionally, the $[M+2Na]^{2+}$ ion yielded more thorough and reliable structural information than $[M+Na]^{+}$ under corresponding optimized condition. Therefore, selection of the doubly charged sodium adduct as the preferred precursor ion allows us to increase our sensitivity by driving our ion population toward sodium adducts.

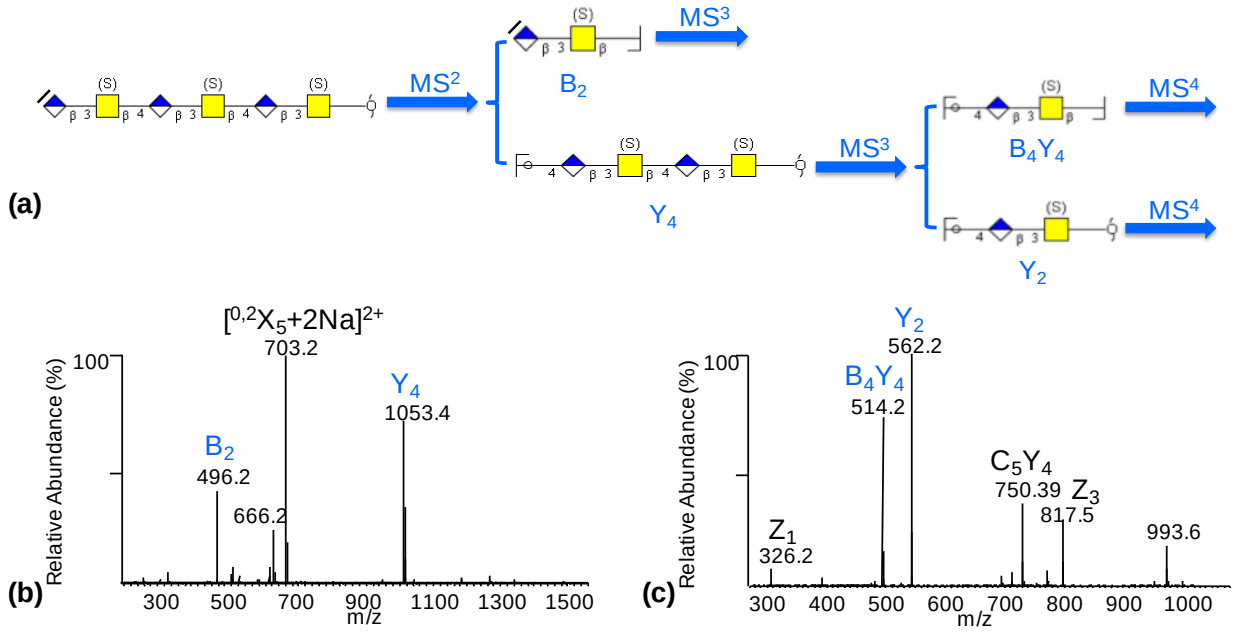


Figure 2.6. Fragmentation path for derivatized CS trisulfated hexasacharides (**a**) and MS^n spectra for derivatized oligosaccharide ΔC6;6;6S-ol: (**b**) MS^2 of $[M+2Na]^{2+}$ (m/z 774.8) and (**c**) MS^3 of $[Y_4+Na]^{+}$ (m/z 1053.4). Symbols: ◆ ΔGlcA, ◆ GlcA, □ GalNAc. All product ions labeled on the spectra are singly charged and mono-sodiated unless otherwise annotated.

As illustrated in **Figure 2.6a**, the product ions $[B_2+Na]^{+}$, $[B_4Y_4+Na]^{+}$ and $[Y_2+Na]^{+}$ with different *m/z* values represent the three disaccharides units of each hexasaccharide, which enable

us to compare the differences between 4- and 6-sulfation on different positions along the oligomeric chain in a systematic, modular scheme by doing MS^n individually of the respective product ions. The MS^2 of $[M+2Na]^{2+}$ gave three major product ions $[^{0,2}X_5+2Na]^{2+}$, $[B_2+Na]^+$ and $[Y_4+Na]^+$ (**Figure 2.6b**) regardless of the original pattern of sulfation. MS^3 of the $[Y_4+Na]^+$ product ion gave two product ions $[B_4Y_4+Na]^+$ and $[Y_2+Na]^+$ (**Figure 2.6c**). The three unique ions representing each disaccharide unit along the hexamer allows for individual determination of the site(s) of sulfation of each disaccharide unit of the oligosaccharide in a systematic manner.

As part of the development of the initial MS^n parameters for 6- and 4-sulfation at each disaccharide unit of the hexasaccharide by direct infusion of these two enriched NMR-characterized standards, identification of diagnostic MS^n ions was performed by on-line RPLC-MS. The additional LC-based separation step was performed prior to diagnostic MS^n product ion analysis since the purity of these standards was only verified by NMR. Low levels of contaminant isomers missed by NMR analysis might lead to incorrect identification of diagnostic product ions in the MS^n spectra. Online LC separation also isolated the target products (hexasaccharides) from those by-products (tetrasaccharides and disaccharides) resulting from the chemical derivatization steps. This also helped to increase the sensitivity of the analysis by reducing ion suppression in the electrospray process from by-products. In order to get $[M+2Na]^{2+}$ as the predominant parent ion, 1mM sodium acetate was added to the LC mobile phases, which did not seem to significantly compromise electrospray stability.

The MS method was set up as six events including one full mass scan followed by five MS^n scans for five ions shown in **Figure 2.6a**: MS^2 of $[M+2Na]^{2+}$, MS^3 of $[B_2+Na]^+$, MS^3 of $[Y_4+Na]^+$, MS^4 of $[B_4Y_4+Na]^+$ and MS^4 of $[Y_2+Na]^+$. Three MS^n spectra of ΔC4;4;4S-ol (**Figure 2.7a-c**) and ΔC6;6;6S-ol (**Figure 2.7d-f**) were compared.

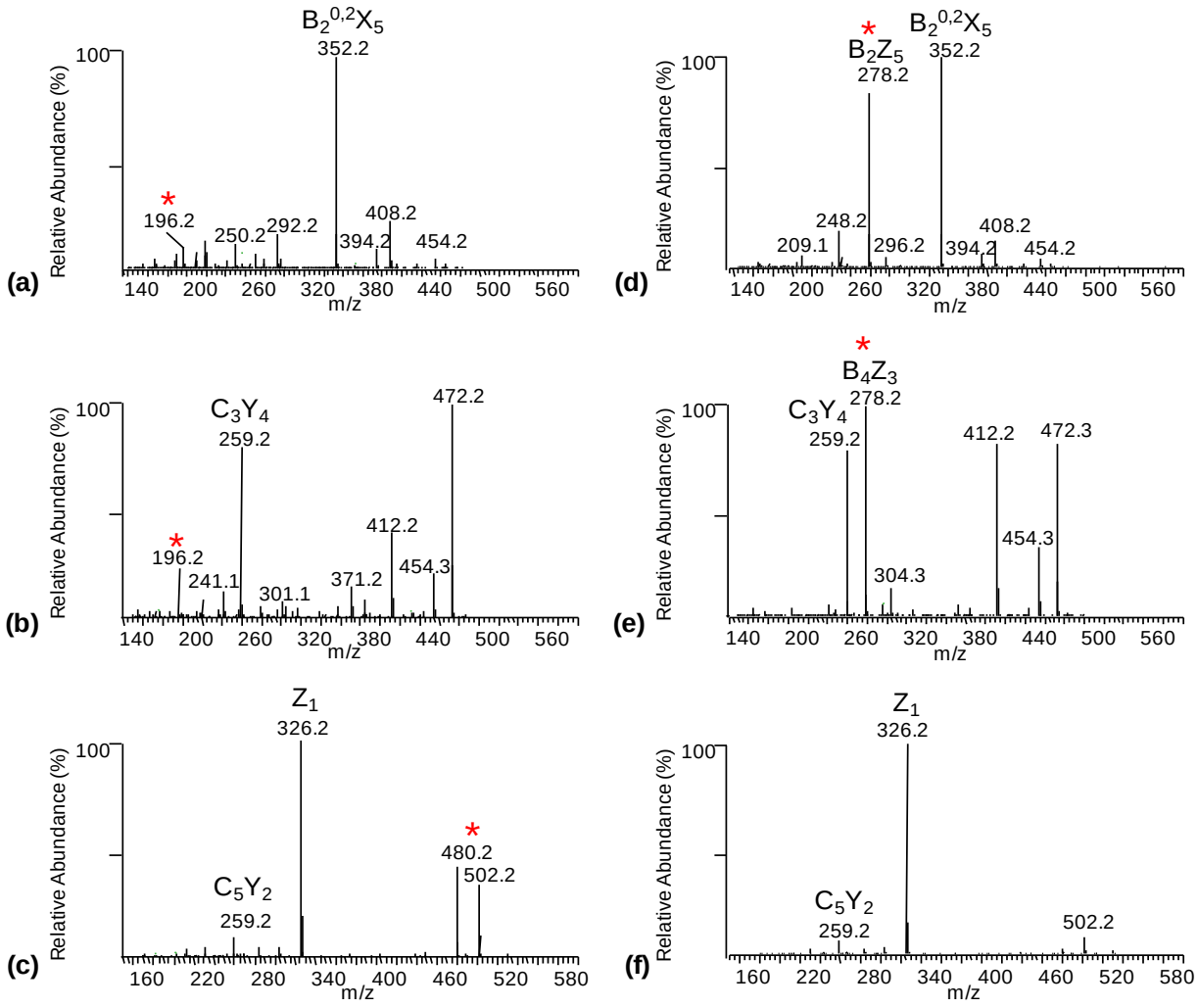


Figure 2.7. MS^n spectra for two derivatized hexasaccharide standards ΔC4;4;4S-ol (a, b, c) and ΔC6;6;6S-ol (d, e, f). From top row to bottom row, they are **(a, d)** MS^3 of $[B_2+Na]^+$ (m/z 496.2), **(b, e)** MS^4 of $[B_4Y_4+Na]^+$ (m/z 514.2) and **(c, f)** MS^4 of $[Y_2+Na]^+$ (m/z 562.2) spectra, representing the three disaccharide units respectively at the non-reducing end, in the middle and at the reducing end. The peak marked with an asterisk represents the diagnostic product ions used for differentiation between 4-sulfated and 6-sulfated GalNAc.

From the MS^3 of $[B_2+Na]^+$, we noticed that the B_2Z_5 ion with *m/z* of 278.2 was abundant only for the 6-sulfated disaccharide on the non-reducing end (**Figure 2.7d**), while 4-sulfation

disaccharide at the non-reducing end resulted in a MS^3 product ion of *m/z* 196.2 (**Figure 2.7a**). Similarly for the internal disaccharide, the B_4Z_3 ion of *m/z* 278.2 was also observed for the 6-sulfated GalNAc (**Figure 2.7e**) on the MS^4 spectra of $[B_4Y_4+Na]^+$ ion and not for the 4-sulfated GalNAc (**Figure 2.7b**) where the product ion of *m/z* 196.2 was observed. The trend did not continue with the disaccharide on the reducing end where the GalNAc was reduced. As shown in **Figure 2.7c**, the MS^4 of $[Y_2+Na]^+$ ion gave two diagnostic product ions with *m/z* of 480.2 and 502.2 for 4-sulfated GalNAc, which were practically undetectable for the 6-sulfated GalNAc at the reducing end (**Figure 2.7f**). For the reduced GalNAc at the reducing end, the acetyl group on the 4-position appeared more likely to be lost during the fragmentation than it did on the 6-position, resulting in the more intense product ions as observed. Overall, by fragmenting the three target product ions corresponding to disaccharides units in a sequential manner, we were able to differentiate between the 4- and 6-sulfated GalNAcs at any disaccharide position along the CS hexasaccharides using a systematic predetermined MS^n scheme.

LC-MS^n Analysis of CS Tri-sulfated Hexasaccharides Mixture

In order to explore the separation capabilities of RPLC for derivatized CS hexasaccharides, as well as demonstrate the ability to differentiate isomers by MS^n on an LC time-scale, a mixture of tri-sulfated CS hexasaccharides was prepared as described in the experimental section, and analyzed by the similar LC-MS^n method used for standards. Unlike the standards ΔC4;4;4S-ol and ΔC6;6;6S-ol, having unsaturated non-reducing end residues (due to lyase digestion), the hexasaccharides mixture used here were prepared by hyaluronidase digestion in order to produce oligosaccharides with saturated non-reducing end residue. This would extend the application of our methodology to CS oligosaccharides produced from different workflows. As listed in **Table 2.1**, among the five ions analyzed in the MS^n method, the

$[M+2Na]^{2+}$ and $[B_2+Na]^+$ ion had different *m/z* values for hexasaccharides having unsaturated and saturated non-reducing ends, while the $[Y_4+Na]^+$, $[B_4Y_4+Na]^+$ and $[Y_2+Na]^+$ ion had same *m/z* values for both species. In order to check if the differentiation of 4- and 6-sulfated for the GalNAc of the saturated disaccharide units would be similar to unsaturated disaccharide units, the MS^3 of $[B_2+Na]^+$ ion (*m/z* 496.2) were extracted and two different types of spectra were observed as shown in **Figure 2.8**. The diagnostic ions with *m/z* of 278.2 and 196.2 were still observed and able to be used for distinguishing the 6- and 4-sulfated GalNAc respectively, even though the MS/MS fragmentation looks different from the standards (**Figure 2.7a *vs.* d**) by missing the $[B_2{}^{0,2}X_5+Na]^+$ ion (*m/z* 352.2) that was believed to be typically produced by retro-Diels-Alder decomposition of cyclohexene-like structure.

Table 2.1. *m/z* of relevant precursor and product ions for derivatized CS hexamer identification.

Parent and product ions[a]	saturated non-reducing end	unsaturated non-reducing end	Mass shift by +/- one sulfation group
M^{2+}	790.8	774.8	+/- 14
$^{0,2}X_5{}^{2+}$	NA[b]	703.2	+/- 14
$B_2{}^+$	528.2	496.2	+/- 28
$Y_4{}^+$	1053.4	1053.4	+/- 28
$B_4Y_4{}^+$	514.2	514.2	+/- 28
$Y_2{}^+$	562.2	562.2	+/- 28
$C_1{}^+$	273.2	NA[b]	+/- 28
$B_2Z_5{}^+$	278.2	278.2	+/- 28
$C_3Y_4{}^+$	259.2	259.2	+/- 28
$B_4Z_3{}^+$	278.2	278.2	+/- 28
$C_5Y_2{}^+$	259.2	259.2	+/- 28
$Z_1{}^+$	326.2	326.2	+/- 28
$B_2{}^{0,2}X_5{}^+$	NA[b]	352.2	+/- 28

[a] All ions were mono-sodiated except for M^{2+} and $^{0,2}X_5{}^{2+}$ which were di-sodiated.
[b] The corresponding product ion was not observed.

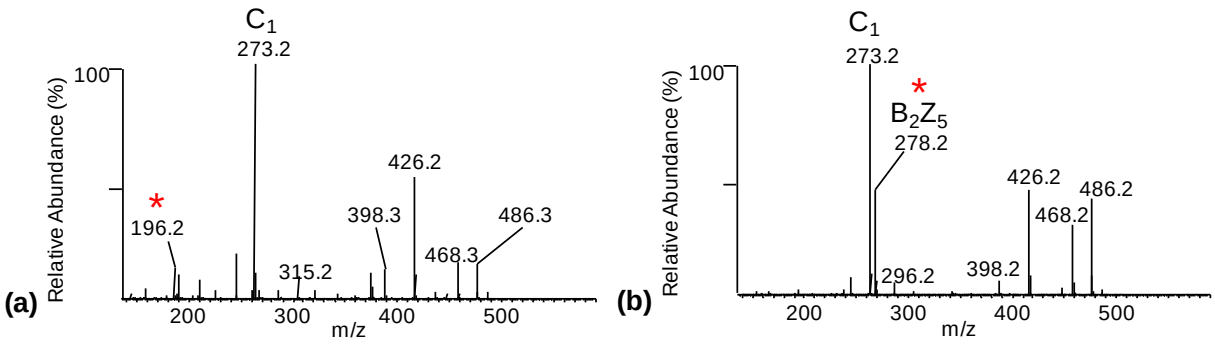


Figure 2.8. Two different MS^3 of $[B_2+Na]^+$ (m/z 528.2) spectra extracted from the LC-MS^n anlaysis of mixture of derivatized trisulfated hexasaccharides with a saturated non-reducing end. While the product ion spectra are notably different from the hexasaccharide standards with an unsaturated non-reducing end (**Figure 2.7 a/d**), the same diagnostic ions were observed at m/z 196.2 and 278.2 representing the 4-sulfated GalNAc (**a**) and the 6-sulfated (**b**) GalNAc. The peak marked with an asterisk represents the diagnostic product ions used for differentiation between 4-sulfated and 6-sulfated GalNAc.

The base peak chromatograms (BPC) of the hexasaccharide mixture for MS^3 of $[B_2+Na]^+$, MS^4 of $[B_4Y_4+Na]^+$ and MS^4 of $[Y_2+Na]^+$ were extracted and shown as a black line in **Figure 2.9** from the top panel to the bottom one. For MS^3 of $[B_2+Na]^+$, the diagnostic product ions with *m/z* of 278.2 (6-sulfated) and 196.2 (4-sulfated) indicate 6-sulfation (**Figure 2.9a**, red line) or 4-sulfation (**Figure 2.9a,** purple line) in GalNAc disaccharide located at the non-reducing end. The same selected ion chromatogram (SIC) were extracted for MS^4 of $[B_4Y_4+Na]^+$ (**Figure 2.9b**), where the same diagnostic product ions with *m/z* of 278.2 and 196.2 were extracted. For MS^4 of $[Y_2+Na]^+$, the ion with *m/z* value of 480.2 was extracted to indicate isomers with 4-sulfated GalNAc on the reducing end (**Figure 2.9c**, red line). However, there was no diagnostic product ion for 6-sulfated isomers; rather, 6-sulfated GalNAc on the reducing end

is signified by a mass consistent with a sulfated reducing end unit combined with a lack of MS^4 product ion of *m/z* 480.2 as shown in **Figure 2.7f**. In order to plot the SIC for the 6-sulfated GalNAc on the reducing end, the intensity of the ion of *m/z* of 326.2 (which is common to both 6-sulfated and 4-sulfated reducing end disaccharide product ions) was subtracted by the sum of the intensities of the ion of *m/z* of 480.2 and 502.2 (which are abundant for the 4-sulfated reducing end disaccharide product ion but not to the 6-sulfated one) to show the 6-sulfated species (**Figure 2.9c**, purple line).

Each SIC in **Figure 2.9** represents all isomers with the annotated modification; therefore, in order to identify a particular isomer, data from all three panels must be superimposed, with the data in panel A indicating the sulfation site at the non-reducing end GalNAc, panel B indicating the sulfation site at the internal GalNAc, and panel C indicating the sulfation site at the reducing end GalNAc. In this manner, even though the chromatography alone is insufficient to resolve the isomers to baseline, a combination of the chromatography with the MS^n data are capable of resolving all of the different isomers in the trisulfated hexamer.

Three major peaks were observed in the BPC shown in **Figure 2.9**. For the first peak, the three SICs that covered the peak area at ~71 minutes for GalNAc at the non-reducing end (panel a), the internal (panel b), and the reducing end (panel c) disaccharides showed that all contain 4-sulfation (purple lines), indicating for the first peak at ~71 minutes that the structure is C4;4;4S-ol.

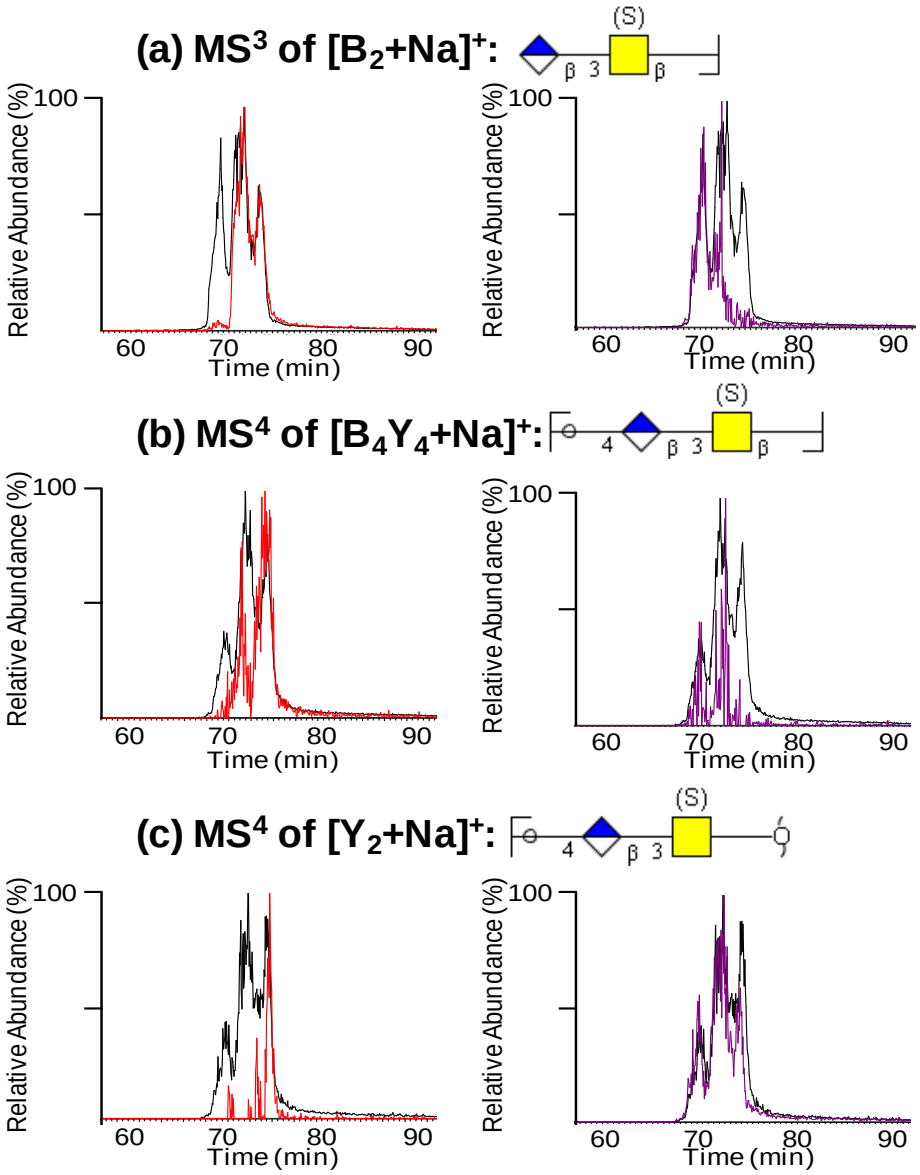


Figure 2.9. LC-MS^n anlaysis for a derivatized minxture of CS trisulfated hexasaccharides with saturated non-reducing ends. Base peak chromatograms were shown for MS/MS spectra of three disaccharide units: (**a**) MS^3 of $[B_2+Na]^+$, (**b**) MS^4 of $[B_4Y_4+Na]^+$ and (**c**) MS^4 of $[Y_2+Na]^+$ **(black)**. Extracted single ion chromatograms (SIC) with m/z of diagnostic ions indicate the appropriate GalNAc is 6-sulfated **(**in red**)** or 4-sulfated **(**in purple**)**. The three sulfation sites for each BPC peak can be assigned separately by the color of the three SIC that cover the corresponding peak area.

The second major peak in the BPC at ~74 minutes seemed to have more than one species based on the various SICs extracted for the non-reducing end (**Figure 2.9a**) and the internal disaccharides (**Figure 2.9b**). The SICs of the non-reducing end indicate the presence of a large amount of 6-sulfation, with a small amount of 4-sulfation that elutes toward the front of the peak. Analysis of the SICs from the internal disaccharide indicate the presence of both the 6- and 4-sulfated internal disaccharide, where the 6-sulfated internal disaccharide co-elutes with the 4-sulfated non-reducing disaccharide, and the 4-sulfated internal disaccharide co-elutes with the 6-sulfated non-reducing disaccharide. No significant overlap was found between the 6-sulfated non-reducing disaccharide and the 6-sulfated internal disaccharide, neither between the 4-sulfated non-reducing disaccharide and the 4-sulfated internal disaccharide. All of the hexasaccharides in the second major peak in the BPC at ~74 minutes exhibited a 4-sulfated reducing end, with no 6-sulfation observed at the reducing end. These results indicate that the second major peak in the BPC at ~74 minutes consists of two analytes: the major analyte is C6,4,4S-ol which eluted in the latter portion of the peak, and a minor component identified as C4,6,4S-ol that eluted at the front of the peak.

For the third major peak in the BPC at ~76 minutes, both the non-reducing end and the internal disaccharide were identified as being solely sulfated at the 6-position. However, the third major peak was determined to consist of two analytes that differed in sulfation at the reducing end: the front of the peak showed the major component to have 4-sulfation at the reducing end, while the tail of the peak showed a relatively minor component with 6-sulfation at the reducing end. These chromatograms allowed the interpretation of the third major peak in the BPC at ~76 minutes as consisting of both C6,6,4S-ol and C6;6;6S-ol. Even though the chromatography did

not separate the CS hexasaccharide isomers to baseline in the BPC, the MS^n indicated sufficient separation capability to clearly identify five slightly different CS hexamer isomers.

In total, five isomers were identified even with two minor spices (C4,6,4S-ol and C6;6;6S-ol) that cannot be chromatographically separated and detected by 2D-NMR without derivatization. Generally, the isomers could be separated by the extension of 6-sulfation, eluting in the order of no 6-sulfate (C4;4;4S-ol), one 6-sulfate (C4,6,4S-ol and C6,4,4S-ol), two 6-sulfates (C6,6,4S-ol) and three 6-sulfates (C6;6;6S-ol).

Structural Analysis of CS Di-sulfated and Tetra-sulfated Hexasaccharides

While the monosulfated A- and C-unit disaccharides are the most commonly detected disaccharide units in most chondroitin sulfate sources, the non-sulfated [GlcA-GalNAc] O-unit and the disulfated [GlcA(2S)-GalNAc(6S)] D-unit and [GlcA-GalNAc(4S6S)] E-unit are also commonly found. To validate the current methodology for recognition of sulfation patterns in CS oligosaccharides with differential sulfation degrees, two other CS saturated hexasaccharides were derivatized and analyzed: the disulfated C6;4;0S-ol and the tetra-sulfated ΔC4;2,6;6S-ol, both also characterized previously by NMR. While the successful identifications of the mono-sulfated disaccharide units, A-unit and C-unit, with either unsaturated or saturated non-reducing end were performed as described above, non-sulfated unit (O-unit) and disulfated units (D-unit and E-unit) were not tested. These units give a characteristic mass shift compared to the mono-sulfated A- and C-units for their relative parent and product ions as listed in **Table 2.1**, and thus can be easily differentiate mono-sulfated disaccharide units based on the disaccharide unit product ion masses. The mass shift caused by changes of sulfation degrees are calculated and listed in **Table 2.1**, where shifts of 28 Da are caused by the mass difference between the acetyl group and the methyl group. Hexasaccharides with one more sulfation group would generate

derivatized products with one more acetyl group along with one less methyl group. This causes the increase of 28 Da for the product ions containing the sulfation site. Similarly, hexasaccharides with one less sulfation group would have a 28 Da less *m/z* value for relative ions. Between the derivatized oligosaccharides with saturated or unsaturated non-reducing end, a mass difference of 32 Da (MeOH) was observed for the parent ion and the $[B_2+Na]^+$ ion, but not for the other product ions that do not contain the non-reducing end. In addition, the $[C_1+Na]^+$ ion was only observed for saturated species (**Figure 2.8a**), while the $[B_2^{0,2}X_5+Na]^+$ ion was only observed for unsaturated species (**Figure 2.7a**).

Following the fragmentation strategy shown in **Figure 2.6a**, we could easily target the product ions having mass shifts and therefore identify the degrees of sulfation for each disaccharide. The MS/MS spectra of hexasaccharides containing O-unit (C6;4;0S-ol) and D-unit (ΔC4;2,6;6S-ol) are shown in **Figure 2.10**. MS^3 of $[Y_4+Na]^+$ for C6;2;0S-ol (**Figure 2.10a**) produced the $[Y_2+Na]^+$ ion with *m/z* of 534.2 having 28 Da less than the $[Y_2+Na]^+$ ion (*m/z* 562.2) shown in **Figure 2.6c**, indicating that the disaccharide at the reducing end bears no sulfation. For the tetra-sulfated ΔC4;2,6;6S-ol, the MS^4 of $[B_4Y_4+Na]^+$ (**Figure 1.10b**) gave the product ion $[B_4Z_3+Na]^+$ with *m/z* of 278.2, indicating that the internal GalNAc is sulfated at the 6-position instead of the 4-position, while the $[C_3Y_4+Na]^+$ ion (*m/z* 287.2) is 28 Da more massive than the one shown in **Figure 2.7e** (*m/z* 259.2), indicating the internal GlcA was sulfated. Thus, the MS^n data indicated the internal disaccharide unit of ΔC4;2,6;6S-ol is disulfated, where one sulfation is on the GalNAc unit and the other sulfation is on the GlcA unit, distinguishing the D-unit from the E-unit (GlcA-GalNAc4,6S) in CS oligosaccharides.

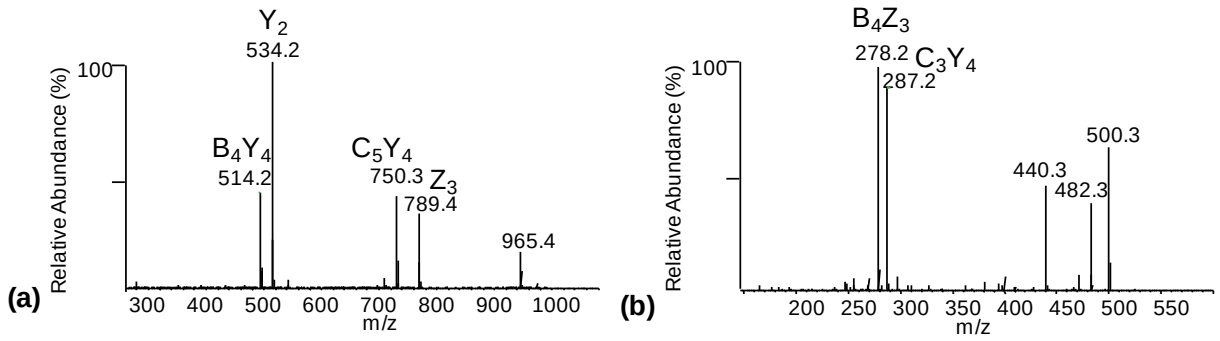


Figure 2.10. Sulfation site identification of disulfated and tetrasulfated hexasaccharides. (**a**) MS^3 of $[Y_4+Na]^+$ (m/z 1025.4) for derivatized C6;4;0S-ol. The m/z for Y_2 ion observed here is 534.2 which is 28 mass units less than 562.2, indicating the GalNAc on the reducing end is non-sulfated. (**b**) MS^4 of $[B_4Y_4+Na]^+$ (m/z 542.2) for derivatized ΔC4;2,6;6S-ol. The m/z for C_3Y_4 ion observed here is 287.2 which is 28 mass units more than 259.2, indicating the internal GlcA is sulfated.

We have shown that the methodology presented here can be successfully used for identification of CS hexasaccharides containing four different types of disaccharide units except for the hexasaccharides with E-unit, which is currently absent in our CS hexameric library. While the differentiation of D-unit and E-unit should be easily obtained by observing the mass difference on the appropriate GlcA and/or GalNAc, we still derivatized and analyzed two disulfated disaccharide standards, the ΔC2,6S and the ΔC46S, to demonstrate the ability of the derivatization scheme to efficiently distinguish the E-unit from the D-unit, and to show the differences of the MS/MS spectra between these two CS disaccharides.

The ΔC2,6S was prepared in the same way as hexasaccharides preparation, while the ΔC46S was purchased from Dextra and analyzed by NMR to verify the structure of the commercial standard (data not shown). Both disaccharides were permethylated, desulfated and

acetylated without previous reduction in order to mimic an MS^n product ion not from the reducing end of an oligosaccharide. The MS^2 spectra of $[M+Na]^+$ (*m/z* 556.2) of both CS dimers are shown in **Figure 2.11**. The *m/z* for Z_1 ion observed for ΔC2,6S was 310.2 (16 Da less than the Z_1 ion shown in **Table 2.1** due to the un-reduced reducing end) indicating only one sulfation site on the GalNAc residue. The Z_1 ion for ΔC46S, on the other hand, with a *m/z* of 338.2 was 28 Da more than the Z_1 ion for ΔC2,6S, indicating two sulfated sites on GalNAc. The $^{0,2}X_1$ ion, a cross-ring cleavage fragment including carbons C1 and C2 of the GlcA and GalNAc, gave the same *m/z* value (412.2) for both disaccharides showing that the second sulfated site for ΔC2,6S was on the C2 position of GlcA. Thus, the results indicate that our methodology can easily differentiate the D-unit from the E-unit, supporting the assignment of CS hexamers bearing any sulfation type at any position.

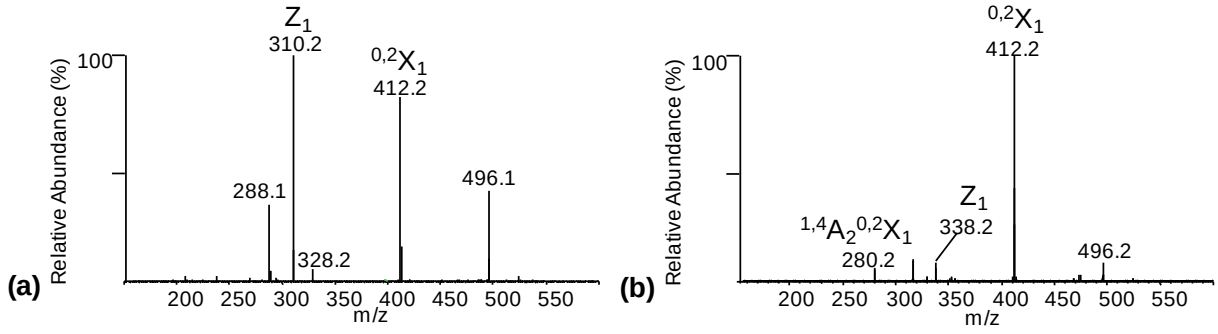


Figure 2.11. MS^2 of $[M+Na]^+$ (m/z 556.2) for derivatized unreduced disaccharide standards. ΔC2,6S (**a**) and ΔC46S (**b**). The m/z for Z_1 ion observed for ΔC2,6S is 310.2 which is 28 Da less than 338.2 for ΔC46S, indicating one less sulfated site on GalNAc. The $^{0,2}X_1$ ion, a cross-ring cleavage fragment ion including carbons C1 and C2 of the GlcA, gives the same m/z value (412.2) showing that the second sulfated site for ΔC2,6S is on the C2 position of GlcA.

In summary, all three units (O-, D- and E-unit) are efficiently derivatized using the protocol detailed above. The derivatized O-unit and D-unit/E-unit can easily be differentiated from each other, as well as from the A-unit/C-unit, based on the masses of the disaccharide product ions in the MS^n fragmentation scheme. While the masses of the D-unit and E-unit disaccharide are identical, our analyses of derivatized disaccharides indicate that these two units can be easily differentiated based on the masses of the abundant monosaccharide product ions present in the final stage of our MS^n protocol, allowing for the accurate identification of all five common disaccharide units in CS hexamers.

Conclusions

In this work, a sequential chemical derivatization strategy allowed the combination of online capillary RPLC separation of isomeric tri-sulfated CS hexasaccharides with subsequent MS^n fragmentation for differentiation of isomers by their sulfation patterns. Online separation of sulfation position isomers was easily achieved by using a standard capillary C18 column using a typical reversed phase gradient coupled to an electrospray interface operated in positive ion mode, which greatly simplifies the analytical procedures and readily supports complex samples of CS hexasaccharides. By using MS^n fragmentations in a linear ion trap, we could fragment hexasaccharides into three distinguishable individual disaccharide units and analyze each separately in a systematic, modular scheme instead of assigning all sulfation sites in just one single complex MS/MS spectra. Therefore, even without producing an abundance of cross-ring cleavage products, assignments based on product ions primarily generated by glycosidic bond cleavages were still easily capable of differentiation between the 4- and 6-sulfated GalNAcs. When the modular MS^n fragmentation scheme is combined with the capillary RPLC separation

of sulfation isomers, the resulting method is capable of handling even complex isomeric mixtures of CS oligomers.

In addition, the methodology presented here has been demonstrated to identify CS sulfation patterns for both saturated and unsaturated hexasaccharides by simple mass shift corrections to the MS^n spectra of the non-reducing terminus. The method described here can be extended to any CS hexasaccharides, regardless of the sulfation pattern, sulfate composition, or non-reducing end saturation/unsaturation, without using specific synthetic standard. Extension to larger oligosaccharides also is a current issue being investigated. In octasaccharides, for example, a non-reducing end disaccharide product ion and a reducing end hexasaccharide product ion would be generated after performing MS-MS, in which the disaccharide product can be identified by its MS^3. The hexasaccharide product can be further fragmented stepwise into positional disaccharides by MS^n, with each disaccharide identified as described in present work. Although, the characterization of oligosaccharides larger than hexasaccharides using such a strategy must require a unique multistage fragmentation path as well as higher stages of tandem mass spectrometry, such an extension should not be prohibitive assuming sufficient signal exists to support the additional layers of MS^n required. Similarly, we are currently investigating the analysis of other sulfated GAGs, including more diverse GAGs such as heparan sulfate, where we anticipate utilizing trideuterated acetyl groups to differentiate between N-acetylated and N-sulfated amino sugars. We anticipate that, with further refinements, this methodology will be expanded to the analysis of sulfation patterns of oligomers from all classes of sulfated GAGs. This is a very efficient analytical tool for characterization of these polysaccharides that are structurally heterogeneous, but widely relevant biomedically.

Acknowledgements

This work was supported by grant from National Center for Research Resources as part of the Research Resources for Integrated Glycotechnology (P41RR005351). Vitor H. Pomin was partially supported by a post-doctoral fellowship (PDE 201019/2008-6) from Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq. We are also grateful to Christian Heiss and Parastoo Azadi for advice on the permethylation of CS oligosaccharides.

References

1. Perrimon, N.; Bernfield, M. *Semin. Cell Dev. Biol.* **2001**, *12*, 65-67.
2. Deepa, S. S.; Yamada, S.; Fukui, S.; Sugahara, K. *Glycobiology* **2007**, *17*, 631-645.
3. Martel-Pelletier, J.; Tat, S. K.; Pelletier, J.-P. *Osteoarthritis and Cartilage* **2010**, *18*, S7-S11.
4. Kirn-Safran, C. B.; D'Souza, S. S.; Carson, D. D. *Semin. Cell Dev. Biol* **2008**, *19*, 187-193.
5. Whitelock, J. M.; Iozzo, R. V. *Chem. Rev.* **2005**, *105*, 2745-2764.
6. Bandtlow, C. E.; Zimmermann, D. R. *Physiol. Rev.* **2000**, *80*, 1267-1290.
7. Linhardt, R. J.; Toida, T. *Acc. Chem. Res.* **2004**, *37*, 431-438.
8. Plaas, A. H.; West, L. A.; Wong-Palms, S.; Nelson, F. R. *J. Biol. Chem.* **1998**, *273*, 12642-12649.
9. Nasser, N. J. *Cell Mol. Lifer Sci.* **2008**, *65*, 1706-1715.
10. Seeberger, P. H.; Werz, D. B. *Nature* **2007**, *446*, 1046-1051.
11. Kinoshita, A.; Sugahara, K. *Anal. Biochem.* **1999**, *269*, 367-378.
12. Yamada, S.; Sugahara, K. *Meth. Mol. Biol.* **2003**, *213*, 71-78.
13. Pomin, V. H.; Sharp, J. S.; Li, X.; Wang, L.; Prestegard, J. H. *Anal. Biochem.* **2010**, *82*, 4078-4088.
14. Dell, A.; Rogers, M. E.; Thomas-Oates, J. E. *Carbohydr. res.* **1988**, *179*, 7-19.
15. Linhardt, R. J.; Wang, H. M.; Loganathan, D.; Lamb, D. J.; Mallis, L. M. *Carbohydr. res.* **1992**, *225*, 137-145.

16. Laremore, T. N.; Murugesan, S.; Park, T. J.; Avci, F. U.; Zagorevski, D. V.; Linhardt, R. J. *Anal. Chem.* **2006**, *78*, 1774-1779.

17. Juhasz, P.; Biemann, K. *Carbohydr. res.* **1995**, *270*, 131-147.

18. Chai, W.; Luo, J.; Lim, C. K.; Lawson, A. M. *Anal. Chem.* **1998**, *70*, 2060-2066.

19. Saad, O. M.; Leary, J. A. *Anal. Chem.* **2003**, *75*, 2985-2995.

20. Yang, H. O.; N, S., Gunay; Toida, T.; Kuberan, B.; G, Y.; Kim, Y. S.; Linhardt, R. J. *Glycobiology* **2000**, *10*, 1033-1039.

21. Wei, W.; Ninonuevo, M. R.; Sharma, A.; Danan-Leon, L. M.; Leary, J. A. *Analytical chemistry* **2011**, *83*, 3703-3708.

22. Zaia, J.; Costello, C. E. *Anal. Chem.* **2003**, *75*, 2445-2455.

23. Saad, O. M.; Ebel, H.; Uchimura, K.; Rosen, S. D.; Bertozzi, C. R.; Leary, J. A. *J. A. Glycobiology* **2005**, *15*, 818-826.

24. Saad, O. M.; Leary, J. A. *Anal. Chem.* **2005**, *77*, 5902-5911.

25. Wolff, J. J.; Chi, L.; Linhardt, R. J.; Amster, I. J. *Anal. Chem.* **2007**, *79*, 2015-2022.

26. Zaia, J. *Mass Spectrom. Rev.* **2009**, *28*, 254-272.

27. Naimy, H.; Leymarie, N.; Bowman, M. J.; Zaia, J. *Biochemistry* **2008**, *47*, 3155-3161.

28. Hitchcock, A. M.; Yates, K. E.; Costello, C. E.; Zaia, J. *Proteomics* **2008**, *8*, 1384-1397.

29. Hitchcock, A. M.; Costello, C. E.; Zaia, J. *Biochemistry* **2006**, *45*, 2350-2361.

30. Estrella, R. P.; Whitelock, J. M.; pacher, N. H.; Karlsson, N. G. *Anal. Biochem.* **2007**, *79*, 3597-3606.

31. Bielik, A. M.; Zaia, J. *Int. J. Mass Spectrom.* **2010**.

32. Solakyildirim, K.; Zhang, Z.; Lindhart, R. J. *Anal. Biochem.* **2010**, *397*, 24-28.

33. Karlsson, N. G.; Schulz, B. L.; Pacher, N. H.; Whitelock, J. M. *J. Chromatogr. B* **2005**, *824*, 139-147.

34. Heiss, C.; Wang, Z.; Azadi, P. *Rapid communications in mass spectrometry : RCM* **2011**, *25*, 774-778.

35. Nagasawa, K.; Inoue, Y.; Tokuyasu, T. *J. Biochem.* **1979**, *86*, 1323-1329.

36. Bendiak, B.; Fang, T. T.; Jones, D. N. M. *Can. J. Chem.* **2002**, *80*, 1032-1050.

37. Taguchi, T.; Iwasaki, M.; Muto, Y.; Kitajima, K.; Inoue, S.; Khoo, K. H.; Morris, H. R.; Dell, A.; Inoue, Y. *Eur. J. Biochem* **1996**, *238*, 357-367.

38. Lei, M.; Mechref, Y.; Novotny, M. V. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1660-1671.

39. Domon, B.; Costello, C. E. *Glycoconj. J.* **1988**, *5*, 397-409.

CHAPTER 3

AN APPROACH FOR SEPARATION AND COMPLETE STRUCTURAL SEQUENCING OF HEPARIN/HEPARAN SULFATE-LIKE OLIGOSACCHARIDES[2]

[2] Huang, R., Liu, J., Sharp, J.S., *Anal. Chem.*, **2013**, *85*, 5787-5795. Reprinted here with permission of publisher.

Abstract

As members of the glycosaminoglycan (GAG) family, heparin and heparan sulfate (HS) are responsible for mediation of a wide range of essential biological actions, most of which are mediated by specific patterns of modifications of regions of these polysaccharides. To fully understand the regulation of HS modification and the biological function of HS through its interactions with protein ligands, it is essential to know the specific HS sequences present. However, the sequencing of mixtures of HS oligosaccharides presents major challenges due to the lability of the sulfate modifications, as well as difficulties in separating isomeric HS chains. Here, we apply a sequential chemical derivatization strategy involving permethylation, desulfation and trideuteroperacetylation to label original sulfation sites with stable and hydrophobic trideuteroacetyl groups. The derivatization chemistry differentiates between all possible heparin/HS sequences solely by glycosidic bond cleavages, without the need to generate cross-ring cleavages. This derivatization strategy combined with LC-MS/MS analysis has been used to separate and sequence five synthetic HS-like oligosaccharides of sizes up to dodecasaccharide, as well as a highly-sulfated Arixtra-like heptamer. This strategy offers a unique capability for the sequencing of microgram quantities of HS oligosaccharide mixtures by LC-MS/MS.

Introduction

Heparan sulfate (HS) and heparin are linear, highly negatively charged polysaccharides that belong to the glycosaminoglycan family, with molecular weights ranging from 5 to 70kDa[1,2]. Through specific binding to a variety of proteins, HS and heparin have been recognized as key factors for mediation of a wide range of biological actions, such as cell growth control, cell signaling, cell adhesion and migration, inflammation, anticoagulation, neural

development and regeneration[3-7]. Their biological significance makes HS and heparin important targets for drug discovery. One of the most studied heparin protein binding motifs is the pentasaccharide responsible for binding antithrombin III and inhibiting the coagulation cascade[8], which has been formulated into an anticoagulant drug (Arixtra). Ongoing efforts are identifying numerous other examples of protein-binding motifs in HS, which could be potential drug candidates[9,10]. Most of the interactions between HS motifs and proteins are structurally specific, requiring a specific sequence of modifications across an oligosaccharide of moderate length. Changes to the HS biosynthesis pathway, whether by regulation or disease state, can alter such interactions leading to change or loss of function. Thus, detailed information on these heparin/HS oligosaccharide sequences is required for a better understanding of their structure-function relationship, as well as for the development of HS/heparin-based drugs[11].

The diversity of heparin/HS modification is what drives the biology, and what makes sequencing of these polysaccharides so challenging. HS and heparin polysaccharide chains are composed of glucosamine (GlcN) and uronic acid (UA) disaccharide repeat units with various types of modifications including acetylation, sulfation and epimerization of the C-5 position of the UA. In the GlcN residue, the amine group can either be a free amine, and acetylated amine, or a sulfated amine. The 6-O position of the GlcN residue can be sulfated, and in uncommon but biologically important instances, GlcNS can be additionally sulfated at the 3-O position. The uronic acid can be either glucuronic acid (GlcA) or iduronic acid (IdoA) differing by the stereochemistry at the C-5 position, and the uronic acid can be sulfated at the 2-O position[2]. The synthesis of these polysaccharides are not template-driven like DNA, but rather are driven by untemplated enzymatic modification of the sugar backbone, which is mediated by a complex and dynamic suite of modification enzymes[12]. The result is a mixture of heparin/HS sequences that

are both polydisperse and heterogeneously modified, often resulting in isomeric sequences that differ widely in biological functions.

The high degree of heterogeneity and negative charge in these polysaccharides caused by the variety of chain lengths and diverse sulfation patterns makes their structural determination a very challenging task. Rapid progresses in mass spectrometry (MS) instrumentation and chromatography separation technique have led to an increasing use of these methodologies in GAG structural studies[13]. Several MS techniques have been used for structural analyses of GAGs including fast-atom bombardment (FAB)[14,15], matrix-assisted laser desorption-ionization (MALDI)[16-18], and electrospray ionization (ESI)[19-21]. ESI is the most commonly used ionization method for its gentle ionization giving minimum in-source fragmentation and sulfate loss. Most methods for the detailed structural analysis of heparin/HS involves either complete or partial depolymerization by either enzymatic or chemical means to obtain disaccharides mixtures for compositional analysis or a range of oligosaccharide fractions with different lengths for oligosaccharide analysis[22,23]. These methods always involve separation of HS/heparin oligosaccharides for either off-line purification or on-line liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis, using capillary electrophoresis (CE)[24] or high-performance liquid chromatography (HPLC)[25] including strong anion-exchange chromatography (SAX)[26,27], size-exclusion chromatography (SEC)[21,28], hydrophilic interaction chromatography (HILIC)[29,30], porous graphitized carbon separation (PGC)[31,32], and ion-pairing reversed-phase (IPRP)[33,34] chromatography.

Disaccharide compositional profiling of HS/heparin derivatized disaccharides is useful and allows for quantitative analysis by using commercially available disaccharides as standards. For differentiation of isomeric disaccharides, collision induced dissociation (CID) MS/MS or

multi-stages tandem mass spectrometry (MS^n) is used to generate diagnostic ions for identifying disaccharides, and recent development in IPRP-HPLC or IPRP-UPLC technique enabled separation of 12 common commercially available HS disaccharides allowing the use of retention time for disaccharide identification[20,33-35]. While disaccharide composition analysis uses relatively little sample and has short analysis times, significant sample preparation is still required. Certain regions of heparin/HS have been shown to be resistant to digestion down to disaccharides, biasing the composition results[36]. Another compositional analysis method was developed by Zaia's group using a HILIC LC-MS platform for on-line separation and MS analysis of oligosaccharides with different lengths derivatized from enzyme digestion[29,30,37]. Compositional information like numbers of hexuronic acid, N-acetylglucosamine, sulfate groups and acetyl groups can be obtained for each observed oligosaccahride based on their accurate mass, and heparin/HS glycomics profiles can be obtained and compared for different samples. While both composition analysis methods are useful for studying patterns and trends in heparin/HS biology, they do not reveal the detailed sequence information of the oligosaccharide units required for most HS-protein interactions.

A major impediment to use tandem mass spectrometry for structural sequencing of heparin/HS oligosaccharides is sulfate loss during fragmentation. As heparin/HS is collisionally activated, one of the most common fragmentation pathways is loss of the sulfate modification, resulting in a loss of sequence information regarding the original site of sulfation. It has been demonstrated that the loss of sulfate groups can be minimized using a combination of charge state manipulation and metal ion adduction[19,38,39]. However, delicate optimization of buffer and ionization conditions is required for each oligosaccharide and on-line separation of isomeric sequences is highly limited, narrowing the applicability of such approach. Saad and Leary

introduced a program called heparin oligosaccharide sequencing tool (HOST) for automated sequencing using the results of tandem mass spectrometry for disaccharides produced by enzyme digestion from the target oligosaccharides[40]. The use of such method is limited in application to structurally homogeneous sample because of its incompatibility with online LC separation. Besides the traditional CID, another tandem mass spectrometry technique, electron detachment dissociation (EDD) has recently been applied for GAG structural studies, with the capability of distinguishing GlcA from IdoA[41]. Like the HOST sequencing method, additional purification step for oligosaccharides or samples with very few contaminants is required in order to generate promising structural information.

While chemical derivatizations, for example, permethylation and acetylation, are not as commonly applied for structural analysis of GAG oligosaccharides as for other carbohydrates (such as N- and O-linked glycans), there were studies reported in 1980s using sequential chemical derivatization for structural sequencing of heparin oligosaccharides with fine structure by FAB MS[42], and combined with chemical depolymerization for monosaccharide analysis of heparin polysaccharides by chemical ionization MS coupled with gas-liquid chromatography[43]. We modified this chemical derivatization scheme to improve versatility and yield, and used it to successfully separate and sequence mixtures of another class of GAGs, chondroitin sulfate, by LC-MS^n [44]. We present here a related approach involving sequential permethylation, desulfation and pertrideuteroacetylation to modify native HS oligosaccharides for on-line separation and structural analysis of mixtures of HS oligosaccharides, preserving the information regarding the original sites of sulfation and allowing for clear sequencing of heparin/HS sequences based solely on glycosidic bond cleavages. By replacing the labile and strongly polar sulfate groups with much more stable and hydrophobic trideuteroacetyl groups, the oligosaccharides can be

separated well by reverse-phase capillary HPLC and fragmented by MS/MS without losing information regarding the sites of modification.

Experimental Section

Materials. HS disaccharide standards were purchased from Dextra (Reading, UK). Chemoenzymatically synthesized Arixtra-like heptamer and five HS-like oligosaccharides (two decamers, two dodecamers and one undecamer) were synthesized following the procedures in previous publications[45,46]. Unless otherwise noted, all chemical reagents used in the chemical derivatization scheme were purchased from Sigma-Aldrich Inc. (St. Louis, MO).

Chemical Derivatization of HS Oligosaccharides. A series of chemical derivatizations were performed to replace the labile and strongly polar sulfate groups with much more stable and hydrophobic trideuteroacetyl groups, which enable the HS oligosaccharides to be retained well and separated by RPLC and fragmented by MS/MS without losing the information about the sulfation modifications. Detailed procedures for the chemical derivatization have been described in our previous work for structural analysis of CS oligosaccharides[44]. To adjust the method for HS oligosaccahride analysis, some modifications have been made. Briefly, HS oligosaccharides were converted to triethylamine (TEA) salts before being permethylated, in order to increase their solubility in dimethyl sulfoxide (DMSO)[47]. For permethylation, the dried TEA salts (10-50 μg) were re-suspended in 200 μL DMSO and 200 μL anhydrous suspension of sodium hydroxide in DMSO (150 μg/μL) followed by addition of 100 μL iodomethane. After 5 min vortexing and 10 min sonicating, the reaction was stopped by adding 2 mL water and sparged with nitrogen to remove iodomethane, followed by desalting using a C18 Sep-Pak cartridge (Waters Co.). The dried permethylated products were then converted to their pyridinium salts for solvolytic desulfation by dissolving in 20μL DMSO containing 10% methanol and incubated for

4h at 95°C to remove the sulfate groups[48]. The solvents were lyophilized and the dried products were re-suspended in 175μL pyridine, 25μL D_6-acetic anhydride and incubated at 50°C overnight to label the original sites of sulfation with trideuteroacetyl groups[49] . The solvents were then removed by using a Speed-Vac concentrator, and the samples were re-suspended in 20% acetonitrile/water at a concentration of 0.2μg/μl for later analysis. An estimate of yield of fully derivatized product based on UV analysis of a dp4 mixture of heparan sulfate was 29.5% (data not shown).

LC-MS/MS Analysis. For structural analysis of each HS oligosaccharides, a RPLC-MS/MS method was used. Buffer A was prepared as water with 1mM sodium acetate, and buffer B was 80% acetonitrile, 20% water with 1mM sodium acetate. Online HPLC was performed on a regular porous capillary C18 column (0.2×50mm, 3μm, 200 Å, Michrom Bioresources, Aubum, CA), using a linear gradient of buffer B from 25%- 100% over 60min, with a flow rate of 4μL/min and a 10μL injection at a sample concentration of 0.2μg/μL. Mass spectrometry was performed on either a Thermo LTQ-FT instrument or Waters Synapt G2 Q-TOF mass spectrometer. Full MS and CID-MS/MS spectra were acquired in positive ion mode, with spray voltage of 2-3kV and capillary temperature 250ºC for LTQ-FT and 80ºC for Q-TOF. The collision energy was set between 40V and 50V.

By mixing the five synthesized HS-like oligomers together, online RPLC separation of HS oligosaccharides using different C18 packing material was also compared. In addition to the regular porous C18 column as we mentioned above, we also used a Halo C18 column with core-shell packing materials (0.2×50mm, 2.6μm, 160 Å, Advanced Material Technology, Wilmington, DE) for online separation. For the porous C18 column, a 70min gradient was used from 40-100% buffer B with flow rate of 4μL/min. For the Halo C18 column, an 18min gradient

was used from 35%-80% buffer B with flow rate of 9µL/min. A10µL injection at a sample concentration of 0.2µg/µL for each of the five oligosaccharides was used for analysis. Mass spectrometry setup was as same as we mentioned above.

Results and Discussion

Chemical Derivatizations of HS Disaccharide Standards. As listed in **Figure 3.1**, the 12 common HS disaccharide standards contain the basic structure ΔUA-GlcN, where the amine group can either be a free amine, acetylated amine, or a sulfated amine. O-sulfation can occur at the 6-O position of the GlcN residue and/or the 2-O position of the ΔUA residue. Five HS disaccharides are selected to illustrate how all possible modifications can be differentiated from each other after derivatization by MS and/or MS/MS analysis. Structures of these disaccharides are presented in **Figure 3.1**, as well as the structure for each derivatized product (with MS and MS/MS spectra of selected ones shown in **Figure 3.2**). Comparison of the derivatized products for ΔUA2S-GlcNAc6S and ΔUA2S-GlcNS6S (**Figure 3.1, I-A and I-S**) illustrate the reason for the use of D_6-acetic anhydride, as the use of trideuteroacetylation allows for the differentiation between native N-acetylation and peracetylation during the derivatization by mass only (**Figure 3.2A and B**).

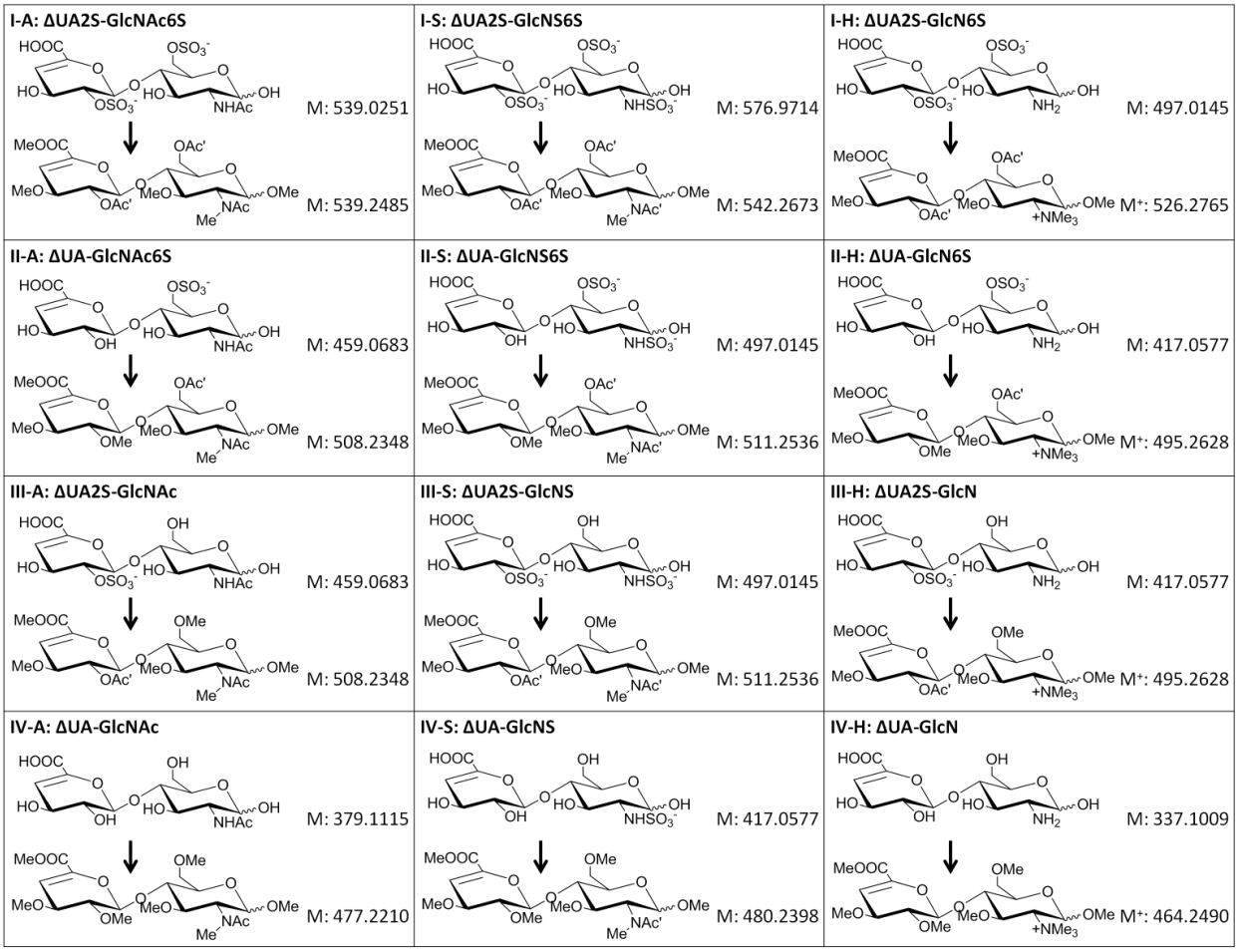


Figure 3.1. Structures of 12 commercially available common HS disaccharides before and after the chemical derivatizations, with each corresponding molecular weight labeled. In the structures, the symbol Me, Ac and Ac' represents for methyl group, acetyl group and trideuteroacetyl group respectively. GlcN, GlcNAc and GlcNS represents for free glucosamine, acetylated glucosamine and N-sulfated glucosamine respectively. 2S and 6S indicate the position of the sulfate group. ΔUA represents for unsaturated uronic acid.

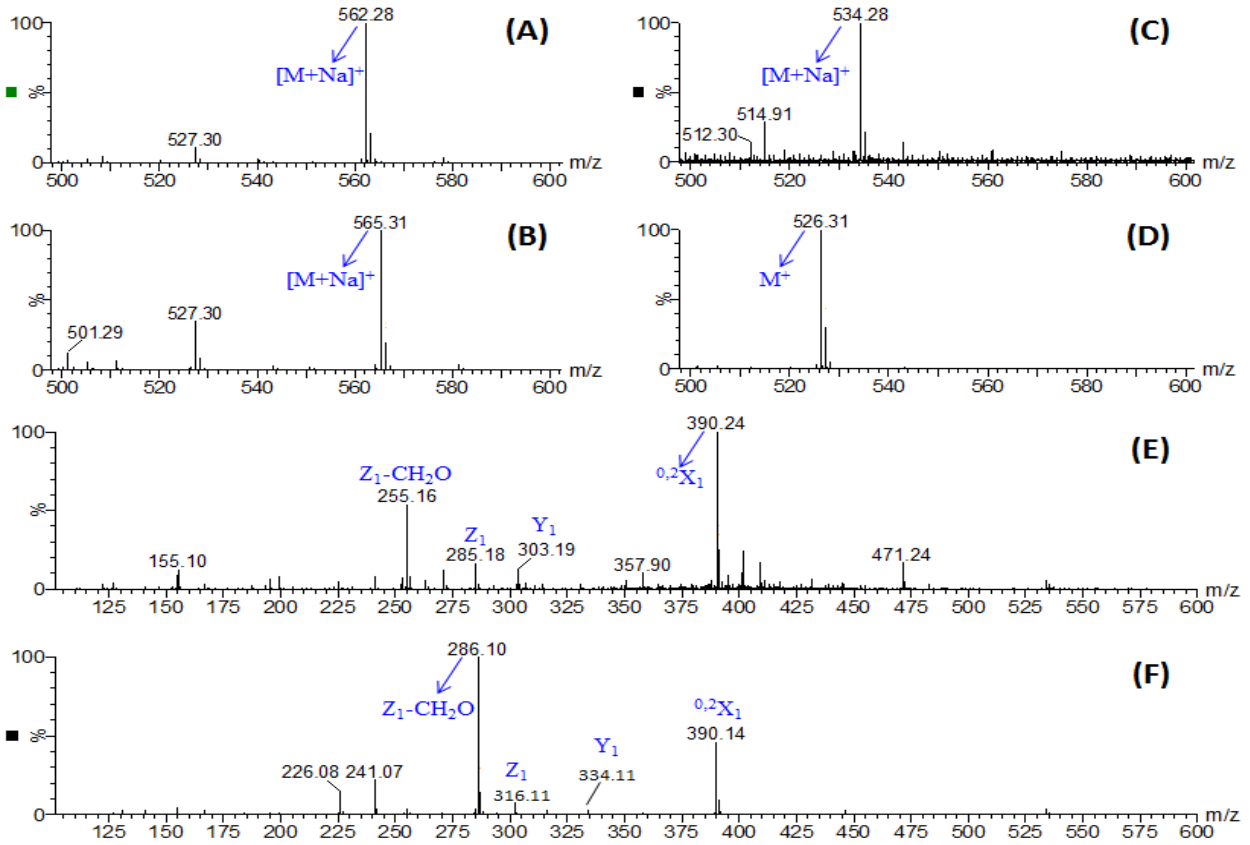


Figure 3.2. MS spectra for derivatized ΔUA2S-GlcNAc6S (A), ΔUA2S-GlcNS6S (B), ΔUA2S-GlcNS (C), ΔUA2S-GlcN6S (D) and MS/MS spectra for ΔUA2S-GlcNS (E), ΔUA-GlcNS6S (F). A mass difference of 3Da were observed for derivatized ΔUA2S-GlcNAc6S and ΔUA2S-GlcNS6S, indicating the use trideuteroacetylation enable the differentiation between native N-acetyl group and trideuteroacetyl group from peracetylation step (spectra **A** and **B**). For derivatized ΔUA2S-GlcNS and ΔUA2S-GlcN6S, a mass difference of 14Da was observed between derivatized ΔUA2S-GlcN6S (detected as M^+) and ΔUA2S-GlcNS (detected as $[M+H]^+$), or a mass difference of 8Da between derivatized ΔUA2S-GlcN6S (detected as M^+) and ΔUA2S-GlcNS (detected as $[M+Na]^+$) (spectra **C** and **D**). For another pair of sulfation positional isomers, the Y_1 and Z_1 ions containing the GlcN residue were 31Da larger for ΔUA-GlcNS6S than ΔUA2S-GlcNS, indicating an original sulfation modification was on GlcN for ΔUA-GlcNS6S and one on ΔUA for ΔUA2S-GlcNS (spectra **E** and **F**).

For sulfation positional isomers, there are three common sulfation sites as we mentioned, including N-sulfation, 2-O-sulfation and 6-O-sulfation. By comparing two isomeric dissacharides ΔUA2S-GlcNS and ΔUA2S-GlcN6S (**Figure 3.1,** disaccharides III-S and I-H), the differentiation between N-sulfation and O-sulfation isomers can be easily achieved only by their mass difference after derivatization(**Figure 3.2C and D**). As it has been reported in other literature about the permethylation of different types of amine groups[50], our result also showed that the free amine group in GlcN residue was converted to trimethyl quaternary amine that could not be further modified, while a sulfated amine was converted to monomethylated sulfoamine that was desulfated and trideuteroactylated afterwards. This resulted in a diagnostic mass difference of 14Da between derivatized ΔUA2S-GlcN6S (detected as M^+) and ΔUA2S-GlcNS (detected as $[M+H]^+$), which in their underivatized forms are isomers.

Differentiation of 2-O-sulfation and 6-O-sulfation of the glucosamine by MS analysis is more difficult. For this isomeric pair, ΔUA-GlcNS6S and ΔUA2S-GlcNS (**Figure 3.1,** disaccharides II-S and III-S), the derivatized products were still isomers with exact same mass which require MS/MS analysis for differentiation. For the derivatized ΔUA-GlcNS6S, there was a trideuteroacetyl group on the 6-O position of the GlcNS residue and a methyl group on the 2-O position of the ΔUA residue. On the opposite side, the derivatized ΔUA2S-GlcNS had a methyl group on the 6-O position of the GlcNS residue and a trideuteroacetyl group on the 2-O position of the ΔUA residue. Compared to the derivatized ΔUA2S-GlcNS, the Y_1 and Z_1 ions of the derivatized ΔUA-GlcNS6S would be 31Da heavier (the mass difference between a methyl group and a trideuteroacetyl group), and the B_1 and C_1 ions would be 31Da lighter (**Figure 3.2E and F)**. One of the major benefits of our derivatization strategy is the greatly increased stability of the trideuteroacetylations as compared to the sulfation, with negligible trideuteroacetyl losses from

the fragment ions upon CID. This stability preserves the information of the site of original sulfation, allowing the site of sulfation to be easily determined. The differentiation of this isomer pair is achieved by simply comparing the glycosidic bond cleavage fragments generated from the MS/MS, without requiring for the generation of cross-ring cleavage fragments as required for chondroitin sulfate[44].

With these initial results from the HS disaccharide analyses, we demonstrated that the chemical derivatization strategy enable the differentiation between all common sulfation patterns based on mass or solely on glycosidic bond cleavages, and therefore we can accurately sequence HS oligosaccharides consisted with these repeating disaccharide units up to any length amenable to glycosidic bond cleavage. However, no disaccharides containing the rare, but biologically important, sulfation at the 3-O position of the GlcN are commercially available. The derivatization products and characteristics of 3-O sulfated GlcN are described below as studied using an Arixtra-like heptamer.

Structural Sequencing and LC Separation of Five Synthesized HS-like oligosaccharides. To evaluate the application of our method to longer HS oligosaccharides, we performed sequential chemical derivatizations and LC-MS/MS analysis of five chemoezymatically synthesized HS-like oligosaccharides with known sequences (**Table 3.1)**. Full derivatization was achieved for all five oligosaccharides with byproducts caused primarily by the β–elimination reaction between the carbon-4 and -5 of the GlcA during the permethylation, with a smaller amount of byproducts apparently formed during the desulfation procedure. As mentioned in our previous work on CS oligosaccharides[44], these byproducts are notable for not introducing species that would lead to the false identification of sites of sulfation;

their major effect is reducing the apparent sensitivity of the technique and result in shortened oligomer sequences (direct infusion MS spectrum shown in **Figure 3.3**).

Table 1. **Sequences for five chemoenzymatic synthesized HS-like oligosaccharides.**

NS-decamer	GlcA-GlcNS-(GlcA-GlcNS)$_3$-GlcA-AnMan
NS6S-decamer	GlcA-GlcNS6S-(GlcA-GlcNS6S)$_3$-GlcA-AnMan
NS-undecamer	GlcNS-(GlcA-GlcNS)$_4$-GlcA-AnMan
NS-dodecamer	GlcA-GlcNS-(GlcA-GlcNS)$_4$-GlcA-AnMan
NS6S-dodecamer	GlcA-GlcNS6S-(GlcA-GlcNS6S)$_4$-GlcA-AnMan

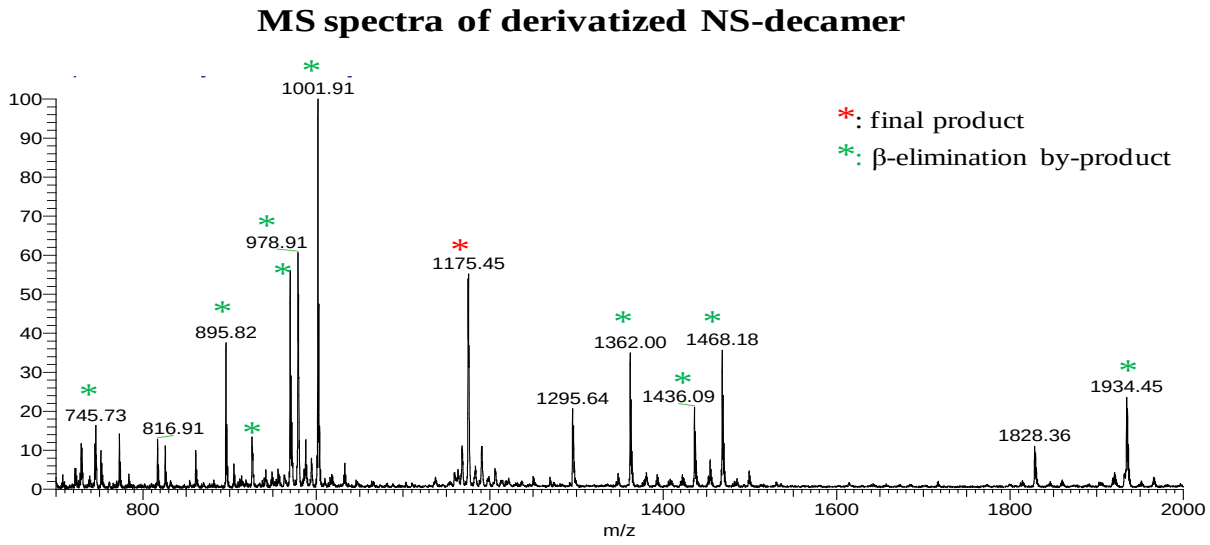


Figure 3.3. Direct infusion MS spectrum of the desalted derivatization products of the NS-decamer. The byproducts are all beta-elimination products that are separated in the reverse phase LC run, and do not result in erroneous sequencing of the product, just in lower sensitivity.

From the MS/MS spectra of the parent ion $[M+2Na]^{2+}$ for each oligosaccharides (with two decamers, two dodecamers and one undecamer shown in **Figure 3.4**), we observed sufficient sequential glycosidic bond cleavage fragments to enable accurate, full structural sequencing of

the modifications. Comparing the series of Y ions for NS-decamer and NS6S-decamer (**Figure 3.4A and B**), the Y ions of the NS6S-decamer gave additional n×31Da mass than the Y ions of the NS-decamer, where n is the number of the basic GlcN residue the Y ion has. For example, Y_2 ions with no GlcN residue shared the same m/z of 447 for each decamer, and Y_4 ions with only one GlcN residue had 31Da mass difference, and Y_6 ions with two GlcN residues had 62Da mass difference, and so on. Together with the similar results obtained for other three longer oligosaccharides, the chemical derivatization strategy was demonstrated to enable successful structural sequencing of HS-like oligosaccharides up to dodecamer (the longest we attempted to sequence) with a single MS/MS experiment using glycosidic bond cleavages only. We did notice detectable losses of MeOH and AcOH from product ions, with these losses correlating strongly with the derivatization state at the 6O-position of each precursor. These neutral mass losses from the glycosidic bond cleavage products are always paired with the intact glycosidic bond cleavage product, making assignment of the product ion straightforward. Additionally, the neutral loss product cannot be mistaken for a different sequence, as the loss of MeOH or AcOH results in a mass not consistent with a fully derivatized glycosidic bond cleavage of another related sequence.

MS/MS of the derivatized HS results primarily in glycosidic bond cleavages. Due to the unique mass differences for each monosaccharide unit after derivatization, interpretation of the MS/MS spectra is similar to interpretation of peptide MS/MS spectra. For example, we refer to the compound analyzed in **Figure 3.4A**. Starting at the non-reducing end, if the monosaccharide is GlcA, we should observe m/z of 1058.7 for Y9 ion; while if it is GlcA2S, an m/z of 1043.2 should be observed. The presence of 1058.7 ion and absence of 1042.7 ion indicate that it is a GlcA residue at the non-reducing end. Following with GlcA, the 2nd residue could be GlcNS,

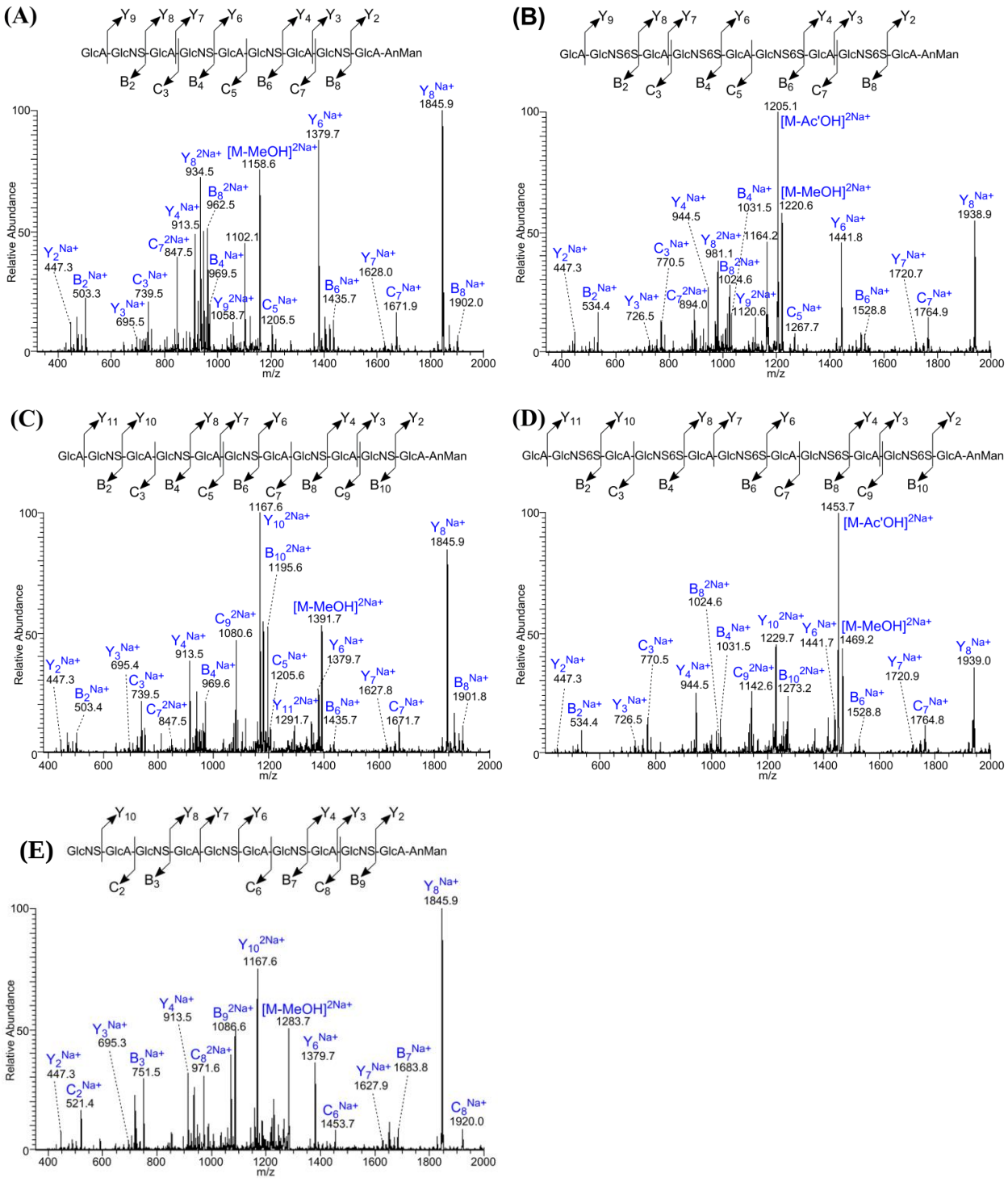


Figure 3.4. MS/MS spectra from the LTQ analyzer of an LTQ-FT of $[M+2Na]^{2+}$ for NS-decamer (**A**), NS6S-decamer (**B**), NS-dodecamer (**C**), NS6S-dodecamer (**D**) and NS-undecamer (**E**) after the chemical derivatizations.

GlcNAc, GlcNS6S, GlcNAc6S, GlcNS3S, GlcNAc3S, GlcNS3S6S, GlcNAc3S6S; and the corresponding B2 ion for each possibility should give m/z of 503.3, 500.3, 534.3, 531.3, 520.3, 517.3, 551.3 and 548.3. From the MS/MS spectra, we only observed the 503.3 ion without clear signs for other ions, which indicate the second residue is GlcNS. For the 3rd residue, we observed C3 ion with m/z of 739.5 but not 770.5, indicating it is a GlcA not GlcA2S. For the 4th residue, we observed B4 ion with m/z of 969.5 but no other possibilities (966.5, 1000.5, 997.5, 986.5, 983.5, 1017.5 and 1014.5), indicating it is a GlcNS. Sequencing for the rest of residues is very similar. The C5 ion with m/z of 1205.5 but not 1236.5 indicates that it is a GlcA at the 5th residue. The B6 ion with m/z of 1435.7 but no other possibilities (1432.7, 1466.7, 1463.7, 1452.7, 1449.7, 1483.7 and 1480.7) indicates that it is a GlcNS at the 6th residue. The C7 ion with m/z of 1671.9 but not 1702.9 indicates that it is a GlcA at the 7th residue. The B8 ion (doubly charged) with m/z of 962.5 but not others (961.0, 978.0, 976.5, 971.0, 969.5, 986.5 and 985.0) indicates that it is a GlcNS at the 8th residue. The last two residues –GlcA-AnMan at the reducing end are fixed according to the synthesis protocol, and is proven by the presences of Y2 ion with m/z of 447.3 for all five oligosaccharides. The absence of 478.3 for Y2 ion also confirmed that the GlcA is not sulfated (see **Figure 3.5** for zoomed in views of the appropriate regions in the MS/MS spectrum).

Initial attempts at separation of the mixture of derivatized HS oligosaccharides was performed on a standard C18 column, by mixing the five HS-like oligosaccharides together for LC-MS-MS analysis. As shown in **Figure 3.6A**, with the optimized LC gradient, 4 of the 5 oligosaccharides were separated from each other, while NS-undecamer and NS-decamer were overlapped. To seek better separation, we moved to a fused core, porous shell Halo C18 column. As shown in **Figure 3.6B**, 5 oligosaccharides were separated, with better peak shapes and

narrower elution times. The core-shell design of the Halo column significantly reduced the back pressure and allowed for a higher flow rate (9µL/min compare to 4µL/min for standard C18 column) to be used under the pressure limit of 400 bar. With their own optimized conditions for these two types of C18 column, the Halo column of the same length and similar particle size was able to generate narrower peak widths and better peak shapes in a shorter time scale. Therefore, for on-line separation of derivatized HS oligosaccharides, the core-shell Halo column is preferred for its higher separation efficiency.

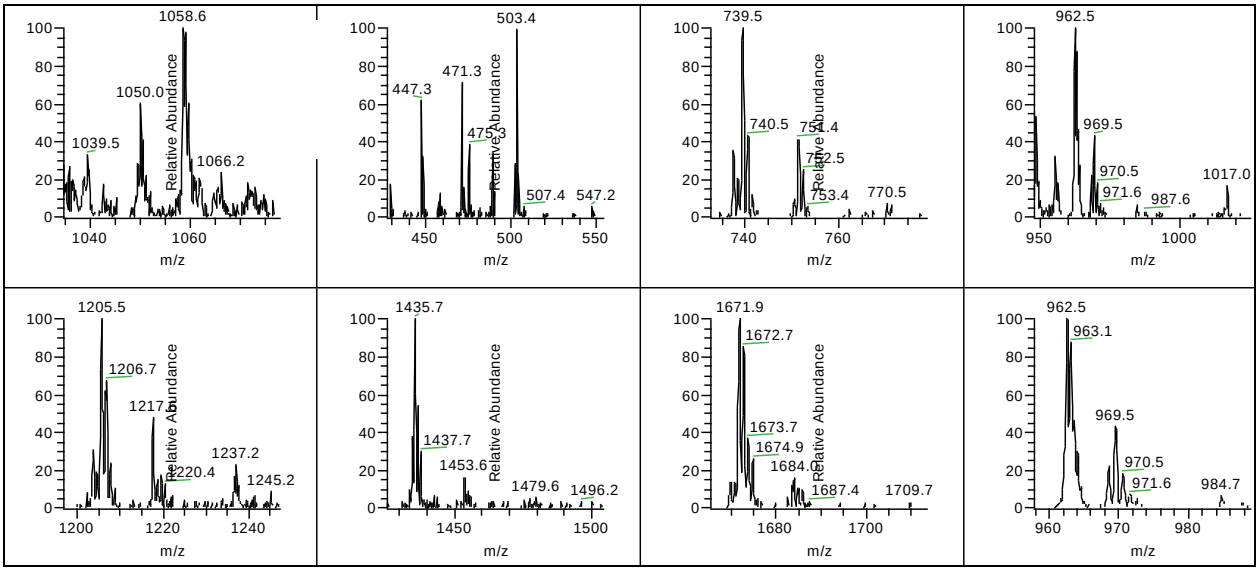


Figure 3.5. Zoomed view of MS/MS spectra for the fully derivatized synthesized NS-decamer, shown in **Figure 3.4A**. These zoomed-in views of the MS/MS spectra clearly show, as described in the manuscript, how an unambiguous interpretation of the MS/MS spectra of fully derivatized HS oligomers can be achieved.

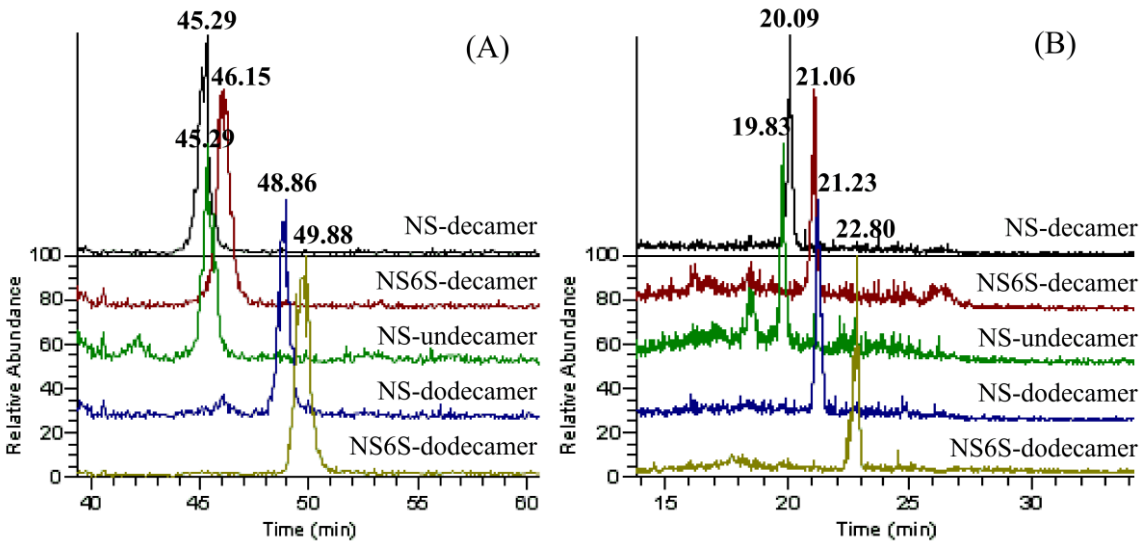


Figure 3.6. LC separation of derivatized products of five synthesized HS-like oligosaccharides by reverse-phase C18 column packed with either traditional porous material (**A**) or fused core-porous shell material (**B**). Same colors in two chromatograms represent the same oligasaccharide as indicated on the right side. For porous C18 column (50mm, 200Å, 3μm): 4 of the 5 oligosaccharides were separated, with NS-hendecamer and NS-decamer co-eluted at 45.29min. For core-shell Halo C18 column (50mm, ~160Å, ~2.6μm): 5 oligosaccharides were able to be separated from each other or at least partially separated.

Structural Analysis of Synthesized Arixtra-like heptamer containing 3-O-sulfation. Compared to N-sulfation and 6-O-sulfation on GlcN residue, 3-O-sulfation is a relatively rare modification for HS oligosaccharides that usually occurs only on N-sulfated GlcN. 3-O sulfation is known to be essential for many HS-protein interactions, including antithrombin, herpes simplex virus 1 glycoprotein D, and growth factor receptor and fibroblast growth factor 7, indicating the important roles of this rare modification in diverse biological functions[51]. For structural analysis of samples with 3-O sulfation present, there is one more isomer pair to be

considered for structural sequencing of HS oligosaccharides, which involved the differentiation between 3-O-sulfated GlcNS and 6-O-sulfated GlcNS. Unfortunately, no commercially-available 3-O sulfated disaccharide is available, and we were unable to purify or synthesize such a disaccharide in any usable quantity. An Arixtra-like heptamer containing a GlcNS6S3S residue (**Figure 3.7A**) was analyzed to study the derivatization properties of 3-O sulfated GlcNS, and the 3-O-sulfated residue was found to have a curious derivatization product resulting from our standard chemical derivatization procedure. After performing complete derivatization for the heptamer, we observed the derivatization product m/z of 901.48 with 14Da less than the theoretical calculated m/z for complete derivatized heptamer (**Figure 3.8**). This underpermethylation was consistent across three attempts, suggesting that the unexpected underpermethylation was not due to improper sample handling. Subsequent derivatization efforts with other heparan sulfate oligosaccharides not containing 3-O sulfation did not exhibit underpermethylation. To determine if the underpermethylation was site-specific, and to locate the position of underpermethylation, we ran a MS/MS experiment on the parent ion with m/z of 901.48 and interpreted the spectra by comparing the observed m/z value with the theoretical m/z value for glycosidic bond cleavage fragments of the fully derivatized product (**Figure 3.7C**). Compared to the theoretical m/z value, the reduction of 14Da started from the Y_5 and B_3 ions, with Y_2, Y_4 and B_2 ions sharing the same m/z with their theoretical values. With these results, we located the under-methylated position to the sulfated amine group on the GlcNS6S3S residue, which contains only one possible position for methylation, the sulfated amine. This also explained why the under-permethylated site was not further trideuteroacetylated, because the free amine group generated after desulfation will only be trideuteroacetylated to a mono-trideuteroacetylated amine group but not to a di-trideuteroacetylated amine group.

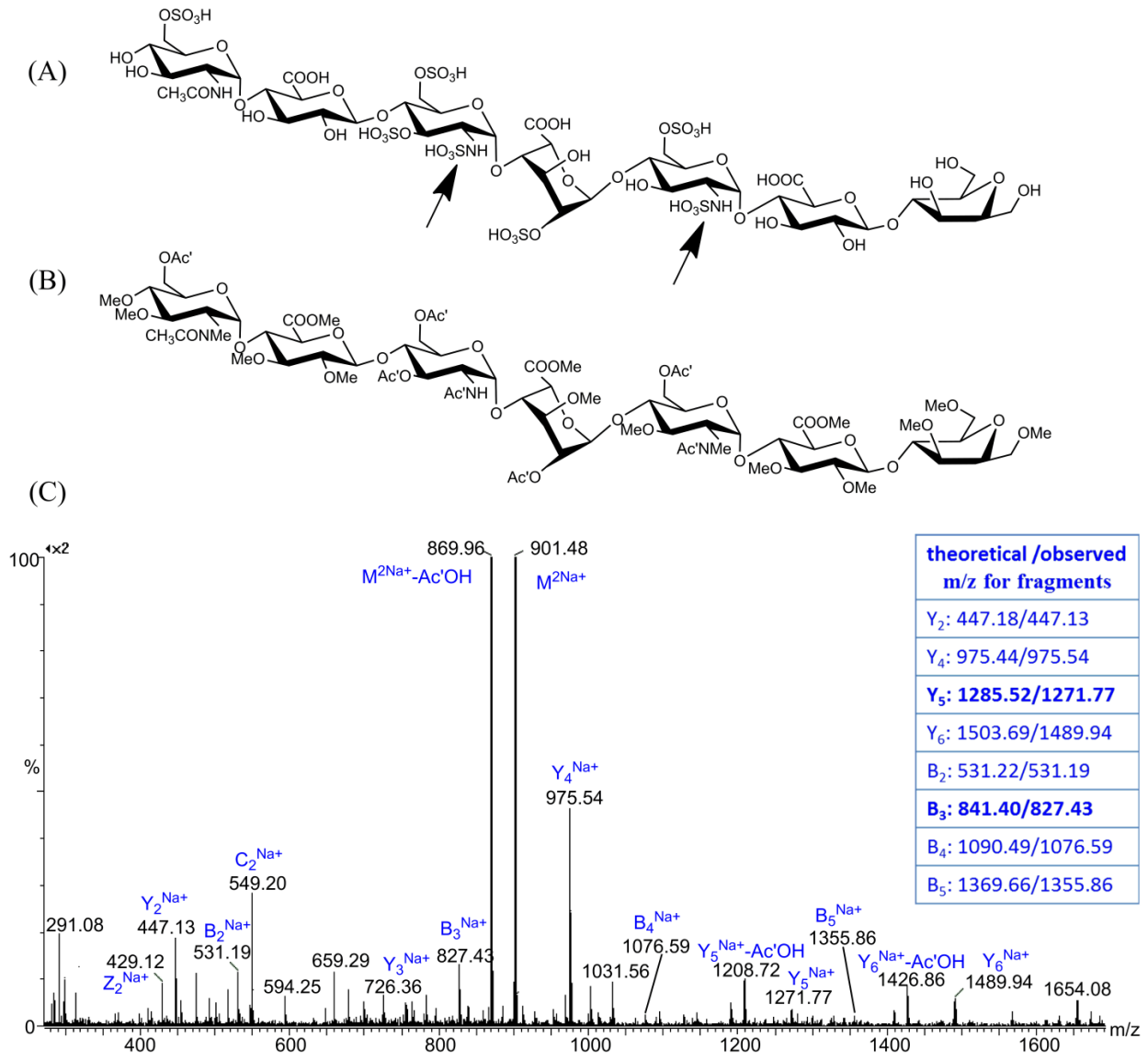


Figure 3.7. Structures of the synthesized Arixtra-like heptamer before (**A**) and after (**B**) the derivatizations. Two black arrows indicated the different behavior during the derivatization between GlcNS6S and GlcNS6S3S residues, which was interpreted from the MS/MS spectra and the comparison between theoretical and observed m/z values for each glycosidic bond cleavage fragment as obtained from a Synapt G2 HDMS Q-TOF (**C**).

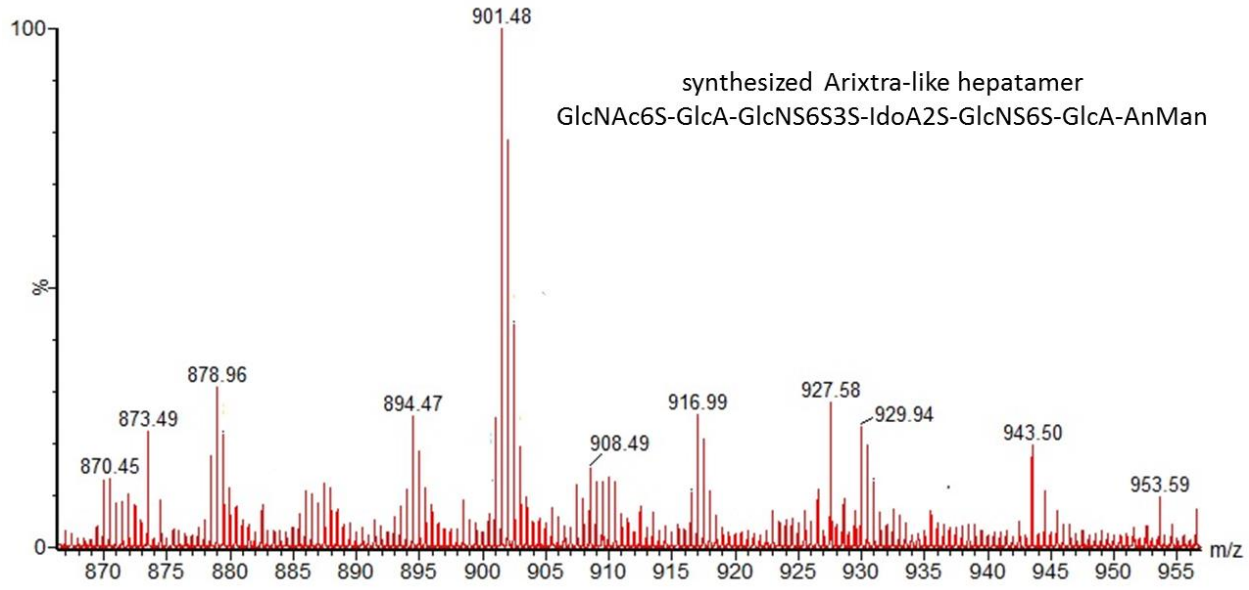


Figure 3.8. MS spectra for the fully derivatized synthesized Arixtra-like hepatamer. As we can see from the spectra, after derivatization the most abundant ion we observed is 901.48, while the theoretical value for a fully permethylated, desulfated and acetylated product should give a m/z of 908.49 for $[M+2Na]^{2+}$ ion. The mass difference between hydrogen and methyl group is also 14Da, indicating that there is one under-permethylation site which could not be further acetylated.

From the structure of the derivatized heptamer (**Figure 3.7B**), we found that the sulfated amine group only became inert when the 3-O position of the same GlcN residue were sulfated, based on the fact that GlcNS6S could be fully methylated while GlcNS6S3S could not. This finding is consistent with previously reported NMR studies showing that the proton on the sulfated amine group of a 3-O sulfated GlcNS is deactivated, exhibiting a much slower exchange rate compared to the amine proton on GlcNS with no 3-O sulfation[52]. With these data, as illustrated in **Figure 3.9**, we conclude that the derivatized 3-O-sulfated and 6-O-sulfated isomers

should be differentiated solely by their 14Da mass difference from the underpermethylation of the 3-O sulfated GlcNS.

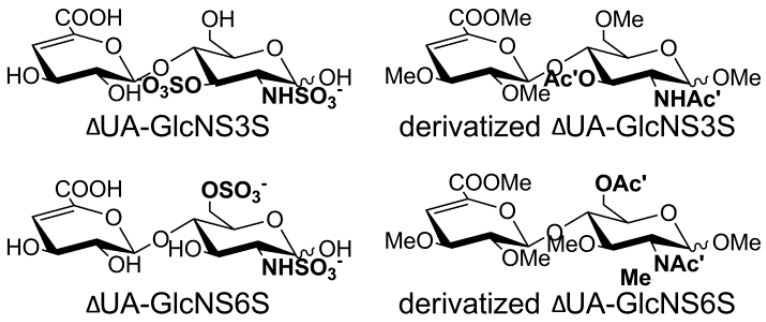


Figure 3.9. Prediction for the structure of the derivatized product of 3-O-sulfated HS disaccharide ΔUA-GlcNS3S, and there is a 14Da mass difference between 3-O-sulfated and 6-O-sulfated isomeric disaccharides after derivatizations.

Conclusions

In this work, we have demonstrated that the sequential chemical derivatization strategy enabled differentiation between all sulfation patterns possible based solely on glycosidic bond cleavages from single-stage MS/MS. Therefore, we can accurately sequence all sulfation and acetylation modifications of oligosaccharides up to any length amenable to glycosidic bond cleavage. We have successfully sequenced synthetic HS oligosaccharides up to dodecamers with a single MS/MS experiment solely using glycosidic bond cleavages, which is the longest oligomer we have attempted to sequence so far. In addition, by replacing the labile and strongly polar sulfate groups with much more stable and hydrophobic trideuteroacetyl groups, the oligosaccharides can be retained well and separated by reverse-phase capillary HPLC, allowing for analysis of mixtures of HS oligosaccharides by automated online LC-MS/MS. The fused

core, porous shell Halo C18 columns have been demonstrated to give better separation than traditional C18 columns using the same gradient but at higher flow rates.

At present, the chemical derivatization protocol combined with LC-MS/MS analysis clearly enables the separation and sequencing of HS oligosaccharides in microgram quantities, but further improvements are possible. One of the limitations is the sample loss and beta-elimination byproducts generated during the permethylation, which requires homogeneous length for oligosaccharides being analyzed. One strategy that is currently being pursued by our group to reduce sample loss is the reduction of the carboxylic acid of the uronic acid residues to alcohols to eliminate the beta-elimination during the permethylation step. Additionally, our current technology does not purport to identify C-5 epimerization of the uronic acid. Preliminary chemometric analysis results suggest that epimerization of the C-5 position of uronic acid may be differentiated by our derivatization and LC-MS/MS protocol, but a broader sample set of analytes is required to ensure robustness of the chemometric analysis.

In applications where a complex mixture of depolymerized heparin/HS is to be analyzed, our methodology provides great promise, even with the issues presented by beta-elimination byproducts unresolved. The beta-elimination byproducts do not shift the sequences of modifications in the byproduct; the remaining sequence is a real sequence that existed in the original mixture, merely truncated from its original size. This truncated sequence can still provide valuable information regarding changes in the sequence. If accurate analysis of full length is required (e.g. affinity purification analyses), then all fragments smaller than the dp fraction size isolated from the size exclusion column can be safely ignored, leaving only full-length sequence data if the researcher so desires.

One notable advantage to our LC-MS/MS protocol is the fact that baseline LC resolution of isomers is not required to address mixtures of isomeric sequences. Partial separation of isomeric sequences by as little as one MS scan in the LC-MS/MS run can successfully identify both sequences by plotting the selection product ion chromatogram specific to each isomeric sequence, which is possible given that our derivatization strategy only relies upon the more abundant glycosidic bond fragments to differentiate between all isomers. While the separation of non-isomeric HS-like oligosaccharides has been achieved in regular HPLC system, and we have previously achieved resolution of isomeric chemically-derived GAG sequences of chondroitin sulfate, the improved resolution granted by a nano-UPLC system would be useful for separations of isomeric oligosaccharides mixtures where much higher resolution is required.

In summary, we have successfully developed and demonstrated a strategy for chemical derivatization combined with standard reverse phase LC-MS/MS analysis using CID fragmentation that enables the structural sequencing for heparin/HS oligosaccharides with different modification patterns. We have used this strategy to analyze samples with as little as 5 μg of total starting material for simple mixtures. This strategy shows tremendous promise for sequencing of mixtures of heparin/HS oligosaccharides obtained from complex biological systems or partial purification strategies, when detailed information about sulfation and acetylation modifications is needed for mixtures of oligosaccharides with certain biological activity or consequence. We anticipate that this methodology will provide a practical strategy for sequencing of the "heparanome" from heparin/HS oligosaccharide mixtures from various systems of biological interest, and represents a valuable follow-up to compositional analysis studies.

Acknowledgements

This research is supported in part by the National Institute of General Medical Sciences-funded "Research Resource for Integrated Glycotechnology" (P41 GM103390) from the National Institutes of Health.

References

1. Gandhi, N. S.; Mancera, R. L. *Chem. Biol. Drug Des.* **2008**, *72*, 455-482.

2. Rabenstein, D. L. *Nat. Prol. Rep.* **2002**, *19*, 312-331.

3. Tumova, S.; Woods, A.; Couchman, J. R. *Int. J. Biochem. Cell Biol.* **2000**, *32*, 269-288.

4. Kresse, H.; Schönherr, E. *J. Cell. Physiol.* **2001**, *189*, 266-274.

5. Lyon, M.; Gallagher, J. T. *Matrix Biol.* **1998**, *17*, 485-493.

6. Li, J.-p.; Vlodavsk, I. *Thromb. Haemost.* **2009**, *102*, 823-828.

7. Dityatev, A.; Schachner, M. *Nat. Rev. Neurosci.* **2003**, *4*, 456-468.

8. Capila, I.; Linhardt, R. J. *Angew. Chem. Int. Ed.* **2002**, *41*, 390-412.

9. Lever, R.; Page, C. P. *Nature Reviews Drug Discovery* **2002**, *1*, 140-148.

10. Theocharis, A. D.; Skandalis, S. S.; Tzanakakis, G. N.; Karamanos, N. K. *FEBS Journal* **2010**, *19*, 3904-3923.

11. Guerrini, M.; Beccati, D.; Shriver, Z.; Naggi, A.; Viswanathan, K.; al., e. *Nat. Biltechnol.* **2008**, *26*, 669-675.

12. Carlsson, P.; Kjellén, L. *Handb. Exp. Pharmacol.* **2012**, *207*, 23-41.

13. Chi, L. L.; Amster, J.; Linhardt, R. J. *Curr. Anal. Chem.* **2005**, *1*, 223-240.

14. Mallis, L. M.; Wang, H. M.; Loganathan, D.; Linhardt, R. J. *Anal. Chem.* **1989**, *61*, 1453-1458.

15. Linhardt, R. J.; Wang, H. M.; Loganathan, D.; Lamb, D. J.; Mallis, L. M. *Carbohydr. Res.* **1992**, *225*, 137-145.

16. Tissot, B.; Gasiunas, N.; Powell, A. K.; Ahmed, Y.; Zhi, Z.-I.; Haslam, S. M.; Morris, H. R.; Turnbull, J. E.; Gallagher, J. T.; Dell, A. *Glycobiology* **2007**, *17*, 972-987.

17. Laremore, T. N.; Linhardt, R. J. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1315-1320.

18. Minamisawa, T.; Hirabayashi, J. *Trends Glycosci. Glyc.* **2006**, *18*, 293-312.

19. Zaia, J.; Costello, C. E. *Anal. Chem.* **2003**, *75*, 2445-2455.

20. Saad, O. M.; Leary, J. A. *Anal. Chem.* **2003**, *75*, 2985-2995.

21. Zaia, J.; Costello, C. E. *Anal. Chem.* **2001**, *73*, 233-239.

22. Jandik, K. A.; Gu, K. A.; Linhardt, R. J. *Glycobiology* **1994**, *4*, 289-296.

23. Shively, J. E.; Conrad, H. E. *Biochemistry* **1976**, *15*, 3932-3942.

24. Volpi, N.; Maccari, F.; Linhardt, R. J. *Electrophoresis* **2008**, *29*, 3095-3106.

25. Zaia, J. *Mass Spectrom. Rev.* **2009**, *28*, 254-272.

26. Vives, R. R.; Goodger, S.; Pye, D. A. *Biochem. J.* **2001**, *354*, 141-147.

27. Bruggink, C.; Maurer, R.; Herrmann, H.; Cavalli, S.; Hoefler, F. *J. Chromatogr. A* **2005**, *1085*, 104-109.

28. Hitchcock, A. M.; Costello, C. E.; Zaia, J. *Biochemistry* **2006**, *45*, 2350-2361.

29. Staples, G. O.; Bowman, M. J.; Costello, C. E.; Hitchcock, A. M.; Lau, J. M.; Leymarie, N.; Miller, C.; Naimy, H.; Shi, X.; Zaia, J. *Proteomics* **2009**, *9*, 686-695.

30. Naimy, H.; Leymarie, N.; Bowman, M. J.; Costello, C. E.; Zaia, J. *Biochemistry* **2008**, *47*, 31553161.

31. Karlsson, N. G.; Schulz, B. L.; Pacher, N. H.; Whitelock, J. M. *J. Chromatogr. B* **2005**, *824*, 139-147.

32. Estrella, R. P.; Whitelock, J. M.; pacher, N. H.; Karlsson, N. G. *Anal. Biochem.* **2007**, *79*, 3597-3606.

33. Yang, B.; Chang, Y.; Weyers, A. M.; Sterner, E. S.; Linhardt, R. J. *J.Chromatogr. A* **2012**, *1225*, 91-98.

34. Yang, B.; Weyers, A. M.; Baik, J. Y.; Sterner, E. S.; Sharfstein, S.; Mousa, S. A.; Zhang, F.; Dordick, J. S.; Linhardt, R. J. *Anal. Biochem.* **2011**, *415*, 59-66.

35. Wei, W.; Niñonuevo, M. R.; Sharma, A.; Danan-Leon, L. M.; Leary, J. A. *Anal. Chem.* **2011**, *83*, 3703-3708.

36. Yamada, S.; Yoshida, K.; Sugiura, M.; Sugahara, K. *J. Biol. Chem.* **1993**, *268*, 4780-4787.

37. Leymarie, N.; McComb, M. E.; Naimy, H.; Staples, G. O.; Zaia, J. *Int. J. Mass Spectrom.* **2012**, *312*, 144-154.

38. Ly, M.; Leach, F. E. r.; Laremore, T. N.; Toida, T.; Amster, I. J.; Linhardt, R. J. *Nat. Chem. Biol.* **2011**, *11*, 827-833.

39. Kailemia, M. J.; Li, L.; Ly, M.; Linhardt, R. J.; Amster, I. J. *Anal. Chem.* **2012**, *84*, 5475-5478.

40. Saad, O. M.; Leary, J. A. *Anal. Chem.* **2005**, *77*, 5902-5911.

41. Wolff, J. J.; Chi, L. L.; Linhardt, R. J.; Amster, J. *Anal. Chem.* **2007**, *79*, 2015-2022.

42. Dell, A.; Rogers, M. E.; Thomas-Oates, J. E. *Carbohydr. Res.* **1988**, *179*, 7-19.

43. Barker, S. A.; Hurst, R. E.; Settine, J.; Fish, F. P.; Settine, R. L. *Carbohydr. Res.* **1984**, *125*, 291-300.

44. Huang, R.; Pomin, V. H.; Sharp, J. S. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1577-1587.

45. Liu, R.; Xu, Y.; Chen, M.; Weiwer, M.; Zhou, X.; Bridges, A. S.; DeAngelis, P. L.; Zhang, Q.; Linhardt, R. J.; Liu, J. *J. Biol. Chem.* **2010**, *285*, 34240-34249.

46. Xu, Y.; Masuko, S.; Takieddin, M.; Xu, H.; Liu, R.; Jing, J.; Mousa, S. A.; Linhardt, R. J.; Liu, J. *Science* **2011**, *334*, 498-501.

47. Heiss, C.; Wang, Z.; Azadi, P. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 774-778.

48. Nagasawa, K.; Inoue, Y.; Kamata, T. *Carbohydr. Res.* **1977**, *58*, 47-55.

49. Bendiak, B.; Fang, T. T.; Jones, D. N. M. *Can. J. Chem.* **2002**, *80*, 1032-1050.

50. Baldwin, M. A.; Stahl, N.; Reinders, L. G.; Gibson, B. W.; Prusiner, S. B.; Burlingame, A. L. *Anal. Biochem.* **1990**, *191*, 174-182.

51. Lindahl, U.; Li, J. P. *Int. Rev. Cell Mol. Biol.* **2009**, *276*, 105-159.

52. Guerrini, M.; Elli, S. S.; Mourier, P.; Rudd, T. R.; Gaudesi, D.; Casu, B.; Boudier, C.; Torri, G.; Viskov, C. *Biochem. J.* **2012**.

CHAPTER 4

THE DE NOVO SEQUENCING OF COMPLEX MIXTURES OF HEPARIN AND HEPARAN SULFATE OLIGOSACCHARIDES[3]

[3] Huang, R., Zong, C., Venot, A., Chiu, Y., Zhou, D., Boons, G. J., and Sharp, J.S., to be submitted for publication.

Abstract

Here, we describe the first sequencing method of a complex mixture of heparan sulfate tetrasaccharides by LC-MS/MS. Heparin and heparan sulfate (HS) are linear polysaccharides that are modified in a complex manner by N- and O-sulfation, N-acetylation, and epimerization of the uronic acid. Heparin and HS are involved in various essential biological processes, including the development and activation of immune cells and cell migration. The structural analysis of these glycosaminoglycans is challenging due to the lability of their sulfate groups, the high heterogeneity of modifications, as well as epimerization of the uronic acids. While advances in liquid chromatography (LC) and mass spectrometry (MS) have enabled compositional profiling of HS oligosaccharide mixtures, on-line separation and detailed structural analysis of isomeric HS mixtures has not been achieved. One of the major challenges is the difficulty in finding mass spectrometry-compatible LC techniques capable of separating HS sulfation positional isomers and stereo-isomers (epimers), as well as a tandem mass spectrometry strategy that can accurately determine the sites of sulfation. Here, we report the development and evaluation of a chemical derivatization and tandem mass spectrometry method that can separate and identify isomeric structures. A series of well-defined synthetic HS tetrasaccharides varying in sulfation patterns and uronic acid epimerization were analyzed by chemical derivatization and LC-MS/MS method. These synthetic compounds made it possible to establish relationships between HS structure and chromatographic behavior and MS/MS fragmentation characteristics. The new method was successfully applied for structural sequencing of HS tetrasacharides mixtures derived from natural sources, showing its capability for analysis of native HS oligosaccharides. The method reported here is an important step towards the sequencing of longer HS oligosaccharide mixtures where detailed structural information is required to determine biological function.

Introduction

Glycosaminoglycans (GAG) are a family of negatively charged linear polysaccharides consisting of repeating disaccharide units, among which heparan sulfate (HS) is the most heterogeneous class with variability in O-sulfation positions, amine group modifications and uronic acid epimerization[1,2]. HS is found on the surface of almost all mammalian cells and in the extracellular matrix[3,4], where it mediates a wide range of key biochemical and developmental processes[5-8], as well as pathological pathways[9-12]. Although a few function-specific heparin/HS motifs have been elucidated, such as the heparin pentasaccharide for binding to antithrombin III and inhibiting coagulation[13,14], there are numerous other heparin/HS sequences responsible for certain cellular functions that are still unknown[15,16]. The need to better understand the structure/function relationships of HS has driven the development of advanced analytical methods for detailed structural characterization of these biomolecules[17-19]. However, the heterogeneous nature of HS, caused by the post-polymerization modifications during biosynthesis, has made their structure determination a highly challenging task.

While advances in liquid chromatography (LC) and mass spectrometry (MS) make it possible to analyze many types of biomolecules in sensitive and high throughput manner, serious challenges remain for structural sequencing of HS[20-22]. This problem is mainly due to the chemical instability of sulfate groups and the high structural heterogeneity leading to considerable difficulties in separating isomeric structures. Efforts have been made to develop MS-based analytical methods for HS disaccharide analysis, HS-oligosaccharide profiling, and tandem mass spectrometric (MS/MS) based sequencing[23]. Disaccharide analysis, a general and useful analytical method for the characterization of GAG populations, involves exhaustive depolymarization of intact HS into disaccharides followed by LC-MS for qualitative and

quantitative analysis of the resulting disaccharides[24,25]. While this method provides compositional information, it does not reveal the sequence information of oligosaccharide domains of the polysaccharide chains, which are considered to be the minimum sequences required by specific protein-HS binding interactions. Oligosaccharide profiling, on the other hand, can provide compositions and abundances of oligosaccharides with moderate lengths by performing partial depolymerization and on-line LC-MS analysis[26,27]. Oligosaccharides can be separated based on their size and degree of sulfation, and composition information (such as chain length, degree of sulfation and number of acetyl groups) can be obtained by accurate mass measurement. However, MS/MS is required in order to obtain detailed sequence information, such as positions of sulfation and acetylation, and types of uronic acid (glucuronic acid or iduronic acid). These specific modifications is crucial, at least in certain biologically important cases if not most protein-HS interactions overall, to the specific biological functions of the HS oligosaccharide[15,19].

Numerous difficulties exist for the MS/MS-based sequencing of HS oligosaccharides, among which sulfate loss during fragmentation is the major obstacle. Several attempts have been reported recently to address this problem. Studies have shown that the sulfate loss during the collision induced dissociation (CID) can be minimized by lowering the degree of protonation through charge state manipulation and/or proton-sodium exchange[28]. An alternative fragmentation method that has been used for MS/MS sequencing of heparin/HS is electron detachment dissociation or negative electron transfer dissociation[29-31], which under proper conditions seems to fragment heparin/HS with minimal loss of sulfates. Another potential solution is the replacement of the labile sulfates with stable acetyl groups by permethylation, desulfation and re-acetylation before CID MS/MS analysis[32,33]. We have demonstrated that this

approach can successfully be employed for the sequencing of isomeric chondroitin sulfate (CS) oligosaccharides and synthetic HS oligosaccharides.

Mass spectrometry is quite capable of separating HS oligosaccharides that are non-isobaric in the gas phase with very high resolution, making the technologies mentioned above useful for the sequencing of such analytes. However, in order to obtain unambiguous sequence identification by MS/MS for complex mixtures of heparin/HS oligosaccharides, such as those that would exist in the natural polymer, isomers need to be separated at least partially before the subsequent fragmentation. Several MS-compatible LC systems have been studied, including size-exclusion chromatography (SEC)[34], hydrophilic interaction chromatography (HILIC)[27,35], porous graphitized carbon (PGC)[36,37] chromatography and reversed phase ion pairing (RPIP)[38,39]. These techniques are effective for either disaccharide analysis or oligosaccharide profiling, however, the separation of isomeric complex mixtures of HS oligosaccharide has not been achieved by these methods. Ion mobility spectrometry (IMS), which are technologies for gas-phase separation of ions, has also been reported recently for the separation of HS oligosaccharides[40,41]. While this type of gas-phase separation can be employed for isomers, the requirement of subtle instrumental tuning limits their application for robust and high throughput analysis. Additionally, there are still unresolved effects of gas-phase conformation and cation adduction state on the measured mobility. While future work shows promise in the application of IMS toward separation of isomers and epimers, they are currently not suited for analysis of complex mixtures of HS oligosaccharides.

As an alternative strategy, a chemical derivatization scheme was developed in our lab that allows the most widely used LC-MS/MS platform for separation and sequencing of HS oligosaccharides. This scheme involves the position-specific replacement of labile, hydrophilic

groups of the heparin/HS oligosaccharide with stable hydrophobic groups, a modification and repurposing of a chemical derivatization scheme that was initially reported by Dell *et al.* for GAG analysis by FAB[42]. While our previous study demonstrated the ability of our method to perform *de novo* sequencing of HS oligosaccharides[32], the analysis of isomeric and epimeric HS oligosacchairdes was not achieved due to the limited availability of isomeric HS oligosaccharide standards. Here, we report the development and validation of this technology for the separation and sequencing of isomeric HS tetrasaccharides. This further development of the technology was possible due to the creation of a library of synthetic HS tetrasaccharides varying in degree and position of sulfation, and uronic acid epimerization. While this library is not exhaustive, it is sufficiently broad to allow the development of LC conditions, as well as the evaluation of fragmentation patterns specific for certain modification positions and epimerizations. The RPLC selectivity of sulfation positional isomers and epimers was systematically studied with the use of a fused core-porous shell C18 stationary phase. The improved separation, coupled with the *de novo* sequencing using the derivatization-CID MS/MS method, made it possible for the first time, to sequence a complex mixture of twenty-one synthetic HS tetrasaccharides, including many sulfation positional isomers and/or epimers. The combination of molecular weight, chromatography behavior and fragmentation pattern makes it possible to identify isomeric structures. The utility of this method was demonstrated through the LC-MS/MS analysis of a mixture of HS tetrasaccharides obtained by enzymatic depolymerization of native HS, which represents the first report of LC-MS/MS to sequence a complex mixture of native HS tetrasaccharides.

Experimental Section

Synthetic HS tetrasaccharides preparation. HS tetrasaccharides were synthesized and purified as previous described[43]. The structures of the synthetic compounds were confirmed by NMR spectroscopy and mass spectrometry. Compounds were prepared with varying degrees and patterns of sulfation as well as uronic acid epimerization, including a series of tetrasaccharides with alkyl linker (**Table 4.1**) and two tetrasaccharides with free reducing end (GlcA-GlcNAc6S-GlcA-GlcNAc6S and GlcA-GlcNAc6S-IdoA-GlcNAc6S).

Table 4.1. Sequences for four sets of synthetic tetrasaccharides with alkyl linker at the reducing end. GlcN, free glucosamine; GlcNAc, N-acetylated glucosamine; GlcNS, N-sulfated glucosamine; GlcA, glucuronic acid; IdoA, iduronic acid; 2S, 2-O-sulfate; 6S, 6-O-sulfate.

observered m/z for precursor ion M^+	Compound # and sequences
1117.596	**1a**: IdoA2S-GlcNAc-GlcA-GlcNAc-$(CH_2)_5NH_2$ **1b**: IdoA-GlcNAc6S-GlcA-GlcNAc-$(CH_2)_5NH_2$ **1c**: GlcA-GlcNAc-IdoA2S-GlcNAc-$(CH_2)_5NH_2$ **1d**: GlcA-GlcNAc-IdoA-GlcNAc6S-$(CH_2)_5NH_2$ **1e**: GlcA-GlcNAc-GlcA2S-GlcNAc-$(CH_2)_5NH_2$
1123.633	**2a**: IdoA2S-GlcNS-GlcA-GlcNS-$(CH_2)_5NH_2$ **2b**: IdoA-GlcNS6S-GlcA-GlcNS-$(CH_2)_5NH_2$ **2c**: GlcA-GlcNS-IdoA2S-GlcNS-$(CH_2)_5NH_2$ **2d**: GlcA-GlcNS-IdoA-GlcNS6S-$(CH_2)_5NH_2$ **2e**: GlcA-GlcNS-GlcA2S-GlcNS-$(CH_2)_5NH_2$
1148.609	**3a**: IdoA2S-GlcNAc6S-GlcA-GlcNAc-$(CH_2)_5NH_2$ **3b**: IdoA-GlcNAc6S-GlcA-GlcNAc6S-$(CH_2)_5NH_2$ **3c**: GlcA-GlcNAc-IdoA2S-GlcNAc6S-$(CH_2)_5NH_2$ **3d**: GlcA-GlcNAc6S-IdoA-GlcNAc6S- $(CH_2)_5NH_2$ **3e**: GlcA-GlcNAc-GlcA2S-GlcNAc6S-$(CH_2)_5NH_2$ **3f**: GlcA-GlcNAc6S-GlcA-GlcNAc6S- $(CH_2)_5NH_2$
1154.648	**4a**: IdoA2S-GlcNS6S-GlcA-GlcNS-$(CH_2)_5NH_2$ **4b**: IdoA-GlcNS6S-GlcA-GlcNS6S-$(CH_2)_5NH_2$ **4c**: GlcA-GlcNS-IdoA2S-GlcNS6S-$(CH_2)_5NH_2$ **4d**: GlcA-GlcNS-GlcA2S-GlcNS6S-$(CH_2)_5NH_2$ **4e**: GlcA-GlcNS6S-GlcA-GlcNS6S- $(CH_2)_5NH_2$

Native HS tetrasaccharide mixture preparation. HS tetrasaccharide mixture was prepared by enzymatic digestion of intact HS (Celsus Inc., Cincinnati, OH), followed by gel filtration purification. Briefly, 100mg HS and 0.05IU heparinase III (IBEX Technologies Inc., Quebec, Canada) were mixed in 1.5mL buffer of 50mM sodium acetate and 0.4mM calcium acetate. The solution was incubated at 37ºC for a total of 48h, with an additional aliquot of 0.05IU heparinase III added after 24h incubation. The enzyme digestion was stopped by heating the solution at 100ºC for 10min. The depolymerized samples were subjected to gel filtration chromatography on a Bio-Gel P-10 column (2.5×100cm) using 10% ethanol solution containing 1 M NaCl as a mobile phase to obtain tetrasaccharide fractions, which was further desalted through a Sephadex G-15 column (1.0×50cm) and dried under vacuum.

Chemical derivatization of HS tetrasaccharides. HS tetrasaccharides, except those with an alkyl linker at the reducing end which cannot be reduced, were first reduced by sodium borohydride, followed by desalting and lyophilization. Complete permethylation, desulfation and pertrideuteroacetylation were performed to replace the original sulfate groups with trideuteroacetyl groups. Detailed procedures have been described in our previous work for structural analysis of synthetic HS oligosaccharides[32]. Briefly, the dried triethylammonium salts of HS tetrasaccharides (10-100 μg) were permethylated according to Ciucanu methylation method using sodium hydroxide and methyl iodide in dimethyl sulfoxide (DMSO), followed by desalting using a C18 Sep-Pak cartridge (Waters Co.). The pyridinium salts of the permethylated products were re-suspended in DMSO containing 10% methanol and incubated for 4h at 95°C to remove the sulfate groups. The dried desulfated products were then trideuteroacetylated by incubating with D_6-acetic anhydride in pyridine at 50°C overnight and solvent was dried under

vacuum. The final derivatized products were re-suspended in 10% acetonitrile/water at specified concentrations for subsequent LC-MS/MS analysis.

LC-MS/MS analysis. For structural analysis of each synthetic tetrasaccharide, an RPLC-MS/MS method was used. Buffer B was prepared as acetonitrile with 0.1% formic acid, and buffer A was 0.1% formic acid in water with or without 1mM sodium formate. Online HPLC was performed on Agilent LC 1100 series using a Halo C18 column with fused core-porous shell packing materials (0.2×50mm, 2.7μm, 160 Å, Advanced Material Technology, Wilmington, DE). A linear gradient of 10%- 60% buffer B over 15min was used, with a flow rate of 9μL/min and a 2μL injection at a sample concentration of 0.1μg/μL. Mass spectrometry was performed on a Thermo LTQ-FT instrument. Full MS in FT mode and CID-MS/MS spectra in ion trap mode were acquired in positive ion mode. External calibration of the instrument produced mass accuracy of <5ppm for full MS spectra, which enabled the use of accurate mass measurement for composition determination of precursor ion. A data dependent MS/MS method was used, with the top 4 abundant precursor ions selected, to trigger CID-MS/MS fragmentation. Instrument parameters were set as: spray voltage at 1-2kV, capillary voltage at 40V, tube lens at 80V and capillary temperature at 250ºC. The collision energy for CID fragmentation was set between 30V and 50V.

For isomeric and/or epimeric HS tetrasaccharide mixture analyses, online RPLC separation were studied on a longer column with similar C18 packing material (0.2×500mm, 5.0μm, 160 Å, Advanced Material Technology, Wilmington, DE). For separation of synthetic HS tetrasaccharides, a 140min gradient was used from 35% to 55% buffer B. For separation of native HS tetrasacharide mixtures, an isocratic elution was also used at 44% buffer B in order to obtain better separation. A flow rate of 2μL/min and a 2μL injection at a sample concentration of

1μg/μL were used for both mixture analyses. Mass spectrometry setup was the same as mentioned above.

Results and Discussion

After derivatization, all hydroxyl groups of the HS tetrasaccharides as well as carboxyl groups and amine groups were methylated, protecting the unsulfated functional groups. Sulfates were replaced in a site-specific manner by trideuteroacetyl groups, which could be differentiated from original N-acetyl groups and used to locate the original sulfation positions. The derivatized tetrasaccharides were sufficiently hydrophobic for retention on a standard C18 RPLC column. For the HS tetrasaccharides with an alkyl linker, a quaternary amine was formed during the permethylation, leaving a permanent positive charge at the reducing-end alkyl-amine chain (**Figure 4.1**).

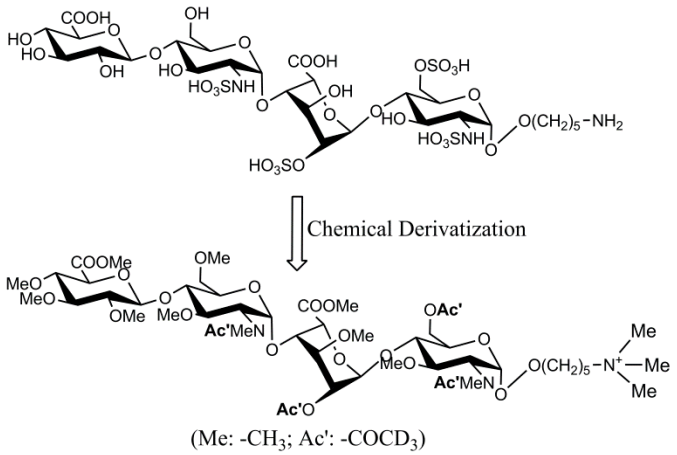


Figure 4.1. Structures of the synthetic tetrasaccharide GlcA-GlcNS-IdoA2S-GlcNS6S-$(CH_2)_5NH_2$ (compound 4c) before and after chemical derivatization. After permethylation, all hydroxyl groups as well as carboxyl groups and amine groups were methylated, where the primary amine of the linker was converted to quaternary amine leaving a fixed positive charge at the reducing end. Original sulfation groups were replaced by trideuteroacetyl groups in the following desulfation and perdeuteroacetylation steps.

MS/MS analysis of synthetic HS tetrasaccharids with alkyl linker. In our previous work, we have demonstrated that chemical derivatization allowed successful sequencing of HS oligosaccharides with sequential glycosidic bond cleavage fragments only, where parent ion $[M+2Na]^{2+}$ was chosen for fragmentation[32]. Here, MS/MS spectra of the parent ion M^+ were taken for each synthetic HS tetrasaccharide as listed in **Table 4.1**, as these ions provided more informative MS/MS spectra for differentiation of isomers and/or epimers rather than parent ion $[M+H]^{2+}$ and $[M+Na]^{2+}$ (data not shown).

After derivatization, sites of N-sulfation can be differentiated from sites of N-acetylation by a mass shift of 3Da. For example, compounds **2a-e** have a 6 Da mass increase for the parent ion M^+ comparing to the M^+ of compounds **1a-e**, and the same is the case for compounds **4a-e** and **3a-f** (**Table 4.1**). The compounds in Group 2 all have two N-sulfates, while similar compounds in Group 1 all have two N-acetyls. As previously reported for HS oligosaccharides, we were able to accomplish *de novo* sequencing of each synthetic tetrasaccharide based solely on glycosidic bond cleavage of the derivatized product. In order to illustrate the differentiation of sulfation positional isomers and epimers, MS/MS spectra of parent ion M^+ for compounds **3a-f** are presented in **Figure 4.2** (MS/MS spectra for compounds **1a-e**, **2a-e** and **4a-e** are also provided in **Figure 4.3A, 4.3B and 4.3C** for reference).

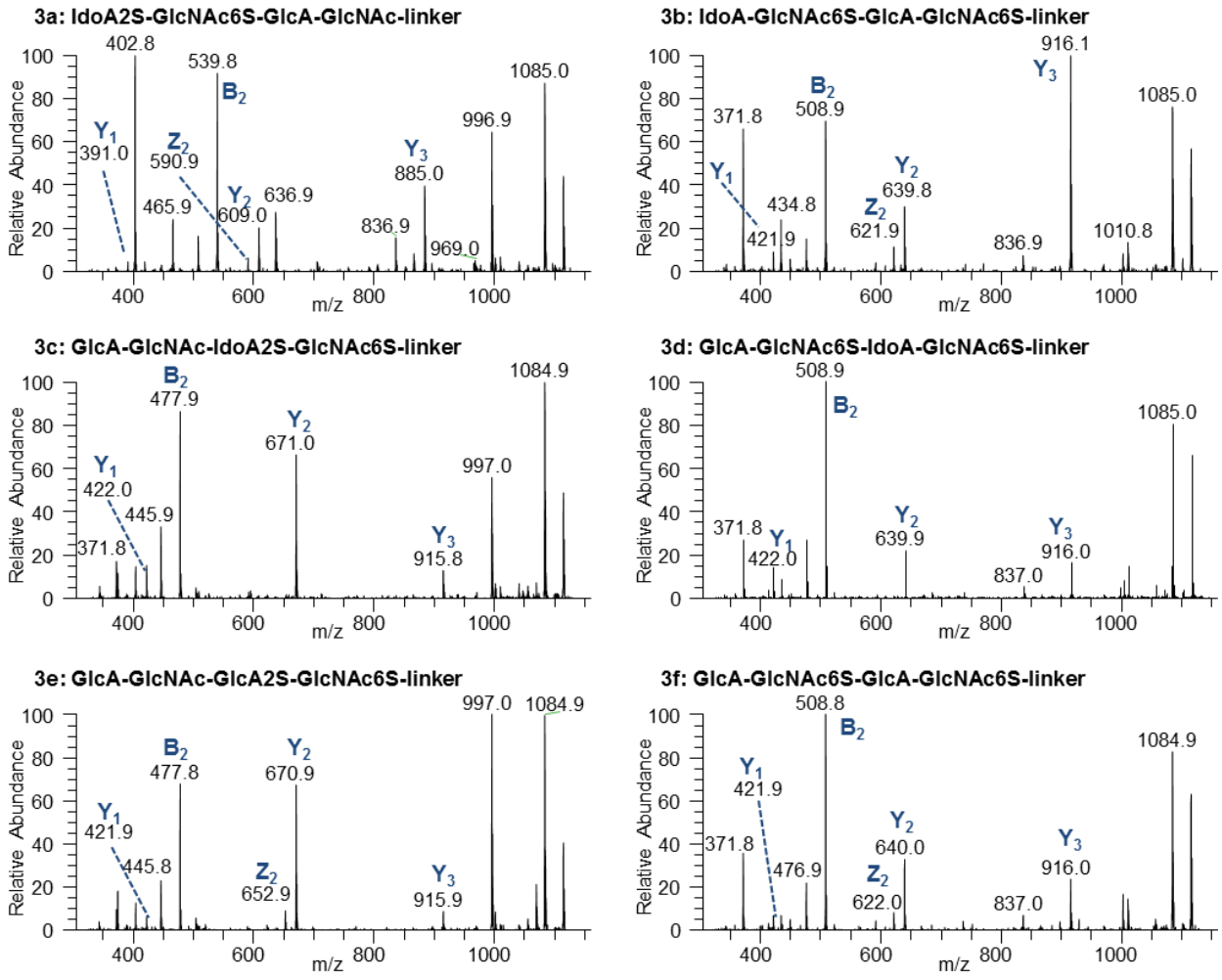


Figure 4.2. CID tandem MS spectra for compounds 3a-f with m/z of 1148.6 for precursor ion M^+. These six tetrasaccharides could be divided into three isomeric sets: [3a], [3b, 3d, 3f] and [3c, 3e], where tetrasaccharides within each set were epimers. Sequential Y ions for each tetrasaccharides were observed which could be used as diagnostic ions to identify sulfation positional isomers. Differences of the relative intensities ratio of certain fragment ions were also observed within epimers. For example, compound 3c-f containing terminal GlcA had Y_3/Y_2 ratio<0.8, whereas compound 3a and 3b containing terminal IdoA shown Y_3/Y_2 ratio>0.8. On the other hand, Z_2 ions were observed for compound 3a, 3b, 3e and 3f containing internal GlcA residue, but not for compound 3c and 3d containing internal IdoA instead.

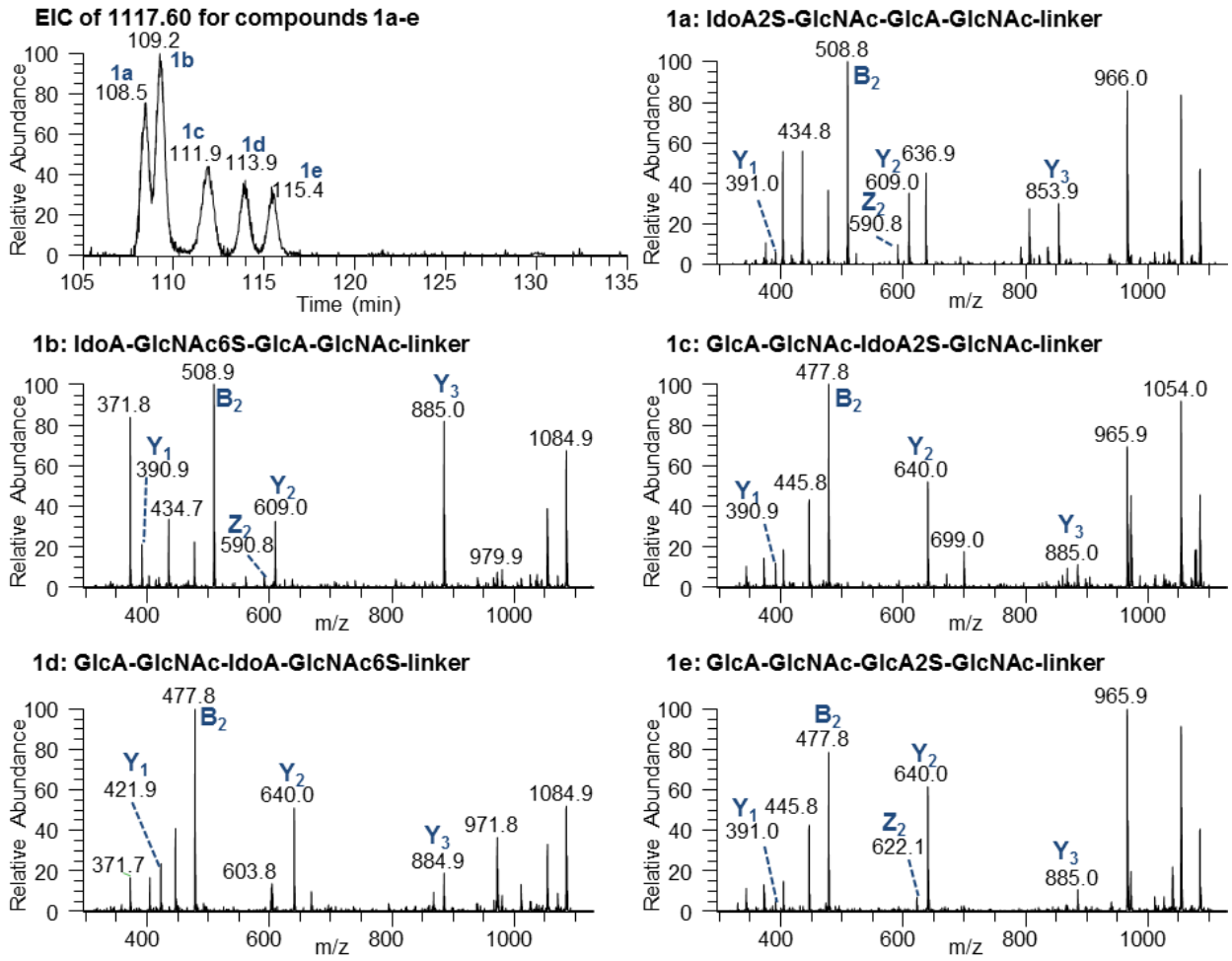


Figure 4.3A. CID tandem MS spectra for compounds 1a-e with m/z of 1117.596 for precursor ion M^+. The structure of these five tetrasaccharides were: 1a: IdoA2S-GlcNAc-GlcA-GlcNAc-*linker*; 1b: IdoA-GlcNAc6S-GlcA-GlcNAc-*linker*; 1c: GlcA-GlcNAc-IdoA2S-GlcNAc-*linker*; 1d: GlcA-GlcNAc-IdoA-GlcNAc6S-*linker*; 1e: GlcA-GlcNAc-GlcA2S-GlcNAc-*linker*. Compounds 1c and 1e were internal epimers, and isomeric with the other three compounds. Sequential Y ions for each tetrasaccharides were observed for identification of sulfation positional isomers. Z_2 ions were observed for compound 1a, 1b and 1e, containing internal GlcA residue, but not for compound 1c and 1d containing internal IdoA instead. Compounds 1c-e containing terminal GlcA had Y_3/Y_2 ratio<0.8, whereas compounds 1a and 1b containing terminal IdoA shown Y_3/Y_2 ratio>0.8.

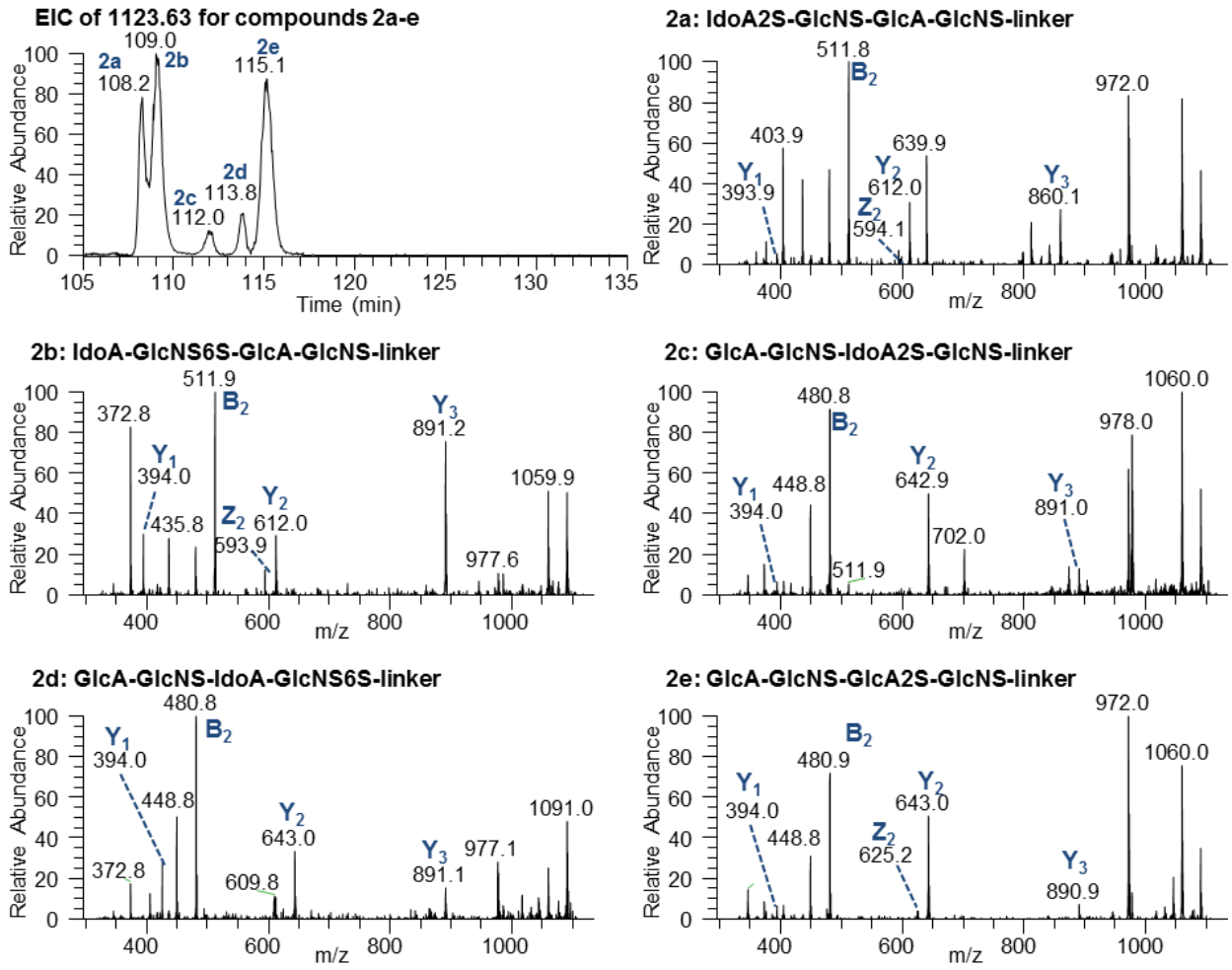


Figure 4.3B. CID tandem MS spectra for compounds 2a-e with m/z of 1123.633 for precursor ion M^+. The structure of these five tetrasaccharides were: 2a: IdoA2S-GlcNS-GlcA-GlcNS-*linker*; 2b: IdoA-GlcNS6S-GlcA-GlcNS-*linker*; 2c: GlcA-GlcNS-IdoA2S-GlcNS-*linker*; 2d: GlcA-GlcNS-IdoA-GlcNS6S-*linker*; 2e: GlcA-GlcNS-GlcA2S-GlcNS-*linker*. Compounds 2c and 2e were internal epimers, and isomeric with the other three compounds. Sequential Y ions for each tetrasaccharides were observed for identification of sulfation positional isomers. Z_2 ions were observed for compound 2a, 2b and 2e, containing internal GlcA residue, but not for compound 2c and 2d containing internal IdoA instead. Compounds 2c-e containing terminal GlcA had Y_3/Y_2 ratio<0.8, whereas compounds 2a and 2b containing terminal IdoA shown Y_3/Y_2 ratio>0.8.

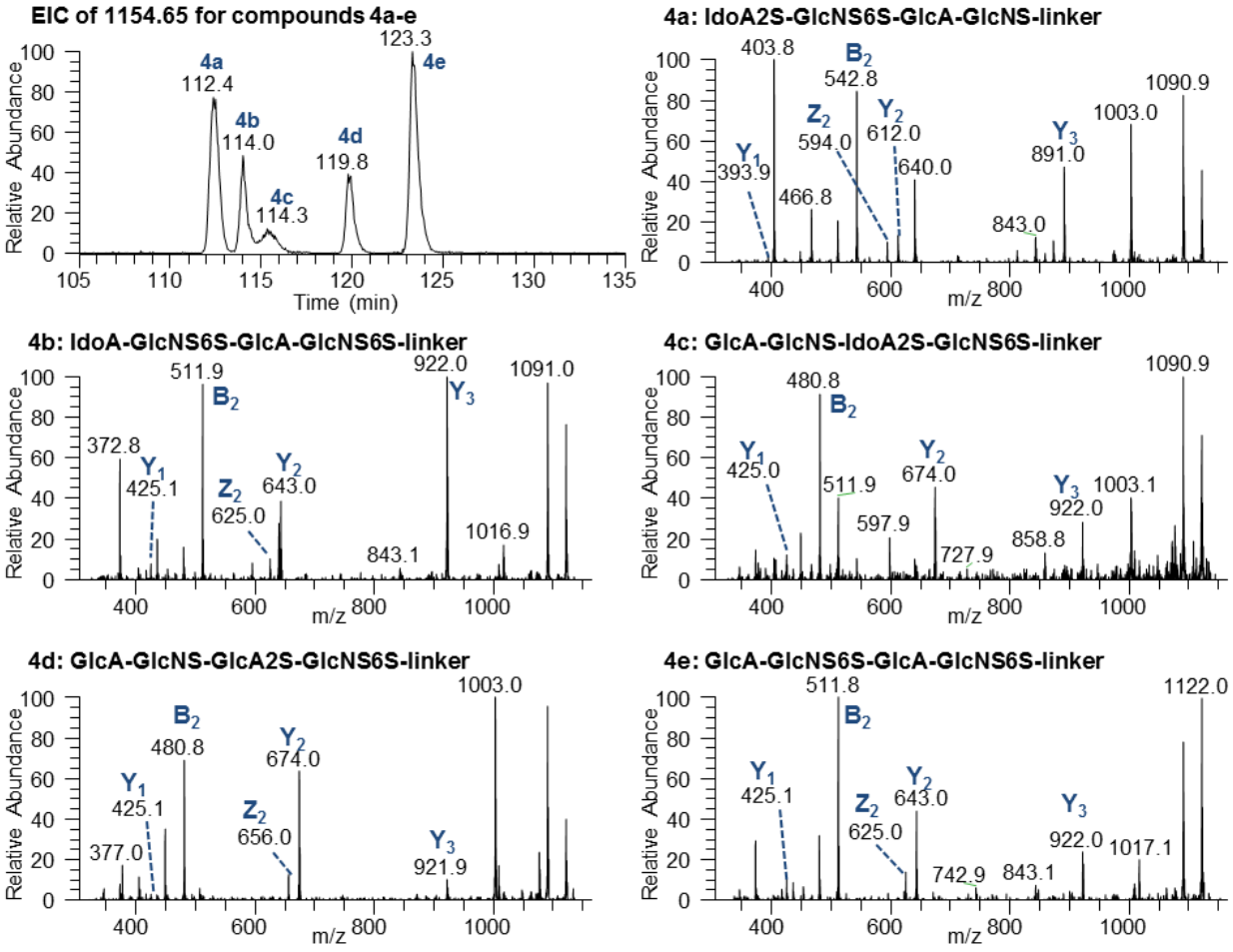


Figure 4.3C. CID tandem MS spectra for compounds 4a-e with m/z of 1154.648 for precursor ion M^+. The structure of these five tetrasaccharides were: 4a: IdoA2S-GlcNS6S-GlcA-GlcNS-*linker*; 4b: IdoA-GlcNS6S-GlcA-GlcNS6S-*linker*; 4c: GlcA-GlcNS-IdoA2S-GlcNS6S-*linker*; 4d: GlcA-GlcNS-GlcA2S-GlcNS6S-*linker*; 4e: GlcA-GlcNS6S-GlcA-GlcNS6S-*linker*. Compounds 4b and 4e were internal epimers whereas compounds 4c and 4d were terminal epimers. Sequential Y ions for each tetrasaccharides were observed for identification of sulfation positional isomers. Z_2 ions were observed for all compounds except 4c that contained internal IdoA residue. Compounds 4c-e containing terminal GlcA had Y_3/Y_2 ratio<0.8, whereas compounds 4a and 4b containing terminal IdoA shown Y_3/Y_2 ratio>0.8.

For synthetic HS tetrasaccharides with m/z of 1148.609 for M^+ ion, the possible sequences would be HexA-GlcNAc-HexA-GlcNAc-*linker*, with two original O-sulfation sites not occurring at the 3-O position of the GlcNAc present on the tetrasaccharide. The sites of O-sulfation have a mass increase of 31 Da after derivatization compared to the originally non-sulfated sugar. For example, the Y_3 ion of compound **3a** has an m/z of 885, indicating that the Y_3 ion has a single O-sulfation site. Since the parent ion has two O-sulfation sites, the non-reducing end uronic acid is O-sulfated. The Y_3 ions for compound **3b-e** were all 31 Da more with m/z of 916, indicating that the Y_3 ion of these tetrasaccharides all contain two O-sulfates, and therefore there is a non-sulfated uronic acid at the non-reducing end. The m/z of Y_2 ion was 609/640/671 for compound **3a**/**3b**,**3d**,**3f**/**3c**,**3e**, respectively, indicating none/one/two original O-sulfation site(s) for the corresponding disaccharide residues at the reducing end. With this information, we could identify the backbone structure for compound **3a** as HexA2S-GlcNAc6S-HexA-GlcNAc-*linker*, and HexA-GlcNAc-HexA2S-GlcNAc6S-*linker* the backbone structure for compound **3c** and **3e**. For compound **3b**, **3d** and **3f**, the Y_1 ion with m/z of 422 instead of 391 indicated the reducing end GlcNAc was originally O-sulfated rather than non-sulfated, which confirmed the sequence was HexA-GlcNAc6S-HexA-GlcNAc6S-*linker* instead of HexA-GlcNAc6S-HexA2S-GlcNAc-*linker*. In every case, the sequence determined by *de novo* analysis of the MS/MS spectrum agreed with the structure of the synthetic compound confirmed by NMR.

While diagnostic ions can be easily observed and used to identify sulfation positional isomers, certain fragment ions make it possible to differentiate epimers by their relative intensities, as shown in **Figure 4.2**. For the terminal epimeric pair **3b** and **3f**, the ratio of relative intensity of Y_3/Y_2 ion may be used to differentiate the terminal GlcA (3f) with Y_3/Y_2 ratio of 0.7 and terminal IdoA (**3b**) with Y_3/Y_2 ratio of 3.3. This also applied to the other four compounds,

where compound **3a** with IdoA2S at the non-reducing end had Y_3/Y_2 ratio of 2.0, but compound 3c, 3d and 3e had Y_3/Y_2 ratio of 0.2, 0.7 and 0.1 respectively by having GlcA at the non-reducing end instead. For the internal epimeric pairs **3c** and **3e**, **3d** and **3f**, we observed Z_2 ions for **3e** and **3f** that contain internal GlcA residue, but not for compounds 3c and 3d that contain internal IdoA instead. In addition, Z_2 ions were also observed for the other two compounds 3a and 3b that contain internal GlcA residue. This indicated that the Z_2 ion may be used as diagnostic ion for differentiation of internal GlcA and IdoA residue, regardless of sulfated uronic acid or non-sulfated one. These trends hold true through for the Group 1, 2, and 4 tetrasaccharide epimeric pairs (**Figure 4.3**). All compounds with terminal GlcA/GlcA2S had Y_3/Y_2 ratio<0.8 and those with terminal IdoA/IdoA2S instead had Y_3/Y_2 ratio>0.8. Z_2 ion was observed for all compounds with internal GlcA/GlcA2S, but not for those with internal IdoA/IdoA2S (< 1% relative abundance). These results indicate that MS/MS identification of epimers through the derivatization process is possible.

LC separation of synthetic HS tetramers with alkyl linker. While MS/MS could differentiate sulfate positions and epimerizations for the pure synthetic tetrasaccharides, the analysis of complex mixtures requires the separation of the isomers prior to MS/MS analysis. For the separation of synthetic HS tetrasaccharides with alkyl linker, an equimolar mixture of synthetic tetrasaccharides (total of twenty one) was subjected to the chemical derivatization process, which was analyzed by the LC-MS/MS approach described in the Experimental section. The reversed phase LC separation was performed on a 500 mm C18 column packed with the HALO-5 fused core-porous shell particles, using a gradient of 30-50%B for 140 min and a flow rate of 2μL/min. As the extracted ion chromatogram (EIC) shows in **Figure 4.4**, base line separation was achieved for most of the species in each group of isomers/epimers, with average

peak width of approximately one minute. For the compounds derived from the N-acetylated tetrasaccharides *vs.* N-sulfated tetrasaccharides, (*e.g.* compound **1a-e** *vs.* compound **2a-e**) the only difference is trideutero *vs.* trihydro labeling of the amines. As shown in **Figure 4.4**, the isotopic labeling did not alternate the chromatography selectivity, leading to the co-elution of compounds **1a-e** and compounds **2a-e**. However, the 6 Da mass shifts for the parent ion M^+ allowed for clean separation of these analytes by mass spectrometry, making the MS/MS analysis able to cleanly differentiate between the two groups of sequences.

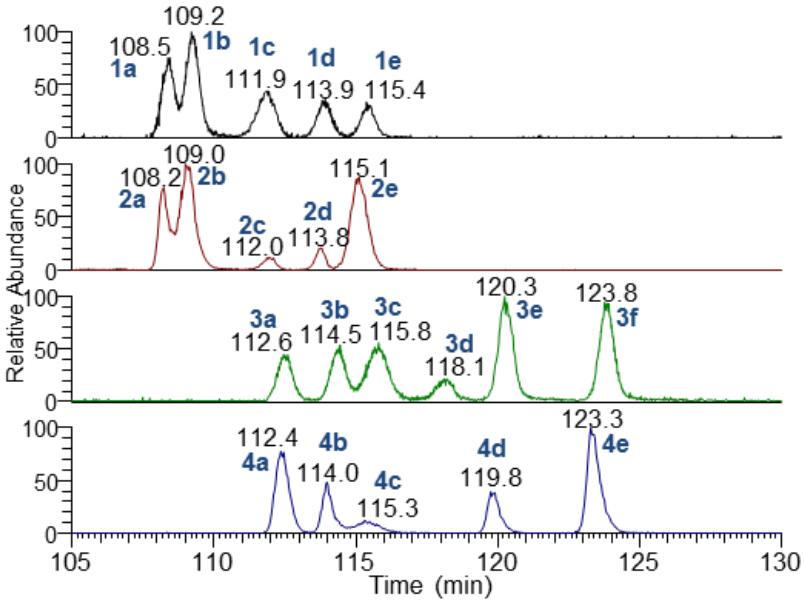


Figure 4.4. Reversed phased LC separation for derivatized synthetic tetrasaccharide mixture with alkyl linkers. Extracted ion charomatograms (EIC) for each of the four precursor ions were presented, with a total of twenty one tetrasaccharides being separated. Peaks are labeled with the compound # of the eluting tetrasaccharide. Refer to Table 1 for tetrasaccharides structures.

We observed patterns in LC selectivity that could be correlated to the original degree of sulfation, sulfation positions and uronic acid epimerization. In general, tetrasaccharides with a higher degree of sulfation elute later. This observation is consistent with what we observed for longer HS oligosaccharides[32]. By comparing the sulfation positions and retention time for isomers **3a** and **3b**, **3c** and **3d**, **3e** and **3f**, it was noticed that tetrasaccharides with 2-O-sulfation on GlcA/IdoA eluted earlier than those modified by 6-O sulfation on GlcNAc/GlcNS, with retention time differences ranging from 1 min to 4 min. For epimer pairs, more clear patterns could be observed for their retention times: tetrasaccharides with all GlcA eluted last, tetrasaccharides with non-reducing terminal IdoA elute first, and tetrasaccharides with an internal IdoA elute in between the two. For example, compound **3d** came out 3.6 min later than compound **3b** and 5.7 min earlier than compound **3f** (**Figure 4.4**). Overall, each structural characteristic contributed to the chromatographic behavior of these synthetic HS tetrasaccharides, where sulfation degrees and uronic acid epimerizations seemed to have larger effect than sulfation positions.

Structural analysis and LC Separation of two epimeric synthetic HS tetrasaccharides with free reducing end. Two synthetic epimers with a free reducing ends (GlcA-GlcNAc6S-GlcA-GlcNAc6S and GlcA-GlcNAc6S-IdoA-GlcNAc6S), which have the same saccharide structure as synthetic tetrasaccharides **3f** and **3d**, were also subjected to chemical derivatization and LC-MS/MS analysis to determine the effect of the alkyl linker on LC selectivity and resolution. The LC chromatogram and MS/MS spectra were presented in **Figure 4.5**. Compared to compound **3d** and compound **3f** (**Figure 4.4**), tetrasaccharides without linker eluted about **20** min later under the same chromatographic conditions (**Figure 4.5A**). The fixed positive charge on the alkyl linker after derivatization made compound **3d** and **3f** relatively more

hydrophilic than the two tetrasaccharides without linker, leading to an earlier elution time for compounds **3d** and **3f**. In addition, the selectivity within the epimer pair was also different. While compound **3d** with internal IdoA eluted about 6 min *earlier* than compound **3f** with internal GlcA, GlcA-GlcNAc6S-IdoA-GlcNAc6S eluted about 3 min *later* than GlcA-GlcNAc6S-GlcA-GlcNAc6S when the alkyl linker was absent from the reducing end. Although the mechanisms causing such differences are unclear, baseline separation was still achieved, demonstrating the capability of our methodology for on-line separation of native epimeric HS tetrasaccharides.

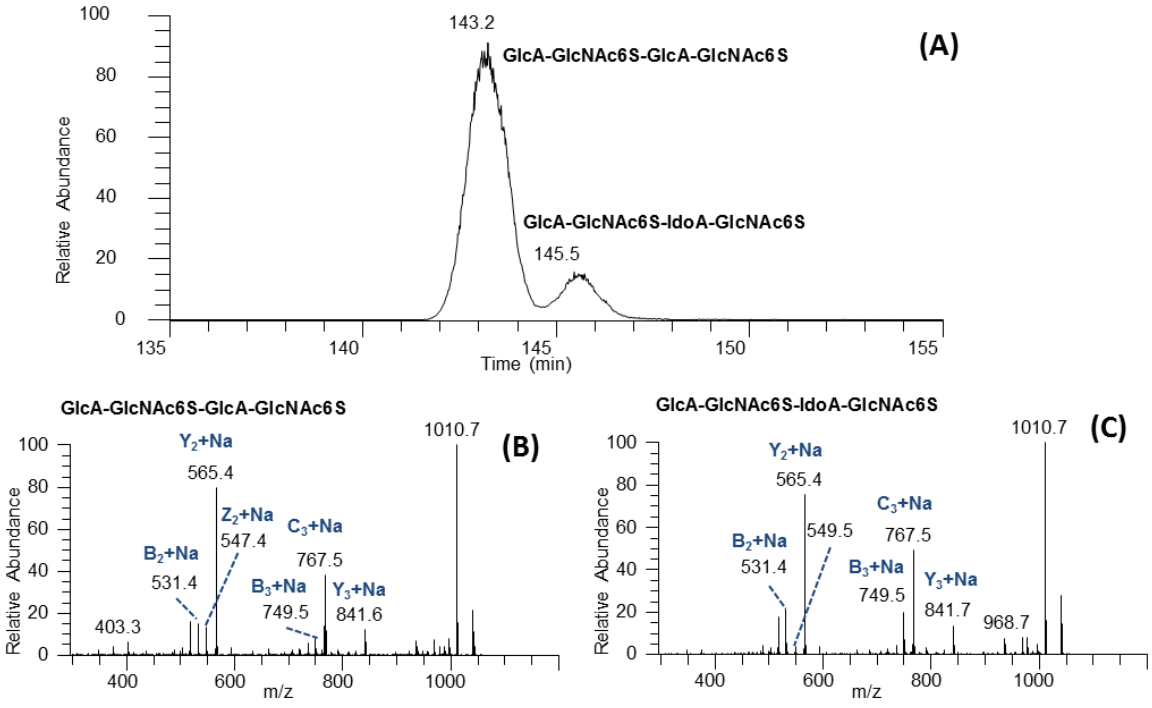
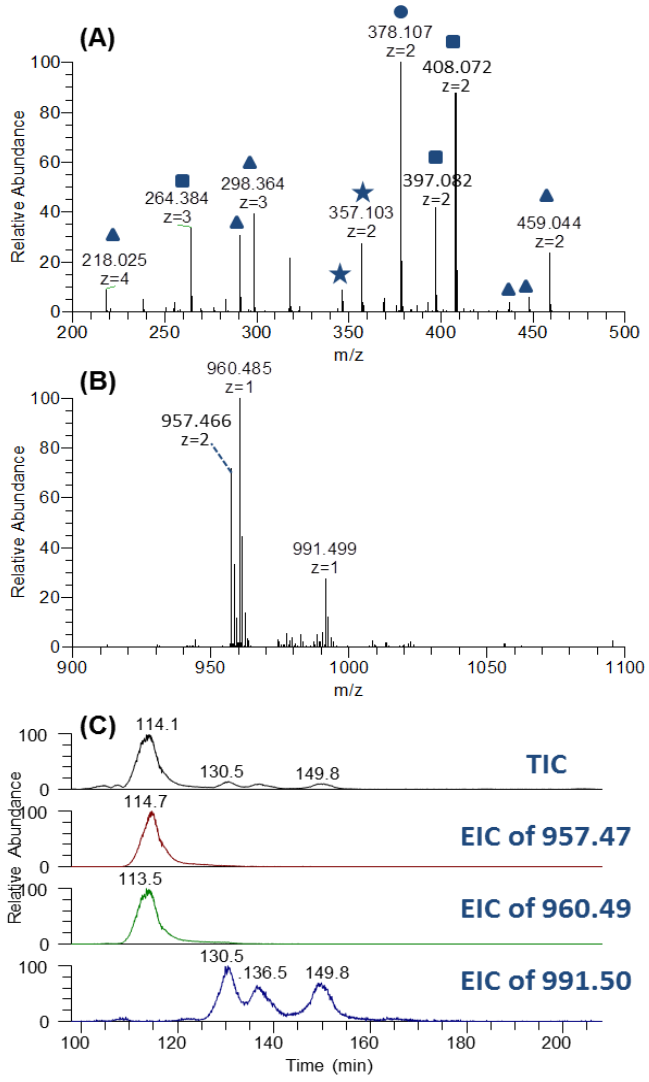


Figure 4.5. **LC chromatogram (A) and MS/MS spectra of the tetrasaccharide standards GlcA-GlcNAc6S-GlcA-GlcNAc6S (B) and GlcA-GlcNAc6S-IdoA-GlcNAc6S (C) of parent ion [M+Na]$^+$.** A series of B ions and Y ions were observed for sequencing of sulfation positions, whereas Z_2 ion with m/z of 547.4 was only observed for GlcA-GlcNAc6S-GlcA-GlcNAc6S with internal GlcA residue.

MS/MS spectra were also compared for GlcA-GlcNAc6S-IdoA-GlcNAc6S and GlcA-GlcNAc6S-GlcA-GlcNAc6S in order to differentiate the epimerization (**Figure 4.5B** and **4.5C**). By selecting the proper parent ion $[M+Na]^+$, similar patterns were observed as for the differentiation between compound **3d** and **3f** (**Figure 4.2**). Z_2 ions were observed for GlcA-GlcNAc6S-GlcA-GlcNAc6S containing internal GlcA residue, but not for GlcA-GlcNAc6S-IdoA-GlcNAc6S. In addition, the intensity ratio of Y_3/Y_2 <0.8 was observed for both MS/MS spectra, which was consistent with what we observed for tetrasaccharides with alkyl linker containing terminal GlcA residue.

With the results obtained from the LC-MS/MS analysis of these two synthetic epimeric standards, we understood their chromatographic selectivity where tetrasaccharides with internal GlcA residue would be eluted earlier than those with internal IdoA residue instead. Also, the Z_2 ions from the fragmentation of $[M+Na]^+$ can be used as diagnostic ion for the existence of internal GlcA residue. These data demonstrate that internal epimers can be distinguished by combining LC separation and fragmentation of the appropriate parent ion.

On-line LC separation and structural sequencing of native HS tetrasaccharides mixture. To evaluate and apply our method for structural analysis of native HS oligosaccharide mixtures, an enzymatic digested HS tetrasaccharide mixture was prepared as described in Experimental section. Prior to chemical derivatizations and LC-MS/MS analysis, an MS spectra was taken in negative ion mode to obtain compositional information (**Figure 4.6A**). For each observed molecular weight (MW), all possible sequences (ignoring epimerizations) are listed together with MW for the corresponding derivatization product (**Figure 4.6D**). After sequential chemical derivatizations, LC-MS/MS analysis was performed and m/z corresponding to each derivatized MW listed in **Figure 4.6D** was extracted from the full MS scan (within 5 ppm).



(D)

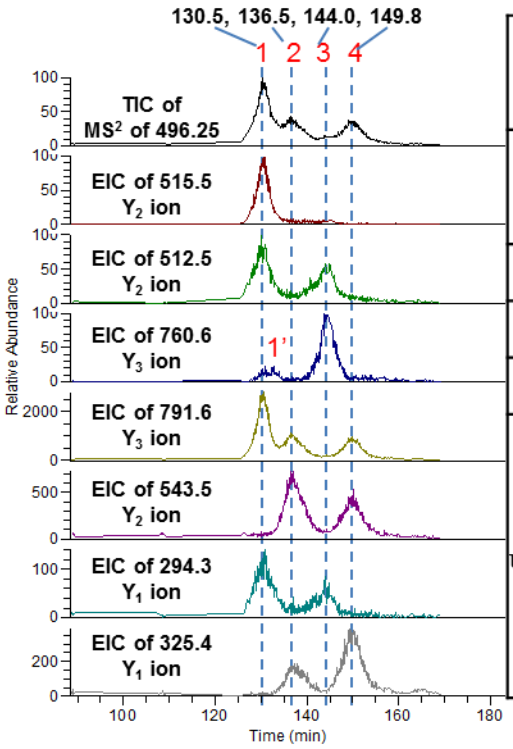
Sequences	Native MW	Derivatized MW
d-HexA-GlcNAc-HexA-GlcN-ol	★ 716.222	943.486
d-HexA-GlcN-HexA-GlcNAc-ol	★ 716.222	943.486
d-HexA-GlcNAc-HexA-GlcNAc-ol	● 758.230	**956.458**
d-HexA-GlcNAc-HexA-GlcNS-ol	796.180	**959.477**
d-HexA-GlcNS-HexA-GlcNAc-ol	796.180	**959.477**
d-HexA2S-GlcNAc-HexA-GlcN-ol	796.180	974.500
d-HexA2S-GlcN-HexA-GlcNAc-ol	■ 796.180	974.500
d-HexA-GlcN6S-HexA-GlcNAc-ol	796.180	974.500
d-HexA-GlcNAc6S-HexA-GlcN-ol	796.180	974.500
d-HexA-GlcNAc-HexA2S-GlcN-ol	796.180	974.500
d-HexA-GlcNAc-HexA-GlcN6S-ol	796.180	974.500
d-HexA-GlcN-HexA2S-GlcNAc-ol	796.180	974.500
d-HexA-GlcN-HexA-GlcNAc6S-ol	796.180	974.500
d-HexA-GlcNAc-HexA-GlcNS3S-ol	876.132	976.475
d-HexA-GlcNS3S-HexA-GlcNAc-ol	876.132	976.475
d-HexA2S-GlcNAc-HexA-GlcNS-ol	876.132	**990.490**
d-HexA2S-GlcNS-HexA-GlcNAc-ol	876.132	**990.490**
d-HexA-GlcNAc6S-HexA-GlcNS-ol	876.132	**990.490**
d-HexA-GlcNAc-HexA2S-GlcNS-ol	876.132	**990.490**
d-HexA-GlcNAc-HexA-GlcNS6S-ol	876.132	**990.490**
d-HexA-GlcNS6S-HexA-GlcNAc-ol	876.132	**990.490**
d-HexA-GlcNS-HexA2S-GlcNAc-ol	▲ 876.132	**990.490**
d-HexA-GlcNS-HexA-GlcNAc6S-ol	876.132	**990.490**
d-HexA2S-GlcN6S-HexA-GlcNAc-ol	876.132	1005.514
d-HexA2S-GlcNAc6S-HexA-GlcN-ol	876.132	1005.514
d-HexA2S-GlcNAc-HexA2S-GlcN-ol	876.132	1005.514
d-HexA2S-GlcNAc-HexA-GlcN6S-ol	876.132	1005.514
d-HexA2S-GlcN-HexA2S-GlcNAc-ol	876.132	1005.514
d-HexA2S-GlcN-HexA-GlcNAc6S-ol	876.132	1005.514
d-HexA-GlcN6S-HexA2S-GlcNAc-ol	876.132	1005.514
d-HexA-GlcN6S-HexA-GlcNAc6S-ol	876.132	1005.514
d-HexA-GlcNAc6S-HexA2S-GlcN-ol	876.132	1005.514
d-HexA-GlcNAc6S-HexA-GlcN6S-ol	876.132	1005.514
d-HexA-GlcNAc-HexA2S-GlcN6S-ol	876.132	1005.514
d-HexA-GlcN-HexA2S-GlcNAc6S-ol	876.132	1005.514

Figure 4.6. Structural analysis of enzymatic digested tetrasaccharide mixture. An MS spectra of the mixture was taken under negative ion mode before derivatization (**A**). After derivatization, the tetrasaccharides mixture was subjected to LC-MS/MS analysis using isocratic condition of 44% B. A zoom-in full MS spectra extracted from 100 to 180min was shown in the range of m/z with charge state of 1(**B**). Three major components were observed with m/z of 957.466, 960.485 and 991.499. Total ion chromatogram (TIC) and extracted ion chromatogram (EIC) of each three major components were also shown (**C**). Each molecular weight (MW) observed in spectra **A** was listed, together with the possible sequences and their corresponding MW after derivatization (**D**).

Three major species were observed: tetrasaccharides with two N-acetyl groups and no O-sulfate group (MW of 956.458); tetrasaccharides with one N-acetyl group, one N-sulfate group and no O-sulfate group (MW of 959.477) and tetrasaccharides with one N-acetyl group, one N-sulfate group and one O-sulfate group (MW of 990.490) (**Figure 4.6B**). The same LC conditions as described in previous section were used for initial test, with all three species eluting between the synthetic tetrasaccharides with alkyl linker and the two without linker (data not shown). This range of retention times was anticipated, as these native tetrasaccharides had free reducing ends instead of alkyl linkers and all three observed species had fewer O-sulfation sites than the two synthetic tetrasaccharides with free reducing ends.

In order to obtain further separation to allow clear interpretation of isomers, isocratic elution was used at 44%B (**Figure 4.6C**). For the species with m/z of 991.499, three major LC peaks were observed coming from a total of eight theoretical possible sequences regardless of the internal HexA epimerizations (**Figure 4.6D**). The TIC of MS/MS of parent ion $[M+2H]^{2+}$ (m/z of 496.252) and a series of EIC of diagnostic Y ions were presented in **Figure 4.7** to illustrate the method for sequencing of sulfation positions. The elution profile of various diagnostic Y ions allows for the identification of isomeric/epimeric sequences that are only partially resolved by LC. If the EIC of two diagnostic sequencing ions perfectly align by LC, then the two sequencing ions can be assigned to a single analyte. However, if they do not align, the sequencing ions are assigned to separate analytes. This allows for the identification of sequences that are only partially resolved by LC, and was previously used in the MS^n analysis of native CS oligosaccharides[33]. The near-perfect EIC co-elution of two mutually exclusive fragment ions can also be used to differentiate between positions of N-sulfation and N-acetylation, which differ after derivatization by only the position of three deuterons that is difficult to separate to any

extend by RPLC. Based on the biosynthetic pathway of HS, the epimerization of GlcA/GlcA2S residues only occurs when they are attached to GlcNS residue at their non-reducing side[44]. Therefore, only twelve of the total sixteen theoretical structures would be biologically possible (as listed in **Figure 4.7**), and only these sequences were considered as possible for *de novo* interpretations.



Peak1	**d-HexA-GlcNAc6S-GlcA-GlcNS-ol and d-HexA-GlcNS6S-GlcA-GlcNAc-ol**
Peak1'	**d-HexA2S-GlcNS-GlcA-GlcNAc-ol and/or d-HexA2S-GlcNAc-GlcA-GlcNS-ol**
Peak2	**d-HexA-GlcNS-GlcA-GlcNAc6S-ol**
Peak3	**d-HexA2S-GlcNS-IdoA-GlcNAc-ol**
Peak4	**d-HexA-GlcNS-IdoA-GlcNAc6S-ol**
unidentified	d-HexA-GlcNS-GlcA2S-GlcNAc-ol
	d-HexA-GlcNS6S-IdoA-GlcNAc-ol
	d-HexA-GlcNS-IdoA2S-GlcNAc-ol
	d-HexA-GlcNAc-GlcA2S-GlcNS-ol
	d-HexA-GlcNAc-GlcA-GlcNS6S-ol

Figure 4.7. LC separation and sequencing of one post-derivatized composition (with MW of 990.490) of the native HS tetrasaccharide mixture. The TIC of MS/MS of parent ion $[M+2H]^{2+}$ (m/z of 496.25) and a series of EICs of diagnostic Y ions were presented. EIC were extracted using the indicated m/z value with a window of 1Da. Three major peaks (1, 2 and 4) and two minor peaks (3 and 1') were observed with retention times labeled above the numbers. A list of total twelve theoretical structures for this specific MW was also attached with specified epimerizations.

Table 4.2. **Total of 10 sequences identified from native HS tetrasaccharides mixture.** "d-" stands for double bond at the non-reducing end and "-ol" stands for free reducing end.

MW for each observed speciece	Identified Sequences
957.466	d-HexA-GlcNAc-GlcA-GlcNAc-ol
960.485	d-HexA-GlcNS-GlcA-GlcNAc-ol d-HexA-GlcNAc-GlcA-GlcNS-ol
990.490	d-HexA-GlcNAc6S-GlcA-GlcNS-ol d-HexA-GlcNS6S-GlcA-GlcNAc-ol d-HexA2S-GlcNS-GlcA-GlcNAc-ol d-HexA2S-GlcNAc-GlcA-GlcNS-ol d-HexA-GlcNS-GlcA-GlcNAc6S-ol d-HexA2S-GlcNS-IdoA-GlcNAc-ol d-HexA-GlcNS-IdoA-GlcNAc6S-ol

A total of ten structures were identified from the three major compositions observed for the enzyme-digested HS tetrasaccharide mixture (**Table 4.2**). Under the optimized isocratic LC condition, three major peaks (peak 1, 2 and 4) were observed with two minor peaks (peak 3 and peak 1') for species with MW of 990.490. As shown in **Figure 4.7**, EIC of 294 for Y_1 ion (GlcNAc), EIC of 512 for Y_2 ion (HexA-GlcNAc) and EIC of 791 for Y_3 ion (GlcNS6S-HexA-GlcNAc) were aligned with peak1, indicating a tetrasaccharide with sequence of d-HexA-GlcNS6S-HexA-GlcNAc. Another tetrasaccharide (d-HexA-GlcNAc6S-GlcA-GlcNS) was co-eluting in peak1, indicated by the alignment of 297 for Y_1 (GlcNS), 515 for Y_2 (HexA-GlcNS) and 791 for Y_3 (GlcNAc6S-HexA-GlcNS). A small peak (peak1') observed in EIC of 760 for Y_3 ion eluted immediately after peak1, which could be a small portion of d-HexA2S-GlcNS-HexA-GlcNAc and/or d-HexA2S-GlcNAc-GlcA-GlcNS. While peak3 could barely observed from the TIC, EIC of 294 for Y_1 ion, EIC of 760 for Y_3 ion and EIC of 512 for Y_2 ion were clearly shown aligned with peak3, indicating a sequence of d-HexA2S-GlcNS-HexA-GlcNAc. For peak2 and peak4, very similar MS/MS spectra were observed (**Figure 4.8**), with EIC of 325 for Y_1 ion

(GlcNAc6S), EIC of 543 for Y_2 ion (HexA-GlcNAc6S) and EIC of 791 for Y_3 ion (GlcNS-HexA-GlcNAc6S) aligned to each peak corresponding, which indicated peak2 and peak4 were epimers sharing a sequence of d-HexA-GlcNS-HexA-GlcNAc6S. There was no noticeable 546 for Y_2 ion (HexA-GlcNS6S), indicating the lack of sequence of d-HexA-GlcNAc-GlcA-GlcNS6S and/or d-HexA-GlcNAc-GlcA2S-GlcNS. To identify the epimerizations, we applied the rules that observed in the differentiation of synthetic tetrasaccharides GlcA-GlcNAc6S-IdoA-GlcNAc6S and GlcA-GlcNAc6S-GlcA-GlcNAc6S discussed in the previous section. With peak2 eluting earlier than peak4 and the relative abundant of Z_2 ion in their MS/MS spectra of $[M+Na]^+$ (**Figure 4.9**), we could confirm that peak2 was d-HexA-GlcNS-GlcA-GlcNAc6S, whereas peak4 was d-HexA-GlcNS-IdoA-GlcNAc6S. The same rules were also applied for the structure d-HexA-GlcNAc6S-GlcA-GlcNS and d-HexA-GlcNS6S-GlcA-GlcNAc in peak1, and d-HexA2S-GlcNS-IdoA-GlcNAc in peak3 (with zoom-in MS/MS spectra of $[M+Na]^+$ also shown in **Figure 4.9**).

Sequences were also identified for species with m/z of 957.466 (dHexA-GlcNAc-GlcA-GlcNAc) and 960.485 (dHexA-GlcNS-GlcA-GlcNAc with little dHexA-GlcNAc-GlcA-GlcNS), with MS/MS spectra and detailed interpretation included in **Figure 4.10**. With isocratic elution, sulfation positional isomers and epimers were fully or partially separated, allowing for accurate subsequent MS/MS analysis. Sulfation positional isomers could be identified with diagnostic Y ions, where the alignment of EIC of each diagnostic Y ion with the TIC provide a great help for assigning partially separated species. The differentiation of the internal epimerizatins was also achieved by applying the rules of elution order and MS/MS fragmentation of $[M+Na]^+$, both of which were consistent with what we observed for synthetic standards with free reducing end.

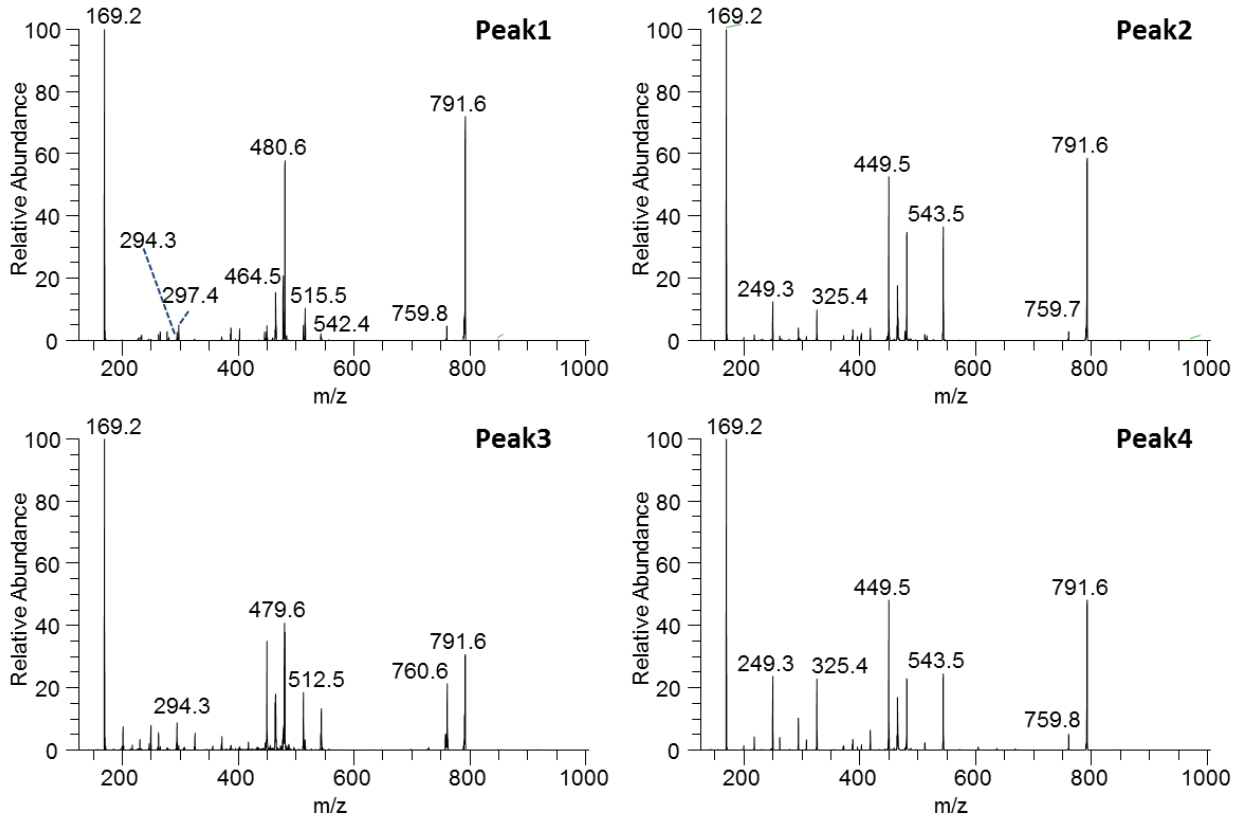


Figure 4.8. MS/MS spectra of parent ion $[M+2H]^{2+}$ (m/z of 496.25) extracted from the major four LC peaks shown in Figure 4.7. A series of B and Y ions were observed and used for the sequencing of sulfation positions. Y_1 ions: 294.3 (-GlcNAc), 297.4 (-GlcNS) and 325.4 (-GlcNAc6S). Y_2 ions: 512.5 (-HexA-GlcNAc), 515.5 (-HexA-GlcNS) and 543.5 (-HexA2S-GlcNAc or -HexA-GlcNAc6S). Y_3 ions: 760.6 (d-HexA2S) and 791.6 (d-HexA). The mass difference between ions with original N-acetyl group and N-sulfation group was 3Da, whereas a mass increase of 31Da would be observed for ions with O-sulfation group comparing to those non-sulfated ones. While sufficient information was obtained for sequencing the sulfation positions, there was no obvious difference for identifying epimers (peak2 *vs.* peak4) from these MS/MS spectra of parent ion $[M+2H]^{2+}$.

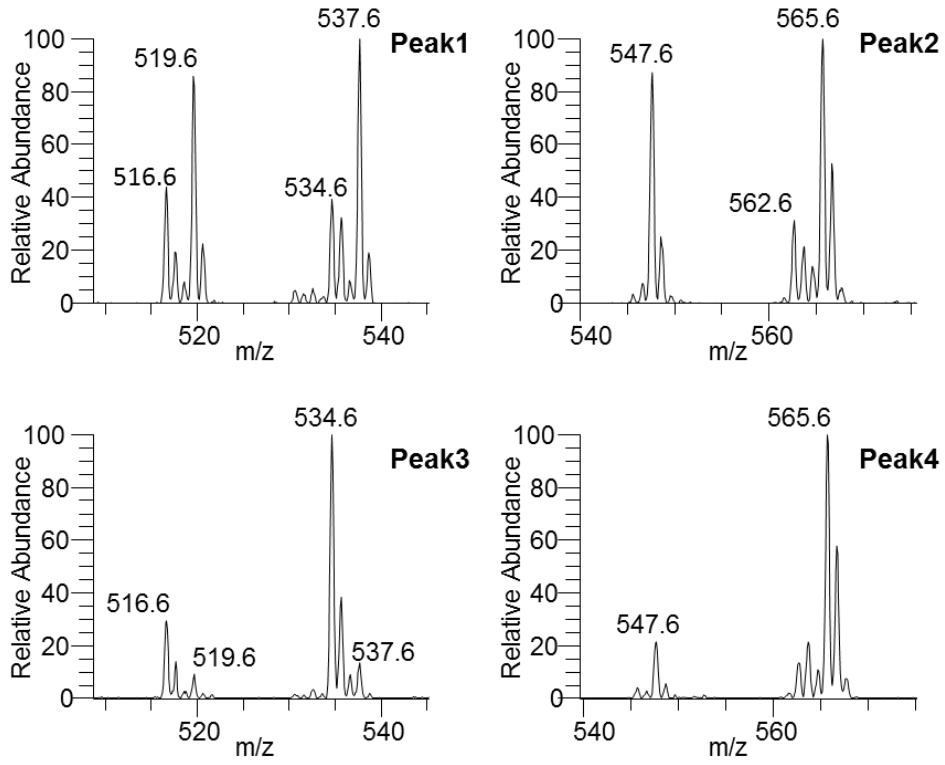


Figure 4.9. Zoom-in MS/MS spectra of parent ion $[M+Na]^+$ (m/z of 1013.48) extracted from the major four LC peaks shown in Figure 4.7. Observed Y_2 ions in each peak were: 537.6, 534.6, 565.6, 534.6 and 565.6, whereas the corresponding Z_2 ions were: 519.6, 516.6, 547.6, 516.6 and 547.6. Z_2 and Y_2 ions with m/z of 537.6 and 519.6 shown in peak1 were from structure d-HexA-GlcNAc6S-GlcA-GlcNS-ol, where a Z_2/Y_2 ratio of about 0.9 was observed. Because the internal uronic acid was adjacent to GlcNAc6S at the non-reducing-end side, it could only be an internal GlcA residue instead of IdoA. The observed Z_2/Y_2 ratio was about 0.9 and 0.2 for peak2 and peak4 respectively, indicating an internal GlcA for peak2 and internal IdoA for peak4.

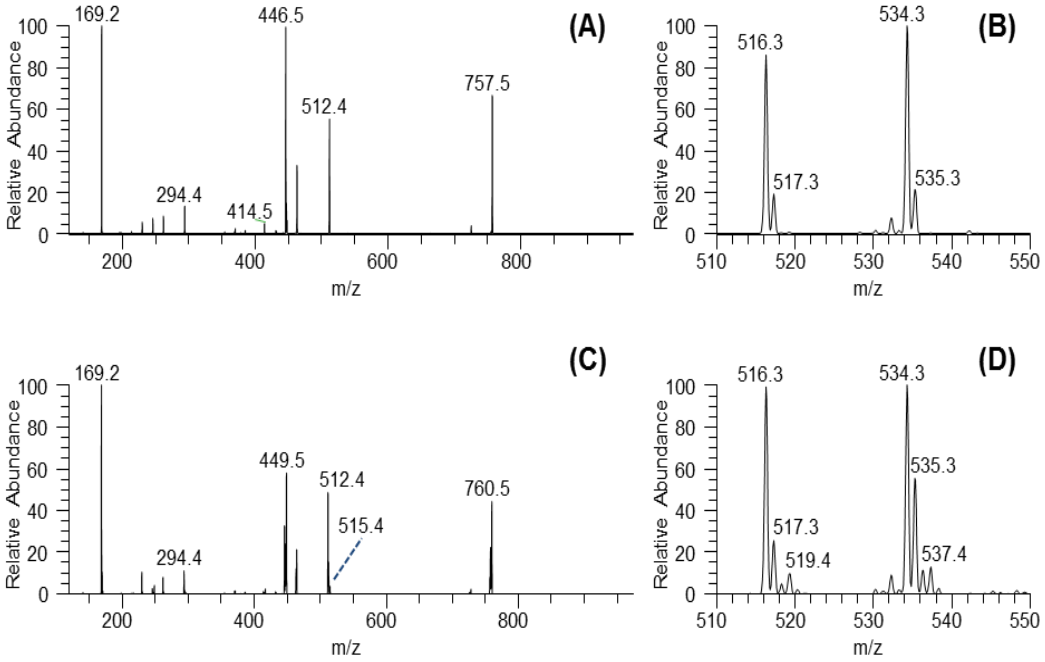


Figure 4.10. MS/MS spectra of parent ion $[M+2H]^{2+}$ and zoom-in MS/MS spectra of parent ion $[M+Na]^{+}$ for species with MW of 956.458 (A and B) and species with MW of 959.477 (C and D). Refer to Figure S-2 for their corresponding EIC and possible sequences. In spectra **A**, observed m/z value of 294 for Y_1 ion (GlcNAc), 512 for Y_2 ion (HexA-GlcNAc) and 757 for Y_3 ion (GlcNAc-HexA-GlcNAc) indicated a tetrasaccharide with sequence of d-HexA-GlcNAc-HexA-GlcNAc. Since the residue to its non-reducing end is GlcNAc, the internal HexA can only be GlcA, which is confirmed by having a Z_2/Y_2 ratio (516/534) of about 0.9 as shown in spectra **B**. For species with MW of 959.477 (spectra **C**), observed m/z value of 294 for Y_1 ion (GlcNAc), 512 for Y_2 ion (HexA-GlcNAc) and 760 for Y_3 ion (GlcNS-HexA-GlcNAc) indicated a tetrasaccharide with sequence of d-HexA-GlcNS-HexA-GlcNAc. With a much lower abundance, Y_2 ion with m/z of 515 was also observed, indicating the presence of d-HexA-GlcNAc-GlcA-GlcNAS. The Z_2/Y_2 ratio (516/534, 519/537) of about 0.9 observed in spectra **D** confirmed that both species had an internal GlcA.

Conclusions

The research described here represents the first successful separation and structural sequencing of HS tetrasaccharides with varying sulfation patterns as well as uronic acid epimerization, including a defined synthetic mixture of twenty-one tetrasaccharides as well as a mixture of tetrasaccharides from natural HS. Whereas successful separation and identification of non-isomeric HS oligosaccharides was previously achieved[32], the current study makes it possible to separate and differentiate sulfation positional isomers and epimers that are widely found in natural HS. The synthesis of a series of HS tetrasaccharides with defined structure made it possible to determine the relationship between structures and their chromatographic selectivity and fragmentation specificity. The chromatographic selectivity can be affected by three factors including degree of sulfation, sulfation position and uronic acid epimerization. While identification of sulfation positional isomers is relatively straightforward using diagnostic glycosidic bond cleavage fragments, distinguishing epimers is more challenging. However, through a combination of LC elution time and MS/MS analysis, epimers can be unambiguously differentiated. With preliminary applications on the native enzymatic digested HS tetrasaccharides mixtures, the method shows great promise for structural sequencing of complex HS oligosaccharides mixtures containing isomers and epimers, where detailed structural information beyond composition is required. Applications for sequencing of mixtures of longer HS oligosaccharides from affinity purification or complex biological systems are currently under investigation.

Acknowledgements

This research is supported in part by the National Institute of General Medical Sciences-funded "Research Resource for Integrated Glycotechnology" (P41 GM103390) from the

National Institutes of Health. The authors gratefully acknowledge Dr. Barry Boyes (Advanced Material Technology, Wilmington, DE) for generously providing us with the custom C18 HALO columns and helpful discussions regarding the reversed phase LC separation.

References

1. Gandhi, N. S.; Mancera, R. L. *Chemical biology & drug design* **2008**, *72*, 455-482.

2. Rabenstein, D. L. *Natural product reports* **2002**, *19*, 312-331.

3. Tumova, S.; Woods, A.; Couchman, J. R. *The international journal of biochemistry & cell biology* **2000**, *32*, 269-288.

4. Kresse, H.; Schonherr, E. *Journal of cellular physiology* **2001**, *189*, 266-274.

5. Fannon, M.; Forsten, K. E.; Nugent, M. A. *Biochemistry* **2000**, *39*, 1434-1445.

6. Lin, X. *Development* **2004**, *131*, 6009-6021.

7. Wu, Z. L.; Zhang, L.; Yabe, T.; Kuberan, B.; Beeler, D. L.; Love, A.; Rosenberg, R. D. *The Journal of biological chemistry* **2003**, *278*, 17121-17129.

8. Perrimon, N.; Bernfield, M. *Seminars in cell & developmental biology* **2001**, *12*, 65-67.

9. Gotte, M. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **2003**, *17*, 575-591.

10. Chen, Y.; Maguire, T.; Hileman, R. E.; Fromm, J. R.; Esko, J. D.; Linhardt, R. J.; Marks, R. M. *Nature medicine* **1997**, *3*, 866-871.

11. Toyoda, H.; Kinoshita-Toyoda, A.; Selleck, S. B. *The Journal of biological chemistry* **2000**, *275*, 2269-2275.

12. Theocharis, A. D.; Skandalis, S. S.; Tzanakakis, G. N.; Karamanos, N. K. *The FEBS journal* **2010**, *277*, 3904-3923.

13. Lindahl, U.; Backstrom, G.; Hook, M.; Thunberg, L.; Fransson, L. A.; Linker, A. *Proceedings of the National Academy of Sciences of the United States of America* **1979**, *76*, 3198-3202.

14. Guerrini, M.; Guglieri, S.; Casu, B.; Torri, G.; Mourier, P.; Boudier, C.; Viskov, C. *The Journal of biological chemistry* **2008**, *283*, 26662-26675.

15. Habuchi, H.; Habuchi, O.; Kimata, K. *Glycoconjugate journal* **2004**, *21*, 47-52.

16. Capila, I.; Linhardt, R. J. *Angewandte Chemie* **2002**, *41*, 391-412.

17. Garg, H. G.; Cindhuchao, N.; Quinn, D. A.; Hales, C. A.; Thanawiroon, C.; Capila, I.; Linhardt, R. J. *Carbohydrate research* **2002**, *337*, 2359-2364.

18. Guglier, S.; Hricovini, M.; Raman, R.; Polito, L.; Torri, G.; Casu, B.; Sasisekharan, R.; Guerrini, M. *Biochemistry* **2008**, *47*, 13862-13869.

19. Nugent, M. A.; Zaia, J.; Spencer, J. L. *Biochemistry. Biokhimiia* **2013**, *78*, 726-735.

20. Zaia, J. *Mass spectrometry reviews* **2009**, *28*, 254-272.

21. Korir, A. K.; Larive, C. K. *Analytical and bioanalytical chemistry* **2009**, *393*, 155-169.

22. Chi, L. L.; Amster, J.; Linhardt, R. J. *Curr Anal Chem* **2005**, *1*, 223-240.

23. Zaia, J. *Molecular & cellular proteomics : MCP* **2013**, *12*, 885-892.

24. Yang, B.; Chang, Y.; Weyers, A. M.; Sterner, E.; Linhardt, R. J. *Journal of chromatography. A* **2012**, *1225*, 91-98.

25. Gill, V. L.; Aich, U.; Rao, S.; Pohl, C.; Zaia, J. *Analytical chemistry* **2013**, *85*, 1138-1145.

26. Maxwell, E.; Tan, Y.; Tan, Y.; Hu, H.; Benson, G.; Aizikov, K.; Conley, S.; Staples, G. O.; Slysz, G. W.; Smith, R. D.; Zaia, J. *PloS one* **2012**, *7*, e45474.

27. Staples, G. O.; Naimy, H.; Yin, H.; Kileen, K.; Kraiczek, K.; Costello, C. E.; Zaia, J. *Analytical chemistry* **2010**, *82*, 516-522.

28. Kailemia, M. J.; Li, L.; Ly, M.; Linhardt, R. J.; Amster, I. J. *Analytical chemistry* **2012**, *84*, 5475-5478.

29. Leach, F. E., 3rd; Arungundram, S.; Al-Mafraji, K.; Venot, A.; Boons, G. J.; Amster, I. J. *International journal of mass spectrometry* **2012**, *330-332*, 152-159.

30. Wolff, J. J.; Leach, F. E., 3rd; Laremore, T. N.; Kaplan, D. A.; Easterling, M. L.; Linhardt, R. J.; Amster, I. J. *Analytical chemistry* **2010**, *82*, 3460-3466.

31. Huang, Y.; Yu, X.; Mao, Y.; Costello, C. E.; Zaia, J.; Lin, C. *Analytical chemistry* **2013**.

32. Huang, R.; Liu, J.; Sharp, J. S. *Analytical chemistry* **2013**, *85*, 5787-5795.

33. Huang, R.; Pomin, V. H.; Sharp, J. S. *Journal of the American Society for Mass Spectrometry* **2011**, *22*, 1577-1587.

34. Hitchcock, A. M.; Costello, C. E.; Zaia, J. *Biochemistry* **2006**, *45*, 2350-2361.

35. Hitchcock, A. M.; Yates, K. E.; Costello, C. E.; Zaia, J. *Proteomics* **2008**, *8*, 1384-1397.

36. Estrella, R. P.; Whitelock, J. M.; Packer, N. H.; Karlsson, N. G. *Analytical chemistry* **2007**, *79*, 3597-3606.

37. Karlsson, N. G.; Schulz, B. L.; Packer, N. H.; Whitelock, J. M. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2005**, *824*, 139-147.

38. Kuberan, B.; Lech, M.; Zhang, L.; Wu, Z. L.; Beeler, D. L.; Rosenberg, R. D. *Journal of the American Chemical Society* **2002**, *124*, 8707-8718.

39. Thanawiroon, C.; Linhardt, R. J. *Journal of chromatography. A* **2003**, *1014*, 215-223.

40. Seo, Y.; Andaya, A.; Leary, J. A. *Analytical chemistry* **2012**, *84*, 2416-2423.

41. Kailemia, M. J.; Park, M.; Kaplan, D. A.; Venot, A.; Boons, G. J.; Li, L.; Linhardt, R. J.; Amster, I. J. *Journal of the American Society for Mass Spectrometry* **2013**.

42. Dell, A.; Rogers, M. E.; Thomas-Oates, J. E. *Carbohydrate research* **1988**, *179*, 7-19.

43. Arungundram, S.; Al-Mafraji, K.; Asong, J.; Leach, F. E., 3rd; Amster, I. J.; Venot, A.; Turnbull, J. E.; Boons, G. J. *J Am Chem Soc* **2009**, *131*, 17394-17405.

44. Rudd, T. R.; Yates, E. A. *Molecular bioSystems* **2012**, *8*, 1499-1506.

CHAPTER 5

STRUCTURAL SEQUENCING OF AFFINITY PURIFIED ROBO1-BOUND HEPARAN SULFATE OCTASACCHARIDES[4]

[4] To be submitted.

Abstract

Binding of the secreted protein Slit to its transmembrane receptor Robo (roundabout) provides important signals involved in the development of CNS (central nervous system) and other organs, as well as in tumor angiogenesis and metastasis. While substantial evidence has been reported that heparan sulfate is strictly required for Slit-Robo signaling by binding to Slit and Robo simultaneously, knowledge of structure specificity of the bound regions of heparan sulfate is limited. To determine which HS sequences have high affinity binding to Robo, affinity purification of a HS octasaccharide library with immobilized human Robo1 is performed to isolate Robo1-bound sequences, the structural compositions of which are determined by high resolution mass spectrometry (HRMS). By combining chemical derivatization, on-line LC separation and MS/MS analysis as described in previous chapters, structure sequencing of the Robo1-bound HS octasaccharides are investigated in order to obtain more detailed structural information. An in-house developed program is also used to facilitate the MS/MS spectra interpretation as well as sequence identification.

Introduction

Robo (roundabout) is a transmembrane protein, documented as the cognate receptors for the secreted axon guidance molecule Slit[1]. In vertebrates, there are four types of Robos[2] (Robo1-4) and three different Slits[3] (Slit1-3). Recognition of Slit by Robo generates signals resulting in repulsion of axons away from the midline[4,5], which is essential for development of the nervous system[3,6]. Besides its neuron-related functions, Slit-Rob signaling is also required for the proper development of other organs including lung, kidney, heart and diaphragm[7-9], whereas aberrant Slit-Robo signaling contributes to human cancers by promoting tumor angiogenesis and metastasis[10-12]. Studies have shown that the Slit-Robo interaction is mediated by the D2 domain

of Slit and the IG1-2 domains of Robo, and heparan sulfate proteoglycans (HSPGs) are strictly required as essential co-receptors to promote such interaction through the formation of a ternary Slit-Robo-HS signaling complex[13,14]. The structure of the ternary complex is still unclear at present, and an understanding of Slit-Robo-HS interactions at the molecular level is of fundamental importance to investigating the mechanism of Slit-Robo signaling.

HS/heparin are linear, anionic polysaccharides composed of repeating disaccharide units with basic structure of [-4GlcAβ1-3GlcNAcα1-] that can undergo various post-polymerization modifications during their biosynthesis, including N-deacetylation and N-sulfation, O-sulfation and GlcA epimerization[15]. Heparin is a more heavily and uniformly modified variant of HS. Due to the differences in localization of HS and heparin, HS is considered to be the more physiologically relevant protein binding ligand. Heparin is produced exclusively in mast cells and sequestered in intracellular granules as free glycan chain[16], whereas HS, as the glycan module of HSPGs, is expressed on cell surface or secreted in extracellular matrix by almost all cell types[17]. Heparin is commonly used in biochemical assays as protein binding ligand for functional studies[18,19], for their high abundance, commercial availability and similar structural property but less complexity compared to HS. While heparin-based studies are quite valuable, HS may be a better representative for investigation of physiological relative protein-ligand interactions due to its localization *in vivo*, and diverse domain sequences provided by the sparse modification of HS are also essential for examination of structural specificity of HS-protein interactions.

A crystal structure of *Drosophila* Robo IG1-2 with a bound heparin octasaccharide has been reported by Hohenester *et al*[14], supporting that Robo is a heparin binding-protein. While the crystallographic studies and structure-based mutagenesis was used to identify Robo residues

involved in heparin binding, the structural specificity of heparin/HS required for their binding to Robo was not investigated. To address this issue, surface plasmon resonance (SPR) analysis is carried out by Wang *et al*[20] to study heparin binding to human Robo1 IG1-2, where SPR competition studies between heparin and several chemically modified heparins (with certain sulfation groups being removed) suggests that fine structure plays an important role in heparin-Robo1 interactions. With these results, we attempt to further explore the structural specificity of HS required for Robo binding by utilizing our chemical derivatization coupled LC-MS/MS strategy for detailed structural sequencing of HS oligosaccharides isolated from Robo1 immobilized affinity column.

Partial depolymerization of heparin/HS coupled with affinity chromatography is commonly used in studies of heparin/HS-protein interactions in order to isolate oligosaccharides with high affinity binding to a target protein[21,22]. Here, with the availability of human Robo1 IG1-2 and a library of HS octasaccharides, we isolate Robo1-bound HS octasaccharides which are subjected to high resolution mass spectrometry (HRMS) analysis followed by chemical derivatization and LC-MS/MS analysis, a strategy similar to that described in **Chapter 4** for the HS tetrasaccharide analysis. Because the structural complexity of HS oligosaccharides increases at an exponential rate as oligosaccharides get larger, manual interpretation of MS/MS spectra and identification of isomeric sequences can be very challenging. Therefore, an in-house developed program, GAG-ID, is also used to facilitate data analysis. By combining HRMS, chemical derivatization coupled with LC-MS/MS and semi-automatic data analysis, we seek to characterize Robo1-bound HS octasaccharides and thereby to determine how post-polymerization modifications affect the binding.

Experimental Section:

Materials. Green fluorescent protein (GFP)-tagged human Robo1 (IG domains 1 and 2), termed Robo1-GFP, was prepared by Dr. Moremen's lab as previously described[20]. Porcine intestinal heparan sulfate was purchased from Celsus Laboratories (Cincinnati, OH). Heparinase III was purchased from IBEX Technologies Inc. (Quebec, Canada). Gel filtration columns and packing materials were purchased from Bio-Rad (Hercules, CA) and packed as instructed by the manufacturer.

Generation of HS octasaccharides library. HS octasaccharide was prepared by enzymatic digestion of intact HS, followed by gel filtration purification. Briefly, 100 mg HS and 0.05 IU heparinase III were mixed in 1.5 mL buffer of 50 mM sodium acetate and 0.4 mM calcium acetate. The solution was incubated at 37ºC for a total of 48 h, with an additional aliquot of 0.05 IU heparinase III added after 24 h incubation. The enzyme digestion was stopped by heating the solution at 100ºC for 10 min. The depolymerized samples were subjected to gel filtration chromatography on a Bio-Gel P-10 column (2.5×100 cm) using 10% ethanol solution containing 1 M NaCl as a mobile phase to obtain octasaccharide fractions, which was further desalted through a Sephadex G-15 column (2.5×50 cm) and dried under vacuum.

Affinity purification of HS octasaccharides. Robo1-GFP or GFP was biotinylated by biotin-protein ligase (Avidity LLC, Aurora, CO) based on the manufacturer's protocol. Robo1-GFP-biotin was bound at 3.2 mg/mL on streptavidin-agarose resin (Thermo Scientific Pierce, Rockford, IL). GFP-biotin was bound at 1.6 mg/mL on streptavidin-agarose resin. 1mL of each resin was poured into empty plastic columns and equilibrated in 0.15 M ammonium acetate buffer (pH 7.4), designated as Robo1-GFP column and GFP column respectively.

An HS octasaccharide library (0.7-0.8 mg) in 1 mL 0.15 M ammonium acetate buffer was loaded to the GFP column and the column was capped and rotated at room temperature (RT) for 2 h. After incubation, the flow-through fraction was collected and pooled with a wash of 3 mL 0.15 M ammonium acetate (with final volume of approximate 4 mL). The column was further washed by additional 10 mL 0.15 M ammonium acetate. The bound HS octasaccharides were eluted and collected in 8 mL 2.0 M ammonium acetate, being designated as *GFP-bound fraction*. The collected flow-through and wash fraction in 4 mL of 0.15 M ammonium acetate was loaded to the Robo1-GFP column and rotated at RT for 2 h. After incubation, the column was extensively washed by 10 mL 0.15 M ammonium acetate. The bound HS octasaccharides were eluted and collected in 8 mL of 2.0 M ammonium acetated, being designated as *Robo1-bound fraction.*

Three affinity purifications were performed in parallel using the same preparation of HS octasaccharide library (referred as dp8 library A); with GFP-bound fractions and Robo1-bound fractions collected from three separate sets of GFP column and Robo1-GFP column. Each fraction (referred as GFP-bound fraction A1/A2/A3 and Robo1-bound fraction A1/A2/A3) was lyophilized to remove ammonium acetate, together with the dp8 library A, were subjected to HRMS by direct infusion for composition analysis. Another preparation of HS octasaccharide library (referred as dp8 library B) was also applied for affinity purification with one of each GFP-bound fraction and Robo1-bound fraction (referred as GFP-bound fraction B and Robo1-bound fraction B) collected; and all three samples were subjected to hydrophilic interaction liquid chromatography (HILIC)-MS for composition analysis.

Composition analysis by high resolution mass spectrometry (HRMS). Mass spectrometry analysis was performed on a 9.4 T Bruker Apex Ultra QeFTMS (Billerica, MA) by

direct infusion, using nano-electrospray ionization (nano-ESI) on negative ion mode. One fifth of GFP-bound fraction A1 and each Robo1-bound fraction (A1, A2 and A3) suspended in 10 µL 50% methanol/water containing 0.2% formic acid was injected, and a continue flow of 50% methanol at 0.25uL/min was used to push the sample through for ionization. 100 acquisitions were signal averaged per mass spectrum. External calibration of mass spectra produced mass accuracy of 5ppm. The dp8 library A (1 µg) was also analyzed under the same condition. Mass spectra were processed in Data Analysis 4.0 (Bruker Daltonics), and deconvoluted mass and corresponding abundance were transferred into an Excel template for composition analysis.

Hydrophilic interaction liquid chromatography (HILIC) LC-MS was performed on a Thermo LTQ Orbitrap XL instrument coupled with Surveyor HPLC System (Waltham, MA), A home-made spray tip (0.08×130 mm,) packed with amide-80 material (3µm, TOSOH Biosciences) to 110 mm long was used for LC separation. Buffer A was 80% 55mM ammonium formate in water (pH 4.4), and buffer B was prepared as 95% acetonitrile and 5% buffer A. A linear gradient of 65%-30% buffer B over 30 min was used, with a flow rate of 135 µL/min set for the pump delivering an actual column flow rate between 0.3~0.4 µL/min after splitting. MS analysis was performed in Orbitrap in the negative high resolution mode. The dp8 library B (5ug), GFP-bound fraction B and Robo1-bound fraction B were suspended in 50 µL 65% buffer B respectively and 10 µL was loaded onto the amide-80 packed tip for LC-MS analysis.

Chemical derivatization of Robo1-bound HS octasaccharides. The remaining three Robo1-bound fractions from dp8 library A were mixed and subjected to the chemical derivatization protocol as described in previous chapters. The remaining Robo1-bound fraction B was also subjected to the chemical derivatization. Briefly, the HS octasaccharides were firstly converted into their triethylammonium (TEA) salts form and lyophilized. The dried HS

octasaccharide TEA salts were permethylated using sodium hydroxide and methyl iodide in dimethyl sulfoxide (DMSO), followed by desalting using a C18 Sep-Pak cartridge (Waters Co.). The pyridinium salts of the permethylated products were re-suspended in DMSO containing 10% methanol and incubated for 4h at 95°C to remove the sulfate groups. The dried desulfated products were then trideuteroacetylated by incubating with D_6-acetic anhydride in pyridine at 50°C overnight and solvent was dried under vacuum.

LC-MS/MS analysis. On-line LC separation was performed on a Thermo Finnigan Surveyor HPLC System, using a home-made spray tip (0.08×130 mm,) packed with C18 resin (5μm, 300Å, Mettler-Toledo, LLC) to 110 mm long. Buffer B was prepared as 80% acetonitrile/water with 0.1% formic acid and 1mM sodium acetate, and buffer A was 0.1% formic acid in water with 1mM sodium acetate. A linear gradient of 20%-100% buffer B over 80 min was used, with a flow rate of 135 μL/min set for the pump delivering an actual column flow rate between 0.3~0.4 μL/min after splitting. The derivatized Robo1-bound HS octasaccharides were suspended in 10 μL of 20% buffer B and 5 μL was loaded onto the C18 packed tip for LC-MS/MS analysis.

Mass spectrometry was performed on a Thermo LTQ Orbitrap XL instrument (Waltham, MA). Both full MS and CID-MS/MS spectrum were acquired by orbitrap on positive ion mode for Robo1-bound fraction A. For Robo1-bound fraction B, the CID-MS/MS spectrum was taken in ion trap mode. A data dependent MS/MS method was used, with the top 6 abundant precursor ions selected, to trigger CID-MS/MS fragmentation. Instrument parameters were set as: spray voltage at 1-2kV, capillary voltage at 40V, tube lens at 80V and capillary temperature at 250ºC. The collision energy for CID fragmentation was set at 40V.

Automatic sequencing by GAG-ID. The LC-MS/MS data collected for the derivatized HS octasaccharides, in "*.raw" file format readable by Xcalibur (Thermo Scientific), was converted into "*.mgf" file format, which would be processed by a program GAG-ID developed in our lab for automatic identification of HS oligosaccharides sequences (as illustrated in **Figure 5.1**). The tolerances for full MS and MS^2 were set as 0.05Da and 0.1Da for Robo1-bound fraction A sample, and 0.05Da and 0.5Da for Robo1-bound fraction B sample. Other parameters were set as: sodium adduct, non-reduced, no tag, and charge up to +2. Separate sequencing was performed against octasaccharides (dp8) database, and hexasaccharides (dp6) and tetrasaccharides (dp4) databases as well to explore the β–elimination by products generated from permethylation of octasaccharides. Further interpretation and confirmation of the identified sequences were done by manually going through the corresponding full MS and MS^2 spectrum, as certain MS spectra might be represented by more than one isomeric structure that were not separated by the LC prior to their fragmentation.

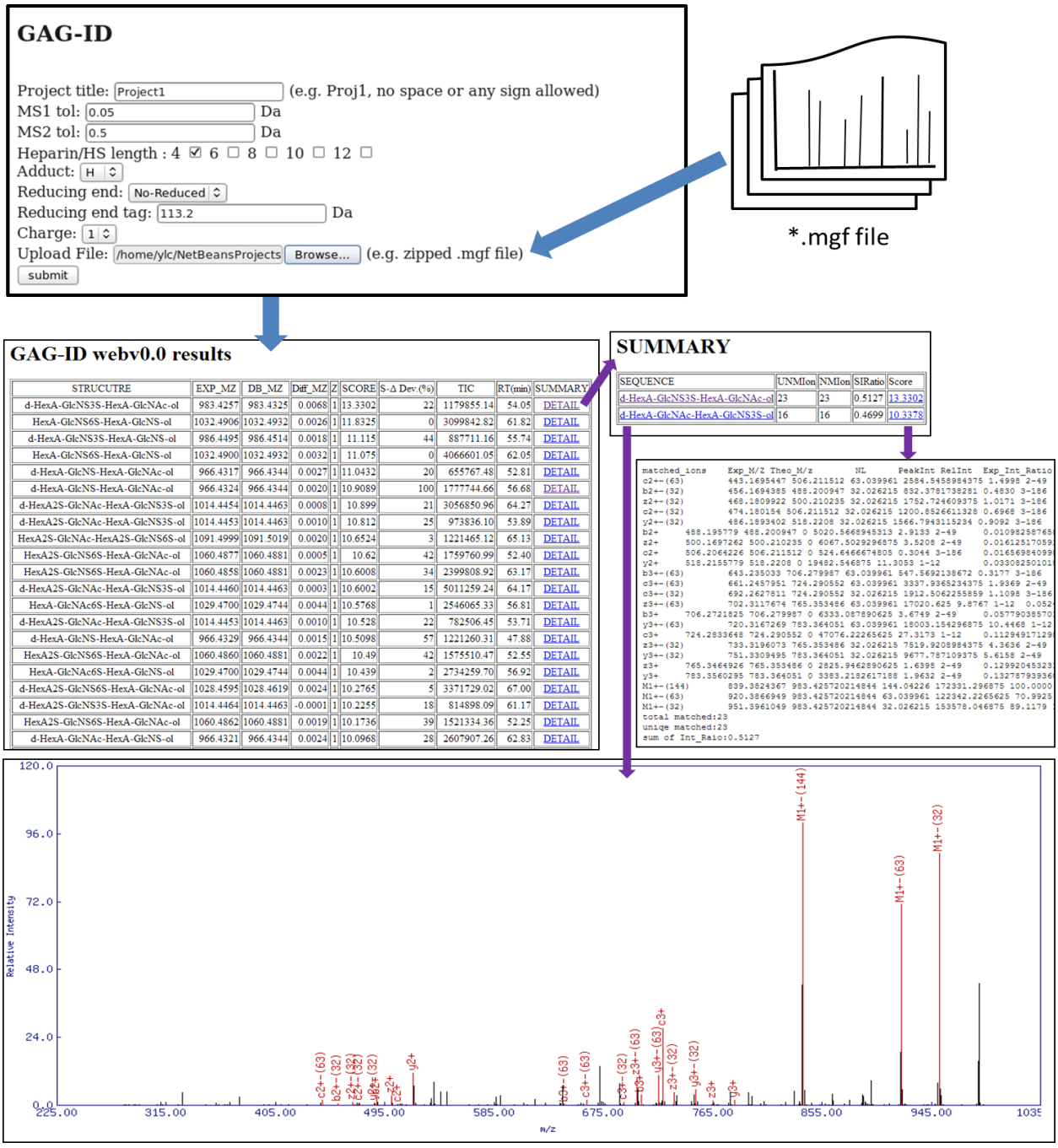


Figure 5.1. Automatic sequencing by program GAG-ID. The interface of GAG-ID for parameters input and mgf file upload was illustrated, together with web-based results display showing a list of identified sequence with top score for each MS/MS spectrum with retention time indicated. Details about spectrum matching and scores for all theoretical possible sequences under the given mass could also be pulled out to facilitate manual confirmation.

Results and Discussion

Affinity purification of Robo1-bound HS octasaccharides. HS octasaccharides derived from partially enzymatic digestion were passed through a GFP affinity column in low-salt buffer, and then applied to Robo1-GFP affinity column, where Robo1-bound fractions were eluted using a high-salt buffer. Because the Robo1 was expressed with GFP tag, the GFP affinity column was used as a control to identify HS octasaccharides that may bind to GFP non-specifically. By comparing the Robo1-bound fractions with GFP-bound fractions, HS octasaccharides compositions detected exclusively or with higher abundance in Robo1-bound fractions would be considered as specific Robo1-bound ligands, and therefore are more important.

Composition analysis by HRMS. While chemical derivatization strategy was useful for detailed structural sequencing, the sensitivity was sacrificed by performing multiple derivatization steps. Therefore, MS analysis of underivatized HS dp8 fractions was performed to obtain composition information prior to chemical derivatizations, which can also narrow down the possible sequences existed within the sample of interest. For the direct infusion HRMS results, 10 ppm mass accuracy was used to determine HS dp8 compositions. The triplicate results for composition analysis of Robo1-bound fractions (A1, A2 and A3) were presented in **Figure 5.2**, with Robo1-bound fraction A1, GFP-bound fraction A1 and dp8 library A presented together in **Figure 5.3**. HS oligosaccharide compositions were presented as [ΔHexA, HexA, GlcN, Ac, SO_3] and determined by processing deconvoluted peak lists in an Excel template modified based on GlycReSoft developed by Zaia *et al*[23]. The most abundant compositions identified from Robo1-bound fractions were [1,3,4,0,X] (X=6,7,8) and [1,3,4,1,X] (X=4,5,6) (**Figure 5.2**), showing low N-acetylation and moderate sulfation densities. In **Figure 5.3**, two compositions [1,3,4,1,5] and [1,3,4,1,6] were observed only in Robo1-bound fraction and the

unfractionated library, but not in GFP-bound fraction, indicating that these structures have specific binding with Robo1.

Composition analysis of dp8 library B, GFP-bound fraction B and Robo1-bound fraction B were also performed by using GlycReSoft to process deconvoluted HILIC-MS data as described in the literature[23]. Four dp8 compositions, [1,3,4,0,8], [1,3,4,1,7], [1,3,4,1,8] and [1,3,4,2,6] were observed in the Robo1-bound fraction but not the GFP-bound fraction (**Figure 5.4**), indicating that these structures have specific binding with Robo1. Differences between the composition analysis results from the two sets of affinity purification experiment may due to two major respects: first, the content of the dp8 library may vary from one preparation to another; second, ionization efficiency is usually composition as well as instrumentation dependent, meaning the relative abundant is not comparable between data collected from different MS instruments.

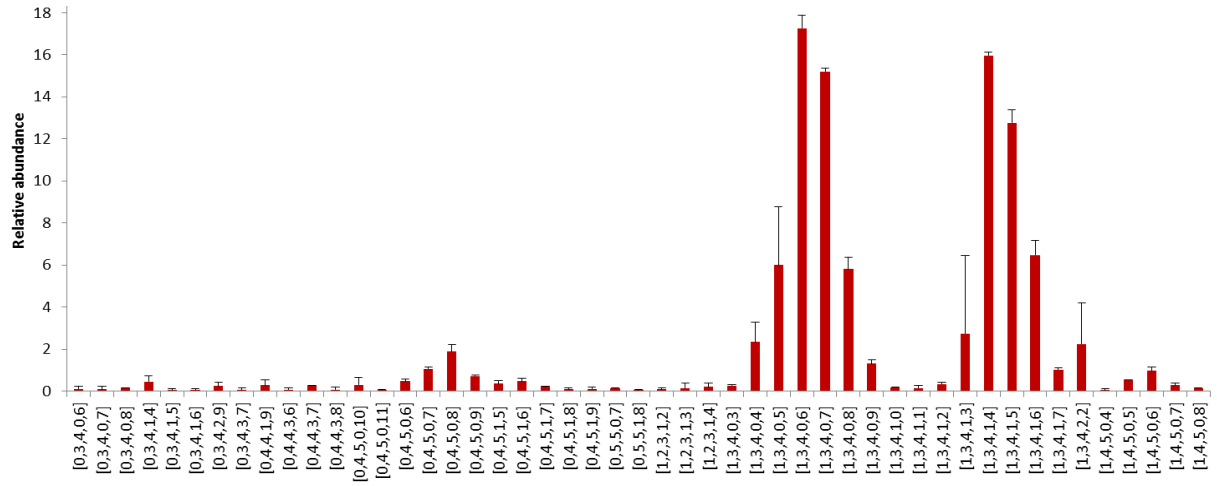


Figure 5.2. Triplicate results for composition analysis of dp8 Robo1-bound fraction A1, A2 and A3. The composition is given as: [ΔHexA, HexA, GlcN, Ac, SO_3] (ΔHexA, unsaturated uronic acid; HexA, saturated uronic acid; GlcN, glucosamine; Ac, acetyl; SO_3, sulfate).

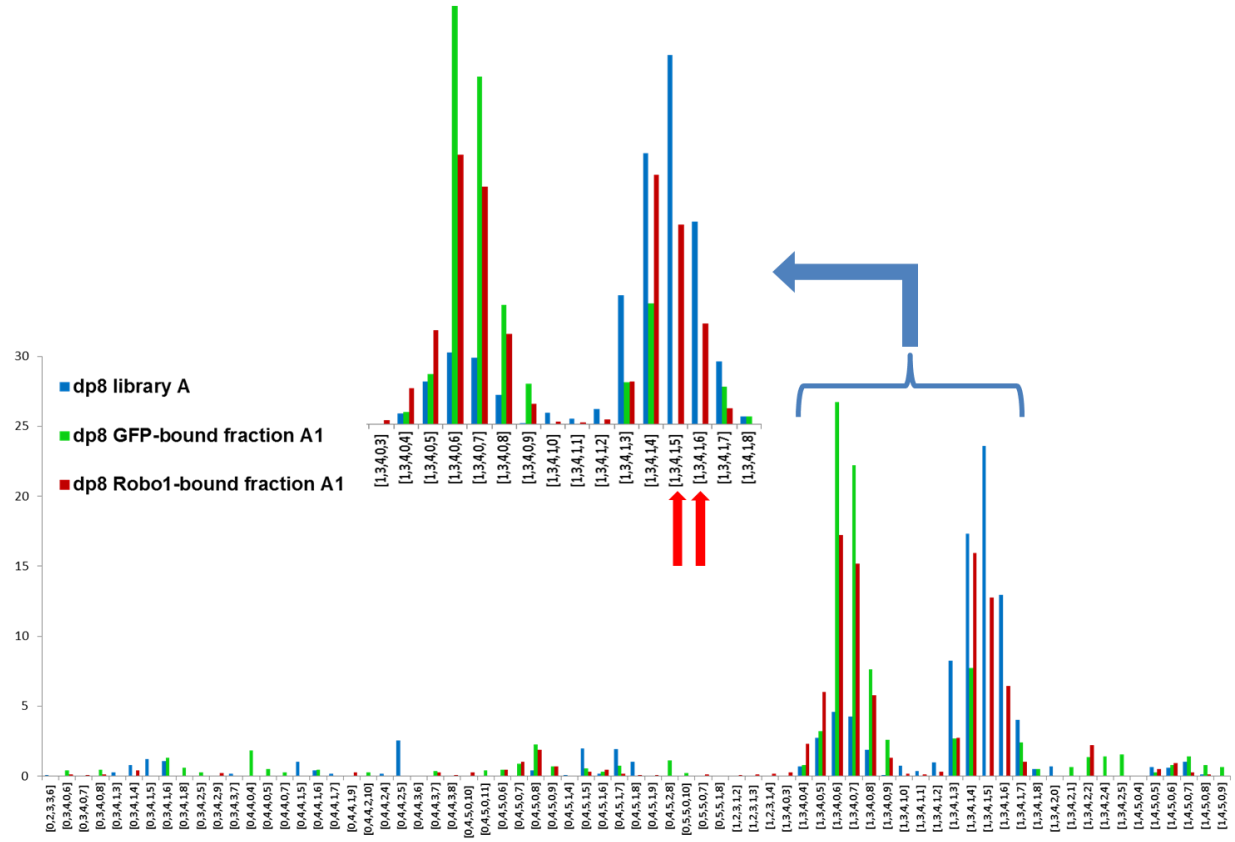


Figure 5.3. Composition analysis of dp8 library A, GFP-bound fraction A1 and Robo1-bound fraction A1. Each composition is given as follows: [ΔHexA, HexA, GlcN, Ac, SO_3]. The region with most abundant compositions was enlarged, where the two compositions only observed in the library and the Robo1-bound fraction were indicated by red arrows.

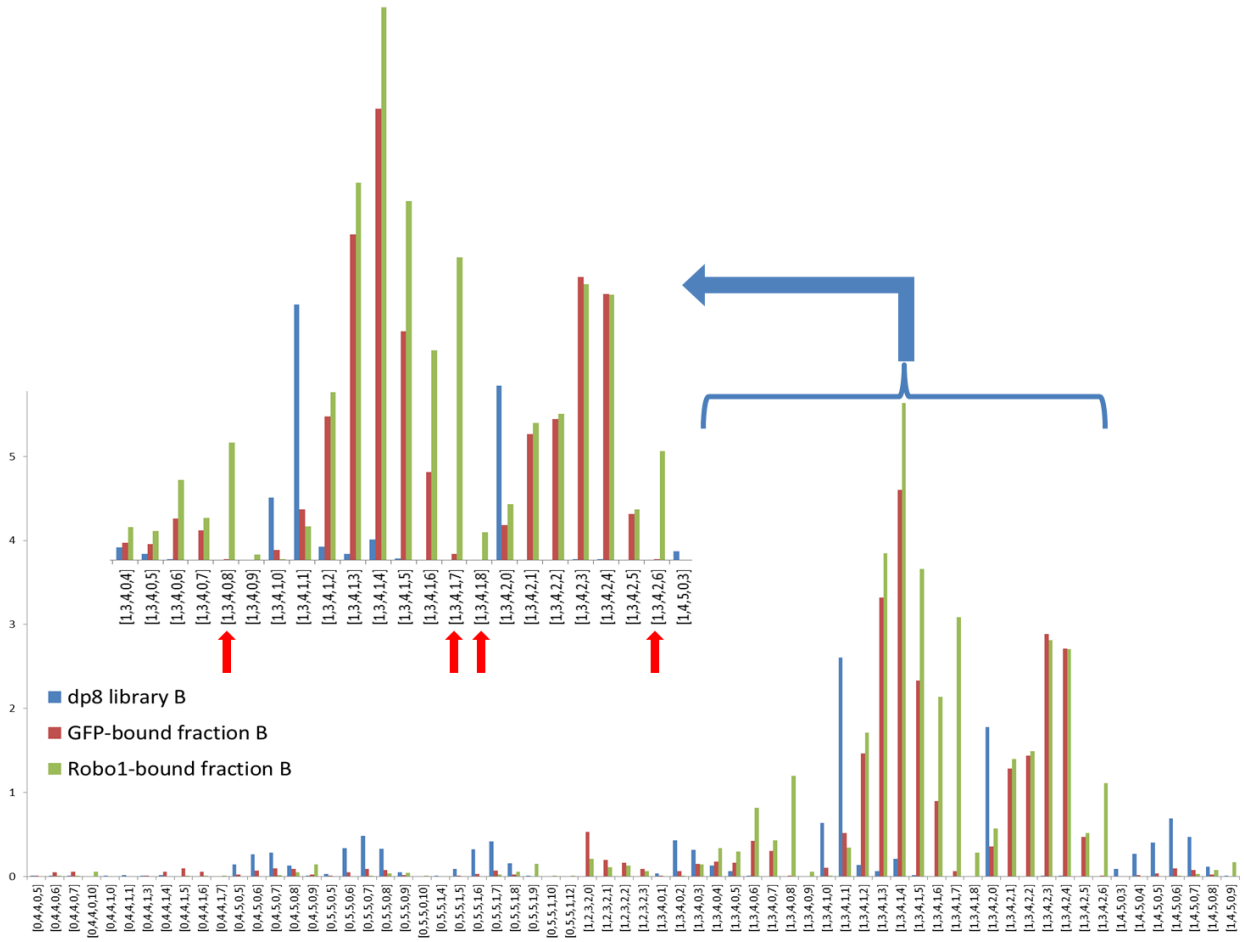


Figure 5.4. Composition analysis of dp8 library B, GFP-bound fraction B and Robo1-bound fraction B. Each composition is given as follows: [ΔHexA, HexA, GlcN, Ac, SO_3]. The region with most abundant compositions was enlarged, where the four compositions only observed in the Robo1-bound fraction were indicated by red arrows.

Chemical derivatization of Robo1-bound HS octasaccharides. While the exact material amount subjected to chemical derivatizations was unknown, the same protocol was used as described in previous chapters including permethylation, desulfation and trideutero-acetylation. Half of the final products were applied for LC-MS/MS analysis. The file was first searched against dp8 database using GAG-ID, but no sequence was identified, as well as for dp6

database (data not shown). HS sequences started to show up when we searched against dp4 database, indicating the existence of several HS tetrasaccharides sequences. We believe that this is caused by the β–elimination at uronic acid residues during the permethylation step as disused in previous chapters, generating shorter sequences as the by-products (**Figure 5.5**). The extent of β–elimination depends on length, derivatization conditions and possibly sequence. While it has been believed that the high basicity involved in permethylation was the reason causing the β–elimination, we have not been able to find a condition that can prevent such β–elimination without scarificing the permethylation efficiency. An alternative approach may be performing carboxylic acid reduction[24] prior to permethylation, which could avoid the β–elimination by converting highly active uronic acid to much less active hexose. This step can replace the original reduction step as the reducing end would be reduced as well under the harsher condition involved in carboxylic acid reduction. Nevertheless, the generated tetrasacchrides represent partial sequences of the original octasaccharides, and therefore are still useful for structural characterization.

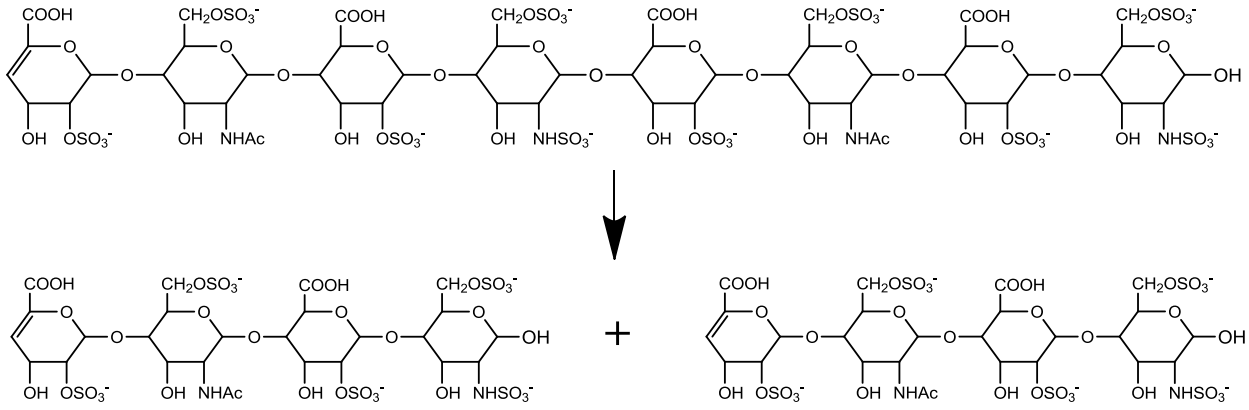
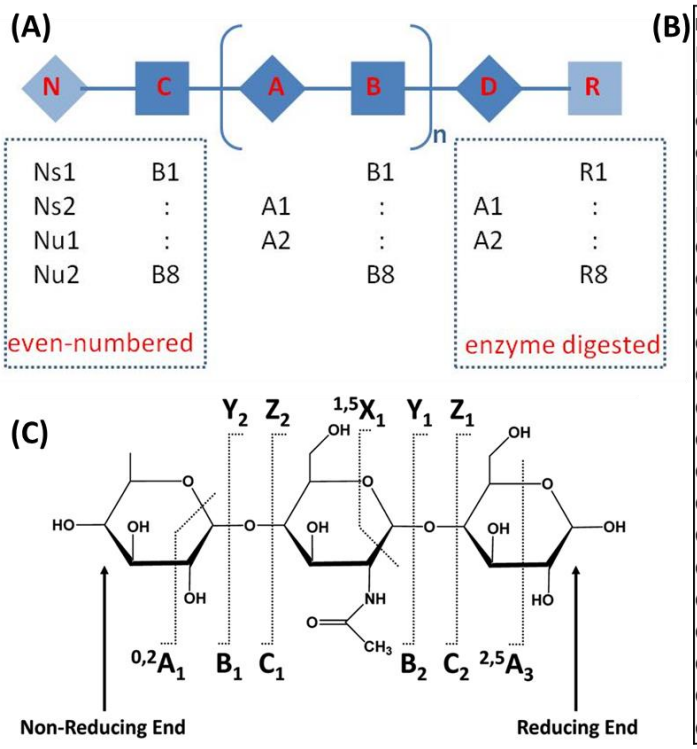


Figure 5.5. Illustration for products generated by β–elimination at uronic acid residues during permethylation step.



(B)

residue	ID	Non-reduced	reduced
HexA	Ns1	249.097430	249.097430
HexA2S	Ns2	280.111176	280.111176
dHexA	Nu1	217.071215	217.071215
dHexA2S	Nu2	248.084961	248.084961
HexA	A1	218.079040	218.079040
HexA2S	A2	249.092786	249.092786
GlcNH	B1	232.154335	232.154335
GlcNH6S	B2	263.168081	263.168081
GlcNAc	B3	245.126324	245.126324
GlcNAc6S	B4	276.140070	276.140070
GlcNS	B5	248.145155	248.145155
GlcNS6S	B6	279.158901	279.158901
GlcNS3S	B7	265.143251	265.143251
GlcNS3S6S	B8	296.156997	296.156997
GlcNH	R1	247.177810	263.209110
GlcNH6S	R2	278.191556	294.222856
GlcNAc	R3	260.149799	276.181099
GlcNAc6S	R4	291.163545	307.194845
GlcNS	R5	263.168630	279.199930
GlcNS6S	R6	294.182376	310.213676
GlcNS3S	R7	280.166726	296.198026
GlcNS3S6S	R8	311.180472	327.211772

Figure 5.6. Database build-up for GAG-ID. (**A**) Residues for building heparin/HS sequences. (**B**) Theoretical mass for each residue: HexA, glucuronic acid or idoronic acid; d, double bond. (**C**) Fragmentation nomenclature of carbohydrates as described by Domon and Costello[25].

Automatic sequencing by GAG-ID. An automatic sequencing program, GAG-ID, is developed in our lab for the chemical derivatization based LC-MS/MS structural analysis of heparin/HS oligosaccharides, which use a database specific designed for derivatized heparin/HS oligosaccharides (**Figure 5.6**). A theoretical database for oligosaccharides with different lengths can be generated as illustrated in **Figure 5.6A**, with theoretical mass for each residue listed in **Figure 5.6B**. Both saturated and unsaturated non-reducing end are considered, while the reducing end can be either reduced or non-reduced before the sequential derivatizations (permethylation, desulfation and trideutero-acetylation). For glucosamine residue, the relatively rare 3-O-sulfation is also considered in addition to regular modifications, which presents as

GlcNS3S or GlcNS3S6S in current version of GAG-ID. Sequential glycosidic-bond cleavage fragments, B, Y, C and Z ions (as illustrated in **Figure 5.6C**), as well as commonly observed neutral losses for derivatized GAGs, are also calculated to generate theoretical fragment lists. Each isomeric sequence under the same mass entry has its own fragment list, which can be used to match against experimental MS^2 peak list and a score would be calculated basing on number and relative intensity of the matched product ions (**Figure 5.1**).

As illustrated in **Figure 5.1**, relative parameters can be used to direct the automatic sequencing. Only those MS^2 with parent ion mass found in defined database within tolerance will be selected for fragments matching. Web-based results are reported by listing the top1 (with highest score) matching sequence for each MS^2, where details about the spectrum matching and score calculating can be easily accessed for each possible isomeric sequence to assist manual confirmation for each identified sequence. With the first entry of the search results for example (**Figure 5.7**), the identified sequence can be further confirmed by pulling out the spectrum view for each possible isomeric sequence. By matching Y_2 and C_3 ions (indicated by red arrows), the top1 sequence has a significant higher score than top2 and is correctly assigned for the corresponding spectrum.

The [S-Δ Dev.] was calculated as $(score_{top1}\text{-}score_{top2})/score_{top1}$ for top1 and top2 isomeric sequences to evaluate if the top1 sequence has significant better fitting. In the example shown in **Figure 5.7**, the [S-Δ Dev.] value is 22, which we know, by interpreting the spectrum, that it represents significant better matching of the top 1 sequence than the top2 one. On the other hand, in cases where the [S-Δ Dev.] value is relatively low, even zero, it usually indicates that either the fragments are not informative enough to differentiate certain isomers (**Figure 5.8**) or the spectrum is extracted from overlapping LC peaks containing more than one isomeric structure

(**Figure 5.9**). In either case, manual interpretation of the MS^2 spectrum by aligning corresponding LC chromatogram, if isomeric structures are separated or partially separated, is required to give more confident results.

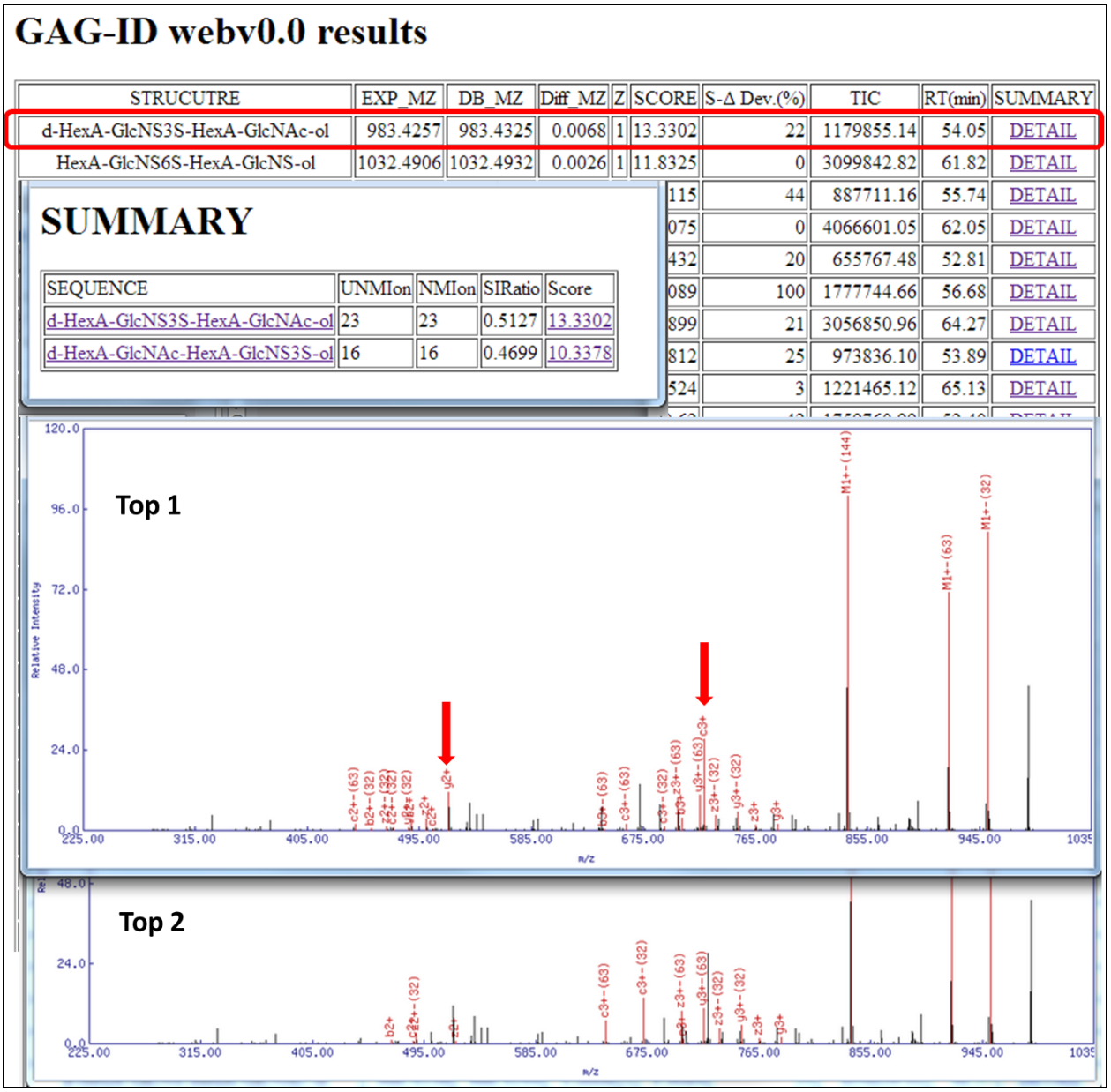


Figure 5.7. A snapshot of the GAG-ID sequencing of tetrasaccharides detected in derivatized Robo1-bound HS octasaccharides. Detail information is pulled out for the first entry listed on the results.

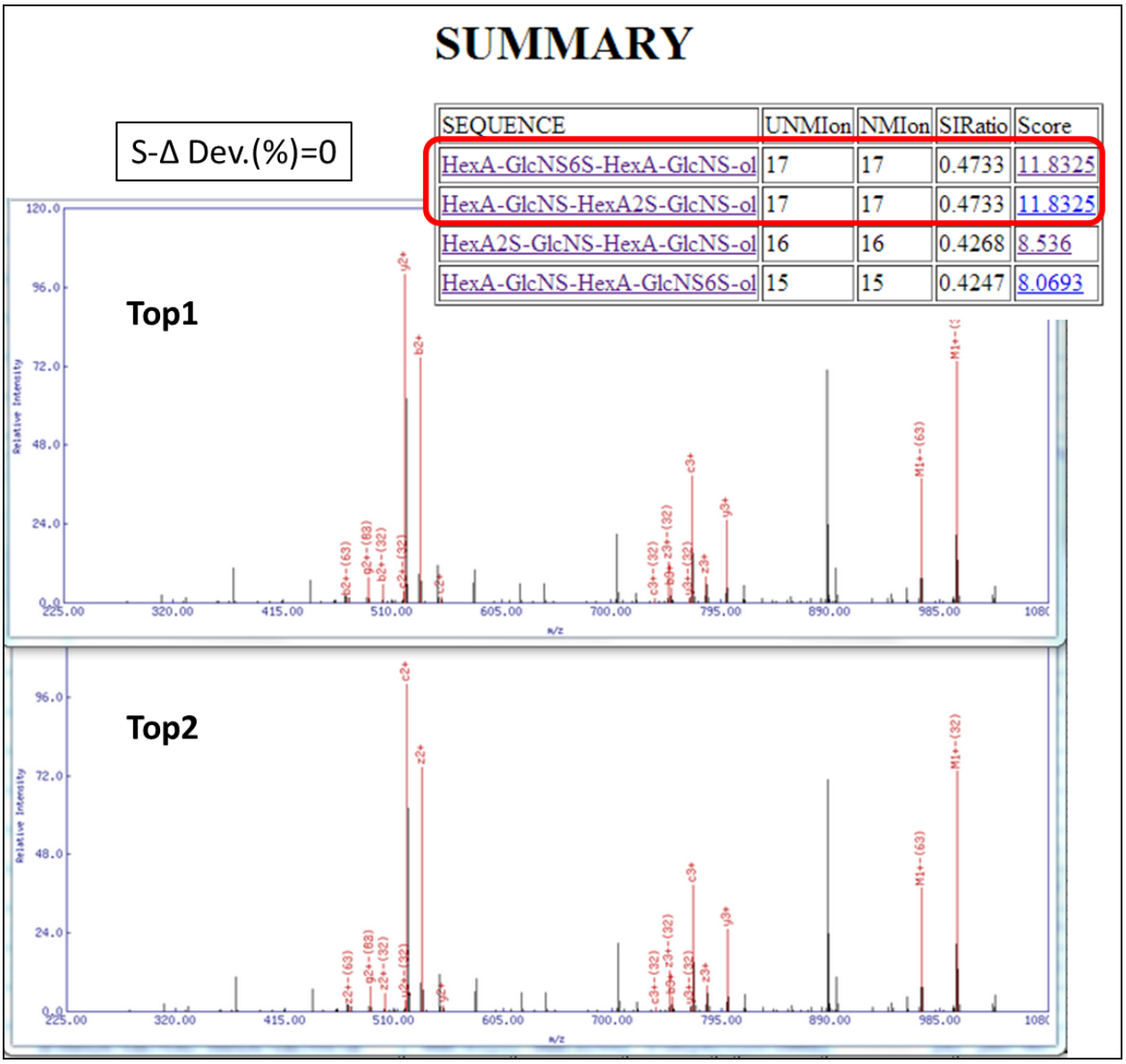


Figure 5.8. An example showing two sequences being assigned to one MS^2 spectrum with same score. In this case, the top1 and top2 sequences share symmetric fragments in terms of mass, where the potential diagnostic ion B_2 and Y_2 ion of top1 sequence share the same m/z with Z_2 and C_2 ion of top2 sequence.

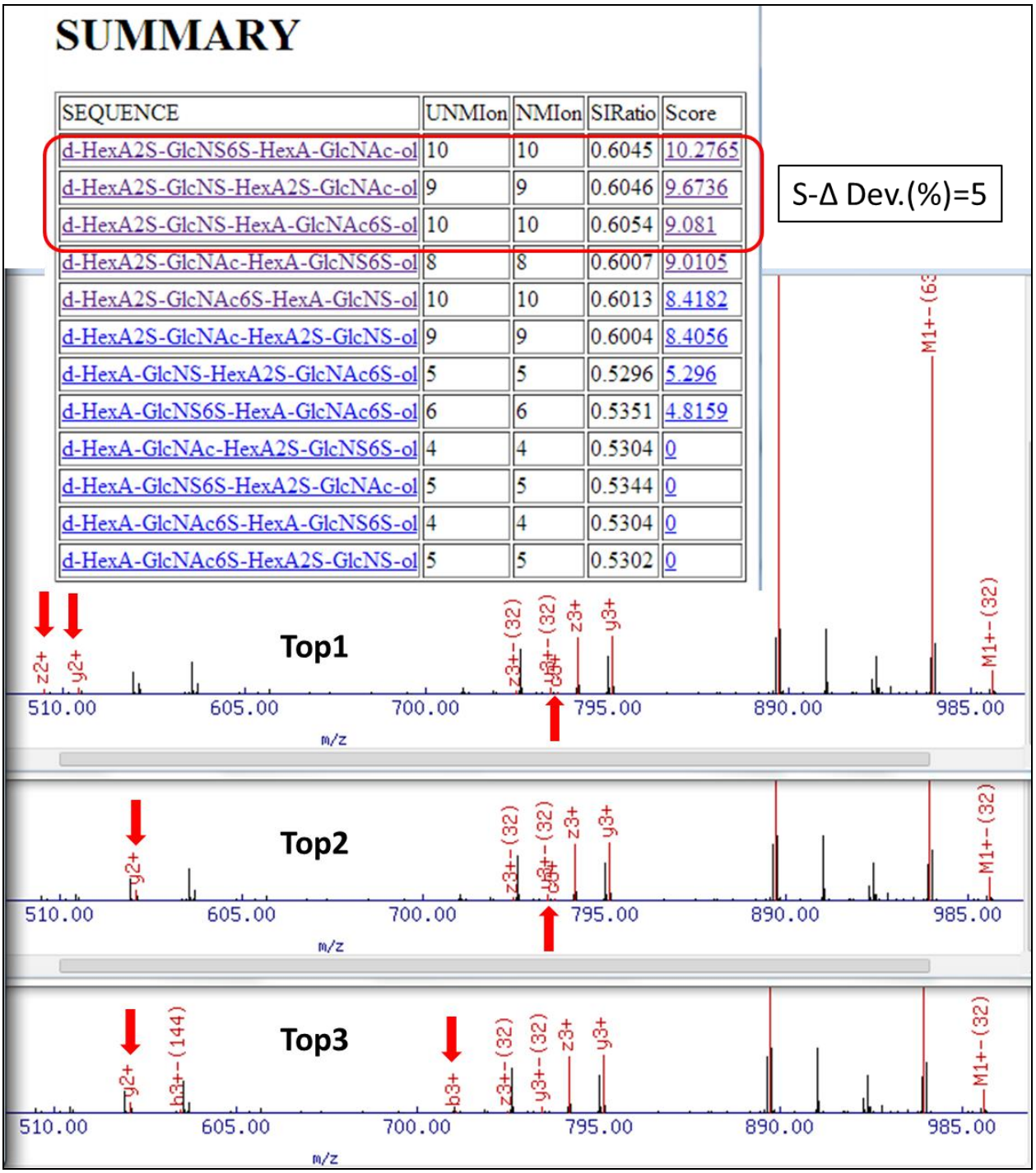


Figure 5.9. A typical MS^2 indicating the existence of more than one isomeric structures. Glycosidic-bond cleavage fragments of each sequence can be observed in relative equal quantities and abundance, explaining the relatively close scores for the first few sequences (only three spectrums are attached for illustration).

The GAG-ID sequencing results for the Robo1-bound fraction A sample included a total of 393 MS^2 spectrums, in which 13 unique HS tetrasaccharides compositions were identified (only those with score higher than 3 are considered) with a total of 43 unique sequences ranked as a top1 sequence for corresponding MS^2. Each of these sequences was manually checked, and only predominant ones identified with confidence were listed in **Table 5.1**, where undistinguishable sequences are all listed for certain saturated species ([0,2,2,0,3] (**Figure 5.8**) and ([0,2,2,1,2]).

Table 5.1. Identified tetrasaccharides sequence from the derivatized HS octasaccharides in the Robo1-bound fraction A sample. ([a,b] undistinguishable sequences are all listed)

STRUCUTRE	DB_MZ	Composition	Total # of possible sequences
HexA2S-GlcNAc-HexA2S-GlcNS6S-ol	1091.502	[0,2,2,1,4]	8
HexA-GlcNAc-HexA2S-GlcNS6S-ol HexA2S-GlcNS6S-HexA-GlcNAc-ol	1060.488	[0,2,2,1,3]	12
HexA-GlcNS-HexA2S-GlcNS-ol HexA-GlcNS6S-HexA-GlcNS-ol	1032.493[a]	[0,2,2,0,3]	4
HexA-GlcNS-HexA2S-GlcNAc-ol HexA-GlcNS6S-HexA-GlcNAc-ol HexA-GlcNAc-HexA-GlcNS6S-ol HexA-GlcNAc-HexA2S-GlcNS-ol HexA-GlcNAc6S-HexA-GlcNS-ol HexA2S-GlcNS-HexA-GlcNAc-ol	1029.474[b]	[0,2,2,1,2]	8
d-HexA2S-GlcNS6S-HexA-GlcNAc6S-ol	1059.476	[1,1,2,1,4]	8
d-HexA2S-GlcNS6S-HexA-GlcNAc-ol d-HexA2S-GlcNS-HexA2S-GlcNAc-ol d-HexA2S-GlcNS-HexA-GlcNAc6S-ol d-HexA2S-GlcNAc-HexA-GlcNS6S-ol	1028.462	[1,1,2,1,3]	12
d-HexA-GlcNS3S-HexA-GlcNAc6S-ol d-HexA2S-GlcNS3S-HexA-GlcNAc-ol	1014.446	[1,1,2,1,3]	8
d-HexA-GlcNS6S-HexA-GlcNS-ol d-HexA2S-GlcNS-HexA-GlcNS-ol	1000.467	[1,1,2,0,3]	4
d-HexA-GlcNAc6S-HexA-GlcNS-ol d-HexA-GlcNS6S-HexA-GlcNAc-ol d-HexA2S-GlcNS-HexA-GlcNAc-ol	997.448	[1,1,2,1,2]	8
d-HexA-GlcNS3S-HexA-GlcNS-ol	986.451	[1,1,2,0,3]	2
d-HexA-GlcNS3S-HexA-GlcNAc-ol	983.433	[1,1,2,1,2]	2
d-HexA-GlcNS-HexA-GlcNS-ol	969.453	[1,1,2,0,2]	1
d-HexA-GlcNS-HexA-GlcNAc-ol	966.434	[1,1,2,1,1]	2

Automatic sequencing results of the Robo1-bound fraction B sample also gave 13 unique HS tetrasaccharides compositions and the predominant sequences identified with confidence were listed in **Table 5.2**.

Table 5.2. Identified tetrasaccharides sequence from the derivatized HS octasaccharides in the Robo1-bound fraction B sample.

STRUCUTRE	DB_MZ	Composition	Total # of possible sequences
HexA2S-GlcNAc6S-HexA-GlcNAc-ol	1057.469	[0,2,2,1,4]	8
d-HexA2S-GlcNS6S-HexA-GlcNAc6S-ol d-HexA2S-GlcNS6S-HexA2S-GlcNAc-ol d-HexA2S-GlcNAc6S-HexA2S-GlcNS-ol	1059.476	[1,1,2,1,4]	8
d-HexA2S-GlcNS6S-HexA-GlcNAc-ol d-HexA2S-GlcNAc6S-HexA-GlcNS-ol	1028.462	[1,1,2,1,3]	12
d-HexA-GlcNS3S-HexA2S-GlcNAc-ol d-HexA2S-GlcNS3S-HexA-GlcNAc-ol	1014.446	[1,1,2,1,3]	8
d-HexA-GlcNS6S-HexA-GlcNS-ol d-HexA2S-GlcNS-HexA-GlcNS-ol d-HexA-GlcNS-HexA2S-GlcNS-ol	1000.467	[1,1,2,0,3]	4
d-HexA-GlcNAc6S-HexA-GlcNS-ol d-HexA-GlcNS-HexA-GlcNAc6S-ol d-HexA-GlcNS-HexA2S-GlcNAc-ol d-HexA2S-GlcNS-HexA-GlcNAc-ol	997.448	[1,1,2,1,2]	8
d-HexA-GlcNAc6S-HexA-GlcNAc-ol	994.429	[1,1,2,2,1]	
d-HexA-GlcNS3S-HexA-GlcNS-ol	986.451	[1,1,2,0,3]	2
d-HexA-GlcNS3S-HexA-GlcNAc-ol	983.433	[1,1,2,1,2]	2
d-HexA-GlcNS-HexA-GlcNS-ol	969.453	[1,1,2,0,2]	1
d-HexA-GlcNS-HexA-GlcNAc-ol d-HexA-GlcNAc-HexA-GlcNS-ol	966.434	[1,1,2,1,1]	2
d-HexA-GlcNAc-HexA-GlcNAc-ol	963.416	[1,1,2,2,0]	
d-HexA-GlcNAc6S-HexA-GlcN-ol d-HexA2S-GlcNAc-HexA-GlcN-ol	958.468	[1,1,2,1,1]	8

An attempt was made to perform complete sequencing for each composition, by aligning extracted ion chromatogram (EIC) of sequential glycosidic bond cleavage product ions as described in **Chapter 4** for analysis of HS dp4. However, there are several factors in this particular case prevent us from getting complete structural sequencing. First of all, the data presented here was performed in a different LC-MS/MS condition, prior to the work presented in

Chapter 4 where the optimized condition for complete sequencing was established. As we discussed in previous chapters, base line separation or at least partial separation is critical for sequencing isomeric structures, as well as selecting appropriate precursor ions for fragmentation. In this work, a self-packed C18 tip was used instead of a 50cm core-shell C18 column, and only $[M+Na]^+$ precursor ion was selected for fragmentation whereas a combination of $[M+2H]^{2+}$ and $[M+Na]^+$ was believed to be required for complete sequencing. Because of the time-consuming preparation and limited amount available for the Robo1-bound HS octasaccharides, there was no material left for us to perform further analysis when this draft was composed. Second, the reduction step was skipped in this work to minimize sample losses, which led to two consequences: peak splitting caused by the anomeric configurations at the reducing end and undistinguishable isomeric species with saturated non-reducing end due to symmetric structures in terms of mass. Both became especially problematic, as presented in this work, when complex mixtures were involved. Reduction appeared to be a necessary step to simplify LC chromatogram as well as to allow differentiation of certain isomeric pairs. Third, although there were several 3-O-sulfation containing sequences identified as listed in both **Table 5.1** and **Table 5.2**, study of standards with 3-O-sulfation for better understanding of the fragmentation patterns of these sequences would be desired to provide higher confident for characterizing these rare species.

While the original plan was to identify HS octasaccharides in the Robo1-bound fraction, only tetrasaccharides and disaccharides were detected after sequential derivatizations due to the β–elimination. Considering that we were able to sequence HS dp4 (20~50 μg starting material was used) as described in the previous chapter, increasing the starting material through scaling up affinity purification protocols might provide a better chance to detect derivatized

octasaccharides. Initial trials on an unfractionated HS dp8 library showed that 50 μg of starting material may be sufficient, while further efforts would be needed to find out the required minimum amount. On the other hand, it has been believed that the permethylation conditions, particularly the base concentration and amount, plays important role in β–elimination[26]. Efforts have been made to optimize the conditions by trying different base concentrations, base to oligosaccharides ratio, as well as solid phase permethylation[27]. Unfortunately, there was no avoiding β–elimination without affecting the permethylation efficiency. To totally avoid such degradation, carboxylic acid reduction[24] may be the only choice, which converts the active uronic acid residue into much more stable hexose.

Although we are not able to direct sequence HS octasaccharides, the identified tetrasaccharides sequences are still useful for structural characterization. For the Robo1-bound fraction A sample (**Figure 5.3 and Table 5.1**), the two octasaccharide compositions detected only in Robo1-bound fractions could be a combination of dp4 compositions [1,1,2,1,3] and [1,1,2,0,2] or [1,1,2,1,2] and [1,1,2,0,3] for dp8 composition [1,3,4,1,5]; and dp4 compositions [1,1,2,1,3] and [1,1,2,0,3] for dp8 composition [1,3,4,1,6]. For the Robo1-bound fraction B sample (**Figure 5.4 and Table 5.2**), the combination of dp4 compositions [1,1,2,1,4] and [1,1,2,0,3] could comprise dp8 composition [1,3,4,1,7]; and dp4 compositions [1,1,2,1,2] and [1,1,2,1,4] or two dp4 composition [1,1,2,1,3] could comprise dp8 composition [1,3,4,2,6]. While these are not as straightforward as octasaccharides sequence, it provides information to narrow down the possible sequences from hundreds to tens. By performing the LC-MS/MS analysis using optimized conditions, more useful structural information can be obtained, such as the epimerization on uronic acid residues, to allow further interpretations. Current biosynthesis knowledge may also be applied to help the characterizing, for example, IdoA(±2S) only attaches

to the reducing end of N-sulfated GlcN species (regardless of their O-sulfation modifications) and 3-O-sulfation is usually found as GlcA-GlcNS(3S±6S), GlcA/IdoA(2S)-GlcNS3S and Ido(2S)-GlcN(3S±6S)[28,29]. Because the GlcN(3S±6S) residue is even more rare compare to GlcNS(3S±6S) species, it is not included in our current GAG-ID database but could be considered for future versions.

We propose a possible octasaccharide sequence candidate for the dp8 composition [1,3,4,1,7], notable for its high abundance among the four compositions identified as specifically binding to Robo1 (**Figure 5.4**). The tetrasaccharide sequence d-HexA2S-GlcNS6S-HexA-GlcNAc6S-ol and d-HexA-GlcNS6S-HexA-GlcNS-ol were chosen from dp4 composition [1,1,2,1,4] and [1,1,2,0,3] respectively, which could form two octasaccharides sequence: d-HexA-GlcNS6S-HexA-GlcNS-HexA2S-GlcNS6S-HexA-GlcNAc6S-ol and d-HexA2S-GlcNS6S-HexA-GlcNAc6S-HexA-GlcNS6S-HexA-GlcNS-ol. Because the heparinase III we used to generate HS dp8 library has predominant specificity for GlcNAc/GlcNS±6Sα1–4IdoAα/GlcAβ[30], the HexA2S is less likely to appear at the non-reducing end. The internal HexA residue for the two tetrasaccharides were both assigned as IdoA due to the observed observed Z_2/Y_2 ratio of approximately 0.3 (**Figure 5.10**), according to the results we had for HS dp4 analysis described in **Chapter 4** (**Figure 4.9**). With the assumption that IdoA is more likely to have 2-O-sulfation than GlcA, we assign IdoA2S to the HexA2S residue. With that, we provide a potential candidate as Robo1-bound HS dp8 sequence: d-HexA-GlcNS6S-IdoA-GlcNS-IdoA2S-GlcNS6S-IdoA-GlcNAc6S, which is currently synthesized in Geert-Jan Boons group for further characterization and study of its affinity binding with Robo1. Because the synthetic sequence would have saturated non-reducing end, we suggest that the non-reducing end residue to be GlcA as the heparinase III favors GlcA more than IdoA.

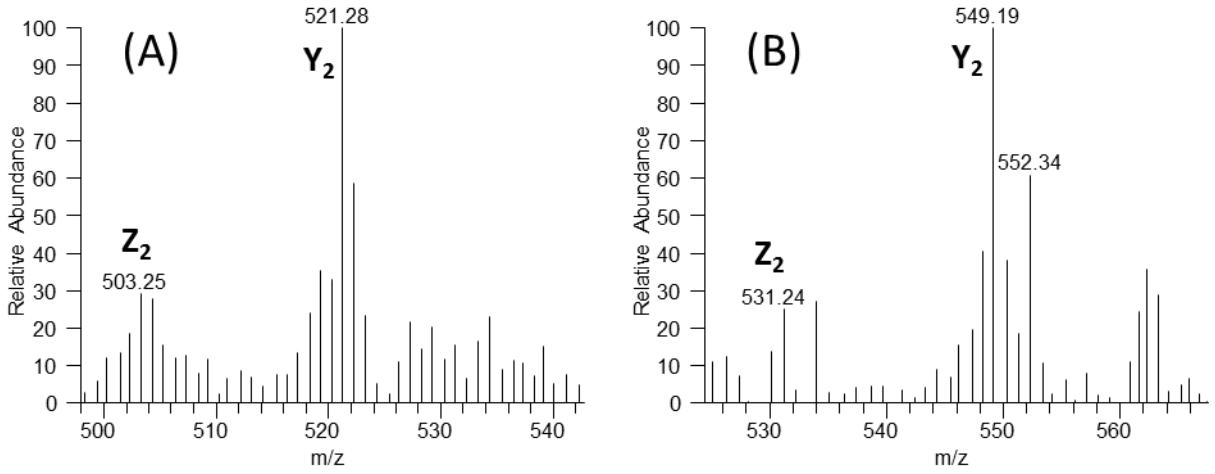


Figure 5.10. Zoom-in MS/MS spectra for parent ion $[M+Na]^+$ with m/z of 1000.467 for dp4 [1,1,2,0,3] (A) and m/z of 1059.476 for dp4 [1,1,2,1,4] (B). The observed Z_2/Y_2 ratio was approximately 0.3 for both spectrums, indicating an internal IdoA instead of GlcA.

Conclusions

The work presented here represents the entire workflow of our initial proposal for study the structural specificity of HS responsible for interactions with Robo1 protein, which includes affinity purification of Robo1-bound HS oligosaccharides, HRMS for composition analysis, chemical derivatizations coupled with LC-MS/MS and semi-automatic sequencing for complete structural characterizing. While complete sequencing of affinity-purified octasaccharides has not been accomplished in the very few trials we tested, lessons have been learned and continual efforts are being made in order to meet the challenges. For sample preparation, an alternative protocol may be used to scale up and accelerate affinity purification procedure. For the chemical derivatization protocol, reducing end reduction needs to be included whereas carboxylic acid reduction may be worthwhile to optimize and add in order to avoid β–elimination; alternative acetylation reagents may be used to increase LC resolution between original N-sulfated and N-

acetylated residues. For LC-MS/MS analysis, the optimized condition used for HS dp4 analysis described in **Chapter 4** should be used in future for characterization of partial tetrasaccharide sequences generated from β–elimination and as a starting point for further optimization for octasaccharide analysis. Data analysis is apparently a great challenge as well, especially for larger oligosaccharides, as we know the number of theoretical sequences grows exponentially as the size increases. The GAG-ID program has been developed into a very useful tool to narrow down thousands of MS^2 spectrums to hundreds and provide automatic sequencing together with valuable information to facilitate manual confirmation of each identified sequence. While only glycosidic bond cleavage is considered, further efforts on finding and adding informative cross-ring cleavage product ions may lead to more confident sequencing.

Acknowledgements

This work was collaborated with Yulun Chiu, who wrote the program GAG-ID. We thank Heather A. Moniz for providing the Robo1 protein and Robo1-immoblized resin for affinity purification. We thank Eduard Condac for performing the affinity purification experiments and analyzing the HRMS data (collected by John K. Muchena) and HILIC LC-MS data (collected by Mayumi Ishihara). The author also acknowledge Lance Wells lab for their instrumentation time. This research is supported in part by the National Institute of General Medical Sciences-funded "Research Resource for Integrated Glycotechnology" (P41 GM103390) from the National Institutes of Health.

References:

1. Brose, K.; Bland, K. S.; Wang, K. H.; Arnott, D.; Henzel, W.; Goodman, C. S.; Tessier-Lavigne, M.; Kidd, T. *Cell* **1999**, *96*, 795-806.

2. Park, K. W.; Morrison, C. M.; Sorensen, L. K.; Jones, C. A.; Rao, Y.; Chien, C. B.; Wu, J. Y.; Urness, L. D.; Li, D. Y. *Developmental biology* **2003**, *261*, 251-267.

3. Dickson, B. J.; Gilestro, G. F. *Annual review of cell and developmental biology* **2006**, *22*, 651-675.

4. Li, H. S.; Chen, J. H.; Wu, W.; Fagaly, T.; Zhou, L.; Yuan, W.; Dupuis, S.; Jiang, Z. H.; Nash, W.; Gick, C.; Ornitz, D. M.; Wu, J. Y.; Rao, Y. *Cell* **1999**, *96*, 807-818.

5. Kidd, T.; Bland, K. S.; Goodman, C. S. *Cell* **1999**, *96*, 785-794.

6. Dickson, B. J. *Science* **2002**, *298*, 1959-1964.

7. Xian, J.; Clark, K. J.; Fordham, R.; Pannell, R.; Rabbitts, T. H.; Rabbitts, P. H. *Proc Natl Acad Sci U S A* **2001**, *98*, 15062-15066.

8. Liu, J.; Zhang, L.; Wang, D.; Shen, H.; Jiang, M.; Mei, P.; Hayden, P. S.; Sedor, J. R.; Hu, H. *Mechanisms of development* **2003**, *120*, 1059-1070.

9. Grieshammer, U.; Le, M.; Plump, A. S.; Wang, F.; Tessier-Lavigne, M.; Martin, G. R. *Developmental cell* **2004**, *6*, 709-717.

10. Wang, L. J.; Zhao, Y.; Han, B.; Ma, Y. G.; Zhang, J.; Yang, D. M.; Mao, J. W.; Tang, F. T.; Li, W. D.; Yang, Y.; Wang, R.; Geng, J. G. *Cancer science* **2008**, *99*, 510-517.

11. Prasad, A.; Fernandis, A. Z.; Rao, Y.; Ganju, R. K. *J Biol Chem* **2004**, *279*, 9115-9124.

12. Bedell, V. M.; Yeo, S. Y.; Park, K. W.; Chung, J.; Seth, P.; Shivalingappa, V.; Zhao, J.; Obara, T.; Sukhatme, V. P.; Drummond, I. A.; Li, D. Y.; Ramchandran, R. *Proc Natl Acad Sci U S A* **2005**, *102*, 6373-6378.

13. Hussain, S. A.; Piper, M.; Fukuhara, N.; Strochlic, L.; Cho, G.; Howitt, J. A.; Ahmed, Y.; Powell, A. K.; Turnbull, J. E.; Holt, C. E.; Hohenester, E. *J Biol Chem* **2006**, *281*, 39693-39698.

14. Fukuhara, N.; Howitt, J. A.; Hussain, S. A.; Hohenester, E. *J Biol Chem* **2008**, *283*, 16226-16234.

15. Varki, A.; Cummings, R.; Esko, J. D.; Freeze, H.; Hart, G. W.; Marth, J. *Essentials of Glycobiology*; Cold Spring Harbor Laboratories Press: New York, 1999.

16. Carlsson, P.; Kjellen, L. *Handbook of experimental pharmacology* **2012**, 23-41.

17. Multhaupt, H. A.; Couchman, J. R. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **2012**, *60*, 908-915.

18. Yu, Y.; Sweeney, M. D.; Saad, O. M.; Crown, S. E.; Hsu, A. R.; Handel, T. M.; Leary, J. A. *J Biol Chem* **2005**, *280*, 32200-32208.

19. Loo, B. M.; Kreuger, J.; Jalkanen, M.; Lindahl, U.; Salmivirta, M. *J Biol Chem* **2001**, *276*, 16868-16876.

20. Zhang, F.; Moniz, H. A.; Walcott, B.; Moremen, K. W.; Linhardt, R. J.; Wang, L. *Biochimie* **2013**, *95*, 2345-2353.

21. Schenauer, M. R.; Yu, Y.; Sweeney, M. D.; Leary, J. A. *J Biol Chem* **2007**, *282*, 25182-25188.

22. Naimy, H.; Leymarie, N.; Zaia, J. *Biochemistry-Us* **2010**, *49*, 3743-3752.

23. Maxwell, E.; Tan, Y.; Hu, H.; Benson, G.; Aizikov, K.; Conley, S.; Staples, G. O.; Slysz, G. W.; Smith, R. D.; Zaia, J. *PloS one* **2012**, *7*, e45474.

24. Barker, S. A.; Hurst, R. E.; Settine, J.; Fish, F. P.; Settine, R. L. *Carbohydr Res* **1984**, *125*, 291-300.

25. Domon, B.; Costello, C. E. *Biochemistry-Us* **1988**, *27*, 1534-1543.

26. Ciucanu, I. *Anal Chim Acta* **2006**, *576*, 147-155.

27. Lei, M.; Mechref, Y.; Novotny, M. V. *J Am Soc Mass Spectrom* **2009**, *20*, 1660-1671.

28. Xu, D.; Tiwari, V.; Xia, G.; Clement, C.; Shukla, D.; Liu, J. *Biochem J* **2005**, *385*, 451-459.

29. Xia, G.; Chen, J.; Tiwari, V.; Ju, W.; Li, J. P.; Malmstrom, A.; Shukla, D.; Liu, J. *J Biol Chem* **2002**, *277*, 37912-37919.

30. Desai, U. R.; Wang, H. M.; Linhardt, R. J. *Arch Biochem Biophys* **1993**, *306*, 461-468.

CHAPTER 6

CONCLUSIONS

Chemical derivatization coupled with liquid chromatography and tandem mass spectrometry (LC-MS/MS) analysis has been performed for two major classes of sulfated glycosaminoglycans (GAGs): chondroitin sulfate (CS) and heparan sulfate (HS)/heparin. The present work demonstrates that chemical derivatization allows the use of standard reversed phase LC (RPLC) for on-line separation and conventional collision induced dissociation (CID) for structural characterization of isomeric sulfated GAGs.

The basic idea of sequential chemical derivatizations is to replace the labile sulfate groups with relatively more stable acetyl groups after protecting other unsulfated sites. The original hydrophilic, negatively charged sulfated GAGs are converted into hydrophobic, neutral products, which allows their separation by RPLC and ionization on positive ion mode. In addition, the original sulfation sites are indicated by acetylation sites that can be determined by CID MS/MS without worrying the sulfate loss during the fragmentation.

CS and HS are the two most abundant sulfated GAGs that can be found in the form of proteoglycans expressed on the cell surface or secreted in the extracellular matrix, and participate in a variety of biological activities. Initial method development was performed on CS oligosaccharides, and an isomeric CS hexasaccharides mixture derived from enzymatic digestion of intact CS was successfully characterized after derivatization by using a combination of RPLC and multi-level CID MS/MS. Extension of the chemical derivatization protocol to HS/heparin analysis required the use of hexadeutero labeled acetic anhydride in order to differentiate original N-acetylation and N-sulfation residues. The analysis of a series of HS disaccharides standards as well as synthetic larger HS oligosaccharides demonstrated that the chemical derivatization strategy enabled the use of MS^2 alone to differentiate all common sulfation positional isomers and possibly for those isomers with rare 3-O-sulfation modifications as well.

While HS shares a lot common structural properties with CS, they differ in the positions of O-sulfation, and HS has variants at N position of glucosamine residues as well as epimerization at C5 position of uronic acid residues that CS lacks. Thus, the structural characterization of HS oligosaccharides raises its own challenges compared to the analysis of CS oligosaccharides. Due to the more diverse modifications, oligosaccharides derived from enzymatic digestion of intact HS are likely to be more heterogeneous mixtures, making the separation and characterization of such mixtures even more challenging. While it is very beneficial to use standards with defined structure for method development, the heterogeneous nature of HS makes it difficult and laborious to enrich pure HS oligosaccharides (larger than disaccharides) in a single isoform with sufficient quality and quantity. Thus, synthetic HS oligosaccharides standards are very valuable by providing defined, customable structures in relative high amounts (up to mg). For example, the availability of a library of synthetic HS tetrasaccharides varying in sulfation density and/or positions as well as epimerizations has allowed us to systematically study the difference in chromatographic behavior and fragmentation patters between various sulfation positional isomers as well as epimers. With that, complete structural sequencing of a HS tetrasaccharide mixture derived from natural sources has been achieved for the first time, showing the capability of our methods to separate and characterize of complex mixtures of HS oligosaccharides.

The difficulties presented in the project for structural sequencing of Robo1-bound HS octasaccharides exposed certain limitations of our current protocol for analysis of larger oligosaccharides mixtures with limited amount available and forced us to seek further improvements. First of all, the yield of the permethylation steps may be increased by avoiding β–elimination through carboxylic acid reduction. It may be beneficial to develop a protocol for

quantitative carboxylic acid reduction as a replacement of the regular reduction, since the reduction of reducing end residues has been proved necessary to simplify LC separation and avoid ambitious fragmentation patterns. Secondly, both optimized LC conditions and appropriate precursor ions are critical for complete structural sequencing of HS oligosaccharides by providing maximum separation and informative fragmentations for isomeric structures. The use of well-calibrated and high resolution MS (HRMS) for full mass scans or even MS/MS scans if possible is also very helpful for accurate assignment of precursor and product ions. Thirdly, the use of GFP-tagged Robo1 protein and limited availability of HS has made it very laborious and time-consuming to get sufficient amount of Robo1-bound HS octasaccharides. The use of commercially available heparin ocatasaccharides and untagged Robo1 with affinity trapping assays may be an alternative option to facilitate the sample preparation in future experiments.

With the assistance of our sequencing program, GAG-ID, data analysis becomes more efficient as manual interpretation alone can be very time-consuming when it comes to complex unknown mixtures. This is because that not only the number of possible structures for a given size oligosaccharides can be hundreds and even tens of thousands, but also there can be tens of possible isomeric structures with the exact same mass and more than 50% similarity. GAG-ID can efficiently screen the correct precursor ions when HRMS is used for full mass scan. It is also helpful to have scores used to evaluate automatic fragmentation matching for MS^2 spectrums against each possible isomeric structure, although manual interpretation is still required to confirm each identified sequence. In addition, an extracted ion chromatogram of diagnostic product ions is very useful to assist manual identification of partially separated or even co-eluted isomeric structures, since the extracted MS^2 spectrum alone may present as a mixture making the interpretation even more complicate.

The chemical derivatization methods have been applied for structural sequencing of almost all common sulfation patterns of both CS and HS in the present work, with limited practice for those with rare sulfation modifications (*e.g.* 3-O-sulfation for HS). It would be worthwhile to study these rare species if there is any standard available in future, as they are likely to be more biological relevant and therefore more important. Besides CS and HS/heparin, the chemical derivatization method should also be extendable to other sulfated GAGs, such as dermatan sulfate (DS) and keratan sulfate (KS). The structural analysis of DS would be similar to that of CS and HS, since it shares the same back bone structure and similar sulfation patterns with CS, and has epimerization on the uronic acid residues just like HS. It may benefit even more from the chemical derivatization method for structural analysis of KS, because β–elimination would not be a problem anymore owe to the lack of uronic acid residues for KS.

With an increasing number of GAG-binding proteins being identified, analytical methods to study the structural specificity of GAGs required for their interactions with proteins are desired for better understanding their structure-function relationships. The present work on CS and HS/heparin demonstrated that the chemical derivatization strategy is a useful tool for detailed structural characterization of potentially all sulfated GAGs. Continued work on improvement of current methods for biologically relevant CS or HS/heparin analysis and extension to other sulfated GAGs is of great importance to offer a valuable analytical tool for GAG oligosaccharides characterization where detailed structural information is required to determine biological functions.