STRUCTURAL CHARACTERIZATION OF OLIGOSACCHARIDES AND UNDERSTANDING HEPARAN SULFATE-PROTEIN INTERACTIONS

Po-Hung Hsieh

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Pharmaceutical Sciences (Chemical Biology and Medicinal Chemistry).

Chapel Hill
2015

Approved by:

Albert Bowers

Jian Liu

Kuo-Hsiung Lee

Lars Pedersen

Michael Jarstfer

ABSTRACT

PO-HUNG HSIEH: Structural characterization of oligosaccharides and understanding heparan sulfate-protein interactions
(Under the direction of Jian Liu)

Heparin is known as a widely used anticoagulant drug since the 1930s. Heparin and heparan sulfate (HS) are highly sulfated polysaccharides that consist of a repeating disaccharide unit of glucosamine and glucuronic (GlcA) or iduronic acid (IdoA). The polydispersity of the saccharide sequence as well as the polyheterogeneity of the sulfation pattern make the study of this type of carbohydrate a formidable task. Instead of traditional extraction from animal mucus, we use the chemoenzymatic approach to synthesize homogeneous heparin oligosaccharides. The synthetic low molecular weight heparin was proven anti-FXa activity with reversible anticoagulant activity. In addition to secure the supply chain, the extra benefits aid future clinical usage of synthetic heparin drug. On the other side, the availability of synthetic pure common and less abundant HS oligosaccharides can serve as heparin component standards to secure the drug safety.

Take the advantage of chemoenzymatic synthesis, the sufficient amount of structurally tailored oligosaccharide reagents are accessible. To investigate the interaction between HS and proteins, several oligosaccharide probes were successfully developed. Firstly, a synthesized heptasaccharide was used to co-crystalize with 2-*O*-sulfotransferase (2-OST) and probe the molecular basis of specificity. 2-OST was found to recognize *N*-sulfo and exclude 6-*O*-sulfo group of substrates, supporting the biosynthetic hypothesis 2-*O*-sulfation occurs after *N*-sulfation

and prior to 6-*O*-sulfation. Secondly, we designed and synthesized the active heparan sulfate oligosaccharide probe carrying a diazoacetyl group. The resultant oligosaccharides demonstrate inhibitory activity and structural selectivity toward HS biosynthetic enzyme 2-OST and 3-OST, respectively. Thirdly, ^{13}C-labeled saccharide was introduced at the desired site of the oligosaccharide to enhance the intensity of NMR signals. The labeled reagents facilitate directly analysis the binding between specific saccharide and protein. Finally, nuclear magnetic resonance analyses were performed to obtain coupling constants and full chemical shift assignments of all synthetic oligosaccharides. Comparing to relatively rigid GlcA, IdoA shows structural plasticity and diverse conformational preference in response to different sulfation patterns, which may be critical for various protein specific binding. The success of these projects assists the fundamental understand of HS-protein interaction as well as the design of next generation heparin-based therapeutics.

To my Dad and Mom, Ming-Hua Hsieh and Hui-Ying Huang, for guiding me through all of the struggles and successes of graduate career.

ACKNOWLEDGEMENTS

With my sincerest gratitude, I would like to thank several people in my graduate career. I am very fortunate to meet all of you in the school and in my life. Without your altruistic and supportive help, these works cannot be completed.

First I would like to thank the dissertation committee and collaborators for the generous and intelligent inputs on these research projects. Thanks to Dr. K. H. Lee, who leads me study abroad and constantly reminds me the big picture in life. Thanks to Dr. Michael Jarstfer, who always patiently guides me when I have any academic and course question. Thanks to Dr. Albert Bowers's valuable suggestions in the advisory and committee meeting. Thanks to Dr. Lars Pedersen, it was unforgettable great experience discussing conformation study with you in New Orleans. Thanks to Dr. David A Keire, who helps to characterize the structure of less abundant heparin in FDA affiliated project. Thanks to Dr. Marco Guerrini, who assists in the NMR study of isotopically labeled oligosaccharide. Thanks to Dr. Robert Woods and David Thieker for aiding in computational study and accommodating me at Complex Carbohydrate Research Center. It is my great pleasure to have the opportunity discuss and work with you.

Next I am deeply thankful the people in the glycobiology laboratory including former members Dr. Vijayakanth Pagadala, Dr. Ryan Bullis, Dr. Kasemsiri Chandarajoti, Dr. Chunhui Liu, Dr. Wenfang Dou, Dr. Wen Zhou as well as current labmates Susan Woody PharmD, Dr. Jianhong Yang, Truong Pham, Tim O'Leary, Katelyn Arnold, Juwong Kwon, and most importantly, Dr. Yongmei Xu. You not only provide the assistance on these works, but also

warm my heart when I struggle in the experiments. I also want to appreciate the people in UNC badminton team and TSA badminton club. Thank you for letting me always has relaxed mind and healthy body for the next day working. Especially thank my friend Jie-Yu Liu, who builds the club and plays the tournaments with me, is the one always being there when I need help.

Finally, I would like sincere thank my advisor, Dr. Jian Liu. He is an exceptional mentor for my graduate career. I greatly appreciate his enthusiasm, inspiring thinking, and erudition in glycobiology. He always pushes me critical thinking, but at the same time patiently guides me to the possible solutions even he is under the extremely tight schedule. As an international student not familiar with the English writing and speaking at early time, I will never forget the moment- he seats side by side and teaches me how to prepare first paper as well as the slides of my first seminar. Any academic successes in my future will largely ascribe to his generously guidance.

TABLE OF CONTENTS

LIST OF TABLES xiii

LIST OF FIGURES xiv

LIST OF ABBREVIATIONS xviii

INTRODUCTION 1

PROTEOGLYCANS 1

Chemical Structure of Heparin 3

Chemical Structure of Heparan Sulfate 4

BIOSYNTHESIS OF HEPARAN SULFATE 5

Biosynthesis Scheme 5

Glucosaminyl Deacetylation/N-sulfation 10

C_5-epimerization 10

2-O-sulfation 11

6-O-sulfation 13

3-O-sulfation 13

HEPARIN PHARMACEUTICAL PRODUCTION 15

Traditional Extraction and Depolymerization 16

Heparin Safety Crisis 16

Chemical Synthesis of Heparin 17

Chemoenzymatic Synthesis 18

STRUCTURAL ANALYSIS OF HEPARAN SULFATE 19

Enzymatic or Chemical Degradation for Disaccharide and Mass Analysis.....19

Nuclear Magnetic Resonance (NMR) Measurement.....23

Molecular Dynamics (MD) Simulation.....24

BIOLOGICAL FUNCTIONS OF HEPARAN SULFATE.....25

Anticoagulation.....26

Cell Proliferation.....30

Tumor Progression.....32

Inflammation.....35

Viral Infection.....36

STATEMENT OF PROBLEM.....39

MATERIALS AND METHODS.....41

PREPARATION OF HEPARAN SULFATE BIOSYNTHETIC COFACTORS.....41

Saccharide Donor.....41

CHEMOENZYMATIC SYNTHESIS OF HEPARAN SULFATE OLIGOSACCHARIDES.....42

KfiA and pmHS$_2$ for Backbone Elongation.....42

GlcNTFA Deacetylation and Subsequent Modification with NST.....42

C$_5$-Epi and 2-OST Modification.....43

6-OST-1, 6-OST-3 and 3-OST Modification.....43

Preparation of ^{35}S-Labeled Heparan Sulfate.....44

PURIFICATION OF SYNTHETIC HEPARAN SULFATE.....45

C$_{18}$ and Q Sepharose Fast Flow™-HPLC Purification and Analysis.....45

HPLC Analysis.....45

STRUCTURAL CHARACTERIZATION OF SYNTHETIC HEPARAN SULFATE.....46

NMR Parameter and Spectroscopy Analysis 46

MS Spectrometry Analysis 47

Disaccharide Analysis 49

MOLECULAR MODELING 49

Oligosaccharide Pre-processing 49

Energy Minimization 49

Molecular Dynamics Simulation 51

Simulation Data Analysis 51

BIOLOGICAL ACTIVITIES OF SYNTHETIC HEPARAN SULFATE 52

Activity Measurement for 2-O-Sulfotransferase 52

Activity Measurement for 3-O-Sulfotransferase 53

In vitro and ex vivo Anti-FXa Activity 54

Binding Affinity to Antithrombin 54

SAFEGUARD HEPARIN SAFETY: REVERSIBLE ANTICOAGULANT ACTIVITY & A LIBRARY OF STANDARDS 56

SYNTHESIS OF LOW MOLECULAR WEIGHT HEPARINS (LMWH) 56

Introduction 56

Targeted Structures and Schematic Synthesis of LMWHs 58

ANTICOAGULANT ACTIVITY AND PROTAMINE REVERSIBILITY OF SYNTHETIC LMWH 61

SYNTHESIS OF LESS ABUNDANT HEPARAN SULFATE OLIGOSACCHARIDE LIBRARY 65

Introduction 65

Chemoenzymatic Synthesis of GlcA2S- and IdoA2S-Containing Hexasaccharides 67

STRUCTURAL CHARACTERIZATION OF 2-O-SULFATED GLUCURONIC ACID (GLCA2S)- AND IDURONIC ACID (IDOA2S)- CONTAINING OLIGOSACCHARIDES 72

NMR Analysis of Hexasaccharides 72

Characterization of GlcA2S-Containing Oligosaccharides 72

Intramolecular effects by N-sulfation, 6-O-sulfation, and 3-O-sulfation 81

ANTITHROMBIN BINDING AND ANTI-XA ACTIVITY OF LESS ABUNDANT HEPARAN SULFATE 83

Contribution of GlcA2S and IdoA2S to the Binding Affinity to Antithrombin 83

DISCUSSION 85

CONCLUSIONS 87

UTILIZE HEPARAN SULFATE BASED PROBE TO INVESTIGATE INTERACTION WITH PROTEINS 89

SUBSTRATE SPECIFICITY REVEALED BY CRYSTAL STRUCTURE OF 2-*O*-SULFOTRANSFERASE 89

Introduction 89

NMR Analysis of Heptasaccharide Substrate 90

Crystal Structure of 2OST with Bound Heptasaccharide 91

SYNTHESIS OF HEPARAN SULFATE OLIGOSACCHARIDES CARRYING DIAZOACETYL GROUP 95

Introduction 95

Synthesis and Characterization of HS-Based Probe and Hydrolysis Product 96

Using GlcNDaz-Containing Oligosaccharides as Active HS Based Probes (AHSBP) 101

Inhibitory Effect on 2-O-Sulfotransferase and 3-O-Sulfotransferase 103

SYNTHESIS OF ^{13}C-ISOTOPE LABELED HEPARAN SULFATE OLIGOSACCHARIDES 105

Introduction.....105

Chemoenzymatic Synthesis of ^{13}C-labeled Cofactor and Oligosaccharide.....106

Determination of the Binding of Labeled Hexasaccharide to AT.....110

Investigation of the Conformation of Heparan Sulfate.....113

CONCLUSIONS.....116

INVESTIGATE THE EFFECT OF DIFFERENT SULFATION ON HEPARAN SULFATE CONFORMATION.....119

SYNTHESIS OF HEPARAN SULFATE WITH DIFFERENT SULFATION PATTERN.....119

Introduction.....119

Chemoenzymatic Synthesis of Oligosaccharide.....122

Structural Characterization.....124

CONFORMATION ANALYSIS AND MOLECULAR DYNAMIC SIMULATION.....133

$^{3}J_{H\text{-}H}$ Coupling Constant Measurement.....134

Shift of IdoA Conformation Population.....137

NOE Intensity Measurement.....139

Possible Mechanism for Shift of Conformation Equilibrium.....141

Synthesis of Large Heparan Sulfate with Multiple IdoA/2S Saccharides.....145

Discussion.....148

CONCLUSIONS.....148

APPENDIX I. SUPPLEMENTARY HPLC, MASS, AND NMR DATA.....151

APPENDIX II. CURRICULUM VITAE.....206

REFERENCES.....209

LIST OF TABLES

Table 1 - Heparin drugs and pharmacokinetic properties 16

Table 2 - Internal restraints applied for iduronic acid during MD simulation 50

Table 3 - Summary of the properties of synthetic LMWHs 61

Table 4 - $^{1}H–^{13}C$ NMR chemical shift assignments (in ppm) of Constructs 1–4 76

Table 5 - $^{1}H–^{13}C$ NMR chemical shift assignments (in ppm) of Constructs 5–8 77

Table 6 - ^{1}H NMR chemical shift difference ($\Delta\delta$). 81

Table 7 - List of ^{13}C-labeled hexasaccharides prepared in the pilot synthesis 110

Table 8 - ^{1}H-^{13}C NMR chemical shift assignments (in ppm.) of construct 1 – 4 132

Table 9 - ^{1}H-^{13}C NMR chemical shift assignments (in ppm) of construct 5 – 8 132

Table 10 - Experimental and theoretical $^{3}J_{H\text{-}H}$ couplings of IdoA/2S 135

Table 11 - Averaged theoretical dihedral angle and $^{3}J_{H\text{-}H}$ value for each conformation of IdoA 136

Table 12 - Experimental and theoretical NOE H2-H5/H4-H5 ratio of IdoA/2S among synthetic constructs 141

Table 13 - The occupancy of possible hydrogen bond during 1us MD simulation 143

Table 14 - List of multi-IdoA-containing oligosaccharide prepared in the pilot synthesis 147

LIST OF FIGURES

Figure 1 - Representation of heparan sulfate proteoglycans 2

Figure 2 - Structure of main type of glycosaminoglycans (GAGs) 3

Figure 3 - Chemical structure difference between heparin and heparan sulfate 5

Figure 4 - Glycosaminoglycan tetrasaccharide linkage to core protein 6

Figure 5 - Polymerization of the heparan sulfate backbone 8

Figure 6 - Heparan sulfate backbone modifications 10

Figure 7 - *N*-deacetylation and *N*-sulfation by NDSTs 10

Figure 8 - C5-epimerization by C5-epimerase 11

Figure 9 - 2-*O*-sulfation by 2-OST 13

Figure 10 - 6-*O*-sulfation by 6-OSTs 13

Figure 11 - 3-*O*-sulfation by 3-OSTs 15

Figure 12 - Anticoagulant pentasaccharide structure 18

Figure 13 - Chemoenzymatic synthesis of oligosaccharide 19

Figure 14 - Enzymatic degradation of heparan sulfate by heparin lyase 21

Figure 15 - Chemical degradation of heparan sulfate by nitrous acid 22

Figure 16 - Structural comparison between over sulfated chondroitin sulfate and heparin 24

Figure 17 - Blood coagulation cascade 27

Figure 18 - Anticoagulation mechanism that heparin activates antithrombin to inhibit Xa 28

Figure 19 - Key structural components of pentasaccharide for AT binding 29

Figure 20 - Anticoagulation mechanism that heparin activates antithrombin to inhibit IIa 30

Figure 21 - Minimum heparan sulfate structure for FGFs binding 32

Figure 22 - Role of HS and heparin in mediating of tumor progression 34

Figure 23 - Role of HS and heparin in mediating of inflammation 36

Figure 24 - Chemical structure of HSV-I binding and entry receptor as well as inhibitor.....38

Figure 25 - Targeted structures and schematic synthesis of synthetic LMWHs.....58

Figure 26 - Chemoenzymatic synthetic schemes of Compound 1 and Compound 6 59

Figure 27 - Determination of the anti-FXa activity and sensitivity to protamine neutralization of synthetic LMWHs.....62

Figure 28 - The synthetic hexasaccharide constructs 1 to 8 67

Figure 29 - Chemoenzymatic synthesis of GlcA2S- and IdoA2S-containing hexasaccharides.....68

Figure 30 - Two major products of epimerization and 2-*O*-sulfation.....69

Figure 31 - Mass and HPLC analysis of construct 1 to 4 70

Figure 32 - Mass and HPLC analysis of construct 5 to 8 71

Figure 33 - ^{1}H-NMR spectra of construct 1, 2, 3 and 4.....73

Figure 34 - ^{1}H-NMR spectra of construct 5, 6, 7 and 8.....74

Figure 35 - 2D ^{1}H-^{13}C HSQC-TOCSY NMR spectrum of construct 3.....75

Figure 36 - ^{1}H-^{13}C HMBC key correlations of construct 3 78

Figure 37 - 2D ^{1}H-^{13}C HSQC NMR spectrum of GlcA2S- and IdoA2S-containing oligosaccharide mixture.....86

Figure 38 - Crystal structure of the 2-OST trimer 90

Figure 39 - Refinement of different conformations of the acceptor saccharide 93

Figure 40 - Modeling of 6-*O*-sulfates on the substrate heptasaccharide.....94

Figure 41 - Structures of the synthesized AHSBP.....96

Figure 42 - Synthesis scheme of active heparan sulfate based probes.....97

Figure 43 - Two potential pathways for activation of the GlcNDaz residue 99

Figure 44 - $^{1}H–^{13}C$ HMBC key correlations of compound 11.....100

Figure 45 - Analysis of the conjugate product of 2-OST and compound 1.....102

Figure 46 - Inhibitory effects of compounds 1 and 2 on 2-OST and 3-OST.....104

Figure 47 - Synthesis of [^{13}C]-UDP-Glucuronic Acid.....107

Figure 48 - Synthesis of ^{13}C-labeled oligosaccharide.....107

Figure 49 - HPLC and HSQC spectra of hexasaccharide A.....108

Figure 50 - HPLC and HSQC spectra of hexasaccharide B.....109

Figure 51 -2D-HSQC NMR spectrum of HS/hexasaccharide A mixture.....111

Figure 52 - The proton-decoupled ^{13}C-NMR.....112

Figure 53 - The anticipated inter-residue glycosidic $^{3}J_{C-H}$ coupling detected by HMBC.....114

Figure 54 - HSQC and J-HMBC spectra of hexasaccharide A.....115

Figure 55 - Transition of simulated dihedral angle to coupling constant by Karplus equation.....121

Figure 56 - Chemoenzymatic synthesized oligosaccharide constructs 1 to 8.....122

Figure 57 - Schematic synthesis of IdoA- and IdoA2S-containing oligosaccharides.....123

Figure 58 - The anion-exchange HPLC chromatogram of the epimerization reaction mixture.....124

Figure 59 - ^{1}H-NMR spectra of construct 1, 2, 3 and 4.....125

Figure 60 - ^{1}H-NMR spectra of construct 5, 6, 7 and 8.....126

Figure 61 - 2D ^{1}H-^{13}C HSQC-TOCSY NMR spectrum of construct 6.....127

Figure 62 - 2D ^{1}H-^{1}H COSY NMR spectrum of construct 6.....128

Figure 63 - 2D ^{1}H-^{13}C HSQC NMR spectrum of construct 6.....129

Figure 64 - 2D ^{1}H-^{13}C HMBC key correlations of construct 6.....130

Figure 65 - Hexuronic acid (GlcA, IdoA/2S) as pivot of conformation equilibrium to response sulfations 137

Figure 66 - The simulated proton distance for three major conformations of Iduronic acid 140

Figure 67 - Superimposition of low energy conformations of construct 4 with iduronic acid as major $^{2}S_{o}$ conformation and minor $^{1}C_{4}$ conformation 142

Figure 68 - Superimposition of low energy conformations of construct 8 with iduronic acid as major $^{1}C_{4}$ conformation and minor $^{2}S_{o}$ conformation 144

Figure 69 - Superimposition of low energy conformations of construct 8 with IdoA as $^{1}C_{4}$ conformation and construct 7 with IdoA as $^{1}C_{4}$ conformation 145

Figure 70 - Superimposition of low energy conformations of construct 4 with IdoA2S as $^{2}S_{o}$ conformation and construct 8 with IdoA as $^{1}C_{4}$ conformation 146

LIST OF ABBREVIATIONS

2-OST	2-*O*-sulfotransferase
3-OST	3-*O*-sulfotransferase
6-OST	6-*O*-sulfotransferase
AT	Antithrombin
C_5-Epi	C_5-epimerase
CS	Chondroitin sulfate
DEAE	Diethyl aminoethyl
DMSO	Dimethyl sulfoxide
DS	Dermatan sulfate
ESI-MS	Electrospray ionization mass spectrometry
FBS	Fetal bovine serum
FGF(R)	Fibroblast growth factor (receptor)
GAG	Glycosaminoglycan
Gal	Galactose
GlcNAc	*N*-acetyl glucosamine
GlcNS6S	*N*-sulfated, 6-*O*-sulfated glucosamine
$GlcNH_2$	Unsubstituted glucosamine
GlcNS	*N*-sulfated glucosamine
GlcNTFA	*N*-trifluoracetylated glucosamine
GlcA	Glucuronic acid
GlmU	Glucosamine-1-phosphate acetyltransferase/*N*-acetylglucosamine-1-phosphate uridyl transferase
HPLC	High performance liquid chromatography

HS	Heparan sulfate
HSV-1	Herpes simplex virus-1
IdoA	Iduronic acid
IdoA2S	2-*O*-sulfated iduronic acid
LMWH	Low molecular weight heparin
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NST	*N*-sulfotransferase
PAMN-HPLC	Polyamine high performance liquid chromatography
PAP	3'-phosphoadenosine 5'-phosphate
PAPS	3'-phosphoadenosine 5'-phosphosulfate
RPIP-HPLC	Reverse phase ion-pairing high performance liquid chromatography
TFA	Trifluoroacetic acid
UDP	Uridine diphosphate

Chapter I

Introduction

Proteoglycans

Carbohydrates, in addition to nucleic acids and proteins, dominate critical and diverse biological roles in living organism. Many physiological processes and protein functions including cell development, differentiation, inflammation, bacteria and virus infection, metastasis and cancer are mediated by proteoglycans, macromolecules composed of carbohydrates [2]. Proteoglycans are ranging from 32 to 500 kDa. The proteoglycans consist of core protein and various glycosaminoglycan (GAG) side chains. The types of core protein include syndecans, glypicans, agrin, perlecans, and serglycin (Figure 1) [3]. Syndecans and glypicans are membrane bound via transmembrane domain and glycosylphosphatidylinositol anchor, respectively [4, 5]. Agrin and perlecans are non-membrane bound and found in the extracellular matrix. Serglycin is the only intracellular proteoglycan and carrying heparin as side chain [6]. The heparin proteoglycan is majorly found in connective tissue mast cell which belongs to secretory granule of hematopoietic cell type [7].

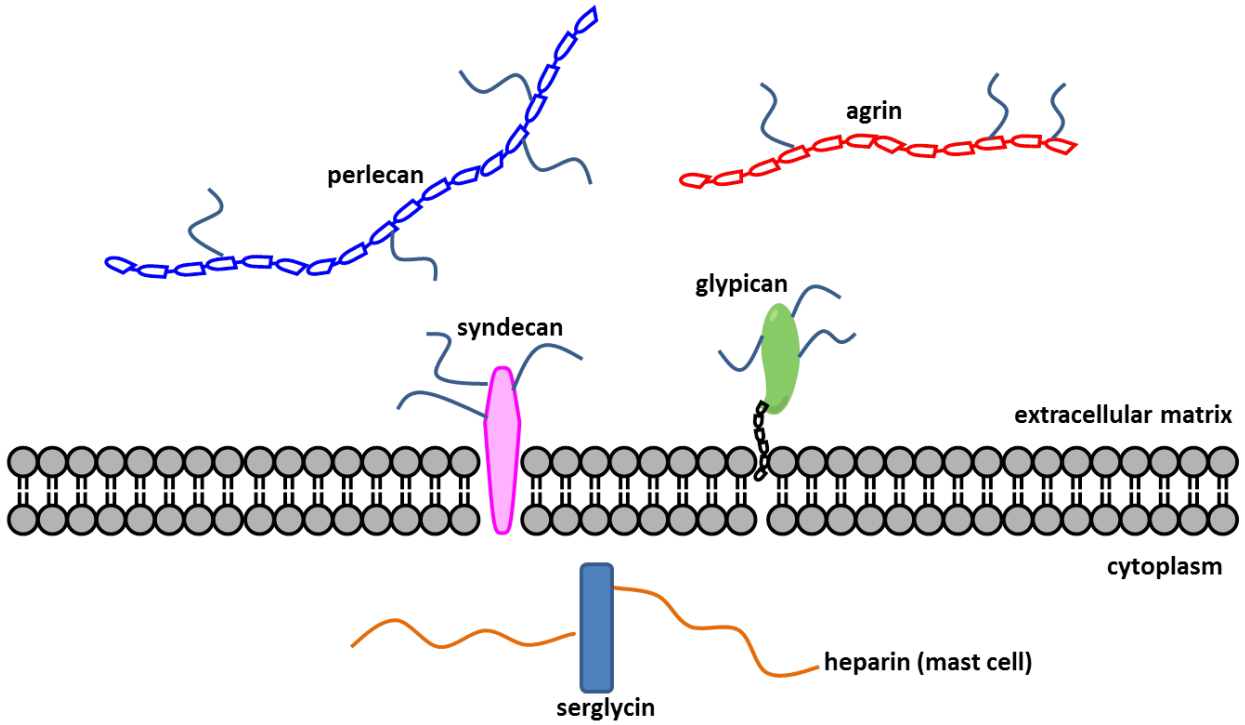


Figure 1. Representation of heparan sulfate proteoglycans. The syndecans and glypicans are membrane bound via transmembrane domain and glycosylphosphatidylinositol anchor, respectively. Agrin and perlecans are non-membrane bound and found in the extracellular matrix. Serglycin is the only intracellular proteoglycan and carrying heparin as side chain. The heparin proteoglycan is majorly found in connective tissue mast cell.

Main glycosaminoglycan types are hyaluronan, chondroitin sulfate, dermatan sulfate, heparan sulfate (HS) and heparin (Figure 2) [8]. Unlike protein translated from template DNA, the GAGs are generated by non-template enzymatic process as well as the array of various saccharide building blocks [9]. However, each type of GAG is made by specific repeated disaccharide and inherited manner of glycosidic linkage with specific anomeric configuration [9]. The usage of designated saccharide and glycosidic linkage render the distinct roles. Heparin and HS are the most widely investigated ones due to capable binding to antithrombin to contribute to clinic antithrombotic application [10].

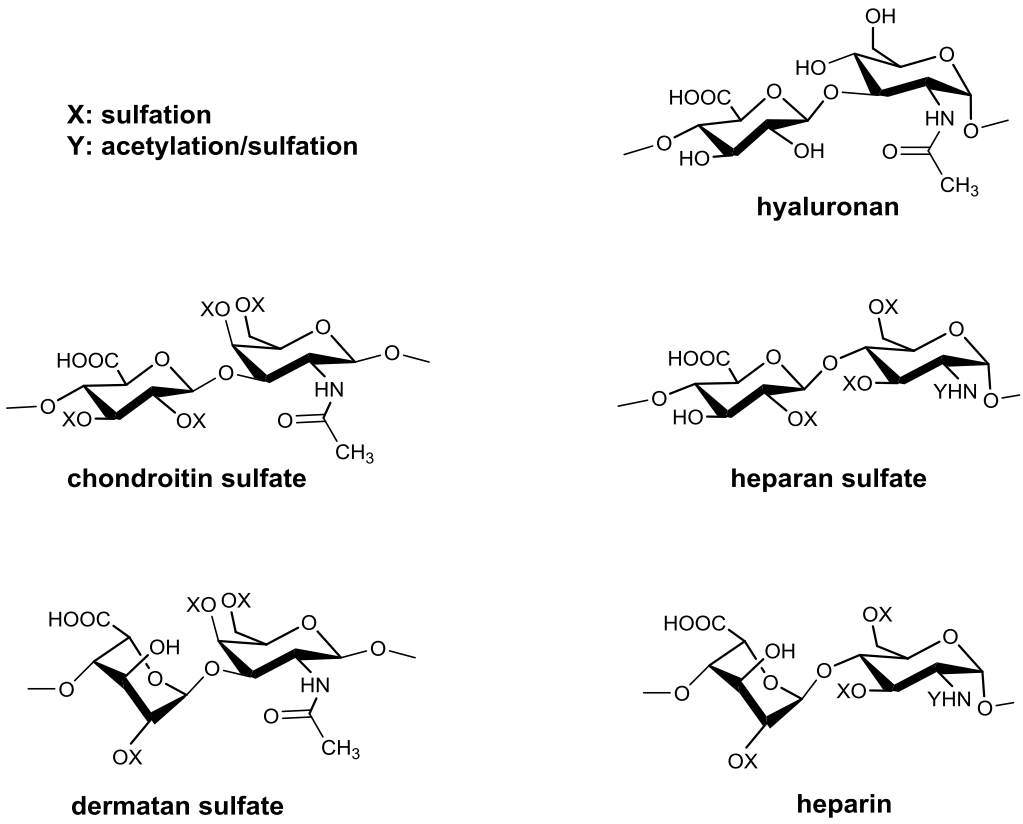


Figure 2. Structure of main type of glycosaminoglycans (GAGs). GAGs are the components of proteoglycans in the extracellular matrix and cell surface. Main GAG types are hyaluronan, chondroitin sulfate, dermatan sulfate, heparan sulfate, and heparin. The structure of common disaccharide repeating units are shown here. The possible sulfation or acetylation/sulfation are indicated as X or Y.

Chemical Structure of Heparin

Heparin, a widely used clinical anticoagulant since the 1930's, was the first biopolymeric drug and is one of the few carbohydrate based drugs [11, 12]. Because of the heterogeneity of heparin, structural characterization remains challenging. Heparin is a mixture of polysaccharides and has high polydispersity and polyheterogeneity [12] as well as flexible structural characteristics [11]. In addition, heparin is a negatively charged polysaccharide under physiological pH. This charged character originates from carboxylate and sulfo groups on the

saccharides. The average molecular weight of heparin is ~17 kDa [13] with a range of chain lengths from 5,000 to 50,000 Da [14]. Heparin has repeating disaccharide units consisting of uronic acid (90% L-iduronic acid, IdoA, and 10% D-glucuronic acid, GlcA) linked 1 → 4 to D-glucosamine (GlcN) [15], where the glucosamine residues can be either *N*-acetylated (GlcNAc) or sulfated (GlcNS). The GlcN monosaccharide can also contain *O*-sulfo groups, including 6-*O*- and 3-*O*-sulfation, while the GlcA and IdoA residues can be 2-*O*-sulfated.

Chemical Structure of Heparan Sulfate

Heparan sulfate (HS) is structurally analogous to heparin. HS has a similar disaccharide repeating structure to heparin with some noticeable differences. Generally, the size of the HS chains is longer with an average MW of 30 kDa [16], and HS has lesser IdoA (< 20%) content and lesser sulfation than heparin [12]. On average, HS contains around 0.6 sulfo group for each disaccharide, and heparin contains 2.6 sulfo groups per disaccharide [4]. In addition, HS has more diverse sulfated saccharide sequences and domain structures compared with heparin (Figure 3). The domain structures can be roughly categorized into NAc domain, NAc/NS domain, and NS domain based on sulfation density.

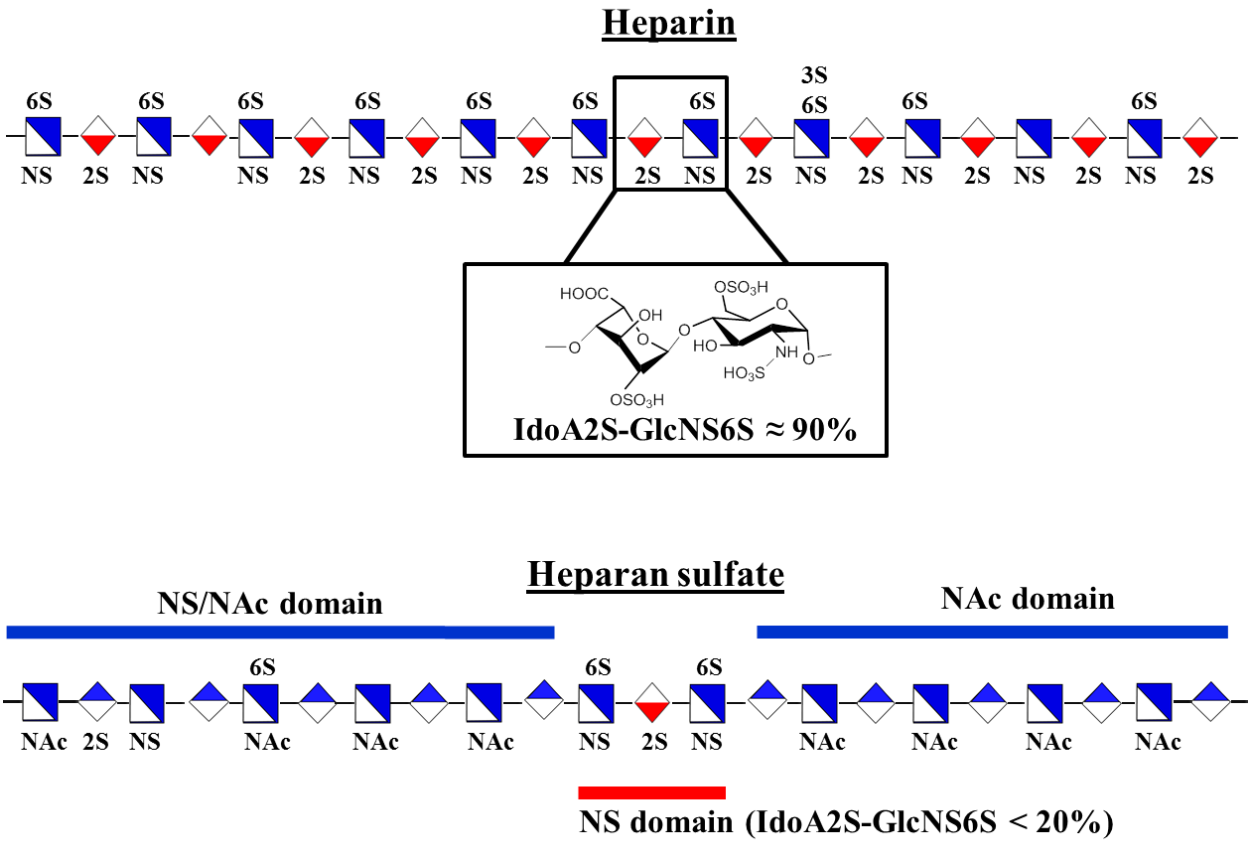


Figure 3. Chemical structure difference between heparin and heparan sulfate. Heparin contains repeating disaccharide units consisting of 90% 2-*O*-sulfated iduronic acid linked 1 → 4 to glucosamine. The glucosamine residues are mainly *N*-sulfated and 6-*O*-sulfated. Heparan sulfate contains lesser IdoA (< 20%) content and lesser sulfation than heparin. HS has more diverse sulfated saccharide sequences and domain structures comparing to heparin, such as NAc domain, NS/NAc domain, and NS domain.

Biosynthesis of Heparan Sulfate

Biosynthesis Scheme

The biosynthesis of HS mainly happens in Golgi apparatus. The participating enzymes cooperately work in the lumen. The biosynthetic processes of heparan sulfate can be categorized into three major steps: initiation, polymerization, and modification. Each step requires specific enzyme for reaction.

Initiation starts from attachment of tetrasaccharide linkage to the core protein (Figure 4). The core protein with the sequence of Ser-Gly/Ala-X(any)-Gly is recognized by enzyme for tetrasaccharide linkage [17]. The tetrasaccharide is common in all kinds of glycosaminoglycan with the sequence of –GlcAβ1,3Galβ1,3Galβ1,4Xylβ1-*O*-[Serine] [18]. Each monosaccharide is transferred from respective UDP-sugar sequentially. Firstly, xylose is transferred from UDP-xylose to serine by xylosyltransferase [19]. Secondly, two galactose residues are subsequently added by β1,4-galactosyltransferase-I and β1,3-galactosyltransferase-II [20, 21]. Finally, the glucuronic acid is attached by β1,3-glucuronyltransferase-I [22]. Some modification, including 4-*O*-, 6-*O*-sulfation on galactose as well as 2-*O*-phosphorylation on xylose, were reported relate to selectively regulation of different GAG biosynthesis [23].

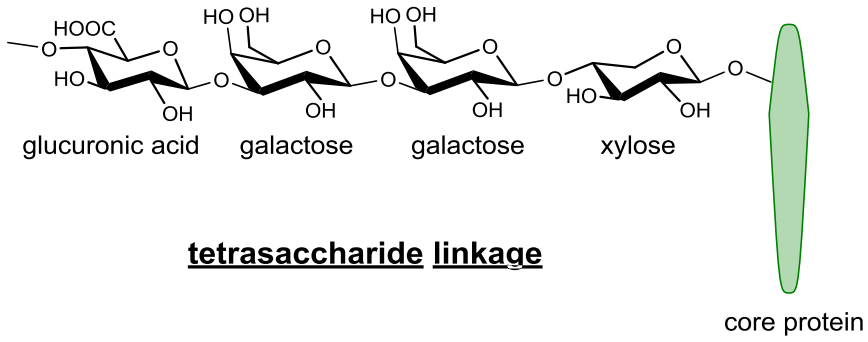


Figure 4. Glycosaminoglycan tetrasaccharide linkage to core protein. The tetrasaccharide with the sequence of GlcAβ1,3Galβ1,3Galβ1,4Xylβ1-*O*-[Serine] is attached to the core protein. Each monosaccharide is transferred from respective UDP-sugar. Xylose is transferred from UDP-xylose to serine by xylosyltransferase. Two galactose residues are subsequently added by β1,4-galactosyltransferase-I and β1,3-galactosyltransferase-II. The final glucuronic acid is attached by β1,3-glucuronyltransferase-I.

Polymerization of backbone chain happens after attachment of the first tetrasaccharide at initiation phase. The monosaccharide GlcNAc and GlcA is alternatively transferred by

exostosin-like and exostosin family of glycosyltransferase including EXTL1, EXTL2, EXTL3, EXT1 and EXT2 (Figure 5) [24, 25]. The first saccharide attachment to the linkage region tetrasaccharide is the divergent point for biosynthesis of CS or HS. Attachment of GalNAc by GalNAcT enzyme results in CS; attachment of GlcNAc leads synthesis of HS. EXTL2 and EXTL3 transfer the first GlcNAc to the GlcA-end of tetrasaccharide linkage [24, 26]. These EXTL enzymes were reported capable of recognizing and interacting with specific sequence of core protein to selectively polymerize HS chain [17, 27]. After the polymerization of GlcA and GlcNAc are performed by EXT1 and EXT2 [25]. Both EXT1 and EXT2 may be able to form a complex and cooperate with other biosynthetic enzymes to regulate the HS biosynthetic machinery [28].

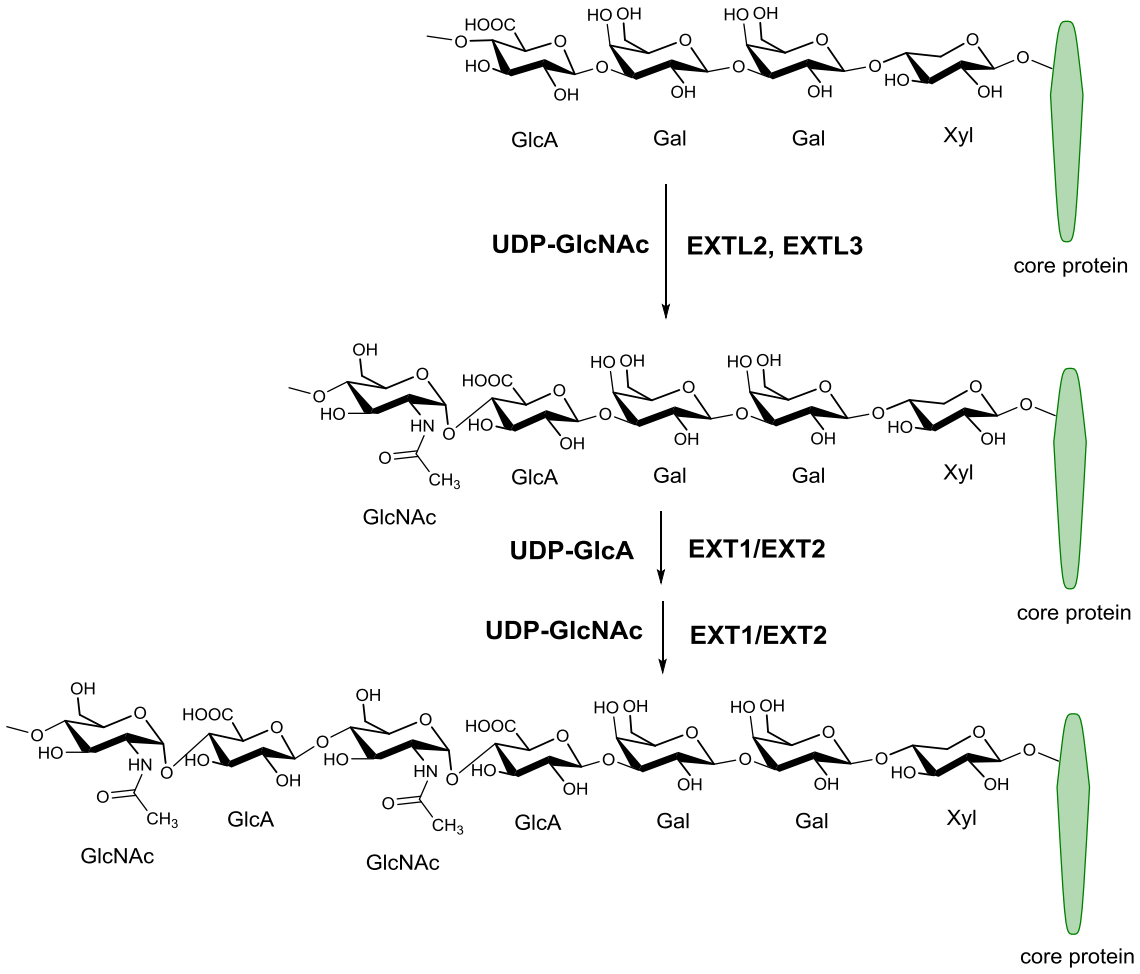


Figure 5. Polymerization of the heparan sulfate backbone. Polymerization starts after attachment of the first tetrasaccharide at initiation phase. EXTL2 and EXTL3 transfer the first GlcNAc to the GlcA-end of tetrasaccharide linkage. The adding of repeating units of GlcA and GlcNAc are performed by EXT1 and EXT2.

HS backbone modification includes deacetylation, epimerization, and 4 types of sulfation. Each reaction requires unique enzyme with different substrate specificity. The feature of substrate recognition leads each enzyme able to perform modification on modified HS backbone sequentially. The HS backbone modification is generally known in the order of deacetylation, *N*-sulfation, epimerization, 2-*O*-sulfation, 6-*O*-sulfation and 3-*O*-sulfation (Figure 6). The sulfotransferases utilize 3'-phosphoadenosine 5'-phosphosulfate (PAPS) as a sulfo

group donor to introduce sulfation to HS backbone. However, incompletion at each modification step leads great structural heterogeneity of HS.

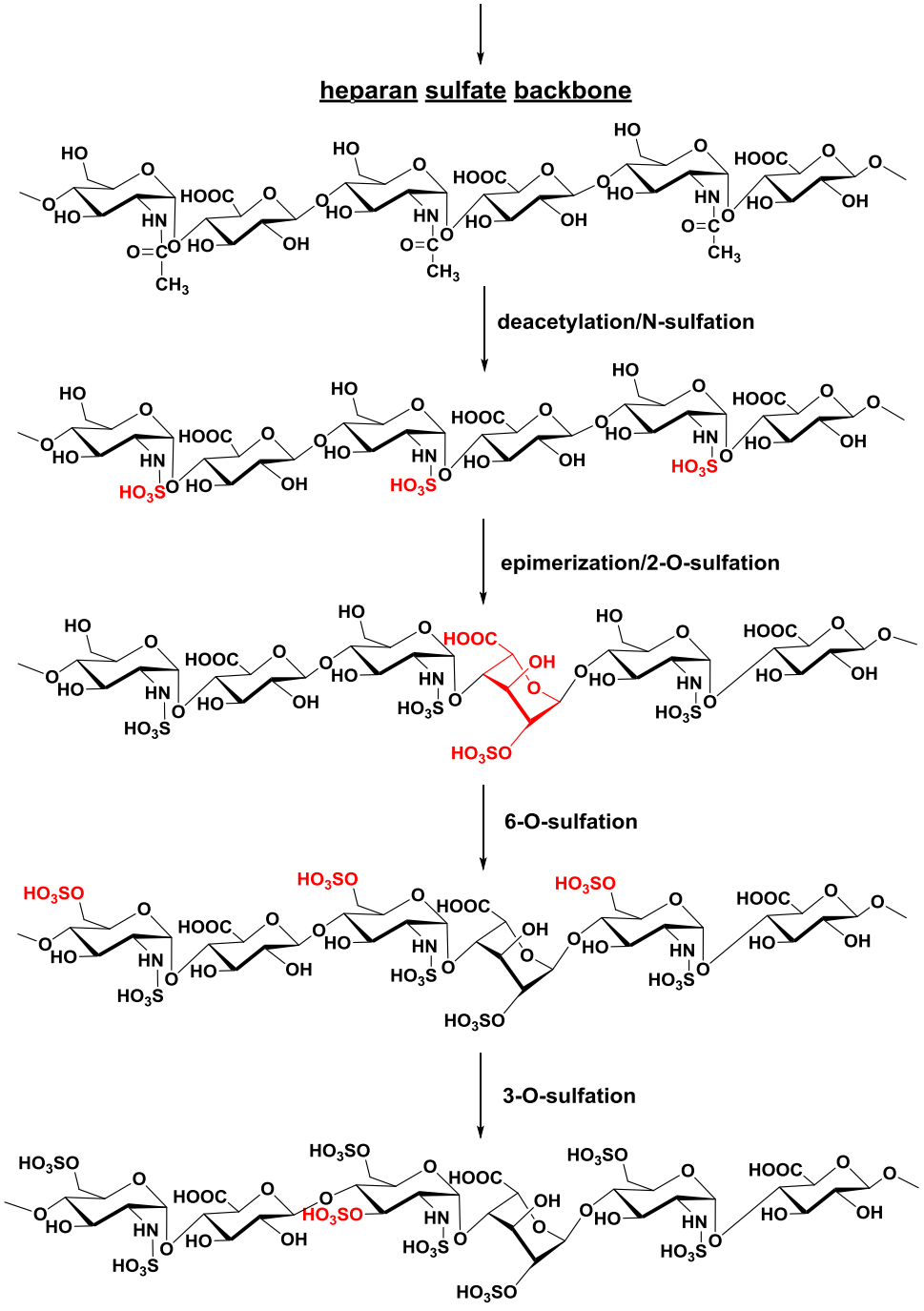


Figure 6. Heparan sulfate backbone modifications. After heparan sulfate backbone polymerization, the biosynthetic modifications are in the order of deacetylation, *N*-sulfation, epimerization, 2-*O*-sulfation, 6-*O*-sulfation and 3-*O*-sulfation. Each reaction requires unique enzyme with different substrate specificity. The modification parts are highlighted in red color.

Glucosaminyl Deacetylation/N-sulfation

Deacetylation and *N*-sulfation of GlcN residue is performed by glucosaminyl *N*-deacetylase/*N*-sulfotransferase (NDST) (Figure 7). NDST contains two domains for deacetylation and *N*-sulfation, respectively. There are four isoforms of NDST with different substrate preference providing different modified HS backbones, which can be selectively recognized by the other cooperative enzymes [29, 30]. NDST-1 and NDST-2 were reportedly critical for HS and heparin biosynthesis. NDST-1 is found in most of the cell. On the contrary, NDST-2 is mainly found in connective tissue mast cell associated with heparin synthesis [31, 32]. On the other hand, the biological significance of NDST-3 and NDST-4 are not well understood [33].

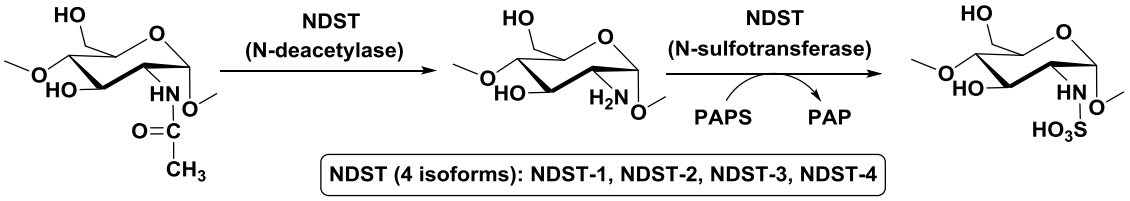


Figure 7. *N*-deacetylation and *N*-sulfation by NDSTs. NDST contains two domains: *N*-deacetylase and *N*-sulfotransferase for glucosamine deacetylation and sulfation, respectively. PAPS is the sulfo donor of the reaction. The four isoforms carry different substrate preference.

C_5-epimerization

Epimerization of $_D$-GlcA to $_L$-IdoA provides HS with greater heterogeneity in addition to different size and sulfation level of backbone. C5-epimerase was found only single isoform in

human genome with 70 kDa molecular weight [34]. The gene knockout mice were neo-natal lethal, proving the essential role of C5-epimerase [35]. The epimerization is happened at the C5-position of GlcA via putative carbanion intermediate [36]. The original $^4C_1$ conformation of GlcA is transformed to relatively stable $^1C_4$ or ^2So conformation of IdoA (Figure 8). This conformational change further increases the structural diversity of HS for various proteins selectively binding. Epimerization is reversible between GlcA and IdoA and preferably having a non-*O*-sulfated GlcNS at adjacent nonreducing-end [36, 37]. However, a recent report found GlcNAc at nonreducing-end 3 residues away from epimerization site can lock epimerization and relatively irreversible forward to IdoA [38]. The phenomenon suggests the substrate structure or sulfation state is critical for the subsequent biosynthetic enzymes to synthesize HS.

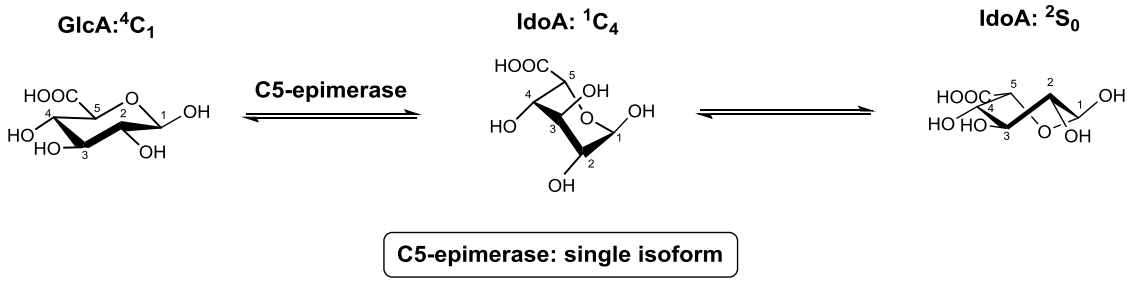
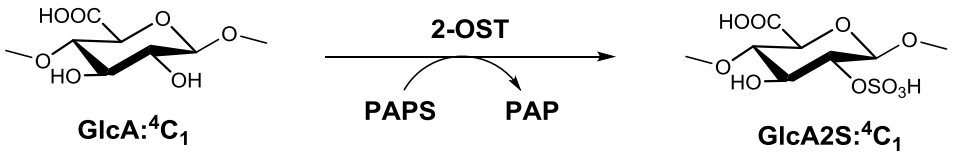


Figure 8. C5-epimerization by C5-epimerase. The epimerization is happened at the C5-position of GlcA via putative carbanion intermediate. A proton is abstracted and added back at C5 position. The original $^4C_1$ conformation of GlcA is transformed to relatively stable $^1C_4$ or ^2So conformation of IdoA.

2-O-sulfation

Uronosyl-2-*O*-sulfotransferase (2-OST) is used to transfer the sulfo group to uronic acid including GlcA and IdoA (Figure 9). 2-OST$^{-/-}$ gene knockout mice led neonatal death supporting

2-OST is essential in sustaining life [39]. Similar to C5-epimerase, there is only one isoform of 2-OST in human genome with 43 kDa molecular weight [40]. C5-epimerase and 2-OST were also been proven highly cooperate with each other, including stability, localization, and function [41]. 2-OST prefers to sulfate IdoA than GlcA [42]. The preference substrate is that an IdoA flanked by two non-6-*O*-sulfated GlcNS residues, suggesting that HS biosynthesis follow an order of reaction steps [43]. *N*-sulfated substrate makes the HS backbone subsequently recognized by epimerase, 2-OST, and finally 6-OST.



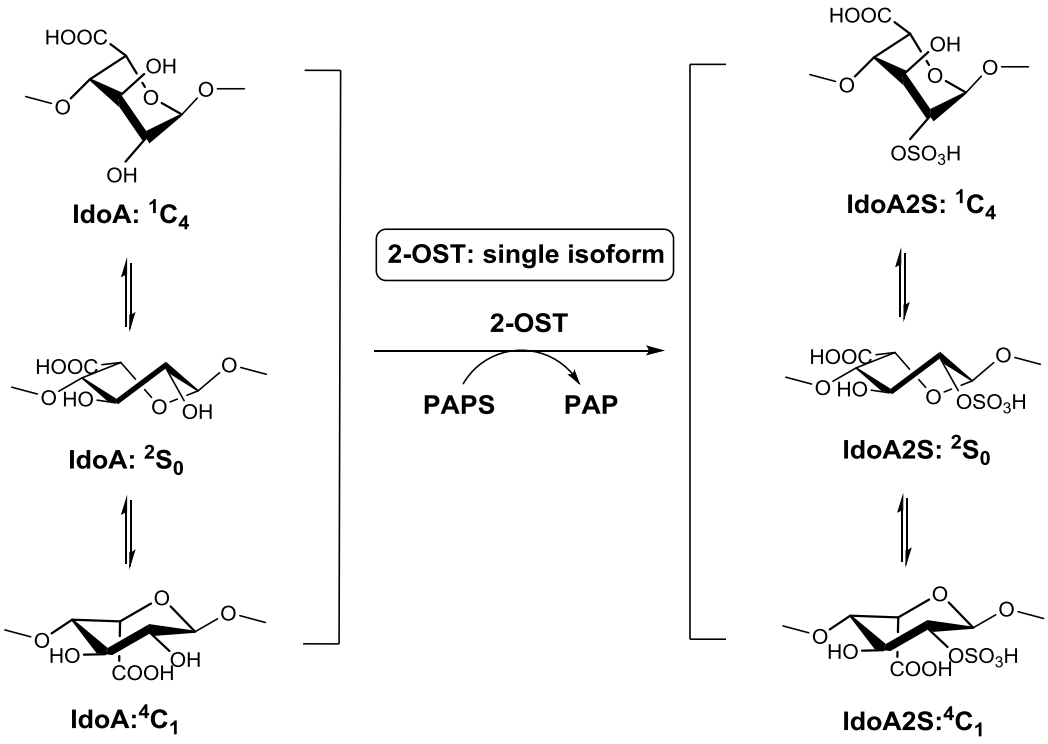


Figure 9. 2-*O*-sulfation by 2-OST. Uronosyl-2-*O*-sulfotransferase (2-OST) is used to transfer the sulfo group to uronic acid including GlcA and IdoA. PAPS is utilized as sulfo donor. IdoA is the preferred substrate than GlcA.

6-O-sulfation

6-*O*-sulfotransferase (6-OST) transfers the sulfo group to GlcNS residue at HS backbone (Figure 10). There are three known isoforms of 6-OST, including 6-OST-1, 6-OST-2, and 6-OST-3 with 50-57% homology and molecular weight 47kDa [44]. The expression of these isoforms may be regulated in tissue-specific manners [44]. These isoforms provide similar but subtly different substrate specificity resulting in structural heterogeneity of HS [44-47]. However, these studies of substrate preference may be ambiguous because of using heterogeneous polysaccharide substrate. The 6-*O*-sulfted HS were reported as prerequisite for different biological functions, such as signaling of FGF1 and FGF2 [44, 48]. The isoforms of 6-OST maybe provide diverse modified structures for the specific downstream bioactivities.

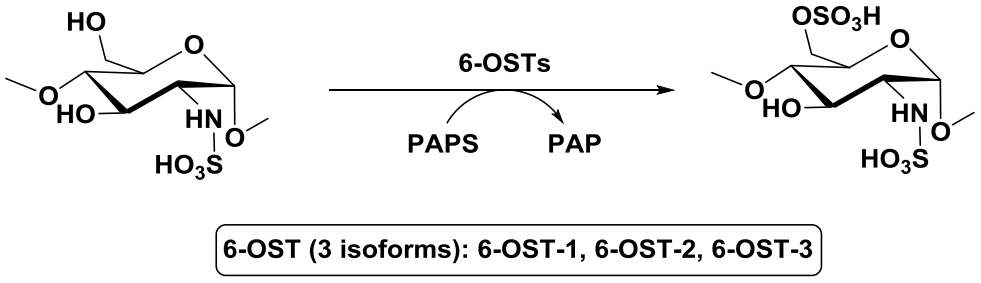
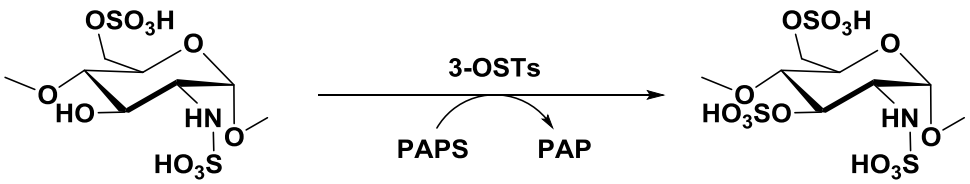


Figure 10. 6-*O*-sulfation by 6-OSTs. 6-*O*-sulfotransferase (6-OST) transfers the sulfo from donor PAPS to GlcNS residue at HS backbone. The isoforms provide the different substrate specificity resulting in highly structural heterogeneity of HS.

3-O-sulfation

The final sulfation modification is performed by 3-*O*-sulfotransferase (3-OST) adding sulfo group to the 3-*O*-hydroxyl of unsubstituted or *N*-/6-*O*-sulfated glucosamine (Figure 11). So far there are 7 isoforms of 3-OST discovered, which are ubiquitously or specifically residing in certain tissues [49, 50]. Although only around 0.5% of total sulfate happened in HS, 3-*O*-sulfation is considered the most critical modification for bioactivities. 3-OST-1 is reported correlate with HS binding to antithrombin to contribute anticoagulant activity; however, 3-OST-3A/B is highly associated with binding of gD-envelope protein of HSV-I to initial the viral entry [51-53]. These isoforms of 3-OST also equip with different substrate specificities. For example, 3-OST-1 prefers binding to GlcNS/6S with no*n*-2-*O*-sulfated GlcA/IdoA at nonreducing-end [50, 54]. The bound IdoA2S, at reducing-end of modified GlcN, adopt the $^{1}C_{4}$ conformation in 3-OST-1 crystal structure complex. On the contrary, 3-OST-3A/B prefers binding to GlcNS/6S with IdoA2S at nonreducing-end. The bound IdoA2S, at reducing-end of modified GlcN, adopt ^{2}So conformation is 3-OST-3 complex instead [55, 56]. Regarding to the rest of isoforms, 3-OST-2 prefers binding to GlcNS [57]. The substrate specificity of 3-OST-4 and 3-OST-6 are not well understood yet [58-60]. The substrate recognition was reported associated with the saccharide at the nonreducing-end of the modified GlcN as well as likely related to the conformation flexibility coming from IdoA entity[61]. These various isoforms of 3-OST greatly enlarge the structural heterogeneity of HS and the corresponding biological functions.



3-OST (7 isoforms): 3-OST-1, 3-OST-2, 3-OST-3A, 3-OST-3B, 3-OST-4, 3-OST-5, 3-OST-6

Figure 11. 3-*O*-sulfation by 3-OSTs. 3-*O*-sulfotransferase (3-OST) adds sulfo group to the 3-*O*-hydroxyl of unsubstituted or *N*-/6-*O*-sulfated glucosamine. The isoforms of 3-OST equips with different substrate specificities. The substrate recognition was reported associated with the saccharide at the nonreducing-end of the modified GlcN as well as likely related to the conformation flexibility coming from reducing-end IdoA entity.

Heparin Pharmaceutical Production

Heparin is a widely used anticoagulant. Based on molecular weight difference, heparin related drugs can generally be categorized into three types: unfractionated heparin (UFH), low molecular weight heparin (LMWH), and synthetic pentasaccharide fondaparinux (Arixtra) (Table 1) [62]. Fondaparinux is purely chemical synthesized with molecular weight of 1508 Da. However, UFH and LMWH are heterogeneous polysaccharide mixture extracted from animal source. The averaged molecular weight of UFH and LMWH is around 14 kDa and 4.5~6.5 kDa, respectively. Different types of heparin drug provide different pharmacokinetic profiles [63]. Healthcare workers can choose the best suitable form of heparin based on clinical needs.

Type of heparin/pharmaceutical properties	Unfractionated Heparin (UFH)	Low molecular weight Heparin (LMWHs)	Synthetic Heparin ULMW (Arixtra®)
MW_{avg}	~ 14000	~ 4500-6500	1508.3
Administration route	intravenous	subcutaneous	subcutaneous
Half life	30-60 mins	5-6 hours	12-15 hours
Side effect	bleeding, HIT (high)	bleeding, HIT (low)	bleeding, HIT (low)
Antidote	protamine	no	no
Indication	surgery, kidney dialysis	DVT, PE	DVT, PE
Excretion	liver	Partially liver	kidney
Renal insufficient patient usage	Yes	Yes; but monitor	No

Table 1. Heparin drugs and pharmacokinetic properties. Heparin related drugs can generally be categorized into three types: unfractionated heparin (UFH), low molecular weight heparin (LMWH), and synthetic pentasaccharide fondaparinux (Arixtra). Different types of heparin drug provide different pharmacokinetic profiles. * Heparin-induced thrombocytopenia (HIT), deep vein thrombosis (DVT), pulmonary embolism (PE).

Traditional Extraction and Depolymerization

Unfractionated heparin (UFH) is isolated from animal source including bovine lung and porcine intestine. Generally, these animal tissues are firstly boiled before collecting the mucosa. The dried mucosa then requires very complicated processes of purification and formulation to obtain the final commercial product for human usage. Low molecular weight heparin (LMWH) is derived from partially chemical or enzymatic digestion of UFH [64]. Based on different manufacturing approaches, there are three common commercialized LMWH products: enoxaparin (averaged MW: 4.5 kDa), dalteparin (averaged MW: 6 kDa), and tinzaparin (6.5 kDa) [65]. Enoxaparin is manufactured from alkaline hydrolysis of UFH and followed by chemical benzylation. Dalteparin is made by nitrous acid depolymerization of UF. Tinzaparin is obtained by heparinase digestion. However, FDA indicates that these LMWH drugs are not interchangeable due to different range of molecular weights and pharmacological profiles [66, 67]. The complicate manufacturing processes from original animal source to final drug also make supply chain relatively vulnerable and insecure.

Heparin Safety Crisis

In 2007-2008, batches of contaminated heparin were discovered in US, which drew significant public attention because the contaminated drug killed 94 patients in US [68]. Hundreds of people got severe allergic effects [69]. Afterwards, nuclear magnetic resonance (NMR) techniques were employed, and rapidly identified the nature of the contaminant to be

over sulfated chondroitin sulfate (OSCS) [70]. OSCS was reported intentionally added for cost reduction reason by manufacturer [71]. Due to OSCS structural similar to heparin, the quality control failed to screen this adulterant out and made the crisis happened [72]. The long process and vulnerable supply chain make the chance of contamination and adulteration. Therefore, the government and regulatory agency are eagerly looking for the alternative production approach to secure the heparin supply.

Chemical Synthesis of Heparin

Synthesis of heparin using a purely chemical approach was considered. However, the method is proved to be difficult to synthesize the homogeneous compound with the size of LMWH ranging from 4.5-6.5 kDa. Instead, scientists have found the pentasaccharide with the sequence of GlcNS6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S is sufficient enough for AT binding to contribute anticoagulant activity (Figure 12) [10]. The total synthesis method to synthesize this sequence of oligosaccharide was developed in 1989 with US Patent 4,818,816, 1989. The synthesis scheme takes overall around 50 steps with only 0.1% yield [73]. Fondaparinux (Arixtra) is the licensed drug. However, this size of pentasaccharide is not long enough to be neutralized by antidote protamine [74]. Furthermore, fondaparinux cannot be used for renal-impaired patient due to the compound requires kidney for clearance [75, 76]. Due to the lengthy chemical scheme and relative low yield, the chemical synthesis is unable to deliver large size oligosaccharides. A high efficient approach capable to synthesize homogeneous larger oligosaccharide resembling LMWH is urgently needed.

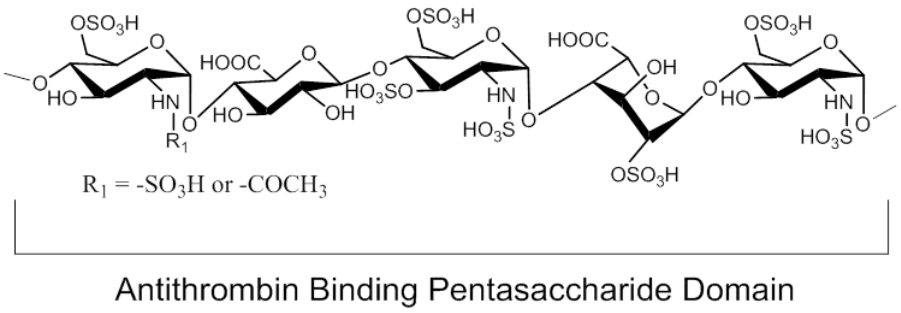


Figure 12. Anticoagulant pentasaccharide structure. The pentasaccharide with the sequence of GlcNS6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S can activate antithrombin to contribute anticoagulant activity.

Chemoenzymatic Synthesis

Recently, a new method using recombinant HS biosynthetic enzymes to prepare structurally homogeneous HS oligosaccharides was developed (Figure 13) [1]. This method allows the synthesis of a series of structurally homogeneous heparin or HS oligosaccharides at the scale up to gram level with high yield. The size of synthetic oligosaccharide more than twenty residues is feasible. Access to highly purified HS oligosaccharides allows for full-scale analyses of these complex molecules by NMR.

Notably this synthetic approach enables the analysis of structurally "rare" oligosaccharides which are difficult to isolate in sufficient quantity for extended biological and structural analysis. For example, the -GlcA2S- residue represents only 1-5% of total uronic acid in heparin isolated from natural sources [77]. The –IdoA- entity without 2-*O*-sulfation is even less abundant within heparan sulfate. To address the knowledge gap of these potentially biological important species, chemoenzymatic approach is capable to synthesize a library of standards efficiently. This approach is also possible to secure the future heparin and heparan sulfate supply chain.

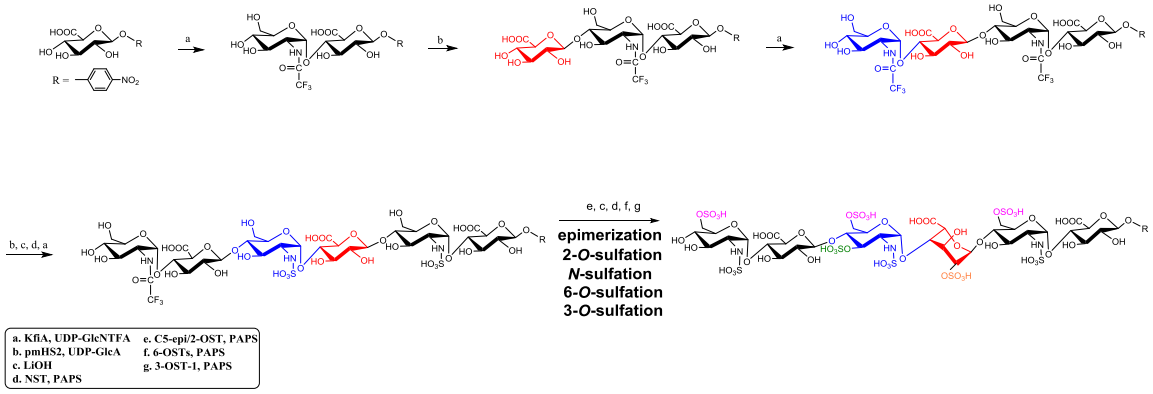


Figure 13. Chemoenzymatic synthesis of oligosaccharide. The recombinant HS biosynthetic enzymes are used to prepare structurally homogeneous HS oligosaccharides with high yield. Rearrangement of the reaction steps can synthesize diverse desired structures. The reactions can be simply repeated to synthesize larger size of oligosaccharides. The structurally rare oligosaccharides, which are difficult to isolate in sufficient quantity, can also be acquired by chemoenzymatic approach.

Structural Analysis of Heparan Sulfate

Accurate structural characterization of heparan sulfate is remaining challenging for decades. The polydispersity of the sequence as well as the polyheterogeneity of the sulfation pattern are formidable issues for structural analysis. The access of sufficient structural defined samples for study is also a big problem. However, it is extremely important to accurately characterize the heparan sulfate structure. The study of structure-activity relationship requires comprehensive understanding the molecular properties. Nowadays there are several of methods used for identifying heparan sulfate structures including disaccharide analysis, Mass spectrometry, and NMR spectroscopy.

Enzymatic or Chemical Degradation for Disaccharide and Mass Analysis

Since heparan sulfate is heterogeneous and ranging from 100 to 400 units of saccharides, depolymerization followed by compositional analysis is an alternative approach to characterize

the structure. Two commonly used degradation methods are enzymatic and chemical degradation. The degraded components are afterward analyzed by reverse phase ion pairing-high performance liquid chromatography (RPIP-HPLC) [78, 79]. The chromatography condition is altered by ion pairing reagent, such as tributylamine or tetrabutylammonium phosphate. This system can well resolve the epimers as well as the degraded disaccharides with different sulfation patterns. The reference standards are co-eluted to identify the components of heparan sulfate. There are overall 22 various disaccharides reported from heparan sulfate compositional analysis [5]. The mass spectrometry is utilized to confirm the molecular weight of degraded fragments. Electrospray ionization mass spectrometry (ESI-MS) is useful to determine the sequence and length of fragments. Tandem mass spectrometry (MS/MS) can further investigate the type and number of modifications based on fragmentation patterns.

Heparin lyase is used for enzymatic degradation. There are three isoforms heparin lyase I~III found from *Flavobacterium heparinum*. Different isoform brings different substrate specificity to depolymerize heparan sulfate [5, 80]. Heparin lyase I prefers the cleavage site with the sequence GlcNS-IdoA2S. Heparin lyase II and III have relative promiscuous substrate preference. Heparin lyase II cleaves the glycosidic linkage with GlcNAc/NS-GlcA/IdoA/2S, and heparin lyase III can degrade the linkage of GlcNAc/NS-GlcA. The β-elimination reaction is performed by heparin lyase to cleave the glycosidic linkage (Figure 14). The degraded fragments carrying the $\Delta^{4,5}$–unsaturated uronic acid with UV 232nm absorbance greatly makes the observation easy. The identities and relative abundance of observed fragments are used to deduce the structure of parent molecule. However, the degraded fragments lose the C5 configuration due to forming of double bond. This disadvantage makes enzymatic degradation approach unable to understand the property of GlcA/IdoA epimer in the parent molecule.

Additionally, the incompletion of enzymatic degradation results in misleading or underestimating of the amount of monitored disaccharide fragments.

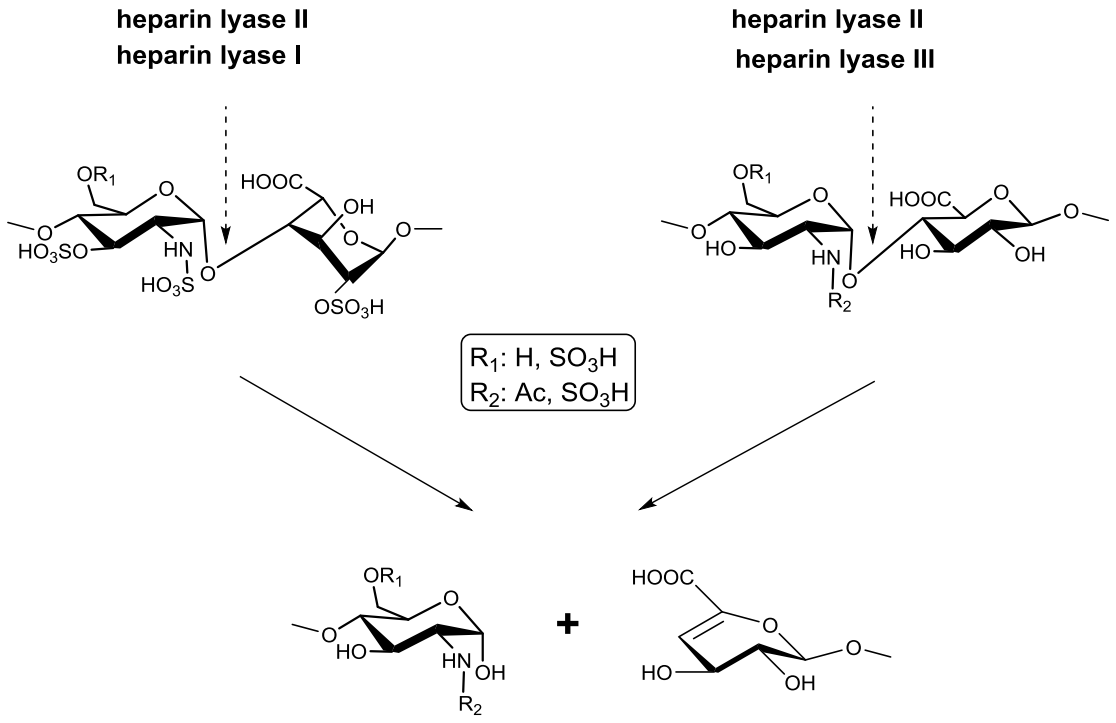


Figure 14. Enzymatic degradation of heparan sulfate by heparin lyase. Heparin lyase isoforms show different substrate specificity to depolymerize heparan sulfate. Heparin lyase I prefers the cleavage site with the sequence GlcNS-IdoA2S. Heparin lyase II and III have relative promiscuous substrate preference. The β-elimination reaction is performed at glycosidic linkage to produce the $\Delta^{4,5}$– unsaturated uronic. The C5 configuration of parent molecule is lost due to forming of double bond.

Chemical degradation by nitrous acid is another approach to depolymerize the heparan sulfate (Figure 15) [81]. The glycosidic linkage between glucosamine and uronic acid can be degraded. The glycosidic linkage of GlcNH2-GlcA/IdoA is the favored substrate at elevated pH 4.5; on the other hand, at pH 1.5 the sequence of GlcNS-GlcA/IdoA is the preferred one. The GlcNAc residue requires using hydrazine to remove acetyl group before nitrous acid treatment. The chemically degraded fragments are followed by reduction utilizing sodium borohydride. The final reduced disaccharide fragments, carrying modified residue 2,5-anhydromannitol, are

co-eluted with standards during RPIP-HLPC analysis. The advantage of this approach is the configuration of uronic acid still remained for inferring the parent structure [5]. However, the radio-labelled sample needs to be provided before degradation since there is no UV absorbance for degraded fragments. Combining use both enzymatic and chemical degradation approach give people good understanding of the identity and relative abundance of components of heparan sulfate.

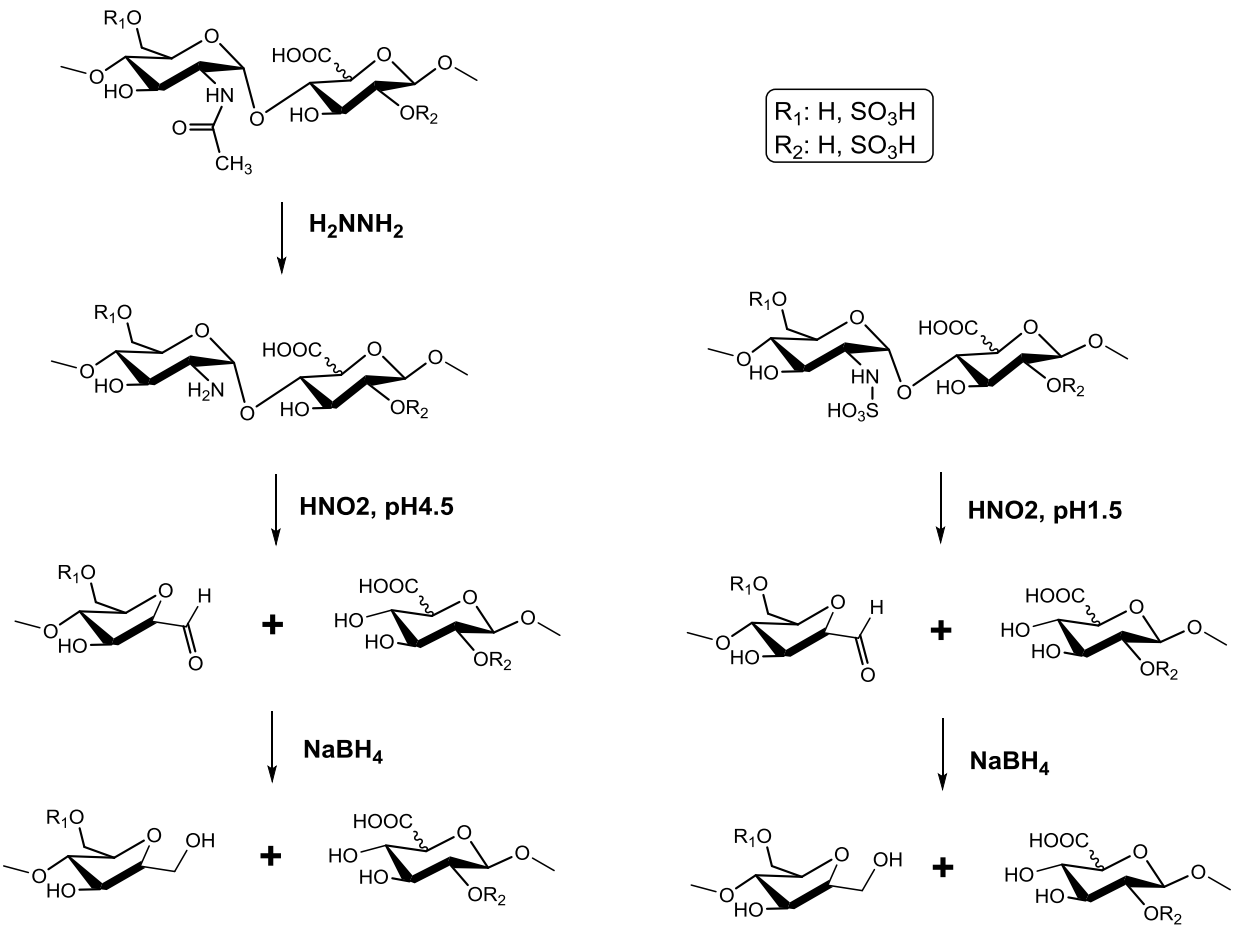


Figure 15. Chemical degradation of heparan sulfate by nitrous acid. The glycosidic linkage between glucosamine and uronic acid can be degraded by nitrous acid. The glycosidic linkage of GlcNH2-GlcA/IdoA is the favored substrate at elevated pH 4.5; on the other hand, at pH 1.5 the sequence of GlcNS-GlcA/IdoA is the preferred one. The GlcNAc residue requires using hydrazine (H_2NNH_2) to remove acetyl group before nitrous acid treatment. The chemically degraded fragments are followed by reduction utilizing sodium borohydride to obtain 2,5-anhydromannitol. the configuration of uronic acid still remained by this chemical degradation approach.

Nuclear Magnetic Resonance (NMR) Measurement

Reliable structural characterization techniques are critically important to safeguard the quality of heparin drugs. In 2008, batches of contaminated heparin were discovered in US, which drew significant public attention because the contaminated drug killed 94 patients in US [68]. Nuclear Magnetic Resonance (NMR) techniques were employed, and rapidly identified the nature of the contaminant to be over sulfated chondroitin sulfate (OSCS), demonstrating the important role of NMR measurements in heparin quality control (Figure 16) [11].

NMR data provides structural information that cannot be obtained by other methods, specifically; epimer recognition, anomeric configuration identification, sulfated position determination, and conformation and secondary structure determination. One limitation in utilizing NMR to analyze the structure of heparin or HS is that these species are a mixture of compounds and pure oligosaccharide standards have not been available. Although certain oligosaccharides have been isolated from partially depolymerized heparin or HS, obtaining sufficient amounts of those highly purified oligosaccharides is a difficult and time consuming process.

The chemoenzymatic approach makes the synthesis of sufficient amount of oligosaccharide constructs possible. Structural characterization by NMR experiments becomes much easier by virtue of better signal intensity. ^{1}H NMR was used to test the purity of constructs as well as to measure the $^{3}J_{\text{H-H}}$ coupling constants. To fully assign the chemical shifts of all the proton and carbon signals, multidimensional homonuclear (COSY, NOESY) and heteronuclear (HSQC, HMBC, HSQC-TOCSY) NMR experiments were used to resolve the overlapped

signals. The chemical shift assignments of constructs were acquired by dissecting the individual residue coupling system, one-bond and three-bond correlation based on NMR spectra.

Since the 2008 contaminated heparin crisis, NMR analysis of heparin drug has become part of the heparin sodium US Pharmacopeia monograph. Because heparin is a heterogeneous polysaccharide mixture, the assignment of all of the signals is difficult. To this end, acquiring the NMR spectra of highly purified heparin oligosaccharide standards provides a map for assigning highly complex heparin spectra, especially those signals from rare components. Chemoenzymatic synthesis used to produce structures for NMR.

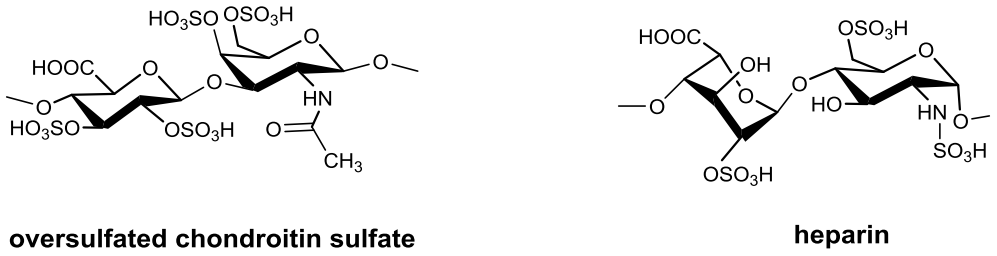


Figure 16. Structural comparison between over sulfated chondroitin sulfate (OSCS) and heparin. Over sulfated chondroitin sulfate was identified as contaminant in heparin drug. The main disaccharide repeating unit is shown here.

Molecular Dynamics (MD) Simulation

In addition to the length of saccharide chain, diversity of sulfation patterns and plasticity of iduronic acid (IdoA) are known correlated to HS biological importance [8, 82]. The flexible glycosidic linkages and IdoA could facilitate heparan sulfate conformational rearrangement. The same structure of HS oligosaccharide may dynamically populate at several conformations [43, 83]. The preference of conformation could attribute to comparable energy states established by local environment. Highly heterogeneous sulfations of heparan sulfate provide the local environment diversity. NMR used to be the best way to accurately identify the molecular

structure of protein. Unfortunately due to highly flexible character, carbohydrate is difficult for NMR analyzed [84]. The rapid dynamically exchange between conformations makes NMR only observe the signals averaged from individual conformations [85]. Alternatively, molecular dynamics (MD) simulation could identify the list of possible low energy conformations.

However, simulation is unable to accurately calculate the equilibrium population of each conformation [86]. Take the advantage of both NMR and simulation analysis, the best manner to calculate the realistic equilibrium population of conformations is to deconvolute the NMR measured parameters into simulated parameters for each conformation with corresponding populations. The measurement of spin-spin coupling constant ($^3J_{H\text{-}H}$) and intensity of nuclear overhauser effects (NOE) are the appropriate parameters representing saccharide conformation [82]. The change of conformation altering the dihedral between vicinal protons as well as the spatial distance between ring protons which are reflecting on shift of $^3J_{H\text{-}H}$ coupling constant and NOE, respectively. The $^3J_{H\text{-}H}$ coupling constants are obtained by direct ^{1}H-NMR experiment or calculated from simulated dihedral using Karplus equation [87, 88]. Intensity of NOE signals are acquired from 2D-NOESY NMR experiment or calculated from simulated proton distance. Both $^3J_{H\text{-}H}$ coupling constant and NOE are used as constraints to monitor the conformational change and corresponding population. Parallel using both NMR and simulation approach enable to capture the structural and dynamic properties of carbohydrate conformation ensembles [85].

Biological Functions of Heparan Sulfate

Modification by sequential biosynthetic enzymes provides high structural diversity to HS. The isoform of enzymes with different substrate specificity further enlarge the array of possible

synthetic structures. The diverse structures of HS associate with various physiological functions. HS has been being reported relate to anticoagulation, cell proliferation, tumor progression, inflammation, viral and bacterial infection. The structure-activity relationship between HS and anticoagulation was well known. However, the studies understanding the role of HS play to the other physiological or pathophysiological functions is still limited or ongoing. The homogeneous HS oligosaccharide is required for detail ligand-protein interaction studies.

Anticoagulation

Blood flow is highly regulated and finely tuned in living organism. The coagulation and anticoagulation require under strictly control and balance. Series of serine protease are lined in a cascade involved in activation or regulation downstream enzymes to achieve final blood clotting (Figure 17) [89]. For example Factor IIa, commonly called thrombin, is the key protease cleaves and activates the fibrinogen to fibrin to form clot. Coagulation cascade can initiate from either intrinsic or extrinsic pathway [90]. The intrinsic pathway is induced by damaged vasculature surface; on the other hand, extrinsic pathway is happened upon body injury. To balance the coagulation cascade, several anticoagulation factors are critical to maintain blood flowing if there is no vasculature damage or body trauma. These anticoagulation factors, also called serpins, are serine protease inhibitors. Serine protease inhibitors include heparin cofactor II [91, 92], protein C inhibitor [93, 94], tissue factor pathway inhibitor [95], and the most essential one – antithrombin-III (AT) [10]. AT is the key regulator of blood coagulation cascade, which gene knockout is lethal to mice model [96, 97].

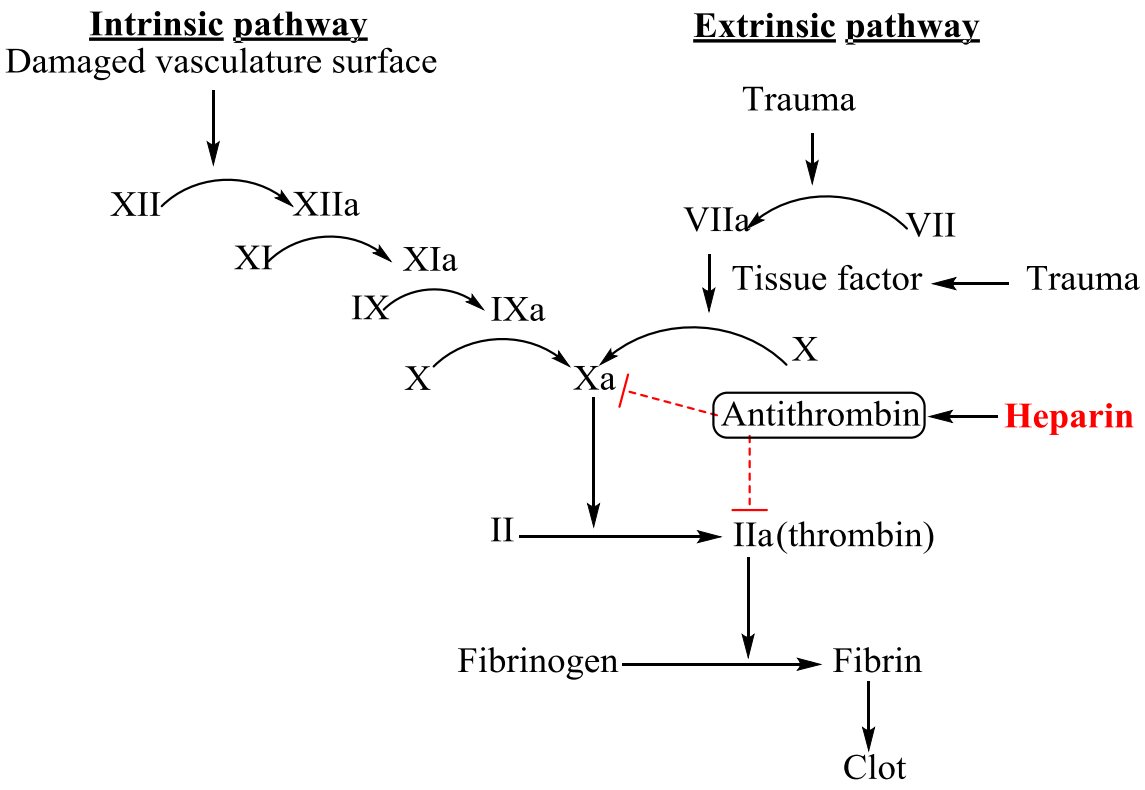


Figure 17. Blood coagulation cascade. Series of serine protease are line in a cascade involved in activation or regulation downstream enzymes to achieve final blood clotting. Coagulation cascade can be initiated from either intrinsic or extrinsic pathway. The intrinsic pathway is induced by damaged vasculature surface; on the other hand, extrinsic pathway is happened upon body injury. Factor Xa is the convergent point of both pathways. Antithrombin is the key regulator of blood coagulation cascade via interaction with heparin.

The interaction mechanism of serpin and the corresponding protease is exquisitely regulated. Generally these serpins consist of a reactive center loop which can be specific recognized by the corresponding protease [98-100]. For example, antithrombin (AT) can inhibit factor-IIa or Xa using reactive center loop extended into the catalytic site of protease. This AT-protease recognition is further regulated by binding to heparan sulfate or heparin. Heparin binding to AT induces conformational change of AT and outward reactive center loop. This conformational change of AT significantly increase the affinity and inhibitory effect to factor-IIa or Xa [98, 101, 102].

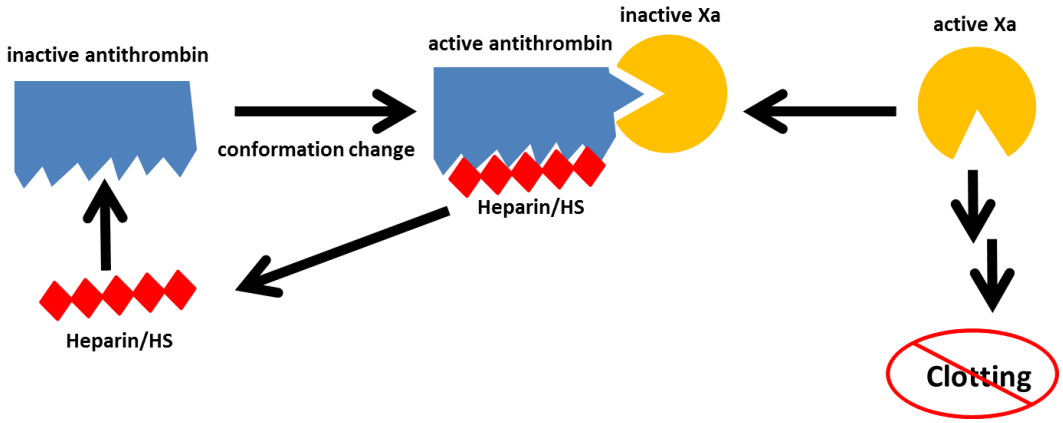


Figure 18. Anticoagulation mechanism that heparin activates antithrombin to inhibit Xa. Heparin binding to AT induces conformational change of AT and outward reactive center loop. This conformational change of AT significantly increase the affinity and inhibitory effect to factor-Xa. The coagulation cascade is held by inhibition of the key protease regulator Xa.

Upon the heterogeneous structure of heparin, the pentasaccharide with the sequence of GlcNS6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S was reported able to bind and activate AT to inhibit the downstream protease in coagulation cascade (Figure 12) [103]. Unless under high concentration, heparin or heparan sulfate without this pentasaccharide sequence weakly binds to AT [104]. AT binding to pentasaccharide is mainly due to electrostatic interactions between basic amino acid residues and acidic sulfate and carboxylate groups [10]. The structure-activity study shows that the 3-*O*-sulfate and nonreducing-end 6-*O*-sulfate are especially essential for AT binding affinity [105, 106]. Additionally, different sulfation also associates with the conformational shift of the saccharide. Within the oligosaccharide fragment the ^{2}So conformation of IdoA residue, rather than $^{1}C_{4}$ form, is critical for AT binding (Figure 19). However, except AT, the mechanism and specific oligosaccharide structure requirement binding to the other serpins is still unclear.

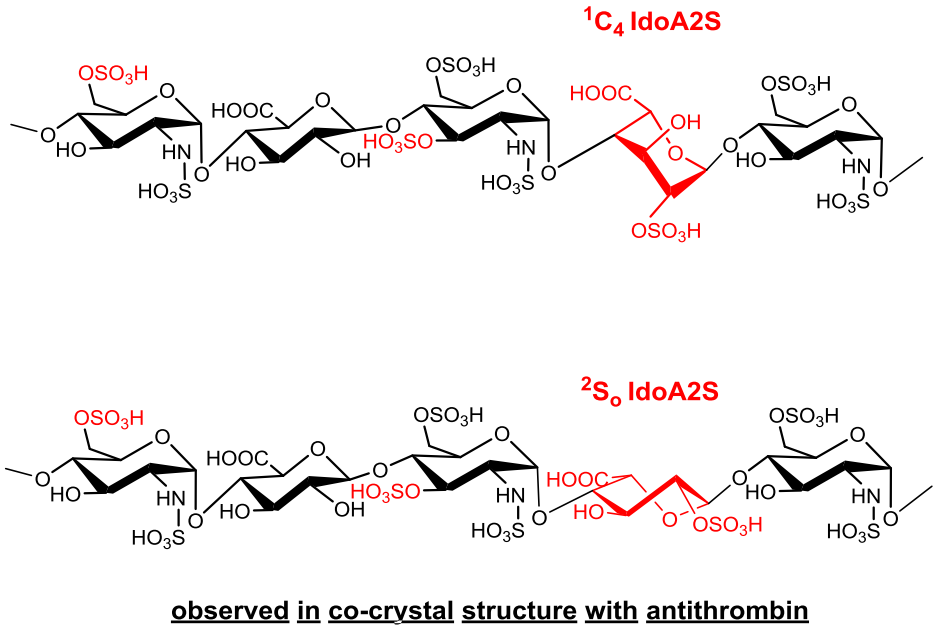


Figure 19. Key structural components of pentasaccharide for AT binding. The pentasaccharide with the sequence of GlcNS6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S was reported bind and activate AT to inhibit coagulation cascade. The IdoA2S as well as 3-*O*-sulfate and nonreducing-end 6-*O*-sulfate, highlighted in red, are especially essential for AT binding. The ^2S_o conformation of IdoA residue, rather than $^1C_4$ form, is observed in crystal structure of HS/AT complex.

In addition to anticoagulant serpins, heparin binds to protease as well. Heparin could serve as bridge bringing serpin and corresponding target protease tightly together (Figure 20). For example, in ternary structure the nonreducing-end of heparin can bind to factor-IIa with reducing-end pentasaccharide binding to AT [107]. Crystal structure elucidates that heparin binds the exosite of factor-IIa and orientates the catalytic site toward anticoagulant AT. Similarly, heparin also acts as bridge between AT and target protease factor-IXa demonstrated in another ternary crystal structure. The role of heparin play is more complicate because of capability forming different ternary complex to mediate the protein-protein interaction at higher level of regulation. On the other hand, smaller fragment of LMWH are not able to functional as

bridge. Simply, LMWH activate AT conformational change to directly bind and inhibit factor-Xa to contribute anticoagulation [108]. The pentasaccharide drug fondaparinux is the factor-Xa inhibitor via this relatively simple mechanism [73].

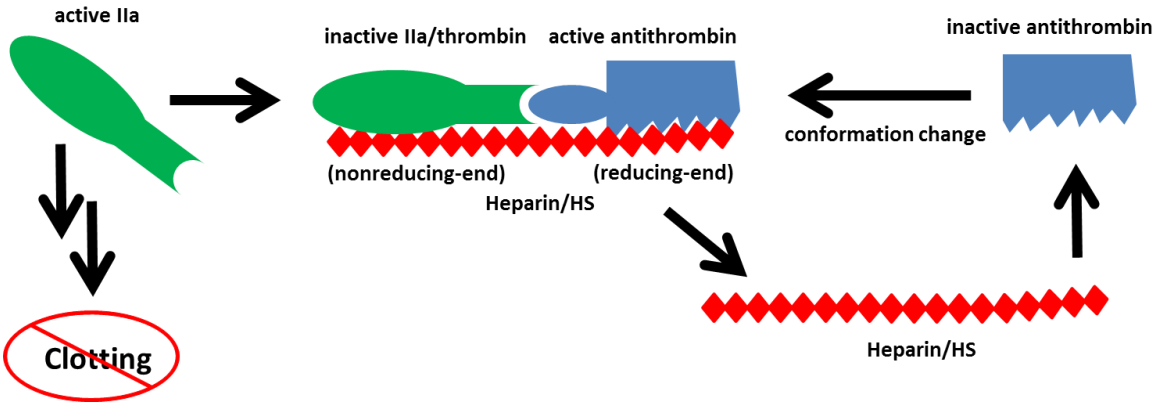


Figure 20. Anticoagulation mechanism that heparin activates antithrombin to inhibit IIa (thrombin). Heparin binding to AT induces conformational change of AT and outward reactive center loop. Heparin also acts as bridge bringing AT and IIa tightly together. In ternary structure the nonreducing-end of heparin can bind to factor-IIa with reducing-end pentasaccharide binding to AT. The inhibition of IXa possibly takes similar mechanism. However, smaller fragment such as fondaparinux or LMWH are not able to functional as bridge to exhibit anti-IIa activity.

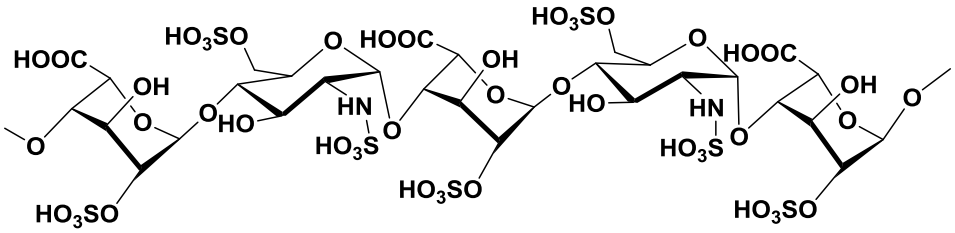
Cell Proliferation

Heparan sulfate play the important role mediating cellular proliferation. Heparan sulfate and heparin can regulate the binding between fibroblast growth factor (FGF) and the receptor tyrosine kinase (FGFR) [109-112]. The binding of FGF to FGFR induces the receptor dimerization, autophosphorylation, and finally leads the downstream signaling for cell differentiation or proliferation [113-115]. Due to highly negatively charged entity, heparan sulfate can stabilize the FGF/HS/FGFR ternary complex via electrostatic interaction with positively charged amino acid residues [116]. The intervention of heparan sulfate greatly enhances the affinity between FGF and FGFR. Nowadays, overall a family of 22 FGFs and

FGFR-1~4 are identified. It's been believed certain selectivity existed among the structure of heparan sulfate and FGF/FGFR. Different set of complex may bring different signaling pathway for specific requested response.

The heparan sulfates binding to FGF1 or FGF2 are the most extensive studied interactions. 2-*O*-sulfated iduronic acid (IdoA2S) and *N*-sulfated glucosamine (GlcNS) are both required for binding to FGF1 or FGF2. However, FGF1 shows different substrate preference versus FGF2 (Figure 21). The repeated trisulfated disaccharide IdoA2S-GlcNS6S with the size larger than pentasaccharide is the preferred substrate for FGF1 [117]. On the other hand, FGF2 prefer binds to non-6-*O*-sulfated HS to mediate the cellular proliferation [118]. The sequence of UA-GlcNS-UA-GlcNS-IdoA2S pentasaccharide was reported minimum HS fragment with high affinity to FGF2 [119]. The substrate specificity of different FGFs may involve multiple factors, such as length of HS, IdoA/GlcA ratio property, and the position of sulfo groups [120]. The comprehensive study of structure-activity relationship between HS and FGFs requires further investigation. Understanding the mechanism of HS binding to FGFs and associated downstream signaling may help design of HS oligosaccharide for suppression of tumor cell proliferation and angiogenesis.

FGF1 minimum binding site



FGF2 minimum binding site

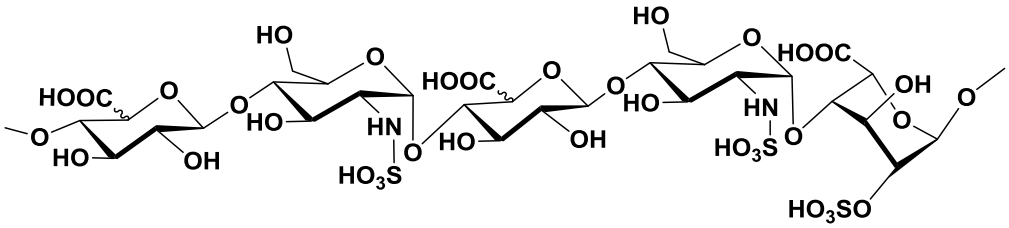


Figure 21. Minimum heparan sulfate structure for FGFs binding. FGF1 shows different substrate preference versus FGF2. The repeated trisulfated disaccharide IdoA2S-GlcNS6S with the size larger than pentasaccharide is the preferred substrate for FGF1. FGF2 prefer binds to non-6-*O*-sulfated HS with the minimum sequence of UA-GlcNS-UA-GlcNS-IdoA2S. The length of HS, IdoA/GlcA ratio property, and the position of sulfo groups may involve in substrate specificity of FGFs binding.

Tumor Progression

HS is known associated with several aspects of tumor progression including tumor growth, angiogenesis, invasiveness, and metastasis [121]. Different modification of HS may either enhance or inhibit tumor progression [122]. On the other hand, cancer cell can modify nearby HS sequence as well as mediate the expression level and distribution of HS-attached proteoglycan [123]. For example, syndecans-1 is down-regulated or absent in several of cancer

cell lines [124]. However, the detail HS proteoglycan (HSPG) structure-activity relationship needs further clarification.

HSPG present on the cell surface and have the ability to interact with proangiogenic factors including fibroblast growth factors (FGFs) and vascular endothelium growth factors (VEGFs). Some sequences of HS can bind to FGFs to mediate dimerization of FGFR and the following tyrosine kinase signaling (Figure 21). The density of these sequences of HS was reported increased in some of cancer lines [125]. The increased abundance of certain HS fragments can enhance FGFs –mediated signaling and the consequent tumor cell proliferation. Similarly, vascular endothelium growth factor is also mediated by HSPG to influence the angiogenesis [126]. Additionally, HS also can modulate tumor metastasis. Tumor cells utilize HS to adhere to the surface of endothelial cells via p-selectin (Figure 22). Once tumor cells adhere and cross into blood stream the metastasis could happen elsewhere. Interestingly, treatment of heparin could inhibit tumor metastasis. Inhibition of tumor growth is possibly due to exogenous heparin competing to binding of p-selectin and epithelial cell [127]. Combined with other inhibitory effects, heparin was reported able to slow down the tumor progression clinically.

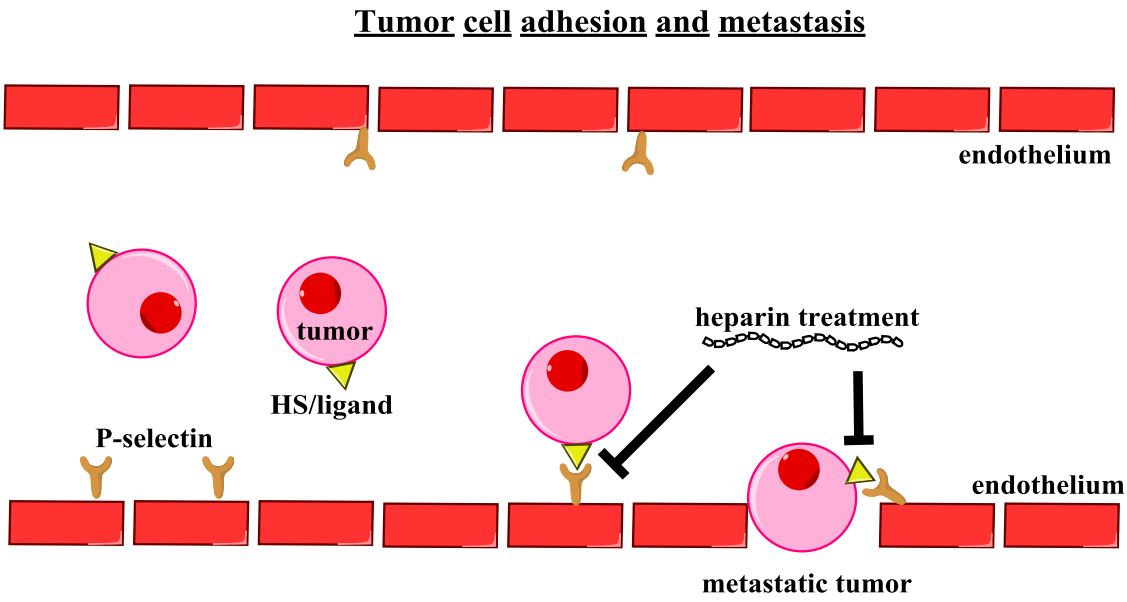


Figure 22. Role of HS and heparin in mediating of tumor progression. Tumor cells utilize HS to adhere to the surface of endothelial cells via p-selectin. Once tumor cells adhere and cross into blood stream the metastasis could happen elsewhere. Treatment of heparin may inhibit tumor adhesion and metastasis by competing binding of p-selectin.

The expression of heparanase is upregulated in several tumor cells [128, 129]. HS within extracellular matrix is suggested capable to sequester FGFs to reduce tumor growth. However, tumor cells upregulated secret heparanase to degrade the HS and to reduce the capability sequestering tumorigenic growth factors [130]. To counter the effect of heparanase, exogenous treatment of HS or heparin may be clinically used to inhibit the cancer growth by sequestering the growth factors. The structural difference of HS sequence for stimulated or inhibitory effect on tumor growth still remains unclear. Homogeneous HS is therefore urgently required to clarify this critical issue of structure-activity relationship. Another promising direction of drug design is to develop anticancer HS-mimetic reagent without anticoagulation effect as concern of unexpected side effect.

Inflammation

Heparan sulfate is a key regulator during the process of inflammation. Inflammation is a physiological response to some malicious scenarios including cancer, skin injury, infection, and so on [131, 132]. Similar to cancer progression, heparanase is upregulated during inflammation [133-135]. The upregulated heparanase degrade the heparin and result in the loss of ability sequestering the inflammatory factors. To counter the heparanase effect, the treatment of exogenous heparin is postulated to attenuate the inflammation.

Regarding to skin injury derived inflammation, HS is critical interacting with several chemoattractant such as chemokines and cytokine [136-138]. Chemokine and cytokine released by injury are bound by negatively charged HS, and subsequently attract neutrophil rolling along vessel (Figure 23) [139]. The L-selectin on the surface of neutrophil is attracted and slowed down the rolling by interaction with this chemoattractant-attached HS [140-142]. The trapped neutrophil further release chemoattractant and recruit more inflammatory responses.

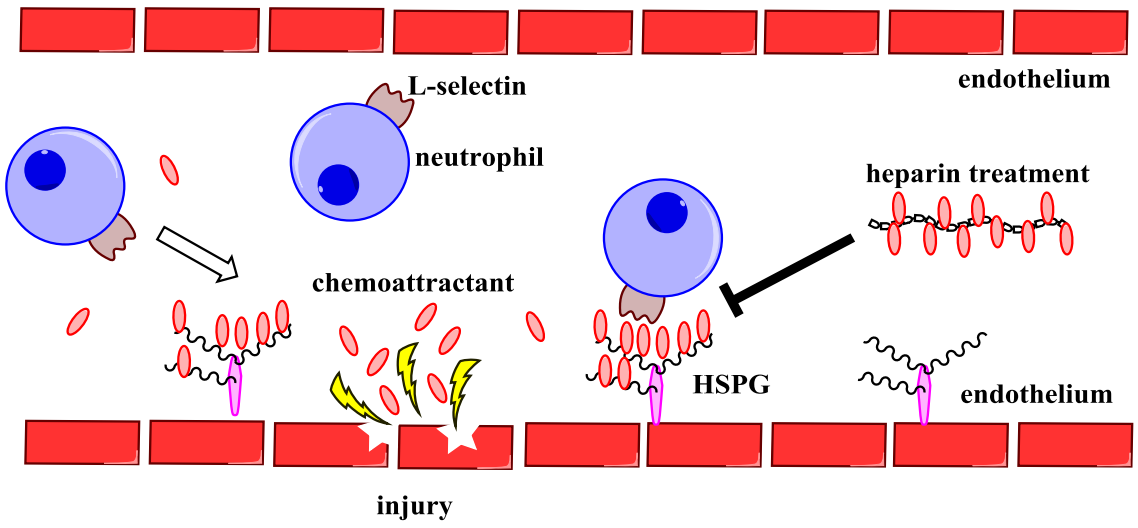


Figure 23. Role of HS and heparin in mediating of inflammation. Chemoattractants including chemokine and cytokine released by injury are bound by negatively charged HS, and subsequently attract neutrophil rolling along vessel. The L-selectin on the surface of neutrophil is attracted and slowed down the rolling by interaction with this chemoattractant-attached HS. Heparin is potentially useful to control inflammation by competing endogenous HS binding to chemoattractant and reducing neutrophil trafficking at the injury site.

HS polysaccharide chain containing several unsubstituted glucosamine residues ($GlcNH_2$) was revealed may interact with L-selectin [143]. Unsubstituted glucosamine is relatively sparse saccharide among heparan sulfate. Possibly this special saccharide play the role to distinguish various chemoattractant and achieve the desired inflammation response. Heparin is potentially useful to control inflammation such as ulcerative colitis, bronchial asthma, and skin injury [144-147]. The explanation is exogenous heparin can compete endogenous HS binding to chemoattractant, and decrease neutrophil trafficking at the injury site bringing further inflammatory response. Additionally, 6-*O*-sulfo group on HS is required for the inhibitory effect [140]. The electrostatic interaction between positively charged chemoattractant and highly negatively charged heparin may drive the binding. However, to develop the anti-inflammatory reagent without anticoagulation effect demands more detail understanding of relation between structure and biofunction.

Viral Infection

HS on cell surface play critical role in mediating multiple virus infection such as human immunodeficiency virus (HIV), hepatitis C, human papillomavirus (HPV), and herpes simplex virus-I (HSV-I) [148-150]. HS is essential to interact with several viral glycoproteins for internalization of the virus into the host cell [53, 151].

The importance of HS for HSV-I infection is one of the most well-studied case [152, 153]. HSV-I infection can generally separate to two phase: attach host cell, and entry host cell. Firstly, attachment of HSV-I requires viral envelop glycoprotein gB and gC to interact with HS-attached host cell [154-156]. The HS responsible for HSV-I attachment is called binding receptor with the sequence of repeated IdoA2S-GlcNS6S more than decasaccharide [157]. Secondly, 3-*O*-sulfate HS serve as entry receptor for viral entry [53, 158]. gB, gH, gL, and the most important gD are used to recognize entry receptors and fuse envelope into host membrane [159-161]. Instead of 3-OST-1 make the AT binding domain, 3-*O*-sulfation of this entry receptor HS was reported made by 3-OST-3 [53]. The implication of distinct 3-*O*-sulfation performed by isoforms suggests that different modified HS may be in charge of different biological functions.

Since 3-*O*-sulfated HS is principal for virus infection, design of 3-*O*-sulfated HS mimic is possibly a good candidate structure for development of anti-viral drug (Figure 24). Accordingly, heparin, which contains scattered 3-*O*-sulfo group, is proven preventing HSV-I infection [151]. An octasaccharide with the sequence of UA2S-GlcNS-IdoA2S-GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S also shows IC50 as 40uM for inhibition of HSV-I infection [162]. The various 3-*O*-sulfated HS structure synthesized by 3-OST isoform may be another possible approach to develop new anti-viral therapy.

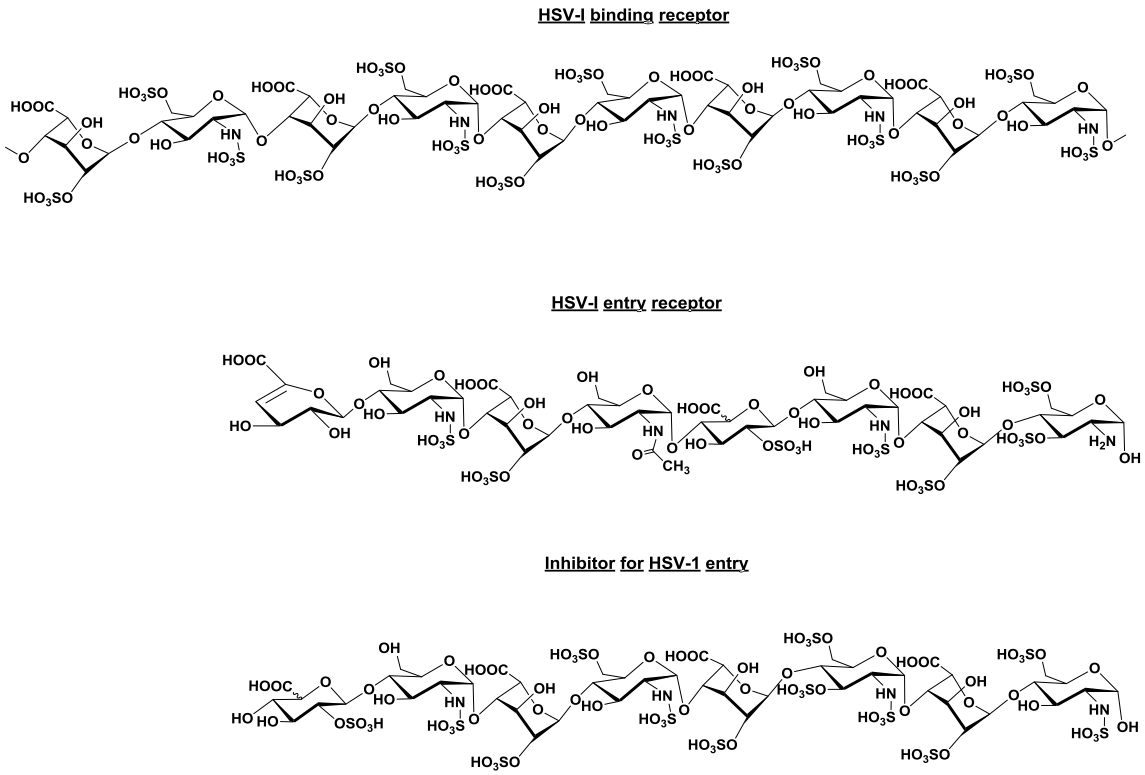


Figure 24. Chemical structure of HSV-I binding and entry receptor as well as inhibitor. The HS responsible for HSV-I attachment is called binding receptor with the sequence of repeated IdoA2S-GlcNS6S more than decasaccharide. 3-*O*-sulfate HS was isolated as entry receptor for viral entry. 3-*O*-sulfated HS mimic, an octasaccharide with the sequence of UA2S-GlcNS-IdoA2S-GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S, shows IC50 as 40uM for inhibition of HSV-I infection.

Statement of Problem

Heparin is a widely used clinical anticoagulant since the 1930s. Heparin was the first biopolymeric drug and one of the few carbohydrate based drugs. The polydispersity of the saccharide sequence as well as the polyheterogeneity of the sulfation pattern make study of heparin and heparan sulfate formidable tasks. The access of sufficient structural defined samples for study is extremely difficult. In this dissertation, we directly address three main issues of heparan sulfate: safeguard drug safety, probe interaction with proteins, and characterize the HS conformations.

To safeguard heparin drug safety, reversible anticoagulant activity and secured heparin components are two important concerns. Low-molecular-weight heparin (LMWH) is a mixture of oligosaccharides (MW_{avg} 3500 to 6000 Da) and the most widely prescribed heparin in the US. However, LMWH is incompletely neutralized with antidote protamine, thereby increasing the risks of bleeding. A structural defined homogeneous LMWH with reversible activity is highly expected in clinical usage. Reliable structural characterization of heparin components is also critical to safeguard the quality of heparin drugs. In 2008, batches of contaminated heparin killed 94 patients in US because adulteration of structurally similar carbohydrate was escaped from government supervision. Therefore, a library of common and less abundant heparin oligosaccharides were synthesized for structural referencing to secure the components of heparin drug.

HS achieves its functions by interacting with a variety of proteins. There is strong a demand for probes to decipher different interactions of specific HS structures with their protein targets. Several of problems are awaiting probes to deal with. Firstly, heparan sulfate

sulfotransferase (2OST) is a key enzyme in the HS biosynthetic pathway; however, the cellular mechanism regulating HS biosynthesis is largely unknown. Secondly, the positions of sulfo groups and the locations of the IdoA residues are critically important for the protein binding specificities, but lack of the probes to selectively target different HS biosynthetic enzymes. Thirdly, HS repeating saccharide or bound by large protein make the NMR signals overlapped and indistinguishable, therefore a tool able to study the specific saccharide and detail picture of HS-protein interaction is in demand. In the dissertation, we developed different kinds of probes to solve these issues.

It has been proposed that the conformation of HS plays a critical role in binding affinity and selectivity toward proteins. In addition to the length of saccharide chain, diversity of sulfation patterns and plasticity of iduronic acid (IdoA) are known possibly associated to the HS conformational change. The flexible glycosidic linkages and IdoA facilitate heparan sulfate conformational rearrangement to allow specific proteins binding. Nevertheless, this hypothesis remains to be unproven due to the lack of a set of compounds that is able to systematically investigate the effect of different sulfation on HS conformation. Unfortunately due to highly flexible character, carbohydrate is difficult to crystalized and NMR-analyzed. Thus an array of various structural defined HS oligosaccharides as well as an innovative method capable to accurately quantitate the saccharide conformations is developed in this study.

Last but not least, the efficient synthetic approach and accurate characterization of the HS oligosaccharide is crucial for all the studies. Understanding the molecular mechanisms of HS biosynthetic enzymes and interaction proteins should greatly aid in the design and production of heparin-based molecules. The success of these projects will assist the development of therapeutics tailored for specific disease targets.

Chapter II

Materials and Methods

Preparation of Heparan Sulfate Biosynthetic Cofactors

Saccharide Donor

UDP-GlcA and UDP-GlcNAc are commercial available products from Sigma-Aldrich while UDP-GlcNTFA is made by chemoenzymatic approach. The preparation of UDP-GlcNTFA was started from glucosamine (Sigma-Aldrich), which was first converted to GlcNTFA by reacting with *S*-ethyl trifluorothioacetate (Sigma-Aldrich) followed the protocol described previously [163]. The resultant GlcNTFA was converted to GlcNTFA-1-phosphate using *N*-acetylhexosamine 1-kinase [164]. The plasmid expressing *N*-acetylhexoamine 1-kinase was a generous gift from Prof. Peng Wang (Georgia State University), and the expression of the enzyme was carried out in *E. coli* as reported [164]. The UDP-GlcNTFA synthesis was completed by transforming GlcNTFA-1-phosphate using glucosamine-1-phosphate acetyltransferase/*N*-acetylglucosamine-1-phosphate uridyltransferase (GlmU) as described [163]. The resultant UDP-GlcNTFA was ready for the elongation reaction involved in using KfiA as described below. Because the impurities in the reaction mixture used to synthesize UDP-GlcNTFA do not interfere with the reactions involved in KfiA, purification of UDP-GlcNTFA was not conducted. The successful synthesis of UDP-GlcNTFA was

confirmed by PAMN-HPLC analysis as a distinct UV 260 nm peak. The intensity of the peak was also used to estimate the amount of UDP-GlcNTFA.

Chemoenzymatic Synthesis of Heparan Sulfate Oligosaccharides

KfiA and pmHS$_2$ for Backbone Elongation

The elongation starts from a commercial material, GlcA-pnp, followed by elongation of GlcNTFA and GlcA with number of desired repeating times. Elongation of GlcA-pNP requires using two bacterial glycosyl transferases, *N*-acetyl glucosaminyl transferase of *E. coli* K5 (KfiA) and heparosan synthase-2 (PmHS2) from *Pasteurella multocida.* To introduce a GlcNTFA residue, GlcA-pnp (1.2 mM) was incubated with KfiA (20 μg/ml) in a buffer containing Tris (25 mM, pH 7.5), $MnCl_2$ (15 mM) and UDP-GlcNTFA (1.5 mM), at room temperature overnight. To introduce a GlcA residue, disaccharide substrate, GlcNTFA-GlcA-pnp (1.2 mM), was incubated with pmHS2 (20 μg/ml) in a buffer containing Tris (25 mM, pH 7.5), $MnCl_2$ (15 mM) and UDP-GlcA (1.5 mM), at room temperature overnight. C_{18} column (0.75 × 20 cm) was used for purification with gradient elution (0-100% acetonitrile in H_2O, 0.1% TFA, 2 ml/min in 60 min). The eluted products were monitored by UV absorbance at 310 nm and characterized by ESI-MS. The additional GlcNTFA and GlcA elongation was repeated under the same condition.

GlcNTFA Deacetylation and Subsequent Modification with NST

Detrifluoroacetylation was completed using 0.1M LiOH at 0°C for two hours. The products were analysis by PAMN-HPLC (polyamine-based high performance liquid

chromatography) and electrospray ionization mass spectrometry (ESI-MS). Upon the completion of detrifluoroacetylation, the pH of the reaction mixture was neutralized to pH 7.0. Afterwards, *N*-sulfation was performed within pH 7.0 MES (50 mM), *N*-sulfotransferase (10 μg/ml), and PAPS (0.5 mM, 1.5 equivalent of substrate amount) at 37°C overnight. The sulfated products were purified by Q-Sepharose column (15 × 200 mm, GE Health Sciences) with linear gradient elution (20–100% 1 M NaCl in 20 mM NaOAc, pH 5.0, flow rate 2 ml/min) and followed by dialysis to obtain highly targeted oligosaccharides or intermediates.

C_5-Epi and 2-OST Modification

Epimerization and 2-*O*-sulfation were performed within the solution contained MES (50 mM, pH 7.0), $CaCl_2$ (2 mM), C_5-epi (10 μg/ml), 2-OST (10 μg/ml), PAPS (0.2 mM), and oligosaccharide intermediate (0.13 mM) at 37^{o}C overnight. The products were analysis by PAMN-HPLC (polyamine-based high performance liquid chromatography) and electrospray ionization mass spectrometry (ESI-MS). The sulfated products were purified by Q-Sepharose column (15 × 200 mm, GE Health Sciences) with linear gradient elution (20–100% 1 M NaCl in 20 mM NaOAc, pH 5.0, flow rate 2 ml/min) and followed by dialysis to obtain highly targeted oligosaccharides or intermediates.

6-OST-1, 6-OST-3 and 3-OST Modification

General 6-*O*-sulfation was performed using PAPS (400 μM), MES (50 mM, pH 7.0), 6-OST-1 (0.1 mg/ml), 6-OST-3 (0.1 mg/ml) overnight at 37°C. For some naturally rare structures,

sometimes 6-*O*-sulfation was pushed in higher concentration of PAPS (800 µM), MES (50 mM, pH 7.0), 6-OST-1 (0.6 mg/ml), 6-OST-3 (0.6 mg/ml) overnight at 37°C. The reaction was monitored by DEAE-HPLC to ensure the desired product reached to the highest level and the substrate was consumed. The reaction mixture was purified by Q-Sepharose column and dialyzed.

3-*O*-sulfation was performed to synthesize constructs under the condition of MES (50 mM, pH 7.0), $MnCl_2$ (10 mM), $MgCl_2$ (5 mM), 3-OST-1 (10 µg/ml) or 3-OST-5 (10 µg/ml), PAPS (600 µM) at 37^{o}C overnight. The extent reaction was also monitored by DEAE-HPLC. The products were purified by Q-Sepharose column to obtain targeted oligosaccharides.

Preparation of ^{35}S-Labeled Heparan Sulfate

UFH (from US Pharmacopeia) and enoxaparin (from local pharmacy) were modified by NST. Reaction consisted of MES (2-(*N*-morpholino)ethanesulfonic acid, Sigma) 50 mM pH 7.0, NST 0.1 mg ml^{-1} and 0.5 nmol [^{35}S]PAPS (specificity activity of [^{35}S]PAPS was 2.2×10^4 cpm pmol^{-1}), 50 µg of UFH or enoxaparin in total 500 µl at 37°C overnight. The products were purified by a DEAE-column.

^{35}S-labeled synthetic constructs were prepared from the synthetic LMWH constructs intermediates, which are followed by further any sulfation using [^{35}S]PAPS**.** Reaction consisted of MES 50 mM pH 7.0, 10 mM $MnCl_2$, 5 mM $MgCl_2$, any sulfotransferase 0.1 mg ml^{-1} and 0.5 nmol [^{35}S]PAPS and oligosaccharide (5 µg) in total 500 µl at 37°C overnight. The ^{35}S-labeled construct was purified by a DEAE-HPLC column.

Purification of Synthetic Heparan Sulfate

C_{18} and Q Sepharose Fast Flow^{TM}-HPLC Purification and Analysis

C_{18} column (0.75 × 20 cm; Biotage) was used for purification with gradient elution (0-100% acetonitrile in H_2O, 0.1% TFA, 2 ml/min in 60 min). The eluted products were monitored by UV absorbance at 310 nm and characterized by ESI-MS.

Purification of oligosaccharides as well as intermediates was conducted by a fast flow Q-Sepharose column (15 × 300 mm; GE Health Care), which was eluted with a linear gradient of 20–100% 1M NaCl in 20 mM NaOAc at pH 5.0 in 2 h at a flow rate of 2 ml/min^{-1}. Purification of those highly sulfated oligosaccharide, was eluted with a linear gradient of 30–100% 2M NaCl in 20 mM NaOAc at pH 5.0 in 2 h at a flow rate of 2 ml/min^{-1}.

HPLC Analysis

Both DEAE-HPLC and polyamine-based anion exchange (PAMN)-HPLC were used to analyze the purity of the products. The elution conditions for the HPLC analysis were described elsewhere [165]. Briefly, for DEAE-HPLC method, the DEAE-NPR column ((0.46 × 7.5 cm, Tosohaas) was eluted with a linear gradient of NaCl in 20 mM sodium acetate buffer (pH 5.0) from 0 to 1 M in 60 min at a flow rate of 0.4 ml/min. For PAMN-HPLC, the column (silica-based polyamine column, 0.45 × 25 cm, YMC) was eluted with a linear gradient of KH_2PO_4 from 0 to 1 M in 40 min at a flow rate of 0.5 ml/min. GPC-HPLC was used to analyze and protein purify and protein-oligosaccharide conjugates. The GPC-HPLC conditions were as

follows: the column G3000SW from Tosoh Bioscience (7.8 mm × 300 mm, 5 μm) the elution was isocratic; the buffer consisted of 25 mM MOPS and 400 mM NaCl; the dual wavelengths were set as 280 nm and 310 nm respectively, the flow rate was 0.2 mL/min.

Structural Characterization of Synthetic Heparan Sulfate

NMR Parameter and Spectroscopy Analysis

The structure of the oligosaccharides constructs and intermediates were analyzed by NMR experiments, including 1D- ("s2pul" pulse sequence ^{1}H or ^{13}C), 2D- (^{1}H-^{1}H "gCOSY" pulse sequence COSY, "TOCSY" pulse sequence TOCSY, ^{1}H-^{13}C "gHSQC" pulse sequence HSQC, "gHMBC" pulse sequence HMBC, and "gHSQCTOXY" pulse sequence HSQC-TOCSY) NMR. NMR experiments were performed at 298 K on a Varian Inova 500 MHz spectrometer equipped with 5mm triple resonance XYZ or broadband PFG probe and processed by VnmrJ 2.2D software. Chemical shifts are referenced to external 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS, Sigma, Co.). Deuterated EDTA (Sigma, Co.) was added to resolve the coupling constants of IdoA2S signals. Samples (5.0 to 10.0 mg) were each dissolved in 0.5 mL D_2O (99.994%, Sigma, Co.) and lyophilized three times to remove the exchangeable protons. The samples were re-dissolved in 0.5 mL D_2O and transferred to NMR microtubes (OD 5 mm, Norrell). 1D ^{1}H-NMR experiments were performed with 256 scans and an acquisition time of 768 msec. 1D ^{13}C-NMR experiments were performed with 40000 scans, 1.0 sec relaxation delay, and an acquisition time of 1000 msec. 2D (^{1}H-^{1}H COSY, TOCSY, ^{1}H-^{13}C HSQC, HSQC-TOCSY) spectra were recorded with GARP carbon decoupling, a 1.5 sec relaxation delay, a 204 msec acquisition time with 500 increments for 48 scans. Forty eight steady state scans were used

prior to the start of acquisition. 2048 points were collected in f2. Gradients were used for coherence transfer selection with sensitivity enhancement in the HSQC experiment. ^{13}C transmitter offset was set at 77.0 ppm. The echo-anti-echo phase cycling scheme was used. The polarization transfer delay was set with a $^{1}J_{C\text{-}H}$ coupling value of 147Hz. 2D ^{1}H-^{13}C HMBC experiments were performed with 72 scans, 1.5 sec relaxation delay, and a 204 msec acquisition time. The delay for evolution of long-range couplings was set with J_{lr} of 7.6 Hz.

NMR experiments were also performed at 298 K on Bruker Avance II 800 MHz spectrometer with Topsin 2.1 software. Samples (3.0 to 6.0 mg) were each dissolved in 0.5 ml D_2O (99.996%, Sigma-Aldrich) and lyophilized three times to remove the exchangeable protons. The samples were re-dissolved in 0.4 ml D_2O and transferred to NMR microtubes (O.D. 5 mm, Norrell). 1D ^{1}H NMR experiments were performed with 256 scans and an acquisition time of 850 msec. 2D ^{1}H-^{1}H COSY experiments were performed with 16 scans, 1.5 sec relaxation delay, and 500 msec acquisition time. 2D ^{1}H-^{13}C HMQC experiments were performed with 16 scans, 1.5 sec relaxation delay, and 250 msec acquisition time.

MS Spectrometry Analysis

The low-resolution MS analyses were performed at a Thermo LCQ-Deca. A syringe pump (Harvard Apparatus) was used to introduce the sample by direct infusion (35 μl min^{-1}). Experiments were carried out in negative ionization mode. Synthetic non-sulfated oligosaccharides were diluted in 200 μl of H_2O, the electrospray source set is 5 KV and 275 °C. Sulfated LMWHs were diluted in 200 μl of 10 mM ammonium bicarbonate, the electrospray source set is 3 KV and 200 °C. Sulfated oligosaccharide (1 μl) was diluted in a different working solution containing 200 μl of 10 mM ammonium bicarbonate. Experiments for sulfated

oligosaccharides were carried out in negative ionization mode with the electrospray source set to 2 KV and 200 °C. The automatic gain control was set to 1×10^7 for full scan MS. The MS data were acquired and processed using Xcalibur 1.3.

High resolution ESI-MS analysis was conducted on Thermo LTQ XL Orbitrap (Breman, Germany) under the following conditions. A Luna hydrophilic liquid interaction chromatography (HILIC) column (2.0×150 mm^2, 200 Å, Phenomenex, Torrance, CA) was used to separate the oligosaccharide mixture. Mobile phase A was 5 mM ammonium acetate prepared with high performance liquid chromatography (HPLC) grade water. Mobile B was 5 mM ammonium acetate prepared in 98% HPLC grade acetonitrile with 2% of HPLC grade water. After injection of 8.0 µL oligosaccharide mixture (1.0 µg µl^{-1}) through an Agilent 1200 autosampler, HPLC binary pump was used to deliver the gradient from 10% A to 35% A over 40 min at a flow rate of 150 µl min^{-1}. The LC column was directly connected online to the standard electrospray ionization source of LTQ-Orbitrap XL Fourier transform (FT) mass spectrometer (MS) (Thermo Fisher Scientific, San-Jose, CA). The source parameters for FT-MS detection were optimized using fondaparinux to minimize the insource fragmentation and sulfate loss and maximize the signal/noise in the negative-ion mode. The optimized parameters included a spray voltage of 4.2 kV, a capillary voltage of −40 V, a tube lens voltage of −50 V, a capillary temperature of 275° C, a sheath flow rate of 30, and an auxiliary gas flow rate of 6. External calibration of mass spectra routinely produced a mass accuracy of better than 3 ppm. All FT mass spectra were acquired at a resolution 60,000 with 300−2000 Da mass range.

Disaccharide Analysis

To determine the size of the [^{35}S]sulfation, the sample was subjected to nitrous acid degradation at pH 1.5 to produce ^{35}S-labeled disaccharide as described by Shively and Conrad [166]. The resultant ^{35}S-labeled disaccharides were resolved using a C_{18} reverse-phase column (0.46 × 25 cm; Vydac) under reverse-phase ion-pairing HPLC conditions [167]. The identities of the disaccharides were determined by co-elution with the appropriate ^{35}S-labeled standards.

Molecular Modeling

Oligosaccharide Pre-processing

The hexasaccharides were created with the GLYCAM06 (version j-1) parameter set [168], neutralized with NA^+ ions, and solvated by the TIP5P water model in a truncated octahedral box 12 Å around the ligand [169]. Ensemble charges were developed for the Glucosamine ($GlcNH_2$) residue according to the standard GLYCAM protocol [168]. Briefly, electrostatic potentials were calculated by GAUSSIAN09 at the HF/6-31G(*) level of theory for 100 snapshots at 0.5 ns intervals (50 ns total). Partial atomic charges were computed using the RESP module of AMBER14 by fitting to these electrostatic potentials with a restraint weight of 0.01 [170].

Energy Minimization

Energy minimizations and MD simulations were performed using the *pmemd.cuda* module from AMBER14 [171]. The system was minimized using the steepest descent method for

the initial 1000 cycles before switching to conjugate gradient for the remaining 24,000 cycles. Three minimizations were performed with varying atomic restraints. Initially, cartesian restraints (10 kcal/mol $Å^2$) were applied to each atom of the hexasaccharide. The second stage involved restraining the ring atoms. All restraints were removed for the final minimization. During the remaining simulations, internal restraints were applied to the iduronic acid residue in order to maintain a $^1C_4$, ^2So, or $^4C_1$ conformation. Details of these restraints are provided in Table 2.

All Rk2=300 Rk3=300				
$^1C_4$	R1	R2	R3	R4
O5-C1-C2-C3	-74.6	-64.6	-44.6	-34.6
C1-C2-C3-C4	33.6	43.6	63.6	73.6
C2-C3-C4-C5	-72.9	-62.9	-42.9	-32.9
C3-C4-C5-O5	34.1	44.1	64.1	74.1
C4-C5-O5-C1	-77.9	-67.9	-47.9	-37.9
C5-O5-C1-C2	36.4	46.4	66.4	76.4
^2S_o	R1	R2	R3	R4
O5-C1-C2-C3	9.7	19.7	39.7	49.7
C1-C2-C3-C4	-77.1	-67.1	-47.1	-37.1
C2-C3-C4-C5	-1.6	8.4	28.4	38.4
C3-C4-C5-O5	24.1	34.1	54.1	64.1
C4-C5-O5-C1	-98.4	-88.4	-68.4	-58.4
C5-O5-C1-C2	14.5	24.5	44.5	54.5
$^4C_1$	R1	R2	R3	R4
O5-C1-C2-C3	27.2	37.2	57.2	67.2
C1-C2-C3-C4	-67.7	-57.7	-37.7	-27.7
C2-C3-C4-C5	35.1	45.1	65.1	75.1
C3-C4-C5-O5	-79.3	-69.3	-49.3	-39.3
C4-C5-O5-C1	45.1	55.1	75.1	85.1
C5-O5-C1-C2	-77.2	-67.2	-47.2	-37.2

Table 2. Internal restraints applied for iduronic acid during MD simulation. The IdoA ring conformations were restrained by limiting the rotational freedom of each dihedral of the 6-membered ring. The restraint is imposed by an energy potential in the shape of a flat-welled parabola which becomes linear past specified torsion angles (R1 and R4). The R2 and R3 parameters define the points where the parabola flattens. The force constants (RK2 and RK3) were set at 300 kcal/mol for each ring conformation.

Molecular Dynamics Simulation

MD simulations were performed with *pmemd.cuda* from AMBER14 [171]. Electrostatic interactions were treated with the Particle-Mesh Ewald algorithm [172]. An 8 Å cutoff for non-bonded interactions. The SHAKE algorithm was included in order to constrain hydrogen-containing bonds, enabling an integration time step of 2 fs. The system was heated to 300 K under NVT conditions over 60 ps by employing the Berendsen thermostat with a coupling time constant of 1 ps, and allowed to equilibrate for a total of 1 ns under NPT conditions. A post-equilibration data set was collected for 1 μs, also under NPT conditions.

Simulation Data Analysis

Structural metrics (i.e. torsion angles, hydrogen bond properties, r.m.s.d. values, and coordinate averaging) were generated using the cpptraj module of AmberTools14 [173]. The Φ- and Ψ-glycosidic torsion angles were defined by the H1-C1-O4-C4 and C1-O4-C4-H4 atomic sequences, respectively. Theoretical J-coupling values were calculated by collecting the H1-H2, H2-H3, H3-H4, and H4-H5 torsion angles for the Iduronic acid ring for the three ring conformations. An average torsion angle from 8 μs of data, one μs per compound, was transformed into a J-coupling value via the reparameterized Haasnoot Karplus equation- $14.63*\cos2(\varphi)-0.78*\cos(\varphi)+0.60+\sum\lambda i\{0.34-2.31\cos2[\sin(\varphi)+18.4|\lambda i|]\}$ [174, 175].

Biological Activities of Synthetic Heparan Sulfate

Activity Measurement for 2-O-Sulfotransferase

Activity measurement for 2-OST using the completely desulfated *N*-sulfated heparin (CDSNS, purchased from Neoparin) substrate was determined by incubating 0.5 μg of 2-OST with 1 μL of CDSNS (10 mg/mL) and 1~5 $\times 10^5$ cpm of [^{35}S]PAPS in 200 μL of buffer containing 50 mM MES(2-(*N*-morpholino)ethanesulfonic acid) (pH 7.2), 10 mM $CaCl_2$ at 37 ^{0}C for 1 h. Reactions were quenched by the addition of 6 M urea, 1mM EDTA (ethylenediaminetetraacetic acid), 0.01% Triton X-100, then subjected to 200-μL DEAE-Sepharose (from Sigma) chromatography to purify the ^{35}S-labeled polysaccharide product. The quantity of [^{35}S]sulfated polysaccharide was determined by liquid scintillation counting. The negative control contained all of the components with the exception of 2-OST enzyme.

To analyze the inhibition effect of test compound on the activity of 2-OST enzyme, 1 μg of 2-OST was deactivated by incubating with various amount (0, 1, 2, 4, 6 and 10 μL) of test compound (5 μM) at pH 5.5 to initiate the inhibition (cross-linking) reaction in 100 μL of 25 mM MES buffer. After 1 h incubation, the inhibition effect of compound on *2-OST* was determined by incubating 10 μL of the deactivated 2-OST protein above, 10 μg of CDSNS heparin and 1~5 $\times 10^5$ cpm of [^{35}S]PAPS in 100 μL of a buffer containing 50 mM MES (pH 7.2), 10 mM $CaCl_2$. All the reactions were incubated at 37 ^{0}C for 1h and quenched by the addition of a buffer containing 6 M urea, 1mM EDTA, and 0.01% Triton X-100. The samples were then loaded onto a 200-μL DEAE-Sepharose (Sigma-Aldrich) column to purify the [^{35}S] CDSNS heparin, which were eluted with 1M NaCl and 0.01% Triton X-100. The quantity of [^{35}S] HS was determined by liquid scintillation counting. The inhibition activity of test compound against 2-OST was showed

by IC_{50} (half maximal inhibitory concentration) value that was calculated by linear regression analysis of the concentration-response curves by means of Sigmaplot 12.5.

Activity Measurement for 3-O-Sulfotransferase

Activity measurement for *3*-OST using the heparan substrate (isolated from bovine kidney) was determined by incubating 0.5 μg of 3-OST with 1 μL of HS (10 mg/mL) and 1~5 $\times 10^5$ cpm of [^{35}S]PAPS in 200 μL of buffer containing 50 mM MES (pH 7.2), 10 mM $MnCl_2$ and 5 mM $MgCl_2$ at 37 ^{0}C for 1 h. Work-up of reaction mixture and activity measurement followed essentially the same procedure as described for 2-OST.

To analyze the inhibition effect of test compound on the activity of 3-OST enzyme, 3-OST (1 μg) was deactivated by incubating with various amount (0, 0.5, 1.0, 1.5 and 2.0 μL) of test compound (2.0 μM) at pH 5.5 to initiate the inhibition (cross-linking) reaction in 100 μL of 25 mM MES buffer. After 1 h incubation, the inhibition effect of compound on 3-OST was determined by incubating 10 μL of the deactivated 3-OST protein above, 10 μg of HS and 1~5$\times 10^5$ cpm of [^{35}S]PAPS in 100 μL of a buffer containing 50 mM MES (pH 7.2), 10 mM $MnCl_2$, 5 mM $MgCl_2$. All the reactions were incubated at 37 ^{0}C for 1h and quenched by the addition of a buffer containing 3 M urea, 1mM EDTA, and 0.01% Triton X-100. The samples were then loaded onto a 200-μL DEAE-Sepharose (Sigma-Aldrich) column to purify the [^{35}S]HS, which were eluted with 1M NaCl and 0.01% Triton X-100. The quantity of [^{35}S] HS was determined by liquid scintillation counting. The inhibition activity of test compound against 3-OST was showed by IC_{50} value that was calculated by linear regression analysis of the concentration-response curves by means of Sigmaplot 12.5.

In vitro and ex vivo Anti-FXa Activity

Assays were based on a previously published method [176, 177]. Briefly, human factor Xa (FXa) (Enzyme Research Laboratories, South Bend, IN) was diluted to 50 U ml^{-1} with PBS. The chromogenic substrates, S-2765 was from Diapharma (Westchester, OH) and made up at 1 mg ml^{-1} in water. UFH (from US Pharmacopea), enoxaparin (Lovenox® from local pharmacy) and synthetic oligosaccharide was dissolved in PBS at various concentrations (3 to 600 µg ml^{-1}). The reaction mixture, which consisted of 20 µl of human plasma (Sigma-Aldrich) and 8 µl of the solution containing the sample, was incubated at room temperature for 5 min. Factor Xa (100 µl) was then added. After incubating at room temperature for 4 min, 30 µl of S-2765 substrate was added. The absorbance of the reaction mixture was measured at 405 nm continuously for 5 min using spectrophotometer. The absorbance values were plotted against the reaction time to measure the reaction rate. The initial reaction rates were used to measure the activity of FXa.

Binding Affinity to Antithrombin

^{35}S-labeled constructs were prepared via any sulfation as described previously, using [^{35}S]PAPS (0.5 nmol, specific activity 2.2×10^4 cpm/pmol) and oligosaccharide precursor (5 µg) in total 500 µl at 37°C overnight. The products were purified by DEAE-HPLC column. Affinity co-electrophoresis [178] was used for measurement of dissociation constant (K_d) of oligosaccharides. ^{35}S-labeled constructs (1500-2500 cpm) were loaded for lanes with AT concentration at 0, 8, 16, 32, 60, 120, 250, 500, and 1000 nM. The agarose gel was run with 300 mA electrophoresis for two hours, and was dried at room temperature overnight. The dried gel

was analyzed by PhosphoImager (Amersham Biosciences, Storm 860). The retardation coefficient was calculated at $R = (M_0-M)/M_0$, where M_0 is the mobility of the oligosaccharide flow through in the lane without AT, and M is the mobility of the sample in an individual lane. The retardation coefficient was then plotted against the retardation coefficient divided by its respective AT concentration. The constant $-1/K_d$ is the slope of the trendline.

Chapter III

Safeguard Heparin Safety:

Reversible Anticoagulant Activity and a Library of Standards[1]

Synthesis of Low Molecular Weight Heparins (LMWH)

Introduction

Heparin is a widely used anticoagulant to prevent and treat arterial and venous thrombosis [179]. There are three approved forms of heparin: unfractionated heparin (UFH, average molecular weight (MW_{avg}) ~14000), LMWH (MW_{avg} 3500 to 6000) and fondaparinux (MW 1508). UFH has a rapid onset action, is safe for renal-impaired patients, and is reversible using the cationic peptide drug, protamine[180], but shows a 1-6% incidence of heparin-induced thrombocytopenia (HIT), a life-threatening complication[181]. Introduced in the 1990's[182], LMWHs are *subcutaneously* administered and have a longer half-life than UFH, permitting their outpatient use and/or self-administration. Because of these advantages, LMWH is the most widely prescribed heparin in the US[183]. However, LMWH can only be used in renal-impaired patients at reduced doses[184] and is incompletely neutralized with protamine, thereby

[1]Part of this chapter previously appeared as two articles in the Journal of Nature Chemical Biology and Glycobiology. The original citations are as follows: Y. Xu, C. Cai, K. Chandarajoti, P. H. Hsieh, L. Li, T. Pham, E. Sparkenbaugh, J. Sheng, N. Key, R. Pawlinski, E. Harris, R. Linhardt, J. Liu, "Homogeneous low-molecular-weight heparins with reversible anticoagulant activity," *Nature Chemical Biology*, 10, 2014, 248–250; P. H. Hsieh, Y. Xu, D. A. Keire, J. Liu, "Chemoenzymatic synthesis and structural characterization of 2-O-sulfated glucuronic acid-containing heparan sulfate hexasaccharides," *Glycobiology*, 24, 2014, 681–692

increasing the risks of bleeding. Fondaparinux, a synthetic pentasaccharide, is *subcutaneously* bioavailable and has reduced risks of HIT and osteoporosis[185]. However, it is primarily excreted through the kidney, and thus, is contraindicated in renal-impaired patients[184], and lacks an antidote. The US Food and Drug Administration (FDA) recently approved generic forms of LMWH and fondaparinux, underscoring the rapid growth in the development of heparin-based drugs.

Heparin consists of a disaccharide-repeating unit of either iduronic acid (IdoA) or glucuronic acid (GlcA) and glucosamine (GlcN) residues, each capable of carrying sulfate groups. The locations of sulfate groups, IdoA and GlcA dictate the anticoagulant activity of heparin[186]. *In vivo*, heparin is synthesized by a series of heparan sulfate (HS) biosynthetic enzymes. HS polymerase catalyzes the formation of the polysaccharide backbone, a repeating disaccharide of GlcA and *N*-acetylated glucosamine (GlcNAc). This backbone is modified by *N*-deacetylase/*N*-sulfotransferase (NDST), C_5-epimerase (C_5-epi), 2-*O*-sulfotransferase (2-OST), 6-*O*-sulfotransferase (6-OST), and 3-*O*-sulfotransferase (3-OST). Recombinant HS biosynthetic enzymes, readily expressed in *E. coli*, display comparable substrate specificities to their mammalian counterparts[187]. These recombinant enzymes offer a new strategy to synthesize heparin oligosaccharides using a chemoenzymatic approach[188-190].

LMWH is a depolymerized derivative of heparin, isolated from porcine intestine. A worldwide contamination of heparin in 2007 impacted the purity and safety of LMWHs[191] and was associated with over 200 deaths in the US[192]. This crisis revealed the fragility of the LMWH supply chain. The cost-effective preparation of a synthetic LMWH could improve drug safety and efficacy[193]. Currently available LMWHs are complex mixtures, having average molecular masses of 3500-6000 Daltons, corresponding to 12-20 saccharide units. Until now the

preparation of a homogeneous LMWH has not been possible due to difficulties in the chemical synthesis. Thus, we sought to design and synthesize a potential lead compound for a new generation of LMWH, particularly to explore reversibility of treatment.

Targeted Structures and Schematic Synthesis of LMWHs

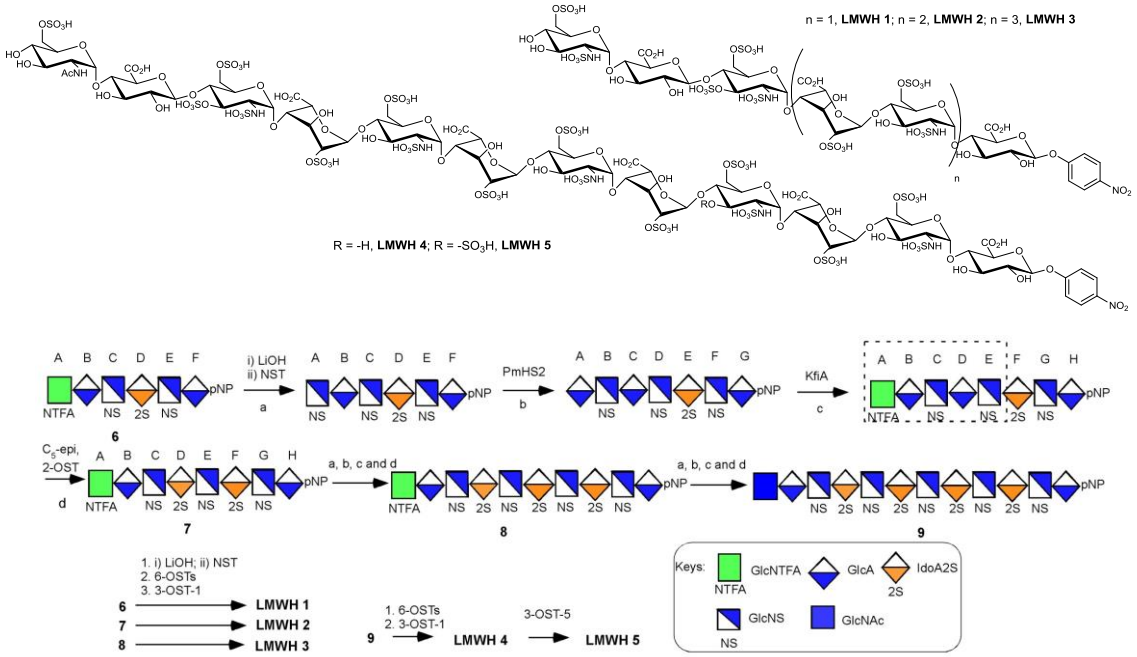


Figure 25. Targeted structures and schematic synthesis of synthetic LMWHs. (**a**) The five synthetic LMWHs prepared in this study. (**b**) Hexasaccharide **6** (Figure 26) was converted to octasaccharide **7**, decasaccharide **8** and dodecasaccharide **9** to synthesize **2**, **3** and **4**, respectively. The conversion of **4** to **5** was achieved by 3-OST-5 modification.

Five synthetic LMWHs (Compound **1-5**, Figure 25), ranging from hexasaccharide to dodecasaccharide, were synthesized. The structures of **1-4** contain a different number of –

IdoA2S-GlcNS6S- (where S is sulfate) repeating units near their reducing ends. **5** differs from **4**, as it has two 3-*O*-sulfate groups. The synthesis of **1** was initiated from a commercially available monosaccharide, 1-*O*-(*para*-nitrophenyl) glucuronide (GlcA-pNP) (Figure 26). Elongation of GlcA-pNP to a hexasaccharide was accomplished using two bacterial glycosyltransferases: KfiA (*N*-acetylglucosaminyl transferase from *E. coli* K5 strain) and heparosan synthase 2 (PmHS2) from *Pasteurella multocida*. *N*-sulfation, *O*-sulfation and epimerization afforded **1**[190]. **1** was expected to have a comparable pharmacological profile to that of fondaparinux as both have very similar structures. The approach also afforded hexasaccharide **6** (461 mg), a key intermediate for subsequent syntheses.

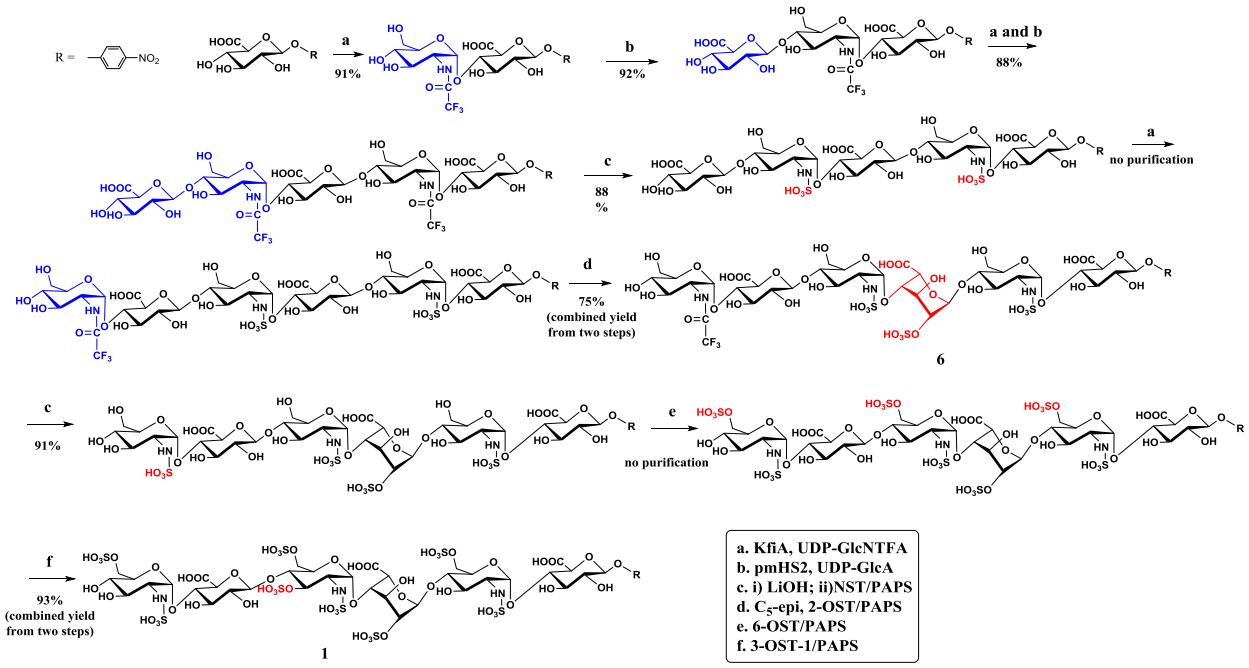


Figure 26. Chemoenzymatic synthetic schemes of Compound 1 and Compound 6. The synthesis started from monosaccharide 1-*O*-(*para*-nitrophenyl) glucuronide (GlcA-pNP), which is commercially available. The recovery yield at each synthetic step is indicated below the reaction arrow.

The syntheses of **2**, **3**, and **4** were initiated from hexasaccharide **6** through intermediates **7**, **8** and **9**, respectively (Figure 25). These intermediates contain multiple –IdoA2S-GlcNS- repeating units, posing a synthetic challenge due to the substrate specificity of C_5-epi[25]. A carefully designed sequence of enzymatic steps was employed for high purity and yields. The conversion of GlcA to IdoA2S involves two steps: C_5-epi catalyzes epimerization of a GlcA to an IdoA; and 2-OST transfers a sulfate group to IdoA. C_5-epi catalyzes both the forward and reverse reactions, leading to the incomplete conversion of GlcA to IdoA2S[194] and a complex mixture of products[195]. The placement of a pentasaccharide domain, GlcN-trifluoroacetyl(TFA)-GlcA-GlcNS-**GlcA**-GlcNS-, into the substrate, directs C_5-epi to relatively irreversible react only with one GlcA residue and avoids incomplete conversion.

The conversion of **6** to **7**, to construct the pentasaccharide domain recognized by C_5-epi, was completed in four steps. The GlcNTFA residue (Residue A) of hexasaccharide **6** was first converted to a GlcNS residue (Step a, Figure 25). The resultant hexasaccharide was then elongated to an octasaccharide in two enzymatic steps (step b and c), to obtain the desired pentasaccharide domain (dashed box, Figure 25). The conversion of GlcA (Residue E) to IdoA2S in **7** was achieved by C_5-epi and 2-OST (step d). The formation of IdoA2S residue removes its reactivity towards further C_5-epi modification. Repeating these steps (steps a to d) once or twice afforded **8** and **9**, respectively. The syntheses of **7**, **8**, and **9** were completed, and structural characterization and purity are shown in Appendix Supplementary Figure 1 – 12**.** The octasaccharide **7** and decasaccharide **8** were converted to **2** and **3** respectively after *N*-sulfation, 6-*O*-sulfation and 3-*O*-sulfation; dodecasaccharide **9** was converted to **4** after 6-*O*-sulfation and 3-*O*-sulfation. The conversion of **4** to **5** was achieved by 3-OST-5 modification. Final product purity was confirmed by high performance liquid chromatography (HPLC). Structures were

determined by mass spectrometry (MS) and nuclear magnetic resonance (NMR) analysis (Appendix Supplementary Figure 13-28). MS-assisted sequence analysis[196] was employed to further prove the structure of **5** (Appendix Supplementary Figure 29-30).

Anticoagulant Activity and Protamine Reversibility of Synthetic LMWH

Compound	Molecular mass (Da)[1]	Amount (mg)	Purity (%)[2]	Affinity to antithrombin (K_d)[3]	Anti-FXa activity (IC_{50}, $ngml^{-1}$)[4]
1	1791.6 ± 0.4 (1791.5)	5	≥ 99	7 ± 2 nM	14
2	2368.4 ± 0.4 (2368.9)	6	≥ 99	8 ± 3 nM	17
3	2945.9 ± 0.4 (2946.4)	6	≥ 99	5 ± 1 nM	15
4	3485.1 ± 0.5 (3485.8)	70	≥ 99	30 ± 7 nM	35
5	3565.2 ± 1.1 (3565.8)	17	≥ 99	28 ± 16 nM	21

1. Molecular mass was determined by low resolution ESI-MS. The calculated molecular mass is shown in the parenthesis.
2. The purity was determined by DEAE-HPLC analysis.
3. The K_d values are the average of two or three determinations. The K_d value for fondaparinux was previously reported to be 5.9±1.5nM [1].
4. IC_{50} represents half maximal inhibitory concentration to inhibit the activity of factor FXa. The IC_{50} value for fondaparinux was previously reported to be 4.5$ngml^{-1}$ [1]. The lower IC_{50} value observed for fondaparinux is ascribed to the different protocols used in the experiments.

Table 3. Summary of the properties of synthetic LMWHs [1].

The anticoagulant activities of synthetic LMWHs were determined. All compounds (**1-5**) displayed strong antithrombin (AT)-binding affinity (K_d 5-30 nM) (Table 3). Unlike UFH and enoxaparin, synthetic LMWHs have no detectable anti-FIIa activity, and thus, these compounds are FXa-specific inhibitors. Our data also suggest that compounds larger than dodecasaccharides are required for anti-FIIa activity. Bemiparin, a "second-generation" LMWH, was recently

approved by the European Medicines Agency (EMA). It has significantly lower anti-FIIa activity than other LMWH drugs, but has very similar clinical utility, suggesting that LMWH anti-FIIa activity is not critical[197].

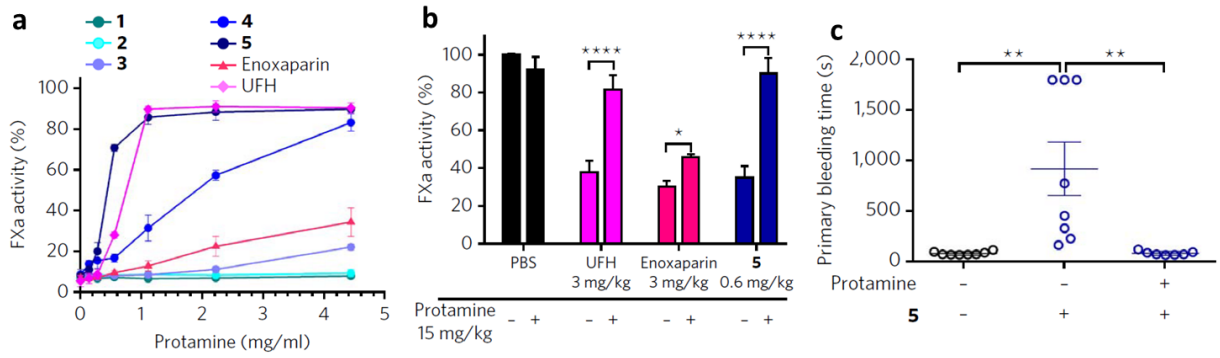


Figure 27. Determination of the anti-FXa activity and sensitivity to protamine neutralization of synthetic LMWHs. (**a**) FXa activity of synthetic LMWHs in the presence of different concentrations of protamine under *in vitro* conditions. (**b**) *Ex vivo* reversibility of anti-FXa activity by protamine. The inhibition of FXa activity by the test compounds was significantly affected in the presence of protamine (*$P < 0.05$ and ****$P < 0.0001$). Data presented in **a**–**b** are the average of three to five determinations ± s.d. (**c**) Effect of **5** and protamine on tail-bleeding time after tail transection. Protamine significantly shortened the primary bleeding time (**$P < 0.01$) that was induced by **5**. Each data point represents the measured value from an individual mouse in the test group. Data are presented as mean ± s.d.

We next determined the protamine reversibility of the anticoagulant activity of five synthetic LMWHs (Figure 27a). In contrast to the activities of **1**–**3**, the activity of **4** was partially reversed by protamine and the activity of **5** was more reversible than enoxaparin and showed the same protamine reversibility as UFH activity (Figure 27a). Using an *ex vivo* mouse model, we confirmed that **5** has similar sensitivity to protamine neutralization as UFH. As expected, enoxaparin was only partially neutralized by protamine (Figure 27b). Finally, using a mouse-tail-clip bleeding model, we demonstrated that protamine shortened the prolongation of bleeding time induced by **5** (Figure 27c), confirming the sensitivity of **5** to protamine neutralization *in vivo*. The US FDA has recently approved three new anticoagulants—dabigatran

etexilate, rivaroxaban and apixaban—however, none of these drugs has an antidote. Although other strategies to inhibit anticoagulant drugs using an engineered FXa-like protein[198] or antithrombin mutant[199] have been reported, these methods remain to be approved by a regulatory agency. Synthetic LMWH **5** can be neutralized by protamine, an FDA-approved antidote for UFH.

Superb sensitivity toward protamine neutralization displayed by synthetic **5** suggests that both sulfation pattern and size contribute to its sensitivity to neutralization. Currently marketed LMWHs, *e.g.* enoxaparin, are a mixture of oligosaccharides with a broad size distribution of chains, having different affinities towards protamine. Our data suggest that protamine only neutralizes chains in LMWHs that are larger than decasaccharides. Smaller chains, pentasaccharide to decasaccharide, can exhibit anticoagulant activity[200], even though they are not protamine reversible. Consequently, only partial neutralization for commercial LMWH can be achieved with protamine. Synthetic **5** has a uniformly high affinity to protamine because it is a structurally homogeneous compound. In addition, our results demonstrate that an extra 3-*O*-sulfate group significantly increases the sensitivity for protamine neutralization, suggesting that the sulfation pattern plays a critical role in protamine binding.

Additional efforts are necessary to further advance our findings to a LMWH drug candidate. First, replacement of pNP aglycone should eliminate concerns that it might be converted to *p*-nitrophenol *in vivo*, leading to a potentially harmful metabolite. Second, a comprehensive structure-activity relationship study will be required to understand the contribution of sulfation pattern to the pharmacological effects.

LMWHs are critical for the practice of modern medicine yet their production still depends on a long supply chain, which is vulnerable to contamination and adulteration. After the

contamination crisis, the US FDA and EMA implemented a series of new approaches to monitor the purity of heparin drugs. Although these efforts have stopped the influx of contaminated heparin into the market, a long term solution should be to manufacture synthetic heparin under highly regulated processes, eliminating the need for animal-sourced heparin[182, 193]. Chemoenzymatic synthesis offers a promising approach towards this goal. A homogeneous product reduces the complexity for the quality control during the manufacturing process, and is compatible with standard approval processes by regulatory agencies[201]. While the structures produced through chemoenzymatic synthesis are limited by the substrate specificities of the enzymes, their synthetic capability can be expanded through better understanding of enzyme properties. As demonstrated in the present study, a firm understanding of C_5-epi specificity resulted in the design of a modification sequence leading to the synthesis of the critical -IdoA2S-GlcNS- repeating domain. The next challenge will be to determine whether a large-scale synthesis of homogeneous LMWH can be accomplished to meet the needs of the anticoagulant drug market. The synthesis of **5** requires 22 synthetic steps with the yield around 7%, but it is significantly more efficient than the synthesis of fondaparinux that requires 50 steps with 0.1% yield [202]. Fondaparinux is now synthesized in kg-scale, and has been a profitable drug for ten years. The success of fondaparinux suggests that further development of the chemoenzymatic approach will result in cost-effective products, accelerating the modernization of LMWH therapeutics.

Synthesis of Less Abundant Heparan Sulfate Oligosaccharide Library

Introduction

Proteoglycans mediate many protein functions and physiological processes ranging from cell development, differentiation, inflammation, viral or parasite infection, metastasis to cancer [2]. Proteoglycans comprise a core protein and various glycosaminoglycans (GAG) side chains. Although the amino acid sequence of the core protein is encoded by a specific gene, the GAG side chains are generated by a non-template enzymatic process [35, 203, 204]. Heparin and heparan sulfate (HS) are members of this polysulfated GAG family.

Heparin, a widely used clinical anticoagulant since the 1930's, was the first biopolymeric drug and is one of the few carbohydrate based drugs [11, 12]. Because of the heterogeneity of heparin sequences, understanding the contribution of a specific sulfated carbohydrate sequence for a given biological activity remains challenging. Heparin is a mixture of polysaccharides and has high polydispersity and polyheterogeneity [12] as well as flexible structural characteristics [11]. In addition, heparin is a highly sulfated polysaccharide with an average molecular weight of ~17 kDa [13] and with a range of chain lengths from 5,000 to 50,000 Da [14]. Heparin has repeating disaccharide units consisting of uronic acid (90% L-iduronic acid, IdoA, and 10% D-glucuronic acid, GlcA) linked 1 → 4 to D-glucosamine (GlcN) [15], where the glucosamine residues can be either *N*-acetylated (GlcNAc) or sulfated (GlcNS). The GlcN monosaccharide can also contain *O*-sulfo groups, including 6-*O*- and 3-*O*-sulfation, while the GlcA and IdoA residues can be 2-*O*-sulfated. Heparan sulfate (HS) is structurally analogous to heparin. HS has a similar disaccharide repeating structure to heparin with some noticeable differences. Generally, the size of the HS chains is longer with an average MW of 30 kDa [16], and HS

contains a higher GlcA content and less sulfation than heparin [12]. In addition, HS has more diverse sulfated saccharide sequences and domain structures compared with heparin.

Reliable structural characterization techniques are critically important to safeguard the quality of heparin drugs. In 2008, batches of contaminated heparin were discovered in US, which drew significant public attention because the contaminated drug killed 94 patients in US [68]. Nuclear Magnetic Resonance (NMR) techniques were employed, and rapidly identified the nature of the contaminant to be over sulfated chondroitin sulfate, demonstrating the important role of NMR measurements in heparin quality control [11]. NMR data provides structural information that cannot be obtained by other methods, specifically; epimer recognition, anomeric configuration identification, sulfated position determination, and conformation and secondary structure determination. One limitation in utilizing NMR to analyze the structure of heparin or HS is that these species are a mixture of compounds and pure oligosaccharide standards have not been available. Although certain oligosaccharides have been isolated from partially depolymerized heparin or HS, obtaining sufficient amounts of those highly purified oligosaccharides is a difficult and time consuming process.

Recently, a new method using recombinant HS biosynthetic enzymes to prepare structurally homogeneous HS oligosaccharides was developed [1]. This method allows the synthesis of a series of structurally homogeneous heparin or HS oligosaccharides at the scale of tens to hundreds of milligram with high yield. Access to highly purified HS oligosaccharides allows for full-scale analyses of these complex molecules by NMR. Notably these synthetic compounds enable the analysis of structurally “rare” oligosaccharides which are difficult to isolate in sufficient quantity for NMR analysis. For example, the -GlcA2S- residue represents only 1-5% of total uronic acid in heparin isolated from natural sources [77].

Here, to address this knowledge gap of the structure characteristics of these potentially important oligosaccharide species, we report the NMR analysis of a series of GlcA2S-containing hexasaccharides with different sulfation patterns. The GlcA2S oligosaccharides were prepared by a chemoenzymatic approach. The GlcA2S NMR data was compared with the data obtained from IdoA2S-containing oligosaccharides synthesized by the same method. The data confirm previous assignments for the NMR signals of GlcA2S saccharides [77].

Chemoenzymatic Synthesis of GlcA2S- and IdoA2S-Containing Hexasaccharides

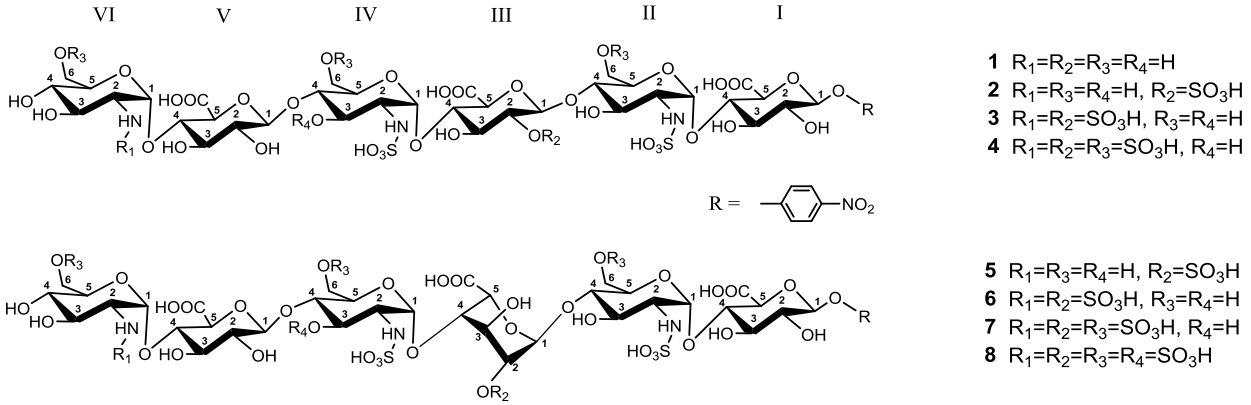


Figure 28. The synthetic hexasaccharide constructs 1 to 8. A total of eight synthetic hexasaccharides were prepared in this study.

Total of eight hexasaccharide constructs **1** to **8** were synthesized in this study using a chemoenzymatic approach (Figure 28). The synthesis was initiated from a commercially available monosaccharide, 1-*O*-(*para*-nitrophenyl) glucuronide (GlcA-pnp), requiring elongation steps, sulfation and epimerization steps (Figure 29) [1].

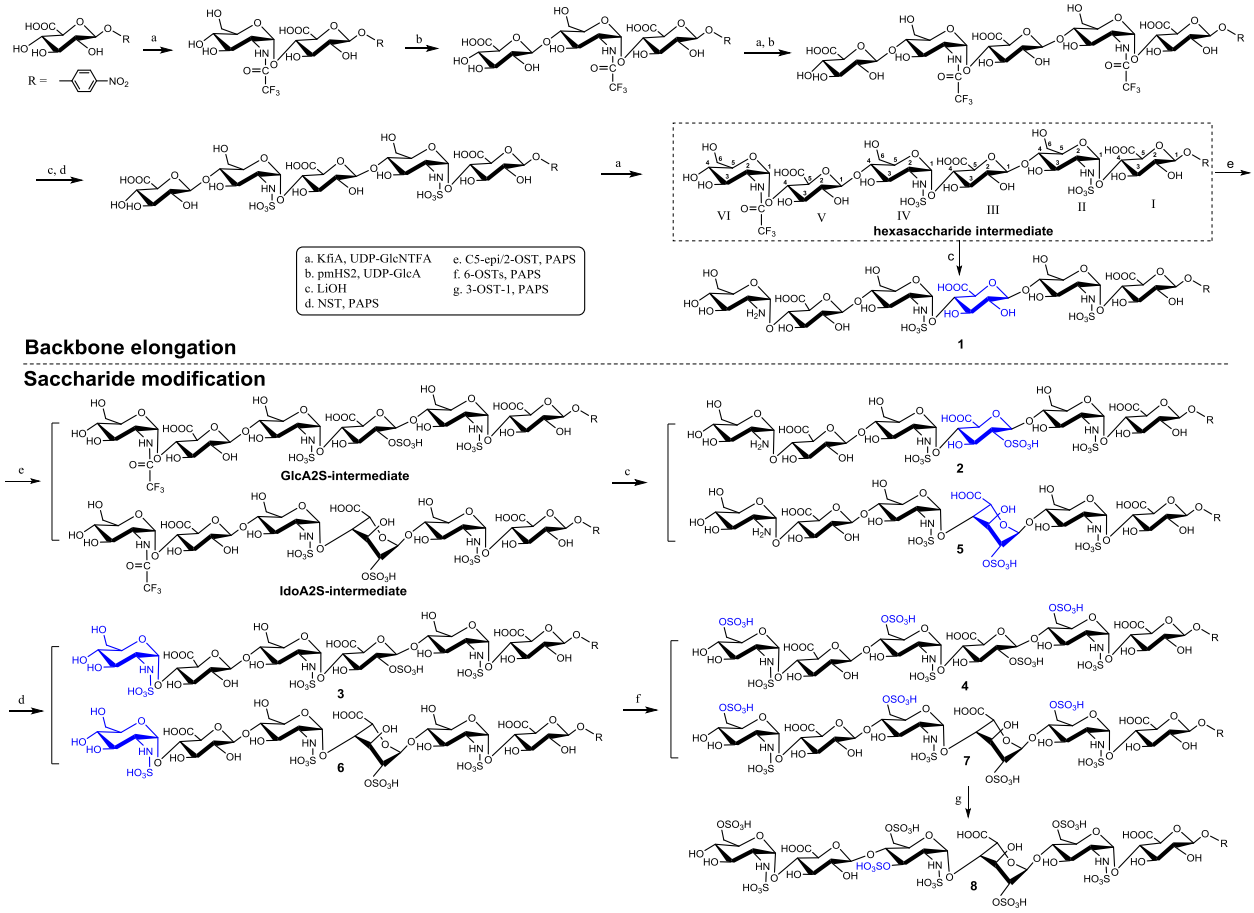


Figure 29. Chemoenzymatic synthesis of GlcA2S- and IdoA2S-containing hexasaccharides. The chemoenzymatic synthetic scheme for hexasaccharides includes the backbone elongation and saccharide modification steps. The synthesis was initiated from a commercially available monosaccharide, 1-*O*-(*para*-nitrophenyl) glucuronide (GlcA-pnp), requiring elongation steps, sulfation and epimerization steps to synthesize hexasaccharide intermediate and construct **1** to **8**. The modification site is highlighted in blue color.

In the synthesis, a hexasaccharide with a structure of GlcNTFA-GlcA-GlcNS-GlcA-GlcNS-GlcA-pnp is a crucial intermediate. The *N*-trifluoroacetyl group of this hexasaccharide intermediate was hydrolyzed to yield construct **1** by treatment with lithium hydroxide. Then this hexasaccharide intermediate was treated with a mixture of C5-epi and 2-OST to yield a mixture of GlcA2S- and IdoA2S-intermediate. An anion-exchange HPLC method was used to isolate and purify those two products with a ratio of 1:9 (Figure 30A).

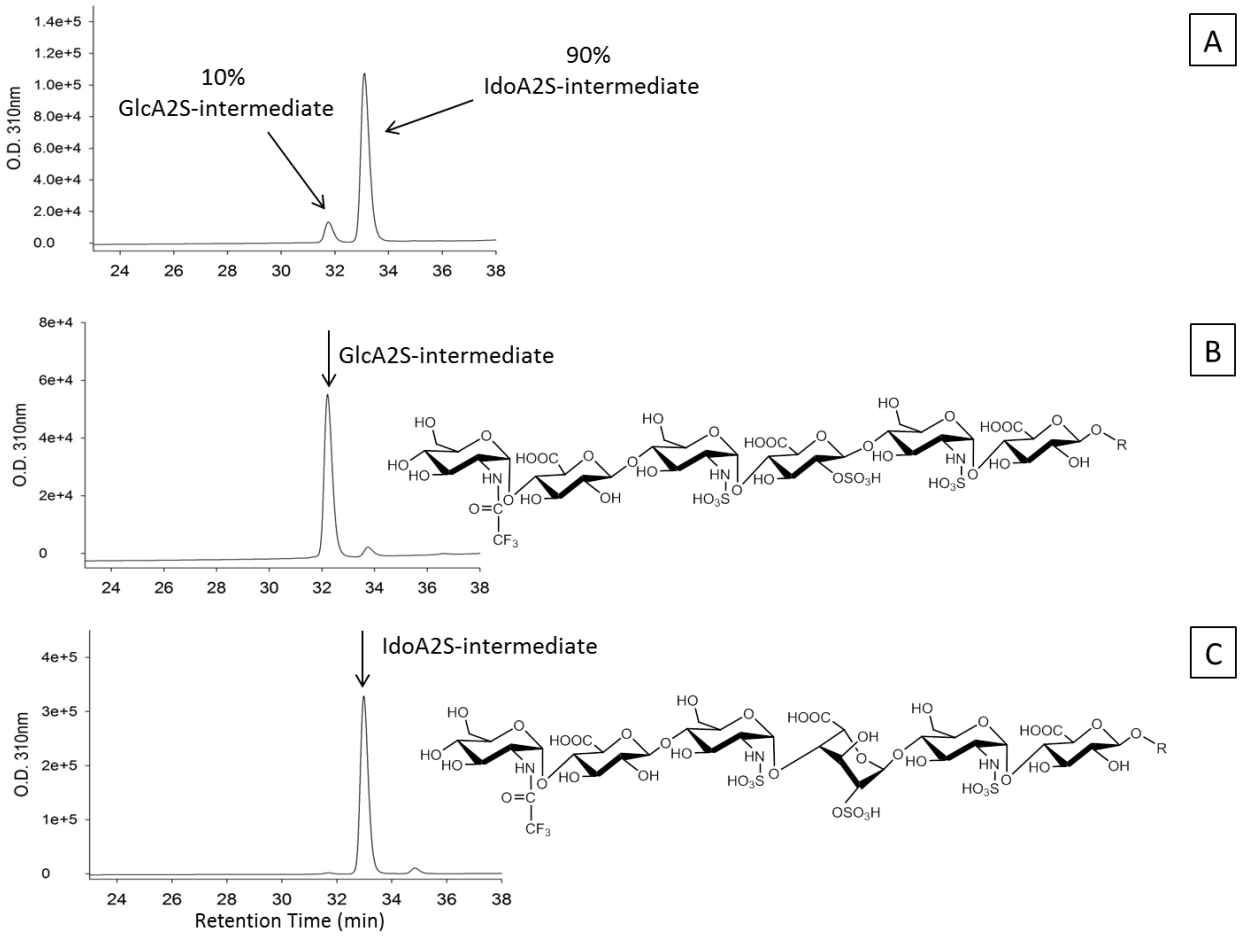


Figure 30. Two major products of epimerization and 2-*O*-sulfation. Panel A shows the anion-exchange HPLC chromatogram after epimerization/2-*O*-sulfation of hexasaccharide intermediate. The product contains 10% GlcA2S-intermediate and 90% IdoA2S-intermediate. Panel B and C show the anion-exchange HPLC chromatogram of GlcA2S-intermediate and IdoA2S-intermediate after purification. The purity is greater than 93% and 96%, respectively. The chemical structure of GlcA2S-intermediate and IdoA2S-intermediate are shown on right of Panel B and C.

After the purification, a GlcA2S- intermediate was obtained with the purity greater than 93% as determined by anion-exchange HPLC (Figure 30B). Similarly, an IdoA2S-intermediate was obtained with the purity greater than 96% (Figure 30C). The GlcA2S-intermediate was further converted to construct **2** by detriflouroacetylation under an alkaline condition. Additional modifications were made, including *N*-sulfation and 6-*O*-sulfation, yielded to construct **3** and **4**,

respectively. Similarly, the IdoA2S-intermediate was transformed to construct **5, 6** and **7**. 3-*O*-sulfation was carried out on construct **7** to form construct **8**. The purity analysis of each synthesized oligosaccharide and mass spectrometric analyses are shown in the Supplementary data (Figure 31-32).

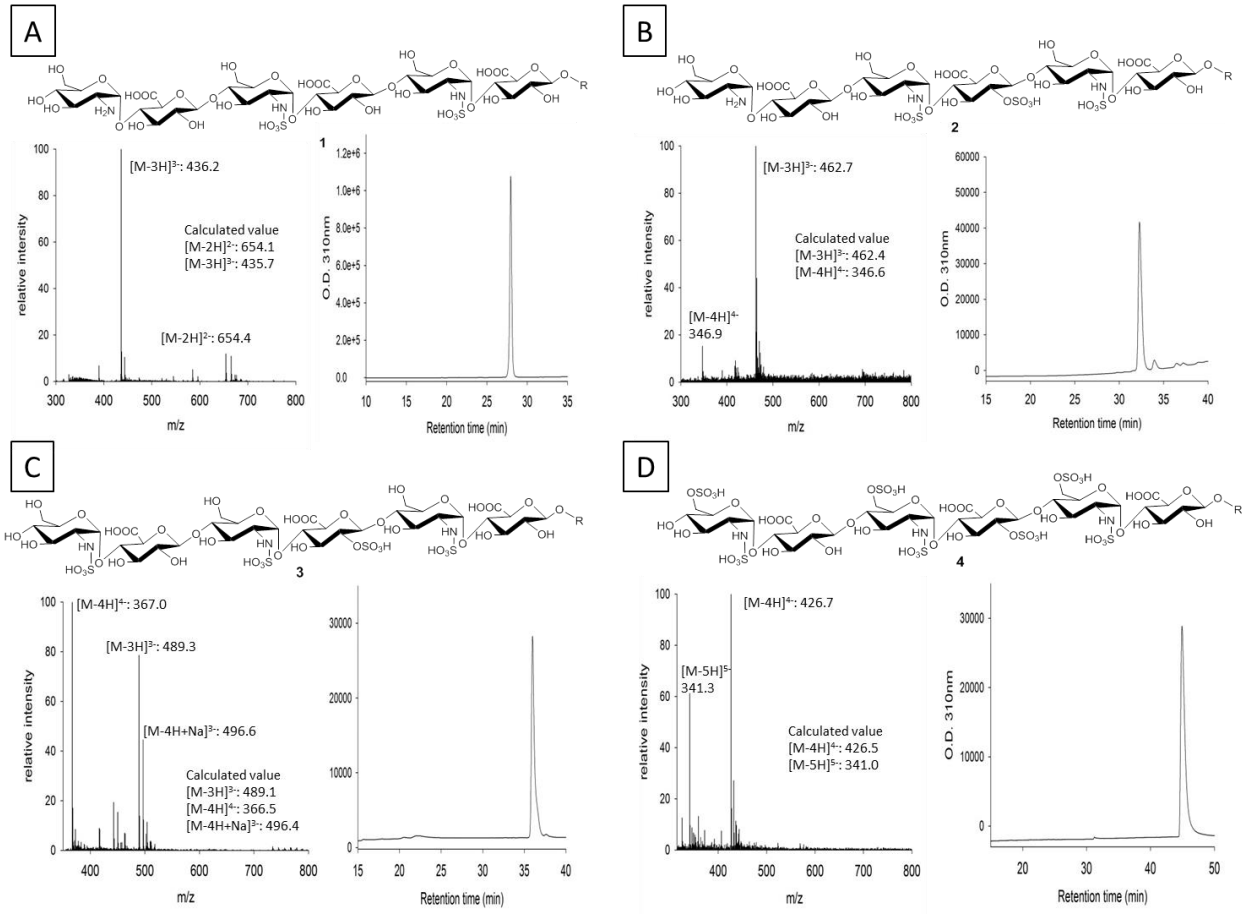


Figure 31. Mass and HPLC analysis of construct 1 to 4. Panel A, B, C and D show the ESI-MS spectrometry and anion-exchange HPLC chromatogram of construct **1**, **2**, **3** and **4**, respectively. The chemical structure is shown on top of each Panel. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states in negative electrospray ionization (ESI) mode.

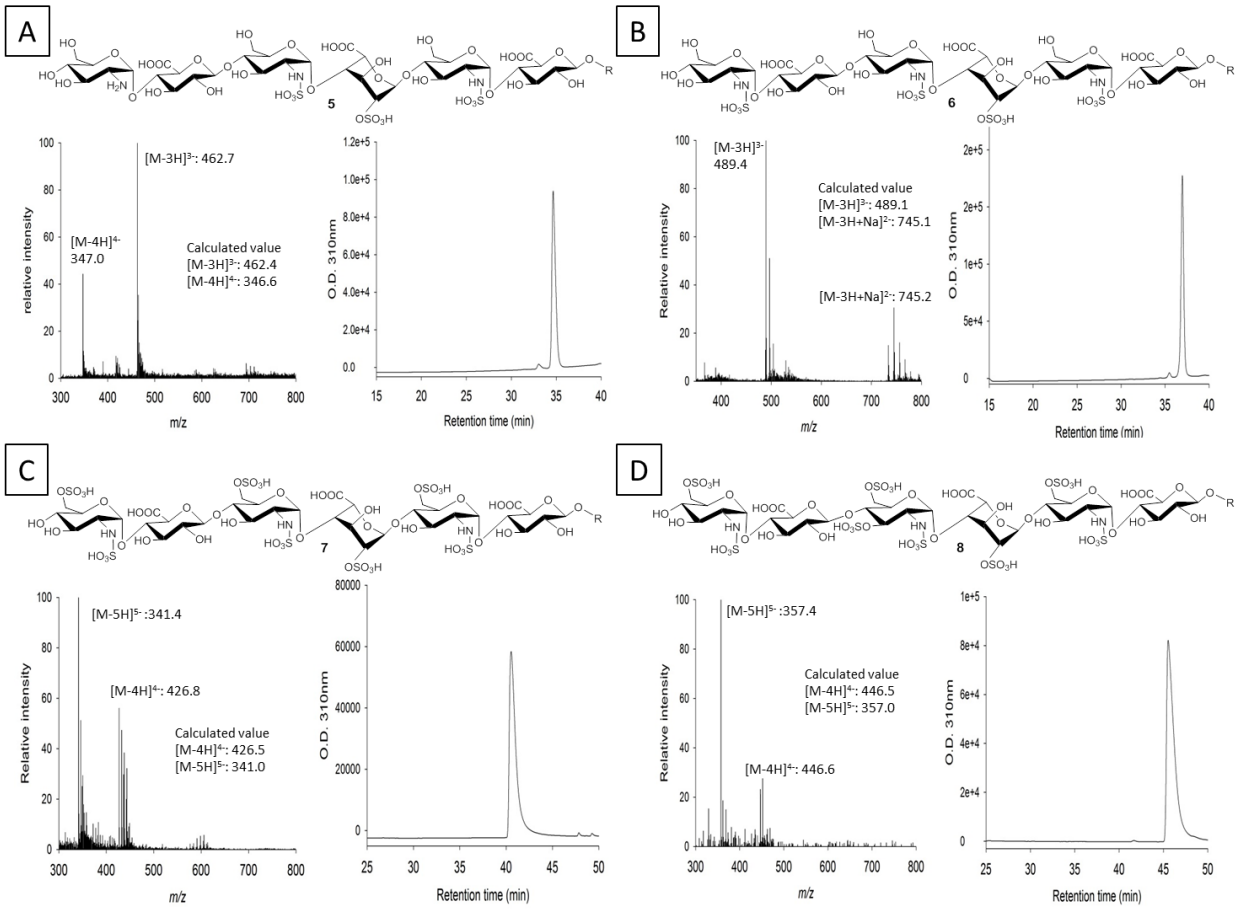


Figure 32. Mass and HPLC analysis of construct 5 to 8. Panel A, B, C and D show the ESI-MS spectrometry and anion-exchange HPLC chromatogram of construct **5**, **6**, **7** and **8**. The chemical structure is shown on top of each Panel. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states in negative electrospray ionization (ESI) mode.

The GlcA2S residue is present in low abundance in heparin, and thus, very limited information about the role of such rare saccharide residues has been obtained. In this work, we demonstrate the synthesis and structural characterization of highly pure oligosaccharides containing GlcA2S. The synthesis of three GlcA2S containing hexasaccharides was completed through a chemoenzymatic approach. The substrate specificity of 2-OST modifies only one of the two GlcA in the hexasaccharides, allowing us to position the GlcA2S residue at the desired site (Figure 29). Although the synthesis of a number of HS hexasaccharides using a chemical and

chemoenzymatic approach has been reported [163, 205], the synthesis of GlcA2S-containing hexasaccharide variants described in this research is novel.

Structural Characterization of 2-O-Sulfated Glucuronic Acid (GlcA2S)- and Iduronic Acid (IdoA2S)- Containing Oligosaccharides

NMR Analysis of Hexasaccharides

The availability of high purity structurally defined hexasaccharides provided us sufficient quantities for NMR analyses. The ^{1}H-NMR spectra of GlcA-containing hexasaccharide (construct **1**) and GlcA2S-containing hexasaccharides, including construct **2**, **3** and **4,** are shown in Figure 33; the ^{1}H-NMR spectra of IdoA2S-containing hexasaccharides, including construct **5** to **8,** are shown in Figure 34. The clustered proton signals from the saccharide residues made the assignments difficult. Therefore, 2D-homonuclear and -heteronuclear correlation spectroscopy analyses (^{1}H-^{1}H COSY, TOCSY, ^{1}H-^{13}C HSQC, and HMBC) were conducted; however the analyses only partially resolved the overlapped signals. Thus, 2D ^{1}H-^{13}C HSQC-TOCSY NMR experiments (Figure 35) was employed to resolve these signals. The fully assigned signals for constructs **1**-**8** are listed in Table 4 and 5.

Characterization of GlcA2S-Containing Oligosaccharides

Construct **3** (with the structure of GlcNS-GlcA-GlcNS-GlcA2S-GlcNS-GlcA-pnp) was used to demonstrate how the signals of GlcA2S were identified in the synthesized oligosaccharides. In the ^{1}H-NMR spectrum of construct **3**, the signals from anomeric protons of

the three GlcNS units are located around 5.6 ppm, and the signals from GlcA/GlcA2S are located around 4.6 ppm (Figure 33-34). The ~5.3 ppm shift of I-1 is attributed to the effect of the nearby electron withdrawing group, *para*-nitrophenyl (pnp). The $^{3}J_{HH}$ coupling constants of I-1, III-1, and V-1 were ~8 Hz (consistent with β linkages), and II-1, IV-1, and VI-1 were ~4 Hz (consistent with α linkages). Other saccharide ring proton signals were highly overlapped and resonated between 3.2 and 4.1 ppm.

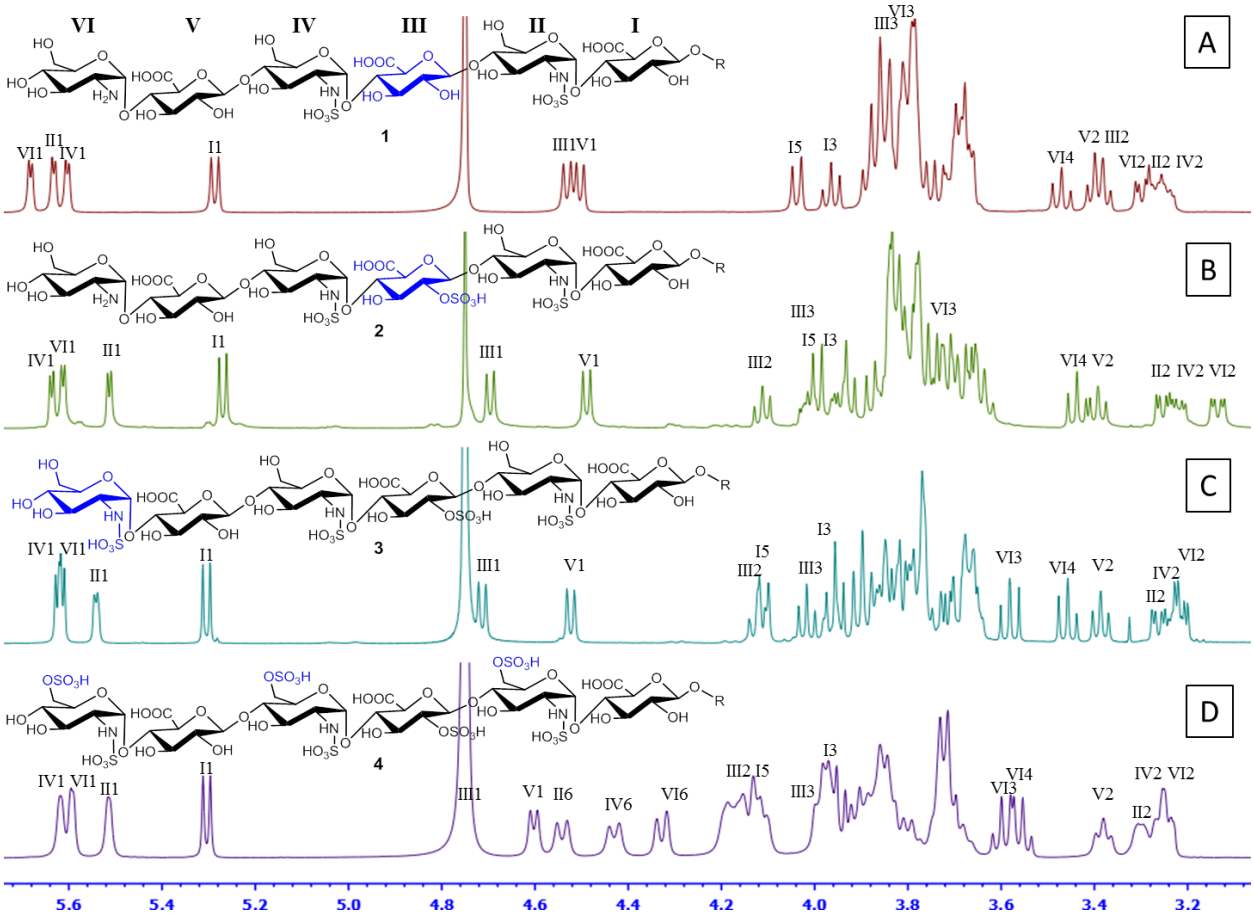


Figure 33. ^{1}H-NMR spectra of construct 1, 2, 3 and 4. Panel A, B, C and D show the ^{1}H-NMR spectrum of construct **1**, **2**, **3** and **4**, respectively. The signals of anomeric protons and some non-overlapping protons are indicated. The chemical structure of construct is shown on top of each panel. The ^{1}H-NMR spectra represent the range between 3.1 and 5.7 ppm.

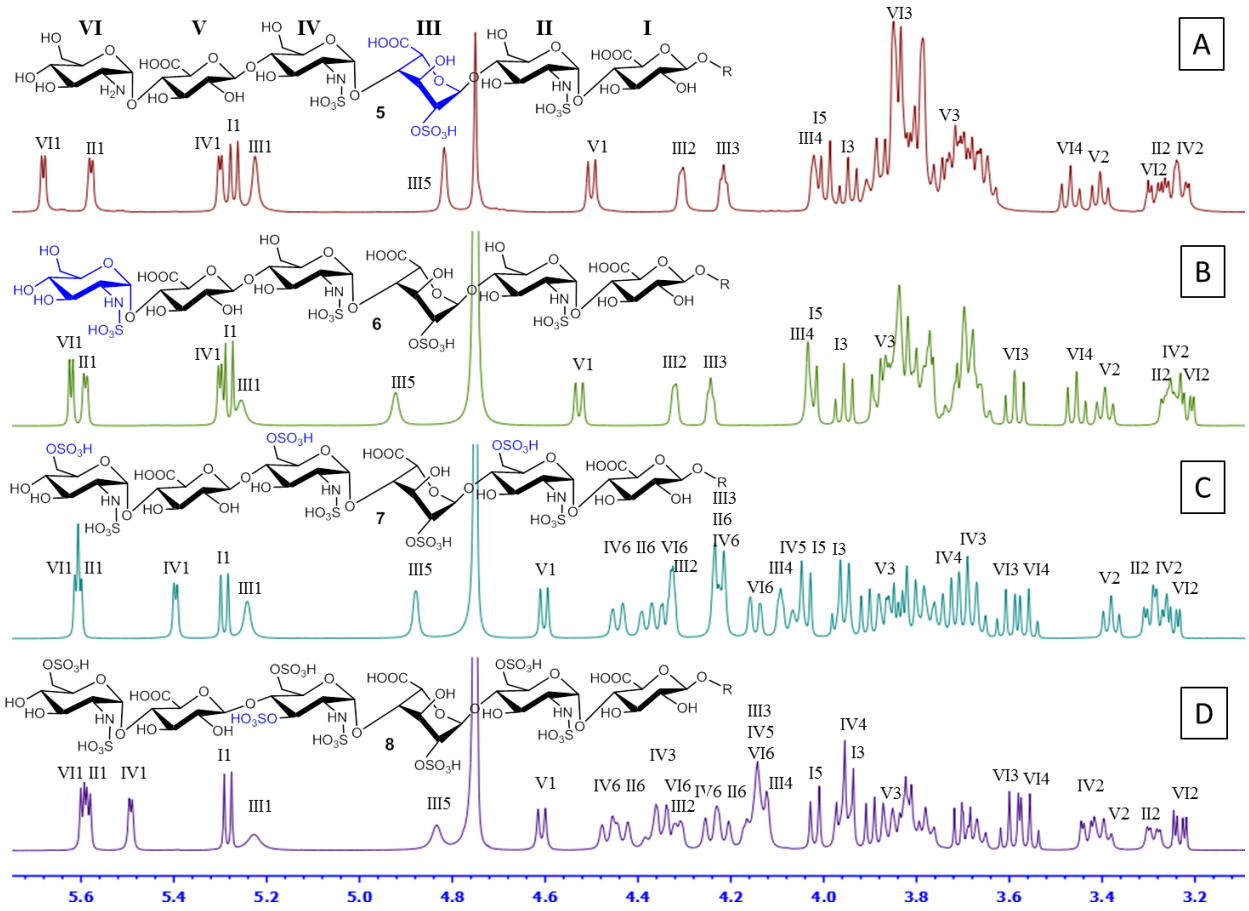


Figure 34. ^{1}H-NMR spectra of construct 5, 6, 7 and 8. Panel A, B, C and D show the ^{1}H-NMR spectrum of construct **5**, **6**, **7** and **8**, respectively. The signals of anomeric protons and some non-overlapping protons are indicated. The chemical structure of construct is shown on top of each panel. The ^{1}H-NMR spectra represent the range between 3.1 and 5.7 ppm.

The increased shift dispersion in 2D ^{1}H-^{13}C HSQC-TOCSY NMR, fully resolved these signals so that assignments could be made (Figure 35). In addition, ^{1}H-^{1}H COSY and ^{1}H-^{13}C HSQC analysis were run with better digital resolution to allow more accurate identification of the proton and carbon assignments (Table 4-5). The assignments for other GlcA2S-containing hexasaccharides were conducted in a similar manner.

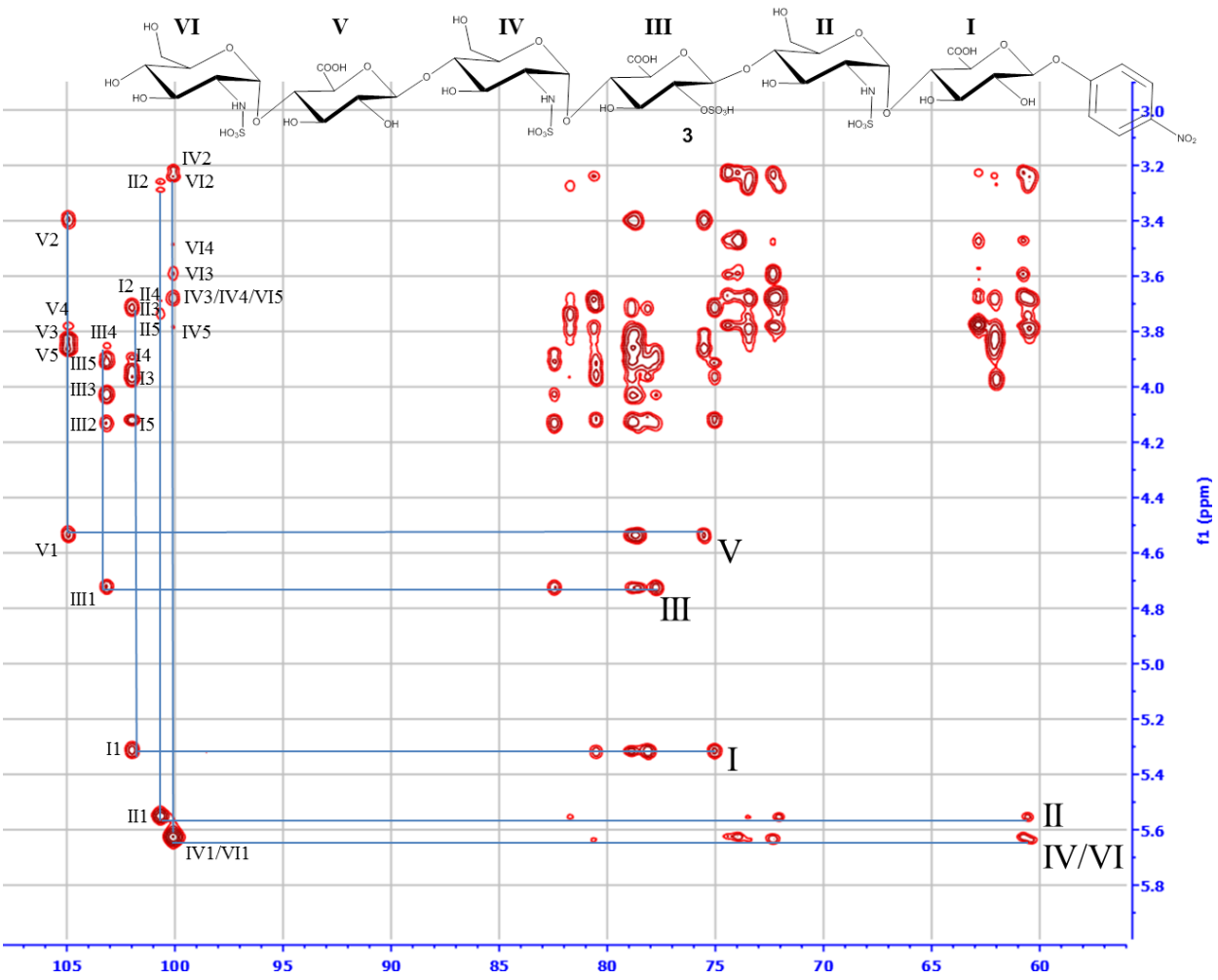


Figure 35. 2D ^{1}H-^{13}C HSQC-TOCSY NMR spectrum of construct 3. 2D ^{1}H-^{13}C HSQC-TOCSY NMR experiment was employed to characterize the structure of construct **3**. The signals of protons are indicated. The signals are connected with lines identifying correlated protons on each saccharide residue. The chemical structure of construct **3** is shown on top of the Figure.

	1	2	3	4	5	6a	6b
Construct **1** ($GlcNH_2$-GlcA-GlcNS-GlcA-GlcNS-GlcA-pnp)							
I	5.29/102.0	3.69/75.1	3.97/78.6	3.89/79.0	4.04/79.1		
II	5.63/99.9	3.28/60.4	3.69/72.1	3.69/80.6	3.81/73.3	3.80/62.1	3.85/62.1
III	4.53/104.8	3.38/75.4	3.85/78.7	3.79/79.0	3.84/78.7		
IV	5.60/99.9	3.25/60.4	3.68/72.1	3.68/80.5	3.79/73.2	3.80/62.1	3.85/62.1
V	4.50/104.8	3.40/75.9	3.75/78.6	3.82/78.7	3.87/78.6		
VI	5.68/97.9	3.30/56.9	3.84/72.1	3.47/71.8	3.71/75.1	3.79/62.4	3.80/62.4
Construct **2** ($GlcNH_2$-GlcA-GlcNS-GlcA2S-GlcNS-GlcA-pnp)							
I	5.27/102.0	3.70/75.1	3.94/78.2	3.88/81.0	4.00/79.8		
II	5.51/100.7	3.25/60.6	3.73/72.2	3.65/81.9	3.82/73.4	3.82/62.1	3.96/62.1
III	4.70/103.3	4.11/82.5	4.02/77.9	3.84/78.9	3.85/79.0		
IV	5.64/99.9	3.22/60.3	3.67/72.4	3.67/80.6	3.81/73.3	3.84/62.1	3.85/62.1
V	4.49/104.9	3.39/75.9	3.75/78.6	3.77/78.7	3.83/78.6		
VI	5.61/98.9	3.15/57.2	3.76/73.3	3.44/72.0	3.73/75.0	3.78/62.6	3.79/62.6
Construct **3** (GlcNS-GlcA-GlcNS-GlcA2S-GlcNS-GlcA-pnp)							
I	5.31/102.0	3.70/75.0	3.95/78.1	3.89/80.5	4.11/78.9		
II	5.54/100.7	3.26/60.5	3.73/72.0	3.66/81.7	3.78/73.5	3.81/62.0	3.96/62.0
III	4.71/103.2	4.12/82.4	4.01/77.7	3.88/78.8	3.90/78.6		
IV	5.62/100.1	3.23/60.4	3.67/72.3	3.67/80.6	3.77/73.4	3.83/62.1	3.86/62.1
V	4.52/104.9	3.39/75.5	3.83/78.6	3.79/79.0	3.86/78.8		
VI	5.61/100.1	3.21/60.8	3.58/73.9	3.45/72.3	3.66/74.4	3.76/62.8	3.78/62.8
Construct **4** (GlcNS6S-GlcA-GlcNS6S-GlcA2S-GlcNS6S-GlcA-pnp)							
I	5.31/102.0	3.72/75.1	3.96/78.2	3.91/81.4	4.12/79.2		
II	5.51/101.0	3.30/60.5	3.73/72.0	3.73/80.3	3.98/71.7	4.18/68.5	4.54/68.5
III	4.76/102.7	4.14/82.4	3.99/77.8	3.86/79.3	3.87/79.1		
IV	5.62/100.2	3.25/60.2	3.69/72.3	3.73/79.7	3.99/71.5	4.19/68.5	4.43/68.5
V	4.60/104.6	3.38/75.6	3.85/78.7	3.81/79.6	3.87/79.0		
VI	5.59/100.3	3.24/60.7	3.61/73.8	3.56/71.7	3.86/72.5	4.15/69.0	4.33/69.0

Table 4. ^{1}H–^{13}C NMR chemical shift assignments (in ppm) of Constructs 1–4 (25°C, D2O, pD 7.0).

	1	2	3	4	5	6a	6b
Construct **5** ($GlcNH_2$-GlcA-GlcNS-IdoA2S-GlcNS-GlcA-pnp)							
I	5.27/102.0	3.69/75.1	3.95/78.5	3.87/79.8	4.00/79.6		
II	5.58/100.3	3.25/61.0	3.65/72.5	3.70/79.7	3.79/73.8	3.83/62.4	3.84/62.4
III	5.23/101.8	4.30/77.6	4.22/70.7	4.03/78.4	4.82/71.1		
IV	5.30/99.8	3.23/60.5	3.71/72.2	3.66/80.5	3.90/73.3	3.84/62.3	3.86/62.3
V	4.50/104.8	3.41/75.9	3.74/78.6	3.83/78.7	3.83/78.8		
VI	5.68/97.9	3.29/56.9	3.83/72.2	3.47/71.9	3.72/75.0	3.78/62.5	3.79/62.5
Construct **6** (GlcNS-GlcA-GlcNS-IdoA2S-GlcNS-GlcA-pnp)							
I	5.28/102.0	3.70/75.1	3.96/78.5	3.88/79.7	4.02/79.4		
II	5.59/100.3	3.26/61.0	3.66/72.4	3.70/79.8	3.80/73.8	3.84/62.3	3.84/62.3
III	5.26/101.8	4.32/77.2	4.24/70.3	4.03/78.4	4.92/70.7		
IV	5.30/100.0	3.25/60.4	3.71/72.1	3.69/80.6	3.85/73.4	3.87/62.1	3.87/62.1
V	4.53/104.9	3.39/75.5	3.84/78.6	3.78/78.9	3.82/79.1		
VI	5.62/100.0	3.22/60.8	3.59/73.9	3.45/72.4	3.68/74.4	3.77/62.9	3.77/62.9
Construct **7** (GlcNS6S-GlcA-GlcNS6S-IdoA2S-GlcNS6S-GlcA-pnp)							
I	5.29/102.0	3.71/75.1	3.95/78.5	3.90/79.7	4.04/79.3		
II	5.60/100.3	3.30/60.7	3.67/72.4	3.77/78.5	3.95/71.9	4.22/69.0	4.38/69.0
III	5.24/101.8	4.32/78.3	4.22/71.5	4.09/78.8	4.88/71.7		
IV	5.40/99.6	3.28/60.2	3.69/72.1	3.74/79.7	4.05/71.6	4.22/68.6	4.44/68.6
V	4.60/104.5	3.38/75.5	3.84/78.6	3.78/79.4	3.82/79.0		
VI	5.61/100.3	3.25/60.6	3.61/73.8	3.56/71.7	3.87/72.5	4.15/69.0	4.33/69.0
Construct **8** (GlcNS6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S-GlcA-pnp)							
I	5.28/102.0	3.70/75.1	3.96/78.5	3.90/80.0	4.02/79.4		
II	5.58/100.4	3.29/60.8	3.67/72.4	3.78/79.7	3.96/72.0	4.22/69.0	4.43/69.0
III	5.23/102.2	4.31/79.6	4.17/72.8	4.14/78.5	4.83/72.8		
IV	5.49/98.9	3.43/59.4	4.37/78.8	3.96/75.7	4.14/72.3	4.24/68.7	4.46/68.7
V	4.61/103.9	3.39/75.5	3.81/78.2	3.82/79.6	3.82/78.8		
VI	5.60/100.3	3.23/60.7	3.60/73.9	3.56/71.7	3.86/72.5	4.14/69.0	4.34/69.0

Table 5. $^{1}H–^{13}C$ NMR chemical shift assignments (in ppm) of Constructs 5–8 (25°C, D2O, pD 7.0).

The 2D-^{1}H-^{13}C HMBC experiment confirmed the glycosidic linkages between each saccharide unit through ^{3}J correlations between the anomeric II-1, III-1, IV-1, V-1 or VI-1 protons show connectivity to the I-4, II-4, III-4, IV-4 or V-4 carbons, respectively (Figure 36). These connectivities allowed sequencing of the hexasaccharides. In addition, the anomeric I-1 proton shows a $^{3}J_{CH}$-correlation with the quaternary carbon of the pnp tag. Furthermore, the II-6, IV-6 or VI-6 protons shows through-bond connectivity to II-4, IV-4 or VI-4 carbons, respectively, supporting the assignments of glucosamine residues. The other GlcA2S-containing hexasaccharides were characterized by following the same approach. In addition to ^{1}H-^{1}H COSY and ^{1}H-^{13}C HSQC experiments, we focused on using 2D ^{1}H-^{13}C HSQC-TOCSY NMR to identify the residue spin systems and 2D ^{1}H-^{13}C HMBC NMR to confirm the linkages.

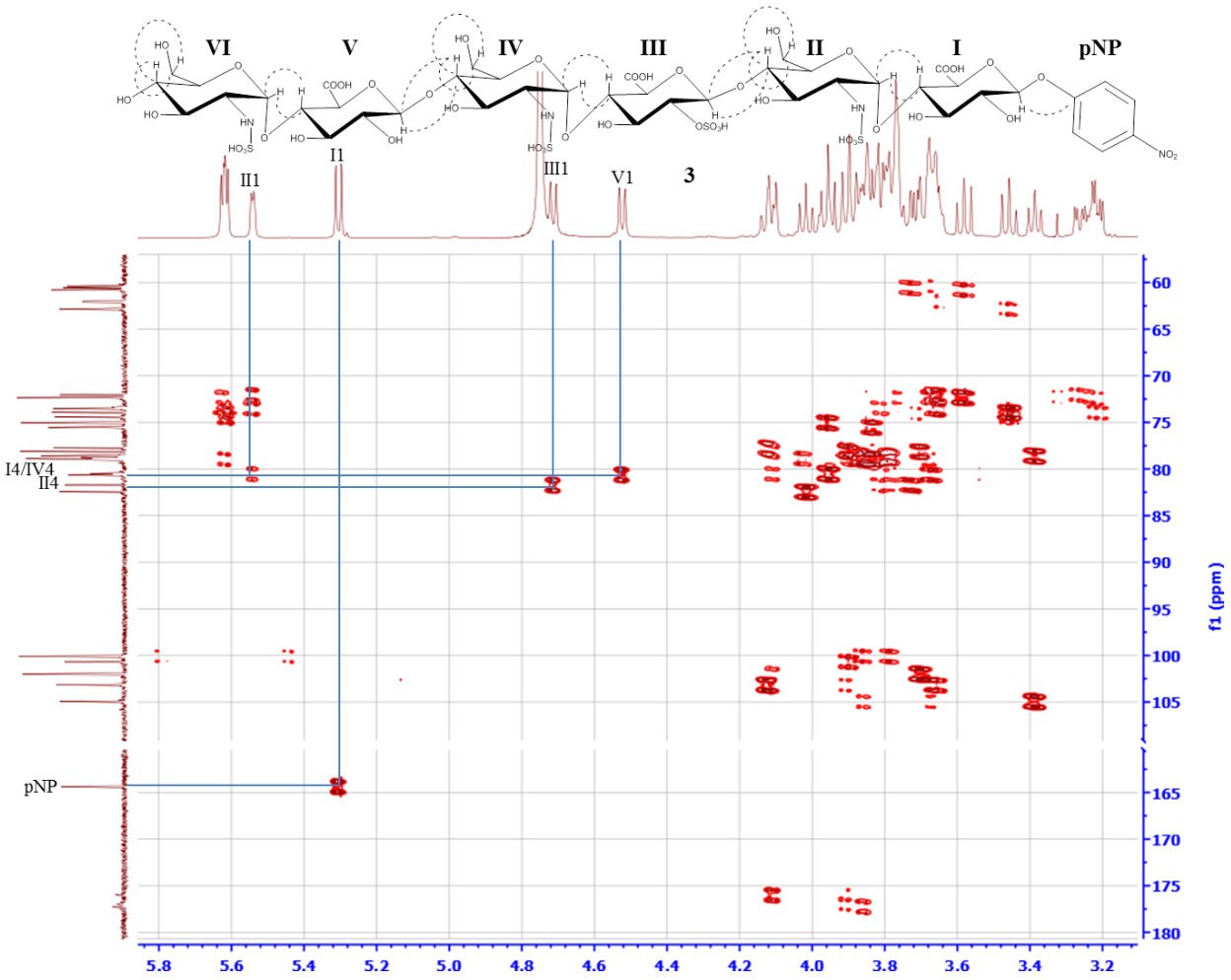


Figure 36. ^{1}H-^{13}C HMBC key correlations of construct 3. 2D ^{1}H-^{13}C HMBC NMR experiment was employed to confirm the glycosidic linkages between each saccharide unit of construct **3**. The signals of anomeric protons are indicated. The peaks are connected with lines identifying correlated carbon and proton. The chemical structure of construct **3** is shown on top of Figure. Some other key $^{3}J_{CH}$-correlations are shown as dashed lines on the chemical structure.

Comparing the NMR spectra of construct **1** with construct **2** provides evidence of the influence of 2-*O*-sulfation on the neighboring residues (Figure 33). The signals of III-1, III-2 and III-3 from construct **2** are shifted +0.17, +0.73 and +0.17 ppm (to 4.70, 4.11, and 4.02 ppm), respectively, due to the attached 2-*O*-sulfo group. Notably, the shift of proton II-1 (-0.12 ppm), II-6b (+0.11 ppm), and VI-2 (-0.15 ppm) are also sensitive to *2-O*-sulfation even though they are

distant from the site of substitution. The chemical shift changes of II residue are attributed to inter-residue influences between this additional 2-*O*-sulfo group. The effect on VI-2 may be due to charge-to-charge interaction between the positively charged unsubstituted amino group and the negatively charged *2-O*-sulfo group. The ~8 Hz axial-axial proton coupling constants of the III residue remains unchanged indicating that the GlcA2S conformation remains predominately in the $^4C_1$ form in Construct **2** and is the same as GlcA in Construct **1** [206].

To reveal the structural flexibility of GlcA2S and IdoA2S residues, we compared the proton spectra of GlcA2S-containing hexasaccharides with IdoA2S-containing hexasaccharides, focusing on the signals of the Residue III. The measurements of $^3J_{HH}$ coupling constant reflect the change of saccharide conformation. The coupling constant (doublet, 8.0 Hz) of III-1 signal of GlcA2S in construct **2** remains same as that in construct **3** and **4**, suggesting that the conformation of GlcA2S is relatively unaffected by the extra sulfo groups (Figure 33). This result is consistent with that GlcA2S predominately maintain the $^4C_1$ conformation as GlcA in construct **1** [206]. We observed that broad signals for IdoA2S (residue III) (Figure 34). The addition of EDTA significantly sharpened the IdoA2S signals. Similarly, the coupling constants of III-5 signals are also sensitive to different sulfation patterns, suggesting that conformational distribution of IdoA2S is affected by different sulfation pattern. For example, the $^3J_{HH}$ coupling constants of III-1 (2.2 Hz) and III-5 (2.2 Hz) signal in construct **5** are increased to (3.7 Hz) and (3.1 Hz) in construct **8** due to the presence of additional sulfo groups. These changes are attributed to greater structural flexibility of IdoA2S residue compared to GlcA2S in these constructs. This result is consistent with the previously reported that IdoA2S equilibrates between $^2S_0$ (skew boat) and $^1C_4$ (chair) conformers [82, 206].

The presence of GlcA2S or IdoA2S also impacted the chemical environment of neighboring saccharide residues. We compared the signals from each proton of GlcA2S- and IdoA2S-containing hexasaccharides (Table 6). As expected, signals from Residue III displayed the largest difference in chemical shifts. Outside Residue III, significant differences in the chemical shifts were observed for the protons of Residue II (II-1, II-3, II-6), and protons of Residue IV (IV-1 and IV-5) (Table 3). For example, comparing the IV-1 proton in construct **2** and **5**, the signals are shifted by 0.34 ppm. Similarly, between construct **3** and **6** the IV-1 proton was shifted by 0.32 ppm and 0.22 ppm between construct **4** and **7**, respectively. These data suggest the presence of GlcA2S or IdoA2S has different effects on the chemical environment of neighboring residues. Beyond Residue II and IV, chemical shift differences between GlcA2S- and IdoA2S-hexasaccharides diminished significantly. Some differences in chemical shifts were observed from Residue VI, an *N*-unsubstituted glucosamine residue in construct **2** and **5**, suggesting that the presence of a primary amino group ($-NH_2$) of this end residue is sensitive to GlcA2S vs IdoA2S residues.

	1	2	3	4	5	6a	6b
Construct **2** vs construct **5**							
I							
II	[+0.07]		[−0.08]				[−0.12]
III	[+0.53]	[+0.19]	[+0.20]	[+0.19]	[+0.97]		
IV	[−0.34]				[+0.09]		
V							
VI	[+0.07]	[+0.14]	[+0.07]				
Construct **3** vs construct **6**							
I					[−0.09]		
II	[+0.05]		[−0.07]				[−0.12]
III	[+0.55]	[+0.20]	[+0.23]	[+0.15]	[+1.02]		
IV	[−0.32]				[+0.08]		
V							
VI							
Construct **4** vs Construct **7**							
I					[−0.08]		
II	[+0.09]		[−0.06]				[−0.16]
III	[+0.48]	[+0.18]	[+0.23]	[+0.23]	[+1.01]		
IV	[−0.22]				[+0.06]		
V							
VI							

Table 6. ^{1}H NMR chemical shift difference (Δδ) by subtracting GlcA2S- from IdoA2S-containing hexasaccharides. Listed values represent chemical shift difference ≥0.05 ppm.

Intramolecular Effects by N-sulfation, 6-O-sulfation, and 3-O-sulfation

Comparing the chemical shifts of the signals from hexasaccharides bearing *N*-unsubstituted versus *N*-sulfated Residue VI suggests the intramolecular interaction between the *N*-sulfo groups and the GlcA residue at its reducing end. The chemical shifts of the V-3 signals in construct **2** and **3** are differed by +0.08 ppm; similarly, the signals of V-3 proton in construct **5** and **6** are different by +0.10 ppm (Table 4 and 5). These effects may attribute to the function of acidic sulfo group masking the basic amino group on VI residue and altering the chemical

environment around the V residue. These observations also support by the molecular dynamics simulation result that an interaction exits between the *N*-sulfo group of Residue VI and the V-3 position of Residue V as proposed by Langeslay and colleagues through the analysis of the structure of fondaparinux [207]. Another interesting observation is the impact of *N*-sulfation on the chemical shift of VI-2 signals. In GlcA2S-containing hexasaccharides, *N*-sulfation shifts the VI-2 signal to higher ppm (+0.06 ppm) probably due to the effects of de-shielding (less local electron density). By contrast, in IdoA2S-containg hexasaccharides, the *N*-sulfation shifts the VI-2 signal to lower ppm (-0.07 ppm) due to increased shielding (more electron density). Such different effects suggest more complicated through space interactions between the saccharide residues, which could involve in the participation of IdoA2S or GlcA2S residue; however, the available techniques do not permit us to dissect these interactions.

As expected, 6-*O*-sulfation clearly shifts the proton resonances of Residues II, IV and VI in both GlcA2S-containg and IdoA2S-containing hexasaccharides (Table 4 and 5). The presence of 6-*O*-sulfation also affects the chemical environments of the neighboring residues. For instance, Residue V in construct **4** is flanked by two GlcNS6S resides (Residue IV and VI)**,** the signal of V-1 is shifted by +0.08 ppm in comparison with the signal of construct **3**. Similarly, a chemical shift difference for the V-1 proton was also observed (+0.07 ppm) between construct **6** and **7** (Table 1 and 2). These data suggest that the impact of 6-*O*-sulfation on the neighboring residues can be observed in both GlcA2S-containing and IdoA2S-containing hexasaccharides.

The availability of 3-*O*-sulfated construct **8** and non-3-*O*-sulfated construct **7** differentiates the impact of 3-*O*-sulfation on the distribution of IdoA2S residue conformers. The 3-*O*-sulfation on Residue IV alters the conformation of its reducing-end IdoA2S residue (Residue III) as demonstrated by changes in the coupling constant of IdoA2S signals. For

example, the $^{3}J_{HH}$ coupling constants of III-1 and III-5 are larger in construct **8** (3.7 Hz, 3.1 Hz) than in construct **7** (3.0 Hz, 2.7 Hz) (Supplementary Fig S5), suggesting an increased proportion of $^{2}S_{0}$ conformation for Residue III in construct **8**. Indeed, Guerrini and colleagues demonstrated that 3-*O*-sulfated glucosamine residue influences the IdoA2S residue conformational equilibrium between $^{1}C_{4}$ and $^{2}S_{0}$ [208].

Antithrombin Binding and Anti-Xa Activity of Less Abundant Heparan Sulfate

Contribution of GlcA2S and IdoA2S to the Binding Affinity to Antithrombin

Heparin binds antithrombin to exert its anticoagulant activity. Within heparin a unique pentasaccharide sequence, -GlcNS6S-GlcA-GlcNS3S6S-**IdoA2S**-GlcNS6S-, has been shown to display high binding affinity. Previous work has suggested that IdoA2S in this pentasaccharide domain contributes to the binding to antithrombin [209]. To confirm this conclusion, the binding affinity to antithrombin of the IdoA2S versus GlcA2S constructs was determined. Both 3-*O*-[^{35}S]sulfated GlcA2S-containing hexasaccharide and IdoA2S-containing hexasaccharide were prepared for the analysis. The result showed that IdoA2S-containing hexasaccharide binds to antithrombin with a dissociation constant (K_d) of 12 ± 3 nM, whereas no binding affinity was measurable for GlcA2S-containing hexasaccharide. Our result confirmed the essential role of IdoA2S in antithrombin binding.

The chemoenzymatic approach allows the construction of diverse hexasaccharide analogs containing antithrombin-binding domain to confirm the contribution of IdoA2S and GlcA2S residue to the binding affinity to antithrombin. Our data suggest that GlcA2S-containing hexasaccharide does not bind to antithrombin despite the fact it carries 3-*O*- and 6-*O*-sulfo

groups, confirming the previously reported essential role of IdoA2S for high affinity binding to antithrombin [10, 210, 211].

The NMR structural analysis data are consistent with the hypothesis that GlcA2S is a relatively structural rigid unit in heparin and HS compared to IdoA2S. Compared with the signals of the IdoA2S residue, the $^{3}J_{HH}$ coupling constants are unchanged for the GlcA2S residue. The $^{3}J_{HH}$ coupling constants of GlcA ring protons in all constructs are about 8.0 Hz as reported previously as $^{4}C_{1}$ form [206]. In our studies, the coupling constant of GlcA2S ring protons in the synthesized hexasaccharide show equal values, supporting the conformation of GlcA2S remains predominately in the $^{4}C_{1}$ form as well. In contrast, the protons of IdoA2S display smaller and various coupling constants in the constructs, consistent with the fact that the conformations of the IdoA2S residue are in the equilibrium between $^{2}S_{0}$ and $^{1}C_{4}$ forms.

The $^{2}S_{0}$ conformation of IdoA2S is known to contribute to antithrombin binding affinity [210-212]. Indeed, IdoA2S-containing hexasaccharide displayed high affinity with a K_{d} value of 12 nM, while no detectable antithrombin-binding affinity for GlcA2S-containing hexasaccharide was observed. Importantly, the conformation of the IdoA2S residue can be affected by the protein that interacts with it as well as the neighboring sulfated saccharide residues. For example, binding of fibroblast growth factor 1 (FGF1) requires $^{1}C_{4}$ form of IdoA2S that is flanked by GlcNS6S residues [213-215]; however, in the absence of nearby GlcNS6S residue, IdoA2S could rearrange its conformation from $^{2}S_{0}$ to $^{1}C_{4}$ form which binds FGF2 [216, 217]. Possibly, the IdoA2S conformation flexibility plays a role in orienting the oligosaccharide to increase its binding affinity to a specific protein.

Discussion

Using highly purified GlcA2S- and IdoA2S-containing hexasaccharides synthesized by the chemoenzymatic approach, we successful fully assign all the proton and carbon chemical shifts of GlcA2S and IdoA2S-containing oligosaccharides through 1D and 2D NMR experiments. We compared the assignments obtained from our study with the previously reported values. The proton assignments of GlcA2S signals in this study are similar to the previous reports, where the GlcA2S residue is flanked by GlcNS6S and GlcNS residue in a tetrasaccharide [208, 218-220]. The anomeric carbon assignment of GlcA2S was also reported [77], and our result is consistent with the reported values. In the reports, only partial assignments were achieved for GlcA2S residues. Here, we provide the full assignments for the GlcA2S residue within different sulfated saccharide contexts. The assignments for IdoA2S are consistent with previous reports; however an IdoA2S residue flanked by two GlcNS3S6S residues show significant difference on proton and carbon assignments [208]. The GlcNS3S6S residues may participate in the change of chemical environment and conformation equilibrium of neighboring IdoA2S residue, causing significant changes in chemical shifts.

Taking advantage of these chemical shift assignments, the presence of GlcA2S can be easily distinguished from mixture of GlcA2S- and IdoA2S-containing oligosaccharides within NMR spectra. 2D ^{1}H-^{13}C HSQC NMR is a powerful tool for fingerprint the GlcA2S residue, and the ^{1}H-^{13}C correlation of III-1 and III-2 position signals from this residue is easily observable (Figure 37). The cross peak of GlcA2S anomeric signal is also consistent to the signal observed in enoxaparin [77] or porcine intestinal heparin [221], demonstrating the capability of using the 2D-NMR fingerprint method to confirm the authenticity of heparin drugs. Very interestingly, the anomeric signals of GlcA2S were reportedly difficult to identify in unfractionated heparin

and low-molecular weight heparins; however, such signals are visible in enoxaparin, a low-molecular weight heparin manufactured by depolymerized unfractionated heparin under basic conditions [77, 222].

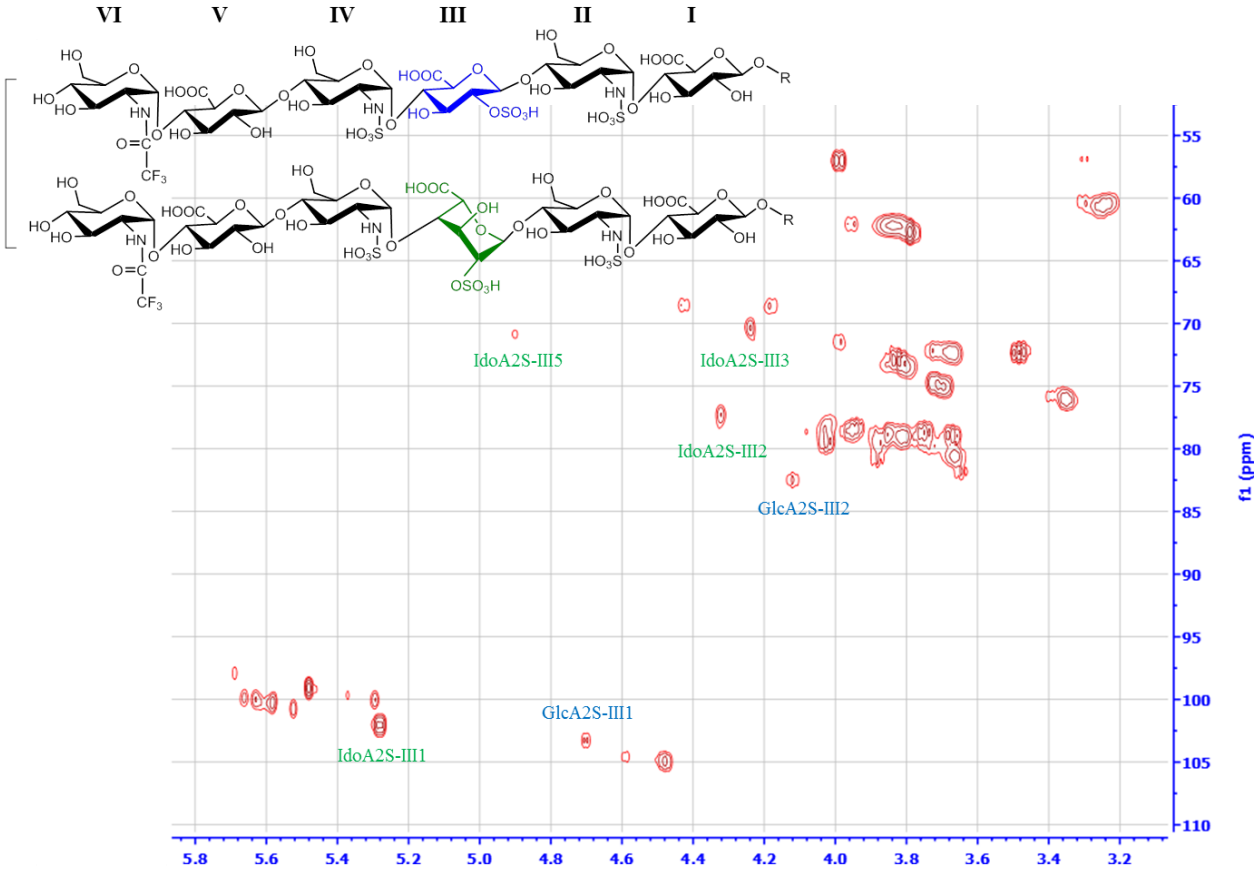


Figure 37. 2D ^{1}H-^{13}C HSQC NMR spectrum of GlcA2S- and IdoA2S-containing oligosaccharide mixture. 2D ^{1}H-^{13}C HSQC NMR experiment was employed to distinguish the presence of GlcA2S and IdoA2S residue from mixture of GlcA2S- and IdoA2S-containing oligosaccharides. The chemical structure of oligosaccharide mixture is shown on top of the Figure. The distinguishable signals of GlcA2S and IdoA2S residue are highlighted in different colors. The GlcA2S residue and its signals are highlighted in blue color. The IdoA2S residue and its signals are highlighted in green color.

Since the 2008 contaminated heparin crisis, NMR analysis of heparin drug has become part of the heparin sodium US Pharmacopeia monograph. Because heparin is a heterogeneous polysaccharide mixture, the assignment of all of the signals is difficult. To this end, acquiring the NMR spectra of highly purified heparin oligosaccharide standards provides a map for assigning

highly complex heparin spectra, especially those signals from rare components. Our study should assist NMR analysis of heparin to safeguard the heparin drug.

Conclusions

Low molecular weight heparins (LMWHs) are carbohydrate-based anticoagulants clinically used to treat thrombotic disorders, but impurities, structural heterogeneity or functional irreversibility can limit treatment options. We report a series of synthetic LMWHs prepared by cost-effective chemoenzymatic methods. The high anti-FXa activity of structural defined synthetic LMWH was reversible *in vitro* and *in vivo* using protamine, demonstrating synthetically accessible constructs can play a critical role in the next generation of LMWHs.

Heparan sulfate and heparin are highly sulfated polysaccharides that consist of a repeating disaccharide unit of glucosamine and glucuronic or iduronic acid. The 2-*O*-sulfated iduronic acid (IdoA2S) residue is commonly found in heparan sulfate and heparin; however, 2-*O*-sulfated glucuronic acid (GlcA2S) and non-sulfated iduronic acid (IdoA) are less abundant monosaccharides (~<5% of total saccharides). Here, we report the synthesis these less abundant HS component molecules including GlcA2S-containing (Chapter III) and IdoA-containing hexasaccharides (Chapter V) using a chemoenzymatic approach. For comparison purposes, additional IdoA2S-containing hexasaccharides were synthesized. NMR analyses were performed to obtain full chemical shift assignments for the GlcA2S-, IdoA- and IdoA2S-hexasaccharides. These data show that GlcA2S is a more structurally rigid saccharide residue than IdoA and IdoA2S. The antithrombin binding affinities of synthetic oligosaccharides were determined by affinity co-electrophoresis. In contrast to IdoA- and IdoA2S-hexasaccharides, the

GlcA2S-hexasaccharide does not bind to antithrombin, confirming that the presence of IdoA or IdoA2S is critically important for the anticoagulant activity. The availability of pure synthetic common and less abundant HS oligosaccharides can not only serve as referencing standards to secure the drug components, but also allow the investigation of the structure and activity relationship of individual molecules of heparin or heparan sulfate.

Chapter IV

Utilize Heparan Sulfate Based Probe to Investigate Interaction with Proteins[1]

Substrate Specificity Revealed by Crystal Structure of 2-O-Sulfotransferase

Introduction

The heparan sulfate 2-*O*-sulfotransferase (2-OST) is a key enzyme in the HS biosynthetic pathway. 2OST is responsible for the transfers of a sulfo group to the 2-OH position of preferentially IdoA, and to a limited extent GlcA, within HS polysaccharides to form 2-*O*-sulfated iduronic acid (IdoA2S) or 2-*O*-sulfated glucuronic acid (GlcA2S). The IdoA2S unit is a common structural motif found in cellular HS, and is required for binding several growth factors and is essential for triggering growth factor–mediated signal transduction pathways [223, 224]. The critical physiological functions of IdoA2S residues have been revealed through numerous studies.

Better understanding of the enzymes used in the chemoenzymatic approach should aid in the production of heparin-based therapeutics tailored for specific disease targets. However, the cellular mechanism regulating HS biosynthesis is largely unknown. Understanding the substrate

[1]Part of this chapter previously appeared as two articles in the Journal of Biological Chemistry and Chemical Communications. The original citations are as follows: C. Liu, J. Sheng, J. Krahn, L. Perera, Y. Xu, P. H. Hsieh, J. Liu, L. Pedersen, "Molecular mechanism of substrate specificity for heparan sulfate 2-O-sulfotransferase, *Journal of Biological Chemistry*," 2014, 13407–13418; W. Zhou‡, P. H. Hsieh‡, Y. Xu, T. R. O'Leary, X. Huang, J. Liu, "Design and synthesis of active heparan sulfate-based probes," *Chemical Communications*, 51, 2015, 11019-11021 (‡contributed equally)

specificities and mechanisms of HS biosynthetic enzymes is crucial for dissecting this fundamental process. In this article, we report a crystal structure of a ternary complex of 2OST with 3'-phosphoadenosine 5'-phosphate (PAP) and a heptasaccharide acceptor substrate (Figure 38).

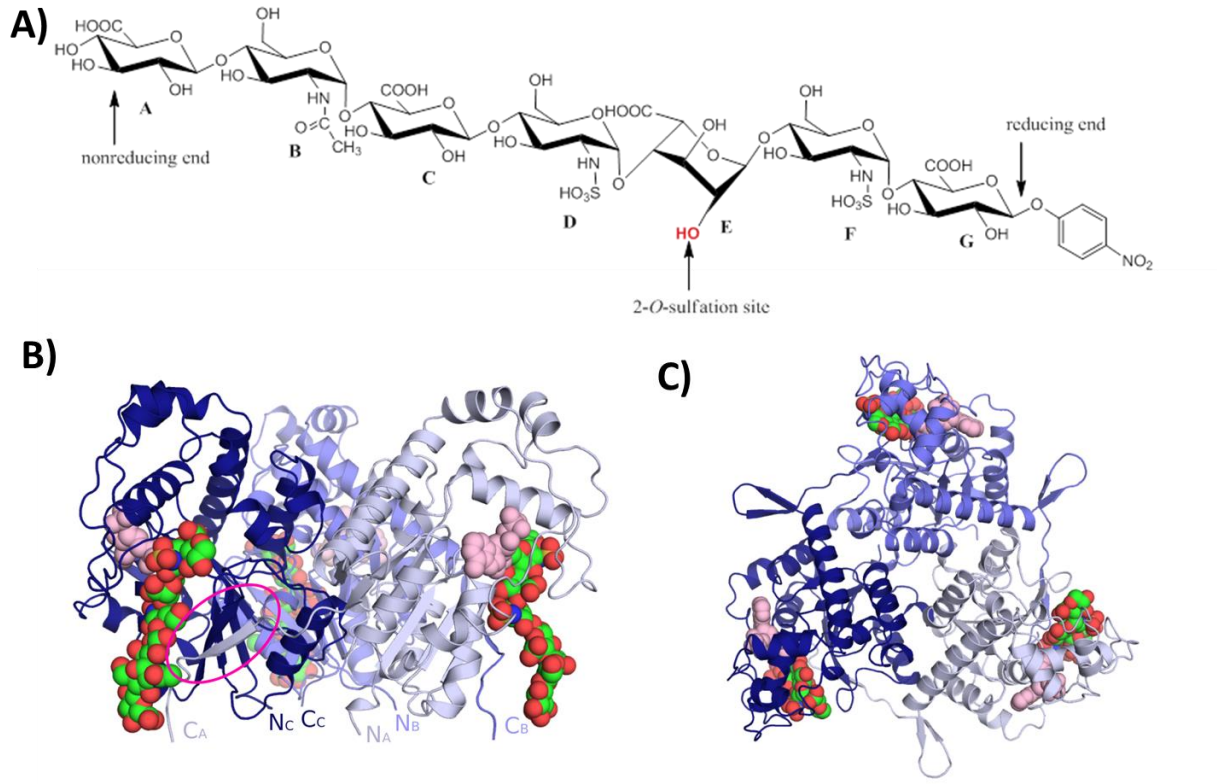


Figure 38. Crystal structure of the 2-OST trimer. A) Chemical structure of the heptasaccharide substrate used in the ternary complex. B) Crystal structure of the 2OST trimer with each monomer (colored dark blue, slate, light blue) bound to a molecule of PAP (pink) and heptasaccharide substrate (green) (monomers D,E and F). The C-terminus of each subunit (circled in magenta) extends into the active site of the neighboring molecule. C) Figure 1B rotated 90° with respect to the horizontal axis in the plane of the paper.

NMR Analysis of Heptasaccharide Substrate

The structure of oligosaccharides constructs and intermediates were analyzed by NMR experiments, including 1D- (^{1}H and ^{13}C), 2D- (^{1}H-^{1}H COSY, TOCSY, ^{1}H-^{13}C HSQC) NMR. NMR experiments were performed at 298 K with a Varian Inova 500 MHz spectrometer

equipped with 5mm triple resonance XYZ or broadband PFG probe and processed by VnmrJ 2.2D software. Samples (2 mg) were each dissolved in 0.5 mL D_2O (99.994%, Sigma, Co.) and lyophilized three times to remove the exchangeable protons. The samples were re-dissolved in 0.5 mL D_2O and transferred to NMR microtubes (OD 5 mm, Norrell). 1D ^{1}H-NMR experiments were performed with 256 scans and an acquisition time of 768 msec. 1D ^{13}C-NMR experiments were performed with 40000 scans, 1.0 sec relaxation delay, and an acquisition time of 1000 msec. 2D (^{1}H-^{1}H COSY, TOCSY, ^{1}H-^{13}C HSQC) spectra were recorded with carbon decoupling, 1.5 sec relaxation delay, during 204 msec acquisition time with 500 increments for 48 scans.

Taking advantage of 1D and 2D NMR experiments, we are able to assign anomeric protons and chemical shifts of saccharide E with GlcA or IdoA in the heptasaccharides. Series of IdoA proton signals are E1 (4.94 ppm, very broad), E5 (4.76 ppm, very broad), E3 (4.09 ppm, broad), E4 (4.04 ppm, broad), and E2 (3.72 ppm, broad) from downfield to upfield. HSQC spectra of compound confirmed the presence of IdoA residue in compound. In the ^{1}H NMR spectrum of compound, the proton peaks of IdoA also showed as smaller coupling constants or broad shape, not typically larger axial-axial coupling constants in GlcA, supporting the existence of IdoA and possibly fast conformational equilibrium among chair ($^{4}C_1$ and $^{1}C_4$) and skew-boat ($^{2}S_0$) conformers.

Crystal Structure of 2OST with Bound Heptasaccharide

Here, we report the crystal structure of the ternary complex of 2OST, 3'-phosphoadenosine 5'-phosphate, and a heptasaccharide substrate. Utilizing site-directed mutagenesis and specific oligosaccharide substrate sequences, we probed the molecular basis of specificity and 2OST's position in the ordered HS biosynthesis pathway.

Based on the best fit to the electron density, the acceptor saccharide was modeled as an IdoA in the $^4C_1$ conformation (Figure 39). Because the resolution (3.45 Å) does not allow for unambiguous identification of the sugar or ring conformation at the acceptor site, molecular dynamic simulations were carried out on all four sugar conformations for 25ns. The conformations $^1C_4$ and $^2S_0$ have been previously reported as preferred conformations for HS structures[187, 225]. Among the three possible conformations for IdoA, the $^4C_1$ conformation displayed the strongest binding energy (ΔE) (Table 2), supporting the crystallographic assignment of saccharide E as an IdoA in the $^4C_1$ conformation while binding 2OST. However, given the relatively small range for the binding energies of the IdoA conformations, coupled with the inherent errors in the calculation of binding energy for this system, it is difficult to rule out any of the three IdoA conformations.

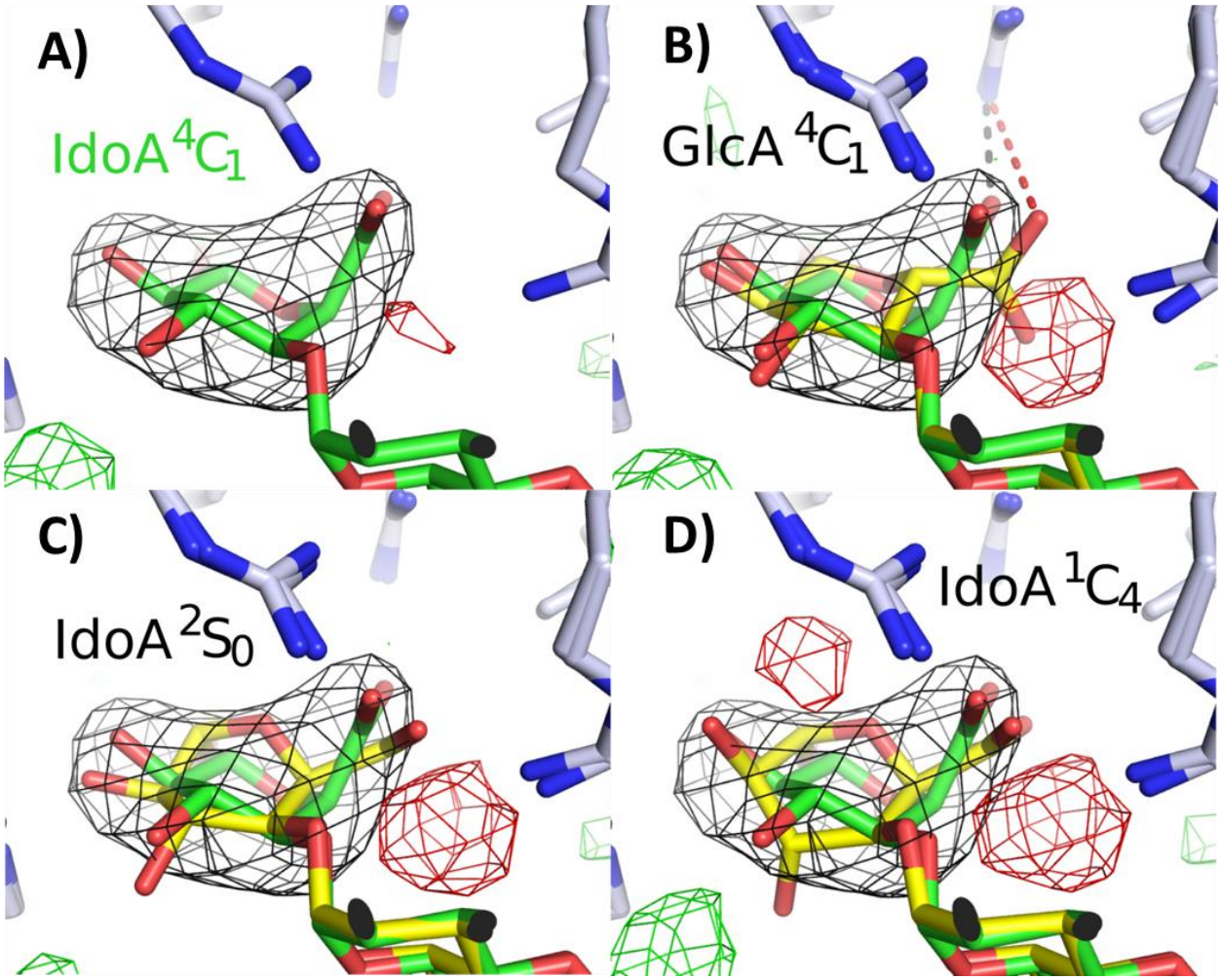


Figure 39. Refinement of different conformations of the acceptor saccharide. A) Refinement of saccharide E as Ido A in the $^4C_1$ conformation (green). B,C,D) Refinement of saccharide E as either GlcA in the $^4C_1$ conformation, IdoA $^2S_0$, or IdoA $^1C_4$ respectively (yellow) superimposed with the IdoA $^4C_1$ structure (green). The simulated annealing omit Fo-Fc map contoured at 4.0 σ is displayed in black. Positive *F*o-*F*c density from each respective refinement is shown in green (3.0 σ) and negative density (3.0 σ) is shown in red. In figure B the black dashed line represents a salt bridge (2.9 Å) between the IdoA and Arg189. For GlcA this distance is 4.0 Å (red dashed line) and thus likely does not form a salt bridge.

This study indicates that a pentasaccharide is the minimum size required for sulfation by 2OST. The crystal structure suggests saccharides C-G of the heptasaccharide form all of the observed specific interactions with the protein. These studies revealed that Arg80, Lys350, and Arg190 of 2OST interact with the *N*-sulfo groups near the modification site, consistent with the dependence of 2OST on *N*-sulfation. Within the required pentasaccharide 2OST recognizes a

trisaccharide motif with the structure of –GlcNS-GlcA-GlcNS- or –GlcNS-IdoA-GlcNS-. The acceptor IdoA or GlcA residue must be flanked by two *N*-sulfo glucosamine residues.

In contrast, 6-*O*-sulfation on the flanking GlcNS residues of the acceptor site abolishes the reactivity to 2OST modification. Based on the crystal structure, 6-*O*-sulfation on saccharide F, on the reducing side of the acceptor saccharide, would create steric clashes with Pro82 and/or Tyr173, interfering with its ability to bind within the substrate binding cleft (Figure 40A). 6-*O*-sulfation is also likely selected against on the nonreducing adjacent glucosamine due to charge repulsion with Glu349 from the neighboring monomer (Figure 40B). 6-*O*-sulfo groups on HS are likely excluded by steric and electrostatic repulsion within the active site supporting the hypothesis that 2-*O*-sulfation occurs prior to 6-*O*-sulfation.

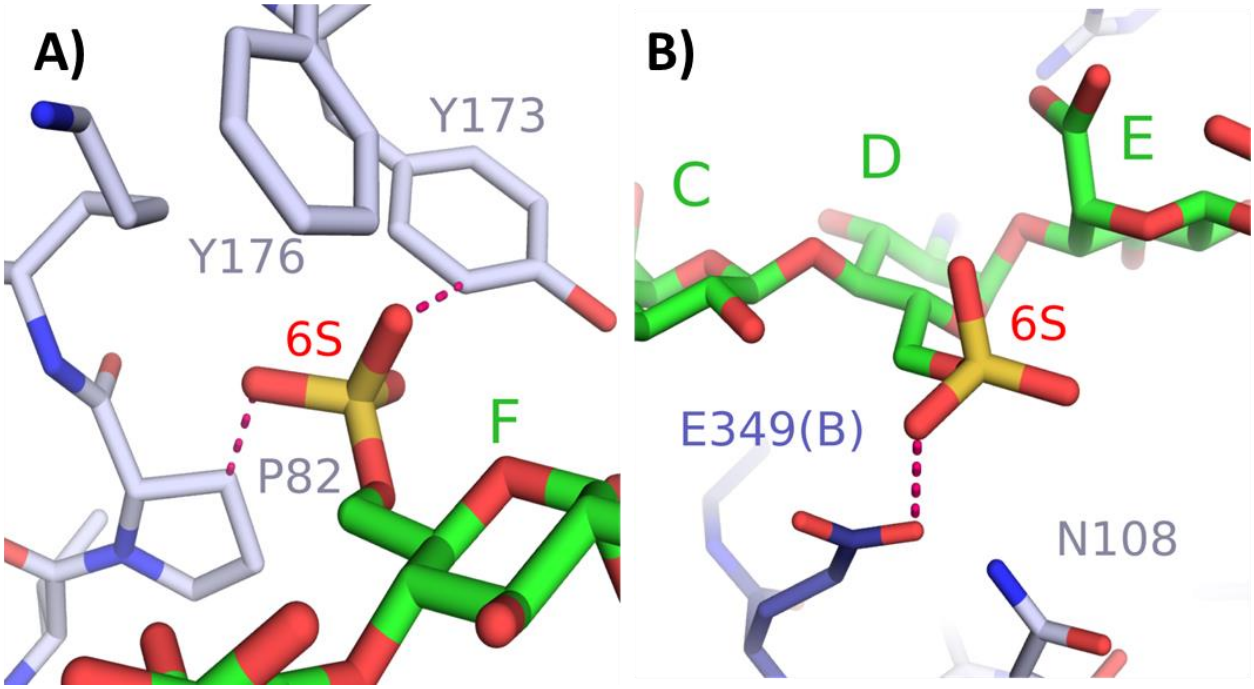


Figure 40. Modeling of 6-*O*-sulfates on the substrate heptasaccharide. Modeling of 6-*O*-sulfates on the substrate heptasaccharide. A) A 6-*O*-sulfate on saccharide F would result in steric clashes with Y173 (2.2 Å) and P82 (2.4 Å). B) 6-*O*-sulfate on saccharide D would likely create electrostatic repulsion with E349 (3.0 Å).

Our conclusion is consistent with the current sequential model for the HS biosynthetic pathway with *N*-sulfation occurring first, followed by C_5-epimerization/2-*O*-sulfation, 6-*O*-sulfation and 3-*O*-sulfation respectively. This study provides the structural evidence for understanding the sequence of enzymatic events in this pathway.

Synthesis of Heparan Sulfate Oligosaccharides Carrying Diazoacetyl Group

Introduction

Heparan sulfate (HS) is a highly sulfated polysaccharide abundantly present on the cell surface and in the extracellular matrix. It participates in a wide range of physiological and pathophysiological functions, including embryonic development,[226] inflammatory responses,[227] blood coagulation,[228] and viral/bacterial infections.[229] HS consists of a disaccharide repeating unit with glucuronic acid (GlcA) or iduronic acid (IdoA) and glucosamine, each of which is capable of carrying sulfo groups. HS achieves its functions by interacting with a variety of proteins.[230] The positions of sulfo groups and the locations of the IdoA residues are critically important for their binding specificities.[231] There is strong demand for methods to decipher the interactions of specific HS structures with their protein targets.

Here, we report a new approach for synthesizing active heparan sulfate based probes (AHSBP). The probe consists of an HS oligosaccharide functionalized with an *N*-diazoacetyl moiety (Figure 41). It utilizes the saccharide motif as a guiding system to find the protein target site. The diazoacetyl group,[232] while stable under neutral pH, upon binding and acidification, could be activated to covalently couple the oligosaccharide with the protein and inhibit protein

activity. The effects of AHSBP on two HS biosynthetic enzymes, including 2-*O*-sulfotransferase (2-OST) and 3-*O*-sulfotransferase (3-OST), were examined to demonstrate the utility of these compounds.

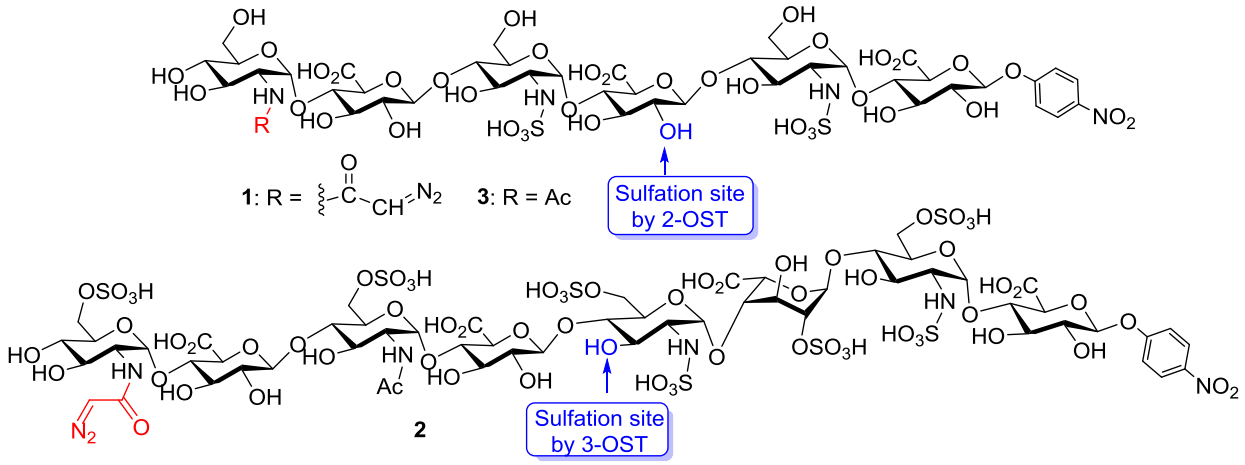


Figure 41. Structures of the synthesized AHSBP. **1** and **2** are the substrates for 2-OST and 3-OST respectively. The hydroxyl groups that can be modified by 2-OST and 3-OST are indicated**.**

Synthesis and Characterization of Heparan Sulfate Based Probe and Hydrolysis Product

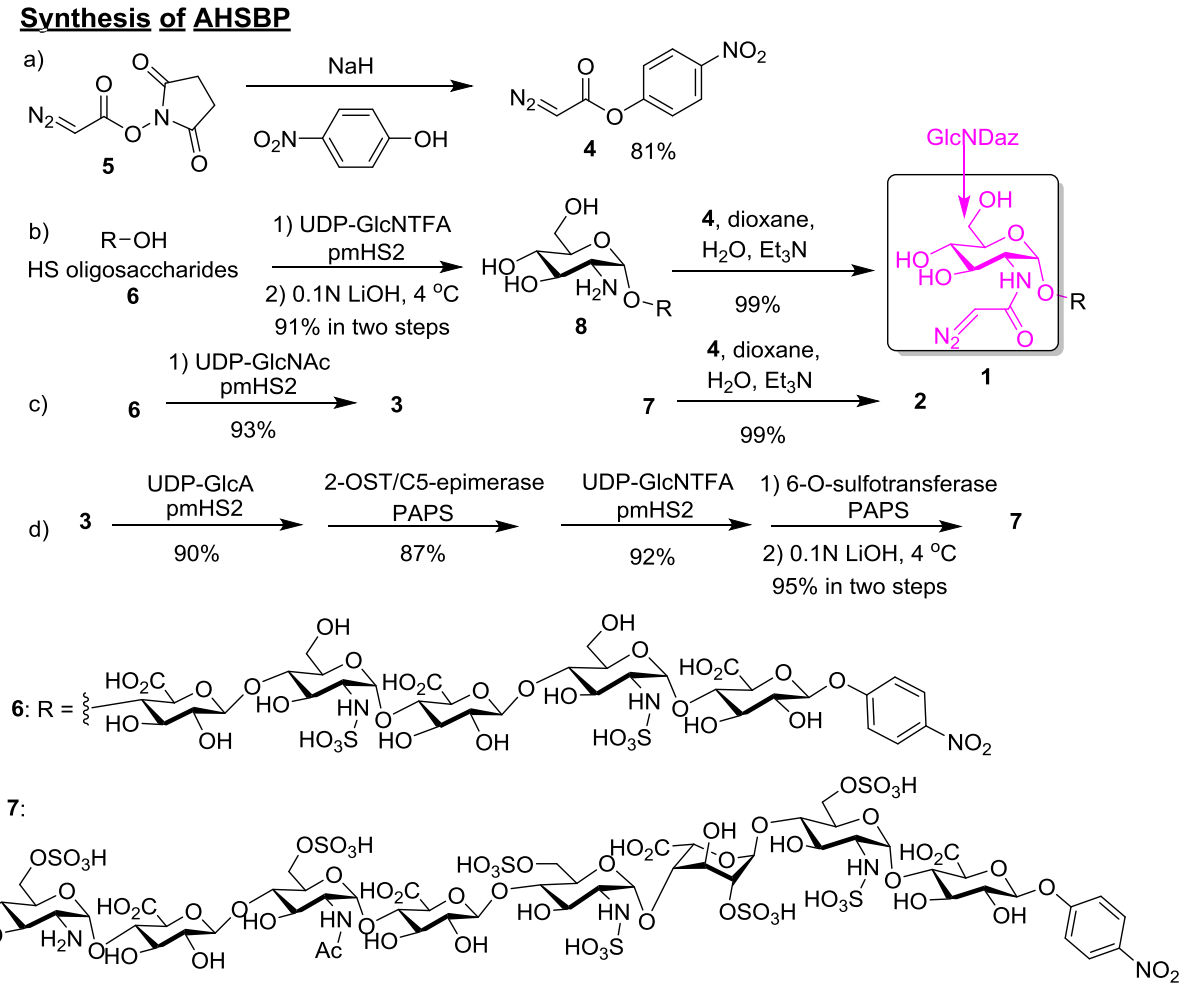


Figure 42. Synthesis scheme of active heparan sulfate based probes (AHSBP).

Synthesis of AHSBP started from the preparation of diazo acylating agent *p*-nitrophenyl diazoacetate **4**, which was obtained from the *N*-hydroxysuccinimide diazoacetate **5**[233] in 81% yield (Figure 42a).[233] The oligosaccharides bearing diazoacetyl functionalized glucosamine (GlcNDaz) residues were synthesized through a chemoenzymatic approach (Figure 42). The optimized condition was to maintain the molar ratio of compound **4** and the HS oligosaccharide at a 10:1 ratio in a mixture of 1,4-dioxane/H_2O (2:1) using trimethylamine as the base. An attempt to use compound **5** to directly react with the *N*-unsubstituted glucosamine (**8**) failed under various conditions. This is probably due to the higher hydrolytic instability of **5** compared

to the *p*-nitrophenyl ester **4**. HS octasaccharide **2** was synthesized from **7** that was prepared from **3** through four enzymatic steps (Figure 42c and 42d). A control compound, hexasaccharide **3**, was also synthesized (Figure 42c). **3** has the identical number of saccharide residue and sulfo groups as **1**, except that a GlcNAc residue was used to substitute the GlcNDaz at the non-reducing end (Appendix Supplementary Figure 31-33).

While diazo compounds can be activated by light, we explored the alternative of their activation by acid. After **1** was incubated in water at pH 5.5 for one hour, one major product was obtained from the solution. Electrospray ionization mass spectrometry (ESI-MS) analysis demonstrated the product had the molar mass of 1369.3 indicating addition of a water molecule and the loss of nitrogen gas (Appendix Supplementary Figure 34-35). There are two possible mechanisms for product formation. The first is that the diazoacetamide underwent Wolff rearrangement with subsequent hydrolysis of the isocyanate leading to a carboxylic acid **10** (Figure 43a).[234, 235] The other pathway is that protonation of the diazoacetamide triggers the release of nitrogen, which is followed by nucleophilic attack by water generating the glycolamide **11** (Figure 43b), which is an isomer of **10**.

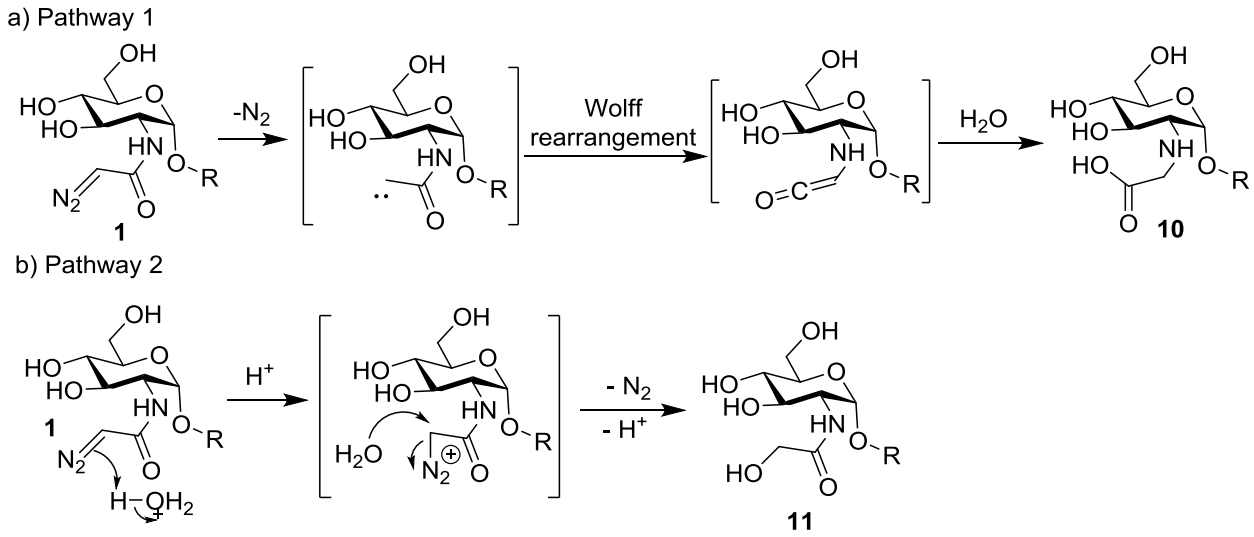


Figure 43. Two potential pathways for activation of the GlcNDaz residue.

To establish the product structure, two disaccharide standards (**12** and **13**) were synthesized (Appendix Supplementary Figure 36-41). **12** has the *N*-carboxymethyl glucosamine residue as in **10**, while **13** bears the *N*-glycolyl glucosamine residue, mimicking the structure of **11**. From ^{1}H-^{13}C-HMBC (homonuclear multiple bond correlation) spectra, key correlations were observed from these disaccharides (Appendix Supplementary Figure 39). The signals from cross peak I-2 (3.97 ppm)/I-7(177.9 ppm) from **13** were similar to that observed from product of **1** (I-2(3.93 ppm)/I-7(177.8 ppm)), hinting that **11** rather than **10** was the structure of main product from **1** (Figure 44). To finally confirm the structure, hexasaccharide **14** was synthesized from **8** (Appendix Supplementary Figure 42-45) and glycolic acid. ^{1}H-NMR spectrum of **14** matched well with that from acid treated **1** (Figure 44). Taken together, these results suggest that the major pathway for acid catalyzed conversion of **1** is through pathway 2 (Figure 43b) rather than Wolff rearrangement.

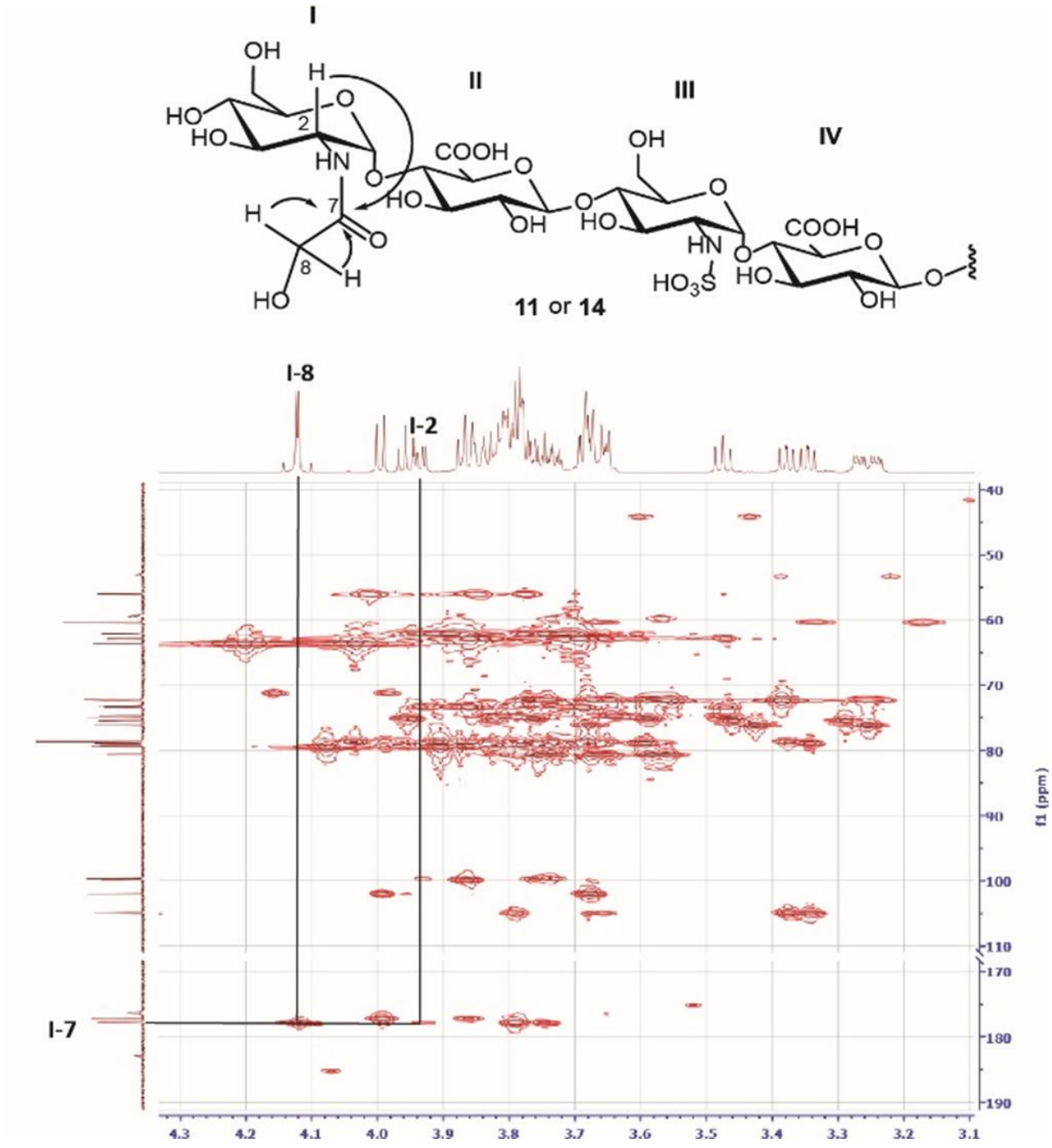


Figure 44. $^{1}H–^{13}C$ HMBC key correlations of compound 11. The signals of featured groups are indicated. The cross peaks are connected with arrowed lines identifying correlated carbon and proton. The corresponding chemical structure is shown on top of each spectrum. The anticipated correlations are indicated in the structures by curved arrows.

Using GlcNDaz-Containing Oligosaccharides as Active HS Based Probes (AHSBP)

The abilities of compounds **1** and **2** to function as AHSBPs (Appendix Supplementary Figure 46-53) were investigated next using two HS biosynthesis enzymes, i.e., 2-OST and 3-OST.[236] 2-OST is an enzyme that transfers sulfo group to the 2-OH position of a glucuronic acid residue that is flanked by two *N*-sulfated glucosamine residues (Figure 41). **1** was incubated with 2-OST at pH 5.5. The reaction was monitored with gel-permeation chromatography HPLC (GPC-HPLC) at two wavelength, 280 nm and 310 nm, characteristic for the protein and *p*-nitrophenyl chromophore, respectively. 2-OST (MW at around 70 KDa) was eluted at 39.3 min (Figure 45A), and **1** (MW = 1379.4 Da) was eluted at 56.3 min (Figure 45B). The reaction mixture yielded a major new peak at 28.6 min, earlier than 2-OST alone (Figure 45C). Compared to 2-OST, the new peak has significantly higher absorbance in 310 nm, suggesting that the coupling of 2-OST and **1** occurred. In a control experiment, incubation of 2-OST with compound **3** bearing no GlcNDaz, did not change the elution time for 2-OST (Figure 45D).

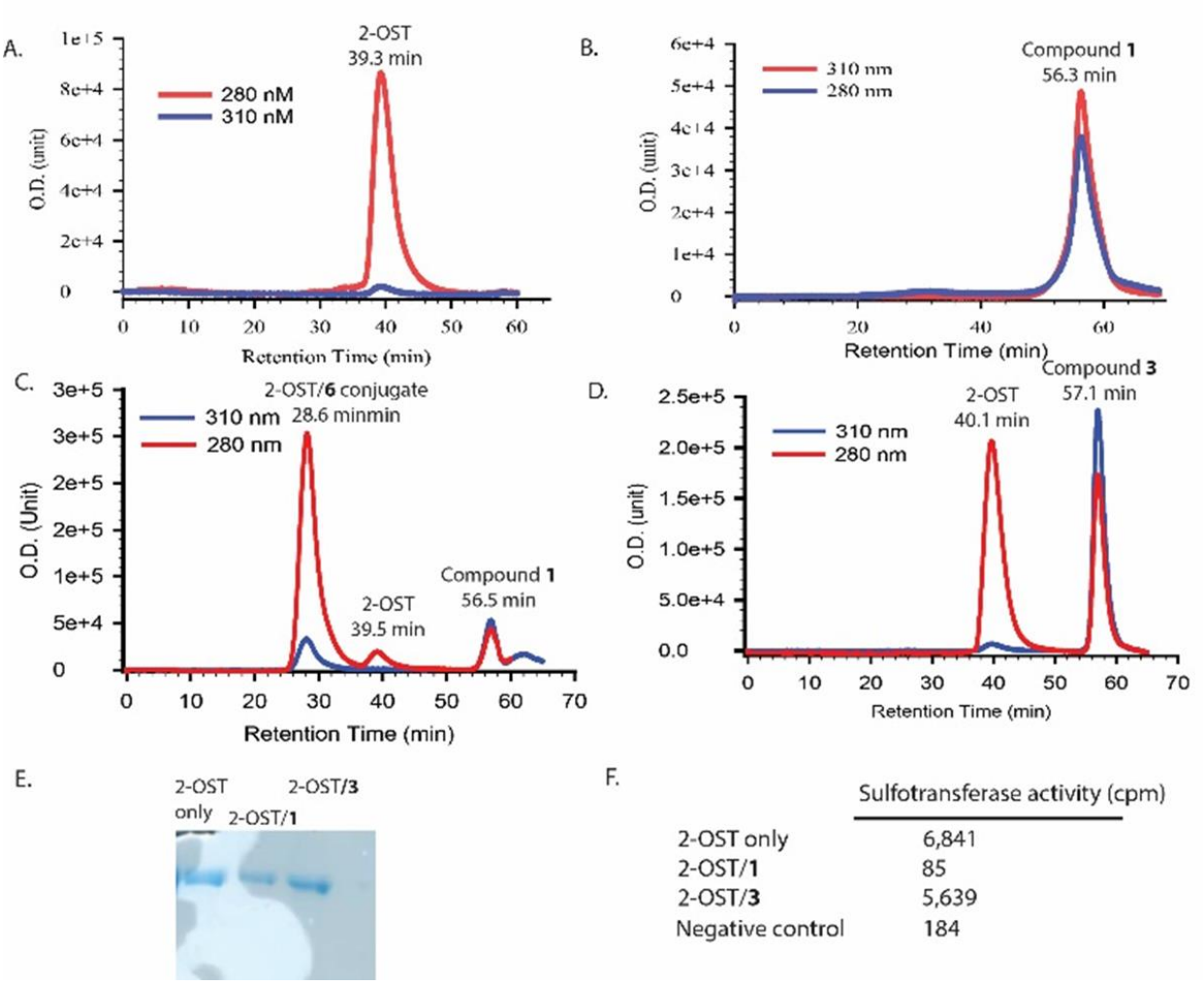


Figure 45. Analysis of the conjugate product of 2-OST and compound 1. Panel A shows the GPC-HPLC chromatogram of 2-OST protein only in the absence of compound **1**. Panel B shows the GPC-HPLC chromatogram of compound **1** only. Panel C shows the GPC-HPLC chromatogram after incubating 2-OST with compound **1**. Panel D shows the chromatogram of 2-OST and compound **3** after coupling reaction. Panel E shows the picture of PAGE analysis of three proteins samples from GPC-HPLC analysis. Panel F shows the results of the sulfotransferase activity measurement of three protein samples. The activity was determined based on the amount ^{35}S-labeled sulfo group transferred to a polysaccharide substrate.

Additional evidence suggesting the cross-linking between 2-OST and **1** was demonstrated by the effect on the enzymatic activity. The protein peaks resolved from GPC-HPLC were collected for the activity measurement. The presence of the desired protein in collected fractions from GPC-HPLC was confirmed by SDS-PAGE (Figure 45E). Both the protein alone and the

protein incubated with **3** displayed excellent sulfotransferase activities, while the sample incubated with **1** lost the enzymatic activity completely (Figure 45F). In contrast, incubation of 2-OST with **3** did not change the sulfotransferase activity. The reason for the lost activity of 2-OST/**1** conjugate is presumably because the hexasaccharide was covalently attached to or near the active site of the enzyme, preventing the substrate binding.

Inhibitory Effect on 2-O-Sulfotransferase and 3-O-Sulfotransferase

We next demonstrate that enzyme inhibition is based on the saccharide structures. 3-OST transfers a sulfo group to the 3-OH position of a glucosamine residue that is linked to a GlcA at the nonreducing end. 3-OST (3-OST-1) recognizes very different saccharide sequence from **1**, preferring substrates with high level of sulfations.[187] Based on the substrate specificity studies and crystal structures, only 2-OST can bind to the hexasaccharide **1**,[237] but not to **2.**[187] In contrast, 3-OST (3-OST-1) should bind to **2**, but not to **1.**[187] In this experiment, **1** or **2** was incubated with 2-OST and 3-OST separately under an acidic condition to allow the compound to react with the enzymes. The activity measurement for each enzyme was then performed to assess the impact of the AHSBP (Figure 46). **1** displayed much higher potency toward the inhibition of 2-OST (IC_{50} = 13 nM) than that for 3-OST (IC_{50} = 334 nM). In contrast, **2** potently inhibited the activity of 3-OST (IC_{50} = 2 nM), but not against 2-OST (IC_{50} > 50 nM). Our data suggest that **1** and **2** selectively inhibit their respective target enzymes. This is the first time that small molecule inhibitors display selective inhibition effects towards different HS biosynthetic enzymes.

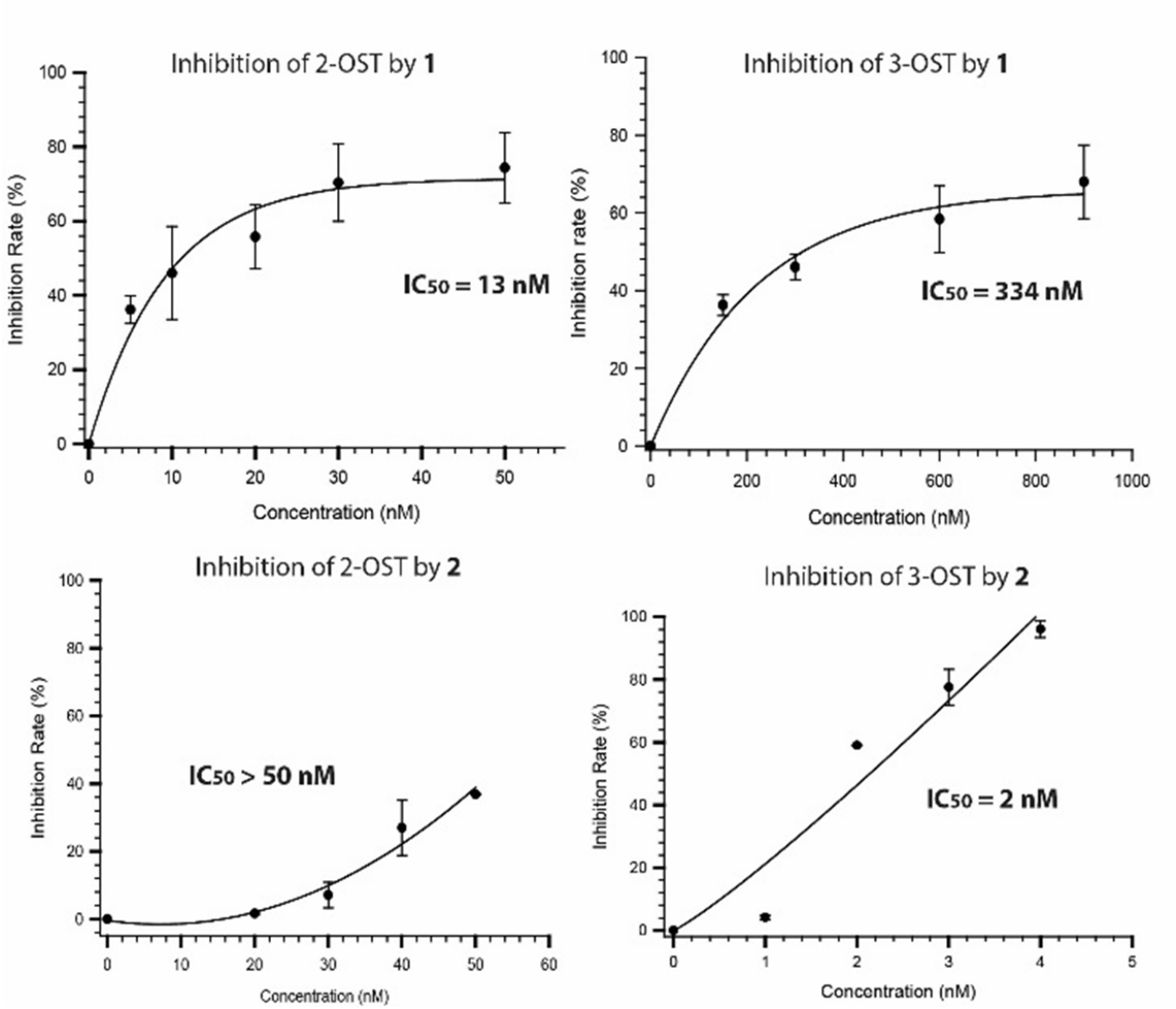


Figure 46. Inhibitory effects of compounds 1 and 2 on 2-OST and 3-OST. 2-OST, or 3-OST, was incubated with different concentrations of compound **1** or **2** in an acidic buffer at 4°C for 1h to allow the cross-linking between the compound and proteins.

Synthesis of ^{13}C-Isotope Labeled Heparan Sulfate Oligosaccharides

Introduction

The goal of this application is to develop a NMR-based method to study the conformation of heparan sulfate (HS) and the interaction of HS with proteins. HS exerts its biological functions by interacting with proteins; however, the mechanism used by HS to display the selectivity during these interactions is only partially understood. It has been proposed that the conformation of HS plays a critical role in binding affinity and selectivity toward proteins [238]; but this hypothesis remains to be unproven due to the lack of a tool that is able to analyze the conformation of HS accurately.

Although nuclear magnetic resonance (NMR) is a widely used method for studying molecular conformations, including carbohydrates[239], several factors limit its application to study the conformation of HS. First, the NMR signals from HS highly overlap because HS is composed of structurally similar repeating units. Second, sufficient amount of structurally defined oligosaccharides for comprehensive NMR analysis are difficult to obtain [240]. Here, we propose a new method to study the conformation of HS. The crucial innovation is to strategically introduce a ^{13}C-labeled residue in an oligosaccharide. The ^{13}C-labeled oligosaccharides were synthesized using a chemoenzymatic approach in adequate amounts. The ^{13}C-labeled saccharide residue improves the detection sensitivity by 100-fold using NMR. This method also simplifies the ^{13}C-signals from the oligosaccharide, provided that only the labeled saccharide residues are detectable by NMR. Computational simulations are employed to relate the NMR parameters to 3D structural properties. This advance will provide an essential tool to study HS and the interaction with proteins utilizing sophisticated NMR methods.

Chemoenzymatic Synthesis of ^{13}C-labeled Cofactor and Oligosaccharide

Two key reagents for preparing ^{13}C-labeled oligosaccharides are UDP-[^{13}C]GlcA and UDP-[^{13}C]GlcNTFA. The synthesis of both sugar nucleotides was completed using enzymatic methods following previous publications [241, 242]. The preparation of UDP-[^{13}C]GlcNTFA was started from uniformly-^{13}C-labeled glucosamine (Omicron, Inc.) followed by the protocol described previously in Chapter II. UDP-GlcNTA is made by chemoenzymatic approach (Figure 47). The preparation of UDP-[^{13}C]GlcA was started from ^{13}C-labeled glucose (Cambridge Isotope Laboratories, Inc.). The ^{13}C-labeled glucose was firstly converted to [^{13}C]Glucose-6-phosphate using hexokinase. The expression of the enzyme was carried out in *E. coli*. The [^{13}C]Glucose-6-phosphate was secondly converted to [^{13}C]Glucose-1-phosphate using phosphoglucomutase. Thirdly, the UDP-[^{13}C]Glucose synthesis was obtained by transforming [^{13}C]Glucose-1-phosphate using UDP-glucose pyrophosphorylase (GalU). Finally, the UDP-[^{13}C]Glucose was converted to UDP-[^{13}C]GlcA by using UDP-glucose dehydrogenase. The resultant UDP-[^{13}C]GlcA as well as UDP-[^{13}C]GlcNTFA were ready for the elongation reaction.

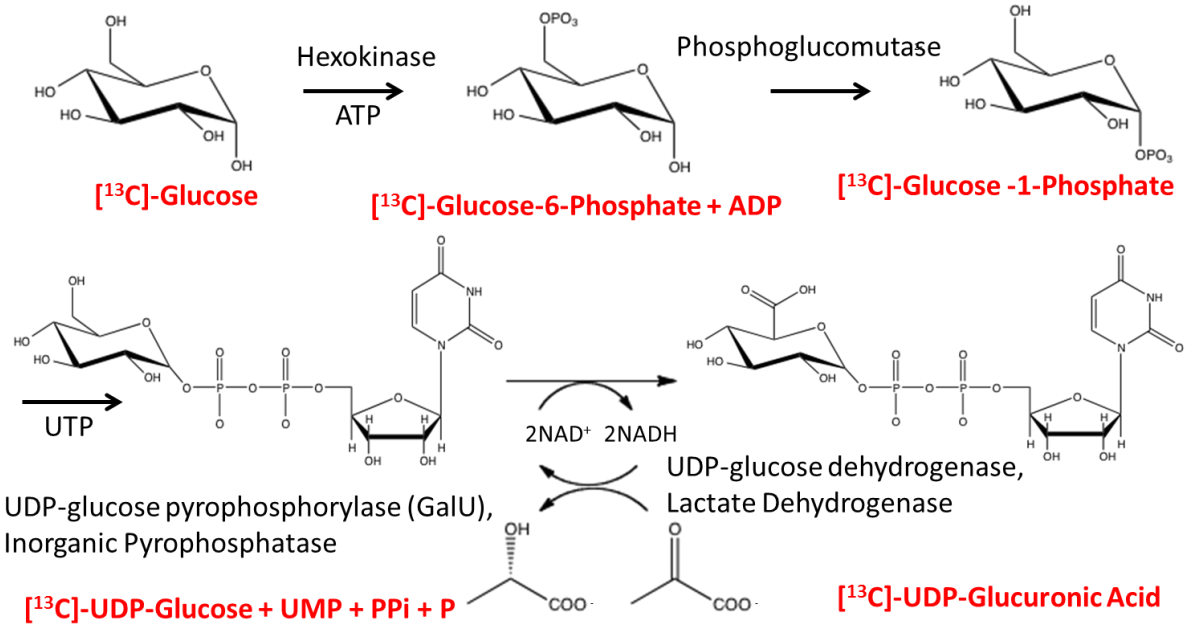


Figure 47. Synthesis of [^{13}C]-UDP-Glucuronic Acid.

The pilot synthesis of two hexasaccharides was completed following the scheme shown in Figure 48. Hexasaccharide A has the ^{13}C-labeled residue at the IdoA2S position; and hexasaccharide B has the ^{13}C-labeled saccharide residues at both GlcNS3S6S and IdoA2S positions. About 6.7 mg of hexasaccharide A and 3.5 mg of hexasaccharide B were obtained by far.

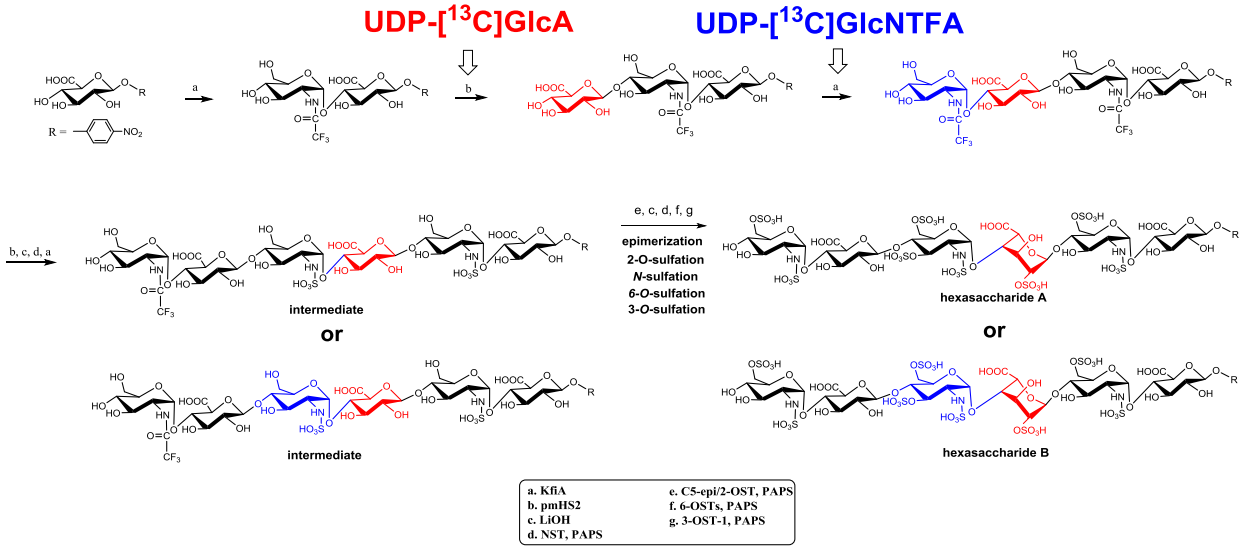


Figure 48. Synthesis of ^{13}C-labeled oligosaccharide. The synthesis was initiated from a monosaccharide GlcA-pNP. The introduction of ^{13}C-saccharide residue was completed by using KfiA and pmHS2 during the backbone synthesis. Introducing UDP-[^{13}C]saccharide at specific steps lead to synthesis of ^{13}C-labeled constructs labeled at desired saccharide residues. After the hexasaccharide intermediates syntheses were completed, both intermediate were subjected to sulfation/epimerization modification steps to form the final hexasaccharide products

MS and NMR analysis confirmed the structures of both hexasaccharide A and hexasaccharide B (Figure 49-50). In NMR spectrum the ^{13}C-labeled carbons show 100 times stronger intensity of signal than the unlabeled ones. As expected, HSQC analysis showed that

hexasaccharide A has only five cross peaks from IdoA2S residue (Figure 49), while hexasaccharide B has only 12 cross peaks from both GlcNS3S6S and IdoA2S residues (Figure 50). Only 150 ug was enough used for HSQC and 1D ^{13}C-NMR analysis for the hexasaccharide, suggesting that our method is able to synthesize sufficient amount of the hexasaccharides for the studies.

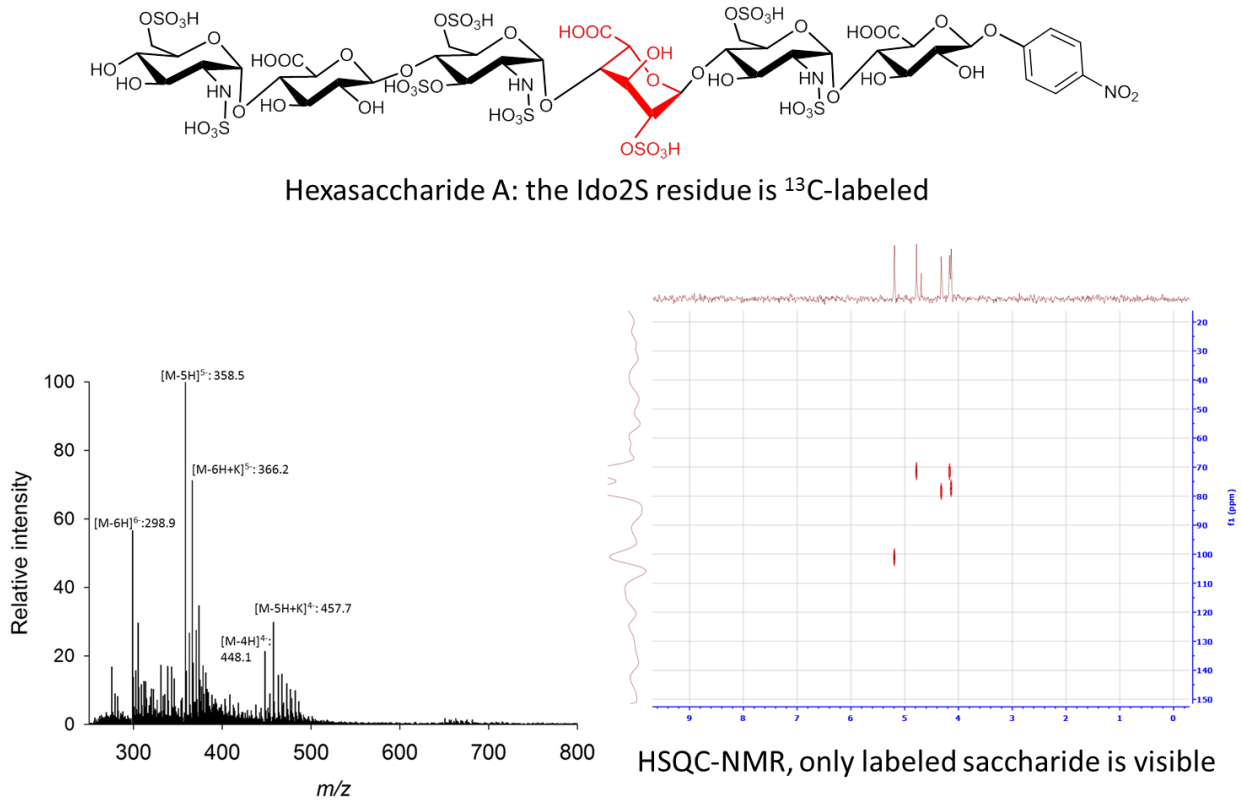


Figure 49. HPLC and heteronuclear single quantum coherence (HSQC) spectra of hexasaccharide A. Hexasaccharide A (0.17 mM) were dissolved in 10 mM phosphate buffer containing 150 mM NaCl. The ^{13}C-labeling sites in hexasaccharide A is colored in red (IdoA2S). The spectra were recorded at 35^{o}C on Bruker 850 MHz instrument.

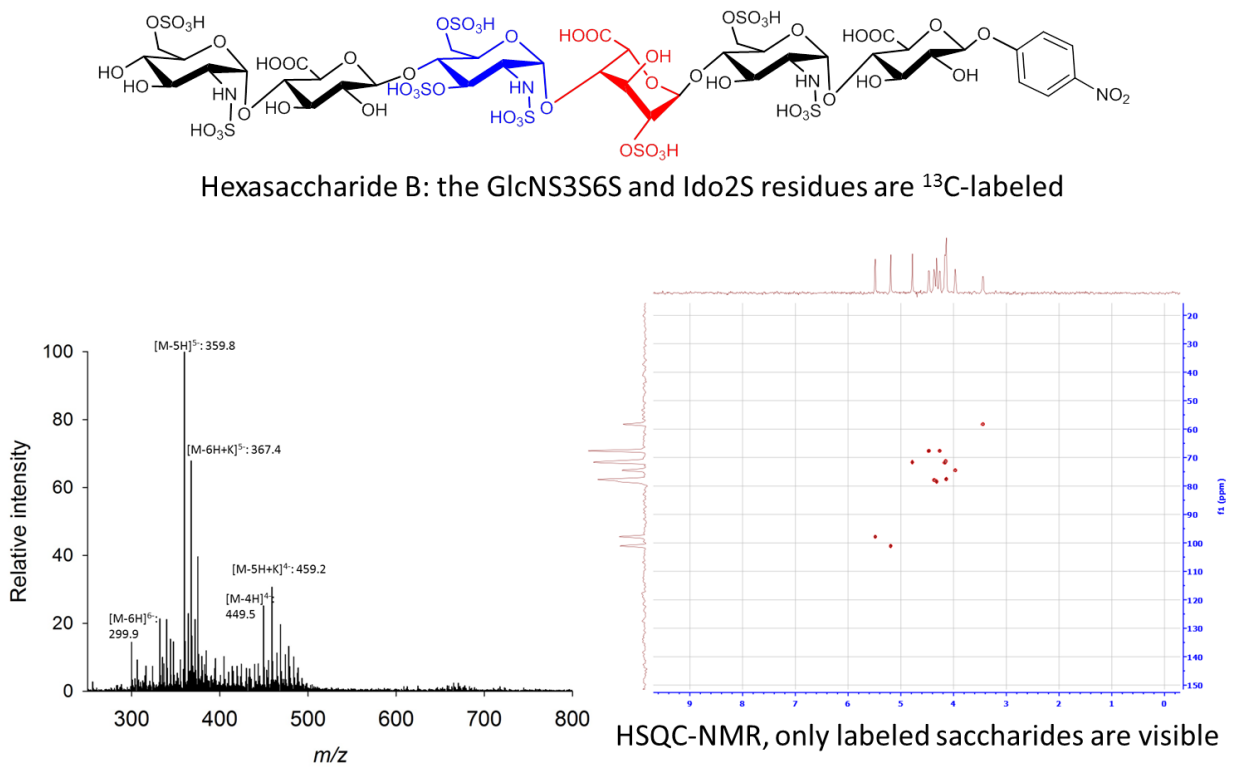


Figure 50. HPLC and heteronuclear single quantum coherence (HSQC) spectra of hexasaccharide B. Hexasaccharide B (0.17 mM) were dissolved in 10 mM phosphate buffer containing 150 mM NaCl. The ^{13}C-labeling sites in hexasaccharide B is colored in blue (GlcNS3S6S) and in red (IdoA2S). The spectra were recorded at 35°C on Bruker 850 MHz instrument.

In addition to hexasaccharide A and B, a series of ^{13}C-labeled hexasaccharides with different sulfation levels were also obtained from the pilot synthesis (Table 7). Among these oligosaccharides, only hexasaccharide A and B bind to AT.

compounds	Abbreviated structures	Measured MW	Calculated MW
Hexa-1A	$GlcNH_2$-GlcA-GlcNS-**[^{13}C]IdoA2S**-GlcNS-GlcA-pNP	1397.5 ± 0.5	1397.1
Hexa-2A	GlcNS-GlcA-GlcNS-**[^{13}C]IdoA2S**-GlcNS-GlcA-pNP	1477.7 ± 0.7	1477.2
Hexa-3A	GlcNS6S-GlcA-GlcNS6S-**[^{13}C]IdoA2S**-GlcNS6S-GlcA-pNP	1717.4 ± 0.6	1717.4
Hexasaccharide A	GlcNS6S-GlcA-GlcNS3S6S-**[^{13}C]IdoA2S**-GlcNS6S-GlcA-pNP	1797.8 ± 1.5	1797.4
Hexa-1B	$GlcNH_2$-GlcA-**[^{13}C]GlcNS**-**[^{13}C]IdoA2S**-GlcNS-GlcA-pNP	1403.4 ± 0.3	1403.1
Hexa-2B	GlcNS-GlcA-**[^{13}C]GlcNS**-**[^{13}C]IdoA2S**-GlcNS-GlcA-pNP	1483.1 ± 0.5	1483.2
Hexa-3B	GlcNS6S-GlcA-**[^{13}C]GlcNS6S**-**[^{13}C]IdoA2S**-GlcNS6S-GlcA-pNP	1723.4 ± 0.6	1723.6

Hexasaccharide B	GlcNS6S-GlcA-[^{13}C]GlcNS3S6S-[^{13}C]IdoA2S-GlcNS6S-GlcA-pNP	1803.3 ± 1.2	1803.4

Table 7. List of ^{13}C-labeled hexasaccharides prepared in the pilot synthesis. The labeled saccharides are marked in blue or red color.

Determination of the Binding of Labeled Hexasaccharide to AT

AT interacts in high affinity with a specific pentasaccharide domain in heparin [200], both hexasaccharide A and B contain this domain. We used of ^{13}C-labeled hexasaccharide A to probe the binding to AT. Firstly, we performed 2D-HSQC experiments using NMR. Under the same concentration of AT and hexasaccharide A (0.17 mM), only signals of AT are visible, signals of hexasaccharide A are not distinguishable (Figure 51A). The invisible signals of hexasaccharide A is due to being tightly bound by AT with molecular weight around 60kDa. To confirm the tight binding of ^{13}C-labeled hexasaccharide A and AT, an additional one equivalent amount of hexasaccharide A was added to observe the expected free-form of hexasaccharide A. At this concentration, the amount of hexasaccharide A is two-folds that of AT (0.34 mM). At this concentration, the signals of free-form hexasaccharide A become visible (Figure 51B). The cross peaks of free-form hexasaccharide A are the same as in Figure 49. The results support the labeled oligosaccharide can tightly bind AT and the binding is detectable using this optimized condition.

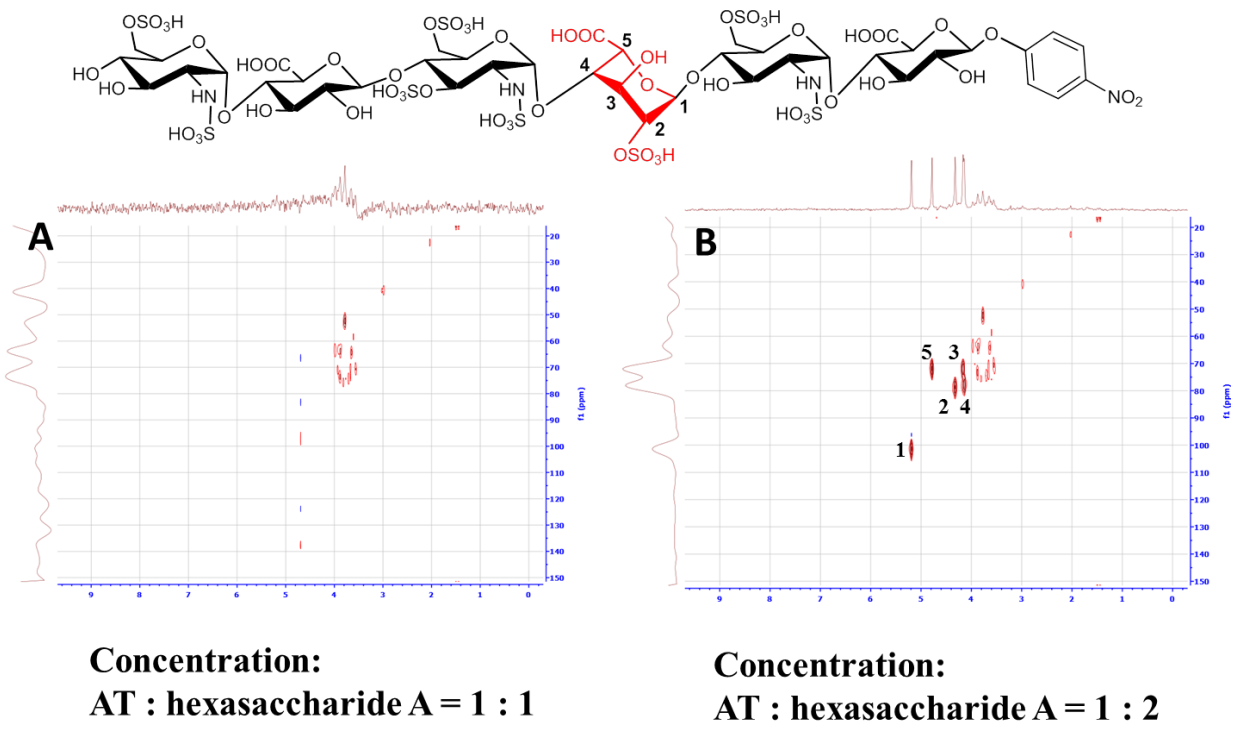


Figure 51. 2D-HSQC NMR spectrum of HS/hexasaccharide A mixture. A) Under the same concentration of AT and hexasaccharide A (0.17 mM), only signals of AT are visible, signals of hexasaccharide A are not distinguishable. B) Under the concentration of hexasaccharide A is two-folds of AT (0.34 mM), the signals of free-form hexasaccharide A becomes visible. The cross peaks from labeled IdoA2S are indicated.

Secondly, three individual proton-decoupled ^{13}C-NMR experiments were performed using the same concentration (0.17 mM) of AT only, hexasaccharide A only, or both AT and hexasaccharide A, respectively (Figure 52). In the experiment containing both AT and hexasaccharide A, the signals of bound-form hexasaccharide are visible because of ^{13}C-isotope labeling (indicated by arrow). The labeled saccharide has 100-fold higher sensitivity leading to the signals being visible even when bound to a much larger protein - 60kDa AT here. The signals of bound hexasaccharide A are broad due to high AT affinity with shorter T2. The significant different chemical shifts between bound and free form of hexasaccharide A are highlighted in red. The signals of C2 and C5 show different chemical shifts indicating that the

sulfate and carboxylate at these two positions are tightly involved in interaction with AT, which is consistent to previous literature reported crystal structures [243, 244]. These results demonstrate a direct way to find the binding epitope of HS, and show the promising potential of isotope-labeled reagents in NMR-related applications.

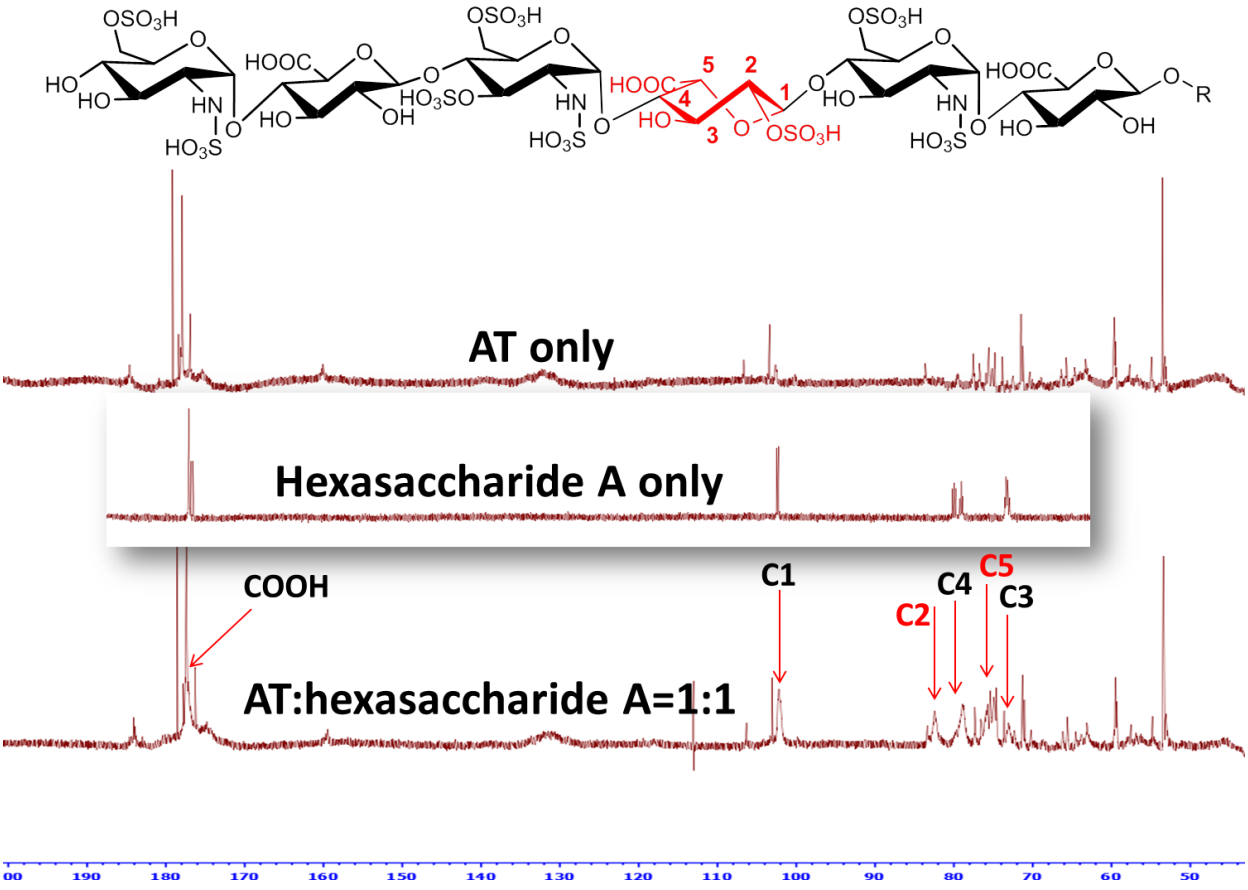


Figure 52. The proton-decoupled ^{13}C-NMR. Three individual proton-decoupled ^{13}C-NMR experiments were performed using the same concentration of AT only, hexasaccharide A only, or both AT and hexasaccharide A. The signals of bound-form hexasacchairde are visible because of ^{13}C-isotope labeling (indicated by arrow). The significant different chemical shifts between bound and free form of hexasaccharide are highlighted in red.

Another goal is to develop 2D-saturation transfer different (STD) experiments to study the interaction of proteins and oligosaccharides. Unlike traditional 1-D STD, the 2D-STD will

provide the option to measure the changes in HSQC spectrum, significantly improving the accuracy and sensitivity for monitoring protein/HS oligosaccharide interactions. The 2D-STD has never been performed on HS/protein complexes, and thus, a careful validation experiment is needed. This part of experiment is still in progress, and will be completed in the near future. Because the co-crystal and solution structures of AT/pentasacccharides are available[243, 244], the results from our studies will validate our approach. Results of 2D-STD analysis will greatly help obtain a high resolution picture for the interaction of AT/oligosaccharide in solution. The same strategy can also extend to the studies of other HS-bound proteins.

Investigation of the Conformation of Heparan Sulfate

The conformations of a given hexasaccharide include the conformation of individual saccharide residue (*i.e.*, chair, skew boat or boat conformations) as well as the two dihedral angles, ϕ and φ, for the glycosidic linkages. Here, we will focus on determining the conformation of the dihedral angles of glycosidic linkages, at the reducing and nonreducing end of ^{13}C-labeled saccharide residue, under the influences of different sulfation patterns. Measurements of proton-carbon long-range coupling constants combined with molecular simulation, will achieve these goals. Introduction of ^{13}C-labeling technique is essential to complete the measurement due to the increased sensitivity of ^{13}C-NMR signals.

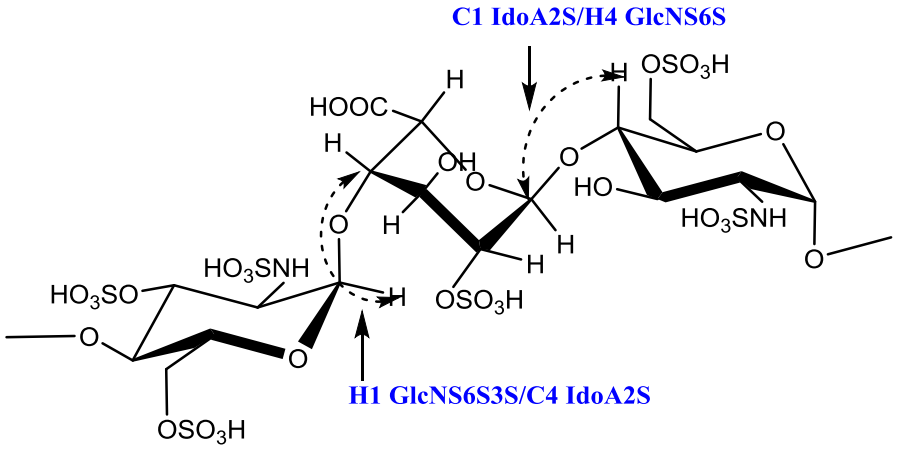


Figure 53. The anticipated inter-residue glycosidic $^{3}J_{C\text{-}H}$ coupling detected by HMBC.

To demonstrate the feasibility, we conducted the J-HMBC and HSQC analysis on hexasaccharide A. The anticipated long range ^{1}H-^{13}C coupling effects measured by heteronuclear multiple-bond correlation (HMBC) are indicated in Figure 53. In addition to those HSQC signals (signals colored in red in Figure 54), two signals were observed in HMBC resulting from reducing end glycosidic linkage of IdoA2S residue (the C1 of IdoA2S and H4 from GlcNS6S residue), as well as the nonreducing end glycosidic linkage of IdoA2S residue (the C4 of IdoA2S and H1 from GlcNS3S6S residue). As shown in the enlarged portion showing H4 GlcNS6S/C1 IdoA2S signal (Figure 54a), both one bond C-C coupling ($^{1}J_{C\text{-}C}$) and three bond C-H ($^{3}J_{C\text{-}H}$) coupling can be detected. Using frequency-selective pulse, we can suppress the one bond C-C bond coupling [245], allowing us to obtain accurate three bond coupling constant ($^{3}J_{C\text{-}H}$) (Figure 54b).

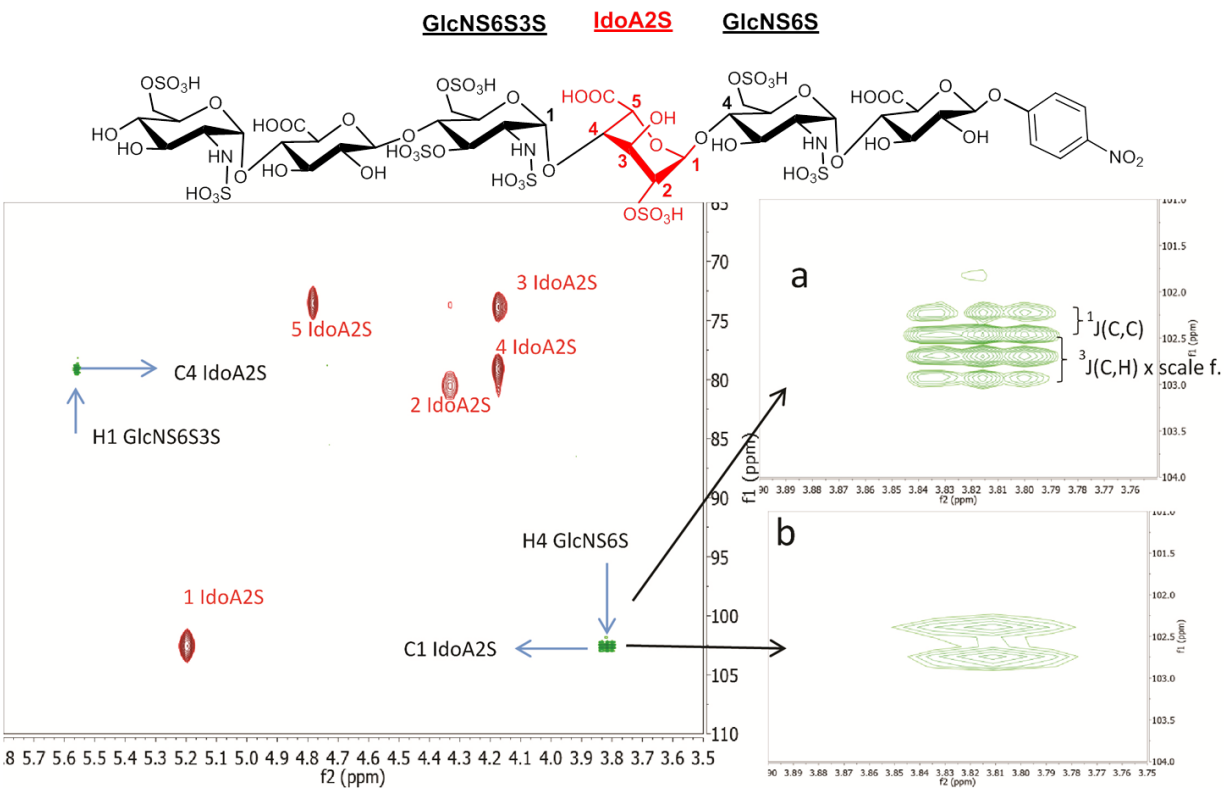


Figure 54. HSQC (red) and J-HMBC (green) spectra of hexasaccharide A. The region related to $^{3}J_{C1\text{-}H4'}$ without (a) and with (b) frequency-selective pulse for suppression of homonuclear carbon–carbon coupling constants.

The coupling constant values determined by NMR can be translated into glycosidic angles according to established computational protocols employed previously by the Woods' lab [246-250] . Briefly, quantum mechanical calculations can be employed to compute the relevant heteronuclear ^{3}J-values for conformations derived from independent molecular dynamics (MD) simulations of the oligosaccharide structures. The experimental ^{3}J-values can then be deconvoluted into a linear combination of the populations of the theoretical conformational states [249]. In the event that MD simulations fail to identify appropriate states, the experimental J-values can be employed as constraints during the MD simulations to obtain a single experimentally-consistent average conformation. This part of experiment is still in progress. We

plan to carry out similar analysis on some of the seven additional oligosaccharides as shown in Table 7, depending on the results from the analysis of hexasaccharide A. The results should provide the desired information on the conformation of glycosidic linkages under different sulfation patterns. We suspect the results will confirm that the sulfation patterns dictate the conformation of HS oligosaccharides.

Conclusions

HS is composed of sulfated saccharides that are biosynthesized through a complex pathway involving multiple enzymes. HS 2-*O*-sulfotransferase (2OST) is a key enzyme in this pathway. Here, we report the crystal structure of the ternary complex of 2OST, 3'-phosphoadenosine 5'-phosphate, and a heptasaccharide substrate. We successfully probed the molecular basis of specificity and 2OST's position in the ordered of the HS biosynthesis pathway. These studies revealed that 2OST interacts with the *N*-sulfo groups near the modification site, consistent with the dependence of 2OST on *N*-sulfation. In contrast, 6-*O*-sulfo groups on HS are likely excluded by steric and electrostatic repulsion within the active site supporting the hypothesis that 2-*O*-sulfation occurs prior to 6-*O*-sulfation. Our results provide the structural evidence for understanding the sequence of enzymatic events in this pathway. In addition, for 2-OST the acceptor IdoA residue in the heptasaccharide appears to be in the $^{4}C_{1}$ conformation. This conformation contradicts previous data that suggests IdoA assumes either the $^{1}C_{4}$ or $^{2}S_{0}$ conformation within the context of HS.

The saccharide sequences of HS determine the selectivity/specificity of their functions. Traditionally, the efforts for the identification of the contribution of specific saccharide structure

to HS biological function have been hindered due to the lack of an effective method. A chemoenzymatic approach for synthesizing active heparan sulfate based probes (AHSBP) with a reactive diazoacetyl saccharide residue is reported. The AHSBP approach takes advantage of the understanding of structural selectivity of HS receptors. Although using diazo group that is conjugated with peptide substrates has been used to study the biochemistry of proteases, utilization of this chemistry in HS oligosaccharides has not been previously reported. The covalent linkage formed between the probe and the target protein enables irreversible inhibition of the activities in a sequence selective manner. The resultant oligosaccharides were demonstrated to serve as specific inhibitors for heparan sulfate sulfotransferases 2-OST or 3-OST, offering a new set of tools to probe the structural selectivity for heparan sulfate-binding proteins. This is the first time that small molecule inhibitors display selective inhibition effects towards different HS biosynthetic enzymes. This approach can be potentially applied to any HS binding proteins or enzymes. Conversely, a specific AHSBP molecule can be utilized to identify previously unknown binding partners of HS. With further development, the AHSBP method can become a powerful method to aid in the study of HS biology.

We also developed an innovative method to directly analyze the binding between HS and target proteins. The crucial innovation of this method is capable of introducing ^{13}C-labeled saccharide at the desired site based on a newly developed and state-of-the-art chemoenzymatic method. Our site-specific labeled oligosaccharides will overcome the NMR signal overlapping issue caused from other repeating saccharides and protein residues, allowing us to analyze the specific saccharide/protein interaction directly. Furthermore, making multiple ^{13}C-labeled saccharides at different sites of single oligosaccharide is accessible as well. The access of these oligosaccharides will open the possibility for the conformational analysis of HS using NMR.

The conformational analysis of glycosidic linkages can take advantages of these labeled saccharides via computing dihedral angles from NMR measured $^{3}J_{C\text{-}H}$ coupling values. This method is particularly valuable given the fact attempts to crystalize a free HS oligosaccharide for X-ray analysis failed due to extremely high water solubility of HS oligosaccharides. This approach will also allow us to systematically investigate the contributions of the size and sulfation patterns on the conformation of HS. In the near future we will study the effect of protein binding on the conformation of HS oligosaccharides.

The success from these three studies will advance the fundamental understanding for the interaction between HS and proteins. In addition, these strategies may be used for the analysis of other biological important glycosaminoglycans, such as dermatan sulfates and chondroitin sulfates.

Chapter V

Investigate the Effect of Different Sulfation on Heparan Sulfate Conformation

Synthesis of Heparan Sulfate with Different Sulfation Pattern

Introduction

HS and heparin are negatively charged polysaccharides consisting of repeating disaccharide unit of α-D-(1,4)-glucosamine and β-D-(1,4)-glucuronic or α-L-(1,4)-iduronic acid with different sulfations.[9] Except antithrombin, heparan sulfates were reported to bind to several other proteins, including fibroblast growth factors and corresponding receptors, chemokines, selectins, platelet factor 4, etc.[6] In addition to the length of saccharide chain, diversity of sulfation patterns and plasticity of iduronic acid (IdoA) are known to be correlated to the biological importance.[8, 82] The flexible glycosidic linkages and IdoA could facilitate heparan sulfate conformational rearrangement to allow diverse proteins binding. The same structure of HS oligosaccharide may dynamically populate at several stable conformations, but only certain conformation binds to protein.[43, 83] The preference of conformation could attribute to comparable energy states established by local environment. Highly heterogeneous sulfation patterns of heparan sulfate provide the local environment diversity. We hypothesize the combining of specific sulfation pattern with certain sequenced oligosaccharide may allow us to

build pre-defined conformations with spatially tuned flexibility for specific protein binding. Larger sizes of oligosaccharide could allow for more defined conformation by more intramolecular constraints including hydrogen bonds and electrostatic interactions.[8] Therefore, the sufficient length of oligosaccharide or multi-flexible-domain may increase the specificity and selectivity for carbohydrate-protein interactions.

Crystallography and NMR are the best ways to accurately identify the molecular structure of HS. Unfortunately due to highly flexible character, carbohydrates are difficult to crystalized and analyzed by NMR. [84] The rapid dynamically exchange between conformations makes NMR only observe the signals averaged from individual conformations.[85] Alternatively, molecular dynamics (MD) simulation could identify the list of possible low energy conformations; however, simulations are unable to accurately calculate the equilibrium population of each conformation[86]. Taking the advantage of both NMR and MD simulation analysis, the best manner to calculate the realistic equilibrium population of conformations is to deconvolute the NMR measured parameters into simulated parameters for each conformation with corresponding populations. The measurement of spin-spin coupling constant ($^3J_{H\text{-}H}$) and intensity of nuclear overhauser effects (NOE) are the appropriate parameters representing saccharide conformation.[82] The change of conformation altering the dihedral between vicinal protons as well as the spatial distance between ring protons which are reflected in a shift of the $^3J_{H\text{-}H}$ coupling constant and NOE, respectively. The $^3J_{H\text{-}H}$ coupling constants are obtained by direct ^{1}H-NMR experiment or calculated from simulated dihedral using the Karplus equation (Figure 55).[87, 88] Intensity of NOE signals are acquired from 2D-NOESY NMR experiment or calculated from simulated proton distance. Both $^3J_{H\text{-}H}$ coupling constant and NOE are used as constraints to monitor the conformational change and corresponding population in this study.

Parallel usage of both NMR and simulation approach enable us to capture the structural and dynamic properties of carbohydrate conformation ensembles.[85]

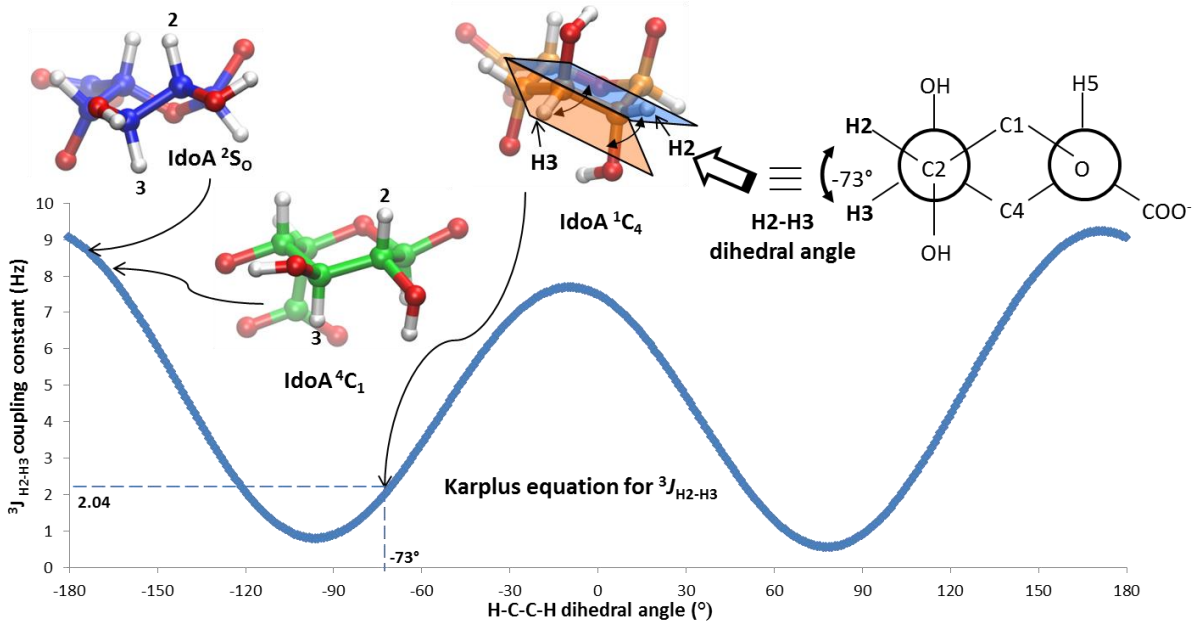


Figure 55. Transition of simulated dihedral angle to coupling constant by Karplus equation. The exemplified dihedral angle of H2-H3 was translated to $^{3}J_{H2\text{-}H3}$ coupling constant for $^{1}C_{4}$, ^{2}So and $^{4}C_{1}$ conformation, respectively, highlighted by arrows. The reparameterized Haasnoot Karplus equations: $^{3}J_{H\text{-}H}$ = 14.63*cos2(φ)-0.78*cos(φ)+0.60+∑λi[0.34-2.31cos2[sin(φ)+18.4|λi|]]
.

This study is to investigate the relation between the various sulfations and the saccharide conformations among heparan sulfate. A total of 8 HS oligosaccharides were synthesized with different sulfation pattern (Figure 56). The size of hexa-saccharide was considered containing sufficient constrains to make molecule at more defined conformation comparing to di-, tri- or tetra-saccharide. The saccharide of interest, IdoA, was designed at the central site of the oligosaccharides (Figure 56). The construct 1 to 4 were used to understand the effect of *N*-sulfation, 6-*O*-sulfation and 3-*O*-sulfation on the conformation of IdoA. In addition, the effect of 2-*O*-sulfation was investigated. The construct 5 to 8, which lack the 2-*O*-sulfo group on iduronic

acid, were used to identify the effect of 2-*O*-sulfation on saccharide conformation with different sulfation level. Our results suggest that the sulfation affects heparan sulfate conformation.

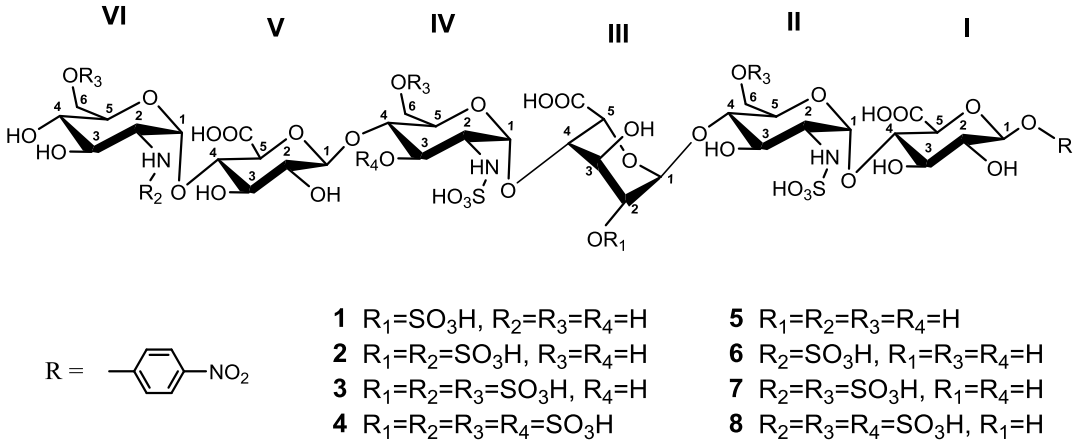


Figure 56. Chemoenzymatic synthesized oligosaccharide constructs 1 to 8.

Chemoenzymatic Synthesis of Oligosaccharide

A series of constructs were designed to determine impact of the surrounding saccharide sequences on the conformation of IdoA. The chemoenzymatic synthesis approach was utilized to synthesize a total of 8 oligosaccharides.[1] The structures are a set of 2-*O*-sulfated oligosaccharide construct 1 to 4 with 4 different sulfation patterns, and another set of oligosaccharides construct 5 to 8 with comparable sulfation patterns but lack 2-*O*-sulfation (Figure 56). Constructs 5 to 8 are new structures that have never been reported synthesized. Starting from commercial monosaccharide, backbone elongation and saccharide modification were carried out in sequential steps (Figure 57). Construct 1-4 were made in the order of

epimerization/2-*O*-sulfation, *N*-sulfation, 6-*O*-sulfation and 3-*O*-sulfation. Similarly, without using 2-*O*-sulfotransferase, constructs 5-8 were obtained.

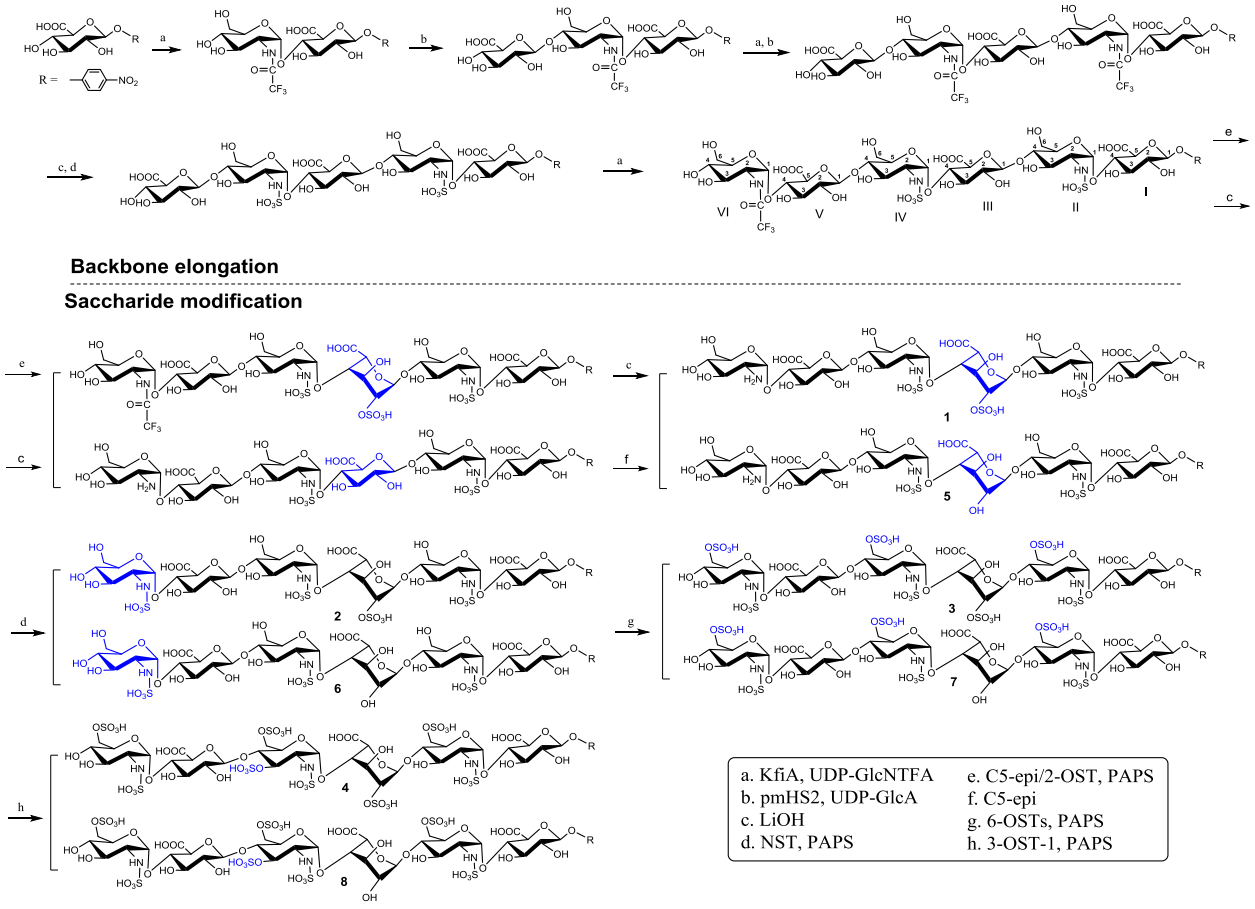


Figure 57. Schematic synthesis of IdoA- and IdoA2S-containing oligosaccharides.

Purification of synthetic oligosaccharides was performed using C_{18} and anion-exchange column chromatography for elongation and modification steps, respectively. Separation of epimerized construct 5 from epimer precursor demands high performance liquid chromatograph (HPLC) system with semi-preparative polyamine column and optimized buffer gradient (Figure 58). The purity of synthesized oligosaccharides was analyzed using HPLC and Mass spectrometry (Appendix Supplementary Figure 54)

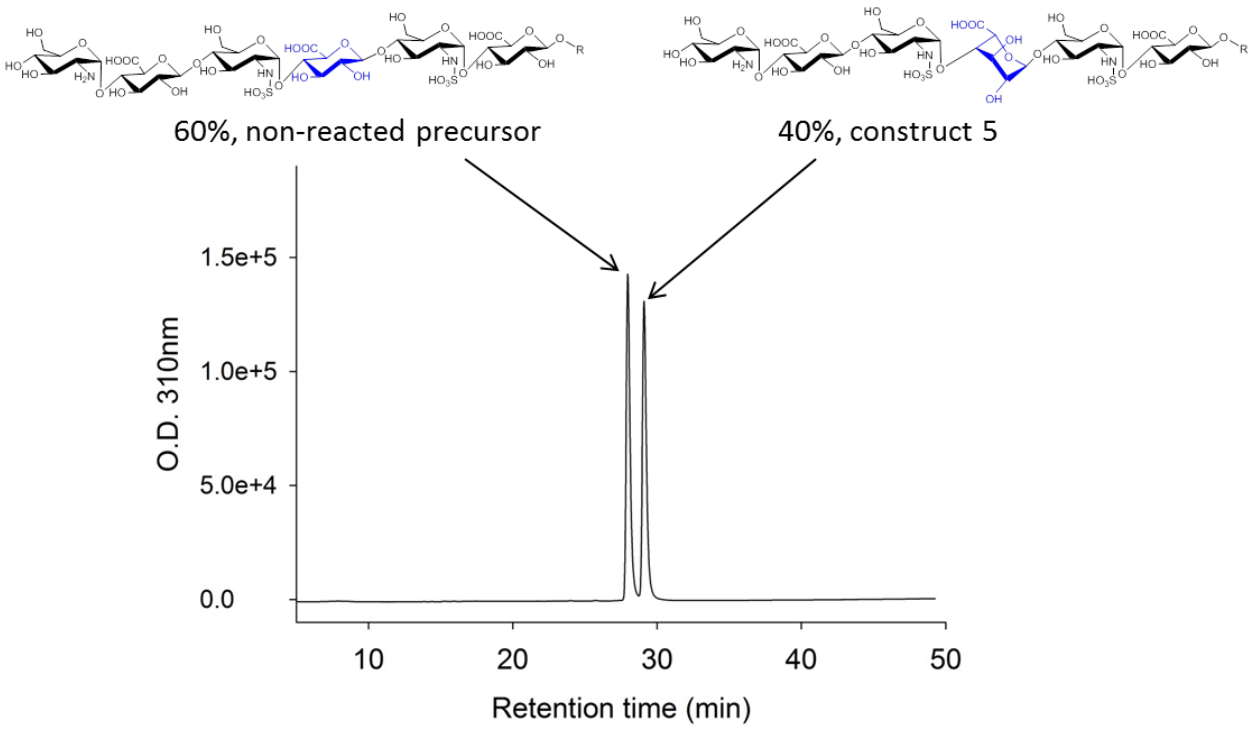


Figure 58. The anion-exchange HPLC chromatogram of the epimerization reaction mixture. The products contain non-reacted precursor (~60 %) and epimerized construct 5 (~40 %) with only 1 min retention time difference in 40 min gradient buffer system. The chemical structure of precursor and construct 5 is shown on top of figure.

Structural Characterization

The chemoenzymatic approach makes the synthesis of sufficient amount of constructs possible. Structural characterization by NMR experiments becomes much easier by virtue of better signal intensity. ^{1}H NMR was used to test the purity of constructs as well as to measure the $^{3}J_{\text{H-H}}$ coupling constants (Figure 59-60).

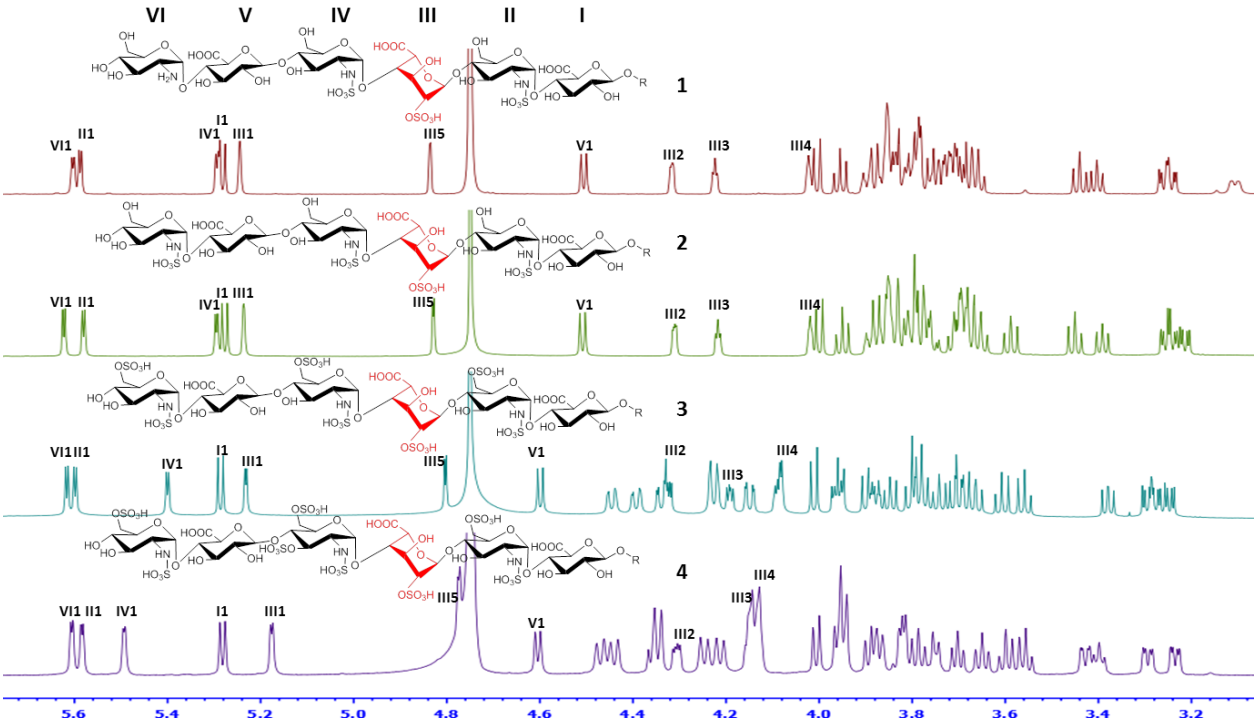


Figure 59. ^{1}H-NMR spectra of construct 1, 2, 3 and 4

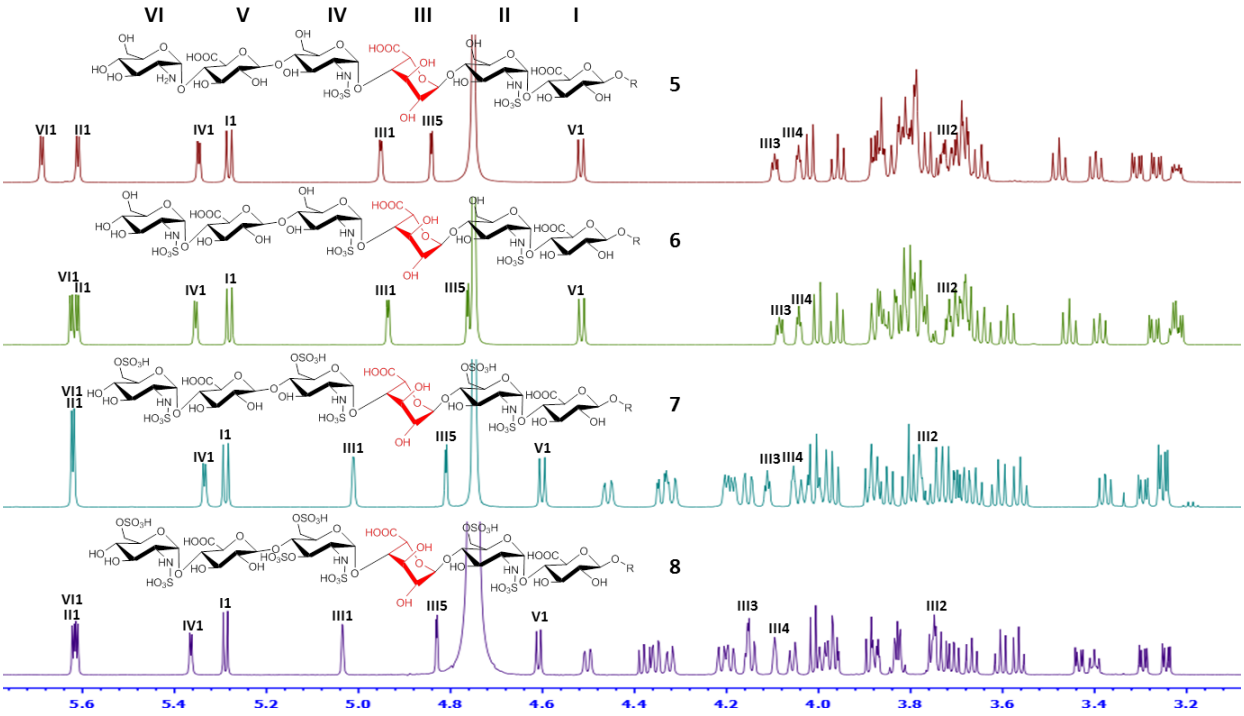


Figure 60. ^{1}H-NMR spectra of construct 5, 6, 7 and 8.

To fully assign the chemical shifts of all the proton and carbon signals, multidimensional homonuclear (COSY, NOESY) and heteronuclear (HSQC, HMBC, HSQC-TOCSY) NMR experiments were used to resolve the overlapped signals. The chemical shift assignments of constructs were acquired by dissecting the individual residue coupling system, one-bond and three-bond correlation based on NMR spectra.

Construct 6 (hexasaccharide with the structure of GlcNS-GlcA-GlcNS-IdoA-GlcNS-GlcA-pnp) was exemplified for structural characterization due to highly overlapped spectrum. In ^{1}H-NMR spectrum, the anomeric protons are located in the range between 4.5 and 5.6 ppm (Figure 60). Proton II-1, IV-1, and VI-1 were identified resonating between 5.3 and 5.7 ppm with $^{3}J_{H\text{-}H}$ coupling constants at 3.7 Hz because of α-(1, 4) linkage of glucosamine (GlcNS) residue. Proton I-1 and V-1 of GlcA residue were assigned at 5.28 and 4.52 ppm due to β-(1, 4) linkage with larger $^{3}J_{H\text{-}H}$ coupling constants at 7.9 Hz. The electron withdrawing group, para-nitrophenyl (pnp), makes proton I-1 resonates at downfield comparing to V-1. The remaining protons at chemical shift 4.94 and 4.76 ppm were assigned as III-1 and III-5 from IdoA residue with $^{3}J_{H\text{-}H}$ at 3.1 and 2.7 Hz, respectively.

The multidimensional NMR experiments were then utilized to assign the overlapped proton signals between 3.2 and 4.1 ppm. 2D ^{1}H-^{13}C HSQC-TOCSY was used to identify coupling network of each residue. This heteronuclear NMR experiment provide better chemical shift dispersion at indirect carbon dimension and resolved the signals. The spectrum analysis starting from anomeric protons, the six saccharide residue coupling networks were identified

(Figure 61). Within each residue coupling network, the proton and carbon signals belonging to the same saccharide were connected by lines (Figure 61).

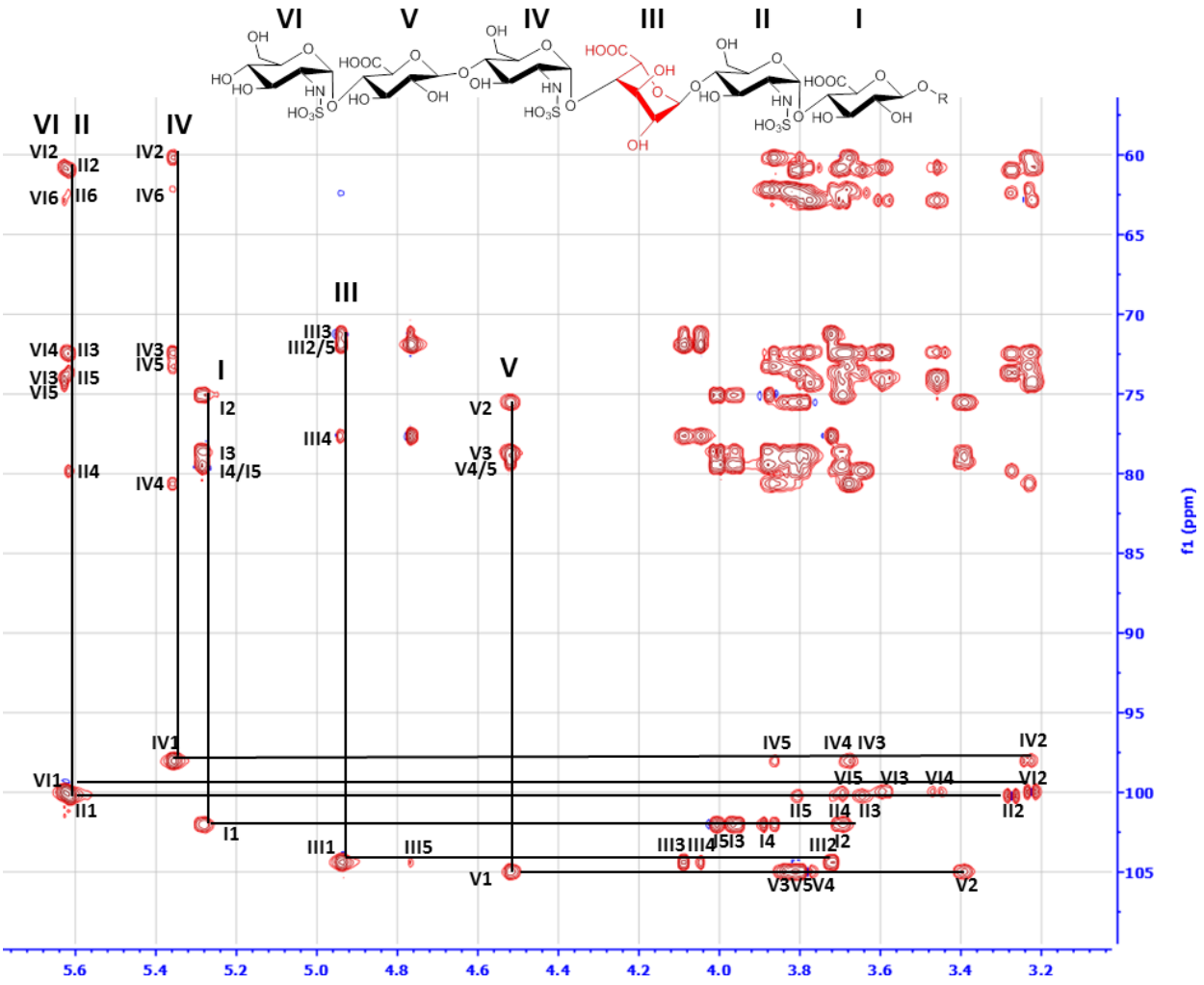


Figure 61. 2D ^{1}H-^{13}C HSQC-TOCSY NMR spectrum of construct 6.

Afterwards, ^{1}H-^{1}H COSY was used to assign coupled protons from one proton to the adjacent one within the same saccharide residue (Figure 62). For example, anomeric protons (I-1, II-1, III-1, IV-1, V-1, VI-1) have cross peaks with nearby protons (I-2, II-2, III-2, IV-2, V-2, VI-2), respectively.

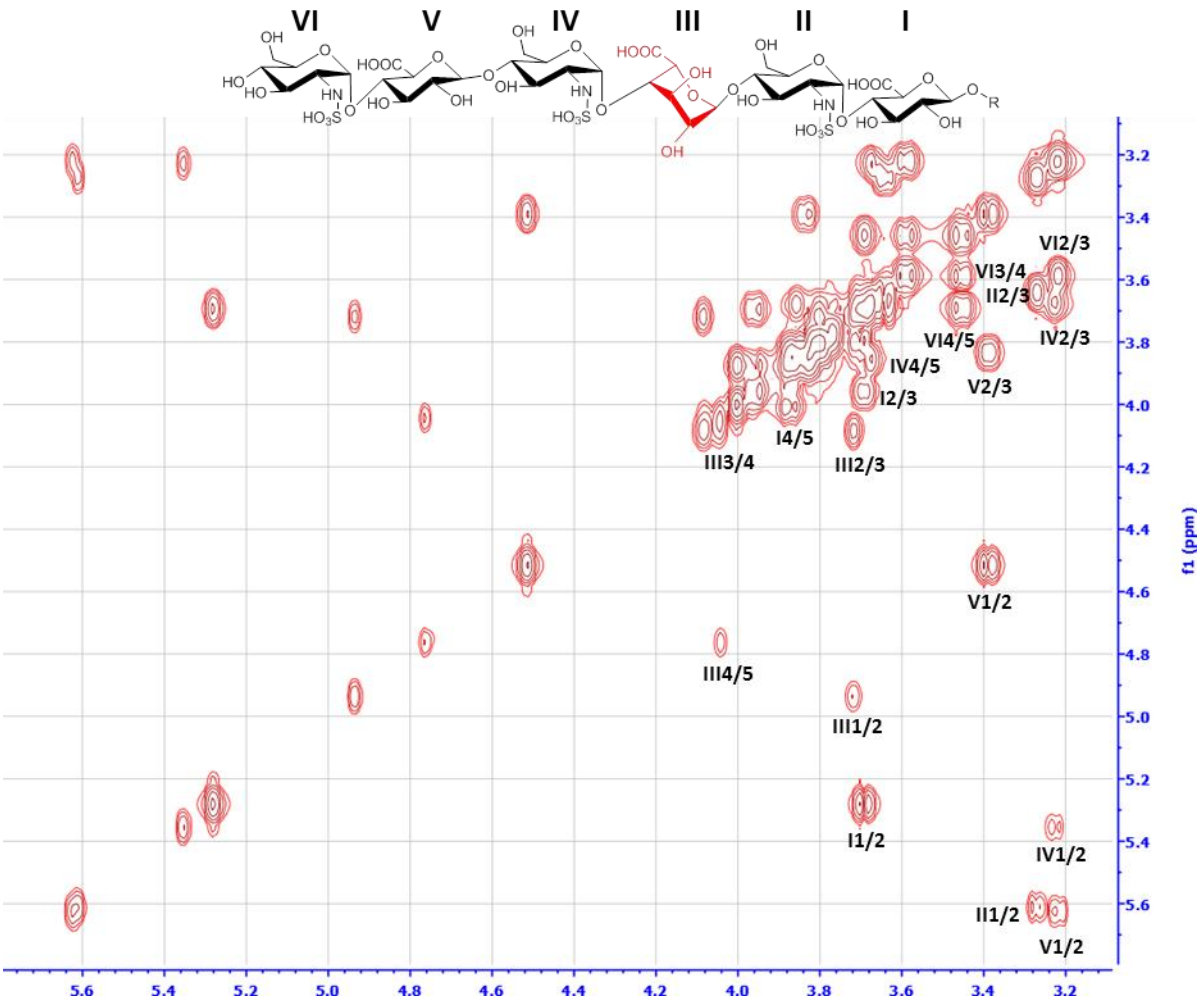


Figure 62. 2D ^{1}H-^{1}H COSY NMR spectrum of construct 6.

^{1}H-^{13}C HSQC acquires the signals of proton with correlated attached carbon (Figure 63). The proton and carbon chemical shift assignment of each individual saccharide residue were obtained.

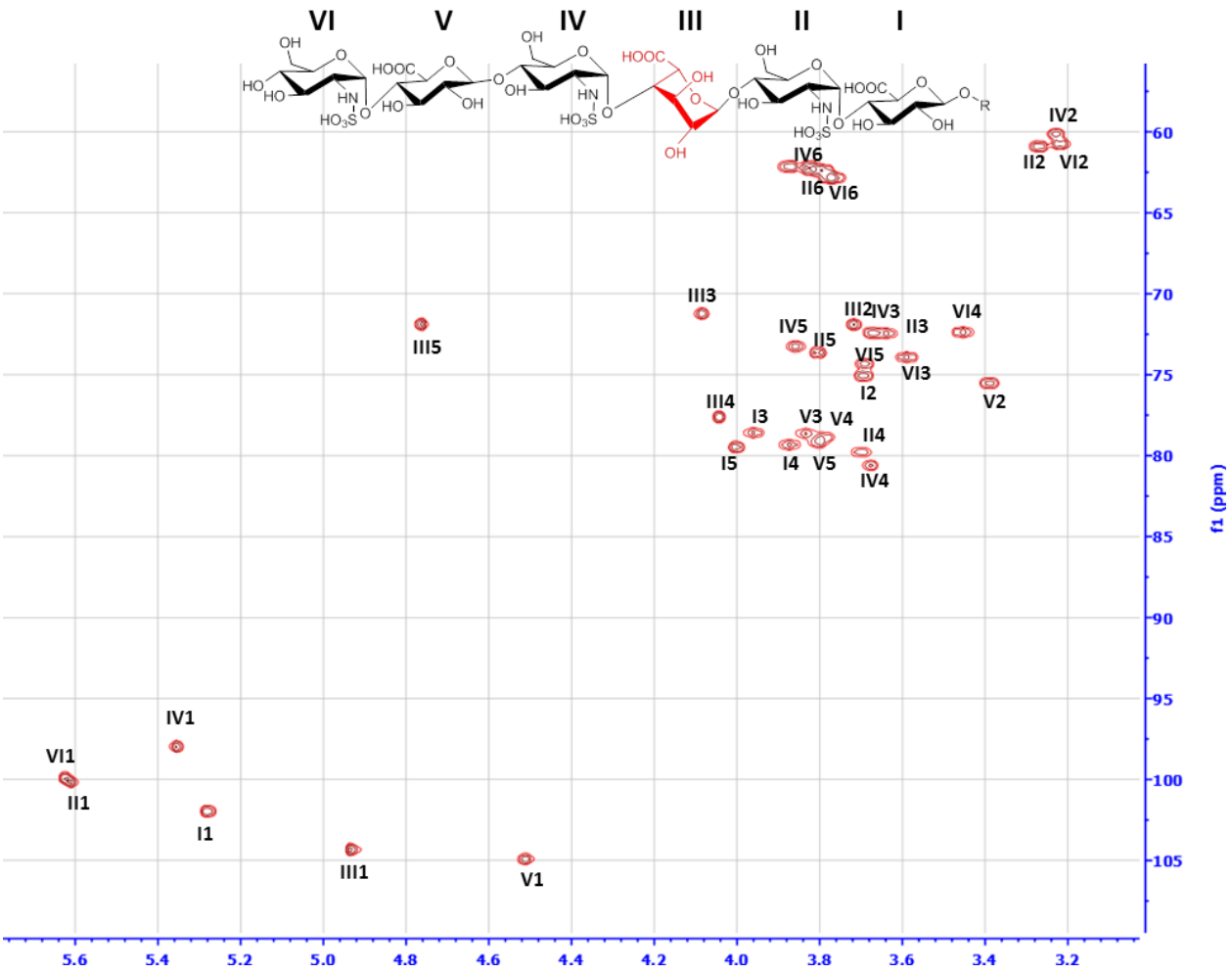


Figure 63. 2D ^{1}H-^{13}C HSQC NMR spectrum of construct 6.

^{1}H-^{13}C HMBC was finally used to designate the glycosidic linkages through three-bond correlation (Figure 64). The signals of anomeric protons (II-1, III-1, IV-1, V-1, VI-1) have cross peaks with the carbons at reducing end residue (I-4, II-4, III-4, IV-4, V-4). The anomeric proton I-1 was correlated to the quaternary carbon of pnp group. These correlations were confirmed from anomeric carbons to the respective protons at reducing end residue. The glucosamine protons II-6, IV-6 and VI-6 also showed the connectivity with correlated carbons II-4, IV-4 and

VI-4, respectively. These multiple bond correlations observed by HMBC enable the sequencing of the residues within the oligosaccharides.

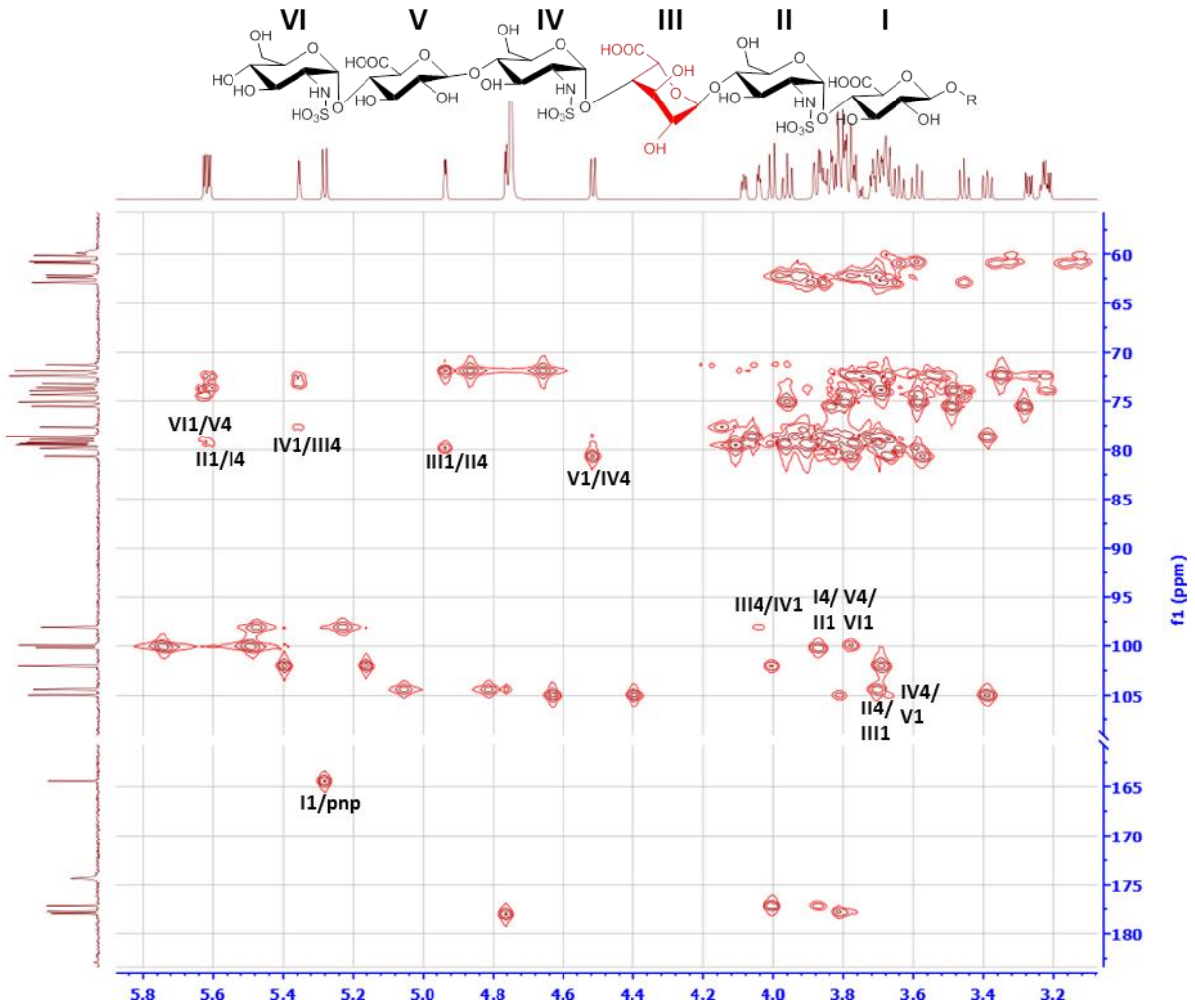


Figure 64. 2D ^{1}H-^{13}C HMBC key correlations of construct 6.

The full chemical shift assignments of construct 6 were made by analysis and cross-validation of these sequential NMR experiments. The structures of other constructs were characterized by similar approaches (Table 8-9). The HSQC-TOCSY, COSY and HSQC experiments were run to assign coupling network of each residue, HMBC were then used to

connect the residues through the glycosidic linkages. Structural characterization of the oligosaccharides using this approach greatly enhances the accuracy and facilitates full chemical shift assignments.

	1	2	3	4	5	6a	6b
Construct 1 (GlcNH2-GlcA-GlcNS-IdoA2S-GlcNS-GlcA-pnp)							
I	5.27	3.69	3.95	3.87	4.00		
	102.0	75.1	78.5	79.7	79.6		
II	5.58	3.26	3.66	3.70	3.80	3.83	3.84
	100.3	61.0	72.4	79.7	73.8	62.4	62.4
III	5.24	4.31	4.22	4.02	4.83		
	101.8	77.3	70.4	78.3	70.9		
IV	5.29	3.24	3.72	3.67	3.90	3.84	3.85
	99.8	60.4	72.1	80.6	73.2	62.3	62.3
V	4.50	3.40	3.75	3.83	3.83		
	104.9	75.8	78.6	78.7	78.8		
VI	5.60	3.10	3.74	3.44	3.72	3.78	3.79
	99.1	57.2	73.5	72.0	75.0	62.6	62.6
Construct 2 (GlcNS-GlcA-GlcNS-IdoA2S-GlcNS-GlcA-pnp)							
I	5.28	3.70	3.96	3.88	4.02		
	102.0	75.1	78.5	79.7	79.4		
II	5.59	3.26	3.66	3.70	3.80	3.84	3.84
	100.3	61.0	72.4	79.8	73.8	62.3	62.3
III	5.24	4.31	4.22	4.02	4.83		
	101.8	77.2	70.3	78.4	70.7		
IV	5.30	3.25	3.71	3.69	3.85	3.87	3.87
	100.0	60.4	72.1	80.6	73.4	62.1	62.1
V	4.53	3.39	3.84	3.78	3.82		
	104.9	75.5	78.6	78.9	79.1		
VI	5.62	3.22	3.59	3.45	3.68	3.77	3.77
	100.0	60.8	73.9	72.4	74.4	62.9	62.9
Construct 3 (GlcNS6S-GlcA-GlcNS6S-IdoA2S-GlcNS6S-GlcA-pnp)							
I	5.29	3.71	3.95	3.90	4.04		
	102.0	75.1	78.5	79.7	79.3		
II	5.60	3.30	3.67	3.77	3.95	4.22	4.38
	100.3	60.7	72.4	78.5	71.9	69.0	69.0
III	5.23	4.32	4.19	4.09	4.80		
	101.8	78.3	71.5	78.8	71.7		
IV	5.40	3.28	3.69	3.74	4.05	4.22	4.44
	99.6	60.2	72.1	79.7	71.6	68.6	68.6
V	4.60	3.38	3.84	3.78	3.82		
	104.5	75.5	78.6	79.4	79.0		
VI	5.61	3.25	3.61	3.56	3.87	4.15	4.33
	100.3	60.6	73.8	71.7	72.5	69.0	69.0
Construct 4 (GlcNS6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S-GlcA-pnp)							
I	5.28	3.71	3.95	3.90	4.00		
	102.0	75.1	78.5	80.0	79.5		
II	5.58	3.30	3.65	3.79	3.95	4.22	4.44
	100.4	60.8	72.4	78.2	72.0	68.9	68.9
III	5.18	4.31	4.15	4.13	4.78		
	102.2	79.6	73.0	78.6	72.8		
IV	5.50	3.43	4.35	3.95	4.15	4.25	4.47
	98.8	59.4	78.9	75.6	72.3	68.7	68.7
V	4.61	3.40	3.83	3.76	3.82		
	103.8	75.5	79.5	79.8	78.8		
VI	5.61	3.24	3.60	3.56	3.87	4.14	4.35
	100.2	60.7	73.8	71.7	72.4	69.0	69.0

Table 8. ^{1}H-^{13}C NMR chemical shift assignments (in ppm) of construct 1 – 4 (25°C, D_2O, pD 7.4, 3mM EDTA).

	1	2	3	4	5	6a	6b
Construct **5** (GlcNH2-GlcA-GlcNS-IdoA-GlcNS-GlcA-pnp)							
I	5.28	3.69	3.96	3.87	4.02		
	102.0	75.1	78.6	79.3	79.3		
II	5.61	3.27	3.65	3.70	3.79	3.79	3.79
	100.2	61.0	72.4	79.9	73.7	62.4	62.4
III	4.95	3.73	4.10	4.04	4.84		
	104.3	71.7	70.9	77.5	71.5		
IV	5.35	3.22	3.69	3.68	3.82	3.83	3.87
	98.2	60.2	72.3	80.6	73.3	62.2	62.2
V	4.52	3.40	3.76	3.81	3.87		
	104.7	75.9	78.6	78.6	78.6		
VI	5.69	3.31	3.84	3.48	3.72	3.79	3.79
	97.9	56.9	72.1	71.8	75.1	62.5	62.5
Construct **6** (GlcNS-GlcA-GlcNS-IdoA-GlcNS-GlcA-pnp)							
I	5.28	3.69	3.96	3.87	4.00		
	102.0	75.1	78.6	79.4	79.5		
II	5.61	3.27	3.64	3.70	3.81	3.80	3.80
	100.2	60.9	72.5	79.8	73.7	62.4	62.4
III	4.94	3.72	4.09	4.04	4.76		
	104.4	71.9	71.3	77.6	71.9		
IV	5.36	3.23	3.67	3.68	3.86	3.83	3.87
	98.0	60.2	72.4	80.6	73.3	62.2	62.2
V	4.52	3.39	3.83	3.78	3.80		
	105.0	75.5	78.6	78.9	79.2		
VI	5.63	3.22	3.59	3.46	3.69	3.77	3.77
	100.0	60.8	73.9	72.4	74.3	62.9	62.9
Construct **7** (GlcNS6S-GlcA-GlcNS6S-IdoA-GlcNS6S-GlcA-pnp)							
I	5.29	3.71	3.97	3.89	4.01		
	102.0	75.1	78.6	79.4	79.4		
II	5.62	3.30	3.66	3.73	3.99	4.20	4.32
	100.2	60.8	72.6	80.2	71.7	68.9	68.9
III	5.01	3.78	4.11	4.05	4.81		
	104.7	71.2	70.2	77.1	71.1		
IV	5.34	3.25	3.68	3.75	4.03	4.19	4.46
	98.1	60.0	72.3	79.7	71.5	68.6	68.6
V	4.60	3.38	3.85	3.78	3.81		
	104.5	75.6	78.6	79.3	79.1		
VI	5.62	3.25	3.61	3.56	3.88	4.15	4.34
	100.2	60.6	73.8	71.7	72.5	69.1	69.1
Construct **8** (GlcNS6S-GlcA-GlcNS3S6S-IdoA-GlcNS6S-GlcA-pnp)							
I	5.29	3.71	3.97	3.89	4.01		
	102.0	75.1	78.6	79.5	79.4		
II	5.62	3.29	3.67	3.73	3.99	4.19	4.32
	100.2	60.9	72.6	80.4	71.7	68.9	68.9
III	5.04	3.75	4.16	4.10	4.83		
	104.7	71.3	69.5	77.3	70.9		
IV	5.37	3.43	4.38	3.97	4.06	4.21	4.50
	98.3	59.3	78.7	75.5	72.3	68.6	68.6
V	4.61	3.40	3.83	3.83	3.75		
	104.0	75.5	78.8	79.5	79.8		
VI	5.61	3.24	3.60	3.56	3.88	4.15	4.35
	100.2	60.7	73.8	71.7	72.4	69.0	69.0

Table 9. ^{1}H-^{13}C NMR chemical shift assignments (in ppm) of construct 5 – 8 (25°C, D2O, pD 7.4, 3mM EDTA).

Conformation Analysis and Molecular Dynamic Simulation

The best approach to accurately identify the saccharide conformations is to observe the three-bond proton-proton coupling constants ($^3J_{H\text{-}H}$) of the sugar residue using NMR. Each three-bond coupling constant corresponds to a dihedral angle of two vicinal protons on the saccharide (Figure 55). The pyranose saccharide ring contains four proton-proton dihedrals with corresponding coupling constants $^3J_{H1\text{-}H2}$, $^3J_{H2\text{-}H3}$, $^3J_{H3\text{-}H4}$ and $^3J_{H4\text{-}H5}$, respectively. The integrated information combining series of these four corresponding proton-proton dihedrals with corresponding $^3J_{H\text{-}H}$ coupling constants designate the specific ring conformation. Except iduronic acid, the glucosamine and glucuronic acid residues are relatively structural rigid in all the 8 constructs in spite of sulfation patterns. The $^3J_{H1\text{-}H2}$ value for anomeric proton of glucosamine and glucuronic acid are similar with α and β linkage as 3.7 and 7.9 Hz, respectively. Comparing to the glucosamine and glucuronic acid, iduronic acid flips among different conformations, including the major stable $^1C_4$, ^2So and $^4C_1$ conformers (Figure 55).[251] The energy barrier between conformations is relatively low, which allows iduronic acid to populate among these three major states.[206] The conformation exchanges on the microsecond time scale.[206] This fast exchange leads real-time NMR to merely measure the averaged $^3J_{H\text{-}H}$ coupling constant over the three populated conformations. Accordingly, all the NMR measured coupling constants, $^3J_{H1\text{-}H2}$, $^3J_{H2\text{-}H3}$, $^3J_{H3\text{-}H4}$ and $^3J_{H4\text{-}H5}$, are averaged by the relative population of three conformations. The best manner to calculate the equilibrium population of conformations is to deconvolute the NMR measured $^3J_{H\text{-}H}$ values into theoretical $^3J_{H\text{-}H}$ values of each conformation with corresponding populations. The IdoA theoretical $^3J_{H\text{-}H}$ values for $^1C_4$, ^2So and

$^4C_1$ conformation were acquired by molecular dynamics (MD) simulation. ^{1}H-NMR was used to measure $^3J_{H\text{-}H}$ values of IdoA for each construct.

$^3J_{H\text{-}H}$ *Coupling Constant Measurement*

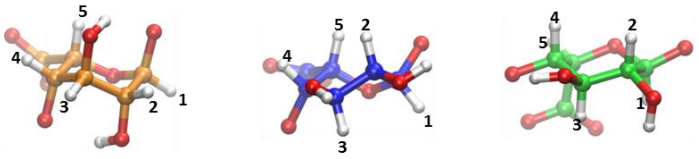
To simplify NMR analysis, only a single IdoA or IdoA2S residue is present in these oligosaccharides. The proton chemical shifts of IdoA and IdoA2S can be recognized on ^{1}H-NMR spectrum after fully assignment of all the signals (Table 8-9). The splitting patterns of coupled protons were able to be identified on ^{1}H-NMR spectrum (Figure 59-60). For example in construct 6, all the five IdoA ring protons, 4.94, 3.72, 4.09, 4.04 and 4.76 ppm marginally belong to AX pattern at the high magnetic field of 700 MHz. The doublet proton III-1 and III-5 were measured with $^3J_{H1\text{-}H2}$ and $^3J_{H4\text{-}H5}$ at 3.1 and 2.7 Hz. The proton III-2, III-3 and III-4 were recognized as doublet of doublets with $^3J_{H2\text{-}H3}$ and $^3J_{H3\text{-}H4}$ at 5.3 and 3.5 Hz. For all the synthetic constructs, the experimental $^3J_{H\text{-}H}$ values of iduronic acid were acquired using the similar approach (Table 10). The $^3J_{H1\text{-}H2}$ and $^3J_{H4\text{-}H5}$ were firstly measured from proton III-1 and III-5 due to the signals away from overlapped proton region; secondly, the $^3J_{H2\text{-}H3}$ and $^3J_{H3\text{-}H4}$ were identified by the other protons III-2, III-3 or III-4. In the case of leaning effect happened among coupled protons, the program 1D DAISY in Topspin was used to simulate the splitting pattern to assist the analysis of coupling constants.

	Experimental $^3J_{H\text{-}H}$-couplings (measured by NMR)				Theoretical $^3J_{H\text{-}H}$-couplings (calculated by MD simulation)				Sum of square difference
	$^3J_{H1\text{-}H2}$	$^3J_{H2\text{-}H3}$	$^3J_{H3\text{-}H4}$	$^3J_{H4\text{-}H5}$	$^3J_{H1\text{-}H2}$	$^3J_{H2\text{-}H3}$	$^3J_{H3\text{-}H4}$	$^3J_{H4\text{-}H5}$	
Construct **1**	2.2	4.3	3.2	2.2	2.7	4.1	3.0	2.1	0.34
Construct **2**	2.2	4.3	3.2	2.3	2.7	4.1	3.0	2.2	0.34
Construct **3**	3.0	5.8	3.6	2.7	3.6	5.5	3.6	2.5	0.49
Construct **4**	3.7	7.4	4.0	3.1	4.5	6.9	4.2	2.8	1.02
Construct **5**	3.0	5.2	3.5	2.7	3.3	5.1	3.4	2.4	0.20
Construct **6**	3.1	5.3	3.5	2.7	3.4	5.2	3.5	2.4	0.19
Construct **7**	2.3	4.3	3.4	2.4	2.8	4.1	3.2	2.2	0.37
Construct **8**	2.3	3.9	3.5	2.2	2.7	3.8	3.2	2.2	0.26

Table 10. Experimental and theoretical $^3J_{H\text{-}H}$ couplings of IdoA/2S among synthetic constructs.

The theoretical $^3J_{H\text{-}H}$ values of iduronic acid for each conformation were calculated from the results of molecular dynamics (MD) simulation. The molecular dynamics simulation of synthetic constructs were performed by AMBER14 with GLYCAM06 parameter sets specifically designed for carbohydrate. Total of 8 microsecond simulation were independently performed for iduronic acid at $^1C_4$, ^2So and $^4C_1$ conformation. The simulation was performed with extended microsecond timescale to secure the representative conformation of HS oligosaccharides. The simulated proton dihedral angles H1-H2, H2-H3, H3-H4 and H4-H5 of iduronic acid for the three different conformations were collected separately. A simulated dihedral angle was transformed into a corresponding $^3J_{H\text{-}H}$ value via the reparameterized Haasnoot Karplus equation which properly considers the electronegative substituent effects.[174, 175] For example, the simulated H2-H3 dihedral angle of the $^1C_4$ conformation is -73°, can be transformed to a $^3J_{H2\text{-}H3}$ equal to 2.04 Hz (Figure 55, Table 11). Each transformed $^3J_{H\text{-}H}$ values were then averaged over 8 microseconds to obtain the theoretical $^3J_{H1\text{-}H2}$, $^3J_{H2\text{-}H3}$, $^3J_{H3\text{-}H4}$ and $^3J_{H4\text{-}}$

$_{H5}$ values for one IdoA conformation. All the three major IdoA conformation, $^{1}C_4$, ^{2}So and $^{4}C_1$ separately acquired their own corresponding theoretical $^{3}J_{H\text{-}H}$ values (Table 11).



Average theoretical dihedral angles (°) of IdoA/2S	1C4	2So	4C1
H1-H2	74.6	146.0	169.0
H2-H3	-73	-172	-167
H3-H4	70.9	141.0	173.0
H4-H5	54.9	42.7	-51
Average theoretical $^{3}J_{H\text{-}H}$ values of IdoA/2S	1C4	2So	4C1
$^{3}J_{H1\text{-}H2}$	1.35	5.55	7.85
$^{3}J_{H2\text{-}H3}$	2.04	8.48	8.03
$^{3}J_{H3\text{-}H4}$	2.19	4.82	8.5
$^{3}J_{H4\text{-}H5}$	1.69	3.14	5.08

Table 11. Averaged theoretical dihedral angle and $^{3}J_{H\text{-}H}$ value for each conformation of IdoA. The reparameterized Haasnoot Karplus equation was used: $^{3}J_{H\text{-}H}$ = 14.63*cos2(φ)-0.78*cos(φ)+0.60+∑λi{0.34-2.31cos2[sin(φ)+18.4|λi|]}.

Instead of maintaining a single conformation state, IdoA can populate three major conformation states during the timescale of an NMR experimental measurement. Percentage fitting of the theoretical $^{3}J_{H\text{-}H}$ values from three individual IdoA conformations into the experimental $^{3}J_{H\text{-}H}$ values can allow one to obtain the population of each IdoA conformation (Figure 65). Different populations of three IdoA conformations were acquired using different construct's experimental $^{3}J_{H\text{-}H}$ values to fit (Figure 65). Based on the populations of three conformations, in order to refer the timescale of NMR experiments, the averaged theoretical IdoA $^{3}J_{H\text{-}H}$ values of each construct were calculated (Table 10). For each construct, the statistic

program R was used to calculate the best fitting with minimum sum of square differences between theoretical and the corresponding experimental $^{3}J_{H\text{-}H}$ values (Table 10). The smaller sum of square difference means the better fitting. The results show that theoretical and experimental $^{3}J_{H\text{-}H}$ values have good consistency comparing to literature reported values. Among all the eight constructs, the maximum $^{3}J_{H\text{-}H}$ value difference is only 0.8 Hz supporting that the MD simulations were performed properly resulting in confident calculation of conformation populations.

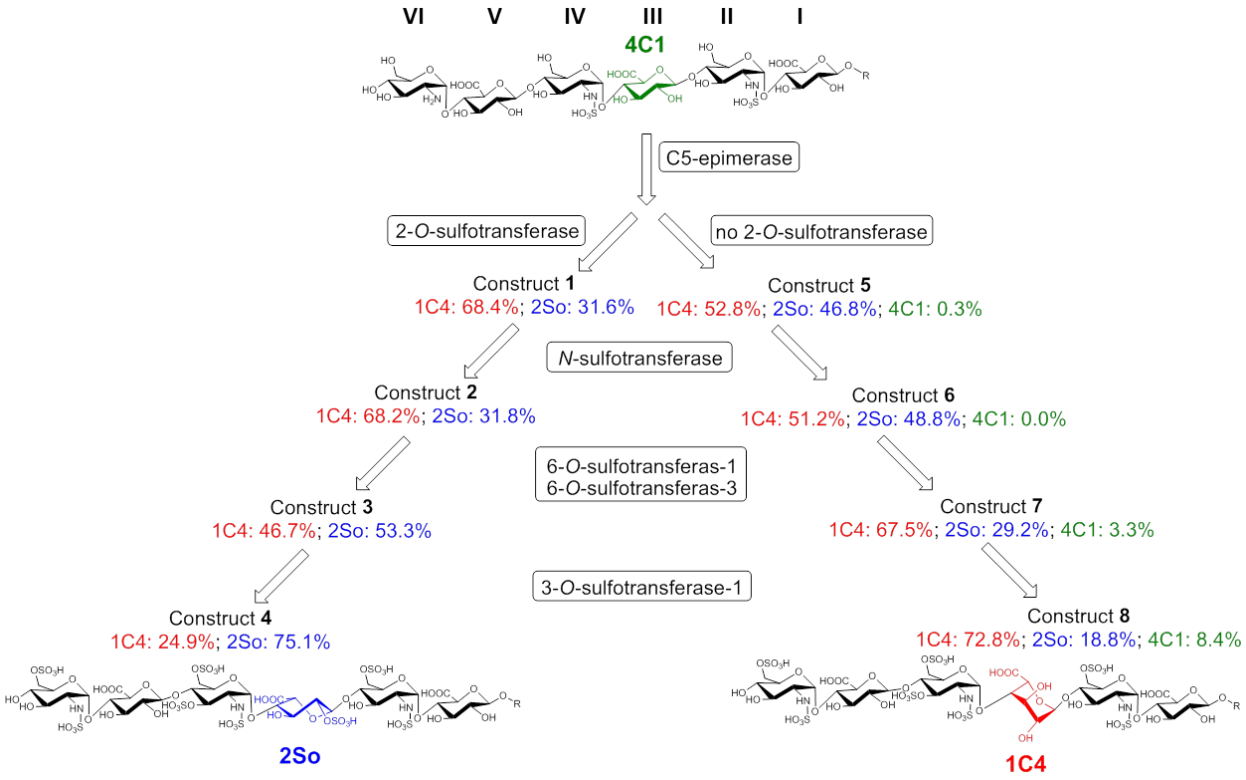


Figure 65. Hexuronic acid (GlcA, IdoA/2S) as pivot of conformation equilibrium to response sulfations.

Shift of IdoA Conformation Population

The shift of IdoA conformation population reflects the effect of different sulfation pattern on heparan sulfate oligosaccharide (Figure 65). The glucuronic acid (GlcA) was observed as $^4C_1$ conformation presented in the precursor HS backbone. Epimerization of GlcA to IdoA renders the accessibility of multi-conformation at desired location of the oligosaccharide. The chemoenzymatic approach introduces the sequential sulfation mimicking the biosynthesis of heparan sulfate, including 2-*O*-sulfation, *N*-sulfation, 6-*O*-sulfation and 3-*O*-sulfation (Figure 65). Along the synthetic steps, the epimerization and 2-*O*-sulfation dramatically change the destiny of iduronic acid conformation equilibrium. For example in construct 1, the 2-*O*-sulfated IdoA populate at major $^1C_4$ conformation (68.4 %) and minor ^{2}So conformation (31.6 %). In construct 5, lacking a 2-*O*-sulfo group, the IdoA is able to populate at three conformations with comparable population of $^1C_4$ (52.8 %) and ^{2}So (46.8 %) as well as a rare $^4C_1$ conformer. However, the following stepwise 6-*O*-sulfation and 3-*O*-sulfation lead IdoA2S toward ^{2}So conformer in construct 4; oppositely, lead IdoA to majorly populate at $^1C_4$ conformation in construct 8. The effect of 6-*O*-sulfation and 3-*O*-sulfation shift ^{2}So population of IdoA2S from 31.8 % in construct 2, to 53.3 % and 75.1 % in construct 3 and 4, respectively. In contrast, these two sulfations shift ^{2}So population of IdoA from 48.8 % in construct 6, to 29.2 % and 18.8 % in construct 7 and 8. These added 6-*O*- and 3-*O*-sulfo groups also gradually increase the $^4C_1$ population of non-sulfated IdoA to 3.3 % and 8.4 % in construct 7 and 8, respectively, but not for $^4C_1$ population of IdoA2S in construct 3 and 4. On the other hand, as expected, the *N*-sulfation of residue VI has no effects on the conformation population of IdoA2S or IdoA at III residue in construct 2 and 6 comparing to construct 1 and 5. These results support iduronic acid with, or without, 2-*O*-sulfation may result in different free energy state or different energy barrier between conformations. Furthermore, IdoA and IdoA2S could display different conformation

equilibrium based on respective surrounding environment in the constructs. The biosynthetic downstream sequential 6-*O*- and 3-*O*-sulfations generate different environments to further divide IdoA and IdoA2S containing oligosaccharides toward different stable conformations (Figure 65). Comparing to relatively rigid glucuronic acid, iduronic acid containing heparan sulfate could show different conformational preference in response to different neighboring sulfation patterns.

NOE Intensity Measurement

Proton-proton nuclear overhauser effect (NOE) is another approach that can be used to confirm the shift of IdoA conformation equilibrium observed using the $^{3}J_{\text{H-H}}$ value measurements [208, 211]. For each $^{1}C_{4}$, ^{2}So or $^{4}C_{1}$ conformation, the spacial distance between the ring protons represents different intensity of NOE. The change of NOE intensity on indicative protons reflects the shift of saccharide conformation. Based on simulation result, in iduronic acid the distance between proton H2 and H5 is significantly shorter for ^{2}So (2.6 Å) than for $^{1}C_{4}$ (4.0 Å) or $^{4}C_{1}$ (4.2 Å) conformation (Figure 66). The distance between proton H4 and H5 are relatively similar among these three conformations (2.4, 2.3, 2.4 Å) (Figure 66).

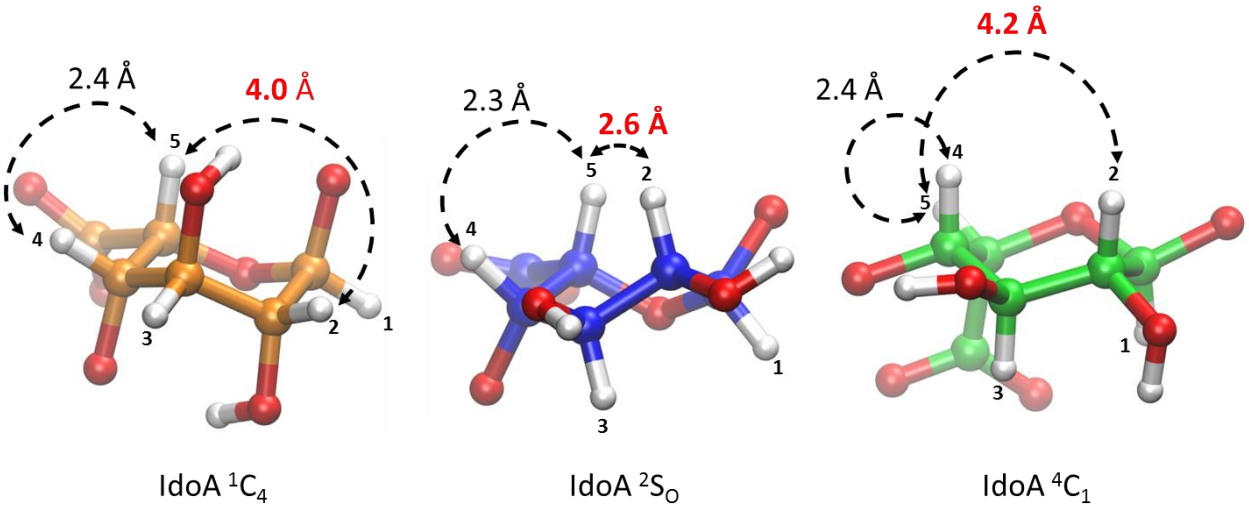


Figure 66. The simulated proton distance for three major conformations of Iduronic acid. The distance between H2 and H5, H4 and H5 is labelled. The featured difference is highlighted in red color.

NOE intensity is inversely proportional to the sixth power of distance, therefore, the NOE intensity ratio of H2-H5/H4-H5 of iduronic acid could be used as an indicator of relative population of ^{2}So to the other $^1C_4$ and $^4C_1$ conformation. The higher population of ^{2}So conformation carry higher NOE ratio of H2-H5/H4-H5 due to larger intensity of NOE of H2-H5. The theoretical NOE ratio of H2-H5/H4-H5 was separately calculated for $^1C_4$, ^{2}So and $^4C_1$ conformation as 4.53 %, 52.73 % and 3.11 %, respectively (Table 12). Since IdoA rapidly populates among these three conformations, multiplication of population of each conformation acquired the averaged theoretical NOE ratio values for each construct (Table 12). The trend of increased theoretical NOE ratio values through construct 1 to 4 shows the increased population of ^{2}So conformation by additional 6-*O*- and 3-*O*-sulfation. In contrast, the decreased theoretical NOE ratio values through construct 5 to 8 displays the reduction of ^{2}So population. On the other hand, NMR 2D-NOESY experiments were performed for each construct. The NOE cross peaks of iduronic acid H2-H5 and H4-H5 were integrated for intensity ratio calculation. Through

construct 1 to 4 as well as construct 5 to 8, the opposite trends of experimental NOE ratio of H2-H5/H4-H5 are nicely consistent with the theoretical NOE ratio of H2-H5/H4-H5. The NOE analysis further supports different sulfation pattern could shift conformation equilibrium of heparan sulfate as observed in previous analysis of $^{3}J_{\text{H-H}}$ values.

	Averaged proton distance of simulations in the respective conformation (Ångstrom)		
	1C4	2So	4C1
H2-H5	4.0	2.6	4.2
H4-H5	2.4	2.3	2.4
	Theoretical [(NOE: H2-H5)/(NOE: H4-H5) Ratio ($\propto$ 1 / r ^6)		
	4.53%	52.73%	3.11%

	Theoretical IdoA/2S Ring Populations			Theoretical NOE H2-H5/H4-H5 Ratio	Experimental NOE H2-H5/H4-H5 Ratio
	1C4	2So	4C1		
Construct **1**	68.4%	31.6%	0.0%	19.8%	19.2%
Construct **2**	68.2%	31.8%	0.0%	19.9%	20.8%
Construct **3**	46.7%	53.3%	0.0%	30.2%	25.5%
Construct **4**	24.9%	75.1%	0.0%	40.7%	34.4%
Construct **5**	52.8%	46.8%	0.3%	27.1%	34.7%
Construct **6**	51.2%	48.8%	0.0%	28.1%	34.8%
Construct **7**	67.5%	29.2%	3.3%	18.6%	21.9%
Construct **8**	72.8%	18.8%	8.4%	13.5%	18.6%

Table 12. Experimental and theoretical NOE H2-H5/H4-H5 ratio of IdoA/2S among synthetic constructs.

Possible Mechanism for Shift of Conformation Equilibrium

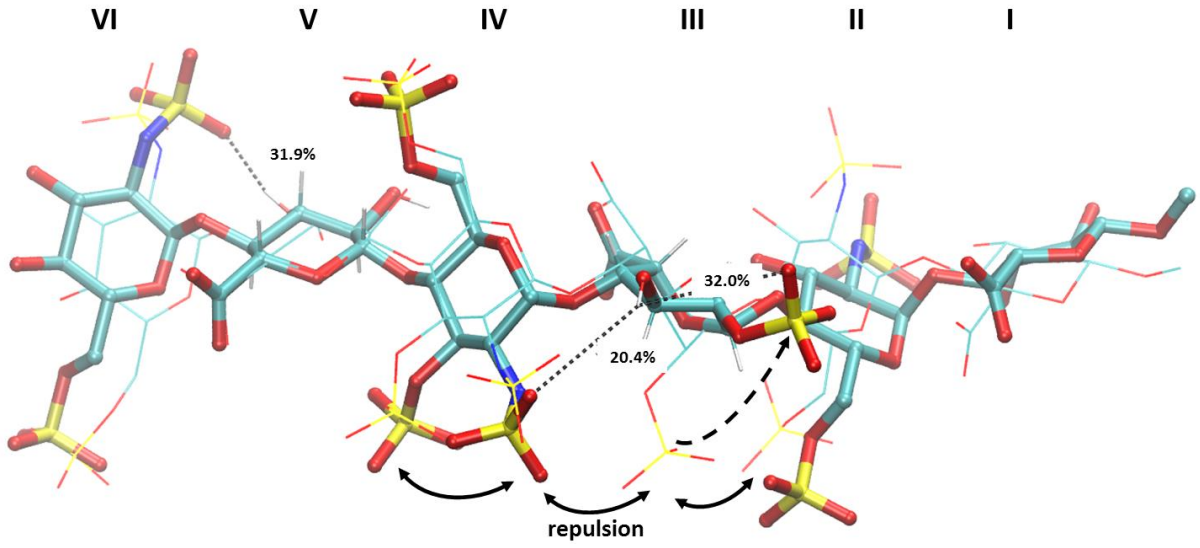


Figure 67. Superimposition of low energy conformations of construct 4 with iduronic acid as major ^{2}So conformation (bold bond) and minor $^{1}C_{4}$ conformation (thin line). The featured possible hydrogen bond and electrostatic repulsion is highlighted by dashes with occupancy and double arrows, respectively. The hydrogen bond is based on atom distance cutoff of 3.5 Å and angle of 120°.

The possible mechanism for the shift of conformation equilibrium among constructs is attributed to hydrogen bond and electrostatic repulsion using MD simulations. For individual $^{1}C_{4}$, ^{2}So and $^{4}C_{1}$ conformation of iduronic acid in each construct, the simulation was performed separately for all 8 constructs. A total of 24 independent simulations were finished with the extended 1 microsecond to acquire the proper averaged low energy conformation. Based on the low energy conformation of construct 4, the 2-*O*-sulfated IdoA prefers ^{2}So conformation rather than $^{1}C_{4}$ conformation possibly due to two main reasons: hydrogen bond and electrostatic repulsion (Figure 67). Firstly, the 2-*O*-sulfo group is capable to form an intra-residue hydrogen bond with 3-*O*-hydroxyl group in the ^{2}So conformation of IdoA2S, but not at $^{1}C_{4}$ conformation (Figure 67). The hydrogen bond occurs frequently among all IdoA2S-containing constructs with occupancy higher than 30 % (Table 13). The orientated 3-*O*-hydroxy group of IdoA2S in the

^{2}So conformation is able to from an inter-residue hydrogen bond with *N*-sulfate at residue IV as well (Figure 67). Secondly, the spacial adjacent 3-*O*- and *N*-sulfo groups at residue IV as well as 6-*O*-sulfo at residue II could contribute repulsion to the 2-*O*-sulfo group of IdoA2S in the $^{1}C_{4}$ conformation. The electrostatic repulsion makes IdoA2S disfavor the $^{1}C_{4}$ conformation and prefer the ^{2}So conformation (Figure 67). These hydrogen bond and electrostatic repulsion effects could synergistically lead IdoA2S conformation equilibrium toward the ^{2}So conformation in construct 4.

	IdoA/2S Conformation	H-bond Acceptor	H-bond Donor	H-bond Occupancy
Construct 1	IdoA2S-2So	SO3 at residue III	3-OH at residue III	38.0%
Construct 2	IdoA2S-2So	SO3 at residue III	3-OH at residue III	38.8%
Construct 3	IdoA2S-2So	SO3 at residue III	3-OH at residue III	30.5%
Construct 4	IdoA2S-2So	SO3 at residue III	3-OH at residue III	32.0%
Construct 5	IdoA-1C4	SO3 at residue IV	2-OH at residue III	16.6%
Construct 6	IdoA-1C4	SO3 at residue IV	2-OH at residue III	16.7%
Construct 7	IdoA-1C4	SO3 at residue IV	2-OH at residue III	15.2%
Construct 8	IdoA-1C4	SO3 at residue IV	2-OH at residue III	47.2%

Table 13. The occupancy of possible hydrogen bond during 1us MD simulation.

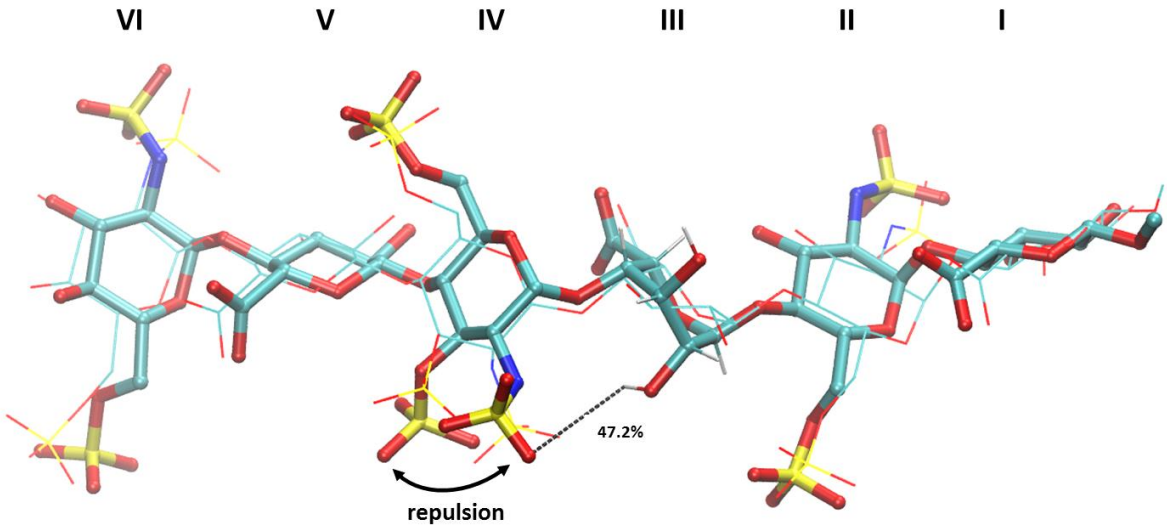


Figure 68. Superimposition of low energy conformations of construct 8 with iduronic acid as major $^{1}C_{4}$ conformation (bold bond) and minor ^{2}So conformation (thin line). The featured possible hydrogen bond and electrostatic repulsion is highlighted by dashes with occupancy and double arrows, respectively. The hydrogen bond is based on atom distance cutoff of 3.5 Å and angle of 120°.

On the other hand, the IdoA in construct 8 lacks 2-*O*-sulfation and majorly populates the $^{1}C_{4}$ instead of the ^{2}So conformation (Figure 68). The possible mechanism is the repulsion from the 3-*O*-sulfo group repulses the nearby *N*-sulfate (residue IV) closer to the 2-*O*-hydroxy group of IdoA (residue III). Without the 2-*O*-sulfo group in construct 8, the 2-*O*-hydroxyl group of IdoA at $^{1}C_{4}$ conformer, rather than ^{2}So conformer, is openly spacial close to *N*-sulfate and capable to form an inter-residue hydrogen bond (Figure 68). The occupancy of this inter-residue hydrogen bond is significantly increased in construct 8 (47.2 %) comparing to construct 7 (15.2 %) further supports the role of added 3-*O*-sulfo group for shift of iduronic acid conformation equilibrium (Figure 69, Table 13). Comparing the conformation population of iduronic acid in construct 4 and 8, given the 2-*O*-sulfation or not could shift the conformation equilibrium of IdoA-containing oligosaccharides toward opposite preference.

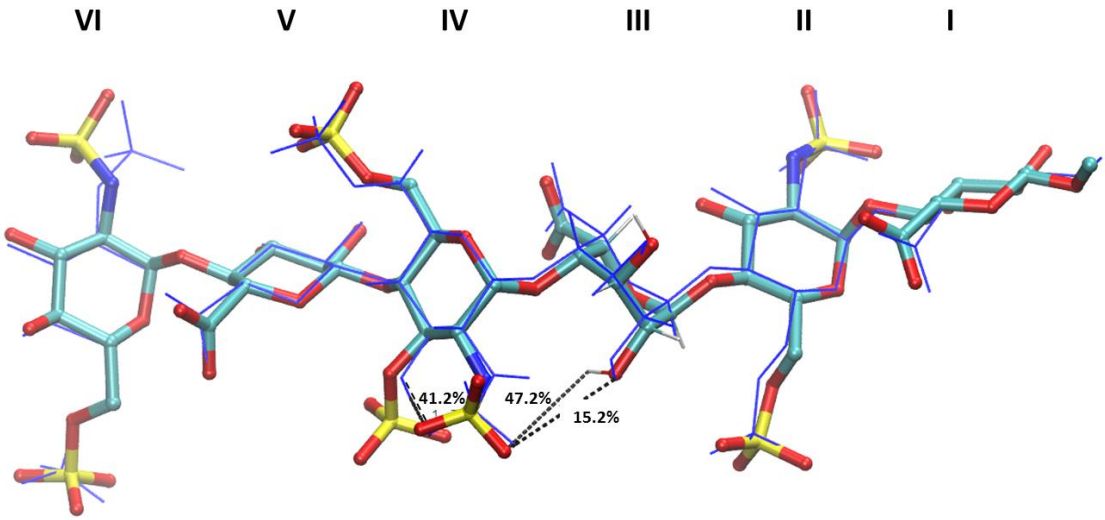


Figure 69. Superimposition of low energy conformations of construct 8 with IdoA as $^1C_4$ conformation (bold bond) and construct 7 with IdoA as $^1C_4$ conformation (thin blue line). The featured possible hydrogen bond is highlighted by dashes with occupancy. Without 3-*O*-sulfaiton, 3-*O*-hydorxy group and *N*-sulfate at residue IV could have intra-residue hydrogen bond with occupancy 41.2 %. The hydrogen bond is based on atom distance cutoff of 3.5 Å and angle of 120°.

Synthesis of Large Heparan Sulfate with Multiple IdoA/2S Saccharides

The presence of IdoA enhances the structural complexity for heparan sulfate. Unlike GlcA and GlcNS which are stable as $^4C_1$ conformation, IdoA was reported to adopt $^1C_4$ conformation and form a kink within linear structure of whole heparan sulfate molecules.[9] Furthermore, we found different sulfation patterns could change the conformational preference of IdoA, and accordingly shifts the substituent and ring position of the IdoA (Figure 70). Comparing to the rest of the residues, the orientation of 3-*O*-hydroxyl and 2-*O*-sulfo group is shifted while IdoA taking either $^1C_4$ or ^{2}So conformation. The relatively ring position of the IdoA is altered as well, which associated with some degree difference of glycosidic dihedral angle (Figure 70).

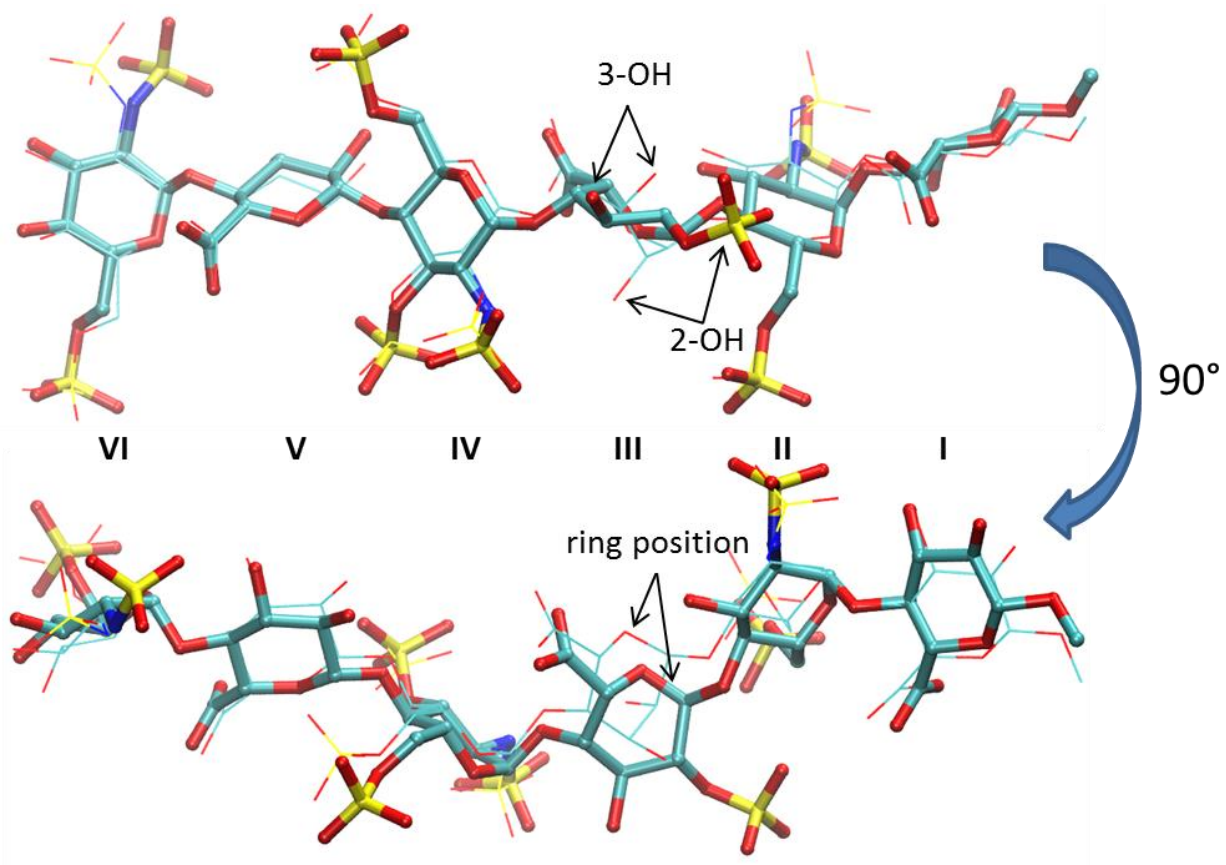


Figure 70. Superimposition of low energy conformations of construct 4 with IdoA2S as ^{2}So conformation (bold bond) and construct 8 with IdoA as $^{1}C_{4}$ conformation (thin line). Construct 4 and construct 8 are similar molecules except IdoA2S in construct 4 and IdoA in construct 8 at residue III. The bottom panel was made by rotate 90° of upper panel. The featured differences between two constructs are highlighted by arrow.

These structural heterogeneities are gradually accumulated in longer or multiple IdoA containing heparan sulfate. The conformational change of IdoA based on sulfation pattern even further enlarges the structural diversity. However, there may be only certain low energy heparan sulfate conformations that are relatively stable. The intramolecular constraint such as hydrogen bond or electrostatic repulsion could possibly establish some defined conformations. The widespread strong electronegative sulfo and polar hydroxyl groups of saccharide may be the

determining factors in the environmental influence on the conformation of IdoA/2S. Protein and nucleic acid are known to have secondary or tertiary structure due to intramolecular interactions with the extended size of these molecules. To understand the presence of any higher level of stable heparan sulfate conformation, an array of large size of IdoA/2S-containing oligosaccharides is urgently required for future study. In a pilot synthesis study, a series of oligosaccharide larger than hexasaccharide were or are to be synthesized consisting of multiple IdoA2S saccharides (Table 14). The isotope labeled saccharide can make the target saccharide signals distinguishable from other repeated saccharides. The conformation of IdoA2S and glycosidic linkage located at different desired sites of HS molecule, such as at reducing- or nonreducing-end, will be analyzed and compared.

compounds	Abbreviated structures
Octasaccharide	GlcNS-GlcA-GlcNS-**IdoA2S**-GlcNS-**IdoA2S**-GlcNS-GlcA-pNP
Nonasaccharide	**[^{13}C]GlcA**-GlcNS-GlcA-GlcNS-**IdoA2S**-GlcNS-**IdoA2S**-GlcNS-GlcA-pNP
Decasaccharide-1	GlcNS-**[^{13}C]GlcA**-GlcNS-**IdoA2S**-GlcNS-**IdoA2S**-GlcNS-**IdoA2S**-GlcNS-GlcA-pNP
Decasaccharide-2	GlcNS6S-**[^{13}C]GlcA**-GlcNS6S-**IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-GlcA-pNP
Decasaccharide-3	GlcNS6S-**[^{13}C]GlcA**-GlcNS3S6S-**IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-GlcA-pNP
Dodecasaccharide-1	GlcNS-GlcA-GlcNS-**[^{13}C]IdoA2S**-GlcNS-**IdoA2S**-GlcNS-**IdoA2S**-GlcNS-**IdoA2S**-GlcNS-GlcA-pNP
Dodecasaccharide-2	GlcNS6S-GlcA-GlcNS6S-**[^{13}C]IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-GlcA-pNP
Dodecasaccharide-3	GlcNS6S-GlcA-GlcNS3S6S-**[^{13}C]IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-**IdoA2S**-GlcNS6S-GlcA-pNP

Table 14. List of multi-IdoA-containing oligosaccharide prepared in the pilot synthesis. The IdoA2S saccharides are marked as bold font. The ^{13}C-labeled saccharides are indicated in red color.

Discussion

In addition to heparan sulfate, IdoA is also present in dermatan sulfate (DS) glycosaminoglycans. The structural differences leads dermatan sulfate to have distinct biological roles.[252] Dermatan sulfate uses β-D-galactosamine and -(1,3)- linkage instead of α-D-glucosamine and -(1,4)- linkage as in heparan sulfate. Additionally, the galactosamine can be 4-*O*-sulfated instead of 3-*O*-sulfation on glucosamine. Here we assert that 3-*O*-sulfation is an important factor in the conformation of the adjacent IdoA via intramolecular interaction. Similar sulfation effect may also impact the conformational preference of dermatan sulfate and even other type of glycosaminoglycans. Similar to the 3-*O*-sulfated pentasaccharide fondaparinux binding to AT for anticoagulation effect, special sulfation pattern may drive glycosaminoglycan conformation for other specific protein binding and biological activity. The carbohydrate microarray made by structure defined oligosaccharides may serve as useful tool for protein binding screening as well as investigating the conformation-activity relationship.

Conclusions

Diversity of sulfation patterns and plasticity of iduronic acid (IdoA) are known to be associated with the protein binding; however, the understanding of the relationship between sulfations and saccharide conformations is not well understood. Here, we developed a method to accurately analyze the saccharide conformation of HS. This study systematically investigates the relation between the various sulfations and the saccharide conformations among heparan sulfate. The chemoenzymatic approach was used to synthesize a total of eight IdoA/2S-containing hexasaccharides with different sulfation patterns. To calculate the realistic equilibrium

population of IdoA/2S conformations, we deconvoluted the NMR measured parameters into simulated parameters for each conformation with corresponding populations. Both spin-spin coupling constant ($^3J_{H\text{-}H}$) and intensity of nuclear overhauser effects (NOE) are used to monitor the conformational change and corresponding population in this study.

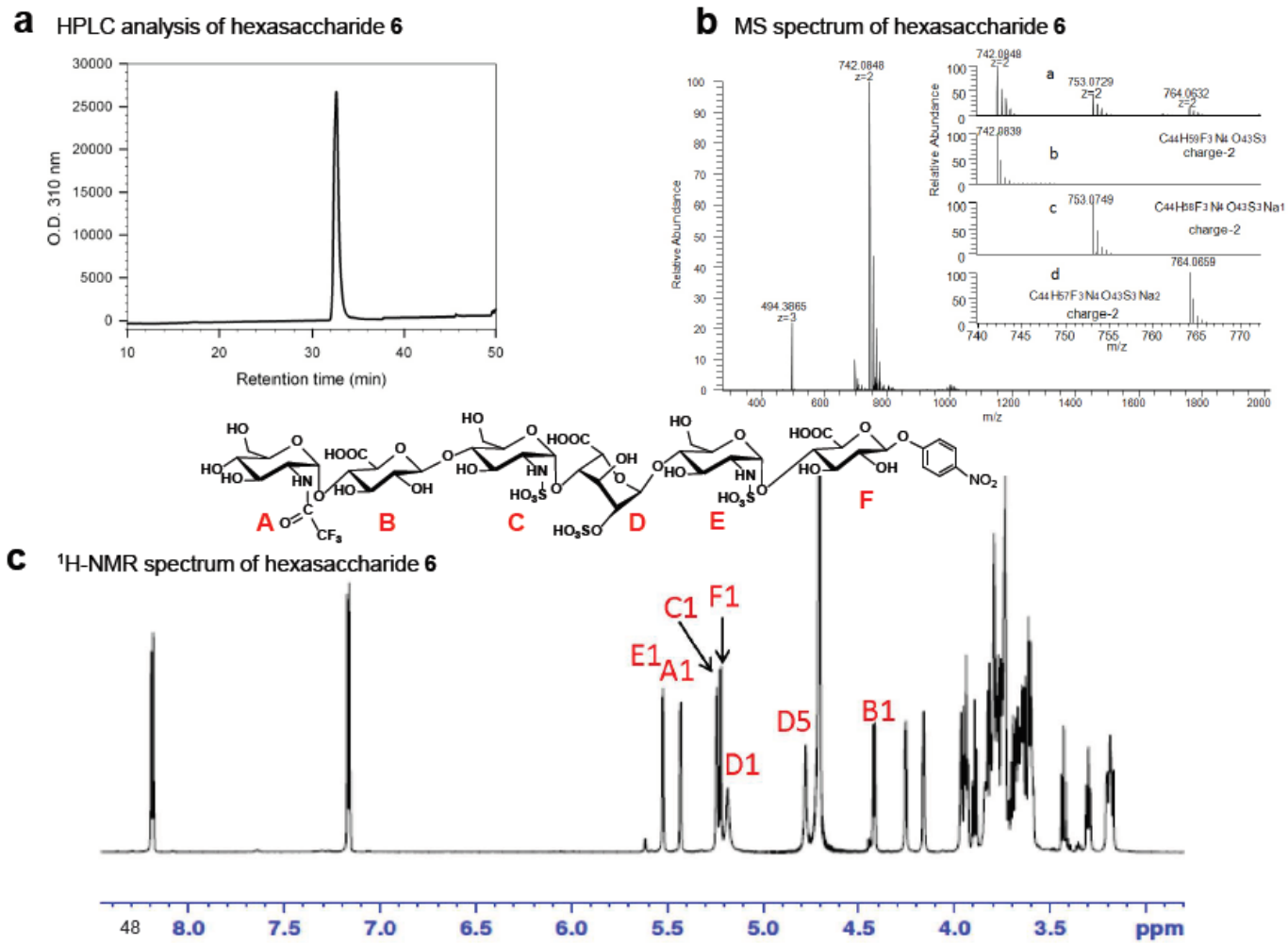
Our results directly suggest that different sulfation influences the heparan sulfate conformation. 2-*O*-sulfation is capable to shift the conformation equilibrium of IdoA. The biosynthetic downstream 6-*O*- and 3-*O*-sulfations provide different environment shifting IdoA2S toward increased ^{2}So conformation, which shifting IdoA population to $^1C_4$ conformation. Comparing to relatively rigid glucuronic acid, IdoA/2S-containing oligosaccharide could show diverse conformational preference in response to different sulfation patterns. This is the first demonstration that 2-*O*-sulfation cooperates with other sulfation to regulate conformational equilibrium of IdoA. The sequential *N*-, 6-*O*- and 3-*O*-sulfation are involved in settling local environment for different conformational preference of heparan sulfates.

The combination of the specific sulfation pattern and certain sequenced oligosaccharide can build pre-defined conformation. The buildup of conformation appears due to hydrogen bonds and electrostatic interactions. Therefore, the larger size and multi-IdoA/2S oligosaccharides are synthesized to investigate the conformation of individual saccharide at different locations of oligosaccharide. The labeled saccharides are introduced to facilitate NMR analysis by enhancing the signals of interested saccharide by 100-folds. Assembling the conformational analysis results from each labeled piece will provide a complete conformation for the oligosaccharide. In our opinion, our method will provide an unique, and the most effective by far, solution to analyze the conformation of HS oligosaccharides. It may be existed the specific oligosaccharide conformation for desired biological function. The understanding of the

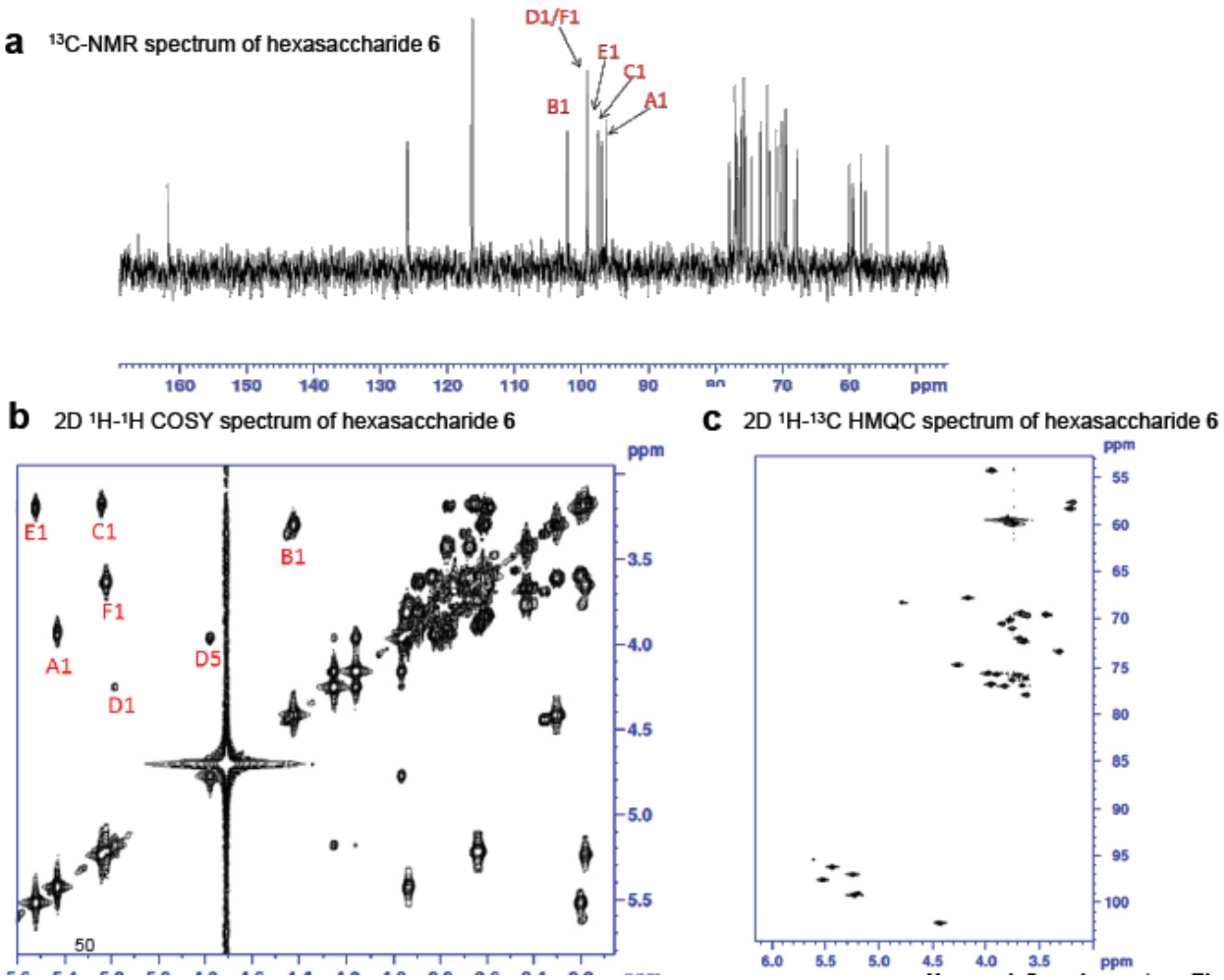
sulfation-conformation relationship will assist the design of next generation heparin-based therapeutics.

APPENDIX I

SUPPLEMENTARY HPLC, MASS, AND NMR DATA



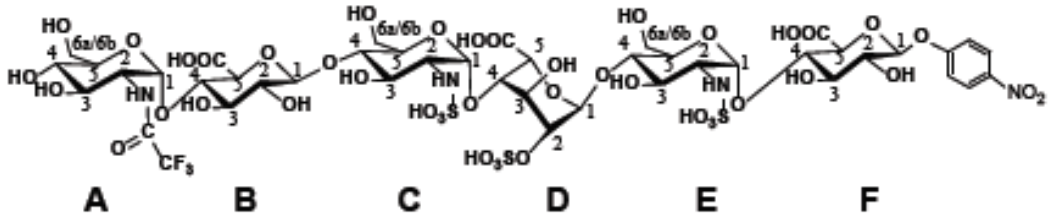
Supplementary Fig 1. **Purity and structural analysis of hexasaccharide 6**. (a) Silica-based polyamine(PAMN)-HPLC chromatogram. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data, and inset b, c and d is simulated theoretical data for H adduct, one Na adduct and two Na adduct. Accurate MS and isotope distribution matched very well (within 5 ppm mass accuracy) with the empirical formulae. (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The signal from H-5 of Residue D is also indicated, demonstrating the presence of an IdoA2S residue. The chemical structure of hexasaccharide **6** is shown on top of (c).



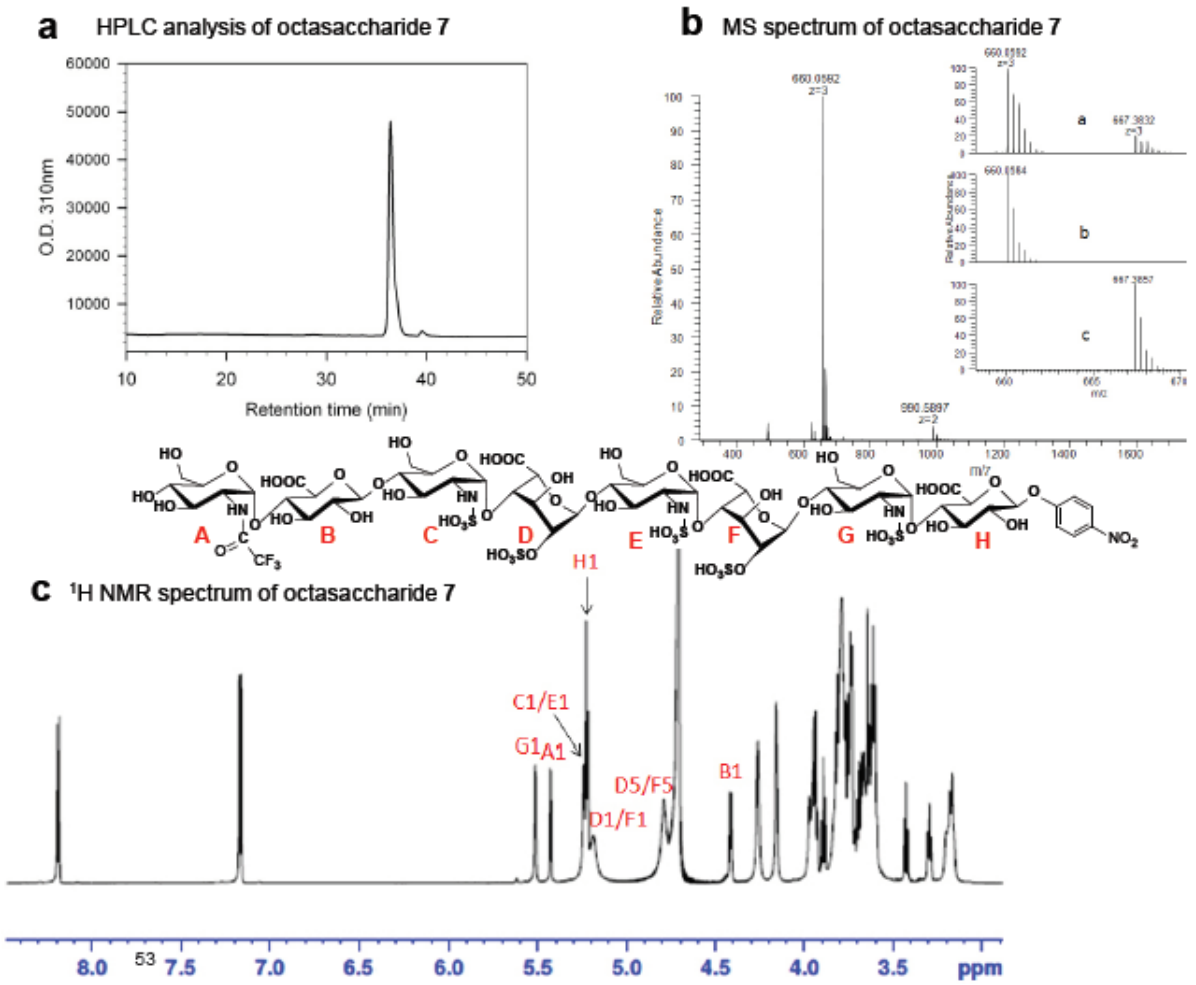
Supplementary Fig 2. **NMR analysis of hexasaccharide 6**. (a) 13C-NMR spectrum of hexasaccharide **6**. Signals from six anomeric carbons are identified. (b) 2D 1H-1H COSY spectrum of hexasaccharide **6**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.52 (d, J = 3.2 Hz, 1H), 5.43 (d, J = 4.0 Hz, 1H), 5.24 (d, J = 3.2 Hz, 1H), 5.22 (d, J = 8.0 Hz, 1H), 5.18 (broad singlet, 1H), and 4.41 (d, J = 8.0 Hz, 1H) ppm. The chemical shift of D1 on IdoA is 5.18 ppm (broad singlet, 1H). (c) 2D 1H-13C HMQC spectrum of hexasaccharide **6**.

^{1}H NMR /^{13}C NMR chemical shift assignments (in ppm) of hexasaccharide **6**

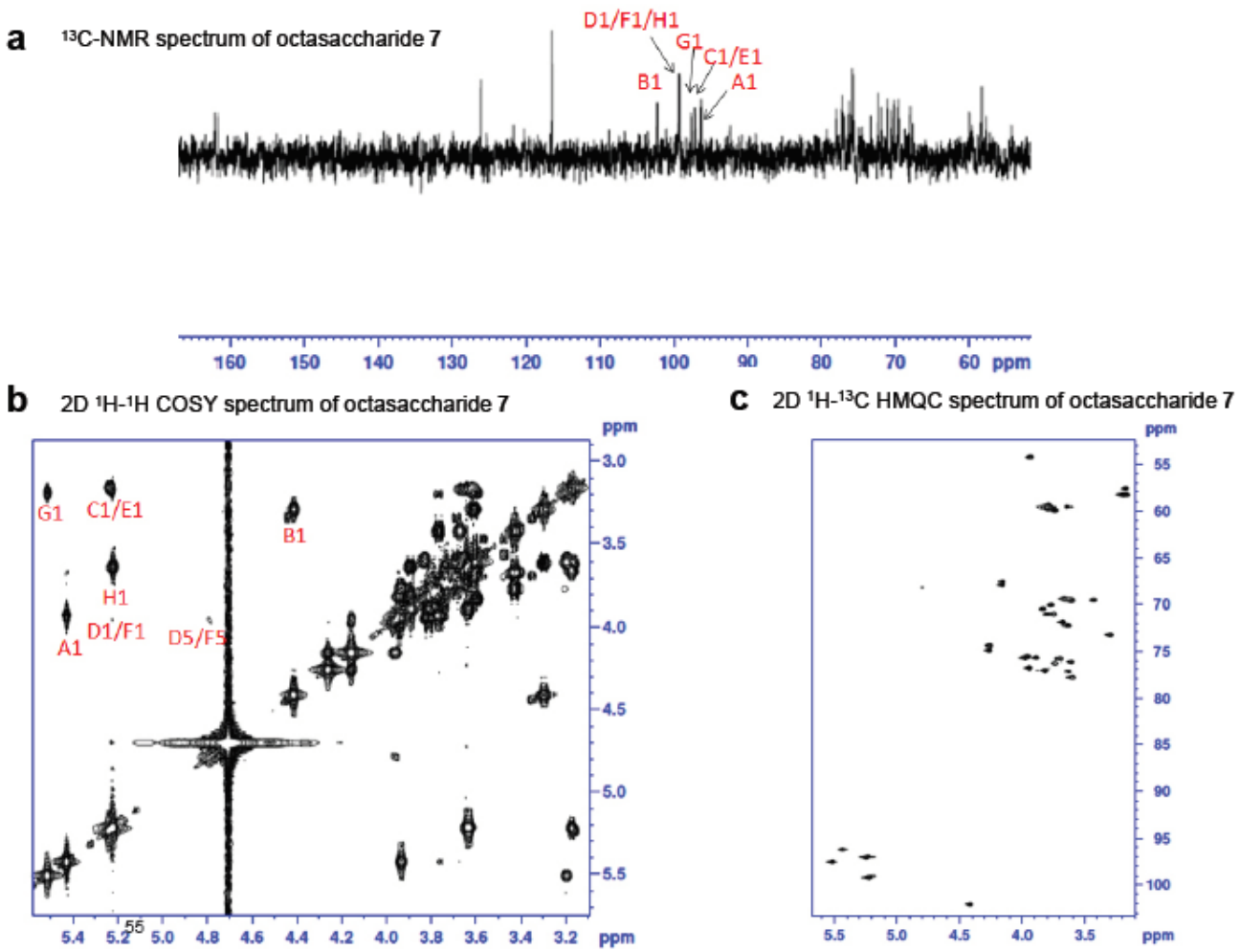
	1	2	3	4	5	6a	6b
A	5.43/96.2	3.93/54.1	3.77/70.0	3.43/69.5	3.67/69.4	3.71/59.9	3.78/59.9
B	4.41/102.1	3.29/73.3	3.61/77.8	3.63/76.1	3.73/76.3		
C	5.24/96.9	3.17/57.6	3.69/71.8	3.70/75.8	3.73/70.9	3.74/59.4	3.79/59.4
D	5.18/98.9	4.26/74.7	4.15/67.7	3.94/76.7	4.77/68.1		
E	5.52/97.4	3.19/58.2	3.60/69.6	3.66/76.9	3.84/70.4	3.73/59.4	3.78/59.4
F	5.22/99.1	3.65/72.1	3.89/75.6	3.81/77.0	3.96/75.5		



Supplementary Fig 3. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of hexasaccharide 6**. The structure of hexasaccharide **6** is shown. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



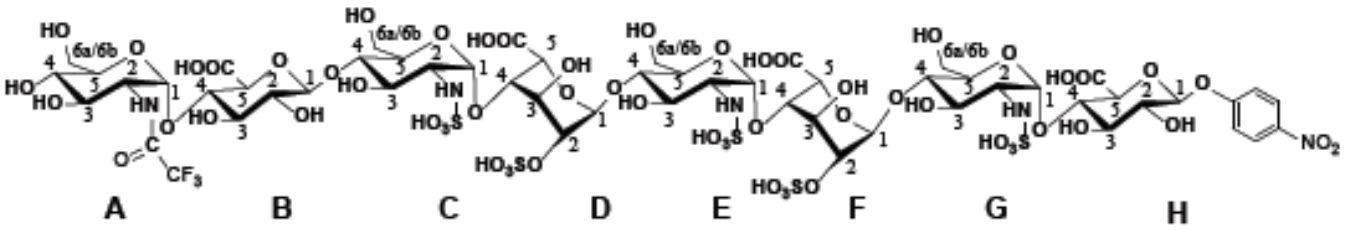
Supplementary Fig 4. **Purity and structural analysis of octasaccharide 7**. (a) PAMN-HPLC chromatogram. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data, and inset b and c is simulated theoretical data for H adduct and one Na adduct for the empirical formulae C56H80F3N5O59S5 and C56H79F3N5O59S5Na, respectively. Accurate MS and isotope distribution matched very well (within 5 ppm mass accuracy) with the empirical formulae. (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The signals from H-5 of Residue D and F are also indicated, demonstrating the presence of two IdoA2S residues. The chemical structure of octasaccharide **7** is shown on top of (c).



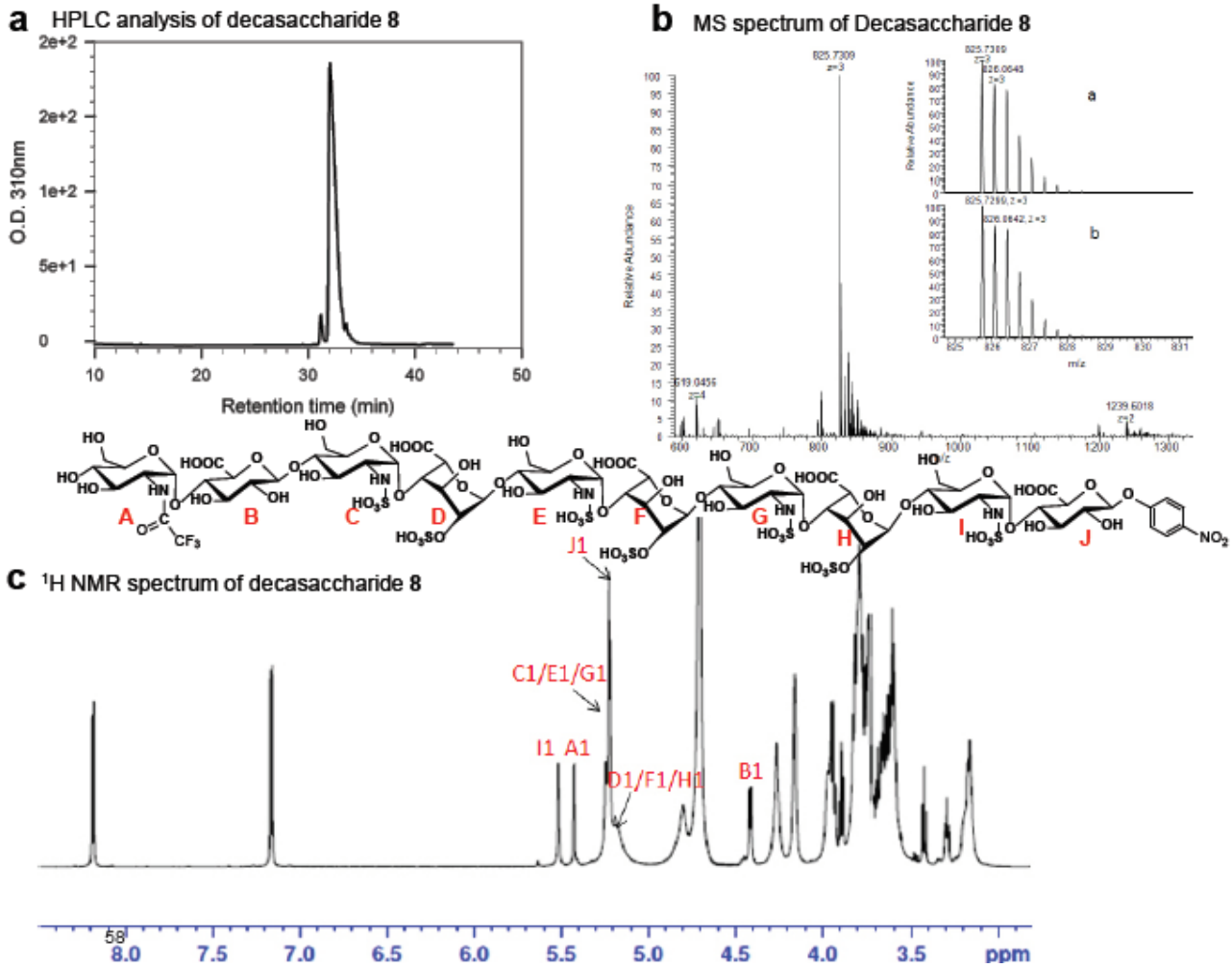
Supplementary Fig 5. **NMR analysis of octasaccharide 7**. (a)13C-NMR spectrum of octasaccharide **7**. Signals from eight anomeric carbons are identified. (b) 2D 1H-1H COSY spectrum of octasaccharide **7**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.51 (d, J = 3.2 Hz, 1H), 5.43 (d, J = 4.0 Hz, 1H), 5.24 (s, 1H), 5.23 (s, 1H), 5.22 (d, J = 8.0 Hz, 1H), 5.18 (broad singlets, 2H), and 4.42 (d, J = 8.0 Hz, 1H) ppm. (c) 2D 1H-13C HMQC spectrum of octasaccharide **7**.

^{1}H NMR/^{13}C NMR chemical shift assignments (in ppm) of octasaccharide **7**

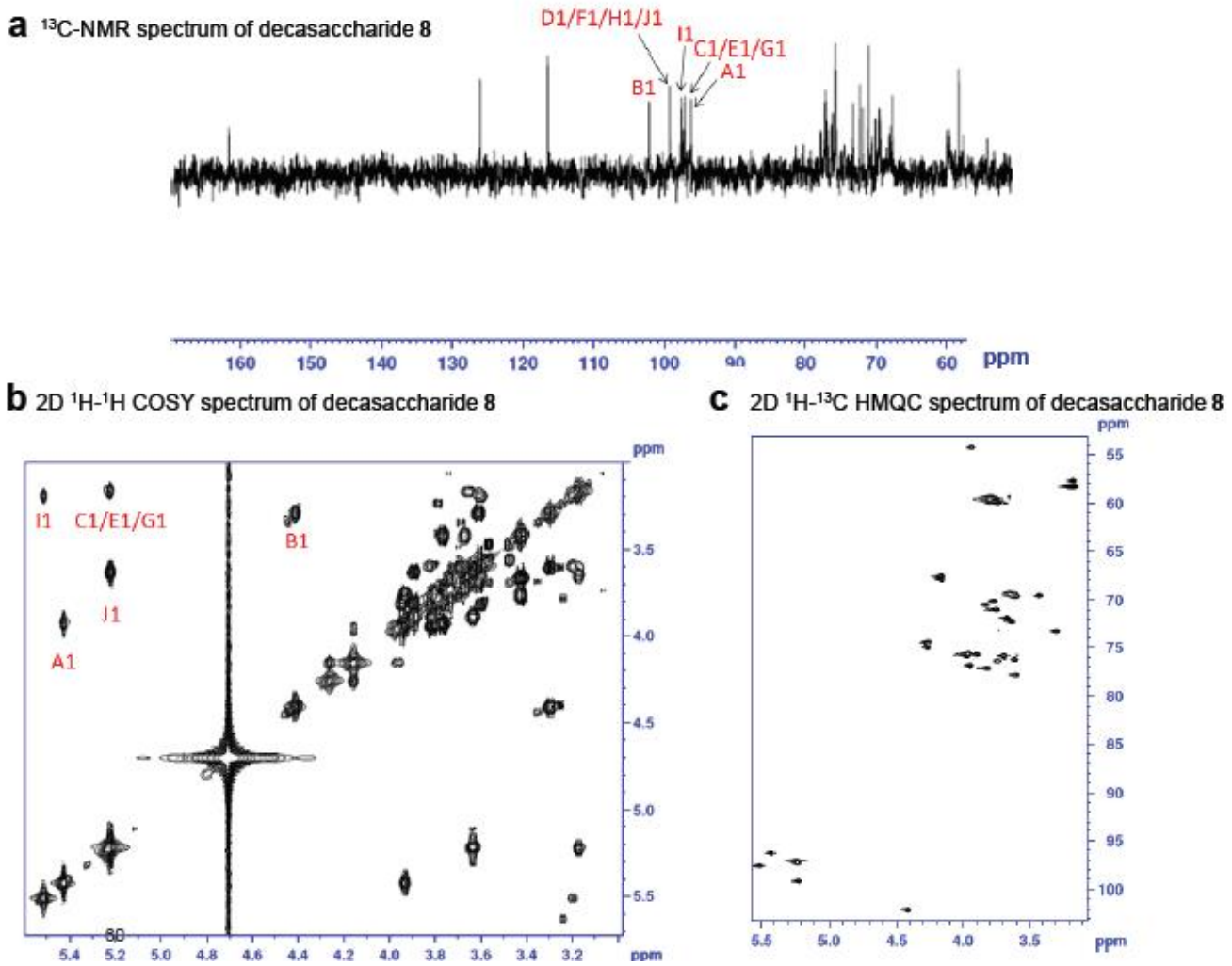
	1	2	3	4	5	6a	6b
A	5.43/96.2	3.93/54.1	3.77/70.0	3.43/69.5	3.67/69.4	3.71/59.9	3.78/59.9
B	4.42/102.1	3.29/73.3	3.61/77.8	3.69/76.1	3.73/76.3		
C	5.23/96.9	3.17/57.6	3.65/71.8	3.70/75.8	3.73/70.9	3.74/59.5	3.79/59.5
D	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1		
E	5.24/97.1	3.17/58.2	3.65/71.8	3.70/75.8	3.73/71.0	3.72/59.8	3.78/59.8
F	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1		
G	5.51/97.4	3.19/58.2	3.60/69.6	3.66/76.9	3.84/70.4	3.73/59.4	3.78/59.4
H	5.22/99.1	3.64/72.1	3.89/75.6	3.81/77.0	3.94/75.5		



Supplementary Fig 6. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of octasaccharide 7**. The structure of octasaccharide **7** is shown. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



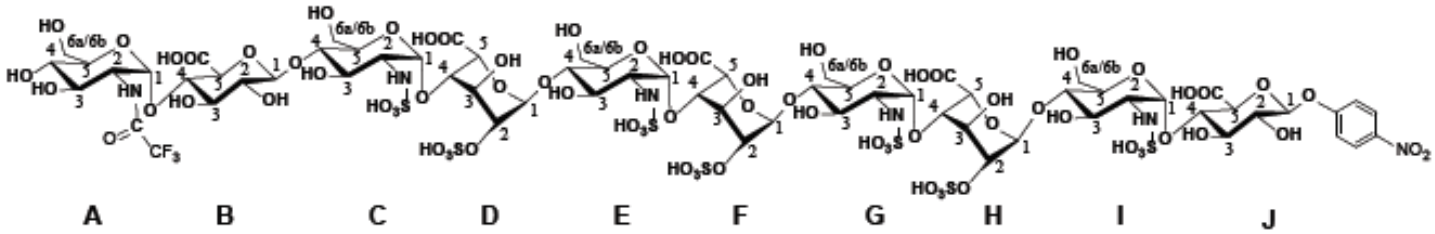
Supplementary Fig 7. **Purity and structural analysis of decasaccharide 8**. (a) PAMN-HPLC chromatogram. A small peak eluted before the major peak is a detrifluoroacetylated decasaccharide **8**. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data, and inset b is simulated theoretical data. Accurate MS and isotope distribution matched very well (within 5 ppm mass accuracy) with the empirical formula (C68H99F3N6O75S7). (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The chemical structure of decasaccharide **8** is shown on top of (c).



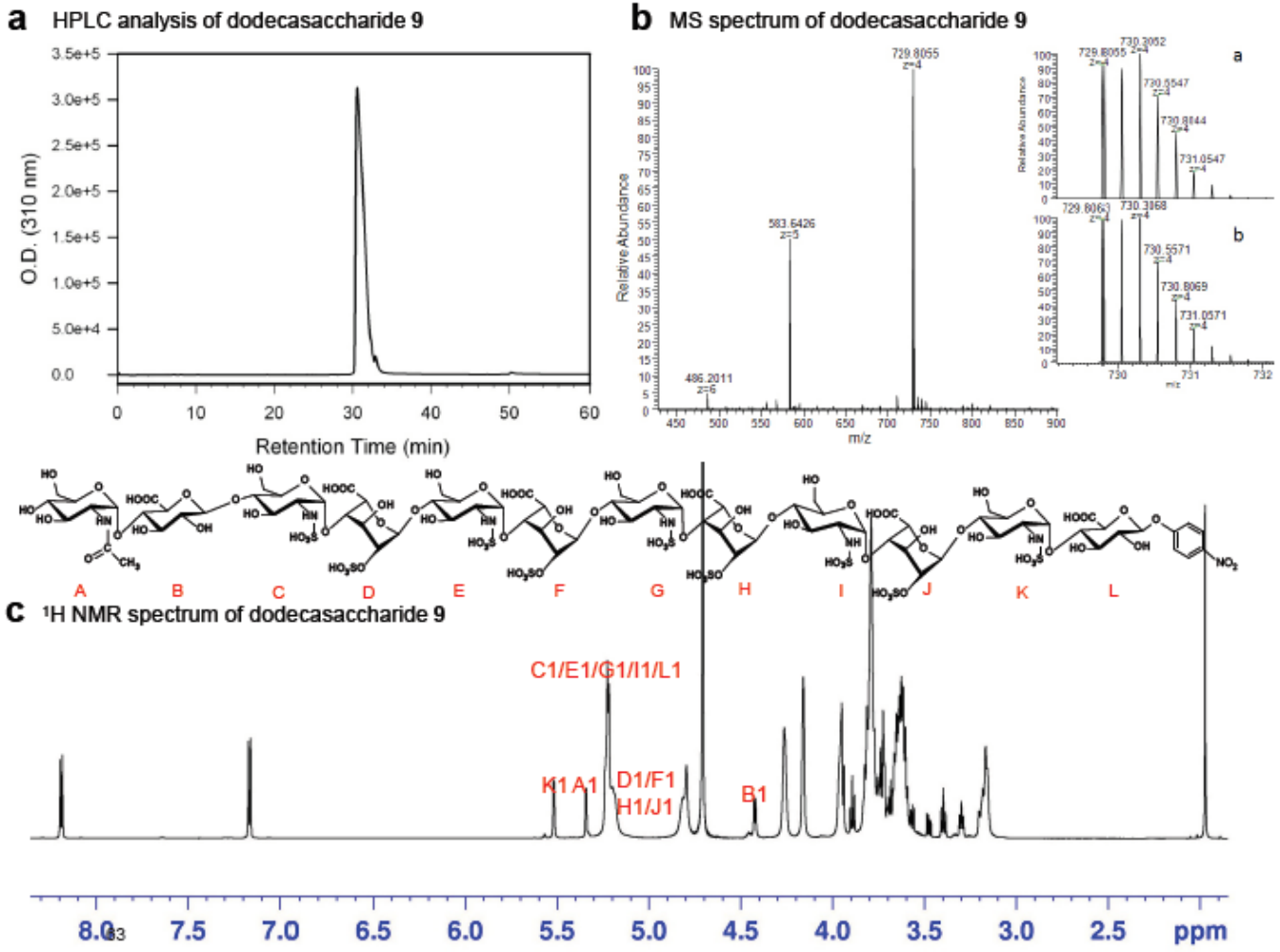
Supplementary Fig 8. **NMR analysis of decasaccharide 8**. (a) 13C-NMR spectrum of decasaccharide **8**. Signals from ten anomeric carbons are identified. (b) 2D 1H-1H COSY spectrum of decasaccharide **8**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.51 (d, J = 3.2 Hz, 1H), 5.43 (d, J = 4.0 Hz, 1H), 5.24 (s, 3H), 5.22 (d, 1H), 5.18 (broad singlets, 3H), and 4.42 (d, J = 7.2 Hz, 1H) ppm. The intensity of the signal at 5.18 ppm is consistent with three protons, suggesting that three IdoA2S residues are present in decasaccharide **8**. (c) 2D 1H-13C HMQC spectrum of decasaccharide **8**.

^{1}H NMR/^{13}C NMR chemical shift assignments (in ppm) of decasaccharide **8**

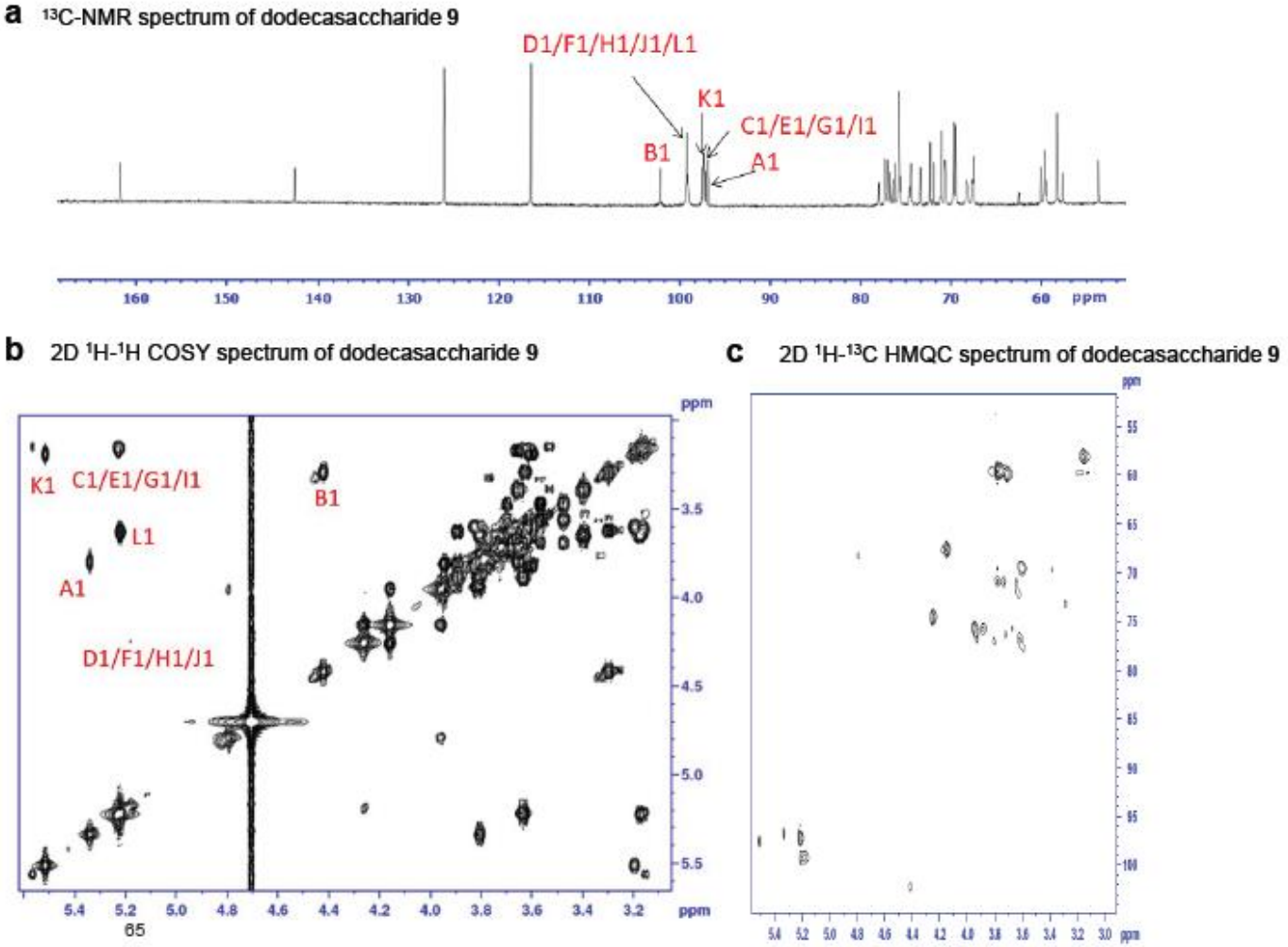
	1	2	3	4	5	6a	6b
A	5.43/96.2	3.93/54.1	3.77/70.0	3.43/69.5	3.67/69.4	3.71/59.9	3.78/59.9
B	4.42/102.1	3.29/73.3	3.61/77.8	3.69/76.1	3.73/76.3		
C	5.23/96.9	3.17/57.6	3.65/71.8	3.70/75.8	3.73/70.9	3.74/59.5	3.79/59.5
D	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1		
E	5.24/97.1	3.17/58.2	3.65/71.8	3.70/75.8	3.73/71.0	3.72/59.8	3.78/59.8
F	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1		
G	5.24/97.1	3.17/58.2	3.65/71.8	3.70/75.8	3.73/71.0	3.72/59.8	3.78/59.8
H	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1		
I	5.51/97.4	3.19/58.2	3.60/69.6	3.66/76.9	3.84/70.4	3.73/59.4	3.78/59.4
J	5.22/99.1	3.64/72.1	3.89/75.6	3.81/77.0	3.94/75.5		



Supplementary Fig 9. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of decasaccharide 8**. The structure of decasaccharide **8** is shown. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



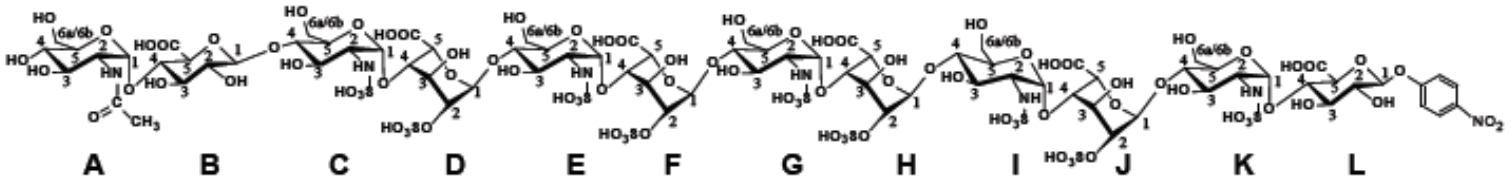
Supplementary Fig 10. **Purity and structural analysis of dodecasaccharide 9**. (a) PAMN-HPLC chromatogram. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data, and inset b is simulated theoretical data. Accurate MS and isotope distribution matched very well (within 5 ppm mass accuracy) with the empirical formula ($C_{80}H_{121}N_7O_{91}S_9$). (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The chemical structure of dodecasaccharide **9** is shown on top of (c).



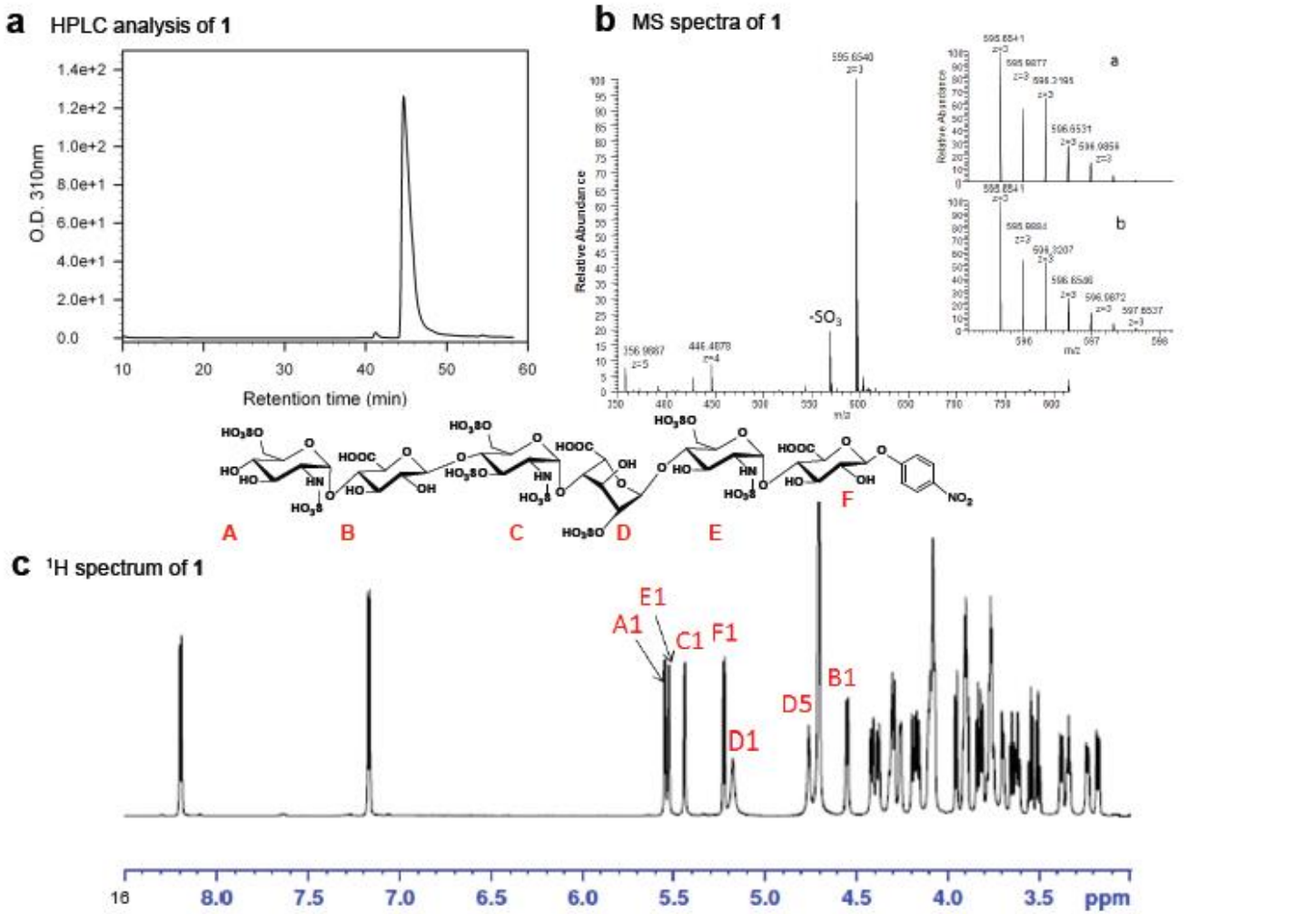
Supplementary Fig 11. **NMR analysis of dodecasaccharide 9**. (a) 13C-NMR spectrum of dodecasaccharide **9**. Signals from twelve anomeric carbons are identified. (b) 2D 1H-1H COSY spectrum of dodecasaccharide **9**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.51 (d, J = 3.2 Hz, 1H), 5.34 (d, J = 4.0 Hz, 1H), 5.24-5.20 (m, 5H), 5.18 (broad singlets, 4H), and 4.42 (d, J = 7.2 Hz, 1H) ppm. The intensity of the signal at 5.18 ppm is consistent with four protons, suggesting that four IdoA2S residues are present in dodecasaccharide **9**. (c) 2D 1H-13C HMQC spectrum of dodecasaccharide **9**.

^{1}H NMR/^{13}C NMR chemical shift assignments (in ppm) of dodecasaccharide **9**

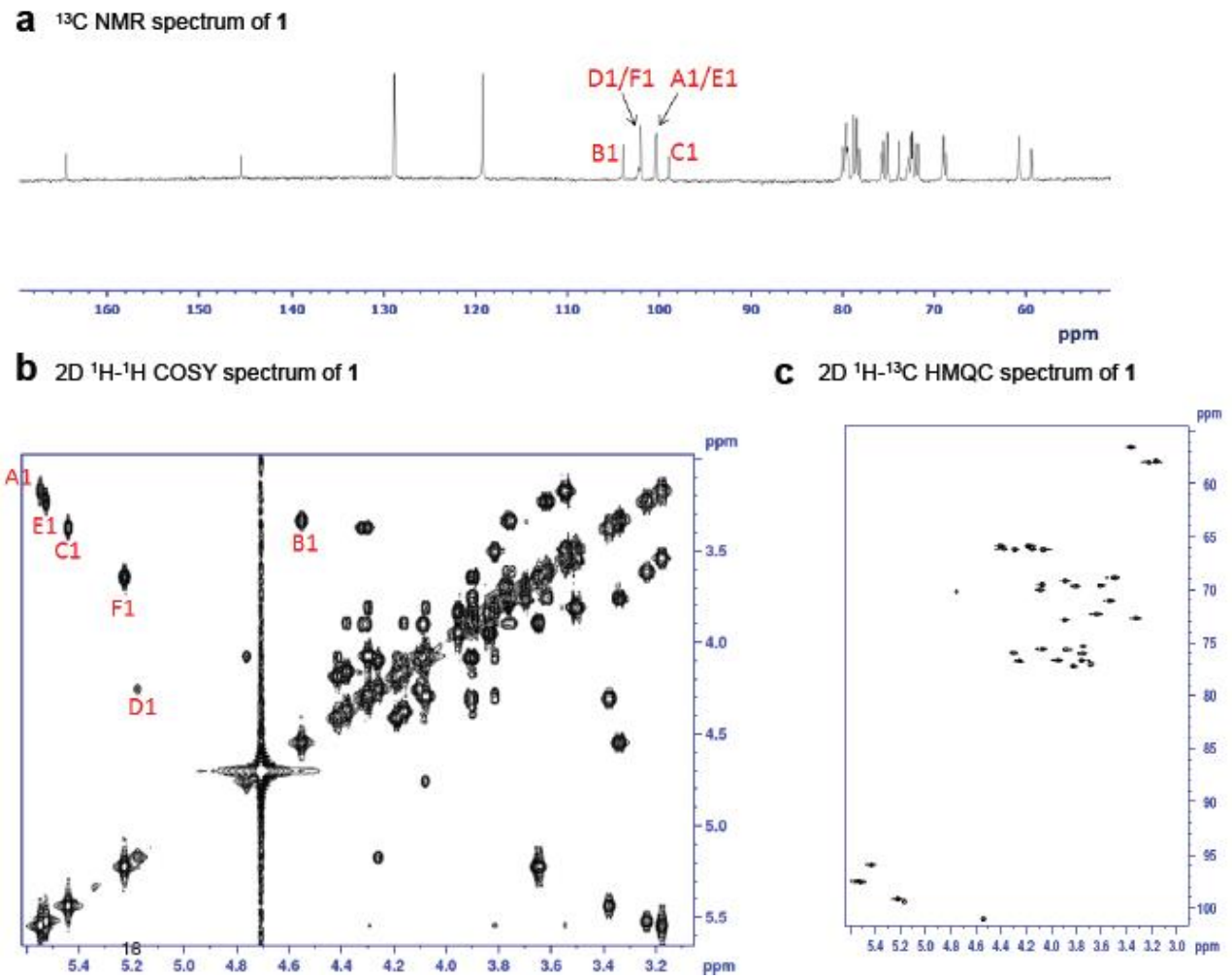
	1	2	3	4	5	6a	6b	-$COCH_3$
A	5.34/96.8	3.80/53.5	3.65/70.9	3.39/69.7	3.68/69.7	3.71/59.6	3.78/59.6	2.03
B	4.42/102.1	3.30/73.3	3.62/77.8	3.67/76.1	3.72/76.3			
C	5.21/97.1	3.17/58.2	3.65/71.8	3.70/75.8	3.73/71.0	3.72/59.8	3.78/59.8	
D	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1			
E	5.21/97.1	3.17/58.2	3.65/71.8	3.70/75.8	3.73/71.0	3.72/59.8	3.78/59.8	
F	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1			
G	5.22/97.1	3.17/58.2	3.65/71.8	3.70/75.8	3.73/71.0	3.72/59.8	3.78/59.8	
H	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1			
I	5.22/97.1	3.17/58.2	3.65/71.8	3.70/75.8	3.73/71.0	3.72/59.8	3.78/59.8	
J	5.18/98.9	4.26/74.7	4.16/67.7	3.96/76.7	4.79/68.1			
K	5.51/97.4	3.19/58.2	3.60/69.6	3.66/76.9	3.84/70.4	3.73/59.4	3.78/59.4	
L	5.22/99.1	3.64/72.1	3.89/75.6	3.81/77.0	3.94/75.5			



Supplementary Fig 12. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of dodecasaccharide 9**. The structure of dodecasaccharide **9** is shown. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



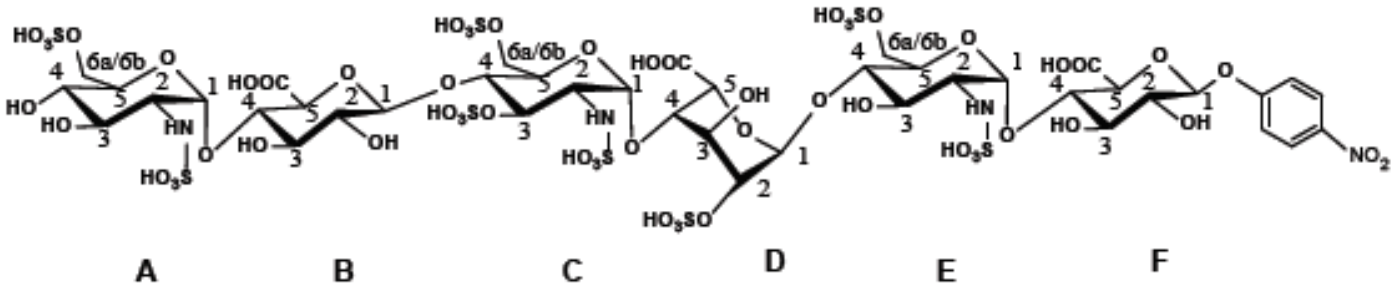
Supplementary Fig 13. **Purity and structural analysis of 1**. (a) DEAE-HPLC chromatogram. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data for **1**, and inset b is simulated theoretical data accurate MS data for **1** at charge 3 in negative electrospray ionization (ESI) mode. Isotope distribution analysis by high resolution MS matched (within 5 ppm mass accuracy) with the empirical formula ($C_{42}H_{62}N_4O_{57}S_8$). "-SO3" indicates the signal of the molecule ion losing a sulfate group. (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The chemical structure of **1** is shown on top of (c).



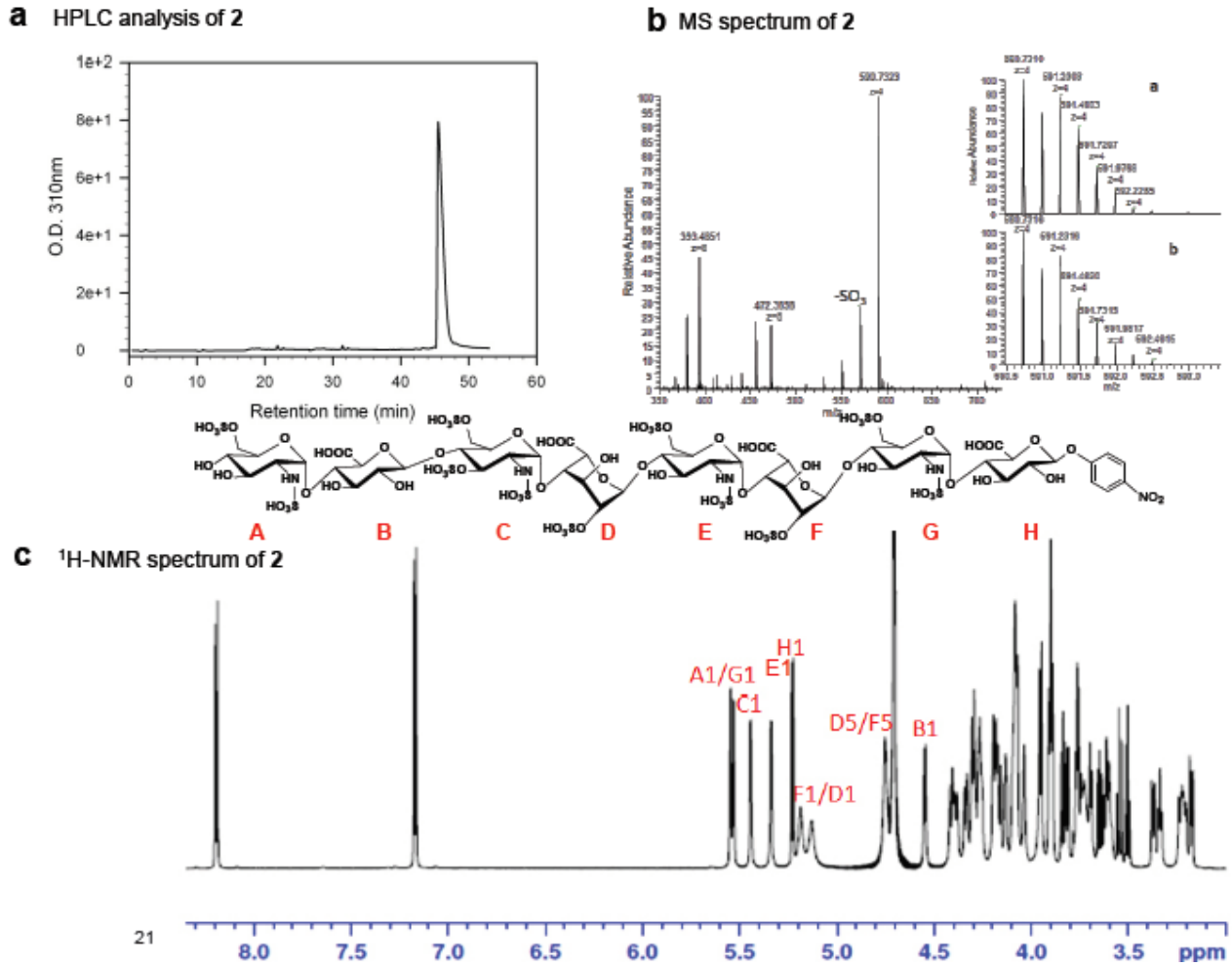
Supplementary Fig 14. **NMR analysis of 1**. (a) 13C-NMR spectrum of **1**. Signals from six anomeric carbons are labeled. (b) 2D 1H-1H Correlation Spectroscopy (COSY) spectrum of **1**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.55 (d, J = 3.4 Hz, 1H), 5.53 (d, J = 3.3 Hz, 1H), 5.44 (d, J = 2.5 Hz, 1H), 5.23 (d, J = 7.8 Hz, 1H), 5.18 (broad singlet, 1H), and 4.55 (d, J = 7.8 Hz, 1H) ppm. The small coupling constants (<4 Hz) of the anomeric protons indicate an α linkage, and the large coupling constants (~8 Hz) indicate a β linkage. (c) 2D 1H-13C Heteronuclear Multiple-Quantum Correlation (HMQC) spectrum of **1**.

^{1}H NMR/^{13}C NMR chemical shift assignments (in ppm) of **1**

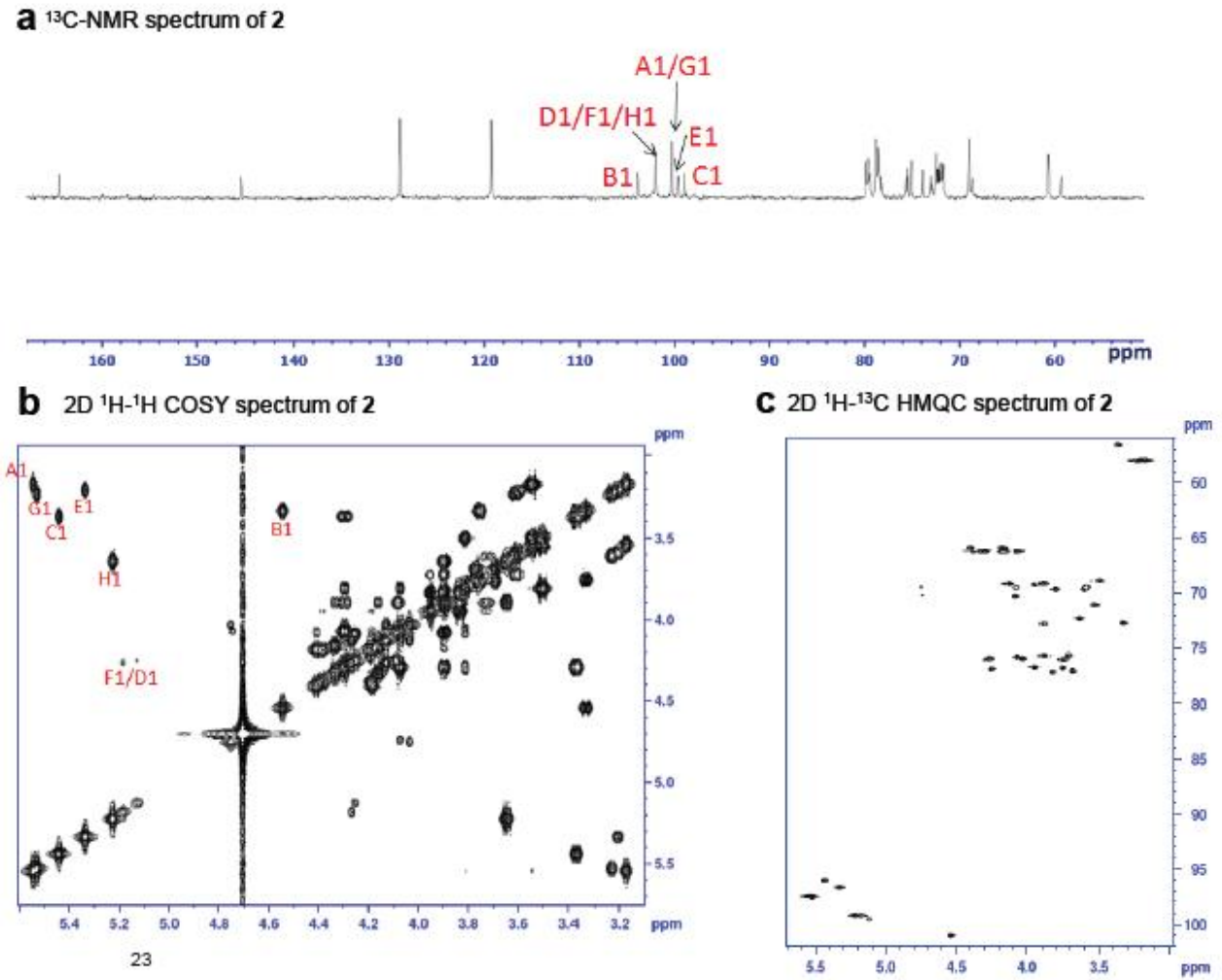
	1	2	3	4	5	6a	6b
A	5.55/97.4	3.18/57.8	3.55/70.9	3.49/68.8	3.81/69.6	4.08/66.0	4.29/66.0
B	4.55/101.1	3.34/72.5	3.76/75.3	3.76/76.6	3.70/76.9		
C	5.44/96.0	3.38/56.4	4.30/75.8	3.89/75.6	4.08/69.5	4.19/65.8	4.41/65.8
D	5.18/99.3	4.26/76.6	4.10/70.0	4.07/75.6	4.76/70.1		
E	5.53/97.4	3.23/57.8	3.62/69.5	3.76/76.0	3.90/69.0	4.16/66.1	4.37/66.1
F	5.23/99.1	3.64/72.2	3.89/75.6	3.82/77.2	3.95/76.6		



Supplementary Fig 15. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of 1.** The structure of **1** is also shown. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



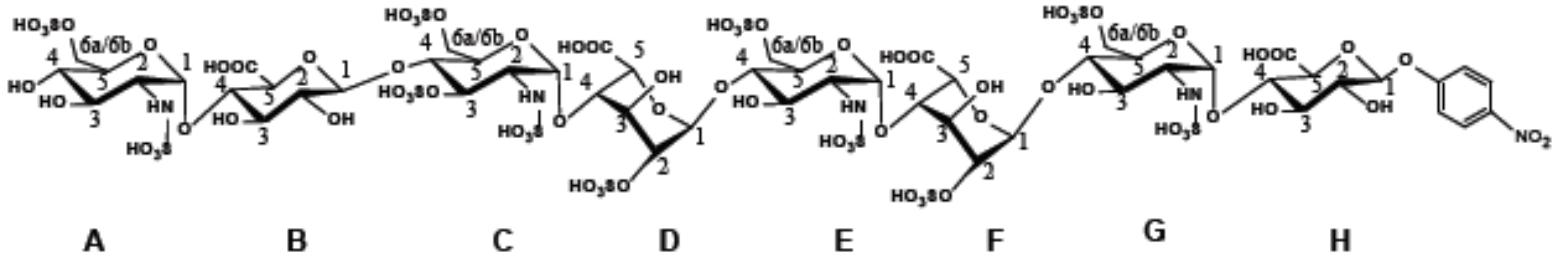
Supplementary Fig 16. **Purity and structural analysis of 2**. (a) DEAE-HPLC chromatogram. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data for **2**, and inset b is simulated theoretical data accurate MS data for **2** at charge 4 in negative ESI mode. Accurate MS and isotope distribution matched very well (within 5 ppm mass accuracy) with the empirical formula ($C_{54}H_{81}N_5O_{76}S_{11}$). (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The chemical structure of **2** is shown on top of (c).



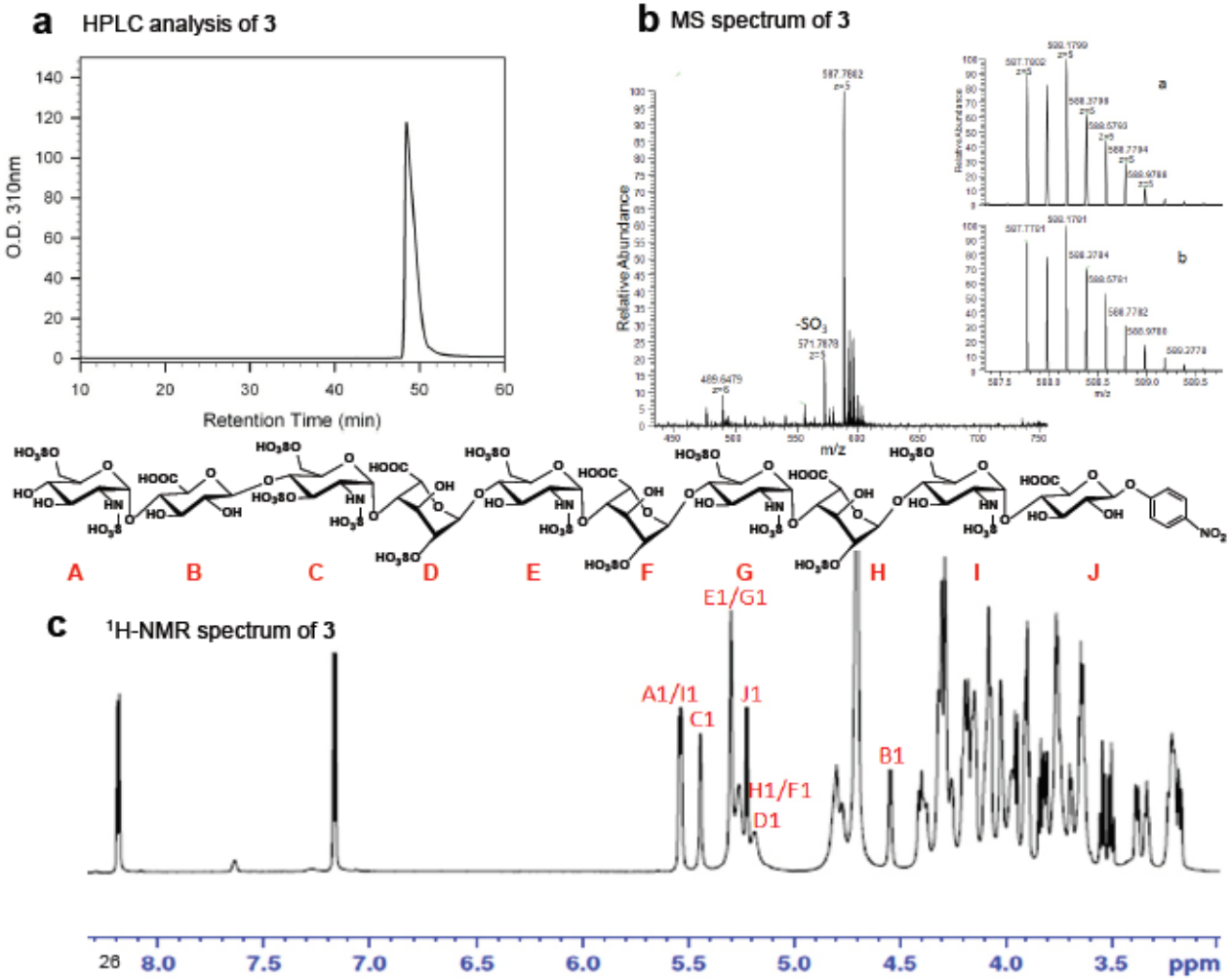
Supplementary Fig 17. **NMR analysis of 2**. (a) 13C-NMR spectrum of **2**. Signals from eight anomeric carbons are clearly identified. (b) 2D 1H-1H COSY spectrum of **2**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.55 (d, J = 3.6 Hz, 1H), 5.53 (d, J = 3.5 Hz, 1H), 5.44 (d, J = 2.8 Hz, 1H), 5.33 (d, J = 2.8 Hz, 1H), 5.23 (d, J = 7.8 Hz, 1H), 5.18, 5.13 (broad singlet, 2H), and 4.55 (d, J = 7.6 Hz, 1H) ppm. (c) 2D 1H-13C HMQC spectrum of **2**

^{1}H NMR/^{13}C NMR chemical shift assignments (in ppm) of **2**

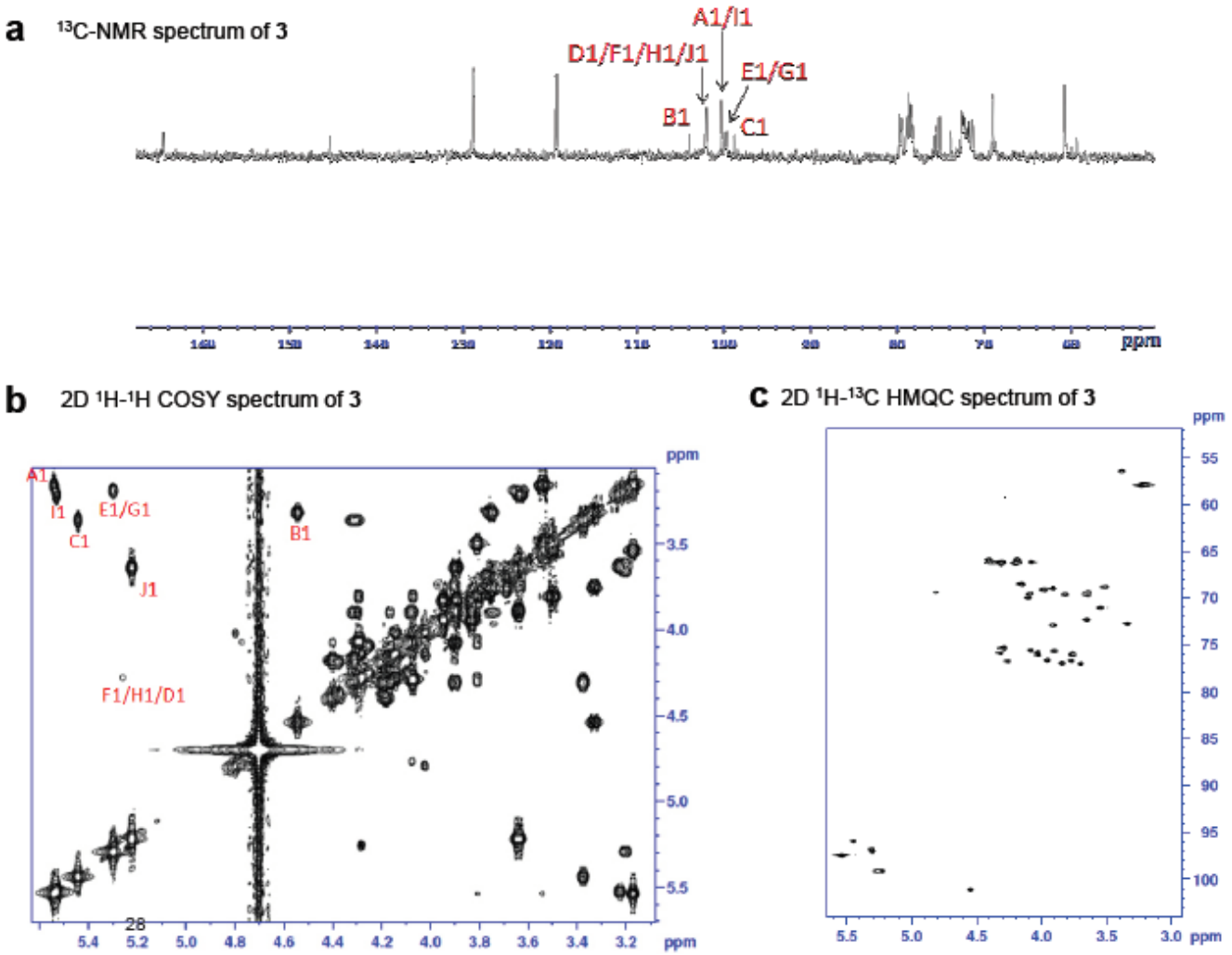
	1	2	3	4	5	6a	6b
A	5.55/97.4	3.18/57.8	3.55/70.9	3.49/68.8	3.81/69.6	4.08/66.0	4.29/66.0
B	4.55/101.1	3.34/72.5	3.76/75.3	3.76/76.6	3.70/76.9		
C	5.44/96.0	3.38/56.4	4.30/75.8	3.89/75.6	4.08/69.5	4.19/65.8	4.41/65.8
D	5.13/99.5	4.25/76.7	4.08/70.1	4.07/75.7	4.74/70.2		
E	5.33/96.6	3.20/57.8	3.59/69.4	3.73/76.0	3.94/69.2	4.19/66.1	4.39/66.1
F	5.18/99.3	4.26/76.6	4.10/70.0	4.07/75.6	4.76/69.4		
G	5.53/97.4	3.23/57.8	3.62/69.5	3.76/76.0	3.90/69.0	4.16/66.1	4.37/66.1
H	5.23/99.1	3.64/72.2	3.89/75.6	3.82/77.2	3.95/76.6		



Supplementary Fig 18. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of 2**. The structure of **2** is shown. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



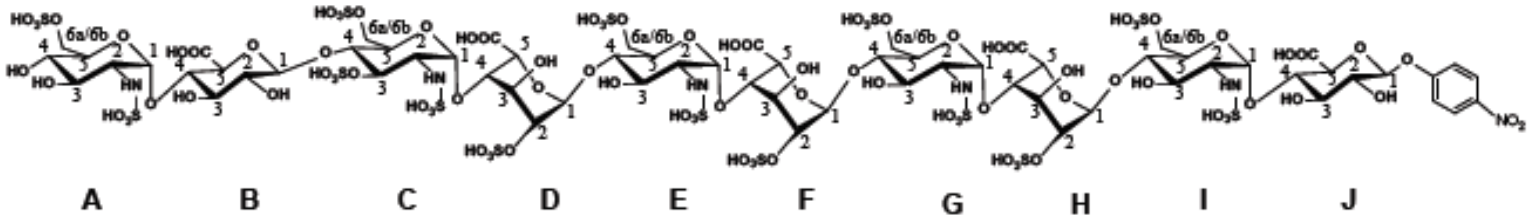
Supplementary Fig 19. **Purity and structural analysis of 3**. (a) DEAE-HPLC chromatogram. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data for **3**, and inset b is simulated theoretical data accurate MS data for **3** at charge 5 in negative ESI mode. Accurate MS and isotope distribution matched very well (within 5 ppm mass accuracy) with the empirical formula ($C_{66}H_{100}N_6O_{95}S_{14}$). (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The chemical structure of **3** is shown on top of (c).



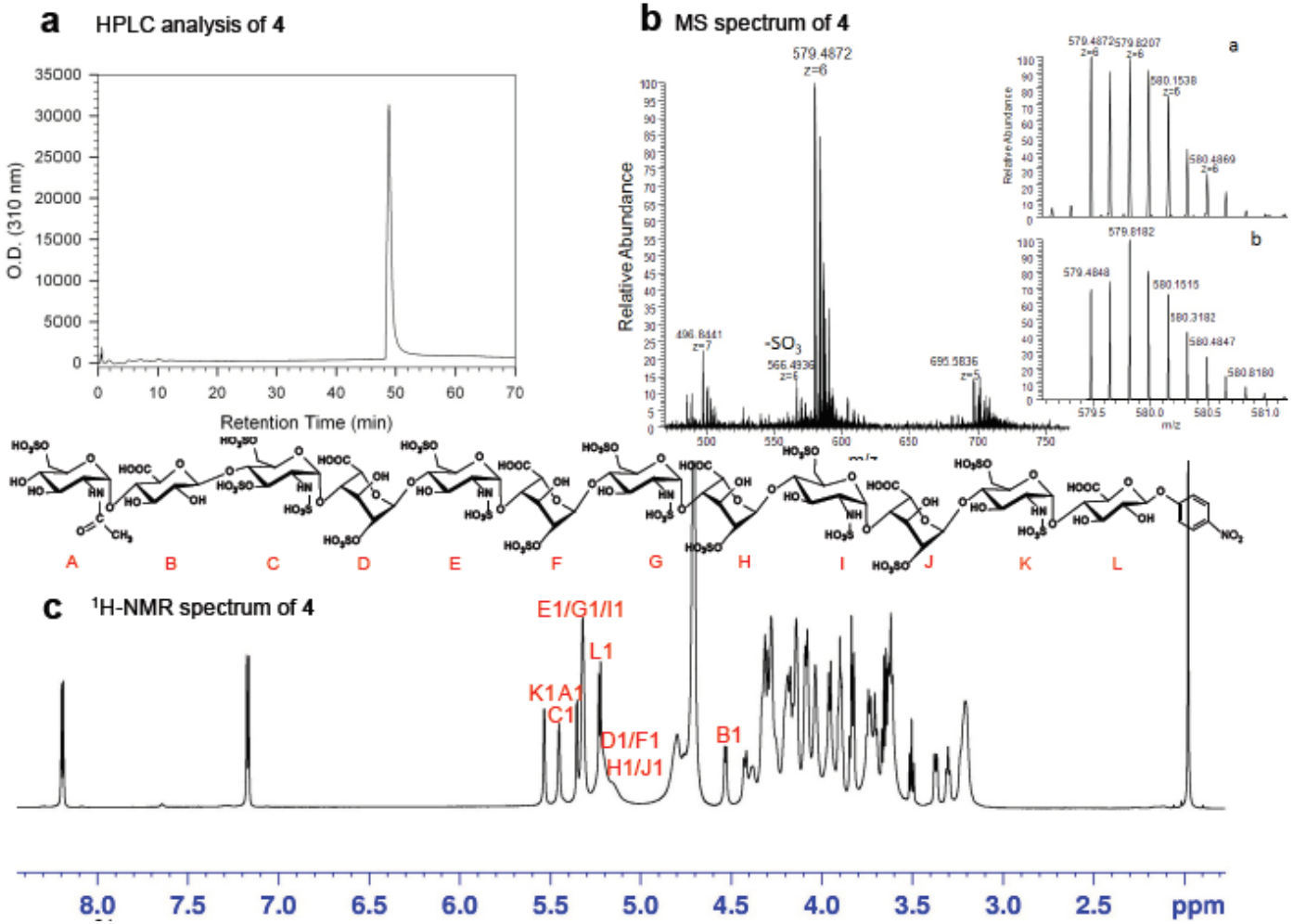
Supplementary Fig 20. **NMR analysis of 3**. (a) 13C-NMR spectrum of **3**. Signals from ten anomeric carbons are identified. (b) 2D 1H-1H COSY spectrum of **3**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.55 (s, 1H), 5.53 (s, 1H), 5.44 (s, 1H), 5.33 (s, 2H), 5.26 (broad singlets, 2H), 5.23 (d, J = 7.8 Hz, 1H), 5.18 (broad singlet, 1H), and 4.55 (d, J = 7.6 Hz, 1H) ppm. (c) 2D 1H-13C HMQC spectrum of **3**.

^{1}H NMR/^{13}C NMR chemical shift assignments (in ppm) of **3**

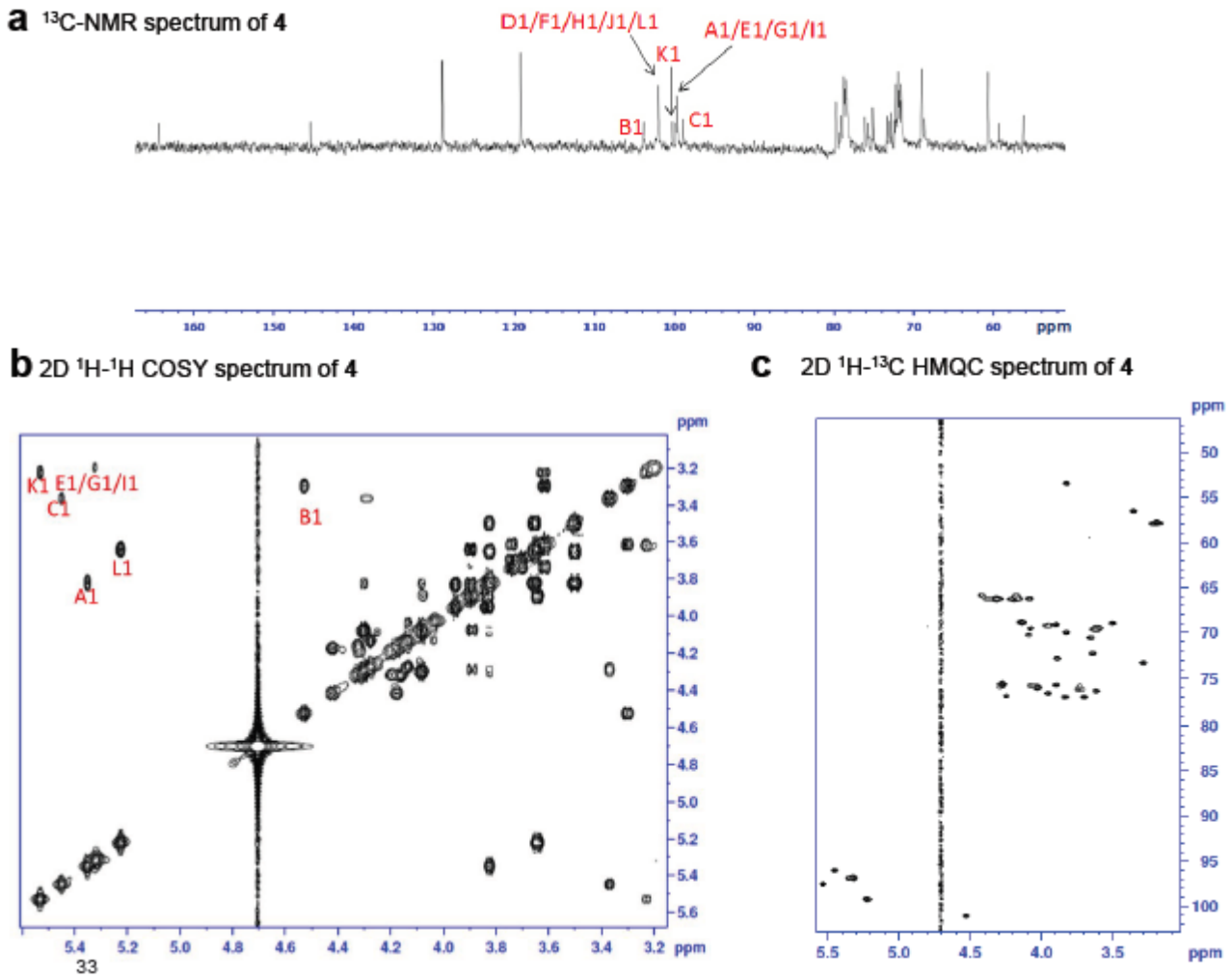
	1	2	3	4	5	6a	6b
A	5.55/97.4	3.18/57.8	3.55/70.9	3.49/68.8	3.81/69.6	4.08/66.0	4.29/66.0
B	4.55/101.1	3.34/72.5	3.76/75.3	3.76/76.6	3.70/76.9		
C	5.44/96.0	3.38/56.4	4.30/75.8	3.89/75.6	4.08/69.5	4.19/65.8	4.41/65.8
D	5.18/99.3	4.25/76.6	4.08/70.0	4.07/75.6	4.77/69.4		
E	5.33/96.6	3.20/57.8	3.59/69.4	3.73/76.0	3.94/69.2	4.19/66.1	4.39/66.1
F	5.26/99.1	4.26/76.7	4.13/68.5	4.03/76.0	4.80/69.4		
G	5.33/96.6	3.20/57.8	3.59/69.4	3.73/76.0	3.94/69.2	4.19/66.1	4.39/66.1
H	5.26/99.1	4.26/76.7	4.13/68.5	4.03/76.0	4.80/69.4		
I	5.53/97.4	3.23/57.8	3.62/69.5	3.76/76.0	3.90/69.0	4.16/66.1	4.37/66.1
J	5.23/99.1	3.64/72.2	3.89/75.6	3.82/77.2	3.95/76.6		



Supplementary Fig 21. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of 3**. The structure of **3** is shown. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



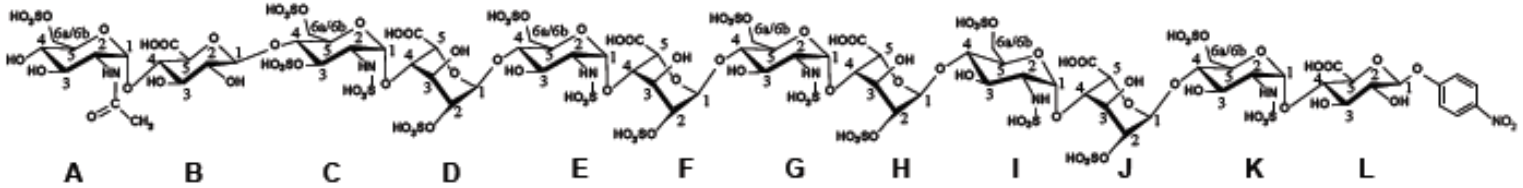
Supplementary Fig 22. **Purity and structural analysis of 4**. (a) DEAE-HPLC chromatogram. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data for **4**, and inset b is simulated theoretical data accurate MS data for **4** at charge 6 in negative ESI mode. Accurate MS and isotope distribution matched very well (within 5 ppm mass accuracy) with the empirical formula ($C_{80}H_{121}N_7O_{112}S_{16}$). (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The chemical structure of **4** is shown on top of (c).



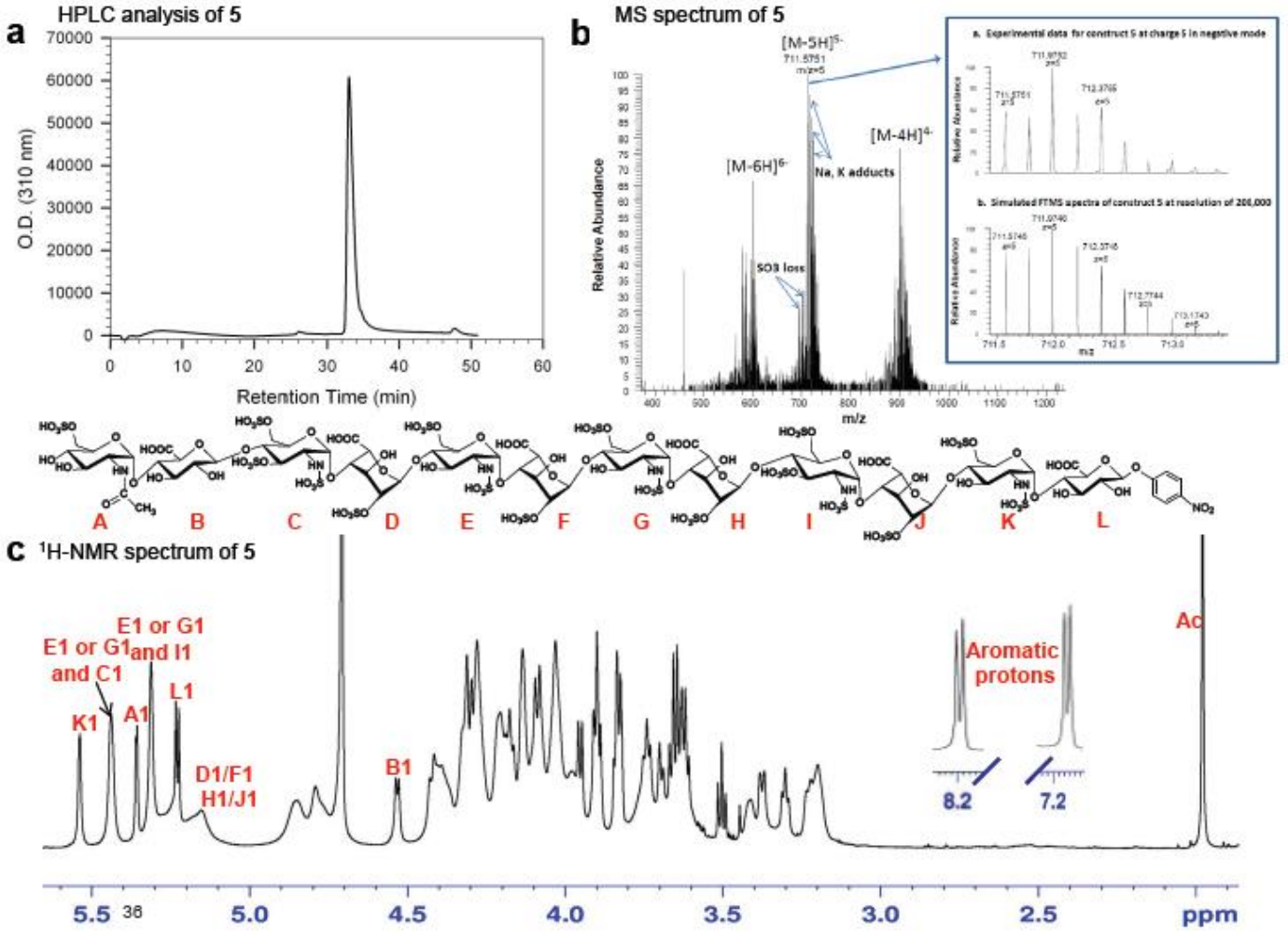
Supplementary Fig 23. **NMR analysis of 4**. (a) 13C-NMR spectrum of **4**. Signals from twelve anomeric carbons are identified. (b) 2D 1H-1H COSY spectrum of **4**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.53 (s, 1H), 5.45 (s, 1H), 5.35 (d, J = 3.0 Hz, 1H), 5.31 (s, 3H), 5.26-5.10 (broad singlets, 4H), 5.23 (d, J = 7.8 Hz, 1H) and 4.53 (d, J = 7.6 Hz, 1H) ppm. (c) 2D 1H-13C HMQC spectrum of **4**.

^{1}H NMR/^{13}C NMR chemical shift assignments (in ppm) of **4**

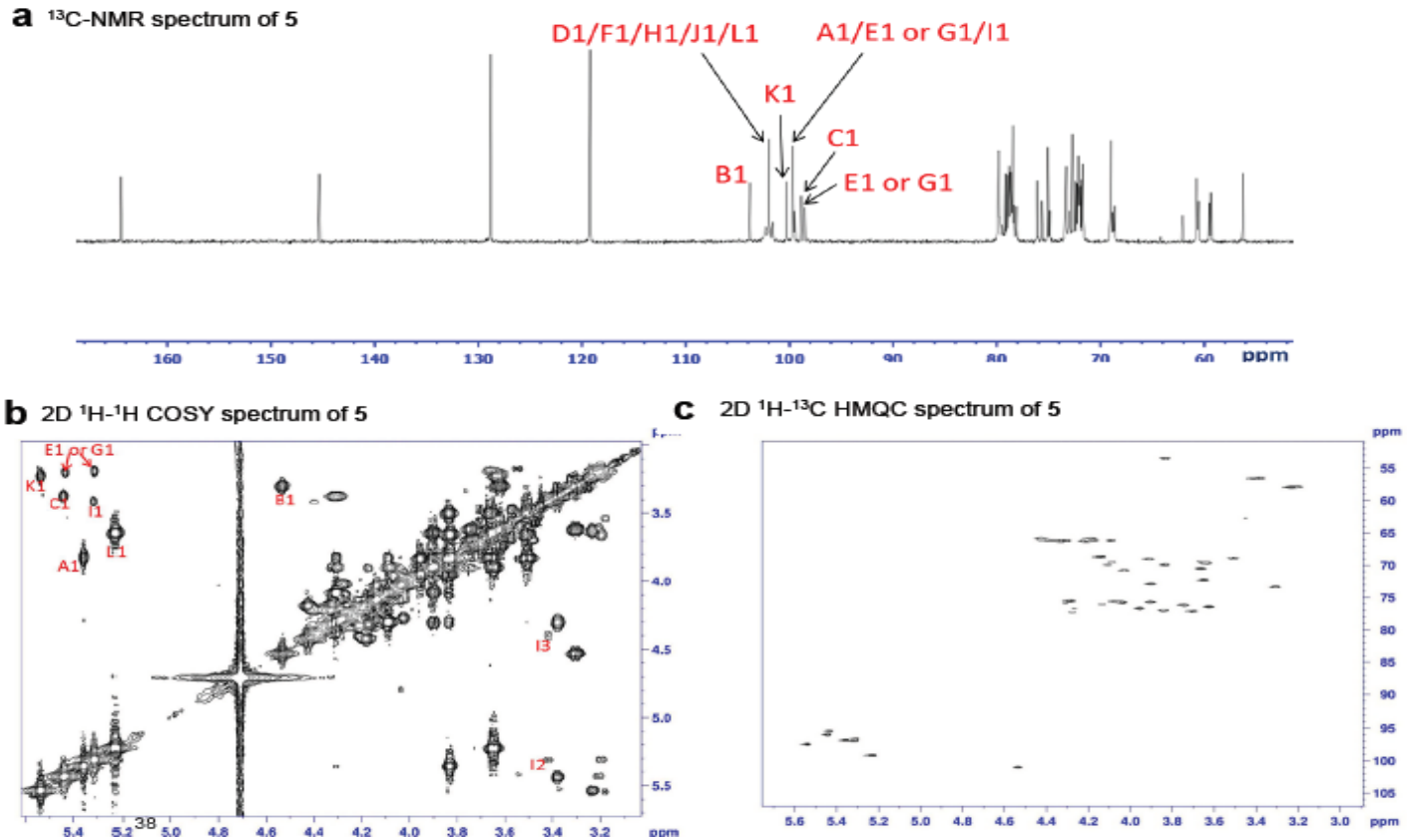
	1	2	3	4	5	6a	6b	-$COCH_3$
A	5.35/96.9	3.82/53.4	3.64/72.1	3.50/68.9	3.82/69.9	4.08/66.1	4.30/66.1	2.03
B	4.53/101.0	3.30/73.3	3.62/76.3	3.73/76.2	3.70/77.0			
C	5.45/96.0	3.36/56.4	4.29/75.8	3.89/75.6	4.08/69.5	4.19/65.8	4.41/65.8	
D	5.13/99.5	4.25/76.7	4.08/70.1	4.07/75.7	4.74/70.2			
E	5.31/96.6	3.20/57.8	3.59/69.4	3.73/76.0	3.94/69.2	4.19/66.1	4.39/66.1	
F	5.20/99.1	4.26/75.5	4.13/68.7	4.03/75.9	4.80/69.4			
G	5.31/96.6	3.20/57.8	3.59/69.4	3.73/76.0	3.94/69.2	4.19/66.1	4.39/66.1	
H	5.20/99.1	4.26/75.5	4.13/68.7	4.03/75.9	4.80/69.4			
I	5.31/96.6	3.20/57.8	3.59/69.4	3.73/76.0	3.94/69.2	4.19/66.1	4.39/66.1	
J	5.20/99.1	4.26/75.5	4.13/68.7	4.03/75.9	4.80/69.4			
K	5.53/97.4	3.23/57.8	3.62/69.5	3.74/76.0	3.90/69.0	4.16/66.3	4.31/66.3	
L	5.23/99.1	3.64/72.2	3.89/75.6	3.82/77.2	3.95/76.6			



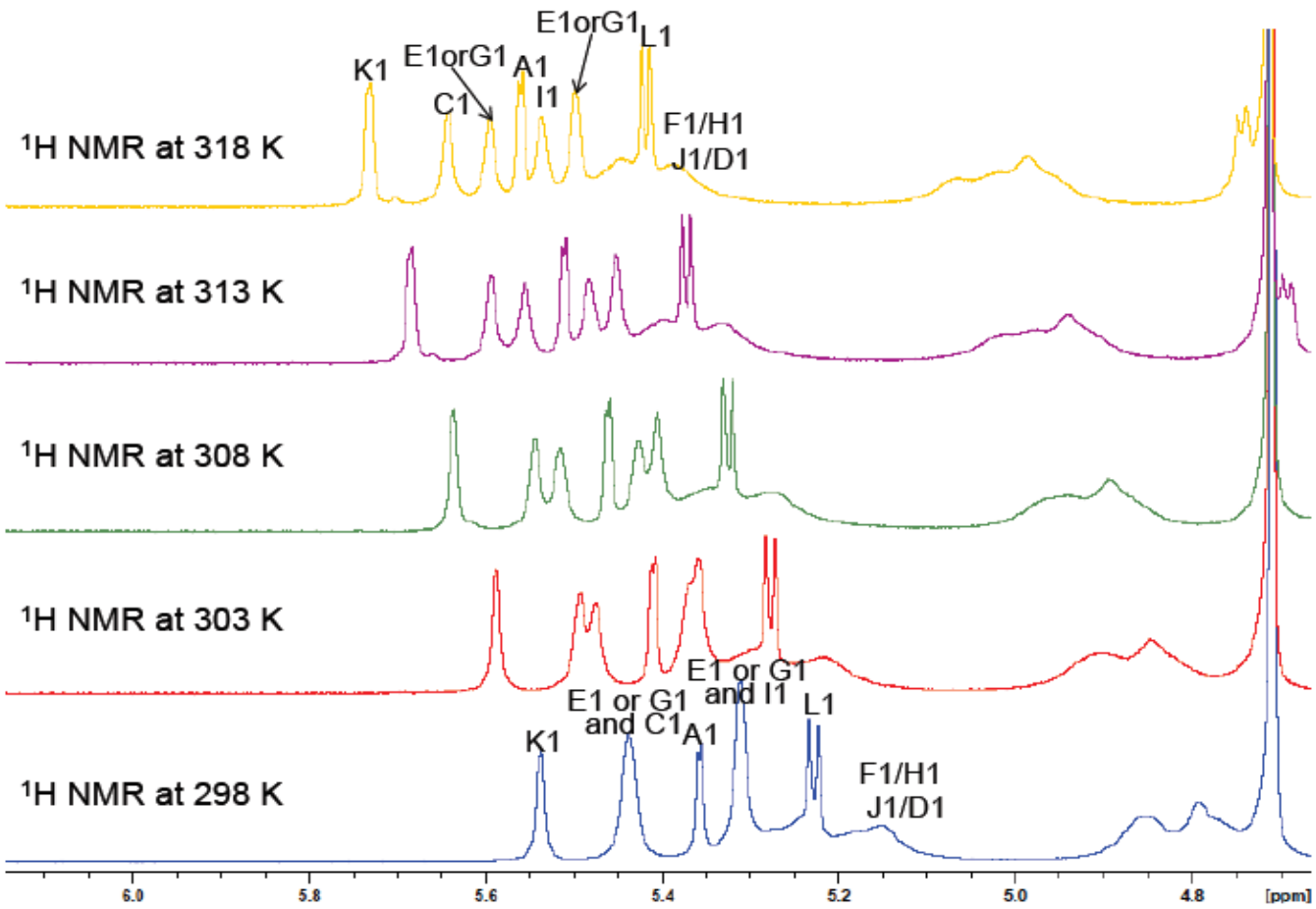
Supplementary Fig 24. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of 4.** The structure of **4** is shown. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



Supplementary Fig 25. Purity and structural analysis of 5. (a) DEAE-HPLC chromatogram. (b) High resolution MS spectrum. Monoisotopic peak mass and charge ratio (m/z) values are shown for different charge states. An isotopic distribution analysis was performed as shown in insets. Inset a is experimental data for **5**, and inset b is the simulated isotopic distribution for **5** at charge -5. Accurate MS and isotope distribution matched very well (within 5 ppm mass accuracy) with the empirical formula ($C_{80}H_{116}N_7O_{115}S_{17}$). (c) 1H-NMR spectrum. The signals of anomeric protons are indicated. The chemical structure of **5** is shown on top of (c).



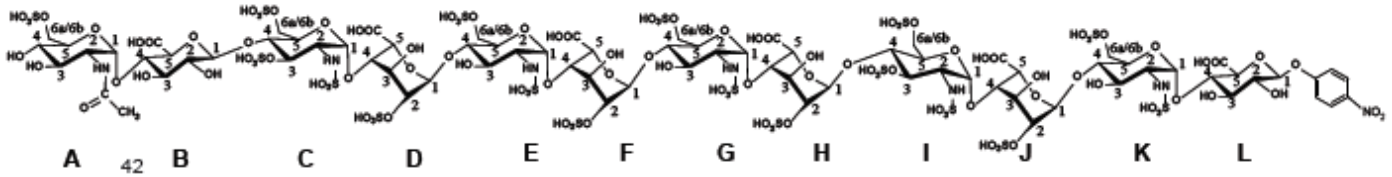
Supplementary Fig 26. NMR analysis of 5. (a) 13C-NMR spectrum of **5**. Signals from twelve anomeric carbons are identified. (b) 2D 1H-1H COSY spectrum of **5**. The corresponding peak assignments of the anomeric protons resonate as doublets or singlets at δ 5.53 (s, 1H), 5.44 (s, 2H), 5.36 (d, J = 3.1 Hz, 1H), 5.31 (s, 2H), 5.26-5.10 (broad singlets, 4H), 5.23 (d, J = 7.8 Hz, 1H) and 4.54 (d, J = 7.4 Hz, 1H) ppm. The small coupling constants (< 4 Hz) of the anomeric protons indicate a α linkage, and the large coupling constants (~8 Hz) indicate a β linkage. Through the 2D 1H-1H COSY, we can identify that one hydroxyl group at the 3-position on GlcNS of E or G or I was sulfated. One of the anomeric protons at 5.31 ppm correlates to the proton at one of the 2-postions of E or G or I resonated at 3.42 ppm (the other two signals for the 2-positions were at 3.20 ppm). The proton at 3-position coupled to the proton at 3.42 ppm resonated at 4.39 ppm (instead of the 3.59 ppm seen for the other two 3-position protons) consistent of its substitution with a 3-*O*-sulfate group. (c) 2D 1H-13C HMQC spectrum of **5**. Two questions remain. The first question is why the residue with the new 3-*O*-sulfate group (inserted using 3-*O*ST-5) has an anomeric proton resonating upfield from the other 3-*O*-sulfate group (inserted using 3-*O*ST-1) and with a GlcNS6S residue not having a 3-*O*-sulfate group (this will be addressed in **Supplementary Fig 27**). The second question is which residue (E, G or I) contains the 3-*O*-sulfate group (this will be addressed in **Supplementary Figs 29-30**).



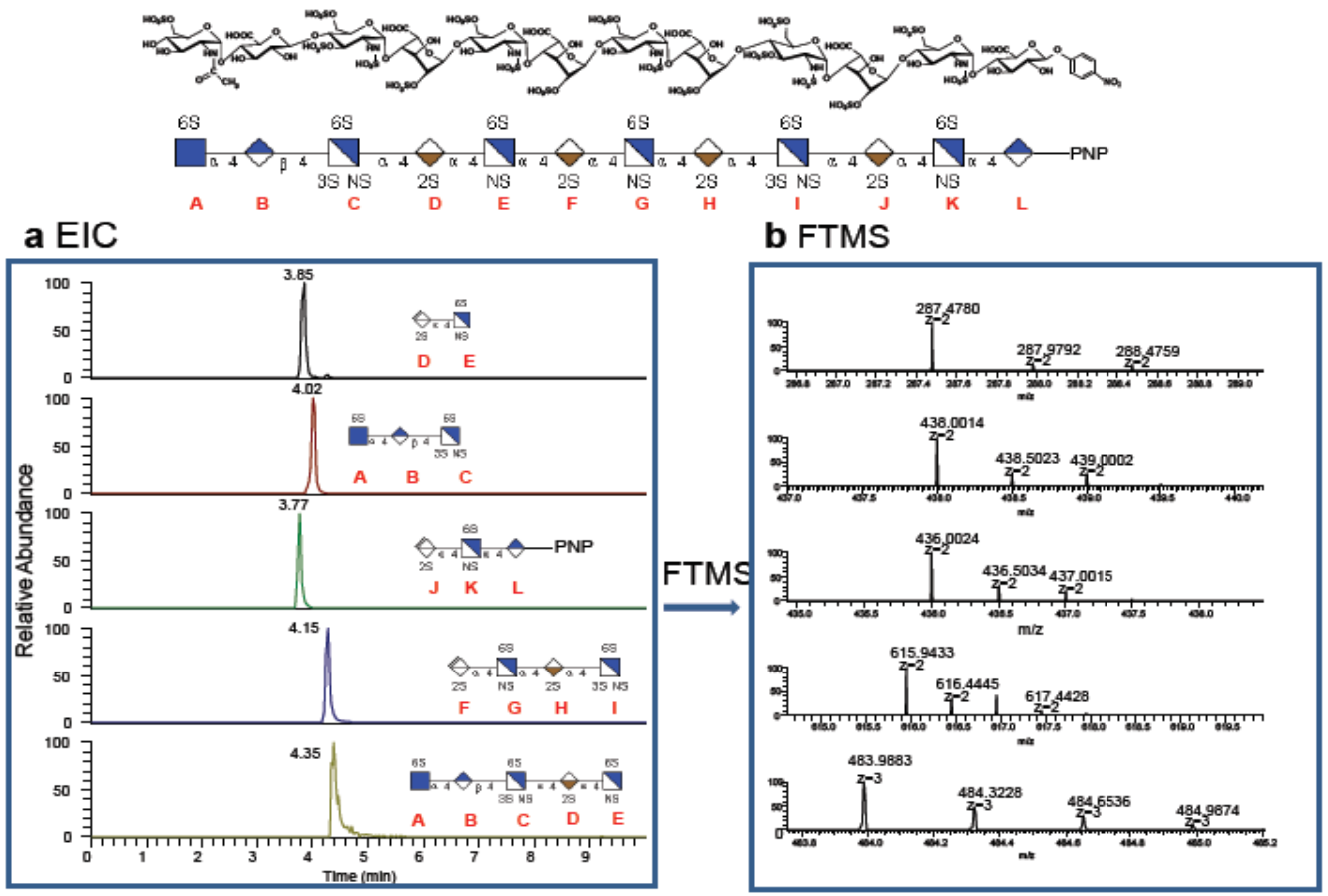
Supplementary Fig 27. 1H-NMR spectra of 5 under different temperatures (298 K to 318 K). The GlcNS3S6S residues (C and I) should show similar shift values for their anomeric protons (as should the GlcNS6S residues of E and G), because of the similar chemical environments for these residues. Since they do not, we hypothesize that the observed differences in shift are due to conformational effects. To test this hypothesis and to resolve these signals, 1D 1H NMR was conducted at temperatures ranging from 298K to 318K. With increasing temperature, the two anomeric resonances for the GlcNS3S6S residues shifted downfield relative to the two GlcNS6S anomeric resonances. In addition, the two anomeric resonances for the GlcNS3S6S residues moved closer to each other as did the two anomeric resonances for the GlcNS6S residues. Finally, all four of these anomeric protons were resolved. This is consistent with the unusual chemical shifts for these anomeric protons being caused by a conformational effect that can be overcome on heating. The assignment of which residue (E,G, or I) contains the 3-*O*-sulfate group is addressed in **Supplementary Figs 29-30**.

^{1}H NMR/^{13}C NMR chemical shift assignments (in ppm) of **5**

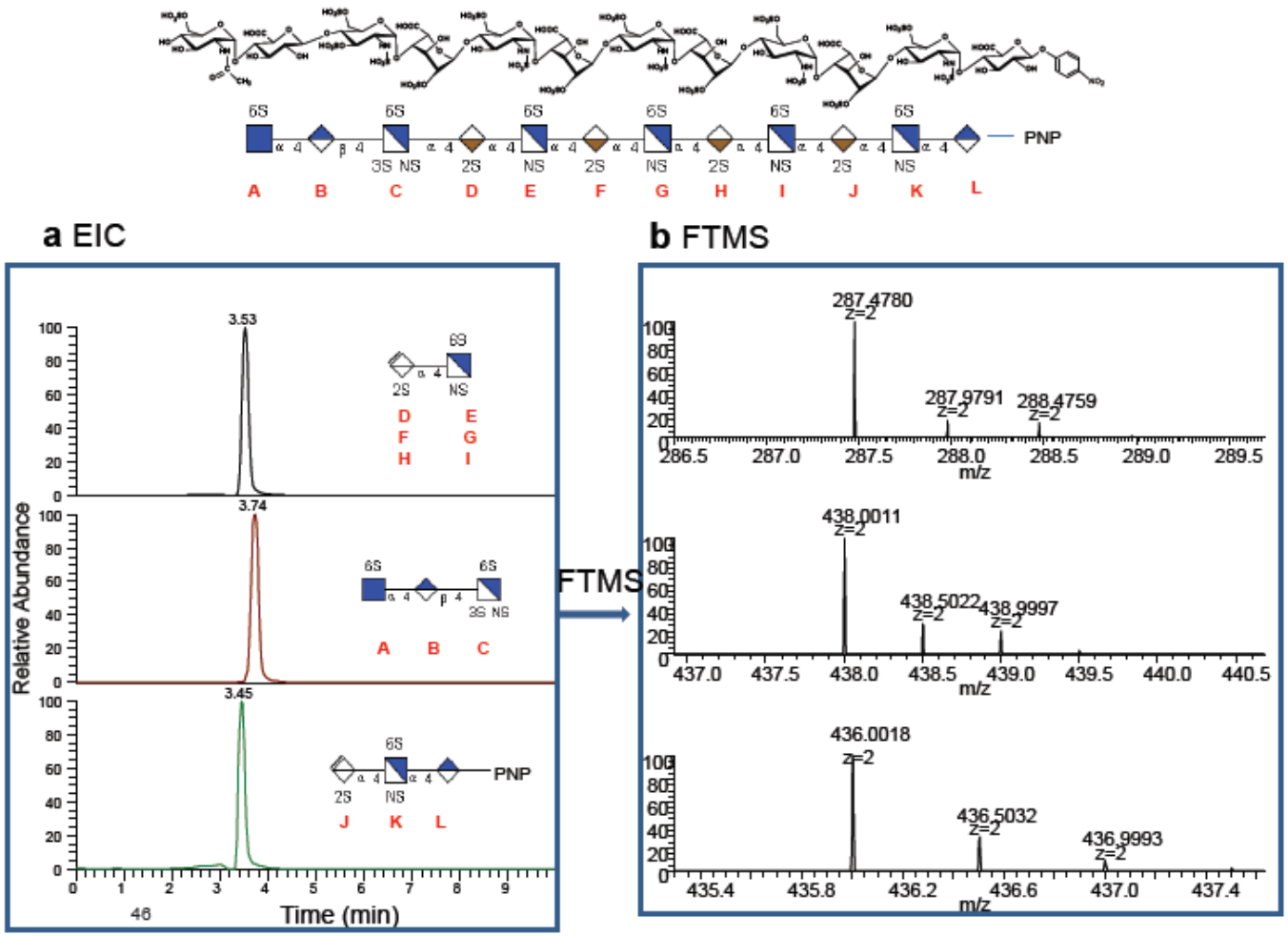
	1	2	3	4	5	6a	6b	-$COCH_3$
A	5.36/96.9	3.82/53.4	3.64/72.1	3.50/68.9	3.82/69.9	4.08/66.1	4.30/66.1	2.03
B	4.54/101.0	3.30/73.3	3.62/76.3	3.73/76.2	3.70/77.0			
C	5.44/96.0	3.36/56.4	4.29/75.8	3.89/75.6	4.08/69.5	4.19/65.8	4.41/65.8	
D	5.20/99.1	4.25/76.7	4.08/70.1	4.07/75.7	4.74/70.2			
E	5.31 or 5.44 /95.1 or 96.9	3.20/57.8	3.59/69.6	3.73/76.3	3.94/69.0	4.19/66.0	4.39/66.0	
F	5.20/99.2	4.26/75.5	4.13/68.7	4.03/75.7	4.80/69.4			
G	5.31 or 5.44 /95.1 or 96.9	3.20/57.8	3.59/69.6	3.73/76.3	3.94/69.0	4.19/66.0	4.39/66.0	
H	5.20/99.2	4.26/75.5	4.13/68.7	4.03/75.7	4.80/69.4			
I	5.31/96.9	3.42/56.6	4.39/75.4	4.09/75.5	3.94/69.0	4.19/66.0	4.39/66.0	
J	5.20/99.2	4.26/75.5	4.13/68.7	4.03/75.7	4.80/69.4			
K	5.53/97.4	3.23/57.8	3.62/69.5	3.74/76.0	3.90/69.0	4.16/66.3	4.31/66.3	
L	5.23/99.1	3.64/72.2	3.89/75.6	3.83/77.2	3.95/76.6			



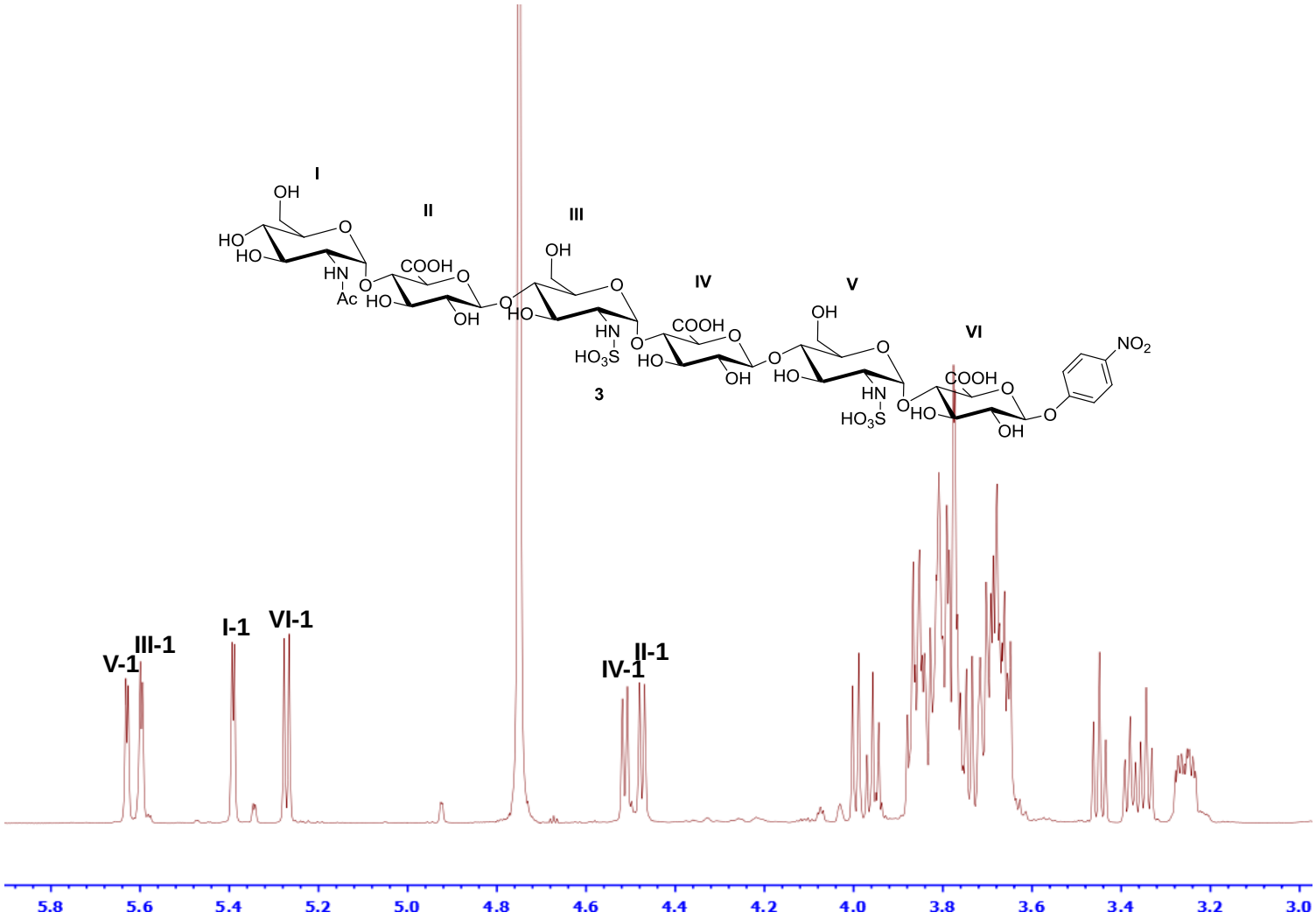
Supplementary Fig 28. **1H-NMR and 13C-NMR chemical shift assignments (in ppm) of 5.** The structure of **5** is shown. The proton assignments for Residue I are colored in red to demonstrate that an extra 3-*O*-sulfate glucosamine residue is present in **5**. The location of the additional 3-*O*-sulfate glucosamine residue was determined by a MS-assisted oligosaccharide sequencing technique as described in **Supplementary Fig 29**. The protons and carbons of pNP group are clearly identifiable because they are outside the ranges of the chemical shifts of carbohydrate protons and carbons. The protons (doublet 7.28 ppm and doublet 8.28 ppm) and carbons (116.4 ppm and 126.1 ppm) of pNP are not listed in the table.



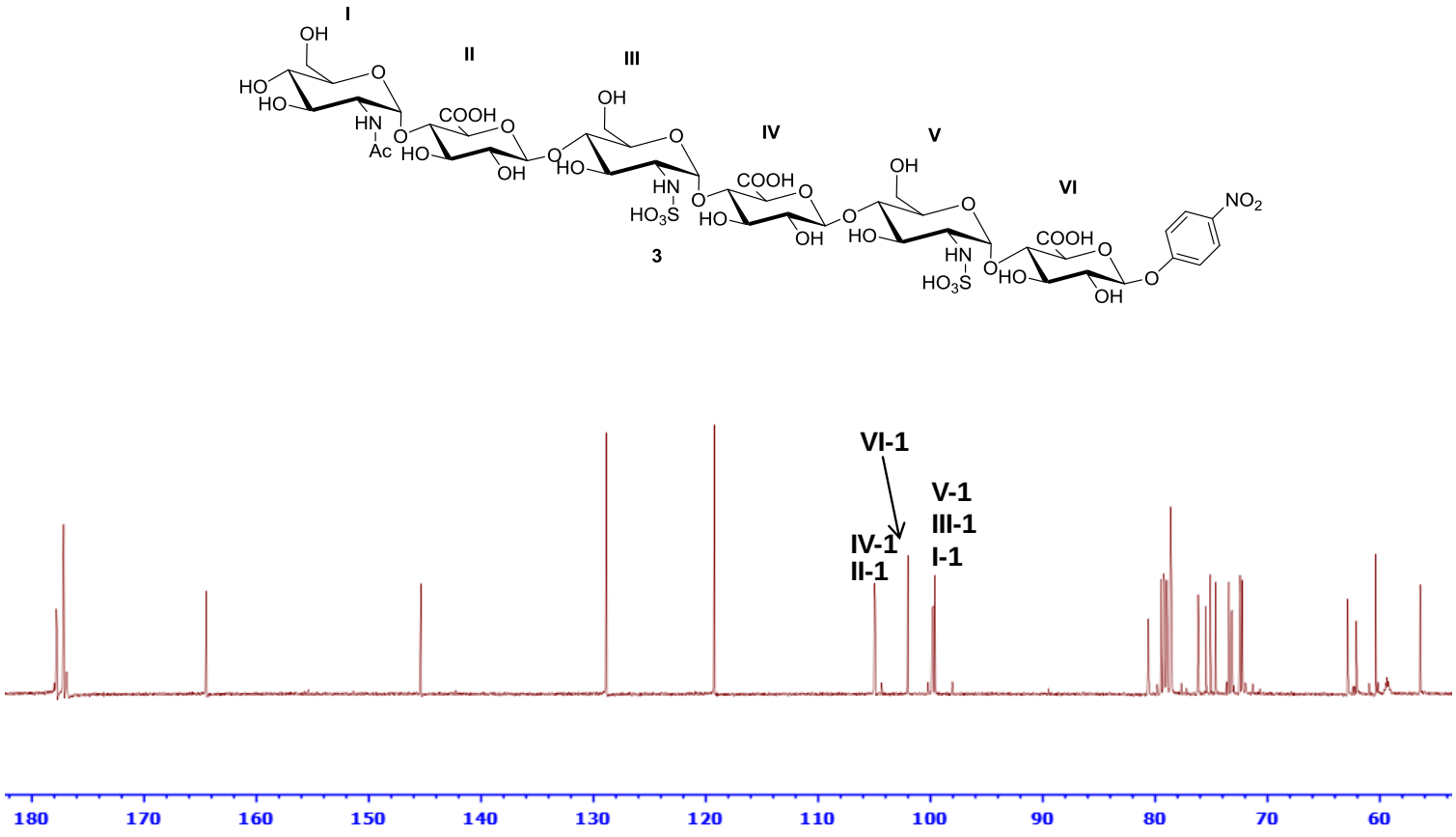
Supplementary Fig 29. MS-assisted sequencing analysis of 5. Due to the complex structure of **5**, NMR and MS analyses were only able to locate one 3-*O*-sulfate group within Residue C. Therefore, MS-assisted sequence analysis was carried out 6. In this experiment, **5** was partially degraded with heparin lyase II. (a) Signals of the degraded products. (b) MS spectra of each signal, proving the structure of the degraded products. MS and LC/MS analyses were then performed to determine the structures of the resulting products, from which the structure of **5** could be definitively established. The signals of products, ranging from disaccharide to pentasaccharide, were observed and their structures were established by MS. The presence of a 3-*O*-sulfate group at Residue I, prevented the heparin lyase II cleavage of the linkage between Residue G and H 7. Consequently, a tetrasaccharide product, representing FGHI, was detected, clearly pinpointing the location of the second 3-*O*-sulfate group at Residue I. The MS-assisted sequencing analysis was also conducted on **4** (missing the 3-*O*-sulfate group on Residue I) and the tetrasaccharide fragment (FGHI) was absent in the products (**Supplementary Fig 30)**.



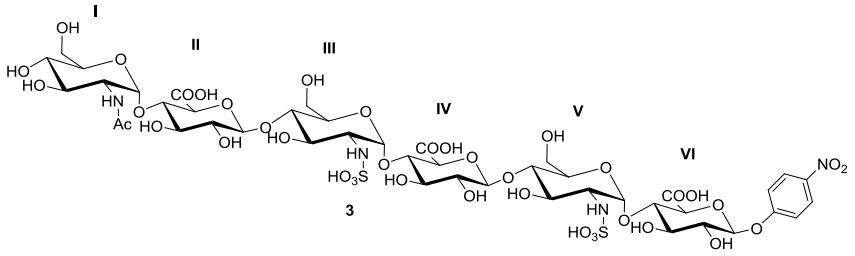
Supplementary Fig 30. MS-assisted sequencing analysis of 4. **4** was partially degraded with heparin lyase II. MS and LC/MS analyses were then performed to determine the structures of the degraded product. Knowing the structures of the degraded products, one can piece together the structure of **4**. (a) Signals of the degraded products. (b) MS spectra of each signal, proving the structure of the degraded products. A representative shorthanded representation and the structure of **4** are shown above (a) and (b). Comparing with the digestion pattern for **5** (shown in **Supplementary Fig 29**), more complete digestion by heparin lyase II was observed. Namely, no tetrasaccharide product, representing Residue FGHI, was observed.

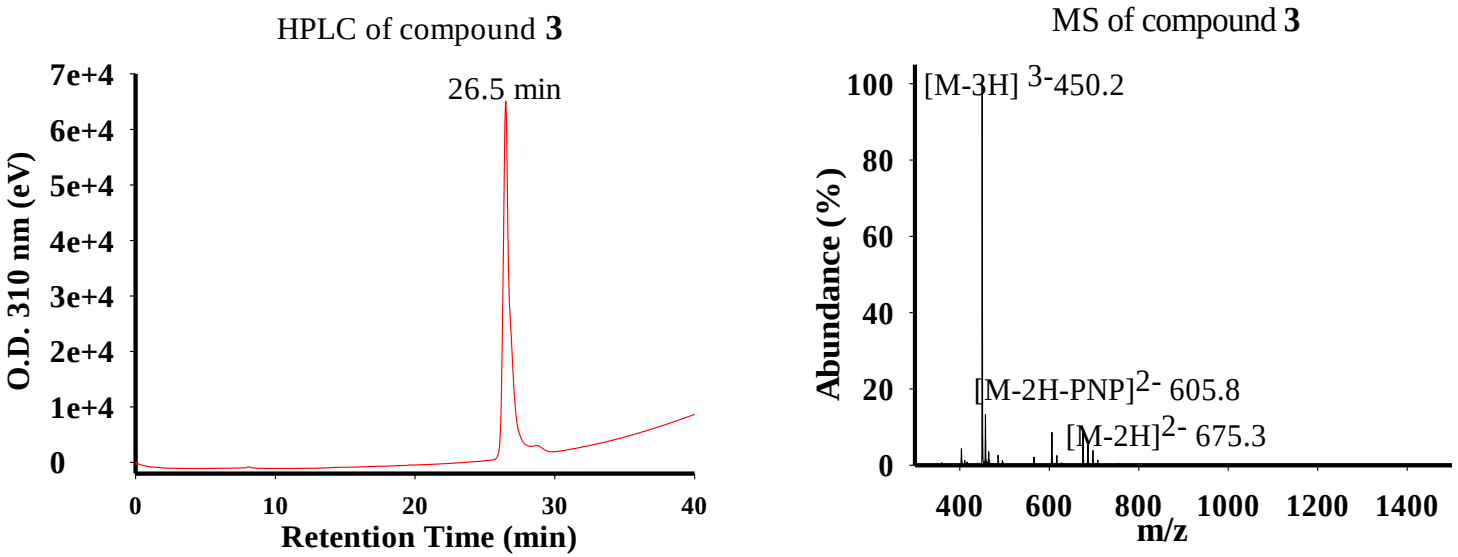


Supplementary Fig 31. ^{1}H-NMR of compound 3 (700 MHz, D_2O). The signals of anomeric protons are indicated. The anomeric protons resonate as doublet at δ 5.63, 5.60, 5.39, 5.27, 4.51, and 4.47 ppm. Chemical structure of compound **3** is shown on top of figure.

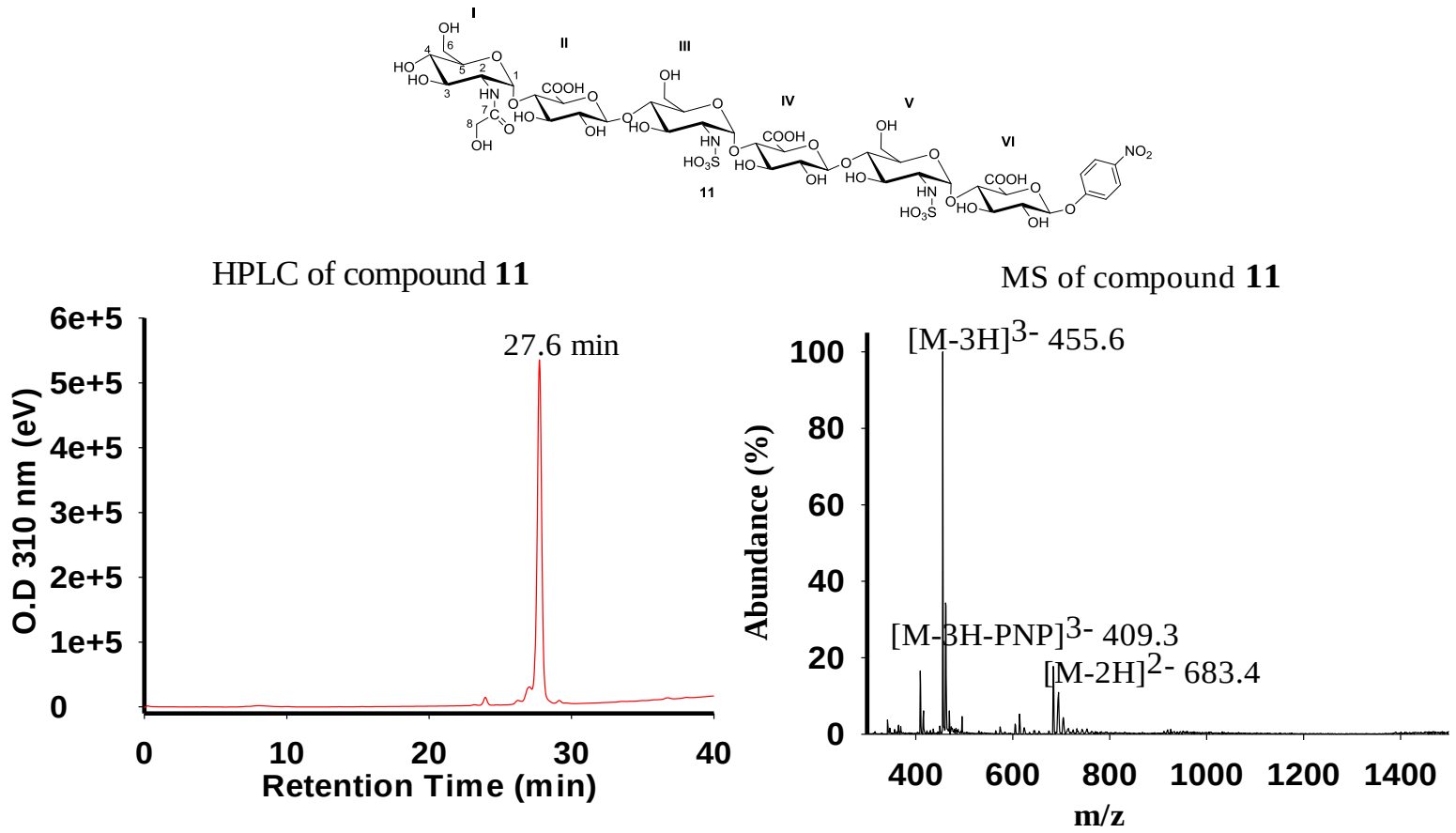


Supplementary Fig 32. ^{13}C-NMR of compound 3 (175 MHz, D_2O). The signals of anomeric carbons are indicated. The anomeric carbons resonate at δ 105.0, 104.9, 102.0, 99.9, 99.8, and 99.6 ppm. Chemical structure of compound **3** is shown on top of figure.

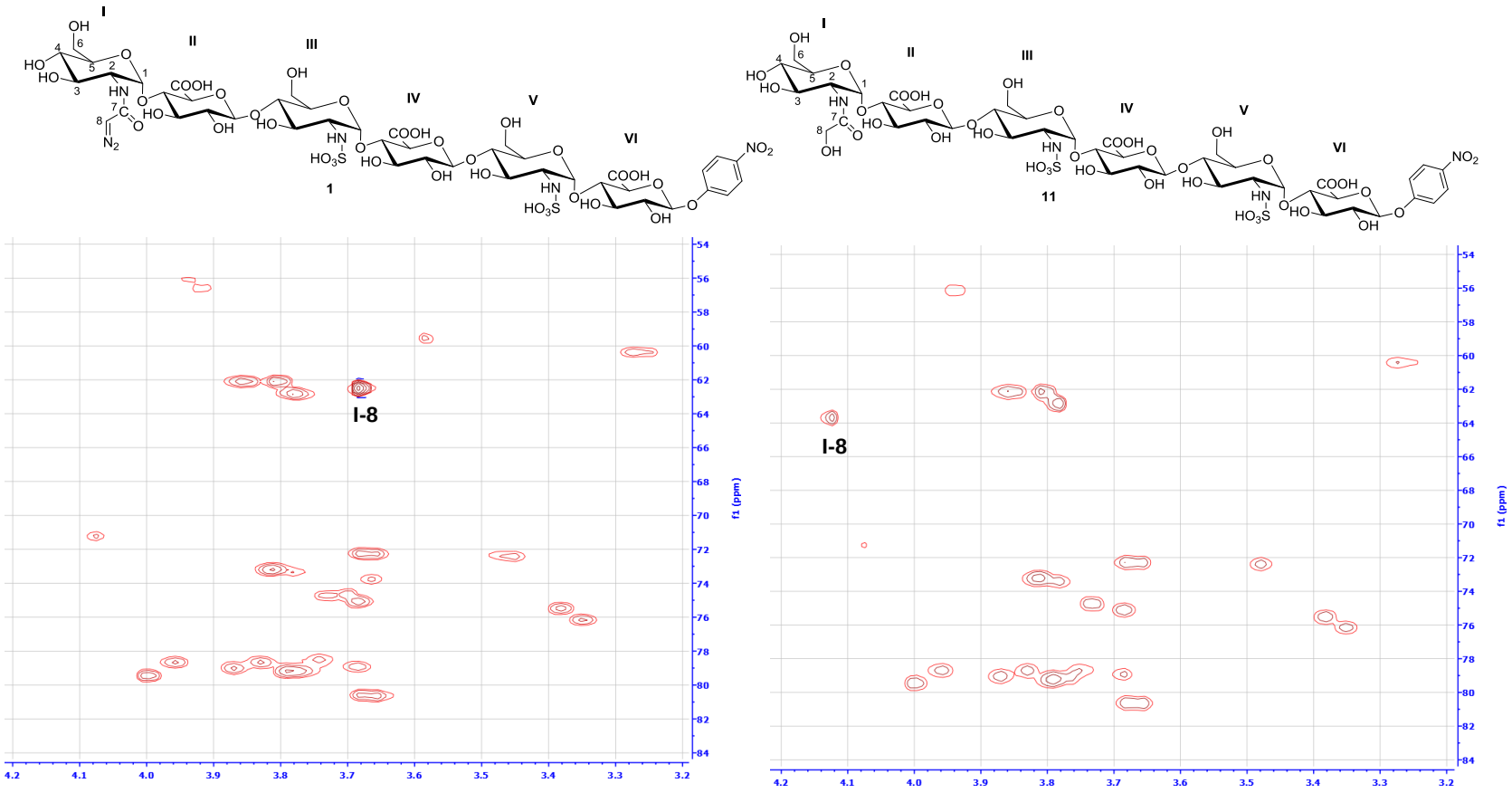




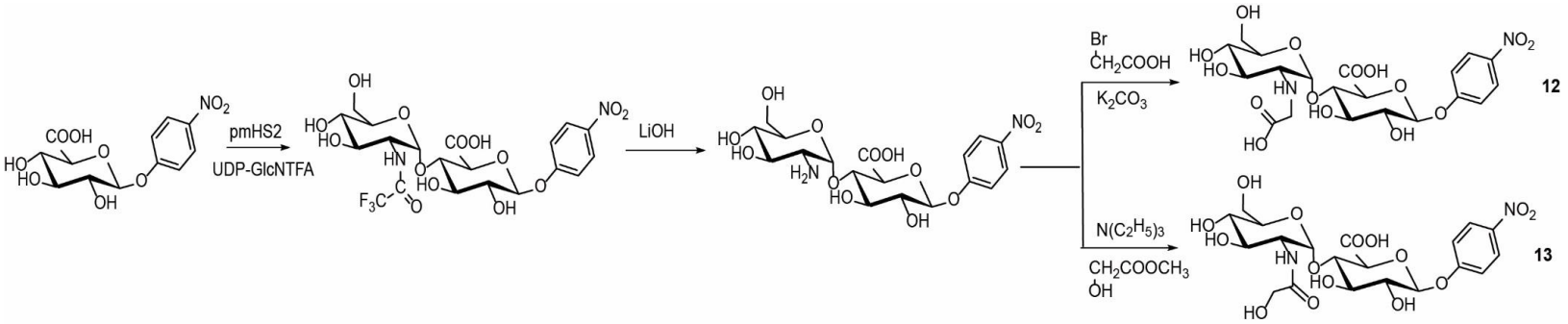
Supplementary Fig 33. Purity and MS analysis of compound 3. Left panel shows the PAMN-HPLC chromatogram of compound **3**. The compound was eluted as a single symmetric peak on PAMN-HPLC, suggesting that compound **3** was pure. Right panel shows the MS spectrum. The measured molecular mass was 1353.6, which is very close to the calculated molecular weight (1353.1). [M-2H-PNP]$^{2-}$ refers to the molecular ion that lost PNP (*p*-nitrophenyl) moiety.



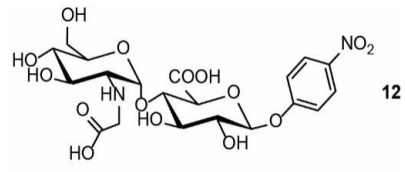
Supplementary Fig 34. Purity and structural analysis of compound 11. Left panel shows the PMAN-HPLC chromatogram of compound **11**. The compound was eluted as a single symmetric peak on PMAN-HPLC, suggesting that **11** was pure. Right panel shows the MS spectrum. The measured molecular mass was 1369.3, which is very close to the calculated molecular weight (1369.1). $[M-3H-PNP]^{3-}$ refers to the molecular ion of compound **11** that lost PNP moiety.

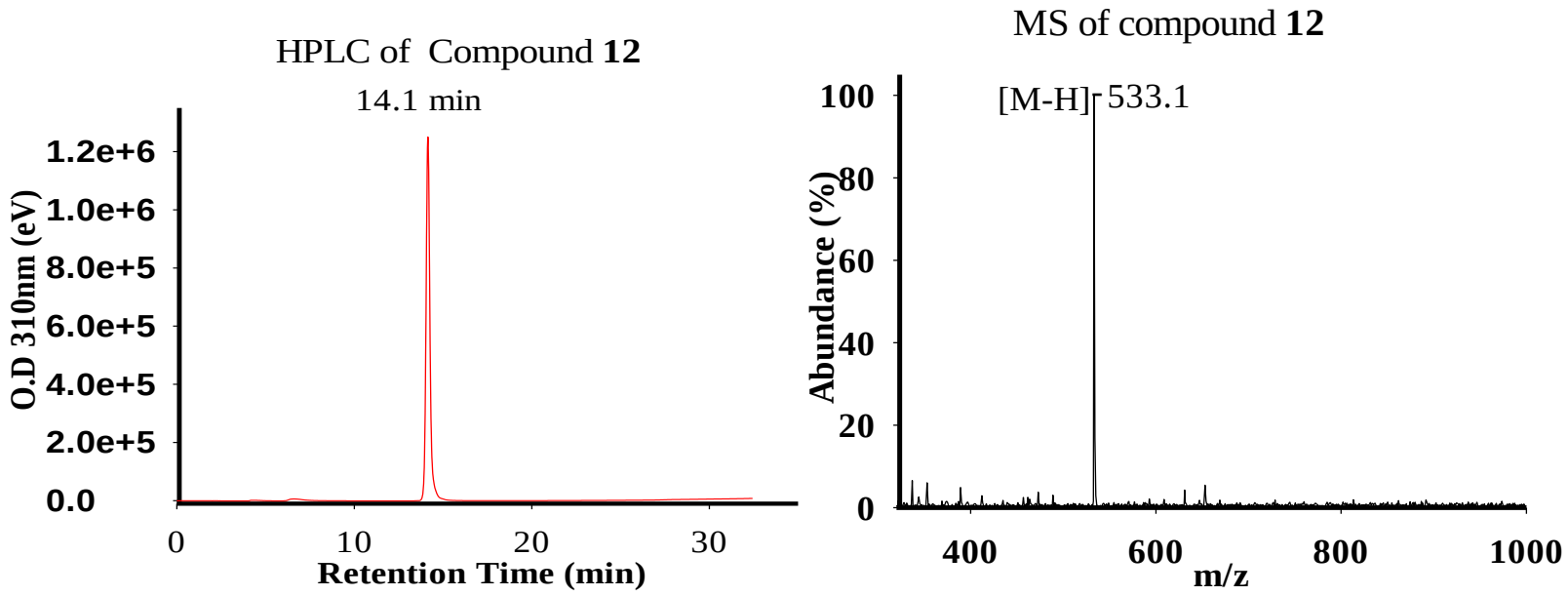


Supplementary Fig 35. ^{1}H-^{13}C HSQC NMR spectra of 1 and 11. Compound **11** is the product after **1** being exposed to water under an acidic condition. Spectra were recorded in D_2O containing 10 mM phosphate buffer at pD7.4 and 25° C. Under this condition, signals of compound **1** were detectable before the diazoacetyl group underwent decomposition reactions. The distinct signals are indicated in spectra.

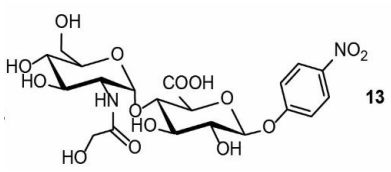


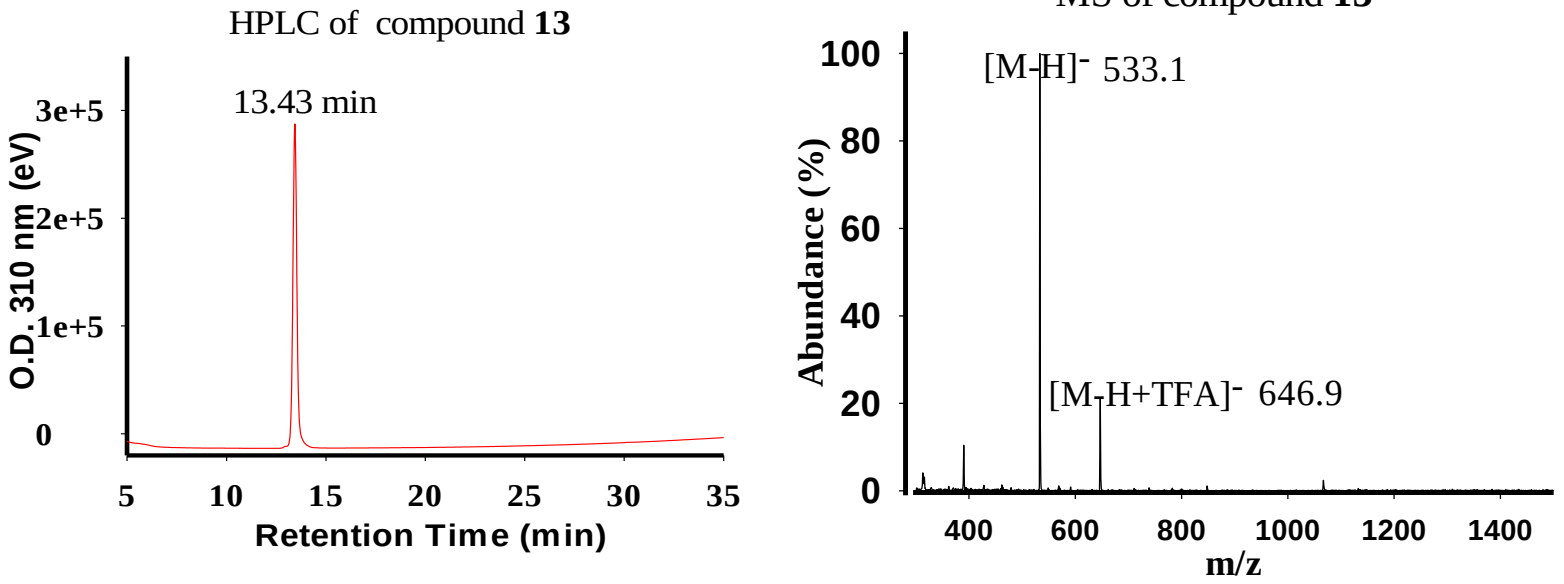
Supplementary Fig 36. Synthetic scheme for compound 12 and 13



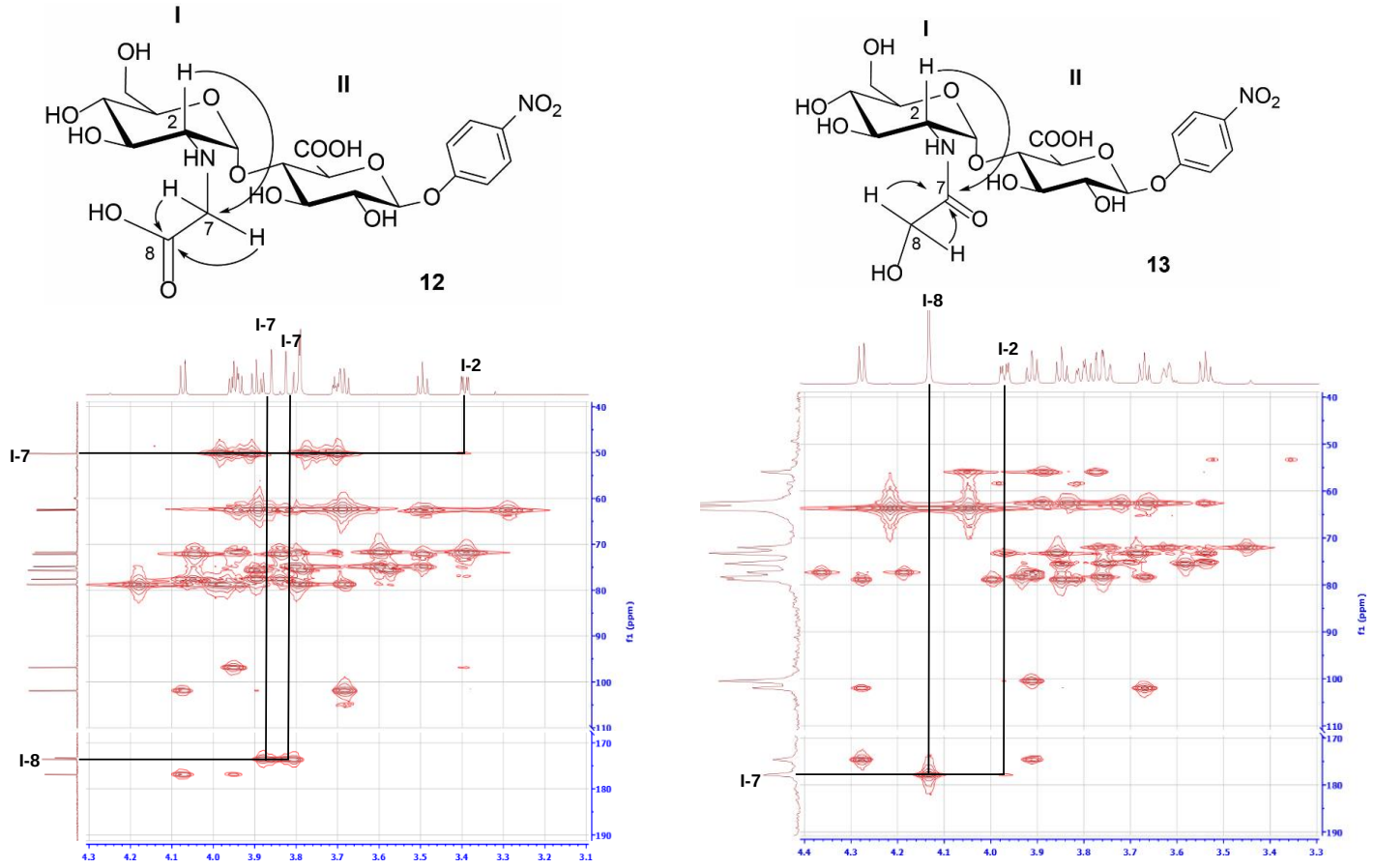


Supplementary Fig 37. Purity and MS analysis of compound 12. Left panel shows the PMAN-HPLC chromatogram of compound **12**. The compound was eluted as a single symmetric peak on PMAN-HPLC, suggesting that compound **12** was pure. Right panel shows the MS spectrum. The measured molecular mass was 534.1, which is very close to the calculated molecular weight (534.4).

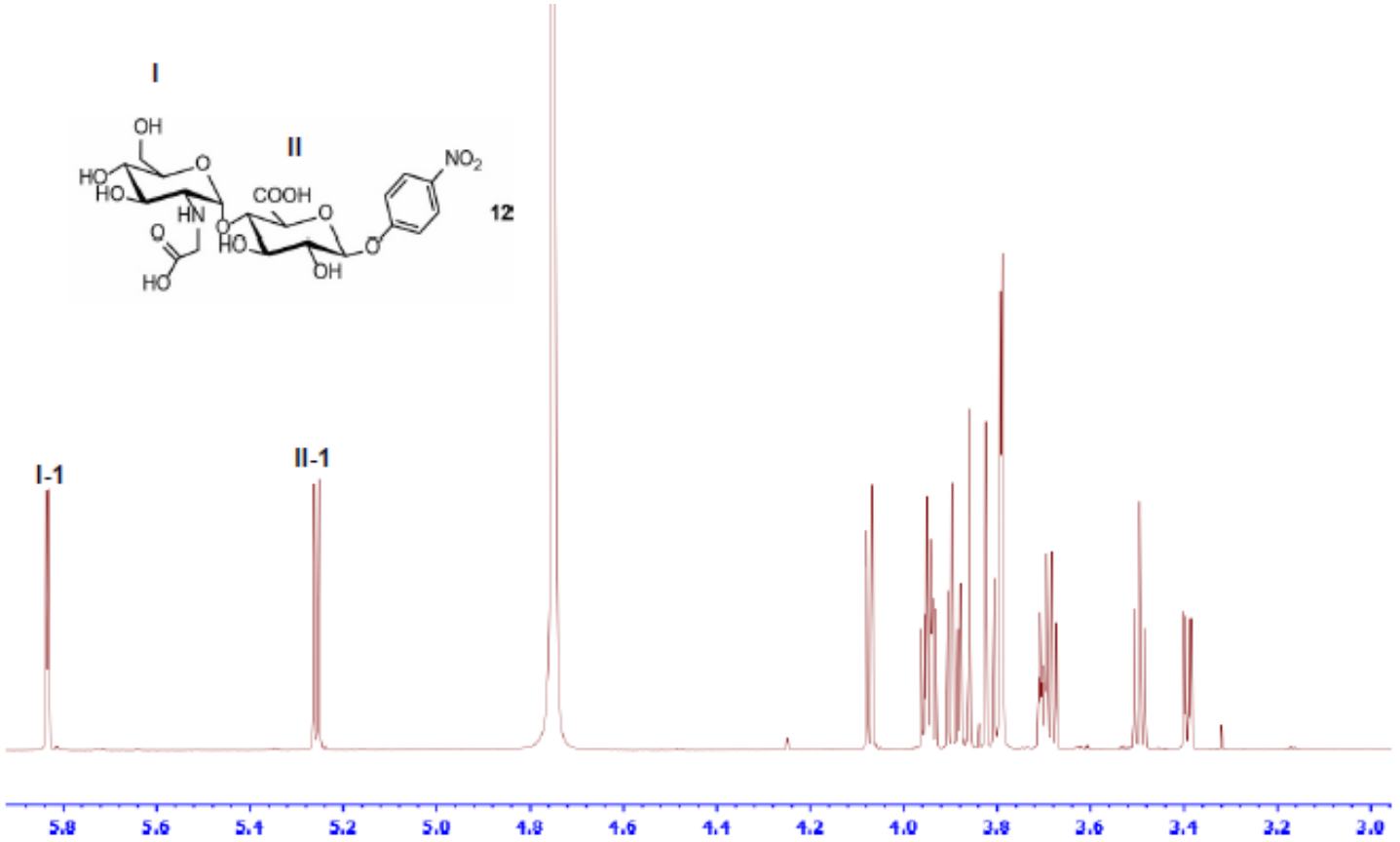




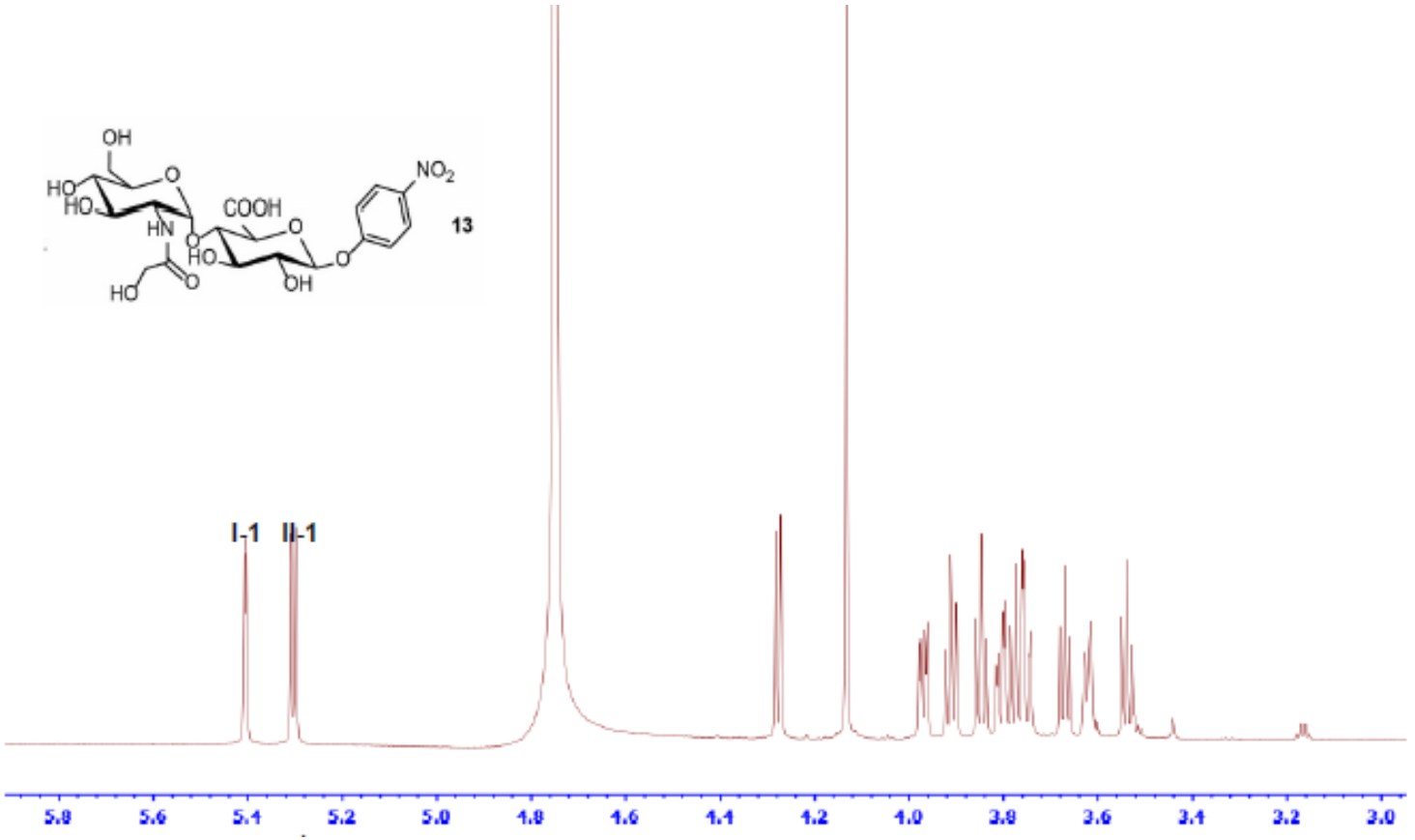
Supplementary Fig 38. Purity and MS analysis of 13. Left panel shows the PMAN-HPLC chromatogram of compound **13**. The compound was eluted as a single symmetric peak on PMAN-HPLC, suggesting that compound **13** was pure. Right panel shows the MS spectrum. The measured molecular mass was 534.1, which is very close to the calculated molecular weight (534.4). $[M-H+TFA]^-$ refers to the molecular ion of compound **13** and trifluoroacetic acid adduct.



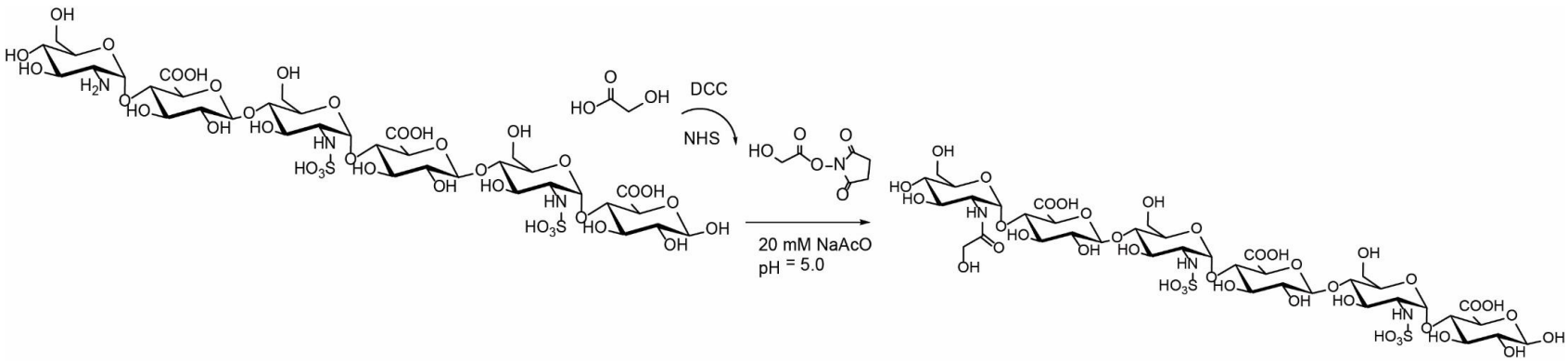
Supplementary Fig 39. ^{1}H-^{13}C HMBC key correlations of 12 and 13. The signals of featured groups are indicated. The cross peaks are connected with arrowed lines identifying correlated carbon and proton. The corresponding chemical structure is shown on top of each spectrum. The anticipated correlations are indicated in the structures by curved arrows.



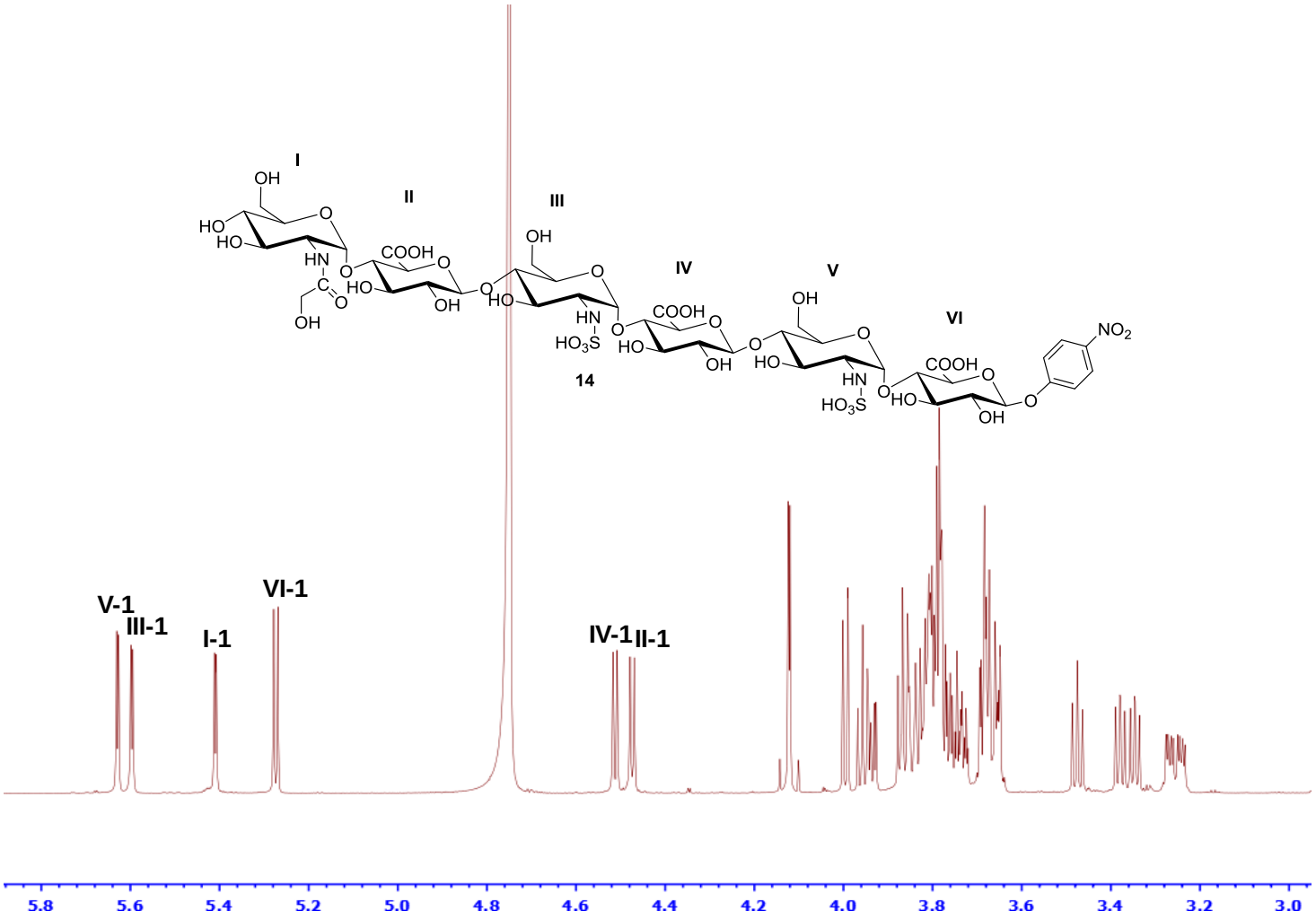
Supplementary Fig 40. 1H-NMR of compound 12(850 MHz, D2O). The signals of anomeric protons are indicated. The anomeric protons resonate as doublet at δ5.83 and 5.26 ppm. The chemical structure of compound **12**is shown on top of figure.



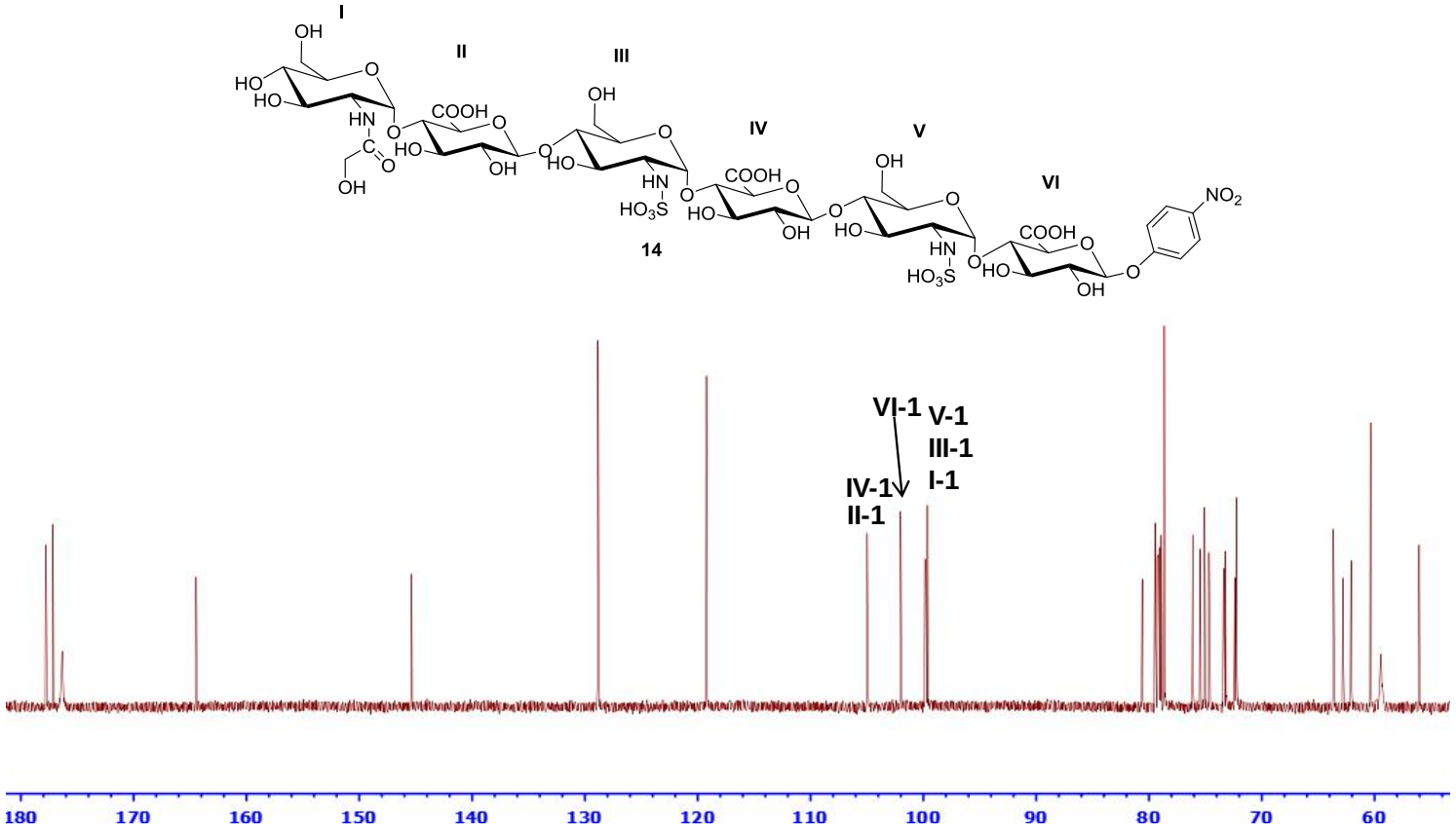
Supplementary Fig 41. 1H-NMR of compound 13(850 MHz, D2O). The signals of anomeric protons are indicated. The anomeric protons resonate as doublet at δ5.41 and 5.30 ppm. The anomeric proton signals from **13**were similar to the comparable ones observed from **14**(I-1, 5.41 ppm/VI-1, 5.27 ppm).The chemical structure of compound **13**is shown on top of figure.



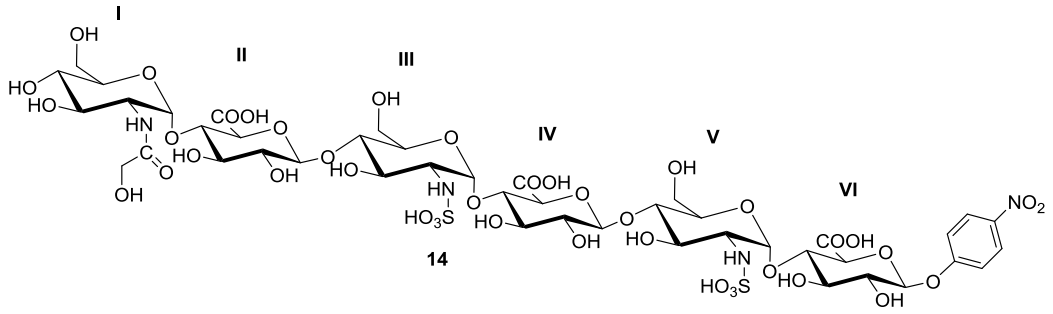
Supplementary Fig. 42. The scheme for the synthesis compound 14. The starting hexasaccharide was an intermediate from the synthesis of compound **1**.

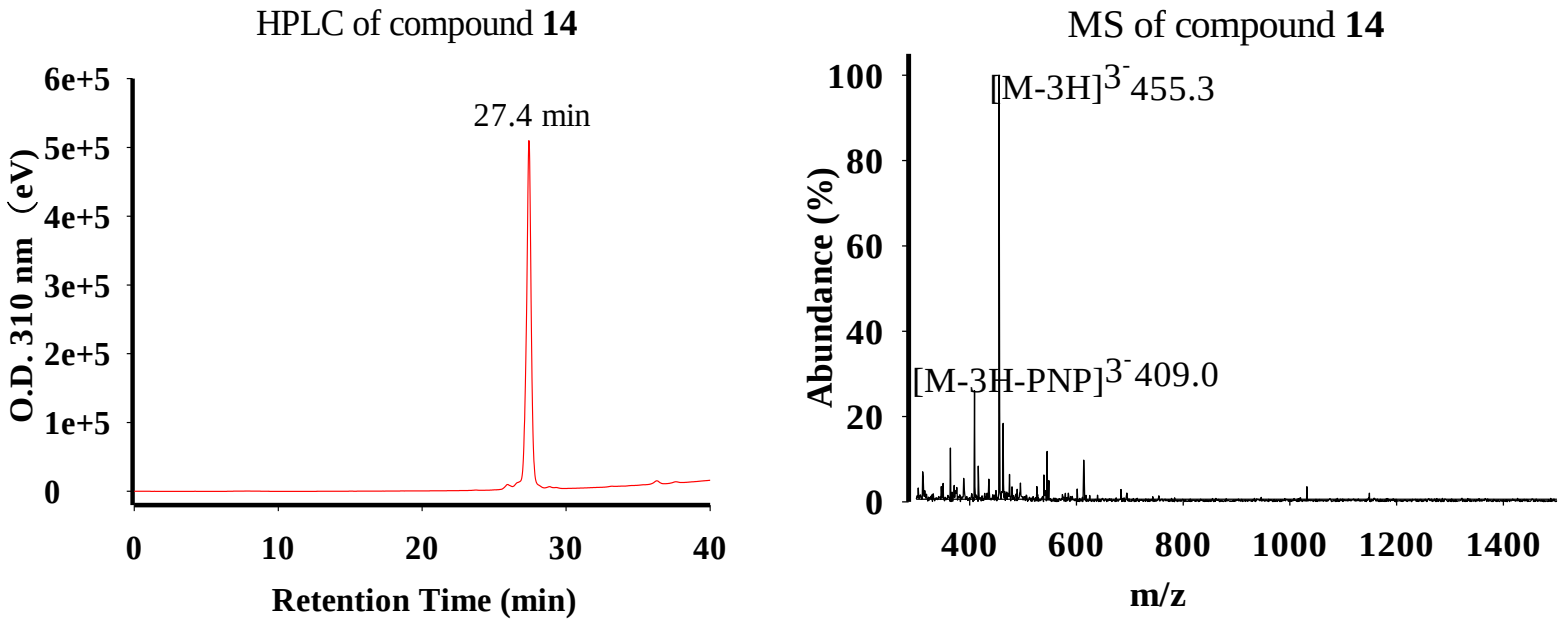


Supplementary Fig 43. ^{1}H-NMR of compound 14 (850 MHz, D_2O). The signals of anomeric protons are indicated. The anomeric protons resonate as doublet at δ 5.63, 5.60, 5.41, 5.27, 4.51, and 4.47 ppm. Chemical structure of compound **14** is shown on top of figure.

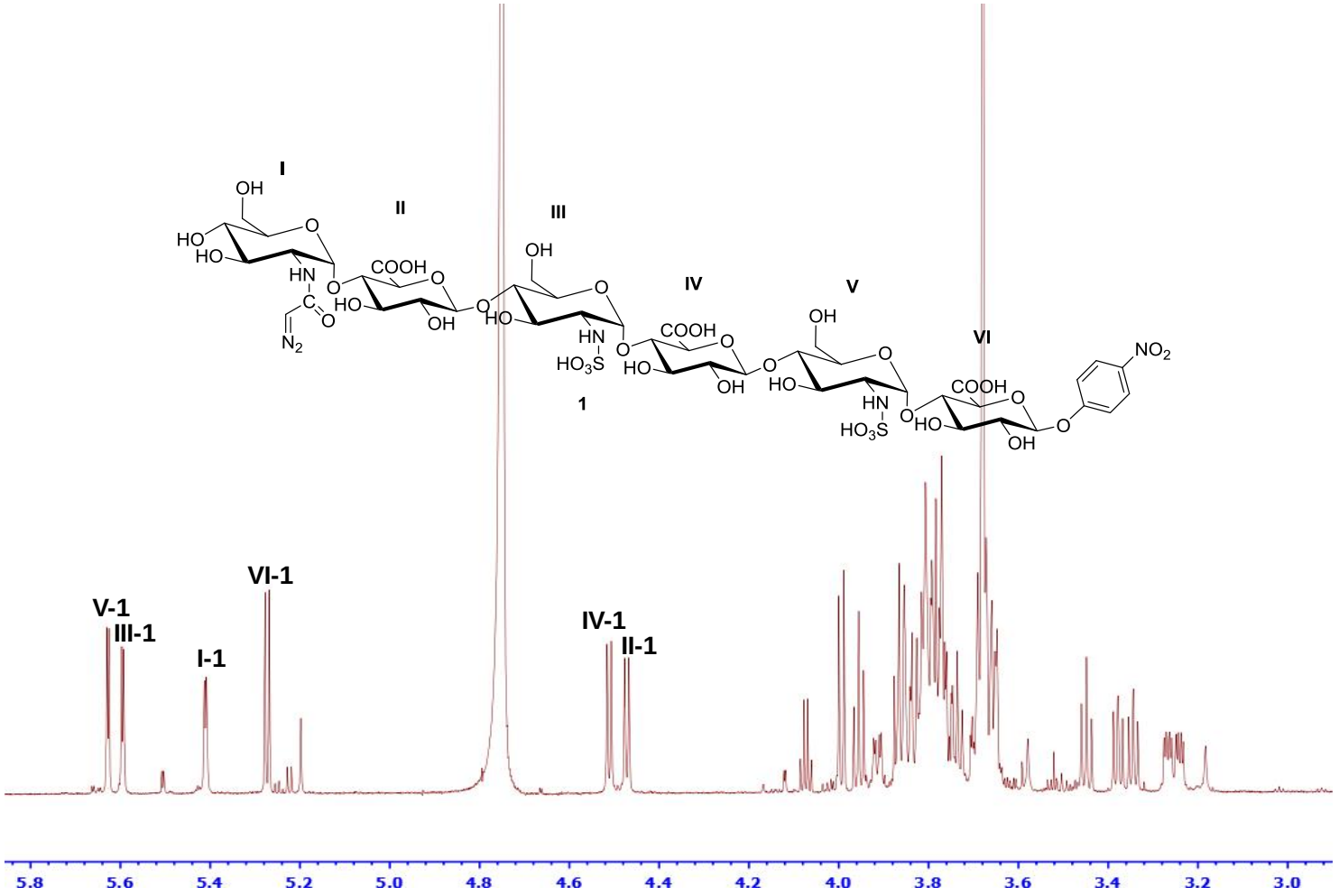


Supplementary Fig 44. ^{13}C-NMR of compound 14 (212.5 MHz, D_2O). The signals of anomeric carbons are indicated. The anomeric carbons resonate at δ 105.0, 104.9, 102.0, 99.9, 99.8, and 99.6 ppm. Chemical structure of compound **14** is shown on top of figure.

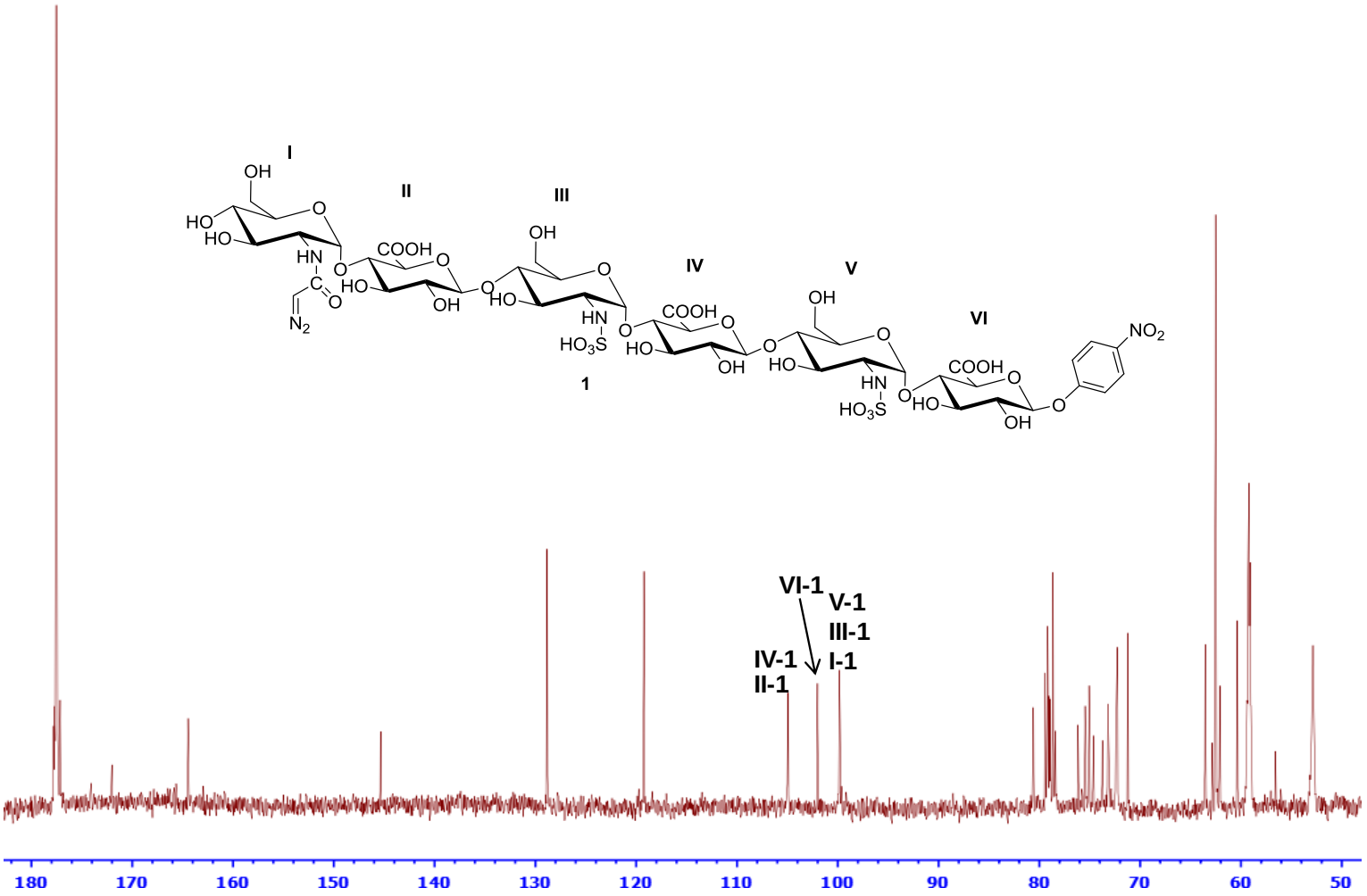




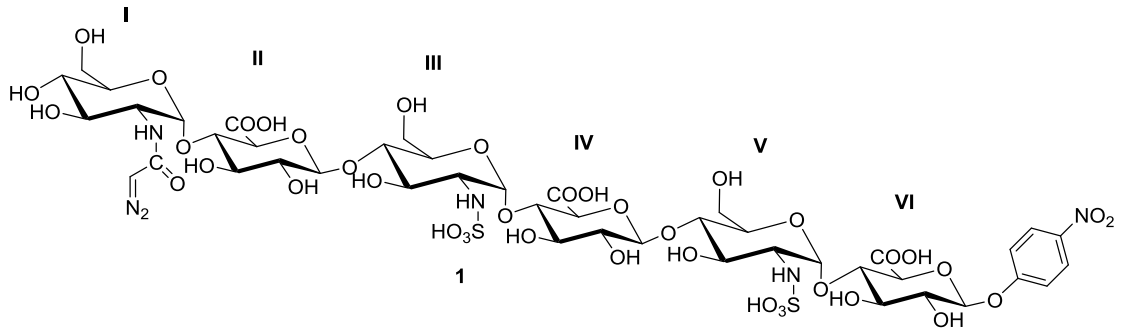
Supplementary Fig. 45. Purity and MS analysis of compound 14. Left panel shows the PMAN-HPLC chromatogram of compound **14**. The compound was eluted as a single symmetric peak on PMAN-HPLC, suggesting that compound **14** was pure. Right panel shows the MS spectrum. The measured molecular mass was 1368.9, which is very close to the calculated molecular weight (1369.1). $[M-3H-PNP]^{3-}$ refers to the molecular ion of compound **14** that lost PNP moiety.

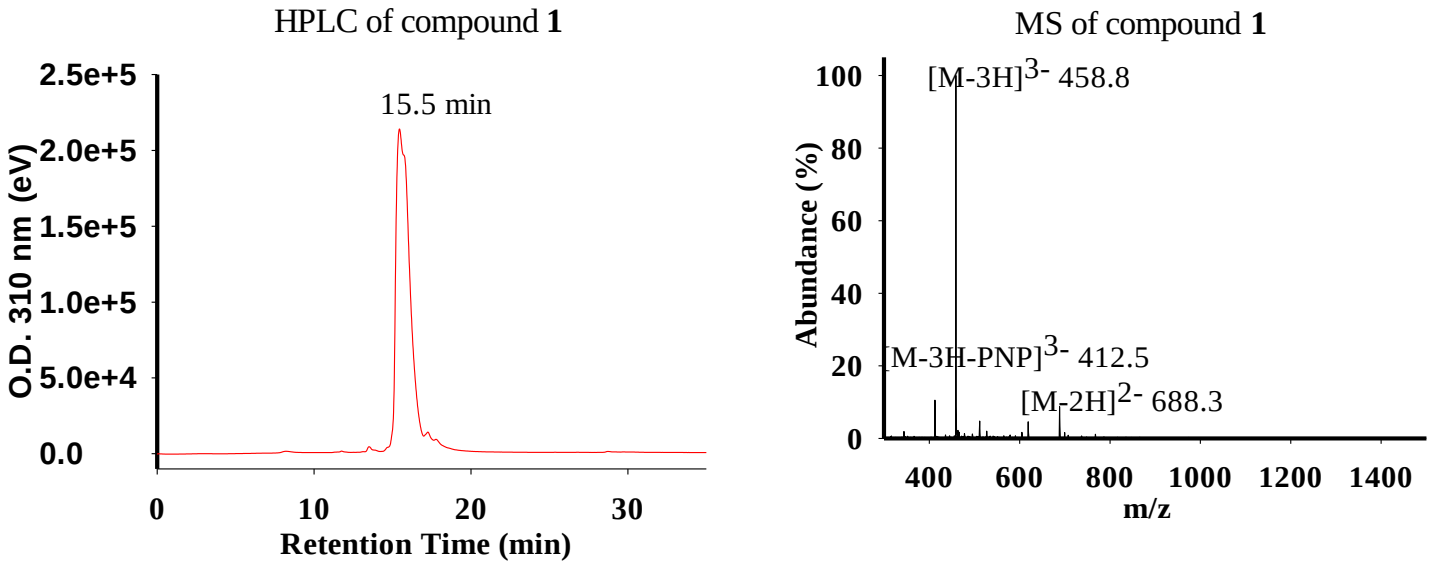


Supplementary Fig 46. ^{1}H-NMR of compound 1 (850 MHz, D_2O). The signals of anomeric protons are indicated. The anomeric protons resonate as doublet at δ 5.63, 5.60, 5.41, 5.27, 4.51, and 4.47 ppm. Chemical structure of compound **1** is shown on top of figure.

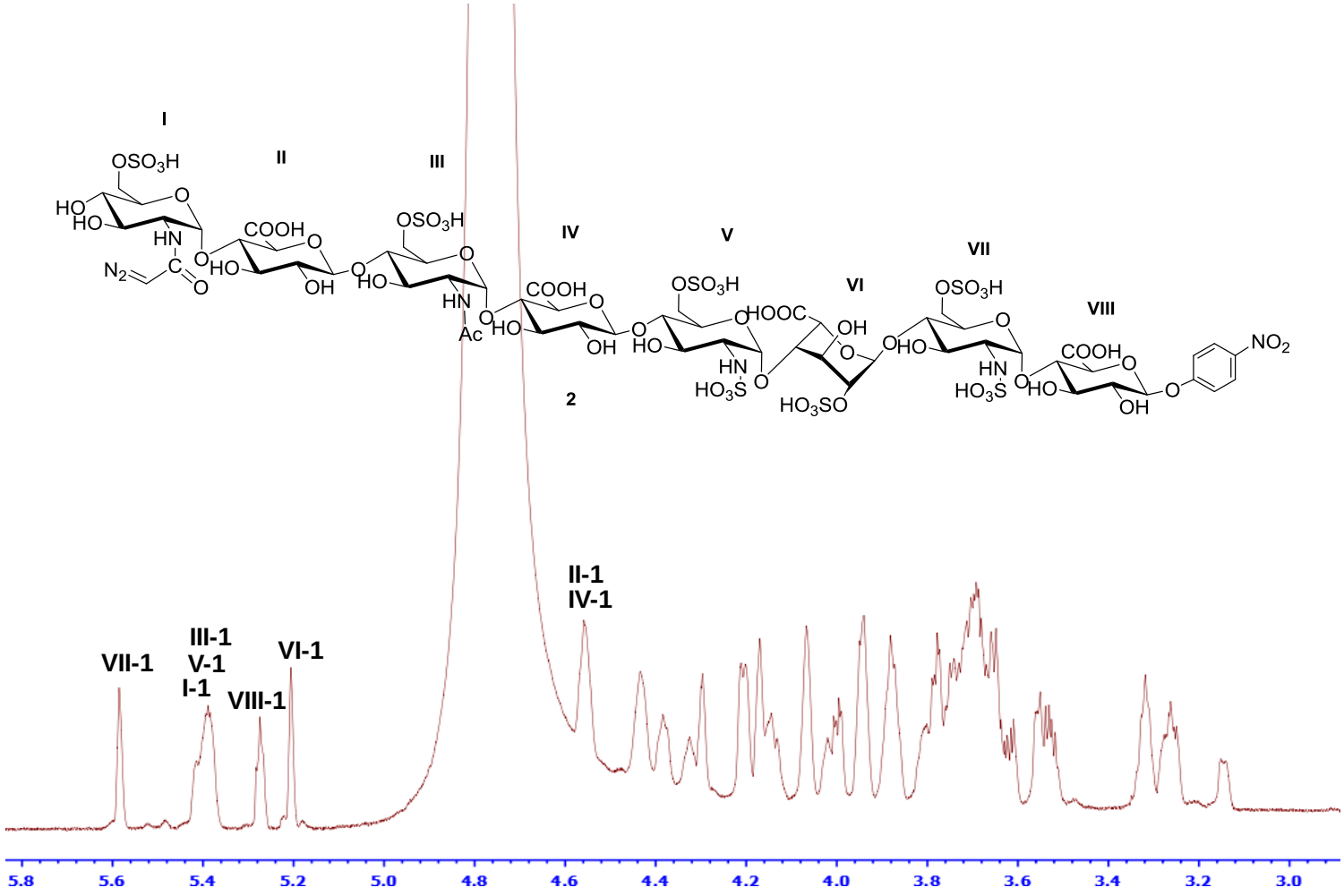


Supplementary Fig 47. ^{13}C-NMR of compound 1 (212.5 MHz, D_2O). The signals of anomeric carbons are indicated. The anomeric carbons resonate at δ 105.0, 104.9, 102.0, 99.8, 99.8, and 99.8 ppm. Chemical structure of compound **1** is shown on top of figure.

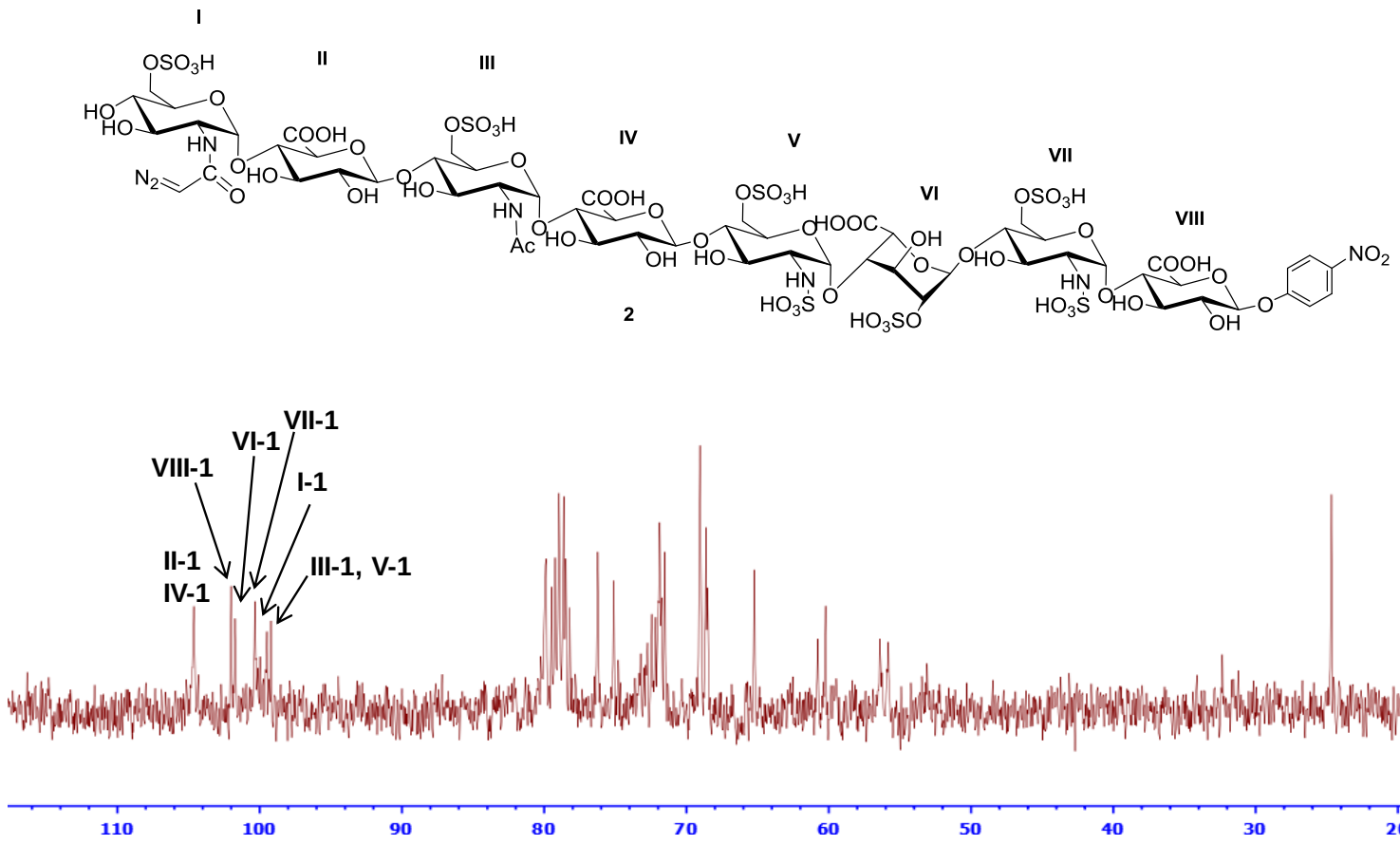




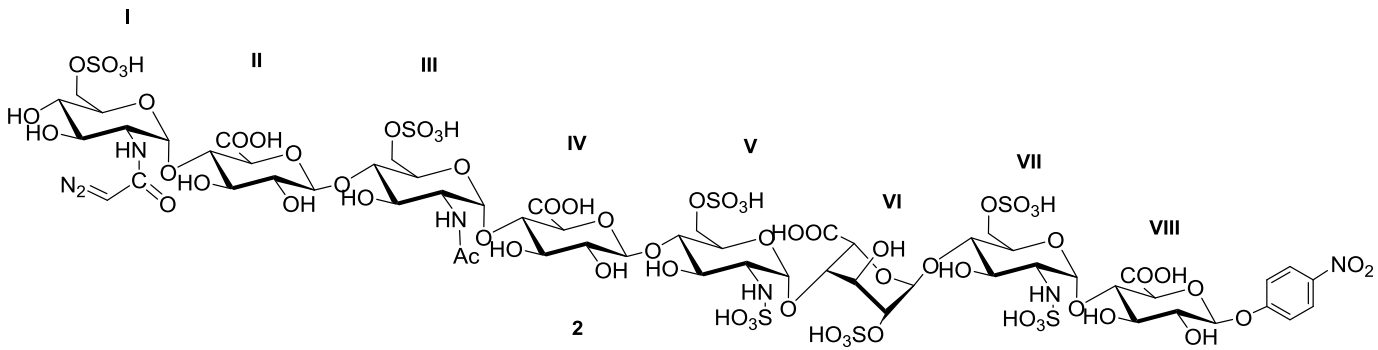
Supplementary Fig 48. Purity and MS analysis of compound 1. Left panel shows the DEAE-HPLC chromatogram of compound **1**. The compound was eluted as a single symmetric peak on DEAE-HPLC, suggesting that compound **1** was pure. Right panel shows the MS spectrum. The measured molecular mass was 1379.4, which is very close to the calculated molecular weight (1379.1). $[M\text{-}3H\text{-}PNP]^{3-}$ refers to the molecular ion that lost PNP (*p*-nitrophenyl) moiety.

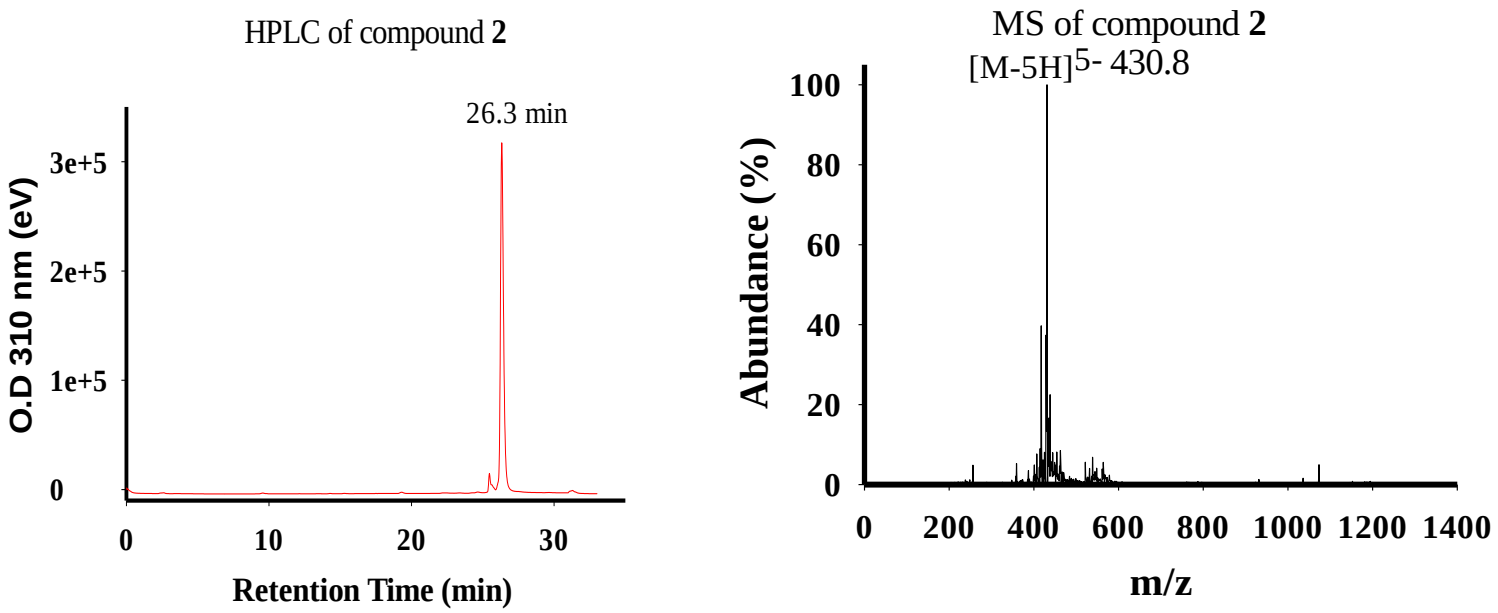


Supplementary Fig 49. ^{1}H-NMR of compound 2 (850 MHz, D_2O). The signals of anomeric protons are indicated. The anomeric protons resonate at δ 5.59, 5.41, 5.40, 5.39, 5.27, 5.21, 4.56, and 4.56 ppm. Chemical structure of compound **2** is shown on top of figure. The spectrum was acquired with 256 scans.

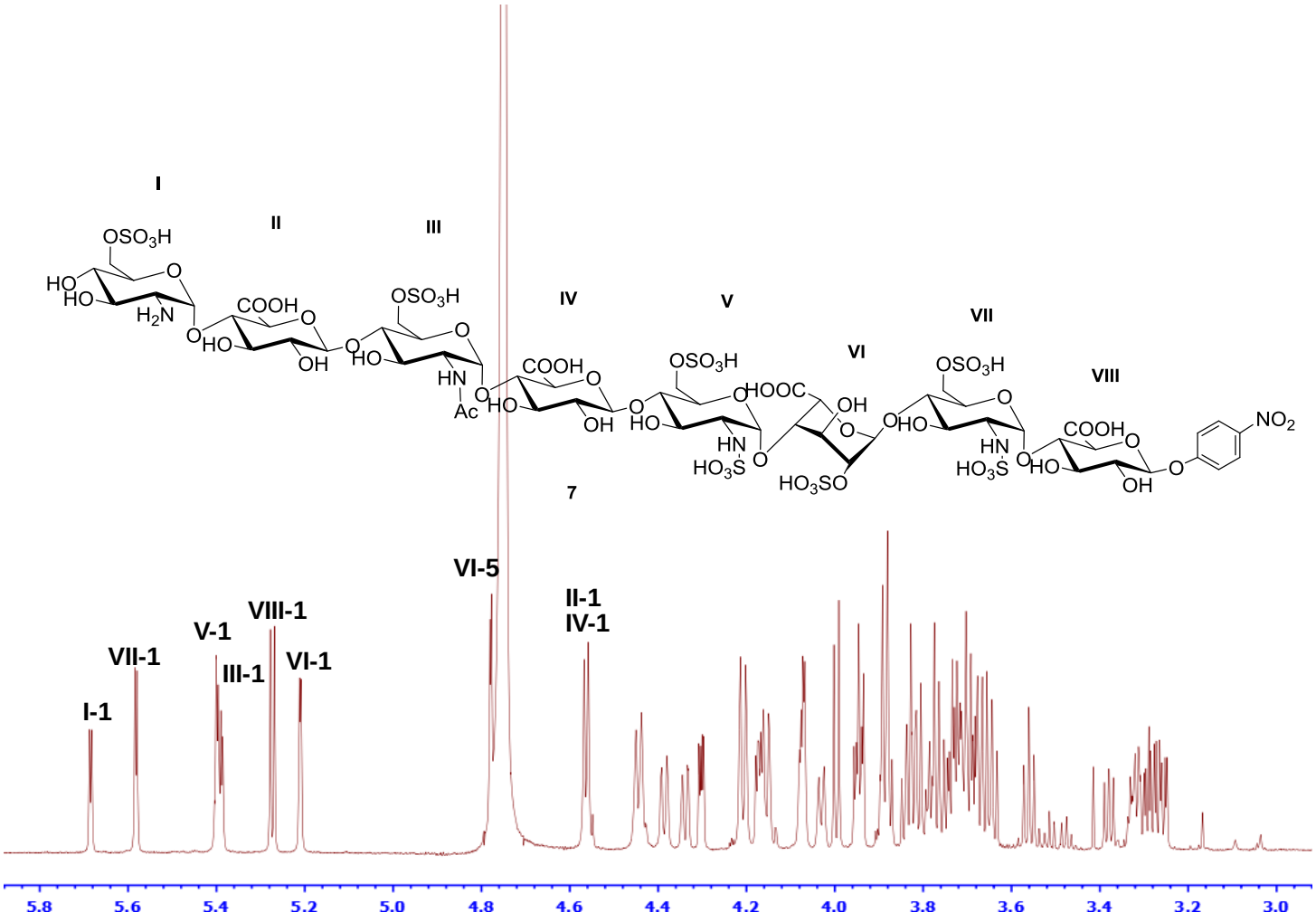


Supplementary Fig 50. ^{13}C-NMR of compound 2 (212.5 MHz, D_2O). The signals of anomeric carbons are indicated. The anomeric carbons resonate at δ 104.6, 104.6, 102.0, 101.7, 100.3, 100.0, 99.5, and 99.2 ppm. The structure of compound **2** is shown on top of figure. The spectrum was acquired using distortionless enhancement by polarization transfer with pulse width for ^{1}H at 45° .

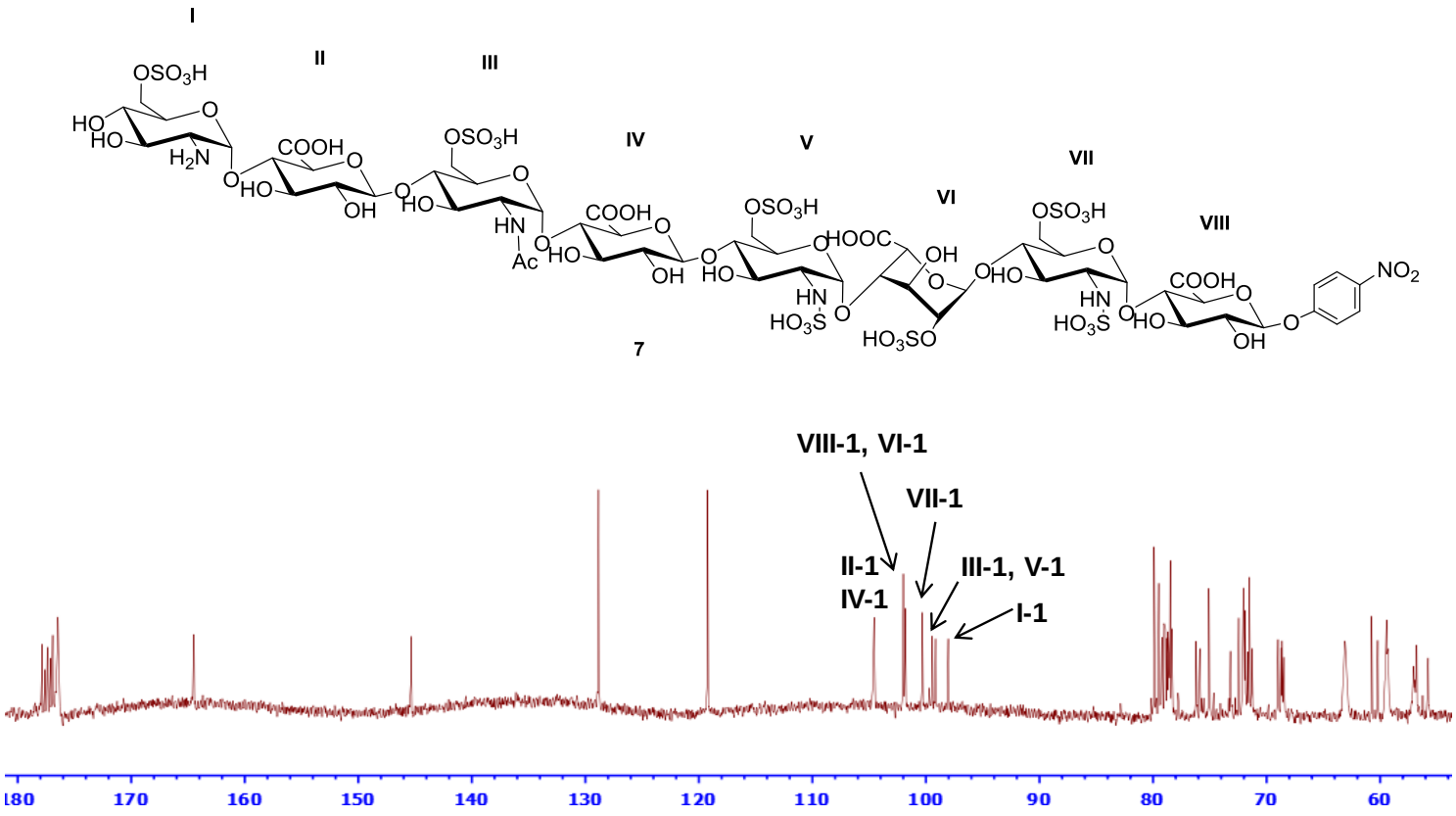




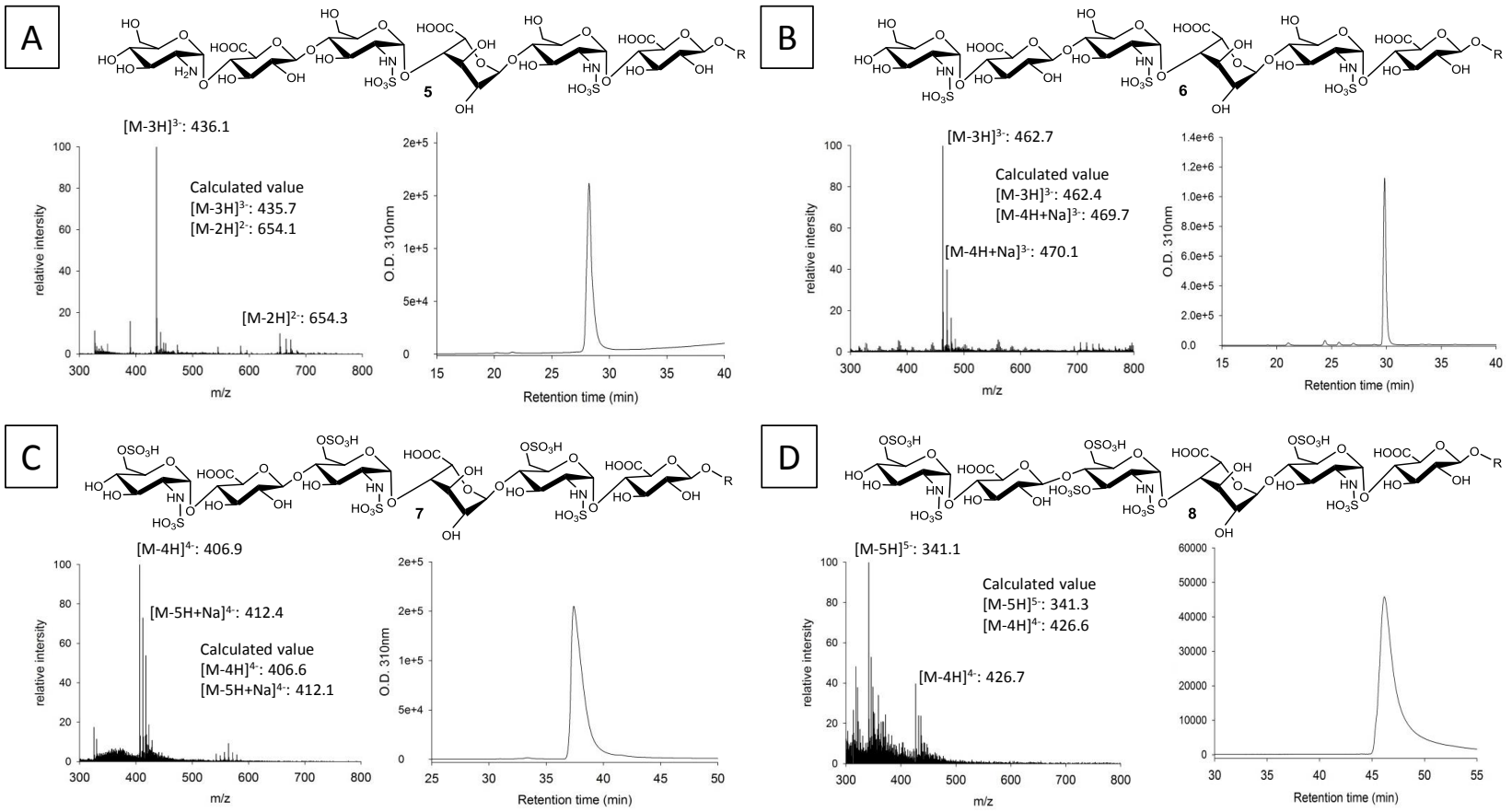
Supplementary Fig 51. Purity and structural analysis of compound 2. Left panel shows the DEAE-HPLC chromatogram of compound **2**. The compound was eluted as a major single symmetric peak on DEAE-HPLC, suggesting that compound **2** was pure. Right panel shows the MS spectrum. The measured molecular mass was 2159.0, which is very close to the calculated molecular weight (2158.7).



Supplementary Fig 52. ^{1}H-NMR of compound 7 (850 MHz, D_2O). The signals of anomeric protons are indicated. The anomeric protons resonate as doublet at δ 5.69, 5.58, 5.40, 5.39, 5.27, 5.21, 4.56, and 4.56 ppm. Chemical structure of compound **7** is shown on top of figure.



Supplementary Fig 53. ^{13}C-NMR of compound 7 (212.5 MHz, D_2O). The signals of anomeric carbons are indicated. The anomeric carbons resonate at δ 104.6, 104.6, 102.0, 101.8, 100.3, 99.5, 99.2, and 98.0 ppm. Chemical structure of compound **7** is shown on top of figure.



Supplementary Fig 54. Mass and HPLC analysis of constructs

APPENDIX II

CURRICULUM VITAE

Hsieh, Po-Hung

919-360-8966 / hsieh.pohung@gmail.com / phsieh@email.unc.edu

Education

University of North Carolina at Chapel Hill **NC, U.S.**

Doctoral study in Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy 2011-2015

National Taiwan University **Taipei, Taiwan**

MS in Division of Medicinal Chemistry, School of Pharmacy 2006-2008

- GPA: 3.93/4.00

BS in Department of Biochemical Science and Technology, School of Life Science 2002-2006

- GPA: 3.81/4.00 (major)

Publications

1. **P. H. Hsieh,** D. Thieker, M. Guerrini, R. Woods, J. Liu, Sulfation pattern dictates the conformation of heparan sulfate, 2015, *manuscript in preparation*

2. **W. Zhou**[‡]**, P. H. Hsieh**[‡], Y. Xu, T. R. O'Leary, X. Huang, J. Liu, Design and synthesis of active heparan sulfate-based probes, ***Chemical Communications***, 2015, 51, 11019-11021, ([‡]contributed equally)

3. **P. H. Hsieh**, Y. Xu, D. A. Keire, J. Liu, Chemoenzymatic synthesis and structural characterization of 2-O-sulfated glucuronic acid-containing heparan sulfate hexasaccharides, ***Glycobiology***, 2014, 24, 681–692

4. Y. Xu, C. Cai, K. Chandarajoti, **P. H. Hsieh**, L. Li, T. Pham, E. Sparkenbaugh, J. Sheng, N. Key, R. Pawlinski, E. Harris, R. Linhardt, J. Liu, Homogeneous low-molecular-weight heparins with reversible anticoagulant activity, ***Nature Chemical Biology***, 2014, 10, 248–250

5. C. Liu, J. Sheng, J. Krahn, L. Perera, Y. Xu, **P. H. Hsieh**, J. Liu, L. Pedersen, Molecular mechanism of substrate specificity for heparan sulfate 2-O-sulfotransferase, ***Journal of Biological Chemistry***, 2014, 13407–13418

6. S. H. Lam, C. T. Ruan, **P. H. Hsieh**, M. J. Su, and S. S. Lee, Hypoglycemic Diterpenoids from Tinospora crispa, ***Journal of Natural Products***, 2012, *75*, 153–159

Master Graduate Thesis

- Title: Preparation of acylated rhamnosyl flavonoid analogs as α-glucosidase inhibitors
- Major work in organic synthesis of flavonoid glycoside analogs and assay development

Work Experiences

RA, School of Pharmacy, UNC-Chapel Hill, NC 2012-2015

- Chemoenzymatic synthesis of heparin oligosaccharide anticoagulant
- Characterization of the oligosaccharide structure using NMR Spectroscopy
- Methodology development for study of heparan sulfate-protein interactions

TA, School of Pharmacy, UNC-Chapel Hill, NC 2011-2012

- TA for Compounding Lab experiments

Researcher, National Science Council, Taiwan 2010-2011

- Development of natural product drugs for diabetes
- Major work in assay-guided natural product isolation and characterization

Draftee, Taiwan 2009-2010

- Served alternative military service

RA, School of Pharmacy, NTU 2006-2008

- Exploration of hypoglycemic agents derived from Cinnamomum and Machilus plants
- Major work in organic synthesis of natural product derivatives and assay development

TA, Dept. of Pharmaceutical Science, School of Pharmacy, NTU 2006-2008

- TA for Pharmaceutical Analysis Lab experiments

RA, Institute of Plant Biology, NTU 2004-2006

- Isolation and analysis the heat-induced genes from heat-tolerant tomato
- Major work in genetic manipulation and protein purification

Honors & Awards

Doctoral and Master Graduate School:

- Ministry of Education Fellowship, Taiwan, 2011-2012
- Honor Student Award, 2006-2008
- Graduate Student Representative, 2008

Undergraduate:

- Cultural Foundation NTU Alumni Association Scholarship, 2006
- Valedictorian of Department, 2006
- Presidential Award, 2005
- Winner of the poster contest, 2005

Collaborations & Projects

Dr. David A Keire (FDA, USA) 2012-2015

- Chemoenzymatic synthesis of oligosaccharide standards

Dr. Lars Pedersen (NIEHS, USA) 2012-2015

- Investigation of molecular mechanism of sulfotransferase

Dr. Greg Young (School of Medicine at UNC) 2013-2015

- Characterization of synthetic carbohydrate structure

Dr. Robert Woods (Complex Carbohydrate Research Center at UGA) 2014-2015

- Performing molecular dynamics simulation of oligosaccharides

Dr. Marco Guerrini (G. Ronzoni Institute, Italy) 2014-2015

- NMR analysis of carbohydrate structure and conformation

REFERENCES

1. Xu, Y., et al., *Chemoenzymatic synthesis of homogeneous ultralow molecular weight heparins.* Science, 2011. **334**(6055): p. 498-501.

2. Bishop, J.R., M. Schuksz, and J.D. Esko, *Heparan sulphate proteoglycans fine-tune mammalian physiology.* Nature, 2007. **446**(7139): p. 1030-7.

3. Iozzo, R.V., *Heparan sulfate proteoglycans: intricate molecules with intriguing functions.* J Clin Invest, 2001. **108**(2): p. 165-7.

4. Peterson, S., A. Frick, and J. Liu, *Design of biologically active heparan sulfate and heparin using an enzyme-based approach.* Nat Prod Rep, 2009. **26**(5): p. 610-27.

5. Liu, J. and S.C. Thorp, *Cell surface heparan sulfate and its roles in assisting viral infections.* Med Res Rev, 2002. **22**(1): p. 1-25.

6. Lever, R. and C.P. Page, *Novel drug development opportunities for heparin.* Nat Rev Drug Discov, 2002. **1**(2): p. 140-8.

7. Abrink, M., M. Grujic, and G. Pejler, *Serglycin is essential for maturation of mast cell secretory granule.* J Biol Chem, 2004. **279**(39): p. 40897-905.

8. Mulloy, B. and M.J. Forster, *Conformation and dynamics of heparin and heparan sulfate.* Glycobiology, 2000. **10**(11): p. 1147-56.

9. Raman, R., V. Sasisekharan, and R. Sasisekharan, *Structural insights into biological roles of protein-glycosaminoglycan interactions.* Chem Biol, 2005. **12**(3): p. 267-77.

10. Jin, L., et al., *The anticoagulant activation of antithrombin by heparin.* Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14683-8.

11. Liu, H., Z. Zhang, and R.J. Linhardt, *Lessons learned from the contamination of heparin.* Nat Prod Rep, 2009. **26**(3): p. 313-21.

12. Linhardt, R.J., *2003 Claude S. Hudson Award address in carbohydrate chemistry. Heparin: structure and activity.* J Med Chem, 2003. **46**(13): p. 2551-64.

13. Sommers, C.D., et al., *Characterization of currently marketed heparin products: analysis of molecular weight and heparinase-I digest patterns.* Anal Bioanal Chem, 2011. **401**(8): p. 2445-54.

14. Ahsan, A., et al., *Molecular profiling and weight determination of heparins and depolymerized heparins.* J Pharm Sci, 1995. **84**(6): p. 724-7.

15. Casu, B., *Structure and biological activity of heparin.* Adv Carbohydr Chem Biochem, 1985. **43**: p. 51-134.

16. Griffin, C.C., et al., *Isolation and characterization of heparan sulfate from crude porcine intestinal mucosal peptidoglycan heparin.* Carbohydr Res, 1995. **276**(1): p. 183-97.

17. Esko, J.D. and L. Zhang, *Influence of core protein sequence on glycosaminoglycan assembly.* Curr Opin Struct Biol, 1996. **6**(5): p. 663-70.

18. Esko, J.D. and S.B. Selleck, *Order out of chaos: assembly of ligand binding sites in heparan sulfate.* Annu Rev Biochem, 2002. **71**: p. 435-71.

19. Kuhn, J., et al., *First isolation of human UDP-D-xylose: proteoglycan core protein beta-D-xylosyltransferase secreted from cultured JAR choriocarcinoma cells.* J Biol Chem, 2001. **276**(7): p. 4940-7.

20. Almeida, R., et al., *A family of human beta4-galactosyltransferases. Cloning and expression of two novel UDP-galactose:beta-n-acetylglucosamine beta1, 4-galactosyltransferases, beta4Gal-T2 and beta4Gal-T3.* J Biol Chem, 1997. **272**(51): p. 31979-91.

21. Amado, M., et al., *A family of human beta3-galactosyltransferases. Characterization of four members of a UDP-galactose:beta-N-acetyl-glucosamine/beta-nacetyl-galactosamine beta-1,3-galactosyltransferase family.* J Biol Chem, 1998. **273**(21): p. 12770-8.

22. Terayama, K., et al., *Cloning and functional expression of a novel glucuronyltransferase involved in the biosynthesis of the carbohydrate epitope HNK-1.* Proc Natl Acad Sci U S A, 1997. **94**(12): p. 6093-8.

23. Gulberti, S., et al., *Phosphorylation and sulfation of oligosaccharide substrates critically influence the activity of human beta1,4-galactosyltransferase 7 (GalT-I) and beta1,3-glucuronosyltransferase I (GlcAT-I) involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans.* J Biol Chem, 2005. **280**(2): p. 1417-25.

24. Kim, B.T., et al., *Human tumor suppressor EXT gene family members EXTL1 and EXTL3 encode alpha 1,4- N-acetylglucosaminyltransferases that likely are involved in heparan sulfate/ heparin biosynthesis.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7176-81.

25. Zak, B.M., B.E. Crawford, and J.D. Esko, *Hereditary multiple exostoses and heparan sulfate polymerization.* Biochim Biophys Acta, 2002. **1573**(3): p. 346-55.

26. Kitagawa, H., H. Shimakawa, and K. Sugahara, *The tumor suppressor EXT-like gene EXTL2 encodes an alpha1, 4-N-acetylhexosaminyltransferase that transfers N-acetylgalactosamine and N-acetylglucosamine to the common glycosaminoglycan-protein linkage region. The key enzyme for the chain initiation of heparan sulfate.* J Biol Chem, 1999. **274**(20): p. 13933-7.

27. Chen, R.L. and A.D. Lander, *Mechanisms underlying preferential assembly of heparan sulfate on glypican-1.* J Biol Chem, 2001. **276**(10): p. 7507-17.

28. Presto, J., et al., *Heparan sulfate biosynthesis enzymes EXT1 and EXT2 affect NDST1 expression and heparan sulfate sulfation.* Proc Natl Acad Sci U S A, 2008. **105**(12): p. 4751-6.

29. Holmborn, K., et al., *Heparan sulfate synthesized by mouse embryonic stem cells deficient in NDST1 and NDST2 is 6-O-sulfated but contains no N-sulfate groups.* J Biol Chem, 2004. **279**(41): p. 42355-8.

30. Aikawa, J., et al., *Multiple isozymes of heparan sulfate/heparin GlcNAc N-deacetylase/GlcN N-sulfotransferase. Structure and activity of the fourth member, NDST4.* J Biol Chem, 2001. **276**(8): p. 5876-82.

31. Forsberg, E., et al., *Abnormal mast cells in mice deficient in a heparin-synthesizing enzyme.* Nature, 1999. **400**(6746): p. 773-6.

32. Kusche-Gullberg, M., et al., *Identification and expression in mouse of two heparan sulfate glucosaminyl N-deacetylase/N-sulfotransferase genes.* J Biol Chem, 1998. **273**(19): p. 11902-7.

33. Pallerla, S.R., et al., *Altered heparan sulfate structure in mice with deleted NDST3 gene function.* J Biol Chem, 2008. **283**(24): p. 16885-94.

34. Crawford, B.E., et al., *Cloning, Golgi localization, and enzyme activity of the full-length heparin/heparan sulfate-glucuronic acid C5-epimerase.* J Biol Chem, 2001. **276**(24): p. 21538-43.

35. Feyerabend, T.B., et al., *Heparan sulfate C5-epimerase is essential for heparin biosynthesis in mast cells.* Nat Chem Biol, 2006. **2**(4): p. 195-6.

36. Hagner-Mcwhirter, A., U. Lindahl, and J. Li, *Biosynthesis of heparin/heparan sulphate: mechanism of epimerization of glucuronyl C-5.* Biochem J, 2000. **347 Pt 1**: p. 69-75.

37. Hagner-McWhirter, A., et al., *Biosynthesis of heparin/heparan sulfate: kinetic studies of the glucuronyl C5-epimerase with N-sulfated derivatives of the Escherichia coli K5 capsular polysaccharide as substrates.* Glycobiology, 2000. **10**(2): p. 159-71.

38. Sheng, J., et al., *Uncovering biphasic catalytic mode of C5-epimerase in heparan sulfate biosynthesis.* J Biol Chem, 2012. **287**(25): p. 20996-1002.

39. Bullock, S.L., et al., *Renal agenesis in mice homozygous for a gene trap mutation in the gene encoding heparan sulfate 2-sulfotransferase.* Genes Dev, 1998. **12**(12): p. 1894-906.

40. Kobayashi, M., et al., *Purification and characterization of heparan sulfate 2-sulfotransferase from cultured Chinese hamster ovary cells.* J Biol Chem, 1996. **271**(13): p. 7645-53.

41. Pinhal, M.A., et al., *Enzyme interactions in heparan sulfate biosynthesis: uronosyl 5-epimerase and 2-O-sulfotransferase interact in vivo.* Proc Natl Acad Sci U S A, 2001. **98**(23): p. 12984-9.

42. Rong, J., et al., *Substrate specificity of the heparan sulfate hexuronic acid 2-O-sulfotransferase.* Biochemistry, 2001. **40**(18): p. 5548-55.

43. Liu, C., et al., *Molecular mechanism of substrate specificity for heparan sulfate 2-O-sulfotransferase.* J Biol Chem, 2014. **289**(19): p. 13407-18.

44. Habuchi, H., et al., *The occurrence of three isoforms of heparan sulfate 6-O-sulfotransferase having different specificities for hexuronic acid adjacent to the targeted N-sulfoglucosamine.* J Biol Chem, 2000. **275**(4): p. 2859-68.

45. Zhang, L., et al., *6-O-sulfotransferase-1 represents a critical enzyme in the anticoagulant heparan sulfate biosynthetic pathway.* J Biol Chem, 2001. **276**(45): p. 42311-21.

46. Jemth, P., et al., *Oligosaccharide library-based assessment of heparan sulfate 6-O-sulfotransferase substrate specificity.* J Biol Chem, 2003. **278**(27): p. 24371-6.

47. Smeds, E., et al., *Substrate specificities of mouse heparan sulphate glucosaminyl 6-O-sulphotransferases.* Biochem J, 2003. **372**(Pt 2): p. 371-80.

48. Brickman, Y.G., et al., *Structural modification of fibroblast growth factor-binding heparan sulfate at a determinative stage of neural development.* J Biol Chem, 1998. **273**(8): p. 4350-9.

49. Cadwallader, A.B. and H.J. Yost, *Combinatorial expression patterns of heparan sulfate sulfotransferases in zebrafish: I. The 3-O-sulfotransferase family.* Dev Dyn, 2006. **235**(12): p. 3423-31.

50. Shworak, N.W., et al., *Molecular cloning and expression of mouse and human cDNAs encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferase.* J Biol Chem, 1997. **272**(44): p. 28008-19.

51. Shworak, N.W., et al., *Pathway-specific regulation of the synthesis of anticoagulantly active heparan sulfate.* J Biol Chem, 1994. **269**(40): p. 24941-52.

52. Colliec-Jouault, S., et al., *Characterization of a cell mutant specifically defective in the synthesis of anticoagulantly active heparan sulfate.* J Biol Chem, 1994. **269**(40): p. 24953-8.

53. Shukla, D., et al., *A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry.* Cell, 1999. **99**(1): p. 13-22.

54. Zhang, L., et al., *The effect of precursor structures on the action of glucosaminyl 3-O-sulfotransferase-1 and the biosynthesis of anticoagulant heparan sulfate.* J Biol Chem, 2001. **276**(31): p. 28806-13.

55. Moon, A.F., et al., *Dissecting the substrate recognition of 3-O-sulfotransferase for the biosynthesis of anticoagulant heparin.* Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5265-70.

56. Moon, A.F., et al., *Structural analysis of the sulfotransferase (3-o-sulfotransferase isoform 3) involved in the biosynthesis of an entry receptor for herpes simplex virus 1.* J Biol Chem, 2004. **279**(43): p. 45185-93.

57. Liu, J., et al., *Expression of heparan sulfate D-glucosaminyl 3-O-sulfotransferase isoforms reveals novel substrate specificities.* J Biol Chem, 1999. **274**(8): p. 5185-92.

58. Xia, G., et al., *Heparan sulfate 3-O-sulfotransferase isoform 5 generates both an antithrombin-binding site and an entry receptor for herpes simplex virus, type 1.* J Biol Chem, 2002. **277**(40): p. 37912-9.

59. Xu, D., et al., *Characterization of heparan sulphate 3-O-sulphotransferase isoform 6 and its role in assisting the entry of herpes simplex virus type 1.* Biochem J, 2005. **385**(Pt 2): p. 451-9.

60. Tiwari, V., et al., *A role for 3-O-sulfotransferase isoform-4 in assisting HSV-1 entry and spread.* Biochem Biophys Res Commun, 2005. **338**(2): p. 930-7.

61. Xu, D., et al., *Engineering sulfotransferases to modify heparan sulfate.* Nat Chem Biol, 2008. **4**(3): p. 200-2.

62. Linhardt, R.J. and J. Liu, *Synthetic heparin.* Curr Opin Pharmacol, 2012. **12**(2): p. 217-9.

63. Hirsh, J. and R. Raschke, *Heparin and low-molecular-weight heparin: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy.* Chest, 2004. **126**(3 Suppl): p. 188S-203S.

64. Linhardt, R.J. and N.S. Gunay, *Production and chemical processing of low molecular weight heparins.* Semin Thromb Hemost, 1999. **25 Suppl 3**: p. 5-16.

65. Merli, G.J. and J.B. Groce, *Pharmacological and clinical differences between low-molecular-weight heparins: implications for prescribing practice and therapeutic interchange.* P T, 2010. **35**(2): p. 95-105.

66. Nightingale, S.L., *From the Food and Drug Administration.* JAMA, 1993. **270**(14): p. 1672.

67. Brieger, D. and J. Dawes, *Production method affects the pharmacokinetic and ex vivo biological properties of low molecular weight heparins.* Thromb Haemost, 1997. **77**(2): p. 317-22.

68. McMahon, A.W., et al., *Description of hypersensitivity adverse events following administration of heparin that was potentially contaminated with oversulfated chondroitin sulfate in early 2008.* Pharmacoepidemiol Drug Saf, 2010. **19**(9): p. 921-33.

69. Kishimoto, T.K., et al., *Contaminated heparin associated with adverse clinical events and activation of the contact system.* N Engl J Med, 2008. **358**(23): p. 2457-67.

70. Guerrini, M., et al., *Oversulfated chondroitin sulfate is a contaminant in heparin associated with adverse clinical events.* Nat Biotechnol, 2008. **26**(6): p. 669-75.

71. Hedlund, K.D., et al., *The heparin recall of 2008.* Perfusion, 2013. **28**(1): p. 61-5.

72. Zhang, Z., et al., *Oversulfated chondroitin sulfate: impact of a heparin impurity, associated with adverse clinical events, on low-molecular-weight heparin preparation.* J Med Chem, 2008. **51**(18): p. 5498-501.

73. Petitou, M. and C.A. van Boeckel, *A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next?* Angew Chem Int Ed Engl, 2004. **43**(24): p. 3118-33.

74. Xu, Y., et al., *Homogeneous low-molecular-weight heparins with reversible anticoagulant activity.* Nat Chem Biol, 2014. **10**(4): p. 248-50.

75. Pempe, E.H., et al., *Probing structural selectivity of synthetic heparin binding to Stabilin protein receptors.* J Biol Chem, 2012. **287**(25): p. 20774-83.

76. Lim, W., et al., *Meta-analysis: low-molecular-weight heparin and bleeding in patients with severe renal insufficiency.* Ann Intern Med, 2006. **144**(9): p. 673-84.

77. Guerrini, M. and G. Torri, *NMR Spectroscopy in Pharmaceutical Analysis*, Holzgrabe U, Diehl B, and Wawer I, Editors. 2008. p. 407-428.

78. Guo, Y.C. and H.E. Conrad, *The disaccharide composition of heparins and heparan sulfates.* Anal Biochem, 1989. **176**(1): p. 96-104.

79. Jones, C.J., S. Beni, and C.K. Larive, *Understanding the effect of the counterion on the reverse-phase ion-pair high-performance liquid chromatography (RPIP-HPLC) resolution of heparin-related saccharide anomers.* Anal Chem, 2011. **83**(17): p. 6762-9.

80. Desai, U.R., H.M. Wang, and R.J. Linhardt, *Substrate specificity of the heparin lyases from Flavobacterium heparinum.* Arch Biochem Biophys, 1993. **306**(2): p. 461-8.

81. Shively, J.E. and H.E. Conrad, *Formation of anhydrosugars in the chemical depolymerization of heparin.* Biochemistry, 1976. **15**(18): p. 3932-42.

82. Ferro, D.R., et al., *Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences.* Carbohydr Res, 1990. **195**(2): p. 157-67.

83. Li, W., et al., *Structure of the antithrombin-thrombin-heparin ternary complex reveals the antithrombotic mechanism of heparin.* Nat Struct Mol Biol, 2004. **11**(9): p. 857-62.

84. Wormald, M.R., et al., *Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling.* Chem Rev, 2002. **102**(2): p. 371-86.

85. DeMarco, M.L. and R.J. Woods, *Structural glycobiology: a game of snakes and ladders.* Glycobiology, 2008. **18**(6): p. 426-40.

86. Gonzalez-Outeirino, J., et al., *Reconciling solvent effects on rotamer populations in carbohydrates - A joint MD and NMR analysis.* Can J Chem, 2006. **84**(4): p. 569-579.

87. Haasnoot, C.A.G., F.A.A.M. Deleeuw, and C. Altona, *The Relationship between Proton-Proton Nmr Coupling-Constants and Substituent Electronegativities .1. An Empirical Generalization of the Karplus Equation.* Tetrahedron, 1980. **36**(19): p. 2783-2792.

88. Altona, C., et al., *Empirical Group Electronegativities for Vicinal Nmr Proton-Proton Couplings Along a C-C Bond - Solvent Effects and Reparameterization of the Haasnoot Equation.* Magnetic Resonance in Chemistry, 1994. **32**(11): p. 670-678.

89. Davie, E.W., K. Fujikawa, and W. Kisiel, *The coagulation cascade: initiation, maintenance, and regulation.* Biochemistry, 1991. **30**(43): p. 10363-70.

90. Gailani, D. and T. Renne, *Intrinsic pathway of coagulation and arterial thrombosis.* Arterioscler Thromb Vasc Biol, 2007. **27**(12): p. 2507-13.

91. O'Keeffe, D., et al., *The heparin binding properties of heparin cofactor II suggest an antithrombin-like activation mechanism.* J Biol Chem, 2004. **279**(48): p. 50267-73.

92. Baglin, T.P., et al., *Crystal structures of native and thrombin-complexed heparin cofactor II reveal a multistep allosteric mechanism.* Proc Natl Acad Sci U S A, 2002. **99**(17): p. 11079-84.

93. Sun, W., A.S. Eriksson, and S. Schedin-Weiss, *Heparin enhances the inhibition of factor Xa by protein C inhibitor in the presence but not in the absence of Ca2+.* Biochemistry, 2009. **48**(5): p. 1094-8.

94. Van Walderveen, M.C., L.R. Berry, and A.K. Chan, *Effect of covalent antithrombin-heparin on activated protein C inactivation by protein C inhibitor.* J Biochem, 2010. **148**(2): p. 255-60.

95. Broze, G.J., Jr., et al., *The lipoprotein-associated coagulation inhibitor that inhibits the factor VII-tissue factor complex also inhibits factor Xa: insight into its possible mechanism of action.* Blood, 1988. **71**(2): p. 335-43.

96. Mann, K.G., S. Butenas, and K. Brummel, *The dynamics of thrombin formation.* Arterioscler Thromb Vasc Biol, 2003. **23**(1): p. 17-25.

97. Ishiguro, K., et al., *Complete antithrombin deficiency in mice results in embryonic lethality.* J Clin Invest, 2000. **106**(7): p. 873-8.

98. Huntington, J.A., R.J. Read, and R.W. Carrell, *Structure of a serpin-protease complex shows inhibition by deformation.* Nature, 2000. **407**(6806): p. 923-6.

99. Silverman, G.A., et al., *Serpins flex their muscle: I. Putting the clamps on proteolysis in diverse biological systems.* J Biol Chem, 2010. **285**(32): p. 24299-305.

100. Whisstock, J.C., et al., *Serpins flex their muscle: II. Structural insights into target peptidase recognition, polymerization, and transport functions.* J Biol Chem, 2010. **285**(32): p. 24307-12.

101. Olson, S.T., et al., *Molecular mechanisms of antithrombin-heparin regulation of blood clotting proteinases. A paradigm for understanding proteinase regulation by serpin family protein proteinase inhibitors.* Biochimie, 2010. **92**(11): p. 1587-96.

102. Whisstock, J.C., et al., *Conformational changes in serpins: II. The mechanism of activation of antithrombin by heparin.* J Mol Biol, 2000. **301**(5): p. 1287-305.

103. Lindahl, U., et al., *Extension and structural variability of the antithrombin-binding sequence in heparin.* J Biol Chem, 1984. **259**(20): p. 12368-76.

104. Streusand, V.J., et al., *Mechanism of acceleration of antithrombin-proteinase reactions by low affinity heparin. Role of the antithrombin binding pentasaccharide in heparin rate enhancement.* J Biol Chem, 1995. **270**(16): p. 9043-51.

105. Atha, D.H., et al., *Contribution of monosaccharide residues in heparin binding to antithrombin III.* Biochemistry, 1985. **24**(23): p. 6723-9.

106. Kuberan, B., et al., *Enzymatic synthesis of antithrombin III-binding heparan sulfate pentasaccharide.* Nat Biotechnol, 2003. **21**(11): p. 1343-6.

107. Al Dieri, R., et al., *The inhibition of blood coagulation by heparins of different molecular weight is caused by a common functional motif--the C-domain.* J Thromb Haemost, 2003. **1**(5): p. 907-14.

108. Gettins, P.G. and S.T. Olson, *Exosite determinants of serpin specificity.* J Biol Chem, 2009. **284**(31): p. 20441-5.

109. Givol, D. and A. Yayon, *Complexity of FGF receptors: genetic basis for structural diversity and functional specificity.* FASEB J, 1992. **6**(15): p. 3362-9.

110. Jaye, M., J. Schlessinger, and C.A. Dionne, *Fibroblast growth factor receptor tyrosine kinases: molecular analysis and signal transduction.* Biochim Biophys Acta, 1992. **1135**(2): p. 185-99.

111. Lee, P.L., et al., *Purification and complementary DNA cloning of a receptor for basic fibroblast growth factor.* Science, 1989. **245**(4913): p. 57-60.

112. Ornitz, D.M. and N. Itoh, *Fibroblast growth factors.* Genome Biol, 2001. **2**(3): p. REVIEWS3005.

113. Lin, X., et al., *Heparan sulfate proteoglycans are essential for FGF receptor signaling during Drosophila embryonic development.* Development, 1999. **126**(17): p. 3715-23.

114. Spivak-Kroizman, T., et al., *Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation, and cell proliferation.* Cell, 1994. **79**(6): p. 1015-24.

115. Yayon, A., et al., *Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor.* Cell, 1991. **64**(4): p. 841-8.

116. Schlessinger, J., et al., *Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization.* Mol Cell, 2000. **6**(3): p. 743-50.

117. Kreuger, J., et al., *Characterization of fibroblast growth factor 1 binding heparan sulfate domain.* Glycobiology, 1999. **9**(7): p. 723-9.

118. Sugaya, N., et al., *6-O-sulfation of heparan sulfate differentially regulates various fibroblast growth factor-dependent signalings in culture.* J Biol Chem, 2008. **283**(16): p. 10366-76.

119. Maccarana, M., B. Casu, and U. Lindahl, *Minimal sequence in heparin/heparan sulfate required for binding of basic fibroblast growth factor.* J Biol Chem, 1993. **268**(32): p. 23898-905.

120. Kreuger, J., et al., *Interactions between heparan sulfate and proteins: the concept of specificity.* J Cell Biol, 2006. **174**(3): p. 323-7.

121. Liu, D., et al., *Dynamic regulation of tumor growth and metastasis by heparan sulfate glycosaminoglycans.* Semin Thromb Hemost, 2002. **28**(1): p. 67-78.

122. Liu, D., et al., *Tumor cell surface heparan sulfate as cryptic promoters or inhibitors of tumor growth and metastasis.* Proc Natl Acad Sci U S A, 2002. **99**(2): p. 568-73.

123. Blackhall, F.H., et al., *Heparan sulfate proteoglycans and cancer.* Br J Cancer, 2001. **85**(8): p. 1094-8.

124. Sanderson, R.D., *Heparan sulfate proteoglycans in invasion and metastasis.* Semin Cell Dev Biol, 2001. **12**(2): p. 89-98.

125. Sasisekharan, R., et al., *Roles of heparan-sulphate glycosaminoglycans in cancer.* Nat Rev Cancer, 2002. **2**(7): p. 521-8.

126. Iozzo, R.V. and J.D. San Antonio, *Heparan sulfate proteoglycans: heavy hitters in the angiogenesis arena.* J Clin Invest, 2001. **108**(3): p. 349-55.

127. Varki, N.M. and A. Varki, *Heparin inhibition of selectin-mediated interactions during the hematogenous phase of carcinoma metastasis: rationale for clinical studies in humans.* Semin Thromb Hemost, 2002. **28**(1): p. 53-66.

128. Okada, Y., et al., *Structural recognition by recombinant human heparanase that plays critical roles in tumor metastasis. Hierarchical sulfate groups with different effects and the essential target disulfated trisaccharide sequence.* J Biol Chem, 2002. **277**(45): p. 42488-95.

129. Simizu, S., K. Ishida, and H. Osada, *Heparanase as a molecular target of cancer chemotherapy.* Cancer Sci, 2004. **95**(7): p. 553-8.

130. Elkin, M., et al., *Heparanase as mediator of angiogenesis: mode of action.* FASEB J, 2001. **15**(9): p. 1661-3.

131. De Larco, J.E., B.R. Wuertz, and L.T. Furcht, *The potential role of neutrophils in promoting the metastatic phenotype of tumors releasing interleukin-8.* Clin Cancer Res, 2004. **10**(15): p. 4895-900.

132. Waugh, D.J. and C. Wilson, *The interleukin-8 pathway in cancer.* Clin Cancer Res, 2008. **14**(21): p. 6735-41.

133. Li, R.W., et al., *Dramatic regulation of heparanase activity and angiogenesis gene expression in synovium from patients with rheumatoid arthritis.* Arthritis Rheum, 2008. **58**(6): p. 1590-600.

134. Waterman, M., et al., *Heparanase upregulation by colonic epithelium in inflammatory bowel disease.* Mod Pathol, 2007. **20**(1): p. 8-14.

135. Yang, Y., et al., *Heparanase enhances syndecan-1 shedding: a novel mechanism for stimulation of tumor growth and metastasis.* J Biol Chem, 2007. **282**(18): p. 13326-33.

136. Salek-Ardakani, S., et al., *Heparin and heparan sulfate bind interleukin-10 and modulate its activity.* Blood, 2000. **96**(5): p. 1879-88.

137. Spillmann, D., D. Witt, and U. Lindahl, *Defining the interleukin-8-binding domain of heparan sulfate.* J Biol Chem, 1998. **273**(25): p. 15487-93.

138. Witt, D.P. and A.D. Lander, *Differential binding of chemokines to glycosaminoglycan subpopulations.* Curr Biol, 1994. **4**(5): p. 394-400.

139. Massena, S., et al., *A chemotactic gradient sequestered on endothelial heparan sulfate induces directional intraluminal crawling of neutrophils.* Blood, 2010. **116**(11): p. 1924-31.

140. Wang, L., et al., *Heparin's anti-inflammatory effects require glucosamine 6-O-sulfation and are mediated by blockade of L- and P-selectins.* J Clin Invest, 2002. **110**(1): p. 127-36.

141. Butcher, E.C., *Leukocyte-endothelial cell recognition: three (or more) steps to specificity and diversity.* Cell, 1991. **67**(6): p. 1033-6.

142. Springer, T.A., *Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm.* Cell, 1994. **76**(2): p. 301-14.

143. Norgard-Sumnicht, K. and A. Varki, *Endothelial heparan sulfate proteoglycans that bind to L-selectin have glucosamine residues with unsubstituted amino groups.* J Biol Chem, 1995. **270**(20): p. 12012-24.

144. Luo, J., et al., *Heparan sulfate and chondroitin sulfate proteoglycans inhibit E-selectin binding to endothelial cells.* J Cell Biochem, 2001. **80**(4): p. 522-31.

145. Ahmed, T., J. Garrigo, and I. Danta, *Preventing bronchoconstriction in exercise-induced asthma with inhaled heparin.* N Engl J Med, 1993. **329**(2): p. 90-5.

146. Saliba, M.J., Jr., *Heparin in the treatment of burns: a review.* Burns, 2001. **27**(4): p. 349-58.

147. Torkvist, L., et al., *Low molecular weight heparin as adjuvant therapy in active ulcerative colitis.* Aliment Pharmacol Ther, 1999. **13**(10): p. 1323-8.

148. Tyagi, M., et al., *Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans.* J Biol Chem, 2001. **276**(5): p. 3254-61.

149. Garson, J.A., et al., *Suramin blocks hepatitis C binding to human hepatoma cells in vitro.* J Med Virol, 1999. **57**(3): p. 238-42.

150. Giroglou, T., et al., *Human papillomavirus infection requires cell surface heparan sulfate.* J Virol, 2001. **75**(3): p. 1565-70.

151. Shieh, M.T., et al., *Cell surface receptors for herpes simplex virus are heparan sulfate proteoglycans.* J Cell Biol, 1992. **116**(5): p. 1273-81.

152. Liu, J., et al., *Characterization of a heparan sulfate octasaccharide that binds to herpes simplex virus type 1 glycoprotein D.* J Biol Chem, 2002. **277**(36): p. 33456-67.

153. WuDunn, D. and P.G. Spear, *Initial interaction of herpes simplex virus with cells is binding to heparan sulfate.* J Virol, 1989. **63**(1): p. 52-8.

154. Herold, B.C., et al., *Glycoprotein C-independent binding of herpes simplex virus to cells requires cell surface heparan sulphate and glycoprotein B.* J Gen Virol, 1994. **75 (Pt 6)**: p. 1211-22.

155. Herold, B.C., et al., *Glycoprotein C of herpes simplex virus type 1 plays a principal role in the adsorption of virus to cells and in infectivity.* J Virol, 1991. **65**(3): p. 1090-8.

156. Trybala, E., et al., *Localization of a functional site on herpes simplex virus type 1 glycoprotein C involved in binding to cell surface heparan sulphate.* J Gen Virol, 1994. **75 (Pt 4)**: p. 743-52.

157. Feyzi, E., et al., *Structural requirement of heparan sulfate for interaction with herpes simplex virus type 1 virions and isolated glycoprotein C.* J Biol Chem, 1997. **272**(40): p. 24850-7.

158. Shukla, D. and P.G. Spear, *Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry.* J Clin Invest, 2001. **108**(4): p. 503-10.

159. Montgomery, R.I., et al., *Herpes simplex virus-1 entry into cells mediated by a novel member of the TNF/NGF receptor family.* Cell, 1996. **87**(3): p. 427-36.

160. Warner, M.S., et al., *A cell surface protein with herpesvirus entry activity (HveB) confers susceptibility to infection by mutants of herpes simplex virus type 1, herpes simplex virus type 2, and pseudorabies virus.* Virology, 1998. **246**(1): p. 179-89.

161. Krummenacher, C., et al., *Herpes simplex virus glycoprotein D can bind to poliovirus receptor-related protein 1 or herpesvirus entry mediator, two structurally unrelated mediators of virus entry.* J Virol, 1998. **72**(9): p. 7064-74.

162. Copeland, R., et al., *Using a 3-O-sulfated heparin octasaccharide to inhibit the entry of herpes simplex virus type 1.* Biochemistry, 2008. **47**(21): p. 5774-83.

163. Liu, R., et al., *Chemoenzymatic design of heparan sulfate oligosaccharides.* J Biol Chem, 2010. **285**(44): p. 34240-9.

164. Zhao, G., et al., *Enzymatic route to preparative-scale synthesis of UDP-GlcNAc/GalNAc, their analogues and GDP-fucose.* Nat Protoc, 2010. **5**(4): p. 636-46.

165. Liu, R., et al., *Chemoenzymatic design of heparan sulfate oligosaccharides.* J Biol Chem, 2010. **285**: p. 34240-34249.

166. Shively, J.E. and H.E. Conrad, *Formation of anhydrosugars in the chemical depolymerization of heparin.* Biochemistry, 1976. **15**: p. 3932-3942.

167. Liu, J., et al., *Expression of heparan sulfate D-glucosaminyl 3-O-sulfotransferase isoforms reveals novel substrate specificities.* J. Biol. Chem., 1999. **274**: p. 5185-5192.

168. Kirschner, K.N., et al., *GLYCAM06: a generalizable biomolecular force field. Carbohydrates.* J Comput Chem, 2008. **29**(4): p. 622-55.

169. Mahoney, M.W. and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions.* The Journal of Chemical Physics, 2000. **112**(20): p. 8910-8922.

170. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model.* The Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.

171. Case, D., et al., *Amber 14.* 2014.

172. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems.* The Journal of chemical physics, 1993. **98**(12): p. 10089-10092.

173. Roe, D.R. and T.E. Cheatham III, *PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data.* Journal of Chemical Theory and Computation, 2013. **9**(7): p. 3084-3095.

174. Haasnoot, C., F.A. de Leeuw, and C. Altona, *The relationship between proton-proton NMR coupling constants and substituent electronegativities—I: an empirical generalization of the Karplus equation.* Tetrahedron, 1980. **36**(19): p. 2783-2792.

175. Altona, C., et al., *Empirical group electronegativities for vicinal NMR proton–proton couplings along a C□ C bond: Solvent effects and reparameterization of the Haasnoot equation.* Magnetic resonance in chemistry, 1994. **32**(11): p. 670-678.

176. Zhang, L., et al., *6-O-sulfotransferase-1 represents a critical enzyme in the anticoagulant heparan sulfate biosynthetic pathway.* J. Biol. Chem., 2001. **276**: p. 42311-42321.

177. Duncan, M.B., et al., *The biosynthesis of anticoagulant heparan sulphate by the heparan sulphate 3-O-sulphotransferase isoform 5.* Biochim Biophys Acta, 2004. **1671**: p. 34-43.

178. Lee, M.K. and A.D. Lander, *Analysis of affinity and structural selectivity in the binding of proteins to glycosaminoglycans: development of a sensitive electrophoretic approach.* Proc Natl Acad Sci U S A, 1991. **88**(7): p. 2768-72.

179. Kahn, S.R., et al., *Prevention of VTE in nonsurgical patients* Chest, 2012. **141 (suppl)**: p. e195S-e226S.

180. Hirsh, J., M.O. O'Donnell, and J.W. Eikelboom, *Beyond unfractionated heparin and warfarin: current and future advances.* Circulation, 2007. **116**: p. 552-560.

181. Arepally, G. and T.L. Ortel, *Heparin-induced thrombocytopenia.* N. Eng. J. Med., 2006. **355**: p. 809-817.

182. Linhardt , R.J. and J. Liu, *Synthetic heparin.* Curr Opin Pharmacol, 2012. **12**: p. 217-219.

183. Melnikova, I., *The anticoagulants market.* Nat Rev Drug Discov., 2009. **8**: p. 353-354.

184. Harder, S., *Renal profiles of anticoagulants.* J. Clin. Pharmacol., 2012. **52**: p. 964-975.

185. Weitz, J.I. and L.A. Linkins, *Beyond heparin and warfrarin: the new generation of anticoagulants.* Expert Opin. Investig. Drugs, 2007. **16**: p. 271-282.

186. Petitou, M. and C.A.A. van Boeckel, *A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next?* Angew. Chem. Int. Ed., 2004. **43**: p. 3118-3133.

187. Moon, A.F., et al., *Dissecting substrate recognition mode of heparan sulfate 3-O-sulfotransferases for the biosynthesis of anticoagulant heparin.* Proc. Natl. Acad. Sci. USA, 2012. **109**: p. 5256-5270.

188. Chen, J., C.L. Jones, and J. Liu, *Using an enzymatic combinatorial approach to identify anticoagulant heparan sulfate structures.* Chem Biol, 2007. **14**: p. 986-993.

189. Zhang, Z., et al., *Solution structure of chemoenzymatically synthesized heparin and its precursors.* J Am Chem Soc, 2008. **130**: p. 12998-13007.

190. Xu, Y., et al., *Chemoenzymatic synthesis of homogeneous ultra-low molecular weight heparin.* Science, 2011. **334**: p. 498-501.

191. Zhang, Z., et al., *Oversulfated chondroitin sulfate: impact of a heparin impurity, associated with adverse clinical events, on low-molecular-weight heparin preparation.* J. Med. Chem., 2008. **51**: p. 5498-5501.

192. Turnbull, J., *Getting the farm out of pharma for heparin production.* Science, 2011. **334**: p. 462-463.

193. Liu, H., Z. Zhang, and R.J. Linhardt, *Lessons learned from the contamination of heparin.* Nat Prod Rep, 2009. **26**: p. 313-321.

194. Sheng, J., et al., *Uncovering biphasic catalytic mode of C5-epimerase in heparan sulfate biosynthesis.* J. Biol. Chem., 2012. **287**: p. 20996-21002.

195. Pempe, E., et al., *Probing the Structural Selectivity of Synthetic Heparin Binding to the Stabilin Receptors.* J. Biol. Chem., 2012. **287**: p. 20774-20783.

196. Li, L., et al., *Top-down approach for the direct characterization of low molecular weight heparins using LC-FT-MS.* Anal. Chem., 2012. **84**: p. 8822-8829.

197. Martinez-Gonzalez, J. and C. Rodriguez, *New challenges for a second-generation low-molecular-weight heparin: focus on bemiparin.* Expert Rev. Cardiovasc. Ther., 2010. **8**: p. 625-634.

198. Lu, G., et al., *A specific antidote for reversal of anticoagulation by direct and indirect inhibitors of coagulation factor Xa.* Nat. Med., 2013. **19**: p. 446-451.

199. Bianchini, E.P., et al., *Development of a recombinant antithrombin variant as a potent antidote to fondaparinux and other heparin derivatives* Blood, 2011. **117**: p. 2054-2060.

200. Atha, D.H., et al., *Contribution of monosaccharide residues in heparin binding to antithrombin III.* Biochemistry, 1985. **24**: p. 6723-6729.

201. Lee, S., et al., *Scientific considerations in the review and approval of generic enoxaparin in the United States.* Nat. Biotechnol., 2013. **31**: p. 220-226.

202. Petitou, M., et al., *Process for the organic synthesis of oligosaccharides and derivatives thereof.* 1989.

203. Carlsson, P., et al., *Heparin/heparan sulfate biosynthesis: processive formation of N-sulfated domains.* J Biol Chem, 2008. **283**(29): p. 20008-14.

204. Victor, X.V., et al., *Investigating the elusive mechanism of glycosaminoglycan biosynthesis.* J Biol Chem, 2009. **284**(38): p. 25842-53.

205. Petitou, M. and C.A. van Boeckel, *Chemical synthesis of heparin fragments and analogues.* Fortschr Chem Org Naturst, 1992. **60**: p. 143-210.

206. Sattelle, B.M., et al., *Free energy landscapes of iduronic acid and related monosaccharides.* J Am Chem Soc, 2010. **132**(38): p. 13132-4.

207. Langeslay, D.J., et al., *Sulfamate proton solvent exchange in heparin oligosaccharides: evidence for a persistent hydrogen bond in the antithrombin-binding pentasaccharide Arixtra.* Glycobiology, 2012. **22**(9): p. 1173-82.

208. Guerrini, M., et al., *An unusual antithrombin-binding heparin octasaccharide with an additional 3-O-sulfated glucosamine in the active pentasaccharide sequence.* Biochem J, 2013. **449**(2): p. 343-51.

209. Das, S.K., et al., *Synthesis of conformationally locked L-iduronic acid derivatives: direct evidence for a critical role of the skew-boat 2S0 conformer in the activation of antithrombin by heparin.* Chemistry, 2001. **7**(22): p. 4821-34.

210. Hricovini, M., et al., *Conformation of heparin pentasaccharide bound to antithrombin III.* Biochem J, 2001. **359**(Pt 2): p. 265-72.

211. Guerrini, M., et al., *Effects on molecular conformation and anticoagulant activities of 1,6-anhydrosugars at the reducing terminal of antithrombin-binding octasaccharides isolated from low-molecular-weight heparin enoxaparin.* J Med Chem, 2010. **53**(22): p. 8030-40.

212. Guerrini, M., et al., *Antithrombin-binding octasaccharides and role of extensions of the active pentasaccharide sequence in the specificity and strength of interaction. Evidence for very high affinity induced by an unusual glucuronic acid residue.* J Biol Chem, 2008. **283**(39): p. 26662-75.

213. DiGabriele, A.D., et al., *Structure of a heparin-linked biologically active dimer of fibroblast growth factor.* Nature, 1998. **393**(6687): p. 812-7.

214. Faham, S., et al., *Heparin structure and interactions with basic fibroblast growth factor.* Science, 1996. **271**(5252): p. 1116-20.

215. Kreuger, J., et al., *Sequence analysis of heparan sulfate epitopes with graded affinities for fibroblast growth factors 1 and 2.* J Biol Chem, 2001. **276**(33): p. 30744-52.

216. Zhang, Z., et al., *Solution structures of chemoenzymatically synthesized heparin and its precursors.* J Am Chem Soc, 2008. **130**(39): p. 12998-3007.

217. Guglier, S., et al., *Minimum FGF2 binding structural requirements of heparin and heparan sulfate oligosaccharides as determined by NMR spectroscopy.* Biochemistry, 2008. **47**(52): p. 13862-9.

218. Limtiaco, J.F., et al., *The efficient structure elucidation of minor components in heparin digests using microcoil NMR.* Carbohydr Res, 2011. **346**(14): p. 2244-54.

219. Ozug, J., et al., *Structural elucidation of the tetrasaccharide pool in enoxaparin sodium.* Anal Bioanal Chem, 2012. **403**(9): p. 2733-44.

220. Yamada, S., et al., *Isolation of the porcine heparin tetrasaccharides with glucuronate 2-O-sulfate. Heparinase cleaves glucuronate 2-O-sulfate-containing disaccharides in highly sulfated blocks in heparin.* J Biol Chem, 1995. **270**(15): p. 8696-705.

221. Yamada, S., et al., *Isolation of the Porcine Heparin Tetrasaccharides with Glucuronate 2-O-Sulfate - Heparinase Cleaves Glucuronate 2-O-Sulfate-Containing Disaccharides in Highly Sulfated Blocks in Heparin.* Journal of Biological Chemistry, 1995. **270**(15): p. 8696-8705.

222. Keire, D.A., L.F. Buhse, and A. Al-Hakim, *Characterization of currently marketed heparin products: composition analysis by 2D-NMR.* Analytical Methods, 2013. **5**(12): p. 2984-2994.

223. Kreuger, J., et al., *Sequence Analysis of Heparan Sulfate Epitopes with Graded Affinities for Fibroblast Growth Factors 1 and 2.* J. Biol. Chem., 2001. **276**: p. 30744-30752.

224. Witt, R.M., et al., *Heparan sulfate proteoglycans containing a glycpican 5 core and 2-O-sulfo-iduronic acid function as sonic hedgehog co-receptors to promote proliferation.* J. Biol. Chem., 2013: p. M112.438937.

225. Moon, A., et al., *Structural analysis of the sulfotransferase (3-OST-3) involved in the biosynthesis of an entry receptor of herpes simplex virus 1.* J. Biol. Chem., 2004. **279**: p. 45185-45193.

226. Patel, V.N., et al., *3-O-sulfated heparan sulfate controls epithelial Kit+FGFR2b+ progenitor expansion via rapid autocrine regulation of 3-O-sulfotransferases.* Dev Cell, 2014. **29**: p. 662-673.

227. Bishop, J., M. Schuksz, and J.D. Esko, *Heparan sulphate proteoglycans fine-tune mammalian physiology.* Nature, 2007. **446**: p. 1030-1037.

228. Liu, J. and R.J. Linhardt *Chemoenzymatic synthesis of heparan sulfate and heparin.* Nat Prod Rep, 2014. **31**: p. 1676-1685.

229. Liu, J. and S.C. Thorp, *Heparan sulfate and the roles in assisting viral infections.* Med. Res. Rev., 2002. **22**: p. 1-25.

230. Xu, D. and J. Esko, *Demystifying heparan sulfate-protein interactions.* Annu Rev Biochem. , 2014. **83**: p. 129-157.

231. Kreuger, J., et al., *Interactions between heparan sulfate and proteins: the concept of specificity.* J. Cell Biol., 2006. **174**: p. 323-327.

232. Powers, J.C., et al., *Irreversible inhibitors of serine, cysteine, and threonine proteases.* Chem Rev, 2002. **102**: p. 4639-4750.

233. Ouihia, A., et al., *A new diazoacylating reagent: Preparation, structure, and use of succinimidyl diazoacetate.* J Org Chem, 1992. **58**: p. 1641-1642.

234. Wentrup, C., et al., *Matrix-IR spectroscopic investigations of the thermolysis and photolysis of diazoamides.* J. Org. Chem., 2013. **78**: p. 10705-10717.

235. Chaimovich, H., R.J. Vaughan, and F.H. Westheimer, *Rearrangement of accompanying the phtolysis of diazoacyl esters.* J. Am. Chem. Soc., 1968. **90**: p. 4088-4093.

236. Liu, J., et al., *Understanding the substrate specificity of the heparan sulfate sulfotransferases by a combined synthetic crystallographic approach.* Curr Opin Struct Biol 2012. **22**: p. 550-557.

237. Liu, C., et al., *Molecular mechanism of substrate specificity for heparan sulfate 2-O-sulfotransferase.* J. Biol. Chem., 2014. **289**: p. 13407-13418.

238. Casu, B., A. Naggi, and G. Torri, *Heparin-derived heparan sulfate mimics to modulate heparan sulfate-protein interaction in inflammation and cancer.* Matrix Biol, 2010. **29**: p. 442-452.

239. Barb, A.W. and J.H. Prestegard, *NMR analysis demosntrates immunoglobulin G N-glycans are accessible and dynamic.* Nat Chem Biol, 2011. **7**: p. 147-153.

240. Guerrini, M., et al., *Antithrombin-binding oligosaccharides: structural diversities in a unique function?* Glycoconj. J., 2014. **31**: p. 409-416.

241. Toone, E.J., E.S. Simon, and G.M. Whitesides, *Enzymatic synthesis of uridine 5'-diphosphoglucuronic acid on a gram scale.* J. Org. Chem., 1991. **56**: p. 5603-5606.

242. Zhao, G., et al., *Enzymatic route to preparative-scale synthesis of UDP-GlcNAc/GalNAc, their analogues and GDP-fucose.* Nat. Protoc., 2010. **5**: p. 636-646.

243. Johnson, D.J., et al., *Antithrombin-S195A factor Xa-heparin structure reveals the allosteric mechanism of antithrombin activation.* EMBO J., 2006. **25**: p. 2029-2037.

244. Hricovíni, M., et al., *Conformation of heparin pentasaccharide bound to antithrombin III.* Biochem. J., 2001. **359**: p. 265-272.

245. Munoz-Garcia, J.C., et al., *Conformations of the iduronate ring in short heparin fragments described by time-averaged distance restrained molecular dynamics.* Glycobiology, 2013. **23**(11): p. 1220-9.

246. Sattelle, B.M., et al., *Dependence of pyranose ring puckering on anomeric configuration: methyl idopyranosides.* J. Phys Chem B, 2012. **116**: p. 6380-6386.

247. DeMarco, M.L. and R.J. Woods, *Atomic-resolution conformational analysis of the GM3 ganglioside in a lipid bilayer and its implications for ganglioside-protein recognition at membrane surfaces.* Glycobiology, 2009. **19**: p. 344-355.

248. Yongye, A.B., et al., *The conformational properties of methyl alpha-(2,8)-di/trisialosides and their N-acyl analogues: implications for anti-Neisseria meningitidis B vaccine design.* Biochemistry, 2008. **47**: p. 12493-12514.

249. Yongye, A.B., B.L. Foley, and R.J. Woods, *On achieving experimental accuracy from molecular dynamics simulations of flexible molecules: aqueous glycerol.* J. Phys. Chem. A, 2008. **112**: p. 2634-2639.

250. Gonzalez-Outeiriño, J., et al., *Reconciling solvent effects on rotamer populations in carbohydrates — A joint MD and NMR analysis.* Can. J. Chem., 2006. **84**: p. 569-579.

251. Casu, B., et al., *Conformational flexibility: a new concept for explaining binding and biological properties of iduronic acid-containing glycosaminoglycans.* Trends Biochem Sci, 1988. **13**(6): p. 221-5.

252. Trowbridge, J.M. and R.L. Gallo, *Dermatan sulfate: new functions from an old glycosaminoglycan.* Glycobiology, 2002. **12**(9): p. 117R-25R.

"It is man's duty to be courageous enough to seek for truth." — Nicolaus Copernicus