

**INSIGHT FROM NMR OBSERVATIONS ON GLYCOSAMINOGLYCAN  
INTERACTIONS WITH THE LINK MODULE OF TSG-6**

by

YOUNGHEE PARK

(Under the Direction of James H. Prestegard)

**ABSTRACT**

Tumor necrosis factor stimulated gene-6 (TSG-6) is a glycosaminoglycan (GAG) binding protein expressed in inflammatory diseases and inflammatory-like processes. The Link module, one domain of TSG-6, is well known for its interactions with hyaluronan (HA), which play important roles in protecting against cartilage matrix destruction during inflammation and stabilizing the HA-rich extracellular matrix (ECM) during ovulation. Interactions with HA, and to some extent heparin/heparan sulfate (Hp/HS), have been structurally studied, indicating that HA and Hp/HS interact with the Link module at totally different binding sites. However, its interactions with chondroitin sulfate (CS) have yet to be adequately described. In this thesis, we isolated a well-defined homogeneous hexasaccharides of CS using enzymatic digestions followed by chromatographic separation and determined their structures using Nuclear Magnetic Resonance (NMR) methods. A spin-labeled analog of a CS hexasaccharide was synthesized as well. These, in turn, were used in further binding studies with the Link module. The interactions between the Link module and CSs were monitored by NMR

chemical shift perturbation (CSP), suggesting possible CS binding sites and multiple binding modes for the Link module. The production of a Link dimer during the binding process was confirmed by measuring relaxation rates of  $^{15}\text{N}$  sites in the Link module as well as by dynamic light scattering (DLS) measurements. Paramagnetic relaxation enhancement (PRE) data from  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectra of the Link module in the absence and presence of the spin labeled analog of CS added to binding site definition. A solution structure of the induced dimer of the Link module was produced, using triple resonance assignment methods for NMR resonance assignments and Nuclear Overhauser Effect (NOE) and Residual Dipolar Coupling (RDC) data for structural constraints. Finally, a dimer model of Link module with the 4,4,4 sulfated, TEMPO adduct of a CS hexasaccharide bound was generated using CSP, PRE, and RDC restraints in docking program (HADDOCK). This provides insights into the cross-linking of CS in the ECM that can play an additional role in the remodeling and stabilization of the ECM during inflammation.

**INDEX WORDS:** TSG-6; Link module; Glycosaminoglycans; Chondroitins sulfate; NMR spectroscopy; Size-exclusion chromatography; Strong anion exchange chromatography; Mass spectrometry; Triple resonance assignment; Residual dipolar coupling; Spin relaxation; Chemical shift perturbation; Paramagnetic relaxation enhancement; TEMPO; HADDOCK; Protein-ligand docking

**INSIGHT FROM NMR OBSERVATIONS ON GLYCOSAMINOGLYCAN  
INTERACTIONS WITH THE LINK MODULE OF TSG-6**

by

**YOUNGHEE PARK**

B.S., Konkuk University, South Korea, 2005

M.S., Konkuk University, South Korea, 2008

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

**DOCTOR OF PHILOSOPHY**

**ATHENS, GEORGIA**

2015

© 2015

YOUNGHEE PARK

All Rights Reserved

**INSIGHT FROM NMR OBSERVATIONS ON GLYCOSAMINOGLYCAN  
INTERACTIONS WITH THE LINK MODULE OF TSG-6**

by

YOUNGHEE PARK

Major Professor: James H. Prestegard

Committee: Robert J. Woods  
Jeffrey L. Urbauer

Electronic Version Approved:

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

## **DEDICATION**

To God for His unconditional love and grace on me in my entire life. To my dear father, Tae-Rak Park who has trusted in me and supported me with his love and prayer. To my dear mother, Kyoung-Ok Nam who is my daily energy and my best friend in my entire life. To my elder brother, Dong-Jin Park for his enduring love and support.

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my major advisor, Professor James Prestegard. I could not fulfill this long academic journey without his guidance and supervision. I will always be grateful for the academic and research opportunities his lab afforded me and his enduring advice throughout my graduate school career. I consider myself incredibly fortunate to have worked with such a talented scientist and generous person. I also would like to thank my committee members, Dr. Robert Woods and Dr. Jeffrey Urbauer for supporting me and sharing their knowledge and time.

I thank Dr. Vitor Pomin for many discussions related to glycosaminoglycans and for co-authoring a manuscript and book chapter. I especially thank Iris Tropp for proof-reading my thesis even during her travels, encouraging me to stick to a schedule, and for her sincere care and concern. I thank Dr. John Glushka for his technical assistance, discussions about NMR concepts, experiments and pulse sequences, and sharing his coffee and time. I thank Laura Morris for her technical assistance with hardware and software issues that came up over the years and for always keeping her door open to me. I thank Dr. Xu Wang for his assistance setting up bench space in the lab and for early discussions of NMR concepts. I thank Dr. Charles Mobley for his thought provoking discussions and for the memorable moments that we have shared through the years. I thank Dr. Kari Pederson for helping me learn NMR experiments and for her explanations about generating NMR structures from specific calculations. I thank Qi Gao for being a great peer in the lab and consistently providing comfort and friendship when needed.

Many thanks to all of the Prestegard group members: Dr. David Live, Dr. Khan Hekmatyar, Dr. Cheng Yu Chen, Dr. You Zhuo, and Josh Franks for being fantastic peers in the lab, and creating a great work environment.

I especially thank my close friends (Hanna Moon, Sungeun Cho, Nahray Han, Isaac Lee, and Jaeyun Sung) for their prayers and support. I have enjoyed every single moment of life with you in Athens – you are invaluable to me and will always be in my heart. I experienced the great love of Christ through my church. I believe all prayers you made for me were lifted up to God throughout this journey. Thank you to all of my friends in Athens for your prayers, encouraging words, and enjoyable attitude. Thank you to my lovely family: Tae-Rak Park, Kyoung-Ok Nam, and Dong-Jin Park. Without your support and concern, I could not have prevailed. Thank you for always encouraging me, believing in me, and reassuring me that this was possible.



## TABLE OF CONTENTS

	Page
<b>ACKNOWLEDGEMENTS</b> .....	v
<b>LIST OF TABLES</b> .....	x
<b>LIST OF FIGURES</b> .....	xi
<b>CHAPTER</b>	
<b>1 INTRODUCTION AND LITERATURE REVIEW</b> .....	1
1.1 Overview .....	1
1.2 Interaction between GAGs and GAG binding proteins .....	3
1.3 The Link module as a GAG binding protein.....	8
1.4 Research questions.....	12
1.5 References .....	13
<b>2 PREPARATION OF A NITROXIDE SPIN-LABELED ANALOG OF CHONDROITIN SULFATE</b> .....	23
2.1 Introduction .....	23
2.2 Methodology for creating $\Delta$ C444S-TEMPO .....	27
2.3 Results.....	29
2.4 Discussion .....	37
2.5 References .....	38
<b>3 NMR RESONANCE ASSIGNMENTS OF UNIFORMLY LABELED LINK MODULE OF TSG-6</b> .....	41

3.1	Introduction .....	41
3.2	Experimental Procedures .....	48
3.3	Results.....	51
3.4	Discussion .....	70
3.5	References .....	71
<b>4</b>	<b>INTERACTION OF LINK MODULE OF HUMAN TSG-6 AND</b>	
	<b>CHONDROITIN SULFATES .....</b>	<b>74</b>
4.1	Abstract .....	75
4.2	Introduction .....	76
4.3	Results.....	80
4.4	Discussion .....	100
4.5	Methods.....	104
4.6	References .....	110
<b>5</b>	<b>CONCLUDING REMARKS .....</b>	<b>119</b>
5.1	References .....	121

## **APPENDICES**

<b>A</b>	<b>EXPLOITING ENZYME SPECIFICITIES IN DIGESTIONS OF</b>	
	<b>CHONDROITIN SULFATES A AND C: PRODUCTION OF WELL-</b>	
	<b>DEFINED HEXASACCHARIDES .....</b>	<b>124</b>
A.1	Abstract .....	125
A.2	Introduction .....	126
A.3	Results.....	130
A.4	Discussion .....	146

A.5	Materials and methods .....	150
A.6	References .....	155

## LIST OF TABLES

	Page
<b>Table 1.1:</b> Different classes of glycosaminoglycans and their repeating units .....	4
<b>Table 1.2:</b> Biological relevance of GAG-protein interactions .....	7
<b>Table 3.1:</b> CA and CB chemical shifts for random coil backbone .....	54
<b>Table 3.2:</b> RDCs of the Link-TSG-6:ΔC444S complex aligned in neutral gel measured in IPAP-HSQC experiment .....	63
<b>Table 3.3:</b> Dihedral angles of Link-TSG-6: ΔC444S complex .....	65
<b>Table 3.4:</b> NMR and structural statistics for the Link module of human TSG-6.....	68
<b>Table 4.1:</b> Estimated distance of resonances experiencing reduction of peak intensities	88
<b>Table 4.2:</b> NMR and structural statistics for human Link-TSG-6.....	92
<b>Table A1:</b> Structure and respective abbreviations used for disaccharides and hexasaccharides obtained from six different digestions.....	129
<b>Table A2:</b> Yields of the disaccharides and the major hexasaccharides obtained from different digestion types within different time courses .....	131

## LIST OF FIGURES

	Page
<b>Figure 1.1:</b> Full length TSG-6 showing the signal peptide (1-19), N-terminal sequence (20-35), Link module (36-133), and CUB module (134-248) followed by C-terminal region (249-277) .....	9
<b>Figure 1.2:</b> 3D structures of Link-TSG-6 determined by x-ray crystallography (A) and NMR spectroscopy (B) .....	11
<b>Figure 2.1:</b> The structure of 4-amino-TEMPO (amino-2,2,6,6-tetramethylpiperidine-1-oxyl).....	25
<b>Figure 2.2:</b> SEC (A) and SAX (B) profiles of chondroitin sulfate hexasaccharides after chondroitinase AC digestion with btCS-A .....	30
<b>Figure 2.3:</b> $^1\text{H}$ - $^{13}\text{C}$ HSQC spectrum of $\Delta\text{C444S}$ after SEC and SAX separations from chondroitinase AC digestion .....	31
<b>Figure 2.4:</b> Size-exclusion (A) and SAX chromatography (B) analyses of the products from chondroitinase C digestion of btCS-A.....	32
<b>Figure 2.5:</b> Reductive amination of $\Delta\text{C444S}$ with 4-amino-TEMPO .....	33
<b>Figure 2.6:</b> SAX-HPLC chromatograph (A) and Mass spectrum (B) of $\Delta\text{C444S}$ -TEMPO .....	34
<b>Figure 2.7:</b> 1D proton spectrum of $\Delta\text{C444S}$ (upper panel) and $\Delta\text{C444S}$ -TEMPO (lower panel).....	36
<b>Figure 2.8:</b> $^1\text{H}$ - $^{13}\text{C}$ HSQC spectra of $\Delta\text{C444S}$ (A) and $\Delta\text{C444S}$ -TEMPO (B) .....	37

<b>Figure 3.1:</b> Simplified schematic of 3D NMR spectroscopy .....	42
<b>Figure 3.2:</b> Dipolar coupled spin pair .....	46
<b>Figure 3.3:</b> $^1\text{H}$ - $^{15}\text{N}$ HSQC spectrum of the Link module of TSG-6 in its $\Delta\text{C444S}$ -bound state.....	51
<b>Figure 3.4:</b> Backbone sequential assignment of the Link module of TSG-6 using 3D experiments of CBCA(CO)NH (magenta) and HNCACB (cyan) (A), and strip plot of CBCA(CO)NH, HNCACB (B) for the residues from V2 to K13 .....	53
<b>Figure 3.5:</b> 3D HNCO experiment .....	56
<b>Figure 3.6:</b> Strip plot of 3D HNCO experiment .....	57
<b>Figure 3.7:</b> Assignment of side chain using HCCH-TOCSY .....	58
<b>Figure 3.8:</b> General scheme of NOEs observed in $^1\text{H}$ - $^{13}\text{C}$ NOESY-HSQC (A) and $^1\text{H}$ - $^{15}\text{N}$ NOESY-HSQC (B) experiments .....	59
<b>Figure 3.9:</b> Strips from $^1\text{H}$ - $^{13}\text{C}$ HSQC-NOESY (first three strip plots from the left) and $^1\text{H}$ - $^{15}\text{N}$ HSQC-NOESY (strip plot, far right) for V2 of the Link-TSG-6: $\Delta\text{C444S}$ complex .....	60
<b>Figure 3.10:</b> Two-stage NMR tube for RDC experiments.....	61
<b>Figure 3.11:</b> Spectra of IPAP-HSQC experiments for isotropic (A) and oriented (B) samples .....	62
<b>Figure 3.12:</b> RDC fitting onto Link module in HA bound state (1O7C.pdb).....	64
<b>Figure 3.13:</b> 20 superimposed models for the NMR structures of the Link-TSG-6 in its $\Delta\text{C444S}$ bound state.....	67
<b>Figure 3.14:</b> Superimposition of NMR structures in free (1O7B.pdb, blue), HA bound (1O7C.pdb, pink), and CS bound (gray) states .....	69

<b>Figure 3.15:</b> Structure comparisons among NMR structures in free (1O7B.pdb, blue), HA bound (1O7C.pdb, pink), and CS bound (gray) states .....	70
<b>Figure 4.1:</b> Superimposed $^1\text{H}$ - $^{15}\text{N}$ HSQC spectra of the Link module titrated with $\Delta\text{C444S}$ compared to concentration of protein: black-control, without ligand; blue-40% ligand; cyan-80%; green-100%; magenta-150%; red-200% .....	84
<b>Figure 4.2:</b> Fitting for intensity of residues in slow exchange (A). Fitting of chemical shift for residues in fast exchange (B).....	84
<b>Figure 4.3:</b> Chemical shift differences ( $\Delta\delta$ ) of CA (black bar) and CB (red bar) between free and ligand-bound Link-TSG-6 plotted as a function of residue number (A) and mapping of shift perturbations (B) .....	85
<b>Figure 4.4:</b> Chemical shift perturbation ( $\Delta\delta$ ) of Link module caused by $\Delta\text{C444S}$ relative to free protein.....	85
<b>Figure 4.5:</b> Mapping fast (red), intermediate (green), and slow (blue) exchanging resonances on the existing X-ray structure of the human Link-TSG-6 (2PF5.pdb).. .....	86
<b>Figure 4.6:</b> Expanded $^1\text{H}$ - $^{15}\text{N}$ HSQC spectra of Link-TSG-6 in the presence of TEMPO Analog.....	88
<b>Figure 4.7:</b> $R_1$ , $R_2$ , and $R_2/R_1$ data for free ( $\Delta$ ) and $\Delta\text{C444S}$ -bound ( $\bullet$ ) Link-TSG-6 at 800MHz, 25 °C .....	90
<b>Figure 4.8:</b> Superimposition of 20 soluble structures of Link-TSG-6 in the presence of $\Delta\text{C444S}$ hexasaccharide .....	93
<b>Figure 4.9:</b> A dimer model of Link-TSG-6.....	96
<b>Figure 4.10:</b> Complex models for the dimer of Link module and $\Delta\text{C444S}$ -TEMPO .....	98

<b>Figure 4.11:</b> Model of the Link-TSG-6 dimer and the chondroitin sulfate hexasaccharide TEMPO, $\Delta$ C444S-TEMPO .....	99
<b>Figure 4.12:</b> Weak binding site resulting in $\Delta$ C444S binding onto Link-TSG-6 .....	100
<b>Figure 4.13:</b> Superimposition of structures of Link-TSG-6 in its HA and $\Delta$ C444S bound states.....	100
<b>Figure A1:</b> Size fractionation on Bio-Gel P-10 column of the products from btCS-A digested with a commercial preparation of chondroitinase ABC from <i>P. vulgaris</i> ...	133
<b>Figure A2:</b> SAX-HPLC fractionation of unsaturated hexasaccharides from digestion of (A) btCS-A, and (B) scCS-C with a commercial preparation of chondroitinase ABC from <i>P. vulgaris</i> .....	136
<b>Figure A3:</b> Size fractionation on Bio-Gel P-10 column of the products from 2-day digestions of scCS-C (-○-), and btCS-A (-●-) with a commercial preparation of chondroitinase C from <i>F. heparinum</i> .....	140
<b>Figure A4:</b> SAX-HPLC fractionation of reduced unsaturated hexasaccharides from 2-day digestions of (A) scCS-C, and (B) btCS-A with a commercial preparation of C lyase from <i>F. heparinum</i> .....	141
<b>Figure A5:</b> Size fractionation on Bio-Gel P-10 column of the products from 2-day digestion of btCS-A (-○-), and scCS-C (-●-) with a commercial preparation of ovine hyaluronidase .....	143
<b>Figure A6:</b> SAX-HPLC fractionation of reduced saturated hexasaccharides from 2-day digestion of (A) btCS-A, and (B) scCS-C with a commercial preparation of ovine hyaluronidase .....	145



## **CHAPTER 1**

### **INTRODUCTION AND LITERATURE REVIEW**

#### **1.1 Overview**

This dissertation is mostly about a protein structure determination of our target protein in its ligand bound state in solution. To determine the structure, we have relied on NMR methodology and utilize other techniques as necessary.

TSG-6 (Tumor necrosis factor stimulated gene-6) is secreted as a 35 kDa glycoprotein and its expression is found at the site of inflammation and injury (1-3). It consists mainly of two domains, Link module and CUB (Complement subcomponents C1r/C1s, Uegf, BMP-1) module. It interacts with various proteins and ligands, such as inter- $\alpha$ -inhibitor, thrombospondin-1, glycosaminoglycans (GAGs), which has important roles in inflammation related diseases and ovulation (4-8). These bind to the Link module of TSG-6 that is also classified as a hyaladherin responsible for HA-binding. Most previous structural and biological studies of the Link module of human TSG-6 have been conducted using hyaluronan (HA) and heparin/heparan sulfate (Hp/HS). Chondroitin sulfate (CS) is uncommonly used though *in vitro* data suggest that the Link module binds to 4-sulfated CS better than that of 6-sulfated isomers (7). Moreover, early studies of the interactions between CS and the Link module demonstrate that 4-sulfated CS isomers could bind to the Link module with a similar affinity to that of HA at pH 5.8 (7). However, later studies showed that 4- and 4,6- sulfated chondroitin sulfates have affinities for TSG-6 similar to HA for interacting with TSG-6 under physiological

conditions (9). 6-sulfated chondroitin sulfate can bind to the protein, but the binding affinity is slightly lower than 4- and 4,6-sulfated chondroitin sulfates (9). This has significant physiological implications since CSs, among other roles, are critical structural components of cartilage. Here they often occur as chains covalently linked to the proteoglycan, aggrecan. The proteolytic attack of aggrecan results in cartilage erosion and promotes arthritis. Mouse models demonstrate that TSG-6 produces a chondroprotective effect during inflammation (10,11). In addition, experimental models of arthritis indicate that CS may be responsible for localization in cartilage and may directly inhibit aggrecan degradation (9). Even though competition binding assays between HA and CS suggest that CSs may bind to the Link module at the HA binding surface (7), detailed structural information about the interaction of the Link module with CSs does not exist. This is a significant gap since CSs play significant roles in the stabilization of ECM during inflammation.

This study describes a method for optimizing the production of a homogeneous target CS using enzyme degradation of polysaccharides of chondroitin sulfate-A. Enzyme treated samples were subjected to size-exclusion chromatography (SEC) followed by strong anion exchange (SAX) chromatography with heterogeneous hexasaccharides of CS separated by SEC. Several homogeneous CS isomers were separated and their structures were assigned using NMR experiments (COSY, TOCSY, NOESY, and  $^{13}\text{C}$ -HSQC). Detailed experimental procedures and data are described in Chapter 2. Our studies confirm that the lyase product,  $\Delta$ chondroitin 4,4,4 sulfated CS ( $\Delta\text{C444S}$ ), where  $\Delta$  indicate the unsaturation left at the cleavage site, binds to the Link module with greater affinity than other isomers and that binding sites for CS overlap with

those of HA. Our studies also indicate that CS may induce dimerization of the protein and that the interface of the dimer could be part of the heparin binding surface. Given these two binding tendencies of the Link module, it is possible that the Link module has multiple modes of binding with CS under physiological conditions.

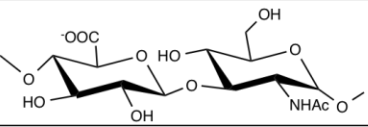
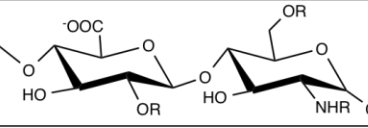
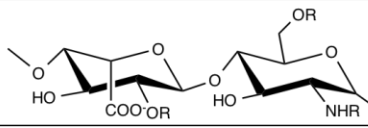
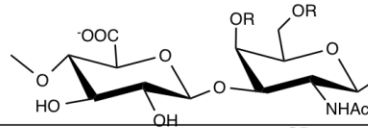
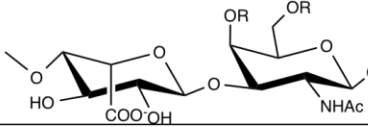
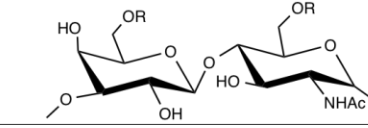
We employed several NMR applications to investigate interactions between the Link module of TSG-6 and  $\Delta$ C444S and the multiple binding modes of Link module in the presence of sugar: chemical shift perturbations (CSP), relaxation rate measurements, paramagnetic relaxation enhancements (PREs), residual dipolar couplings (RDCs), coupled with non-NMR based data, including dynamic light scattering (DLS), and HADDOCK. These are described in Chapter 4. In Chapter 3, a newly determined Link module structure in the presence of  $\Delta$ C444S using triple resonance experiments is introduced because the existing structures of Link-TSG-6 could not fit RDC data, and the binding sites for CS are not identical to that for HA. The induction of dimerization of the Link module by  $\Delta$ C444S is a novel discovery and suggests a need for the development of a new model to illustrate accurately the mechanism underlying the Link module's affinity for and binding to CS.

## **1.2 Interaction between GAGs and GAG binding proteins**

The interactions between glycosaminoglycans (GAGs) and proteins have been studied extensively because they make significant contributions to biological processes such as growth (12,13), cell adhesion (14), development (15-19), coagulation/thrombosis (20,21), angiogenesis, cancer (22-25), microbial pathogenesis (26-30), and inflammation (31-34). This biological activity of GAGs is generally potentiated through the localization, stabilization, activation or deactivation of corresponding proteins.

GAGs are linear polysaccharides composed of repeating disaccharide units of uronic acid (glucuronic acid or iduronic acid) or galactose and an amino sugar (N-acetylglucosamine or N-acetylgalactosamine). There are two main types of GAGs, non-sulfated and sulfated. The major non-sulfated GAG is hyaluronic acid (HA) whereas sulfated GAGs include heparin/heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), and keratan sulfate (KS). Structures for these GAGs are provided in Table 1.1.

**Table 1.1:** Different classes of glycosaminoglycans and their repeating units.

GAGs	Structures	Disaccharide units
Hyaluronic acid		GlcA- $\beta$ (1-3)-GlcNAc- $\alpha$ (1-4)
Heparan sulfate/ Heparin		GlcA- $\beta$ (1-4)-GlcNAc- $\alpha$ (1-4)
		IdoA- $\beta$ (1-4)-GlcNAc- $\alpha$ (1-4)
Chondroitin sulfate		GlcA- $\beta$ (1-3)-GalNAc- $\beta$ (1-4)
Dermatan sulfate		IdoA- $\beta$ (1-3)-GalNAc- $\beta$ (1-4)
Keratan sulfate		Gal- $\beta$ (1-4)-GalNAc- $\beta$ (1-3)

OR: OH or OSO<sub>3</sub><sup>-</sup>

NHR: NHCOCH<sub>3</sub> or NHOSO<sub>3</sub><sup>-</sup>

Structural diversity can be created by modifying the residues in these GAGs; sulfation at a uronic acid and hexosamine can be used to produce a highly negatively charged property. For example, heparan sulfates potentially may contain up to 48 distinct

disaccharide building blocks in contrast to DNA which contains 4 building blocks and proteins which contain 20 building blocks. Given this biological diversity of GAGs, it is not surprising that GAGs are involved in numerous physiological processes.

GAG polymers are made from nucleotide sugars synthesized in the cytosol and transported to the Golgi with a final stage in the trans-Golgi to complete the polymerization in GAG polymers where they are simultaneously modified by enzymes such as sulfotransferases and epimerases. Then, at the end, those are localized in the extracellular membrane for interaction with various proteins and giving physical strength to maintain the extracellular matrix (ECM).

Most examples of GAG-protein interactions in recent literature focus on those with Hp/HS due to their clinical significance and ready availability of Hp as an HS mimic. Other GAGs, namely CS, DS, and KS have been studied less extensively. Table 1.2 lists protein-GAG interactions and their biological significance. Hp has well-documented anti-coagulant properties and is used in a standard pharmaceutical treatment for thrombosis, thrombophlebitis, and embolism. Moreover, structures of heparin that are highly conserved in a broad range of vertebrate and invertebrate organisms (35) are easily accessed. From the known Hp/HS binding protein structures, amino acid consensus sequences for Hp-binding sites have been described as XBBXBX or XBBBXXBX, where B is a lysine or arginine and X is a hydrophobic residue (36). Interactions between Hp and fibroblast growth factors (FGF-1, and -2) have been thoroughly studied, and show Hp stabilizes the resulting complexes. Those interactions and binding sites have been studied by crystallography. In the case of Antithrombin III (AT-III) heparin pentasaccharide binding induces a conformational change of the protein that leads to the

expulsion of the reactive site loop of AT-III. The result is an increased binding affinity for Factor Xa (37). In several cases, for instance FMDV, annexin V, the NK1 domain of HGF, and VCP, crystal structures of the sugars at the binding surface have been documented. These studies show that heparin-protein complexes vary based on conformational changes, GAG sulfation patterns, the presence or absence of metal ions and the effect of pH on binding affinity (38). This reinforces the necessity of obtaining structural information to better understand the underlying geometry of protein-GAG binding.

CCL5, RANTES (regulated on activation, normal T cell expressed and secreted), is a common example that demonstrates the advantage of conducting structural investigations of protein-GAG complexes. CCL5 is a chemokine that interacts with GAGs found on the proteoglycan components of the cell surface and ECM. Immobilized chemokines interact with its receptor on leukocytes. The concentration gradient of chemokines allows the cell to penetrate the cell surface and reach inflammatory sites. The binding of GAGs to CCL5 is important for establishing a haptotactic chemokine gradient on the cell surface (39-41). Utilizing structural investigation tools such as, nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS) and mass spectrometry (MS), the tetramer structure of CCL5 was determined, and the mechanism underlying the interaction between GAG, CCL5 and the protein-protein complex was explained (42).

One final example of GAG-protein interaction is the DBL3x domain in the VAR2CSA protein encoded by the *vars2csa* gene, which is expressed in and transformed on the surface of infected erythrocytes from *Plasmodium falciparum*. This organism is known to cause malaria and the interaction with CS is responsible for adhesion to the

placenta of pregnant women. Placental adhesion of DBL3x causes the sequestration of mature infected erythrocytes in placental blood, leading to maternal anemia (43) and underweight neonates (44).

**Table 1.2:** Biological relevance of GAG-protein interactions.

Protein	GAG	Biological roles
NCAM	HS	Adhesion protein
Fibronectin	HS	Adhesion protein
IL-8	Hp	Chemokine
PF-4	Hp	Chemokine
RANTES	Hp/HS	Chemokine
Annexin V	Hp	Extracellular protein
Annexin A2	Hp	Extracellular protein
Amyloid P (AP)	Hp	Glycoprotein
Fibroblast growth factors-1,2	HS	Growth factors
FGF receptors	HS	Growth factor/receptor
HB-GAM	Hp	Growth factor
Thrombin	Hp	Protease
SLPI	Hp	Protease
AT-III	Hp	Serpin
AT-III/factor Xa	Hp	Serpin/protease
Cardiotoxin A3, A5, M4, and M1	Hp/HS	Toxin
HIV-1-gp-120	Hp	Viral pathogen
Dengue viral envelope protein	Hp/HS	Viral pathogen
FMDV	Hp	Viral pathogen
VCP	Hp	Viral pathogen
HGF	DS	Viral pathogen
Midkine	CS	Viral pathogen
Pleotrophin	CS	Viral pathogen
DBL	CS	Viral pathogen
FGF-2,7	DS	Growth factor
Heparin cofactor-II	DS	Anticoagulant
Link module	HA	Cell adhesion
IaI	HA	Protease inhibitor

X-ray structures in the presence of CSA have been produced and several CSA binding sites have been suggested based on *in vitro* experiments (45). These putative CSA binding sites provide insights into the binding mechanism and have led to the development of better vaccines to prevent malaria during pregnancy.

Even though there are well-established techniques for measuring affinities and distinguishing binding sites, the mechanism of protein-GAG interactions remains unclear due to a lack of structural information. NMR applications and other analytical techniques can provide this sort of detailed structural information to better understand how a GAG-protein complex functions. Employing these methods to study interactions between the Link module domain of human TSG-6 and chondroitin sulfate are discussed below.

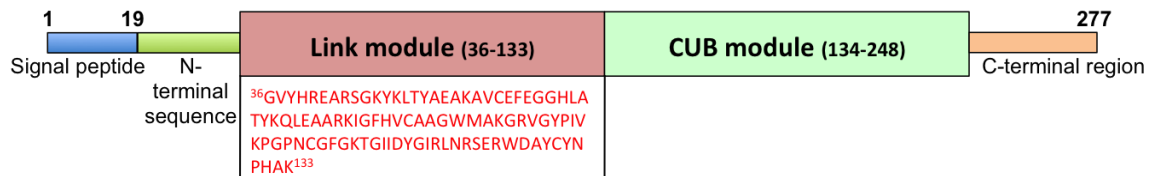
### **1.3 The Link module as a GAG binding protein**

As stated previously, the interaction between GAGs and GAG binding proteins is an important physiological process in the ECM. This study focuses on the Link module of human tumor necrosis factor stimulated gene-6 (TSG-6). The Link module binds to hyaluronan, a non-sulfated GAG and for this reason, a protein that contains Link module is called a “hyalohedrin”. There are three types of superfamily of Link module containing proteins according to the size of the HA-binding domain (46). The type A domain contains a single folded Link module; TSG-6 is a typical example (47,48). The type B domain is also formed from a single Link module but the  $\beta$ -strand is extended by N- and C-terminal flanking sequences that are essential for the correct folding and functioning of the protein (49). CD44, a cell surface receptor for HA, belongs to this domain. Type C consists of two contiguous Link modules, HA-binding domains (HABDs) and a G1 domain of a chondroitin sulfate proteoglycan (CSPG); aggrecan,



versican, neurocan, and brevican are classified as type C (46,50).

TSG-6 is a 35 kDa protein composed of two contiguous modules, Link and CUB (Fig. 1.1). This protein maps to human chromosome 2q23.3 (51,52) and was originally identified as a cDNA prepared from TNF-treated human fibroblasts (1,53). It is not constitutively expressed in healthy adult tissues, but its protein expression is induced in response to inflammatory mediators and growth factors that are tightly associated with numerous disease and physiological processes. These include Rheumatoid arthritis and osteoarthritis (54,55), Kawasaki disease (56), systemic lupus erythematosus (57), and asthma (58). In mouse models of arthritic inflammation TSG-6 exhibits potent anti-inflammatory effects that are protective against cartilage matrix destruction (59-61). TSG-6 also is a potent inhibitor of osteoclast-mediated bone erosion (2,62) and proffers cardioprotective effects during myocardial infarction (63). Moreover, TSG-6 appears to play a role in the reduction of inflammatory corneal damage following injury (64) and the attenuation of zymosan-induced peritonitis (65). Recent studies of direct interaction between TSG-6 and the chemokine CXCL8 demonstrated that TSG-6 inhibits CXCL8-induced transendothelial migration of human neutrophils (66).



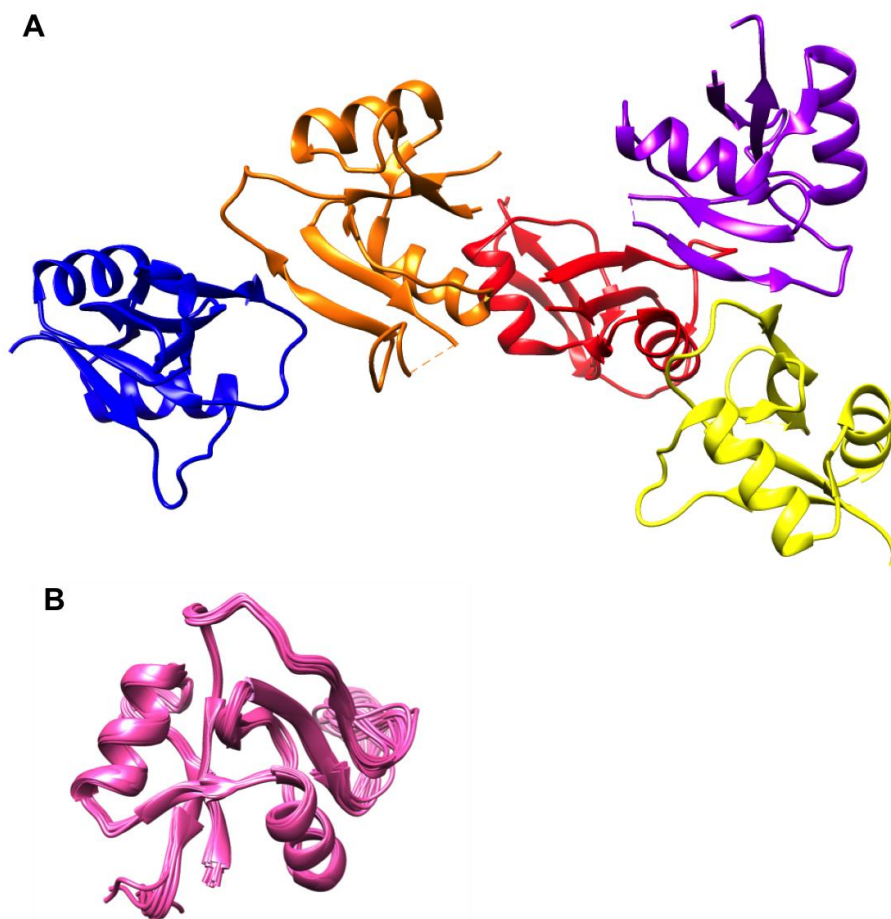
**Figure 1.1:** Full length TSG-6 showing the signal peptide (1-19), N-terminal sequence (20-35), Link module (36-133), and CUB module (134-248) followed by C-terminal region (249-277).

Inflammation-like processes such as blood vessel wall injury (67), ovulation (67-71), and cervical ripening (72) also have been known to induce the expression of TSG-6 in a normal physiological environment. The interaction of TSG-6 with inter- $\alpha$ -inhibitor (I $\alpha$ I), a serine protease inhibitor, is well known and an important factor in female fertility. TSG-6 is critical for the cumulus matrix expansion, the assembly of an HA-rich ECM around the oocyte prior to ovulation (69,71). I $\alpha$ I consists of three polypeptides (heavy chain 1, 2 and bikunin) and TSG-6 acts as a cofactor that transfers heavy chains onto HA (73) at inflammation sites (54,58)

TSG-6 interacts with several components of the ECM via the Link module, such as hyaluronan (HA) (5,7,74-76), heparin and chondroitin sulfate (7), the G1 domain of aggrecan (77), and pentraxin-3 (52,69,71,73,78-80). Most structural studies of the Link module of TSG-6 utilize HA as a main binding partner although in a physiological system there are many binding molecules. 3D presentations of the Link module generated by x-ray crystallography without HA (2PF5.pdb) (81) and solution NMR in the absence and presence of HA octasaccharide (1O7B.pdb and 1O7C.pdb respectively) (75) are shown in Fig. 1.2.

The Link module is a domain of 98 amino acid residues and forms a compact structure comprised of two  $\beta$ -sheets flanked by two  $\alpha$ -helices in a correct folding. According to existing structural studies, the main binding forces of HA and the Link module are CH- $\pi$  stacking interactions between sugar rings and two adjacent tyrosine residues of the Link module, and salt bridges between glucuronic acids in the sugars and basic residues of the protein. As HA binds to the protein, the structure opens to create a shallow binding groove (75), which may occur due to the movement of the C47-C68

disulfide bridge and a rearrangement of the  $\beta 4/\beta 5$  loop which exposes tyrosine residues that are necessary for HA binding (81). In addition to the structural data given above, site-directed mutagenesis data confirm that K11, Y12, Y59, Y78, and R81 are the key HA-binding residues within the groove of the Link module.



**Figure 1.2:** 3D structures of Link-TSG-6 determined by x-ray crystallography (A) and NMR spectroscopy (B). A: Five chains in a unit cell are shown - red (chain A), orange (chain B), yellow (chain C), blue (chain D), and purple (chain E) (2PF5.pdb). B: 20 deposited structures are drawn with magenta colored ribbons (1O7C.pdb).

The binding affinity of the HA octasaccharide for the Link module is approximately 17.83  $\mu$ M. In addition to HA, Hp is often studied with the Link module because it is known to accommodate the Link module using a different binding mechanism. Site-directed mutagenesis and docking predictions show that, unlike HA, the Hp binding site

houses K20, K34, R40, K41, K54, R56 and R84 (5). This binding site is distinct from the HA binding surface. Not surprisingly, the interaction between the Link module and heparin likely involves extensive electrostatic contacts because heparin includes many sulfate moieties that give the sugar a highly negatively charged character. The binding affinity of Link-TSG-6 to Hp octasaccharide is  $\sim 6$ -fold higher ( $3.14 \mu\text{M}$ ) than its interaction with the HA octasaccharide under the same conditions (5). However, HA and Hp are unable to bind to the Link module simultaneously even though their binding surfaces are distinct from one another and do not overlap. This might be due to allosteric effects. One possible explanation is that once Hp binds to Link-TSG-6 it induces a conformational change in the closed state that makes HA-binding residues less accessible for interactions with HA (75). Conversely, HA binding that stabilizes the protein in an open conformation may impact the relative orientation of the amino acids involved in Hp binding (75). Unlike HA, which binds to the Link module in a 1:1 ratio, Hp octasaccharide has  $\sim 0.5$  of the stoichiometry for interaction with Link-TSG-6, indicating that one Hp octasaccharide binds to two Link-TSG-6 modules to form a dimer (5). It has been shown that Hp binds to the Link module and causes dimerization of the protein and a dimer of the Link module with Hp interacts with the bikunin module of  $\text{I}\alpha\text{I}$ . In this regard, Hp binding to Link-TSG-6 affects its potentiation of the antiplasmin activity of  $\text{I}\alpha\text{I}$  by inducing changes in the relative orientations of Kunitz domains and making the protease binding sites more accessible.

#### **1.4 Research questions**

The main focus of this study is using NMR to clarify the binding mode of the Link module with chondroitin sulfate. In order to characterize structurally the Link

module with respect to CS binding using NMR applications, homogeneous CSs were required to acquire high quality experimental data. Due to extreme structural heterogeneity in CS polymers, it was a challenge in obtaining well defined oligosaccharides of CS. However, it was possible to meet this challenge by using enzymatic degradation of CS polymers from natural sources. An improved understanding of digestive enzymes and their activities was also achieved in the course of the production. The compositions of the oligosaccharides from the digestions benefitted from a structural assignment with NMR.

To begin the process, the binding sites of CS and the Link module were examined because there are two known binding sites for HA and Hp/HS that have been confirmed by X-ray, NMR, molecular modeling, and mutagenesis methods. Early on, it appeared that the binding of CS to the Link module occurred very differently than its interactions with HA. The CS binding residues are slightly different to those of HA, even though the binding pocket of Link module is the same. Another distinction is that the Link module and the CS hexasaccharide form a dimer. In order to adequately employ an accurate dimer model, new soluble structures for this bound state complex had to be determined. As a result, additional analytical tools were employed to generate the dimer model.

## **1.5 References**

1. Lee, T. H., Wisniewski, H. G., and Vilcek, J. (1992) A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member of the family of hyaluronate binding proteins, closely related to the adhesion receptor CD44. *The Journal of cell biology* **116**, 545-557
2. Mahoney, D. J., Swales, C., Athanasou, N. A., Bombardieri, M., Pitzalis, C., Kliskey, K., Sharif, M., Day, A. J., Milner, C. M., and Sabokbar, A. (2011) TSG-6 inhibits osteoclast activity via an autocrine mechanism and is functionally synergistic with osteoprotegerin. *Arthritis Rheum* **63**, 1034-1043

3. Nagyeri, G., Radacs, M., Ghassemi-Nejad, S., Tryniszewska, B., Olasz, K., Hutás, G., Gyorfy, Z., Hascall, V. C., Glant, T. T., and Mikecz, K. (2011) TSG-6 protein, a negative regulator of inflammatory arthritis, forms a ternary complex with murine mast cell tryptases and heparin. *J Biol Chem* **286**, 23559-23569
4. Getting, S. J., Mahoney, D. J., Cao, T., Rugg, M. S., Fries, E., Milner, C. M., Perretti, M., and Day, A. J. (2002) The link module from human TSG-6 inhibits neutrophil migration in a hyaluronan- and inter-alpha -inhibitor-independent manner. *J Biol Chem* **277**, 51068-51076
5. Mahoney, D. J., Mulloy, B., Forster, M. J., Blundell, C. D., Fries, E., Milner, C. M., and Day, A. J. (2005) Characterization of the interaction between tumor necrosis factor-stimulated gene-6 and heparin: implications for the inhibition of plasmin in extracellular matrix microenvironments. *J Biol Chem* **280**, 27044-27055
6. Kuznetsova, S. A., Day, A. J., Mahoney, D. J., Rugg, M. S., Mosher, D. F., and Roberts, D. D. (2005) The N-terminal module of thrombospondin-1 interacts with the link domain of TSG-6 and enhances its covalent association with the heavy chains of inter-alpha-trypsin inhibitor. *J Biol Chem* **280**, 30899-30908
7. Parkar, A. A., and Day, A. J. (1997) Overlapping sites on the Link module of human TSG-6 mediate binding to hyaluronan and chondroitin-4-sulphate. *FEBS Lett* **410**, 413-417
8. Marson, A., Robinson, D. E., Brookes, P. N., Mulloy, B., Wiles, M., Clark, S. J., Fielder, H. L., Collinson, L. J., Cain, S. A., Kielty, C. M., McArthur, S., Buttle, D. J., Short, R. D., Whittle, J. D., and Day, A. J. (2009) Development of a microtiter plate-based glycosaminoglycan array for the investigation of glycosaminoglycan-protein interactions. *Glycobiology* **19**, 1537-1546
9. Wisniewski, H. G., Snitkin, E. S., Mindrescu, C., Sweet, M. H., and Vilcek, J. (2005) TSG-6 protein binding to glycosaminoglycans: formation of stable complexes with hyaluronan and binding to chondroitin sulfates. *J Biol Chem* **280**, 14476-14484
10. Mindrescu, C., Thorbecke, G. J., Klein, M. J., Vilcek, J., and Wisniewski, H. G. (2000) Amelioration of collagen-induced arthritis in DBA/1J mice by recombinant TSG-6, a tumor necrosis factor/interleukin-1-inducible protein. *Arthritis Rheum* **43**, 2668-2677
11. Bardos, T., Mikecz, K., Finnegan, A., Zhang, J., and Glant, T. T. (2002) T and B cell recovery in arthritis adoptively transferred to SCID mice: antigen-specific

activation is required for restoration of autopathogenic CD4<sup>+</sup> Th1 cells in a syngeneic system. *Journal of immunology* **168**, 6013-6021

12. Beaulieu, J. F., Vachon, P. H., and Chartrand, S. (1991) Immunolocalization of extracellular matrix components during organogenesis in the human small intestine. *Anat Embryol (Berl)* **183**, 363-369
13. Cohn, R. H., Cassiman, J. J., and Bernfield, M. R. (1976) Relationship of transformation, cell density, and growth control to the cellular distribution of newly synthesized glycosaminoglycan. *The Journal of cell biology* **71**, 280-294
14. de Aguiar C, L.-S. B., Alvarez-Silva M, Trentin AG. (2005) Glycosaminoglycans modulate C6 glioma cell adhesion to extracellular matrix components and alter cell proliferation and cell migration. . *Bmc Cell Biology* **6**
15. Bradbury, E. J., Moon, L. D., Popat, R. J., King, V. R., Bennett, G. S., Patel, P. N., Fawcett, J. W., and McMahon, S. B. (2002) Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature* **416**, 636-640
16. Perrimon, N., and Bernfield, M. (2000) Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **404**, 725-728
17. Mizuguchi, S., Uyama, T., Kitagawa, H., Nomura, K. H., Dejima, K., Gengyo-Ando, K., Mitani, S., Sugahara, K., and Nomura, K. (2003) Chondroitin proteoglycans are involved in cell division of *Caenorhabditis elegans*. *Nature* **423**, 443-448
18. Hwang, H. Y., Olson, S. K., Esko, J. D., and Horvitz, H. R. (2003) *Caenorhabditis elegans* early embryogenesis and vulval morphogenesis require chondroitin biosynthesis. *Nature* **423**, 439-443
19. Lin, X. (2004) Functions of heparan sulfate proteoglycans in cell signaling during development. *Development* **131**, 6009-6021
20. de Mattos, D. A., Stelling, M. P., Tovar, A. M., and Mourao, P. A. (2008) Heparan sulfates from arteries and veins differ in their antithrombin-mediated anticoagulant activity. *Journal of thrombosis and haemostasis : JTH* **6**, 1987-1990
21. He, L., Giri, T. K., Vicente, C. P., and Tollefsen, D. M. (2008) Vascular dermatan sulfate regulates the antithrombotic activity of heparin cofactor II. *Blood* **111**, 4118-4125

22. Sasisekharan, R., Shriver, Z., Venkataraman, G., and Narayanasami, U. (2002) Roles of heparan-sulphate glycosaminoglycans in cancer. *Nature reviews. Cancer* **2**, 521-528
23. Liu, D., Shriver, Z., Venkataraman, G., El Shabrawi, Y., and Sasisekharan, R. (2002) Tumor cell surface heparan sulfate as cryptic promoters or inhibitors of tumor growth and metastasis. *Proc Natl Acad Sci U S A* **99**, 568-573
24. Vlodavsky, I., and Goldshmidt, O. (2001) Properties and function of heparanase in cancer metastasis and angiogenesis. *Haemostasis* **31 Suppl 1**, 60-63
25. Casu, B., Guerrini, M., Guglieri, S., Naggi, A., Perez, M., Torri, G., Cassinelli, G., Ribatti, D., Carminati, P., Giannini, G., Penco, S., Pisano, C., Belleri, M., Rusnati, M., and Presta, M. (2004) Undersulfated and glycol-split heparins endowed with antiangiogenic activity. *J Med Chem* **47**, 838-848
26. Ganesh, V. K., Smith, S. A., Kotwal, G. J., and Murthy, K. H. (2004) Structure of vaccinia complement protein in complex with heparin and potential implications for complement regulation. *Proc Natl Acad Sci U S A* **101**, 8924-8929
27. Fry, E. E., Lea, S. M., Jackson, T., Newman, J. W., Ellard, F. M., Blakemore, W. E., Abu-Ghazaleh, R., Samuel, A., King, A. M., and Stuart, D. I. (1999) The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex. *EMBO J* **18**, 543-554
28. Trowbridge, J. M., and Gallo, R. L. (2002) Dermatan sulfate: new functions from an old glycosaminoglycan. *Glycobiology* **12**, 117R-125R
29. Liu, J., Shriver, Z., Pope, R. M., Thorp, S. C., Duncan, M. B., Copeland, R. J., Raska, C. S., Yoshida, K., Eisenberg, R. J., Cohen, G., Linhardt, R. J., and Sasisekharan, R. (2002) Characterization of a heparan sulfate octasaccharide that binds to herpes simplex virus type 1 glycoprotein D. *J Biol Chem* **277**, 33456-33467
30. Alkhalil, A., Achur, R. N., Valiyaveetil, M., Ockenhouse, C. F., and Gowda, D. C. (2000) Structural requirements for the adherence of Plasmodium falciparum-infected erythrocytes to chondroitin sulfate proteoglycans of human placenta. *J Biol Chem* **275**, 40357-40364
31. Doodes, P. D., Cao, Y., Hamel, K. M., Wang, Y., Rodeghero, R. L., Kobezda, T., and Finnegan, A. (2009) CCR5 is involved in resolution of inflammation in proteoglycan-induced arthritis. *Arthritis Rheum* **60**, 2945-2953



32. Handel, T. M., Johnson, Z., Crown, S. E., Lau, E. K., and Proudfoot, A. E. (2005) Regulation of protein function by glycosaminoglycans--as exemplified by chemokines. *Annual review of biochemistry* **74**, 385-410
33. Kaplan, C. D., O'Neill, S. K., Koreny, T., Czipri, M., and Finnegan, A. (2002) Development of inflammation in proteoglycan-induced arthritis is dependent on Fc gamma R regulation of the cytokine/chemokine environment. *J Immunol* **169**, 5851-5859
34. Koninger, J., Giese, N. A., Bartel, M., di Mola, F. F., Berberat, P. O., di Sebastiano, P., Giese, T., Buchler, M. W., and Friess, H. (2006) The ECM proteoglycan decorin links desmoplasia and inflammation in chronic pancreatitis. *J Clin Pathol* **59**, 21-27
35. Nader, H. B., Chavante, S. F., dos-Santos, E. A., Oliveira, T. W., de-Paiva, J. F., Jeronimo, S. M., Medeiros, G. F., de-Abreu, L. R., Leite, E. L., de-Sousa-Filho, J. F., Castro, R. A., Toma, L., Tersariol, I. L., Porcionatto, M. A., and Dietrich, C. P. (1999) Heparan sulfates and heparins: similar compounds performing the same functions in vertebrates and invertebrates? *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.]* **32**, 529-538
36. Cardin, A. D., and Weintraub, H. J. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **9**, 21-32
37. Jin, L., Abrahams, J. P., Skinner, R., Petitou, M., Pike, R. N., and Carrell, R. W. (1997) The anticoagulant activation of antithrombin by heparin. *Proc Natl Acad Sci U S A* **94**, 14683-14688
38. Gandhi, N. S., and Mancera, R. L. (2008) The structure of glycosaminoglycans and their interactions with proteins. *Chem Biol Drug Des* **72**, 455-482
39. Johnson, Z., Kosco-Vilbois, M. H., Herren, S., Cirillo, R., Muzio, V., Zaratini, P., Carbonatto, M., Mack, M., Smailbegovic, A., Rose, M., Lever, R., Page, C., Wells, T. N., and Proudfoot, A. E. (2004) Interference with heparin binding and oligomerization creates a novel anti-inflammatory strategy targeting the chemokine system. *Journal of immunology* **173**, 5776-5785
40. Murooka, T. T., Wong, M. M., Rahbar, R., Majchrzak-Kita, B., Proudfoot, A. E., and Fish, E. N. (2006) CCL5-CCR5-mediated apoptosis in T cells: Requirement for glycosaminoglycan binding and CCL5 aggregation. *J Biol Chem* **281**, 25184-25194

41. Proudfoot, A. E., Handel, T. M., Johnson, Z., Lau, E. K., LiWang, P., Clark-Lewis, I., Borlat, F., Wells, T. N., and Kosco-Vilbois, M. H. (2003) Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1885-1890
42. Wang, X., Watson, C., Sharp, J. S., Handel, T. M., and Prestegard, J. H. (2011) Oligomeric structure of the chemokine CCL5/RANTES from NMR, MS, and SAXS data. *Structure* **19**, 1138-1148
43. Viebig, N. K., Gamain, B., Scheidig, C., Lepolard, C., Przyborski, J., Lanzer, M., Gysin, J., and Scherf, A. (2005) A single member of the Plasmodium falciparum var multigene family determines cytoadhesion to the placental receptor chondroitin sulphate A. *EMBO Rep* **6**, 775-781
44. Duffy, M. F., Maier, A. G., Byrne, T. J., Marty, A. J., Elliott, S. R., O'Neill, M. T., Payne, P. D., Rogerson, S. J., Cowman, A. F., Crabb, B. S., and Brown, G. V. (2006) VAR2CSA is the principal ligand for chondroitin sulfate A in two allogeneic isolates of Plasmodium falciparum. *Mol Biochem Parasitol* **148**, 117-124
45. Kavita Singh, A. G. G., Phuc Nguyen, D Channe Gowda, Louis H Miller, and David N Garboczi. (2008) Structure of the DBL3x domain of pregnancy-associated malaria protein VAR2CSA complexed with chondroitin sulfate A. *Nat Struct Mol Biol.* **15**, 932
46. Day, A. J., and Prestwich, G. D. (2002) Hyaluronan-binding proteins: tying up the giant. *J Biol Chem* **277**, 4585-4588
47. Milner, C. M., and Day, A. J. (2003) TSG-6: a multifunctional protein associated with inflammation. *J Cell Sci* **116**, 1863-1873
48. Wisniewski, H. G., and Vilcek, J. (2004) Cytokine-induced gene expression at the crossroads of innate immunity, inflammation and fertility: TSG-6 and PTX3/TSG-14. *Cytokine & growth factor reviews* **15**, 129-146
49. Teriete, P., Banerji, S., Noble, M., Blundell, C. D., Wright, A. J., Pickford, A. R., Lowe, E., Mahoney, D. J., Tammi, M. I., Kahmann, J. D., Campbell, I. D., Day, A. J., and Jackson, D. G. (2004) Structure of the regulatory hyaluronan binding domain in the inflammatory leukocyte homing receptor CD44. *Mol Cell* **13**, 483-496
50. Spicer, A. P., Joo, A., and Bowling, R. A., Jr. (2003) A hyaluronan binding link protein gene family whose members are physically linked adjacent to chondroitin

sulfate proteoglycan core protein genes: the missing links. *J Biol Chem* **278**, 21083-21091

51. Lee, T. H., Klampfer, L., Shows, T. B., and Vilcek, J. (1993) Transcriptional regulation of TSG6, a tumor necrosis factor- and interleukin-1-inducible primary response gene coding for a secreted hyaluronan-binding protein. *J Biol Chem* **268**, 6154-6160
52. Nentwich, H. A., Mustafa, Z., Rugg, M. S., Marsden, B. D., Cordell, M. R., Mahoney, D. J., Jenkins, S. C., Dowling, B., Fries, E., Milner, C. M., Loughlin, J., and Day, A. J. (2002) A novel allelic variant of the human TSG-6 gene encoding an amino acid difference in the CUB module. Chromosomal localization, frequency analysis, modeling, and expression. *J Biol Chem* **277**, 15354-15362
53. Lee, T. H., Lee, G. W., Ziff, E. B., and Vilcek, J. (1990) Isolation and characterization of eight tumor necrosis factor-induced gene sequences from human fibroblasts. *Mol Cell Biol* **10**, 1982-1988
54. Wisniewski, H. G., Maier, R., Lotz, M., Lee, S., Klampfer, L., Lee, T. H., and Vilcek, J. (1993) TSG-6: a TNF-, IL-1-, and LPS-inducible secreted glycoprotein associated with arthritis. *J Immunol* **151**, 6593-6601
55. Bayliss, M. T., Howat, S. L., Dudhia, J., Murphy, J. M., Barry, F. P., Edwards, J. C., and Day, A. J. (2001) Up-regulation and differential expression of the hyaluronan-binding protein TSG-6 in cartilage and synovium in rheumatoid arthritis and osteoarthritis. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* **9**, 42-48
56. Nomura, I., Abe, J., Noma, S., Saito, H., Gao, B., Wheeler, G., and Leung, D. Y. (2005) Adrenomedullin is highly expressed in blood monocytes associated with acute Kawasaki disease: a microarray gene expression study. *Pediatric research* **57**, 49-55
57. Wisniewski, H. G., and Vilcek, J. (1997) TSG-6: an IL-1/TNF-inducible protein with anti-inflammatory activity. *Cytokine & growth factor reviews* **8**, 143-156
58. Forteza, R., Casalino-Matsuda, S. M., Monzon, M. E., Fries, E., Rugg, M. S., Milner, C. M., and Day, A. J. (2007) TSG-6 potentiates the antitissue kallikrein activity of inter-alpha-inhibitor through bikunin release. *Am J Respir Cell Mol Biol* **36**, 20-31
59. Glant, T. T., Kamath, R. V., Bardos, T., Gal, I., Szanto, S., Murad, Y. M., Sandy, J. D., Mort, J. S., Roughley, P. J., and Mikecz, K. (2002) Cartilage-specific constitutive expression of TSG-6 protein (product of tumor necrosis factor alpha-

stimulated gene 6) provides a chondroprotective, but not antiinflammatory, effect in antigen-induced arthritis. *Arthritis Rheum* **46**, 2207-2218

60. Mindrescu, C., Dias, A. A., Olszewski, R. J., Klein, M. J., Reis, L. F., and Wisniewski, H. G. (2002) Reduced susceptibility to collagen-induced arthritis in DBA/1J mice expressing the TSG-6 transgene. *Arthritis Rheum* **46**, 2453-2464
61. Szanto, S., Bardos, T., Gal, I., Glant, T. T., and Mikecz, K. (2004) Enhanced neutrophil extravasation and rapid progression of proteoglycan-induced arthritis in TSG-6-knockout mice. *Arthritis Rheum* **50**, 3012-3022
62. Mahoney, D. J., Mikecz, K., Ali, T., Mabileau, G., Benayahu, D., Plaas, A., Milner, C. M., Day, A. J., and Sabokbar, A. (2008) TSG-6 regulates bone remodeling through inhibition of osteoblastogenesis and osteoclast activation. *J Biol Chem* **283**, 25952-25962
63. Lee, R. H., Pulin, A. A., Seo, M. J., Kota, D. J., Ylostalo, J., Larson, B. L., Semprun-Prieto, L., Delafontaine, P., and Prockop, D. J. (2009) Intravenous hMSCs improve myocardial infarction in mice because cells embolized in lung are activated to secrete the anti-inflammatory protein TSG-6. *Cell stem cell* **5**, 54-63
64. Oh, J. Y., Roddy, G. W., Choi, H., Lee, R. H., Ylostalo, J. H., Rosa, R. H., Jr., and Prockop, D. J. (2010) Anti-inflammatory protein TSG-6 reduces inflammatory damage to the cornea following chemical and mechanical injury. *Proc Natl Acad Sci U S A* **107**, 16875-16880
65. Choi, H., Lee, R. H., Bazhanov, N., Oh, J. Y., and Prockop, D. J. (2011) Anti-inflammatory protein TSG-6 secreted by activated MSCs attenuates zymosan-induced mouse peritonitis by decreasing TLR2/NF-kappaB signaling in resident macrophages. *Blood* **118**, 330-338
66. Dyer, D. P., Thomson, J. M., Hermant, A., Jowitt, T. A., Handel, T. M., Proudfoot, A. E., Day, A. J., and Milner, C. M. (2014) TSG-6 inhibits neutrophil migration via direct interaction with the chemokine CXCL8. *J Immunol* **192**, 2177-2185
67. Ye, L., Mora, R., Akhayani, N., Haudenschild, C. C., and Liau, G. (1997) Growth factor and cytokine-regulated hyaluronan-binding protein TSG-6 is localized to the injury-induced rat neointima and confers enhanced growth in vascular smooth muscle cells. *Circ Res* **81**, 289-296
68. Carrette, O., Nemade, R. V., Day, A. J., Brickner, A., and Larsen, W. J. (2001) TSG-6 is concentrated in the extracellular matrix of mouse cumulus oocyte

- complexes through hyaluronan and inter-alpha-inhibitor binding. *Biol Reprod* **65**, 301-308
69. Fulop, C., Szanto, S., Mukhopadhyay, D., Bardos, T., Kamath, R. V., Rugg, M. S., Day, A. J., Salustri, A., Hascall, V. C., Glant, T. T., and Mikecz, K. (2003) Impaired cumulus mucification and female sterility in tumor necrosis factor-induced protein-6 deficient mice. *Development* **130**, 2253-2261
  70. Mukhopadhyay, D., Hascall, V. C., Day, A. J., Salustri, A., and Fulop, C. (2001) Two distinct populations of tumor necrosis factor-stimulated gene-6 protein in the extracellular matrix of expanded mouse cumulus cell-oocyte complexes. *Arch Biochem Biophys* **394**, 173-181
  71. Ochsner, S. A., Russell, D. L., Day, A. J., Breyer, R. M., and Richards, J. S. (2003) Decreased expression of tumor necrosis factor-alpha-stimulated gene 6 in cumulus cells of the cyclooxygenase-2 and EP2 null mice. *Endocrinology* **144**, 1008-1019
  72. Fujimoto, T., Savani, R. C., Watari, M., Day, A. J., and Strauss, J. F., 3rd. (2002) Induction of the hyaluronic acid-binding protein, tumor necrosis factor-stimulated gene-6, in cervical smooth muscle cells by tumor necrosis factor-alpha and prostaglandin E(2). *Am J Pathol* **160**, 1495-1502
  73. Rugg, M. S., Willis, A. C., Mukhopadhyay, D., Hascall, V. C., Fries, E., Fulop, C., Milner, C. M., and Day, A. J. (2005) Characterization of complexes formed between TSG-6 and inter-alpha-inhibitor that act as intermediates in the covalent transfer of heavy chains onto hyaluronan. *J Biol Chem* **280**, 25674-25686
  74. Blundell, C. D., Almond, A., Mahoney, D. J., DeAngelis, P. L., Campbell, I. D., and Day, A. J. (2005) Towards a structure for a TSG-6.hyaluronan complex by modeling and NMR spectroscopy: insights into other members of the link module superfamily. *J Biol Chem* **280**, 18189-18201
  75. Blundell, C. D., Mahoney, D. J., Almond, A., DeAngelis, P. L., Kahmann, J. D., Teriete, P., Pickford, A. R., Campbell, I. D., and Day, A. J. (2003) The link module from ovulation- and inflammation-associated protein TSG-6 changes conformation on hyaluronan binding. *J Biol Chem* **278**, 49261-49270
  76. Kohda, D., Morton, C. J., Parkar, A. A., Hatanaka, H., Inagaki, F. M., Campbell, I. D., and Day, A. J. (1996) Solution structure of the link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration. *Cell* **86**, 767-775

77. Parkar, A. A., Kahmann, J. D., Howat, S. L., Bayliss, M. T., and Day, A. J. (1998) TSG-6 interacts with hyaluronan and aggrecan in a pH-dependent manner via a common functional element: implications for its regulation in inflamed cartilage. *FEBS Lett* **428**, 171-176
78. Wisniewski, H. G., Burgess, W. H., Oppenheim, J. D., and Vilcek, J. (1994) TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable complex with the serum protein inter-alpha-inhibitor. *Biochemistry* **33**, 7423-7429
79. Mukhopadhyay, D., Asari, A., Rugg, M. S., Day, A. J., and Fulop, C. (2004) Specificity of the tumor necrosis factor-induced protein 6-mediated heavy chain transfer from inter-alpha-trypsin inhibitor to hyaluronan: implications for the assembly of the cumulus extracellular matrix. *J Biol Chem* **279**, 11119-11128
80. Sanggaard, K. W., Karring, H., Valnickova, Z., Thogersen, I. B., and Enghild, J. J. (2005) The TSG-6 and I alpha I interaction promotes a transesterification cleaving the protein-glycosaminoglycan-protein (PGP) cross-link. *J Biol Chem* **280**, 11936-11942
81. Higman, V. A., Blundell, C. D., Mahoney, D. J., Redfield, C., Noble, M. E., and Day, A. J. (2007) Plasticity of the TSG-6 HA-binding loop and mobility in the TSG-6-HA complex revealed by NMR and X-ray crystallography. *J Mol Biol* **371**, 669-684

## **CHAPTER 2**

### **PREPARATION OF A NITROXIDE SPIN-LABELED ANALOG OF CHONDROITIN SULFATE**

#### **2.1 Introduction**

Throughout this research, chondroitin sulfate has been exploited to study interactions with the Link module of TSG-6. This chapter describes the synthesis of a spin-labeled chondroitin sulfate that was used to generate more quantitative distance constraints on how the Link module and CS interact, a critical step toward generating a reliable model.

Glycosaminoglycans (GAGs) are linear polysaccharides mostly found on animal cell surfaces and in extracellular matrices (1). The synthesis of sulfated GAGs begins in the endoplasmic reticulum (ER) by adding a xylose (Xyl) residue to specific serine residues in core proteins. Two galactoses (Gal) and one glucuronic acid (GlcA) are then transferred sequentially by specific transferases in medial- and trans-Golgi compartments. These tetrasaccharides are also commonly found in HS. After this transfer, N-acetyl galactosamine (GalNAc) and GlcA are repeatedly added in sequence by various glycosyltransferases including GalNAcT-I, GalNAcT-II, GlcAT-II, ChSy-1, and ChPF (2-4). The synthesis of CS occurs in the trans-Golgi with numerous modifications including sulfation, and epimerization. The average size of a polysaccharide of CS is approximately 20 kDa (~40 disaccharides). Unlike HA, CS is usually found linked to proteins as part of a proteoglycan (chondroitin sulfate proteoglycans, CSPGs). This is a very important structural component of cartilage as it strengthens tissue and regulates

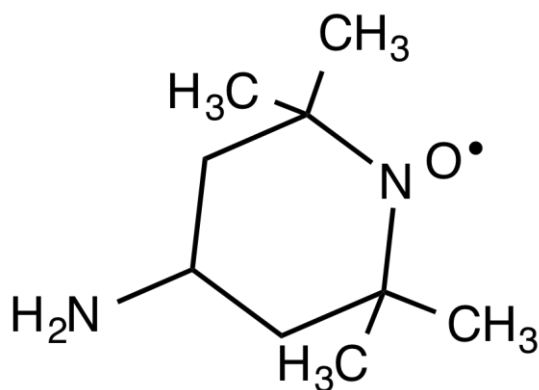
cell-cell and cell-matrix interactions (5). The structural organization of extracellular matrices and regulation of growth factors, cytokines, chemokines, enzymes, and the immune system rely heavily on the interplay between GAGs and GAG-binding proteins (6-10). In theory these interactions can be monitored with various analytical techniques, such as x-ray crystallography, nuclear magnetic resonance spectroscopy, molecular modeling, mass spectrometry, and surface plasma resonance, and various *in vitro* techniques. However, few techniques are available to provide detailed structural information about the protein-GAG complex and its binding sites. NMR spectroscopy is one of the best analytical techniques for this purpose due to its ability to study dynamic interactions between proteins and GAGs. Unfortunately, NMR relies mainly on the samples that contain NMR active nuclei with  $\frac{1}{2}$  nuclear spin such as  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$ . In the current study, when a protein sample that is complex with GAG is isotopically labeled with  $^{13}\text{C}$  or  $^{15}\text{N}$ , or both for acquiring spectra, it yields predominantly structure information on the protein and incomplete information about the GAG. The chemical shifts of non-exchangeable protons on GAGs mostly overlap with the chemical shifts of protons from aliphatic regions of amino acids of the target protein (especially  $\alpha$ -protons). This makes monitoring resonances from sugars bound to the isotope labeled protein very challenging. Understanding how the protein functions in the presence of ligand during binding is dependent upon knowing the geometry of its ligand. Therefore, additional protocols are needed to improve the precision and sophistication of obtaining structural information from GAGs.

Most GAG oligosaccharide structures have a reducing terminus with a reactive anomeric center that can be modified in a number of ways (11). Because of this, the



modification of the reducing end is commonly used with a number of reactions such as Kochetkov amination, glycosylhydrazide formation, and reductive amination (12-17). Functionalized GAG oligosaccharides can be used for sugar analysis, protein purification, and binding affinity studies (12,14,18-20).

Here, we introduced the use of an organic spin-label molecule, 4-amino TEMPO (amino-2,2,6,6-tetramethylpiperidine-1-oxyl) which possesses a nitroxide moiety (Fig. 1). TEMPO is a red-orange chemical compound that is widely used as a stable radical marker and a structural probe (21). In our study, it was incorporated into the reducing end of CS hexasaccharide to synthesize a spin labeled analog of CS hexasaccharide.



**Figure 2.1:** The structure of 4-amino-TEMPO (amino-2,2,6,6-tetramethylpiperidine-1-oxyl).

TEMPO has an unpaired electron in its nitroxide moiety that provides a paramagnetic effect to neighboring spins as far away as 20 Å, yielding long distance constraints as well as RDCs. Thus, PREs can complement NOEs that provide short-range distance constraints, roughly from 2 to 5 Å. This reduces the intensity of the resonances in most NMR experiments. In this research,  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the Link module were collected in reduced and oxidized forms of spin-labeled CS hexasaccharide.

The difference of the intensity changes with reduced and oxidized TEMPO analogs provides distance restraints according to Equation 1.

$$\ln\left(\frac{I_{nsl}}{I_{sl}}\right) = f \times t \times \left(\frac{K}{r^6}\right) \times (4\tau_c + 3\tau_c \times [1 + \omega_H^2 \tau_c^2]^{-1})$$

**Equation 1.**

$I_{nsl}$  and  $I_{sl}$  are the peak intensities of resonances in the presence and absence of a spin label, respectively.  $f$  is the fraction of the protein bound ligand, and  $t$  is the total time during INEPT (insensitive nuclei enhanced by polarization transfer) and refocusing periods of the HSQC pulse sequence.  $K$  is a constant related to spin properties of the system ( $K=1/15*S(S+1)\gamma^2g^2B^2=1.23 \times 10^{-32} \text{ cm}^6\text{s}^{-2}$ ),  $r$  is the distance between the nitroxide and a spin of interest,  $\tau_c$  is the effective correlation time for tumbling of the protein, and  $\omega_H$  (800 MHz  $2\pi$ ) is the precession frequency for the amide proton. Along with RDCs and NOEs, PREs provide useful distance constraints that will hugely improve the quality of the complex compared to the one refined only with NOEs.

To functionalize the reducing end of the GAG oligosaccharide, chondroitin sulfate hexasaccharide ( $\Delta C444S$ ), we chose a reductive amination with an appropriate amine (4-amino TEMPO) to give the corresponding secondary amine (11,12,14). Reductive amination of GAG oligosaccharides typically require extended reaction times at low pH in the presence of an excess of amine to form an initial imine, which is then reduced with an excess of reducing agent (6). The spin label derivative was separated by HPLC. The molecular weight was confirmed using mass spectrometry followed by acquiring 1D  $^1\text{H}$  and 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra to validate whether TEMPO remained in its oxidized form on the sugar in solution. This chapter introduces the synthesis of the spin label molecule; further applications of the interaction assay with Link-TSG-6 are described in Chapter 4.

## **2.2 Methodology for synthesizing ΔC444S-TEMPO**

### **2.2.1 Materials and reagents**

Chondroitin sulfate A from bovine trachea (btCS-A) and chondroitin AC lyase from *F. heparinum* (EC 4.2.2.5) were purchased from Sigma-Aldrich (St. Louis, MO, USA). 4-amino-TEMPO was purchased from Acros organics. Sodium cyanoborohydride was purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetic acid, methanol, and acetonitrile were used to dissolve solutes for reaction and product analysis. Deuterium oxide “100%” (D 99,96%) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). Bio-Gel P-10, a resin in the form of fine polyacrylamide beads and a polypropylene chromatographic column (1200 x 15 mm) for size-exclusion chromatography and a 500 x 10 mm column for desalting were purchased from Bio-Rad Life Science (Hercules, CA, USA). Sephadex G-15 resin and 1,9-dimethyl-methyleneblue (dye content 80%) were purchased from Sigma-Aldrich Co (St. Louis, MO). A pre-packed Spherisorb S10 SAX column (250 x 10 mm, 5 μm) was obtained from the Waters Corporation (Milford, MA).

### **2.2.2 Preparation of starting GAG material**

The digestion of btCS-A with lyase AC was carried out by incubating 150 mg of btCS-A with 0.33 units of enzyme in 5 ml of digestion buffer (50 mM Tris-HCl, pH 8, 150 mM sodium acetate, 100 μg/ml of BSA) at 37 °C for 210 min. After digestion was completed, the sample was heated at 100 °C for 15 min to quench the activity of the enzyme. A 5 ml digested sample was subjected to size-exclusion chromatography on a Bio-Gel P-10 column (1200x 15 mm) with an elution solution of 1 M NaCl containing 10 % ethanol at a flow rate of 2 ml/15 min/fraction. The separation was monitored by

UV absorbance at  $\lambda = 232$  nm. The tubes of the peak corresponding to the size of a hexasaccharide were pooled and subjected to a G-15 column separation from salt using an elution solution of ddH<sub>2</sub>O with a flow rate at 1 ml/ 7.5 min/fraction monitored at  $\lambda = 232$  nm. The hexasaccharide fractions were lyophilized and dissolved in water and subjected to SAX chromatography (250 x 10 mm) employing a Waters HPLC system using a linear gradient from 0 to 1 M of NaCl in H<sub>2</sub>O. Elution was monitored at  $\lambda = 232$  nm with a flow rate at 3 ml/min. The target peaks were collected and subjected to a G-15 desalting column to remove salt for further analysis. Detail procedures are explained in a published paper included as an Appendix.

### **2.2.3 Reductive amination reaction**

CS hexasaccharides and sodium cyanoborohydride were dissolved in 80 % methanol and 4-amino TEMPO was dissolved in 80 % methanol and 16 % glacial acetic acid. 1 mg  $\Delta$ C444S-nr (non-reduced form) was added to 0.4 mL of solution with 73 mM 4-amino TEMPO and 250 mM sodium cyanoborohydride (NaCNBH<sub>3</sub>). This was heated at 65 °C for 3 h under anaerobic conditions, and allowed to cool to 25 °C. The products were subjected to a G-15 desalting column and the fractions were monitored at  $\lambda = 232$  nm of UV absorbance. The fractions of the target sugar were pooled and freeze-dried.

### **2.2.4 Separation of spin-labeled analogues using chromatography**

The product from the reductive amination reaction was dissolved in 100  $\mu$ L of water and subjected to SAX-HPLC on a 250 x 10 mm column using a linear NaCl gradient from 0 to 2 M in H<sub>2</sub>O (pH previously adjusted to 5.0 with HCl). Elution occurred over a 60 min period at a flow rate of 3.0 mL/min. The separations were monitored by UV absorption at  $\lambda = 232$  nm for lyase-derived products. The peaks were

collected separately, concentrated, desalted on a Sephadex G-15 column (500 x 10 mm), and lyophilized.

### **2.2.5 Mass spectrometry**

The samples in 50 % acetonitrile were subjected to nanospray ionization, and analyzed in the negative ion mode with an LTQ-FT hybrid mass spectrometer (Thermo Scientific, Waltham, MA) operating in the linear ion trap only mode. The sample was infused through a hand-pulled silica capillary at a flow rate of ~0.6  $\mu\text{L}/\text{min}$ , with an electrospray voltage of 2.6 kV. Spectra were accumulated for one minute and averaged to increase the signal-to-noise ratio.

### **2.2.6 NMR spectroscopy**

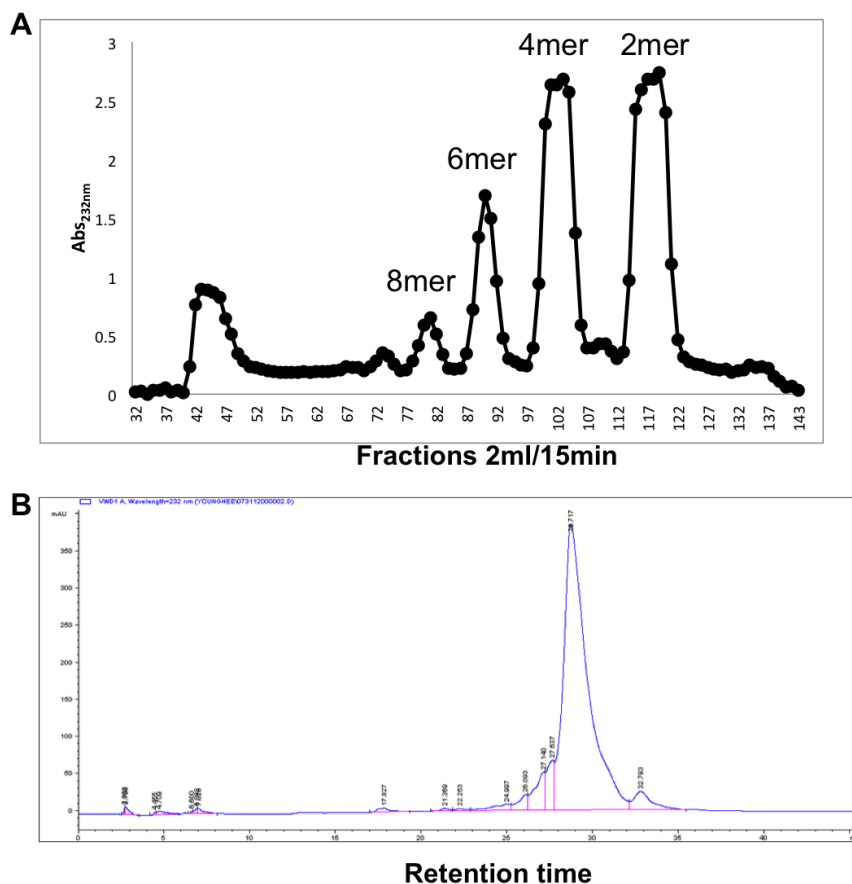
1D proton and 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC experiments were acquired for the chondroitin sulfate and spin-labeled analog at 25 °C on a Varian/Agilent INOVA/DD2 800 MHz spectrometer using a standard sequence from BioPack (Varian/Agilent). 1 mg of  $\Delta\text{C444S}$  and 0.2 mg of  $\Delta\text{C444S-TEMPO}$  dissolved in 99.99 %  $\text{D}_2\text{O}$  in a 3 mm Shigemi NMR tube (Shigemi.inc) were used to collect NMR data.

## **2.3 Results**

### **2.3.1 Preparation of hexasaccharides of chondroitin sulfate**

Representative data are presented here; detailed experimental data are shown in the Appendix. Chondroitinase AC has been shown to exhibit a random endolytic action pattern on chondroitin-6-sulfate (22). Hexasaccharides were separated from the sample after treatment of btCS-A with AC lyase using size-exclusion chromatography (Bio-Gel P-10, Fig. 2.2). The length of the digestion time is short like lyase ABC while the digestion activity is similar to lyase C. The separation profile from SEC is quite simple

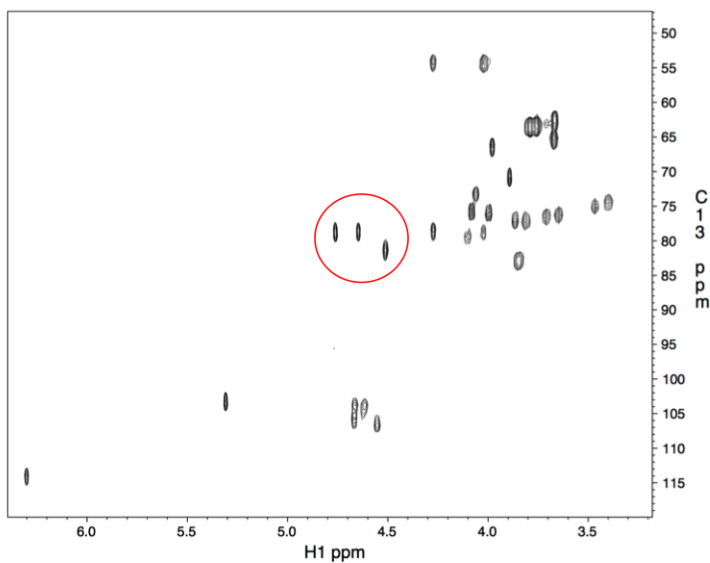
showing that most of the digested oligomers belong to di- and tetra- oligosaccharides; few hexasaccharides were obtained from the digestion (Fig. 2.2). The length of the hexasaccharide was confirmed using mass spectrometry (data not shown). The mixture of hexasaccharides was further separated by strong anion exchange chromatography (SAX), and the separation profile is shown in Fig. 2.2B.



**Figure 2.2:** SEC (A) and SAX (B) profiles of chondroitin sulfate hexasaccharides after chondroitinase AC digestion with btCS-A.

Unexpectedly, only a single major peak was detected during separation. This was further analyzed by 2D NMR spectroscopy in order to confirm its sulfation pattern along the hexasaccharide (Fig. 2.3). As mentioned in the Appendix, a cross peak of C4-H4 at the sulfated carbon on GalNAc is distinct and diagnostic for sulfated carbon. Fig. 2.3 confirms that three C4-H4 cross peaks of GalNAc from the non-reducing to the reducing

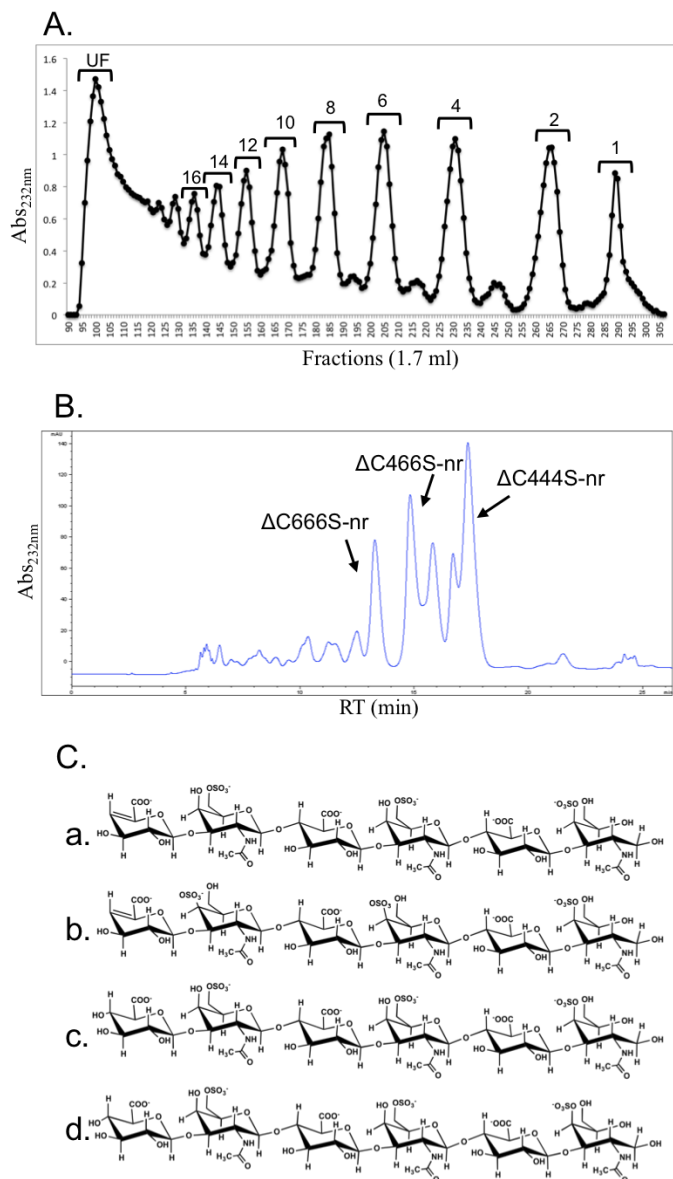
end are observed with the order of the peaks from downfield to upfield in proton dimension, respectively. Given this, it is likely that the hexasaccharide from lyase AC digestion is mostly comprised of C4 sulfated GalNAc. This isomer is exactly what we desired to separate for the study with protein



**Figure 2.3:**  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of  $\Delta\text{C444S}$  after SEC and SAX separations from chondroitinase AC digestion. The cross peaks for C4-H4 of 4-sulfated GalNAc are circled in red.

$\Delta\text{C444S}$  and other isomers of chondroitin sulfate hexasaccharides were prepared for the study with proteins in order to find the specificity of binding preferences for the Link module. Before chondroitinase AC was used for btCS-A digestion, other enzymes such as chondroitinase ABC and C and hyaluronidase had been used to depolymerize native btCS-A and scCS-C. A representation of this digestion and the subsequent production of several hexasaccharide isomers are shown in Fig. 2.4. The SEC profile for chondroitinase C digestion of btCS-A in Fig. 2.4A shows that this enzyme depolymerizes polysaccharides into higher order of longer oligomers than lyase AC (Fig. 2.2). Using SAX-HPLC and based on relative charges, hexasaccharides from SEC were separated

into three major isomers,  $\Delta C666S$ ,  $\Delta C466S$  and  $\Delta C444S$  (Fig. 2.4B). Sulfation patterns for each isomer were confirmed using NMR assignment (see Appendix).



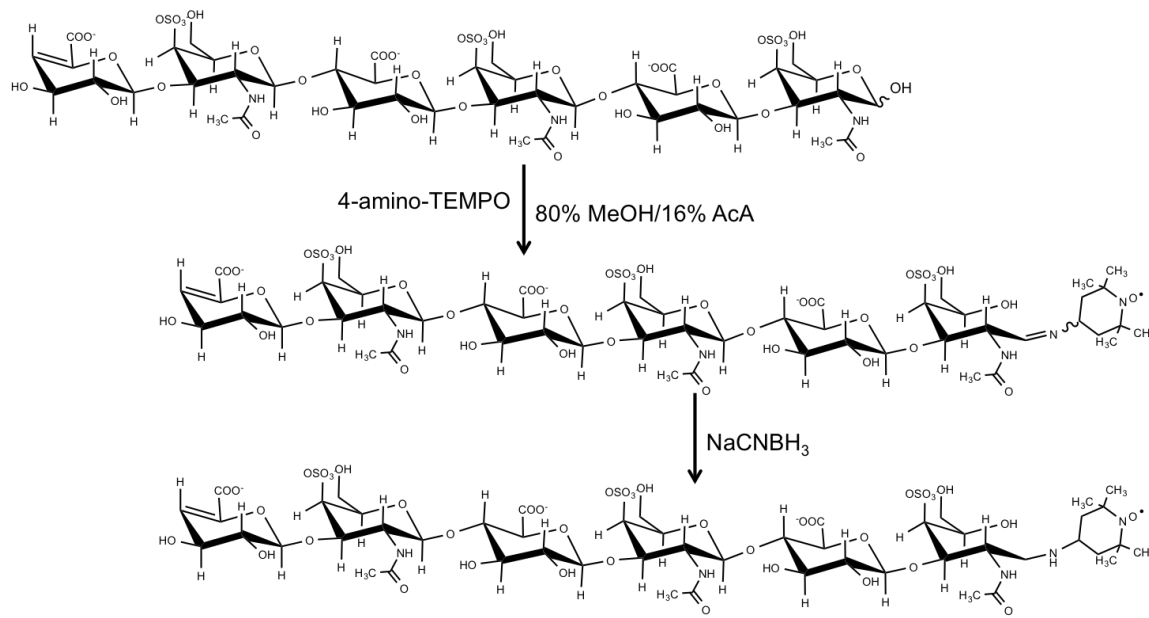
**Figure 2.4:** Size-exclusion (A) and SAX chromatography analyses of the products from chondroitinase C digestion of btCS-A. (A) Gel filtration chromatography on Bio-Gel P-10 of the CS digested over 2 Days. The numbers on top of each peak represent the number of sugar rings. (B) SAX-HPLC of the fraction of hexasaccharides from size-exclusion chromatography. Homogeneous hexasaccharides were obtained with a gradient of 2 M NaCl from 25 to 45%. (nr: non-reduced end). (C) Structures of sugars are drawn in ChemDraw (used for titrations and structure determination): a- $\Delta C664S$ ; b- $\Delta C444S$ ; c- $C664S$ ; d- $C444S$ .



From these enzyme degradation assays, four isomers of chondroitin sulfate were obtained and assigned for the study of the interaction with the Link module as described in Chapter 4. The structures of four CS hexasaccharides are shown in Fig 2.4C (a-d). More detailed methods and results are provided in the Appendix.

### 2.3.2 Synthesis of spin-labeled glycosaminoglycans

4-amino-TEMPO is a very stable spin-labeled material that has an unpaired electron. Due to the solubility limitations of the sugar, 80% MeOH was used to dissolve all materials that were used in this reaction. Since better yields require low pH 4-5, 16% acetic acid was added to the TEMPO preparation. As in most reductive amination reactions, an excess of amine was added to the sugar to form a Schiff base; approximately 100-fold molar excess 4-amino-TEMPO was used. The imine was reduced to the corresponding aminodeoxyalditol using  $\text{NaCNBH}_3$  (Fig. 2.5).

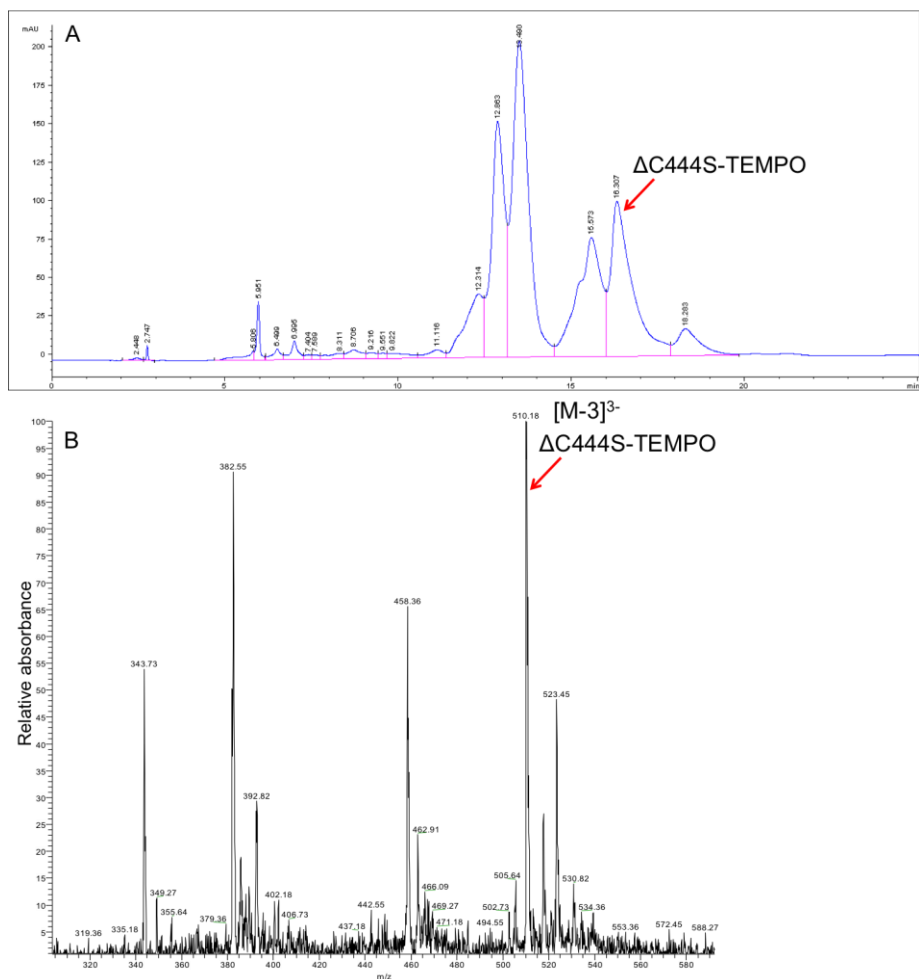


**Figure 2.5:** Reductive amination of  $\Delta\text{C444S}$  with 4-amino-TEMPO.

The mixture of the sugar, 4-amino-TEMPO, and NaCNBH<sub>3</sub> was heated at 65 °C for 3 h, then cooled to room temperature and passed through a desalting column. Figure 2.5 depicts this reductive amination reaction.

### 2.3.3 Separation of spin-labeled-GAG

Lyophilized material from the desalting column was dissolved in 100 µl H<sub>2</sub>O and the sample was subjected to SAX chromatography. Figure 2.6A shows the separation of the reactants from the reductive amination reaction. Because TEMPO materials were retained on the column longer than the corresponding sugars without TEMPO, they were eluted later than sugar as shown in Fig. 2.6A.

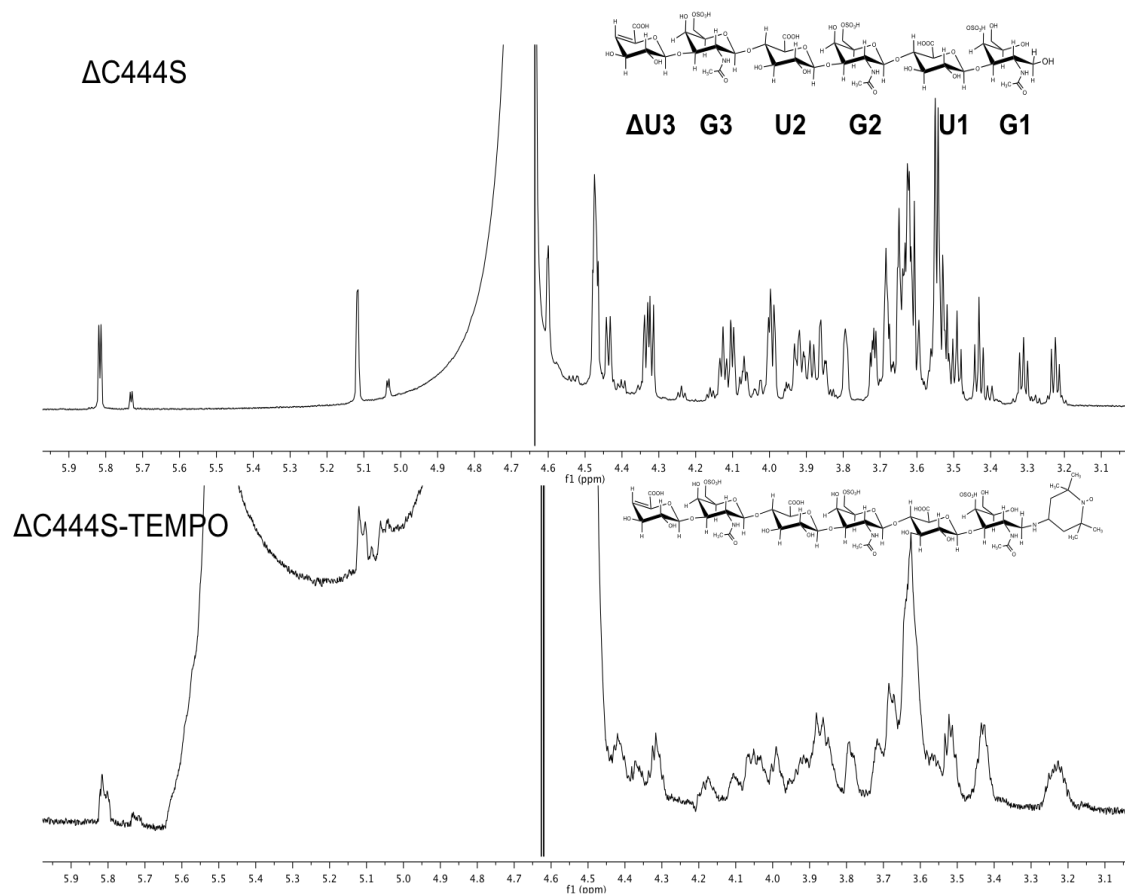


**Figure 2.6:** SAX-HPLC chromatograph (A) and Mass spectrum (B) of ΔC444S-TEMPO. The red arrow indicates the target peak, ΔC444S-TEMPO.

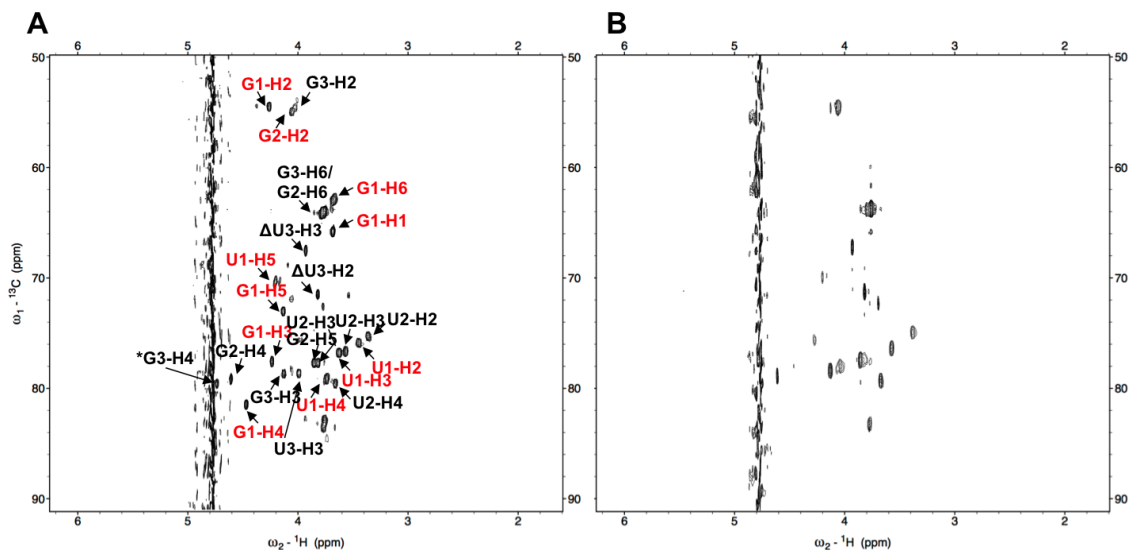
The peak of the spin-labeled derivative  $\Delta$ C444S-TEMPO is indicated by the red arrow. ESI-MS confirmed its mass to be 1533.54 (Fig. 2.6B). Using mass spectrometry, the remaining peaks from the SAX column were confirmed to be isomers of  $\Delta$ C444S or the Schiff-base form of  $\Delta$ C444S-TEMPO. The first step in the Schiff-base reaction is reversible; this reaction is normally conducted under anaerobic conditions. Although the yield of this reaction typically is more than 90 % in anaerobic conditions with dry solvent, a poor yield (40%) was achieved in this study. This might have resulted from solubility issues of the GAG in 80 % MeOH, producing an unfavorable environment for this reaction.

The existence of the TEMPO moiety at the reducing end of  $\Delta$ C444S was examined by 1D  $^1\text{H}$  and 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC experiments. The spectra show poor water suppression around peak 4.7 ppm. This is due to incomplete water removal when the sample was prepared by exchanging water to 99 %  $\text{D}_2\text{O}$ . However, the peaks from the samples are observable and analyzable. As shown in Fig. 2.7, a 1D proton spectrum for  $\Delta$ C444S was acquired with decent resolution showing clear sharp peaks. Otherwise, most of the peaks from  $\Delta$ C444S experienced line broadening due to the paramagnetic enhancement relaxation effects of the free radical on TEMPO. Several peaks completely disappeared which are mostly from the rings close to the reducing end. This was clearly confirmed by 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra (Fig. 2.8). A complete assignment of  $\Delta$ C444S with labels is given in Fig. 2.8A. The region of  $^{13}\text{C}$  dimension from 50 to 92 ppm and of  $^1\text{H}$  dimension from 1.6 to 6.2 ppm is enlarged, so that the C1-H1 cross peaks from the all sugar rings except  $\Delta$ U3, and C4-H4 and C1-H1 cross peaks from  $\Delta$ U3 are excluded. As expected, most peaks from the G1 and U1 rings were perturbed and lost their cross peaks

(Fig. 2.8B). Some of the peaks from the G2 and U2 rings experienced PRE effects from TEMPO, showing reduced peak intensity as well as peak broadening. As previously mentioned, the TEMPO derivative remained in its oxidized form in solution. As shown in Fig 2.7 and 2.8, it is likely that almost all resonances were perturbed due to the PRE effect.



**Figure 2.7:** 1D proton spectrum of  $\Delta C444S$  (upper panel) and  $\Delta C444S\text{-TEMPO}$  (lower panel). Insets at the right corner in A and B show the structures for  $\Delta C444S$  and  $\Delta C444S\text{-TEMPO}$ .



**Figure 2.8:**  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra of  $\Delta\text{C444S}$  (A) and  $\Delta\text{C444S-TEMPO}$  (B). Peak assignments are labeled for  $\Delta\text{C444S}$  and peaks perturbed by TEMPO are labeled in red. Asterisk indicates that the peak is suppressed with water.

## 2.4 Discussion

Homogeneous hexasaccharides were separated successfully using enzymatic degradation of btCS-A and scCS-C. With these approaches, not only fully 4-sulfated chondroitin sulfate but also other isomers of various combination of 4- and 6-sulfated chondroitin sulfate were homogeneously prepared. Also, we found that lyase AC treatment of btCS-A yielded primarily fully 4-sulfated chondroitin hexasaccharide. This satisfied our need for fully 4-sulfated hexasaccharide of CS to be used in various experiments in the study of Link module - CS interaction.

Even though reductive amination reactions are a standard method of producing amines in both academic and pharmaceutical laboratories, there are only a few cases in which TEMPO-labeled glycans have been used to study a glycan-protein complex (23). Here, we successfully synthesized a TEMPO-labeled chondroitin sulfate hexasaccharide. Before working with the hexasaccharide, the reductive amination reaction was tested with

a disaccharide of chondroitin sulfate using the methods described above (data not shown). In this case, the yield of the reaction was more than 80 % and the separation of the products using SAX-HPLC was clear. However, there were difficulties dissolving the hexasaccharide in dry ethanol. To increase the solubility of the hexasaccharide, we used 80 % MeOH. As the results show, the yield was still low. However, NMR confirmed that a TEMPO analog was successfully synthesized. More extensive work is needed to increase the yield for this reaction with unsaturated GAGs. The separated CS isomers and synthesized TEMPO derivative were used for the protein and GAG binding studies and structure determinations presented in Chapters 3 and 4.

## **2.5 References**

1. Linhardt, R. J., and Toida, T. (2004) Role of glycosaminoglycans in cellular communication. *Acc Chem Res* **37**, 431-438
2. Silbert, J. E., and Sugumaran, G. (1995) Intracellular membranes in the synthesis, transport, and metabolism of proteoglycans. *Biochim Biophys Acta* **1241**, 371-384
3. Silbert, J. E. (1996) Organization of glycosaminoglycan sulfation in the biosynthesis of proteochondroitin sulfate and proteodermatan sulfate. *Glycoconj J* **13**, 907-912
4. Kitagawa, H., Izumikawa, T., Uyama, T., and Sugahara, K. (2003) Molecular cloning of a chondroitin polymerizing factor that cooperates with chondroitin synthase for chondroitin polymerization. *J Biol Chem* **278**, 23666-23671
5. Baeurle, S. A., Kiselev, M.G., Makarova, E.S., and Nogovitsin, E.A. (2009) Effect of the counterion behavior on the frictional-compressive properties of chondroitin sulfate solutions. *Polymer* **50**, 1813
6. Gemma, E., Meyer, O., Uhrin, D., and Hulme, A. N. (2008) Enabling methodology for the end functionalization of glycosaminoglycan oligosaccharides. *Molecular bioSystems* **4**, 481-495
7. Handel, T. M., Johnson, Z., Crown, S. E., Lau, E. K., and Proudfoot, A. E. (2005) Regulation of protein function by glycosaminoglycans--as exemplified by chemokines. *Annual review of biochemistry* **74**, 385-410

8. Volpi, N. (2006) Therapeutic applications of glycosaminoglycans. *Current medicinal chemistry* **13**, 1799-1810
9. Herbert, A. P., Uhrin, D., Lyon, M., Pangburn, M. K., and Barlow, P. N. (2006) Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. *J Biol Chem* **281**, 16512-16520
10. Prosser, B. E., Johnson, S., Roversi, P., Herbert, A. P., Blaum, B. S., Tyrrell, J., Jowitt, T. A., Clark, S. J., Tarelli, E., Uhrin, D., Barlow, P. N., Sim, R. B., Day, A. J., and Lea, S. M. (2007) Structural basis for complement factor H linked age-related macular degeneration. *J Exp Med* **204**, 2277-2283
11. Rice, K. G. (2000) Derivatization strategies for preparing N-glycan probes. *Anal Biochem* **283**, 10-16
12. Anumula, K. R. (2006) Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates. *Anal Biochem* **350**, 1-23
13. Monsigny, M., Quetard, C., Bourgerie, S., Delay, D., Pichon, C., Midoux, P., Mayer, R., and Roche, A. C. (1998) Glycotargeting: the preparation of glyco-amino acids and derivatives from unprotected reducing sugars. *Biochimie* **80**, 99-108
14. Seyfried, N. T., Blundell, C. D., Day, A. J., and Almond, A. (2005) Preparation and application of biologically active fluorescent hyaluronan oligosaccharides. *Glycobiology* **15**, 303-312
15. Shaklee, P. N., and Conrad, H. E. (1984) Hydrazinolysis of heparin and other glycosaminoglycans. *Biochem J* **217**, 187-197
16. Yu, F., Wolff, J. J., Amster, I. J., and Prestegard, J. H. (2007) Conformational preferences of chondroitin sulfate oligomers using partially oriented NMR spectroscopy of <sup>13</sup>C-labeled acetyl groups. *J Am Chem Soc* **129**, 13288-13297
17. Jackson, P. (1994) High-resolution polyacrylamide gel electrophoresis of fluorophore-labeled reducing saccharides. *Methods in enzymology* **230**, 250-265
18. Mourier, P. A., and Viskov, C. (2004) Chromatographic analysis and sequencing approach of heparin oligosaccharides using cetyltrimethylammonium dynamically coated stationary phases. *Anal Biochem* **332**, 299-313

19. Vynios, D. H., Karamanos, N. K., and Tsiganos, C. P. (2002) Advances in analysis of glycosaminoglycans: its application for the assessment of physiological and pathological states of connective tissues. *J Chromatogr B Analyt Technol Biomed Life Sci* **781**, 21-38
20. Drummond, K. J., Yates, E. A., and Turnbull, J. E. (2001) Electrophoretic sequencing of heparin/heparan sulfate oligosaccharides using a highly sensitive fluorescent end label. *Proteomics* **1**, 304-310
21. Barriga, S. (2001) 2,2,6,6-Tetramethylpiperidine-1-oxyl (TEMPO). *Synlett* **4**, 563
22. Jandik, K. A., Gu, K., and Linhardt, R. J. (1994) Action pattern of polysaccharide lyases on glycosaminoglycans. *Glycobiology* **4**, 289-296
23. Macnaughtan, M. A., Kamar, M., Alvarez-Manilla, G., Venot, A., Glushka, J., Pierce, J. M., and Prestegard, J. H. (2007) NMR structural characterization of substrates bound to N-acetylglucosaminyltransferase V. *J Mol Biol* **366**, 1266-1281



## **CHAPTER 3**

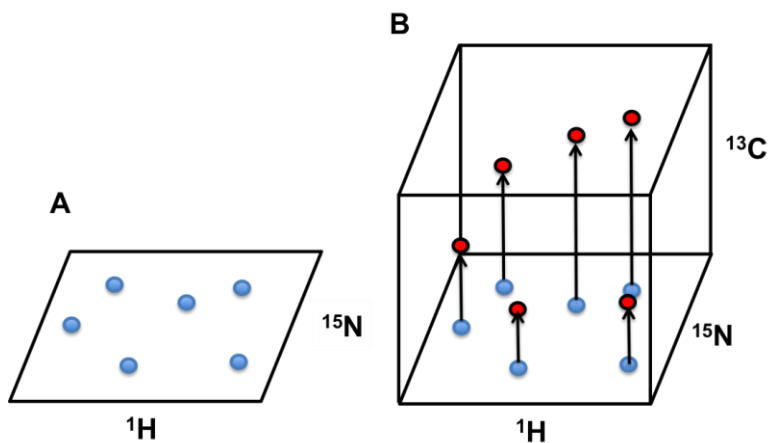
### **NMR RESONANCE ASSIGNMENTS OF UNIFORMLY LABELED LINK**

#### **MODULE OF TSG-6**

#### **3.1 Introduction**

Since 1970s, more than 100,000 structures of proteins and protein-nucleic acid complexes have been deposited in Protein Data Bank (<http://www.rcsb.org/>). There are three main approaches for determining protein structures. The most widely used is X-ray crystallography. Once a sample is crystallized it is often straightforward to solve the structure. However, not all systems form suitable crystals and if dynamics as well as structure are of interest methods based on NMR may be appropriate. Nuclear Magnetic Resonance (NMR) spectroscopy is the second most frequently used structural method. Approximately 10,000 structures currently deposited in the Protein Data Bank have been determined by NMR spectroscopy. As the importance of understanding the effect of protein dynamics on physiological function grows, NMR continues to become a method of choice. However, the weaknesses of NMR include its size limitation and low sensitivity. Early applications of protein NMR were based on  $^1\text{H}$  homonuclear NMR spectroscopy; for example, COSY (CORrelation SpectroscopY), TOCSY (TOfal Correlation SpectroscopY), and NOESY (Nuclear Overhauser Effect SpectroscopY), were utilized to determine protein structure. As protein size increases beyond 20 kDa, overlap and broadening of signals due to increasing T2 relaxation rates are observed. The spectrum becomes crowded and assignment accuracy is compromised. These issues have been resolved by adapting a triple resonance assignment strategy that allow  $^1\text{H}$  spectra to

be edited with respect to  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts and simplified. A 3D experiment is obtained by concatenating two 2D experiments, so that the 2D is extended into a third dimension. For instance, the HNCO experiment, is basically a 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum, where x and y axes are  $^1\text{H}$  and  $^{15}\text{N}$  of i residue, respectively with  $^{13}\text{C}$  chemical shift of carbonyl carbons added in an indirectly detected z dimension. The peaks on the 2D plane move along the z axis to the  $^{13}\text{C}$  ppm value for the CO group of residue i-1. This carbon is selectively detected due to the large scalar coupling of the directly bonded C and N atoms in the amide bond. The rest of the 3D experiments rely on similar principles with different observable nuclei at x, y and z axes depending on the researcher's objective. These 3D NMR techniques as well as 4D experiments have enhanced the accuracy of determining protein structures.



**Figure 3.1:** Simplified schematic of 3D NMR spectroscopy. A) representation of a 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum. Blue circles indicate individual cross peaks of amide proton and nitrogen on a protein backbone. B) representation of 3D HNCO - cross peaks have moved along the axis for carbon.

Most structural data for NMR experiments have come from Nuclear Overhauser Effects (NOEs) where connectivities are established by the  $1/r^6$  dependence of magnetization transfer from one proton to another. A 3D  $^{15}\text{N}$  edited NOESY spectrum has one  $^{15}\text{N}$  dimension and two  $^1\text{H}$  dimension.

A method of growing importance in determining a structure is molecular modeling. It has been used in cases of the failure to achieve structures from both X-ray crystallography and NMR. This approach uses known structures (templates) with >20 % sequence identity to produce a structural model of a target protein since a high level of sequence similarity between template and target usually implies significant structure similarity (1). Electron microscopy may also be used in combination with X-ray crystallography or NMR spectroscopy to define atomic information for macromolecular structures.

As mentioned previously, the Link module is an HA binding protein with structures determined by X-Ray crystallography and NMR spectroscopy (2PF5.pdb and 1O7C.pdb, respectively) (2,3). In these structures, Link module has been extensively studied in ligand free forms (both X-Ray and NMR studies) and HA bound forms (NMR study). In addition, docking predictions and site-directed mutagenesis show that the interaction between Link-TSG-6 and heparin/heparan sulfate (Hp/HS) occurs at a distinct site from the HA binding groove (4). In addition to HA and Hp/HS, chondroitin sulfates (CSs) are another type of glycosaminoglycan that binds to Link in the ECM. Although an interaction has been confirmed by previous studies (5-9), there is a lack of structural information to explain the mechanism of this binding.

There are three reasons to pursue further structural studies of the Link module in complex with CS. Firstly, the binding sites for CS hexasaccharide that we defined with chemical shift perturbation experiments appear to differ slightly from those of HA. Secondly, the experimental RDCs that we obtained did not fit well into existing structures, suggesting that there is a huge possibility of a structural change in the Link module in the

presence of CS. Lastly, we found that binding of CS hexasaccharide to Link module seems to induce dimerization of the protein. Taken all together, there is a need for a new structure in the presence of a well-defined CS hexasaccharide to understand the behavior of Link module in the interaction with CS. Herein, we used a NMR methodology to determine the Link module structure in solution.

Link module of TSG-6 is of a size, 11 kDa that is amenable to study using NMR structural methods. To begin, standard 2D and 3D experiments,  $^1\text{H}$ - $^{15}\text{N}$  HSQC, HNCOC, CBCA(CO)NH, HNCACB,  $^{13}\text{C}$ -edited HSQC-NOESY,  $^{15}\text{N}$ -edited HSQC-NOESY and HCCH-TOCSY were collected. The name of the experiments implies the route for magnetization transfer among the nuclei of interest. Most protein assignments begin with the  $^1\text{H}$ - $^{15}\text{N}$  HSQC (heteronuclear single quantum coherence) experiment which gives cross peaks of amide protons and nitrogen in the peptide bond. Each residue except proline (that lacks an HN group) produces an individual signal in this spectrum, observing the peak dispersion and counting the number of peaks give a quick assessment of the folding status of a protein and the quality of the sample. Optimized parameters and settings are needed for each experiment because the relative sensitivity varies with respect to the type according to the experiments.

For 3D experiments, protein backbone assignments are conducted first using the appearance of CA and CB peaks of identical chemical shift connected with H-N crosspeaks of *i* and *i*-1 residues. For this,  $^1\text{H}$ - $^{15}\text{N}$  HSQC, HNCACB, and CBCA(CO)NH spectra are acquired and screened to find matching peaks. HNCACB shows correlations between amide peaks of the residue (*i*) and the CA and CB of the same residue (*i*) as well as those of the preceding residue (*i*-1). CBCA(CO)NH is a companion experiment to

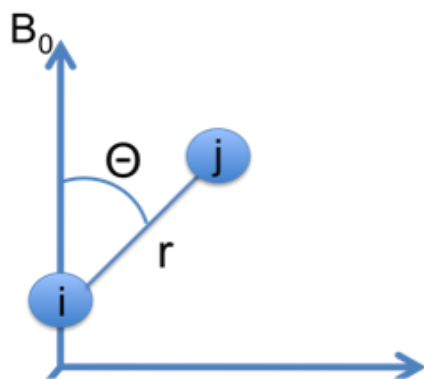
HNCACB and identifies the CA and CB of the preceding residue (i-1). The HNCO experiment provides correlations between amide protons and nitrogen and the carbonyl carbon of preceding residues (i-1). Therefore, carbonyl carbon chemical shifts can be generally assigned in HNCO experiments. With these experiments, sequential assignment of the protein backbone can be completed. For the side chains of amino acids, the HCCH-TOCSY experiment provides  $^1\text{H}$  and  $^{13}\text{C}$  connectivity to side chain sites. The detailed analysis based on experimental data is provided in the results section of this chapter.

For all experiments described above, magnetization transfers occur through large heteronuclear one-bond couplings. However, as more magnetization transfer steps are set in the pulse sequences of the experiment, a reduced amount of returning magnetization is detected, resulting in a poor signal-to-noise ratio which can affect the quality of the spectrum. To compensate, concentrated samples of protein are required to increase the sensitivity of these experiments.

Once the chemical shift assignment is completed, distance restraints are collected using  $^1\text{H}$ - $^{15}\text{N}$ -edited HSQC-NOESY and  $^1\text{H}$ - $^{13}\text{C}$ -edited HSQC-NOESY experiments. These are comprised of a homonuclear pulse sequence of NOESY and an HSQC pulse sequence for  $^1\text{H}$ - $^{15}\text{N}$  and  $^1\text{H}$ - $^{13}\text{C}$  spins. In addition, these resolve crosspeaks between protons ( $^1\text{H}$ ) according to the chemical shift of the heteronuclei bonded directly to one of the protons. In  $^1\text{H}$ - $^{15}\text{N}$ -edited HSQC-NOESY spectra, HA-HN and HN-HN cross peaks are dominantly seen as well as cross peaks for other pairs of protons close in space. In the  $^1\text{H}$ - $^{13}\text{C}$ -edited HSQC-NOESY spectrum, NOE cross peaks between any H and aliphatic Hs are observed. These are the standard methods determining structures in

solution, providing that the NOE cross peaks have distance between nuclei within 2-5 Å. Moreover, these will be used in structural calculations for generating reasonable structures.

Residual Dipolar Coupling (RDC) is also informative for structure determination and has been used to restrict the orientations of bond vectors between pairs of spin  $\frac{1}{2}$  nuclei to produce data that describe the relative geometry of remote parts of the ligand and the orientation of a ligand relative to a protein. Due to Brownian motion, dipolar couplings average to zero under isotropic conditions in solution, but they can be observed under anisotropic conditions in alignment media such as bicelles, bacteriophages polyacrylamide gels, or alkyl-polyethyleneglycol ether (PEG) dispersions. The selection of a particular medium is chosen based on the properties of the protein. Dipolar couplings are potentially quite large interactions and are caused by the magnetic field produced by one nucleus (i) affecting the energy of another nuclear spin state (j) (Fig. 3.2).



**Figure 3.2:** Dipolar coupled spin pair. i and j indicate NMR active nuclei with a distance of r.  $\theta$  is an angle between an internuclear vector and the magnetic field ( $B_0$ ).

RDCs are useful for structure characterization because they depends on distance, orientation, and dynamics. The orientational dependence of RDCs stems from a

contribution to couplings that is proportional to the average of  $(3\cos^2\theta - 1)/2$  where  $\theta$  is the angle between the external magnetic field ( $B_0$ ) and the vector of interest (Eq.1). The average angular dependence of the RDC for a pair of spin 1/2 nuclei can be described by Equation 1

$$D_{ij} = -\frac{\mu_0\gamma_i\gamma_j h}{(2\pi r_{ij})^3} \cdot \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle$$

**Equation 1.**

where  $D_{ij}$  is the residual dipolar coupling in Hz between nuclei  $i$  and  $j$ ,  $\gamma_i$  and  $\gamma_j$  are nuclear gyromagnetic ratios,  $r_{ij}$  is the internuclear distance, and  $h$  is Planck's constant. The brackets denote a time average of the quantity. From a series of  $D_{ij}$ s, the information on relative angles  $\theta$ , are provided. Many measurements of RDCs are made between pairs of bonded nuclei, so that  $r$  is fixed. RDCs have, thus, been used primarily to provide angular information. Knowing angle  $\theta$  is useful for defining a molecule structure. There have been many applications of RDCs for determining solution structures of proteins without a ligand whereas there have been few applications to determine solution structures of protein bound ligands (10). RDCs can define the average orientation of the bound ligand with respect to the protein. Establishing the orientation of a ligand with respect to a protein may make it possible to determine complexed protein structures. RDCs can be measured by various NMR experiments. Since RDC between two J-coupled spins,  $i$  and  $j$ , has an important property that  $D_{ij}$  is added to the isotropic coupling constant,  $J$ , effective coupling constants,  $J_{\text{eff}} = J + D_{ij}$  can be measured. Therefore, measurements of  $J$  in isotropic solution and  $J_{\text{eff}}$  in anisotropic solution are needed to derive a  $D_{ij}$  value for structure determination. RDCs are an ideal complement to NOEs

that are traditionally used as distance constraints for structure determination.

With all assignments made and restraints collected, structural calculation and refinement can be conducted. The chemical shifts of HA, CA, CB, CO, and N in a protein backbone rely on secondary structures and are used to predict the secondary structure of the protein. Normally, once all backbone heavy atoms are assigned, TALOS (Torsion Angle Likelihood Obtained from Shift and Sequence Similarity, <http://spin.niddk.nih.gov/bax/nmrserver/talos/>) (11) is used to predict phi (N-C-CA-N) and psi (C-CA-N-C) backbone torsion angles based on the chemical shifts (HN, HA, HB, CA, CB, and N) and sequence information. The Link module structure bound to CS hexasaccharide was calculated using CYANA (Combined Assignment and Dynamics Algorithm for NMR Applications) Version 3.0 (12-15). This program is for automated structure calculation on the basis of conformational restraints made by NMR. Link module soluble structures were further refined using XPLOR-NIH version 2.28 (16); all structure calculations were carried out with NOEs, torsion angles, and RDC data. Structural statistics and global structure quality were computed using the PSVS (Protein Structure Validation Server) version 1.5 software package (17). This evaluates the quality of a protein structure and provides quality scores. The final ensemble of 20 refined models with lowest total energy were chosen and deposited in the Biological Magnetic Resonance Data Bank (BMRB).

### **3.2 Experimental Procedures**

#### **3.2.1 Expression, purification and refolding of $^{13}\text{C}$ , $^{15}\text{N}$ labeled Link-TSG-6**

The engineered gene of Link-TSG-6 provided by Dr. Tony Day's laboratory in the pRK172 expression vector was transformed into the expression host



BL21(DE3)pLysS. The starter cultures were grown in LB medium containing 100 µg/ml ampicillin, overnight. 10 ml of grown cells were transferred to 1 L M9 medium enriched with 1 g  $^{15}\text{N}$  labeled  $\text{NH}_4\text{Cl}$  and 2 g  $^{13}\text{C}$  labeled glucose and protein expression was induced by adding 0.1 mM IPTG when  $\text{OD}_{600\text{nm}}$  was 0.4. Cells were harvested 4 h after induction by centrifugation for 15 min at 5000 g and stored at  $-20\text{ }^\circ\text{C}$  in lysis buffer (50 mM Tris-HCl, pH 7.8, 150 mM NaCl, and 1 mM EDTA). Cells were lysed by three passages through a French press cell then inclusion bodies were gathered by centrifugation for 45 min at 20000 g (18) and solubilized in 6 M Guanidine-HCl containing 50 mM Tris-HCl, pH8.0, and 100 mM DTT. Solubilized inclusion bodies were loaded onto a Superdex 75 (16 x 260 mm, Pharmacia) exclusion column, equilibrated and run with a flow rate at 1ml/min in the same buffer as above without DTT. Fractions containing the target protein were injected onto a C4 column (10 x 250 mm, YMC) and equilibrated with 100%  $\text{H}_2\text{O}$  containing 0.01% TFA, with a flow rate of 3ml/min. After 5 min, the protein was eluted by linear gradients from 0 to 80% acetonitrile containing 0.1% TFA over 40 min. The eluent was monitored at 220nm continuously. Eluent containing the Link module was collected and lyophilized.

Since the Link module protein was induced in *E. coli* it could not be produced as a 100% folded protein. Link-TSG-6, therefore, was refolded from its denatured state by resuspending it in 50 mM ammonium acetate at pH 6.0 to a concentration of 500 µg/ml (45.78 µM, 40 ml final volume). A 100-fold molar excess of  $\beta$ -mercaptoethanol (12.8 µl) was added, and the refolding solution was incubated at  $25\text{ }^\circ\text{C}$  for 2 days under aerobic conditions without stirring. After 2 days, the refolding solution was stored at  $4\text{ }^\circ\text{C}$  for 5 days. The folded Link-TSG-6 protein was purified by HPLC-C4 column.

### **3.2.2 Preparation of chondroitin sulfate hexasaccharide**

The procedure for the preparation of  $\Delta$ C444S can be found in Chapter 2 and Appendix.

### **3.2.3 NMR spectroscopy**

$^1\text{H}$ - $^{15}\text{N}$  HSQC, HNCOC, CBCA(CO)NH, NHCACB and HCCH-TOCSY NMR spectra were acquired on a Varian 600 spectrometer at 25 °C.  $^{15}\text{N}$ ,  $^{13}\text{C}$ -edited HSQC-NOESYs were acquired on a Varian 800 spectrometer at 25 °C. Spectra were processed with nmrPipe and analyzed using Sparky. Complete backbone and side chain resonance assignments were obtained with 0.5 mM  $^{13}\text{C}$  and  $^{15}\text{N}$  enriched Link-TSG-6 in a buffer of 50 mM MES at pH 6 in the presence of 0.6 mM  $\Delta$ C444S using the following heteronuclear 2D and 3D experiments;  $^1\text{H}$ - $^{15}\text{N}$  HSQC, HNCOC, CBCA(CO)NH, NHCACB,  $^{15}\text{N}$ ,  $^{13}\text{C}$ -edited HSQC-NOESYs, and HCCH-TOCSY. RDCs were measured on a 0.5 mM Link-TSG-6 sample in the presence 1 mM sugar a buffer of 50 mM MES at pH 6 aligned in stretched neutral polyacrylamide gel using a two-stage gel NMR tube described in Liu et al. (19). RDC values were measured using an IPAP-HSQC (in-phase and anti-phase) experiment (20).

### **3.2.4 Structure calculations**

Chemical shift assignment was carried out using Sparky (21) with a series of spectra,  $^1\text{H}$ - $^{15}\text{N}$  HSQC, HNCOC, CBCA(CO)NH, NHCACB,  $^{15}\text{N}$ ,  $^{13}\text{C}$ -edited HSQC-NOESYs, and HCCH-TOCSY. These assignments are essential for structure determination using structure determination programs. Protein backbone torsion angles of the protein backbone were calculated by TALOS and measured chemical shift for HA, CA, CB, CO, and N atoms (<http://spin.niddk.nih.gov/bax/nmrserver/talos/>). TALOS also

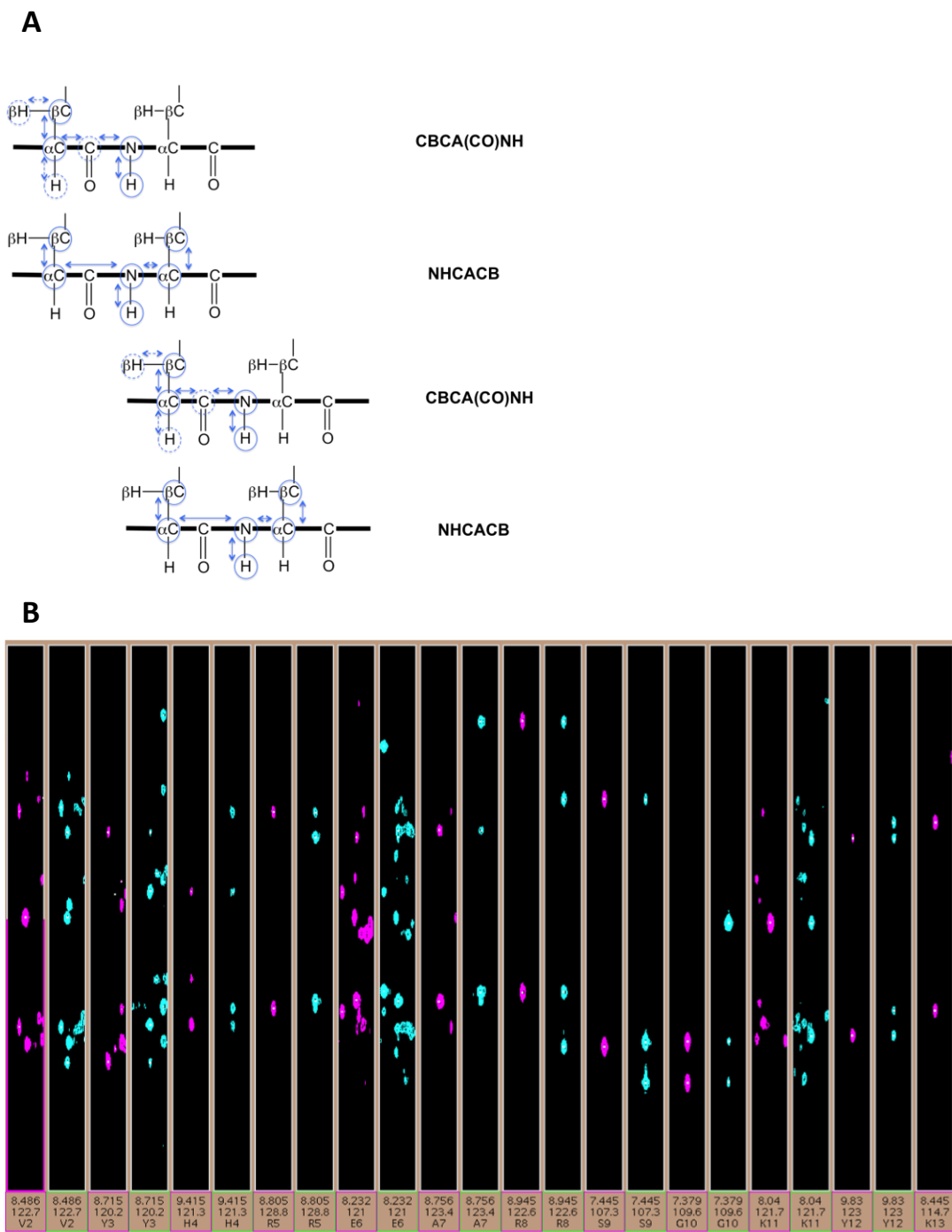


The sequence of the Link module is given below.

G<sup>1</sup>VYHREARSGKYKLTYAEAKAVCEFE<sup>98</sup>GGHLATYKQLEAARKIGFHVCAAGWM  
AKGRVGYPIVKPGPNCGFGKTGIIDYGIRLNRSERWDAYCYNPHAK<sup>98</sup>.

Excluding prolines (four prolines exist in Link), 94 cross peaks should have been observed. However, only 88 H-N cross peaks were assigned. The resonances of G1, V62, G65, G69, F70, G71, and L82 were not observed. This may have been due to chemical exchange broadening or from chemical shift occurring in a region with extensive overlap or interference from the water resonance.

The HNCACB correlates each NH group with the CA and CB chemical shifts of its own residue (intense peaks) and of the preceding residue (weak peaks). Then, the CBCA(CO)NH confirms the assignment by showing only the correlation between a given H-N group and the CA and CB chemical shifts of the preceding residue (Fig. 3.4A). The strip plots of CBCA(CO)NH and HNCACB spectra for residues from V2 to K13 are shown in Fig 3.4B. In the same NH strip for the V2 resonance (8.48 and 122.7 ppm), one peak was found in the CBCA(CO)NH to be the CA (43.39 ppm) of G1, and three peaks were observed in HNCACB. One was CA of the preceding residue, G1, and two were CA (60.54 ppm) and CB (33.3 ppm) of V2. Y3 was assigned by searching for a H-N cross peak that has CA (60.54 ppm) and CB (33.3 ppm) correlations of V2 in the CBCA(CO)NH spectrum. These were found at 8.715 and 120.19 ppm for H and N, respectively. Then this confirmed by the NHCACB whether it had CA and CB cross peaks that belongs to Y3. Reasonable values of chemical shifts for CA and CB of Tyrosine were observed (56.02 and 40.32 ppm, respectively) so that assignment could be confirmed. The remaining residues were assigned in the same manner.



**Figure 3.4:** Backbone sequential assignment of the Link module of TSG-6 using 3D experiments of CBCA(CO)NH (magenta) and HNCACB (cyan) (A), and strip plot of CBCA(CO)NH, HNCACB (B) for the residues from V2 to K13. Blue circles and arrows indicate the transferring of magnetization through bonds. Dotted circles and arrows indicate undetected nuclei and transferring magnetization between detected and undetectable nuclei, respectively.

Average values and statistics for random coil backbone chemical shifts for individual amino acids are shown in Table 3.1.

**Table 3.1:** CA and CB chemical shifts for random coil backbone (22).

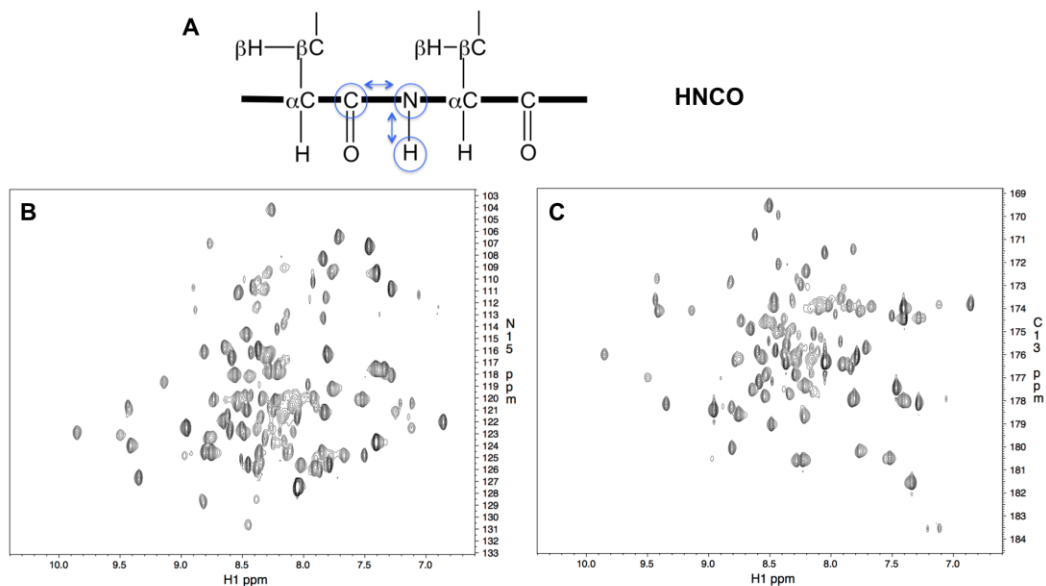
Name of residue	CA (ppm)	CB (ppm)
Ala	52.1	19.3
Cys (Reduced/Oxidized)	58.2/55.3	29.4/40.5
Asp	53.8	41.2
Glu	56.3	30.3
Phe	57.2	30.3
Gly	45.2	
His	55.3	30.1
Ile	60.4	38.7
Lys	56.2	32.8
Leu	54.5	42.5
Met	55.4	33.7
Asn	53.0	38.9
Pro	62.6	31.9
Gln	55.5	29.4
Arg	55.9	31.0
Ser	58.1	64.1
Thr	60.9	69.7
Val	61.4	32.8
Trp	57.3	30.4
Tyr	57.6	39.4

Even though this is specific for residues in a random coil, it suggests possible chemical shift ranges for residues of a folded protein and was used as a reference for all assignments in this study. Moreover, the presence of a disulfide bonds can be confirmed by the chemical shifts of CA and CB. In general, chemical shifts of CA and CB of the cysteine in a disulfide bond are observed further downfield than those with a free cysteine. The Link module has four cysteines at residues 23, 47, 68 and 92, and forms two disulfide bonds between 23 and 92, and 47 and 68. Each CA and CB of each cysteine

agrees with the chemical shift ranges provided by BMRB (22): C23 (55.22/33.56), C47 (55.39/44.58), C68 (54.46/44.32), and C92 (50.57/37.74) were assigned. If there is no proline between residues, a sequential assignment can be carried out in single procession. However, four prolines exist at positions 60, 64, 66 and 95 in the sequence of the Link module. Because of this, as well as the absence of certain cross peaks in  $^1\text{H}$ - $^{15}\text{N}$  HSQC, making assignments for the region between 60 and 70 was challenging. Chemical shifts of CA and CB for disappearing resonances (G1, V62, G71, and L82) could be assigned using data from the CBCA(CO)NH strip of the  $i+1$  residue. Unfortunately, the absence of HSQC cross peaks for G69, F70, and G71 prevents assignment of G69 and F70. CA and CB chemical shifts for P62, P66, and P95 were assigned from the CBCA(CO)NH strip of the following residues. However, those of P64 could not be assigned because G65 was missing.

An HNCO experiment was acquired to assign carbonyl carbons on the peptide linkage of the backbone. Figures 3.5 and 3.6 demonstrate how carbonyl carbons were assigned. The magnetization transfer begins from an amide proton to an amide nitrogen and then is transferred to the carbonyl carbon of a preceding residue via selective N-CO J coupling. The magnetization is transferred back to the amide proton to be detected. From the 3D HNCO experiment, 2D projection plots of amide protons and amide nitrogen, and amide protons and carbonyl carbons can be obtained by processing data as shown in Fig 3.5. Carbonyl carbons can be assigned by matching cross peaks in the 2D amide proton – amide nitrogen (HN-N) projection of the HNCO (Fig. 3.5C) to an assigned  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum match. The CO cross peak visible on a column rising from a well resolved HN-N cross peak belongs to the preceding residue.

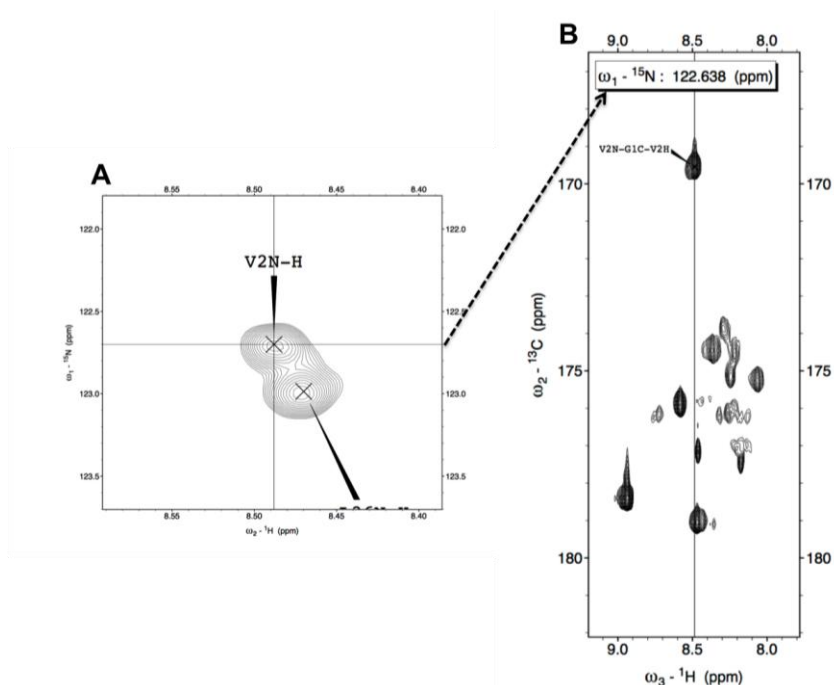
In the case of overlapping HN-N cross peaks resolution can often be achieved in the CO dimension of a 3D HNCO experiment.



**Figure 3.5:** 3D HNCO experiment. (A) illustrates how magnetization is passed from H to  $^{15}\text{N}$  and then selectively to the carbonyl carbon of the preceding residue through N-C J coupling. Blue circles and arrows indicate the detection of nuclei and magnetization flows. (B) and (C) show 2D spectra of H-N and H-CO, respectively. The 2D plot of H-N (B) is identical to the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum. H-CO plot (C) has cross peaks of amide protons correlated with carbonyl carbon via selective N-CO J coupling.

An illustration of CO assignment using the SPARKY software package (21) is shown in Fig. 3.6. The valine of residue 2 (label with V2 in the figure) is selected in the HN-N projection (Fig. 3.6A) and a strip plot of a section of the amide proton and carbonyl carbon plane at a corresponding amide nitrogen chemical shift is chosen (Fig. 3.6B). Although several cross peaks are observed in the strip plot, there is only one matching the CO cross peak of the preceding residue G1. In this manner, the rest of the CO resonances were assigned.

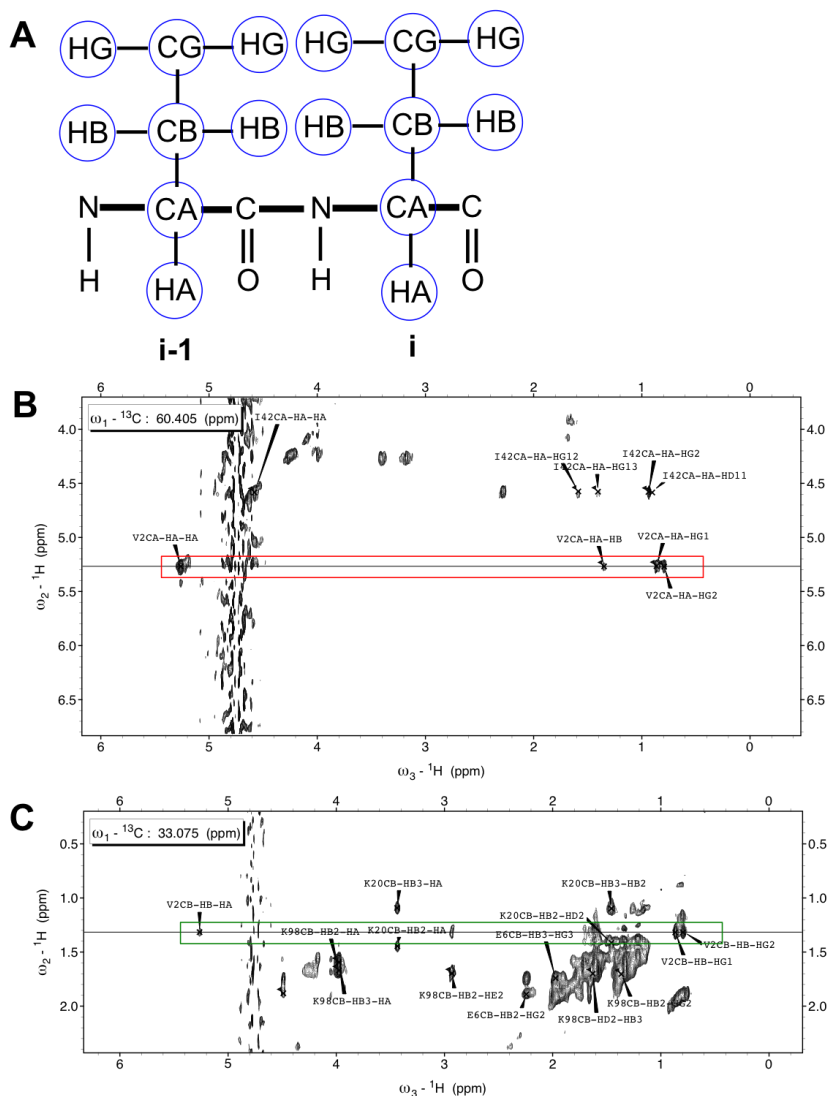




**Figure 3.6:** Strip plot of 3D HNCO experiment. (A) shows the cross peak for V2 in  $^1\text{H}$ - $^{15}\text{N}$  HSQC projection. (B) shows a strip from an  $^1\text{H}$ - $^{13}\text{C}$  plane at the  $^{15}\text{N}$  chemical shift of V2 with a line at the V2  $^1\text{HN}$  chemical shift. It shows one CO cross peak belonging to the preceding residue, G1.

Side chain assignments were carried out using an HCCH-TOCSY experiment. The resulting 3D spectrum shows crosspeaks at the chemical shifts for a particular proton carbon pair (C-H) in one projection and crosspeaks at proton chemical shifts for any sites on the same sidechain fall on vertical lines emanating from a selected proton the selected C-H crosspeak.. Magnetization transfer begins from the side chain hydrogens and passes to their attached carbons for evolution. After mixing among carbons, the magnetization is transferred back to attached hydrogens for detection (Fig. 3.7A). Here V2 is presented as an example (Fig 3.7 B and C). The chemical shifts of CA (60.54 ppm) and CB (33.3 ppm) were assigned using CBCA(CO)NH and HNCACB experiments, above. The strip plot at 60.54 ppm (the shift of CA) of HCCH-TOCSY was selected. Then cross peaks of HB and HG correlating with HA were observed: HB=1.35 ppm; HG1=0.865 ppm; HG2=0.794 ppm (Fig. 3.6B). At 33.3 ppm of CB strip plot, cross peaks of HA, HG1, and

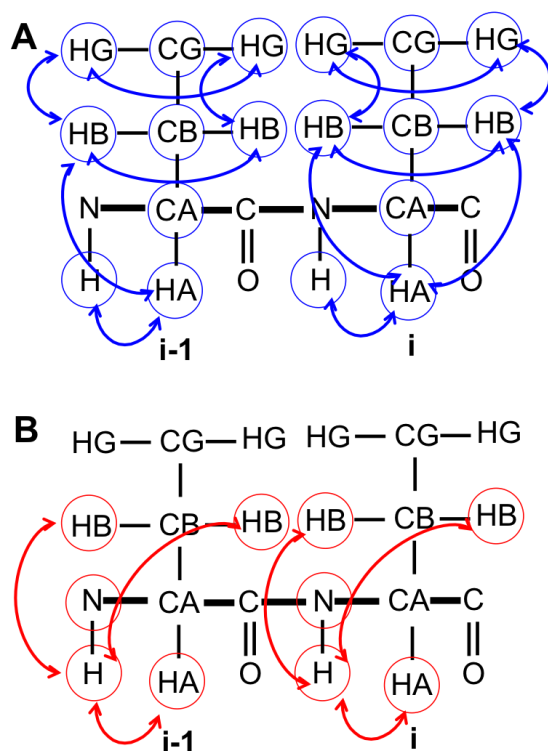
HG2 were observed, correlating with HB (Fig. 3.7C). Rest of the side chain assignments for remaining residues was carried out in the same manner.



**Figure 3.7:** Assignment of side chain using HCCH-TOCSY. (A) shows the magnetization transfer flow between side chains circled in blue for *i* and *i-1* residues. (B) shows the CH strip at the CA chemical shift of V2, cross peaks are highlighted and labeled in the red box. (C) shows a CH strip at the chemical shift of CB of V2, cross peaks correlating with HB are highlighted and labeled in the green box.

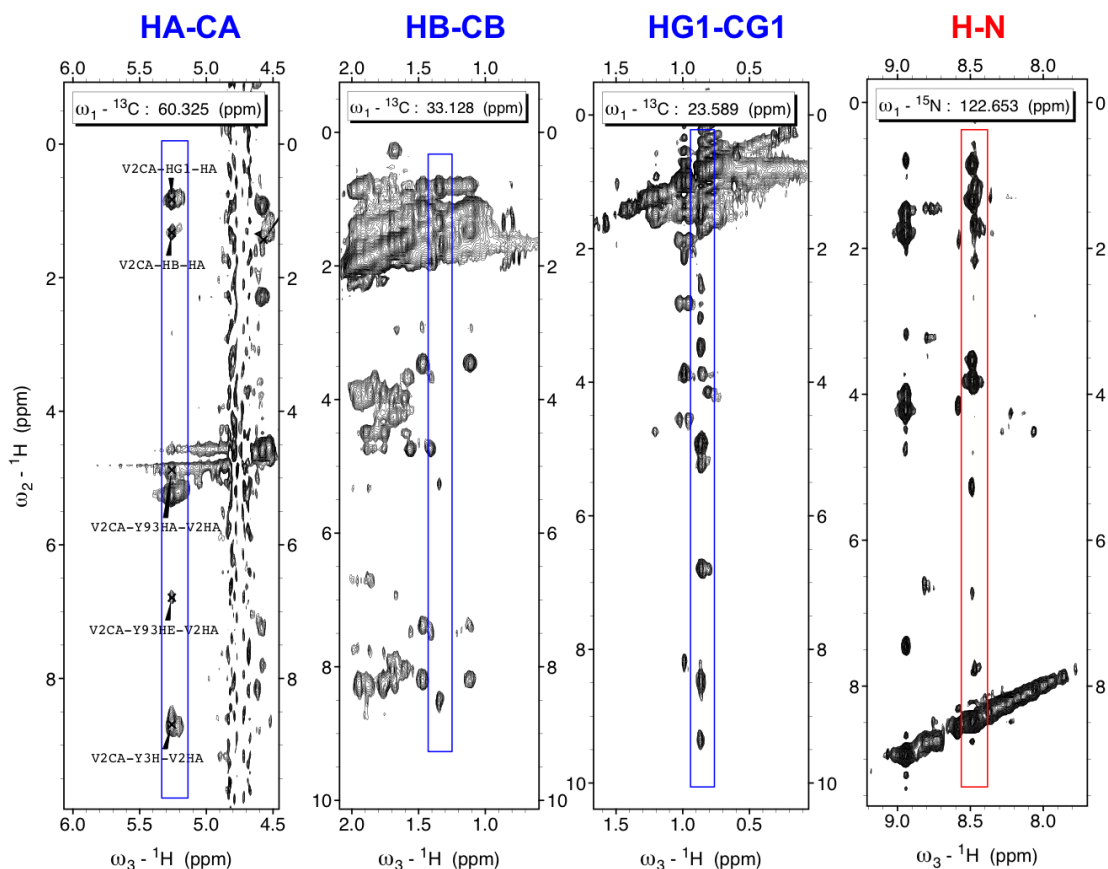
Based on the above stated assignments of backbone and side chains, heteronuclear-edited NMR experiments were exploited such as  ${}^1\text{H}$ - ${}^{15}\text{N}$  HSQC-NOESY and  ${}^1\text{H}$ - ${}^{13}\text{C}$  HSQC-NOESY. The spectra were used to obtain NOE restraints for structure

refinement. As expected from previous experiments, the magnetization during  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY and  $^1\text{H}$ - $^{13}\text{C}$  HSQC-NOESY began with protons, passed to directly bonded  $^{15}\text{N}$  and then  $^{13}\text{C}$  nuclei. This magnetization was exchanged between all protons close together within 6 Å and transferred to neighboring  $^{15}\text{N}$  or  $^{13}\text{C}$  and back to  $^1\text{H}$  attached to  $^{15}\text{N}$  or  $^{13}\text{C}$  for detection (Fig. 3.8). In  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY, any proton nuclei close to the proton being detected within 5-6 Å can be observed and provide useful distance restraint for protein structure calculation.



**Figure 3.8:** General scheme of NOEs observed in  $^1\text{H}$ - $^{13}\text{C}$  HSQC-NOESY (A) and  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY (B) experiments. Blue circles and arrows in (A) indicate possible intra-residue NOEs which can be detected in  $^1\text{H}$ - $^{13}\text{C}$  HSQC-NOESY. Red circles and arrows represent possible intra-residue NOEs which can be detected  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY. Additional inter-residue NOEs will also be observed. 4096 x 144 x 128 data points were collected in each dimension in  $^1\text{H}$ - $^{13}\text{C}$  HSQC-NOESY experiment in which length of acquisition was about 24 hr. 1722 x 144 x 80 data points were collected in  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY experiment for 50 hr. NMR spectroscopy was carried out on 800 MHz spectrometer.

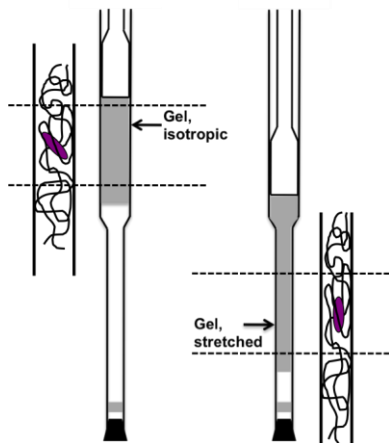
A sample set of spectra of V2 is provided in Fig. 3.9. The first H-H projection at the  $^{13}\text{C}$  chemical shift of 60.33 shows NOE cross peaks of V2HA correlating with Y3HN, Y93HE, Y93HA, V2HB, and V2HG1. Even though Y93HE and Y93HA are not close to the V2HA in the sequence, they are close enough to correlate through dipolar coupling in space. NOEs for other projections in Fig. 3.9 were assigned in the same manner.



**Figure 3.9:** Strips from  $^1\text{H}$ - $^{13}\text{C}$  HSQC-NOESY (first three strip plots from the left) and  $^1\text{H}$ - $^{15}\text{N}$  HSQC-NOESY (strip plot, far right) for V2 of the Link-TSG-6: $\Delta\text{C444S}$  complex. The cross peaks in  $^1\text{H}$ - $^1\text{H}$  projections at CA, CB, and CG1 chemical shifts are presented in blue boxes. The red box indicates cross peaks at the NH chemical shift of V2. The labels are on the first projection of CA.

In addition to NOEs, residual dipolar coupling restraints were obtained from IPAP-HSQC experiments for the oriented and isotropic phase. A neutral polyacrylamide

gel was used to align the molecule for RDC measurement. Because the protein tends to have a positive charge while the sugar has a negative charge at pH 6, the only possibility was to place the protein-ligand complex in a neutral gel to avoid unwanted charge-charge interactions between the molecules and alignment media. Two-stage NMR tubes were used to acquire spectra for the RDCs (19). A neutral gel containing the protein-ligand complex sample was placed into the upper part of the NMR tube for acquiring isotropic data as shown in Fig 3.10 (left). The sample was then drawn into the lower narrower part in order to stretch the gel and align the molecule in the gel (Fig. 3.10, right tube).



**Figure 3.10:** Two-stage NMR tube for RDC experiments.

The experiment was performed by acquiring two sets of data in an interleaved manner. One set of data gives a doublet with the same sign and is referred to as in-phase (IP). Another set of data has a doublet with the opposite sign that is referred to as anti-phase (AP). The addition and subtraction of the IP and AP spectra provides two spectra with the appearance of normal  $^1\text{H}$ - $^{15}\text{N}$  HSQC but with the  $^{15}\text{N}$  chemical shifts displaced by either  $+J_{\text{NH}}/2$  or  $-J_{\text{NH}}/2$  as shown in Fig. 3.11.



oriented samples are measured in the same way as the J coupling. D values were assigned by subtracting J from J+D. Resonances that either overlapped with other resonances or experienced peak broadening in at least one spectrum were excluded.

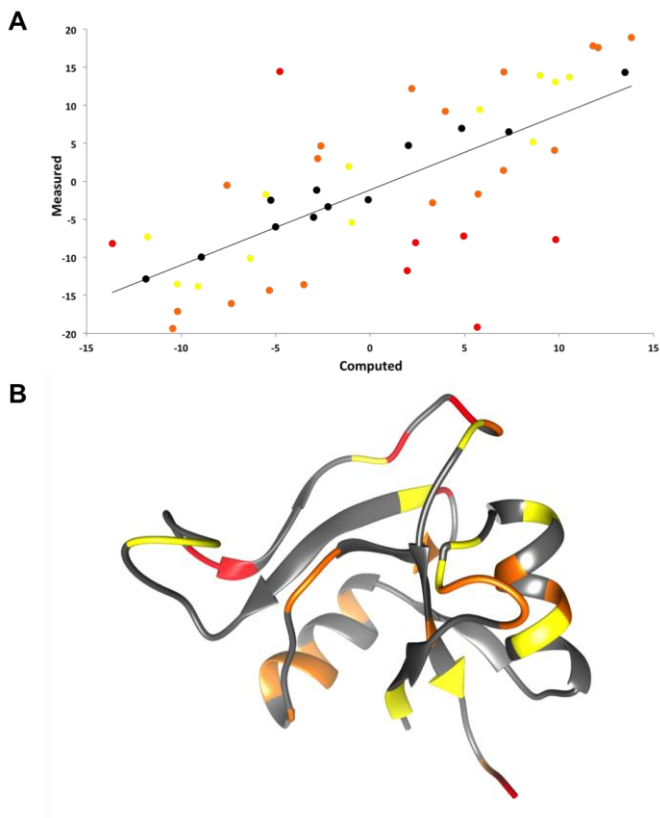
In the end, 46 RDCs were obtained with an experimental error of +/- 1.6 Hz as shown in Table 3.2. These were fitted onto existing structure in its HA bound state (first model of the ensemble was selected, 1O7C.pdb) in order to validate RDCs.

**Table 3.2:** RDCs of the Link-TSG-6: $\Delta$ C444S complex aligned in neutral gel measured with IPAP-HSQC.

Residue	RDC (Hz)	Residue	RDC (Hz)
V2	-5.2	F44	-4.7
Y3	-13.1	A48	-3.0
R8	-17.8	M52	16.1
S9	-12.2	A53	1.2
G10	-1.4	K54	3.4
K11	-13.7	G55	19.2
K13	10.7	R56	1.7
L14	-6.5	N67	7.4
T15	6.0	C68	-9.4
A17	13.5	T73	13.6
E18	12.9	G74	7.2
A19	19.4	I75	4.7
K20	10.0	G79	5.4
A21	8.2	I80	11.7
V22	13.8	N83	-14.4
L30	-14.3	S85	17.1
Y33	-7.0	E86	-2.0
K34	-14.4	R87	2.5
A38	-17.6	Y91	0.5
R40	-9.2	Y93	-13.9
K41	-18.9	H96	1.7
I42	-14.1	A97	8.1
G43	14.3	K98	7.6

As shown in Fig 3.12A, RDCs back calculated using the existing NMR structure, 1O7C.pdb, do not fit well to experimental RDCs. (Quality factor is 0.7 and RMSD

between measured and calculated RDCs is 7.7 Hz), indicating that measured RDCs in the presence of CS do not match the existing structure in its HA bound state. Especially, the residues that have RDC difference over 10 Hz (red in Fig.3.12) are mostly in areas lacking well-formed secondary structure. Several deviations in the 5-10 Hz range are also in areas with recognizable secondary structure.



**Figure 3.12:** RDC fitting to the Link module in the HA bound state (1O7C.pdb). A shows back calculated RDC fitting to the existing structure. The averaged experimental error is  $\pm 1.6$  Hz. Residues that show RDC differences more than 10 Hz, 10-5, and 5-3 Hz between measured and calculated are colored in red, orange, and yellow in the plot. B shows these residues depicted onto the ribbon structure of Link module with same colors.

This suggests that CS binding to the Link module may cause structural changes that are different from those caused by HA binding. This suggests the need for a structure determination in the presence of CS. Total 46 RDCs were used in structures calculations.



Dihedral angles were predicted by TALOS as shown in Table 3.3. These were used in structure calculation programs, CYANA and XPLOR-NIH. The structure calculation using CYANA version 3.0 was carried out first. In this program, NOE cross peaks were calibrated according to the intensity or volume of the peaks whose distance is already known. NOE peaks in  $\alpha$ -helical secondary structure regions were used as standard peaks for the fixed distance. Statistically, the HN-HN distance between  $i$  and  $i+1$  residues in the  $\alpha$ -helix is 2.8 Å and the distance between the HA-HN of  $i$  and  $i+1$  is 3.5 Å. Structure calculations were performed with calibrated NOEs and dihedral restraints. 20 structures were produced using a simulated annealing protocol. The structures were further refined in XPLOR-NIH with additional restraints of RDCs.

**Table 3.3:** Dihedral angles of the Link-TSG-6:  $\Delta$ C444S complex.

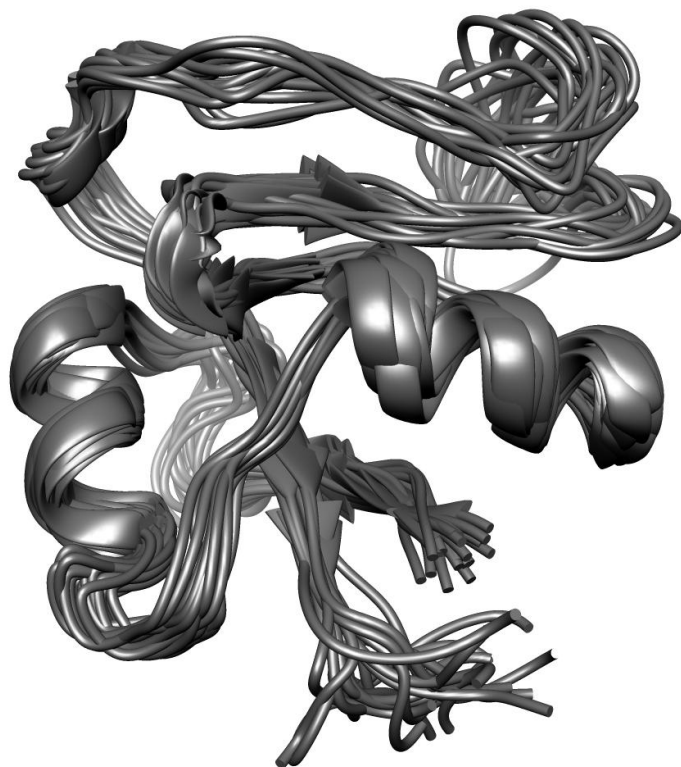
Residue	Phi		Psi	
	Angle_low	Angle_high	Angle_low	Angle_high
V2	-162.35	-84.26	123.1	161.58
Y3	-168.51	-102.23	124.21	175.01
H4	-148.37	-75.22	98.61	148.48
R5	-145.88	-102.88	116.37	168.47
E6	-167.86	-94.55	112.36	186.34
A7	-89.24	-54.49	107.62	181.49
R8	-80.75	-40.13	-50.08	-10.86
S9	-111.79	-68.45	-40.19	36.91
K13	-174.09	-69.13		
L14	-144.2	-66.53	110.23	164.84
T15	-99.69	-72.8	160.96	174.72
Y16	-70.05	-46.19	-59.25	-26.25
A17	-74.02	-45.78	-56.9	-29.06
E18	-74.49	-57.41	-47.52	-31.35
A19	-77.99	-51.55	-60.16	-31.1
K20	-67.97	-52.6	-60.8	-31.95
A21	-74.84	-52.15	-52.62	-22.71
V22	-92.22	-60.93	-59.76	-16.15
C23	-167.55	-106.55	140.4	174.14
F25	-84.1	-25.29	-60.26	-18.85
E26	-106	-68.06	-18.73	27.75
G27	42.64	136.76	-24.56	49.72
G28	-152.8	-48.37	87.49	213.75
H29	-173.48	-102.81	120.62	178.73
L30	-100.77	-43.94	116.72	157.06

A31	-143.89	-43.68	73.52	176.16
T32	-121	-60	149.5	178
Y33	-66.77	-54.67	-51.62	-23.81
K34	-74.59	-47.54	-57.6	-21.38
Q35	-70.44	-62.14	-49.91	-27.63
L36	-72.68	-55.02	-51.51	-29.32
E37	-75.06	-51.5	-49.12	-32.13
A38	-77.42	-47.86	-57.01	-27.56
A39	-80.83	-53.49	-52.15	-23.81
R40	-72.49	-57.56	-57.78	-22.12
K41	-90.06	-52.28	-45.16	-4.02
I42	-114.53	-71.18	-16.3	27.34
G43	59.55	114.75	-28.93	38.84
F44	-131.97	-56.5	105.39	159.46
H45	-154.97	-79.59	106.89	165.78
V46	-138.36	-98.36	107.85	156.89
C47	-155.49	-94.52		
A48	-179.82	-69.1	94.37	149.36
A49	-132.63	-45.3	96.8	163.49
W51	-171.8	-36.44	108.34	168.23
M52	-182.58	-97.66	137.77	188.5
K54	48.63	64.86	16.3	49.41
G55	58.13	104.94	-19.2	16.15
R56	-138.32	-70.14		
V57			110.3	175.76
I61	-166.59	-60.27	93.14	191.94
N67	-138.37	-68.19	-44.14	43.7
C68	-177.78	-92.12	112.89	202.47
T73	-179.79	-96.61	125.19	171.88
D77	-135.49	-61.76	101.16	161.02
Y78	-142.65	-68.69		
R81	-110.7	-63.17	40.65	180.12
L82	-83.74	-52.22	-40.54	3.19
N83	-112.8	-75.55	-19.42	15.88
R84	-116.25	-49.76	-80.86	47.95
S85	-100.96	-48.15	-77.78	21.77
E86	-111.85	-60.4		
R87	-171.47	-79.85	110.04	176.45
W88	-153.47	-109.67	136.34	173.08
A90	-189.61	-106.64	122.09	169.38
Y91	-148.63	-76.17	104.92	160.18
C92	-162.14	-116.84	145.41	188.13
Y93	-157.05	-109.88	113.13	150.82
N94	-152.33	-67.11	80.52	168.99

20 final structures were determined for  $\Delta$ C444S bound Link-TSG-6 using 1329

NOE distance restraints, 46 RDC restraints and 135 dihedral restraints. This provided

15.3 restraints per residue. All 20 models are superimposed as shown in Fig. 3.13. The average backbone root mean square deviation of ordered residues in the final ensemble is 0.8 Å. Backbone dihedral angle analysis indicated that 88 % of the residues fall into the most favored regions of the Ramachandran plot. Structural statistics are summarized in Table 3.4. Two double-stranded anti-parallel  $\beta$ -sheets, two  $\alpha$ -helices and two disulfide bonds (C23-C92 and C47-C68) were observed in (Several models are missing a  $\beta$ -sheet).



**Figure 3.13:** 20 superimposed models for the NMR structures of the Link-TSG-6 in its  $\Delta$ C444S bound state.

**Table 3.4:** Structural statistics for the human Link module of TSG-6. Statistics were computed for the ensemble of 20 NMR structures deposited in the Protein Data Bank. r.m.s., root mean square; r.m.s.d., root mean square deviation.

Conformationally restricting constraints <sup>a</sup>	
Distance constraints	
Total	1329
Intraresidue ( $i = j$ )	260
Sequential ( $ i - j  = 1$ )	394
Medium range ( $1 < i - j \leq 5$ )	204
Long range ( $ i - j  > 5$ )	471
Distance constraints per residue	13.6
Dihedral angle constraints	135
No. of constraints per residue	15.0
No. of long-range constraints per residue	4.8
Residual constraint violations <sup>a</sup>	
Average no. of distance violations per structure	
0.1–0.2 Å	11.45
0.2–0.5 Å	7.55
>0.5 Å	0.35
Average r.m.s. distance violations/constraint (Å)	0.03
Maximum distance violation (Å)	0.88
Average no. of dihedral angle violations per structure	
1–10°	18.9
>10°	0.3
Average r.m.s. dihedral angle violation/constraints (°)	1.49
Maximum dihedral angle violation (°)	16.80
r.m.s.d. from average coordinates (all/ordered) <sup>a,b</sup>	
Backbone atoms (Å)	1.2/0.8
Heavy atoms (Å)	1.8/1.4
Ramachandran statistics for ordered residues (Richardson MolProbity) <sup>a,b</sup>	
Most favored regions (%)	87.9
Additionally allowed regions (%)	11.0
Disallowed regions (%)	1.1
Global quality scores (raw/Z-score) <sup>a</sup>	
Verify3D	0.37
ProsaII	0.54
PROCHECK ( $\phi$ – $\psi$ )	-0.60
PROCHECK (all)	-0.42
MolProbity Clash	21.93
RDC statistics <sup>c</sup>	
No. of D <sub>NH</sub> constraints	46
R	0.98
Q <sub>r.m.s.</sub>	0.013

<sup>a</sup> Values were calculated using PSVS Version 1.5. Average distance violations were calculated using the sum over  $r^{-6}$

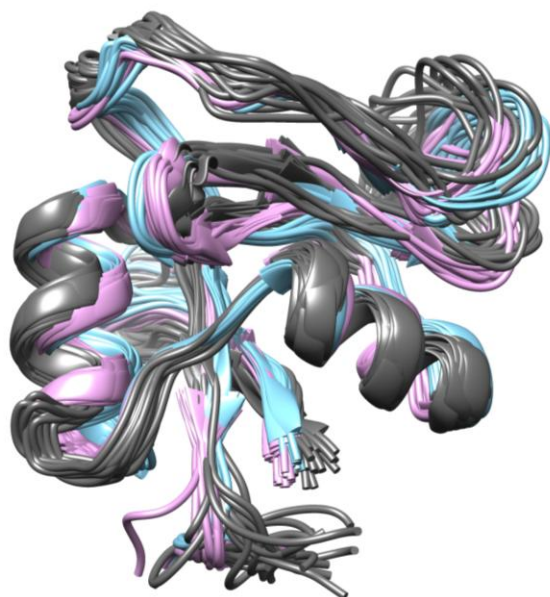
<sup>b</sup> *Ordered residue ranges: 2-9,13-46,48-57,60-62,81-93*

Residues selected based on: Dihedral angle order parameter, with  $S(\phi)+S(\psi)\geq 1.8$

<sup>c</sup> RDC statistics were computed by PALES

### **3.3.2 Comparison of the Link module of TSG-6 in its free, HA octasaccharide-bound and $\Delta$ C444S-bound states**

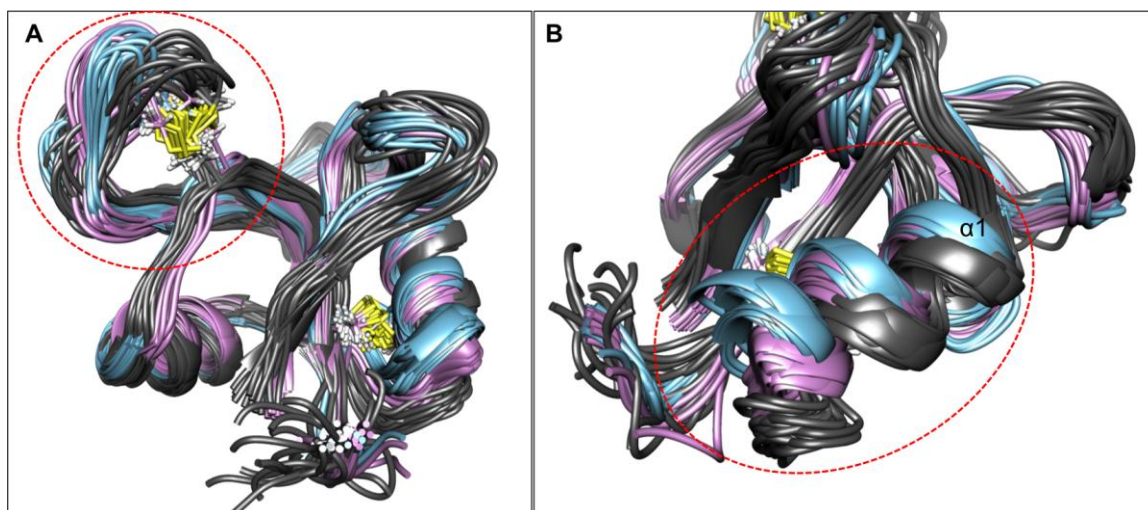
As mentioned in previous chapters, the structure of Link-TSG-6 has been studied previously. Comparison of our structure to these structures is appropriate (1O7B, 1O7C). The average backbone r.m.s.d was about 1.2 Å in each case. As shown in Fig. 3.14, the overall structure is well overlapped with few exceptions.



**Figure 3.14:** Superimposition of NMR structures in free (1O7B.pdb, blue), HA bound (1O7c.pdb, pink), and  $\Delta$ C444S bound (gray) states.

However, one difference in structures between HA and bound CS is the large loop shown in Fig. 3.15A. This loop in the  $\Delta$ C444S bound form is located nearer to the protein core than in the other structures. It is involved in HA binding (structure 1O7C) where it has moved slightly toward the core in comparison to the free models. Its more extreme

movement suggests that the same loop is involved in  $\Delta$ C444S binding but that sulfated CS binds to Link module somewhat differently. A detailed discussion and docking analysis are provided in Chapter 4. A gradual movement of helix  $\alpha 1$  was also observed from the free Link models (Blue ribbons) to HA-bound models (Pink ribbons) and finally the  $\Delta$ C444S bound models (Gray ribbons) (Fig. 3.15B). This movement could be associated with chondroitin sulfate binding, even though this helix is not at the binding site.



**Figure 3.15:** Structure comparisons of NMR structures in free (1O7B.pdb, blue), HA bound (1O7C.pdb, pink), and  $\Delta$ C444S bound (gray) states. Cysteines forming disulfide bonds are present in all conformations. A difference between loops among conformations is shown in the red dotted circle in A. A difference in the  $\alpha 1$  helix secondary structure is highlighted in the red dotted circle in B.

### 3.4 Discussion

An ensemble of structures for Link-TSG-6 in the presence of  $\Delta$ C444S in solution has been successfully produced by the methods described above. Due to the binding of  $\Delta$ C444S to the Link module, signal broadening was observed, resulting in the disappearance of some resonances in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum. Even though backbone nuclei chemical shifts disappeared, some side chain nuclei could be assigned

with the aid of other 3D experiments. As shown above, structural comparisons were made between the Link modules in its free form (1O7B.pdb), HA bound state (1O7C.pdb) and  $\Delta$ C444S bound showing an overall a RMSD of 1.2 Å. A closer look at the binding groove for HA and  $\Delta$ C444S suggests that the site is shared, but binding is somewhat different. Among the indicators are a more exaggerated move of a binding loop and the more remote the  $\alpha$ 1 helix. These structural changes were studied more extensively by NMR spectroscopy and a molecular docking study. Chapter 4 provides the details of this further investigation.

### **3.5 References**

1. Chothia, C., and Lesk, A. M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J* **5**, 823-826
2. Higman, V. A., Blundell, C. D., Mahoney, D. J., Redfield, C., Noble, M. E., and Day, A. J. (2007) Plasticity of the TSG-6 HA-binding loop and mobility in the TSG-6-HA complex revealed by NMR and X-ray crystallography. *J Mol Biol* **371**, 669-684
3. Blundell, C. D., Mahoney, D. J., Almond, A., DeAngelis, P. L., Kahmann, J. D., Teriete, P., Pickford, A. R., Campbell, I. D., and Day, A. J. (2003) The link module from ovulation- and inflammation-associated protein TSG-6 changes conformation on hyaluronan binding. *J Biol Chem* **278**, 49261-49270
4. Mahoney, D. J., Mulloy, B., Forster, M. J., Blundell, C. D., Fries, E., Milner, C. M., and Day, A. J. (2005) Characterization of the interaction between tumor necrosis factor-stimulated gene-6 and heparin: implications for the inhibition of plasmin in extracellular matrix microenvironments. *J Biol Chem* **280**, 27044-27055
5. Parkar, A. A., and Day, A. J. (1997) Overlapping sites on the Link module of human TSG-6 mediate binding to hyaluronan and chondroitin-4-sulphate. *FEBS Lett* **410**, 413-417
6. Parkar, A. A., Kahmann, J. D., Howat, S. L., Bayliss, M. T., and Day, A. J. (1998) TSG-6 interacts with hyaluronan and aggrecan in a pH-dependent manner via a common functional element: implications for its regulation in inflamed cartilage. *FEBS Lett* **428**, 171-176

7. Wisniewski, H. G., Snitkin, E. S., Mindrescu, C., Sweet, M. H., and Vilcek, J. (2005) TSG-6 protein binding to glycosaminoglycans: formation of stable complexes with hyaluronan and binding to chondroitin sulfates. *J Biol Chem* **280**, 14476-14484
8. Milner, C. M., and Day, A. J. (2003) TSG-6: a multifunctional protein associated with inflammation. *J Cell Sci* **116**, 1863-1873
9. Higman, V. A., Briggs, D. C., Mahoney, D. J., Blundell, C. D., Sattelle, B. M., Dyer, D. P., Green, D. E., DeAngelis, P. L., Almond, A., Milner, C. M., and Day, A. J. (2014) A refined model for the TSG-6 link module in complex with hyaluronan: use of defined oligosaccharides to probe structure and function. *J Biol Chem* **289**, 5619-5634
10. Jain, N. U., Noble, S., and Prestegard, J. H. (2003) Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings. *J Mol Biol* **328**, 451-462
11. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-223
12. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**, 283-298
13. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**, 209-227
14. Guntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-378
15. Guntert, P. (2009) Automated structure determination from NMR spectra. *European biophysics journal : EBJ* **38**, 129-143
16. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**, 65-73
17. Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778-795



18. Day, A. J., Aplin, R. T., and Willis, A. C. (1996) Overexpression, purification, and refolding of link module from human TSG-6 in *Escherichia coli*: effect of temperature, media, and mutagenesis on lysine misincorporation at arginine AGA codons. *Protein expression and purification* **8**, 1-16
19. Liu, Y., and Prestegard, J. H. (2010) A device for the measurement of residual chemical shift anisotropy and residual dipolar coupling in soluble and membrane-associated proteins. *J Biomol NMR* **47**, 249-258
20. Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* **131**, 373-378
21. Kneller, T. D. G. a. D. G. SPARKY 3
22. [http://www.bmrb.wisc.edu/published/Ikura\\_cs\\_study/part2.html](http://www.bmrb.wisc.edu/published/Ikura_cs_study/part2.html)

## **CHAPTER 4**

### **NMR INSIGHT INTO THE MULTIPLE GLYCOSAMINOGLYCAN BINDING MODES OF THE LINK MODULE FROM HUMAN TSG-6<sup>1</sup>**

---

<sup>1</sup>Younghee Park, Anthony J. Day and James H. Prestegard  
To be submitted to *Biochemistry*

## 4.1 **Abstract**

Tumor necrosis factor-stimulated gene-6 (TSG-6) is a hyaluronan (HA) binding protein that is essential for stabilizing and remodelling the extracellular matrix (ECM) during inflammation. TSG-6 plays a role in numerous inflammatory disease processes and physiological events such as arthritis and ovulation. The Link module, one of the domains of TSG-6, is responsible for binding hyaluronan and other glycosaminoglycans (GAGs) found in the ECM. In this study, we used well-defined chondroitin sulfate (CS) hexasaccharides ( $\Delta$ C444S) to determine the structure of the Link module in solution in its chondroitin sulfate bound state. A variety of NMR techniques were employed, including chemical shift perturbation, residual dipolar couplings (RDCs), NOEs, spin relaxation measurements, and paramagnetic relaxation enhancements (PREs) from a spin-labeled analog of  $\Delta$ C444S. As expected, the binding site for  $\Delta$ C444S on the Link module overlapped with that of HA. Surprisingly, this  $\Delta$ C444S binding induced dimerization of the Link module, and a second binding site that partially overlapped with a previously postulated heparan sulfate site was occupied. A dimer model was generated using chemical shift perturbations and RDCs as restraints in the docking program HADDOCK. We postulate that the molecular cross-linking enhanced by the multiple binding modes of the Link module may be critical for stabilizing the ECM during inflammation and under other conditions.

*Keywords:* Link module, Dimerization, TSG-6, Chondroitin sulfate, NMR spectroscopy, Chemical shift perturbation, Glycosaminoglycans, HADDOCK, RDC, PRE

## **4.2 Introduction**

Link modules are domains found in a number of proteins (hyaladherins) that interact with glycosaminoglycans (GAGs) in the extracellular matrix (ECM) of mammalian cells. These interactions are suggested to stabilize the matrix while still allowing remodeling in response to physiological changes. These properties can be beneficial, as in the repair of blood vessels after injury, or detrimental when they contribute to the progression of debilitating diseases (1-7). Characterizing the structural nature of the interactions between Link modules and GAGs can provide insights into the development of GAG mimics that can be used to alter these interactions in the treatment of injury and disease (2,8-12). Here, we present a structural investigation of the interaction between the Link module of TSG-6 and well-defined oligomers of chondroitin sulfate using nuclear magnetic resonance (NMR) methodology. These studies confirm that this interaction occurs in two distinct sites and induces protein dimerization in the process. The data build on previous studies using other GAGs, and contribute to a complex cross-linking model that explains how Link modules stabilize and remodel the extracellular matrix.

The Link module from TSG-6 is well studied and has become a model for the large family of modules from other hyaladherins. Link module is a small domain comprised of approximately 100 amino acids and X-ray crystallography and solution NMR structures exist of this domain (13-18). Its interaction with hyaluronan (HA) is best studied; the family name, hyaladherin stems from HA's well-known interaction with the Link module (13,16-20). HA is an excreted copolymer of glucuronic acid (GlcA) and N-acetyl glucosamine (GlcNAc) and is found in abundance in the ECM. Its binding site has

been characterized by site-directed mutagenesis and NMR and likely involves residues K11, Y12, Y59, F70, and Y78 (19,21). Other types of GAGs that are found in the ECM, such as heparan sulfate (HS), a copolymer of repeating disaccharide units of glucuronic acid (GlcA) or iduronic acid (IdoA) and N-acetyl glucosamine (GlcNAc), and chondroitin sulfate (CS), a copolymer of glucuronic acid (GlcA) and N-acetyl galactosamine (GalNAc) are known to also bind to Link-TSG-6 (20,22,23).

Heparin/HS binds to a site distinct from that of HA, namely one involving residues K20, K34, R40, K41, K54, R56 and R84 (15,22). While the identification of this site is primarily based on computational modeling, its existence is supported by mutagenesis studies (22). Less information is known about CS binding sites. However, competition assays between CS and HA and binding assays with chimeric CS and HA monitored by NMR suggest that CS prefers to attach to the HA binding surface, despite the fact that CS and HS are both characterized by sulfation and an associated increase in negative charge (1,14,20,23,24). GAGs are extremely heterogeneous materials; for CS and HS this is primarily due to the possibility of sulfation at a number of distinct positions on the constituent sugar rings. HS can be N-sulfated, with the sulfate replacing the acetyl groups on the initial GlcNAc residues, as well as O-sulfated on the 3 and 6 positions of GlcNAc and the 2 position of iduroic acid residues. CS is less heterogeneous than HS with possibilities for sulfation at the 4 and 6 positions of the GalNAc residue.

A hexasaccharide of CS is an ideal way to study the interaction with a Link module since it is large enough to fill known binding sites and small enough to allow isolation or synthesis of a single well-defined hexasaccharide. However, even a hexasaccharide of CS has 64 possible sulfation patterns. Isolating a specific oligomer

was overcome here by choosing a polymer source that is rich in 4 sulfation (CS-A) for digestion with lyase and hydrolase enzymes, and treating digests using size separation and ion exchange chromatography. This method allowed for the isolation of four CS candidates:  $\Delta$ C664S,  $\Delta$ C444S, C664S, and C444S (in which the  $\Delta$  indicates the lyase product and the numbers indicate the sites of sulfation on the three GalNAc residues). The most strongly binding oligomer,  $\Delta$ C444S, was selected for structural studies of the Link module because it binds with a similar affinity as HA (20).

In addition to GAGs, TSG-6 has various binding partners such as the G1 domains of the proteoglycans aggrecan and versican, and a number of other extracellular membrane proteins, including inter-alpha-inhibitor (I $\alpha$ I), pantraxin 3 (PTX3), and thrombospondin 1 (6,20-22,25-39). Using solution NMR methods to obtain structural information about how the Link module of TSG-6 binds to CS would offer a three-dimensional model of the interactive surfaces of the Link module and reveal possible interactions between GAGs located in the ECM, as well as these other binding partners.

Measuring chemical shift perturbations in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -labeled proteins provides one means of identifying amino acids that are potentially involved in binding. This information is very qualitative, resulting from a mix of ligand-induced changes in backbone conformation and through-space interactions. Cautious interpretations of these data are prudent since induced conformation changes can be remote or originate from other phenomena such as ligand-induced dimer formation (40). However, once resonance assignments are made, data are easily collected on relatively small amounts of sample, and analysis of shifts as a function of ligand concentration can provide binding constants. After collecting ligand induced chemical shift changes, a

paramagnetically tagged version of a  $\Delta$ C444S oligomer that carries a TEMPO group at the reducing end was used to add more quantitative data to a structural description of ligand binding. Spin relaxation of amide protons at  $^{15}\text{N}$  labeled sites in the protein occurs with approximately an inverse sixth power dependence on the distance between the TEMPO group and the labeled sites and provides a definitive location for the reducing terminus of the hexasaccharide.

Residual dipolar couplings (RDCs) provide detailed quantitative data about changes in protein conformation through their dependence on the average of functions involving the angle various  $^1\text{H}$ - $^{15}\text{N}$  vectors make with the NMR magnetic field in partially ordered media (41,42). RDCs can be collected with a simple extension of the basic HSQC sequence, the IPAP experiment, which allows measurement of coupling between  $^1\text{H}$ - $^{15}\text{N}$  pairs (43). These data help define the conformational changes and subsequent dimer formation that occur during ligand binding.

Computational programs can produce reliable structures for ligand-protein complexes based on NMR data. Our studies employed the docking program, HADDOCK, to generate models for the dimer and protein-sugar complex. HADDOCK incorporates a variety of quantitative and qualitative data in a simulated annealing search for a best structure along with options to refine structures with force fields representing the energetics of molecular interactions (44,45). We found that human Link-TSG-6 has a primary CS binding surface that partially overlaps with HA. We also discovered that ligand binding induces the formation of a dimer that has a second binding site that partially overlaps the suggested HS site. A detailed presentation of data leading to a molecular model for this complex follows.

### **4.3 Results**

#### **4.3.1 NMR-monitored titration of Link-TSG-6 with CS**

The isomers of CS hexasaccharide (GlcA-GalNAc-GlcA-GalNAc-GlcA-GalNAc) selected for our studies included  $\Delta$ C664S,  $\Delta$ C444S, C664S, and C444S. CS binding sites on Link-TSG-6 were mapped using chemical shift perturbations in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra as a function of ligand concentration up to a twofold molar excess over protein concentration. Figure 4.1 shows the superimposition of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of 240  $\mu\text{M}$  Link-TSG-6 in the absence and presence of 0.096, 0.192, 0.29, 0.38, and 0.48 mM of the  $\Delta$ C444S hexasaccharide. Similar behavior was observed for other hexasaccharides (data provided in Supplement). Assignments of crosspeaks in the HSQC spectra were made using triple resonance experiments on a  $^{15}\text{N}$ ,  $^{13}\text{C}$  labeled sample as described in the Methods section. Assignments for the most significantly shifted crosspeaks are indicated on Figure 4.1. Three distinct types of behavior were observed in terms of a chemical shift exchange rate between complexed and uncomplexed states. When ligand was added, some crosspeaks progressively decreased in intensity as other peaks at nearby positions progressively increased in intensity. This is characteristic of exchange between complexed and uncomplexed states that are on a slow timescale compared to the inverse of line widths of the resonance. Crosspeaks belonging to residues H4, L14, H45, C47, A48, A49, G58, K63, C68, K72, G74, and I76 fall into this category. A few crosspeaks belonging to G65, G59, F70, and G71 also exhibited similar behavior (data not shown). In this case, the crosspeaks disappeared but could not be correlated with newly appearing peaks due to overlap or additional broadening in the complexed state. Because intensity changes in the limit of slow exchange reflect the population of complexed and



uncomplexed states, they can be used to determine a dissociation constant for this process. Figure 4.2A shows a fit to the normalized average intensity changes for residues A48, G58, and G74. The equations 1-3 below were derived based on the assumption that the exchange broadening exists for slow exchange residues. The system was settled in and parameters iterated to achieve the fit displayed in Fig. 2A.

$$[C] = K_1 \times [L] \times [P]$$

Equation 1

$$[PT] = [P] + [C] \quad \text{Equation 2}$$

$$[LT] = [L] + [C] \quad \text{Equation 3}$$

where [C] is concentration of the complex.  $K_1$  indicates a binding constant for ligand binding. [P], [PT], [L], and [LT] represent concentrations of protein, total protein, ligand, and total ligand, respectively. Although intensity changes depart somewhat from ideal behavior due to intensity losses during HSQC transfers from peaks with residual lifetime broadening, it was possible to determine a dissociation constant of approximately  $K_d = 10 \mu\text{M}$ . Slow exchanging residues are a part of a shallow binding groove to which HA binds. The following residues were previously identified as being involved in HA binding: K11, Y12, Ty59, F70, and Y78. CH- $\pi$  stacking interactions of saccharide rings against aromatic planes are suggested to play an important role (18). These data suggest that CS and HA interact similarly with the Link module. However, even though  $\Delta\text{C444S}$  and HA share a binding groove, the participating residues are different. Positively charged residues, such as H4, H45, K63, and K72 are perturbed substantially indicating that they may be associating with sulfation groups on the GalNAc of  $\Delta\text{C444S}$ . Another piece of

evidence to show  $\Delta C444S$  shares the HA binding site is the perturbation of C47 and C68 resonances. These are important residues that form a disulfide bridge close to the HA binding surface. A change in geometry resulting in a rearrangement of the  $\beta 4$ - $\beta 5$  loop (residue 61-74) may cause this perturbation (16). As shown in Figure 4.3, the chemical shifts of  $\alpha$  and  $\beta$  carbons in residues 60-78 are perturbed to a significant degree as would be expected with a perturbation of the protein backbone.

For a second class of perturbed peaks crosspeaks displayed some chemical shift change and signal broadening as the ligand concentration increased. In most cases, this resulted in the disappearance of crosspeaks when the total ligand concentration was approximately one-half of the total protein concentration. These crosspeaks then reappeared at much higher ligand concentration. This is characteristic of an exchange process occurring on a somewhat faster intermediate timescale. Residues E18, G50, N67, I75, and W88 exhibited this behavior. The changes in chemical shifts were, in general, smaller than those for residues showing slow exchange. They may well be involved in the same ligand binding process but be near parts of the bound ligand that still have significant internal mobility.

A third type of behavior occurs when peaks exhibit small changes in chemical shift and line width at lower ligand concentrations, then show larger chemical shift changes at higher ligand concentrations. Crosspeaks for K11, E24, H29, L30, Y33, W51, A53, Y91, and Y93 belong in this category. Their behavior is typical of weaker binding with the onset of a second binding event occurring after binding at the tight site is completed. Under rapid exchange, chemical shifts are population-weighted averages of complexed and uncomplexed states. Chemical shift changes for various resonances can

be represented by the equation  $\Delta\delta = [(\Delta\delta_{\text{HN}})^2 + (\Delta\delta_{\text{N}} \times 0.17)^2]^{1/2}$  which combines shifts for amide protons and nitrogen (Figure 4). Here, the scale factor 0.17 was taken from the chemical shift range for  $^1\text{H}$  and  $^{15}\text{N}$  (46). The set of equations 4-6 below was derived based on the assumption that only the products of slow exchange are capable of participating in fast exchange. The system was settled in and parameters iterated to achieve the fits displayed in Fig. 4.2B. This is built-in equation in Maple program, providing the calculation of shift perturbation for two sites with dimerization required for binding to a second site.

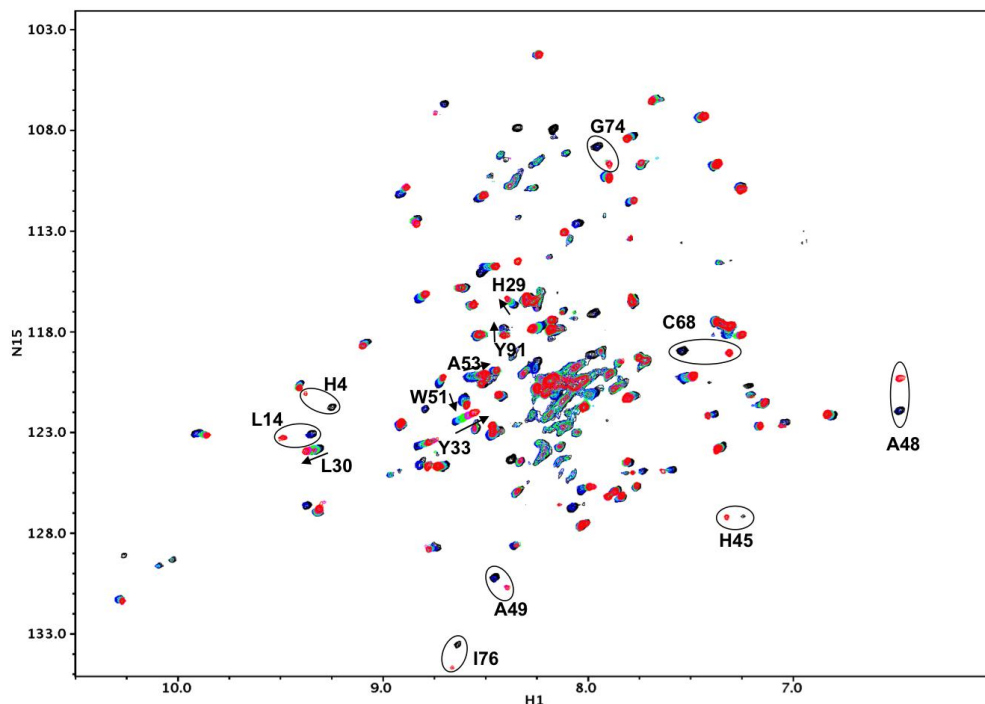
$$[C2] = K_2 \times [L] \times K_D \times (K_1 \times [L] \times [P])^2$$

[Equation 4]

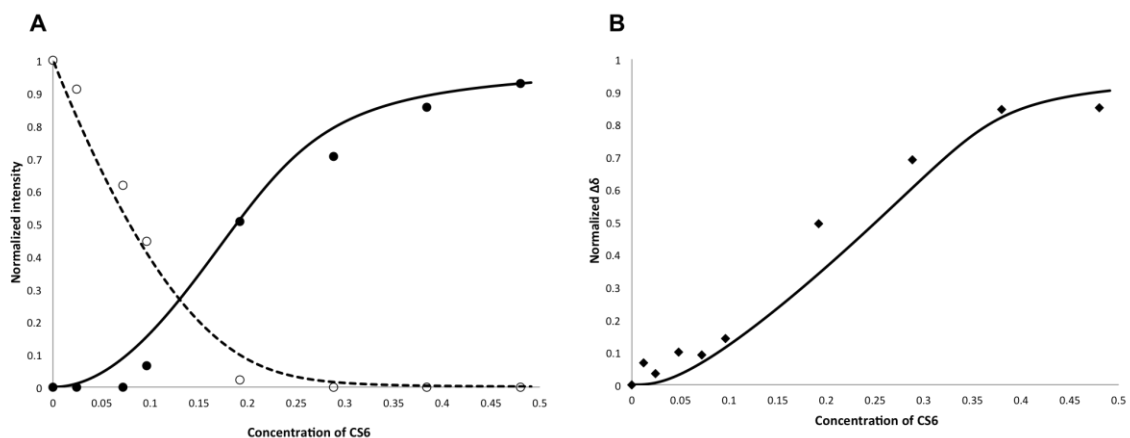
$$[PT] = [P] + 2 \times [C2] + K_1 \times [L] \times [P] \quad \text{[Equation 5]}$$

$$[LT] = [L] + K_1 \times [L] \times [P] + 3 \times [C2] \quad \text{[Equation 6]}$$

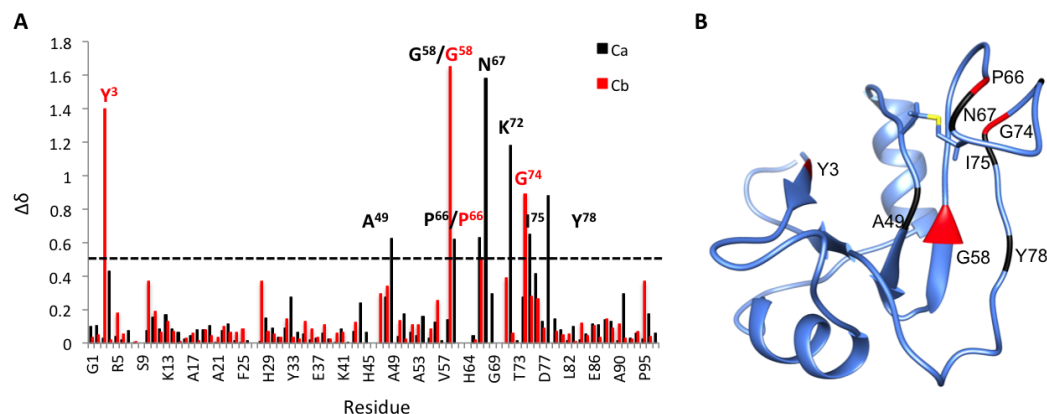
where [C1] and [C2] are defined to be  $[C1] = K_1 \times [P] \times [L]$  and  $[C2] = K_2 \times [C1] \times [L]$ , respectively. [C1] and [C2] indicate the concentrations of first and second complexes. [PT] and [LT] are total concentration of protein and ligand, respectively.  $K_1$  represents a binding constant for slow exchange residues involving in sugar binding.  $K_2$  indicates a binding constant for the second binding complex. Based on these, the simulations were carried out with a series of binding constants provided in advance in order to find the best fitting to the experimental data. [PT] was set to be 0.24 mM. 50 points with increasing ligand concentration were used. This equation allows  $K_D$  to be fit from measured values of the chemical shift at different concentrations of ligand and the estimated tight binding constant. The dissociation constant extracted is about 100  $\mu\text{M}$ .



**Figure 4.1:** Superimposed  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the Link module titrated with  $\Delta\text{C444S}$  compared to concentration of protein: black-control, without ligand; blue-40% ligand; cyan-80%; green-100%; magenta-150%; red-200%. Solid arrows represent residues showing fast exchange and circles show resonances experiencing slow exchange in the presence of a sugar bond.

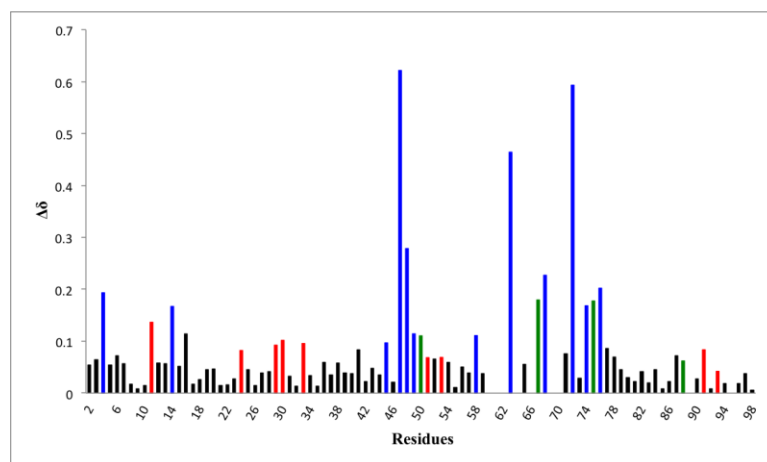


**Figure 4.2:** Fitting for intensity of residues in slow exchange (A). Fitting of chemical shift for residues in fast exchange (B). In A, averaged normalized intensities of residues, A48, G58, and G74 were used for uncomplexed (open circle with dotted fitting line) and complexed (closed circle with solid fitting line). The fitting lines are just smoothed curves and have no theoretical significance. For B normalized chemical shift change for residues W51, A53, and Y91 were used (closed diamond with solid fitting line). The fitting line was generated using equation 1-3 and 4-6 for slow and fast exchanges, respectively.



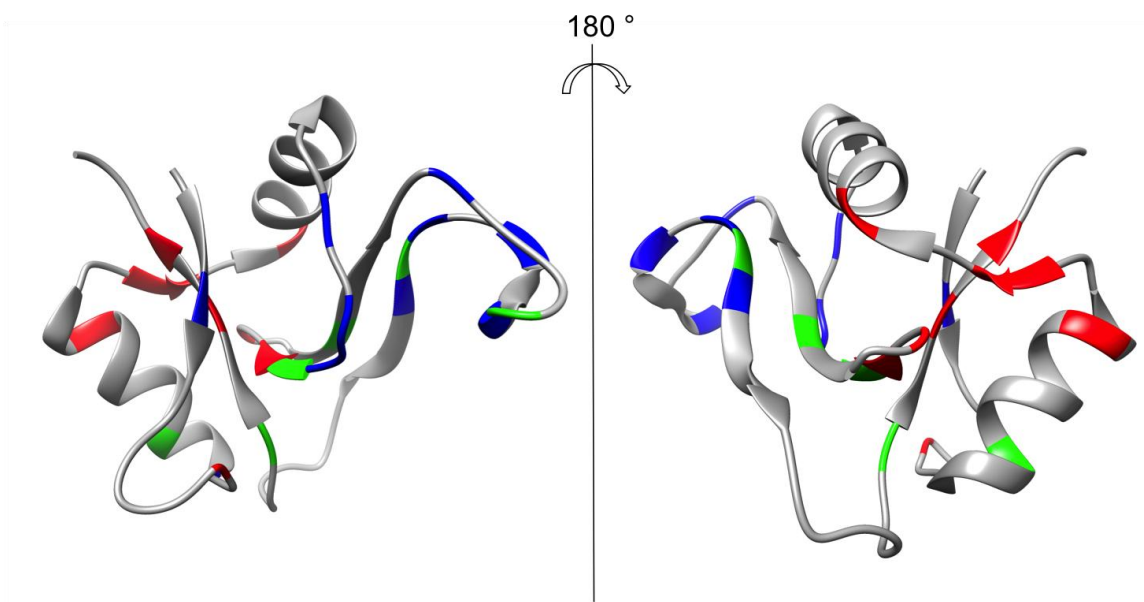
**Figure 4.3:** Chemical shift differences ( $\Delta\delta$ ) of CA (black bar) and CB (red bar) between free and ligand-bound Link-TSG-6 plotted as a function of residue number (A) and mapping of shift perturbations (B). The horizontal dotted line represents the arbitrary threshold levels chosen to indicate a significantly perturbed residue (CA and CB > 0.5 ppm). The residues above threshold for CA (black) and CB (red) are labeled with residue name and number. These residues are mapped onto the Link-TSG-6: $\Delta$ C444S model (B). C47-C68 disulfide bond is depicted on the protein. Yellow stick represents sulfur atom.

Figure 4.4 shows combined changes in  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for all assigned residues. The color-coded bars represent the residues involved in fast, intermediate and slow exchange when a ligand is added.



**Figure 4.4:** Chemical shift perturbation ( $\Delta\delta$ ) of Link module caused by  $\Delta$ C444S relative to free protein. Chart shows the absolute chemical shift difference in the presence of  $\Delta$ C444S compared with free protein. Red, green, and blue indicate residues belonging to fast, intermediate and slow exchanges, respectively.

In Figure 4.5, these residues are highlighted on an existing structure for the Link module of human TSG-6 (2PF5.pdb). The exchanges clearly involve somewhat different regions of the protein. As is described more thoroughly in the discussion, the slow exchange residues overlap extensively with the previously identified HA sites. It is less clear what processes might exhibit intermediate and fast timescale effects.



**Figure 4.5:** Mapping fast (red), intermediate (green), and slow (blue) exchanging resonances on the existing X-ray structure of the human Link-TSG-6 (2PF5.pdb). The backbone structure is shown with rounded ribbon in dark gray.

#### **4.3.2 Binding of Link-TSG-6 with a TEMPO derivative of $\Delta$ C444S**

In order to define the position of the bound  $\Delta$ C444S hexasaccharide, a paramagnetic version of  $\Delta$ C444S ( $\Delta$ C444S-TEMPO), which includes a nitroxide carrying TEMPO group attached to the reducing end, was used to generate additional quantitative data (sugar sequence; GlcA1-GalNAc2-GlcA3-GalNAc4-GlcA5-GalNAc6-TEMPO). A description of the reductive amination used to prepare this derivative is included in the Methods section. Figure 4.6 shows a comparison of a sample of 0.17 mM Link-TSG-6

with 0.34 mM  $\Delta$ C444S-TEMPO in oxidized (4.6A, paramagnetic) and reduced (4.6B, diamagnetic) states; the reduced state was achieved by the addition of twelve equivalents of ascorbic acid (47). With few exceptions, chemical shifts were identical suggesting that similar modes of binding occur with the TEMPO derivative compared to  $\Delta$ C444S. The time scales of exchange also appear to be similar. For example, A48 still shows slow exchange when it interacts with  $\Delta$ C444S-TEMPO.

In the oxidized state certain crosspeaks lose intensity due to the paramagnetically enhanced decay of transverse magnetization during the transfer and refocusing periods of the HSQC experiment. The changes in intensity between reduced and oxidized TEMPO derivatives can be converted to distances between the nitroxide oxygen and a residue of interest using equation 7 below (48-50).

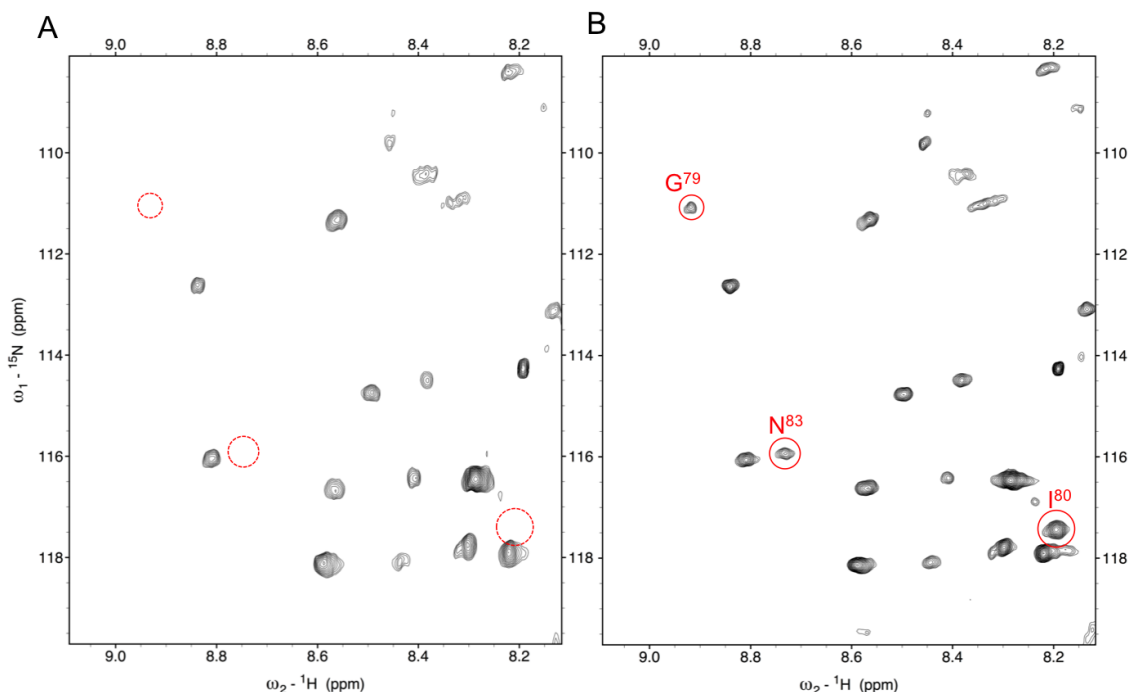
$$\ln\left(\frac{I_{red}}{I_{ox}}\right) = f \times t \times \left(\frac{K}{r^6}\right) \times \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2\tau_c^2}\right) \quad [\text{Equation 7}]$$

Here,  $I_{red}$  and  $I_{ox}$  are the peak intensities of the crosspeaks in the presence of reduced and oxidized TEMPO derivatives,  $f$  is the fraction of the protein bound to ligand, and  $t$  is the total time during INEPT and the refocusing periods of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC pulse sequence ( $t=9.78$  ms).  $K$  is a constant related to spin properties of the system ( $K=1/15 \times S(S+1)\gamma^2 g^2 \beta^2 = 1.23 \times 10^{-23} \text{cm}^6 \text{S}^{-2}$ ),  $r$  is the distance between the nitroxide and amide proton of the crosspeak of interest,  $\tau_c$  is the correlation time for tumbling of the protein-ligand complex, and  $\omega_H$  ( $800 \text{ MHz} \times 2\pi$ ) is the precession frequency for the amide proton. Given all values, and taking  $f$  and  $\tau_c$  to be 1 and 10.15 ns, respectively, the following distances were obtained.

**Table 4.1:** Estimated distance of resonances experiencing reduction of peak intensities.

Resonance	Link-TSG-6 (no ligand)	With spin label (oxidized)	With no spin label (reduced)	Distances (Å)
G79	1.00	0.12	0.65*	$12 \pm 4$
I80	1.00	0.00	0.89*	$<10$
N83	1.00	0.19	0.43*	$13 \pm 5$

\*Incomplete nitroxide reduction in the treatment of ascorbate



**Figure 4.6:** Expanded  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of Link-TSG-6 in the presence of TEMPO analogs. Comparison of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of Link-TSG-6 with two-fold molar excess of  $\Delta\text{C444S}$ -TEMPO oxidized (A), and reduced (B) by 4 mM ascorbic acid. Solid red circles indicate peaks reappearing in reduced forms of nitroxide and dotted red circles represent the same residues in oxidized forms of TEMPO. Red labels indicate the corresponding residue names.

These effects are purely distance dependent and relate only to the positioning of the terminal TEMPO group. Residues G79, I80, and N83 are the most affected by the presence of the TEMPO group, with the crosspeak for I80 completely disappearing. Even though a direct correlation with residues showing chemical shift perturbation was not necessarily expected, G79 and I80 are near the segment 74-76 which also showed

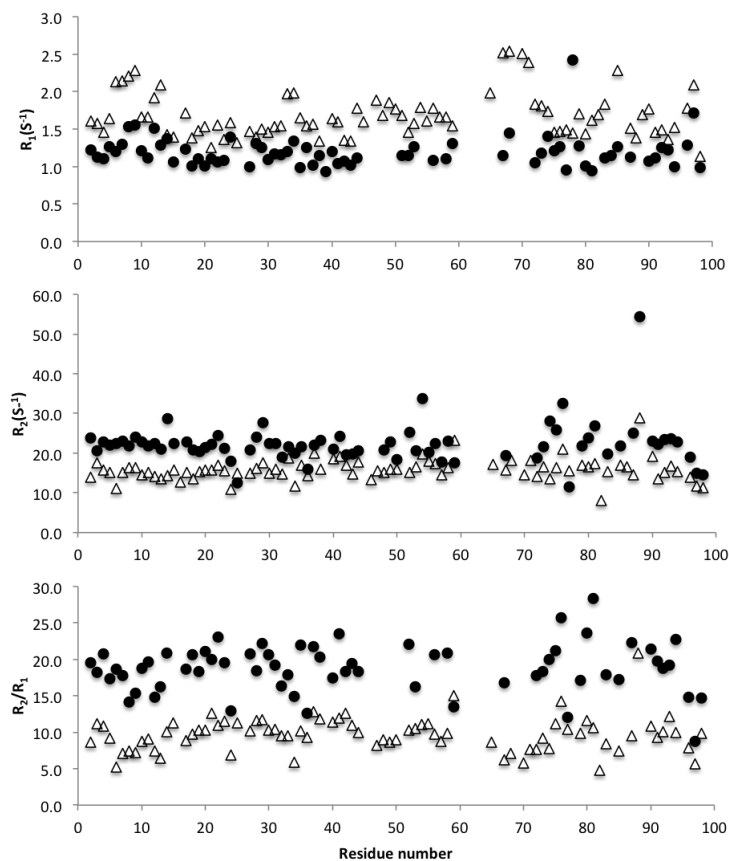


slow exchange, suggesting that distances involving these residues would be useful for positioning a ligand in the strongest binding site. Given that the TEMPO is attached to the reducing end of the hexasaccharide data are consistent with the findings by Higman et al. (2014), that the reducing end of the HA binds close to Y78 and R81 (14). The derived distances can be used as constraints in docking exercises. To allow for the presence of motion of the TEMPO moiety, large ranges were used with lower limits set to 3 Å and upper limits set to the measured distance plus 3 Å. The TEMPO moiety may have a high degree of mobility because it is linked to the reducing end of the open ring of the first GalNAc.

#### **4.3.3 Link module of TSG-6 forms a dimer in the presence of $\Delta$ C444S**

It is possible that some of the observed chemical shift perturbations arise from interactions that do not directly involve the ligand, but are interactions in remote parts of the protein that are induced by ligand binding. One possibility is ligand induced dimer formation with perturbed residues being involved in a dimer interface. NMR spin relaxation measurements provide one way of assessing the possibility of dimer formation.  $R_1$  and  $R_2$  relaxation rates for  $^{15}\text{N}$  nuclei in amide groups were measured in free and  $\Delta$ C444S-bound states. The ratio of  $R_2/R_1$  yields a rotational correlation time for a macromolecule, provided the amide sites measured do not experience substantial internal motion.  $R_1$ ,  $R_2$ , and  $R_2/R_1$  ratios are reported in Fig. 4.7. At 25 °C in dilute aqueous solution the correlation times are expected to be approximately equal to half the molecular weight in kDa. Hence, the values of somewhat less than 7.0 ns for Link-TSG-6 at a concentration of 0.26 mM in the absence of ligand are in reasonable agreement with the expected molecular weight for a monomer (10.9 kDa). In the presence of

$\Delta$ C444S, the correlation times rise nearly to 10.2 ns, a value consistent with significant dimer content in the solution. In addition, dynamic light scattering (DLS, data not shown) indicated the protein and sugar complex to be approximately 15.7 kDa in contrast to the monomer at 10.6 kDa.



**Figure 4.7:**  $R_1$ ,  $R_2$ , and  $R_2/R_1$  data for free ( $\Delta$ ) and  $\Delta$ C444S-bound ( $\bullet$ ) Link-TSG-6 at 800MHz, 25 °C. All samples were dissolved in 50 mM MES, pH 6.0.

Since  $\Delta$ C444S addition induces dimer formation, it is possible that some of the ligand induced chemical shifts attributed to  $\Delta$ C444S-protein interactions are the direct result of dimer formation. We therefore attempted to gain information about possible dimer contacts by examining  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra in the absence of  $\Delta$ C444S, but at higher concentrations. Comparing chemical shifts at 0.1 mM and 0.5 mM there were a number of measureable changes, including H4, R5, E6, A7, Y12, H29, L30, H45, and

H96. Two of these, H29 and L30, were discounted since these were apart from other residues and the fact that histidine ionization can be altered with slight changes in pH and ionic strength as the samples are concentrated. However, some of the other observed changes do correspond to those seen with the addition of ligand. We therefore discount these in identifying direct ligand interaction sites. Getting a structure for the dimer now becomes important in understanding how ligands bind to Link-TSG-6.

#### **4.3.4 Link-TSG-6:ΔC444S solution NMR structure**

The clear existence of a dimer and the possibility that ligand binding induces additional structural changes near the binding site put the use of existing structures determined in the absence of ligand into question. We, therefore, pursued solution structure determinations using conventional triple resonance assignment methods combined with NOE, dihedral angle, and RDC restraints. The process of data collection is presented in the Methods section, and a statistical summary of the data is included in Table 4.2. The initial structure for a monomer of Link-TSG-6 in the presence of ΔC444S was generated using CYANA version 3 with 1198 manually assigned NOEs out of 1329 total NOEs and 135 dihedral angles predicted from TALOS (51). In addition, two disulfide bonds were created and restrained to normal bond distances, based on their occurrence in all other Link module structures. The structure was then further refined using XPLOR-NIH with additional NOEs iteratively assigned based on initial structures from CYANA calculations, and with added RDC restraints. The final stage included refinement in the presence of discrete water molecules as solvent. 20 models having the lowest total energy were chosen and are represented in Figure 4.8.

**Table 4.2:** NMR and structural statistics for human Link-TSG-6. Structural statistics were computed for the ensemble of 20 NMR structures deposited in the Protein Data Bank. r.m.s., root mean square; r.m.s.d., root mean square deviation.

<b>Conformationally restricting constraints<sup>a</sup></b>	
Distance constraints	
Total	1329
Intraresidue ( $i = j$ )	260
Sequential ( $ i - j  = 1$ )	394
Medium range ( $1 < i - j \leq 5$ )	204
Long range ( $ i - j  > 5$ )	471
Distance constraints per residue	13.6
Dihedral angle constraints	135
No. of constraints per residue	15.0
No. of long-range constraints per residue	4.8
<b>Residual constraint violations<sup>a</sup></b>	
Average no. of distance violations per structure	
0.1–0.2 Å	11.45
0.2–0.5 Å	7.55
>0.5 Å	0.35
Average r.m.s. distance violations/constraint (Å)	0.03
Maximum distance violation (Å)	0.88
Average no. of dihedral angle violations per structure	
1–10°	18.9
>10°	0.3
Average r.m.s. dihedral angle violation/constraints (°)	1.49
Maximum dihedral angle violation (°)	16.80
<b>r.m.s.d. from average coordinates (all/ordered)<sup>a,b</sup></b>	
Backbone atoms (Å)	1.2/0.8
Heavy atoms (Å)	1.8/1.4
<b>Ramachandran statistics for ordered residues (Richardson MolProbity)<sup>a,b</sup></b>	
Most favored regions (%)	87.9
Additionally allowed regions (%)	11.0
Disallowed regions (%)	1.1
<b>Global quality scores (raw/Z-score)<sup>a</sup></b>	
Verify3D	0.37
ProsaII	0.54
PROCHECK ( $\phi$ - $\psi$ )	-0.60
PROCHECK (all)	-0.42
MolProbity Clash	21.93
<b>RDC statistics<sup>c</sup></b>	
No. of D <sub>NH</sub> constraints	46
R	0.99
Q <sub>r.m.s.</sub>	0.013

<sup>a</sup> Values were calculated using PSVS Version 1.5. Average distance violations were calculated using the sum over  $r^{-6}$

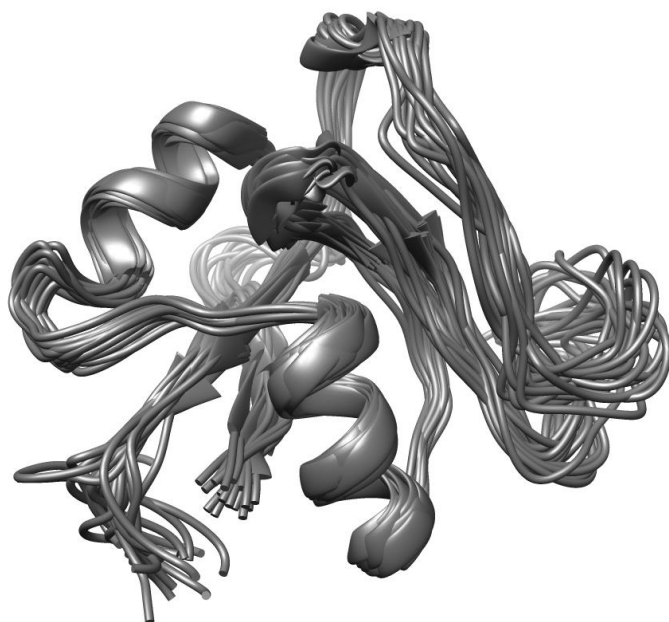
<sup>b</sup> *Ordered residue ranges:* 2-9,13-46,48-57,60-62,66-68, 81-94

Residues selected based on dihedral angle order parameter, with  $S(\phi) + S(\psi) \geq 1.8$

<sup>c</sup> RDC statistics were computed by PALES

This model has two identifiable alpha helices and a small beta sheet of two parallel and one antiparallel strand. One concern is that some NOEs may have arisen

from inter-monomer contacts in the dimer rather than intra-monomer contacts. After the generation of a dimer structure from the initial monomer structure (see below), inter-monomer contacts less than 5 Å for proton pairs were compared to NOE lists. 1736 contacts were found and 9 NOEs corresponding to those contacts were excluded in a second determination. The removal of these 9 NOEs no significant changes in the structure presented in Fig. 4.8.



**Figure 4.8:** Superimposition of 20 soluble structures of Link-TSG-6 in the presence of  $\Delta C444S$  hexasaccharide. Backbone structures appear as dark gray.

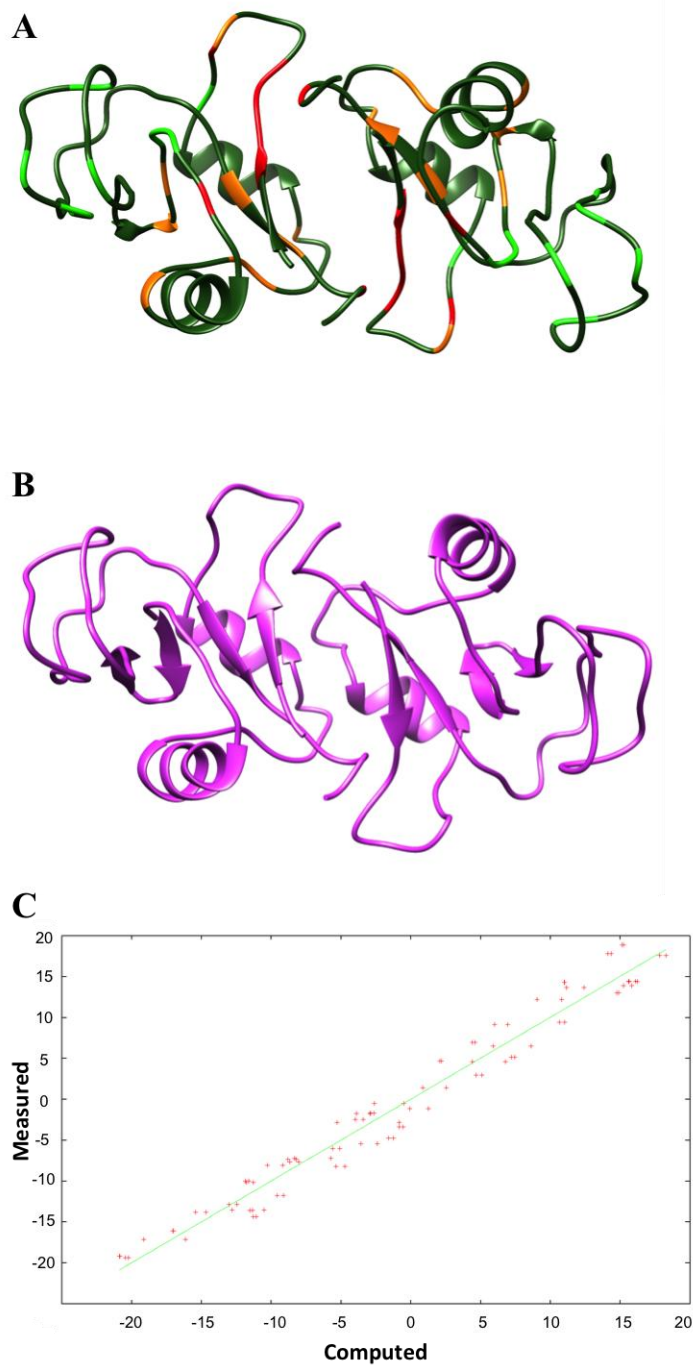
Comparing the 20 structures of Figure 4.8, significant variations among the structures were noted in the loop region between 63-73. This might be due to missing backbone resonances for several residues (I61, V62, G65, G59, F70, and G71) and few NOEs observed around the loop. The variations are also large for the C-terminal region. Here resonances and assignments exist but no NOEs are observed, consistent with a floppy C-terminal tail. Assignments, NOEs and RDCs of the complex of Link-TSG-

6:ΔC444S have been deposited into the Biological Magnetic Resonance Data Bank (BMRB) with accession number 25663 and the structures have been deposited into the Protein Data Bank (PDB) (2N40.pdb).

After generating the initial monomer structure, we addressed the issue of dimer formation. RDCs in combination with a reasonable monomer structure can be used to independently generate a set of possible structures for symmetric dimers (52). A principal alignment frame can be determined in which one axis must coincide with the  $C_2$  symmetry axis of the dimer. This works best when at least two sets of alignment conditions can be obtained and coincidence of one of the frame axes can be used to identify the  $C_2$  axis. Three possible sets of dimer structures can then be generated by rotating the monomer  $180^\circ$  about the x, y, or z axes followed by translation in the plane perpendicular to the rotation axis in order to find the best interface for dimer. The residues identified from concentration dependent chemical shifts as possibly involved in the dimer interface could then be used to select the best translational position from each set (52). In our case, working with a positively charged Link module and negatively charged ligand greatly restricted a choice of alignment media, leaving only a single set of RDCs. With three possible choices for  $C_2$ , this results in a much larger set of possible structures. In light of the large number, we chose a different approach, one that used a docking program to generate a dimer and then compared this dimer structure to RDC-based models.

The docking program, HADDOCK 2.1, was used to generate a dimer structure (44,45). We began with the lowest energy monomer structure from those shown in Figure 4.8. 46 RDC restraints and a symmetry restraint were given to guide dimer

formation. These residues were chosen based on the chemical shift changes of the titration as a function of protein concentration. Residues 9–12 were kept flexible during docking and refinement to encourage development of the best interface. Comparisons of a docking model with active residues of H4, R5, E6, A7, Y12, H45 and H96 for docking at the lowest total energy and one generated using rotation about the y principal axis of the alignment tensor determined from fitting RDCs are shown in Figures 4.9A and 4.9B. The correlation plot of experimental versus back-calculated RDCs using the best set of alignment parameters for the dimer model from HADDOCK is shown Figure 4.9C. The Q-factor was 0.18 with principal order parameters,  $1.68 \times 10^{-4}$ ,  $7.28 \times 10^{-4}$ ,  $-8.96 \times 10^{-4}$  and Euler angles, -179.55, 91.5, 1.33. The structure appears to have good stabilizing contacts with identifiable inter-monomer hydrogen bonds (Y3-R5, H4-H96, R5-E26, R5-N94, R8-E26, R8-Y93), four ion pairs (two from E26-R8 and two from R5-E26), several hydrophobic contacts (Y3-Y3, P95-R8, H4-H96), and a decrease in the solvent accessible surface (SAS) of 723.3 Å.



**Figure 4.9:** A dimer model of Link-TSG-6. (A) shows the Link-TSG-6 dimer with dark green ribbon. The dimer interface designated for docking in HADDOCK included H4, R5, E6, A7, Y12, H29, and H96 (red).  $\Delta$ C444S binding sites are colored green; fast exchange residues are highlighted in orange. The dimer interface is opposite to the  $\Delta$ C444S binding sites. (B) shows a dimer model of Link-TSG-6 generated by the RDC-based method. (C) shows back calculated RDC fitting to dimer model in A. The Q-factor is 0.18 and the experimental error is  $\pm 1.6$  Hz.

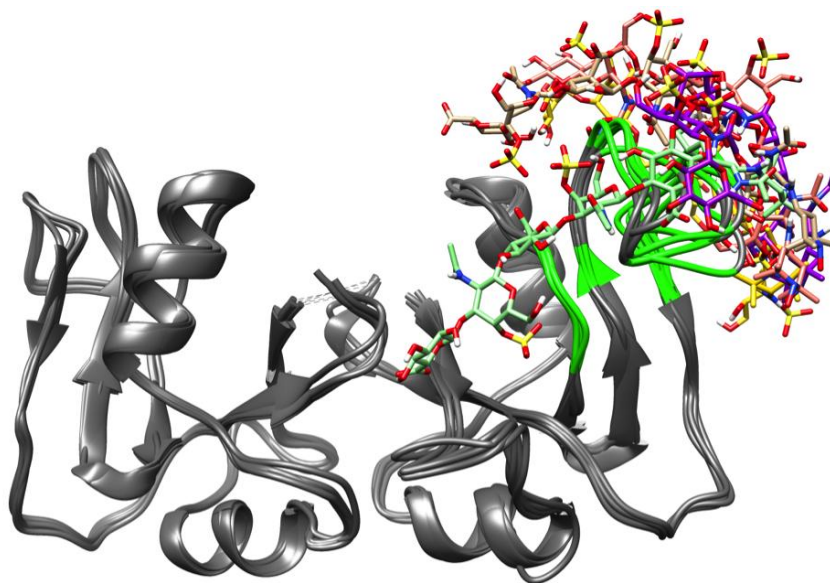


#### **4.3.5 Docking of $\Delta$ C444S to the primary site in the Link-TSG-6 Dimer**

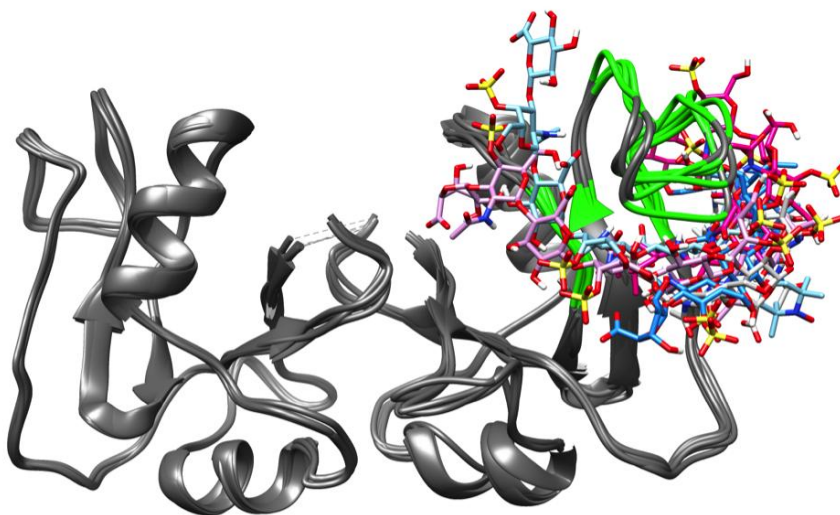
$\Delta$ C444S-TEMPO was docked to the above-described dimer model using the following ambiguous interaction restraints (AIR); C47-A49, G58, K63-G65, and C68-I76 for the protein and all residues for  $\Delta$ C444S-TEMPO. Unambiguous distance restraints coming from PREs are given in Table 4.1 and the 46 RDC restraints on protein  $^{15}\text{N}$ - $^1\text{H}$  vectors were provided as well. The top 10 scoring complexes from HADDOCK were selected and clustered based on where they docked on the protein. Two clusters were classified as shown Figures 4.10A and B. In Figure 4.10A, a set of 5 models is chosen that has the sugars sitting on the protein surface with a broad range of conformations. In Figure 4.10B, a set of 5 models docked into a shallow groove overlapping the HA binding site. This cluster makes better contacts with the perturbed resonances. The model with the lowest energy among the top 3 scoring models, based on the sum totals of  $E_{\text{tot}}$ ,  $E_{\text{bond}}$ ,  $E_{\text{angle}}$ ,  $E_{\text{improper}}$ ,  $E_{\text{dihed}}$ ,  $E_{\text{vdw}}$ , and  $E_{\text{elec}}$ , is a member of cluster 2. It also shows lowest electrostatic energy ( $E_{\text{elec}}$ ) among the top 3 models. It was chosen for further analysis of specific interactions between protein and  $\Delta$ C444S (Figure 4.11). 6 potential hydrogen bonds ( $\text{H}\dots\text{A}$  distance  $< 2.5\text{\AA}$ ) were observed between the protein and  $\Delta$ C444S-TEMPO (GlcA1-O5:H45-HE2, GalNAc2-O3: H45-HE2, GlcA3-O5:N67-HD21, GlcA3-OE2:N67-HD21, GlcA3-OE1H:C68SG and Glc5-O5:G69-NH). In addition 115 atom pairs are within  $0.2\text{\AA}$  of van der Waals contact across the protein-sugar interface. Residues K11, H45-A49, Y59, P64, N67-G59, F70 K72, I76 and Y78 were involved in these contacts. Sulfate groups on two GalNAc residues were in close contact with the side chain of positively charged amino acids (K11 and H45). Other sulfate groups are located outside of the binding groove where they can make effective contact with water.

The nitroxide on the TEMPO group showed distance to G79 (11.76 Å), I80 (12.58 Å), and N83 (9.39 Å) that are consistent with distance restraints from the PRE data. Figure 4.12 shows the second binding site (fast exchange effects) resulting from CSP. This suggests that second binding site is close to dimer interface.

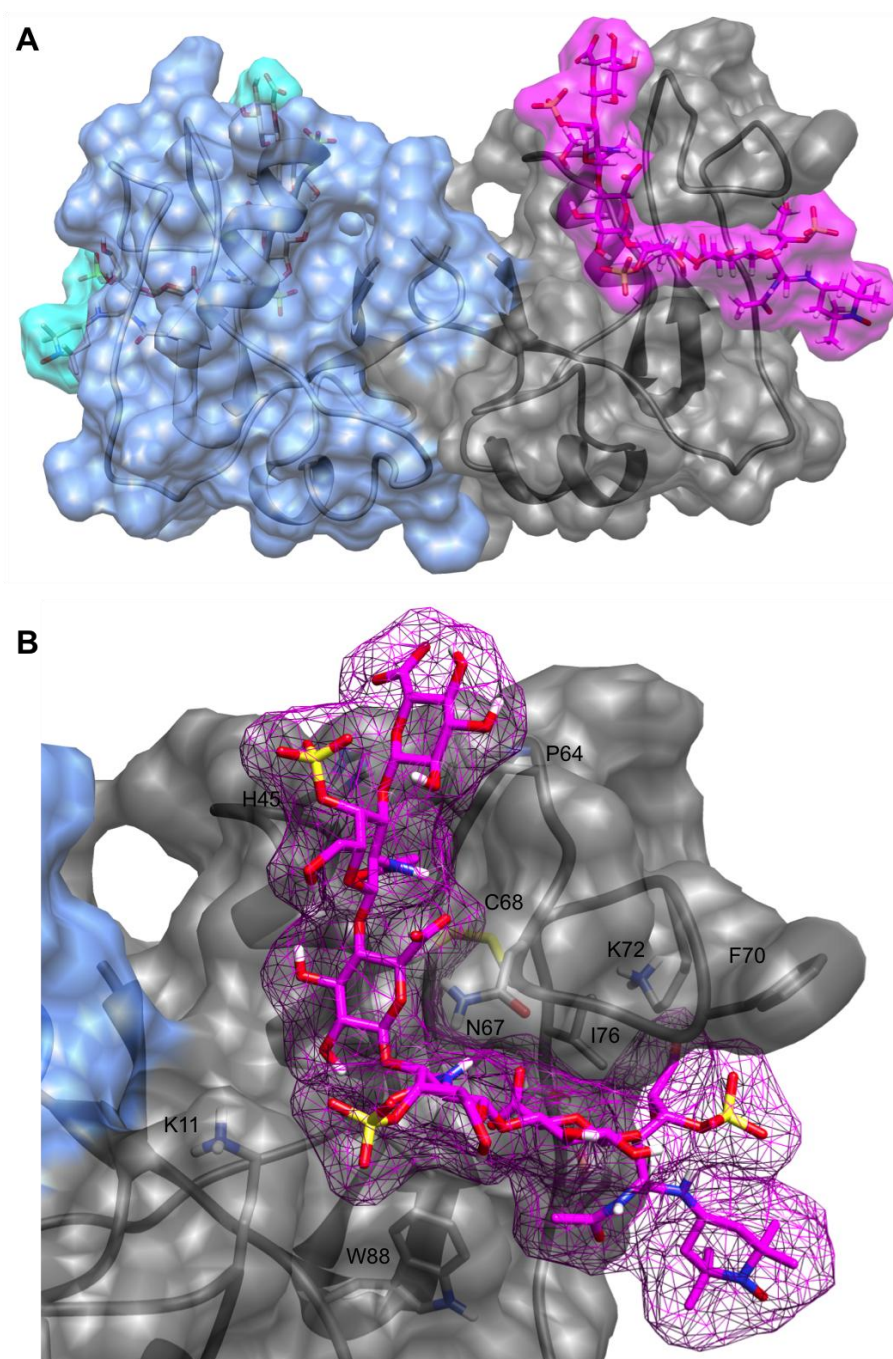
**A**



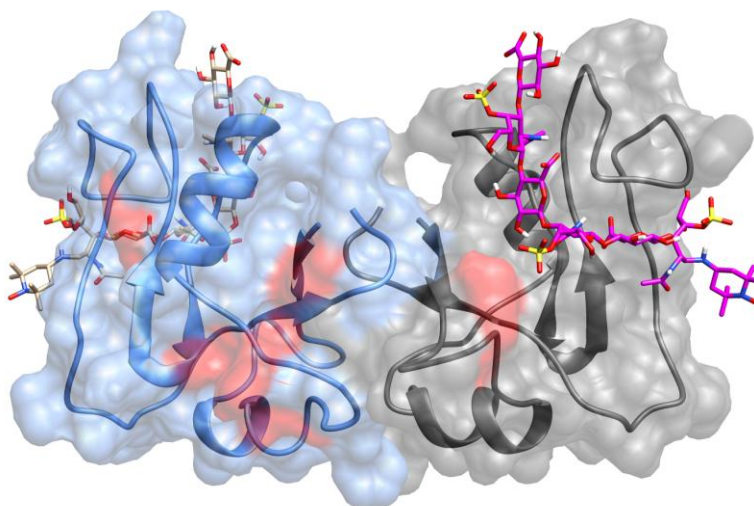
**B**



**Figure 4.10:** Complex models for the dimer of Link module and  $\Delta$ C444S-TEMPO. 10 top-scored models are shown which are classified into two groups, A and B. Group A shows 5 models with ligand binding to the protein surface rather than the HA groove. Group B shows 5 models with ligand docking to the binding surface composed of the slow exchange residues from titration experiment. The protein backbone is shown in dark grey. Designated ligand binding sites for the docking are shown in green.

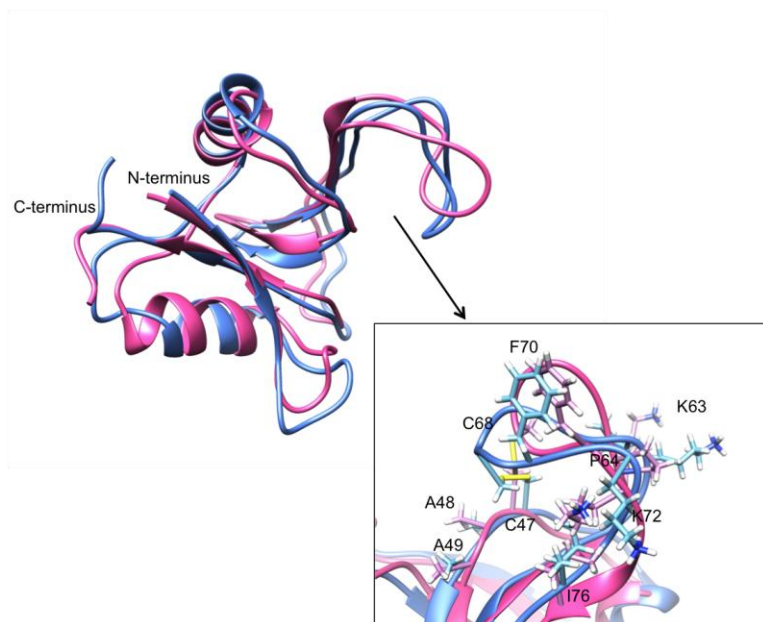


**Figure 4.11:** Model of the Link-TSG-6 dimer and the chondroitin sulfate hexasaccharide-TEMPO,  $\Delta$ C444S-TEMPO. (A) shows the dimer model with a ribbon and surface structure (shown in gray and blue for each monomer). The two docked ligands are shown as stick structures with surfaces shown in magenta and cyan. (B) shows an expansion of binding interface between the Link module and  $\Delta$ C444S-TEMPO. The residues in close contact to sugar are shown with the labels.



**Figure 4.12:** Weak binding site for  $\Delta C444S$  on Link-TSG-6. Each monomer is represented with blue and gray ribbons with surfaces shown as semi-transparent. Tight-binding ligands are docked as in Figure 4.11. Link-TSG-6 residues showing fast exchange effects on  $\Delta C444S$  binding are colored in red.

#### 4.4 Discussion



**Figure 4.13:** Superimposition of structures of Link-TSG-6 in its HA and  $\Delta C444S$  bound states. Backbone structures of HA (1O7C.pdb, pink ribbon) and  $\Delta C444S$  (blue ribbon) are shown with N- and C-terminals labeled. The inset shows some of the putative  $\Delta C444S$  binding residues.

Structural aspects of the interactions of Link-TSG-6 with both HA and heparin (which is closely related to matrix heparan sulfate) have been studied previously (13,16-20,22,23). There has also been competition assays that suggest that HA and CS-4 bind to similar sites on the Link module (20). And, more recently, there has been some work with hybrid HA oligomers containing desulfated chondroitin sulfate segments that suggested these segments can be placed into a narrow binding groove of the Link module in the same manner as HA. Here we used a highly sulfated hexasaccharide ( $\Delta$ C444S) model to provide additional structural evidence of Link-TSG-6 interacting with chondroitin sulfate (CS). Several modes of interaction were discovered. One displays slow exchange on the NMR timescale and an affinity similar to that found for HA (22). Docking of  $\Delta$ C444S on the surface of Link-TSG-6 still shows contacts with residues that overlap significantly with residues that are found for the HA binding (19,21). The deviation from an exact overlap in binding is likely due to the presence of sulfate groups on each GalNAc residue of  $\Delta$ C444S, and some repositioning required for additional interactions with these highly charged groups.

The binding region of the protein also seems to experience some adjustments on interaction with  $\Delta$ C444S. This can be seen in Figure 4.13 where our lowest energy structure in the presence of  $\Delta$ C444S is compared to a previously determined structure (1O7C) with an HA oligomer bound (Fig. 4.13). The overall differences are small with a RMSD for backbone atoms of 1.34 Å. However, the large loop (residue 63-73) involved in binding  $\Delta$ C444S or HA shows a larger difference. The binding groove appears to be closed more tightly when  $\Delta$ C444S is present than when HA is present.

A second mode of interaction of  $\Delta C444S$  with the Link module exhibiting fast exchange and a  $K_d > 100 \mu M$  was also observed. Interestingly, the residues perturbed in this interaction are near some of the residues previously identified as important for heparin binding (K20, K34, K41, and K54) (15,22). As shown in Fig. 4.5, the residues perturbed in this weaker interaction are physically separated from those exhibiting slow exchange.

One notable discovery in this study is that binding of  $\Delta C444S$  to the Link module induces dimer formation. The dimer model presented in Figure 4.11 was determined using RDC data and a list of residues showing substantial concentration dependent chemical shift changes (H4, R5, E6, A7, Y12, H45, and H96). The dimer structure has an interface distinctly different from the possible dimer contacts suggested by crystal structure determined in the absence of ligands (15). It also differs from a dimer found in the drosophila crystal structure where a piece of heparin bridges two monomers (22). However, the reduction in solvent accessible surface areas for the three possible dimers in the non-liganded crystal structure are not significantly different (625.34, 884.15, 612.36  $\text{\AA}^2$ ) from that observed here 723.3  $\text{\AA}^2$ . It is possible that other dimer structures can exist, but populations of various dimers are influenced by crystal packing.

In our model, primary  $\Delta C444S$  binding sites exist on each monomer, nearly on opposite sides of the dimerization interface, and cannot play a bridging role in dimer formation. It is possible that, in the absence of a bridging ligand, electrostatic effects are largely responsible for dimerization. Link-TSG-6 is a positively charged domain ( $pI = 9.48$ , at lower pH where histidines are also protonated). The binding of a negatively charged ligand would then reduce the repulsion that exists in the unliganded monomers



and allow dimerization. One of the more interesting aspects of dimerization is that residues that seem to be involved in the second (fast exchange) sites on each monomer are closer to one another in the dimer (along the bottom left of the dimer structure in Fig. 4.11A, and Fig. 4.12), raising the possibility that a longer oligomer could bind across the interface with higher avidity than observed with our hexasaccharide. However, more study is needed to clarify the role of the second binding sites. We can also not rule out the possibility that longer oligomers may connect even the primary binding sites by extending through the channel that can be seen at the top middle of the dimer structure. The picture that emerges is a module that could be very effective in crosslinking HA and other glycosaminoglycans found in the extracellular matrix, greatly contributing to the stability and specific geometries necessary for interactions with other protein components.

There is substantial precedent for dimerization of link modules, or the molecules that carry them, as well cross-linking of ECM components by the resulting dimers. And there are suggested biological consequences. Heparin, for example, is known to induce dimer formation of Link-TSG-6 in a manner that may potentiate the anti-plasmin activity of inter- $\alpha$ -inhibitor (I $\alpha$ I) (22). Cross linking of hyaluronan by TSG-6 dimers has also been proposed in the literature (53). These studies were conducted with Link-TSG-6 and full length TSG-6 in which a CUB module provides the primary dimer interface. The introduction of the cross-link between HA by TSG-6 is suggested to bring neighboring HAs closer and reduce the mesh size of ECM during inflammation (53). Moreover, it has been proposed that the cable-like structure formed as a result of cross-link may provide a protective mechanism in promoting adhesion of leukocytes to the hyaluronan complex rather than contacting inflammation-associated proteins at inflammation sites (54). Our

studies with a truncated Link module demonstrate that the Link module itself can bind to  $\Delta C444S$  and form a dimer structure which cross links neighboring glycosaminoglycans. Thus, the Link module's propensity for dimerization and ability to interact with multiple ECM components could play a significant role in stabilizing and remodeling the ECM. These observations are an important step toward discovering novel mechanisms for stabilizing and reconstructing the ECM during injury and inflammation. Specific information related to how the Link module interacts with extended GAGs and how cross links occur in the ECM require additional studies. As mentioned earlier, obtaining homogeneous preparations of longer GAGs oligomers can be challenging; the heterogeneity of the GAGs exponentially increase with oligomer size. Our techniques successfully used bacterial lyases to depolymerize polysaccharides of CS-A and CS-C and separate oligomers up to hexasaccharides (55). Other preparative protocols combined with more sophisticated structural methodology may allow these studies in the future.

## **4.5 Methods**

### **4.5.1 Expression, purification and refolding processes of Link-TSG-6**

Expression and refolding followed published procedures (56). More specifically, the engineered gene of Link-TSG-6 in the pRK172 expression vector was transformed into expression host BL21(DE3)pLysS. 10 ml starter cultures were grown overnight in LB medium containing 100  $\mu\text{g/ml}$  ampicillin. Grown cells were then transferred to 1 L M9 medium containing isotopes,  $^{15}\text{N-H}_4\text{Cl}$ ,  $^{13}\text{C-glucose}$  from CIL (Cambridge Isotope Laboratories) and protein expression was induced by adding IPTG to 0.1 mM when the  $\text{OD}_{600\text{nm}}$  reached 0.4. Cells were harvested 4 h after induction by centrifugation for 15



min at 5000g and stored at 20 °C in lysis buffer (50 mM Tris-HCl, pH 7.8, 150 mM NaCl, and 1mM EDTA). Cells were lysed by three passages through a French press and cells and inclusion bodies were gathered by centrifugation for 45 min at 20000 g. (56) Inclusion bodies were solubilized in 6M Guanidine-HCl containing 50mM Tris-HCl, pH 8.0, and 100 mM DTT and the sample was loaded onto a Superdex 75 (16 x 260 mm, Pharmacia) exclusion column, equilibrated and run with a flow rate at 1ml/min in the same buffer without DTT. Fractions containing the target protein were injected onto a C4 column (10 x 250 mm, YMC<sup>®</sup>), equilibrated with H<sub>2</sub>O containing 0.01% TFA. After 5 min of washing with the same solvent at 3 ml/min, the protein was eluted over 40 min using a linear gradient from 0 to 80% acetonitrile containing 0.1% TFA. The eluent was monitored at 220nm continuously. Eluent containing the denatured Link module was collected and lyophilized.

To accomplish the folding of Link-TSG-6, the protein was re-suspended in 50 mM ammonium acetate at pH 6.0 at a concentration of 500 µg/ml (40 ml final volume). A 100-fold molar excess of  $\beta$ -mercaptoethanol was added, and the refolding solution was incubated at 25 °C for 2 days under aerobic condition without stirring. After 2 days the refolding solution was stored at the 4 °C for 5 additional days. The folded Link-TSG-6 protein was purified by HPLC in the same manner as described above.

#### **4.5.2 Preparation of hexasaccharides of chondroitin sulfate**

The method described below is based on the works in Pomin *et al.* (55) 150 mg of chondroitin sulfate A from bovine trachea (btCS-A) was incubated with either 5 IU of chondroitin C lyase from *Flavobacterium heparinum* or 0.33 IU of chondroitin ABC lyase from *Proteus vulgaris* in 5 ml digesting buffer (50 mM Tris-HCl, pH 8.0, 150 mM

sodium acetate, 100 µg/ml of BSA) at 37 °C for 36 h and 150 min, respectively. Hyaluronidase digestion of btCS-A was carried out by incubation of 150 mg of btCS-A with 10 mg of enzyme in 3 ml of digesting buffer (50 mM sodium phosphate, 150 mM NaCl, pH 6.0) at 37 °C for 48 h. The digested samples were heated to quench enzyme activities and subjected to separation on a Bio-Gel P-10, size exchange column (15 x1200 mm) with elution buffer (1 M NaCl, 10% ethanol) at a flow rate of 1.7 ml/15 min/fraction. Eluent was monitored at 232nm and separated peaks in the chromatogram were desalted on a Sephadex G-15 column (10 x 500 mm) and examined by mass spectrometry (MS) to identify those peaks having molecular weights corresponding to hexasaccharides. To eliminate complexities from anomeric equilibrium at the reducing end, hexasaccharides were reduced in the presence of one equivalent of sodium borohydride (NaBH<sub>4</sub>) in 1 ml of water for 3 h and the reaction was quenched by adding a molar equivalent of acetic acid for 1 h in an ice-bath, followed by desalting. Hexasaccharides were then purified on a strong anion exchange column (SAX, 250 x 10 mm) using a linear gradient from 25 to 45% of 2 M NaCl, 25 mM phosphate, pH 4.5, for 50 min at a flow rate of 3 ml/min. Each peak from the SAX chromatography was desalted and lyophilized for other experiments. Structures were identified by NMR spectroscopy.

#### **4.5.3. Synthesis of ΔC444S-TEMPO**

4-amino-TEMPO (4-amino-2,2,6,6-tetramethylpiperidine-1-oxyl) was purchased from Acros Organics. The reductive amination reaction was carried out on a 400 µl sample containing 1 mg ΔC444S, 73 mM of 4-amino-TEMPO and 250 mM sodium cyanoborohydride in 80% MeOH at 65 °C for 3h. The sample was desalted and further purification was performed using SAX-HPLC monitored at 232nm.

#### 4.5.4. NMR spectroscopy

NMR spectra for titrations of protein with several isomers of CS were recorded at 25 °C on Varian 800- and 900-MHz spectrometers and processed with NMRPipe and Sparky (57,58). Titration experiments were carried out using 0.2-0.5 mM uniformly  $^{15}\text{N}$ -labeled Link module with increasing amounts of CSs ( $\Delta\text{C664S}$ ,  $\Delta\text{C444S}$ , C664S, and C444S) in 50 mM MES buffer, pH 6.0, 0.02%  $\text{NaN}_3$ , 10%  $\text{D}_2\text{O}$ . A series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were recorded after adding sugar at 0.096, 0.192, 0.29, 0.38, and 0.48 mM to the protein sample. The degree of chemical shift change for amide proton and nitrogen resonances was calculated using the empirical formula,  $\Delta\delta = [(\Delta\delta_{\text{HN}})^2 + (\Delta\delta_{\text{N}} \times 0.17)^2]^{1/2}$  where  $\Delta\delta_{\text{HN}}$  and  $\Delta\delta_{\text{N}}$  are the observed chemical shift changes for  $^1\text{H}$  and  $^{15}\text{N}$ , respectively. The weighting factor of 0.17 reflects the difference in chemical shift dispersion of  $^1\text{H}$  and  $^{15}\text{N}$  in folded proteins. (59) For relaxation experiments using a TEMPO labeled version of  $\Delta\text{C444S}$ , a 0.17 mM  $^{15}\text{N}$ -labeled LinK-TSG-6 sample was prepared and spectra were recorded at 25 °C on a Varian 800-MHz spectrometer.

RDCs were measured on a 0.5 mM  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled Link-TSG-6 sample in the presence of 1 mM  $\Delta\text{C444S}$  aligned in stretched neutral polyacrylamide gel. Gels were cast initially in a 4.5 mm diameter glass tube overnight for polymerization. The polymerized gels were washed for two or three cycles in deionized water overnight, followed by a washing with protein buffer to equilibrate the pH. Finally, the gels were washed again with deionized water to remove residual protein buffer. Swollen gels were then trimmed to a length of 26 mm and dried at room temperature for 2 days. The dried gel was placed in the upper stage of an NMR tube and protein sample was added to cover the gel and let it swell for 2 days. Spectra were recorded using a two-stage NMR tube as

described in Liu et al. (60). RDC values were measured using a  $^1\text{H}$ - $^{15}\text{N}$  HSQC IPAP (in-phase and anti-phase) experiment under both isotropic and anisotropic conditions (43).

Complete backbone and side chain resonance assignments were obtained with 0.5 mM  $^{13}\text{C}$  and  $^{15}\text{N}$  enriched Link-TSG-6 in the presence of 0.6 mM  $\Delta\text{C444S}$  using the following heteronuclear 2D and 3D experiments:  $^1\text{H}$ - $^{15}\text{N}$  HSQC, CBCA(CO)NH, HNCACB, HNCO and HCCH-TOCSY (61-68). All spectra for assignment were collected at 25 °C on Varian 600- and 900-MHz spectrometers.  $^{15}\text{N}$ , and  $^{13}\text{C}$ -edited NOESY-HSQCs were collected on same sample at 800-MHz and used for structure determination (69,70).

In order to measure  $^{15}\text{N}$   $T_1$  and  $T_2$  relaxation rates and evaluate the level of dimerization, spectra were recorded at 25 °C on a Varian 800-MHz spectrometer using standard sequences from BioPack (Varian/Agilent).  $^{15}\text{N}$   $T_1$  values were measured from the spectrum with different delay times,  $T = 10, 20, 30, 40, 50, 70, 90, 110, 150, 200, 250, 300, 400,$  and  $500$  ms.  $T_2$  values were determined from spectra recorded with delays  $T = 10, 30, 50, 70, 90, 110,$  and  $150$  ms.  $T_1$  and  $T_2$  values were extracted from log plots of peak intensities in 2D spectra as a function of delay duration.

#### **4.5.5. Dynamic Light scattering (DLS)**

260  $\mu\text{M}$  protein and 1 mM protein in the presence of 0.5 mM sugar were used. The samples were spun down for 30 min at 5000 g and 20  $\mu\text{l}$  was placed into a 1.5 mm cuvette. Light scattering experiments were carried out for both standard and Link-TSG-6 samples at 25 °C in a PROTEIN SOLUTIONS DynaPro 99.

#### **4.5.6. NMR solution structure calculation**

All NOESY crosspeaks were picked using Sparky and assigned manually followed by several rounds of automatic assignment. 1329 NOEs and 135 backbone torsion angle constraints derived from TALOS (71) based on the assigned chemical shifts of HA, CA, CB, CO, and N were used to calculate the solution structure. CYANA version 3.0 calculations were performed and 20 structures with the lowest energies were selected for analysis and further refinement (72,73). In the refinement stage, residual dipolar couplings were incorporated starting with Da and Rh values calculated from principle order parameters determined in REDCAT (-12.15 and 0.433, respectively). Structural refinement was then carried out using NOE distances, dihedral angles, and RDC values with Xplor-NIH (74).

#### **4.5.7. Computational docking**

HADDOCK version 2.1 was used for protein-protein and protein-ligand docking. The newly determined NMR structure of the protein was used for docking. The C444S structure was built in Glycam Biomolecular Builder (75) (also <http://dev.glycam.org>) and the TEMPO group was added manually using UCSF Chimera (76). Residues H4, R5, E6, A7, Y12, H45 and H96, were set as highly ambiguous interaction restraints (AIRs) for protein-protein docking. Residues, C47-A49, G58, K63-G65, and C68-I76, were used as AIRs for dimer- $\Delta$ C444S-TEMPO docking. In addition, PRE data from spin labeled ligands were used as unambiguous interaction restraints for dimer- $\Delta$ C444S-TEMPO docking. Initially 1,000 structures were determined by rigid-body docking. Then simulated annealing (SA) was carried out with the 200 lowest-energy structures using default force field parameters. Specified residues were allowed to be fully or semi

flexible during various stages of simulated annealing. For ligand docking  $\Delta$ C444S-TEMPO was allowed to be fully flexible and residue segments of the protein 45-49, 63-65, and 68-76, were allowed to be semiflexible. For monomer-monomer docking to produce a dimer protein residues 9-12 near dimer interface were specified as semiflexible during the SA. Semiflexible residues were allowed to have side chains move during the SA. Fully flexible residues were allowed to be fully flexible throughout the entire docking protocol except for the rigid body minimization. RDC restraints were given during docking simulations and a tensor was included in the structures calculations. Finally the structures were subjected to a water refinement stage. Models were clustered based on similarity of poses and ranked based on total score. The top five in each cluster were chosen for display and the lowest energy model of the top three scores was chosen for more detailed analysis. The best structures were chosen from the refined structures based on lowest total energy and docking score.

#### **4.6 References**

1. Milner, C. M., and Day, A. J. (2003) TSG-6: a multifunctional protein associated with inflammation. *J Cell Sci* **116**, 1863-1873
2. Ye, L., Mora, R., Akhayani, N., Haudenschild, C. C., and Liao, G. (1997) Growth factor and cytokine-regulated hyaluronan-binding protein TSG-6 is localized to the injury-induced rat neointima and confers enhanced growth in vascular smooth muscle cells. *Circ Res* **81**, 289-296
3. Carrette, O., Nemade, R. V., Day, A. J., Brickner, A., and Larsen, W. J. (2001) TSG-6 is concentrated in the extracellular matrix of mouse cumulus oocyte complexes through hyaluronan and inter-alpha-inhibitor binding. *Biol Reprod* **65**, 301-308
4. Fulop, C., Kamath, R. V., Li, Y., Otto, J. M., Salustri, A., Olsen, B. R., Glant, T. T., and Hascall, V. C. (1997) Coding sequence, exon-intron structure and chromosomal localization of murine TNF-stimulated gene 6 that is specifically expressed by expanding cumulus cell-oocyte complexes. *Gene* **202**, 95-102

5. Mukhopadhyay, D., Hascall, V. C., Day, A. J., Salustri, A., and Fulop, C. (2001) Two distinct populations of tumor necrosis factor-stimulated gene-6 protein in the extracellular matrix of expanded mouse cumulus cell-oocyte complexes. *Arch Biochem Biophys* **394**, 173-181
6. Ochsner, S. A., Day, A. J., Rugg, M. S., Breyer, R. M., Gomer, R. H., and Richards, J. S. (2003) Disrupted function of tumor necrosis factor-alpha-stimulated gene 6 blocks cumulus cell-oocyte complex expansion. *Endocrinology* **144**, 4376-4384
7. Ochsner, S. A., Russell, D. L., Day, A. J., Breyer, R. M., and Richards, J. S. (2003) Decreased expression of tumor necrosis factor-alpha-stimulated gene 6 in cumulus cells of the cyclooxygenase-2 and EP2 null mice. *Endocrinology* **144**, 1008-1019
8. Oh, J. Y., Roddy, G. W., Choi, H., Lee, R. H., Ylostalo, J. H., Rosa, R. H., Jr., and Prockop, D. J. (2010) Anti-inflammatory protein TSG-6 reduces inflammatory damage to the cornea following chemical and mechanical injury. *Proc Natl Acad Sci U S A* **107**, 16875-16880
9. Bardos, T., Kamath, R. V., Mikecz, K., and Glant, T. T. (2001) Anti-inflammatory and chondroprotective effect of TSG-6 (tumor necrosis factor-alpha-stimulated gene-6) in murine models of experimental arthritis. *Am J Pathol* **159**, 1711-1721
10. Glant, T. T., Kamath, R. V., Bardos, T., Gal, I., Szanto, S., Murad, Y. M., Sandy, J. D., Mort, J. S., Roughley, P. J., and Mikecz, K. (2002) Cartilage-specific constitutive expression of TSG-6 protein (product of tumor necrosis factor alpha-stimulated gene 6) provides a chondroprotective, but not antiinflammatory, effect in antigen-induced arthritis. *Arthritis Rheum* **46**, 2207-2218
11. Mindrescu, C., Thorbecke, G. J., Klein, M. J., Vilcek, J., and Wisniewski, H. G. (2000) Amelioration of collagen-induced arthritis in DBA/1J mice by recombinant TSG-6, a tumor necrosis factor/interleukin-1-inducible protein. *Arthritis Rheum* **43**, 2668-2677
12. Mindrescu, C., Dias, A. A., Olszewski, R. J., Klein, M. J., Reis, L. F., and Wisniewski, H. G. (2002) Reduced susceptibility to collagen-induced arthritis in DBA/1J mice expressing the TSG-6 transgene. *Arthritis Rheum* **46**, 2453-2464
13. Kahmann, J. D., O'Brien, R., Werner, J. M., Heinegard, D., Ladbury, J. E., Campbell, I. D., and Day, A. J. (2000) Localization and characterization of the hyaluronan-binding site on the link module from human TSG-6. *Structure* **8**, 763-774

14. Higman, V. A., Briggs, D. C., Mahoney, D. J., Blundell, C. D., Sattelle, B. M., Dyer, D. P., Green, D. E., DeAngelis, P. L., Almond, A., Milner, C. M., and Day, A. J. (2014) A refined model for the TSG-6 link module in complex with hyaluronan: use of defined oligosaccharides to probe structure and function. *J Biol Chem* **289**, 5619-5634
15. Higman, V. A., Blundell, C. D., Mahoney, D. J., Redfield, C., Noble, M. E., and Day, A. J. (2007) Plasticity of the TSG-6 HA-binding loop and mobility in the TSG-6-HA complex revealed by NMR and X-ray crystallography. *J Mol Biol* **371**, 669-684
16. Blundell, C. D., Almond, A., Mahoney, D. J., DeAngelis, P. L., Campbell, I. D., and Day, A. J. (2005) Towards a structure for a TSG-6.hyaluronan complex by modeling and NMR spectroscopy: insights into other members of the link module superfamily. *J Biol Chem* **280**, 18189-18201
17. Kohda, D., Morton, C. J., Parkar, A. A., Hatanaka, H., Inagaki, F. M., Campbell, I. D., and Day, A. J. (1996) Solution structure of the link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration. *Cell* **86**, 767-775
18. Blundell, C. D., Mahoney, D. J., Almond, A., DeAngelis, P. L., Kahmann, J. D., Teriete, P., Pickford, A. R., Campbell, I. D., and Day, A. J. (2003) The link module from ovulation- and inflammation-associated protein TSG-6 changes conformation on hyaluronan binding. *J Biol Chem* **278**, 49261-49270
19. Mahoney, D. J., Blundell, C. D., and Day, A. J. (2001) Mapping the hyaluronan-binding site on the link module from human tumor necrosis factor-stimulated gene-6 by site-directed mutagenesis. *J Biol Chem* **276**, 22764-22771
20. Parkar, A. A., and Day, A. J. (1997) Overlapping sites on the Link module of human TSG-6 mediate binding to hyaluronan and chondroitin-4-sulphate. *FEBS Lett* **410**, 413-417
21. Getting, S. J., Mahoney, D. J., Cao, T., Rugg, M. S., Fries, E., Milner, C. M., Perretti, M., and Day, A. J. (2002) The link module from human TSG-6 inhibits neutrophil migration in a hyaluronan- and inter-alpha -inhibitor-independent manner. *J Biol Chem* **277**, 51068-51076
22. Mahoney, D. J., Mulloy, B., Forster, M. J., Blundell, C. D., Fries, E., Milner, C. M., and Day, A. J. (2005) Characterization of the interaction between tumor necrosis factor-stimulated gene-6 and heparin: implications for the inhibition of plasmin in extracellular matrix microenvironments. *J Biol Chem* **280**, 27044-27055



23. Wisniewski, H. G., Snitkin, E. S., Mindrescu, C., Sweet, M. H., and Vilcek, J. (2005) TSG-6 protein binding to glycosaminoglycans: formation of stable complexes with hyaluronan and binding to chondroitin sulfates. *J Biol Chem* **280**, 14476-14484
24. Lee, T. H., Wisniewski, H. G., and Vilcek, J. (1992) A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member of the family of hyaluronate binding proteins, closely related to the adhesion receptor CD44. *The Journal of cell biology* **116**, 545-557
25. Kuznetsova, S. A., Day, A. J., Mahoney, D. J., Rugg, M. S., Mosher, D. F., and Roberts, D. D. (2005) The N-terminal module of thrombospondin-1 interacts with the link domain of TSG-6 and enhances its covalent association with the heavy chains of inter-alpha-trypsin inhibitor. *J Biol Chem* **280**, 30899-30908
26. Parkar, A. A., Kahmann, J. D., Howat, S. L., Bayliss, M. T., and Day, A. J. (1998) TSG-6 interacts with hyaluronan and aggrecan in a pH-dependent manner via a common functional element: implications for its regulation in inflamed cartilage. *FEBS Lett* **428**, 171-176
27. Salustri, A., Garlanda, C., Hirsch, E., De Acetis, M., Maccagno, A., Bottazzi, B., Doni, A., Bastone, A., Mantovani, G., Beck Peccoz, P., Salvatori, G., Mahoney, D. J., Day, A. J., Siracusa, G., Romani, L., and Mantovani, A. (2004) PTX3 plays a key role in the organization of the cumulus oophorus extracellular matrix and in in vivo fertilization. *Development* **131**, 1577-1586
28. Wisniewski, H. G., and Vilcek, J. (2004) Cytokine-induced gene expression at the crossroads of innate immunity, inflammation and fertility: TSG-6 and PTX3/TSG-14. *Cytokine & growth factor reviews* **15**, 129-146
29. Kuznetsova, S. A., Issa, P., Perruccio, E. M., Zeng, B., Sipes, J. M., Ward, Y., Seyfried, N. T., Fielder, H. L., Day, A. J., Wight, T. N., and Roberts, D. D. (2006) Versican-thrombospondin-1 binding in vitro and colocalization in microfibrils induced by inflammation on vascular smooth muscle cells. *J Cell Sci* **119**, 4499-4509
30. Wisniewski, H. G., Burgess, W. H., Oppenheim, J. D., and Vilcek, J. (1994) TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable complex with the serum protein inter-alpha-inhibitor. *Biochemistry* **33**, 7423-7429
31. Nentwich, H. A., Mustafa, Z., Rugg, M. S., Marsden, B. D., Cordell, M. R., Mahoney, D. J., Jenkins, S. C., Dowling, B., Fries, E., Milner, C. M., Loughlin, J., and Day, A. J. (2002) A novel allelic variant of the human TSG-6 gene encoding

- an amino acid difference in the CUB module. Chromosomal localization, frequency analysis, modeling, and expression. *J Biol Chem* **277**, 15354-15362
32. Fulop, C., Szanto, S., Mukhopadhyay, D., Bardos, T., Kamath, R. V., Rugg, M. S., Day, A. J., Salustri, A., Hascall, V. C., Glant, T. T., and Mikecz, K. (2003) Impaired cumulus mucification and female sterility in tumor necrosis factor-induced protein-6 deficient mice. *Development* **130**, 2253-2261
  33. Mukhopadhyay, D., Asari, A., Rugg, M. S., Day, A. J., and Fulop, C. (2004) Specificity of the tumor necrosis factor-induced protein 6-mediated heavy chain transfer from inter-alpha-trypsin inhibitor to hyaluronan: implications for the assembly of the cumulus extracellular matrix. *J Biol Chem* **279**, 11119-11128
  34. Sanggaard, K. W., Karring, H., Valnickova, Z., Thogersen, I. B., and Enghild, J. J. (2005) The TSG-6 and I alpha I interaction promotes a transesterification cleaving the protein-glycosaminoglycan-protein (PGP) cross-link. *J Biol Chem* **280**, 11936-11942
  35. Rugg, M. S., Willis, A. C., Mukhopadhyay, D., Hascall, V. C., Fries, E., Fulop, C., Milner, C. M., and Day, A. J. (2005) Characterization of complexes formed between TSG-6 and inter-alpha-inhibitor that act as intermediates in the covalent transfer of heavy chains onto hyaluronan. *J Biol Chem* **280**, 25674-25686
  36. Roberts, S., Evans, H., Menage, J., Urban, J. P., Bayliss, M. T., Eisenstein, S. M., Rugg, M. S., Milner, C. M., Griffin, S., and Day, A. J. (2005) TNFalpha-stimulated gene product (TSG-6) and its binding protein, IalphaI, in the human intervertebral disc: new molecules for the disc. *Eur Spine J* **14**, 36-42
  37. Enghild, J. J., Thogersen, I. B., Cheng, F., Fransson, L. A., Roepstorff, P., and Rahbek-Nielsen, H. (1999) Organization of the inter-alpha-inhibitor heavy chains on the chondroitin sulfate originating from Ser(10) of bikunin: posttranslational modification of IalphaI-derived bikunin. *Biochemistry* **38**, 11804-11813
  38. Colon, E., Shytuhina, A., Cowman, M. K., Band, P. A., Sanggaard, K. W., Enghild, J. J., and Wisniewski, H. G. (2009) Transfer of inter-alpha-inhibitor heavy chains to hyaluronan by surface-linked hyaluronan-TSG-6 complexes. *J Biol Chem* **284**, 2320-2331
  39. Leali, D., Inforzato, A., Ronca, R., Bianchi, R., Belleri, M., Coltrini, D., Di Salle, E., Sironi, M., Norata, G. D., Bottazzi, B., Garlanda, C., Day, A. J., and Presta, M. (2012) Long pentraxin 3/tumor necrosis factor-stimulated gene-6 interaction: a biological rheostat for fibroblast growth factor 2-mediated angiogenesis. *Arteriosclerosis, thrombosis, and vascular biology* **32**, 696-703

40. Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy* **73**, 1-16
41. Losonczi, J. A., Andrec, M., Fischer, M. W., and Prestegard, J. H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* **138**, 334-342
42. Prestegard, J. H., Bougault, C. M., and Kishore, A. I. (2004) Residual dipolar couplings in structure determination of biomolecules. *Chemical reviews* **104**, 3519-3540
43. Yao, L., Ying, J., and Bax, A. (2009) Improved accuracy of <sup>15</sup>N-<sup>1</sup>H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *J Biomol NMR* **43**, 161-170
44. de Vries, S. J., van Dijk, A. D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A. M. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726-733
45. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-1737
46. Farmer, B. T., 2nd, Constantine, K. L., Goldfarb, V., Friedrichs, M. S., Wittekind, M., Yanchunas, J., Jr., Robertson, J. G., and Mueller, L. (1996) Localizing the NADP<sup>+</sup> binding site on the MurB enzyme by NMR. *Nature structural biology* **3**, 995-997
47. Kveder, M., Pifat, G., Pecar, S., Schara, M., Ramos, P., and Esterbauer, H. (1997) Nitroxide reduction with ascorbic acid in spin labeled human plasma LDL and VLDL. *Chemistry and physics of lipids* **85**, 1-12
48. Battiste, J. L., and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355-5365
49. Jain, N. U., Venot, A., Umemoto, K., Leffler, H., and Prestegard, J. H. (2001) Distance mapping of protein-binding sites using spin-labeled oligosaccharide ligands. *Protein Sci* **10**, 2393-2400

50. Johnson, P. E., Brun, E., MacKenzie, L. F., Withers, S. G., and McIntosh, L. P. (1999) The cellulose-binding domains from *Cellulomonas fimi* beta-1, 4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations. *J Mol Biol* **287**, 609-625
51. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-223
52. Wang, X., Bansal, S., Jiang, M., and Prestegard, J. H. (2008) RDC-assisted modeling of symmetric protein homo-oligomers. *Protein Sci* **17**, 899-907
53. Baranova, N. S., Nileback, E., Haller, F. M., Briggs, D. C., Svedhem, S., Day, A. J., and Richter, R. P. (2011) The inflammation-associated protein TSG-6 cross-links hyaluronan via hyaluronan-induced TSG-6 oligomers. *J Biol Chem* **286**, 25675-25686
54. Day, A. J., and de la Motte, C. A. (2005) Hyaluronan cross-linking: a protective mechanism in inflammation? *Trends Immunol* **26**, 637-643
55. Pomin, V. H., Park, Y., Huang, R. R., Heiss, C., Sharp, J. S., Azadi, P., and Prestegard, J. H. (2012) Exploiting enzyme specificities in digestions of chondroitin sulfates A and C: Production of well-defined hexasaccharides. *Glycobiology* **22**, 826-838
56. Day, A. J., Aplin, R. T., and Willis, A. C. (1996) Overexpression, purification, and refolding of link module from human TSG-6 in *Escherichia coli*: effect of temperature, media, and mutagenesis on lysine misincorporation at arginine AGA codons. *Protein expression and purification* **8**, 1-16
57. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293
58. Kneller, T. D. G. a. D. G. SPARKY 3.
59. Roder, H., and Wuthrich, K. (1986) Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. *Proteins* **1**, 34-42
60. Liu, Y., and Prestegard, J. H. (2010) A device for the measurement of residual chemical shift anisotropy and residual dipolar coupling in soluble and membrane-associated proteins. *J Biomol NMR* **47**, 249-258

61. Bodenhausen, G., and Ruben, D. J. (1980) Natural Abundance N-15 Nmr by Enhanced Heteronuclear Spectroscopy. *Chem Phys Lett* **69**, 185-189
62. Grzesiek, S., and Bax, A. (1992) Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr. *Journal of the American Chemical Society* **114**, 6291-6293
63. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach for Assigning H-1 and C-13 Spectra of C-13-Enriched Proteins. *Journal of magnetic resonance* **88**, 425-431
64. Olejniczak, E. T., Xu, R. X., and Fesik, S. W. (1992) A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins. *Journal of biomolecular NMR* **2**, 655-659
65. Wittekind, M., and Mueller, L. (1993) Hncacb, a High-Sensitivity 3d Nmr Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. *J Magn Reson Ser B* **101**, 201-205
66. Grzesiek, S., and Bax, A. (1992) Improved 3d Triple-Resonance Nmr Techniques Applied to a 31-Kda Protein. *Journal of magnetic resonance* **96**, 432-440
67. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins. *Journal of magnetic resonance* **89**, 496-514
68. Muhandiram, D. R., and Kay, L. E. (1994) Gradient-Enhanced Triple-Resonance 3-Dimensional Nmr Experiments with Improved Sensitivity. *J Magn Reson Ser B* **103**, 203-216
69. Zuiderweg, E. R., and Fesik, S. W. (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* **28**, 2387-2391
70. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* **28**, 6150-6156

71. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302
72. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**, 283-298
73. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* **24**, 171-189
74. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**, 65-73
75. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **29**, 622-655
76. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612

## **CHAPTER 5**

### **CONCLUDING REMARKS**

GAGs and GAG binding protein interactions are important for the regulation of a wide range of physiological processes including cell-cell interactions and cell signaling (1-6). These interactions greatly impact numerous physiological events including development, cell growth, cell adhesion, inflammation, coagulation, tumorigenesis, and interactions with pathogens (7-22). Gaining insights into the mechanisms of these interactions is necessary to development of pharmaceutical interventions to treat disease. There are a variety of experimental tools to study these interactions, providing a wide range of information according to the data they produce for the analysis. For structural studies at the molecular level of the interactions between the GAG and GAG binding protein, advanced analytical approaches including SPR, Mass spectrometry, electron microscopy, SAXS, NMR spectroscopy and X-ray crystallography have been widely utilized. However, obtaining a sample of GAG and GAG binding protein for the study, producing the data, and analyzing it can be challenging. This requires several conditions: 1) pure samples of the GAGs and the proteins, 2) advanced analytical tools to collect and analyze data, and 3) expert skills to deal with data in order to produce high-resolution structures.

In this thesis, we have chosen NMR spectroscopy to study the interaction between a chondroitin 4-sulfated hexasaccharide and the Link module of human TSG-6. Isolating homogeneous GAGs is challenging due to the heterogeneity of GAGs in which sugars

can be modified with sulfation in either uronic acids or amino sugars, and epimerization of glucuronic acid to iduronic acid can occur in the case of HS and Hp. Because of given possible difficulties, the structure determination of GAG and GAG binding protein has been hindered. I have chosen to isolate homogeneous chondroitin sulfate hexasaccharides by enzymatic digestions for study in this thesis. Chondroitin sulfate-A (4-sulfation dominant) and -C (6-sulfation dominant) were treated with chondroitinase ABC, AC, C, and hyaluronidase to produce various oligosaccharides. Their sequences and structures were characterized using mass spectrometry and NMR spectroscopy. Based on this approach, 15 hexasaccharides were isolated and 4 of those were selected for binding assays with the Link module of TSG-6.

I then went on to study the structural aspects of the Link-TSG-6: $\Delta$ C444S complex using NMR applications including chemical shift perturbation, relaxation experiments, paramagnetic relaxation enhancement and residual dipolar coupling. Even though structures of the Link module of TSG-6 exist, there were reasons to pursue new solution state structures: (1) binding of CS to the Link module shows multiple binding modes that were not previously suspected, (2) RDC data fits poorly with respect to existing structures, and (3) CS induces dimerization of the Link module for which there is no structure. According to chemical shift perturbation for homogeneously well-defined CS hexasaccharides (C664S, C444S,  $\Delta$ C664S,  $\Delta$ C444S), it is likely that all isomers bind to the Link module at the same binding site for HA with similar affinities. The 4-sulfated hexasaccharide,  $\Delta$ C444S, was used as a ligand for structural studies because it showed the strongest affinity among CS isomers.

The solution structure of Link-TSG-6 in its  $\Delta$ C444S bound state was then



determined using heteronuclear resonance assignment methods and structural constraints from NOEs, RDCs and PRE. Combined with several NMR applications and other analytical protocols, it is confirmed that  $\Delta$ C444S binds near a previously identified HA binding groove, but in a different geometry which may better accommodate sulfate group interactions with protein residues. A preliminary set of dimer models was generated by computational docking. A final model was chosen based on scoring and total energy analysis. This model is compatible with experimental data and provides a unique and reliable dimer model for future studies.

The observations with the Link module suggest that the Link module itself can interact with chondroitin sulfate as well as form dimer conformation that is a novel characteristic of the Link module. I believe that it is an important discovery to understand novel mechanism of the interaction between sulfated-CS and Link module for stabilizing and reconstructing the ECM during inflammation and ovulation.

## **5.1 References**

1. Linhardt, R. J. a. T., T. (1997) Heparin analogs—development and applications in *In Carbohydrates as Drugs* (eds. Z. B. Witczak and K. A. Nieforth) Marcel Dekker, New York. pp 277-341
2. Lindahl, U., and Hook, M. (1978) Glycosaminoglycans and their binding to biological macromolecules. *Annual review of biochemistry* **47**, 385-417
3. Sasisekharan, R., Raman, R., and Prabhakar, V. (2006) Glycomics approach to structure-function relationships of glycosaminoglycans. *Annual review of biomedical engineering* **8**, 181-231
4. Imberty, A., Lortat-Jacob, H., and Perez, S. (2007) Structural view of glycosaminoglycan-protein interactions. *Carbohydr Res* **342**, 430-439
5. Raman, R., Sasisekharan, V., and Sasisekharan, R. (2005) Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chem Biol* **12**, 267-277

6. Gandhi, N. S., and Mancera, R. L. (2008) The structure of glycosaminoglycans and their interactions with proteins. *Chem Biol Drug Des* **72**, 455-482
7. Perris, R., Perissinotto, D., Pettway, Z., Bronner-Fraser, M., Morgelin, M., and Kimata, K. (1996) Inhibitory effects of PG-H/aggrecan and PG-M/versican on avian neural crest cell migration. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **10**, 293-301
8. Pettway, Z., Domowicz, M., Schwartz, N. B., and Bronner-Fraser, M. (1996) Age-dependent inhibition of neural crest migration by the notochord correlates with alterations in the S103L chondroitin sulfate proteoglycan. *Exp Cell Res* **225**, 195-206
9. Inatani, M., Haruta, M., Honjo, M., Oohira, A., Kido, N., Takahashi, M., Honda, Y., and Tanihara, H. (2001) Upregulated expression of N-syndecan, a transmembrane heparan sulfate proteoglycan, in differentiated neural stem cells. *Brain Res* **920**, 217-221
10. Tully, S. E., Mabon, R., Gama, C. I., Tsai, S. M., Liu, X., and Hsieh-Wilson, L. C. (2004) A chondroitin sulfate small molecule that stimulates neuronal growth. *J Am Chem Soc* **126**, 7736-7737
11. Beaulieu, J. F., Vachon, P. H., and Chartrand, S. (1991) Immunolocalization of extracellular matrix components during organogenesis in the human small intestine. *Anat Embryol (Berl)* **183**, 363-369
12. Cohn, R. H., Cassiman, J. J., and Bernfield, M. R. (1976) Relationship of transformation, cell density, and growth control to the cellular distribution of newly synthesized glycosaminoglycan. *The Journal of cell biology* **71**, 280-294
13. de Aguiar C, L.-S. B., Alvarez-Silva M, Trentin AG. (2005) Glycosaminoglycans modulate C6 glioma cell adhesion to extracellular matrix components and alter cell proliferation and cell migration. . *Bmc Cell Biology* **6**
14. Handel, T. M., Johnson, Z., Crown, S. E., Lau, E. K., and Proudfoot, A. E. (2005) Regulation of protein function by glycosaminoglycans--as exemplified by chemokines. *Annual review of biochemistry* **74**, 385-410
15. Kaplan, C. D., O'Neill, S. K., Koreny, T., Czipri, M., and Finnegan, A. (2002) Development of inflammation in proteoglycan-induced arthritis is dependent on Fc gamma R regulation of the cytokine/chemokine environment. *J Immunol* **169**, 5851-5859

16. Doodes, P. D., Cao, Y., Hamel, K. M., Wang, Y., Rodeghero, R. L., Kobezda, T., and Finnegan, A. (2009) CCR5 is involved in resolution of inflammation in proteoglycan-induced arthritis. *Arthritis Rheum* **60**, 2945-2953
17. de Mattos, D. A., Stelling, M. P., Tovar, A. M., and Mourao, P. A. (2008) Heparan sulfates from arteries and veins differ in their antithrombin-mediated anticoagulant activity. *Journal of thrombosis and haemostasis : JTH* **6**, 1987-1990
18. He, L., Giri, T. K., Vicente, C. P., and Tollefsen, D. M. (2008) Vascular dermatan sulfate regulates the antithrombotic activity of heparin cofactor II. *Blood* **111**, 4118-4125
19. Cattaruzza, S., and Perris, R. (2005) Proteoglycan control of cell movement during wound healing and cancer spreading. *Matrix biology : journal of the International Society for Matrix Biology* **24**, 400-417
20. Muramatsu, T., and Muramatsu, H. (2008) Glycosaminoglycan-binding cytokines as tumor markers. *Proteomics* **8**, 3350-3359
21. Ruiz-Arguello, M. B., Smith, V. P., Campanella, G. S., Baleux, F., Arenzana-Seisdedos, F., Luster, A. D., and Alcami, A. (2008) An ectromelia virus protein that interacts with chemokines through their glycosaminoglycan binding domain. *J Virol* **82**, 917-926
22. van Putten, J. P., Hayes, S. F., and Duensing, T. D. (1997) Natural proteoglycan receptor analogs determine the dynamics of Opa adhesin-mediated gonococcal infection of Chang epithelial cells. *Infection and immunity* **65**, 5028-5034

## **APPENDIX**

### **A. EXPLOITING ENZYME SPECIFICITIES IN DIGESTIONS OF CHONDROITIN SULFATES A AND C: PRODUCTION OF WELL- DEFINED HEXASACCHARIDES<sup>2</sup>**

---

<sup>2</sup> 2012, *Glycobiology*, 22, 826-838, Vitor H. Pomin, Younghee Park, Rongrong Huang, Christian Heiss, Joshua S. Sharp, Parastoo Azadi, and James H. Prestegard  
Reprinted here with permission from publisher, 07/14/15

## **A.1 Abstract**

Interactions between proteins and glycosaminoglycans (GAGs) of the extracellular matrix are important to regulation of cellular processes including growth, differentiation, and migration. Understanding these processes can benefit greatly from the study of protein-GAG interactions using GAG oligosaccharides of well-defined structure. Materials for such studies have, however, been difficult to obtain because of challenges in synthetic approaches and the extreme structural heterogeneity in GAG polymers. Here it is demonstrated that diversity in structures of oligosaccharides derived by limited enzymatic digestion of materials from natural sources can be greatly curtailed by proper selection of combinations of source materials and digestive enzymes, a process aided by an improved understanding of the specificities of certain commercial preparations of hydrolases and lyases. Separation of well-defined oligosaccharides can then be accomplished by size-exclusion chromatography followed by strong anion-exchange chromatography. We focus here on two types of chondroitin sulfate (CS) as starting material (CS-A, and CS-C) and the use of three digestive enzymes with varying specificities (testicular hyaluronidase, and bacterial chondroitinases ABC, and C). Analysis using NMR and mass spectrometry focuses on isolated CS disaccharides and hexasaccharides. In all, 15 CS hexasaccharides have been isolated and characterized. These serve as useful contributions to growing libraries of well-defined GAG oligosaccharides that can be used in further biophysical assays.

*Keywords:* Chondroitin sulfate; chondroitinase; enzymatic specificity; hyaluronidase; oligosaccharides.

## **A.2 Introduction**

Complexes involving glycosaminoglycans (GAGs) of the extracellular matrix and various proteins are important in regulating cell–cell interactions and cell-signaling events (Lindahl and Hook 1978; Handel et al. 2005; Raman et al. 2005; Sasisekharan et al. 2006; Imberty et al. 2007; Gandhi and Mancera 2008). Those interactions have innumerable physiological consequences including organogenesis/growth control (Cohn et al. 1976; Beaulieu et al. 1991), cell adhesion (de Aguiar et al. 2005), coagulation/thrombosis (De Mattos et al. 2008; He et al. 2008), regeneration/wound healing (Gorio et al. 1997; Cattaruzza and Perris 2005), tumorigenesis/metastasis (Cattaruzza and Perris 2005; Muramatsu and Muramatsu 2008), morphogenesis (Thesleff et al. 1988; Domowicz et al. 2000), inflammation (Kaplan et al. 2002; Handel et al. 2005; Koninger et al. 2006; Doodes et al. 2009) and neural development/regeneration (Perris et al. 1996; Pettway et al. 1996; Inatani et al. 2001; Tully et al. 2004). They also play an important role in infection by pathogens (vanPutten et al. 1997) and in mediation of prion internalization (Warner et al. 2002; Ben-Zaken et al. 2003; Horonchik et al. 2005). Involvement of GAGs in such a variety of processes is perhaps not surprising given the diversity of structures and their extracellular location (Handel et al. 2005; Sasisekharan et al. 2006; Gandhi and Mancera 2008).

There are many different types of GAGs: heparan sulfate, heparin (Hp), chondroitin sulfate (CS), dermatan sulfate, keratan sulfate and hyaluronic acid (HA), all of them physiologically active and biomedically important. They differ in terms of composing units (uronic acids, or galactose, and hexosamines). These units occur in disaccharides of one uronic acid/galactose and one hexosamine linked to make a linear

polymer. In all except HA extensive sulfation adds to their highly anionic character and structural diversity.

This structural diversity in GAGs is essential to the specificity and variety of interactions with proteins mediating the processes listed above. For example, a specific Hp pentasaccharide with a rare 3-sulfation is clinically exploited due to its high affinity for antithrombin and consequent action in controlling the clotting process (Jin et al. 1997; Richard et al. 2009). Specific sequences of CS having a mix of non- and 4-sulfated N-acetylgalactosamine (GalNAc) residues are believed to mediate the binding of erythrocytes infected by the pathogen responsible for placental malaria (Achur et al. 2008; Singh et al. 2008). Understanding these GAG–protein interactions can clearly be of importance to human health (Kaplan et al. 2002; Handel et al. 2005; Horonchik et al. 2005; Raman et al. 2005; Imberty et al. 2007; Gandhi and Mancera 2008; Muramatsu and Muramatsu 2008; Singh et al. 2008; Richard et al. 2009). However, studies leading to these understandings require the availability of GAG oligosaccharides with well-defined structure (Jin et al. 1997; Muthusamy et al. 2004; Singh et al. 2008; Richard et al. 2009). Studies presented here on enzymatically digested products from natural sources of CS polysaccharides illustrate a means of producing these well-defined oligosaccharides. CS is the most abundant GAG in the body with numerous biological functions (Sugahara et al. 2003). We have chosen to initially explore the production of CS oligosaccharides, since CS is the most homogeneous- sulfated GAG type composed of just alternating GalNAc and glucuronic acid (GlcA) units that are differentially sulfated. This facilitates separation and structure determination.

Among many combinations of enzymes and CS substrates available for study (Ernst et al. 1995), here we choose three commonly used commercially available CS degrading enzyme preparations, the CS lyases, chondroitinase ABC from *Proteus vulgaris* and chondroitinase C from *Flavobacterium heparinum* and the hydrolase, testicular hyaluronidase from sheep testes. We choose two CS standards for investigation, bovine tracheal CS-A (btCS-A, mostly 4-sulfated) and shark cartilage CS-C (scCS-C, predominantly 6-sulfated). Neither enzyme nor substrate preparations are particularly pure, but they are readily available in quantities suitable for preparative work, justifying a systematic investigation of the digestive properties of these systems. A comparison of yields and structures of the major disaccharide and hexasaccharide products from the six digestion combinations is presented. After proper isolation by a combination of size-exclusion chromatography (SEC) and strong anion exchange (SAX) chromatography as successfully performed in previous works (Deepa et al. 2007; Pothacharoen et al. 2007), all oligosaccharide species were characterized by a combination of nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). The number of distinct species produced for analysis is strikingly small and point to specificities in the enzyme preparations that can be exploited in devising approaches to production of well-defined oligosaccharides.



**Table A1:** Structure and respective abbreviations used for disaccharides and hexasaccharides obtained from six different digestions.

	Structural description <sup>a</sup>	Abbreviation <sup>b</sup>
Disaccharides	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc}$	$\Delta\text{C0S}$
	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc4S}$	$\Delta\text{C4S}$
	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc6S}$	$\Delta\text{C6S}$
	$\Delta\text{UA2S}(\beta 1 \rightarrow 3)\text{GalNAc6S}$	$\Delta\text{C2,6S}$
Hexasaccharides	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S-ol}$	$\Delta\text{C6;6;4S-ol}$
	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S-ol}$	$\Delta\text{C6;4;4S-ol}$
	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S-ol}$	$\Delta\text{C4;6;4S-ol}$
	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S-ol}$	$\Delta\text{C4;4;4S-ol}$
	$\Delta\text{UA2S}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}$	$\Delta\text{C2,6;0;6S}$
	$\Delta\text{UA2S}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S}$	$\Delta\text{C2,6;0;4S}$
	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S-ol}$	$\Delta\text{C6;6;6S-ol}$
	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S-ol}$	$\Delta\text{C4;6;6S-ol}$
	$\Delta\text{UA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA2S}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S-ol}$	$\Delta\text{C4;2,6;6S-ol}$
	$\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc-ol}$	$\text{C6;6;0S-ol}$
	$\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc-ol}$	$\text{C6;4;0S-ol}$
	$\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S-ol}$	$\text{C6;4;4S-ol}$
	$\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S-ol}$	$\text{C4;4;4S-ol}$
	$\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S-ol}$	$\text{C6;6;4S-ol}$
	$\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA2S}(\beta 1 \rightarrow 3)\text{GalNAc6S-ol}$	$\text{C6;6;2,6S-ol}$
	$\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA2S}(\beta 1 \rightarrow 3)\text{GalNAc4S6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc6S-ol}$	$\text{C6;2,6;6S-ol}$
	$\text{GlcA2S}(\beta 1 \rightarrow 3)\text{GalNAc6S}(\beta 1 \rightarrow 4)\text{GlcA}(\beta 1 \rightarrow 3)\text{GalNAc4S}(\beta 1 \rightarrow 4)\text{GlcA2S}(\beta 1 \rightarrow 3)\text{GalNAc6S-ol}$	$\text{C2,6;4;2,6S-ol}$

<sup>a</sup> Nomenclature:  $\Delta\text{UA} = \Delta^{4,5}$ unsaturated uronic acid; GalNAc = N-acetyl galactosamine; GlcA = glucuronic acid; S = sulfation group, where the numbers before “S” represent the ring position; -ol stands for reduced sugars (open rings at the reducing-ends).

<sup>b</sup> In the structural codes, the comma (,) was used to separate hexoses, whereas semicolon (;) was used to separate disaccharide units. The digits 0, 4, and 6 denote respective positions of sulfation in the GalNAc units, whereas the 2 denotes 2-sulfation at uronic acid units.

### **A.3 Results**

Disaccharide analysis of extensively digested btCS-A and scCS-C In order to set a standard for further analysis of disaccharides released under more limited digestion conditions, a nearly complete digestion of btCS-A and scCS-C was attempted (Supplementary data, Figure S1A and B). After a week-long digestion of btCS-A polymers (8–50 kDa, Supplementary data, Figure S2) during which enzymes were replaced periodically, high-pressure liquid chromatography (HPLC) analysis showed disaccharide content to be 8%  $\Delta C0S$ , 42%  $\Delta C6S$  and 50%  $\Delta C4S$  (Supplementary data, Figure S1C and Tables I and II). This is in reasonable agreement with previous works (Mucci et al. 2000; Muthusamy et al. 2004; 10%  $\Delta C0S$ , 40%  $\Delta C6S$  and 50%  $\Delta C4S$ ). Small deviations depending on sources are not unexpected.

HPLC disaccharide analysis of similarly digested scCS-C polymers (10–50 kDa, Supplementary data, Figure S2) showed disaccharide content to be 3%  $\Delta C0S$ , 49%  $\Delta C6S$ , 26%  $\Delta C4S$  and 22%  $\Delta C2,6S$  (Supplementary data, Figure S1D and Table A2). Again this is in reasonable agreement with the literature (2%  $\Delta C0S$ , 49%  $\Delta C6S$ , 29%  $\Delta C4S$ , 17%  $\Delta C2,6S$  and 3%  $\Delta C4,6S$ ; Sorrell et al. 1993; Mucci et al. 2000). Data confirm the heterogeneity of both CS standards. However, quantities of recovered mono- (>5%) and disaccharides (80–85%) show the digestions to be only 85–90% complete (Table A2), suggesting that both CS polymers may contain sequences resistant to even very extensive digestion conditions.

**Table A2:** Yields of the disaccharides and the major hexasaccharides obtained from different digestion types within different time courses.

Digestion		Disaccharides		Hexasaccharides	
Type	Time	Absolute yield <sup>a</sup>	Structure and relative yield <sup>b</sup>	Absolute yield <sup>a</sup>	Structure and relative yield <sup>b</sup>
near-complete digested btCS-A	7 days	~80 %	$\Delta$ C0S (8 %); $\Delta$ C4S (50 %); $\Delta$ C6S (42 %)	< 1%	Not determined
near-complete digested scCS-C	7 days	~85 %	$\Delta$ C0S (3 %); $\Delta$ C4S (26 %); $\Delta$ C6S (49 %); $\Delta$ C2,6S (22 %)	< 1%	Not determined
ABC lyase + btCS-A	2 h and 30 min	~ 70 %	$\Delta$ C0S (7 %); $\Delta$ C4S (60 %); $\Delta$ C6S (33 %)	~ 3 %	$\Delta$ C6;6;4S-ol (20 %);  $\Delta$ C6;4;4S-ol or $\Delta$ C4;6;4S-ol (35 %);  $\Delta$ C4;4;4S-ol (46 %)
ABC lyase + scCS-C	2 h and 30 min	~ 70 %	$\Delta$ C0S (6 %); $\Delta$ C4S (37 %); $\Delta$ C6S (41 %); $\Delta$ C2,6S (16 %)	~ 3 %	$\Delta$ C2,6;0;4S (34 %);  $\Delta$ C2,6;0;6S (33 %)
C lyase + btCS-A	2 days	~ 8 %	$\Delta$ C0S (12 %); $\Delta$ C4S (43 %); $\Delta$ C6S (45 %)	~ 10 %	$\Delta$ C6;6;6S-ol (8 %);  $\Delta$ C4;6;6S-ol (20 %);  $\Delta$ C4;4;4S-ol (57 %)
C lyase + scCS-C	2 days	~ 8 %	$\Delta$ C0S (14 %); $\Delta$ C4S (15 %); $\Delta$ C6S (68 %);	~ 10 %	$\Delta$ C6;6;6S-ol (22 %);  $\Delta$ C4;2,6;6S-ol (39 %)

			$\Delta$ C2,6S (3 %)		
hyaluronidase + btCS-A	2 days	< 1 %	Not determined	~ 15 %	C6;6;0S-ol (10 %);  C6;4;0S-ol (8 %);  C6;4;4S-ol (30 %);  C4;4;4S-ol (19 %);  C6;6;4S-ol (32 %)
hyaluronidase + scCS-C	2 days	< 1 %	Not determined	~ 10 %	C6;6;4S-ol (23 %);  C6;4;4S-ol (18 %);  C6;6;2,6S-ol (20 %);  C6;2,6;6S-ol (19 %);  C2,6;4;2,6S-ol (7 %)

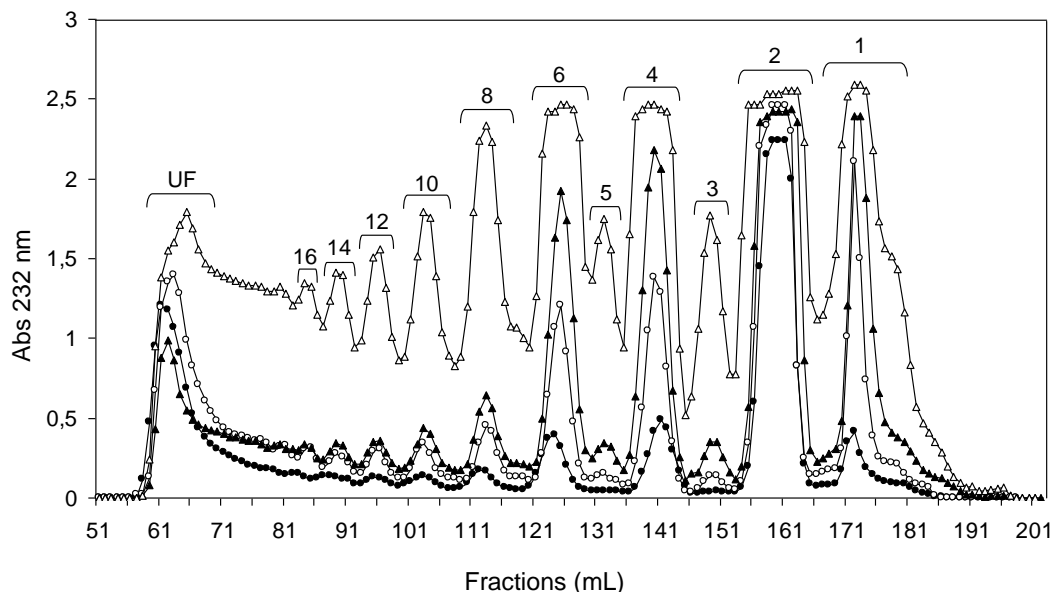
<sup>a</sup> The absolute yield was estimated as the percentage of weight (mg of sample) recovered from the peaks 2 (disaccharides), and 6 (hexasaccharides) from the SEC (Bio-Gel P-10 column) of each digestion type. The values are relative to ~ 75 mg of loaded material into the column.

<sup>b</sup> The relative yield was determined as percentage of weight (mg of material) recovered as pure isomers fractionated by SAX-HPLC chromatography of the heterogeneous mixture of disaccharides or hexasaccharides from each digestion type. The values are compared to ~ 5 mg of material loaded into the column.

### **A.3.1 Limited chondroitinase ABC digestion of btCS-A**

SEC profiles of btCS-A digestion products from shorter periods of ABC lyase treatment are shown in Figure A1. MS has been used to determine the size of oligosaccharides eluted in representative peaks (Supplementary data, Figure S3), and the degree of polymerization is indicated above each peak in Figure A1. Disaccharides are the major products for short periods of digestion, consistent with the previously suggested presence of substantial exolytic cleavage activity (Hardingham et al. 1994). HPLC analysis of disaccharide content of samples digested for 2.5 h (Figure A1, open triangles) showed composition to be: 7%  $\Delta$ C0S, 33%  $\Delta$ C6S and 60%  $\Delta$ C4S

(Supplementary data, Figure S4 and Table A2). In comparison with the extensively digested sample (Supplementary data, Figure S1C), more  $\Delta$ C4S disaccharide is clearly generated suggesting a preference for cleavage at a 4-sulfated GalNAc.



**Figure A1:** Size fractionation on Bio-Gel P-10 column of the products from btCS-A digested with a commercial preparation of chondroitinase ABC from *P. vulgaris*. Data from different digestion times are shown: 10 min (-●-), 30 min (-○-), 1 h (-▲-), and 2.5 h (-Δ-). At the top of each peak the degree of polymerization is indicated. UF stands for unfractionated material.

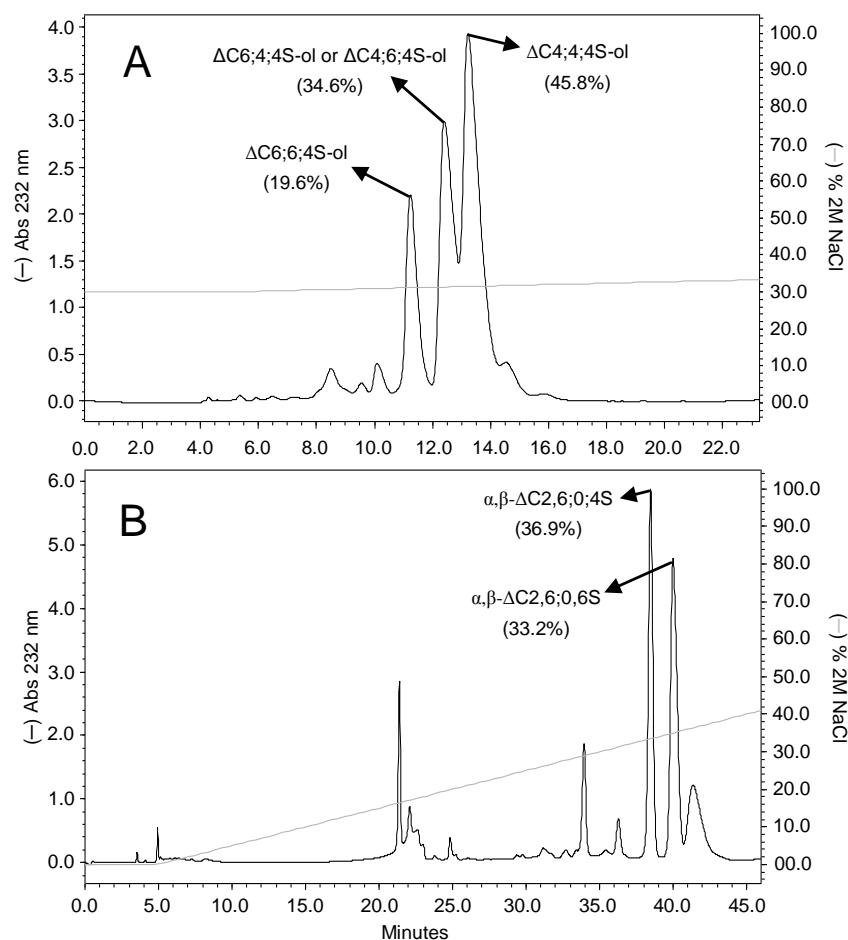
Although dominant production of disaccharides in early periods of digestion is in full agreement with the expected high level of exolytic activity (Hardingham et al. 1994), tetrasaccharides and hexasaccharides are significant at the initial periods as well (Figure A1, filled and open circles) and become progressively more abundant with longer digestions (Figure A1, filled and open triangles). Higher order oligosaccharides (longer than 6-mers) become gradually measurable as well. The commercial preparation of ABC lyase is in fact known to be a mixture of two enzymes suspected to differ in endolytic and exolytic activities, lyases I and II (Hamai et al. 1997; Zhang et al. 2009). Therefore, it is reasonable to consider the production profile as a result of their combined and competing

activities (see Materials and methods). If the exolytic activity dominated, the high-order oligosaccharides could be simply end-products of multiple exolytic cleavages of chains with different lengths in the starting material. To test this possibility, a digestion of a size-selected fraction of btCS-A, 39 kDa or 164 residues for short (10 min) and long (2.5 h) periods, was performed (Supplementary data, Figure S5A). Although large amounts of disaccharides were still observed in both time courses, significant amounts of tetrasaccharides and hexasaccharides were seen (24% each for the 10-min digestion) and the amounts of larger oligosaccharides were small. If only exolytic digestion had occurred in the 10 min digestion and only 38% disaccharides had been produced, the average length of the remaining fragments should have averaged 24.1 kDa or 100 residues in length. Hence, there is significant endolytic activity in the mixture (Jandik et al. 1994) forming mostly tetrasaccharides and hexasaccharides (Hardingham 1994). As the time of digestion proceeds (4.5h in Supplementary data, Figure S5B), the amount of high-order oligosaccharides progressively decreases, likely due to continued exolytic action, until massive disaccharide amounts are formed as final products (Supplementary data, Figure S1A).

There is a general consensus that the ABC lyase exolytic activity proceeds from the non-reducing end (Hardingham et al. 1994). This directionality was confirmed through the MS characterization of digestion products from a size-selected CS fraction modified at its reducing end (Supplementary data, Figure S6). Hence, the existence of sites less prone to exolytic activity may be reflected in the non-reducing end structure of some oligosaccharide products. The hexasaccharide fraction from limited ABC lyase digestion of unfractionated btCS-A was, therefore, further analyzed by separation on a

SAX column followed by NMR and MS analysis. Prior to SAX-HPLC, the mixture of hexasaccharides (Figure A1, peak 6, open triangles) was reduced via treatment with NaBH<sub>4</sub> to avoid possible complications from  $\alpha$ -/ $\beta$ -anomeric mutarotation. This reaction was only used to facilitate structural characterization and can be skipped when oligosaccharide production is the objective. Only three major peaks appeared in the SAX chromatogram (Figure A2A). Among the three fractions, only the two leading and trailing peaks gave <sup>13</sup>C-gradient heteronuclear single-quantum coherence (gHSQC) spectra sufficiently free of contamination from species eluting at either side to allow structure determination. The first peak was characterized as  $\Delta$ C6;6;4S-ol (Table A1). Its <sup>13</sup>C -gHSQC spectrum showed two <sup>1</sup>H/<sup>13</sup>C-signals characteristic of 6-sulfation and one cross-peak typical of 4-sulfation (Supplementary data, Figure S7A and Supplementary data, Table SI). This last 4-sulfate-related signal is connected in the correlation spectroscopy (COSY) spectrum to a resonance of a far upfield methylene group ( $\delta_H$  = 3.69 ppm), as opposed to an anomeric resonance (Supplementary data, Figure S7B). This methylene derives from the open sugar ring coming from the reduction reaction, thus proving that the 4-sulfated GalNAc unit is located at the reducing terminus of the hexasaccharide. The last hexasaccharide from the SAX-HPLC was identified as the entirely 4-sulfated hexasaccharide ( $\Delta$ C4;4;4S-ol, Table A1) based on intense <sup>1</sup>H4-<sup>13</sup>C4 downfield cross-peaks characteristic of 4-sulfated GalNAc residues (Supplementary data, Figure S8A and Supplementary data, Table S1). The central elution fraction, which could not be completely characterized by <sup>13</sup>C-gHSQC, was further subjected to analysis using <sup>15</sup>N-gHSQC. This has previously proven to be very diagnostic for sulfation types in CS oligosaccharides (Pomin et al. 2010). Besides the three cross-peaks showing 4-sulfation

from the contaminating entirely 4-sulfated hexasaccharides (Supplementary data, Figure S8B), an additional intense cross-peak typical of a non-reducing end or middle 6-sulfated GalNAc residue is observed. This suggests that the middle SAX-HPLC fraction (Figure A2A) is either  $\Delta C6;4;4S\text{-ol}$  or  $\Delta C4;6;4S\text{-ol}$ .



**Figure A2:** SAX-HPLC fractionation of unsaturated hexasaccharides from digestion of (A) btCS-A, and (B) scCS-C with a commercial preparation of chondroitinase ABC from *P. vulgaris*. Mixtures of hexasaccharides (reduced, and unreduced for btCS-A, and scCS-C, respectively) were obtained from peak 6 of their respective Bio-Gel P-10 chromatograms of 2.5 h digestions. The two fractionated isomers as characterized by NMR and MS spectroscopy were (A)  $\Delta C6;6;4S\text{-ol}$ , and  $\Delta C4;4;4S\text{-ol}$  for btCS-A, and (B)  $\alpha, \beta\text{-}\Delta C2,6;0;6S$  and  $\alpha, \beta\text{-}\Delta C2,6;0;4S$  for scCS-C. (A) The middle peak could be either  $\Delta C6;4;4S\text{-ol}$  or  $\Delta C4;6;4S\text{-ol}$ . The percentage of material in each peak is indicated in parentheses. The NaCl gradient is shown with the continuous light grey line.



The appearance of hexasaccharides with 6-sulfation at central sites and the non-reducing terminus, and 4-sulfation at the reducing terminus, is expected based on a lower preference for exolytic cleavage at 6-sulfated sites. Exolytic cleavage beginning at the non-reducing end would have been slow at the 6-sulfated sites allowing more time for endolytic cleavage at an upstream 4-sulfated site. The high yield of a  $\Delta C4;4;4S$ -ol oligosaccharide (Table A2) is likely a simple consequence of the higher percentage of 4-sulfation in the starting material.

### **A.3.2 Limited chondroitinase ABC digestion of scCS-C**

The oligosaccharide distribution using chondroitinase ABC to digest the atypical scCS-C substrate over a 2.5-h digestion period (Supplementary data, Figure S9) is similar to that seen for btCS-A at 1 h (Figure A1, filled triangles). The slower progression of the digestion is consistent with a preference of the ABC lyase for 4-sulfation sites and the smaller percentage of 4-sulfation in the scCS-C substrate. scCS-C is 26% 4-sulfated as opposed to 50% in btCS-A (Supplementary data, Figures S1C vs D and Table A2).

Analyses of the disaccharide fractions were distinctive in two respects. First, a new HPLC peak was detected (Supplementary data, Figure S10A). Its MS spectrum (Supplementary data, Figure S10B) indicated an MW of 539 Da, typical of a disulfated unsaturated CS-derived disaccharide. The  $^{13}C$ -gHSQC spectrum confirmed the presence of 6-sulfation on GalNAc, with additional 2-sulfation on GlcA (Supplementary data, Figure S10C).

Second, even though native scCS-C possesses only 26% 4-sulfation (Supplementary data, Figure S1D), larger amounts (37%) of 4-sulfated disaccharides were produced (Supplementary data, Figure S10A and Table A2). This is consistent with

a preference for cleavage in 4-sulfated regions of CS molecules by the exolytic component of ABC lyase preparation.

The presence of hexasaccharides, although comprising <5% of the digested material after 2.5 h digestion (Table A2), is again consistent with significant endolytic activity of the commercial preparation of the ABC lyase from *P. vulgaris*. Fractionation by SAX-HPLC of the unreduced hexasaccharides revealed two well-separated major peaks (Figure A2B); both were further characterized by  $^{13}\text{C}$ -gHSQC spectra (Supplementary data, Figure S11A and B). Positions of sulfate-related  $^1\text{H}/^{13}\text{C}$  crosspeaks suggest that the first fraction to be  $\Delta\text{C}2,6;0;4\text{S}$  (Table A1) and the second fraction to be  $\Delta\text{C}2,6;0;6\text{S}$  (Table A1).

The high representation of 2-sulfated GlcAs and non-sulfated GalNAcs in the hexasaccharide fraction is in sharp contrast to the lower amounts of disaccharides containing these sulfation patterns under near-complete digestion conditions (22 and 3%, respectively; Table A2). This sulfation pattern,  $\Delta\text{C}2,6;0$ , may slow the exolytic action of the 4-sulfation-preferring ABC lyase allowing endolytic activity to produce hexasaccharides with an unexpectedly small number of isomers. 4-sulfation at the reducing end is more prevalent as expected based on preference for cleavage at a 4-sulfated site. The appearance of a hexasaccharide with 6-sulfation at this position likely reflects the abundance of these sites in the starting material.

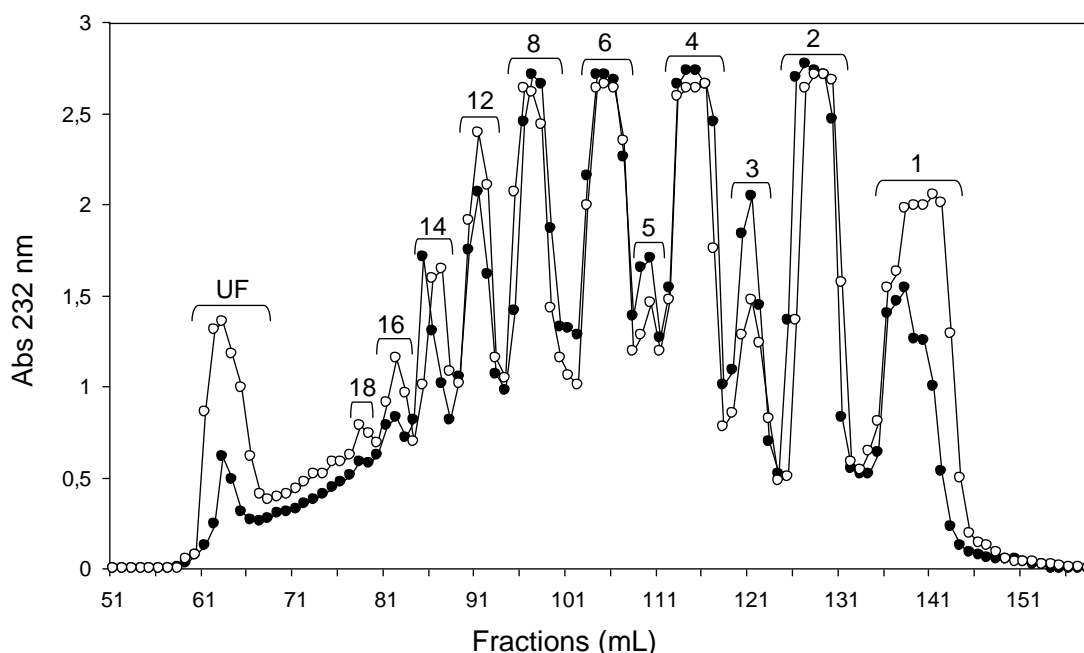
### **A.3.3 Limited chondroitinase C digestion of scCS-C**

The commercial preparation of chondroitinase C is expected to display enhanced activity toward CS-C substrates, but both the chondroitinase C preparation used here and the related chondroitinase AC preparation are known to display activity toward CS-A

products as well (Michelacci and Dietrich 1976; Aguiar et al. 2003). The AC preparation is well characterized and activities toward 4-sulfated and 6-sulfated regions reside in a single enzyme (Gu et al. 1995). The mixed activities in the chondroitinase C preparation reflect a similar activity profile, but one that may result from the presence of multiple enzyme components (see Materials and methods). The activity of chondroitinase C was observed to be lower than that of chondroitinase ABC and hence longer digestion times were used (48 h in Figure A3 vs 2.5 h in Figure A1, open triangles). Only the results after 2-day digestion are presented in Figure A3, and this plot is used to compare the size distribution results on digesting the preferred 6-sulfation-rich substrate, scCS-C, to the less preferred 4-sulfation-rich substrate, btCS-A. The distribution of oligosaccharide sizes appears similar for both substrates as confirmed by the quantitation of amounts of disaccharides and hexasaccharides presented in Table A2. This suggests a substantial amount of 4-sulfation-dependent activity. The higher amounts of hexasaccharide and other oligosaccharides when compared with disaccharides for both digestions suggest a dominant endolytic activity for this enzyme preparation.

Disaccharide fractions after 2-day digestion of scCS-C showed 68% of the structures to have a single 6-type sulfation (Supplementary data, Figure S12A and Table A2), whereas the starting material has 49% (Supplementary data, Figure S1D). Some 6-sulfation preference therefore exists. The specificity of the C lyase preparation as seen in scCS-C digestions can be better assessed by the examination of hexasaccharides isolated by SAX-HPLC combined with analysis by NMR and MS (Figure A4A and Supplementary data, Figures S13 and S14A and B). The large amount of hexasaccharide bearing 6-sulfated units at reducing ends, such as  $\Delta C6;6;6S$ -ol and  $\Delta C4;2,6;6S$ -ol (Table

A1) would support specificity for cleavage at a 6-sulfated site, but may also be explained as a consequence of the preponderance of 6-sulfation in the starting material. A comparison with a C lyase digestion of btCS-A, as presented below, helps to differentiate these possibilities. The different numbers of sulfates in oligosaccharide products from scCS-C when 2-sulfation of the GlcA occurs along with sulfation of the GalNAc are distinctive and facilitate the isolation of homogeneous hexasaccharides from this digestion type.

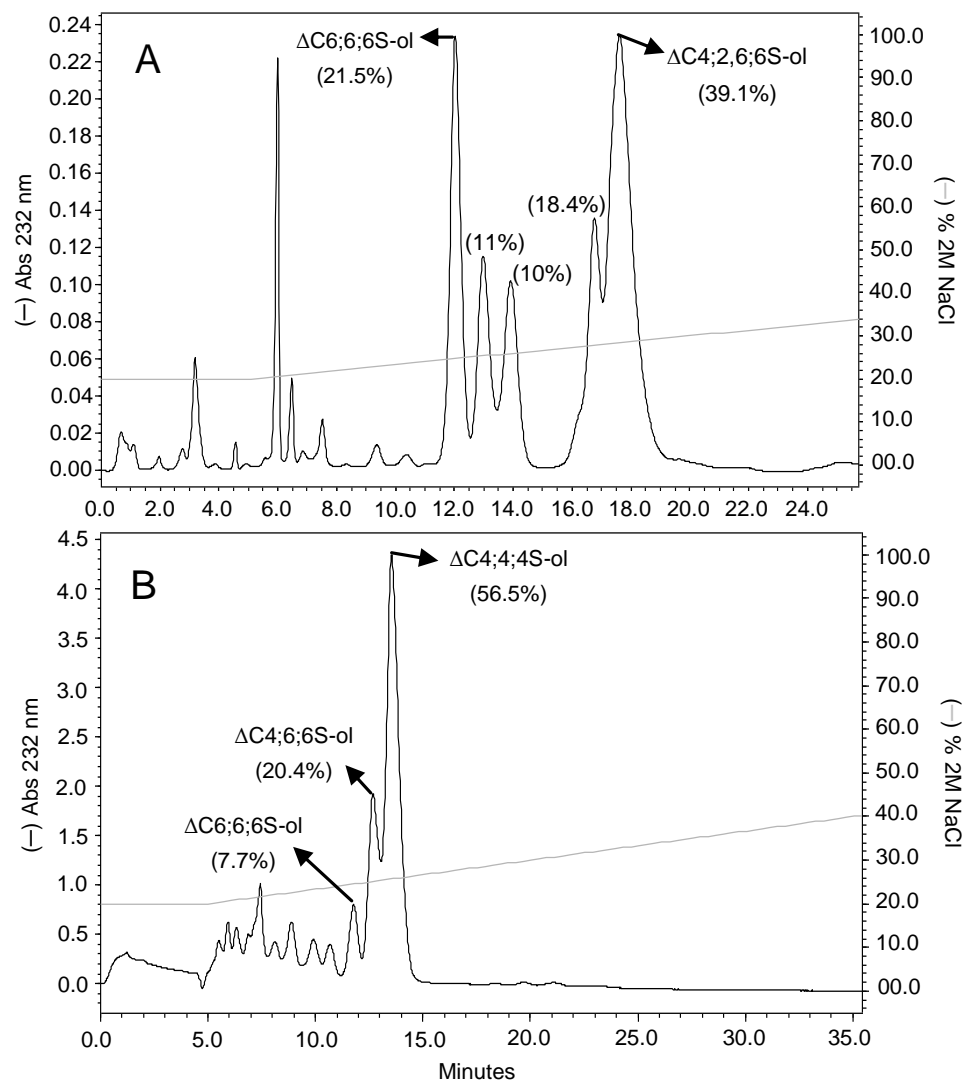


**Figure A3:** Size fractionation on Bio-Gel P-10 column of the products from 2-day digestions of scCS-C (-o-), and btCS-A (-●-) with a commercial preparation of chondroitinase C from *F. heparinum*. The degree of polymerization is given at the top of each peak. UF stands for unfractionated material.

#### **A.3.4 Limited chondroitinase C digestion of btCS-A**

The slight activity preference of C lyase for 6-sulfated regions is noticeable when the disaccharides derived from the atypical substrate btCS-A are analyzed. Disaccharide analysis of products from 2-day incubation revealed nearly equivalent amounts of 4 and 6-sulfated disaccharides (43 and 45%; Supplementary data, Figure S12B and Table A2),

when compared with the excess of 4-sulfated disaccharides (50 vs 42%) seen on near-complete digestion of btCS-A (Supplementary data, Figure S1C and Table A2).



**Figure A4:** SAX-HPLC fractionation of reduced unsaturated hexasaccharides from 2-day digestions of (A) scCS-C, and (B) btCS-A with a commercial preparation of C lyase from *F. heparinum*. Both mixtures of hexasaccharides were obtained from peak 6 of their respective Bio-Gel P-10 chromatograms. The fractionated isomers are (A)  $\Delta C6;6;6S\text{-ol}$ , and  $\Delta C4;2,6;6S\text{-ol}$  for scCS-C, and (B)  $\Delta C6;6;6S\text{-ol}$ ,  $\Delta C4;6;6S\text{-ol}$ , and  $\Delta C4;4;4S\text{-ol}$  for btCS-A. The mean the percentage of material in each peak is given in parentheses. The NaCl gradient is shown with the continuous light grey line.

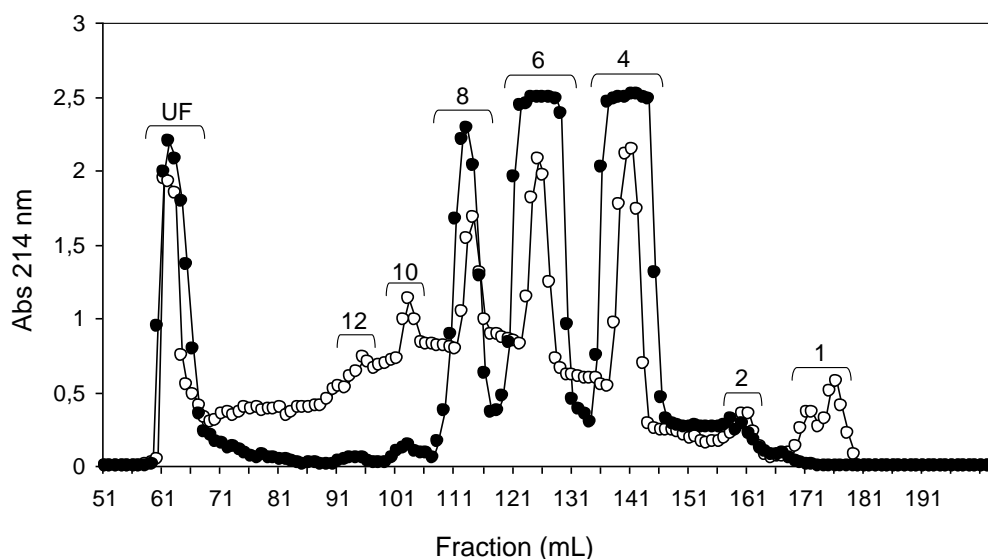
As in the case of scCS-C, the amount of hexasaccharide is relatively high in btCS-A digestions showing substantial endolytic activity. Again a small number of isomers

(three) were obtained from SAX-HPLC (Figure A4B) of the mixture of reduced hexasaccharides (peak 6, open triangle, in Figure A1). These were characterized by both  $^{13}\text{C}$ -gHSQC NMR (Supplementary data, Figure S15) and MS (Supplementary data, Figure S14C and D), giving the following structures:  $\Delta\text{C6;6;6S-ol}$ ,  $\Delta\text{C4;6;6S-ol}$  and  $\Delta\text{C4;4;4S-ol}$  (Table A1). The 57% occurrence of a  $\Delta\text{C4;4;4S-ol}$  hexasaccharide was unexpectedly high (Table A2). Occurrence may have been elevated somewhat by the slight preference for cleavage at 6-sulfated sites, but it can also be a consequence of clustering of 4-sulfation in the starting material. The presence of the  $\Delta\text{C6;6;6S-ol}$  hexasaccharide was likewise surprising, even though in small amounts. Random distributions of 6-sulfated sites in even a slightly 4-sulfate-rich sample would make the occurrence of three in a row rare. Evidence supporting the occurrence of sulfation domains rather than random distribution in CS backbones does exist (Sorrell et al. 1993). In any case, preference for cleavage in 6-sulfated regions over 4-sulfated regions is small when btCS-A is used as a substrate for the C lyase preparation.

#### **A.3.5 Hyaluronidase digestion of btCS-A and scCS-C**

It is clear from the above results that longer oligosaccharides are produced when endolytic activity dominates and production of certain products can be enhanced when non-reducing end processing is reduced by a less preferred sulfation pattern. It is therefore of interest to examine products from digestion with other enzymes for which endolytic activity naturally dominates. Commercial preparations of ovine hyaluronidase, a hydrolase named for its activity toward nonsulfated HA, is known to have such specificity (Takagaki et al. 1994). Preference for cleavage at 4-sulfated or non-sulfated sites (Knudson et al. 1984) should complement observations made using the

chondroitinase C preparation where cleavage at 6-sulfated sites is slightly preferred. This hydrolase would also add saturated oligosaccharides to the growing library of unsaturated oligosaccharides produced by the lyases. Sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE) of the enzyme shows a number of bands, but these are believed to be proteolysis products of a single enzyme (see Materials and methods). Figure A5 presents gel-permeation results for digestions with hyaluronidase on both CS substrates. The small amounts of disaccharide and heavy distribution toward medium sized-oligosaccharides (ranging from 4 to 12 residues, Figure A5) support the predicted effects of high endolytic activity. Surprisingly, hyaluronidase appears to show some enhanced activity toward a 6-sulfation-rich substrate, as scCS-C appears to have been digested faster than btCS-A (Figure A5).



**Figure A5:** Size fractionation on Bio-Gel P-10 column of the products from 2-day digestion of btCS-A (-o-), and scCS-C (-●-) with a commercial preparation of ovine hyaluronidase. The degree of polymerization is given at the top of each peak. UF stands for unfractionated material.

The apparent acceleration of digestion may, however, reflect differences in accessibility due to secondary and tertiary structure as opposed to specificity at the cleavage site.

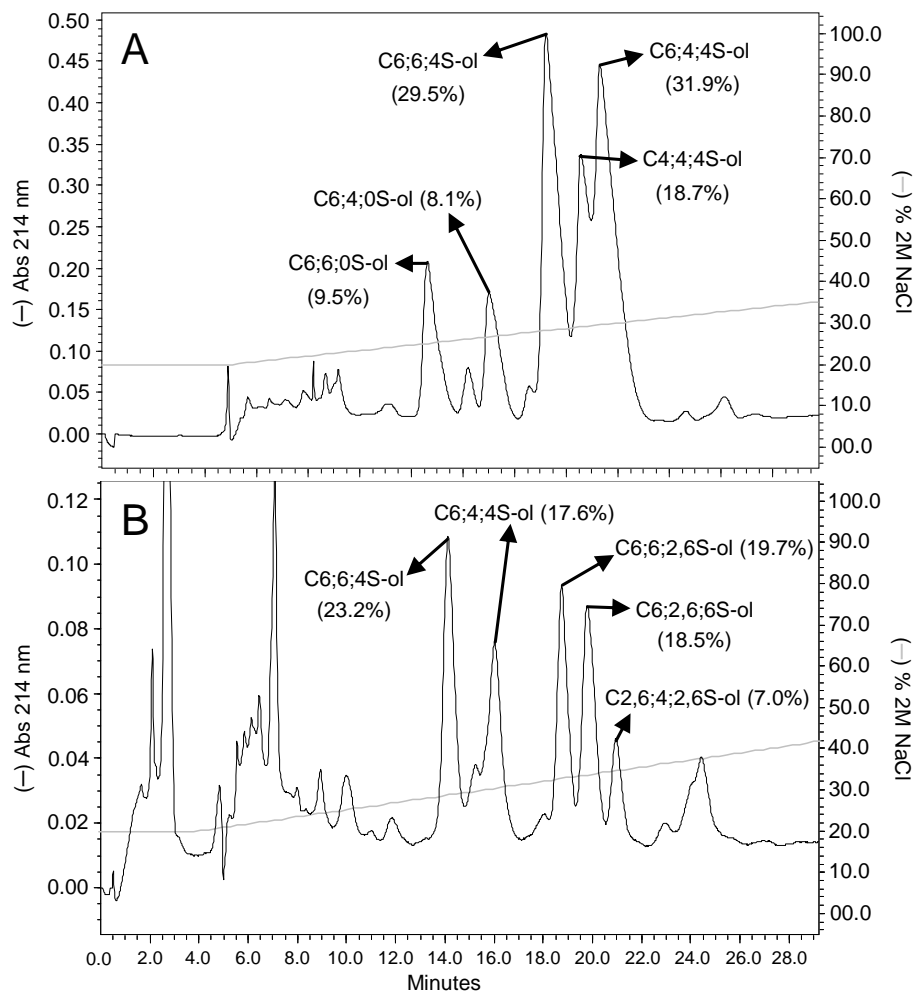
Additional information can again be obtained by separation and characterization of hexasaccharide fractions.

Digestions of both btCS-A and scCS-C showed two well-separated groups of reduced hexasaccharides on SAX-HPLC (Figure A6). In the case of btCS-A, <sup>13</sup>C-gHSQC NMR (Supplementary data, Figure S16) and MS spectra (Supplementary data, Figure S17) showed the first group to be C6;6;0S-ol and C6;4;0S-ol, and the second group to be C6;6;4S-ol, C4;4;4S-ol and C6;4;4S-ol (Figure A6A, Table A2). The groups of HPLC peaks differ in the extent of sulfation. The hexasaccharides in the first group, with a non-sulfated GalNAc at the reducing end, were in fact anticipated based on the expected ability of hyaluronidase to cleave at non-sulfated sites (Knudson et al. 1984). The relatively high percentage of hexasaccharides bearing 6-sulfation in both groups is consistent with the suggestion that 6-sulfated regions may be more accessible to the enzyme. It is noteworthy, however, that all hexasaccharides of the second group have 4-sulfation at the reducing ends, confirming a preference for cleavage between a 4-sulfated GalNAc and a GlcA (Knudson et al. 1984).

The major hexasaccharides from the hyaluronidase digestion of scCS-C were characterized as follows: C6;6;4S-ol, C6;4;4S-ol, C6;6;2,6S-ol, C6;2,6;6S-ol and C2,6;4;2,6S-ol (Figure A6B and Supplementary data, Figure S18 for NMR, and Supplementary data, Figure S19 for MS spectra). Curiously, these products do not contain any non-sulfated GalNAc units. However, the occurrence of non-sulfated GalNAc residues is lower than in btCS-A (3%, Supplementary data, Figure S1D). More prevalent are products eluting at higher salt concentrations containing 2,6-disulfated units. This shows some enhanced hyaluronidase activity for regions with 2,6-disulfated sites in



scCS-C. Products from hyaluronidase digestion bearing 2-sulfation have been observed previously (Nadanaka and Sugahara 1997). Aside from some enhanced activity toward 2,6-disulfated regions, the large amount of hexasaccharides bearing C-type sulfation is consistent with the abundance in the starting material together with the suggested enzymatic accessibility to 6-sulfation-rich regions.



**Figure A6:** SAX-HPLC fractionation of reduced saturated hexasaccharides from 2-day digestion of (A) btCS-A, and (B) scCS-C with a commercial preparation of ovine hyaluronidase. The percentage of material in each peak is given in parentheses. The NaCl gradient is shown with the continuous light grey line.

The occurrence of hexasaccharide products with 2–5 and potentially 6-sulfation sites on digestion with hyaluronidase facilitates separation of discrete products on SAX-

HPLC. Caution must be exercised, however, in assuming that hexasaccharide structures observed represent sequences abundant in native polymers. The transglycosidase activity of hyaluronidase can clearly generate sequences not present or abundant in its polymeric substrates (Takagaki et al. 1994, 1999). However, our observations are reproducible and remain relevant to the potential for defined product retrieval from btCS-A and scCS-C.

#### **A.4 Discussion**

Well-defined, higher-order, GAG oligosaccharides can be of a tremendous value in the investigation of the activities of the numerous proteins involved in GAG metabolism and modulation of cellular function. It would initially seem that digestion of naturally occurring GAG polymers followed by separation on the basis of size and charge would offer a route to the preparation of such oligosaccharides. However, high exolytic activities of certain enzyme preparations lead to the production of primarily disaccharides, and simple statistical predictions of structural diversity, even for oligosaccharides as small as hexasaccharides, would suggest unmanageable numbers of isomers. Work presented here shows that the non-random distribution of sulfation in CS chains, together with cleavage preferences of commercial preparations of digestive enzymes, contribute to a limited diversity in products and enhanced utility of a digest and separate preparation protocol.

The current research was restricted to a study of the activity of just three commonly used commercial enzyme preparations, and their actions on just two readily available CS standards. However, comparative analysis of digestion products from the two differentially sulfated substrates provided some insight into the relationship of cleavage specificities of the enzyme preparations and the limited number of

hexasaccharides produced. The hexasaccharide fractions were chosen for further characterization in this work due to a combination of potential biological interest and feasibility in HSQC-based structural determination. Endolytic activity contributes substantially to the production of the biologically interesting medium-sized products in digestions with C lyase and hyaluronidase, whereas the substantial exolytic activity in ABC lyase preparations limit the production of larger oligosaccharides. However, for the small amounts of larger oligosaccharides produced, specificities for exolytic cleavage seem to contribute to the production of limited numbers of isomers, possibly by enhancing opportunities for endolytic cleavage when regions of low exolytic activity are encountered. For example, the commercial preparation of ABC lyase from *P. vulgaris*, whose activity results from a competition between two enzymes (endolyase and exolyase), produced a very limited set of hexasaccharide structures, mostly with a central and/or non-reducing end 6-sulfated GalNAc or a 2,6;0S pattern penultimate to the reducing end (Table A2). This is likely due to its 4-sulfation cleavage preference at the reducing end (Hardingham et al. 1994). Exploiting this preference, 2 mg of the uncommon hexasaccharide,  $\Delta$ C2,6;0;4S was prepared by treating 75 mg of scCS-C with ABC lyase.

The commercial preparation of chondroitinase C from *F. heparinum* showed predominantly endolytic action. Only 8% disaccharides were produced with limited digestion as opposed to the 70% with limited ABC lyase digestion (Table A2), something that could obscure any effect of exolytic stalling on hexasaccharide product distributions. Analysis of the small amount of generated disaccharide products did show effective cleavage at 4-sulfated GalNAc units as well as 6-sulfated units. There is, in fact, only a

slight elevation of 6-sulfation in disaccharides isolated after short digestions of btCS-A when compared with disaccharides produced on near-complete digestion. The 4,4,4-hexasaccharide isolated from a btCS-A digestion was also produced in large amounts (57%, Table A2) showing cleavage ability at a 4-sulfated site, but all other hexasaccharides from either btCS-A or scCS-C digestion have a 6-sulfated GalNAc unit located at their reducing ends. In addition, the SEC profiles of both substrates were observed to be similar (Figure A3). Hence, specificity for both 4- and 6-sulfated sites does exist. We cannot exclude that this results from contamination of the chondroitinase C preparation with an ABC type enzyme, but both AC and C isolates have previously been shown to have substantial activities toward CS-A and CS-C (Gu et al. 1995; Aguiar et al. 2003). In the former case, the enzyme is quite pure (Gu et al. 1995).

As expected, the commercial hyaluronidase preparation from sheep testes exhibits major endolytic activity producing medium- (4- to 8-mers, Figure A6) and large-sized oligosaccharides (peaks UF in Figure A6) as primary products. As expected, it also displays a slightly enhanced ability to cleave after both 4-sulfated and non-sulfated GalNAc residues (Knudson et al. 1984). For digestion of scCS-C, there is an enhanced abundance of hexasaccharides that carry 2- and 6-disulfated disaccharide units, indicating a possible preference for regions bearing additional sulfation on GlcAs as well. The number of hexasaccharide types is larger than that produced by the lyases studied here, but the number is still below statistical expectations. The occurrence of hexasaccharides with both less and more sulfation than one per GalNAc residue also facilitates the separation of products by SAX in this case.

Using enzyme preparations that have substantial endolytic activities combined

with a consideration of enzyme specificities and a judicious choice of substrates can clearly produce significant amounts of well-defined CS oligosaccharides. For example, treatment of btCS-A with chondroitinase C produced more absolute amounts of hexasaccharides than digestions with chondroitinase ABC (Table A2). The SAX-HPLC-based separation of hexasaccharides from the chondroitinase C digestion of btCS-A revealed only three major isomers (Figure A4B). High levels of larger oligosaccharides were also obtained from hyaluronidase digestions of either btCS-A or scCS-C (Figure A5). Separation of hexasaccharides from hyaluronidase digestion on SAX-HPLC shows well-separated sets of structures bearing two, three or more sulfates, and again no more than three isomers in each group (Figure A6). The structure, C6;6;4S-ol, represents 32% of all hyaluronidase produced hexasaccharides from btCS-A, and the hexasaccharide amount represents 15% of total digested material (Table A2). This enabled ready production of 3 mg of a very specific hexasaccharide from 75mg of readily available starting material.

The data reported in Table A1 is far from a complete survey of production possibilities. But it shows that with a proper choice of enzyme and substrate the diversity in structures produced can be minimized and certain target structures can be produced with reasonable yields. There is some additional supporting data in the literature. A library of CS octasaccharides obtained from the hyaluronidase digestion of scCS-C has, for example, been presented (Deepa et al. 2007). Digestion conditions are different and structures are deduced by additional enzymatic degradation as opposed to NMR/MS analyses. Many of the basic specificities observed in our studies can, however, be seen in the octasaccharide structures as well. Some of the products of such research have already

proven useful in development and use of epitope recognition monoclonal antibodies (Caterson et al. 1990; Pothacharoen et al. 2007). Hopefully, the combination of data presented here and elsewhere will suggest procedures for the production of targeted oligosaccharides and guide additional research into other substrate–enzyme combinations for preparation of biologically active GAG oligosaccharides with well-defined structures.

## **A.5 Materials and methods**

### **A.5.1 Materials and reagents**

The sodium salt of btCS-A, the sodium salt of scCS-C, hyaluronidase from sheep testes (type V) (EC 3.2.1.35), chondroitin ABC lyase from *P. vulgaris* (EC 4.2.2.4), chondroitin C lyase from *F. heparinum* (EC 4.2.2.X), Sephadex G-15 resin (fractionation range of dextrans <1.5 kDa) and 1,9-dimethylmethylene blue (dye content 80%) were purchased from Sigma-Aldrich Co. (St Louis, MO). A pre-packed Spherisorb S10 SAX column (10 × 250 mm, 5 µm) was from Waters Corporation (Milford, MA). Bio-Gel P-10 gel resin (fractionation range of dextrans from 1.5 to 20 kDa) in fine polyacrylamide beads, Bio-Gel P-60 (fractionation range of dextrans from 10 to 60 kDa) in medium polyacrylamide beads, the polypropylene chromatographic columns (120 × 1.5 and 120 × 3.0 cm for SEC and 1.0 × 50 cm for desalting) were purchased from Bio-Rad Life Science (Hercules, CA). Deuterium oxide “100%” (D 99.96%) was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA), and HPLC-grade methanol was purchased from Fisher Scientific (Fair Lawn, NJ).

Analysis of commercial hyaluronidase and chondroitinase ABC and C preparations The purity of enzymes was checked by SDS–PAGE as presented in Supplementary data, Figure S20, and several bands were excised for MS analysis of

peptides from a trypsin digest. All enzyme preparations show multiple bands. Partially resolved bands near 120 kDa in the chondroitinase ABC sample are consistent with an expected mixture of two isoforms (endolyase and exolyase), and the presence of the endolyase was confirmed by MS. Although one preparation examined was designated bovine serum albumin free (BSA free), the bands near 55 kDa show peptides consistent with BSA. BSA is frequently used as a stabilizer in these preparations, and when this is done, the amounts are much higher. The chondroitinase C preparation also showed several bands. While not specified as either BSA containing or BSA-free in the product literature, the preparation does contain BSA as a stabilizer, accounting for the intense band near 55 kDa. The additional bands above 100 kDa appear to be associated with BSA (peptides from the distinct band at 120kDa were shown by MS to be consistent with the BSA sequence). The 74-kDa band yields peptides consistent with the 700 amino acid AC lyase, but the band at 62kDa could not be identified.

The commercial preparation of hyaluronidase from sheep testes type V showed a major band at 28 kDa an additional band at 22 kDa and a minor band at 56 kDa. The bands at 28 and 56 kDa had originally been interpreted as dimers and tetramers (Khorlin et al. 1973). However, recent structural work on homologous hyaluronidases shows single polypeptides of 55kDa having a catalytic domain of 35kDa (Stern and Jedrzejak 2006). Tests for activity on size separated fractions of our material showed activity in both the 56- and 28-kDa materials. Both the 33- and 58-kDa bands from bovine testicular hyaluronidase have been shown to have activity previously (Oettl et al. 2003). Given the lack of unique activities for the separated fractions, commercial preparations of all enzymes were used in the in the following digestions. SDS-PAGE data shown in

Supplementary data, Figure S20 can be used in the future to alert investigators to possible variations in commercial preparations and consequent deviations from results reported here.

#### **A.5.2 Limited enzymatic digestions of btCS-A and scCS-C**

Five ABC lyase digestions of btCS-A were individually performed incubating 150 mg of the polysaccharide with 1 mg of chondroitinase ABC (0.33 IU) in a 5 mL digesting buffer (50 mM Tris-HCl, pH 8.0, 150 mM sodium acetate, 100 µg/mL of BSA) at 37°C, for 10, 30, 60, 150 or 270 min. After the respective periods of incubation, each digested sample was heated (boiled) at 100°C for 15 min to stop the reaction. One ABC lyase digestion of scCS-C and two C lyase digestions of btCS-A and scCS-C were performed using the same protocol described at the beginning of this section; however, only the period of 150 min for the ABC lyase digestion of scCS-C was executed, whereas 48 h was used for C lyase digestions with 8.7 µg of enzyme (2.5 IU). Hyaluronidase digestions of both btCS-A and scCS-C were performed by incubation of each CS type (150 mg) with 10 mg of hyaluronidase ( $1.5 \times 10^4$  IU) in 3 mL of 50 mM sodium phosphate, 150 mM NaCl (pH 6.0) at 37°C, for 48 h each. The digested samples were then heated (boiled) at 100°C for 15 min.

#### **A.5.3 Size fractionation of CS-digested products**

A 2.5-mL sample (75 mg) from each digestion was subjected to SEC on a Bio-Gel P-10 column (120×1.5cm) using an elution solution of 10% ethanol in 1 M aqueous NaCl at a flow rate of 1 mL/15 min/fraction. The separations were monitored by ultraviolet absorption at  $\lambda = 214$  nm or  $\lambda = 232$  nm for the hydrolase- and lyase-derived products, respectively (Horonchik et al. 2005). Some digestion types were subjected to



size separation more than once (Figure A1). The lyase- or hydrolase-digested disaccharide and hexasaccharide fractions were selected for further separation and analyses. Tubes corresponding to their peaks were pooled, concentrated and desalted on a Sephadex G-15 column ( $1.0 \times 50$  cm) using distilled water as eluent. The respective desalted fractions were freeze-dried and stored until use. Additional fractionations of native btCS-A and scCS-C samples on Bio-Gel P-60 were carried out under similar conditions, using a larger column ( $120 \times 3.0$  cm) with loadings of 120 mg.

#### **A.5.4 Reduction of CS hexasaccharides**

To provide samples having terminal sugars reduced to their corresponding N-acetylgalactosaminitol forms (-ol), the desalted and lyophilized hexasaccharides from the Bio-Gel P-10 column (except those from the ABC lyase digestion of scCS-C) were treated with an equivalent weight of sodium borohydride ( $\text{NaBH}_4$ ) in 1 mL of water for 3 h. The reduction reactions were then stopped by adding a molar equivalent of acetic acid and stirring for 1 h in an ice-bath. This was followed by desalting on a Sephadex G-15 column ( $1.0 \times 50$  cm).

#### **A.5.5 SAX chromatography for isolation of CS isomers**

About 5 mg of the different lyase- and hydrolase-digested hexasaccharide or disaccharide fractions were dissolved in 100  $\mu\text{L}$  of water and individually subjected to SAX-HPLC on a  $10 \text{ mm} \times 0.25 \text{ cm}$  column using a linear NaCl gradient from 0 to 2 M in  $\text{H}_2\text{O}$  (pH previously adjusted to 5.0 with HCl). Elution occurred over a 60-min period at a flow rate of 3.0 mL/min. The separations were monitored by UV absorption at  $\lambda = 214$  nm or  $\lambda = 232$  nm for the hydrolase- and lyase-derived products, respectively. The peaks were collected separately, concentrated, desalted on a Sephadex G-15 column ( $1.0 \times 50$

cm), lyophilized and weighed.

#### **A.5.6 Overdigestion of btCS-A and scCS-C**

In order to provide the highest yields of disaccharides for composition analysis, 75 mg of both CS substrates were individually dissolved in 2.5 mL of the digesting buffer described above. Then, both chondroitinases (1.5 mg of 0.5 IU ABC lyase and 3.5 µg of 1 IU C lyase) were added together every 12 h, and the mixtures were kept at 37°C for 7 days. Both overdigested samples were size-fractionated on a Bio-Gel P-10 column and monitored by UV absorbance at 232 nm. The disaccharide peaks were individually pooled, concentrated, desalted, lyophilized and weighed. About 57 and 61 mg of btCS-A and scCS-C disaccharides were recovered. Correcting for water loss in lyase reactions, the best yields of disaccharides for btCS-A and scCS-C were 79 and 84%, respectively; less than 5% was recovered as monomers. The disaccharide types from each substrate were subsequently analyzed by SAX-HPLC. The integrals of the peaks were used to determine relative amounts of disaccharide types.

#### **A.5.7 Structure determination and yields**

The disaccharides and hexasaccharides were analyzed by a combination of MS and one-dimensional (1D) and two-dimensional (2D) NMR experiments, including 1D  $^1\text{H}$  (data not shown),  $^1\text{H}/^1\text{H}$  double-quantum filtered COSY,  $^1\text{H}/^1\text{H}$  total COSY (data not shown) and  $^1\text{H}/^{13}\text{C}$  gHSQC spectroscopy. The locations of sulfate groups were assigned primarily on the basis of characteristic changes in  $^1\text{H}/^{13}\text{C}$  -chemical shifts of cross-peaks in the  $^1\text{H}/^{13}\text{C}$  gHSQC spectra. The identities of the isolated products along with a description of nomenclature used are presented in Table A1. The absolute and relative yields of the disaccharides, and hexasaccharides, are tabulated in Table A2. A table of

<sup>1</sup>H- and <sup>13</sup>C-chemical shifts of all disaccharides studied, as well as those for the ΔC6;6;4S-ol hexasaccharide as an illustrative example is presented in Supplementary data (Table S2). The parameters and conditions for NMR and MS experiments, including specifications of the instruments, are also described in Supplementary data.

#### **A.5.8 Supplementary data**

Supplementary data for this article is available online at [http:// glycob.oxfordjournals.org/](http://glycob.oxfordjournals.org/).

#### **A.6 References**

Achur RN, Kakizaki I, Goel S, Kojima K, Madhunapantula SV, Goyal A, Ohta M, Kumar S, Takagaki K, Gowda DC. 2008. Structural interactions in chondroitin 4-sulfate mediated adherence of *Plasmodium falciparum* infected erythrocytes in human placenta during pregnancy-associated malaria. *Biochemistry*. 47:12635–12643.

Aguiar JAK, Lima CR, Berto AGA, Michelacci YM. 2003. An improved methodology to produce *Flavobacterium heparinum* chondroitinases, important instruments for diagnosis of diseases. *Biotechnol Appl Biochem*. 37:115–127.

Beaulieu JF, Vachon PH, Chartrand S. 1991. Immunolocalization of extracellular-matrix components during organogenesis in the human small-intestine. *Anat Embryol*. 183:363–369.

Ben-Zaken O, Tzaban S, Tal Y, Horonchik L, Esko JD, Vlodavsky I, Taraboulos A. 2003. Cellular heparan sulfate participates in the metabolism of prions. *J Biol Chem*. 278:40041–40049.

Caterson B, Griffin J, Mahmoodian F, Sorrell JM. 1990. Monoclonal- antibodies against chondroitin sulfate isomers—their use as probe for investigating proteoglycan metabolism. *Biochem Soc Trans*. 18:820–823.

Cattaruzza S, Perris R. 2005. Proteoglycan control of cell movement during wound healing and cancer spreading. *Matrix Biol*. 24:400–417.

Cohn RH, Cassiman JJ, Bernfield MR. 1976. Relationship of transformation, cell density, and growth-control to cellular distribution of newly synthesized glycosaminoglycan. *J Cell Biol*. 71:280–294.

de Aguiar C, Lobao-Soares B, Alvarez-Silva M, Trentin AG. 2005. Glycosaminoglycans modulate C6 glioma cell adhesion to extracellular matrix components and alter cell proliferation and cell migration. *BMC Cell Biol*. 6.

De Mattos DA, Stelling MP, Tovar AMF, Mourao PAS. 2008. Heparan sulfates from arteries and veins differ in their antithrombin-mediated anticoagulant activity. *J Thromb*

Haemost. 6:1987–1990.

Deepa SS, Yamada S, Fukui S, Sugahara K. 2007. Structural determination of novel sulfated octasaccharides isolated from chondroitin sulfate of shark cartilage and their application for characterizing monoclonal antibody epitopes. *Glycobiology*. 17:631–645.

Dodgson KS, Lloyd AG. 1958. Degradation of cartilage chondroitin sulphate by the chondroitinase of *Proteus vulgaris*. *Biochem J*. 68:88–94.

Domowicz M, Mangoura D, Schwartz NB. 2000. Cell specific-chondroitin sulfate proteoglycan expression during CNS morphogenesis in the chick embryo. *Int J Dev Neurosci*. 18:629–641.

Doodes PD, Cao YX, Hamel KM, Wang YM, Rodeghero RL, Kobezda T, Finnegan A. 2009. CCR5 is involved in resolution of inflammation in proteoglycan-induced arthritis. *Arthritis Rheum*. 60:2945–2953.

Ernst S, Langer R, Cooney CL, Sasisekharan R. 1995. Enzymatic degradation of glycosaminoglycans. *Crit Rev Biochem Mol Biol*. 30:387–444.

Gandhi NS, Mancera RL. 2008. The structure of glycosaminoglycans and their interactions with proteins. *Chem Biol Drug Des*. 72:455–482.

Gorio A, Lesma E, Vergani L, DiGiulio AM. 1997. Glycosaminoglycan supplementation promotes nerve regeneration and muscle reinnervation. *Eur J Neurosci*. 9:1748–1753.

Gu K, Linhardt RJ, Laliberté M, Gu k, Zimmerman J. 1995. Purification, characterization and specificities of chondroitin lyases and glycuronidase from *Flavobacterium heparinum*. *Biochem J*. 312:569–577.

Hamai A, Hashimoto N, Mochizuki H, Kato F, Makigushi Y, Horie K, Suzuki S. 1997. Two distinct chondroitin sulfate ABC lyases: An endoeliminase yielding tetrasaccharides and an exoeliminase preferentially acting on oligosaccharides. *J Biol Chem*. 272:9123–9130.

Handel TM, Johnson Z, Crown SE, Lau EK, Sweeney M, Proudfoot AE. 2005. Regulation of protein function by glycosaminoglycans—as exemplified by chemokines. *Ann Rev Biochem*. 74:385–410. Hardingham TE, Fosang AJ, Hey NJ, Hazell PK, Kee WJ, Ewins RJF. 1994.

The sulfation pattern in chondroitin sulfate chains investigated by chondroitinase ABC and ACIL digestion and reactivity with monoclonal antibodies. *Carbohydr Res*. 255:241–254.

He L, Giri TK, Vicente CP, Tollefsen DM. 2008. Vascular dermatan sulfate regulates the antithrombotic activity of heparin cofactor II. *Blood*. 111:4118–4125.

Horonchik L, Tzaban S, Ben-Zaken O, Yedidia Y, Rouvinski A, Papy-Garcia D,

- Barritault D, Vlodavsky I, Taraboulos A. 2005. Heparan sulfate is a cellular receptor for purified infectious prions. *J Biol Chem.* 280:17062–17067.
- Imberty A, Lortat-Jacob H, Perez S. 2007. Structural view of glycosaminoglycan-protein interactions. *Carbohydr Res.* 342:430–439.
- Inatani M, Haruta M, Honjo M, Oohira A, Kido N, Takahashi M, Honda Y, Tanihara H. 2001. Upregulated expression of N-syndecan, a transmembrane heparan sulfate proteoglycan, in differentiated neural stem cells. *Brain Res.* 920:217–221.
- Jandik KA, Gu KA, Linhardt RJ. 1994. Action pattern of polysaccharide lyases on glycosaminoglycans. *Glycobiology.* 4:289–296.
- Jin L, Abrahams JP, Skinner R, Petitou M, Pike RN, Carrell RW. 1997. The anticoagulant activation of antithrombin by heparin. *Proc Natl Acad Sci USA.* 94:14683–14688.
- Kaplan CD, O'Neill SK, Koreny T, Czipri M, Finnegan A. 2002. Development of inflammation in proteoglycan-induced arthritis is dependent on Fc gamma R regulation of the cytokine/chemokine environment. *J Immunol.* 169:5851–5859.
- Khorlin AY, Vikha IV, Milishnikov AN. 1973. Subunit structure of testicular hyaluronidase. *FEBS Lett.* 31:107–110.
- Koninger J, Giese NA, Bartel M, di Mola FF, Berberat PO, di Sebastiano P, Giese T, Buchler MW, Friess H. 2006. The ECM proteoglycan decorin links desmoplasia and inflammation in chronic pancreatitis. *J Clin Pathol.* 59:21–27.
- Knudson W, Gundlach MW, Schimid TM, Conrad HE. 1984. Selective hydrolysis of chondroitin sulfates by hyaluronidase. *Biochemistry.* 23:368–375.
- Lindahl U, Hook M. 1978. Glycosaminoglycans and their binding to biological macromolecules. *Ann Rev Biochem.* 47:385–417.
- Lunin VV, Li YG, Linhardt RJ, Miyazono H, Kyogashima M, Kaneko T, Bell AW, Cygler M. 2004. High-resolution crystal structure of *Arthrobacter aurescens* chondroitin AC lyase: An enzyme-substrate complex defines the catalytic mechanism. *J Mol Biol.* 337:367–386.
- Michel G, Pojasek K, Li YG, Sulea T, Linhardt RJ, Raman R, Prabhakar V, Sasisekharan R, Cygler M. 2004. The structure of chondroitin B lyase complexed with glycosaminoglycan oligosaccharides unravels a calcium-dependent catalytic machinery. *J Biol Chem.* 279:32882–32896.
- Michelacci YM, Dietrich CP. 1976. Chondroitinase C from *Flavobacterium heparinum*. *J Biol Chem.* 251:1154–1158.
- Mucci A, Schenetti L, Volpi N. 2000. H-1 and C-13 nuclear magnetic resonance

identification and characterization of components of chondroitin sulfates of various origin. *Carbohydr Polym.* 41:37–45.

Muramatsu T, Muramatsu H. 2008. Glycosaminoglycan-binding cytokines as tumor markers. *Proteomics.* 8:3350–3359.

Muthusamy A, Achur RN, Valiyaveetil M, Madhunapantula SV, Kakizaki I, Bhavanandan VP, Gowda CD. 2004. Structural characterization of the bovine tracheal chondroitin sulfate chains and binding of *Plasmodium falciparum*-infected erythrocytes. *Glycobiology.* 14:635–645.

Nadanaka S, Sugahara K. 1997. The unusual tetrasaccharide sequence GlcA  $\beta$ 1-3GalNAc(4-sulfate) $\beta$ 1-4GlcA(2-sulfate) $\beta$ 1-3GalNAc(6-sulfate) found in the hexasaccharides prepared by testicular hyaluronidase digestion of shark cartilage chondroitin sulfate D. *Glycobiology.* 7:253–263.

Oettl M, Hoechstetter J, Asen I, Bernhardt G, Buschauer A. 2003. Comparative characterization of bovine testicular hyaluronidase and a hyaluronate lyase from *Streptococcus agalactiae* in pharmaceutical preparations. *Eur J Pharm Sci.* 18:267–277.

Ototani N, Yosizawa Z. 1979. Purification of chondroitinase-B and chondroitinase-C using glycosaminoglycan-bound AH-Sepharose-4B. *Carbohydr Res.* 70:295–306.

Perris R, Perissinotto D, Pettway Z, BronnerFraser M, Morgelin M, Kimata K. 1996. Inhibitory effects of PG-H/aggreacan and PG-M/versican on avian neural crest cell migration. *FASEB J.* 10:293–301.

Pettway Z, Domowicz M, Schwartz NB, BronnerFraser M. 1996. Age-dependent inhibition of neural crest migration by the notochord correlates with alterations in the S103L chondroitin sulfate proteoglycan. *Exp Cell Res.* 225:195–206.

Pomin VH, Sharp JS, Li XY, Wang LC, Prestegard JH. 2010. Characterization of glycosaminoglycans by N-15 NMR spectroscopy and in vivo isotopic labeling. *Anal Chem.* 82:4078–4088.

Pothacharoen P, Kalayanamitra K, Deepa SS, Fukui S, Hattori T, Fukushima N, Hardingham T, Kongtawelert P, Sugahara K. 2007. Two related but distinct chondroitin sulfate mimotope octasaccharide sequences recognized by monoclonal antibody WF6. *J Biol Chem.* 282:35232–35246.

Raman R, Sasisekharan V, Sasisekharan R. 2005. Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chem Biol.* 12:267–277.

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

Rigden DJ, Jedrzejewski MJ. 2003. Structures of *Streptococcus pneumoniae* hyaluronate

lyase in complex with chondroitin and chondroitin sulfate disaccharides—insights into specificity and mechanism of action. *J Biol Chem.* 278:50596–50606.

Sasisekharan R, Raman R, Prabhakar V. 2006. Glycomics approach to structure–function relationships of glycosaminoglycans. *Ann Rev Biomed Eng.* 8:181–231.

Shaya D, Hahn BS, Bjerkan TM, Kim WS, Park NY, Sim JS, Kim YS, Cygler M. 2008. Composite active site of chondroitin lyase ABC accepting both epimers of uronic acid. *Glycobiology.* 18:270–277.

Singh K, Gittis AG, Nguyen P, Gowda DC, Miller LH, Garboczi DN. 2008. Structure of the DBL3x domain of pregnancy-associated malaria protein VAR2CSA complexed with chondroitin sulfate A. *Nat Struct Mol Biol.* 15:932–938.

Sorrell JM, Carrino DA, Caplan AI. 1993. Structural domains in chondroitin sulfate identified by anti-chondroitin sulfate monoclonal-antibodies— immunosequencing of chondroitin sulfates. *Matrix.* 13:351–361.

Stern R, Jedrzejewski MJ. 2006. Hyaluronidases: Their genomics, structures, and mechanisms of action. *Chem Rev.* 106:818–839.

Sugahara K, Mikami T, Uyama T, Mizuguchi S, Nomura K, Kitagawa H. 2003. Recent advances in structural biology of chondroitin sulfate and dermatan sulfate. *Curr Opin Struct Biol.* 13:612–620.

Takagaki K, Munakata H, Majima M, Endo M. 1999. Enzymatic reconstruction of a hybrid glycosaminoglycan containing 6-sulfated, 4-sulfated, and unsulfated N-acetylgalactosamine. *Biochem Biophys Res Commun.* 258:741–744.

Takagaki K, Nakamura T, Izumi J, Saitoh H, Endo M, Kojima K, Kato I, Majima M. 1994. Characterization of hydrolysis and transglycosylation by testicular hyaluronidase using ion-spray mass-spectrometry. *Biochemistry.* 33:6503–6507.

Thesleff I, Jalkanen M, Vainio S, Bernfield M. 1988. Cell-surface proteoglycan expression correlates with epithelial mesenchymal interaction during tooth morphogenesis. *Dev Biol.* 129:565–572.

Tully SE, Mabon R, Gama CI, Tsai SM, Liu XW, Hsieh-Wilson LC. 2004. A chondroitin sulfate small molecule that stimulates neuronal growth. *J Am Chem Soc.* 126:7736–7737.

vanPutten JPM, Hayes SF, Duensing TD. 1997. Natural proteoglycan receptor analogs determine the dynamics of Opa adhesin-mediated gonococcal infection of Chang epithelial cells. *Infect Immun.* 65:5028–5034.

Volpi N. 1994. Fractionation of heparin, dermatan sulfate, and chondroitin sulfate sequential precipitation—a method to purify a single glycosaminoglycan species from a mixture. *Anal Biochem.* 218:382–391.

Warner RG, Hundt C, Weiss S, Turnbull JE. 2002. Identification of the heparan sulfate binding sites in the cellular prion protein. *J Biol Chem.* 277:18421–18430.

Xie HX, Nie P, Chang MX, Liu Y, Yao WJ. 2005. Gene cloning and functional analysis of glycosaminoglycan-degrading enzyme chondroitin AC lyase from *Flavobacterium columnare* G(4). *Arch Microbiol.* 184:49–55.

Zhang ZQ, Park Y, Kemp MM, Zhao WJ, Im AR, Shaya D, Cygler M, Kim YS, Linhardt RJ. 2009. Liquid chromatography-mass spectrometry to study chondroitin lyase action pattern. *Anal Biochem.* 385:57–64.