STRUCTURES AND INTERACTIONS OF N-ACETYLATED OLIGOSACCHARIDES: NEW NMR APPROACHES

by

FEI YU

(Under the Direction of James H. Prestegard)

ABSTRACT

Oligosaccharides are of central importance in the development and maintenance of biological systems. They are found free in solution and as constituents of proteins (glycoproteins) and lipids (glycolipids) on the cell surface; they have been implicated in the mechanisms of action of diverse cellular processes including inflammation, cell-cell recognition and adhesion, protein turnover, and bacterial infection. Oligosaccharides, or molecular mimics thereof, are also becoming more frequent targets in drug design for the treatment of human illness and disease. As cellular functions of oligosaccharides are intimately related to their three-dimensional shapes or conformations, generating 3D structures of these molecules and their complexes with receptors is an essential step in deciphering their function.

This dissertation aims to develop NMR-based probes of conformation and protein interaction for N-acetylated oligosaccharides. Efforts focus largely on utilization of information that can be retrieved from ^{13}C enriched acetyl groups. Novel means of introducing ^{13}C-enriched sites, post isolation in natural carbohydrate products, are introduced and the ability to acquire sufficient information to characterize the conformation of the carbohydrate targets is illustrated. Structural information is obtained from chemical shift anisotropy offsets of ^{13}C carbonyl

resonances and ^{13}C-^{13}C dipolar couplings between the labeled methyl and carbonyl carbons of the acetyl groups. Three applications are presented: a structure determination of a chitin dimer in free solution, a structure determination of a chondroitin sulfate (CS) pentamer in free solution, and a preliminary study of CS hexamers bound to a chemokine (CCL5) important to T-cell migration.

INDEX WORDS: Nuclear Magnetic Resonance (NMR), Mass Spectroscopy (MS), Residue Dipolar Coupling (RDC), Chemical Shift Anisotropy (CSA), Glycosaminoglycan (GAG), Chondroitin sulfate (CS), Chemokine (C-C motif) Ligand 5 (CCL5)

STRUCTURES AND INTERACTIONS OF N-ACETYLATED OLIGOSACCHARIDES: NEW NMR APPROACHES

by

FEI YU

B.S. University of Science and Technology of China, P.R. China, 2003

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

2008

STRUCTURES AND INTERACTIONS OF N-ACETYLATED OLIGOSACCHARIDES: NEW NMR APPROACHES

by

FEI YU

Major Professor: James H. Prestegard

Committee: I Jonathan Amster
Jeffrey L. Urbauer

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2008

DEDICATION

I dedicate this dissertation to my grandparents and my parents, who have been through so much but still gave their children the best they can

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Dr. James H. Prestegard, for his support, guidance, and making my past five years in UGA an enjoyable experience. With his support, I was able to learn a variety of techniques. His brilliant guidance made the research process smooth and enjoyable. I would also like to thank Dr. I Jonathan Amster and Dr. Jeffrey L. Urbauer for serving as members of my research committee.

I wish to thank the following people who help me. Without them, I could not have completed this dissertation. Anita I. Kishore, for her priceless support and encouragement when I started in the lab; Xu Wang, for his insightful discussion and being a such good teacher of biochemistry; John Glushka, for his patience in answering all my questions about NMR; Azusa Adams; for helping me prepare the CS samples; Yizhou Liu and Fang Tian, who share their NMR expertise with me; Laura Morris, for her computer support; Jeremy Wolff from the Amster research group, for his assistance in MS; Xinghai Ning from the Boons research group, for his valuable suggestion on organic chemistry; Iris Tropp, for editing my thesis. I also want to thank the whole Prestegard group for creating the harmonious environment for working and learning.

Most of all, I thank my wife, Yanwen Chen, for her unfailing love, support, understanding, and encouragement.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION TO NUCLEAR MAGNETIC RESONANCE FOR STRUCTURAL INVESTIGATION OF CARBOHYDRATES

1.1 Structural Investigation of Carbohydrates and the Need for a New NMR Strategy

Carbohydrates that occur naturally as parts of glycoproteins, glycolipids, and glycosaminoglycans play crucial roles in biological processes that range from cell differentiation, to inflammation, fertilization, protein turnover, and pathogen infection [1-4]. Very often, specific recognition of structural features of carbohydrates by proteins is a key step in these processes and structural characterization of the carbohydrate becomes important for understanding and controlling of these interactions. However, carbohydrates still remain the least exploited among the three major classes of biomolecules, mainly due to the fact that structural characterization of carbohydrate has not been easy. Many carbohydrates do not crystallize easily and NMR spectra of carbohydrates are complex because of the functional similarity of their sugars. ^{13}C NMR and multidimensional NMR spectra using a ^{13}C dimension -offer higher resolution than ^{1}H spectra, but without enrichment with ^{13}C, sensitivity is inadequate. Enrichment in ^{13}C has been accomplished in a few cases [5,6], but unlike proteins and nucleic acids, carbohydrates cannot be easily produced biosynthetically on a large scale with designed sequences. In addition, the similarity of functional substituents (primarily hydroxyl groups) and the numerous chiral centers make the chemical synthesis of complex carbohydrates extremely difficult. Therefore, isolating native polysaccharides, followed by a chemical or enzymatic degradation of the polymer, is the desired method for obtaining many complex oligosaccharides. In order to obtain uniformly ^{13}C-

labeled oligosaccharides during this process, the ^{13}C isotopes would have to be introduced into the polysaccharide precursors at a very early stage with the aid of *in vivo* or *in vitro* metabolic systems[7,8]. This would require a large amount of isotopic reagent and would limit the usefulness of this method.

The objective of this research was to develop a new strategy for labeling isolated carbohydrates with ^{13}C, and develop an NMR-based protocol for structural characterization of these labeled carbohydrates. We capitalized on the widely-distributed N-acetyl groups in the native carbohydrates and developed a standard methodology for introducing ^{13}C-labeled sites into complex carbohydrates by exchanging acetyl groups in natural products after their isolation.

^{13}C-labeled acetyl groups are actually rich in structural information. Each of them can provide two pieces of structural information based on ^{13}C observed spectra. A ^{13}C carbonyl carbon has a large chemical shift anisotropy (CSA) which yields orientational information, and groups labeled at both the methyl and carbonyl carbons exhibit ^{13}C-^{13}C dipolar coupling that provides complementary information. Figure 1.1 shows the structural information observable in ^{13}C spectra. In both cases, the preferential orientation of the molecule in media such as aqueous dispersions of lipid or detergent bicelles is required to measure these interactions.

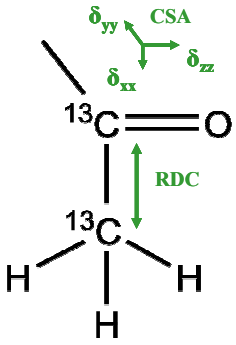


Figure 1.1 Structural information from ^{13}C enriched acetyl group

1.2 Structural Biology of Carbohydrates

Carbohydrates are critical components of all living organisms and are the most abundant class of biological molecules. The name carbohydrate arises from the basic molecular formula (CH2O) n, which can be written (C·H2O) n to show that these substances are hydrates of carbon, where n=3 or more. Carbohydrates constitute a versatile class of molecules. Carbohydrates store the energy from the sun captured by green plants, algae, and some bacteria during photosynthesis. They are the metabolic precursors of virtually all other biomolecules. Carbohydrates linked to lipid molecules, called glycolipids, are common components of biological membranes. Proteins that have covalently linked carbohydrates are called glycoproteins. These two classes of biomolecules, together called glycoconjugates, are important components of cell walls and extracellular structures in plants, animals, and bacteria [9,10]. In addition to the structural roles such molecules play, they also serve in a variety of processes involving recognition between cell types or recognition of cellular structures by other molecules. Recognition events are important in normal cell growth, fertilization, transformation of cells, and other processes[11-13].

1.2.1 Chitin

Chitin is the second most abundant polysaccharide in nature after cellulose. It is the main constituent of the protective shell in insects and crustaceans, like shrimp, lobster and crab[14]. The polysaccharide is composed of β(1-4) linked N-acetyl-D-glucosamine residues (Figure 1.2).

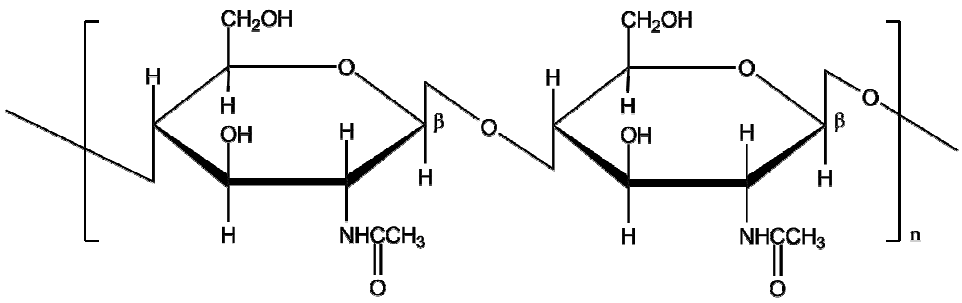


Figure 1.2 Structure of chitin. Chitin is a β (1-4) linked homopolymer of N-acetylated-D-glucosamine.

Chitin is used as binders in the paper, dye, textiles and adhesives industries[15,16]. It's also used as filtration aids, particularly in the waste water treatment industry[17]. Modified chitin has been manipulated to produce media with controlled pore size for chemical separations[14]. In the food industry chitins are used as gelling, thickening and encapsulating agents[18].

1.2.2 Glycosaminoglycans

The extracellular spaces, particularly those of connective tissue such as cartilage, tendon, and blood vessel walls, consist of collagen and elastin fibers embedded in a gel-like matrix known as ground substance. Ground substance is composed largely of glycosaminoglycans, unbranched polysaccharides of alternating uronic acid and hexosamine residues. Members of the glycosaminoglycan family vary in the type of hexosamine, hexose or hexuronic acid unit they contain (Figure 1.3).

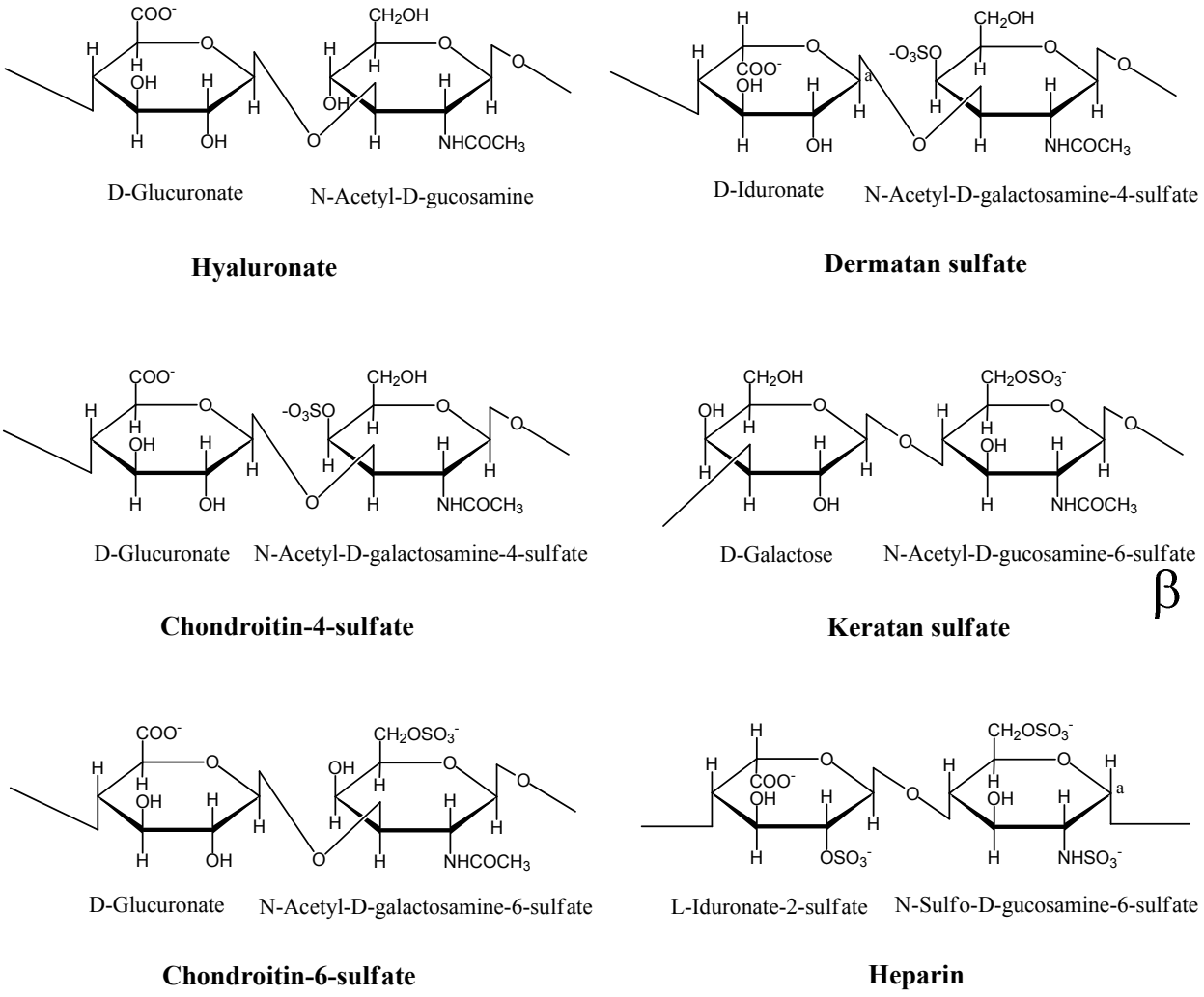


Figure 1.3 The disaccharide repeating units of the common glycosaminoglycans.

Chondroitin sulfate (CS) and dermatan sulfate (DS) have attracted less attention than heparin sulfate (HS) despite their wide distribution and common synthetic pathway. Growing recent evidence suggests, however, that CS and DS chains have intriguing functions in central nervous system development [19], wound repair [20], infection [21], growth factor signaling [22], morphogenes [23] and cell division [24], in addition to their conventional structural roles.

As a major component of the extracellular matrix, chondroitin sulfate plays a critical role in maintaining the structural integrity of the tissue. As part of aggrecan, chondroitin sulfate is a

major component of cartilage. The electrostatic repulsion generated by tightly packed and highly charged sulfate groups of chondroitin sulfate provides much of the resistance of cartilage to compression. Loss of chondroitin sulfate from the cartilage is a major cause of osteoarthritis.

Chondroitin sulfate proteoglycans, as one of the major barrier-forming molecules, affect cell migration patterns and axon path finding in the central nervous system. Chondroitin sulfate chains function as a neural stem cell marker and neurogenic molecules involved in neural stem cell proliferation [25]. These chondroitin sulfate chains are also involved in formation of the neural network by capturing and presenting heparin-binding growth factors to stem cells or neuronal cells [26].

It has been noted that chondroitin sulfate is up-regulated in granulation tissue during the wound-healing process. Compared with the normal gingival fibroblasts, fibroblasts derived from granulation tissue have much higher expression levels of versican, a CS proteoglycan [27]. It was also found that CS, along with other sulfated glycosaminoglycans, has a high concentration in human wound fluid [28]. Moreover, injection of glycylhistidyl- lysine-Cu2+, an activator of wound healing, into full-thickness rat skin wounds results in accumulation of CS and stimulates wound tissue production [29].

It has been discovered that quite a number of pathogens, including viruses, parasites and bacteria, attach to host cells and tissues through interaction with cell-surface CS and DS [30,31]. The adhesion of malaria parasite-infected erythrocytes to placenta requires a low-sulfated CS-A domain consisting of at least six disaccharide units [32,33]. In the efficient infection of herpes simplex virus 1, glycoprotein C of the virus binds CS chains [34] and CS-E is a potent inhibitor of virus infectivity [35]. The spirochete Borrelia burgdorferi, the causative agent of Lyme disease in humans, shows strain selectivity in GAG binding. A certain strain recognizes HS and DS,

whereas another binds only DS [36]. The human bacterial pathogen *Streptococcus pyogenes* expresses a large number of different types of M proteins, which bind DS, HS and heparin [37].

Recent studies have revealed that chondroitin is essential in cytokinesis and morphogenesis at early embryonic stages of *Caenorhabditis elegans* [38]. Mutations in eight squashed vulva (sqv) genes in *C. elegans* cause defects in cytokinesis during embryogenesis and in vulval morphogenesis during postembryonic development, and seven of the eight sqv genes have been shown to control the biosynthesis of the chondroitin and heparin sulphate [39]. It was proposed that chondroitin secretion into the extracellular space results in the hydration of the extracellular matrix and causes epithelial invagination.

1.3 NMR Techniques for the Structural Characterization of Carbohydrates

A number of techniques are currently available for structural elucidation of glycoprotein-derived oligosaccharides including HPLC, SDS-PAGE, MS, capillary electrophoresis and some others. Among them, mass spectrometry (MS) has proven to be a rapid and efficient technique for oligosaccharide sequencing. Often, MS can lead to a full structural characterization of oligosaccharides. However, the crucial issue of oligosaccharide stereochemistry cannot be solved by the current methodology used in routine mass spectroscopy. NMR, on the other hand, provides a way to determine the stereochemistry of oligosaccharides in solution. Because of this, NMR-based structure elucidation is very often combined with MS data for the structural characterization of carbohydrates.

1.3.1 Sequencing Oligosaccharides by Traditional Carbohydrate NMR

Despite its importance for studying the stereochemistry of oligosaccharides, obtaining NMR spectra of carbohydrates is challenging since the sugar components, specifically the hexoses of the oligosaccharides, have very similar structures. In addition, the ^{1}H signals are

observed together in a very narrow region of the ^{1}H NMR spectrum (3.3-4.1ppm), except for the anomeric proton signals, which are observed downfield at around 5ppm. Therefore, standard NMR strategies for sequencing oligosaccharides are based on protons observed in two dimensions. These techniques include: ^{1}H-^{13}C HSQC, ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H NOESY, ^{1}H-^{1}H ROESY and ^{1}H-^{13}C HMBC. HSQC spectra are particularly valuable since they offer excellent dispersion and resolution with which to begin making resonance assignments. Moreover, HSQC spectra usually have a well-defined number of cross-peaks which helps to determine the size of oligosaccharides, and the negative correlation from DEPT-HSQC for CH_2 can be used to identify exo-cyclic methylenes (C6 for pyranoses) on each monosaccharide residue. With the aid of HSQC spectra, the assignments for ^{1}H signals from each individual monosaccharide residue are primarily preformed through combination with TOCSY and COSY spectra. It is possible to measure the size of the coupling constants and the correlations to reveal the identity of the monosaccharide residue through TOCSY spectra. The TOCSY and HSQC data may also be obtained simultaneously by combining the sequences in a 3D HSQC-TOCSY experiment. It is useful and gives additional dispersion in the carbon dimension, which may facilitate the assignment of individual spin systems. The assignment of protons in a sequential manner is eventually made through a COSY experiment. The COSY spectra or double quantum filtered (DQF) COSY spectra, enable most signals for each carbohydrate residue to be assigned starting with the well-resolved signals for anomeric protons. With the protons from each sugar residue assigned, the final step to determine the sequence of the oligosaccharides is to identify the glycosidic linkage. Both NOESY and HMBC offer valuable information about the linkage, but it should be kept in mind that the strongest NOE might not be between the protons directly across the glycosidic linkage, and very often the inter-residues NOE cross-peaks are overlapped

by some intra-residue NOE cross-peaks. Therefore a standard NMR strategy for sequencing oligosaccharides involves first identification of the monosaccharide residues present in the oligosaccharide using scalar connectivities to the relatively well-resolved resonances for the anomeric protons in COSY and TOCSY followed by determination of their sequence with NOESY or HMBC spectra.

1.3.2 Determination of the Spatial Structure of Oligosaccharides

The cellular functions of oligosaccharides are intimately related to their three-dimensional (3D) shapes or conformations and knowledge of these 3D structures, both free in solution and in the bound state, is essential to deciphering the underlying factors that control recognition between oligosaccharides and specific cellular receptors. An oligosaccharide chain exhibits flexibility throughout the whole molecule on a short time scale by fast vibrations of bonds and angles, but hese result in relatively small changes in geometry. On a longer time scale conformation varies by changes of the dihedral angles. Therefore, the determination of the spatial structure of an oligosaccharide is normally focusing on the glycosidic dihedral angles. The conformationally relevant dihedral angles φ and ψ are indicated in Figure 1.4.

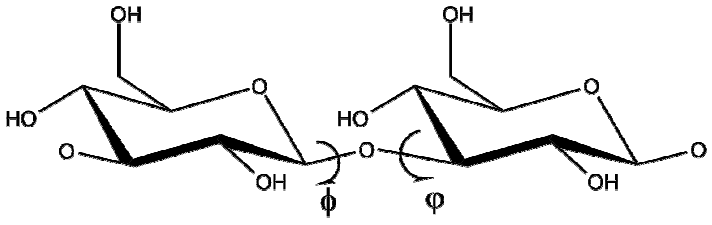


Figure 1.4 Conformational angles φ and ψ

Traditional NMR methods used to determine the spatial structure of carbohydrates are normally based on NOEs and scalar couplings. However these experiments typically yield a

limited number of structural constraints, and data interpretation is often complex, especially in the presence of internal motion. A different technique based on the measurement of residual dipolar coupling constants (RDCs) in partially-aligned media has been developed during the last decade [40,41]. The method can give information not only about short-range interactions as NOEs, but also on the relative orientation of subunits or residues. However, as in the use of NOEs for conformational studies of carbohydrates, the accurate interpretation of the residual dipolar coupling constants must take into account the internal flexibility.

1.3.2.1 Transferred Nuclear Overhauser Enhancement

Application of a weak r.f. field at the Larmor frequency of one of the spins, for a sufficiently long time, has a strong effect on the longitudinal magnetization of the non-irradiated spins, and in some cases, even enhances the magnetization of those spins. This is called the nuclear Overhauser effect (NOE). Present studies of oligosaccharide conformation rely heavily on ^{1}H-^{1}H NOEs for information on the proximity of carbon-bound protons, from which a 3D structure is deduced. This approach works well when there are sufficient measureable NOEs that are sensitive to molecular shape, as is the case of proteins. In oligosaccharides, linkages within rings largely determine the shape of individual residues. Inter-residue NOEs show the torsional angles, φ and ψ, which define conformation about the O-glycosidic linkages. Unfortunately, the number of inter-ring ^{1}H-^{1}H NOEs observed in most oligosaccharides is small, and those that are observed are often weak due to unfavorable rotational correlation times in water at room temperature. Although the latter factor can sometimes be remedied by changing solution viscosity and temperature, or by collecting data in the rotating frame, it is nevertheless true that conformational assignments based on NOE are not unequivocal. Furthermore, the effect of conformational averaging on the NOE is complex; due to the r^{6} dependence of the NOE on the

^{1}H-^{1}H inter-nuclear distance, NOE averaging will not be linear, and this nonlinearity can lead to faulty interpretations of NOE data for oligosaccharides.

1.3.2.2 Scalar Coupling Constants

Because of the shortage of available and measurable parameters, NOE data generally must be combined with other sources of structural constraints, such as those obtained from scalar coupling constants. It has long been recognized that $^{3}J_{HH}$ values obey a Karplus-type relationship, $^{3}J = A\cos^{2}\theta + B\cos\theta + C$, where θ is the dihedral angle, and A, B, and C are empirically-derived parameters whose values depend on the atoms and substituents involved. For a flexible structure, scalar coupling constants have an advantage over NOEs because that they are simple linear averages. However, the use of scalar coupling constants to investigate O-glycosidic linkage conformation is hampered by the lack of inter-residue $^{3}J_{H\text{-}H}$ values and $^{3}J_{C\text{-}H}$ values are small and hard to measure.

1.3.2.3 Residue Dipolar Couplings

Through-space magnetic dipole-dipole couplings have proven to be of great importance for the investigation of molecular conformations [40-42]. These interactions depend on the spin-spin distances and on the orientations of the inter-nuclear vectors with respect to the external magnetic field. This means that the residual dipolar coupling (RDC) is a valuable probe of long-range order and molecular structure. For oligosaccharides, the techniques offer great advantages because valuable restraints can be added to the otherwise scarce experimental information. The dipolar coupling can be described by the equation

$$D_{ij} = \xi_{ij}\left\langle \frac{\left(3\cos^{2}\theta - 1\right)}{2} \right\rangle \left\langle {}^{1}\!/\!_{r^{3}} \right\rangle = D_{\max}\left\langle P_{2}\left(\cos\theta\right)\right\rangle$$

with $D_{\max} = \xi_{ij} \cdot \left({}^{1}\!/\!_{r^{3}}\right)$ and $P_{2}\left(\cos\theta\right) = 1/2\left(3\cos^{2}\theta - 1\right)$ Eq. 1.1.

where ξij is a constant that dependents on the properties of nuclei i and j. The dipolar couplings contain both short-range (distance-dependent) and long-range (angular-dependent) structural information. In traditional NMR experiments, the internuclear vector samples orient isotropically as a result of free tumbling in aqueous solution, thus the $\left(3\cos^2-1\right)$ term is averaged to zero and a direct observation of dipolar coupling is impossible. However, with the presence of certain liquid crystals, like bicelles, the orientations the internuclear vector samples are no longer isotropically distributed due to the orientation of the bicelles in the external magnetic field. As a result, the dipolar couplings do not average to zero and splittings are observed between the dipolar coupled spin pairs. If the internuclear distances are known, as they are for directly bonded pairs, the dipolar couplings can provide new angle-dependent information for investigating biomolecule structure and dynamics.

Even though angle dependency makes dipolar couplings important tools for obtaining long-range structural information, the dynamic properties of the samples complicates the analysis of these anisotropic interactions. A simple approach to obtaining structural and motional information from such parameters is to represent their anisotropic averaging in a Cartesian coordinate system by the Saupe order matrix.

For a rigid molecule, a unit vector $\hat{r}$ in the internuclear direction can be expressed by coordinates (cx, cy, cz) that are fixed relative to an arbitrary, time-dependent molecular coordinate system given by the unit vectors $\hat{e}_x$, $\hat{e}_y$, $\hat{e}_z$.

$$\hat{r} = c_x \cdot \hat{e} + c_y \cdot \hat{e}_y + c_z \cdot \hat{e}_z = \begin{pmatrix} \hat{e}_x & \hat{e}_y & \hat{e}_z \end{pmatrix} \cdot \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix} \qquad \text{Eq. 1.2}$$

The residual dipolar coupling is proportional to $P_2(\cos\theta)$ where cosθ is the scalar product between $\hat{r}$ and a unit vector $\hat{b}$ parallel to the magnetic field axis:

$$\cos\theta = \hat{b}\cdot\hat{r} = \begin{pmatrix}\hat{b}\cdot\hat{e}_x & \hat{b}\cdot\hat{e}_y & \hat{b}\cdot\hat{e}_z\end{pmatrix}\cdot\begin{pmatrix}c_x\\c_y\\c_z\end{pmatrix} = \begin{pmatrix}C_x & C_y & C_z\end{pmatrix}\cdot\begin{pmatrix}c_x\\c_y\\c_z\end{pmatrix} = \sum_{i=x,y,z} C_i c_i \qquad \text{Eq. 1.3}$$

where $C_{x,y,z}$ describe the direction of $\hat{b}$ relative to the coordinate system $\hat{e}_x$, $\hat{e}_y$, $\hat{e}_z$.

$\langle P_2(\cos\theta)\rangle$ then reduces to the form

$$\langle P_2(\cos\theta)\rangle = 3/2\left\langle\left(\sum_{i=x,y,z} C_i c_i\right)^2\right\rangle - 1/2 = 3/2\sum_{\substack{i=x,y,z\\j=x,y,z}}\langle C_i C_j\rangle c_i c_j - 1/2 \qquad \text{Eq. 1.4}$$

With the help of $\langle C_i C_j\rangle = \langle C_j C_i\rangle$, $\langle C_x^{\,2}\rangle + \langle C_v^{\,2}\rangle + \langle C_z^{\,2}\rangle = 1$, and $c_x^{\,2} + c_y^{\,2} + c_z^{\,2} = 1$, it is easy to show that this is equivalent to

$$\langle P_2(\cos\theta)\rangle = \sum_{\substack{i=x,y,z\\j=x,y,z}} c_i S_{ij} c_j \qquad \text{Eq. 1.5}$$

where S_{ij}

$$S_{ij} = 3/2\langle C_i C_j\rangle - 1/2\,\delta_{ij} \qquad \text{Eq. 1.6}$$

Are elements of the Saupe order matrix and δ_{ij} is the Kronecker delta (δ_{ij}=1 if i=j; otherwise δ_{ij}=0).

The Saupe matrix is real, symmetric and traceless. Therefore it contains five independent elements. If the structure of a molecule is known, these five elements can be determined from at least five independent dipolar couplings. In practice, many more couplings usually are measured. The five elements of the Saupe matrix can be derived from a linear fit procedure [43,44].

Because the Saupe order matrix S is real and symmetric, it is always possible to find a molecular coordinate system where S is diagonal. In this system, equation 1.1 reduces to

$$D = D_{\max}\langle P_2(\cos\theta)\rangle = D_{\max}\left[\tilde{S}_{xx}\tilde{c}_x^{\,2} + \tilde{S}_{yy}\tilde{c}_y^{\,2} + \tilde{S}_{zz}\tilde{c}_z^{\,2}\right] \qquad \text{Eq. 1.7}$$

where the tilde expresses the elements of S and c in the principal axis system.

There are some alternate practices to represent anisotropic averaging. One of them uses a polar system, in which $\tilde{\Phi}$ and $\tilde{\Theta}$ provide a description of the internuclear distance unit vector in the principal axis system. Equation 1.7 then becomes:

$$D = D_{\max}\tilde{S}_{zz}\left[P_2(\cos\tilde{\Theta}) + \eta/2\sin^2\tilde{\Theta}\cos 2\tilde{\Phi}\right] \qquad \text{Eq. 1.8}$$

with η being the rhombicity defined as $\eta = \left(\tilde{S}_{xx} - \tilde{S}_{yy}\right)/\tilde{S}_{zz}$.

Many of these calculations become simpler when irreducible representations of these matrices and vectors are used. For example, equation 1.5, involving two 3-dimensional vectors and a 3x3 matrix, becomes an equation involving only two 5-dimensional vectors:

$$\langle P_2(\cos\theta)\rangle = \sum_{\substack{i=x,y,z\\ j=x,y,z}} S_{ij}c_i c_j = \frac{4\pi}{5}\sum_{i=-2,2} S_i^* Y_{2,i}(\Theta,\Phi) \qquad \text{Eq. 1.9}$$

where $Y_{2,i}(\Theta,\Phi)$ is a second-order spherical harmonic and S*-2,-1,…2 are the complex conjugates of the Saupe order matrix in irreducible form with:

$$S_0 = \sqrt{\frac{5}{4\pi}} S_{zz}$$

$$S_1 = -\sqrt{\frac{5}{4\pi}} \sqrt{\frac{2}{3}} \left(S_{xz} + i S_{yz} \right)$$

Eq. 1.10

$$S_2 = \sqrt{\frac{4}{5\pi}} \left(\sqrt{\frac{1}{6}} S_{xx} - \sqrt{\frac{1}{6}} S_{yy} + i \sqrt{\frac{2}{3}} S_{xy} \right)$$

$$S_{-i}^{*} = (-1)^{i} S_i$$

1.3.2.5 Chemical Shift Anisotropy

Other partial alignment effects on NMR spectra display similar angular dependencies and can be used to complement RDCs. Chemical shift anisotropy (CSA) is one of these effects. The chemical shielding of nuclei in most molecular groups varies with orientation in a magnetic field due to the anisotropy of the group's electronic distribution. Normally, in solution NMR, only the isotropic average of the resonance position is seen. However, offsets to the isotropic chemical shift under partial alignment can be significant for nuclei in chemical groups with particularly large CSAs (^{13}C in a carbonyl group, for example). These offsets can be used to place constraints on the geometries of molecular models.

The chemical shift observed in an oriented system differs from the isotropic chemical shift, $\delta_{iso} = \frac{1}{3}(\delta_{11} + \delta_{22} + \delta_{33})$, by an anisotropic contribution that can be expressed as:

$$\delta_{CSA} = \left(2/3\right) \sum_{ijk} S_{ij} \cos(\theta_{ik}) \cos(\theta_{jk}) \delta_{kk}$$

$$= \left(1/3\right) \begin{bmatrix} 2\delta_{xx} \sum_{ij} S_{ij} \cos(\theta_{ix}) \cos(\theta_{jx}) \\ + 2\delta_{yy} \sum_{ij} S_{ij} \cos(\theta_{iy}) \cos(\theta_{jy}) \\ + 2\delta_{zz} \sum_{ij} S_{ij} \cos(\theta_{iz}) \cos(\theta_{jz}) \end{bmatrix}$$

Eq. 1.11

Here the δ_{kk} are the principal elements of the chemical shift tensor, and the $\cos(\theta_{ik})$ are the angles of the CSA principal frame axis, k, relative to the molecular frame axes i and j. The CSA has a similar $\cos(\theta)$ dependent form as an RDC, and the three terms in the equation can be recast in terms of the orientation of three pseudo RDC vectors. These pseudo RDC vectors have the same direction as the axis of the principle CSA tensors. The D_{max} value for each pseudo RDC is equal to two-thirds of the corresponding chemical shift offset element, $2/3\delta_{ii}$. The similarity of the RDC and CSA expressions makes it possible to directly supplement structural calculations based on RDC with CSA offsets. However, one of the primary difficulties in the use of CSA offsets is to choose the correct chemical shift reference for both the isotropic reference spectrum and an aligned spectrum. It is well known that the chemical shift perturbation is not only affected by the orientation of the nucleus but also by the variations in the environment (e.g. temperature change). Thus it is important to find a proper reference with which the chemical shift perturbation due to an environmental change can be minimized. The ideal is to keep the sample environment identical for both isotropic and aligned samples, for example, to use stretched and un-stretched polyacrymide gels. However, most alignment media can not be switched between an isotropic and liquid crystal phase without changing their chemical composition. One interesting approach to this problem is to use magic angle spinning (MAS). MAS is a widely used technique in solid state NMR to eliminate strong dipolar coupling. By applying this technique in solution NMR, the ordering of most liquid crystalline media can be interrupted and an isotropic spectrum can be obtained.

1.4 ^{13}C Enrichment of Carbohydrates

The structural information from ^{1}H observed spectra such as ^{1}H-^{1}H NOE, ^{1}H-^{13}C RDC and $^{3}J_{H\text{-}H}$ normally can provide adequate constraints for the structural characterization of small

oligosaccharides free in solution. However, when it comes to studying oligosaccharides in a large biological system, such as long oligosaccharide molecules or protein-bound carbohydrate ligands, the proton signal very often becomes too broad to provide adequate resolution for analysis. In addition, the background signal from other biomolecules can easily interfere with the proton signal of carbohydrates. ^{13}C observed spectra, on the other hand, usually give sharper signals due to their slow relaxation and large chemical shift dispersions.-However, without ^{13}C enrichment, the sensitivity is inadequate. Therefore, in order to obtain accurate NMR-data with ^{13}C spectra, enrichment of the carbohydrate sample with ^{13}C normally is required. Currently, total synthesis and bio-synthesis are the two major ways to introduce ^{13}C isotopes into carbohydrate molecules.

1.4.1 Total Synthesis: Starting from Monosaccharides

One of the most common ways to obtain ^{13}C-labeled carbohydrates is through total synthesis starting from simple monosaccharide building blocks. This is because the ^{13}C-labeled building blocks required for total synthesis (e.g. ^{13}C-glucose) are relatively cheap and commercially available. With the current development of organic and enzymatic synthesis of carbohydrates, synthesis of some extremely complex carbohydrates has been achieved. Figure 1.5 shows the pathway for synthesis of heparin-like hexamers, GlcNS(β1-4)IdoANS(α1-4)GlcNAc6S(β1-4)IdoA(α1-4) GlcNS(β1-4)IdoANS-OMe.[45] As indicated in Figure 1.5, the synthesis started from two monosaccharide building blocks 2-azido-2-deoxy-D-glucopyranosyl trichloroacetimidate 10 and the iduronic acid glycosyl acceptor 11 which can be achieved by starting from glucose and iduronic acid. The whole synthesis consists of 40 steps and the final yield is less than 1%. The benefit of this labeling method is that one can specify the target

molecules as well as the isotopicly-enriched site. The limitations, however, are also obvious due to a long synthetic process and extremely low yield.

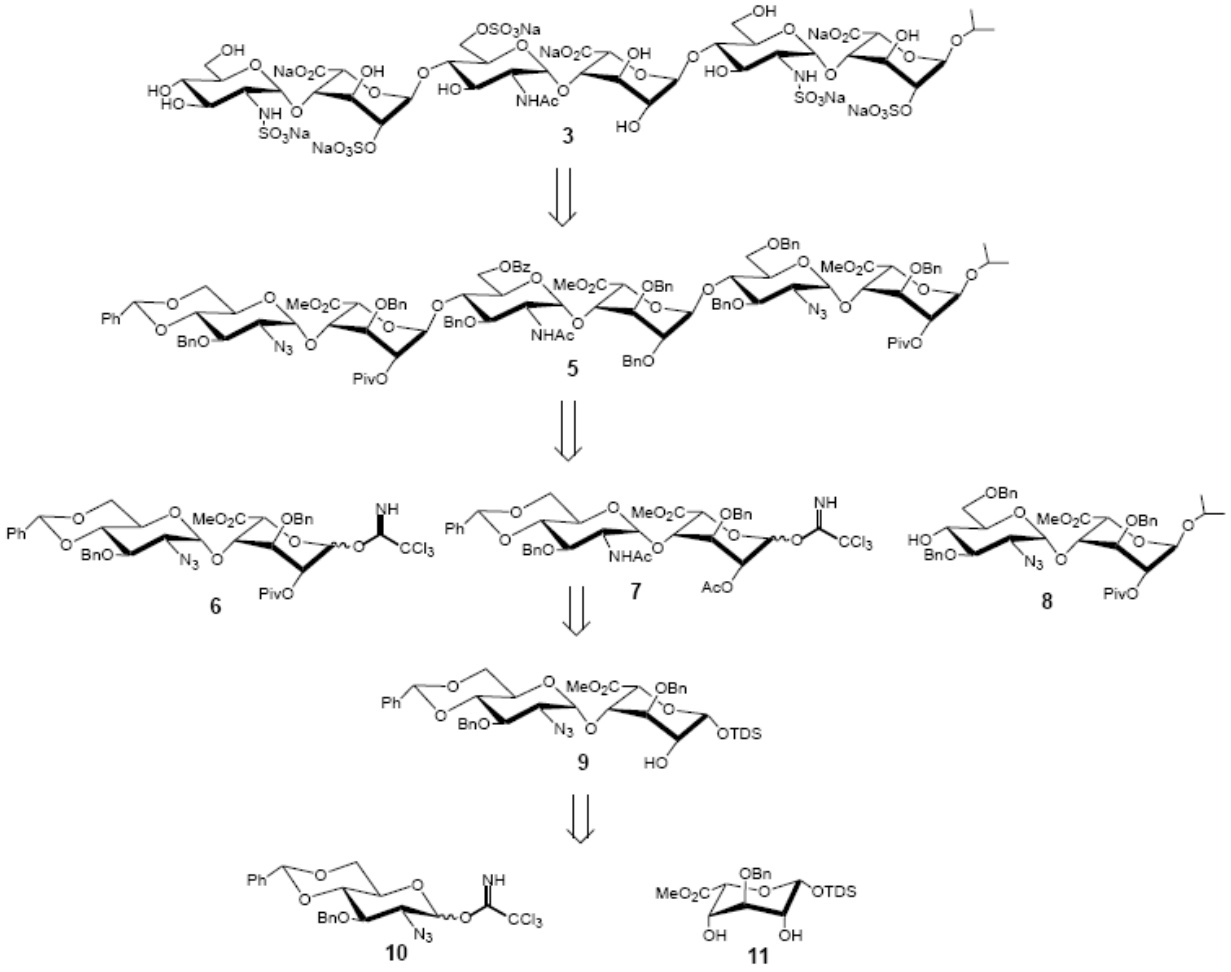


Figure 1.5 Synthetic pathway for a heparin like hexamers, GlcNS(β1-4)IdoANS(α1-4)GlcNAc6S(β1-4)IdoA(α1-4) GlcNS(β1-4)IdoANS-OMe

1.4.2 Biosynthesis: Starting from Polysaccharides

An alternative way for labeling complex carbohydrates with ^{13}C is through biosynthesis. It has been reported that uniformly ^{13}C-labeled high molecular weight HA can be produced by *Escherichia coli* transfected with recombinant HA synthase, fermented in a medium supplemented with $^{13}C_6$-D-glucose [46]. With this method, HA oligomers have been obtained by isolating the cell surface polysaccharides followed by an enzymatic digestion. The final yield for

HA hexamers (HA_6:glucose) is about 20%, which is higher than for total synthesis, and the whole process is much shorter.. However, the lack of other appropriate synthases limits the application of this method. HA is currently the only glycosaminoglycan that can be prepared in large amounts, although small amounts of labeled heparan sulfate have been prepared [47]. It must be noted that the high yield presented here is partially due to the fact that HA is one of most homogenous GAGs. For many other GAGs with a high degree of heterogeneity, a much lower yield is expected.

1.4.3 ^{13}C Enrichment in N-Acetly Groups: Target Specified

The research presented here describes new methodology for labeling complex carbohydrates with ^{13}C. Based on the fact that quite a large percentage of naturally occurring carbohydrates contain N-acetylated groups, we explored the possibility of introducing ^{13}C isotopes into carbohydrates by replacing the native N-acetyl groups with ^{13}C labeled ones. The whole labeling process, shown in the Figure 1.6, primarily consists of two steps including the N-deacetylation and N-acetylation of the carbohydrates.

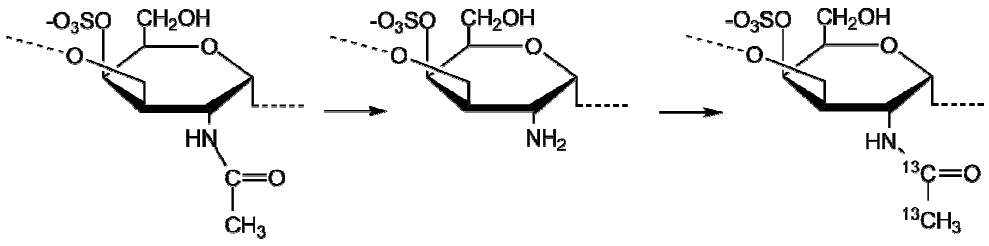


Figure 1.6 The labeling strategy for N-acetylated carbohydrates

Because of the stability of the amide bond, the chemistry required for N-deactylation is generally intense. One of the most common reactions for breaking down amides is base-catalyzed hydrolysis. This reaction is normally undertaken at high temperature with a high

concentration of strong base solution. Unfortunately, for the sulfated carbohydrates like GAGs, the O-sulfate groups are not stable in strong base solution, and strong base cannot be used [48]. Therefore, an alternative method based on hydrazinolysis is adopted for the N-deacetylation of GAGs. Hydrazinolysis of GAGs has been used to aid the deaminative cleavage of GAGs, in particular heparan sulfate (HS) [49]. It was also used to label the N-acetyl groups in the degraded oligomers with radioisotopes in order to monitor their separation [50]. The reaction retained O-sulfation patterns and produced product in high yield. Some other methodologies, like chemoenzymatic N-deacetylation, can be also used. For example, chitin can be N-deacetylated by chitin deacetylase[51]. However, these deacetylases are normally target-specific and the lack of specific enzymes limits the application of this methodology.

Unlike the chemistry required for N-deacetylation, the acetylation of free amines can be carried out under very mild conditions. By nucleophilic acyl substitution, acetyl chlorides and acetic anhydrides can both react with amines in the cold to form N-acetyl groups.

1.5 Structural Investigation of Carbohydrates with Oriented N-Acetyl Groups

The following chapters describe an NMR-based methodology for the structural investigation of N-acetylated carbohydrates. This methodology includes the development of new protocols for partial ^{13}C enrichment of carbohydrates, and the use of limited structural information from ^{13}C spectra to determine molecular conformation. This labeling approach relies on an ability to remove and replace native acetyl groups with ^{13}C-enriched forms. The ^{13}C-^{13}C RDCs and C(O) CSAs are the two pieces of major structural data for conformational analysis. Three different projects were conducted to finally reach the long term goal of structural determination for carbohydrates in a large biomolecules system.

Chapter II describes the investigation of the conformation of a chitin dimer in free solution. Chitin was chosen as an initial target molecule because of its relatively simple structure and common availability. In this project, we explore the possibility of sparsely labeling N-acetylated oligosaccharides with ^{13}C and acquiring sufficient information to select appropriate conformational models from among energetically allowed sets. This project involves the replacement of native N-acetyl groups with ^{13}C-labeled acetyl groups in a simple disaccharide derivative, $(GlcNAc)_2$-OBu, via strong base catalyzed hydrolysis. The assignment of the two acetyl groups introduced is based on a novel combination of NMR and mass spectrometry data. The ^{13}C-^{13}C RDCs and C(O) CSAs are extracted from an observed 1D spectrum of the carbon of the labeled molecule in a liquid crystal solution. Although this project was conducted in a relatively simple system, it lays the ground work for applying this methodology to more complex carbohydrate systems.

Chapter III explores the possibility of applying this methodology to GAGs which is a much more complex family of carbohydrates than chitin. The target molecule is a CS pentasaccharide, GalNAc6*S*(β1-4)GlcA(β1-3)GalNAc4*S*(β1-4)GlcA(β1-3)GalNAc4S-ol, prepared by the enzymatic hydrolysis of chondroitin sulfate. Other than using the base catalyzed hydrolysis reaction, the N-deacetylation is undertaken via hydrazinolysis to protect the O-sulfate groups. The ^{13}C-^{13}C RDCs and C(O) CSAs are used in conjunction with more traditional NMR structural data to derive the geometry of the molecule in free solution. The result is compared to that for similar molecules that have been reported in the literature and prospects for using this new data to study of protein-bound oligosaccharides are discussed.

Chapter IV explores the possibility of applying this method to carbohydrates in protein-bound carbohydrates, a large biomolecule system. This project targets chondroitin sulfate ligands

bound to the chemokine, CCL5. The binding to cell surface GAGs is key to the chemokines exerting their biological function. Studying the nature of the interaction between GAGs and chemokines can help to decipher the underlying factors that control the recognition between GAGs, ligands, and chemokines, and facilitates the development of therapeutic agents for chemokine-related diseases, such as HIV [52]. Two CS ligands, GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol and GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol, were prepared and ^{13}C sparsely labeled through hydrazinolysis. An E66S mutant of chemokine, CCL5, is expressed and purified as the target protein. The ^{13}C-^{13}C RDCs and C(O) CSAs are extracted from a ligand-protein complex aligned in a neutral polyacrymide gel. With the order parameters calculated from the aligned ^{15}N-CCL5, the RDCs and CSAs were back-calculated and compared with the ones observed in ^{13}C spectra. With this comparison, preferred structures for the CS ligand bound to CCL5 can be selected from among a series of docked structures produced by computational means.

References

(1) Haltiwanger, R. S.; Lowe, J. B. Role of glycosylation indevelopment. *Annual Review of Biochemistry* **2004**, *73*, 491-537.

(2) Helenius, A.; Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry* **2004**, *73*, 1019-1049.

(3) Dube, D. H.; Bertozzi, C. R. Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery* **2005**, *4*, 477-488.

(4) Rudd, P. M.; Wormald, M. R.; Dwek, R. A. Sugar-mediated ligand-receptor interactions in the immune system. *Trends in Biotechnology* **2004**, *22*, 524-530.

(5) Martin-Pastor, M.; Bush, C. A. Refined structure of a flexible heptasaccharide using H-1-C-13 and H-1-H-1 NMR residual dipolar couplings in concert with NOE and long range scalar coupling constants. *Journal of Biomolecular Nmr* **2001**, *19*, 125-139.

(6) Colebrooke, S. A.; Blundell, C. D.; DeAngelis, P. L.; Campbell, I. D.; Almond, A. Exploiting the carboxylate chemical shift to resolve degenerate resonances in spectra of C-13-labelled glycosaminoglycans. *Magnetic Resonance in Chemistry* **2005**, *43*, 805-815.

(7) Bouzier, A. K.; Goodwin, R.; de Gannes, F. M. P.; Valeins, H.; Voisin, P.; Canioni, P.; Merle, M. Compartmentation of lactate and glucose metabolism in C6 glioma cells - A C-13 and H-1 NMR study. *Journal of Biological Chemistry* **1998**, *273*, 27162-27169.

(8) Kauppinen, R. A.; Pirttila, T. R. M.; Auriola, S. O. K.; Williams, S. R. Compartmentation of Cerebral Glutamate in-Situ as Detected by H-1/C-13 Nmr. *Biochemical Journal* **1994**, *298*, 121-127.

(9) McCann, M. C.; Milioni, D.; Stacey, N. J.; Carpita, N. C.; Hofte, H.; Reiter, W. D.; Roberts, K. Molecular architecture of the plant cell wall. *Abstracts of Papers of the American Chemical Society* **2000**, *219*, U259-U259.

(10) Latge, J. P. The cell wall: a carbohydrate armour for the fungal cell. *Molecular Microbiology* **2007**, *66*, 279-290.

(11) Kopitz, J.; von Reitzenstein, C.; Andre, S.; Kaltner, H.; Uhl, J.; Ehemann, V.; Cantz, M.; Gabius, H. J. Negative regulation of neuroblastoma cell growth by carbohydrate-dependent surface binding of galectin-1 and functional divergence from galectin-3. *Journal of Biological Chemistry* **2001**, *276*, 35917-35923.

(12) Mourao, P. A. S. A carbohydrate based mechanism of species recognition in sea urchin fertilization. *Brazilian Journal of Medical and Biological Research* **2007**, *40*, 5-17.

(13) Hayman, M. J.; Enrietto, P. J. Cell-Transformation by the Epidermal Growth-Factor Receptor and V-Erbb. *Cancer Cells-a Monthly Review* **1991**, *3*, 302-307.

(14) Prashanth, K. V. H.; Tharanathan, R. N. Chitin/chitosan: modifications and their unlimited application potential - an overview. *Trends in Food Science & Technology* **2007**, *18*, 117-131.

(15) Synowiecki, J.; Al-Khateeb, N. A. Production, properties, and some new applications of chitin and its derivatives. *Critical Reviews in Food Science and Nutrition* **2003**, *43*, 145-171.

(16) Annadurai, G.; Lee, J. F. Equilibrium studies on the adsorption of acid dye into chitin. *Environmental Chemistry Letters* **2008**, *6*, 77-81.

(17) Minamisawa, H.; Arai, N.; Okutani, T. Preconcentration of Cu(Ii)-Nitrosonaphtholdisulfonate Complex on the Chitin/Metal-Furnace Aas. *Bunseki Kagaku* **1993**, *42*, 767-771.

(18) Benjakul, S.; Visessanguan, W.; Tanaka, M.; Ishizaki, S.; Suthidham, R.; Sungpech, O. Effect of chitin and chitosan on gelling properties of surimi from barred garfish (Hemiramphus far). *Journal of the Science of Food and Agriculture* **2001**, *81*, 102-108.

(19) Purushothaman, A.; Fukuda, J.; Mizumoto, S.; ten Dam, G. B.; van Kuppevelt, T. H.; Kitagawa, H.; Mikami, T.; Sugahara, K. Functions of chondroitin sulfate/dermatan sulfate chains in brain development - Critical roles of E and iE disaccharide units recognized by a single chain antibody GD3G7. *Journal of Biological Chemistry* **2007**, *282*, 19442-19452.

(20) Zou, X. H.; Foong, W. C.; Cao, T.; Bay, B. H.; Ouyang, H. W.; Yip, G. W. Chondroitin sulfate in palatal wound healing. *Journal of Dental Research* **2004**, *83*, 880-885.

(21) Buffet, P. A.; Gamain, B.; Scheidig, C.; Baruch, D.; Smith, J. D.; Hernandez-Rivas, R.; Pouvelle, B.; Oishi, S.; Fujii, N.; Fusai, T.; Parzy, D.; Miller, L. H.; Gysin, J.; Scherf, A. Plasmodium falciparum domain mediating adhesion to chondroitin sulfate A: A receptor for human placental infection. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 12743-12748.

(22) Fthenou, E.; Zafiropoulos, A.; Tsatsakis, A.; Stathopoulos, A.; Karamanos, N. K.; Tzanakakis, G. N. Chondroitin sulfate A chains enhance platelet derived growth factor-mediated signalling in fibrosarcoma cells. *International Journal of Biochemistry & Cell Biology* **2006**, *38*, 2141-2150.

(23) Shannon, J. M.; Mccormick-Shannon, K.; Burhans, M. S.; Shangguan, X. F.; Srivastava, K.; Hyatt, B. A. Chondroitin sulfate proteoglycans are required for lung growth and morphogenesis in vitro. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **2003**, *285*, L1323-L1336.

(24) Dietrich, C. P.; Sampaio, L. O. Changes of Acidic Mucopolysaccharides and Mucopolysaccharides during Histogenesis and Fetal Development - Possible Involvement of Chondroitin Sulfate-C and Hyaluronidase with the Processes of Differentiation and Cell-Division. *Archivos De Biologia Y Medicina Experimentales* **1980**, *13*, 441-441.

(25) Ida, M.; Shuo, T.; Hirano, K.; Tokita, Y.; Nakanishi, K.; Matsui, F.; Aono, S.; Fujita, H.; Fujiwara, Y.; Kaji, T.; Oohira, A. Identification and functions of chondroitin sulfate in the milieu of neural stem cells. *Journal of Biological Chemistry* **2006**, *281*, 5982-5991.

(26) Faissner, A.; Clement, A.; Lochter, A.; Streit, A.; Mandl, C.; Schachner, M. Isolation of a Neural Chondroitin Sulfate Proteoglycan with Neurite Outgrowth-Promoting Properties. *Journal of Cell Biology* **1994**, *126*, 783-799.

(27) Hakkinen, L.; Westermarck, J.; Kahari, V. M.; Larjava, H. Human granulation-tissue fibroblasts show enhanced proteoglycan gene expression and altered response to TGF-beta 1. *Journal of Dental Research* **1996**, *75*, 1767-1778.

(28) Penc, S. F.; Pomahac, B.; Winkler, T.; Dorschner, R. A.; Eriksson, E.; Herndon, M.; Gallo, R. L. Dermatan sulfate released after injury is a potent promoter of fibroblast growth factor-2 function. *Journal of Biological Chemistry* **1998**, *273*, 28116-28121.

(29) Simeon, A.; Wegrowski, Y.; Bontemps, Y.; Maquart, F. X. Expression of glycosaminoglycans and small proteoglycans in wounds: Modulation by the tripeptide-copper complex glycyl-L-histidyl-L-lysine-Cu2+. *Journal of Investigative Dermatology* **2000**, *115*, 962-968.

(30) Trowbridge, J. M.; Gallo, R. L. Dermatan sulfate: new functions from an old glycosaminoglycan. *Glycobiology* **2002**, *12*, 117r-125r.

(31) Menozzi, F. D.; Pethe, K.; Bifani, P.; Soncin, F.; Brennan, M. J.; Locht, C. Enhanced bacterial virulence through exploitation of host glycosaminoglycans. *Molecular Microbiology* **2002**, *43*, 1379-1386.

(32) Gamain, B.; Gratepanche, S.; Miller, L. H.; Baruch, D. I. Molecular basis for the dichotomy in Plasmodium falciparum adhesion to CD36 and chondroitin sulfate A. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 10020-10024.

(33) Chai, W. G.; Beeson, J. G.; Lawson, A. M. The structural motif in chondroitin sulfate for adhesion of Plasmodium falciparum-infected erythrocytes comprises disaccharide units of 4-O-sulfated and non-sulfated N-acetylgalactosamine linked to glucuronic acid. *Journal of Biological Chemistry* **2002**, *277*, 22438-22446.

(34) Mardberg, K.; Trybala, E.; Tufaro, F.; Bergstrom, T. Herpes simplex virus type 1 glycoprotein C is necessary for efficient infection of chondroitin sulfate-expressing gro2C cells. *Journal of General Virology* **2002**, *83*, 291-300.

(35) Bergefall, K.; Trybala, E.; Johansson, M.; Uyama, T.; Naito, S.; Yamada, S.; Kitagawa, H.; Sugahara, K.; Bergstrom, T. Chondroitin sulfate characterized by the E-disaccharide unit is a potent inhibitor of herpes simplex virus infectivity and provides the virus binding sites on gro2C cells. *Journal of Biological Chemistry* **2005**, *280*, 32193-32199.

(36) Parveen, N.; Caimano, M.; Radolf, J. D.; Leong, J. M. Adaptation of the Lyme disease spirochaete to the mammalian host environment results in enhanced glycosaminoglycan and host cell binding. *Molecular Microbiology* **2003**, *47*, 1433-1444.

(37) Frick, I. M.; Schmidtchen, A.; Sjobring, U. Interactions between M proteins of Streptococcus pyogenes and glycosaminoglycans promote bacterial adhesion to host cells. *European Journal of Biochemistry* **2003**, *270*, 2303-2311.

(38) Mizuguchi, S.; Uyama, T.; Kitagawa, H.; Nomura, K. H.; Dejima, K.; Gengyo-Ando, K.; Mitani, S.; Sugahara, K.; Nomura, K. Chondroitin proteoglycans are involved in cell division of Caenorhabditis elegans. *Nature* **2003**, *423*, 443-448.

(39) Bulik, D. A.; Wei, G.; Toyoda, H.; Kinoshita-Toyoda, A.; Waldrip, W. R.; Esko, J. D.; Robbins, P. W.; Selleck, S. B. sqv-3 -7, and -8, a set of genes affecting morphogenesis in Caenorhabditis elegans, encode enzymes required for glycosaminoglycan biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 10838-10843.

(40) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 9279-9283.

(41) Tjandra, N.; Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **1997**, *278*, 1111-1114.

(42) Prestegard, J. H.; Bougault, C. M.; Kishore, A. I. Residual dipolar couplings in structure determination of biomolecules. *Chemical Reviews* **2004**, *104*, 3519-3540.

(43) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance* **1999**, *138*, 334-342.

(44) Sass, J.; Cordier, F.; Hoffmann, A.; Cousin, A.; Omichinski, J. G.; Lowen, H.; Grzesiek, S. Purple membrane induced alignment of biological macromolecules in the magnetic field. *Journal of the American Chemical Society* **1999**, *121*, 2047-2055.

(45) Ojeda, R.; Angulo, J.; Nieto, P. M.; Martin-Lomas, M. The activation of fibroblast growth factors by heparin: Synthesis and structural study of rationally modified heparin-like oligosaccharides. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **2002**, *80*, 917-936.

(46) Blundell, C. D.; DeAngelis, P. L.; Day, A. J.; Almond, A. Use of N-15-NMR to resolve molecular details in isotopically-enriched carbohydrates: sequence-specific observations in hyaluronan oligomers up to decasaccharides. *Glycobiology* **2004**, *14*, 999-1009.

(47) 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.; Sasisekharan, R. Characterization of a heparan sulfate octasaccharide that binds to herpes simplex virus type 1 glycoprotein D. *Journal of Biological Chemistry* **2002**, *277*, 33456-33467.

(48) Fraidenraich, D.; Waisman; Cirelli, A. F. Selective O-Desulfation of Heparin in Triethylamine. *Carbohydrate Polymers* **1992**, *17*, 111-114.

(49) Maccarana, M.; Casu, B.; Lindahl, U. Minimal Sequence in Heparin Heparan-Sulfate Required for Binding of Basic Fibroblast Growth-Factor. *Journal of Biological Chemistry* **1993**, *268*, 23898-23905.

(50) Vives, R. R.; Pye, D. A.; Salmivirta, M.; Hopwood, J. J.; Lindahl, U.; Gallagher, J. T. Sequence analysis of heparan sulphate and heparin oligosaccharides. *Biochemical Journal* **1999**, *339*, 767-773.

(51) Araki, Y.; Ito, E. Chitin Deacetylase. *Methods in Enzymology* **1988**, *161*, 510-514.

(52) Kuhmann, S. E.; Hartley, O. Targeting chemokine receptors in HIV: A status report. *Annual Review of Pharmacology and Toxicology* **2008**, *48*, 425-461.

CHAPTER 2

STRUCTURAL MONITORING OF OLIGOSACCHARIDES THROUGH ^{13}C ENRICHMENT AND NMR OBSERVATION OF ACETYL GROUPS[1]

Abstract

Structural characterization of biomolecules by NMR methods frequently requires the introduction or enrichment of magnetically active isotopes at particular molecular sites. Introduction is usually achieved biosynthetically, through the use of bacterial cultures grown on isotopically enriched media, but for certain types of molecules, cell-surface carbohydrates of mammalian origin for example, this is not practical. Here we explore a means of introducing ^{13}C enriched sites, post-isolation of natural carbohydrate products, and illustrate an ability to acquire sufficient information to select appropriate conformational models from among energetically allowed sets. The application presented involves replacement of native N-acetyl groups with ^{13}C-labeled acetyl groups in a simple disaccharide derivative, $(GlcNAc)_2$-OBu or O-butyl-chitobiose. The assignment of the two acetyl groups introduced is based on a novel combination of NMR and mass spectrometry data. Structural information is obtained from chemical shift anisotropy offsets of ^{13}C carbonyl resonances and ^{13}C-^{13}C dipolar couplings between the labeled methyl and carbonyl carbons of the acetyl groups. While the application is to a relatively simple system, it lays the ground work for application to biologically important complex carbohydrate systems.

[1] Yu, F; Prestegard, JH. *Biophys J.* 2006;91:1952-1959.
Reprinted here with permission of publisher.

2.1 Introduction

Carbohydrates are of central importance in the development and maintenance of biological systems. They play key roles in biological recognition processes, in development of diseases, and in many other important biological processes[1-4]. The cellular functions of carbohydrates are intimately related to their three-dimensional shapes or conformations; knowledge of these 3D structures, both free in solution and in the bound state, is essential to deciphering the underlying factors that control recognition between carbohydrates and specific cellular receptors. However, structural characterization of carbohydrate has not been easy[5,6]. Many carbohydrates don't crystallize easily and NMR spectra are complex because of the functional similarity of the sugars that combine to make complex carbohydrates. ^{13}C NMR spectra, and multidimensional versions of these spectra, offer higher resolution than ^{1}H spectra, but without enrichment in ^{13}C, sensitivity is inadequate. Enrichment in ^{13}C has been accomplished in a few cases[7,8], but for many carbohydrates, complexity and species specificity prohibits synthetic or biosynthetic introduction of ^{13}C.

Here we illustrate a procedure for introducing ^{13}C labeled sites in complex carbohydrates through the process of exchange of acetyl groups in isolated natural products. We also include a novel assignment strategy for cases with multiple acetyl groups; this strategy is based on a combination of NMR and mass spectrometry methods. We also demonstrate that adequate structural information can be obtained from ^{13}Cs in a few enriched sites to distinguish energetically reasonable from unreasonable structural models. Application is to a simple disaccharide of N-acetyl glucosamine (chitobiose) anchored to a model membrane through the addition of a butyl chain at the reducing end.

The structural characterization exploits orientationally dependent parameters in a manner similar to our recent work on a galactosyl-mannosyl disaccharide[9], except that here we focus on acetylated saccharides and spectra from their ^{13}C-labeled acetyl groups. Acetyl groups are widely distributed in carbohydrates. In glycosaminoglycans, for example, potentially every second sugar in the polymer, or derived oligomer, could carry an acetyl group[10,11]. The problem is removing these acetyls and replacing them with a ^{13}C enriched version. Here we use a carefully controlled base catalyzed hydrolysis reaction to remove acetyls and we replace them with acetyls from ^{13}C-enriched acetic anhydride. It is however likely that suitable enzyme catalyzed steps for removal or addition could replace these steps in applications to more sensitive targets.

^{13}C labeled acetyl groups are actually rich in structural information. A group labeled at the carbonyl carbon has a large chemical shift anisotropy that can provide orientational information[12]. Likewise, a group labeled at both methyl and carbonyl carbons can exhibit ^{13}C-^{13}C dipolar coupling that provides complementary information[13,14]. In both cases, preferential orientation of the molecule in media such as aqueous dispersions of lipid or detergent bicelles is required to make these interactions measurable. For illustration purposes, we also found it advantageous to enhance orientation of our disaccharide by promoting association with bicelles through the addition of a butyl chain.

The structural information, while rich, is not so abundant that introduction of a single group is adequate, so multiple acetyl groups (two in our case) are introduced. This immediately presents an assignment problem. Usually NMR assignments are accomplished by through-bond magnetization transfer (TOCSY or COSY, for example) to make sequential connections between acetyl atoms and a series of atoms in the backbones of the carbohydrate[6,15]. Without uniform ^{13}C enrichment these assignment strategies are often impractical especially for larger homo-

oligomers and oligomers with resonances broadened by association with proteins or lipid membranes. To avoid this problem, a novel approach based on distinguishing the different isotope enrichment level of each acetyl group was introduced. Both MS and NMR are capable of detecting the presence of ^{13}C isotopes. NMR does this through resonance intensity, MS by mass shifts. NMR provides in addition the structural information, but MS can distinguish residues, even in homo-polymers through well characterized fragmentation patterns[16].

The data acquired prove adequate to distinguish reasonable from unreasonable models for the structure of our $(GlcNAc)_2$-OBu disaccharide. While this characterization falls short of a de novo structural determination method, it sets a precedent for a new structural strategy that may be of particular utility in the investigation of protein interactions with complex carbohydrate systems.

2.2 Materials and Methods

2.2.1 Synthesis

Acetic anhydride-$^{13}C_4$ was purchased from Cambridge Isotope Laboratories and iatrobeads were purchased from Iatron Laboratories. All the other reagents used in preparation of the NMR samples and used in the synthesis of the glucosides were purchased from Sigma-Aldrich Co.

O-*(2-acetamido-3,4,6-tri*-O-*acetyl-2-deoxy-β-D-glucopyranosyl)-(1→4)-2-acetamido-1,3,6-tri*-O-*acetyl-2-deoxy-α, β-D-glucopyranose* **(2)**. The synthesis of the butyl analog of chitobiose **1** followed approximately the procedure outlined in Hare et al[17] (Figure 2.1). To a solution of compound **1** (50 mg, 0.12 mmol) in pyridine (1 ml) was added Ac_2O (0.4 mL, 4.2 mmol) and the solution was stirred at room temperature overnight. The residue was put on a silica gel column and eluted (EtOAc/methanol = 10:1) to yield **2** (74 mg, 0.11 mmol, 90%).

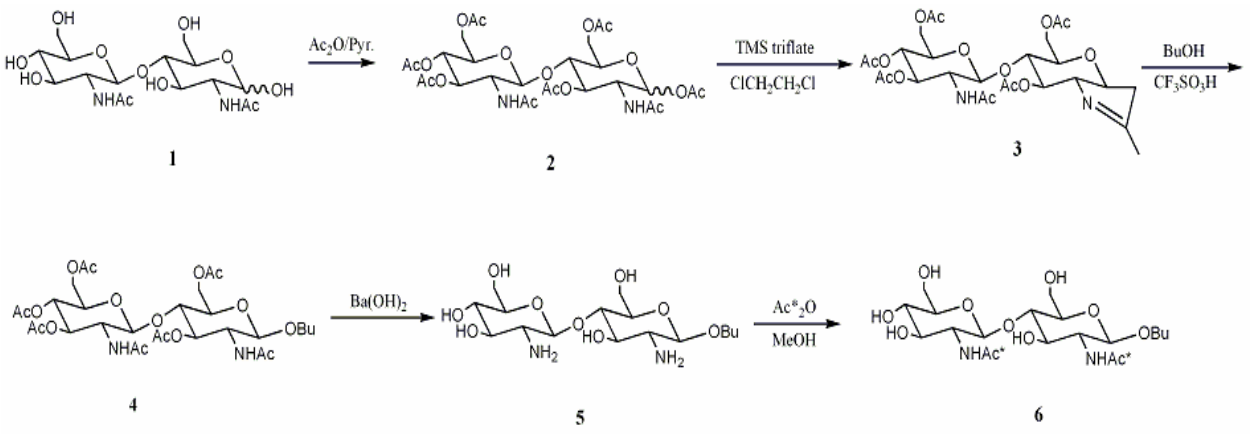


Figure 2.1 Synthesis of ^{13}C $(GlcNAc)_2$-OBu from chitobiose. *Denotes ^{13}C label.

O-*(2-acetamido-3,4,6-tri*-O-*acetyl-2-deoxy-β-D-glucopyranosyl)-(1→4)-2-methyl-(3,6-di*-O-*acetyl-1,2-dideoxy-α-D-glucopyrano)*-[*2,1-d*]-*2-oxazoline* **(3)**. To a solution of compound **2** (74 mg, 0.11 mmol) in dry dichloroethane was added TMS triflate (19 µL, 0.11 mmol, 1 equiv). The reaction flask was flushed with dry argon gas, and the reaction was stirred for 5h at 90^o. Triethylamine (43 µL, 0.33 mmol, 3 equiv) was added at 0^o to halt the reaction and the solvent was evaporated with additions of toluene. It must be noted that compound **3** is not very stable in the presence of water, and can be lost if eluted on a silica gel column. To improve yield, the crude product was moved to the next step without purification.

Butyl-O-*(2-acetamido-3,4,6-tri*-O-*acetyl-2-deoxy-β-D-glucopyranosyl)-(1→4)-2-acetamido-3,6-di*-O-*acetyl-2-deoxy- β-D-glucopyranoside* **(4)**. To a solution of **3** in dry n-butanol (1 mL) trifluoromethanesulfonic acid (2.65 µL, 0.03 mmol) was added. The reaction was protected by overlaying dry argon gas and was stirred at 60^o for 20 min. Pyridine (1.1 mL) was added to halt the reaction and the solution was evaporated with additions of toluene. To improve yield, the crude product was again moved to the next step without purification.

Butyl-O-*(2-amine-2-deoxy-β-D-glucopyranosyl)-(1→4)-2-amine-2-deoxy--β-D-glucopyranoside* **(5).** To a solution of **4** in H_2O (1 mL) was added anhydrous BaO (77 mg) and the solution was stirred at 90° overnight to promote hydrolysis of acetyl groups[18]. H_2SO_4/H_2O (0.5 N) was added at 0° until pH=7 to quench the reaction. Removal of N-acetyls at this point is not complete, leading to a percentage of ^{12}C-N-acetyls in the final product. Precipitated $BaSO_2$ was removed by filtration and the supernatant was freeze dried to give **5** as an amorphous solid (16 mg).

Butyl-O-*(2-*[*1,2-*$^{13}C_2$]acetamido-*2-deoxy-β-D-glucopyranosyl)-(1→4)-2-*[*1,2-*$^{13}C_2$]*acetamido-2-deoxy- β-D-glucopyranoside* **(6).** To a solution of **5** in dry methanol (2 mL) was added acetic anhydride-$^{13}C_4$ (7.5 μL, 0.09 mmol, 1.5 equil). The reaction was stirred for 3 hrs and halted when complete as judged by TLC. The product was purified by iatrobeads column chromatography ($CHCl_3$ /methanol=2:1) to give **6** as colorless solid (19 mg, 95%).

2.2.2 Preparation of NMR Samples

To introduce alignment in the sample and mimic a lipid bilayer environment that would promote association of the alkylated sugar, a liquid crystal solvent was prepared from pentaethylene glycol monododecyl ether (C12E5)[19]. An 8%(w/w) alignment medium was made by adding 47 uL C12E5 and 17 uL hexanol in 500uL D2O. 2.4 mg (GlcNAc)-OBu was added to make a 10 mM field alignable sample. We also made another 10mM sample in pure D_2O for isotropic measurements. 5mg tetramethylammonium bromide was added to both samples as a chemical shift reference.

2.2.3 NMR Methods

The NMR spectra were recorded on a Varian Inova spectrometer operating at 125.67 MHz for ^{13}C. Proton homonuclear decoupling was achieved using WALTZ-16 at a decoupling

power of 37W. For the sample in D_2O, spectra were acquired under near fully relaxed conditions since the signal intensity of each N-acetyl group was used to determine the isotopic ratio of ^{13}C to ^{12}C in acetyl groups of each sugar residue,. To achieve this, the repetition delay was set to 4s. For the sample in the aligned liquid crystal environment, the acquisition time was kept short (100ms) and the relaxation delay relatively long (2s) in order to avoid heating of the sample by the decoupling field.

2.2.4 MS Methods

MALDI-MS spectra were obtained on an Applied Biosystems 4700 Proteomics Analyzer (TOF/TOF), equipped with a nitrogen laser (337nm). 0.1M 2,5-Dihydroxybenzoic acid in acetonitrile was used as the ionization matrix. Collision-induced dissociation (CID) was applied to the molecular ions generated by first TOF stage to fragment the ions for residue specific analysis. The mass selection window was set relatively wide to allow isotope analysis of the fragmented ions in the second stage.

2.2.5 Molecular Simulations

Molecular structures to be used in analysis of NMR data were first generated with a web-based carbohydrate building tool[20]. Structures were then energy minimized using AMBER 7[21] using the GLYCAM_04 force field[20]. This program was also used to generate an energy map as a function of ϕ and ψ glycosidic torsion angles so that energetically unreasonable structures could be eliminated from consideration. To produce structures with variations in glycosidic torsion angles the bond rotation tools in the program CHIMERA were used[22].

2.3 Results

2.3.1 Synthesis

The first step in our strategy for the structural characterization of the $(GlcNAc)_2$-OBu disaccharide is the introduction of ^{13}C-labeled acetyl groups. Both enzymatic and chemical methods are available for the removal of the N-acetyl groups[23,24]. The methodology we illustrated in Figure 2.1 was based on strong base deacetylation. The primary side reaction involves degredation from the reducing end. Hence, we found it both necessary and convenient to protect the reducing end by preparing an O-alkyl derivative before deacetylation. In our case the introduction of an O-butyl group also provides a means of increasing the alignment of the sample[9] (Figure 2.2). The methodology based on the oxazoline chemistry allowed synthesis in a reasonable yield and gave almost exclusively the β-anomer as product.

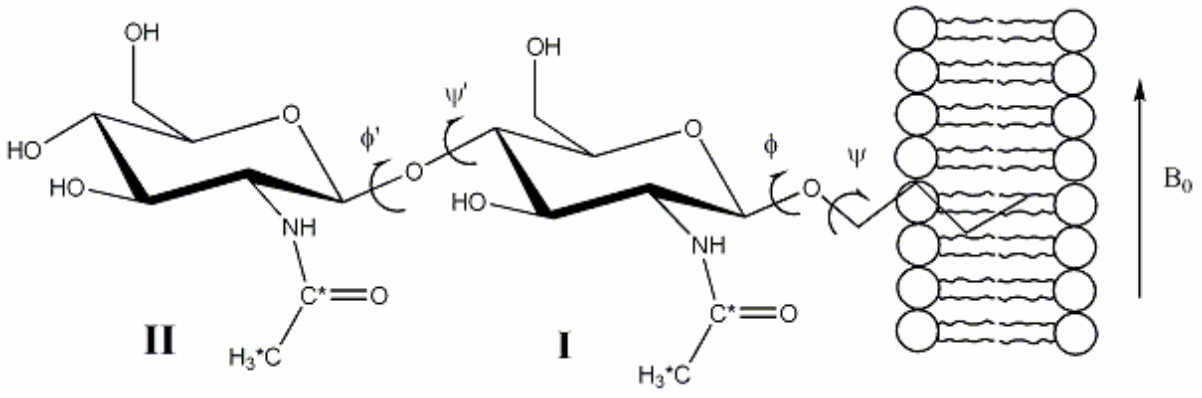


Figure 2.2 Association of the $(GlcNAc)_2$-OBu with oriented bicelles.

To facilitate assignment, it was desirable to have the $^{13}C/^{12}C$ isotopic ratio for each N-acetyl group different. Thus, partial deacetylation, relying on inherent differences in reactivity was required. Here, the temperature for the deacetylation reaction was kept lower than 90^o and the reaction time was limited to less than 12 hrs. Reacetylation with doubly ^{13}C-labeled acetic

anhydride was accomplished with preferential acetylation of the amino group based on the higher reactivity of the amino group and the higher stability of the amide bond.

2.3.2 Mass Spectrometry Data

Assaying the extent of ^{13}C-label introduction was done via mass spectrometry. Expansions of mass regions corresponding to B_1 and Y_1 ions of the natriated form of glycoside are shown in Figure 2.3-a and 2.3-b respectively[16]. These ions contain single N-acetylglucosamine rings from the non-reducing and alkylated parts of the molecule respectively. The monoisotopic peaks at lower mass represent ^{12}C containing molecules; the peaks two mass units higher represent those containing a $^{13}C_2$-acetyl group. Considering the small number of the carbon atoms in the molecule, the contribution from natural abundance ^{13}C to the higher mass peak can be ignored. Therefore, the intensity of the later peak comes primarily from the doubly ^{13}C-labeled N-acetyl group. The N-acetylglucosamine at the non-reducing end is also highly enriched. Fitting the data suggests the isotopic ratio for two N-acetyl groups are 68% and 47% for the non-reducing (II) and alkylated (I) end sugars respectively.

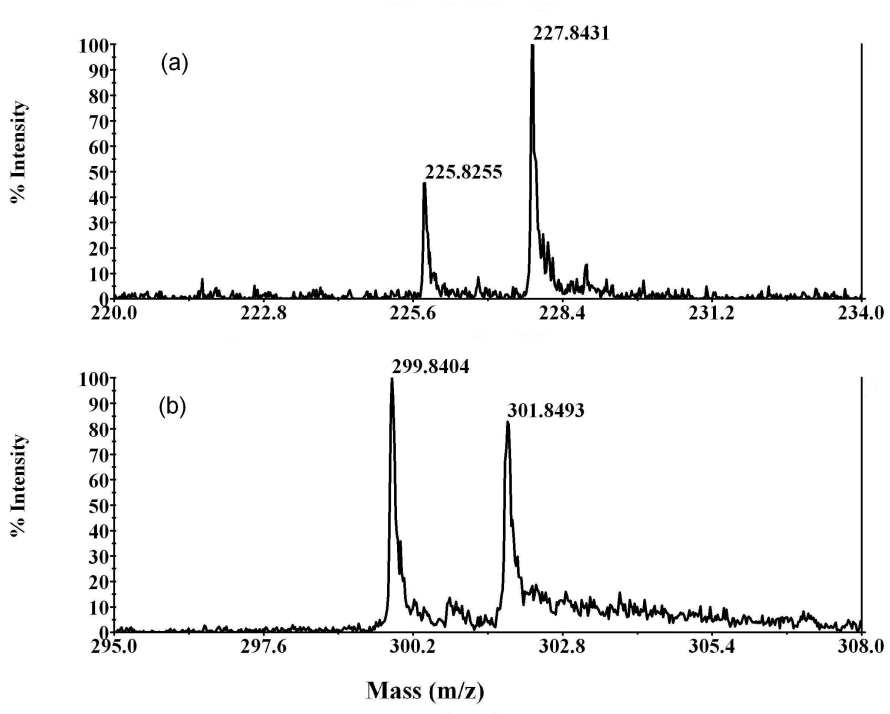


Figure 2.3 Mass spectra of B_1 (a) and Y_1 (b) ions of the natriated form of the oligosaccharide. These show different isotopic ratios for each sugar residue.

2.3.3 NMR Spectra of N-Acetyl Groups

The carbonyl region of ^{13}C NMR spectra for the sample in an isotropic D_2O solution is presented in Figure 2.4-a. Doublets with 50.2 Hz splittings from the N-acetyl carbonyls of each residue were observed at 174.70 and 174.81 ppm respectively. These arise from coupling to a second ^{13}C in the directly bonded acetyl methyl group. The intensities of these doublets are different, with the lower field doublet being approximately 1.5 times as intense. Given that in well relaxed spectra intensities reflect isotopic abundance, and given that the mass spectrometry data show the N-acetylglucosamine at the non-reducing end of the disaccharide to be more highly enriched, the down-field doublet can be assigned to the non-reducing or terminal site (II). The same region of the spectrum from a sample in an aligned C12E5 liquid crystal medium is presented in Figure 2.4-b. Both doublets are shifted down field, but by different amounts (0.29

and 0.39 ppm respectively). Also close inspection shows that the splittings of the doublets have changed (by -6.8 and -10.50 Hz respectively). These changes are typical of the effects of incomplete averaging of anisotropic parameters in aligned media. The change in splitting arises from a residual dipolar coupling and the change in chemical shift arises from a chemical shift anisotropy offset. Changes can also be seen in the acetyl methyl region of the spectrum (not show). However, the changes in splittings are identical to those measured from the carbonyl resonances, and the changes in chemical shift are negligible because of the small chemical shift anisotropy of the methyl carbons.

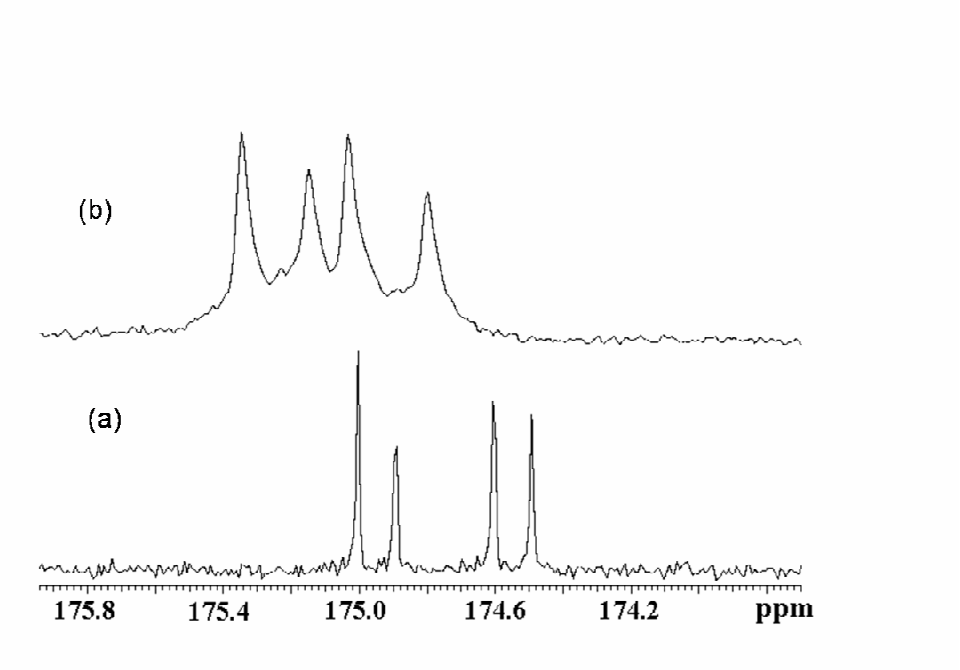


Figure 2.4 ^{1}H-decoupled ^{13}C-NMR spectrum (125MHz) of $(GlcNAc)_2$-OBu ^{13}C labeled in the acetyl carbons. The carbonyl regions of the spectrum are shown. (a) Sample in D2O (b) Sample in oriented medium. The total acquisition time for b was approximately 6 hrs.

2.4 Discussion.

2.4.1 Theory

For a doubly ^{13}C-labeled N-acetyl group, the residual dipolar couplings and chemical shift anisotropy offsets measured above potentially contain a great deal of structural information. In principle, both are functions of the molecular structure and the nature of motional averaging. The latter can, in general, be anisotropic and requires specification of both a principal order parameter and an asymmetry parameter. It can also reflect averaging by internal motions that make order different for different rings of an oligosaccharide. Because of the limited amount of data in our study, and our primary objective of illustrating principles of analysis, we will make the simplifying assumption of axially symmetric motion of a single rigid entity for our bicelle associated disaccharide. Under this assumption, the ^{13}C-^{13}C RDCs measured between the ^{13}C-labeled carbonyl and methyl carbons may be related to structural and order variables as follows[14]:

$$\mathbf{D_{ij}} = \frac{-\boldsymbol{\mu}_0 \mathbf{h}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_j}{(2\boldsymbol{\pi}\mathbf{r_{ij}})^3}\mathbf{S_{DD}}\left(\frac{3\cos^2\boldsymbol{\theta}_{ij}-1}{2}\right) \qquad \text{Eq. 2.1}$$

Here γ_i and γ_j are the gyromagnetic ratios of the two interacting nuclei, r_{ij} is the distance between the two bonded nuclei. S_{DD} is an order parameter describing the degree of motional order. θ is the angle between a given internuclear vector ij and the axial averaging axis. The order parameter, S_{DD}, and the angles θ_{ij} are the only unknowns.

Similar arguments pertain to chemical shift anisotropy (CSA) offsets of the N-acetyl group[14]. The deviation of the observed chemical shift of a nucleus from its isotropic value, δ_{CSA}, in a molecule undergoing axially symmetric motion is given by

$$\boldsymbol{\delta}_{CSA} = \frac{2}{3}\mathbf{S_{CSA}}\left(\boldsymbol{\delta}_{zz}-\frac{1}{2}\boldsymbol{\delta}_{xx}-\frac{1}{2}\boldsymbol{\delta}_{yy}\right) \qquad \text{Eq. 2.2}$$

where S_{CSA} is an order parameter analogous to that described above, and δ_{zz}, δ_{xx} and δ_{yy} are the diagonal elements of the chemical shift tensor for the carbonyl carbon of an N-acetyl group

written in a frame with the z axis coincident with the axial averaging axis. For any orientation of the N-acetyl group, δ_{zz}, δ_{xx} and δ_{yy} can be calculated by transforming the chemical shift tensor, from its principal frame for a given acetyl group, to the molecular averaging frame. Principal values for the shift tensor can be taken from model compounds in the literature. In our case these were taken from the experimentally determined ^{13}C carbonyl chemical shift tensor for glycinylglycine hydrogen chloride monohydtate[25]. The principal values for δ'_{xx} δ'_{yy} and δ'_{zz} are 244, 171, and 95 ppm, respectively. Two of the principal axes, corresponding to the principal values δ'_{xx} and δ'_{yy}, lie in the peptide plane with δ'_{yy} aligned approximately along the C=O bond. The orientation of the principal axis corresponding to δ'_{zz} is nearly perpendicular to the peptide plane. Once transformed to the molecular averaging frame, using molecular geometry information, a value for S_{CSA} can be calculated from measured CSA offsets.

For our molecule we assume that each sugar residue is rigid and that the alkyl chain at the reducing end anchors the disaccharide to the lipid bilayer with the alkyl chain extending along the bilayer normal (Figure 2.2). Moreover, we assume that the orientation of the acetyl group relative to the sugar ring is fixed in a geometry consistent with three bond scalar couplings between the amide proton and the sugar ring 2-proton. Thus, the orientation of residue I and its attached acetyl group are determined by the terminal torsion angles φ and ψ, defined as O5(I)-C1(I)-O1(I)-C1' and C1(I)-O1(I)-C1'-C2', respectively. Here C1' and C2' refer to the two carbons in the butyl chain and the numbers in parentheses refer to residue I. Likewise, the orientation of residue II and its attached acetyl group are determined by the combination of the terminal torsion angles and the glycosidic angles φ' and ψ' which are defined as O5(II)-C1(II)-O1(II)-C4(I) and C1(II)-O1(II)-C4(I)-C3(I) respectively.

If particular geometries for the sugars in our disaccharide (sets of ϕ and ψ angles) are chosen, the only unknowns left in equation 1 and 2 are the order parameters S_{DD} and S_{CSA}. So, once the dipolar coupling and the CSA offsets are measured, for a given N-acetyl group, we can calculate the order parameters S_{DD} and S_{CSA} for each sugar residue in a particular geometry. It is important to note that for any proper orientation of an approximately rigid disaccharide, the numeric values of the two order parameters must be identical. We can define a structural deviation factor which measures the variation in these parameters, as given below, and use this factor as an indicator of acceptable geometries (f close to 1).

$$\mathbf{f_i} = 1 - 2\left|\frac{\mathbf{S_{CSAi}} - \mathbf{S_{DDi}}}{|\mathbf{S_{CSAi}}| + |\mathbf{S_{DDi}}|}\right| \qquad \text{Eq. 2.3}$$

2.4.2 Conformational Analysis

To generate structures and calculate structural deviation factors, we need a molecular structure for our $(GlcNAc)_2$-OBu disaccharide. The initial structure was obtained from the online carbohydrate builder on the CCRC website (http://glycam.ccrc.uga.edu/AMBER/index.html) and this structure was energy minimized using AMBER[21]. The H-N-C2-H torsion angle connecting the N-acetyl group to the glucose backbone, 142^o is consistent with the crystal structural torsion angle of 140^o (Pardi et al., 1984). However the glycosidic torsion angles -60/60 deviate significantly from X-ray crystal data[26]. Here we initially chose a structure consistent with the X-ray data (-80/110) for the intra-disaccharide glycosidic torsion angles and generated structures by incrementing the alkylation torsion angles, ϕ and ψ, in 10^o increments from 0^o to 360^o. Figure 2.5-a shows the contour level diagram of the structural deviation factor, *f*, for residue I as a function of the alkylation torsion angle.

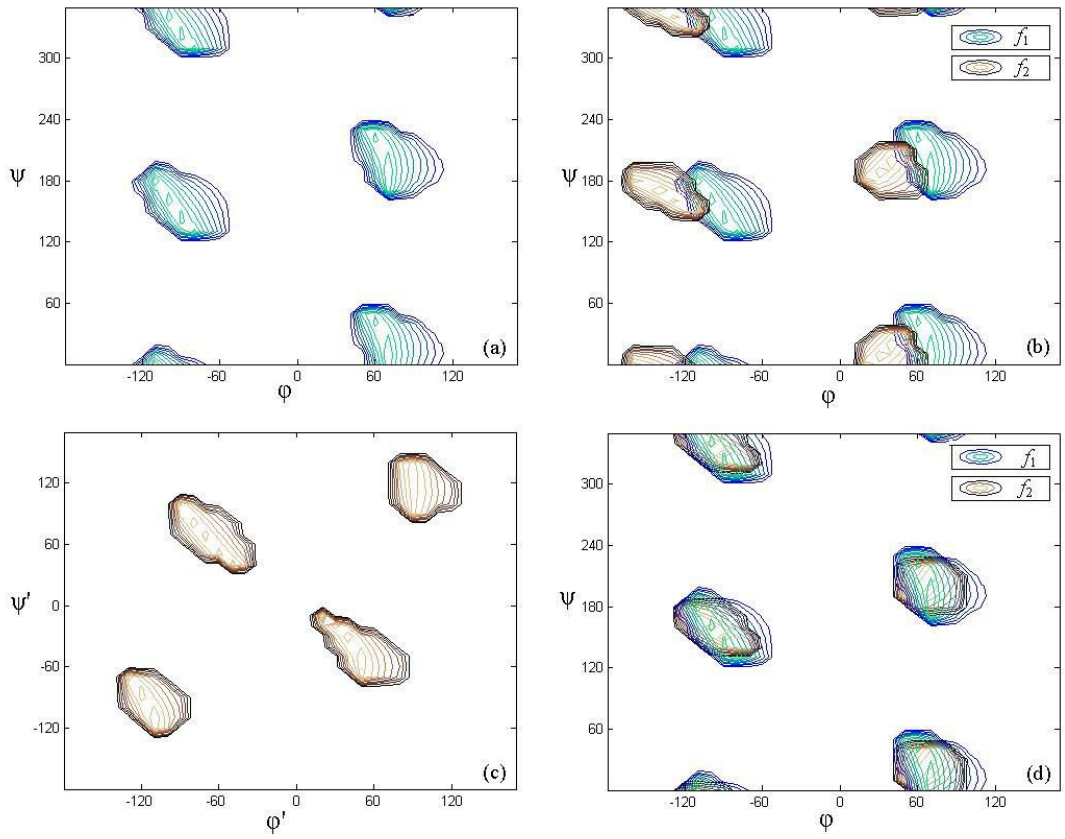


Figure 2.5 Contour diagrams for the structural deviation factors, *f*, as the function of the torsion angles (a) shows *f* for residue I as the function of sugar-chain torsion angles. (b) shows the superimposed diagram of f_2 using data from residue II (black) as a function of sugar-chain torsion angles superimposed on f_1 using data from residue I (blue) (c) shows the *f* for residue II as the function of the glycosidic torsion angles with the terminal ones fixed at -100/160 (d) shows the superimposed diagram of *f* from residue I and II after the optimization of the gycosidic torsion angles

From this figure we see that there are four possible torsion angle sets that are consistent with RDC and CSA offset data for the first sugar ring. One of these (80/20) proves unreasonable based on energy calculations (see Figure 2.6), and another (-100/320) is moderately high in energy and would direct the saccharide portion into the hydrophobic portion of the bicelle. After excluding these, two possible torsion angle regions are left. They are around -110/160 and 70/190. Both fall in low energy regions of an energy calculation diagram[17] (Figure 2.6).

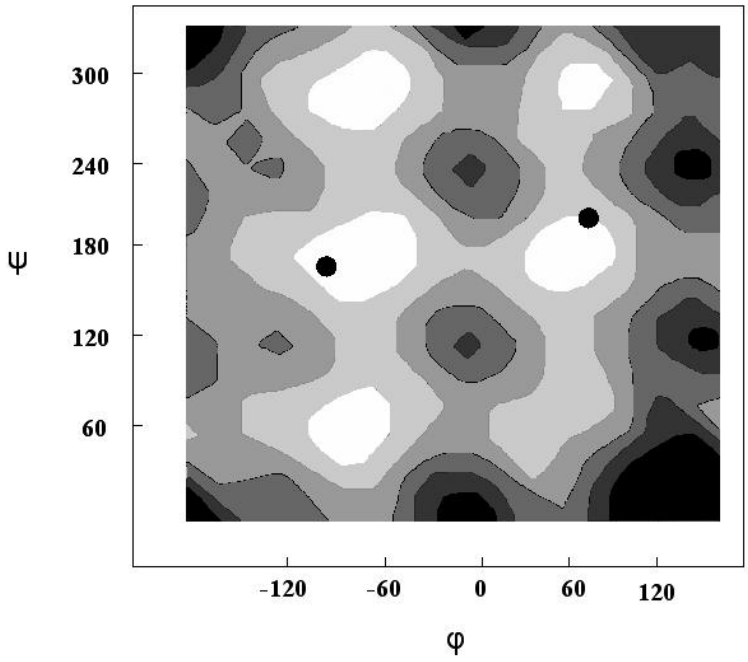


Figure 2.6 Experimental solutions of the molecular conformations (·) superimposed on a potential energy map as the function of terminal torsion angles generated by AMBER

We also calculated the structural deviation factor from residue II, f_2, using the intra-disaccharide gycosidic torsion angles as found in the crystal structure and varying the sugar-chain torsion angles. The contour diagram for f_2 is shown in black overlaid on f_1 in gray (Figure 2.5-b). The agreement is not ideal. While, this may reflect imperfections in the model, such as the assumption of axially symmetric motion or rigidity of the disaccharide unit, we decided to first explore the possibility of slight alterations in the intra-sugar glycosidic torsion angle. To do this, we fixed the terminal torsion angles at one of the two possibilities (-110/160) and calculated *f* values for ring II as a function of the glycosidic torsion angles φ' and ψ' (Figure 2.5-c). Again there are four possible torsion angle sets based on the calculation. None of these agree with the crystal structure set. Three of the four torsion angle sets lead to energetically unreasonable glycosidic conformations. But the fourth represents a torsion angles set (-70/70) that is very close to the -60/60 set of the original energy minimized structure. The same process for

optimization of the glycosidic torsion angles was performed with the terminal torsion angles fixed at 70/190. This also gave an allowed set of intra-disaccharide glycosidic torsion angles near -60/60. Given the two possible terminal torsion angle sets and one possible intra-disaccharide glycosidic torsion angle set, two possible conformations of the $(GlcNAc)_2$-OBu disaccharide, as anchored to a bilayer analog, exist. These conformations are shown in Figure 2.7.

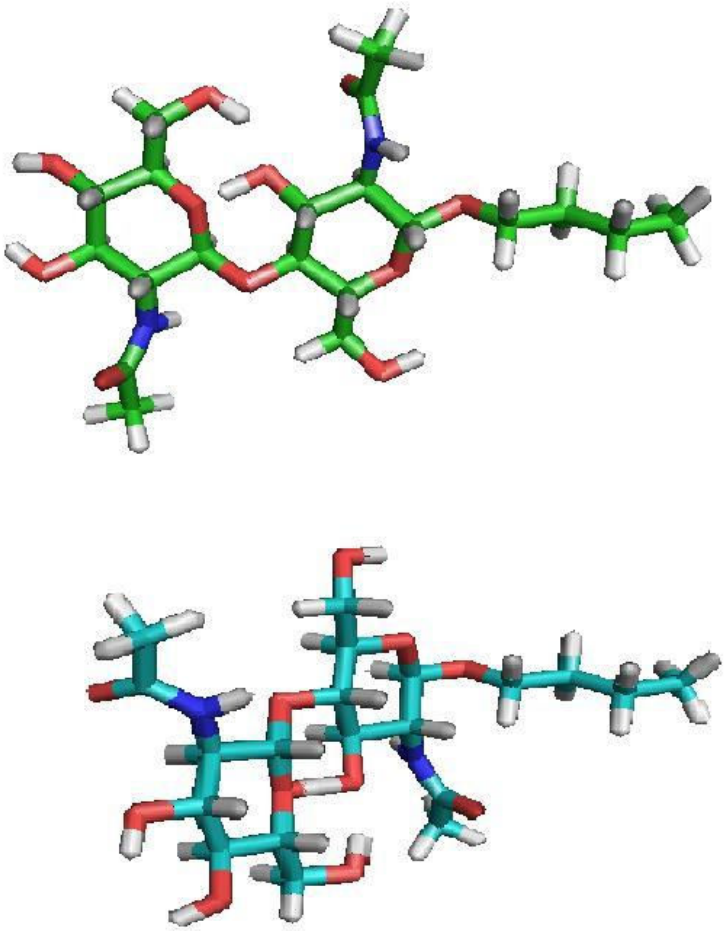


Figure 2.7 Two possible conformations of $(GlcNAc)_2$-OBu.

The order parameter values calculated for each residue, for each of the above geometries, are given in the table 2.1. For conformation II, the order parameters are nearly identical for both rings. This is consistent with modeling based on a nearly rigid intra-ring glycosidic bond. For conformation I, the order parameters are slightly different for each ring, and in fact slightly larger for the more remote ring. To the extent that we would expect the more remote ring to be less ordered we would exclude conformation I. The differences are however minor and may well be within experimental/computational errors given our assumptions of axial symmetry.

Table 2.1 The value of order parameters for the two possible conformations with the intra-disaccharide glycosidic torsion angles optimized at -70/70.

Order parameters	Conformation I (φ=-100,ψ=160)	Conformation II (φ=70,ψ=190)
S_{DD}(I)	0.0041±0.0002	0.0063±0.0010
S_{CSA}(I)	0.0040±0.0002	0.0061±0.0010
S_{DD}(II)	0.0055±0.0002	0.0061±0.0001
S_{CSA}(II)	0.0056±0.0002	0.0060±0.0004

2.5 Conclusion

Hence we are able, with a small amount of NMR orientational information, to select reasonable conformational representations of an alkylated disaccharide bound to a lipid bilayer-like surface. This is in itself not surprising; there have been several other similar accomplishments reported in the literature[27-29]. However, in the case presented here, the information comes from just two $^{13}C_2$-N-acetyl groups that replaced natural N-acetyl groups in an alkylated oligosaccharide. We have also provided a novel way of assigning resonances from these groups based on measuring differential isotope incorporation by mass spectrometry.

The introduction and assignment of isotopically labeled acetyl groups is important because of the widespread distribution of N-acetyl groups in naturally occurring oligosaccharides and the difficulties normally encountered in synthesizing analogs of these oligosaccharides in isotopically labeled forms. Developing methodology applicable to acetyl replacement in gangliosides, N-glycosides, and glycosaminoglycans may provide routes to structural characterization of these molecules as they exist in disease-relevant complexes with proteins that recognize these molecules.

References

(1) Haltiwanger, R. S.; Lowe, J. B. Role of glycosylation indevelopment. *Annual Review of Biochemistry* **2004**, *73*, 491-537.

(2) Helenius, A.; Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry* **2004**, *73*, 1019-1049.

(3) Dube, D. H.; Bertozzi, C. R. Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery* **2005**, *4*, 477-488.

(4) Rudd, P. M.; Wormald, M. R.; Dwek, R. A. Sugar-mediated ligand-receptor interactions in the immune system. *Trends in Biotechnology* **2004**, *22*, 524-530.

(5) Wormald, M. R.; Petrescu, A. J.; Pao, Y. L.; Glithero, A.; Elliott, T.; Dwek, R. A. Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography, and molecular modelling. *Chemical Reviews* **2002**, *102*, 371-386.

(6) Duus, J. O.; Gotfredsen, C. H.; Bock, K. Carbohydrate structural determination by NMR spectroscopy: Modern methods and limitations. *Chemical Reviews* **2000**, *100*, 4589-+.

(7) Martin-Pastor, M.; Bush, C. A. Refined structure of a flexible heptasaccharide using H-1-C-13 and H-1-H-1 NMR residual dipolar couplings in concert with NOE and long range scalar coupling constants. *Journal of Biomolecular Nmr* **2001**, *19*, 125-139.

(8) Colebrooke, S. A.; Blundell, C. D.; DeAngelis, P. L.; Campbell, I. D.; Almond, A. Exploiting the carboxylate chemical shift to resolve degenerate resonances in spectra of C-13-labelled glycosaminoglycans. *Magnetic Resonance in Chemistry* **2005**, *43*, 805-815.

(9) Yi, X. B.; Venot, A.; Glushka, J.; Prestegard, J. H. Glycosidic torsional motions in a bicelle-associated disaccharide from residual dipolar couplings. *Journal of the American Chemical Society* **2004**, *126*, 13636-13638.

(10) Raman, R.; Sasisekharan, V.; Sasisekharan, R. Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chemistry & Biology* **2005**, *12*, 267-277.

(11) Linhardt, R. J.; Toida, T. Role of glycosaminoglycans in cellular communication. *Accounts of Chemical Research* **2004**, *37*, 431-438.

(12) Lipsitz, R. S.; Tjandra, N. Carbonyl CSA restraints from solution NMR for protein structure refinement. *Journal of the American Chemical Society* **2001**, *123*, 11065-11066.

(13) Lipsitz, R. S.; Tjandra, N. Residual dipolar couplings in NMR structure analysis. *Annual Review of Biophysics and Biomolecular Structure* **2004**, *33*, 387-413.

(14) Prestegard, J. H.; Bougault, C. M.; Kishore, A. I. Residual dipolar couplings in structure determination of biomolecules. *Chemical Reviews* **2004**, *104*, 3519-3540.

(15) Bendiak, B.; Fang, T. T.; Jones, D. N. M. An effective strategy for structural elucidation of oligosaccharides through NMR spectroscopy combined with peracetylation using doubly C-13-labeled acetyl groups. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **2002**, *80*, 1032-1050.

(16) Zaia, J. Mass spectrometry of oligosaccharides. *Mass Spectrometry Reviews* **2004**, *23*, 161-227.

(17) Hare, B. J.; Rise, F.; Aubin, Y.; Prestegard, J. H. C-13 Nmr-Studies of Wheat-Germ-Agglutinin Interactions with N-Acetylglucosamine at a Magnetically Oriented Bilayer Surface. *Biochemistry* **1994**, *33*, 10137-10148.

(18) Horton, D.; Weckerle, W. Synthesis of 3-amino-2,3,6-trideoxy-D-ribo-hexose hydrochloride. *Carbohydrate Research* **1976**, *46*, 227-235.

(19) Ruckert, M.; Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society* **2000**, *122*, 7793-7797.

(20) Woods, R. J. GLYCAM_04.

(21) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005**, *26*, 1668-1688.

(22) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, *25*, 1605-1612.

(23) Kafetzopoulos, D.; Martinou, A.; Bouriotis, V. Bioconversion of Chitin to Chitosan - Purification and Characterization of Chitin Deacetylase from Mucor-Rouxii. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 2564-2568.

(24) Methacanon, P.; Prasitsilp, M.; Pothsree, T.; Pattaraarchachai, J. Heterogeneous N-deacetylation of squid chitin in alkaline solution. *Carbohydrate Polymers* **2003**, *52*, 119-123.

(25) Oas, T. G.; Hartzell, C. J.; Dahlquist, F. W.; Drobny, G. P. The Amide N-15 Chemical-Shift Tensors of 4 Peptides Determined from C-13 Dipole-Coupled Chemical-Shift Powder Patterns. *Journal of the American Chemical Society* **1987**, *109*, 5962-5966.

(26) Muraki, M.; Harata, K.; Sugita, N.; Sato, K. Protein-carbohydrate interactions in human lysozyme probed by combining site-directed mutagenesis and affinity labeling. *Biochemistry* **2000**, *39*, 292-299.

(27) Tian, F.; Al-Hashimi, H. M.; Craighead, J. L.; Prestegard, J. H. Conformational analysis of a flexible oligosaccharide using residual dipolar couplings. *Journal of the American Chemical Society* **2001**, *123*, 485-492.

(28) Martin-Pastor, M.; Bush, C. A. The use of NMR residual dipolar couplings in aqueous dilute liquid crystalline medium for conformational studies of complex oligosaccharides. *Carbohydrate Research* **2000**, *323*, 147-155.

(29) Lycknert, K.; Maliniak, A.; Widmalm, G. Analysis of oligosaccharide conformation by NMR spectroscopy utilizing H-1,H-1 and H-1,C-13 residual dipolar couplings in a dilute liquid crystalline phase. *Journal of Physical Chemistry A* **2001**, *105*, 5119-5122.

CHAPTER 3

CONFORMATIONAL PREFERENCES OF CHONDROITIN SULFATE OLIGOMERS USING PARTIALLY ORIENTED NMR OF ^{13}C-ACETYL LABELED OLIGOMERS[2]

Abstract

A new method is presented for the retrieval of information on the conformation of glycosaminoglycan oligomers in solution. The method relies on the replacement of acetyl groups in isolated native oligomers with ^{13}C labeled acetyl groups and the extraction of orientational constraints from residual dipolar couplings (RDCs) and chemical shift anisotropy (CSA) offsets observed in NMR spectra of partially oriented samples. A novel method for assignment of resonances based on the correlation of resonance intensities with isotope ratios determined from mass spectrometric analysis is also presented. The combined methods are used in conjunction with more traditional NMR structural data to determine the solution structure of a pentasaccharide, GalNAc6*S*(β1-4)GlcA(β1-3)GalNAc4*S*(β1-4)GlcA(β1-3)GalNAc4S-ol, derived by enzymatic hydrolysis of chondroitin sulfate. The geometry derived is compared to that for similar molecules that have been reported in the literature and prospects for use of the new types of data in the study of protein-bound oligosaccharides are discussed.

3.1 Introduction

Chondroitin sulfates (CSs) are widely distributed in extracellular matrices, and at the surfaces of mammalian cells. Here, in addition to their structural role in connective tissues, they play crucial roles in neural development and regeneration, wound healing, infection, growth

[2]Yu F, Wolff JJ, Amster IJ, Prestegard JH. *J. Am. Chem. Soc*. 2007; 129(43):13288-13297
Reprinted here with permission of publisher.

factor signaling, morphogenesis and cell division.[1-7] They are also receptors for various pathogens.[8] Despite their ubiquitous nature and important roles, structural characterization of chondroitin sulfates, and other glycosaminoglycan (GAG) polymers has been challenging.[9] Here we present a new approach to structural characterization based on a novel strategy for introducing ^{13}C labels into acetyl groups of isolated CS oligomers and observation of orientation dependent NMR parameters from these groups.

Part of the reason for the challenge in obtaining structural data is the heterogeneity of these polymers and their derived oligomers. Chondroitin sulfate is a linear, highly sulfated, polysaccharide composed of repeating disaccharide units.[9] Like other members of the glycosaminoglycan (GAG) family, each unit contains an N-acetylated sugar and an acidic sugar; in CS these sugars are galactosamine and glucuronic acid. While such structures would seem to be simple, the galactosamine residues are differentially sulfated at C4 and/or C6 carbons and sometimes the glucuronic acid is sulfated at the 2 position, leading to significant structural diversity.

NMR, in particular proton NMR, has been used to characterize some CS oligomers in the past [10,11], and this would be the natural choice for characterization in solution, However, proton NMR spectra are very complex, even when structurally well-defined oligomers can be isolated from digests of the normally heterogeneous polymers.[12] Recently, several structures of octasaccharides have been reported based on observations of nuclear Overhauser effects (NOEs), but interpretation is dependent to some extent on computer simulations.[13-15] The complexity of analysis only increases when the conformation of a CS-oligomer in complex with a CS-binding protein is of interest, and lines are broadened by this interaction. Resolution of proton spectra then degrades, and use of resonance assignment strategies based on three-bond coupling of ^{1}H

spin pairs becomes impractical. ^{13}C NMR spectra, and multidimensional versions of these spectra, offer higher resolution,[16] but without enrichment in ^{13}C, sensitivity is usually inadequate. There have been a few cases of isotope enrichment of GAGs, some based on fully synthetic methods [17] and some based on biosynthetic methods.[18-21] However, synthesis is time consuming and separation of the heterogeneous products produced in biosynthetic schemes can be difficult. Also, even when enriched products are produced, parameters directly connected with ^{13}C observation do not offer the wealth of structural information that normally comes from NOEs in ^{1}H spectra.

Here we present a partial solution to the sensitivity and structural information limitations of ^{13}C based approaches. The solution offered relies on introduction of observable ^{13}C labeled sites by replacing native acetyl groups in chondroitin sulfate oligomers, derived from natural products, with ^{13}C-labled acetyl groups. The resulting ^{13}C spectra return structural information through the orientational dependence of residual dipolar couplings (RDCs) and chemical shift anisotropy (CSA) offsets. We illustrate an assignment strategy for resonances in these spectra based on correlation of enrichment levels at various sites, as seen in NMR spectra, and isotope ratios seen in mass spectra of oligomers. Finally, we illustrate a solution structure determination for a simple CS pentasaccharide, GalNAc6*S*(β1-4)GlcA(β1-3)GalNAc4*S*(β1-4)GlcA(β1-3)GalNAc4S-ol, using a combination of ^{1}H and ^{13}C data. We will refer to this molecule as CS5.

Homogeneous oligomers suitable for structural chatracterization have been produced from native polymers a number of times before by digestion with hyaluronidase and separation by HPLC.[12,22] We shall follow these procedures in preparation of our starting material. The method we have chosen for the introduction of ^{13}C capitalizes on the presence of N-acetyl groups on every second sugar in chondroitin sulfate. These groups can be removed by either synthetic

or biosynthetic methods, [23-25] providing a route to reintroduction of ^{13}C enriched N-acetyl groups. We have previously explored the possibility of introducing ^{13}C labeled acetyls in a dimer of N-acetylglucosamine using a carefully controlled base catalyzed hydrolysis reaction to remove the original N-acetyl group, and reaction with ^{13}C acetic anhydride to reacetylate the dimer.[26] However, for chondroitin sulfate, the O-sulfate groups are not stable in strong base solution,[27] and improved methodology is required. The new methodology is based on hydrazinolysis. Hydrazinolysis of glycosaminoglycans has been used to aid in deaminative cleavage of GAGs, in particular heparan sulfate (HS).[23,28] It was also used to radio isotope label the N-acetyl groups in the degraded oligomers in order to monitor their separation.[23,29] The methodology proved to retain O-sulfation patterns and produce product in high yield.

The introduction of multiple acetyl labels does introduce an NMR resonance assignment problem, particularly if the complexity of proton spectra prevents unambiguous connection of acetyl resonances to resonances of a particular ring in an oligosaccharide. We will take advantage of the fact that hydrazinolysis, while efficient, does occur at different rates for different acetyl sites. If reacetylation with ^{13}C acetyl groups is done without complete hydrazinolysis, each acetyl contains a different $^{12}C/^{13}C$ isotope ratio and these ratios can be used to assign resonances. In NMR spectra the ratios are reflected in the different intensities of acetyl resonances; in MS/MS spectra the ratios are reflected in intensities of mass peaks for ^{12}C and ^{13}C isotopic peaks of the fragment ions. MS/MS of the oligomer of interest provides fragments in a manner that allows placement of various N-acetylgalactosamines in the oligomer sequence and correlation of NMR derived isotope ratios with MS derived isotope ratios allows assignment of the NMR resonances.

^{13}C labeled acetyl groups carry a significant amount of structural information. When ^{13}C is in both the methyl and carbonyl groups, there is a ^{13}C-^{13}C residual dipolar coupling that can be measured. Dipolar couplings are orientation dependent though the $(1-3\cos^2\theta)$ function that characterizes through-space coupling of magnetic dipoles. The function averages to zero for molecules that sample orientations uniformly as they tumble in solution, but when partially ordered, the function does not average to zero, and the resulting contributions to splittings of resonances returns information on the angle, θ, between the ^{13}C-^{13}C bond vector and the magnetic field.[30]

CSA offsets provide additional orientational data. The offsets arise from the fact that chemical shifts are not isotropic parameters, but depend on orientation of a molecule in the magnetic field. The carbonyl group has a particularly large shielding anisotropy.[31] Again the anisotropy averages to zero under uniform sampling of orientations, but in partially ordered systems they manifest themselves as offsets of resonances from their isotropic positions.[32] Interestingly, CSA offsets can be written in the form of two RDC-like equations.[33] The vectors used in describing the offsets are at angles of approximately 20° and 70° relative to the C-C bond vector. Thus, CSA offsets are quite complementary to the RDC data.

A great deal of structural information is required to fully determine the conformation of an oligosaccharide in solution, or in complex with a protein. If determination were to be done solely on the basis of orientational information from RDCs and CSA offsets, at least five pieces of structural information per rigid entity would be needed. We will have six pieces of data from the RDCs and CSAs of acetyls in the pentasaccharide studied here, (GalNAc6*S*(β1-4)GlcA(β1-3)GalNAc4*S*(β1-4)GlcA(β1-3)GalNAc4S-ol), but these are distributed over three residues. Two of these are rigid to a reasonable approximation, but even these are connected by flexible

glycosidic linkages. In applications to protein bound CS oligomers the plan would be to supplement acetyl data with principal order parameters determined from the protein and with oligomer structural constraints derived from modeling. However, in the test case presented here (modeling of the preferred conformation in solution), we will supplement the acetyl data with a number of ^{13}C-^{1}H RDCs from ring carbons, ^{1}H-^{1}H RDCs within sugar rings, NOE distance constraints from ^{1}H-^{1}H trans-glycosidic pairs and J coupling constraints from H2 to HN. The structure determined, using an iterative calculation of alignment parameters from an assumed structure and optimization of the structure using a simulated annealing protocol, gives an opportunity to understand the solution conformational preferences of CS5.

3.2 Materials and Methods

The sodium salt of chondroitin sulfate A (~70%, the balance is chondroitin sulfate C) from bovine trachea, hyaluronidase from sheep testes (type V), anhydrous hydrazine, hydrazine sulfate, iodic acid, hydroiodic acid, Sephadex G-15 resin, and pentaethylene glycol monododecyl ether (E_5 C_{12}) were purchased from Sigma-Aldrich Co (St. Louis MO). Acetic anhydride-$^{13}C_4$ was purchased from Cambridge Isotope Laboratories (Andover, MA) and iatrobeads were purchased from Iatron Laboratories (Tokyo, Japan). A pre-packed strong anion exchange (SAX) column was purchased from VWR Scientific (Rochester, NY).

3.2.1 Preparation of Chondroitin Sulfate Oligomers

Preparation of chondroitin sulfate oligomers was adapated from a literature procedure.[22] Chondroitin sulfate A (1g) was digested with 100mg hyaluronidase (451units/mg) in 50mM phosphate buffer, pH 6, containing 150mM NaCl at 37^{o}C for 48h. The digest was crudely separated into fractions I-III on an iotrabeads column with H_2O/ACN as an elution solvent. Fraction II, which was confirmed by MS to mostly consist of tetramer and pentamer, was treated

with 0.1ml $NaBH_4$ in H_2O to reduce the terminal sugars to the corresponding galactitol form, and then fraction II was subfractionated on a strong anion exchange (SAX) HPLC column (4.6X250 mm) using a linear NaCl gradient from 0 to 700 mM over a 50 min period at a flow rate 3.0 mL/min. The separation was monitored by absorption at 215 nm. Fractions V and VI (elution times of 85 mins and 87 mins) were collected, desalted on a Sephadex G-15 column (1.5X100cm), and lyophilized.

Fraction VI was analyzed by mass spectrometry and NMR. It proved to be a pentamer of mass 1221.20. This is consistent with the following molecule: GalNAc6*S*(β1-4)GlcA(β1-3)GalNAc4*S*(β1-4)GlcA(β1-3)GalNAc4S-ol, denoted CS5. Terminal residues and positions of sulfates were determined by NMR analysis as described below.

3.2.2 Isotopic Labeling of CS Oligomers

Isotope labeling at the galactosamine acetyls relied on the selective removal of N-acetyl groups using a procedure adapted from Shaklee et al,[24] followed by re-acetlyation with ^{13}C-labeled acetic anhydride. The reduced CS5 pentamer (5mg) was dissolved in 1mL anhydrous hydrazine containing 1% (w/v) hydrazine sulfate. The solution was sealed in a 1ml vial and heated at 90°C for 5hrs. After that, the hydrazine was evaporated in vacuum overnight, the dry sample was dissolved in 3ml H_2O, and moved to a 15mL separatory funnel where 500 μL 0.2M HIO_3 solution in H_2O was added slowly at room temperature to convert the uronic acid hydrazide back to a uronic acid residue. The excessive HIO_3 was then reduced by adding 0.2M HI/H_2O in 10 μL increments until no more I_2 formed in the aqueous layer. 5ml diethyl ether was added to extract I_2 formed. The ether layer was removed and 2ml more diethyl ether was added to extract any remaining I_2. The aqueous solution was collected and the diethyl ether layers were

washed with two 1mL portions of water to retrieve any additional deacetylated product. All the aqueous solutions were combined and lyophilized.

The deacetylated sample was dissolved in 1mL water and $NaCO_3$ was added until the pH reached 11. 10μL acetic anhydride-$^{13}C_4$ was added at 0°C and the solution was stirred for 2hrs. The product was purified on a Sephadex G-15 column (1.5X100cm) monitored by absorption at 215nm.

3.2.3 Preparation of NMR Samples

All the NMR samples were buffered to the same pH value of 7 with 20mM phosphate. For extraction of residual dipolar couplings and chemical shift anisotropy offsets from ^{13}C spectra of the N-acetyl group, a sample in an anisotropic medium was prepared by dissolving isotopically labeled CS5 in 500uL of 10% (w/w) alkylethyleneglycol detergent, E_5C_{12}, and 20μL hexanol in aqueous phosphate buffer. 5mg tetramethylammonium bromide was added as a chemical shift reference. The isotropic reference sample was prepared by simply adding extra hexanol (2uL) to the anisotropic sample to transform the aligned medium into an isotropic one. For the measurement of C-H RDCs and H-H RDCs from other parts of the sugar residues, an anisotropic sample was prepared by dissolving non-labeled CS5 in a lower concentration of E_5C_{12} ,medium (4%). Another 5mM unlabeled sample in D_2O was made for the isotropic measurement of the same C-H and H-H couplings. For 1D HN-H2 J coupling measurements an additional, 5mM unlabeled sample was prepared in 10% D_2O/H_2O.

3.2.4 NMR Spectroscopy

All NMR experiments were recorded on a Varian Inova spectrometer operating at 800MHz for ^1H and 200 MHz for ^{13}C. For ^{13}C spectra, proton homonuclear decoupling was achieved using WALTZ-16 with a power level of 30 dB. The recycling delay, d1, was set as long

as 30 seconds to allow the magnetization to fully relax before each scan. 1D spectra were recorded with a spectral width of 50 kHz and acquisition time of 1s. The FIDs accumulated over 3 hrs were apodized with a Gaussian weighting function (0.5) before Fourier transformation.

Heteronuclear single quantum coherence (HSQC) spectra were acquired without proton decoupling in the indirect dimension at spectral widths of 5000Hz for the direct proton dimension and 8000Hz for the indirect carbon dimension using 80 scans per t1 increment to achieve a time domain matrix of 4096 x 512 complex points. This was linear predicted to 4096 x 800, apodized in both dimensions with a 90ºshifted sinebell, and zero filled to 4096 x 2048. The couplings were extracted directly from the indirect frequency domain.

The H-H homonuclear couplings were measured by using the magnitude form of the constant time COSY (CT-COSY) experiment.[31] The data were acquired using the same sequence as the phase sensitive CT-COSY except processed in a magnitude way ($spectra = \sqrt{\mathrm{Re}^2 + \mathrm{Im}^2}$). The spectrum was acquired with a spectral width of 1800 Hz for both direct and indirect proton dimensions, 8 scan per t1 increment, 1024 X 64 complex points. Spectra were then linear predicted to 1024 x 128, apodized in both dimensions with a 90ºshifted sinebell, and zero filled to 1024 x 256. The constant time delays were 40-340 ms with a 20 ms increment. The curve fitting of data as a function of delay was performed using the curve fitting tool in MATLAB.

The NOESY spectra used to supplement structural data were acquired with a spectral width of 1800Hz for both direct and indirect proton dimensions, 64 scan per t1 increment, 512 x 400 complex points. These were linear predicted to 1024 x 512, apodized in both dimensions with a 90ºshifted squared sinebell, and zero filled to 1024 x 1024 .

The 1D proton spectra for HN-H2 J coupling measurements were acquired using a Watergate sequence with a spectral width of 12 kHz and acquisition time of 1s. Assignment of proton resonances of ring protons were based on standard COSY, TOCSY and NOE spectra.

3.2.5 Mass Spectrometry

Experiments were performed with a 7 T Bruker Apex IV QeFTMS (Billerica, MA) fitted with an Apollo II ESI source and a CO_2 laser for infrared multiphoton dissociation (IRMPD). Solutions of each oligossaccharide were made at a concentration of 0.1 mg/mL in 50:50 methanol:H_2O and ionized by nanospray using a pulled fused silica tip (model# FS360-75-15-D-5, New Objective, Woburn, MA). The sample solutions were infused at a rate of 10 μL/hour. All oligosaccharides were examined in negative ion mode. For MS/MS experiments, precursor ions were isolated in the external quadrupole and then transferred to the ICR analyzer cell. The isolation/cell fill was repeated up to 6 times per acquisition. The selection of the precursor ion was further refined by using in-cell isolation with a coherent harmonic excitation frequency (CHEF) event.[34] Precursor ions were then fragmented by IRMPD inside the analyzer cell. 24 acquisitions were signal averaged per mass spectrum. For each mass spectrum, 512 Kpoints were acquired, padded with one zero fill, and apodized using a sinebell window. All fragments are reported using the Domon and Costello nomenclature.[35]

3.2.6 Structure Calculation.

Structure determination was accomplished using a combination of the program, REDCAT (residual dipolar coupling analysis tool) to estimate alignment parameters[36] and XPLOR-NIH to optimize the structure using a simulated annealing protocol.[37,38] REDCAT was originally designed for the analysis of RDCs. Coordinates for pairs of atoms (taken from a trial structure) are entered along with a maximum coupling calculated from nuclear moments

separated by 1A (-60,400 for a C-H pair, -15,200 for a C-C pair and -240,200 for a H-H pair) and an experimental RDC. Error limits are also entered and are usually set to 10% of the range of measured couplings, a number that adequately accounts for structural variations in geometry that are not well modeled by common force fields. The program returns a large set of allowed order tensor solution, and a best set that can be interpreted in terms of principal elements describing the level and asymmetry of order and Euler angles describing the orientation of the principal frame relative to the original molecular frame. It can also back-calculate RDCs for a given structure and order parameter set.

The program was not originally intended to handle CSA offsets. However, based on the similarity of equations relating observables to orientational constraints for RDCs and CSA offsets, it is possible to use REDCAT for the analysis of CSA offsets as well. The dependence of dipolar coupling on orientation is described by equation 3.1.

$$D = D_{\max}(r)^{-3}\sum_{ij} S_{ij}\cos(\theta_i)\cos(\theta_j) \qquad \text{Eq. 3.1}$$

Here D_{max} is a constant that depends on the properties of nuclei *k* and *l*. The S_{ij} are elements of an order matrix written in an arbitrary molecular frame, and the $cos(\theta_i)$ are angles of the internuclear vector relative to molecular axes, i,j.

The dependence of the CSA offset on orientation is described by equation 3.2.

$$\delta_{CSA} = \left(\tfrac{2}{3}\right)\sum_{ijk} S_{ij}\cos(\theta_{ik})\cos(\theta_{jk})\delta_{kk} = \left(\tfrac{1}{3}\right)\begin{bmatrix} 2\delta_{xx}\sum_{ij} S_{ij}\cos(\theta_{ix})\cos(\theta_{jx}) \\ +2\delta_{yy}\sum_{ij} S_{ij}\cos(\theta_{iy})\cos(\theta_{jy}) \\ +2\delta_{zz}\sum_{ij} S_{ij}\cos(\theta_{iz})\cos(\theta_{jz}) \end{bmatrix} \qquad \text{Eq. 3.2}$$

Here the δ_{kk} are principal elements of the chemical shift tensor, and the $cos(\theta_{ik})$ are angles of the CSA principal frame axes *k* relative to molecular frame axes *i,j*. Comparing the expression for RDCs and CSAs a similar $cos(\theta)$ dependent form is observed, and the three terms in equation 2

can be expressed as three pseudo RDC entries that result in a measured average offset. The D_{max} entered for the RDCs is simply replaced by twice the chemical shift anisotropy component and the CSA principal tensor vectors replace coordinates for RDC internuclear vectors. The values used originated with data on peptides systems, were converted to Hz at 200Hz/ppm (-29,600Hz, -400Hz, 30,000Hz)[39] and assumed to be the same for all acetyl groups. In proteins variations can reach 20% of the range.[31] Using the average tool integrated in REDCAT, the CSA data were input in the form of RDCs. An entry of "AVG" is placed for the first two pseudo RDC values and the observed chemical shift offset in Hz is input for the last value. The back-calculated CSA offset (using the back-calculation tool in REDCAT) is equal to the average value of the three pseudo RDCs.

One additional subtly in application is that scaling of both the entry and D_{max} can be used to vary the weight of various types of data in the REDCAT calculation. This is particularly important when experimental data have very different magnitudes. In our case ^{13}C-^{13}C RDCs were scaled up by a factor of 10 to achieve similar weighting of different types of data, or data with differences in experimental precision. The observed CSAs and chemical shift tensor elements δ_i were not scaled as the conversion to Hz put them on in appropriate range.

REDCAT calculations were carried out for a number of structures selected from low energy conformers identified in the literature. In each case an RMSD of back calculated RDCs and CSAs relative to experimental data was returned. The structure with minimum RMSD was chosen as the starting model in the following XPLOR-NIH refinement, and the order parameters determined for that structure were converted to the anisotropy (Da) and Rhombicity (Rh) parameters typically used in XPLOR-NIH. They are related to the REDCAT order parameters as follows: $Da = D_{max} * ½ Szz$ and $Rh = 2/3\eta$.

XPLOR-NIH can accomplish simulated annealing under restraints from experimental data including RDCs, CSA offsets, NOE derived distances and NH-H2 J coupling. A modification was made to the topology file for carbohydrates[40] to allow application to a chondroitin sulfate oligomer. Additional force field parameters needed were obtained from the GLYCAM website (http://glycam.ccrc.uga.edu/AMBER/index.html). The starting model from the above REDCAT calculation was heated to temperatures of 3,500 K, and slowly cooled in steps of 25K to 300 K. The process was repeated 2000 times and 10 structures with the lowest energies were selected and analyzed in VMD.[41] Again, RDCs and CSAs were back-calculated using REDCAT on Xplor refined structure. A new set of alignment parameters was calculated and simulated annealing in XPLOR-NIH was repeated.

3.3 Results

3.3.1 Selection of the Oligomer.

The particular chondroitin sulfate oligomer to be studied here was selected based on its modest size and modest degree of structural heterogeneity. Long oligomers may well provide too much of a challenge in resolution and assignment of labeled acetyl groups and too short an oligomer would limit utility as a basis for future studies of protein-condroitin sulfate interactions. The primary structure of the oligomer, GalNAc6*S*(β1-4)GlcA(β1-3)GalNAc4*S*(β1-4)GlcA(β1-3)GalNAc4S-ol was deduced from the NMR resonances found in ^{13}C-^{1}H HSQC spectra of the native material. The residues were classified as glucuronic or N-acetylgalactosamine by patterns in TOCSY/COSY spectra and the sequence of residues was established by trans-glycosidic NOE patterns. The sulfation sites are easily identified by a characterisitic downfield shift (~0.5 ppm) of the H4 resonance for C4 sulfation and the H6 resonance for C6 sulfation. Considering the inhomogeniety of the polymer precursor (~30% chondroitin sulfate C), the occurrence of some

6-sulfation is not unexpected. A pentamer is, however, unanticipated as hyaluronidase is expected to have specificity for hydrolysis after an N-acetylgalactosamine residue leading to a series of oligomers with an even number of residues. We note that, particularly under long periods of hydrolysis, the fraction of odd oligomers rises, suggesting that production of odd numbered oligomers may be due to the presence of minor amounts of other hydrolytic enzymes.

3.3.2 Isotopic Labeling

Our strategy for the introduction of ^{13}C-labeled acetyl groups to the chondroitin sulfate pentamer (CS5) involves two steps: deacetylation and reacetylation. The deacetylation strategy, which is shown in Figure 3.1, was based on the hydrazinolysis mechanism. Hydrazinolysis has proven to be a high yield reaction that leaves the O-sulfate and carboxyl groups that characterize glycosaminoglycans intact.[24] However, two side reactions possible in the hydrazinolysis of GAGs are of some concern. One is the reaction between hydrazine and the aldehyde exposed at the reducing end of an oligosaccharide on opening of the pyranose ring.[42] To prevent this side reaction, the sugar was reduced to the corresponding galactol before the hydrazinolysis reaction. Leaving this as a galactol in the final product also helps reduce the complexity of the NMR spectra due to elimination of the α/β anomeric equilibrium at the reducing end of the oligomer. The other side reaction results in the formation of a hydrazide derivative of the carboxyl moieties in the glucuronic acid residues.[24] To remove products of this second reaction the initial hydrazinolysis product was treated with HIO_3 to convert the uronic acid hydrazides back into uronic acid residues.

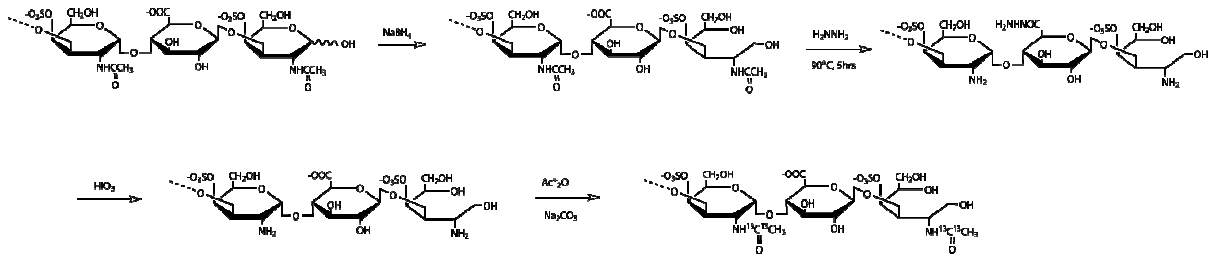


Figure 3.1 ^{13}C enrichment strategy for isolated CS oligomers

Conversion of the amino sugar products back to ^{13}C-labeled N-acetyl sugars using ^{13}C-acteic anhydride was nearly quantitative. To facilitate assignment, however, it was desirable to have different $^{13}C/^{12}C$ isotopic ratios for each N-acetyl group. To accomplish this, we choose to partly remove the native N-acetyl groups at the hydrazinolysis stage, leaving different amounts of ^{12}C-actetyls. Due to the chemical environment difference for each N-acetyl group, the deacetylation ratio is different. Here we optimized the hydrazinolysis reaction conditions (90°C for 5hrs) to achieve deacetylation ratios ranging from 90% to 30%. Deacetylation was followed by fully reacetylating the amino sugars with 99% ^{13}C-acetic anhydride. The overall recovery of CS was ~50%.

3.3.3 ^{13}C-^{13}C RDCs and Carbonyl Carbon CSA Offsets

The carbonyl regions of the ^{13}C NMR spectra for the CS5 sample in an isotropic phase and anisotropic phase of an aqueous solution of E_5C_{12} are presented in Figure 3.2. Both spectra were acquired with the resonance from tetramethyl ammonium bromide as chemical shift reference. Three doublets with 50-60 Hz splittings are observed in both spectra. These correspond to the three N-acetyl groups with splittings arising from the coupling to ^{13}Cs in the directly bonded acetyl methyl groups. In comparing spectra it is clear that the doublets in the aligned spectrum (2b) have both shifted and changed splittings. The doublet labeled I is shifted upfield by 54.6 Hz, the doublet labeled II is shifted upfield by 20.5 Hz, and the doublet labeled

III is shifted downfield by 18.2 Hz. The splittings of the doublets have changed by 7.3 Hz for doublet I, 4.8 Hz for doublet II and 4.9 Hz for doublet III. The change in splitting arises from a residual dipolar coupling and the change in chemical shift arises from a chemical shift anisotropy offset. The changes are typical of the effects of incomplete averaging of anisotropic parameters in aligned media. Changes can also be seen in the acetyl methyl region of the spectrum (not shown). However, the changes in splittings are identical to those measured from the carbonyl resonances, and the changes in chemical shift are negligible because of the small chemical shift anisotropy of the methyl carbons. The data, as summarized in Table 3.1, clearly contain structural information on the average orientation of each acetyl group, but this information is not useful without assignment of doublets to specific acetyls.

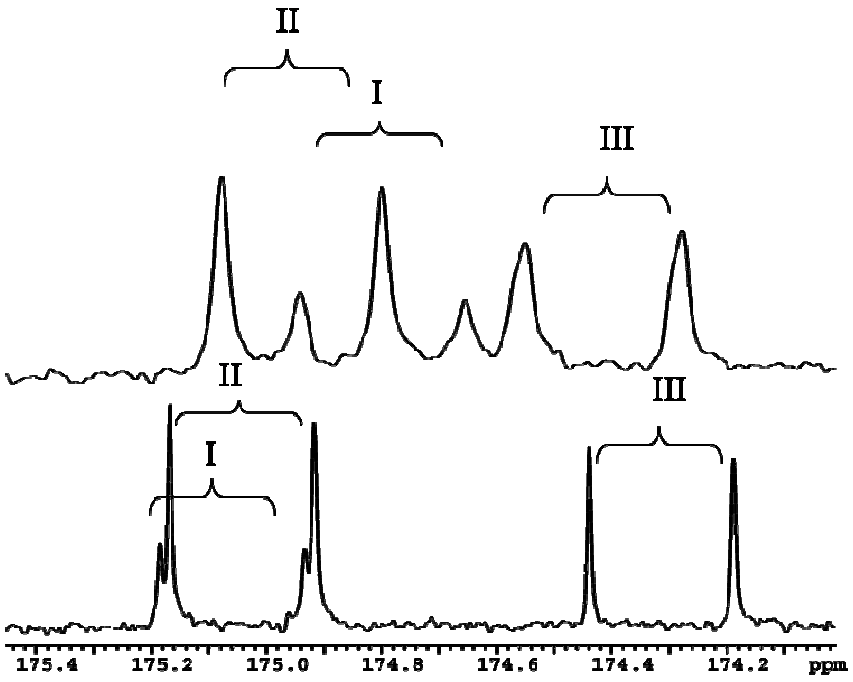


Figure 3.2 Carbonyl region of ^{13}C spectra for an isotopically labeled CS pentamer in 10% $C_{12}E_5$/hexanol/water. The bottom one corresponds to the spectra acquired for sample in isotropic phase and the top one corresponds to the spectra acquired for the sample in anisotropic phase(with extra hexanol). Three doublets with different intensity represent three acetyl groups with different isotopic ratios.

Table 3.1 Observed and unified ^{13}C-^{13}C RDCs and carbonyl carbon CSA offsets

Doublet/Residue	Observed (10% C12E5)		Unified (4% C12E5)	
	CSA (Hz)	RDC (Hz)	CSA (Hz)	RDC (Hz)
II / A	-20.5	4.8	-1.6	0.39
I / C	-54.6	7.3	-4.4	0.58
III / E	18.2	4.9	1.5	0.39

3.3.5 Assignment of ^{13}C Spectra.

Assignments of carbonyl resonances would normally be undertaken by transfer of carbonyl ^{13}C magnetization to the amide proton two bonds away using an INEPT sequence and then transfer to other sugar ring protons using TOCSY techniques. Since all sugars containing acetyls are N-acetylglactosamines, they would need to be distinguished by sequentially linking them with NOE or HMBC transfers across the glycosidic linkages. While this would be likely to work for simple solution samples, the small coupling between the carbonyl carbon and the amide proton make this a difficult procedure and one that is not likely to work for the broader lines encountered in protein complexes, or the broader lines encountered when CS5 is aligned in a E_5C_{12} medium.

Here we use a new assignment strategy based on the different isotopic labeling ratio of each N-acetyl. This was introduced previously in our work on an N-acetylglucosamine disaccharide.[26] In the ^{13}C spectra of Figure 3.2, the difference in the intensity of the doublets reflects the different extent of ^{13}C label introduction. Integration of the peaks shows the relative extent of ^{13}C label introduction for the three N-acetyl groups I:II:III to be 3:10:8. To decide which doublet belongs to which N-acetyl group, we correlate these intensity ratios with mass ratios found for each sugar residue in MS/MS spectra of the CS5 oligomer.

Expansions of the mass regions of MS spectra corresponding to B_2, Y_1 and Y_3 ions of CS5 are shown in Figure 3.3. These were collected in a negative ion mode on an FT/MS spectrometer after fragmentation using IRMPD. The B_2 and Y_1 ions correspond to fragments containing just a single N-acetylglactosamine residue. In these regions two sets of mass peaks were observed. The monoisotopic peaks at lower mass, M, represents ^{12}C containing products. The peaks two mass units higher, M+2, primarily represent those molecules containing an acetyl group with two of ^{13}C atoms. Their ratio is, in principle, correlated with the NMR intensity for a particular doublet seen in Figure 3.2. It must be noted, however, that the natural abundance ^{13}C isotopes from all non-labeled carbons also contribute to the M+2 peaks. Therefore, to extract an accurate isotopic labeling ratio, the contribution from natural abundance ^{13}C was simulated and subtracted from the peaks labeled M+2. The isotopic labeling ratio was then calculated as $R = I'_{M+2}/(I_M + I'_{M+2})$, where I'_{M+2} represents the intensity of the corrected M+2 peaks and I_M represents the intensity of M peaks. The calculated results suggest the isotopic ratio for the N-acetyl groups in residue A (B_2 ion) and residue E (Y_1 ion) are 61% and 72% respectively.

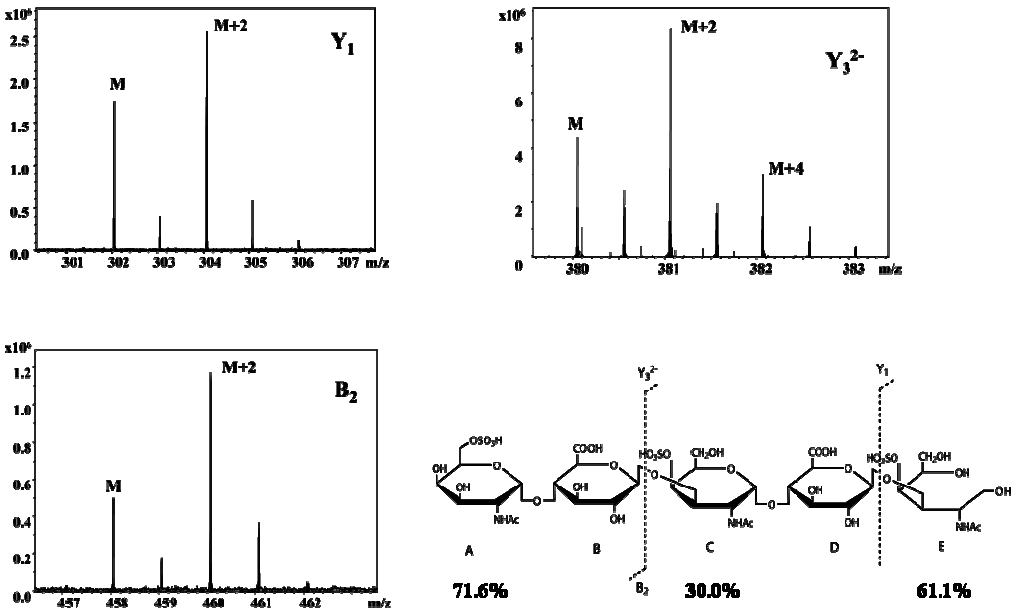


Figure 3.3 MS/MS spectra of Y_1 Y_3^{2-} and B_2 ions. The different isotopic ratio for each N-acetylated residue is reflected in intensities of peaks M/M+2/M+4. After fitting the data, the isotopic labeling ratios calculated for the three N-acetyl groups A, C and E were 72%, 31% and 61%.

Deducing the ^{13}C content of residue C is more complicated. The Y_3 product ion reveals the presence of two partially labeled N-acetyl groups. Here we observe three sets of peaks, M, M+2 and M+4. The peak M represents the molecules with two ^{12}C-acetyl groups. Peak M+2 represents the Y_3 product with only one ^{13}C-acetyl group (in either residue C or residue D). Peak M+4 represents the product with two ^{13}C-acetyl groups. The natural abundance refinement was also performed to improve the accuracy of the isotopic ratio calculation. Since we already knew the isotopic ratio for residue E to be 72%, the isotopic ratio for residue C can be calculated from any pair among the M, M+2 and M+4 peaks. Here we performed the calculation through these three combinations, M & M+2, M+2 & M+4, and M & M+4. The isotopic ratios from a fit of the data are 29%, 32%, and 33%, or an average of 31% for residue C. In summary, the isotopic labeling ratios for the three N-acetyl groups were calculated as 72%, 31% and 61%. Correlating

this information with the intensity ratios seen in the ^{13}C spectra, we can easily make the assignments shown in Table 3.1. For the anisotropic spectra, the assignment was performed in a completely analogous way.

3.3.6 Complementary Structural Information

While the ^{13}C labeled N-acetyl groups provide useful structural information with high sensitivity and resolution, to independently determine the molecular conformation of CS5 in solution, at least five pieces of RDC and CSA data would be required for each rigid unit.[43] This is less of a problem in the case of structural characterization of protein-associated carbohydrate because protein bound carbohydrates share the order parameters Szz and η with the protein, and these can usually be obtained from protein spectra. This reduces the number of pieces of structural information needed for each unit to three, and any restriction on glycosidic torsion angles from modeling the interaction will reduce the number further. Here, since we are going to probe the structure of CS5 in free solution, some other structural information, including C-H RDCs, H-H RDCs, and NOEs are employed as complimentary data.

The one bond C-H RDCs were obtained from a series of sensitivity enhanced constant time HSQC spectra (SE-CT-HSQC) taken under partially oriented and isotropic conditions. These RDCs are listed in Table 3.2. For a sample in the 4% E_5C_{12} medium, values ranging from +9.3 to -10.0 Hz are observed. The spread of the data reflect the different orientations of each sugar residue in the oligomer. For the glucuronic acid, the RDCs for the C1-H1, C2-H2, C3-H3 and C4-H4 are of similar size and sign because of the nearly parallel orientation of these vectors. The same thing is true for the galactosamine residue except for the RDC of the equatorial C4-H4 vector. While the data are of high quality, the near degeneracy in vectors still makes it difficult to get five independent RDC measurements.

Table 3.2 Complementary structural information.

Vector	RDC (Hz)	Vector	RDC (Hz)
Residue A		Residue D	
C1H1	6.1	C1H1	7.9
C2H2	5.8	C2H2	7.7
C4H4	-10.0	C3H3	9.3
C5H5	7.3	C4H4	8.1
H1H2	0.6	H1H2	1.5
Residue B			
C1H1	4.4	CH3(A)	-1.3
C2H2	5.0	CH3(C)	-2.1
C3H3	4.8	CH3(E)	-1.3
C4H4	4.2		
H1H2	-0.2		NOE(Å)
Residue C		H1(A)-H4(B)	2.20
C1H1	7.9	H1(B)-H3(C)	2.47
C2H2	5.7	H1(C)-H4(D)	2.21
C3H3	7.1		
C4H4	-6.5	Residue	JH2-HN(Hz)
H1H2	1.8	A	9.7
H3H4	0.8	C	9.5
		E	8.5

The vectors connecting proximate protons of the sugar rings are not typically parallel to any of the one-bond C-H vectors, and represent additional independent pieces of data. The homonuclear H-H coupling can be accurately measured by a simple magnitude form of the constant time COSY (CT-COSY) experiment. Here several spectra with various constant time delays were recorded and data were fit to an equation relating the intensity ratios of cross-peaks and auto-peaks to a function of the constant time delay (t).

$$\frac{I_{cross}}{I_{auto}} = \left| A \tan\left(\pi\left(J + D\right)t\right)\right| \qquad \text{Eq. 3.3}$$

The extracted 1H-1H RDC are also listed in Table 3.2. Due to the interference from the E_5C_{12} signal, only the coupling for H1-H2 of each residue and the H3-H4 of residue C are reported. The signs of couplings cannot be determined from the CT-COSY experiments unless large scalar

couplings of known sign exist for the pair of protons. Here, the magnitude of the scalar couplings for H1-H2 and H3-H4 are more than 8 Hz, which is large enough to resolve issues of sign ambiguity.

Besides the RDC data, other complementary structural data included NOE data on the distance between proton pairs across the glycosidic linkages and the H2-HN scalar coupling between the H2 ring proton of galactosamine residues and the amide proton of the acetyl amide bond. The glycosidic NOE distances function as additional constraints on the glycosidic torsion angles, and the H2-HN couplings function as constraints on the relative orientation of N-acetyl groups to the galactosamine rings. The NOE distance constraints are shown in the Table 3.2.

The configurations of the C2-N torsion angles in the N-acetylgalactosamine residues are particularly critical to our use of data from acetyl groups in our structure determination. These torsions can be constrained using scalar couplings from HN to H2 of the N-acetylgalactosamine residues. The three H2-HN J couplings measured for residue A, C and E are 9.7 Hz, 9.5Hz and 8.5Hz respectively (shown in table 3.2). The relative smaller value for residue E compared with the other two might be due to the loss of ring structure for this residue. Using the Karplus equation with coefficients from Wang & Bax[44], the calculated torsion angle for the H-N-C2-H2 angle for residue A and C would be 169° and 164°, values close to the trans configuration found in previous MD calculations.[13] The values were thus used to justify selection of a trans orientation for acetyls in all starting structures and an average value of 9.6 Hz was used to restrain the N-acetyl groups in refinement steps.

3.3.7 Merging CSA and RDC data

Because of resolution problems in observing proton spectra, the CSA offset and ^{13}C-^{13}C RDC data were acquired under different conditions than C-H and H-H RDC data, namely 10%

and 4% E_5C_{12}. To perform the structural analysis using all the information simultaneously, the two sets were unified, assuming that the only difference would be a scaling of the measured coupling and CSA offsets. In an acetyl group, the dipolar coupling between the methyl protons and the methyl carbon (D_{CH3}) actually are a redundant measure of the orientation of the C-C bond obtained from the carbonyl carbon – methyl carbon RDC (D_{CC}). This is a result of the fast rotation of the methyl group about the C-C axis. Therefore, the D_{CH3} can be expressed in terms of D_{CC} according to the equation, ${}^1D_{CC} = {}^1D_{CH_3}(-3\gamma_C/\gamma_H)(r_{CH}^3/r_{CC}^3)$.[30] Based on this relationship, the three D_{CH3} values observed in 4% E_5C_{12} were transformed to three D_{CC}' as 0.38, 0.61 and 0.38 Hz respectively. When compared with the corresponding D_{CC} observed in 10% E_5C_{12}, the scale factors obtained were 12.6, 12.0 and 12.9 respectively. An average of 12.5, was then used to scale all 10% E_5C_{12} orientational data to 4% E_5C_{12} conditions. The unified RDCs and CSAs are shown in table 3.1.

3.3.8 Structure Determination

One of the most widely practiced approaches to the utilization of a variety of experimental data in a structure determination is to represent those data in terms of pseudo-energy functions added to the molecular energy representations of molecular dynamics programs. A search can then be done for a minimum energy structure using simulated annealing or Monte Carlo methods. The XPLOR-NIH program has been developed to accommodate a variety of NMR data, including RDCs and CSA offsets.[32,45] However, in order to utilize the RDC and CSA data in XPLOR-NIH, an input of anisotropy and rhombicity parameters for the molecular order is required. In applications to proteins, values for these parameters are normally obtained from analysis of the distributions of large numbers of data. However, with the limited number of data available for oligosaccharides, parameter estimation from distributions is not

reliable. Therefore we adopted a hybrid approach in which a second program, REDCAT (residual dipolar coupling analysis tool), is used to estimate alignment parameters using an initial structure, and then XPLOR-NIH and REDCAT are used iteratively to arrive at a refined structure. REDCAT calculates the best least squares fit of an order matrix to experimental data, given an initial structure. It too can handle RDCs and CSAs in evaluations of structural models. The order tensor REDCAT returns is transformed to a principal frame and the principal order and asymmetry parameters (Szz and η, simply related to anisotropy and rhombicity) are extracted. REDCAT also provides a very convenient way to evaluate a structure by back calculating RDCs and CSAs for comparison to experimental data.

To begin the process reasonable initial structures must be selected. Recently P. Kamerling and coworkers carried out a combined NMR/computational study on a set of eight CS oligomers[11]. This work does not include the molecule studied here, but the set does include similar linkages. Based on energy diagrams for the common linkages we can select reasonable starting conformers for our analysis. These diagrams typically have several local minima. Table 3.3 shows the torsion angles for the local minima of the first three glycosidic bonds of our pentamer (the fourth, to the reduced sugar was not considered because of the expected flexibility of the open ring structure). Structures were initially generated using the GLYCAM utilites[46] and conformations were adjusted to the required torsions using PYMOL[47]. The error limits for data used in REDCAT were initially adjusted to values large enough to allow order matrix solutions for each conformation (typically the error limit for C-C RDCs and C=O CSAs were set to 5 Hz), and the best set of order parameters was used to calculate an RMSD relative to experimental data. The same procedure was repeated for the random structural sampling around the minimum RMSD region. This time, the errors limits were reduced (typically 2.5 Hz for C-C RDCs and

CSAs). All combinations for the first two glycosidic bonds were evaluated. The best conformation for the first two glycosidic bonds was then chosen and all possibilities for the third torsion were explored in combination with this conformation. The results showed the structure with torsion angles in the region b-a-b to achieve a minimum RMSD. The final structure (structure-1) has glycosidic torsion angles of 270/250 305/105 and 270/250 and an RMSD of 0.96 Hz. The final numbers we obtained for order tensor Szz and η are 3.87e-4 and 7.17e-2 respectively.

Table 3.3 The glycosidic torsion angles φ/ψ sampled for the structural approximation by REDCAT.

	α	β	γ	δ	ε
A-B β(1-4)	260/190	280/250	280/70	60/240	80/270
B-C β(1-3)	280/90	280/130	60/110	270/280	
C-D β(1-4)	260/210	280/250	270/60	60/240	80/270

The order parameters from the initial structure were transformed to the anisotropy (Da) and rhombicity (Rh) parameters used in XPLOR-NIH and a simulated annealing routine implemented. Ten structures with minimum final energies were collected from 2000 generated structures. The low energy structure (structure-2) had torsion angles 291/220 298/97 and 301/219. REDCAT analysis of this structure gave an S_{zz} of 3.10E-4 and an η of 2.10E-1 with a back-calculated RMSD relative to experiment of 0.82 Hz. The new order parameters were used in a second round of XPLOR-NIH refinement. The new best structure (structure-3) torsion angles were 290/239 299/109 and 296/234 with an Szz of 3.26E-4 and η = 2.01E-4. This last

refinement resulted in changes of angles by less than 20° and changes in order parameters by less than 5%. Convergence was judged adequate at this point. Figure 3.4 shows the superposition of the ten minimum energy structures from the last refinement. Table 3.4 summarizes the course of the structure search. Back-calculated values of CSA offsets and RDCs have been produced by REDCAT at each stage.

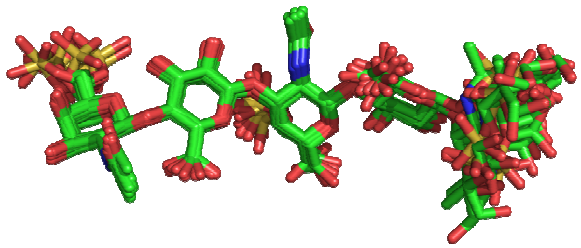


Figure 3.4 Superposition of 10 structures with minimum energy from the XPLOR-NIH simulated annealing calculation.

Table 3.4 Back-calculated values of CSA offsets and RDCs produced by REDCAT at each structure refining stage.

Residue	Vector	Experimental (Hz)	Structure-1 (Hz)	Structure-2 (Hz)	Structure-3 (Hz)
A	C1H1	6.1	4.6	5.3	6.0
	C2H2	5.8	5.8	5.5	6.6
	C4H4	-10.0	-11.7	-11.5	-9.2
	C5H5	7.3	6.7	5.6	6.4
	H1H2	0.6	0.5	0.5	0.5
B	C1H1	4.4	4.1	4.9	3.7
	C2H2	5.0	5.7	5.4	4.6
	C3H3	4.8	6.0	5.2	5.6
	C4H4	4.2	3.4	4.8	4.0
	H1H2	-0.2	0.1	0.1	-0.1
C	C1H1	7.9	7.2	6.8	7.3
	C2H2	5.7	7.3	6.8	7.1
	C3H3	7.1	6.9	6.5	6.8
	C4H4	-6.5	-5.7	-6.7	-7.2
	H1H2	1.8	0.7	0.7	0.8
	H3H4	0.8	0.4	1.8	1.7
D	C1H1	7.9	6.6	7.8	8.2
	C2H2	7.7	8.5	8.1	8.4
	C3H3	9.3	8.2	8.1	8.4
	C4H4	8.1	7.6	7.9	8.2
	H1H2	1.5	0.9	1.0	1.1
A	CC X 10	3.9	2.0	3.3	3.4
C	CC X 10	5.8	6.3	5.3	6.1
A	C(O)	-1.6	-2.8	-3.1	-2.8
C	C(O)	-4.4	-5.0	-4.6	-4.9
RMSD			0.96	0.82	0.68

3.4 Discussion

The combined data clearly lead to a reasonably well defined description of a conformation of our CS pentamer in solution. The minimum energy structures resulting from simulated annealing under experimental restraints cluster well except for residue E. The RMSD for all heavy atoms in residues A-D is 0.53Å; with residue E included the RMSD rises to 2.58Å. The increased deviation in E results from the limited number of RDCs and absence of an NOE

distance constraint between the anomeric proton of residue D. However, the loss of ring structure for this residue on reduction introduces considerable motional freedom and attempts to define a single structure would have been inappropriate in any event.

The validity of the structures depicted in Figure 3.5 is best discussed in comparison to data in the literature. A comparison of glycosidic torsion angles found in our structure and others is presented in Table 3.5. Direct comparisons are not actually possible because of the differences in sulfation, differences in environment such as the presence or absence of counter ions, or differences due to the presence or absence of protein binding (chondroitinase B). However, in each case, molecules share a β(1-3) bond between a glucuronic acid residue and an N-acetylgalactosamine residue and a β(1-4) bond between an N-acetylglactosamine residue and a glucuronic acid residue. In the table we have entered data corresponding to the analogous disaccharide linkage.

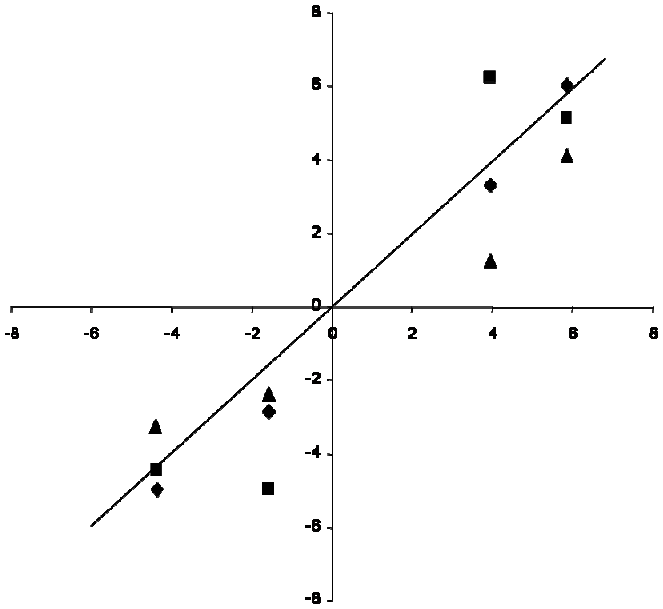


Figure 3.5 Correlation of experimental and calculated ^{13}C-^{13}C RDCs and C(O) CSA offsets for simulated docked CS structures. The NMR structure is assumed to be the correct docked structure (♦), and order parameters for this structure were used to back-calculate data for all structures. The other two docked structures are the X-ray structures from Cael et al(■) [48] and

Michel et al(▲)[49]. Residues B (glucouronic acid residues) in these two structures were superimposed on residue B in the NMR structure before back-calculating data.

Table 3.5 Comparisons of the NMR structure of CS5 to structures in the literature

Structure	Torsions	A-B β(1-4)	B-C β(1-3)	C-D β(1-4)
NMR CS5	φ	-70	-61	-64
	ψ	-121	109	-126
Xray (fiber) [48]	φ	-98	-80	-98
	ψ	-174	107	-174
Xray (chondroitinase) [49]	φ	-69*	-89	-69
	ψ	-108*	108	-108
Computation (MM3) [15]	φ	-79	-79	-78
	ψ	-110	90	-111
Computation (CHARMM25) [50]	φ	-70	-70	-70
	ψ	-120	90	-120
Computation/NMR [13]	φ	-80	-80	-80
	ψ	-110	90	-110

The earliest experimental data in the literature come from structures modeled to X-ray fiber diffraction data.[48] Coordinates for the CS4 tetramer have been deposited (PDB code 2C4S).[48] There are significant departures from the ϕ/φ angles we observe for both the β(1-4) and β(1-4) linkages (28/53 and 34/48 degrees). However, the angles in this model also depart significantly in several cases from other observations in the literature. This may well be the result of intermolecular packing of polymeric chains or the presence of Ca^{2+} in these structures. Our data for the β(1-4) linkage agree well with other structures; for example, ϕ and ψ depart by just 0/1 and 6/6 degrees from values reported by Almond and Sheehan.[50] Our values for the ϕ/φ angles of the β(1-3) linkage depart more significantly from other structures (9/19 degrees from the value given by Almond and Sheehan and 21/19 degrees from the value given by Blanchard et al[13]). However, examination of the energy maps suggests that the energy predicted for our conformer is less than 2 kcal higher that the minimum energy structure; plus the energy maps

provided are for a non-sulfated analog N-acetylgalactosamine in one case and an unsaturated analog of the glucuronic acid in the other case.

It is also important to point out that our experimental data are averages over conformations sampled by a potentially flexible molecule. We have already pointed to the fact that the effects would be severe for the E residue, which we don't attempt to model. For the rest of the residues, however, the fact that a large amount of data fits to a single representation of the conformation argues against this being a large contributor.

More important than the identification of a solution conformation for CS5 is the establishment of a protocol for the evaluation of conformations of CS oligomers bound to proteins. This is an important issue as chondroitin sulfates (and other glycosaminoglycans) are known to modulate signaling by chemokines[51] as well as play roles in modulation of neural development through interactions with growth factors.[52] The proposed protocol would rely primarily on C-C (or C-H) RDCs and CSA offsets from ^{13}C-labeled acetyl groups introduced into natural products. These are parameters that can be extracted from spectra even when resonances are broadened by strong interactions with macromolecules. The NMR/MS assignment strategy that we have introduced also allows specific assignment of resonances from multiple acetyl groups without the aid of triple or double resonance experiments that require passing magnetization though multiple scalar coupling pathways.

A key concern for potential application to protein-bound oligomers is whether the limited amount of data available from ^{13}C-^{13}C labeled acetyl groups can provide sufficient information to guide the identification of protein bound conformers. We do not as yet have data on protein bound oligomers, but we can assess our ability to distinguish sets of conformers that might be generated by docking programs. We can simulate this situation using some of the representative

conformers in table 3.5. We will mimic docking by superimposing the glucouronic acid residue between the two N-acetyl galactosamine of tetramers constructed using torsions listed in the table 3.5. We will also assume that we have obtained an alignment tensor from data on the protein that coincided with the tensor determined from data on CS5 presented above. The acetyl RDCs and CSA offsets were then back-calculated for each of the structures using this alignment tensor in the program REDCAT. The correlation plots are presented in Figure 3.5. As expected, our structure shows a very good fit. Both of the other structures show sizable deviations. While this is purely an exercise in simulation, it does illustrate an ability to select among a set of acceptable conformations based on limited data available from acetyl ^{13}C-^{13}C RDCs and CSA offsets introduced into isolated glycosaminoglycan oligomers. Given an assignment strategy that does not require observation and resolution of resonances from other NMR active sites on the residues, we believe this strategy will be applicable to rather large protein complexes and oligomers in the hexasaccaride to decasaccharide range.

References

(1) Viapiano, M. S.; Matthews, R. T. From barriers to bridges: chondroitin sulfate proteoglycans in neuropathology. *Trends in Molecular Medicine* **2006**, *12*, 488-496.

(2) Mikami, T.; Sugahara, K. The biological importance of specific sulfation of chondroitin sulfate/dermatan sulfate in their functional expression. *Trends in Glycoscience and Glycotechnology* **2006**, *18*, 165-183.

(3) Gowda, C.; Achur, R.; Muthusamy, A.; Takagaki, K. Unusually low-sulfated chondroitin 4-sulfate of human placenta and its role in placental malaria. *Trends in Glycoscience and Glycotechnology* **2004**, *16*, 407-420.

(4) Fthenou, E.; Zafiropoulos, A.; Tsatsakis, A.; Stathopoulos, A.; Karamanos, N. K.; Tzanakakis, G. N. Chondroitin sulfate A chains enhance platelet derived growth factor-mediated signalling in fibrosarcoma cells. *International Journal of Biochemistry & Cell Biology* **2006**, *38*, 2141-2150.

(5) Shannon, J. M.; McCormick-Shannon, K.; Burhans, M. S.; Shangguan, X. F.; Srivastava, K.; Hyatt, B. A. Chondroitin sulfate proteoglycans are required for lung growth and morphogenesis in vitro. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **2003**, *285*, L1323-L1336.

(6) Wang, T. W.; Sun, J. S.; Wu, H. C.; Tsuang, Y. H.; Wang, W. H.; Lin, F. H. The effect of gelatin-chondroitin sulfate-hyaluronic acid skin substitute on wound healing in SCID mice. *Biomaterials* **2006**, *27*, 5689-5697.

(7) Matani, P.; Sharrow, M.; Tiemeyer, M. Ligand, modulatory, and co-receptor functions of neural glycans. *Frontiers in Bioscience* **2007**, *12*, 3852-3879.

(8) Morugova, T.; Morugova, I. The role of dyslipidemia and glycosaminoglycans in diabetic nephropathy pathogenesis. *Diabetologia* **2005**, *48*, A382-A382.

(9) Imberty, A.; Lortat-Jacob, H.; Perez, S. Structural view of glycosaminoglycan-protein interactions. *Carbohydrate Research* **2007**, *342*, 430-439.

(10) Sugahara, K.; Tanaka, Y.; Yamada, S. Preparation of a series of sulfated tetrasaccharides from shark cartilage chondroitin sulfate D using testicular hyaluronidase and structure determination by 500 MHz H-1 NMR spectroscopy. *Glycoconjugate Journal* **1996**, *13*, 609-619.

(11) Yamada, S.; Yoshida, K.; Sugiura, M.; Sugahara, K. One-Dimensional and 2-Dimensional H-1-Nmr Characterization of 2 Series of Sulfated Disaccharides Prepared from Chondroitin Sulfate and Heparan-Sulfate Heparin by Bacterial Eliminase Digestion. *Journal of Biochemistry* **1992**, *112*, 440-447.

(12) Kinoshita-Toyoda, A.; Yamada, S.; Haslam, S. M.; Khoo, K. H.; Sugiura, M.; Morris, H. R.; Dell, A.; Sugahara, K. Structural determination of five novel tetrasaccharides containing 3-O-sulfated D-glucuronic acid and two rare oligosaccharides containing a beta-D-glucose branch isolated from squid cartilage chondroitin sulfate E. *Biochemistry* **2004**, *43*, 11063-11074.

(13) Blanchard, V.; Chevalier, F.; Imberty, A.; Leeflang, B. R.; Basappa; Sugahara, K.; Kamerling, J. P. Conformational studies on five octasaccharides isolated from chondroitin sulfate using NMR spectroscopy and molecular modeling. *Biochemistry* **2007**, *46*, 1167-1175.

(14) Mikhailov, D.; Linhardt, R. J.; Mayo, K. H. NMR solution conformation of heparin-derived hexasaccharide. *Biochemical Journal* **1997**, *328*, 51-61.

(15) Rodriguez-Carvajal, M. A.; Imberty, A.; Perez, S. Conformational behavior of chondroitin and chondroitin sulfate in relation to their physical properties as inferred by molecular modeling. *Biopolymers* **2003**, *69*, 15-28.

(16) Scott, J. E.; Heatley, F. Hyaluronan forms specific stable tertiary structures in aqueous solution: A C-13 NMR study. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 4850-4855.

(17) Martin-Pastor, M.; Canales-Mayordomo, A.; Jimenez-Barbero, J. NMR experiments for the measurement of proton-proton and carbon-carbon residual dipolar couplings in uniformly labelled oligosaccharides. *Journal of Biomolecular Nmr* **2003**, *26*, 345-353.

(18) Milton, M. J.; Harris, R.; Probert, M. A.; Field, R. A.; Homans, S. W. New conformational constraints in isotopically (C-13) enriched oligosaccharides. *Glycobiology* **1998**, *8*, 147-153.

(19) Blundell, C. D.; DeAngelis, P. L.; Day, A. J.; Almond, A. Use of N-15-NMR to resolve molecular details in isotopically-enriched carbohydrates: sequence-specific observations in hyaluronan oligomers up to decasaccharides. *Glycobiology* **2004**, *14*, 999-1009.

(20) Tracy, B. S.; Avci, F. Y.; Linhardt, R. J.; DeAngelis, P. L. Acceptor specificity of the Pasteurella hyaluronan and chondroitin synthases and production of chimeric glycosaminoglycans. *Journal of Biological Chemistry* **2007**, *282*, 337-344.

(21) DeAngelis, P. L. In *Polymer Biocatalysis and Biomaterials* 2005; Vol. 900, p 232-245.

(22) Kinoshita, A.; Yamada, S.; Haslam, S. M.; Morris, H. R.; Dell, A.; Sugahara, K. Novel tetrasaccharides isolated from squid cartilage chondroitin sulfate E contain unusual sulfated disaccharide units GlcA(3-O-sulfate)beta 1-3GalNAc(6-O-sulfate) or GlcA(3-O-sulfate)beta 1-3GalNAc(4,6-O-disulfate). *Journal of Biological Chemistry* **1997**, *272*, 19656-19665.

(23) Rabenstein, D. L. Heparin and heparan sulfate: structure and function. *Natural Product Reports* **2002**, *19*, 312-331.

(24) Shaklee, P. N.; Conrad, H. E. Hydrazinolysis of Heparin and Other Glycosaminoglycans. *Biochemical Journal* **1984**, *217*, 187-197.

(25) Grobe, K.; Ledin, J.; Ringvall, M.; Holmborn, K.; Forsberg, E.; Esko, J. D.; Kjellen, L. Heparan sulfate and development: differential roles of the N-acetylglucosamine N-deacetylase/N-sulfotransferase isozymes. *Biochimica Et Biophysica Acta-General Subjects* **2002**, *1573*, 209-215.

(26) Yu, F.; Prestegard, J. H. Structural monitoring of oligosaccharides through C-13 enrichment and NMR observation of acetyl groups. *Biophysical Journal* **2006**, *91*, 1952-1959.

(27) Fraidenraich, D.; Waisman; Cirelli, A. F. Selective O-Desulfation of Heparin in Triethylamine. *Carbohydrate Polymers* **1992**, *17*, 111-114.

(28) Maccarana, M.; Casu, B.; Lindahl, U. Minimal Sequence in Heparin Heparan-Sulfate Required for Binding of Basic Fibroblast Growth-Factor. *Journal of Biological Chemistry* **1993**, *268*, 23898-23905.

(29) Vives, R. R.; Pye, D. A.; Salmivirta, M.; Hopwood, J. J.; Lindahl, U.; Gallagher, J. T. Sequence analysis of heparan sulphate and heparin oligosaccharides. *Biochemical Journal* **1999**, *339*, 767-773.

(30) Prestegard, J. H.; Bougault, C. M.; Kishore, A. I. Residual dipolar couplings in structure determination of biomolecules. *Chemical Reviews* **2004**, *104*, 3519-3540.

(31) Burton, R. A.; Tjandra, N. Residue-specific C-13' CSA tensor principal components for ubiquitin: Correlation between tensor components and hydrogen bonding. *Journal of the American Chemical Society* **2007**, *129*, 1321-1326.

(32) Lipsitz, R. S.; Tjandra, N. Carbonyl CSA restraints from solution NMR for protein structure refinement. *Journal of the American Chemical Society* **2001**, *123*, 11065-11066.

(33) Choy, W. Y.; Tollinger, M.; Mueller, G. A.; Kay, L. E. Direct structure refinement of high molecular weight proteins against residual dipolar couplings and carbonyl chemical shift changes upon alignment: an application to maltose binding protein. *Journal of Biomolecular Nmr* **2001**, *21*, 31-40.

(34) Heck, A. J. R.; Dekoning, L. J.; Pinkse, F. A.; Nibbering, N. M. M. Mass-Specific Selection of Ions in Fourier-Transform Ion-Cyclotron Resonance Mass-Spectrometry - Unintentional Off-Resonance Cyclotron Excitation of Selected Ions. *Rapid Communications in Mass Spectrometry* **1991**, *5*, 406-414.

(35) Domon, B.; Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate Journal* **1988**, *5*, 397-409.

(36) Valafar, H.; Prestegard, J. H. REDCAT: a residual dipolar coupling analysis tool. *Journal of Magnetic Resonance* **2004**, *167*, 228-241.

(37) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **2003**, *160*, 65-73.

(38) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2006**, *48*, 47-62.

(39) Oas, T. G.; Hartzell, C. J.; Dahlquist, F. W.; Drobny, G. P. The Amide N-15 Chemical-Shift Tensors of 4 Peptides Determined from C-13 Dipole-Coupled Chemical-Shift Powder Patterns. *Journal of the American Chemical Society* **1987**, *109*, 5962-5966.

(40) Weis, W. I.; Brunger, A. T.; Skehel, J. J.; Wiley, D. C. Refinement of the influenza virus hemagglutinin by simulated annealing. *J Mol Biol* **1990**, *212*, 737-61.

(41) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33-&.

(42) Tipson, R. S. D-Ribose Hydrazone. *Journal of Organic Chemistry* **1962**, *27*, 2272-&.

(43) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Quarterly Reviews of Biophysics* **2000**, *33*, 371-424.

(44) Wang, A. C.; Bax, A. Determination of the backbone dihedral angles phi in human ubiquitin from reparametrized empirical Karplus equations. *Journal of the American Chemical Society* **1996**, *118*, 2483-2494.

(45) Tjandra, N.; Marquardt, J.; Clore, G. M. Direct refinement against proton-proton dipolar couplings in NMR structure determination of macromolecules. *Journal of Magnetic Resonance* **2000**, *142*, 393-396.

(46) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraserreid, B. Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides .1. Glycam-93 Parameter Development. *Journal of Physical Chemistry* **1995**, *99*, 3832-3846.

(47) DeLano, W. L. The PyMOL Molecular Graphics System. *http://www.pymol.org* **2002**.

(48) Cael, J. J.; Winter, W. T.; Arnott, S. Calcium Chondroitin 4-Sulfate - Molecular-Conformation and Organization of Polysaccharide Chains in a Proteoglycan. *Journal of Molecular Biology* **1978**, *125*, 21-42.

(49) Michel, G.; Pojasek, K.; Li, Y. G.; Sulea, T.; Linhardt, R. J.; Raman, R.; Prabhakar, V.; Sasisekharan, R.; Cygler, M. The structure of chondroitin B lyase complexed with glycosaminoglycan oligosaccharides unravels a calcium-dependent catalytic machinery. *Journal of Biological Chemistry* **2004**, *279*, 32882-32896.

(50) Almond, A.; Sheehan, J. K. Glycosaminoglycan conformation: do aqueous molecular dynamics simulations agree with x-ray fiber diffraction? *Glycobiology* **2000**, *10*, 329-338.

(51) Proudfoot, A. E. I. The biological relevance of chemokine-proteoglycan interactions. *Biochemical Society Transactions* **2006**, *34*, 422-426.

(52) Snow, D. M.; Mullins, N.; Hynds, D. L. Nervous system-derived chondroitin sulfate proteoglycans regulate growth cone morphology and inhibit neurite outgrowth: A light, epifluorescence, and electron microscopy study. *Microscopy Research and Technique* **2001**, *54*, 273-286.

CHAPTER 4

NMR CHARACTERIZATION OF THE INTERACTION BETWEEN CCL5 AND A CHONDROITIN SULFATE HEXAMER

4.1 Introduction

One of the primary motivations for developing the labeled glycosaminoglycan oligomers described in the previous chapters was to use them as a tool for characterizing protein-cell surface interactions that mediate important biological processes. In this chapter, we illustrate the use of these oligosaccharides in the characterization of an interaction with a small protein called CCL5 or RANTES (Regulated upon Activation, Normal T-cell Expressed and Secreted). CCL5 is a chemokine involved in directing T-cell migration. It utilizes interactions with glycosaminoglycans to maintain a concentration gradient needed to direct the migration. Chemokines are a family of small cytokines with chemotactic properties. They are small and soluble proteins (with a mass of 8-16 kDa), and are produced and released by a variety of cell types during the initial phase of host responses to injury, allergens, antigens, or invading microorganisms [1-4]. Chemokines share a sequence homology and possess four cysteines in conserved locations. These form two disulfide linkages that are key to their tertiary structure and stability. The need to form these disulfide bonds presents some special problems in properly folding over-expressed proteins for biophysical characterization.

Members of the chemokine family are categorized into four groups depending on the spacing of their first two cysteine residues. The CC chemokines or β-chemokines (MIP-1α, MIP-1β, RANTES, MCP-1, MCP-2, MCP-3) constitute the largest family and are characterized by the

presence of two contiguous cysteines near their N-terminus. The members of the CXC or α-chemokine family (Il-8, IP-10, NAP-2, PF-4) have two cysteines separated by one amino acid near the N-terminus. The C chemokines (lymphotactin-α and lymphotactin-β) contain a single cysteine near the N-terminus. A fourth group, in which members have three amino acids between the two cysteines, also has been discovered. These are known as CX_3C chemokines (or δ-chemokines). Chemokines share a structural topology that includes a flexible N-terminal region followed by a triple stranded β-sheet, arranged in a Greek key motif [5] and crossed by a C-terminal α-helix (Figure 4.1). While some chemokines are monomeric at physiological concentrations (e.g., MCP-1 and SDF-1α)[6,7], most of them associate as dimers, tetramers or higher-order polymers.

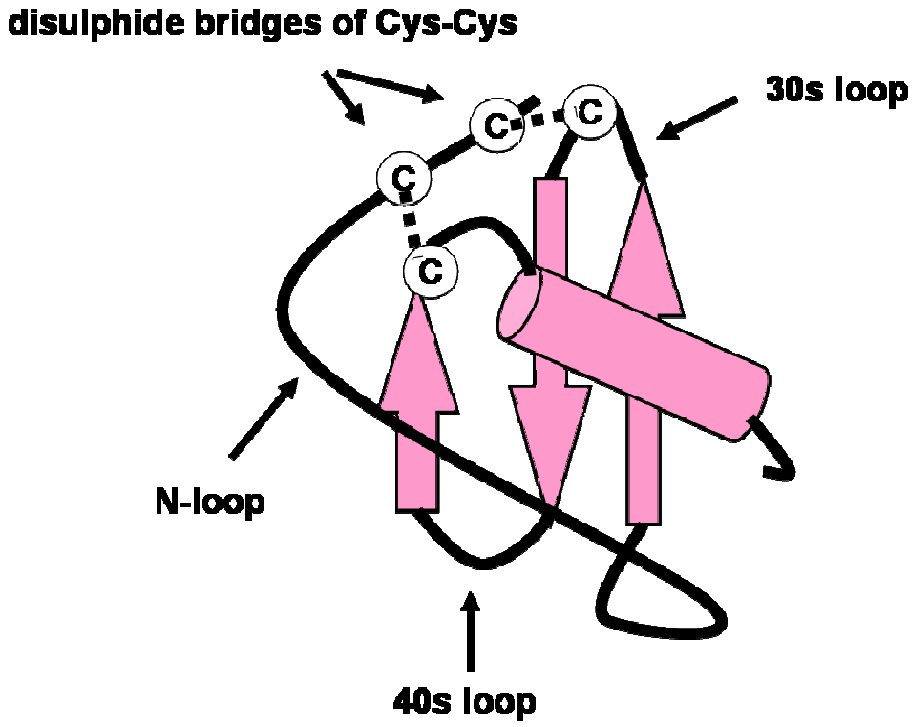


Figure 4.1 Structural model of chemokines[8]

Many of the above chemokines exert their biological effect by mediating the migration of immune cells to sites of action. CCL5, the primary target of our research mediates the migration of leukocytes via interaction with a G-protein coupled receptor (CCR5) during immuno-surveillance and inflammation[9]. Cells with receptors migrate along a chemotatic gradient towards the chemokine releasing site. During this cell transmigration process, the immobilization of chemokines by cellular surface glycosaminoglycans (GAGs) is believed to be a key step in the formation of a chemotactic gradient[10]. Figure 4.2 presents a simple illustration of this chemokine-induced migration process.

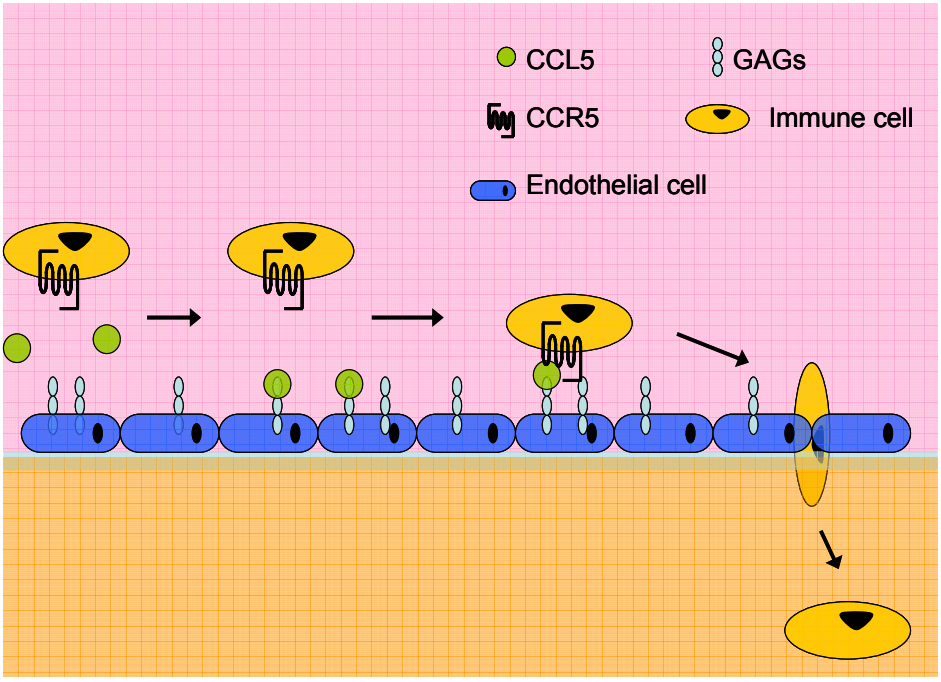


Figure 4.2 Chemokines require immobilization by cell-surface GAGs in order to initiate the transmigration process

Chemokines are known to be stored and released from cytolytic granules of T-lymphocytes complexed to GAGs[11]. In addition, CCL4 bound to tissue matrix GAGs has been shown to induce adhesion of T-cells to endothelium[12]. These observations suggest that chemokines bind to GAGs to create a high local concentration of chemokines. With such a

mechanism, chemokine accumulation is able to resist the disruption of flow in blood vessels. Interaction with endothelium GAGs may help create a chemotactic gradient and direct migration of leukocytes. GAG binding also has been shown to influence chemokine structure and promote aggregation *in vitro* [13]. GAG-induced aggregation also may increase local concentrations of chemokines and increase their *in vivo* half-life, thus protecting them from degradation. Indeed, cell surface GAGs have been reported to facilitate the binding of chemokines to their receptors[13] and the removal of GAGs by glycosidases results in the loss of intracellular Ca^{2+} mobilization[14].

The interactions of chemokines with GAGs have been addressed in a number of studies, especially for CCL5. It has been reported that the ^{44}RKNR^{47} cluster on the 40S loop of CCL5 plays a critical role in heparin binding[16]. This type of basic cluster is found in other heparin-binding proteins and has been named a BBXB motif where B is a basic amino acid.

Despite extensive studies on chemokines, the way GAGs exert their biological function during the course of chemokine-based cell recruitment still remains unclear. The objective of the work presented in this chapter is to study the interactions of CCL5 with GAGs and elucidate the role of GAGs in chemokine function. By optimizing the chondroitin sulfate protocol described in Chapter III, we have obtained two CS hexasaccharides with different sulfation patterns. To avoid the influence of the self-aggregation of wild type CCL5, an E66S mutant was prepared by following a published chemokine protocol[17]. Our initial objective was to use the method for determining the bound conformation of the CS ligand presented in chapter III for the structural characterization of N-acetylated carbohydrates[18]. This method primarily relies on ^{13}C-^{13}C RDCs and CSA offsets from ^{13}C-labeled acetyl groups introduced into CS hexamers. These parameters can be extracted from spectra while other ^{1}H observed resonances from ligands are broadened by strong interactions with macromolecules. In addition to structural information, they can give an

alignment tensor and ultimately a preferred orientation of the oligosaccharide, as it exists in the protein binding site. To complement this, an alignment tensor needs to be extracted for the ^{15}N-labeled protein in the protein-ligand complex.

We have not completely achieved this goal, but have demonstrated a stepwise ability to acquire the necessary data. We also have acquired a set of more traditional structure determination data using chemical shift mapping, NOE spectroscopy, and light scattering, that have allowed us to derive a preliminary structural model.

4.2 Materials and Methods

The clone of CCL5 and BL21(DE3) competent cells were purchased from Invitrogen Inc. The QuickChange® site-directed mutagenesis kit was from Stratagene Inc. The pET-23a vector was from Novagen Inc. The sodium salt of chondroitin sulfate A from bovine trachea (~70%, the balance was chondroitin sulfate C) and hyaluronidase from sheep testes (type V) were purchased from Sigma-Aldrich Co. The ZORBAX® strong anion exchange (SAX) columns were purchased from Agilent Technoligies Inc. (Santa Clara, CA).

4.2.1 Preparation of the CCL5 E66S Mutant

The gene for CCL5 was subcloned from the purchased clone and annealed into pET-23a vectors. An E66S mutant was then created from pET-23a using the QuickChange® site-directed mutagenesis kit with designed primers. We followed the published procedure to prepare and purify the E66S[19]. To prepare ^{15}N labeled and ^{13}C-^{15}N doubly labeled E66S protein, the starter cultures of BL21(DE3), containing the pET-23a plasmid, were grown in M9 2XYT medium, containing 100μg/ml ampicillin, at 37ºC, to give an OD_{600nm} of ~0.1. 5ml of the starter cultures were then inoculated into 1L M9 medium [1.28% (w/v) Na2HPO4·7H2O, 0.3% (w/v) KH2PO4, 0.05% (w/v) NaCl, 0.1% (w/v) ^{15}NH4Cl, 100mM MgSO4, 2mM CaCl2, 0.4% glucose(^{13}C-

glucose for preparation of ^{13}C-^{15}N doubly labeled protein), 0.5% (v/v) 200X vitamins, 0.1% (v/v) 1000x trace minerals] containing 100μg/ml ampicillin. Cultures were grown in a shaking incubator (250 rpm) at 30ºC and induced by the addition of IPTG (0.1mM final concentration) when the OD_{600nm} was ~0.8. Cells were harvested (~12hrs after induction) and stored in 30mL lysis buffer (20mM phosphate buffer, 5% (v/v) glycerol, pH 8.0). The cells were then lysed by three passages though a French pressure cell. The resulting suspension was centrifuged and the pellet (inclusion bodies) was solubilized in 10mL inclusion body extraction buffer (6M guanidine/HCl; 0.1M Tris/HCl, pH 8.0; 1mM DTT) followed by homogenization with a Dounce homogenizer. The solution was then applied to a 1.6 x 60cm column of Superdex 75 equilibrated with inclusion body extraction buffer. The column was run at a flow rate of 1mL/min (Figure 4.3). The CCL5 containing fraction was collected and the proteins were renatured by a 10-fold dilution into renaturation buffer (0.1 M Tris/HCl, pH 8.0; 10μM GSSG; 100μM GSH). Guanidine/HCl solution was then added drop-wise into the renaturation buffer using a peristaltic pump until a concentration of 0.6 M was reached to refold the protein and the solution was stirred overnight at 4ºC. The final solution appeared cloudy due to the precipitation of unfolded protein. After removing the precipitate by filtration, the solution was dialyzed against two changes of 1% acetic acid and one change of 0.1% trifluoroacetic acid (TFA) and lyophilized. To obtain pure protein for NMR spectroscopy, an additional purification step was carried out. The protein in lyophilized powder form was redissolved in 50mM sodium acetate buffer (pH 4.5) followed by filtration through a 0.22μm sterilizing filter. The solution was then applied to a 2 x 5mL HiTrap^{TM} SP HP column. The absorbed protein was eluted with a linear 0-2M NaCl gradient at a flow rate of 2mL/min (Figure 4.4). The fraction containing the protein was again dialyzed against two changes of 1% acetic acid and one change of 0.1% TFA and lyophilized.

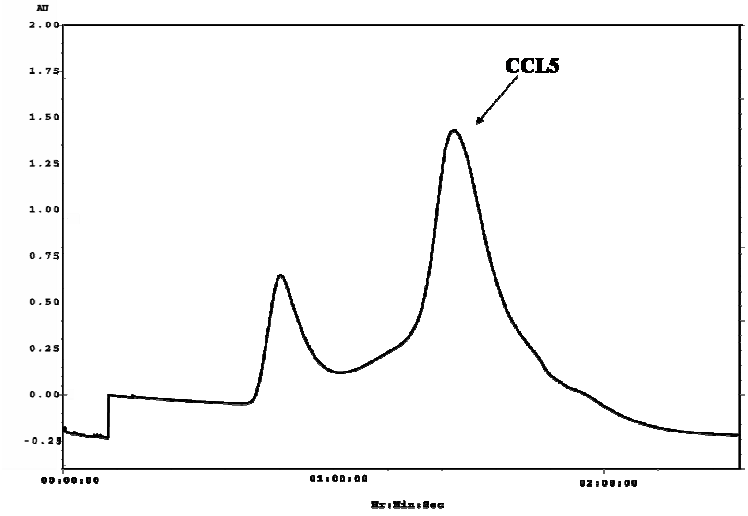


Figure 4.3 Elusion profile of denatured CCL5 on a Superdex 75 column

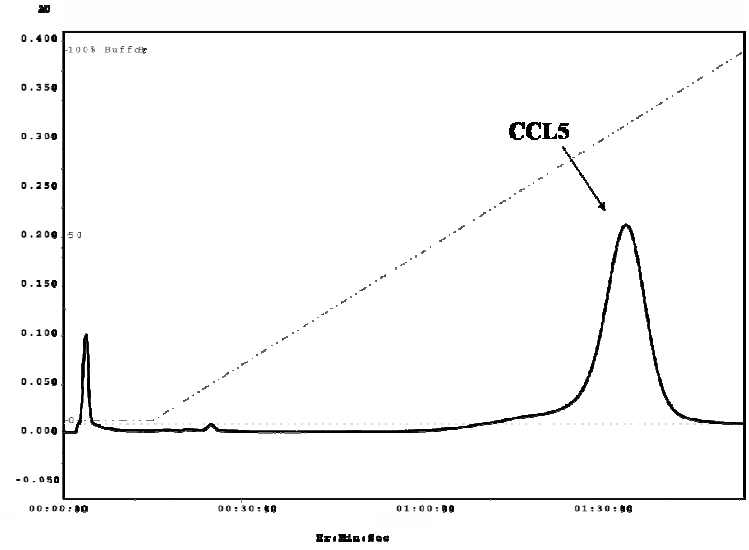


Figure 4.4 Elusion profile of renatured CCL5 on a HiTrap^{TM} SP column

4.2.2 Preparation and ^{13}C-labeling of Chondroitin Sulfate Oligomers

The chondroitin sulfate-A polymers (1g) from bovine trachea were digested with hyaluronidase (451 units/mg) at 37ºC in 10mL digesting buffer [50mM phosphate buffer (pH 7.0)] for 5 hrs. This produced a simple hydrolysis product with an unmodified non-reducing end sugar and a reducing end sugar undergoing anomeric equilibrium. The reducing end sugar was

reduced with $NaBH_4$ to eliminate spectral complexities associated with the anomeric equilibrium. The same procedure as described in chapter III was used to isolate the CS oligomers except that the crude separation of the digests was carried out by size exclusion chromatography on a G-25 column instead of a silica gel column. We also used the same protocol described in chapter III for ^{13}C-labeling the N-acetyl of the oligomers.

4.2.3 Preparation of Stretched Polyacrylamide Gels

Acrylamide gels were used to orient the protein samples for the collection of residual dipolar couplings. An acrylamide solution was prepared by 6-fold dilution of a stock of 29.2% w/v acrylamide and 0.78% w/v N,N'-methylenebisacrylamide into TBE buffer. 600μL aliquots of the solution were cast into a 0.6cm I.D. x 3.5cm height plastic tubes right after the addition of 0.1%w/v ammoniumpersulfate and 0.5% w/v tetramethylethylenediamine (TEMED). After a 6-hour polymerization, the gel was squeezed into a waterbath with the application of gentle water pressure. The gel was then washed with two exchanges of deionized water and cut into a cylinder with smooth edges. After drying in a Teflon™ coated container, the gel cylinder shrank to a 1mm x 6 mm cylinder. The dried gel was then placed into a 2.0mL Eppendorf™ microtube and 0.5mL of the NMR sample was added. The Eppendorf™ tube was placed horizontally in a refrigerator for as long as two days to make sure the gel fully swelled back to its original size. For making an aligned (stretched) gel, the swelled gel was squeezed into an open-ended 5mm NMR tube with the self-made device shown in Figure 4.5. The NMR tube was pre-treated with dimethyldichlorosilane (DMDCS) to obtain a smoother inner surface. The small funnel shown in the figure was made by cutting the sharp end of a 1mL disposable pipette tip so that the I.D. of this end was equal to 4.2 mm, the I.D. of the 5mm NMR tube. By attaching the funnel to one end of the NMR tube with PVC tubing, the swelling gel was sucked into the NMR tube from the

funnel by gently pulling a syringe from the other end of the NMR tube. The NMR tube was then sealed using a PTFE vortex plug.

Figure 4.5 Photos of the self-made device for making the stretched gel sample

4.2.4 NMR Spectroscopy

NMR resonance assignments of CS hexamers were performed with double quantum ^{1}H-^{1}H correlation spectroscopy (DQCOSY), ^{1}H-^{1}H total correlation spectroscopy (TOCSY), single quantum ^{1}H-^{13}C correlation spectroscopy (HSQC) and multiple-bond ^{1}H-^{13}C correlation spectroscopy (HMBC). The sulfation sites on the galactosamine residues were determined based on characteristic downfield shifts both in the ^{1}H dimension (~0.5ppm) and ^{13}C dimension (5~10ppm) of the HSQC spectra. Examples of the spectra for one of the CS hexamers we identified, GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol, are shown in the Figure 4.6.

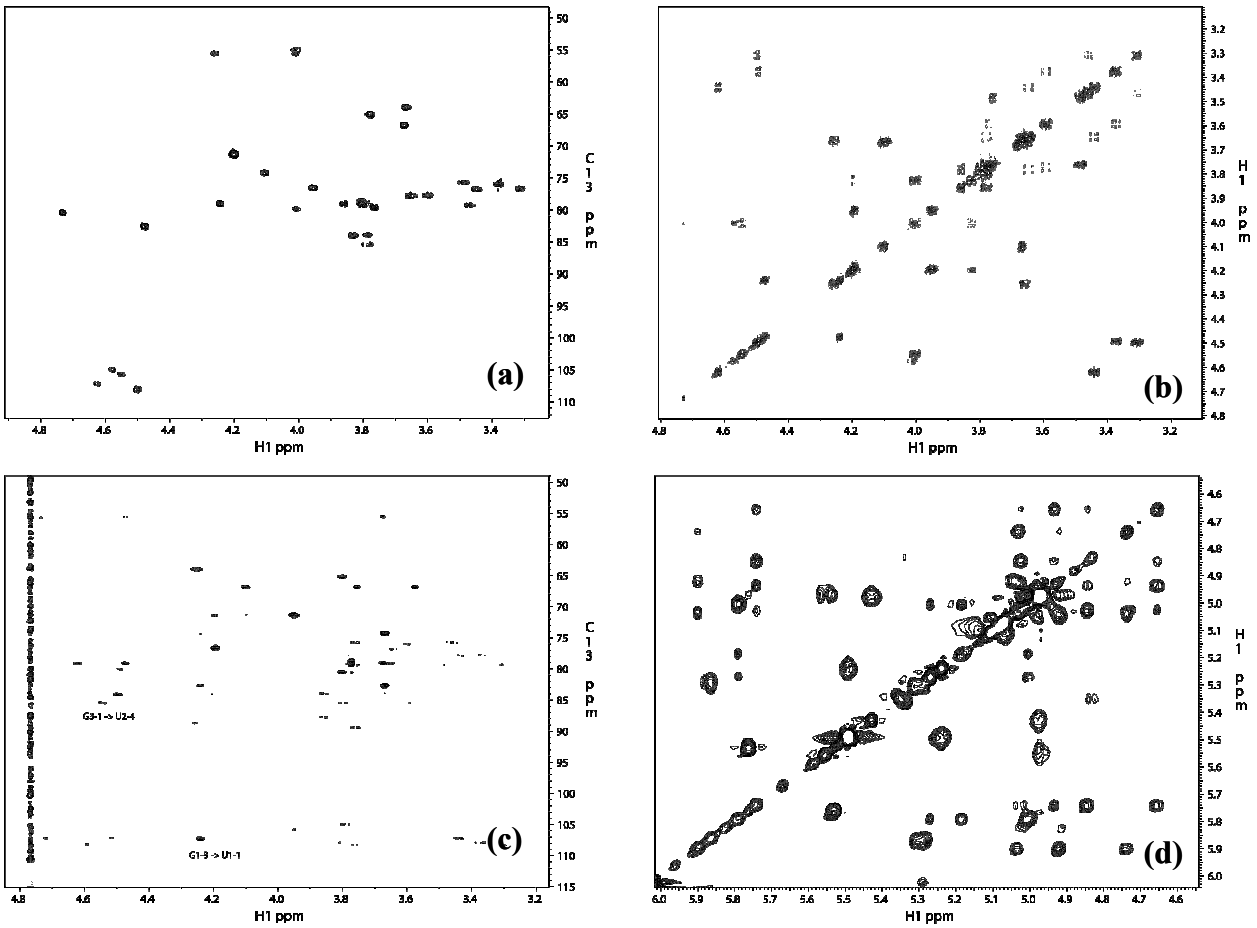


Figure 4.6 NMR spectroscopy for a CS hexamer, GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol. (a) ^{1}H-^{13}C HSQC, (b) ^{1}H-^{1}H DQCOSY, (c) ^{1}H-^{13}C HMBC, (d) ^{1}H-^{1}H TOCSY

Back-bone and side-chain assignments of the E66S were performed using conventional triple resonance protein NMR experiments. The ^{13}C and ^{15}N doubly-labeled protein sample was prepared in 50mM acetate buffer (pH 4.5), 12.5% D_2O and 0.01% (w/v) sodium azide. ^{15}N-HSQC, HNCACB and CBCACONH from the Biopack distributed by Varian were recorded to assist in the backbone assignment. To make side-chain assignments, Biopack versions of HN-TOCSY and HCCH-TOCSY also were used. Table 4.1 lists the detailed experimental parameters.

Table 4.1 3D NMR experiments parameters

Experiment	^{1}H freq. (MHz)	t_1 pts	t_2 pts	t_3 pts	F1 SW (Hz)	F2 SW (Hz)	F3 SW (Hz)	Mixing time (ms)
HNCACB	900	64	32	1024	18104	2189	14535	
CBCACONH	900	64	32	1024	13576	2189	14535	
TOCSYNHSQC	900	64	32	1024	10684	2189	14535	
HCCH-TOCSY	900	90	45	1024	8999	6788	14535	
F1 ^{13}C-filtered F3 ^{13}C-edited NOESYHSQC	900	90	45	1024	8999	6788	14535	150
NOESYNHSQC	900	128	16	1024	8999	2900	14535	150

The binding sites of CS oligomers to E66S were investigated by using chemical shift mapping as observed in ^{15}N-HSQC experiments. The binding affinity and binding stoichiometry also were determined in the same experiments. All the spectra were recorded on a Varian Inova 900MHz spectrometer. The protein was dissolved in 50mM acetate buffer (pH 4.5), 12.5% D_2O and 0.01% (w/v) sodium azide to obtain a 340μM sample. The ligand was titrated into the protein in increments of 34μM. The titration was continued until a final molar ratio of 1:1 (protein: ligand) was reached. The spectra were recorded after each titration with 8 transients per t1 increment to obtain a complex data matrix comprised of 1024 x 64 points.

The CS binding sites on E66S also were determined by NOE experiments. The sample was prepared in the same buffer shown above with 10% D_2O. The concentrations of the protein and ligand were controlled at 500μM. The ^{13}C-NOESYHSQC from Biopack was re-coded into a F1 ^{13}C-filtered F3 ^{13}C-edited NOESYHSQC sequence, and this sequence was used to identify the binding sites on the side chains of the protein. Another Biopack experiment, ^{15}N-NOESYHSQC,

was used to identify the binding sites on the backbone of the protein. The details are listed above in Table 4.1.

4.2.5 Light Scattering

The size of E66S and its complex with CS oligomers were determined by dynamic light scattering using a Protein Solutions Dynapro-99 instrument. All the samples were prepared in acetate buffer (pH 4.5) with a final concentration of ~300μM and pre-centrifuged at 6000 rpm for 30 minutes. The complex sample was prepared with an equal molar amount of the protein and ligand. The measurements were performed with 20 segments of scanning for 100 sec each at a power level of 50%. The data were analyzed the program, Dynamics V6

4.3 Results

4.3.1 Preparation and Isolation of CS Oligomers

The digestion of the CS-A polymer was carried out for a shorter time period (5 hrs) than that used in chapter III (12 hrs) in order to increase the amount of hexamers and the amount of larger oligomers in the digest. The enzyme and small molecules were removed from the digest by gel filtration on a G-25 column and mass spectroscopy was performed to analyze the sample. The largest CS oligomers in the sample were 12-mers, and all the oligomers had an even-number of residues. In contrast, with a 12-hour digestion, we did not obtain CS molecules larger than octamers and almost half of the digest was odd-numbered oligosaccharides. The latter observation supports our earlier conjecture that the odd numbered oligomers arise from a low level of activity from a contaminating enzyme that manifests itself more as time progresses. The sample was then separated by SAX HPLC and the profile shown in Figure 4.7(a) was observed. Twelve major peaks in total were observed and fractions IX and X were judged to contain hexamers and octamers respectively, based on mass spectroscopy. Even though the collected

fractions appeared homogenous based on mass spectrometry, the results from NMR spectroscopy indicated that the molecules were not uniform in their sulfation patterns. This is not surprising since there can be a large number of equal mass isoforms even for a hexamer. If one considers a molecule with three sulfation sites and allows only the 6 and 4 positions to be sulfated, there are twenty possible isoforms. In order to obtain a homogenous CS oligomer, another SAX HPLC separation was attempted. However, simply repeating the SAX HPLC did not improve the separation. This is in line with our previous experience with GAG chromatography[18]. Therefore, a simple chemical modification was carried out. We reduced the molecules from fraction IX with sodium borohydride and re-applied them to SAX HPLC. This did not change the basic charge properties of the oligomers, but the more flexible reduced terminus may have allowed better differentiation of sulfate positions. With the anomeric residue being reduced, fraction IX was revolved into three fractions, annotated by a, b and c as shown in Figure 4.7(b).

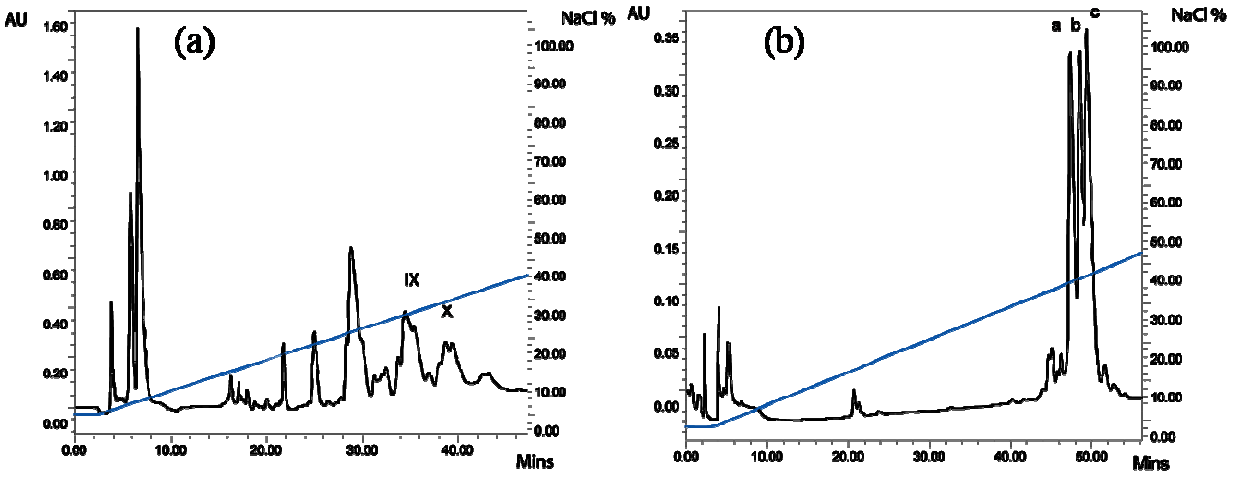


Figure 4.7 Chromatography profile of SAX HPLC before (a) and after (b) the reduction on CS. The blue line shows the time course of the gradient in NaCl concentration as a percentage of 2M NaCl in the buffer.

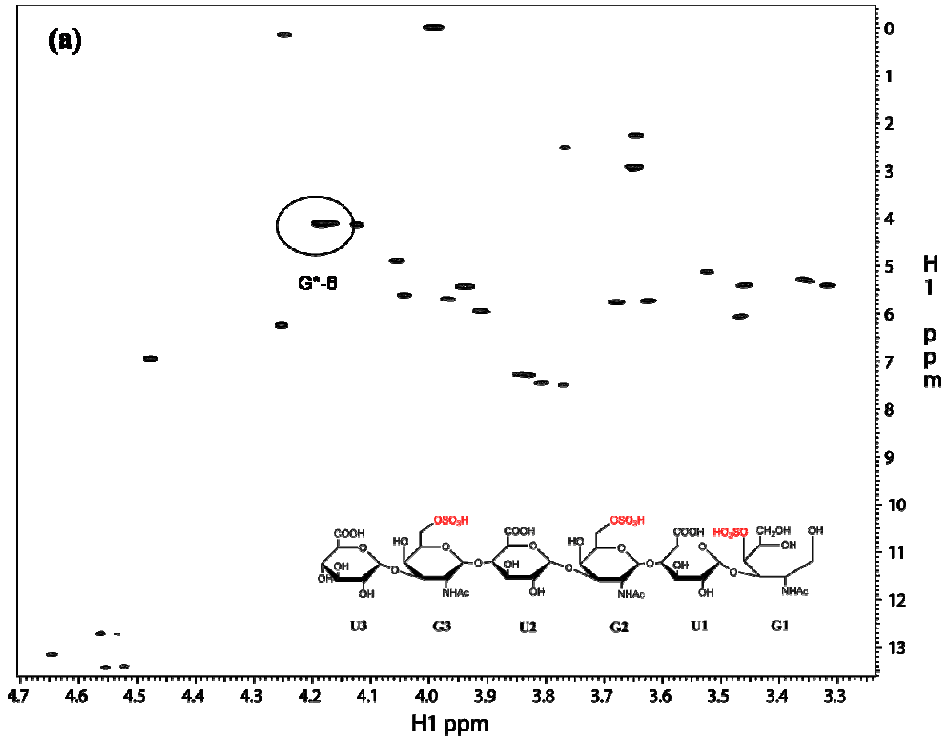
4.3.2 NMR Analysis of CS Hexasaccharides

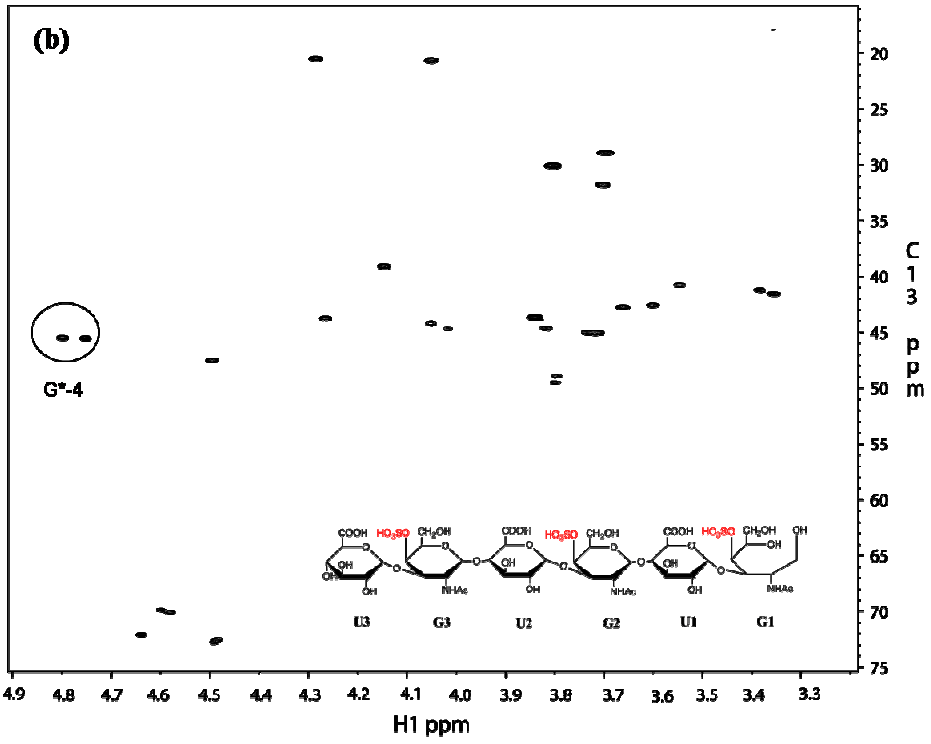
The hexasaccharides from fractions a, b and c were characterized by 2D NMR spectroscopy. Figure 4.7 shows the ^{1}H-^{13}C HSQC spectra of these molecules. Among them, the hexasaccharide c was fully assigned based on DQCOSY, TOCSY and HMBC. The other two hexasaccharides were partially assigned based on DQCOSY and TOCSY. As judged by the HSQC spectra, all three samples show high degrees of homogeneity. With the NMR assignments, the three hexasaccharides, a, b and c, were identified as GlcA-GalNAc6S-GlcA-GAlNAc6S-GlcA-GalNAc4S-ol, GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol and GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol respectively, where "-ol" stands for the reduced anomeric residue of the oligosaccharide. Once resonances were assigned to sites within each ring, and sets of resonances belonging to each residue were sequentially connected, the sulfation sites of the saccharides were simply determined by the characteristic downfield shift of the ^{1}H and ^{13}C resonances of the sulfation site. Resonance assignments are summarized in Table 4.2

Table 4.2 ^{1}H and ^{13}C chemical shift data of GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol

Residue	^{1}H (ppm)	^{13}C (ppm)		Residue	^{1}H (ppm)	^{13}C (ppm)
G1-1	3.68	66.77		U2-1	4.51	108.10
G1-2	4.27	55.55		U2-2	3.39	75.94
G1-3	3.47	79.29		U2-3	3.61	77.71
G1-4	4.49	82.61		U2-4	3.79	85.40
G1-5	4.12	74.23		U2-5	3.82	78.77
G1-6	3.68	63.99		G3-1	4.56	105.72
U1-1	4.63	107.19		G3-2	4.02	54.99
U1-2	3.46	76.76		G3-3	3.85	84.01
U1-3	3.66	77.75		G3-4	4.21	71.29
U1-4	3.80	83.90		G3-5	4.25	76.58
U1-5	3.88	79.07		G3-6	4.21	71.29
G2-1	4.59	104.97		U3-1	4.51	108.10
G2-2	4.02	54.99		U3-2	3.32	76.68
G2-3	3.85	84.01		U3-3	3.48	79.26
G2-4	4.74	80.43		U3-4	3.50	75.70
G2-5	4.02	79.84		U3-5	3.77	79.63
G2-6	3.79	65.12				

As shown in the Figures 4.8a-c, the three HSQC spectra have very similar resonance patterns. Indeed, most of the reduced chondroitin sulfate oligomers (CS-ol) share the same characteristic resonance regions in HSQC spectra which are summarized in Figure 4.8(c) and Table 4.3. Five resonance regions were identified and circled in Figure 4.8(c). Region 1 is the anomeric proton region common for carbohydrates. The number of resonances in this region is always a good indicator of the size of the oligosaccharide as well as the homogeneity of the sample. Resonances from region 2 include the H-4 from a 4-sulfated sugar residue. Region 3 is normally composed of H-6 from 6-sulfated sugar residues. Most of the resonances from the reduced GalNAc residue, along with H-6 from non-6-sulfated residues, are located in region 4. And region 5, another common region for carbohydrates, is composed of most of the H-2 resonances.


(a)
G*-β
U3
G3
U2
G2
U1
G1
H1 ppm
H1 ppm


(b)
G*-4
U3
G3
U2
G2
U1
G1
H1 ppm
C13 ppm

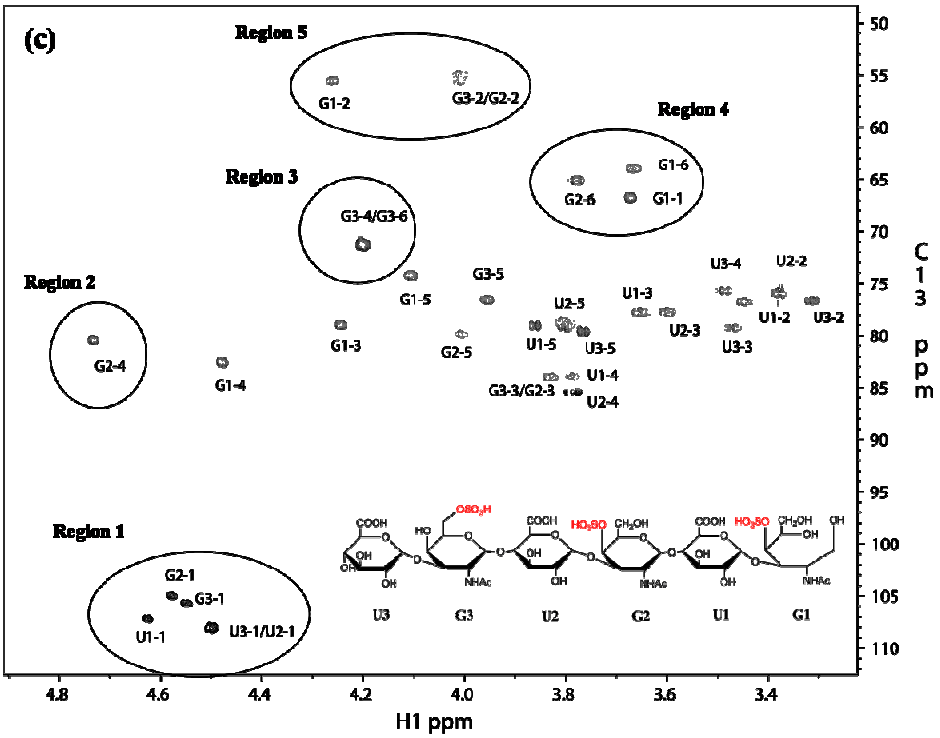


Figure 4.8 ^{13}C-HSQC spectra of CS hexamers a, b and c.

Table 4.3 Characteristic resonance regions in ^{13}C-HSQC spectra of CS oligosaccharides

Region	^{1}H (ppm)	^{13}C (ppm)	Resonances
1	4.4-4.7	65-75	H-1
2	4.7-4.9	40-45	H-4 (sulfated)
3	~4.1	~70	H-6 (sulfated)
4	3.4-3.7	60-70	H-1 (reduced residue) H-6
5	3.5-4.3	55-60	H-2

4.3.3 NMR Assignment of E66S

Backbone assignment of the E66S was performed using conventional backbone directed experiments including HNCACB and CBCACONH. All of the amino acids observable in the ^{15}N-HSQC spectrum were assigned except for the missing Leu19. Figure 4.9 shows the ^{15}N-

HSQC of E66S annotated with the residue number and assignment of each peak. Residue Val58 (5.67ppm in ^{1}H dimension) is not included. The assignments also are given in Table 4.4. Resonance assignments agree with published assignments[20] for approximately 95% of the residues. This level of agreement is reasonable, given the additional mutation and differences in pH of the sample.

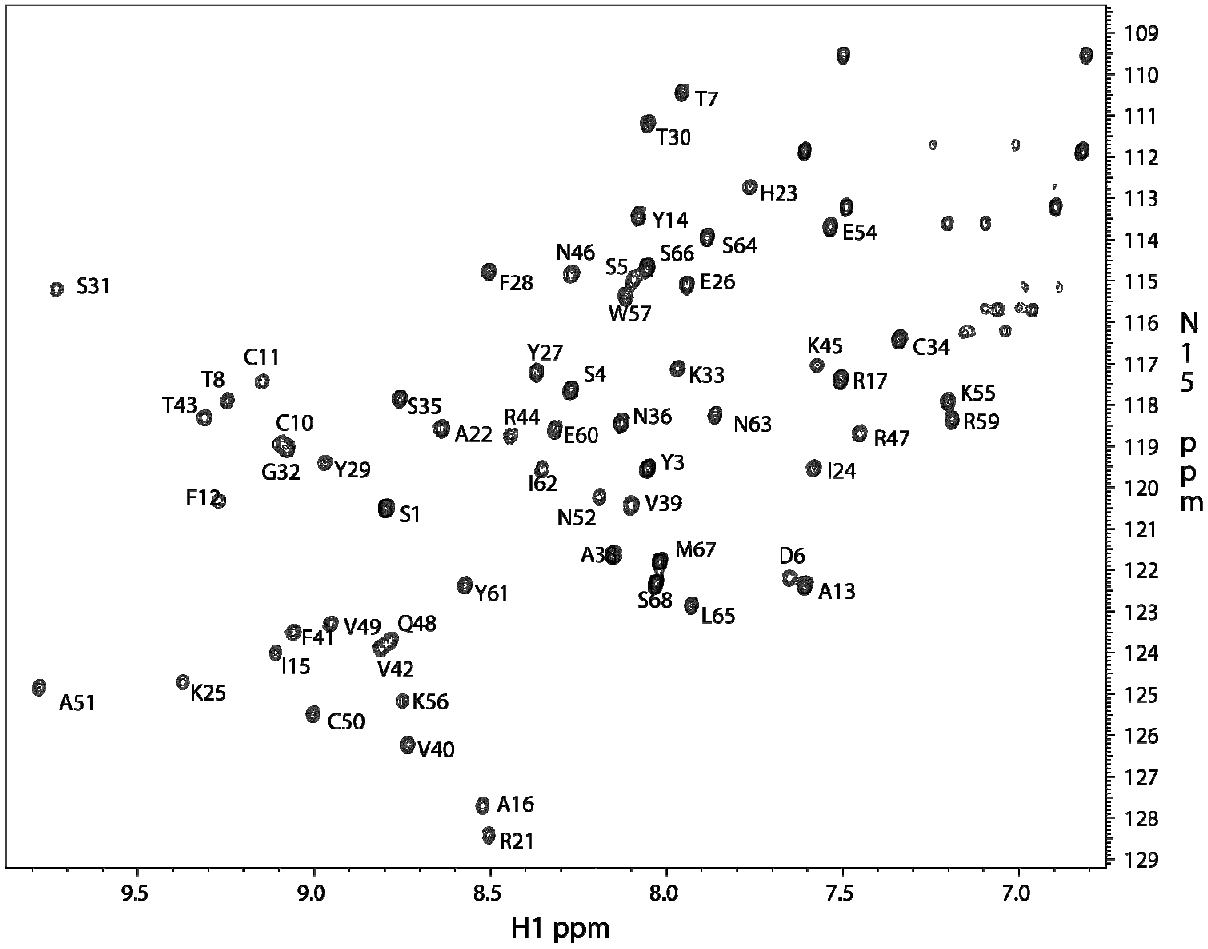


Figure 4.9 Amide ^{1}H-^{15}N assignments of E66S

Table 4.4 ^{15}N and ^{1}H chemical shift of E66S at 297K, pH 4.5

Residue	^{15}N (ppm)	NH (ppm)		Residue	^{15}N (ppm)	NH (ppm)
S1	120.50	8.78		S35	117.88	8.74
P2				N36	118.49	8.11
Y3	119.64	8.04		P37		
S4	117.76	8.26		A38	121.67	8.14
S5	114.91	8.07		V49	120.49	8.08
D6	122.29	7.63		V50	126.32	8.72
T7	110.44	7.94		F41	123.53	9.04
T8	117.88	9.22		V42	124.07	8.81
P9				T43	118.28	9.31
C10	118.92	9.08		R44	118.69	8.43
C11	117.50	9.13		K45	117.01	7.57
F12	120.36	9.26		N46	114.91	8.24
A13	122.35	7.59		R47	118.63	7.43
Y14	113.41	8.07		Q48	123.72	8.78
I15	124.00	9.09		V49	123.30	8.92
A16	127.69	8.51		C50	125.46	8.99
R17	117.39	7.50		A51	124.84	9.76
P18				N52	120.24	8.16
L19				P53		
P20				E54	113.87	7.52
R21	128.41	8.49		K55	117.89	7.19
A22	118.62	8.62		K56	127.69	8.51
H23	113.15	7.73		W57	115.25	8.08
I24	119.62	7.56		V58	122.48	5.61
K25	124.73	9.36		R59	118.49	7.17
E26	115.13	7.92		E60	118.62	8.29
Y27	117.14	8.35		Y61	122.35	8.54
F28	114.78	8.48		I62	119.62	8.34
Y29	119.40	8.96		N63	118.25	7.84
T30	111.25	8.03		S64	114.02	7.86
S31	115.27	9.71		L65	122.97	7.90
G32	119.08	9.06		S66	114.77	8.04
K33	117.13	7.95		M67	121.86	8.00
C34	116.51	7.32		S68	122.52	8.02

4.3.4 Chemical Shift Mapping

^{15}N-HSQC experiments were used to monitor the backbone amide chemical shift perturbation in E66S upon the addition of two different CS hexamers, GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol and GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol. Amide chemical shifts are usually sensitive to the perturbations of connected side-chains that interact with the ligand, but they also may be affected at remote postions by conformational changes in the protein induced by ligand binding. This can complicate the process of locating the binding site.

The first chemical shift mapping experiment was applied to the CS ligand GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol. An equal molar amount of the ligand was added into the protein sample in a 10-step titration. A ^{15}N-^{1}H HSQC spectrum was recorded after every step of titration. Figure 4.10 shows the overlay of the ^{15}N-^{1}H HSQC spectrum of E66S without the ligand (red) and the ^{15}N-^{1}H HSQC spectrum after the final step of the titration (blue). Many of the NH resonances are indeed affected by the addition of the CS ligand. They move progressively as the ligand is added, indicating that the ligand is in fast exchange on and off the protein. Since we already made a back-bone assignment of free E66S, the assignment of most ligand-bound E66S resonances was easily made by tracking the resonance shifts during the course of the titration. It should be noted that several resonances (Ser4, Ser5, Ala16 and Arg44) disappear immediately after the first step of titration. These are very likely broadened by the chemical exchange on a time scale that is approaching the reciprocal of the difference in chemical shift beween resonances in the protein with and without ligand. An expression for the extent of broadening is given by, $\Delta\mu \propto P_A P_B \Delta\omega^2$ (P_A and P_B are the percentage of molecules in free and bound states; $\Delta\omega$ is the chemical shift difference between these two states in rad s^{-1}), for

the fast-exchange situation. As predicted, the broadening will be most extreme for residues with large chemical shift changes. We see other evidence of broadening and intensity loss and the level of loss seems to correlate well with the magnitude of the chemical shift changes. Also, broadening of these peaks is more pronounced at midpoints of the titration, something that is characteristic of the Pa*Pb population dependence of fast-exchange broadening. A plot of the weighted chemical shift perturbation from ^{1}H and ^{15}N [$\Delta\delta = \left(\Delta\delta_H^2 + 0.2\Delta\delta_N^2\right)^{1/2}$] along the protein sequence for the resonances that do not disappear is depicted in Figure 4.11. The positions of residues for which resonances disappear are indicated with an asterisk. If we assume that perturbations correlate with the proximity of ligand binding, the plot shows that CS binding mostly occurs on an amino acid cluster, Arg44-Gln48, within the 40S loop of CCL5. This observation agrees with the previously reported GAG binding site on CCL5 based on a mutation screen of the protein; the 40S loop on CCL5 is where a BBXB motif (^{44}RKNR^{47}) common to heparin binding proteins is found. However, not only the 40S loop is perturbed; we also observed strong perturbations in another two regions, Ser1-Thr7 and Ala16-Arg17. Along with the 40S loop, these three regions happen to be on the same side of the dimer form of E66S seen in the crystal structure[21]. Altogether, they might constitute a binding pocket for the CS ligand which is shown in Figure 4.12.

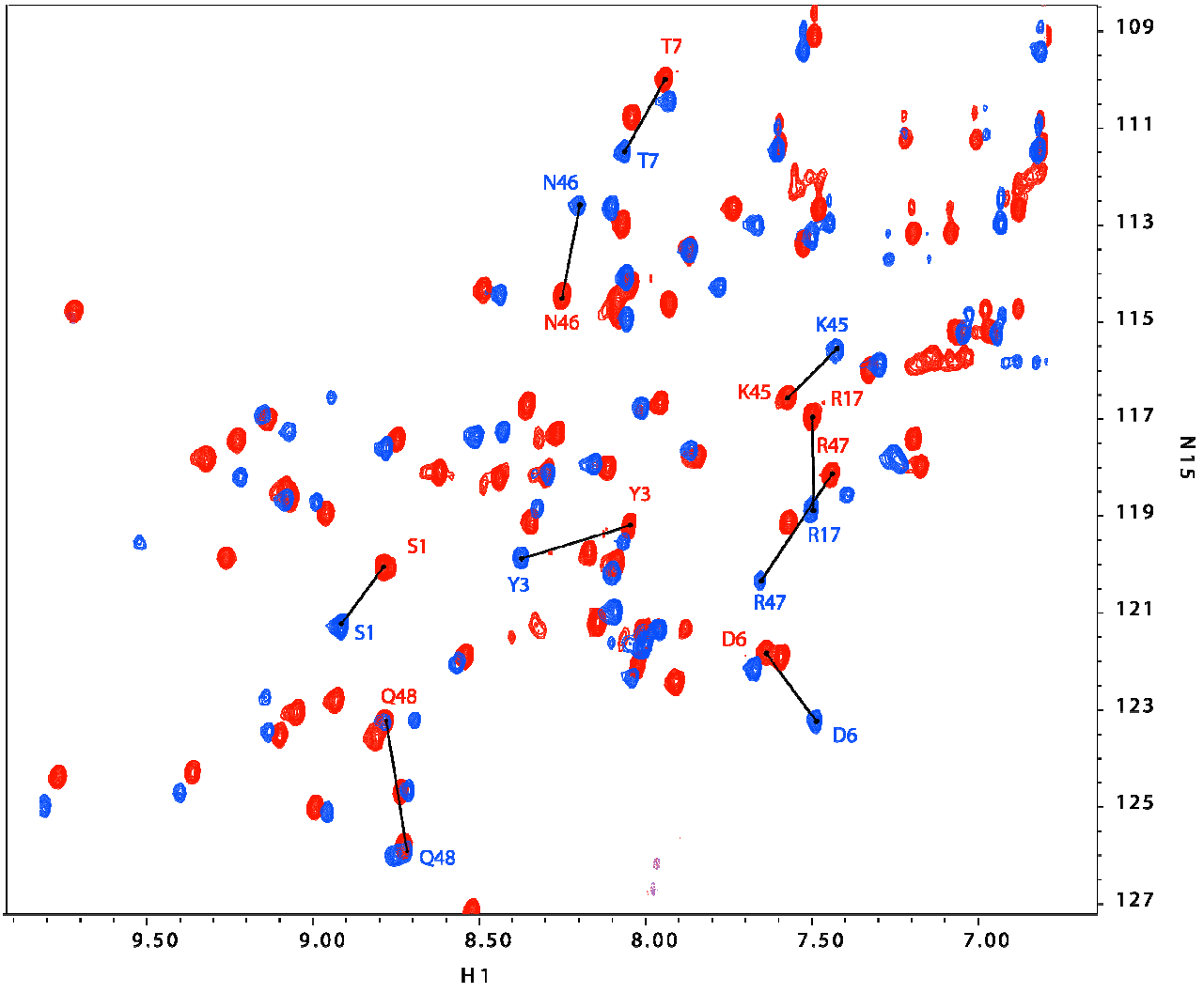


Figure 4.10 Overlaid ^{15}N-HSQC spectra of free (red) and bound (blue) E66S. The resonances with dramatic shifts are labeled with residue numbers.

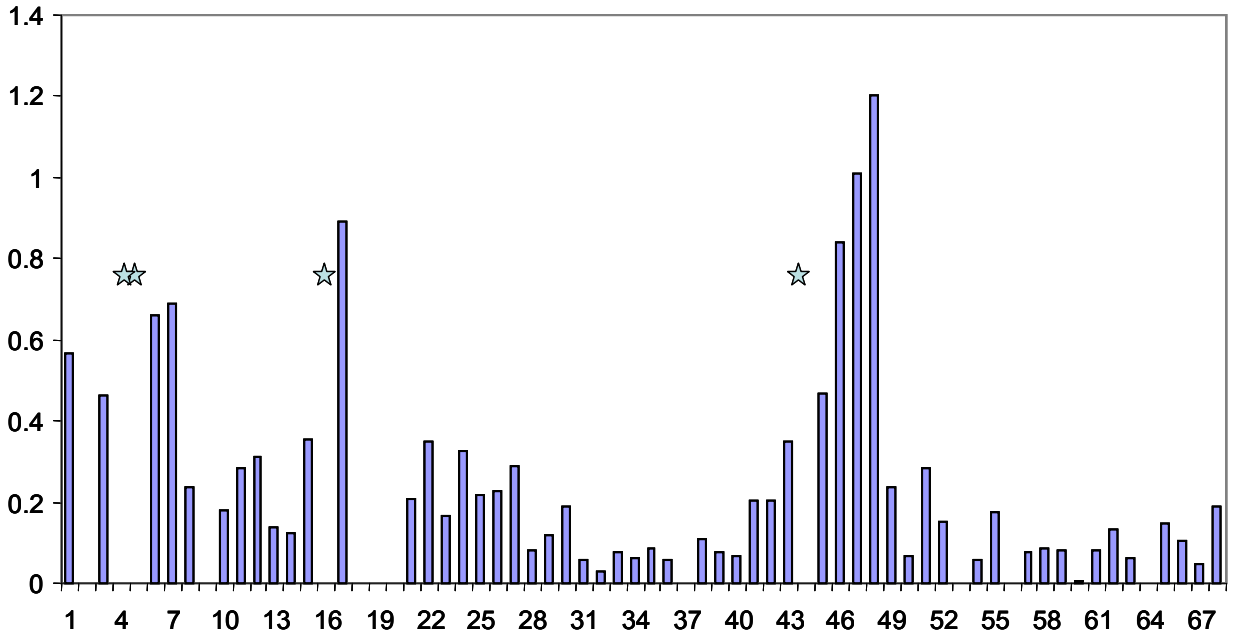


Figure 4.11 Weighted average chemical shift perturbation along the protein sequence. The positions of residues for which resonances disappear are indicated with an asterisk.

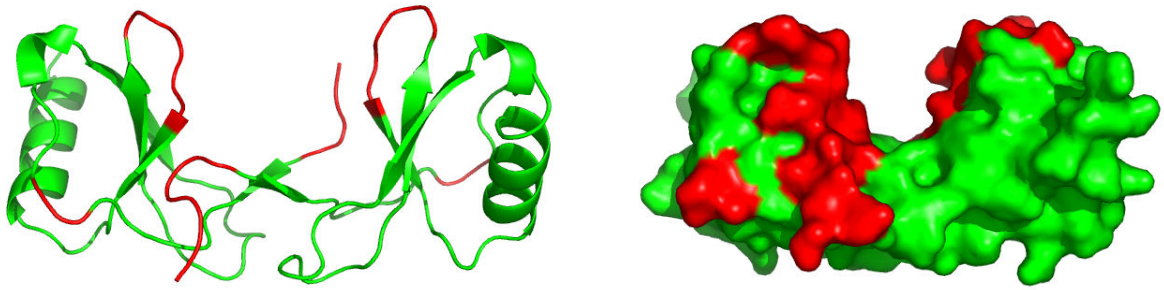


Figure 4.12 Potential binding sites (red) on a CCL5 dimer based on chemical shift mapping

The same mapping experiment was attempted on another CS ligand GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol. Interestingly, the resonances from HSQC did not change in chemical shift but in intensity. The S/N of the spectrum decreased during the course of the

titration and a flocculent precipitation was formed (Figure 4.13). A possible interpretation is that the addition of GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol induces aggregation of E66S and leads to the precipitation of the protein. This suggests that both the size of the GAG ligand and the sulfation pattern play a critical role in the oligomerization of chemokines.

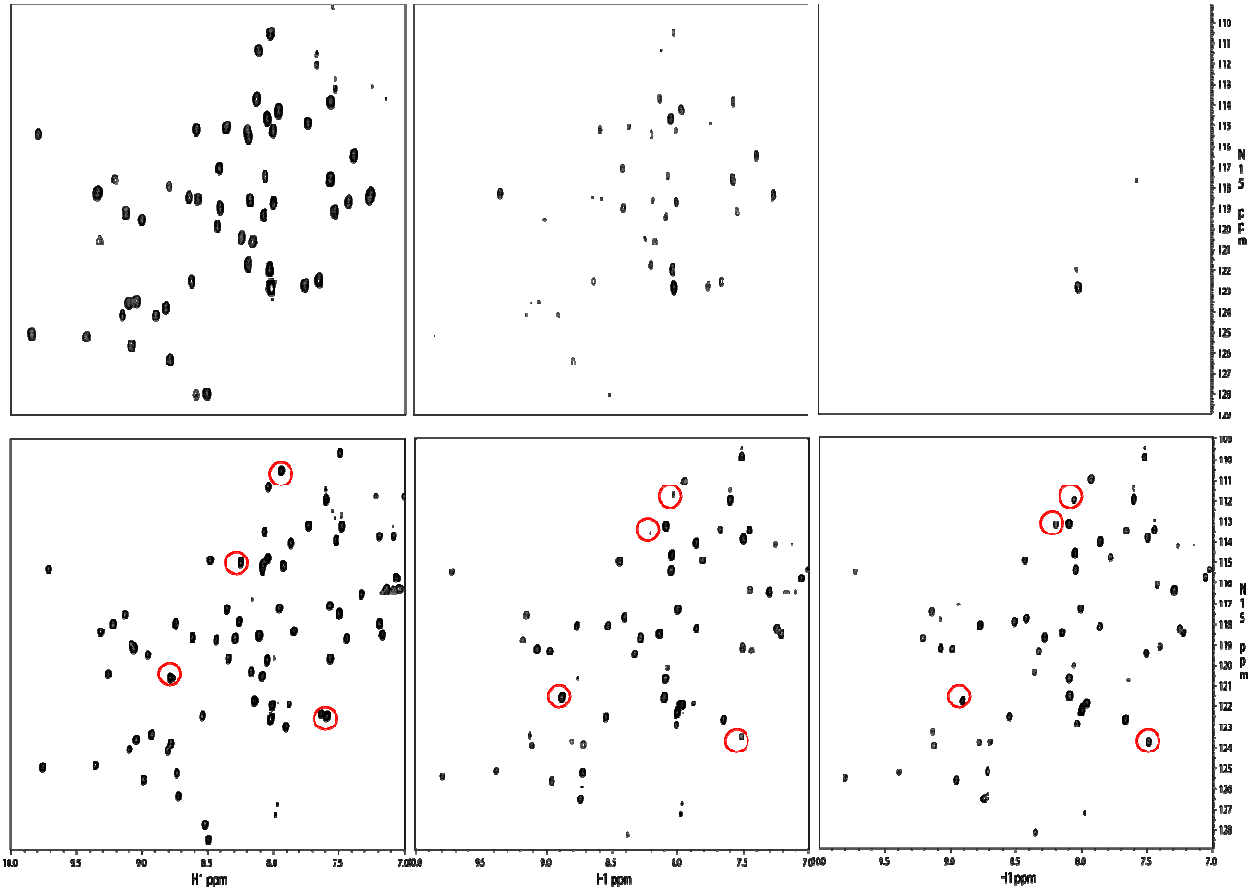


Figure 4.13 Comparison of NMR titrations with CS ligands GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol (top) and GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol (bottom). The molar ratio of CS to E66S monomer at the three steps is 0%, 80% and 120%, respectively. The circles highlight a few peaks showing large chemical shift changes

4.3.5 Light Scattering

Previous literature suggests that native CCL5 tends to aggregate upon the addition of GAGs but that the E66S mutant of CCL5 continues to exist as a dimer[19]. The above observation

of GAG-induced aggregation is therefore both surprising and interesting. Light scattering measurements were used to further characterize the oligomerization status of E66S in the absence and presence of the fast exchanging CS ligand, GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol. Figure 4.14 shows the overlaid regularization graph for light scattering measurements of free (red) and bound (blue) E66S. The two graphs share a similar size distribution pattern and four peaks were observed. Peaks 1, 3 and 4 were believed to be from a contamination because they also were observed in a reference graph from the buffer. The peaks labeled 2 are indicative of the average hydrodynamic size of the E66S molecules. For the free protein, the radius of the molecule is 2.1 nm and the molecular weight was calculated to be 20 kDa. This number suggests that the free E66S exists as a dimer which has a predicted molecular weight of 16 kDa. For the ligand-protein complex, the radius of the molecule is 2.2 nm and the molecular weight was calculated to be 21 kDa. This result shows that the protein remained a dimer upon the addition of this particular ligand which has a molecular weight of 1.4 kDa.

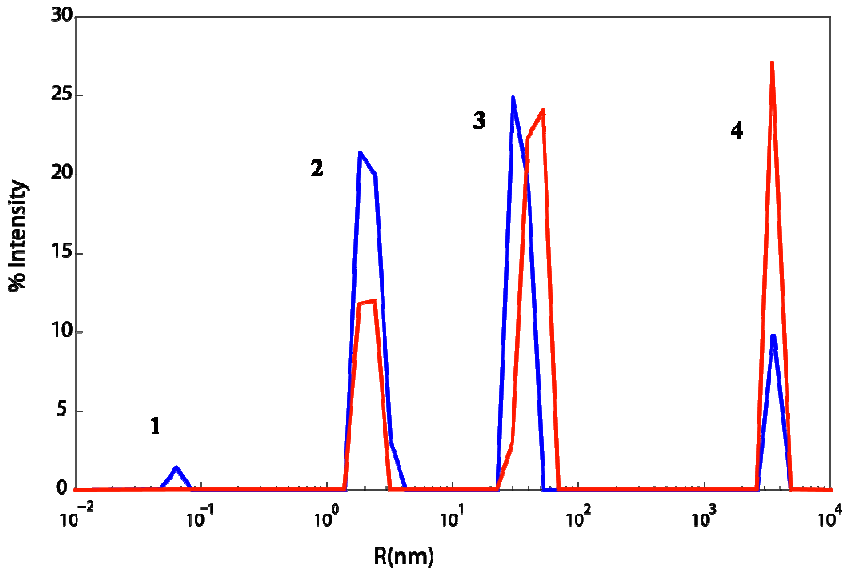


Figure 4.14 Overlay of the regularization graph for the light scattering measurement of free (red) and bound (blue) E66S

4.3.6 Determination of Binding Constants and Stoichiometry by NMR Titration

The NMR titration experiment also was used to determine the binding constant of the CS ligand GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol to CCL5. The residues that show good line shape and large chemical shift changes like Ser1, Tyr3, Phe12, Lys45, Asn46, Glu48 and Leu65, were chosen for a calculation of binding affinity. The weighted average chemical shift change given by $\Delta\delta = \left(\Delta\delta_H^2 + 0.2\Delta\delta_N^2\right)^{1/2}$ from each residue was normalized by $\Delta\delta / \Delta\delta_{\max} = 1.0$, where $\Delta\delta_{\max}$ is the maximum observed shift difference, used for plotting a titration curve. Fig 4.15 shows the overlay of titration curves for the chosen residues. The high degree of agreement for these curves suggests that we are looking at a single binding process and confirms the reliability of the titration experiment. From these experimental curves, an average curve was generated and superimposed to a set of standard curves simulated based on different protein-ligand stoichiometry, 1:1 (Figure 4.16a) and 2:1 (Figure 4.16b). The comparison of the two figures shows that the experimental curve fits the simulations in figure b much better than in figure a. This result further confirms our conclusion that the E66S exists as a dimer, and also suggests every E66S dimer accommodates one CS ligand. Based on the stoichiometry of 2:1 (protein:ligand) binding affinity of the CS ligand (GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol) to CCL5 was calculated to be 4μM through curve fitting.

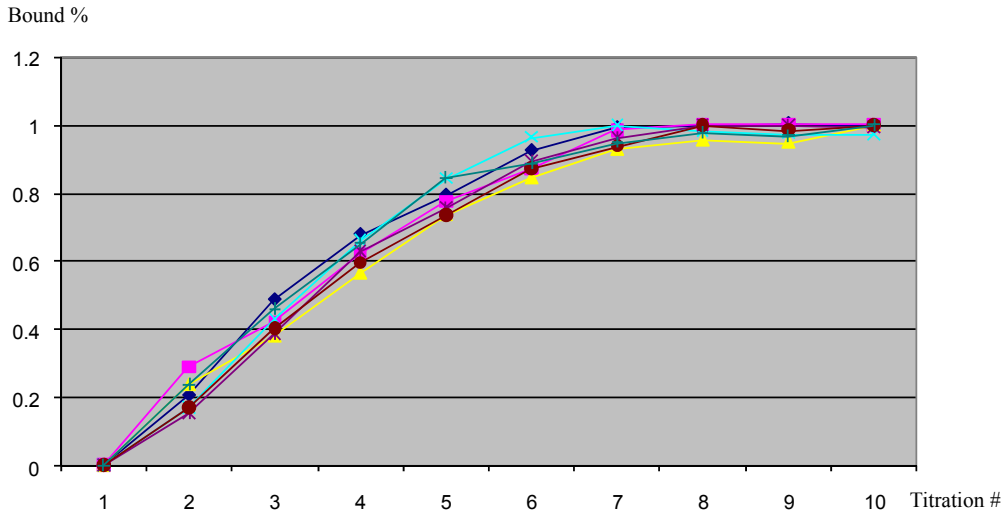


Figure 4.15 Overlay of titration cures for Ser 1, Tyr 3, Phe 12, Lys 45, Asn 46, Gln 48 and Leu 65.

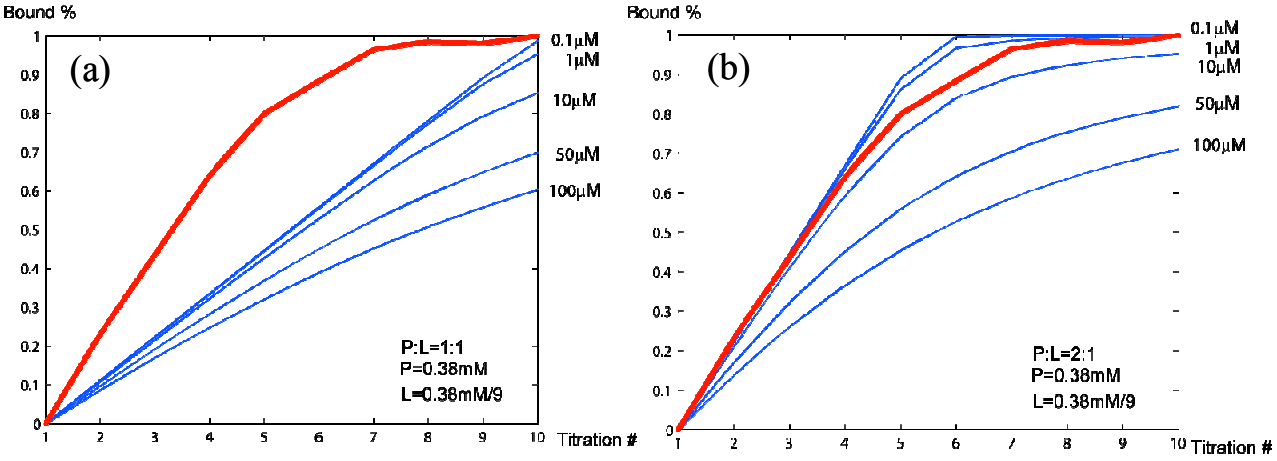


Figure 4.16 Experimental titration curve (red) superimposed on standard curves simulated based on a protein-ligand stoichiometry, of 1:1 (a) and 2:1 (b)

4.3.7 Mapping the Binding Sites on the Back-Bone of CCL5

^{15}N edited NOE spectroscopy, ^{15}N-NOESYHSQC, was used to search for short distance interactions between the ligand and the backbone of CCL5. Figure 4.17 shows a comparison of the spectrum for the free (blue) and bound (red) protein. The residues identified as those

potentially in the binding site were isolated and strips emanating from the HSQC cross peaks are shown. The peaks in these strips identify resonance positions for those coming from protons within 5 Å of the amide proton generating the cross peak. Unfortunately the ligand bound spectrum is of lower quality. However, other than intensity changes and small deviations in shifts, the red and blue pairs are, for the most part, very similar and devoid of extra peaks that might be attributed to the close approach of a ligand proton. One exception is the pair for Arg17. An NH-NH cross peak appears on the Arg17 plane of the bound protein spectrum that is in quite a different position (8.0 ppm) from that assigned to the Ala16 crosspeak in the free protein (8.6 ppm). This peak is unlikely to arise from an inter-residue (Arg17-Ala16) NOE. First, the Ala16 cross-peak does not shift 0.6 ppm upon adding the ligand (this would be a very large shift for protons in any event). And second, the corresponding NH-NH cross-peak cannot be located on the Ala16 strip of the spectrum. This cross-peak also cannot be an Arg17-Pro18 NOE simply because proline does not have an amide proton. The fact that we do not see an Ala16-Arg17 connection in the presence of the ligand is not surprising as the Arg17 auto-peak is weak due to exchange broadening and the Ala16 peak is unobservable in the presence of ligand. Therefore, the most likely possibility for this signal is that it belongs to an NOE between Arg17 and one of the NH groups from the CS ligand. These peaks do occur in the proper region of the spectrum and could give rise to such a cross-peak.

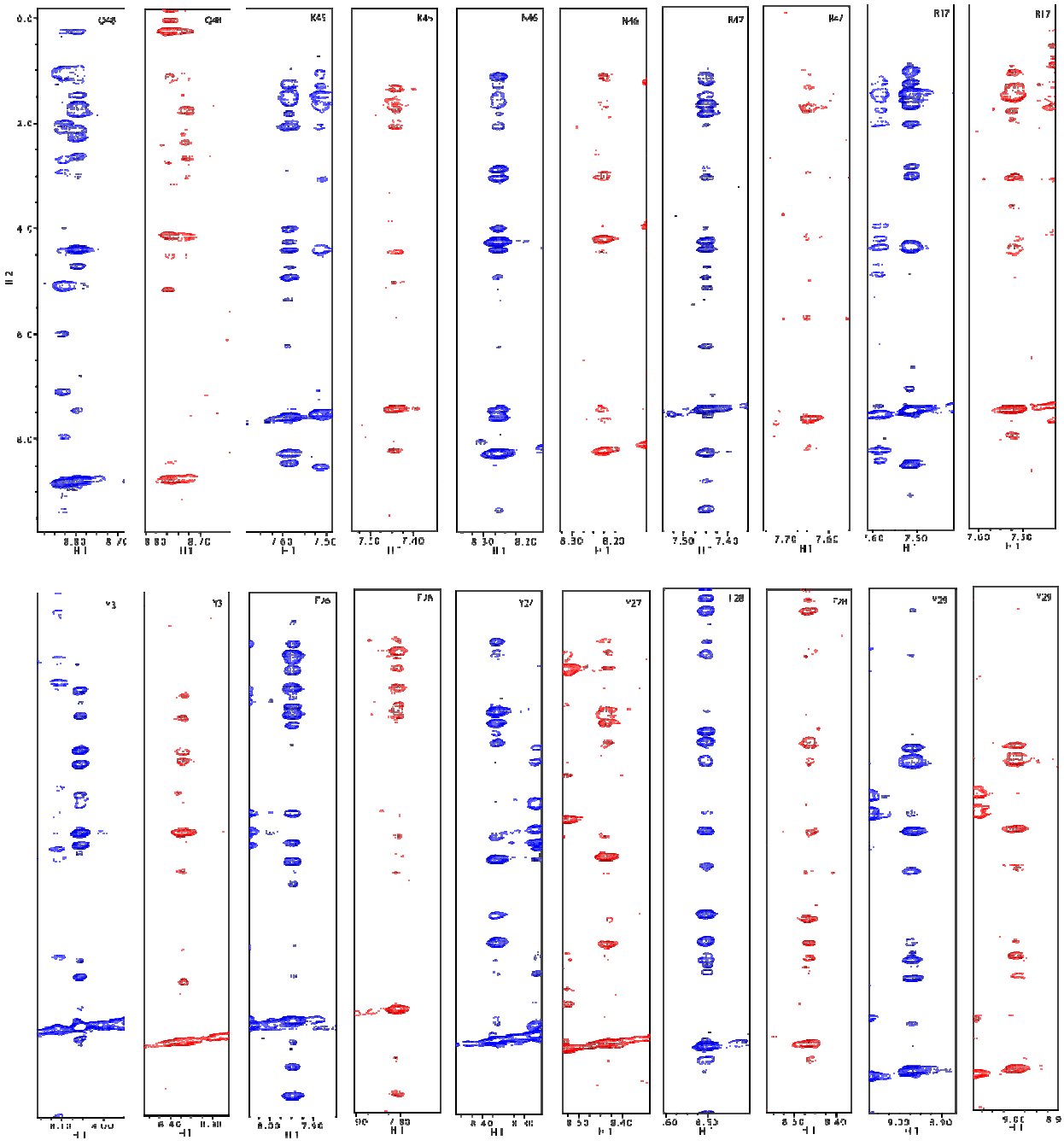


Figure 4.17 2D strips from a 3D ^{15}N-edited NOE spectrum of the residues in the potential binding site for the free (blue) and bound (red) protein

4.3.8 Mapping the Binding Site on the Side-Chains of CCL5

The F1 ^{13}C-filtered F3 ^{13}C-edited NOESYHSQC experiment was used to search for short distance interactions between the side-chain of CCL5 and the CS ligand. The ^{13}C labeled E66S

was used in order to filter out most of the intramolecular NOE signals arising from ^{13}C-connected protons. Only the intermolecular NOEs between the unlabeled ligand and the labeled protein and intramolecular NOEs arising from amide protons are observable. Since most of the NH signals appear far downfield (>7ppm) compared with the signals from carbohydrate (<5ppm), it is very easy to distinguish them. The NOE cross-peaks with characteristic chemical shifts for carbohydrates (3-5ppm) were isolated and are shown in the 2D strips extracted from the 3D spectrum (Figure 4.18). The corresponding diagonal signals arising from ^{13}C-E66S were assigned with the aid of another HCCH-TOCSY spectrum taken from the same sample. The results indicate that the methyl protons of Ala13, Ile15 and Ala51 have short distance contacts with the protons on the ligand.

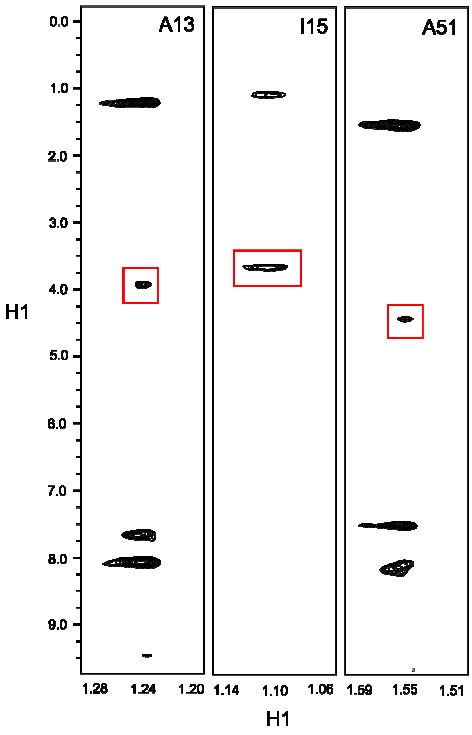


Figure 4.18 2D strips from a 3D ^{15}N-edited NOE spectrum of the NOE cross signal with the characteristic chemical shift for carbohydrates.

4.3.9 CS Ligand Binding Site on CCL5

We have used three independent experiments including chemical shift mapping, ^{15}N-edited NOESYHSQC, and F1 ^{13}C-filtered F3 ^{13}C-edited NOESYHSQC, to identify the residues from CCL5 involved in CS ligand binding. None of these experiments are definitive in location of the binding site. For the chemical shift mapping experiment, the chemical shifts of amides are not only sensitive to the perturbations of connected side-chains that interact with the ligand, but also sensitive to the conformational changes in the protein induced by ligand binding. The NOE experiments are limited by an inability to make definitive assignments to carbohydrate protons. In order to minimize the ambiguity introduced by these experiments, their results are compared in Table 4.5. The locations of these residues in the X-ray structure of wild type CCL5 dimer are shown in Figure 4.19. The residues are shown in different colors based on the experiment through which they are identified (red for chemical shift mapping, yellow for ^{15}N-edited NOESYHSQC and blue for F1 ^{13}C-filtered F3 ^{13}C-edited NOESYHSQC). The results from these comparisons are promising. Most of the NOE interactions occur close to the Ala16-Arg17 binding region determined by chemical shift mapping. Therefore, the residues involved in the CS ligand binding can still be grouped into three regions (Ser1-Thr7 for region I, Ala13-Arg14 and Ala51 for region II, Arg44-Gln48 for region III). Among these three regions, regions II and III are most likely to be the real CS binding sites on CCL5 because they either contain the well-recognized GAG binding motif (BBXB motif) or have NOE interactions with the CS ligand. However, the binding site on region I is doubtful. The chemical shift changes that occur in this region could be due to a conformation perturbation of the flexible N-terminal loop which is close to one of the CS binding sites, region II.

Table 4.5 CS binding involved residues distinguished by different experiments

	Residues from CCL5 involved in CS binding
chemical shift mapping	Ser1-Thr7, Ala16-Arg17, Arg44-Gln48
^{15}N-edited NOESYHSQC	Arg17
F1 ^{13}C-filtered F3 ^{13}C-edited NOESYHSQC	Ala13, Ile15 and Ala51

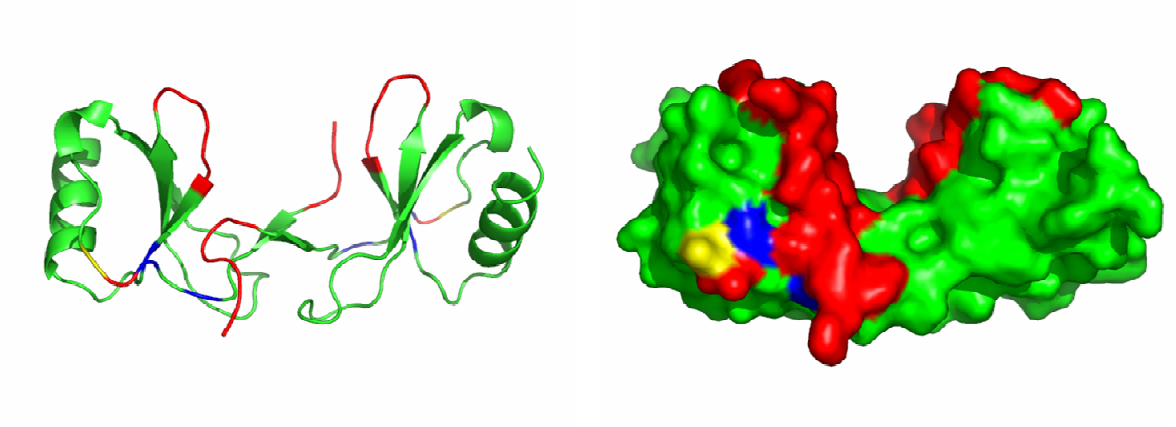


Figure 4.19 CS binding residues distinguished by different experiments are located in an X-ray structure of the CCL5 dimmer with different colors (red for chemical shift mapping, yellow for ^{15}N-edited NOESYHSQC and blue for F1 ^{13}C-filtered F3 ^{13}C-edited NOESYHSQC)

4.3.10 ^{13}C-^{13}C RDCs and Carbonyl Carbon CSA Offsets

To determine the bound conformation of CS ligand to CCL5, we have made our first attempt to obtain orientational information from the ^{13}C-labeled acetyl. The ^{13}C labeled CS ligand, GlcA-GalN^{13}Ac4S-GlcA-GAlN^{13}Ac4S-GlcA-GalN^{13}Ac4S-ol, along with ^{15}N labeled E66S, was aligned in a stretched neutral polyacrymide gel. The ^{15}N-^{1}H RDCs from the protein were measured with gNhsqc-IPAP experiments and values ranging from -10Hz to 15Hz were observed. The ^{13}C spectra of the aligned sample was acquired and compared with isotropic ^{13}C spectra of the free ligand (Figure 4.20). The chemical shifts were referenced to TMAB. Both the carbonyl region (a) and the methyl region (b) of the ^{13}C spectra are presented. The three doublets

arising from three ^{13}C-acetyls in the isotropic spectra are at the bottom (the doublet annotated by asterisks is from a contaminant). The aligned spectrum on the top, especially the carbonyl region, has poor resolution, which makes it difficult to distinguish the resonances. This is mainly because of the poor shimming of the stretched gel sample (some air bubbles remained inside the gel) and the interference from background signals arising from protein carbonyls further complicated the spectrum. We can roughly tell that RDCs are in a range of ±2Hz. Considering the maximum value of ^{15}N-^{1}H RDCs is 15Hz, 2Hz for a ^{13}C-^{13}C RDC is reasonable because taking bond lengths and the ratios of gammas into account, an ^{13}C-^{13}C RDC should be 18% of a ^{15}N-^{1}H RDC for which the orientation is the same. Since the linewidth of the aligned spectra is about 10Hz, which is larger than the observed RDCs, it is almost impossible to obtain accurate RDCs from this preliminary alignment system. The CSA offsets appear to be larger. The most upfield doublet clearly moves further downfield in the methyl region than in the carbonyl region. Shifts in the isotropic and aligned spectra shown cannot be compared directly because of possible environmental effects of the polyacrylamide on either our oligosaccharide or the reference compound. Therefore, in order to make the CSA offsets more promising as a source of alignment information, comparing the shifts of the oligosaccharide in stretched and un-stretched gels is more preferred. In fact, using the ^{15}N RDCs as a standard we calculate that the range of CSA offsets should be somewhat larger than 0.08 PPM.

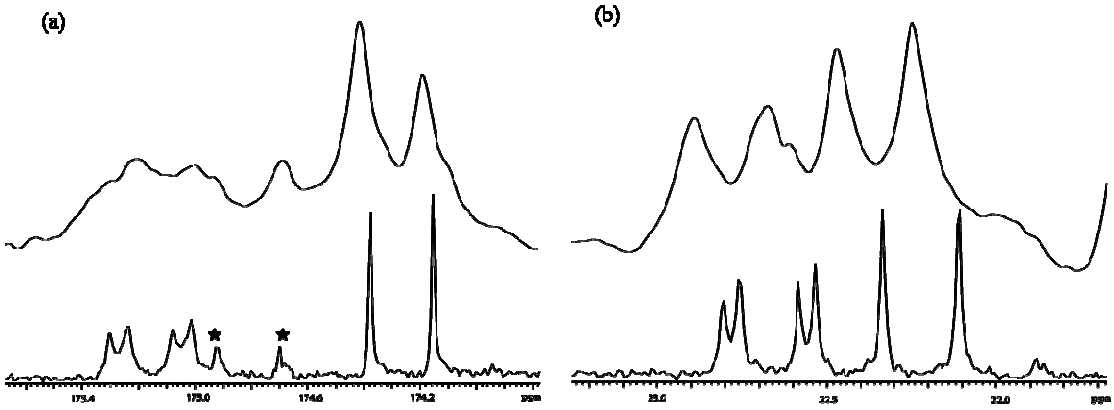


Figure 4.20 ^{13}C spectra of aligned E66S-CS complex in stretched gel (top) and CS ligand in free solution (bottom). Both carbonyl (a) and methyl (b) regions of the ^{13}C spectra are presented

4.4 Discussion

In conclusion, we have isolated two CS hexamers with different sulfation patterns for the investigation of the CS-CCL5 interaction. The different behavior for the two CS ligands in the interaction with CCL5 made us believe that not only the size, but also the sulfation patterns, of the GAG ligands play critical roles in the GAGs-chemokine interaction. With one of the ligands we isolated, GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol, we have identified the CS binding sites in CCL5 by using the chemical shift mapping experiment and transfer NOE spectroscopy. The results prove that the BBXB motif in the 40S loop of the chemokine is the principal site for CS binding. We also found that some other residues (Ala13-Arg14 and Ala51) are involved in CS binding. Data from chemical shift titrations were successfully used to estimate the binding affinity of GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol to CCL5. Both light scattering and chemical shift titration experiments indicate that the E66S mutant of CCL5 exists in the form of a dimer upon binding to the CS ligand, GlcA-GalNAc4S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol.

4.5 Future Directions

While we have made significant progress toward characterizing the geometry of a GAG-CCL5 complex, there remains ample opportunity to improve the precision of the characterization. We made one attempt at adding the RDC/CSA orientational information from acetyl-labeled ligands. However, both the sensitivity and level of alignment need to be improved to take full advantage of these data. Our choice of medium was initially deemed to be limited because charged media interact too strongly with either the positively-charged protein or the negatively-charged GAG. Neutral polyacrylamide stretched gels produced some alignment, but not enough. There are certainly other options including neutral bicelles, and paramagnetic alignment[22]. These could be pursued in the future, but the latter would require a redesign of the protein expression construct. NOE data can also be used more effectively. This will require assignment of carbohydrate resonances in the bound state. Proton-only methods are not likely to be applicable because of the broad lines. ^{13}C labeling of the carbohydrate remains an option. We had hoped that NOE contacts to the acetyls might be observed so that we could further exploit their labels. However, no contacts could to the methyl groups could be identified. Hence, a mores extensive ^{13}C labeling would have to be undertaken.

We also will try to map the CS binding site on CCL5 by using mass spectroscopy to compare with the conclusions we made based on NMR experiments. The MS-based hydroxyl-radical footprinting method[23] will be applied. This method has the ability to probe protein structure, and interactions in solution, and is possibly applicable to the aggregating species observed for the CS ligand, GlcA-GalNAc6S-GlcA-GAlNAc4S-GlcA-GalNAc4S-ol. These experiments are in progress, with the assistance of Dr. Sharp of the CCRC and Ms. Watson of the UGA Chemistry Department.

References

(1) DeVries, M. E.; Hosiawa, K. A.; Cameron, C. M.; Bosinger, S. E.; Persad, D.; Kelvin, A. A.; Coombs, J. C.; Wang, H.; Zhong, R.; Cameron, M. J.; Kelvin, D. J. The role of chernokines and chemokine receptors in allo antigen-independent and allo antigen-dependent transplantation injury. *Seminars in Immunology* **2003**, *15*, 33-48.

(2) Garcia, G.; Humbert, M.; Capel, F.; Rimaniol, A. C.; Escourrou, P.; Emilie, D.; Godot, V. Chemokine receptor expression on allergen-specific T cells in asthma and allergic bronchopulmonary aspergillosis. *Allergy* **2007**, *62*, 170-177.

(3) Rot, A. Contribution of Duffy antigen to chemokine function. *Cytokine & Growth Factor Reviews* **2005**, *16*, 687-694.

(4) Retz, M.; Sidhu, S. S.; Dolganov, G. M.; Lehmann, J.; Carroll, P. R.; Basbaum, C. Chemokine receptor CXCR4 mediates migration and invasion of bladder cancer cells. *Journal of Urology* **2004**, *171*, 192-192.

(5) Vaughn, D. E.; Bjorkman, P. J. The (Greek) key to structures of neural adhesion molecules. *Neuron* **1996**, *16*, 261-273.

(6) Paavola, C. D.; Hemmerich, S.; Grunberger, D.; Polsky, I.; Bloom, A.; Freedman, R.; Mulkins, M.; Bhakta, S.; McCarley, D.; Wiesent, L.; Wong, B.; Jarnagin, K.; Handel, T. M. Monomeric monocyte chemoattractant protein-1 (MCP-1) binds and activates the MCP-1 receptor CCR2B. *Journal of Biological Chemistry* **1998**, *273*, 33157-33165.

(7) Baryshnikova, O. K.; Sykes, B. D. Backbone dynamics of SDF-1 alpha determined by NMR: Interpretation in the presence of monomer-dimer equilibrium. *Protein Science* **2006**, *15*, 2568-2578.

(8) Fernandez, E. J.; Lolis, E. STRUCTURE, FUNCTION, AND INHIBITION OF CHEMOKINES. *Annual Review of Pharmacology and Toxicology* **2002**, *42*, 469-499.

(9) Weber, C. Novel mechanistic concepts for the control of leukocyte transmigration: specialization of integrins, chemokines, and junctional molecules. *Journal of Molecular Medicine-Jmm* **2003**, *81*, 4-19.

(10) Proudfoot, A. E. I. The biological relevance of chemokine-proteoglycan interactions. *Biochemical Society Transactions* **2006**, *34*, 422-426.

(11) Wagner, L.; Yang, O. O.; Garcia-Zepeda, E. A.; Ge, Y. M.; Kalams, S. A.; Walker, B. D.; Pasternack, M. S.; Luster, A. D. beta-chemokines are released from HIV-1-specific cytolytic T-cell granules complexed to proteoglycans. *Nature* **1998**, *391*, 908-911.

(12) Tanaka, Y.; Adams, D. H.; Hubscher, S.; Hirano, H.; Siebenlist, U.; Shaw, S. T-Cell Adhesion Induced by Proteoglycan-Immobilized Cytokine Mip-1-Beta. *Nature* **1993**, *361*, 79-82.

(13) Hoogewerf, A. J.; Kuschert, G. S. V.; Proudfoot, A. E. I.; Borlat, F.; ClarkLewis, I.; Power, C. A.; Wells, T. N. C. Glycosaminoglycans mediate cell surface oligomerization of chemokines. *Biochemistry* **1997**, *36*, 13570-13578.

(14) Burns, J. M.; Gallo, R. C.; DeVico, A. L.; Lewis, G. K. A new monoclonal antibody, mAb 4A12, identifies a role for the glycosaminoglycan (GAG) binding domain of RANTES in the antiviral effect against HIV-1 and intracellular Ca2+ signaling. *Journal of Experimental Medicine* **1998**, *188*, 1917-1927.

(15) Kuschert, G. S. V.; Coulin, F.; Power, C. A.; Proudfoot, A. E. I.; Hubbard, R. E.; Hoogewerf, A. J.; Wells, T. N. C. Glycosaminoglycans interact selectively with chemokines and modulate receptor binding and cellular responses. *Biochemistry* **1999**, *38*, 12959-12968.

(16) Proudfoot, A. E. I.; Fritchley, S.; Borlat, F.; Shaw, J. P.; Vilbois, F.; Zwahlen, C.; Trkola, A.; Marchant, D.; Clapham, P. R.; Wells, T. N. C. The BBXB motif of RANTES is the principal site for heparin binding and controls receptor selectivity. *Journal of Biological Chemistry* **2001**, *276*, 10620-10626.

(17) Proudfoot, A. E. I.; Wells, T. N. C.; Power, C. A. *Chemokine Protocols*, 2000; Vol. 138.

(18) Yu, F.; Wolff, J. J.; Amster, I. J.; Prestegard, J. H. Conformational preferences of chondroitin sulfate oligomers using partially oriented NMR spectroscopy of C-13-labeled acetyl groups. *Journal of the American Chemical Society* **2007**, *129*, 13288-13297.

(19) Czaplewski, L. G.; McKeating, J.; Craven, C. J.; Higgins, L. D.; Appay, V.; Brown, A.; Dudgeon, T.; Howard, L. A.; Meyers, T.; Owen, J.; Palan, S. R.; Tan, P.; Wilson, G.; Woods, N. R.; Heyworth, C. M.; Lord, B. I.; Brotherton, D.; Christison, R.; Craig, S.; Cribbes, S.; Edwards, R. M.; Evans, S. J.; Gilbert, R.; Morgan, P.; Randle, E.; Schofield, N.; Varley, P. G.; Fisher, J.; Waltho, J. P.; Hunter, M. G. Identification of amino acid residues critical for aggregation of human CC chemokines macrophage inflammatory protein (MIP)-1 alpha, MIP-1 beta, and RANTES - Characterization of active disaggregated chemokine variants. *Journal of Biological Chemistry* **1999**, *274*, 16077-16084.

(20) Chung, C. W.; Cooke, R. M.; Proudfoot, A. E. I.; Wells, T. N. C. The 3-Dimensional Solution Structure of Rantes. *Biochemistry* **1995**, *34*, 9307-9314.

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

(22) Rodriguez-Castaneda, F.; Haberz, P.; Leonov, A.; Griesinger, C. Paramagnetic tagging of diamagnetic proteins for solution NMR. *Magnetic Resonance in Chemistry* **2006**, *44*, S10-S16.

(23) Tullius, T. D.; Dombroski, B. A. Hydroxyl Radical Footprinting - High-Resolution Information About DNA Protein Contacts and Application to Lambda-Repressor and Cro Protein. *Proceedings of the National Academy of Sciences of the United States of America* **1986**, *83*, 5469-5473.

CHAPTER 5

CONCLUDING REMARKS

5.1 Conformational Investigation of N-Acetylated Carbohydrates by NMR: Possible Improvements and Remaining Challenges

The structural characterization of complex carbohydrates represents a major challenge for NMR. However, the methodology described in the previous chapters has made NMR more applicable to complex carbohydrates. The most novel aspects of this methodology rely on partially-oriented NMR of ^{13}C-acetyl labeled carbohydrates and the measurement of anisotropic parameters such as chemical shift anisotropy offsets and residue dipolar couplings. I have developed N-deacetylation protocols that should facilitate the ^{13}C-acetyl labeling of carbohydrates and have introduced a NMR-MS method that should facilitate NMR assignment of ^{13}C-labeled N-acetyl groups. In addition, I have made applications to several target molecules such as a chitin dimer in free solution, a chondroitin sulfate pentamer in free solution and CCL5-bound chondroitin sulfate hexamers. However, there are clearly improvements that can be made to these methods.

In order to retain the pyranose structure at the reducing end of the oligosaccharide, it is necessary to avoid the reduction of oligosaccharides prior to the hydrazinolysis. This would allow the use of the N-acetyl group at the reducing end to obtain extra structural information. Instead of reducing the oligosaccharide ligands, a new ^{13}C labeling protocol could start with protecting the reducing end of the ligand with a short alkyl chain such as a methyl group. This

would eliminate the formation of the aldehyde group and prevent the side reaction between hydrazine and the aldehyde without destroying the terminal pyranose structure.

Another potential improvement relates to the particular C-13 NMR experiments used. So far, the ^{13}C-^{13}C RDCs and C(O) CSAs have been extracted from 1D ^{13}C spectra with a simple s2pul sequence. Very often, the ^{13}C resonances, especially carbonyl resonances, experience interference from the background signals from the protein or alignment medium. By extending the C-13 NMR experiment to a version that used a double-quantum coherence filter[1], the background resonance arising from the isolated ^{13}C sites found in natural abundance material would be suppressed. The suppression, based on the need for a directly bonded pair of ^{13}C sites to excite double quantum coherence, would be a factor or ~10,000 as opposed to the factor of ~100 coming from enrichment of a single site.

An ideal target for the methods described here is the CCL5 bound chondroitin sulfate ligand. However, obtaining proper alignment of the system has presented significant challenges. The electrostatic nature of the interactions between the protein and GAG ligands limits the choices of the alignment medium because most alignment media are charged and interact too strongly with either the positively charged protein or negatively charged GAG. Neutral polyacrylamide stretched gels produced some alignment, but not enough. There are certainly other options including neutral bicelles, and paramagnetic alignment. These could be pursued in the future, but the latter would require a redesign of the protein expression construct.

5.2 Prospects for other Protein-Associated Chondroitin Sulfate Systems

The methods developed here will provide tools not only to study the CCL5-bound CS ligand but also may be applied to some other protein-bound GAG ligands. The essential role of

protein–GAG interactions in the regulation of various physiological processes has been recognized for several decades, but the nature of these interactions is still poorly understood. Here, we propose the potential application of ^{13}C-acetyl labeling of CS oligomers to two protein-bound CS systems. One of them involves CS ligands bound to *Plasmodium falciparum*-infected erythrocytes membrane protein 1 (PfEMP1). Adhesive PfEMP1 is displayed on the surface of malaria-infected red blood cells (IRBCs)[2]. The interaction between CS and PfEMP1 plays a critical role in the disease process. For example, a unique feature of *Plasmodium falciparum* infection in pregnant women is that IRBCs sequester extensively in the placenta, causing placental malaria and in turn poor pregnancy outcomes, severe maternal anemia, and death. While CD36 on the endothelial cell surface is the major receptor for IRBC sequestration in the microvascular capillaries[3], chondroitin 4-sulfate (C4S) chains of uniquely low-sulfated chondroitin sulfate proteoglycans (CSPGs) seem to mediate placental IRBC accumulation through the interaction with PfEMP1[4,5]. A molecular understanding of the C4S–PfEMP1 interaction could be important for developing a vaccine for placental malaria.

Another potential application is to the CS ligands bound to the link module[6] of TNF[7]-stimulated gene 6 (TSG-6) protein. TSG-6 is a multifunctional protein that is up-regulated in many pathological and physiological contexts where it plays important roles in inflammation and tissue remodeling[8,9]. For example, it is a potent inhibitor of neutrophil migration and can modulate the protease network through the inhibition of plasmin. TSG-6 binds a wide range of GAGs (glycosaminoglycans) such as hyaluronan, chondroitin 4-sulphate, dermatan sulfate, heparin, and heparan sulfate, through the link module domain[10-13] and these interactions can influence the activities of TSG-6. For example, through its association with HA, TSG-6 can mediate HA cross-linking via several different mechanisms, some of which promote leukocyte

adhesion[14]. Binding to heparin, however, enhances the ability of TSG-6 to potentiate the anti-plasmin activity of inter-alpha-inhibitor, which binds noncovalently to TSG-6 via its bikunin chain[13].

Finally, the protocols we developed in this research for the geometric study of chitin and chondroitin sulfate provide the groundwork for future structural characterization of other biologically important N-acylated carbohydrates.

References

(1) Shaka, A. J.; Freeman, R. Simplification of Nmr-Spectra by Filtration through Multiple-Quantum Coherence. *Journal of Magnetic Resonance* **1983**, *51*, 169-173.

(2) Horrocks, P.; Pinches, R. A.; Chakravorty, S. J.; Papakrivos, J.; Christodoulou, Z.; Kyes, S. A.; Urban, B. C.; Ferguson, D. J. P.; Newbold, C. I. PfEMP1 expression is reduced on the surface of knobless Plasmodium falciparum infected erythrocytes. *Journal of Cell Science* **2005**, *118*, 2507-2518.

(3) Ockenhouse, C. F.; Klotz, F. W.; Tandon, N. N.; Jamieson, G. A. Sequestrin, a Cd36 Recognition Protein on Plasmodium-Falciparum Malaria-Infected Erythrocytes Identified by Antiidiotype Antibodies. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88*, 3175-3179.

(4) Gowda, A. S. P.; Madhunapantula, S. V.; Achur, R. N.; Valiyaveettil, M.; Bhavanandan, V. P.; Gowda, D. C. Chondroitin 4-sulfate mediated adherence of plasmodium falciparum in pregnancy associated malaria: design of novel photoactivable reagents for the identification of parasite adhesive proteins. *American Journal of Tropical Medicine and Hygiene* **2006**, *75*, 149-149.

(5) Gowda, A. S. P.; Madhunapantula, S. V.; Achur, R. N.; Valiyaveettil, M.; Bhavanandan, V. P.; Gowda, D. C. Structural basis for the adherence of plasmodium falciparum-infected erythrocytes to chondroitin 4-sulfate and design of novel photoactivable reagents for the identification of parasite adhesive proteins. *Journal of Biological Chemistry* **2007**, *282*, 916-928.

(6) Barta, E.; Deak, F.; Kiss, I. Evolution of the Hyaluronan-Binding Module of Link Protein. *Biochemical Journal* **1993**, *292*, 947-949.

(7) Carswell, E. A.; Old, L. J.; Kassel, R. L.; Green, S.; Fiore, N.; Williamson, B. Endotoxin-Induced Serum Factor That Causes Necrosis of Tumors. *Proceedings of the National Academy of Sciences of the United States of America* **1975**, *72*, 3666-3670.

(8) Milner, C. M.; Day, A. J. TSG-6: a multifunctional protein associated with inflammation. *Journal of Cell Science* **2003**, *116*, 1863-1873.

(9) Mahoney, D. J.; Mikecz, K.; Ali, T.; Mabilleau, G.; Benayahu, D.; Plaas, A.; Milner, C. M.; Day, A. J.; Sabokbar, A. TSG-6 Regulates Bone Remodeling through Inhibition of Osteoblastogenesis and Osteoclast Activation. *J. Biol. Chem.* **2008**, *283*, 25952-25962.

(10) Kahmann, J. D.; O'Brien, R.; Werner, J. M.; Heinegard, D.; Ladbury, J. E.; Campbell, I. D.; Day, A. J. Localization and characterization of the hyaluronan-binding site on the Link module from human TSG-6. *Structure* **2000**, *8*, 763-774.

(11) Wisniewski, H. G.; Snitkin, E. S.; Mindrescu, C.; Sweet, M. H.; Vilcek, J. TSG-6 protein binding to glycosaminoglycans - Formation of stable complexes with hyaluronan and binding to chondroitin sulfates. *Journal of Biological Chemistry* **2005**, *280*, 14476-14484.

(12) Lee, T. H.; Wisniewski, H. G.; Vilcek, J. 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. *Journal of Cell Biology* **1992**, *116*, 545-557.

(13) Milner, C. M.; Higman, V. A.; Day, A. J. TSG-6: a pluripotent inflammatory mediator? *Biochemical Society Transactions* **2006**, *34*, 446-450.

(14) Wright, A. J.; Day, A. J. Hyaluronan in immune processes. *Glycobiology and Medicine* **2005**, *564*, 57-69.