

# Synthesis, NMR Solution Structure, and Neuritogenic Activity of Chondroitin Sulfate D and E

Thesis by  
Kuang-Wei Yang

In Partial Fulfillment of the Requirements for  
the degree of  
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2018  
(Defended 23 March 2018)

© 2018

Kuang-Wei Yang

All Rights Reserved

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Linda Hsieh-Wilson, for her constant support and guidance. Without Linda, I would not have been able to finish such exciting projects. I would also like to thank my committee members, Dr. Dennis Dougherty, Dr. William Goddard, and Dr. Sarah Reisman, for their insightful comments and suggestions.

I would like to thank Dr. Andrew Almond at University of Manchester, my collaborator for the NMR structural study. Our discussion of science and world news is certainly enjoyable. Also, I need to thank Dr. Scott Virgil, Dr. David VanderVelde, Dr. Mona Shahgholi and Dr. Jeffery Pelton for their assistance. Additionally, I would like to thank all Hsieh-Wilson lab members, past and present, for your scientific input and your day-to-day companionship.

Last but not least, I would like to thank my mom and dad. Thank you for always supporting and believing in me and thank you for your unconditional love. I hope I have made you proud.

## ABSTRACT

Chondroitin sulfate are ubiquitously expressed linear, sulfated polysaccharides that play critical roles in neuronal development and regeneration growth factor signaling, morphogenesis, and virus invasion. The diverse sulfation patterns presented by chondroitin sulfate has been suggested to regulate its activity, but the structural complexity and heterogeneity have hampered the understanding of structure-activity relationship. Therefore, we envisioned that chemically synthesized chondroitin sulfate oligosaccharide may provide a unique opportunity to specifically study the functions of sulfation patterns.

Here, we report the synthesis of a CS-D and CS-E tetrasaccharide in a step-efficient manner. By generating a disaccharide precursor from hydrolysis of polysaccharides, we were able to streamline the synthesis and reduce the number of steps by one-third comparing to the traditional synthesis without losing versatility of the synthetic route and functionality of the final product. With the structurally defined molecules, we were able to determine the NMR solution structure of CS-D and CS-E. In this work, we accomplished the first structural study of CS-D tetrasaccharide and the most thorough study of CS-E to date. Furthermore, we also discovered the existence of a second conformer in CS-D, which is the first time for such behavior to be observed experimentally in chondroitin sulfate. The electrostatic potential surface constructed based on the NMR structure presented unique structural features that may allow proteins to interact specifically.

The CS-D and CS-E tetrasaccharide, along with a CS-D disaccharide, was investigated for their neuritogenic activity. We discovered that the CS-D tetrasaccharide specifically stimulates dendritic growth whereas the CS-E tetrasaccharide preferentially promoted axonal growth, revealing the potential critical role chondroitin sulfate with specific sulfation patterns may play in the nervous system. The lack of activity of the CS-D disaccharide suggested that the minimum motif required for activity of CS-D is a tetrasaccharide.

## TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract .....	iv
Table of Contents.....	v
List of Illustrations and/or Tables.....	vi
Chapter 1: Introduction to Glycosaminoglycans .....	1
Chapter 2: Chemically Synthesis of the CS-D and CS-E Tetrasaccharides.....	15
Appendix for Chapter 2: Relevant Spectral Data .....	67
Chapter 3: NMR Solution Structure of the CS-D and CS-E Tetrasaccharides .....	111
Appendix for Chapter 3: Relevant Data for Structure Determination .....	158
Chapter 4: Neuritogenic Activity of CS-D and CS-E .....	166

## LIST OF FIGURES, SCHEMES AND TABLES

	<i>Page</i>
F1.1 Structure of the five GAG families .....	3
F1.2 Structure of the four major sulfation pattern of CS .....	3
F1.3 Structure of Arixtra .....	5
F1.4 Biosynthesis of CS, DA HS and heparin .....	5
F1.5 The brush-like structure of HA-PG complex .....	7
T2.1 Disaccharide composition analysis of commercially available polysaccharides .....	16
F2.1 Target CS oligosaccharides .....	18
S2.1 Retrosynthesis of the CS-D tetrasaccharide and disaccharide .....	19
S2.2 Synthesis of the key disaccharide intermediate .....	21
S2.3 Synthesis of the disaccharide donor .....	22
T2.2 Optimization of <i>O</i> -acetate deprotection .....	23
S2.4 Synthesis of diol <b>21</b> .....	23
T2.3 Optimization of trichloroacetamide reduction .....	24
S2.5 Synthesis of the CS-D disaccharide .....	25
S2.6 Synthesis of the disaccharide acceptor .....	25
T2.4 Optimization of the coupling reaction .....	26
S2.7 Synthesis of the CS-D tetrasaccharide .....	27
S2.8 Retrosynthesis of the CS-E tetrasaccharide .....	36
S2.9 Synthesis of the CS-E disaccharide acceptor .....	28
T2.5 Optimization of the coupling reaction .....	30
S2.10 Synthesis of the CS-E tetrasaccharide .....	31
S2.11 Derivatization towards the CS-A and CS-C tetrasaccharides .....	31
F3.1 Schematic representation of the torsion angles of interest .....	114
F3.2 Spherical mapping of pyranose conformations represented by the Cremer-Pople polar coordinates .....	114

F3.3 Structure of the CS-D and CS-E tetrasaccharides.....	118
F3.4 Distribution of glycosidic angles of CS-D and CS-E .....	118
T3.1 Average glycosidic torsion angles of the CS-D tetrasaccharide	119
T3.2 Average glycosidic torsion angles of the CS-E tetrasaccharide	119
T3.3 Population of puckers of the CS-D tetrasaccharide .....	119
T3.4 Population of puckers of the CS-E tetrasaccharide .....	120
T3.5 Experimental coupling constants of CS-D .....	122
T3.6 Experimental coupling constants of CS-E .....	122
T3.7 Comparison of coupling constants of CS-D .....	125
T3.8 Comparison of coupling constants of CS-E .....	125
T3.9 Coupling constants of most abundant puckers of CS-D .....	125
T3.10 Average glycosidic torsion angles in the calculated CS-E structure .....	129
F3.5 Ensemble of the top 25 structures of CS-E .....	129
F3.6 The GalNAc2-GlcA1 linkage of the calculated structure of CS-D .....	130
T3.11 Average glycosidic torsion angles in the calculated CS-D N structure .....	134
F3.6 Ensemble of the top 25 structures of CS-D N .....	135
T3.12 Average glycosidic torsion angles in the calculated CS-D I structure .....	135
F3.6 Ensemble of the top 25 structures of CS-D I.....	135
T3.13 Low-energy conformers of CS-D .....	137
T3.13 Low-energy conformers of CS-E.....	137
F3.8 Free energy landscape contour maps of CS-D and CS-E.....	138
F3.9 Comparison of inter-residual interaction between CS-D and CS-E .....	143
F3.10 Structure of the NMR based GalNAc2-GlcA1 linkage of CS-D I.....	144
F3.11 Structure of the low-energy conformer in CS-D.....	145

F3.12 Structure of the low-energy conformer in CS-D.....	146
F3.13 Structure of the low-energy conformer in CS-E .....	147
F3.14 Structure of the low-energy conformer in CS-E .....	147
T3.15 Comparison of glycosidic torsion angles in the present study with selected published CS structures.....	148
F3.14 Electrostatic potential surface of CS-D and CS-E .....	149
TA3.1 Ring puckering definitions .....	159
TA3.2 Inter-residual NOE restraints of CS-D N .....	160
TA3.3 Inter-residual “no-NOE” restraints of CS-D N.....	161
TA3.4 Inter-residual NOE restraints of CS-D I.....	161
TA3.5 Inter-residual “no-NOE” restraints of CS-D I.....	162
TA3.6 Inter-residual NOE restraints of CS-D I.....	163
TA3.7 Inter-residual “no-NOE” restraints of CS-D I.....	164
F4.1 Representative images of CS-D and CS-E stimulated neurite outgrowth.....	171
F4.2 Quantitative analysis of neuritogenic activity of CS-D and CS-E .....	172



*Chapter 1*

## INTRODUCTION TO GLYCOSAMINOGLYCANS

## *Chapter 1*

### INTRODUCTION TO GLYCOSAMINOGLYCANS

#### **The glycosaminoglycan family**

Glycosaminoglycans (GAGs) are a diverse class of polysaccharides that are covalently attached to core proteins, forming proteoglycans (PGs). After GAGs are synthesized in the Golgi apparatus, GAGs are inserted into the plasma membrane, secreted into the extracellular matrix (ECM), or stored in intracellular granules. Although GAGs were once thought of as simply structural components of connective tissue, they are now known to mediating a myriad of cellular processes including but not limited to neuronal development<sup>1-3</sup>, neuroregeneration<sup>4-5</sup>, cancer metastasis<sup>6-9</sup> and viral invasion<sup>10-12</sup>. Structurally speaking, GAGs are linear polysaccharides composed of repeating disaccharide units. The combination of variable length, disaccharide constituent and sulfation pattern of GAGs creates extraordinary structural diversity that enables them to interact with a wide range of biological molecules and participate in diverse biological activities.

#### **Families of GAGs**

Based on the disaccharide constituents, five families of GAGs (figure 1.1) have been established, namely chondroitin sulfate (CS), dermatan sulfate (DS), heparin and heparan sulfate (HS), keratan sulfate (KS) and hyaluronic acid (HA). CS is the most abundant family in brain and in cartilage<sup>13</sup> and participates in numerous biological processes including neuronal development and regeneration<sup>5</sup>, growth factor signaling<sup>14</sup>,

morphogenesis<sup>15</sup>, and virus infection<sup>16</sup>. It is composed of repeating glucuronic acid (GlcA)-*N*-acetylgalactosamine (GalNAc) disaccharide units joined by  $\beta(1\rightarrow3)$  and  $\beta(1\rightarrow4)$  linkages, respectively. There are four major sulfation patterns of CS: CS-A, -C, -D, and -E (figure 1.2) which contain sulfate groups at the 2-*O* position of GlcA or the 4-*O* or 6-*O* positions of GalNAc. Although 3-*O* sulfation of GlcA is known as well<sup>17-18</sup>, it is less common. The expression of different CS sulfation pattern is regulated spatiotemporally by the corresponding sulfotransferases (STs), however, little is known about the exact mechanism of regulation.

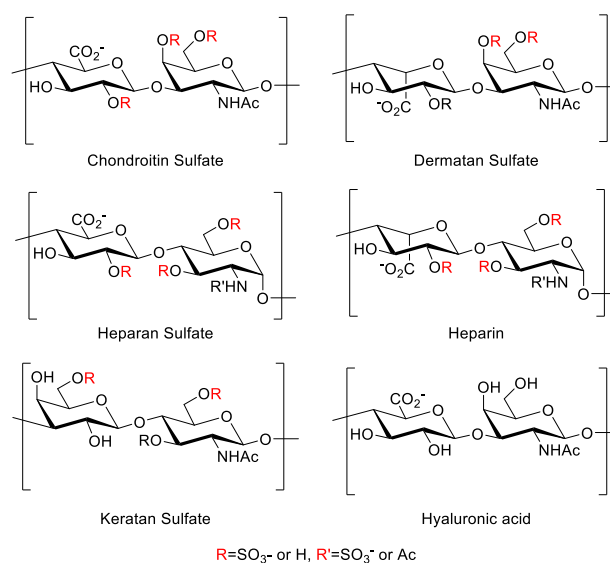


Figure 1.1. Structures of the five GAG families.

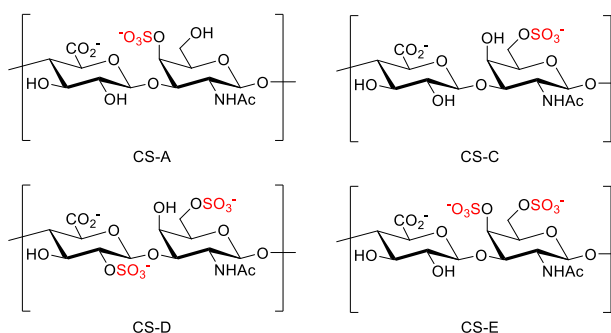


Figure 1.2. Structures of the four major sulfation pattern of CS.

The next GAG family, DS, is the predominant GAG present in skin. DS has been reported to play roles in coagulation<sup>19</sup>, cardiovascular disease<sup>20</sup>, and bacterial invasion<sup>21</sup>. It is composed of repeating iduronic acid (IdoA)-GalNAc disaccharide units joined by  $\beta(1\rightarrow3)$  and  $\beta(1\rightarrow4)$  linkages, which is identical to CS except for only one different stereocenter orientation at C5 of uronic acid. DS can be sulfated at the 2-*O* of GlcA, 4-*O*, or 6-*O* of GalNAc like CS. However, the 2-*O* sulfation of GlcA and 4-*O* sulfation of GalNAc are more common in DS<sup>22</sup>.

Together heparin and HS constitute the most well studied family of GAGs that are particularly important in anticoagulation<sup>23</sup>, cell growth and development<sup>24</sup>, angiogenesis<sup>25-26</sup>, and viral invasion<sup>27-28</sup>. Arixtra (figure 1.3), a methylated heparin pentasaccharide, is an anticoagulant medication that has a market of ~\$500 million. Heparin and HS are both composed of alternating  $\alpha(1\rightarrow4)$  *N*-acetylglucosamine (GlcNAc) and  $\beta(1\rightarrow4)$  GlcA units or alternating  $\alpha(1\rightarrow4)$  GlcNAc and  $\alpha(1\rightarrow4)$  IdoA units. They also share the same available positions for modifications, namely the sulfation at the 3-*O* and 6-*O* position of GlcNAc, sulfation of the 2-*O* position of GlcA or IdoA and *N*-sulfation or *N*-acetylation. Although heparin and HS are structurally related, they are still different in some key aspects<sup>29</sup>. Heparin is only produced by connective tissue type mast cells and has a smaller size (7 – 20 kDa), while HS is expressed ubiquitously and larger in size (10 – 70 kDa). Although both IdoA and GlcA are present in both heparin and HS, the IdoA content is markedly higher in heparin (>70%) than in HS (30% - 50%). The extent of sulfation is also different in that the number of sulfates/disaccharide ratio is 0.8 – 1.8 in HS and 1.8 – 2.6 in heparin. Heparin thereby possesses the highest negative charge density of any known biological

macromolecule<sup>30</sup>. Additionally, N-sulfation is much more common in heparin (>80%) than in HS (40% - 60%).

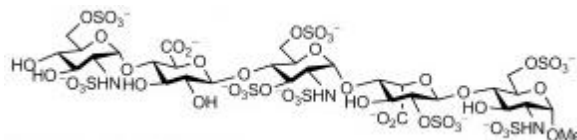


Figure 1.3. Structure of Arixtra.

CS, DS, heparin and HS are linked to the core proteins with the tetrasaccharide linker GlcA-Gal-Gal-Xylose (Xyl). As shown in figure 1.4, after the linker unit was covalently linked to a serine residue of the core protein, addition of the first hexosamine commits the GAG chain to either CS and DS (GalNAc) or HS and heparin (GlcNAc). Subsequent chain elongation, epimerization, deacetylation (HS and heparin), and sulfation of the GAG chains then generates the corresponding GAGs.

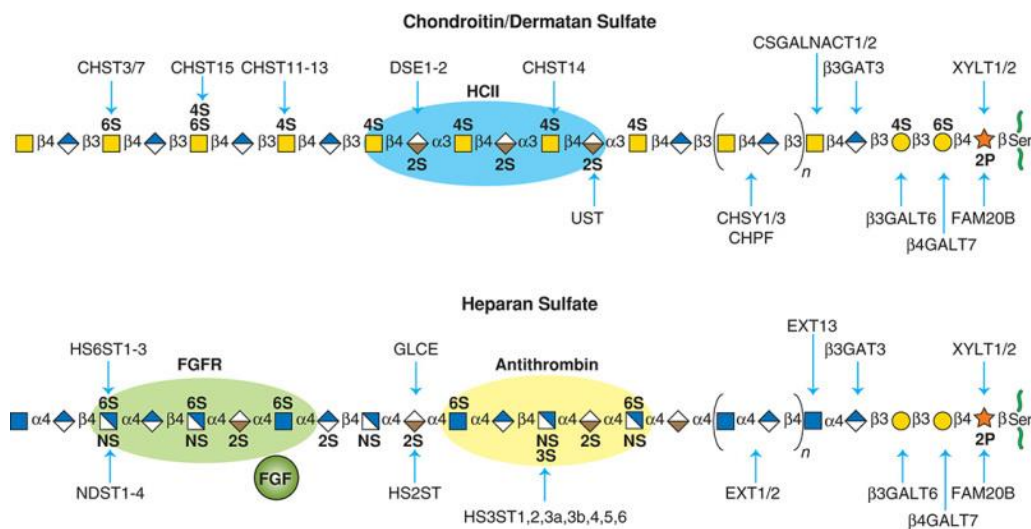


Figure 1.4. Biosynthesis of CS, DA, HS, and heparin.

KS, the last family of sulfated GAGs, is composed of repeating disaccharide units of galactose (Gal)-GlcNAc joined by  $\beta(1\rightarrow4)$  and  $\beta(1\rightarrow3)$  linkages, respectively. KSs can be classified in three categories, KS-I, -II, and -III, based on the GAG-core protein linkage<sup>31</sup>: KS-I is composed of *N*-linked chains that are abundant in cornea. KS-II is composed of chains *O*-linked through GalNAc and identified in cartilage. KS-III is extended from *O*-linked mannose and has been isolated from brain tissue. KS is related to maintaining the proper hydration levels of the cornea, which is relevant to keep the transparency of the tissue<sup>32</sup>. Aside from the principal function in cornea, KS also participates in developmental biology, cellular signaling, and migration, like CS, DS, and HS<sup>32</sup>. Both GlcNAc and Gal in KS can be sulfated at C6, but GlcNAc sulfation is most abundant<sup>33</sup>. Since KS only have two sulfation sites, the structural diversity of KS is significantly less than that of CS and HS.

HA is the only family of GAGs that is not sulfated. Interestingly, it is also the only GAG family that does not covalently bind to core proteins to form PGs. Instead, HA binds to PGs through link proteins. A single HA chain can be linked with multiple PGs and form a brush-like structure (figure 1.5). Normally, HA are composed of 2,000–25,000 disaccharide units of  $\beta(1\rightarrow3)$  GlcA  $\beta(1\rightarrow4)$  GlcNAc, which corresponds to polysaccharides with relative molecular masses of  $10^6$ – $10^7$  and polymer lengths of 2–25  $\mu\text{m}$ . HA is abundant not only in soft connective tissues, but also in epithelial and neural tissues. Due to the size and anionic nature of HA, it has a unique capacity in retaining water and is important to retaining moisture in the skin and lubricating movable parts of the body, such as joints and muscles.

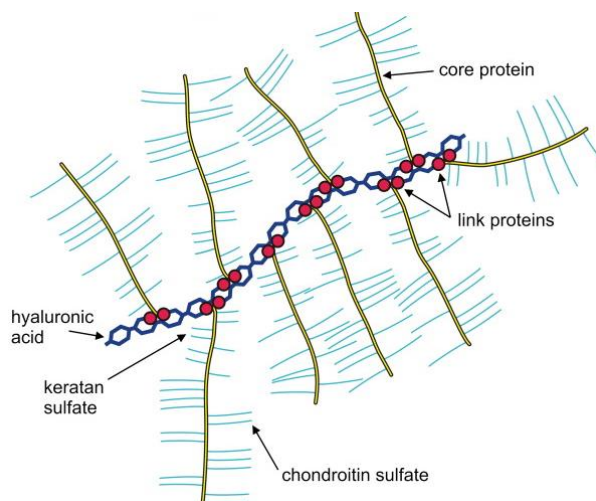


Figure 1.5. The brush-like structure of HA-PG complex.

### Tools to study the role of sulfation

Among the modifications of GAG chains, namely epimerization, deacetylation, and sulfation, sulfation is the most important, as sulfation alone can give rise to hundreds of possible tetrasaccharide structures. Also, the sulfate groups provide a charged surface which allow GAGs to electrostatically interact with proteins. Indeed, GAGs are known to bind to several hundreds of proteins, and the sulfation pattern of GAGs allow selective interactions with different molecules<sup>34</sup>. For example, in the well-studied HS-fibroblast growth factor (FGF) interaction, the binding of particular FGFs requires specific sulfation patterns. While FGF-2 requires 2-*O*-sulfation but not 6-*O*-sulfation for HS binding, FGF-10 has the reverse preference, and FGF-1 requires both 2-*O*-sulfation and 6-*O*-sulfation<sup>35</sup>. Signaling molecules in the central nervous system (CNS) such as semaphoring 3A<sup>36</sup>, midkine<sup>37-38</sup>, pleiotrophin<sup>39</sup>, and neurotrophins<sup>40</sup> have also been reported to preferentially bind to CS with specific sulfation patterns .

While diverse sulfation patterns allow GAGs to be formed and interact specifically with different proteins, complexity of the sulfation pattern also hampers the understanding of the structure-activity relationship (SAR) of GAGs. Although a number of tools are available, not all them allow the study of strictly defined sulfation motif.

Genetic and biochemical approaches such as known-down of STs or digestion of polysaccharide by lyases have established important biological implications for GAGs. However, STs do not function independently and the removal of one ST may lead to global changes of the sulfation profile, making it difficult to specify the activity of a particular sulfation pattern. On the other hand, while GAG lyases can be used to study the function of each GAG family, the impact of specific sulfation patterns cannot be determined. GAG polysaccharides and oligosaccharides can also be isolated from natural sources. However, although polysaccharides enriched in a sulfation patterns are available, they are still far from homogeneous. The oligosaccharides obtained by partial digestion of polysaccharides also suffer from heterogeneity due to the nature of lyases and the source polysaccharides. Furthermore, the inherent sulfation motifs in available polysaccharides and selectivity of the available digestion enzymes limits the variety of sulfation patterns available for study.

Our group envisioned that chemical synthesis can provide a powerful solution to these challenges as the length, sulfation pattern, and stereochemistry are all fully under control<sup>38,41-43</sup>. Aside from the ability to obtain defined oligosaccharides, the freedom of chemical synthesis also facilitates the generation of GAG-conjugated tools that are otherwise unavailable. For example, microarrays conjugated with GAG oligosaccharides allows rapid identification of GAG-protein interaction<sup>38,44</sup>. Glycopolymers with GAG



oligosaccharides attached to a polymeric backbone enable the study of specific sulfation patterns in a multivalent and polysaccharide-like context<sup>42-43,45-46</sup>. GAG oligosaccharides conjugated to immunogenic carrier proteins allow the generation of the most sulfation pattern-specific antibodies<sup>4,38,44,47</sup>.

In the present study, we achieved streamlined chemical synthesis of CS-D and CS-E tetrasaccharides. Next, we successfully calculated the solution structure of the tetrasaccharides based on nuclear magnetic resonance (NMR) data. This is the first NMR structural study of a CS-D tetrasaccharide and the most thorough study of a CS-E tetrasaccharide to date. Neuritogenic activity in hippocampal neurons of the tetrasaccharides was also investigated. We discovered that the CS-D tetrasaccharide specifically stimulates dendritic outgrowth and the CS-E tetrasaccharide specifically stimulates axonal outgrowth.

## References

- (1) Sugahara, K.; Mikami, T.; Uyama, T.; Mizuguchi, S.; Nomura, K.; Kitagawa, H., Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. *Curr. Opin. Struct. Biol.* **2003**, *13* (5), 612.
- (2) Bovolenta, P.; Feraud-Espinosa, I., Nervous system proteoglycans as modulators of neurite outgrowth. *Progress in Neurobiology* **2000**, *61* (2), 113.
- (3) Yamaguchi, Y.; Inatani, M.; Matsumoto, Y.; Ogawa, J.; Irie, F., Roles of Heparan Sulfate in Mammalian Brain Development: Current Views Based on the Findings from Ext1 Conditional Knockout Studies. *Glycosaminoglycans in Development, Health and Disease* **2010**, *93*, 133.
- (4) Brown, J. M.; Xia, J.; Zhuang, B.; Cho, K. S.; Rogers, C. J.; Gama, C. I.; Rawat, M.; Tully, S. E.; Uetani, N.; Mason, D. E.; Tremblay, M. L.; Peters, E. C.; Habuchi, O.; Chen,

D. F.; Hsieh-Wilson, L. C., A sulfated carbohydrate epitope inhibits axon regeneration after injury. *Proc Natl Acad Sci U S A* **2012**, *109* (13), 4768.

(5) Laabs, T.; Carulli, D.; Geller, H. M.; Fawcett, J. W., Chondroitin sulfate proteoglycans in neural development and regeneration. *Current Opinion in Neurobiology* **2005**, *15* (1), 116.

(6) Afratis, N.; Gialeli, C.; Nikitovic, D.; Tseganidis, T.; Karousou, E.; Theocharis, A. D.; Pavao, M. S.; Tzanakakis, G. N.; Karamanos, N. K., Glycosaminoglycans: key players in cancer cell biology and treatment. *Febs Journal* **2012**, *279* (7), 1177.

(7) Knelson, E. H.; Nee, J. C.; Blobe, G. C., Heparan sulfate signaling in cancer. *Trends in Biochemical Sciences* **2014**, *39* (6), 277.

(8) Kumar, A. V.; Gassar, E. S.; Spillmann, D.; Kiesel, L.; Yip, G. W.; Gotte, M., Specific sulfation patterns in heparan sulfate promote a proinvasive phenotype of breast cancer cells via upregulation of Wnt and MAPK signaling. *Experimental and Clinical Endocrinology & Diabetes* **2013**, *121* (3).

(9) Sanderson, R. D., Heparan sulfate proteoglycans in invasion and metastasis. *Seminars in Cell & Developmental Biology* **2001**, *12* (2), 89.

(10) Aquino, R. S.; Park, P. W., Glycosaminoglycans and infection. *Frontiers in Bioscience-Landmark* **2016**, *21*, 1260.

(11) Leistner, C. M.; Gruen-Bernhard, S.; Glebe, D., Role of glycosaminoglycans for binding and infection of hepatitis B virus. *Cellular Microbiology* **2008**, *10* (1), 122.

(12) Jinno, A.; Park, P. W., Role of glycosaminoglycans in infectious disease. *Methods Mol Biol* **2015**, *1229*, 567.

(13) Hascall, V. C.; Sajdera, S. W., Physical Properties and Polydispersity of Proteoglycan from Bovine Nasal Cartilage. *Journal of Biological Chemistry* **1970**, *245* (19), 4920.

(14) Kim, S. H.; Turnbull, J.; Guimond, S., Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor. *Journal of Endocrinology* **2011**, *209* (2), 139.

- (15) Tanaka, M.; Maeda, N.; Noda, M.; Marunouchi, T., A chondroitin sulfate proteoglycan PTP zeta/RPTP beta regulates the morphogenesis of Purkinje cell dendrites in the developing cerebellum. *Journal of Neuroscience* **2003**, 23 (7), 2804.
- (16) Banfield, B. W.; Leduc, Y.; Esford, L.; Visalli, R. J.; Brandt, C. R.; Tufaro, F., Evidence for an Interaction of Herpes-Simplex Virus with Chondroitin Sulfate Proteoglycans during Infection. *Virology* **1995**, 208 (2), 531.
- (17) 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 (34), 11063.
- (18) Sugahara, K.; Tanaka, Y.; Yamada, S.; Seno, N.; Kitagawa, H.; Haslam, S. M.; Morris, H. R.; Dell, A., Novel sulfated oligosaccharides containing 3-O-sulfated glucuronic acid from king crab cartilage chondroitin sulfate K - Unexpected degradation by chondroitinase ABC. *Journal of Biological Chemistry* **1996**, 271 (43), 26745.
- (19) Fernandez, J. A.; Petaja, J.; Griffin, J. H., Dermatan sulfate and LMW heparin enhance the anticoagulant action of activated protein C. *Thrombosis and Haemostasis* **1999**, 82 (5), 1462.
- (20) Kovanen, P. T.; Pentikainen, M. O., Decorin links low-density lipoproteins (LDL) to collagen: A novel mechanism for retention of LDL in the atherosclerotic plaque. *Trends in Cardiovascular Medicine* **1999**, 9 (3-4), 86.
- (21) Schmidtchen, A.; Frick, I. M.; Bjorck, L., Dermatan sulphate is released by proteinases of common pathogenic bacteria and inactivates antibacterial alpha-defensin. *Molecular Microbiology* **2001**, 39 (3), 708.
- (22) da Costa, D. S.; Reis, R. L.; Pashkuleva, I., Sulfation of Glycosaminoglycans and Its Implications in Human Health and Disorders. *Annual Review of Biomedical Engineering*, Vol 19 **2017**, 19, 1.
- (23) Liu, J.; Pedersen, L. C., Anticoagulant heparan sulfate: structural specificity and biosynthesis. *Applied Microbiology and Biotechnology* **2007**, 74 (2), 263.
- (24) Perrimon, N.; Bernfield, M., Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **2000**, 404 (6779), 725.

- (25) d'Angelo, I.; Oliviero, O.; Ungaro, F.; Quaglia, F.; Netti, P. A., Engineering strategies to control vascular endothelial growth factor stability and levels in a collagen matrix for angiogenesis: The role of heparin sodium salt and the PLGA-based microsphere approach. *Acta Biomaterialia* **2013**, 9 (7), 7389.
- (26) Folkman, J.; Taylor, S.; Spillberg, C., The Role of Heparin in Angiogenesis. *Ciba Foundation Symposia* **1983**, 100, 132.
- (27) Hilgard, P.; Stockert, R., Heparan sulfate proteoglycans initiate dengue virus infection of hepatocytes. *Hepatology* **2000**, 32 (5), 1069.
- (28) O'Donnell, C. D.; Shukla, D., The Importance of Heparan Sulfate in Herpesvirus Infection. *Virol Sin* **2008**, 23 (6), 383.
- (29) Esko, J. D.; Kimata, K.; Lindahl, U., Proteoglycans and Sulfated Glycosaminoglycans. In *Essentials of Glycobiology*, nd; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
- (30) Shriver, Z.; Capila, I.; Venkataraman, G.; Sasisekharan, R., Heparin and heparan sulfate: analyzing structure and microheterogeneity. *Handbook of Experimental Pharmacology* **2012**, (207), 159.
- (31) Funderburgh, J. L., Keratan sulfate biosynthesis. *IUBMB Life* **2002**, 54 (4), 187.
- (32) Funderburgh, J. L., Keratan sulfate: structure, biosynthesis, and function. *Glycobiology* **2000**, 10 (10), 951.
- (33) Prydz, K., Determinants of Glycosaminoglycan (GAG) Structure. *Biomolecules* **2015**, 5 (3), 2003.
- (34) Esko, J. D.; J, H. P.; Linhardt, R. J., Proteins That Bind Sulfated Glycosaminoglycans. In *Essentials of Glycobiology*, rd; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor (NY), 2015; pp 493.
- (35) Ashikari-Hada, S.; Habuchi, H.; Kariya, Y.; Itoh, N.; Reddi, A. H.; Kimata, K., Characterization of growth factor-binding structures in heparin/heparan sulfate using an octasaccharide library. *Journal of Biological Chemistry* **2004**, 279 (13), 12346.

- (36) Dick, G.; Tan, C. L.; Alves, J. N.; Ehlert, E. M. E.; Miller, G. M.; Hsieh-Wilson, L. C.; Sugahara, K.; Oosterhof, A.; van Kuppevelt, T. H.; Verhaagen, J.; Fawcett, J. W.; Kwok, J. C. F., Semaphorin 3A Binds to the Perineuronal Nets via Chondroitin Sulfate Type E Motifs in Rodent Brains. *Journal of Biological Chemistry* **2013**, 288 (38), 27384.
- (37) Tamura, J.; Tsutsumishita-Nakai, N.; Nakao, Y.; Kawano, M.; Kato, S.; Takeda, N.; Nadanaka, S.; Kitagawa, H., Synthesis and interaction with midkine of biotinylated chondroitin sulfate tetrasaccharides. *Bioorganic & Medicinal Chemistry Letters* **2012**, 22 (3), 1371.
- (38) Gama, C. I.; Tully, S. E.; Sotogaku, N.; Clark, P. M.; Rawat, M.; Vaidehi, N.; Goddard, W. A., 3rd; Nishi, A.; Hsieh-Wilson, L. C., Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Biol* **2006**, 2 (9), 467.
- (39) Maeda, N.; Fukazawa, N.; Hata, T., The binding of chondroitin sulfate to pleiotrophin/heparin-binding growth-associated molecule is regulated by chain length and oversulfated structures. *Journal of Biological Chemistry* **2006**, 281 (8), 4894.
- (40) Rogers, C. J.; Clark, P. M.; Tully, S. E.; Abrol, R.; Garcia, K. C.; Goddard, W. A., 3rd; Hsieh-Wilson, L. C., Elucidating glycosaminoglycan-protein-protein interactions using carbohydrate microarray and computational approaches. *Proc Natl Acad Sci U S A* **2011**, 108 (24), 9747.
- (41) Tully, S. E.; Mabon, R.; Gama, C. I.; Tsai, S. M.; Liu, X.; Hsieh-Wilson, L. C., A chondroitin sulfate small molecule that stimulates neuronal growth. *Journal of the American Chemical Society* **2004**, 126 (25), 7736.
- (42) Oh, Y. I.; Sheng, G. J.; Chang, S. K.; Hsieh-Wilson, L. C., Tailored glycopolymers as anticoagulant heparin mimetics. *Angewandte Chemie, International Edition in English* **2013**, 52 (45), 11796.
- (43) Sheng, G. J.; Oh, Y. I.; Chang, S. K.; Hsieh-Wilson, L. C., Tunable heparan sulfate mimetics for modulating chemokine activity. *Journal of the American Chemical Society* **2013**, 135 (30), 10898.
- (44) Tully, S. E.; Rawat, M.; Hsieh-Wilson, L. C., Discovery of a TNF-alpha antagonist using chondroitin sulfate microarrays. *Journal of the American Chemical Society* **2006**, 128 (24), 7740.

- (45) Rawat, M.; Gama, C. I.; Matson, J. B.; Hsieh-Wilson, L. C., Neuroactive chondroitin sulfate glycomimetics. *Journal of the American Chemical Society* **2008**, *130* (10), 2959.
- (46) Lee, S. G.; Brown, J. M.; Rogers, C. J.; Matson, J. B.; Krishnamurthy, C.; Rawat, M.; Hsieh-Wilson, L. C., End-functionalized glycopolymers as mimetics of chondroitin sulfate proteoglycans. *Chemical Science* **2010**, *1* (3), 322.
- (47) Gama, C. I.; Hsieh-Wilson, L. C., Chemical approaches to deciphering the glycosaminoglycan code. *Current Opinion in Chemical Biology* **2005**, *9* (6), 609.

## *Chapter 2*

Chemical synthesis of the CS-D and CS-E tetrasaccharides

## Chapter 2

### CHEMICAL SYNTHESIS OF THE CS-D AND CS-E TETRASACCHARIDES

#### *Introduction*

While the sulfation patterns of CS are essential to regulate specific biological functions, the chemical complexity and inherent heterogeneity of CS polysaccharides have hampered the detailed study of SAR. Although CS polysaccharides enriched in a specific sulfation motif are commercially available, the polysaccharides are far from homogeneous, as shown in table 2.1. While CS-A, CS-C, and CS-E have around 50% of the desired sulfation motifs, the CS-D polysaccharide contains only 20% of the D motif, highlighting the need for methods to obtain homogeneous CS-D.

	CS-A	CS-C	CS-D	CS-E	CS-0 <sup>a</sup>
CS-A (Sigma) <sup>b</sup>	62%	30%	ND <sup>c</sup>	0.3%	7%
CS-C (Sigma) <sup>b</sup>	27%	49%	20%	2%	3%
CS-D (Seikagaku) <sup>1</sup>	36%	43%	20%	1%	ND <sup>c</sup>
CS-E (Seikagaku) <sup>1</sup>	19%	8%	ND <sup>c</sup>	56%	8%

**Table 2.1.** Disaccharide composition analysis of commercially available polysaccharides. Suppliers are given in parenthesis. <sup>a</sup>Unsulfated chondroitin. <sup>b</sup>Unpublished data provided by Dr. Sheldon Cheung. <sup>c</sup>Not detectable.

Available methods to prepare chondroitin sulfate oligosaccharide include enzymatic digestion, chemoenzymatic synthesis, and organic synthesis. The digestion approach<sup>2-5</sup> offers a straightforward way to obtain multiple CS oligosaccharides with a few steps. The sulfation pattern of the product oligosaccharide can also be tuned by the choice of enzyme and reactant polysaccharide to some extent. However, separation of the product oligosaccharides is challenging in that high purity samples are not always available.



Furthermore, the available oligosaccharide structures are still limited by the available selectivity of enzymes and the source structure. The unsaturated reducing end generated by chondroitinase digestion may also negatively impact the SAR determination.

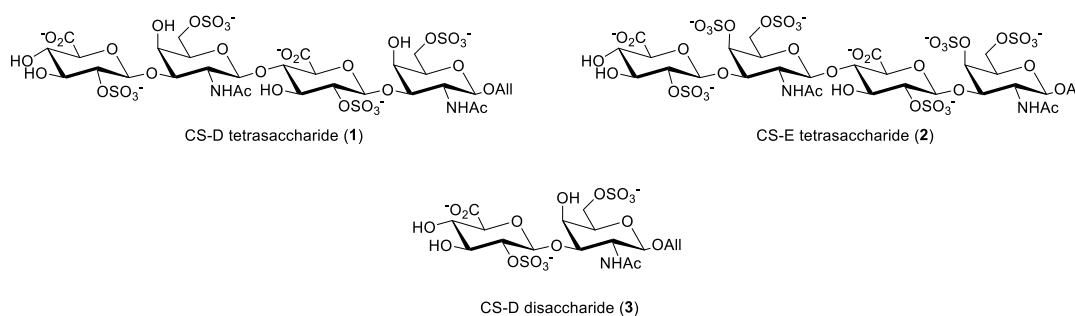
While chemoenzymatic synthesis of heparin oligosaccharide has been more extensively studied<sup>6-7</sup>, the chemoenzymatic synthesis of homogeneous CS oligosaccharide was not reported until recently<sup>8</sup>. Although the chemoenzymatic approach offers shorter synthetic routes and higher overall yields, available structures are still limited by available enzyme specificity<sup>9</sup>.

Organic synthesis of CS oligosaccharides is still the most powerful method, as virtually any desired CS structure can be prepared, even those with non-natural motifs<sup>10</sup>. However, the synthetic routes are usually long and challenging. The first reported synthesis of CS-D tetrasaccharide took 38 steps in total<sup>11</sup>, highlighting the need for improved methods to access CS oligosaccharides. The synthesis of the CS-A, CS-C, and CS-E tetrasaccharide previously reported by our group<sup>10,12</sup>, while modular and efficient, still required 27 steps to obtain the CS-E tetrasaccharide. Traditionally, the synthesis of CS begins with monosaccharide precursors

In this study, we took advantage of the selective acid hydrolysis of CS polysaccharide first reported by Levene<sup>13</sup> and Meyer and Davidson<sup>14</sup> and improved by the Jacquinet group<sup>15</sup> to rapidly access disaccharide starting material, as opposed to the traditional CS synthesis that begins with monosaccharide precursors. Part of the protecting group strategy developed by the Jacquinet group<sup>15-16</sup> was also adapted. The disaccharide starting material significantly streamlined the synthesis by reducing the number of steps required to obtain our target

molecule, the CS-D and CS-E tetrasaccharides, by about one-third. We also installed an allyl group at the reducing end of the tetrasaccharides to serve as a versatile chemical handle, which has yet to be done for CS-D. Although a CS-D tetrasaccharide with a dithiolane moiety has been reported<sup>17</sup>, no description of source, preparation, or characterization is available in the publication.

To study the 3-D structure and neuritogenic activities of the CS-D and CS-E sulfation motifs, we chose to chemically synthesize the CS-D tetrasaccharide (**1**) and CS-E tetrasaccharide (**2**). The CS-D disaccharide (**3**) was also synthesized to determine the minimum length of the CS-D motif required for activity. The CS-E disaccharide has been synthesized in a previous study and reported to have no significant neuritogenic activity<sup>12</sup>, and hence is not included in the present study.

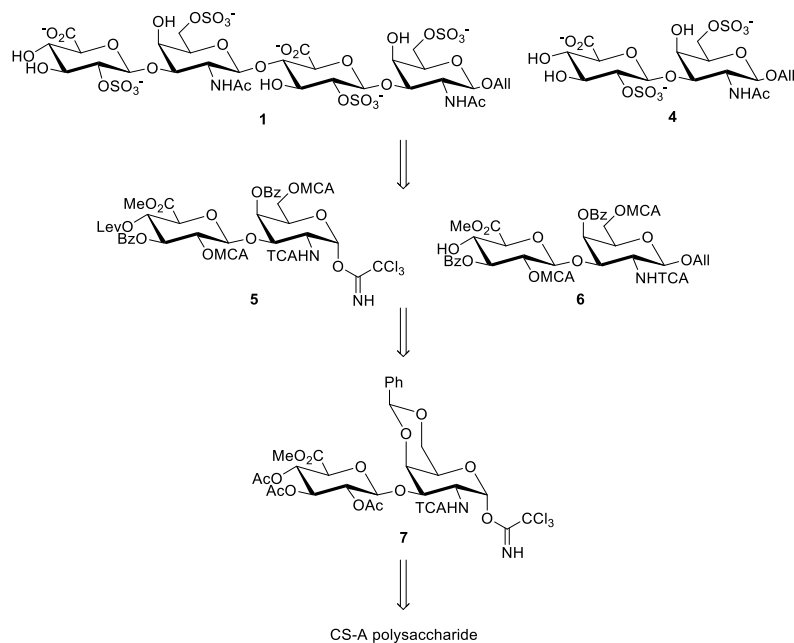


**Figure 2.1.** Target CS oligosaccharides. All = allyl.

### *Synthetic Design of the CS-D tetrasaccharide and disaccharide*

To synthesize the desired CS-D disaccharide and tetrasaccharide, we decided to utilize the polysaccharide digestion method<sup>13-14,16</sup> to obtain the key disaccharide intermediate **7**. The key disaccharide intermediate will be further derivatized to the corresponding disaccharide donor **5** and acceptor **6** for the coupling reaction. The CS-D disaccharide **4** can

also easily be derivatized from the disaccharide donor **5**. We chose to install an allyl group at the reducing end of the oligosaccharides as a convenient chemical handle for conjugation to proteins, small molecules, or surfaces, as previously developed by our group<sup>12,18</sup>. To ensure proper stereochemistry, regioselectivity, and orthogonality, the protecting groups were carefully chosen to facilitate the synthesis.



**Scheme 2.1.** Retrosynthesis of the CS-D tetrasaccharide and disaccharide. All = allyl, Me = methyl, Lev = levuliny, Bz = benzoyl, MCA = chloroacetyl, TCA = trichloroacetyl, Ph = phenyl, Ac = acetyl.

For the most important step in the synthesis, the coupling reaction, the Schmidt trichloroacetimidate method<sup>19-20</sup> was adopted for its mild reaction conditions, good stereocontrolling ability, and convenient donor preparation. To ensure  $\beta$ -glycoside formation, the neighboring participating group trichloroacetamide has been reported to be a powerful stereocontrolling auxiliary. Also, trichloroacetamides can be easily reduced to acetamides by tributyltin hydride in one step and was therefore chosen to be installed at the C2 position of GalNAc. 2-naphthylmethyl group was used to temporarily protect the C1

position of GalNAc for the selective and convenient deprotection with 2,3-Dichloro-5,6-dicyano-1,4-benzoquinone (DDQ).

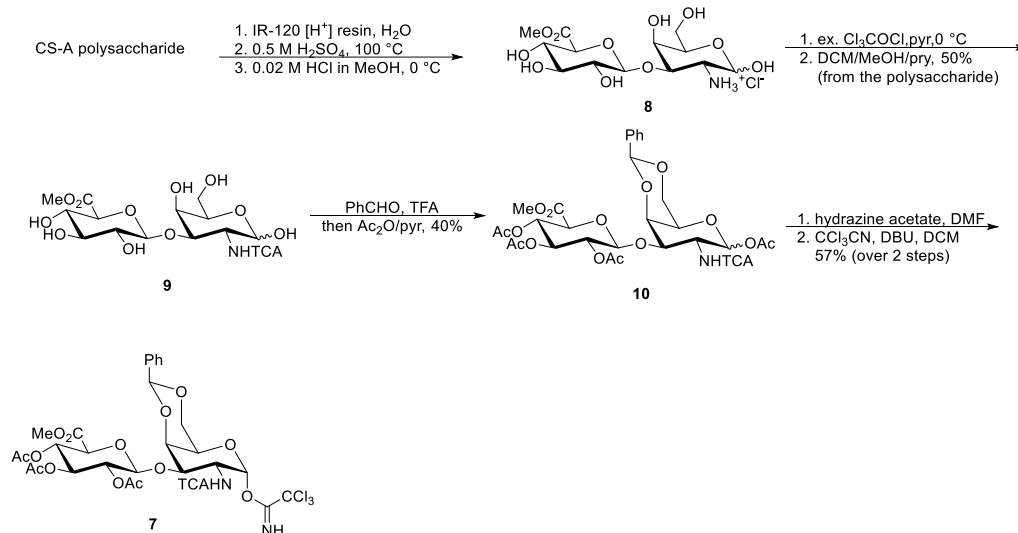
Benzoyl groups were used to protect the free hydroxyl groups of CS for its robustness because they were removed at the latest stage of the synthesis. The C4 position of GlcA is the glycosidic linkage site so a levulinyl group that can be selectively removed by hydrazine hydrate was used. The levulinyl group can also be removed by the same condition used to remove the benzoyl groups, saving one step at the final stage. For the sulfation sites of CS-D, the C2 position of GlcA, and C6 position of GalNAc, chloroacetyl groups that can be selectively cleaved by thiourea were installed. Finally, a methyl group was chosen to mask the carboxylic acid that can be deprotected along with other ester groups in the final stage.

#### *Synthesis of the CS-D disaccharide donor*

Starting from the rigorous acid hydrolysis, chondroitin polysaccharide was digested by sulfuric acid to produce the GlcA-GalNAc disaccharide with all the sulfates and acetyl groups removed. Esterification of the carboxylic acid with dilute methanolic HCl then rendered the methyl ester **8**. *N*-Trichloroacetylation was achieved by pertrichloroacetylation with an excess of trichloroacetyl chloride followed by methanolysis of the *O*-trichloroacetyl esters. Purification of compound **9** was difficult, presumably because of the high polar and the degraded products from hydrolysis. However, further experiments showed that this product doesn't need to be completely purified at this stage.

The first differentiation of the hydroxyl groups was achieved by the selective formation of the 4,6-benzilidene acetal in the presence of three other hydroxyl groups. The following peracetylation then gave the temporary protection to the rest of the hydroxyl

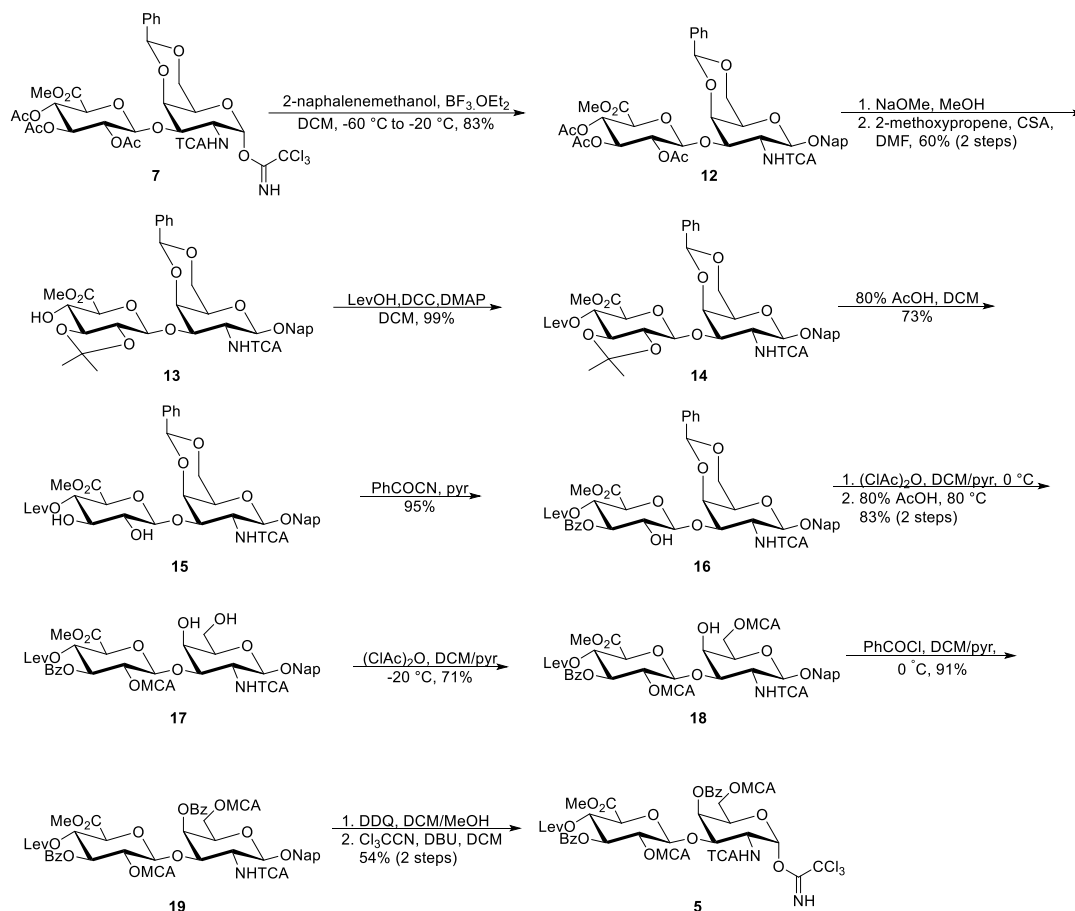
groups. Selective deprotection of the anomeric acetate with hydrazine acetate followed by treatment with 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and trichloroacetonitrile gave the  $\alpha$ -acetimidate **7** exclusively under kinetic control.



**Scheme 2.2.** Synthesis of the key disaccharide intermediate **7**. Me = methyl, TCA = trichloroacetyl, Ph = phenyl, Ac = acetyl.

Glycosylation of **7** with 2-naphthalenemethanol by catalytic Lewis acid boron trifluoride etherate gave the  $\beta$ -glycoside exclusively with good yield. 2,3,4-*O*-acetate deprotection was carried out under Zemplén transesterification condition with sodium methoxide. We tried several base loadings and reaction times after we found that the trichloroactamide could also be deprotected by sodium methoxide. The best result was obtained at 20 min duration and 0.33 equivalence of base per acetate (table 2.2). Preferential 2,3-isopropylidene formation of the triol **19** could be achieved by 2-methoxypropene and camphorsulfonic acid (CSA). Treatment of **13** with levulinic acid, dicyclohexylcarbodiimide (DCC), and 4-dimethylaminopyridine (DMAP) afforded **14** with excellent yield. The isopropylidene acetal was then removed by mild acid hydrolysis. Since removal of benzylidene acetal was observed, we quenched the reaction prematurely to avoid tetraol

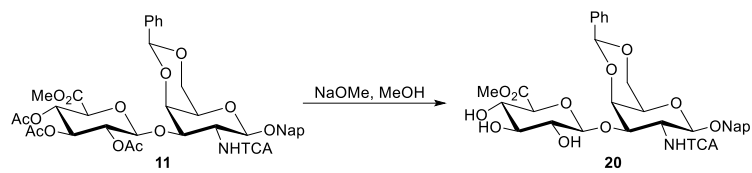
formation and recovered starting material. Therefore, the apparent yield 55% can be increased to an overall yield of 81% with two extra hydrolysis of recovered **14**.



**Scheme 2.3.** Synthesis of the disaccharide donor **5**. Me = methyl, Lev = levuliny, Bz = benzoyl, MCA = chloroacetyl, TCA = trichloroacetyl, Ph = phenyl, Ac = acetyl, Nap = 2-naphylmethyl.

Regioselective benzylation of the 3-hydroxyl group was achieved by benzylation without forming 2-*O*-benzoate. Chloroacetic anhydride was then used to mask the 2-*O*-sulfation site. After a rigorous hydrolysis, the benzylidene acetal was removed to give diol **17**. Selective chloroacetylation of the primary alcohol was achieved by slow addition of chloroacetic anhydride at low temperature. The last hydroxyl group was then protected by benzoyl chloride, affording the fully protected disaccharide **19**. Consecutive DDQ and

trichloroacetonitrile/DBU treatment then gave the disaccharide donor **5** activated at the C1 position. The CS-D disaccharide **3** and disaccharide acceptor **6** were then synthesized from the donor **5**.

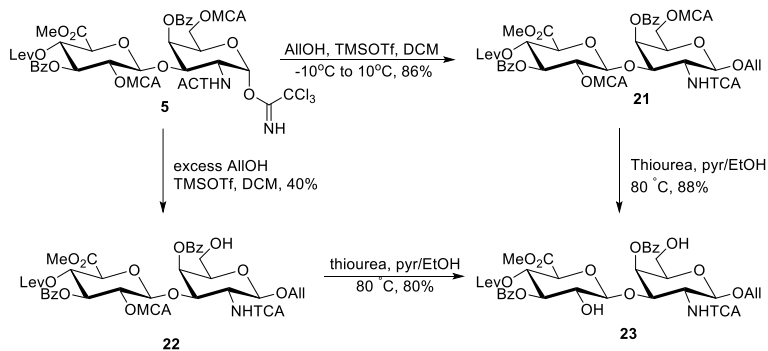


Entry	NaOMe	Reaction time	Yield
<b>1</b>	2 eq	3 h	15%
<b>2</b>	2 eq	2 h	27%
<b>3</b>	1.5 eq	1 h	40%
<b>4</b>	1 eq	40 min	70%
<b>5</b>	<b>1 eq</b>	<b>20 min</b>	<b>76%</b>

**Table 2.2.** Optimization of *O*-acetate deprotection of **12**. Me = methyl, Ac = acetyl, TCA = trichloroacetyl, Ph = phenyl, Nap = 2-naphthylmethyl.

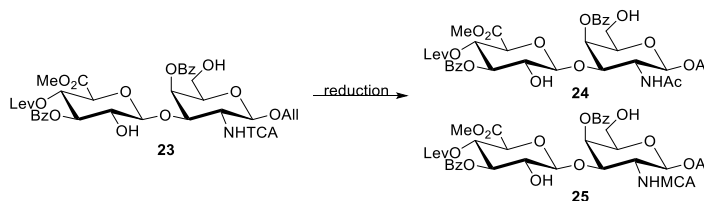
### Synthesis of the CS-D disaccharide

The disaccharide donor **5** was first treated with allyl alcohol and trimethylsilyl trifluoromethanesulfonate (TMSOTf) to give the allyl glycoside **21**. Interestingly, the 6-*O*-chloroacetate was deprotected upon the use of excess allyl alcohol, but the 2-*O*-chloroacetate remained intact. Both allyl glycoside **21** and **22** can be transformed into the diol **23** by thiourea.



**Scheme 2.4.** Synthesis of diol **21**. Me = methyl, TCA = trichloroacetyl, MCA = chloroacetyl, Lev = levulinyl, Bz = benzoyl, All = allyl.

Initial attempts to reduce the trichloroacetyl groups were unsuccessful due to the formation of partially reduced byproduct **25**. Both tributyltin hydride/azobisisobutyronitrile (AIBN) and Zn/Cu conditions were explored (table 2.3). The best result was obtained with a high equivalence of tributyltin hydride and AIBN at elevated temperature in toluene.

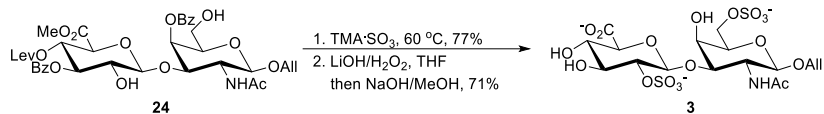


Entry	Reactant	Reagent	Condition	Product
1	<b>23</b>	6 eq. HSnBu <sub>3</sub> , 0.15 eq AIBN	80 °C in benzene, 3 h	<b>25</b>
2	<b>25</b>	10 eq. HSnBu <sub>3</sub> , 0.15 eq AIBN	80 °C in benzene, 6 h	<b>25</b>
3	<b>23</b>	15 eq. Zn/Cu	55 °C in AcOH, 24 h	<b>25/24</b> (4:1) <sup>a</sup>
4	<b>25</b>	15 eq. Zn/Cu	55 °C in AcOH, 24 h	<b>25/24</b> (4:1) <sup>a</sup>
5	<b>23 + 25</b>	15 eq. Zn/Cu + 15 eq. Zn/Cu after 24h	75 °C in AcOH, 60 h	<b>25/24</b> (1:4) <sup>ab</sup>
6	<b>23</b>	12 eq HSnBu <sub>3</sub> , 1.2 eq AIBN	110 °C in toluene	<b>24</b> (74%)

**Table 2.3.** Optimization of trichloroacetamide reduction. Me = methyl, Ac = acetyl, TCA = trichloroacetyl, Lev = levuliny, Bz = benzoyl, All = allyl. <sup>a</sup>Product ratio is determined by LC-MS. <sup>b</sup> Unidentified side product was found.

Complete sulfation was then achieved by a large excess of sulfur trioxide trimethylamine complex in 48 h. A two-step saponification process was then carried out with sequential addition of LiOH/H<sub>2</sub>O<sub>2</sub> and NaOH to avoid potential  $\beta$ -elimination at the GlcA<sup>21</sup>, affording the desired CS-D disaccharide **3**.



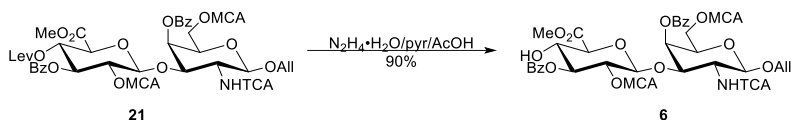


**Scheme 2.5.** Synthesis of the CS-D disaccharide **3**. Me = methyl, Ac = acetyl, TCA = trichloroacetyl, Lev = levulinyl, Bz = benzoyl, All = allyl

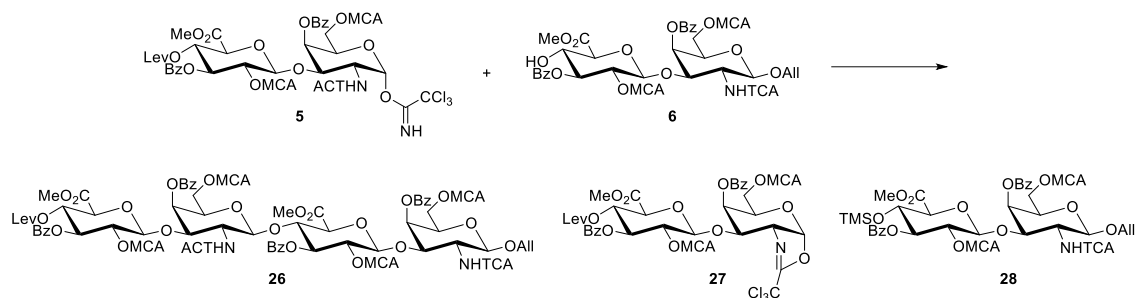
Structure of the CS-D disaccharide **3** was confirmed by NMR (see appendix for chapter 2). The doublet for the anomeric proton of GalNAc and GlcA has a  $J$  value of 7.7 and 8.0 Hz, respectively, demonstrating the  $\beta$ -glycosidic bond. The C-2 proton of GlcA and the C-6 protons of GalNAc are at 4.07 ppm and 4.16 – 4.21 ppm, respectively, indicating the sulfation sites<sup>22</sup>.

### *Synthesis of the CS-D tetrasaccharide*

The allyl glycoside **21** was first transformed to the disaccharide acceptor **6** by treatment of hydrazine hydrate, which selectively removed the levulinyl group. With the donor and acceptor in hand, we first attempted the coupling reaction with TMSOTf at room temperature. However, the donor was transformed into the oxazoline and was unable to react with the acceptor. We then tried the reaction with different Lewis acids, catalyst loading, equivalence of acceptor, temperature, and reaction time (table 2.4), but the desired tetrasaccharide was still not observed. Trimethylsilyl ether **28** was observed with TMSOTf treatment in elevated temperature (entry 3). High Bu<sub>2</sub>BOTf loading with long reaction time transformed the oxazoline to the hydrolyzed and decomposed product as determined by LC-MS (entry 10).



**Scheme 2.6.** Synthesis of the disaccharide acceptor **6**. Me = methyl, TCA = trichloroacetyl, MCA = chloroacetyl, Lev = levulinyl, Bz = benzoyl, All = allyl.



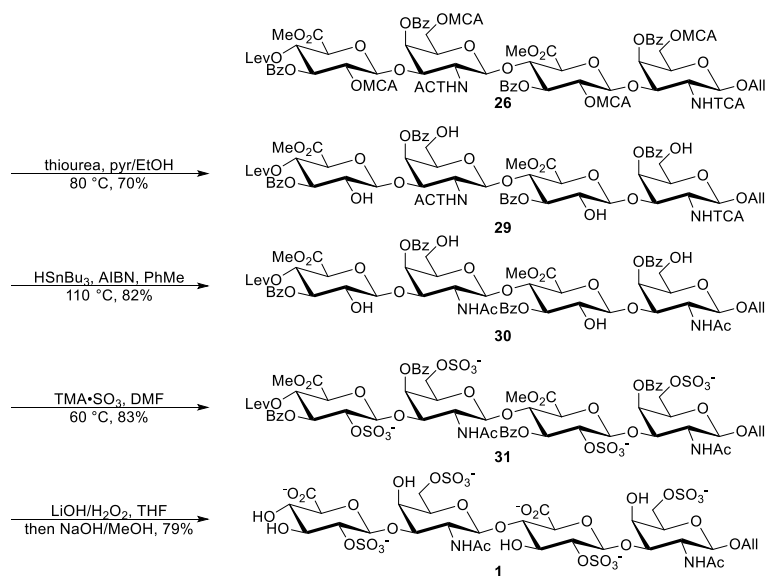
Entry	Donor <b>5</b>	Catalyst	Condition	Product
<b>1</b>	1.5 eq	TMSOTf (0.2 eq)	DCM, RT, 1 h	<b>26</b> + <b>6</b>
<b>2</b>	1.2 eq	TMSOTf (0.2 eq)	DCM, RT, 3 h	<b>26</b> + <b>6</b>
<b>3</b>	1.2 eq	TMSOTf (0.2 eq)	DCE, 40 °C, 3 h	<b>26</b> + <b>6</b> + <b>7</b>
<b>4</b>	1.2 eq	TBSOTf (0.2 eq)	DCM, RT, 3 h	<b>26</b> + <b>6</b>
<b>5</b>	1.2 eq	TBSOTf (0.2 eq)	DCE, 40 °C, 3 h	<b>26</b> + <b>6</b>
<b>6</b>	1.2 eq	BF <sub>3</sub> ·Et <sub>2</sub> O (1.0 eq)	DCE, 40 °C, 3 h	<b>26</b> + <b>6</b>
<b>7</b>	1.2 eq	Bu <sub>2</sub> BOTf (0.1 eq)	DCM, -45 °C to -20 °C, 3h	<b>26</b> + <b>6</b> + <b>5</b>
<b>8</b>	1.2 eq	Bu <sub>2</sub> BOTf (0.1 eq)	DCM, RT, 3 h	<b>26</b> + <b>6</b>
<b>9</b>	1.2 eq	Bu <sub>2</sub> BOTf (0.3 eq)	DCE, 40 °C, 3 h	<b>26</b> + <b>6</b>
<b>10</b>	1.5 eq	Bu <sub>2</sub> BOTf (0.3 eq + 0.6 eq after 12 h)	DCE, 40 °C, 16 h	<b>6</b> + byproducts <sup>a</sup>
<b>11</b>	1.2 eq	TMSOTf (0.2 eq)	DCM, RT, 2 h, in glove box	<b>25</b> (40%) + <b>6</b>
<b>12</b>	1.2 eq	TMSOTf (0.2 eq)	DCM, RT, 2 h, in Schlenk flask	<b>25</b> (58%) + <b>6</b>

**Table 2.4.** Optimization of the coupling reaction. Me = methyl, MCA = chloroacetyl, TCA = trichloroacetyl, Lev = levuliny, Bz = benzoyl, All = allyl, TMS = trimethylsilyl. <sup>a</sup>- Oxazoline derived byproducts as determined by the disappearance of oxazoline spot on TLC or peak in HPLC.

Since the initial attempts were all carried out with the presence of 4Å molecular sieves (entry 1-10), we next tried the reaction without molecular sieves in a glove box to avoid moisture. To our delight, the desired tetrasaccharide was observed and isolated (entry 11). Further attempts to carry the reaction with a Schlenk flask in a common hood were also successful (entry 12). While molecular sieves are commonly used in glycosidation reactions to prevent hydrolysis of the acetimidates resulted from trace amount of water, their slightly basic nature of may be the cause of oxazoline formation. Fortunately, by achieving strictly

dry conditions with Schlenk flasks in the absence of molecular sieves, we were able to prevent both hydrolysis and oxazoline formation.

After obtaining the tetrasaccharide **26**, thiourea treatment then selectively removed the chloroacetyl groups, exposing the sulfation sites. Radical-mediated reduction of the *N*-trichloroacetyl group with tributyltin hydride and AIBN provided **29**. Complete sulfation was achieved by a large excess of sulfur trioxide trimethylamine complex in 48 h in one step without partially sulfated product as determined by HPLC. Sequential addition of LiOH/H<sub>2</sub>O<sub>2</sub> and NaOH resulted in complete ester hydrolysis and gave the target CS-D tetrasaccharide **1**.



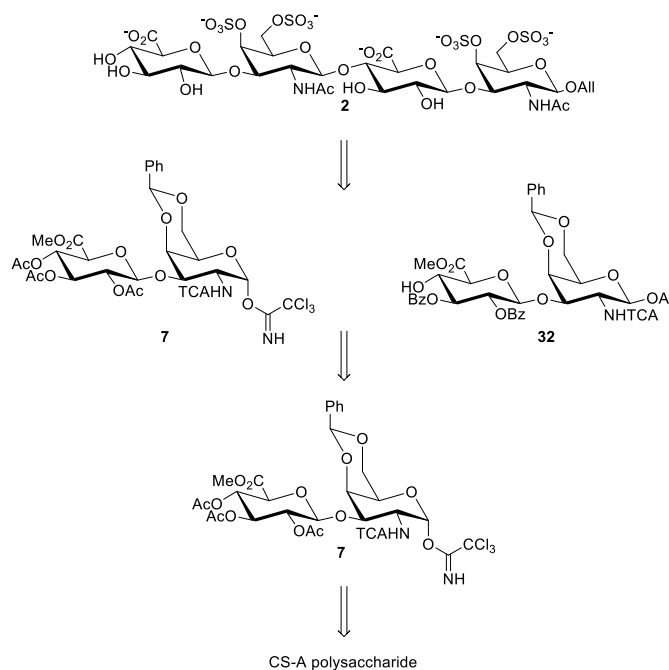
**Scheme 2.7.** Synthesis of the CS-D tetrasaccharide **1**. Me = methyl, MCA = chloroacetyl, TCA = trichloroacetyl, Lev = levulinyl, Bz = benzoyl, All = allyl.

Structure of the CS-D tetrasaccharide **1** was also confirmed by NMR (see appendix for chapter 2). The doublet for the anomeric protons of the GalNAc and GlcA units have *J* values of 8.4, 8.0, 7.3, and 7.6 Hz, all indicating the  $\beta$ -glycosidic bond of each linkage. The C-2 protons of GlcA are at 4.06 and 4.12 ppm, confirming the 2-*O*-sulfation. The C-6 protons

of GalNAc are at 4.16-4.19 ppm and 4.20 – 4.22 ppm, also fully consistent with the chemical shift of 6-*O*-sulfated C-6 protons<sup>22</sup>.

### *Synthetic design of the CS-E tetrasaccharide*

The Schmidt glycosidation and the neighboring participating group trichloroacetamide was adopted for the synthesis of CS-E tetrasaccharide as well. Since we do not attempt to synthesize oligosaccharides longer than tetrasaccharides in this study, we envisioned that the key disaccharide intermediate can be directly used as the disaccharide donor, which saved a great amount of material. The sulfation sites, the C4 and C6 position of GalNAc were already protected with a benzylidene acetal that can be selectively removed. The rest of the hydroxyl groups can stay protected with acetyl groups and do not require further differentiation. If longer oligosaccharides are desired, the donor used in the coupling reaction can still be derived from an intermediate in the synthesis of disaccharide acceptor **32** without the need to develop an independent route.



**Scheme 2.8.** Retrosynthesis of the CS-E tetrasaccharide. All = allyl, Me = methyl, Lev = levulinyl, Bz = benzoyl, TCA = trichloroacetyl, Ph = phenyl, Ac = acetyl.

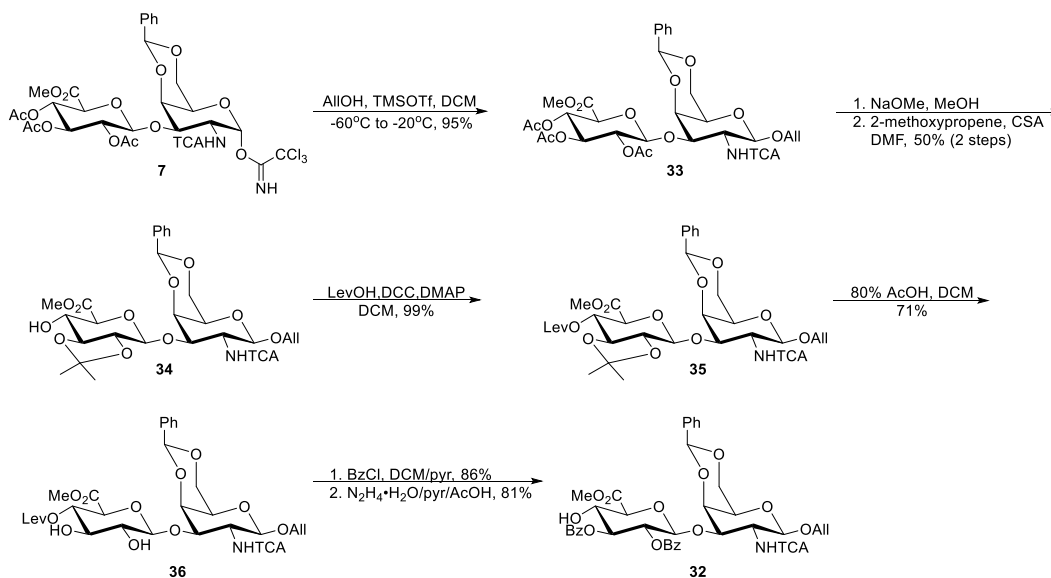
The acceptor was designed accordingly. The C4 position of GlcA was protected with a levulinyl group to facilitate the (1→4) linkage. C2 and C3 position of GlcA were protected with benzoyl groups. The C4 and C6 position of GalNAc was masked with a benzylidene acetal to allow a 1-step deprotection to expose all the sulfation sites of the CS-E tetrasaccharide. This protecting group strategy also allows the synthesis of the CS-A and CS-C tetrasaccharide, as the primary alcohol can be either selectively sulfated or protected after hydroxyl groups at the C4 and C6 position are exposed.

### *Synthesis of the CS-E tetrasaccharide*

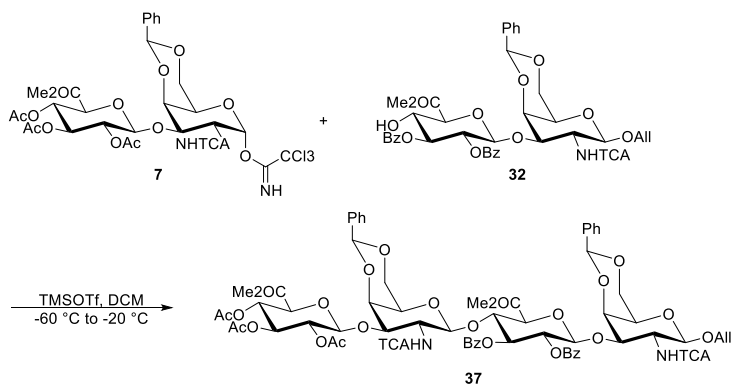
Glycosylation with allyl alcohol by catalytic TMSOTf gave the  $\beta$ -glycoside **33** exclusively with excellent yield. Deprotection of the acetate group and preferential 2,3-isopropylidene formation gave alcohol **34** smoothly. Protection of the C4 position of **34** with levulinyl group was achieved with excellent yield. The 2,3-isopropylidene was then removed with mild acid hydrolysis. Similar to the synthesis of CS-D, benzylidene deprotection was observed during the reaction. Benzoylation of the free 2,3-hydroxyl groups and subsequent deprotection of the levulinyl group then afforded the disaccharide acceptor **32**.

Optimization of the coupling reaction (table 2.5) successfully allowed the tetrasaccharide **37** to be obtained with moderate yield. Next, hydrolysis under mild acidic condition removed the benzylidene groups, exposing the sulfation sites. Treatment of  $\text{HSnBu}_3$  and AIBN in toluene/*N,N*-dimethylacetamide then reduced the trichloroacetamide group. Exhaustive sulfation was achieved by a large excess of sulfur trioxide trimethylamine complex in 48 h to give the sulfated tetrasaccharide **39**. The CS-A and CS-C tetrasaccharides can be easily derived from tetrasaccharide **39** with 3 and 2 steps, respectively, as shown in

scheme 2.11. Saponification was then carried out with LiOH/H<sub>2</sub>O<sub>2</sub> treatment followed by NaOH/MeOH to yield the CS-E tetrasaccharide **2**. It should be noted that the sulfated tetrasaccharide **39** is very acid sensitive. Near-complete decomposition was observed when the compound was left in an acidic condition (pH  $\approx$  3.5) for 12 h.



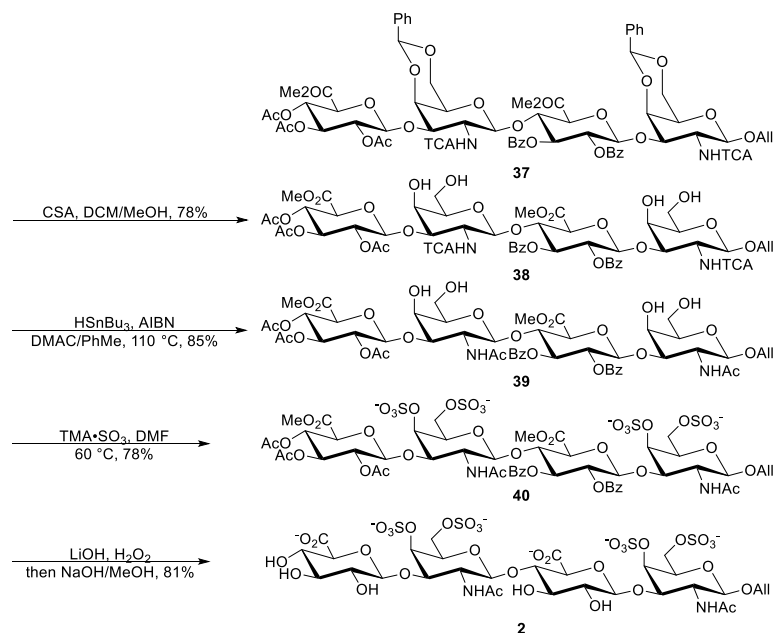
**Scheme 2.9.** Synthesis of the CS-E disaccharide acceptor **32**. All = allyl, Me = methyl, Lev = levulinyl, Bz = benzoyl, MCA = chloroacetyl, TCA = trichloroacetyl, Ph = phenyl, Ac = acetyl.



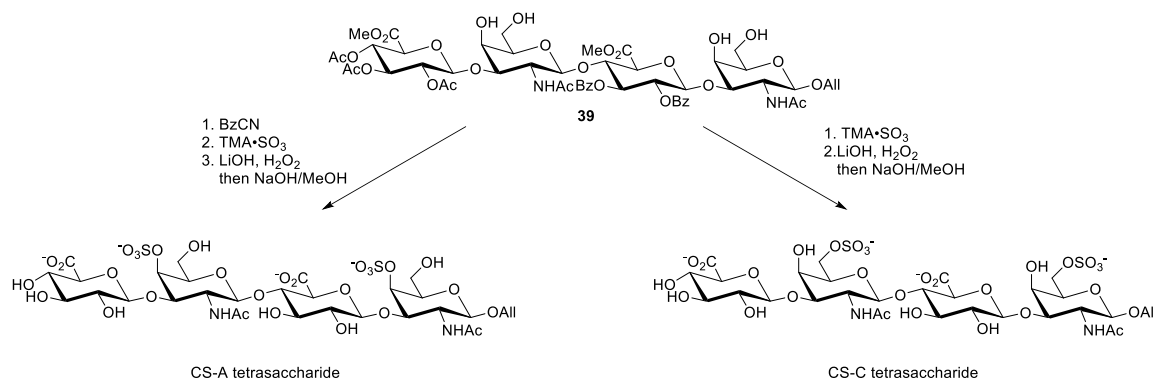
Entry	Donor 7	TMSOTf	Yield
<b>1</b>	1.3 eq	0.2 eq	20%
<b>2</b>	1.6 eq	0.2 eq	28%
<b>3</b>	2.0 eq	0.2 eq	29%
<b>4</b>	2.0 eq	0.4 eq	35%
<b>5</b>	<b>2.0 eq</b>	<b>0.6 eq</b>	<b>50%</b>

**Table 2.5.** Optimization of the coupling reaction. Me = methyl, Ac = acetyl, TCA = trichloroacetyl, Ph = phenyl, TCA = trichloroacetyl, Bz = benzoyl.

Structure of the CS-E tetrasaccharide **2** was confirmed by NMR (see appendix for Chapter 2). The doublet for the anomeric protons of the GalNAc and GlcA units have  $J$  values of 8.5, 7.9, 7.3, and 7.8 Hz, all indicating the  $\beta$ -glycosidic bond of each linkage. The characteristic C4 protons of GalNAc units are at 4.83 and 4.77 ppm, confirming the 4-*O*-sulfation<sup>22</sup>. The C6 protons of GalNAc are at 4.16–4.21 ppm and 4.24 – 4.29 ppm, also fully consistent with the chemical shift of 6-*O*-sulfated C-6 protons<sup>22</sup>.



**Scheme 2.10.** Synthesis of the CS-E tetrasaccharide **1**. Me = methyl, MCA = chloroacetyl, TCA = trichloroacetyl, Lev = levuliny, Bz = benzoyl, All = allyl.



**Scheme 2.11.** Derivatization of **39** towards the CS-A and CS-C tetrasaccharides.

## ***Conclusion***

In summary, we developed synthetic routes for CS-D and CS-E tetrasaccharides equipped with an allyl group functional handle in a step-efficient manner. By taking advantage of the disaccharide precursor generated by hydrolysis, the number of steps required to obtain the tetrasaccharides are reduced to 26 steps and 19 steps, which are roughly one-third less than the 38 steps and 27 steps required in traditional synthesis that utilized monosaccharide precursors. While the synthesis of the CS-E tetrasaccharide is shortened, our protecting group strategy still allows rapid access to the CS-A and CS-C tetrasaccharide like our previous synthesis does<sup>10</sup>. The chemically synthesized homogenous compounds allowed us to investigate the 3-D structure and neuritogenic activity of CS with the D and E sulfation pattern.

## ***Experimental Methods***

### **General Methods**

Unless stated otherwise, reactions were performed in flame dried glassware under an argon atmosphere using dry solvents (distilled or passed over a column of activated alumina). Molecular sieves were flame dried prior to use. All other commercially obtained reagents were used as received, unless otherwise noted. Thin-layer chromatography (TLC) was performed using E. Merck silica gel 60 F254 precoated plates (0.25 mm). Visualization of the developed chromatogram was performed by UV, cerium ammonium molybdate and ninhydrin stain, as necessary. ICN silica gel (particle size 0.032 – 0.063 mm) was used for flash chromatography. <sup>1</sup>H, <sup>13</sup>C NMR experiments were recorded on Varian Inova 500 (at 500 MHz), or Bruker AVANCE AV400 (at 400 MHz) and are reported relative to residual solvent peaks. Data for <sup>1</sup>H are reported as follows: chemical shift (δ ppm), multiplicity (s =



singlet, d = doublet, t = triplet, m = multiplet), coupling constant in Hz, and integration.  $^{13}\text{C}$  NMR spectra were obtained on a Varian Inova 500 (at 125 MHz) or Bruker AVANCE AV400 (at 100 MHz) and are reported in terms of chemical shift relative to residual solvent peaks. Chemical shift of  $^{13}\text{C}$  spectra were measured in aqueous solvent were determined using the absolute reference method. When necessary, proton and carbon assignments were assigned by means of  $^1\text{H}$ - $^1\text{H}$  COSY, and  $^1\text{H}$ - $^{13}\text{C}$  HSQC. High-resolution mass spectrometry was performed at the Mass Spectrometry Facility at the California Institute of Technology on either a JEOL JMS-600H High Resolution Mass Spectrometer or a UPLC-LCT Premier XT TOF Mass Spectrometer.

***O*-(Methyl  $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-amino-2-deoxy-D-galactopyranose hydrochloride (8).**

A solution of chondroitin sulfate (50 g) in water (500 ml) was acidified with Amberlite® IR120 Hydrogen form resin to pH 1.6. The resin was filtered off and washed with water (4 x 100 ml). The volume of the filtrate was adjusted to 970 ml. Concentrated  $\text{H}_2\text{SO}_4$  (18 M, 27.8 ml) was added, and the mixture was stirred for 6 h at 100 °C. After the solution was cooled down,  $\text{Ba}(\text{OH})_2$  was added portionwise under vigorous stirring to pH 3.5. The slurry was allowed to sit overnight. The solids were then filtered off through a celite pad, washed with water, and the yellow filtrate was concentrated to approximately 500 ml and slowly applied to a column of Amberlite IR-120 [ $\text{H}^+$ ] resin (500 ml, settled volume). The column was washed with water (1 L), AcOH/water (3:1, 1 L), then with aqueous HCl (1M, 3 L). The fractions containing ninhydrin-positive material were collected, concentrated, evaporated with water (2 x 500 ml), and dried under vacuum. The residue was treated with

methanolic HCl (0.02 M, 500 ml) for 4 d at 0 °C. Co-evaporation of the solution with absolute EtOH several times gave the crude product as greyish powder (36.0 g).

***O*-(Methyl  $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-deoxy-2-trichloroacetamido-D-galactopyranose (9).**

Trichloroacetyl chloride (85 ml, 0.76 mol) was slowly added to a solution of crude **8** (27 g, 0.65 mol) in pyridine (300 ml) at 0 °C and the solution was stirred at 0 °C for 1 h. Water (35 ml) was then slowly added to the solution by an addition funnel. The solution was then diluted with CH<sub>2</sub>Cl<sub>2</sub> (600 ml) and washed quickly with cold water and brine. A solution of the residue in MeOH/CH<sub>2</sub>Cl<sub>2</sub>/pyridine (1:1:1, 200 ml) was stirred for 4 h at RT and was then concentrated. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:4) gave the product (18.3 g, 50% from the polysaccharide) as a brown foam. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O, internal H<sub>2</sub>O,  $\delta$ H=4.79):  $\delta$  = 5.30 (d, J=3.8 Hz; GalNAc H-1 $\alpha$ ), 4.72 (d, J =6.3Hz; GalNAc H-1 $\beta$ ), 4.40-4.00 (m, 5H), 3.90-3.80 (m, 2H), 3.85 (s, 3H; COOCH<sub>3</sub>), 3.75- 3.35 (m, 6H).

***O*-(Methyl 2,3,4-tri-*O*-acetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-1-Oacetyl-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido-D-galactopyranose (10).**

A mixture of **9** (14.9 g, 28.9 mmol), trifluoroacetic acid (5.6 ml) and benzaldehyde (110 ml) was stirred for 24 h at RT. Anhydrous NaOAc (9.2g, 112 mmol), pyridine (110 ml) and acetic anhydride (70 ml) was then added sequentially and the mixture was stirred for 16 h. The mixture was poured into crushed ice and stirred for 2h. The mixture was then extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 x 200 ml) and the organic layer was washed with water, saturated aqueous NaHCO<sub>3</sub>, and water. The organic phase was dried over MgSO<sub>4</sub> and concentrated. Flash silica chromatography (acetone/CH<sub>2</sub>Cl<sub>2</sub> 1:15 to 1:12) gave the product as a white powder (9.0g,

40%).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 7.50-7.30 (m, 5H), 6.80 (d,  $J=7.8$  Hz, 1H; GalNAc NH), 6.47 (d,  $J_{1,2}=3.4$  Hz, 1H; GalNAc H-1), 5.53 (s, 1H; CHPh), 5.30-5.19 (m, 3H) 5.08 (m, 1H), 4.94 (d, 1H), 4.64 (m, 1H), 4.47 (m, 2H), , 4.07 (m, 3H; 2GalNAc H-6, GlcA H-5), 3.84 (s, 1H), 3.73 (s, 3H;  $\text{COOCH}_3$ ), 2.17 (s, 3H), 2.02 (s, 3H), 2.01 (s, 6H).

***O*-(Methyl 2,3,4-tri-*O*-acetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-Obenzylidene-2-deoxy-2-trichloroacetamido-1-*O*-trichloroacetimidoyl- $\alpha$ -D-galactopyranose (7).**

A mixture of **10** (11.5 g, 14.9 mmol) and hydrazine acetate (2.4 g, 25 mmol) in DMF (100 ml) was stirred for 30 min at RT. The mixture was then diluted with EtOAc (300 ml) and washed with water, brine, and water. The organic phase was dried over  $\text{MgSO}_4$  and concentrated. A mixture of the residue,  $\text{CCl}_3\text{CN}$  (13 ml, 130 mmol) and DBU (350  $\mu\text{l}$ , 2.3 mmol) in  $\text{CH}_2\text{Cl}_2$  (100 ml) was stirred for 1 h. Then more DBU was added and stirred until TLC showed completion. The solvent was removed. Flash silica chromatography (EtOAc/ $\text{CH}_2\text{Cl}_2$  1:10) and recrystallization from diethyl ether gave the product (7.4g, 57%) as a white powder.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 8.76 (s, 1H; NH), 7.55–7.30 (m, 5H; Ar-H), 6.78 (d,  $J=8.2$  Hz, 1H; GalNAc NH), 6.64 (d,  $J=3.4$  Hz, 1H; GalNAc H-1), 5.57 (s, 1H; CHPh), 5.22 (m, 2H), 5.09 (t,  $J=8.0$  Hz, 1H), 4.95 (d,  $J=7.7$  Hz, 1H), 4.77 (m, 1H), 4.55 (d,  $J=3.4$  Hz, 1H), 4.39 (dd,  $J=11.0$  Hz, 3.5 Hz, 1H), 4.33 (d, 12.7 Hz, 1H), 4.12–4.05 (m, 2H), 3.92 (s, 1H), 3.72 (s, 3H;  $\text{COOCH}_3$ ), 2.02 (s, 3H;  $\text{COCH}_3$ ), 2.01 (s, 3H;  $\text{COCH}_3$ ), 2.00 (s, 3H;  $\text{COCH}_3$ ).

**2-Naphthylmethyl *O*-(methyl 2,3,4-tri-*O*-acetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (12).**

A mixture of the trichloroacetimidate **7** (20.1 g, 23.1 mmol), 2-naphthalene methanol (5.7 g, 36.1 mmol) and 4 Å molecular sieves (2.50 g) in CH<sub>2</sub>Cl<sub>2</sub> (180 ml) was stirred under argon at RT for 1 h. The mixture was then cooled down to -60 °C before borontrifluoride etherate (300 µl, 2.43 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (2.3 ml) was added. The mixture was slowly warmed to -20 °C in 3 h and then quenched with triethylamine, filtered, and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 1:10) and gave the product (16.7 g, 83%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 7.80–7.30 (m, 12H; Ar-H), 7.05 (d, J=6.9 Hz, 1H; GalNAc NH), 5.62 (s, 1H; CHPh), 5.17 (m, 3H, 5.04(dd, J=8.8, 8.2 Hz, 1H), 4.95 (ABq, 2H; CH<sub>2</sub>Ar), 4.90 (d, 1H, J=7.6 Hz), 4.67 (dd, J=11.2, 3.5 Hz, 1H), 4.46 (d, 1H; J=3.4 Hz), 4.41 (dd, 1H; J=12.5, 1.6 Hz), 4.03 (d, J=1.8 Hz, 1H), 4.00 (d, J=9.9 Hz, 1H) 3.90 (m, 1H;), 3.71(s, 3H; COOCH<sub>3</sub>), 3.55 (s, 1H), 2.01 (s, 3H; COCH<sub>3</sub>), 2.00 (s, 3H; COCH<sub>3</sub>), 1.99 (s, 3H; COCH<sub>3</sub>).

**2-Naphthylmethyl O-(methyl 2,3-O-isopropylidene-β-D-glucopyranosyluronate)-(1→3)-4,6-O-benzylidene-2-deoxy-2-trichloroacetamido-β-D-galactopyranoside (13).**

0.5 M Methanolic sodium methoxide (2.5 ml, 1.25 mmol) was added to a solution of **12** (1.1g, 1.25 mmol) in dry THF/methanol (1:4, 12.5 ml) and the solution was stirred for 30 mins. The mixture was then neutralized with Amberlite IR-120 [H<sup>+</sup>] resin, filtered, and concentrated to give the crude triol **20**. 2-Methoxypropene (150 µl, 1.58 mmol) were added every 20 min to a solution of the crude triol **20** and CSA (88 mg, 0.38 mmol) in DMF (10 ml). Triethylamine (0.5 ml) was added and the mixture was concentrated. The residue was dissolved in EtOAc (30 ml) and washed with saturated aqueous NaHCO<sub>3</sub>, brine, and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica chromatography (EtOAc/Hexane 4:6 to 6:4, containing 0.5% triethylamine) gave the product as a white solid (587 mg, 60%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 7.85-7.30 (m, 12H; Ar-H), 7.21 (d, J=6.6 Hz, 1H; GalNAc

NH), 5.60 (s, 1H; CHPh), 5.33 (d,  $J=8.2$  Hz, 1H), 4.93 (ABq, 2H; CH<sub>2</sub>Ar), 4.79 (d,  $J=11.7$  Hz, 1H), 4.74 (dd,  $J=11.3, 3.6$  Hz, 1H), 4.47 (d,  $J=3.4$  Hz, 1H), 4.41 (dd,  $J=12.4, 1.6$  Hz, 1H), 4.13 (dd,  $J=12.4, 1.7$  Hz, 1H), 4.07 (m, 1H), 3.90 (m, 1H), 3.83-3.80 (m, 4H), 3.57 (s, 1H), 3.58-3.39 (m, 2H), 3.21 (d,  $J=2.3$  Hz, 1H), 1.41 (s, 3H), 1.38 ppm (s, 3H)

**2-Naphthylmethyl O-(methyl 2,3-*O*-isopropylidene-4-*O*-levulinoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (14).**

DCC (1.74 g, 8.3 mmol) was added portionwise to a mixture of **13** (5.22 g, 6.7 mmol), levulinic acid (661  $\mu$ l, 8.3 mmol) and DMAP (230 mg, 1.8 mmol) in dry CH<sub>2</sub>Cl<sub>2</sub> (83 ml) and stirred for 5 h. The solids were filtered off and washed with CH<sub>2</sub>Cl<sub>2</sub>/hexane (1:1) several times. The filtrate was washed with saturated aqueous NaHCO<sub>3</sub> and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 1:10, containing 0.5% triethylamine) gave the product (5.80 g, 99%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  = 7.82–7.30 (m, 12H; Ar-H), 7.17 (d,  $J=6.6$  Hz, 1H; NH), 5.62 (s, 1H; CHPh), 5.27 (m, 2H), 4.95 (ABq, 2H; CH<sub>2</sub>Ar), 4.88 (d,  $J=7.3$  Hz, 1H), 4.71 (dd,  $J=11.2$  Hz,  $J=3.3$  Hz, 1H), 4.42 (d,  $J=3.5$  Hz, 1H), 4.40 (dd,  $J=1.7, 12.5$  Hz, 1H), 4.13 (dd,  $J=1.7, 12.5$  Hz, 1H), 3.92 (d,  $J=8.3$  Hz, 1H), 3.90 (m, 1H), 3.69 (s, 3H; COOCH<sub>3</sub>), 3.62–3.52 (m, 3H), 2.85–2.55 (m, 4H), 2.18 (s, 3H; COCH<sub>3</sub>), 1.39 (s, 3H), 1.35 (s, 3H).

**2-Naphthylmethyl O-(methyl 4-*O*-levulinoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (15).**

A solution of **14** (6.90 g, 7.8 mmol) in CH<sub>2</sub>Cl<sub>2</sub>/AcOH/water (5:4:1, 140 ml) was stirred for 20 h at RT. The solution was then concentrated and azeotroped with toluene three

times. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:19, containing 0.5% triethylamine) gave the product (4.8 g, 73%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 7.85–7.32 (m, 12H; Ar-H), 7.08 (d, J=6.9 Hz, 1H; NH), 5.60 (s, 1H; CHPh), 4.96 (ABq, 2H; CH<sub>2</sub>Ar), 5.14 (d, J=8.2 Hz, 1H), 5.03 (dd, J=9.2 Hz, J=10.0 Hz, 1H), 4.58 (dd, J=3.5, 11.0 Hz, 1H), 4.51 (d, J=7.8 Hz, 1H), 4.42 (d, J=3.5 Hz, 1H), 4.40 (dd, J=1.7, 12.4 Hz, 1H), 4.12 (dd, J=1.7, 12.4 Hz, 1H), 3.96 (m, 1H), 3.92 (d, J=10 Hz, 1H), 3.72 (s, 3H; COOCH<sub>3</sub>), 3.66 (t, J=9.3 Hz, 1H), 3.54 (m, 2H), 2.80–2.55 (m, 4H), 2.17 (s, 3H; COCH<sub>3</sub>).

**2-Naphthylmethyl *O*-(methyl 3-*O*-benzoyl-4-*O*-levulinoyl-β-D-glucopyranosyluronate)-(1→3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido-β-D-galactopyranoside (16).**

A mixture of **15** (4.30g, 5.11 mmol) and benzoyl cyanide (1.68g, 12.78 mmol) in anhydrous pyridine (80 ml) was stirred for 6 h at RT. Methanol (10 ml) was then added and stirred for 5 min. The mixture was then concentrated. Flash silica chromatography (acetone/CH<sub>2</sub>Cl<sub>2</sub> 1:11, containing 0.5% triethylamine) gave the product (4.8g, 95%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 8.00–7.30 (m, 17H; Ar-H), 7.15 (d, J=6.9 Hz, 1H; NH), 5.61 (s, 1H; CHPh), 5.30 (m, 2H), 5.17 (d, J=8.0 Hz, 1H), 4.94 (ABq, 2H; CH<sub>2</sub>Ar), 4.67 (d, J=7.7 Hz, 1H), 4.62 (dd, J=3.4, 11.1 Hz), 4.45 (d, J=3.6 Hz, 1H), 4.40 (dd, J=1.7, 12.5 Hz, 1H), 4.14 (dd, J=1.7, 12.5 Hz, 1H), 4.06 (m, 1H), 3.97 (m, 1H), 3.80 (m, 1H), 3.74 (s, 3H; COOCH<sub>3</sub>), 3.56 (s, 1H), 2.65–2.30 (m, 4H), 2.03 (s, 3H; COCH<sub>3</sub>).

**2-Naphthylmethyl *O*-(methyl 3-*O*-benzoyl-2-*O*-chloroacetyl-4-*O*-levulinoyl-β-D-glucopyranosyluronate)-(1→3)-2-deoxy-2-trichloroacetamido-β-D-galactopyranoside (17).**

A mixture of **16** (4.31g, 4.55 mmol) and chloroacetic anhydride (3.1g, 18.2 mmol) in dry pyridine/ CH<sub>2</sub>Cl<sub>2</sub> (1:4, 75 ml) was stirred for 2 h at RT. Methanol (5 ml) was then added

and stirred for 5 min. The mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (100 ml) and washed with water, saturated aqueous NaHCO<sub>3</sub> and water, dried with MgSO<sub>4</sub> and concentrated. A mixture of the crude intermediate in water/AcOH (1:4) was stirred for 30 min at 80 °C. The mixture was then cooled down and concentrated. The residue was recrystallized from hot ethanol to give the product (3.53g, 83%) as a white powder. <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD): δ = 8.0–7.45 (m, 12H; Ar-H), 5.47 (t, J=9.7 Hz, 1H), 5.30 (t, J=9.7 Hz, 1H), 5.26 (dd, J=8, 9.7 Hz, 1H), 4.96 (d, J=8 Hz, 1H), 4.94 (ABq, 2H; CH<sub>2</sub>Ar), 4.64 (d, J=8.4 Hz, 1H), 4.31 (d, J= 10.0 Hz, 1H), 4.25 (dd, J=8.4, 10.8 Hz, 1H), 4.12 (ABq, 2H; COCH<sub>2</sub>Cl), 4.08 (d, J=3.2 Hz, 1H), 4.02 (dd, J=3.2, 10.8 Hz, 1H), 3.84(m, 2H), 3.75 (s, 3H; COOCH<sub>3</sub>), 3.57 (m, 1H), 2.66–2.32 (m, 4H), 2.01 (s, 3H).

**2-Naphthylmethyl O-(methyl 3-O-benzoyl-2-O-chloroacetyl-4-O-levulinoyl-β-D-glucopyranosyluronate)-(1→3)-6-O-chloroacetyl-2-deoxy-2-trichloroacetamido-β-D-galactopyranoside (18).**

A solution of chloroacetic anhydride (37 mg, 0.22 mmol) in dry CH<sub>2</sub>Cl<sub>2</sub> (0.7 ml) was added slowly at -20 °C within 50 min to a solution of **17** (137 mg, 0.16 mmol) in dry pyridine/CH<sub>2</sub>Cl<sub>2</sub> (1:10, 3.3 ml). Methanol (0.1 ml) was then added and stirred for 5 min. The mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> and washed with water, saturated aqueous NaHCO<sub>3</sub>, and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 1:9, containing 0.5% triethylamine) gave the product (105 mg, 71%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 8.00–7.35 (m, 12H; Ar-H), 6.93 (d, J=7.3 Hz, 1H; NH), 5.45 (t, J=9.3 Hz, 1H), 5.38 (dd, J=9.3 Hz, 1H), 5.22 (dd, J=7.5, 9.3 Hz, 1H), 4.94 (d, J=8.4 Hz, 1H), 4.90 (ABq, 2H; CH<sub>2</sub>Ar), 4.88 (d, J=7.5 Hz, 1H), 4.52 (m, 3H), 4.16 (m, 2H), 4.11 (s,

2H; COCH<sub>2</sub>Cl), 3.95 (ABq, 2H; COCH<sub>2</sub>Cl), 3.81 (m, 2H), 3.76 (s, 3H; COOCH<sub>3</sub>), 2.90 (s, 1H), 2.65-2.30 (m, 4H; CH<sub>2</sub>CO), 2.04 ppm (s, 3H; COCH<sub>3</sub>).

**2-Naphthylmethyl O-(methyl 3-O-benzoyl-2-O-chloroacetyl-4-O-levulinoyl-β-D-glucopyranosyluronate)-(1→3)-4-O-benzoyl-6-O-chloroacetyl-2-deoxy-2-trichloroacetamido-β-D-galactopyranoside (19).**

Benzoyl chloride (1.51 ml, 13.0 mmol) was added slowly to a solution of the **18** (3.80 g, 3.76 mmol) in dry pyridine/CH<sub>2</sub>Cl<sub>2</sub> (3:5, 40 ml) and stirred for 1 h at 0 °C. Methanol (2 ml) was then added and stirred for 5 min at 0 °C. The mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (80 ml) and washed with water, saturated aqueous NaHCO<sub>3</sub>, and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 5:95, containing 0.5% triethylamine) gave the product (3.83 g, 91%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 8.10–7.30 (m, 17H; Ar-H), 6.94 (d, J=7.7 Hz, 1H; NH), 5.67 (dd, J=3.3 Hz, 1H; GalNAc H-4), 5.38 (t, J=9.5 Hz, 1H; GlcA H-3), 5.31 (t, J=9.5 Hz, 1H; GlcA H-4), 5.10 (m, 1H; GlcA H-2), 4.97 (ABq, 2H; CH<sub>2</sub>Ar), 4.96 (d, J<sub>1,2</sub>=8.3 Hz, 1H; GalNAc H-1), 4.85 (d, J=8.0 Hz, 1H; GlcA H-1), 4.65 (dd, J=3.3, 11.0 Hz, 1H; GalNAc H-3), 4.35 (m, 2H; 2GalNAc H-6), 4.11 (s, 2H; COCH<sub>2</sub>Cl), 4.08 (m, 3H; GlcA H-5 GalNAc H-2, H-5), 3.92 (ABq, 2H; COCH<sub>2</sub>Cl), 3.72 (s, 3H; COOCH<sub>3</sub>), 2.65–2.30 (m, 4H; CH<sub>2</sub>CO), 2.02 (s, 3H; COCH<sub>3</sub>).

**O-(Methyl 3-O-benzoyl-2-O-chloroacetyl-4-O-levulinoyl-β-D-glucopyranosyluronate)-(1→3)-4-O-benzoyl-6-O-chloroacetyl-2-deoxy-2-trichloroacetamido-1-O-trichloroacetimidoyl-α-D-galactopyranose (5).**

A mixture of **19** (3.60 mg, 3.23 mmol) and DDQ (2.3 g, 10 mmol) in MeOH/CH<sub>2</sub>Cl<sub>2</sub> (1:9, 43 ml) was stirred for 24 h, then was diluted with CH<sub>2</sub>Cl<sub>2</sub> (10 ml), washed with water, saturated aqueous NaHCO<sub>3</sub>, and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica



chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 2:8, containing 0.5% triethylamine) gave the intermediate alcohol as a pale yellow solid. A mixture of the intermediate alcohol (2.70 g, 2.89 mmol), CCl<sub>3</sub>CN (2.9 ml, 29 mol) and DBU (92  $\mu$ l, 0.62  $\mu$ mol) in CH<sub>2</sub>Cl<sub>2</sub> (23.4 ml) was stirred for 1 h. Then more DBU (92  $\mu$ l, 0.62  $\mu$ mol) was added and stirred until TLC showed completion. The solvent was removed. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 1:10, containing 0.5% triethylamine) gave the product (1.94 g, 54%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  = 8.83 (s, 1H; NH), 8.10–7.30 (m, 10H; Ar-H), 7.09 (d, J=7.8 Hz, 1H; NH), 6.69 (d, J=3.5 Hz, 1H; GalNAc H-1), 5.64 (d, J<sub>3,4</sub>=3.0 Hz, 1H; GalNAc H-4), 5.47 (t, J=9.5 Hz, 1H; GlcA H-3), 5.30 (m, 2H; GlcA H-2, H-4), 4.97 (d, J=8.5 Hz, 1H; GlcA H-1), 4.77 (m, 1H; GalNAc H-2), 4.60 (dd, J=3.0, 11.0 Hz, 1H; GalNAc H-3), 4.53 (d, J=6.7 Hz, 1H; GalNAc H-5), 4.34–4.16 (m, 3H; 2GalNAc H-6 GlcA H-5), 4.09 (ABq, 2H; COCH<sub>2</sub>Cl), 4.08 (s, 2H; COCH<sub>2</sub>Cl), 3.73 (s, 3H; COOCH<sub>3</sub>), 2.68–2.30 (m, 4H; CH<sub>2</sub>CO), 2.05 ppm (s, 3H; COCH<sub>3</sub>).

**Allyl                      O-(methyl                      3-O-benzoyl-2-O-chloroacetyl-4-O-levulinoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4-O-benzoyl-6-O-chloroacetyl-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (21).**

A mixture of acetimidate **5** (320 mg, 0.286 mmol), allyl alcohol (24.3  $\mu$ l, 0.357 mmol), and 4 Å molecular sieves (270 mg) in CH<sub>2</sub>Cl<sub>2</sub> (8.8 ml) was stirred under argon at RT for 1 h. The mixture was then cooled down to -10 °C and TMSOTf in CH<sub>2</sub>Cl<sub>2</sub> (1 M, 72  $\mu$ l) was added. The mixture was slowly warmed to 10 °C within 1 h and then filtered and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 1:10) and gave the product (250 mg, 86%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  = 8.05–7.35 (m, 10H; Ar-H), 7.00 (d, J=7.5 Hz, 1H; NH), 5.88 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.69 (d, J=3.0 Hz, 1H; GalNAc H-4),

5.41 (t,  $J=9.5$  Hz, 1H; GlcA H-3), 5.33 (t,  $J=9.5$  Hz, 1H; GlcA H-4), 5.27 (m, 2H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.12 (dd,  $J=8.0, 9.5$  Hz, 1H; GlcA H-2), 4.96 (d,  $J=8.0$  Hz, 1H, GlcA H-1), 4.89 (d,  $J=7.9$  Hz, 1H; GalNAc H-1), 4.76 (dd,  $J=10.9, 3.0$  Hz, 1H; GalNAc H-3), 4.38 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.36 – 4.25 (m, 2H; GalNAc 2H-6), 4.13 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.12 (d,  $J=9.5$  Hz, 1H, GlcA H-5), 4.10 (s, 2H; COCH<sub>2</sub>Cl), 4.06 (m, 1H; GalNAc H-5), 3.95 (m, 1H; GalNAc H-2), 3.93 (ABq, 2H; COCH<sub>2</sub>Cl), 3.75 (s, 3H; COOCH<sub>3</sub>), 2.65–2.30 (m, 4H; CH<sub>2</sub>CO), 2.03 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>):  $\delta$  = 205.55, 171.02, 166.91, 166.53, 165.96, 165.55, 165.54, 162.23, 133.67, 133.48, 133.08, 133.03, 129.91, 129.11, 128.53, 128.48, 128.41, 118.70, 99.05, 98.24, 73.21, 72.66, 72.18, 72.00, 71.16, 70.64, 69.14, 68.87, 63.49, 55.59, 53.14, 40.64, 40.51, 37.57, 29.54, 27.59. HRMS:  $m/z$  calcd for [C<sub>37</sub>H<sub>40</sub>Cl<sub>3</sub>NO<sub>16</sub>+Cl]<sup>+</sup>: 1046.0533; found: 1046.0544.

**Allyl *O*-(methyl 3-*O*-benzoyl-4-*O*-levulinoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4-*O*-benzoyl-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (23).**

A mixture of **21** (26.7 mg, 0.0263 mmol) and thiourea (8.0 mg 0.105 mmol) in pyridine/EtOH (1:1, 0.4 ml) was stirred for 2 h at 80 °C, then was cooled down and concentrated. A solution of the residue in CH<sub>2</sub>Cl<sub>2</sub> (5 ml) was washed with water, saturated aqueous NaHCO<sub>3</sub>, and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:19, containing 0.5% triethylamine) gave the product (20 mg, 88%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  = 8.13 – 7.36 (m, 10H; Ar-H), 7.16 (d,  $J=7.6$  Hz, 1H, 1H; NH), 5.88 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.67 (d,  $J=3.4$  Hz, 1H; GalNAc H-4), 5.45 – 5.06 (m, 4H; GlcA H-3, H-4, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.91 (d,  $J=8.3$  Hz, 1H; GlcA H-1), 4.67 (d,  $J=7.9$  Hz, 1H; GalNAc H-1), 4.65 (dd,  $J=11.0, 3.4$  Hz, 1H; GalNAc H-3), 4.39 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.19 – 3.98 (m, 3H; GlcA H-5, GalNAc H-2,

OCH<sub>2</sub>CHCH<sub>2</sub>), 3.85 (m, 1H; GalNAc H-5), 3.80 – 3.65 (m, 5H; GlcA H-2, GalNAc H-6, COOCH<sub>3</sub>), 3.55 (m, 1H; GalNAc H-6), 2.65–2.30 (m, 4H; CH<sub>2</sub>CO), 2.00 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>): δ = 205.55, 171.02, 166.91, 166.53, 165.96, 165.55, 165.54, 162.23, 133.67, 133.48, 133.08, 133.03, 129.91, 129.11, 128.53, 128.48, 128.41, 118.70, 99.05, 98.24, 73.21, 72.66, 72.18, 72.00, 71.16, 70.64, 69.14, 68.87, 63.49, 55.59, 53.14, 40.64, 40.51, 37.57, 29.54, 27.59. HRMS: m/z calcd for [C<sub>37</sub>H<sub>40</sub>Cl<sub>3</sub>NO<sub>16</sub> + Cl]<sup>+</sup>: 894.1101; found: 894.1116.

**Allyl *O*-(methyl 3-*O*-benzoyl-β-D-glucopyranosyluronate)-(1→3)-4-*O*-benzoyl-2-deoxy-2-acetamido-β-D-galactopyranoside (24).**

A mixture of tributylstannane (34 μl, 0.126 mmol), AIBN (2.1 mg, 12.6 μmol) and **23** (10.0 mg, 11.6 μmol) in toluene (0.75 ml) was) was purged with dry Ar gas for 1 hour. The mixture was then heated to 110 °C and stirred for 1 h. The mixture was cooled down, diluted with acetonitrile, and washed with pentane. The acetonitrile layer was concentrated. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:15, containing 0.5% triethylamine) gave the product (6.5 mg, 74%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 8.18 – 7.36 (m, 10H; Ar-H), 5.98 (d, *J* = 7.3 Hz, 1H, 1H; NH), 5.90 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.61 (d, *J* = 3.3 Hz, 1H; GalNAc H-4), 5.41 – 5.15 (m, 4H; GlcA H-3, H-4, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.87 (d, *J* = 8.3 Hz, 1H; GalNAc H-1), 4.67 – 4.58 (m, 2H; GlcA H-1, GalNAc H-3), 4.37 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.14 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.09 (d, *J* = 9.7 Hz, 1H; GlcA H-5), 4.04 (m, 1H; GalNAc H-2), 3.90 (m, 1H; GalNAc H-2), 3.82 (m, 1H; GalNAc H-5), 3.77 – 3.62 (m, 5H; GlcA H-2, GalNAc H-6, COOCH<sub>3</sub>), 3.53 (m, 1H; GalNAc H-6), 3.10 (m, 1H; OH), 2.62–2.32 (m, 5H; CH<sub>2</sub>CO, OH), 2.01 (s, 3H; COCH<sub>3</sub>), 1.99 (s, 3H; NHCOCH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>): δ = 205.61, 171.73, 171.24, 167.63, 166.97, 166.39, 133.66, 133.63,

133.48, 130.21, 129.96, 129.11, 129.07, 128.46, 128.44, 117.95, 103.35, 99.66, 76.27, 74.60, 73.53, 72.45, 71.67, 70.28, 70.19, 68.92, 60.24, 54.18, 52.91, 37.67, 29.50, 27.73, 23.70. HRMS:  $m/z$  calcd for  $[C_{37}H_{43}NO_{16}+H]^+$ : 758.2660; found: 758.2661.

**Allyl (2-*O*-sulfonato- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-6-*O*-sulfonato-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside trisodium salt. (3)**

A mixture of diol (6.4 mg, 8.4  $\mu$ mol) and sulfur trioxide/trimethylamine complex (29.5 mg, 211  $\mu$ mol) was azeotroped with toluene 5 times and dried under high vacuum overnight. The mixture was then dissolved in anhydrous DMF (0.2 ml) and stirred for 48 h at 60 °C. After the mixture was cooled down, methanol (0.1 ml) was added, and the mixture was stirred for another 30 min. The mixture was then directly applied and eluted from a column (3x30 cm) of Sephadex LH-20 with 1:1  $CH_2Cl_2$ /MeOH. Flash silica chromatography (MeOH/ $CH_2Cl_2$  1:5, containing 0.5% triethylamine) followed by a column (1x15 cm) of Amberlite® IR120 Sodium form resin (1:1  $CH_2Cl_2$ /MeOH) then gave the sulfate disaccharide as a white solid (5.9 mg, 77%). NMR (500 MHz,  $CD_3OD$ ):  $\delta$  = 8.13 – 7.32 (m, 10H; Ar-H), 5.91 (m, 1H;  $OCH_2CHCH_2$ ), 5.84 (d,  $J$  = 2.8 Hz, 1H; GalNAc H-4), 5.55 (dd,  $J$  = 9.1, 8.1 Hz, 1H; GlcA H-3), 5.32 (dq,  $J$  = 17.3, 1.8 Hz, 1H;  $OCH_2CHCH_2$ ), 5.22 – 5.13 (m, 2H;  $OCH_2CHCH_2$ , GlcA H-4), 5.04 (d,  $J$  = 6.8 Hz, 1H; GalNAc H-1), 4.71 (d,  $J$  = 8.4 Hz, 1H; GlcA H-1), 4.47 (dd,  $J$  = 8.1, 6.8 Hz, 1H; GlcA H-2), 4.41 – 4.33 (m, 2H;  $OCH_2CHCH_2$ , GalNAc H-3), 4.29 (d,  $J$  = 9.8 Hz, 1H; GlcA H-5), 4.22 – 4.06 (m, 4H; GalNAc H-2, H-5, 2GalNAc H-6), 3.97 (dd,  $J$  = 10.8, 7.7 Hz, 1H; GlcA H-4), 3.60 (s, 3H;  $COOCH_3$ ), 2.66–2.29 (m, 4H;  $CH_2CO$ ), 2.05 (s, 3H;  $NHCOCH_3$ ), 2.02 (s, 3H;  $COCH_3$ ).  $^{13}C$  NMR (126 MHz,  $CD_3OD$ ):  $\delta$  = 208.55, 174.73, 172.78, 169.54, 167.33, 167.18, 135.43, 134.19, 134.01, 131.25, 131.14, 131.07, 129.38, 129.22, 117.20, 102.20, 101.29, 78.82, 78.67, 74.41, 73.69,

73.45, 71.32, 70.97, 69.72, 67.95, 58.32, 53.79, 53.28, 38.36, 29.45, 28.83, 23.43. HRMS:  $m/z$  calcd for  $[C_{37}H_{41}NNa_2O_{22}S - Na]^-$ : 948.0165; found: 948.0153.

The sulfated disaccharide (1.8 mg, 2.0  $\mu$ mol) was dissolved in THF (133  $\mu$ L) and H<sub>2</sub>O (67  $\mu$ L) and cooled to 0 °C. To this were added 1 M aq. LiOH (18  $\mu$ L) and 30% H<sub>2</sub>O<sub>2</sub> (8  $\mu$ L). The reaction was stirred at 0 °C for 1 h and at rt for 12 h. At this time, 4 M NaOH (20  $\mu$ L) and MeOH (130  $\mu$ L) were added and the reaction stirred for another 12 h. It was then neutralized with Amberlite® IR120 Hydrogen form resin, filtered, and lyophilized to afford a white solid. The product was purified by a Sephadex G-10 column (0.5x30 cm) and lyophilized to give the CS-D disaccharide **3** (0.9 mg, 71%). <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O):  $\delta$  = 5.87 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.29 (dq,  $J$  = 17.3, 1.6 Hz, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.23 (dq,  $J$  = 10.5, 1.4 Hz, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.70 (d,  $J$  = 7.7 Hz, 1H; GlcA H-1), 4.60 (d,  $J$  = 8.0 Hz, 1H; GalNAc H-1), 4.31 (ddt,  $J$  = 13.2, 5.2, 1.4 Hz, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.25 – 4.12 (m, 4H; OCH<sub>2</sub>CHCH<sub>2</sub>, GalNAc H-3, 2GalNAc H-6), 4.07 (dd,  $J$  = 9.0, 7.7 Hz, 1H; GlcA H-2), 3.98 – 3.86 (m, 3H; GalNAc H-2, H-4, H-5), 3.77 – 3.63 (m, 2H; GlcA H-3, H-5), 3.57 (t,  $J$  = 9.5 Hz, 1H; GlcA H-4). <sup>13</sup>C NMR (100 MHz, D<sub>2</sub>O):  $\delta$  = 175.72, 174.75, 133.42, 118.32, 101.26, 100.38, 79.89, 79.71, 76.06, 74.64, 73.56, 72.63, 71.52, 70.53, 67.81, 67.25, 51.18, 22.49. HRMS:  $m/z$  calcd for  $[C_{17}H_{25}NNa_2O_{18}S_2 - 2Na]^{2-}$ : 297.5262; found: 297.5272.

**Allyl *O*-(methyl 3-*O*-benzoyl-2-*O*-chloroacetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4-*O*-benzoyl-6-*O*-chloroacetyl-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (**6**).**

A mixture of pyridine/AcOH/hydrazine hydrate (6:4:0.5, 5 ml) was added to a solution of **21** (250 mg, 0.247 mmol) in pyridine (0.83 ml) and stirred for 2.5 min at RT. The mixture was then diluted with CH<sub>2</sub>Cl<sub>2</sub> (50 ml), washed with water, saturated aqueous NaHCO<sub>3</sub>, and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica chromatography

(EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 2:8, containing 0.5% triethylamine) gave the **6** (185 mg, 90%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 8.16 – 7.30 (m, 10H), 7.01 (d, *J* = 7.3 Hz, 1H; NH), 5.89 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.73 (d, *J* = 3.3 Hz, 1H; GalNAc H-4), 5.35 – 5.19 (m, 3H; GlcA H-3, OCH<sub>2</sub>CHCH<sub>2</sub>), 5.07 (dd, *J* = 9.6, 7.7 Hz, 1H; GlcA H-2), 5.02 (d, *J* = 8.2 Hz, 1H; GalNAc H-1), 4.85 (d, *J* = 7.8 Hz, 1H; GlcA H-1), 4.79 (dd, *J* = 10.9, 3.3 Hz, 1H; GalNAc H-3), 4.44 – 4.23 (m, 3H, GalNAc 2H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.14 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.08 (s, 2H; COCH<sub>2</sub>Cl), 4.10-4.04 (m, 2H; GlcA H-4, GalNAc H-5), 3.99 (d, *J* = 9.8 Hz, 1H; GlcA H-5), 3.93 (ABq, 2H; COCH<sub>2</sub>Cl), 3.90 (m, 1H; GalNAc H-2), 3.85 (s, 3H; COOCH<sub>3</sub>), 3.26 (s, 1H; OH). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>): δ = 168.56, 166.92, 166.42, 165.92, 165.68, 162.26, 133.68, 133.48, 133.10, 130.00, 129.93, 129.19, 128.62, 128.52, 128.49, 118.72, 99.78, 97.94, 92.19, 74.64, 74.37, 73.49, 72.03, 71.33, 70.65, 70.14, 69.32, 63.86, 55.93, 53.08, 40.60, 40.47. HRMS: *m/z* calcd for [C<sub>36</sub>H<sub>36</sub>Cl<sub>5</sub>NO<sub>16</sub> + Cl]<sup>-</sup>: 948.0165; found: 948.0153.

**Allyl (methyl 3-*O*-benzoyl-2-*O*-chloroacetyl-4-*O*-levulinoyl-β-D-glucopyranosyluronate)-(1→3)-(4-*O*-benzoyl-6-*O*-chloroacetyl-2-deoxy-2-trichloroacetamido-β-D-galactopyranoside)-(1→4)-(methyl 3-*O*-benzoyl-2-*O*-chloroacetyl-β-D-glucopyranosyluronate)-(1→3)-4-*O*-benzoyl-6-*O*-chloroacetyl-2-deoxy-2-trichloroacetamido-β-D-galactopyranoside (26).**

TMSOTf in CH<sub>2</sub>Cl<sub>2</sub> (100 μl, 0.23 M) was added to a mixture of **5** (134 mg, 0.12 mmol) and **6** (92 mg, 0.10 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (2 ml) and stirred for 2 hours. The mixture was quenched by triethylamine (0.05 ml) and concentrated. Flash silica chromatography (EtOAc/hexane/CH<sub>2</sub>Cl<sub>2</sub> 3:2:1) containing 0.5% triethylamine) gave the product (110 mg, 58%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 8.05 – 6.98 (m, 20H), 7.06 (d, *J* =

7.2 Hz, 1H; NH), 6.93 (d,  $J = 7.8$  Hz, 1H; NH), 5.87 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.71 (d,  $J = 3.3$  Hz, 1H; GalNAc<sup>I</sup> H-4), 5.41 (d,  $J = 3.3$  Hz, 1H; GalNAc<sup>II</sup> H-4), 5.38 – 5.21 (m, 5H; 2GlcA H-3, GlcA<sup>II</sup> H-4, 2OCH<sub>2</sub>CHCH<sub>2</sub>), 5.08 – 5.00 (m, 3H; 2GlcA H-2, GalNAc H-1), 4.98 (d,  $J = 8.4$  Hz, 1H; GalNAc H-1), 4.85 (d,  $J = 7.8$  Hz, 1H; GlcA H-1), 4.79 (dd,  $J = 10.5, 3.3$  Hz, 1H; GalNAc<sup>I</sup> H-3), 4.76 (d,  $J = 7.9$  Hz, 1H; GlcA H-1), 4.41 – 4.35 (m, 3H; GalNAc<sup>II</sup> H-3, GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.26 – 4.19 (m, 2H; GlcA<sup>I</sup> H-4, GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.15 – 3.99 (m, 10H; 4CH<sub>2</sub>Cl, OCH<sub>2</sub>CHCH<sub>2</sub>, 2GalNAc H-2, 2GlcA H-5, GalNAc H-5), 3.94 – 3.70 (m, 12H; 4CH<sub>2</sub>Cl, 6COOCH<sub>3</sub>, GalNAc H-5), 3.38 (m, 1H; GalNAc H-6), 3.13 (m, 1H; GalNAc H-6), 2.62 – 2.30 (m, 4H; CH<sub>2</sub>CO), 2.01 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>):  $\delta = 205.57, 170.95, 168.03, 166.91, 166.54, 166.46, 165.89, 165.81, 165.70, 165.50, 165.19, 165.18, 162.28, 161.71, 133.60, 133.52, 133.43, 133.40, 133.05, 130.01, 129.96, 129.85, 129.41, 129.17, 129.05, 128.93, 128.51, 128.47, 128.43, 128.40, 128.18, 118.78, 100.22, 99.25, 98.84, 97.68, 92.48, 92.11, 75.49, 74.10, 73.77, 73.65, 72.58, 72.08, 72.02, 71.90, 71.75, 71.41, 71.01, 70.63, 69.47, 69.20, 68.53, 64.11, 62.26, 55.98, 54.87, 53.39, 53.04, 40.69, 40.56, 40.44, 40.36, 37.54, 29.52, 27.55$ . HRMS:  $m/z$  calcd for [C<sub>74</sub>H<sub>72</sub>Cl<sub>10</sub>N<sub>2</sub>O<sub>33</sub>-H]<sup>+</sup>: 1865.0824; found: 1865.0750.

**Allyl (methyl 3-*O*-benzoyl-4-*O*-levulinoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-(4-*O*-benzoyl-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside)-(1 $\rightarrow$ 4)-(methyl 3-*O*-benzoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4-*O*-benzoyl-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (29).**

A mixture of **26** (11.2 mg, 6.0  $\mu$ mol), thiourea (3.7 mg, 48  $\mu$ mol) in pyridine/ethanol (1:1, 0.2 ml) was stirred at 80 °C for 2 h. The mixture was cooled down and concentrated. The residue was dissolved in CH<sub>2</sub>Cl<sub>2</sub> and washed with water and

saturated aqueous NaHCO<sub>3</sub>. The organic layer was dried with MgSO<sub>4</sub>, filtered, and concentrated. Flash silica chromatography (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 24:1) containing 0.5% triethylamine) gave the product (6.9 mg, 70%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 8.12 – 6.93 (m, 20H), 7.13 (d, *J* = 7.4 Hz, 1H; NH), 7.08 (d, *J* = 7.6 Hz, 1H; NH), 5.87 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.73 (d, *J* = 3.3 Hz, 1H; GalNAc<sup>I</sup> H-4), 5.39 (d, *J* = 3.2 Hz, 1H; GalNAc<sup>II</sup> H-4), 5.32 – 5.16 (m, 5H; 2GlcA H-3, GlcA<sup>II</sup> H-4, 2OCH<sub>2</sub>CHCH<sub>2</sub>), 4.94 (d, *J* = 8.4 Hz, 1H; GalNAc<sup>I</sup> H-1), 4.91 (d, *J* = 8.4 Hz, 1H; GalNAc<sup>II</sup> H-1), 4.67 (d, *J* = 7.7 Hz, 1H; GlcA<sup>II</sup> H-1), 4.64 (dd, *J* = 11.0, 3.3 Hz, 1H; GalNAc<sup>I</sup> H-3), 4.51 (d, *J* = 7.7 Hz, 1H; GlcA<sup>I</sup> H-1), 4.37 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.19 (dd, *J* = 10.8, 3.3 Hz, 1H; GalNAc<sup>II</sup> H-3), 4.15 – 3.91 (m, 6H; GlcA<sup>I</sup> H-4, 2GlcA H-5, 2GalNAc H-2, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.85 (s, 3H; COOCH<sub>3</sub>), 3.83 (m, 1H; GalNAc<sup>I</sup> H-5), 3.73 (m, 1H; GalNAc<sup>I</sup> H-6), 3.69 (s, 3H; COOCH<sub>3</sub>), 3.68 – 3.58 (m, 2H; 2Glc H-2), 3.56 – 3.42 (m, 2H; GalNAc<sup>II</sup> H-5, GalNAc<sup>I</sup> H-6), 3.24 (dd, *J* = 9.5, 5.8 Hz, 1H; GalNAc<sup>I</sup> OH), 2.98 – 2.90 (m, 2H; GalNAc<sup>II</sup> H-6, GlcA OH), 2.79 (dd, *J* = 8.6, 6.5 Hz, 1H; GalNAc<sup>II</sup> OH), 2.62 – 2.20 (m, 6H; GalNAc<sup>II</sup> H-6, 4CH<sub>2</sub>CO, GlcA OH), 2.00 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>): δ = 205.62, 171.11, 168.72, 167.59, 167.01, 166.70, 166.00, 165.81, 162.56, 162.50, 133.79, 133.72, 133.34, 133.19, 133.00, 130.18, 129.90, 129.59, 129.51, 129.14, 128.97, 128.84, 128.56, 128.37, 128.00, 118.46, 103.69, 103.56, 99.23, 98.76, 92.55, 92.32, 75.29, 74.75, 73.94, 73.77, 73.61, 73.58, 72.58, 71.57, 71.46, 70.51, 70.26, 70.24, 69.56, 69.12, 59.91, 59.29, 55.35, 54.87, 53.21, 52.84, 37.66, 29.49, 27.71. HRMS: *m/z* calcd for [C<sub>66</sub>H<sub>68</sub>Cl<sub>6</sub>N<sub>2</sub>O<sub>29</sub> - H]<sup>-</sup>: 1561.1961; found: 1561.1893.

**Allyl (methyl 3-*O*-benzoyl-4-*O*-levulinoyl-β-D-glucopyranosyluronate)-(1→3)-(2-acetamido-4-*O*-benzoyl-2-deoxy-β-D-galactopyranoside)-(1→4)-(methyl 3-*O*-benzoyl-**



**$\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-acetamido-4-O-benzoyl-2-deoxy- $\beta$ -D-galactopyranoside (30).**

A mixture of tributylstannane (44.0  $\mu$ l, 0.162 mmol), AIBN (3.0 mg, 18.0  $\mu$ mol) and **29** (14.0 mg, 9.0  $\mu$ mol) in toluene (1.1 ml) was purged with dry Ar gas for 1 hour. The mixture was then heated to 110 °C and stirred for 1 h. The mixture was cooled down, diluted with acetonitrile, and washed with pentane. The acetonitrile layer was concentrated. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:15, containing 0.5% triethylamine) gave the product (10.0 mg, 82%). <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD):  $\delta$  = 8.15 – 6.98 (m, 20H), 5.92 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.65 (d,  $J$  = 3.4 Hz, 1H; GalNAc<sup>I</sup> H-4), 5.44 (d,  $J$  = 3.2 Hz, 1H; GalNAc<sup>II</sup> H-4), 5.36 – 5.20 (m, 3H; 2GlcA H-3, OCH<sub>2</sub>CHCH<sub>2</sub>), 5.17 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.10 (t,  $J$  = 9.8 Hz, 1H; GlcA<sup>II</sup> H-4), 4.65 – 4.56 (m, 4H; 2GalNAc H-1, 2GlcA H-1), 4.37 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.29 (dd,  $J$  = 11.0, 3.3 Hz, 1H; GalNAc<sup>II</sup> H-2), 4.23 – 4.02 (m, 6H; 2GalNAc H-3, 2 GlcA H-5, Glc<sup>I</sup> H-4, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.89 (s, 3H; COOCH<sub>3</sub>), 3.86 – 3.76 (m, 2H; GalNAc<sup>I</sup> H-2, GalNAc<sup>I</sup> H-5), 3.68 (s, 3H; COOCH<sub>3</sub>), 3.67 – 3.42 (m, 5H; 2Glc H-2, GalNAc<sup>II</sup> H-5, 2GalNAc<sup>I</sup> H-6), 2.77 (m, 2H; 2GalNAc<sup>II</sup> H-6), 2.58 – 2.25 (m, 4H; 4CH<sub>2</sub>CO), 1.97 (s, 3H; COCH<sub>3</sub>), 1.96 (s, 3H; NHCOCH<sub>3</sub>), 1.94 (s, 3H; NHCOCH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>):  $\delta$  = 208.36, 174.28, 174.16, 172.90, 169.64, 169.15, 167.43, 167.32, 167.31, 167.28, 135.46, 134.38, 134.36, 134.34, 133.78, 131.58, 131.27, 131.09, 131.04, 131.01, 130.96, 130.87, 130.80, 129.52, 129.48, 129.47, 129.01, 117.22, 105.97, 105.21, 102.09, 101.71, 80.06, 78.43, 76.95, 76.43, 75.92, 75.89, 75.59, 75.28, 73.50, 72.59, 72.47, 71.55, 71.16, 70.97, 70.86, 62.34, 61.08, 54.28, 53.50, 53.43, 53.27, 38.28, 29.38, 28.75, 23.36, 23.09. HRMS:  $m/z$  calcd for [C<sub>66</sub>H<sub>74</sub>N<sub>2</sub>O<sub>29</sub> + Na]<sup>+</sup>: 1381.4275; found: 1381.4285.

**Allyl (methyl 3-*O*-benzoyl-4-*O*-levulinoyl-2-*O*-sulfonato- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-(2-acetamido-4-*O*-benzoyl-2-deoxy-6-*O*-sulfonato- $\beta$ -D-galactopyranoside)-(1 $\rightarrow$ 4)-(methyl 3-*O*-benzoyl-2-*O*-sulfonato- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-acetamido-4-*O*-benzoyl-6-*O*-sulfonato-2-deoxy- $\beta$ -D-galactopyranoside tetrasodium salt (31).**

A mixture of **30** (3.5 mg, 2.6  $\mu$ mol) and sulfur trioxide/trimethylamine complex (13.2 mg, 95  $\mu$ mol) was azeotroped with toluene 5 times and dried under high vacuum overnight. The mixture was then dissolved in anhydrous DMF (0.1 ml) and stirred for 48 h at 60 °C. After the mixture was cooled down, methanol (0.05 ml) was added and the mixture was stirred for another 30 min. The mixture was then directly applied and eluted from a column (3x30 cm) of Sephadex LH-20 with 1:1 CH<sub>2</sub>Cl<sub>2</sub>/MeOH. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:5, containing 0.5% triethylamine) followed by a column (1x15 cm) of Amberlite® IR120 Sodium form resin (1:1 CH<sub>2</sub>Cl<sub>2</sub>/MeOH) then gave the product as a white solid (3.8 mg, 83%). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  = 8.18 – 7.02 (m, 20H; Ar-H), 6.00 – 5.87 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.83 (d, *J* = 3.2 Hz, 1H; GalNAc<sup>I</sup> H-4), 5.67 (d, *J* = 3.2 Hz, 1H; GalNAc<sup>II</sup> H-4), 5.57 – 5.44 (m, 2H; 2GlcA H-3), 5.39 – 5.28 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.22 – 5.12 (m, 2H; OCH<sub>2</sub>CHCH<sub>2</sub>, GlcA<sup>II</sup> H-4), 5.03 (d, *J* = 6.5 Hz, 1H; GlcA H-1), 4.95 (d, *J* = 6.8 Hz, 1H; GlcA H-1), 4.73 (d, *J* = 8.4 Hz, 1H; GalNAc<sup>II</sup> H-1), 4.57 (d, *J* = 8.4 Hz, 1H; GalNAc<sup>I</sup> H-1), 4.48 – 4.36 (m, 3H; OCH<sub>2</sub>CHCH<sub>2</sub>; 2GlcA H-2), 4.36 – 4.29 (m, 2H; 2GalNAc H-3), 4.29 – 4.09 (m, 7H; OCH<sub>2</sub>CHCH<sub>2</sub>; GalNAc<sup>I</sup> H-5, H-6, GalII H-2, GlcA<sup>I</sup> H-4, 2GlcA H-5), 3.98 (dd, *J* = 10.6, 7.7 Hz, 1H; GalNAc<sup>I</sup> H-6), 3.85 – 3.79 (m, 2H; GalNAc<sup>I</sup> H-2, GalNAc<sup>II</sup> H-5), 3.77 (s, 3H; COOCH<sub>3</sub>), 3.61 (s, 3H; COOCH<sub>3</sub>), 2.65 – 2.30 (m, 4H; CH<sub>2</sub>CO), 2.07 (s, 3H; COCH<sub>3</sub>), 2.05 (s, 3H; NHCOCH<sub>3</sub>), 2.03 (s, 3H; NHCOCH<sub>3</sub>). <sup>13</sup>C

NMR (101 MHz, CD<sub>3</sub>OD):  $\delta$  = 208.60, 174.79, 174.75, 172.90, 170.05, 169.80, 167.70, 167.31, 167.18, 167.04, 135.40, 134.32, 134.11, 134.02, 133.81, 131.41, 131.19, 131.17, 131.15, 131.12, 131.08, 131.04, 129.55, 129.32, 129.22, 129.09, 117.26, 102.30, 102.15, 101.86, 101.11, 78.81, 78.56, 78.41, 77.97, 77.22, 75.30, 74.74, 74.37, 73.64, 73.29, 73.00, 71.28, 70.98, 69.94, 68.85, 67.93, 66.08, 53.98, 53.66, 53.63, 53.40, 38.31, 29.44, 28.79, 23.64, 23.52. HRMS:  $m/z$  calcd for [C<sub>66</sub>H<sub>70</sub>N<sub>2</sub>Na<sub>4</sub>O<sub>41</sub>S<sub>4</sub> - 4Na]<sup>4-</sup>: 418.5590; found: 418.5597.

**Allyl (2-*O*-sulfonato- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-(2-acetamido-2-deoxy-6-*O*-sulfonato- $\beta$ -D-galactopyranoside)-(1 $\rightarrow$ 4)-(2-*O*-sulfonato- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-acetamido-6-*O*-sulfonato-2-deoxy- $\beta$ -D-galactopyranoside hexasodium salt (1).**

**31** (5.5 mg, 3.1  $\mu$ mol) was dissolved in THF (320  $\mu$ L) and H<sub>2</sub>O (230  $\mu$ L) and cooled to 0 °C. To this were added 1 M aq. LiOH (46  $\mu$ L) and 30% H<sub>2</sub>O<sub>2</sub> (23  $\mu$ L). The reaction was stirred at 0 °C for 1 h and at rt for 12 h. At this time, 4 M NaOH (32  $\mu$ L) and MeOH (230  $\mu$ L) were added and the reaction stirred for another 12 h. It was then neutralized with Amberlite® IR120 Hydrogen form resin, filtered, and lyophilized to afford a white solid. The product was purified by a Sephadex G-25 column (0.5x30 cm) and lyophilized to give the product (3.1 mg, 79%). <sup>1</sup>H NMR (900 MHz, D<sub>2</sub>O)  $\delta$  = 5.91 – 5.84 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.31 – 5.27 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.26 – 5.22 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.71 (d,  $J$  = 7.6 Hz, 1H; Glc<sup>II</sup> H-1), 4.71 (d,  $J$  = 7.3 Hz, 1H; GlcA<sup>I</sup> H-1), 4.60 (d,  $J$  = 8.4 Hz, 1H; GalNAc<sup>I</sup> H-1), 4.57 (d,  $J$  = 7.7 Hz, 1H; GalNAc<sup>II</sup> H-2), 4.34 – 4.28 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.24 (s, 1H; GalNAc<sup>II</sup> H-4), 4.23 – 4.13 (m, 6H; OCH<sub>2</sub>CHCH<sub>2</sub>, GalNAc<sup>I</sup> H-4, 2GalNAc<sup>I</sup> H-6, 2GalNAc<sup>II</sup> H-6), 4.12 (t,  $J$  = 8.3 Hz, 1H; GlcA<sup>I</sup> H-2), 4.06 (t,  $J$  = 8.3 Hz, 1H; GlcA<sup>II</sup> H-2), 3.98 – 3.87 (m, 6H; GalNAc<sup>I</sup> H-2, H-3, H-5, GalNAc<sup>II</sup> H-2, H-3, H-5), 3.84 (t,  $J$  = 9.0 Hz, 1H; GlcA<sup>I</sup> H-4), 3.79 (t,  $J$  = 8.7 Hz, 1H; GlcA<sup>I</sup> H-3), 3.75 –

3.69 (m, 3H; GlcA<sup>I</sup> H-5, GlcA<sup>II</sup> H-3, H-5), 3.57 (t,  $J = 9.6$  Hz, 1H; GlcA<sup>II</sup> H-4). <sup>13</sup>C NMR (101 MHz, D<sub>2</sub>O):  $\delta = 175.74, 174.92, 174.80, 174.29, 133.41, 118.31, 101.73, 101.22, 101.17, 100.42, 79.73, 79.69, 79.56, 79.52, 76.56, 76.02, 74.69, 72.85, 72.64, 72.32, 71.54, 70.53, 67.89, 67.42, 67.16, 66.99, 51.21, 50.78, 22.78, 22.50$ . HRMS:  $m/z$  calcd for [C<sub>31</sub>H<sub>42</sub>N<sub>2</sub>Na<sub>6</sub>O<sub>35</sub>S<sub>4</sub> - 2Na]<sup>2-</sup>: 611.0026; found: 611.0010.

**Allyl *O*-(methyl 2,3,4-*O*-acetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (33).**

A mixture of acetimidate **5** (800.0 mg, 0.87 mmol), allyl alcohol (80.0  $\mu$ l, 1.2 mmol) and 4 Å molecular sieves (600.0 mg) in CH<sub>2</sub>Cl<sub>2</sub> (30 ml) was stirred under argon at RT for 1 h. The mixture was then cooled down to -60 °C and TMSOTf (50  $\mu$ l, 0.26) was added. The mixture was slowly warmed to RT within 1 h and then filtered and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 1:10) and gave the product (686.0 mg, 95%) as a white solid. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta = 7.54 - 7.32$  (m, 5H;Ar-H), 7.15 (d,  $J = 6.8$  Hz, 1H;NH), 5.91 – 5.80 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.57 (s, 1H;CHPh), 5.28 – 5.12 (m, 4H; 2OCH<sub>2</sub>CHCH<sub>2</sub>,GlcA H-3, H-4), 5.07 (d,  $J = 8.3$  Hz, 1H;GalNAc H-1), 5.03 (t,  $J = 8.1$  Hz, 1H;glcA H-2), 4.90 (d,  $J = 7.6$  Hz, 1H; GlcA H1), 4.69 (dd,  $J = 11.2, 3.5$  Hz, 1H; GalNAc H-3), 4.43 (dd,  $J = 3.5, 0.9$  Hz, 1H; GalNAc H-4), 4.37 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.32 (dd,  $J = 12.4, 1.6$  Hz, 1H;GalNAc H-6), 4.12 – 4.04 (m, 2H; OCH<sub>2</sub>CHCH<sub>2</sub>, GalNAc H-6), 4.01 (d,  $J = 9.6$  Hz, 1H; GlcA H-5), 3.77 (ddd,  $J = 11.2, 8.2, 6.8$  Hz, 1H; GalNAc H-2), 3.70 (s, 3H; COOCH<sub>3</sub>), 3.49 (d,  $J = 1.1$  Hz, 1H; GalNAc H-5), 2.00 (s, 3H; COCH<sub>3</sub>), 1.99 (s, 3H; COCH<sub>3</sub>), 1.98 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>):  $\delta = 169.99, 169.39, 169.17, 167.06, 162.15, 137.60, 133.51, 128.78, 128.03, 126.12, 118.08, 100.57, 100.16, 97.46,$

92.16, 75.66, 74.02, 72.30, 71.86, 71.13, 70.27, 68.97, 66.51, 55.12, 52.87, 20.76, 20.52, 20.44. HRMS:  $m/z$  calcd for  $[C_{31}H_{36}Cl_3NO_{15} + NH_4]^+$ : 785.1495; found: 785.1481.

**Allyl *O*-(methyl 2,3-*O*-isopropylidene- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (**34**).**

0.5 M Methanolic sodium methoxide (0.23 ml, 0.0115 mmol) was added to a solution of **33** (700.0 mg, 0.91 mmol) in dry THF/methanol (1:4, 9 ml) and the solution was stirred for 30 mins. The mixture was then neutralized with Amberlite IR-120  $[H^+]$  resin, filtered, and concentrated to give the crude triol. 2-Methoxypropene (95  $\mu$ l, 1.0 mmol) were added every 20 min to a solution of the crude triol **20** and CSA (56.0 mg, 0.24 mmol) in DMF (6.5 ml). Triethylamine (0.5 ml) was added and the mixture was concentrated. The residue was dissolved in EtOAc (30 ml) and washed with saturated aqueous  $NaHCO_3$ , brine, and water, dried with  $MgSO_4$ , and concentrated. Flash silica chromatography (EtOAc/Hexane 4:6 to 6:4, containing 0.5% triethylamine) gave the product as a white solid (310.0 mg, 50%).  $^1H$  NMR (500 MHz,  $CDCl_3$ ):  $\delta$  = 7.56 – 7.30 (m, 5H; Ar-H), 7.22 (d,  $J$  = 6.6 Hz, 1H; NH), 5.93 – 5.84 (m, 1H;  $OCH_2CHCH_2$ ), 5.58 (s, 1H; CHPH), 5.30 – 5.25 (m, 1H;  $OCH_2CHCH_2$ ), 5.22 (d,  $J$  = 8.2 Hz, 1H; GalNAc H-1), 5.21 – 5.17 (m, 1H;  $OCH_2CHCH_2$ ), 4.94 (d,  $J$  = 7.7 Hz, 1H; GlcA H-1), 4.76 (dd,  $J$  = 11.2, 3.5 Hz, 1H; GalNAc H-3), 4.46 (dd,  $J$  = 3.5, 1.0 Hz, 1H; GalNAc H-4), 4.43 – 4.37 (m, 1H;  $OCH_2CHCH_2$ ), 4.35 (dd,  $J$  = 12.3, 1.6 Hz, 1H; GalNAc H-6), 4.15 – 4.05 (m, 3H;  $OCH_2CHCH_2$ ; GalNAc H-6, GlcA H-6), 3.87 – 3.78 (m, 5H;  $COOCH_3$ , GalNAc H-3, GlcA H-5), 3.56 – 3.54 (m, 1H; GalNAc H-5), 3.52 (t,  $J$  = 9.5 Hz, 1H; GlcA H-3), 3.43 (dd,  $J$  = 9.3, 7.7 Hz, 1H; GlcA H-2), 3.24 (d,  $J$  = 2.2 Hz, 1H; GlcA 4-OH), 1.43 (s, 3H;  $CH_3$ ), 1.40 (s, 3H;  $CH_3$ ).  $^{13}C$  NMR (126 MHz,  $CDCl_3$ ):  $\delta$  = 169.71, 162.34, 137.66, 133.62, 128.88, 128.01, 126.26, 118.08, 112.30, 102.64, 100.74, 97.39,

92.21, 79.34, 76.03, 75.68, 75.56, 74.59, 70.77, 70.41, 69.16, 66.57, 55.50, 52.95, 26.65, 26.47. HRMS:  $m/z$  calcd for  $[C_{28}H_{34}Cl_3NO_{12} + NH_4]^+$ : 699.1491; found: 699.1460.

**Allyl *O*-(methyl 2,3-*O*-isopropylidene-4-*O*-levulinoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (35).**

DCC (305 mg, 1.48 mmol) was added portionwise to a mixture of **34** (805.6 mg, 1.18 mmol), levulinic acid (150  $\mu$ l, 1.48 mmol) and DMAP (36 mg, 0.3 mmol) in dry  $CH_2Cl_2$  (23 ml) and stirred for 5 h. The solids were filtered off and washed with  $CH_2Cl_2$ /hexane (1:1) several times. The filtrate was washed with saturated aqueous  $NaHCO_3$  and water, dried with  $MgSO_4$ , and concentrated. Flash silica chromatography (EtOAc/ $CH_2Cl_2$  1:10, containing 0.5% triethylamine) gave the product (971 mg, 99%) as a white solid.  $^1H$  NMR (400 MHz,  $CDCl_3$ ):  $\delta$  = 7.57 – 7.30 (m, 5H; Ar-H), 7.18 (d,  $J$  = 6.6 Hz, 1H; NH), 5.94 – 5.82 (m, 1H;  $OCH_2CHCH_2$ ), 5.60 (s, 1H), 5.32 – 5.23 (m, 2H;  $OCH_2CHCH_2$ , GlcA H-4), 5.23 – 5.15 (m, 2H;  $OCH_2CHCH_2$ , GalNAc H-1), 4.90 (d,  $J$  = 7.3 Hz, 1H; GlcA H-1), 4.73 (dd,  $J$  = 11.2, 3.4 Hz, 1H; GalNAc H-3), 4.43 – 4.36 (m, 2H;  $OCH_2CHCH_2$ , GalNAc H-4), 4.32 (dd,  $J$  = 12.4, 1.6 Hz, 1H; GalNAc H-6), 4.14 – 4.07 (m, 2H;  $OCH_2CHCH_2$ , GalNAc H-6), 3.94 (d,  $J$  = 8.3 Hz, 1H; GlcA H-5), 3.85 – 3.78 (m, 1H; GalNAc H-2), 3.70 (s, 3H;  $COOCH_3$ ), 3.65 – 3.51 (m, 3H; GalNAc H-5, GlcA H-2, H-3), 2.86 – 2.53 (m, 4H;  $CH_2O$ ), 2.18 (s, 3H;  $COCH_3$ ), 1.40 (s, 3H;  $CH_3$ ), 1.37 (s, 3H;  $CH_3$ ).  $^{13}C$  NMR (101 MHz,  $CDCl_3$ )  $\delta$  = 206.06, 171.58, 167.71, 162.39, 137.68, 133.61, 128.90, 128.03, 126.39, 118.16, 112.46, 102.48, 100.81, 97.37, 92.18, 75.87, 75.78, 74.46, 74.39, 70.93, 70.44, 69.06, 66.59, 55.55, 52.97, 37.58, 29.89, 27.64, 26.64, 26.50. HRMS:  $m/z$  calcd for  $[C_{33}H_{40}Cl_3NO_{14} + NH_4]^+$ : 797.1859; found: 797.1823.

**Allyl *O*-(methyl 4-*O*-levulinoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (36).**

A solution of **35** (916.8 mg, 1.2 mmol) in CH<sub>2</sub>Cl<sub>2</sub>/AcOH/water (5:4:1, 20 ml) was stirred for 60 h at RT. The solution was then concentrated and azeotroped with toluene three times. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:19, containing 0.5% triethylamine) gave the product (613.5 mg, 71%) as a white solid. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  = 7.57 – 7.30 (m, 5H; Ar-H), 7.23 (d, *J* = 6.8 Hz, 1H; NH), 5.95 – 5.81 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.57 (s, 1H; CHPh), 5.33 – 5.24 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.22 – 5.15 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.09 (d, *J* = 8.2 Hz, 1H; GalNAc H-1), 5.02 (dd, *J* = 10.0, 9.2 Hz, 1H; GlcA H-4), 4.65 (dd, *J* = 11.1, 3.5 Hz, 1H; GalNAc H-3), 4.51 (d, *J* = 7.7 Hz, 1H; GlcA H-1), 4.43 – 4.36 (m, 2H; GalNAc H-4, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.33 (dd, *J* = 12.4, 1.6 Hz, 1H; GalNAc H-6), 4.15 – 4.05 (m, 2H; GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.93 (d, *J* = 10.0 Hz, 1H; GlcA H-5), 3.83 (ddd, *J* = 11.2, 8.2, 6.8 Hz, 1H; GalNAc H-2), 3.73 (s, 3H; COOCH<sub>3</sub>), 3.65 (t, *J* = 9.3 Hz, 1H; GlcA H-3), 3.58 – 3.48 (m, 2H; GalNAc H-5, GlcA H-2), 2.92 – 2.39 (m, 4H; CH<sub>2</sub>O), 2.17 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>):  $\delta$  = 207.23, 171.98, 167.65, 162.43, 137.59, 133.50, 129.00, 128.12, 126.43, 118.33, 103.29, 101.02, 97.50, 92.20, 75.75, 74.91, 73.90, 73.07, 72.32, 71.43, 70.40, 69.03, 66.58, 55.37, 52.84, 38.04, 29.83, 27.86. HRMS: *m/z* calcd for [C<sub>30</sub>H<sub>36</sub>Cl<sub>3</sub>NO<sub>14</sub>+NH<sub>4</sub>]<sup>+</sup>: 757.1546; found: 757.1516.

**Allyl *O*-(methyl 2,3-*O*-benzoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (32).**

Benzoyl chloride (37  $\mu$ l, 0.32 mmol) was added slowly to a solution of the **36** (99.0 mg, 0.13 mmol) in dry pyridine/CH<sub>2</sub>Cl<sub>2</sub> (5:8, 1.04 ml) and stirred for 1 h at 0 °C. Methanol (0.1 ml) was then added and stirred for 5 min at 0 °C. The mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub>

and washed with water, saturated aqueous NaHCO<sub>3</sub>, and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 1:9, containing 0.5% triethylamine) gave the dibenzoylated product (109.0 mg, 86%) as a white solid. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ = 7.95 – 7.29 (m, 15H; Ar-H), 7.00 (d, *J* = 6.6 Hz, 1H; NH), 5.94 – 5.73 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.62 (t, *J* = 9.2 Hz, 1H; GlcA H-3), 5.56 (s, 1H; CHPh), 5.53 – 5.42 (m, 2H; GlcA H-2, H-4), 5.28 – 5.19 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.18 (d, *J* = 7.4 Hz, 1H; GlcA H-1), 5.17 – 5.12 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.11 (d, *J* = 8.2 Hz, 1H; GalNAc H-1), 4.78 (dd, *J* = 11.2, 3.5 Hz, 1H; GalNAc H-3), 4.47 (dd, *J* = 3.5, 0.9 Hz, 1H; GalNAc H-4), 4.39 – 4.28 (m, 2H; GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.21 (d, *J* = 9.9 Hz, 1H; GlcA H-5), 4.06 (m, 2H; GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.80 – 3.69 (m, 4H; COOCH<sub>3</sub>, GalNAc H-2), 3.51 (d, *J* = 1.2 Hz, 1H; GalNAc H-5), 2.65 – 2.32 (m, 4H; CH<sub>2</sub>O), 2.04 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ = 205.62, 171.21, 167.18, 165.49, 164.80, 162.22, 137.66, 133.54, 133.40, 133.37, 129.84, 129.79, 128.87, 128.73, 128.62, 128.38, 128.04, 126.10, 117.99, 100.51, 100.19, 97.44, 92.05, 75.74, 73.48, 72.45, 72.07, 71.37, 70.32, 69.38, 68.98, 66.53, 55.29, 53.06, 37.50, 29.58, 27.60. HRMS: *m/z* calcd for [C<sub>44</sub>H<sub>44</sub>Cl<sub>3</sub>NO<sub>16</sub> + NH<sub>4</sub>]<sup>+</sup>: 965.2070; found: 965.2035.

A mixture of pyridine/AcOH/hydrazine hydrate (6:4:0.5, 1.1 ml) was added to a solution of the dibenzoylated product (53.9 mg, 0.57 mmol) in pyridine (0.21 ml) and stirred for 2.5 min at RT. The mixture was then diluted with CH<sub>2</sub>Cl<sub>2</sub> (10 ml), washed with water, saturated aqueous NaHCO<sub>3</sub>, and water, dried with MgSO<sub>4</sub>, and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 12:88, containing 0.5% triethylamine) gave the product (39.0 mg, 81%) as a white solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ = 7.96 – 7.29 (m, 15H; Ar-H), 7.01 (d, *J* = 6.6 Hz, 1H; NH), 5.93 – 5.77 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.53 (s, 1H;



CHPh), 5.52 – 5.43 (m, 2H; GlcA H-2, H-3), 5.27 – 5.19 (m, 2H; GlcA H-1, OCH<sub>2</sub>CHCH<sub>2</sub>), 5.19 – 5.11 (m, 2H; GalNAc H-1, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.76 (dd,  $J = 11.2, 3.5$  Hz, 1H; GalNAc H-3), 4.44 (d,  $J = 3.4$  Hz, 1H; GalNAc H-4), 4.39 – 4.31 (m, 2H; GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.21 (t,  $J = 9.2$  Hz, 1H; GlcA H-4), 4.14 – 4.00 (m, 2H; GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.85 (s, 3H; COOCH<sub>3</sub>), 3.73 (ddd,  $J = 11.2, 8.1, 6.5$  Hz, 1H; GalNAc H-2), 3.51 (s, 1H; GalNAc H-5), 3.39 (s, 1H; GlcA 4-OH). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>):  $\delta = 169.11, 166.43, 165.06, 162.15, 137.58, 133.55, 133.42, 133.37, 129.86, 129.83, 128.96, 128.85, 128.40, 128.37, 128.12, 125.98, 117.98, 100.48, 97.40, 92.09, 75.85, 75.07, 74.04, 73.75, 71.09, 70.35, 69.05, 66.51, 55.30, 53.05$ . HRMS:  $m/z$  calcd for [C<sub>39</sub>H<sub>38</sub>Cl<sub>3</sub>NO<sub>14</sub> + NH<sub>4</sub>]<sup>+</sup>: 867.1702; found: 867.1696.

**Allyl (methyl 2,3,4-*O*-acetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-(4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside)-(1 $\rightarrow$ 4)-(methyl 2,3-*O*-benzoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-4,6-*O*-benzylidene-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (37).**

A mixture of donor **7** (39 mg, 0.045 mmol), acceptor **32** (19 mg, 0.022 mmol) was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (30 ml). The mixture was then cooled down to -60 °C and TMSOTf was added. The mixture was slowly warmed to RT within 1 h and then filtered and concentrated. Flash silica chromatography (EtOAc/CH<sub>2</sub>Cl<sub>2</sub> 1:10) and gave the product (17.5 mg, 50%) as a white solid. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta = 7.93 - 7.13$  (m, 20H; Ar-H), 6.93 (d,  $J = 7.1$  Hz, 1H; NH), 6.90 (d,  $J = 6.7$  Hz, 1H; NH), 5.83 (s, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.62 (t,  $J = 8.5$  Hz, 1H; GlcA<sup>I</sup> H-3), 5.52 (s, 1H; CHPh), 5.40 (dd,  $J = 8.4, 7.0$  Hz, 1H; GlcA<sup>I</sup> H-2), 5.34 (s, 1H; CHPh), 5.25 – 5.05 (m, 7H; 2GalNAc H-1, GlcA<sup>I</sup> H-1, GlcA<sup>II</sup> H-3, H-4, 2OCH<sub>2</sub>CHCH<sub>2</sub>), 4.95 (dd,  $J = 8.7, 7.4$  Hz, 1H; GlcA<sup>II</sup> H-2), 4.82 (d,  $J = 7.4$  Hz, 1H; GlcA<sup>II</sup> H-1), 4.72 (dd,  $J$

= 11.1, 3.5 Hz, 1H; GalNAc H-3), 4.57 (t,  $J$  = 9.0 Hz, 1H; GlcA<sup>I</sup> H-4), 4.44 – 4.30 (m, 4H; GalNAc H-3, H-4, H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.22 (d,  $J$  = 3.5 Hz, 1H; GalNAc H-4), 4.17 (d,  $J$  = 9.4 Hz, 1H; GlcA<sup>I</sup> H-5), 4.10 – 4.01 (m, 2H; GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.95 (d,  $J$  = 9.9 Hz, 1H; GlcA<sup>II</sup> H-5), 3.86 – 3.78 (m, 4H; GalNAc H-6, COOCH<sub>3</sub>), 3.76 – 3.67 (m, 6H; GalNAc 2H-2, H-6, COOCH<sub>3</sub>), 3.50 (s, 1H; GalNAc H-5), 3.14 (s, 1H; GalNAc H-5), 1.99 (s, 3H; COCH<sub>3</sub>), 1.98 (s, 3H; COCH<sub>3</sub>), 1.96 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>):  $\delta$  = 170.03, 169.38, 169.11, 168.44, 167.15, 165.37, 165.06, 162.13, 161.78, 137.71, 137.66, 133.55, 133.43, 133.38, 133.27, 132.95, 129.90, 129.84, 129.74, 129.18, 129.01, 128.83, 128.56, 128.34, 128.21, 128.12, 127.83, 126.26, 126.14, 125.99, 118.01, 100.47, 100.35, 100.32, 99.86, 98.10, 97.43, 92.40, 92.03, 75.53, 75.33, 74.77, 73.93, 73.89, 73.76, 73.03, 72.30, 71.82, 71.10, 70.35, 69.07, 69.04, 68.23, 66.66, 66.51, 20.79, 20.56, 20.47. HRMS:  $m/z$  calcd for [C<sub>67</sub>H<sub>68</sub>Cl<sub>6</sub>N<sub>2</sub>O<sub>28</sub>+NH<sub>4</sub>]<sup>+</sup>: 1576.2435; found: 1576.2409.

**Allyl (methyl 2,3,4-*O*-acetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-(2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside)-(1 $\rightarrow$ 4)-(methyl 2,3-*O*-benzoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-deoxy-2-trichloroacetamido- $\beta$ -D-galactopyranoside (38).**

A mixture of **37** (32.5 mg, 6.0  $\mu$ mol), CSA (11.1 mg, 48  $\mu$ mol) in DCM/methanol (1:1, 4 ml) was stirred at rt for 10 h. The mixture was cooled down and concentrated. The residue was dissolved in CH<sub>2</sub>Cl<sub>2</sub> and washed with water and saturated aqueous NaHCO<sub>3</sub>. The organic layer was dried with MgSO<sub>4</sub>, filtered, and concentrated. Flash silica chromatography (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 19:1) containing 0.5% triethylamine) gave the product (19.9 mg, 78%) as a white solid. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  = 7.96 – 7.35 (m, 10H; Ar-H), 5.94 – 5.74 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.57 (t,  $J$  = 7.8 Hz, 1H; GlcA<sup>I</sup> H-3), 5.33 (t,  $J$

= 7.6 Hz, 1H; GlcA<sup>I</sup> H-2), 5.22 – 5.16 (m, 3H; GlcA<sup>I</sup> H-1, GlcA<sup>II</sup> H-3, OCH<sub>2</sub>CHCH<sub>2</sub>), 5.11 – 5.03 (m, 2H; GlcA<sup>II</sup> H-4, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.96 (dd,  $J = 9.5, 8.0$  Hz, 1H; GlcA<sup>II</sup> H-2), 4.85 (d,  $J = 8.0$  Hz, 1H; GlcA<sup>II</sup> H-1), 4.79 (d,  $J = 8.1$  Hz, 1H; GalNAc H-1), 4.56 – 4.49 (m, 2H; GalNAc H-1, GlcA<sup>I</sup> H-4), 4.34 (d,  $J = 7.9$  Hz, 1H; GlcA<sup>I</sup> H-5), 4.30 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.19 (d,  $J = 9.9$  Hz, 1H; GlcA<sup>II</sup> H-5), 4.17 (dd,  $J = 2.9, 1.0$  Hz, 1H; GalNAc H-4), 4.13 – 3.97 (m, 4H; GalNAc, 2H-2, H-3, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.94 – 3.89 (m, 2H; GalNAc H-3, H-4), 3.77 (d,  $J = 6.1$  Hz, 5H; GalNAc 2H-6, COOCH<sub>3</sub>), 3.69 (s, 3H; COOCH<sub>3</sub>), 3.53 (m, 1H; GalNAc H-5), 3.27 – 3.16 (m, 2H; GalNAc 2H-6), 2.03 (s, 3H; COCH<sub>3</sub>), 1.99 (s, 3H; COCH<sub>3</sub>), 1.96 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>):  $\delta = 169.38, 169.25, 168.99, 168.04, 166.96, 164.87, 162.00, 161.74, 133.15, 132.30, 132.23, 129.13, 128.81, 128.75, 128.37, 127.29, 127.23, 115.19, 100.14, 99.84, 99.51, 99.48, 92.13, 91.82, 78.78, 77.18, 74.18, 74.13, 73.57, 72.52, 71.27, 70.92, 70.13, 68.82, 68.70, 66.82, 66.23, 60.13, 59.03, 52.93, 52.27, 51.49, 51.22, 18.90, 18.32, 18.24$ . HRMS:  $m/z$  calcd for [C<sub>53</sub>H<sub>60</sub>Cl<sub>6</sub>N<sub>2</sub>O<sub>28</sub> + Na]<sup>+</sup>: 1405.1361; found: 1405.1639.

**Allyl (methyl 2,3,4-*O*-acetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-(2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside)-(1 $\rightarrow$ 4)-(methyl 2,3-*O*-benzoyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-deoxy-2-acetamido- $\beta$ -D-galactopyranoside (39).**

A mixture of tributylstannane (52.0  $\mu$ l, 0.19 mmol), AIBN (3.0 mg, 18.0  $\mu$ mol) and **38** (12.7 mg, 9.2  $\mu$ mol) in toluene (1.6 ml) and *N,N*-dimethylacetamide was purged with dry Ar gas for 1 hour. The mixture was then heated to 110 °C and stirred for 20 min. The mixture was cooled down, diluted with acetonitrile and washed with pentane. The acetonitrile layer was concentrated. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:11, containing 0.5% triethylamine) gave the product (9.2 mg, 85%). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta = 8.10 -$

7.31 (m, 10H; Ar-H), 5.88 – 5.74 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.65 (t,  $J = 8.6$  Hz, 1H; GlcA<sup>I</sup> H-3), 5.33 (dd,  $J = 8.8, 7.8$  Hz, 1H; GlcA<sup>I</sup> H-2), 5.27 (t,  $J = 9.4$  Hz, 1H; GlcA<sup>II</sup> H-3), 5.24 – 5.16 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.13 – 5.00 (m, 3H; GlcA<sup>I</sup> H-1, GlcA<sup>II</sup> H-4, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.93 (dd,  $J = 9.5, 8.0$  Hz, 1H; GlcA<sup>II</sup> H-2), 4.84 (d,  $J = 8.0$  Hz, 1H; GlcA<sup>II</sup> H-1), 4.55 (d,  $J = 8.2$  Hz, 1H; GalNAc H-1), 4.46 – 4.37 (m, 2H; GalNAc H-1, GlcA<sup>I</sup> H-4), 4.32 – 4.23 (m, 2H; GlcA<sup>I</sup> H-5, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.20 (d,  $J = 10.0$  Hz, 1H; GlcA<sup>II</sup> H-5), 4.12 (dd,  $J = 2.8, 1.1$  Hz, 1H; GalNAc H-4), 4.05 – 3.94 (m, 2H; GalNAc H-2, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.90 (dd,  $J = 2.8, 1.0$  Hz, 1H; GalNAc H-4), 3.88 – 3.77 (m, 5H; GalNAc H-2, H-3, COOCH<sub>3</sub>), 3.77 – 3.70 (m, 3H; GalNAc H-3, 2H-6), 3.68 (s, 3H; COOCH<sub>3</sub>), 3.54 – 3.46 (m, 1H; GalNAc H-5), 3.26 – 3.14 (m, 2H; GalNAc H-5, H-6), 3.10 – 3.02 (m, 1H; GalNAc H-6), 2.02 (s, 3H, COCH<sub>3</sub>), 1.99 (s, 3H; COCH<sub>3</sub>), 1.97 (s, 6H; 2COCH<sub>3</sub>), 1.26 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>):  $\delta = 173.29, 173.21, 171.45, 171.16, 171.14, 169.77, 169.19, 166.91, 166.54, 135.55, 134.54, 134.43, 131.06, 130.93, 130.85, 130.62, 129.54, 129.45, 116.84, 103.38, 102.72, 102.35, 102.07, 82.81, 81.96, 80.87, 77.64, 76.19, 75.80, 75.57, 74.81, 73.42, 72.93, 72.51, 70.80, 70.49, 69.14, 68.09, 62.33, 60.80, 53.60, 53.34, 52.95, 52.23, 23.57, 22.42, 20.77, 20.47, 20.39$ . HRMS:  $m/z$  calcd for [C<sub>53</sub>H<sub>66</sub>N<sub>2</sub>O<sub>28</sub>+Na]<sup>+</sup>: 1201.3701; found: 1201.3696.

**Allyl (methyl 2,3,4-*O*-acetyl- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-(2-acetamido-2-deoxy-4,6-*O*-sulfonato- $\beta$ -D-galactopyranoside)-(1 $\rightarrow$ 4)-(methyl 2,3-*O*-benzoyl-2-*O*-sulfonato- $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-acetamido-4,6-*O*-sulfonato-2-deoxy- $\beta$ -D-galactopyranoside tetrasodium salt (40).**

A mixture of **39** (5.6 mg, 4.8  $\mu$ mol) and sulfur trioxide/trimethylamine complex (33.5 mg, 24  $\mu$ mol) was azeotroped with toluene 5 times and dried under high vacuum overnight. The mixture was then dissolved in anhydrous DMF (0.37 ml) and stirred for 48

h at 60 °C. After the mixture was cooled down, methanol (0.05 ml) was added and the mixture was stirred for another 30 min. The mixture was then directly applied and eluted from a column (3x30 cm) of Sephadex LH-20 with 1:1 CH<sub>2</sub>Cl<sub>2</sub>/MeOH. Flash silica chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub> 1:5, containing 0.5% triethylamine) followed by a column (1x15 cm) of Amberlite® IR120 Sodium form resin (1:1 CH<sub>2</sub>Cl<sub>2</sub>/MeOH) then gave the product as a white solid (6.1 mg, 78%). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ = 8.03 – 7.32 (m, 10H; Ar-H), 5.88 – 5.76 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.73 (t, *J* = 8.9 Hz, 1H; GlcA<sup>I</sup> H-3), 5.44 (dd, *J* = 8.6, 7.4 Hz, 1H; GlcA<sup>I</sup> H-2), 5.27 (t, *J* = 9.4 Hz, 1H; GlcA<sup>II</sup> H-3), 5.24 – 5.15 (m, 2H; GlcA<sup>II</sup> H-4, OCH<sub>2</sub>CHCH<sub>2</sub>), 5.13 – 5.06 (m, 2H; GlcA<sup>I</sup> H-1, OCH<sub>2</sub>CHCH<sub>2</sub>), 5.03 (dd, *J* = 9.3, 8.1 Hz, 1H; GlcA<sup>II</sup> H-2), 4.99 (d, *J* = 2.9 Hz, 1H; GalNAc H-4), 4.75 (d, *J* = 2.9 Hz, 1H; GalNAc H-4), 4.67 (d, *J* = 8.3 Hz, 1H; GalNAc H-1), 4.63 (t, *J* = 9.5 Hz, 1H; GlcA<sup>I</sup> H-4), 4.48 (d, *J* = 9.8 Hz, 1H; GlcA<sup>I</sup> H-5), 4.45 – 4.37 (m, 2H; GalNAc H-1, H-6), 4.33 – 4.24 (m, 2H; GalNAc H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 4.20 (d, *J* = 9.9 Hz, 1H; GlcA<sup>II</sup> H-5), 4.13 – 4.00 (m, 3H; GalNAc H-2, H-6, OCH<sub>2</sub>CHCH<sub>2</sub>), 3.99 – 3.78 (m, 9H; GalNAc H-2, 2H-3, 2H-5, H-6, COOCH<sub>3</sub>), 3.73 (s, 3H; COOCH<sub>3</sub>), 2.00 (s, 3H; COCH<sub>3</sub>), 2.00 (s, 3H; COCH<sub>3</sub>), 1.95 (s, 3H; COCH<sub>3</sub>), 1.88 (s, 3H; COCH<sub>3</sub>), 1.42 (s, 3H; COCH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD): δ = 173.66, 173.59, 171.65, 171.52, 171.25, 170.69, 169.88, 167.44, 167.16, 135.42, 134.63, 134.33, 131.23, 131.01, 130.73, 130.48, 129.58, 129.48, 117.08, 103.16, 102.18, 101.94, 101.07, 80.12, 77.89, 77.04, 76.80, 76.09, 75.32, 74.28, 74.03, 73.72, 73.53, 73.01, 72.69, 70.72, 70.65, 69.17, 68.14, 53.81, 53.58, 53.39, 52.39, 23.41, 22.56, 20.90, 20.49, 20.43. HRMS: *m/z* calcd for [C<sub>53</sub>H<sub>62</sub>N<sub>2</sub>Na<sub>4</sub>O<sub>40</sub>S<sub>4</sub> + Na]<sup>+</sup>: 1609.1250; found: 1609.1287.

**Allyl ( $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-(2-acetamido-2-deoxy-4,6-O-sulfonato- $\beta$ -D-galactopyranoside)-(1 $\rightarrow$ 4)-( $\beta$ -D-glucopyranosyluronate)-(1 $\rightarrow$ 3)-2-acetamido-4,6-O-sulfonato-2-deoxy- $\beta$ -D-galactopyranoside hexasodium salt (2).**

Tetrasaccharide **40** (6.1 mg, 4.1  $\mu$ mol) was dissolved in THF (450  $\mu$ L) and H<sub>2</sub>O (225  $\mu$ L) and cooled to 0 °C. To this were added 1 M aq. LiOH (64  $\mu$ L) and 30% H<sub>2</sub>O<sub>2</sub> (32  $\mu$ L). The reaction was stirred at 0 °C for 1 h and at rt for 12 h. At this time, 4 M NaOH (70  $\mu$ L) and MeOH (475  $\mu$ L) were added and the reaction stirred for another 12 h. It was then neutralized with Amberlite® IR120 Hydrogen form resin, filtered, and lyophilized to afford a white solid. The product was purified by a Sephadex G-25 column (0.5x30 cm) and lyophilized to give the product (3.9 mg, 81%). <sup>1</sup>H NMR (900 MHz, D<sub>2</sub>O, 286 K)  $\delta$  = 5.88 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.30 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 5.27 – 5.23 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.83 (s, 1H; GalNAc<sup>II</sup> H-4), 4.77 (d,  $J$  = 3.0 Hz, 1H; GalNAc<sup>I</sup> H-4), 4.58 (d,  $J$  = 8.0 Hz, 1H; GalNAc<sup>II</sup> H-1), 4.57 (d,  $J$  = 8.4 Hz, 1H; GalNAc<sup>I</sup> H-1), 4.47 (d,  $J$  = 7.8 Hz, 1H; GlcA<sup>I</sup> H-1), 4.45 (d,  $J$  = 7.8 Hz, 1H; GlcA<sup>II</sup> H-1), 4.35 – 4.31 (m, 1H; OCH<sub>2</sub>CHCH<sub>2</sub>), 4.27 (m, 2H; GalNAc<sup>I</sup> H-6, GalNAc<sup>II</sup> H-6), 4.23 – 4.16 (m, 3H; OCH<sub>2</sub>CHCH<sub>2</sub>, GalNAc<sup>I</sup> H-6, GalNAc<sup>II</sup> H-6), 4.12 (dd,  $J$  = 8.8, 2.7 Hz, 1H; GalNAc<sup>II</sup> H-5), 4.09 – 4.03 (m, 3H; GalNAc<sup>I</sup> H-2, H-5, GalNAc<sup>II</sup> H-3), 4.01 (dd,  $J$  = 11.0, 3.0 Hz, 1H; GalNAc<sup>I</sup> H-3), 3.74 (dd,  $J$  = 11.1, 7.4 Hz, GlcA<sup>I</sup> H-4), 3.66 (d,  $J$  = 9.6 Hz, 1H; GlcA<sup>I</sup> H-5), 3.66 – 3.63 (d,  $J$  = 9.4 Hz, 1H; GlcA<sup>II</sup> H-5), 3.59 (dd,  $J$  = 11.0, 7.2 Hz, 1H; GlcA<sup>I</sup> H-3), 3.50 (t,  $J$  = 9.5 Hz, 1H; GlcA<sup>II</sup> H-4), 3.45 (t,  $J$  = 9.2 Hz, 1H; GlcA<sup>I</sup> H-3), 3.39 (dd,  $J$  = 9.5, 7.9 Hz, 1H; GlcA<sup>II</sup> H-2), 3.32 (dd,  $J$  = 9.4, 7.8 Hz, 1H; GlcA<sup>I</sup> H-2), 2.02 (s, 3H; NHCOCH<sub>3</sub>), 1.99 (s, 3H; NHCOCH<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, D<sub>2</sub>O):  $\delta$  = 175.87, 174.93, 174.76, 174.07, 133.08, 118.58, 103.44, 103.40, 101.55, 99.83, 82.21, 76.63, 76.25, 76.08, 76.01, 75.02, 74.82, 73.75, 72.47, 72.29, 72.22, 71.86, 71.80, 70.53, 67.87,

67.62, 51.44, 51.36, 22.55, 22.15. HRMS:  $m/z$  calcd for  $[C_{31}H_{42}N_2Na_6O_{35}S_4 - Na]^-$ : 1244.9945; found: 1244.9948.

## References

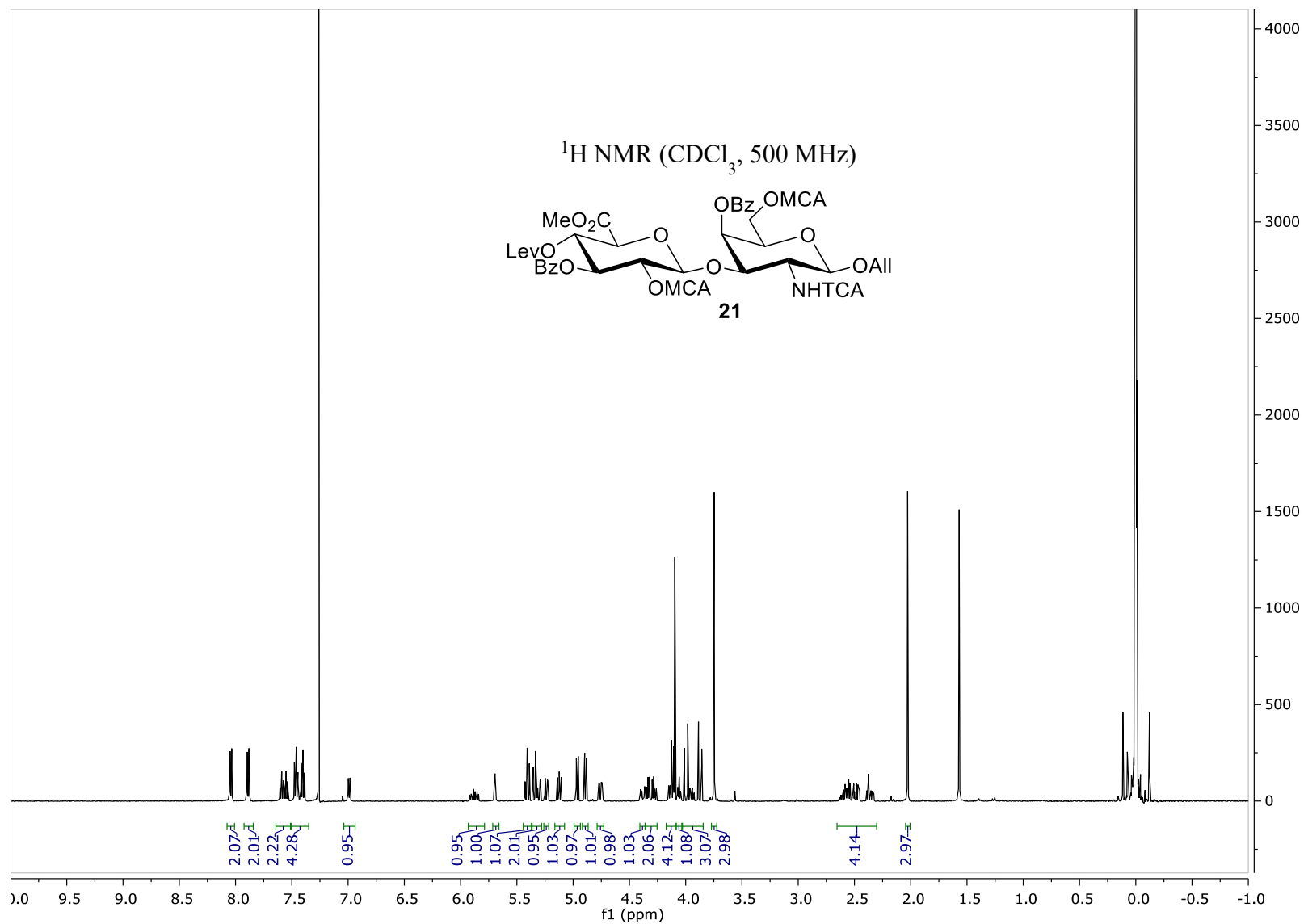
- (1) Hikino, M.; Mikami, T.; Faissner, A.; Vilela-Silva, A. C.; Pavao, M. S.; Sugahara, K., Oversulfated dermatan sulfate exhibits neurite outgrowth-promoting activity toward embryonic mouse hippocampal neurons: implications of dermatan sulfate in neuritogenesis in the brain. *Journal of Biological Chemistry* **2003**, 278 (44), 43744.
- (2) Mizumoto, S.; Murakoshi, S.; Kalayanamitra, K.; Deepa, S. S.; Fukui, S.; Kongtawelert, P.; Yamada, S.; Sugahara, K., Highly sulfated hexasaccharide sequences isolated from chondroitin sulfate of shark fin cartilage: Insights into the sugar sequences with bioactivities. *Glycobiology* **2013**, 23 (2), 155.
- (3) Nadanaka, S.; Clement, A.; Masayama, K.; Faissner, A.; Sugahara, K., Characteristic hexasaccharide sequences in octasaccharides derived from shark cartilage chondroitin sulfate D with a neurite outgrowth promoting activity. *Journal of Biological Chemistry* **1998**, 273 (6), 3296.
- (4) Blanchard, V.; Chevalier, F.; Imberty, A.; Leeftang, 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 (5), 1167.
- (5) Ito, Y.; Hikino, M.; Yajima, Y.; Mikami, T.; Sirko, S.; von Holst, A.; Faissner, A.; Fukui, S.; Sugahara, K., Structural characterization of the epitopes of the monoclonal antibodies 473HD, CS-56, and MO-225 specific for chondroitin sulfate D-type using the oligosaccharide library. *Glycobiology* **2005**, 15 (6), 593.
- (6) Xu, Y. M.; Masuko, S.; Takieddin, M.; Xu, H. M.; Liu, R. P.; Jing, J.; Mousa, S. A.; Linhardt, R. J.; Liu, J., Chemoenzymatic Synthesis of Homogeneous Ultralow Molecular Weight Heparins. *Science* **2011**, 334 (6055), 498.
- (7) Xu, Y. M.; Cai, C.; Chandarajoti, K.; Hsieh, P. H.; Li, L. Y.; Pham, T. Q.; Sparkenbaugh, E. M.; Sheng, J. Z.; Key, N. S.; Pawlinski, R.; Harris, E. N.; Linhardt, R. J.; Liu, J., Homogeneous low-molecular-weight heparins with reversible anticoagulant activity. *Nature Chemical Biology* **2014**, 10 (4), 248.
- (8) Li, J.; Su, G. W.; Liu, J., Enzymatic Synthesis of Homogeneous Chondroitin Sulfate Oligosaccharides. *Angewandte Chemie-International Edition* **2017**, 56 (39), 11784.

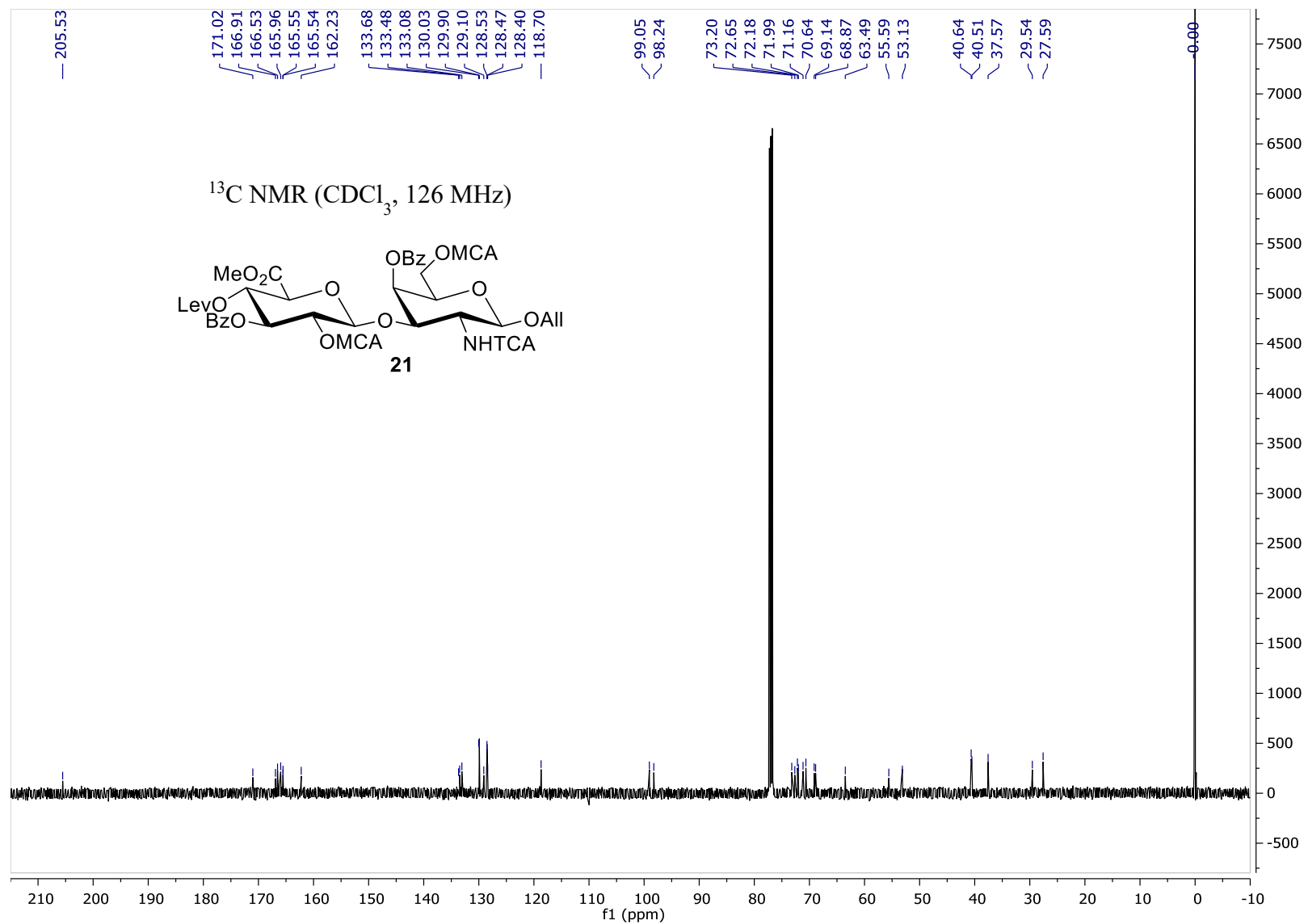
- (9) Sugiura, N.; Shioiri, T.; Chiba, M.; Sato, T.; Narimatsu, H.; Kimata, K.; Watanabe, H., Construction of a Chondroitin Sulfate Library with Defined Structures and Analysis of Molecular Interactions. *Journal of Biological Chemistry* **2012**, 287 (52), 43390.
- (10) Gama, C. I.; Tully, S. E.; Sotogaku, N.; Clark, P. M.; Rawat, M.; Vaidehi, N.; Goddard, W. A., 3rd; Nishi, A.; Hsieh-Wilson, L. C., Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Biol* **2006**, 2 (9), 467.
- (11) Karst, N.; Jacquinet, J. C., Stereocontrolled total synthesis of shark cartilage chondroitin sulfate D-related tetra- and hexasaccharide methyl glycosides. *European Journal of Organic Chemistry* **2002**, (5), 815.
- (12) Tully, S. E.; Mabon, R.; Gama, C. I.; Tsai, S. M.; Liu, X.; Hsieh-Wilson, L. C., A chondroitin sulfate small molecule that stimulates neuronal growth. *Journal of the American Chemical Society* **2004**, 126 (25), 7736.
- (13) Levene, P. A., On chondrosin. *Journal of Biological Chemistry* **1941**, 140 (1), 267.
- (14) Davidson, E. A.; Meyer, K., Structural Studies on Chondroitin Sulfuric Acid .I. The Nature of Chondrosine. *Journal of the American Chemical Society* **1954**, 76 (22), 5686.
- (15) Jacquinet, J. C.; Lopin-Bon, C.; Vibert, A., From Polymer to Size-Defined Oligomers: A Highly Divergent and Stereocontrolled Construction of Chondroitin Sulfate A, C, D, E, K, L, and M Oligomers from a Single Precursor: Part 2. *Chemistry-a European Journal* **2009**, 15 (37), 9579.
- (16) Vibert, A.; Lopin-Bon, C.; Jacquinet, J. C., From Polymer to Size-Defined Oligomers: A Step Economy Process for the Efficient and Stereocontrolled Construction of Chondroitin Oligosaccharides and Biotinylated Conjugates Thereof: Part 1. *Chemistry-a European Journal* **2009**, 15 (37), 9561.
- (17) Miyachi, K.; Wakao, M.; Suda, Y., Syntheses of chondroitin sulfate tetrasaccharide structures containing 4,6-disulfate patterns and analysis of their interaction with glycosaminoglycan-binding protein. *Bioorganic & Medicinal Chemistry Letters* **2015**, 25 (7), 1552.
- (18) Gama, C. I.; Hsieh-Wilson, L. C., Chemical approaches to deciphering the glycosaminoglycan code. *Current Opinion in Chemical Biology* **2005**, 9 (6), 609.

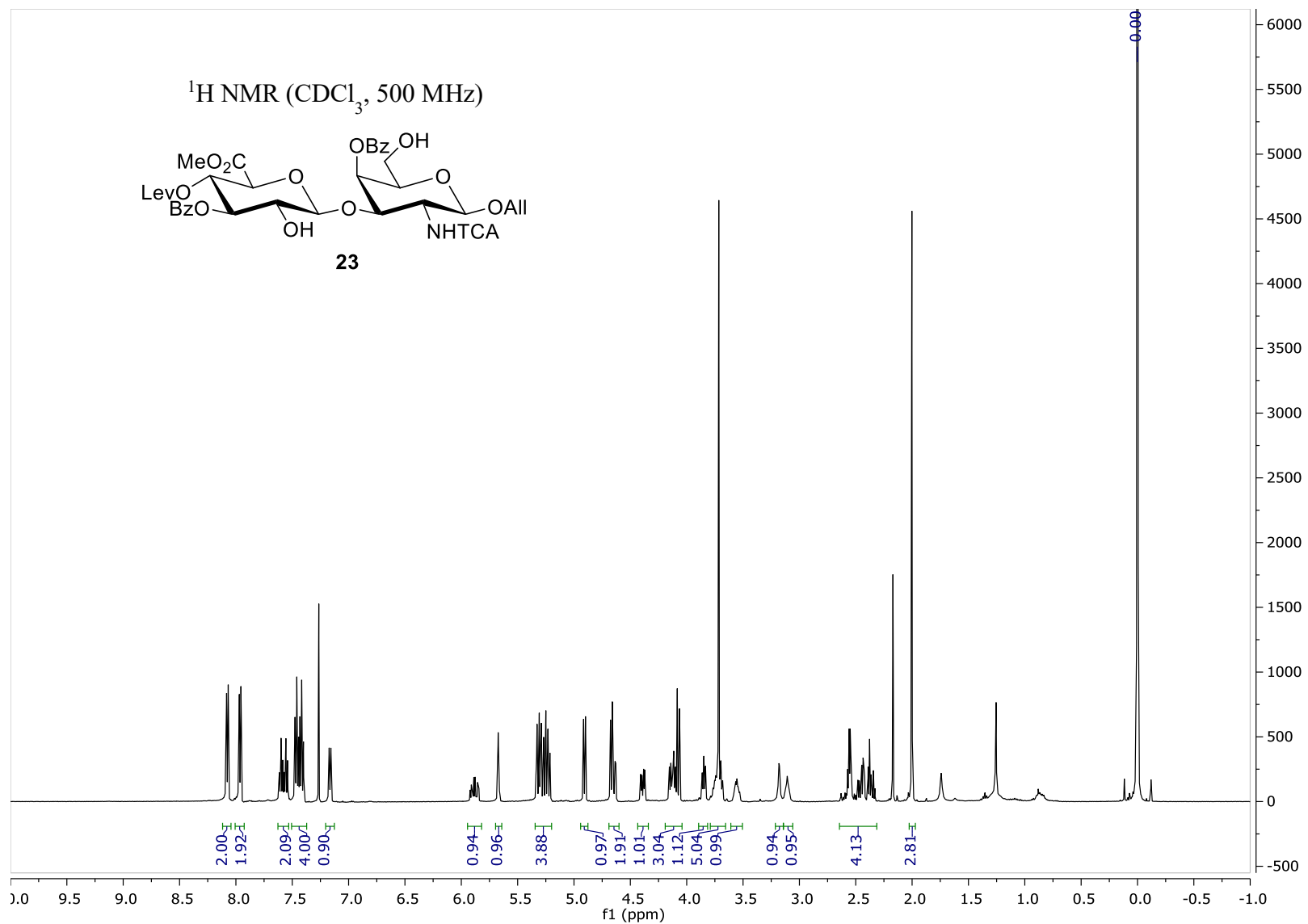


- (19) Schmidt, R. R., New Methods for the Synthesis of Glycosides and Oligosaccharides - Are There Alternatives to the Koenigs-Knorr Method. *Angewandte Chemie-International Edition in English* **1986**, 25 (3), 212.
- (20) Lassaletta, J. M.; Carlsson, K.; Garegg, P. J.; Schmidt, R. R., Total synthesis of sialylgalactosylgloboside: Stage-specific embryonic antigen 4. *Journal of Organic Chemistry* **1996**, 61 (20), 6873.
- (21) Lucas, H.; Basten, J. E. M.; Vandinther, T. G.; Meuleman, D. G.; Vanaelst, S. F.; Vanboeckel, C. A. A., Syntheses of Heparin - Like Pentamers Containing Opened Uronic-Acid Moieties. *Tetrahedron* **1990**, 46 (24), 8207.
- (22) Pomin, V. H., NMR Chemical Shifts in Structural Biology of Glycosaminoglycans. *Analytical Chemistry* **2014**, 86 (1), 65.

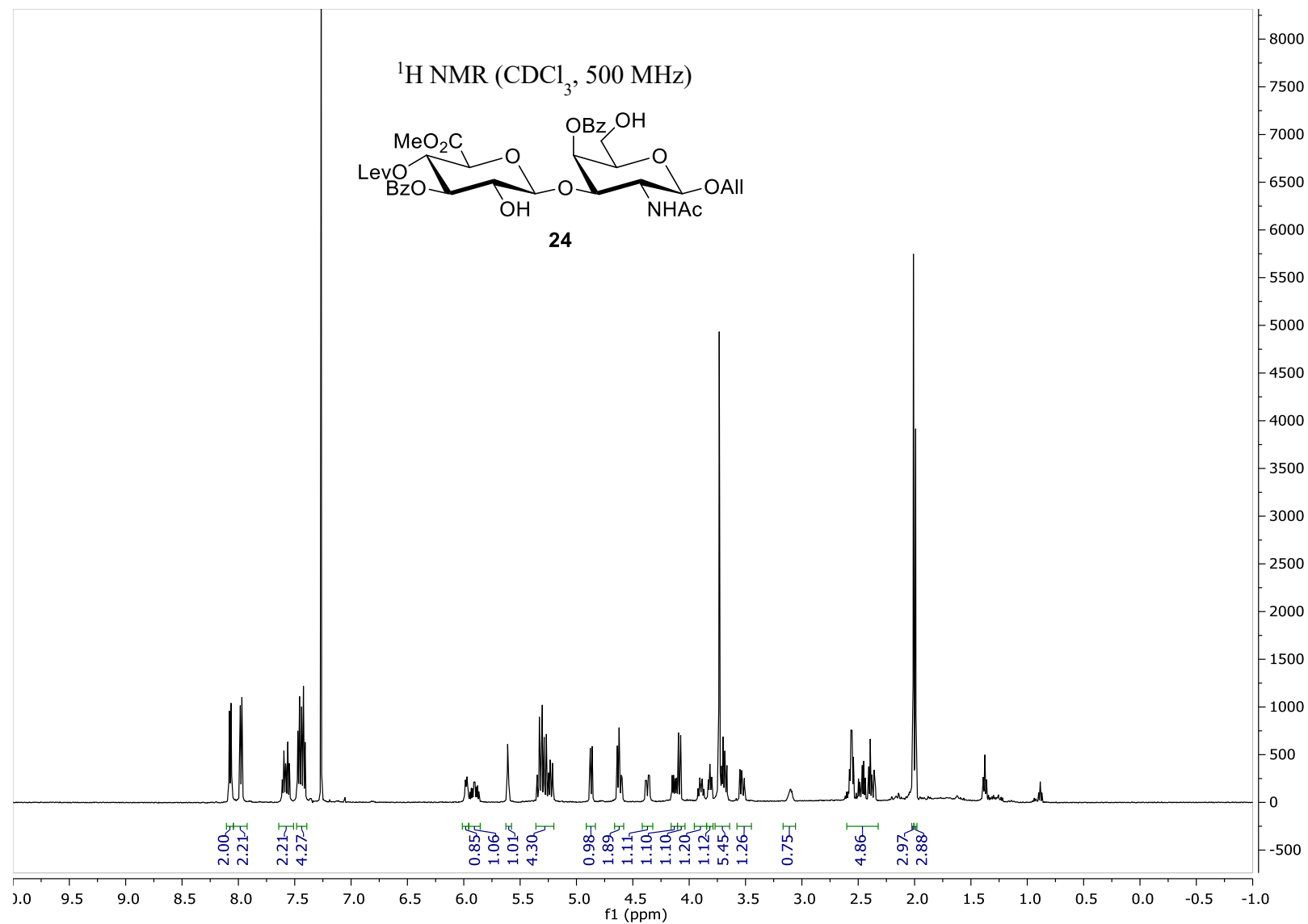
*Appendix for Chapter 2*  
*RELEVANT SPECTRAL DATA*

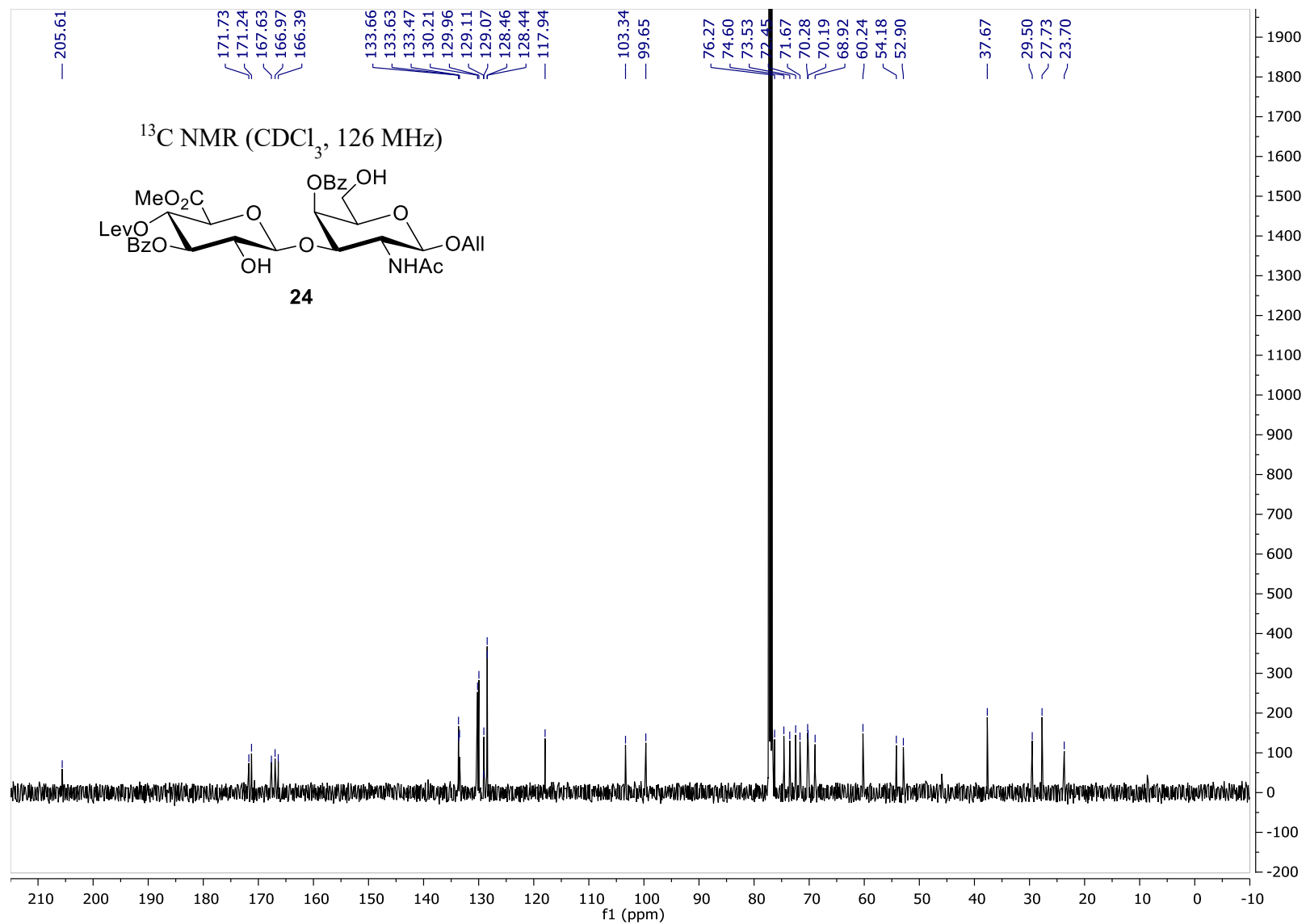




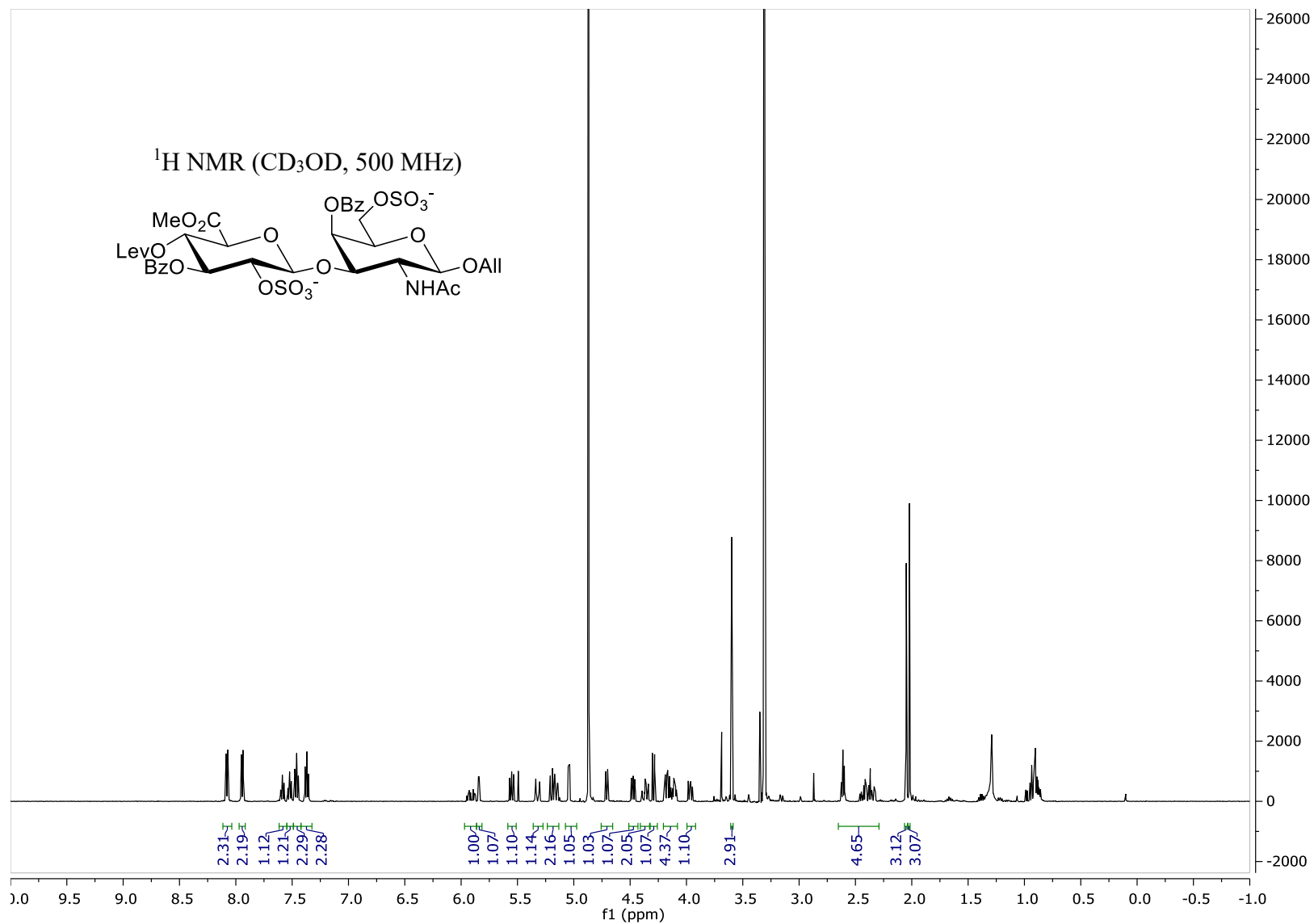


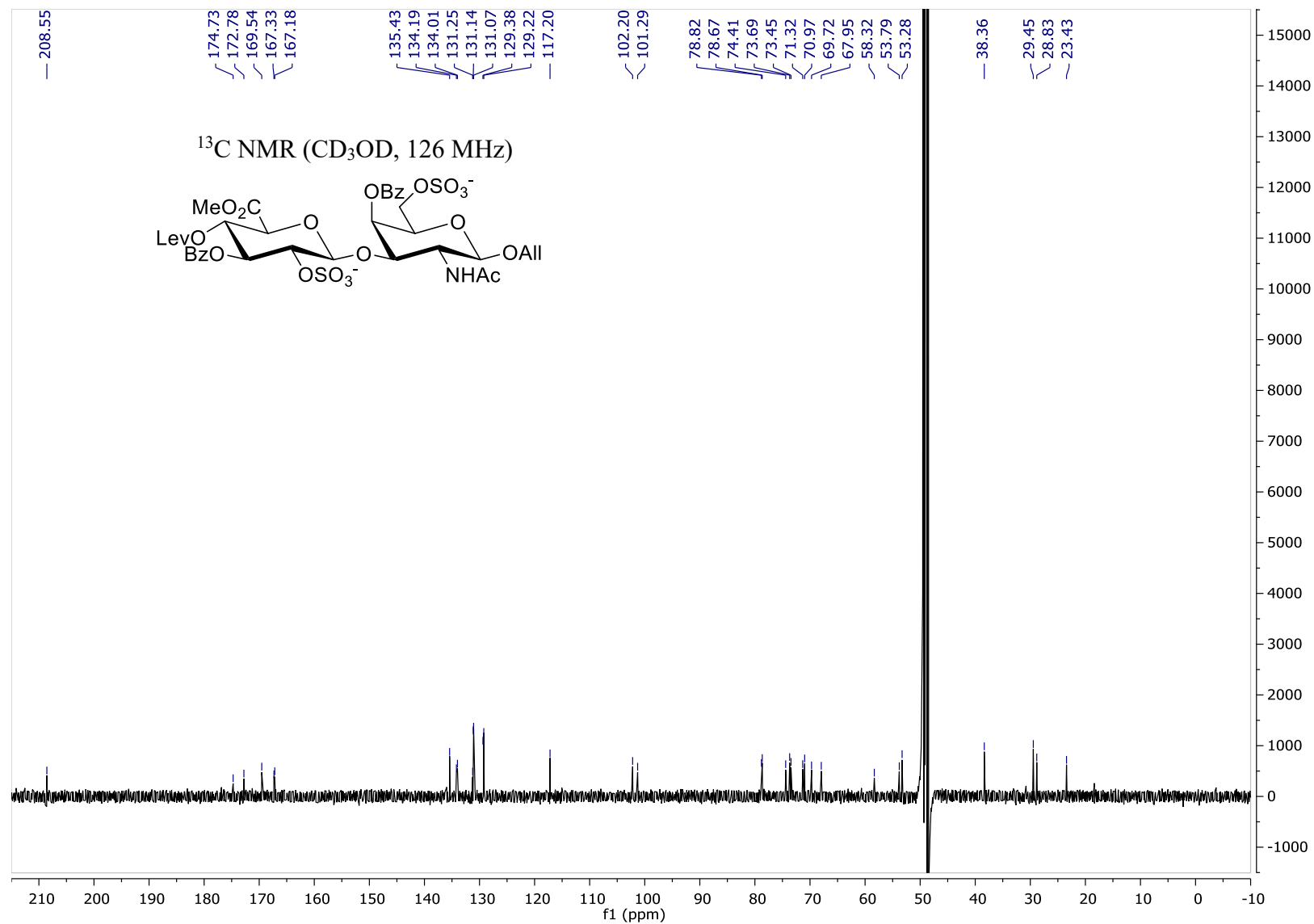


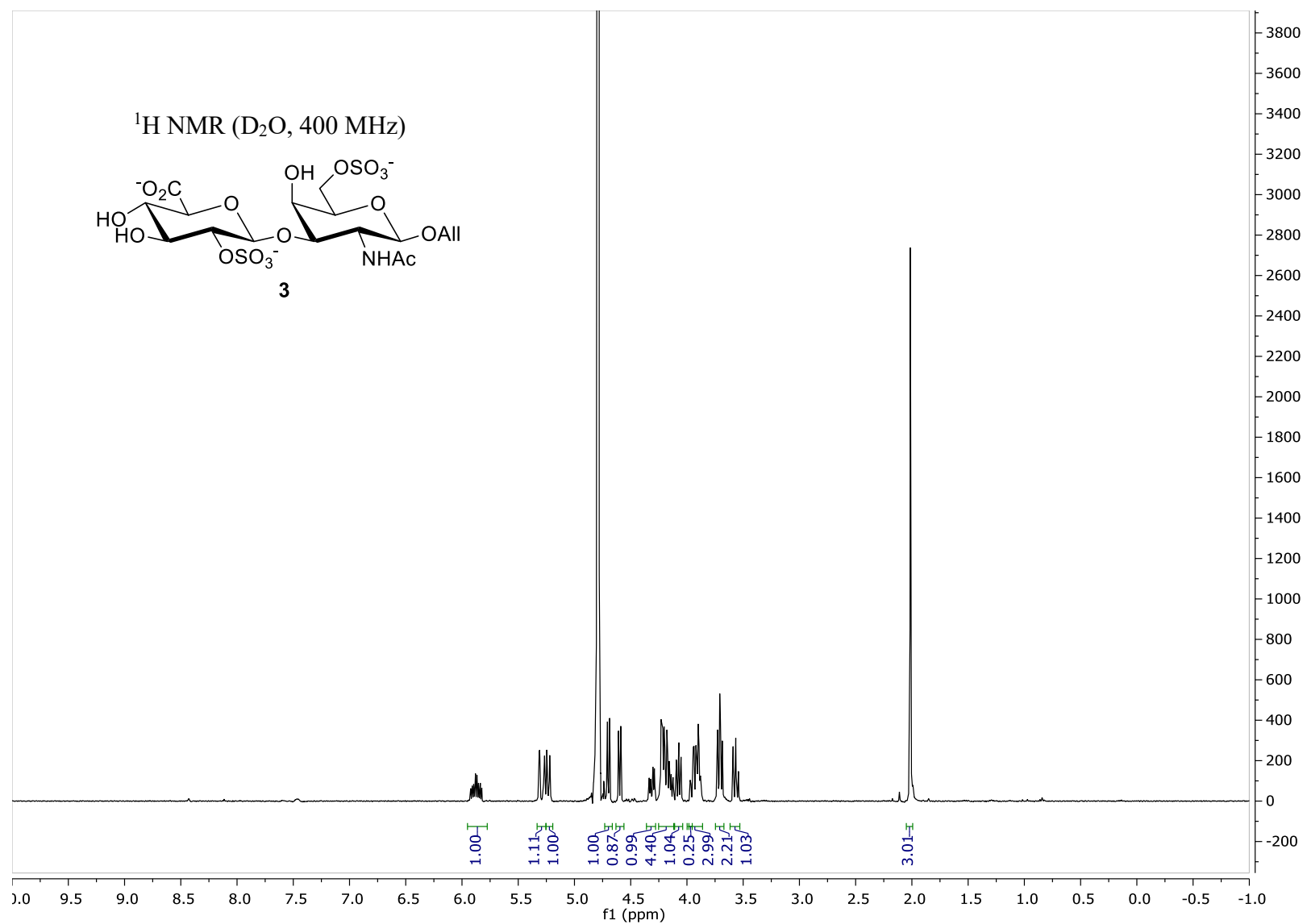


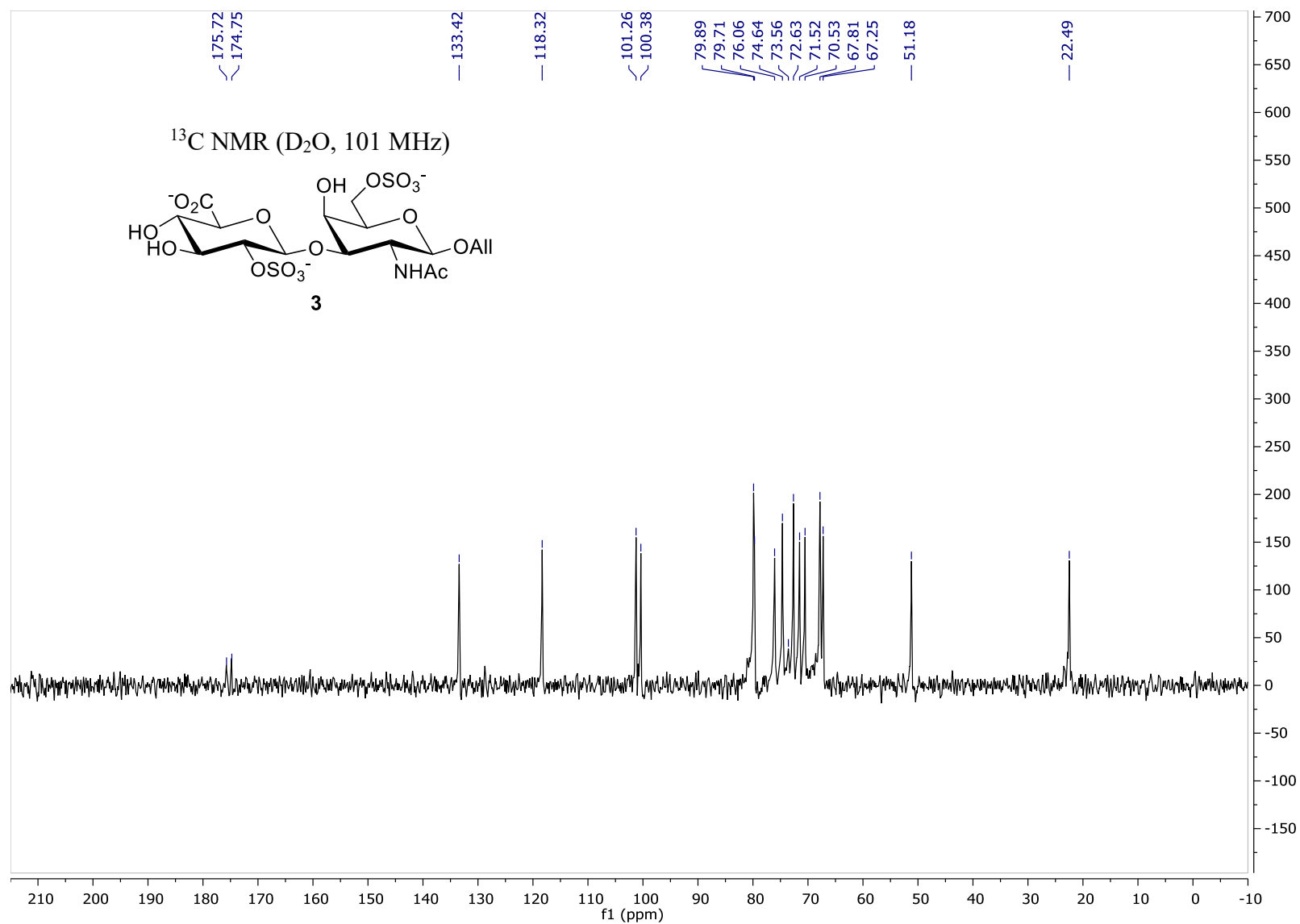


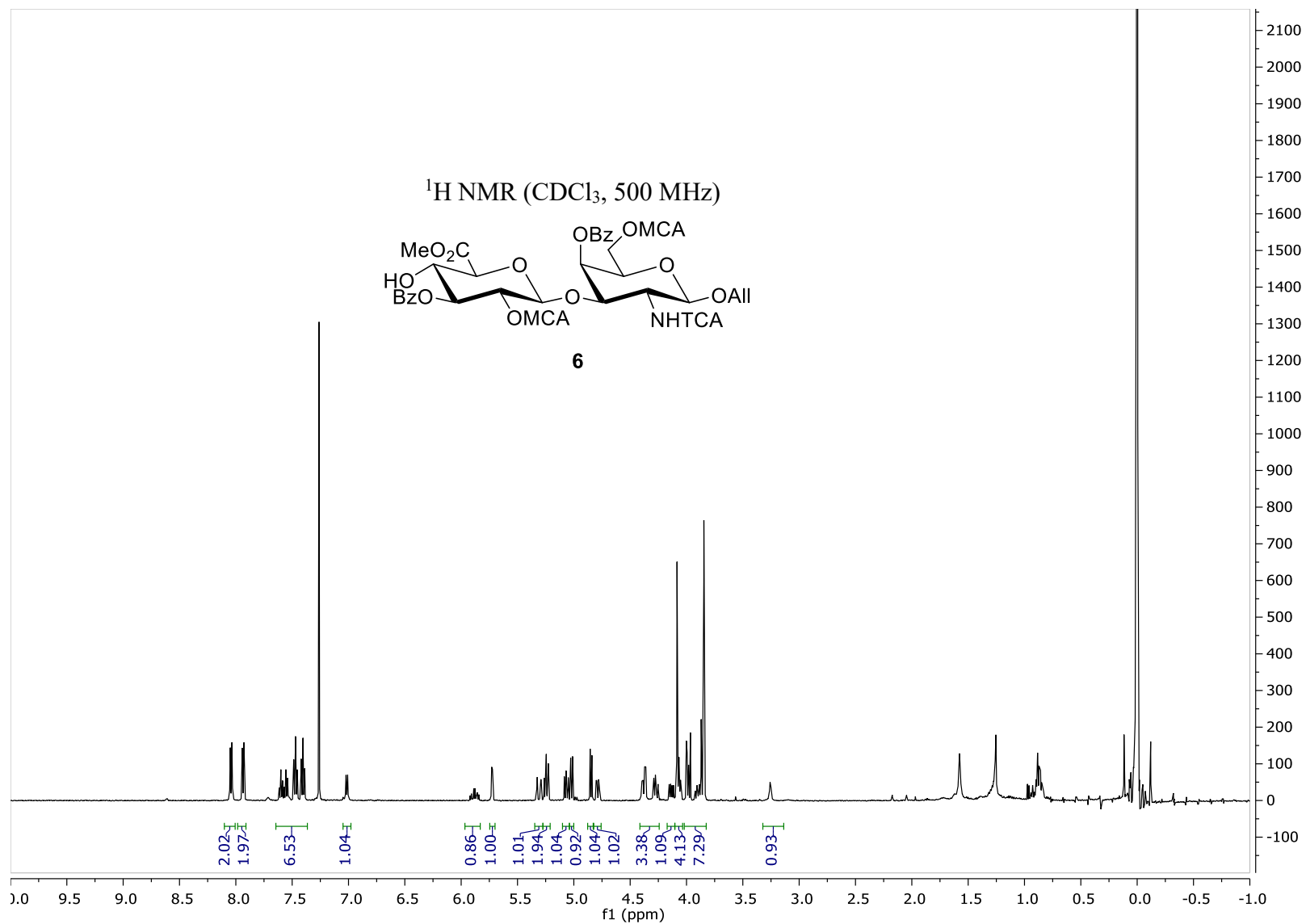


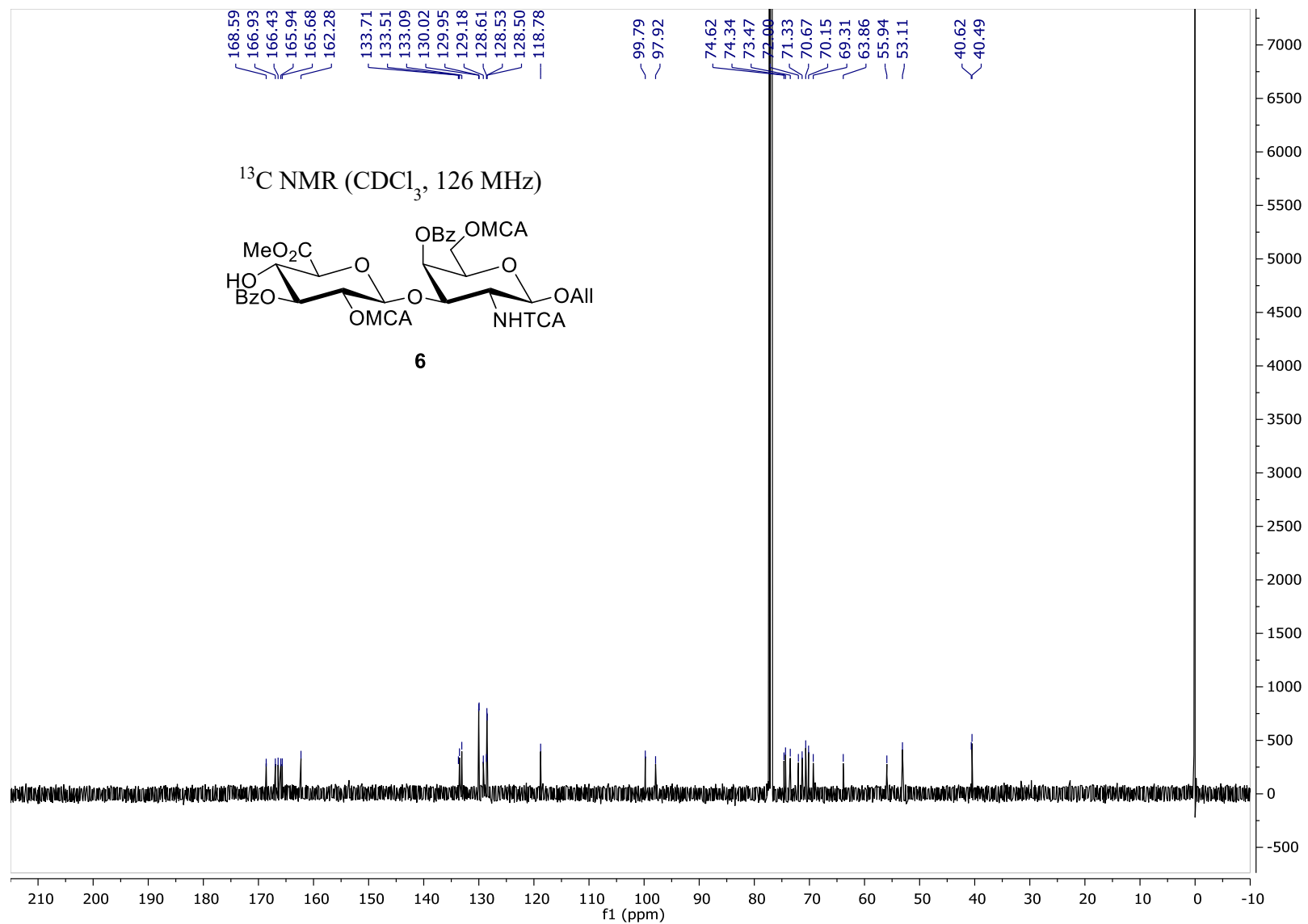


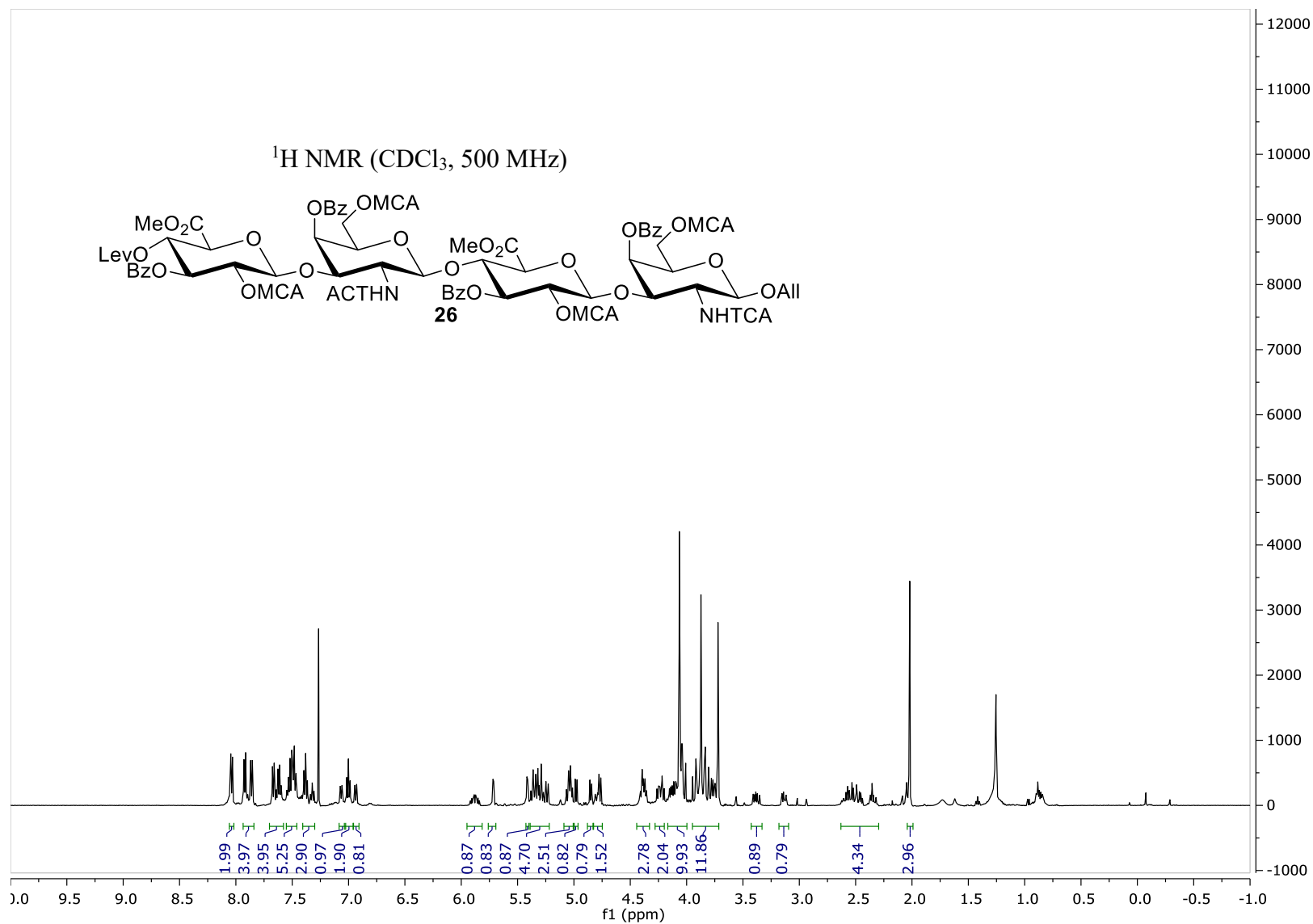


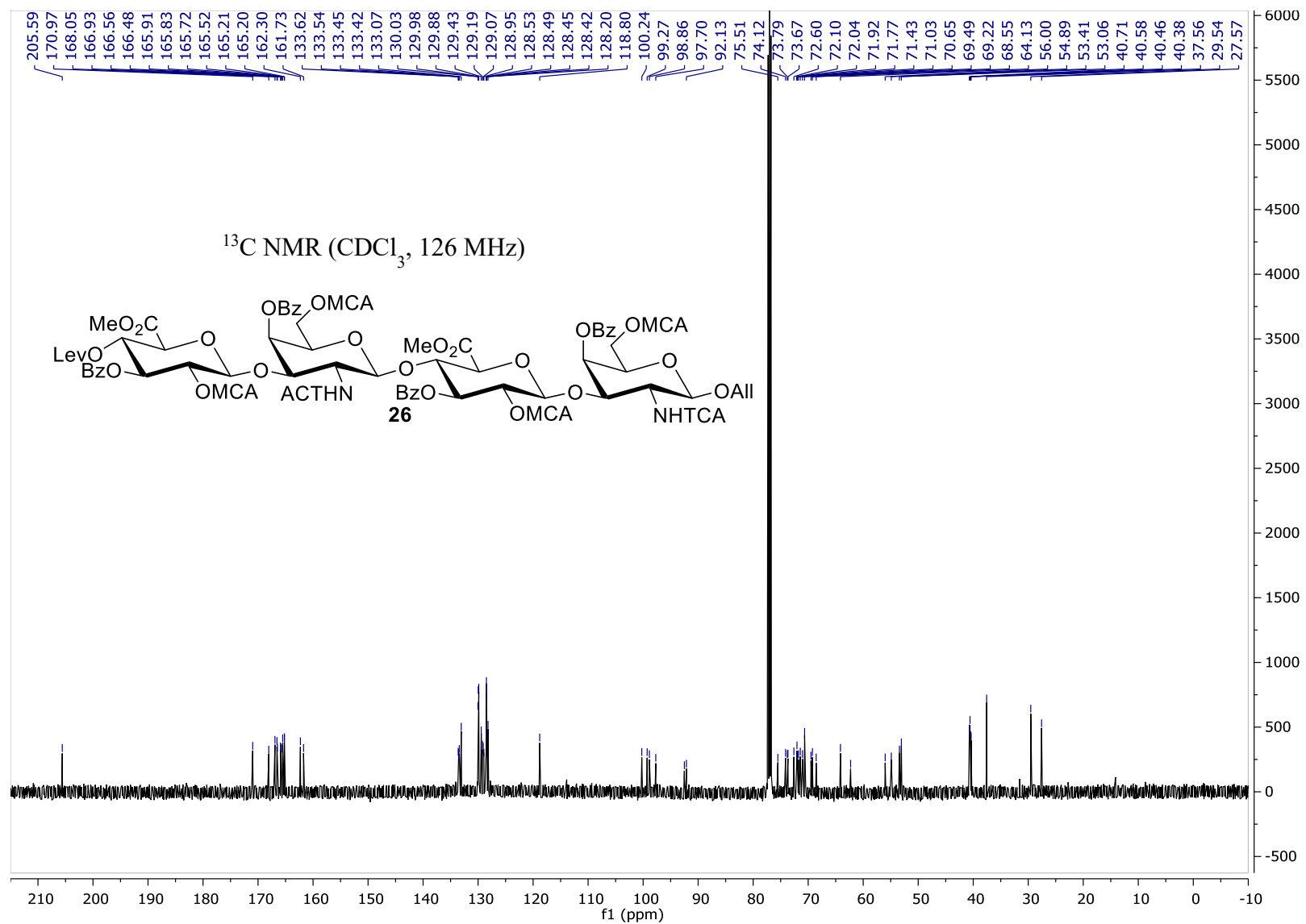




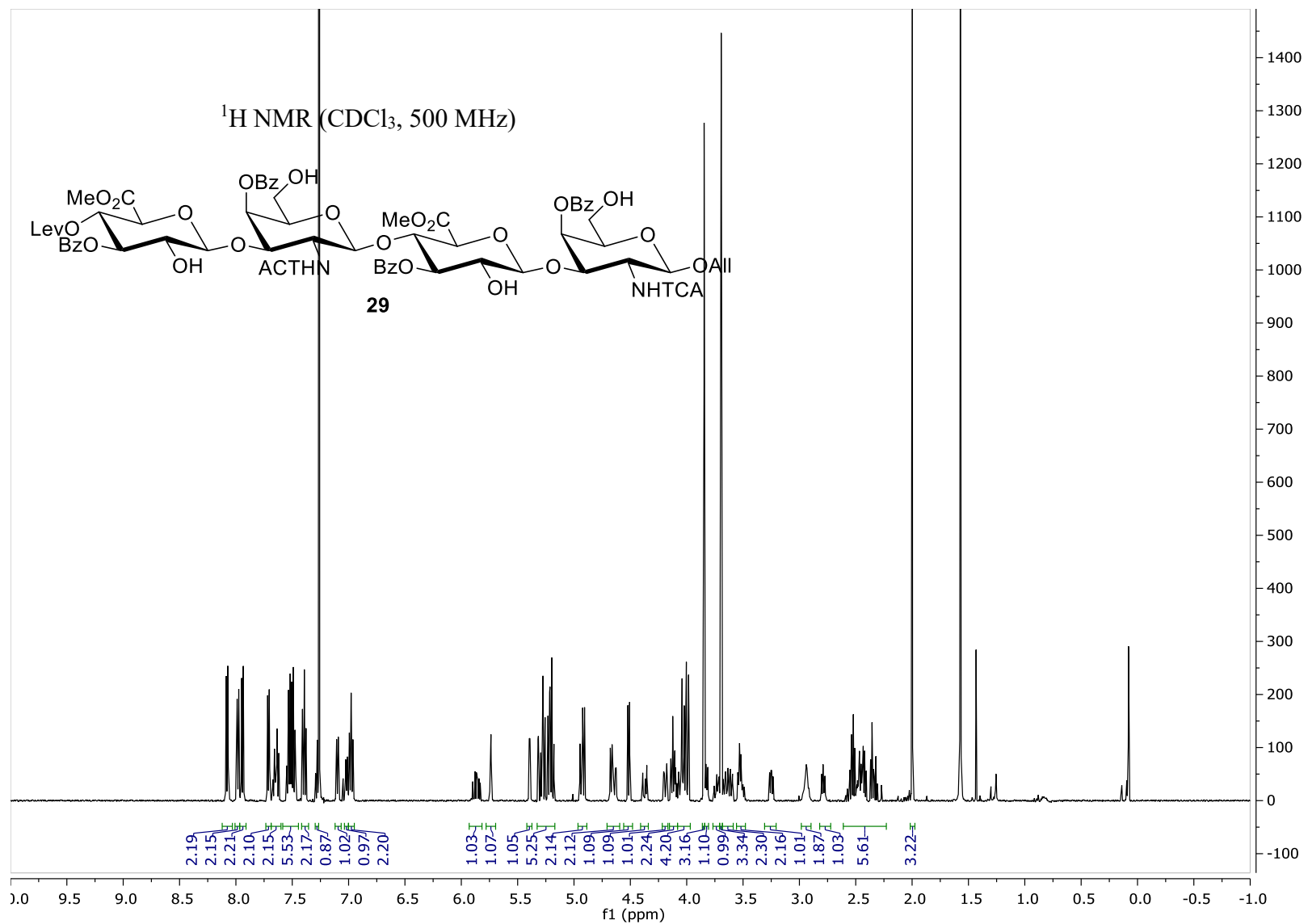


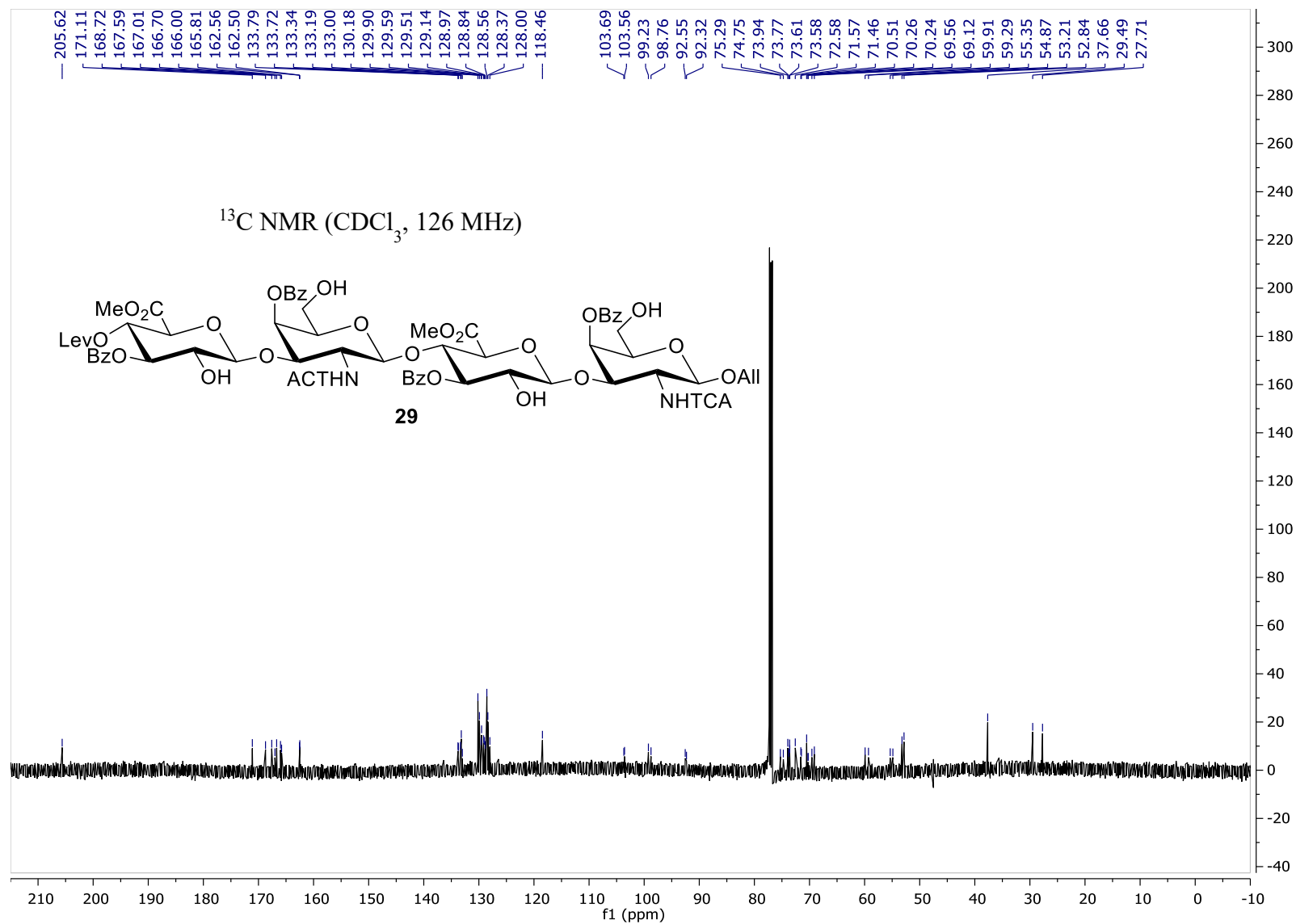


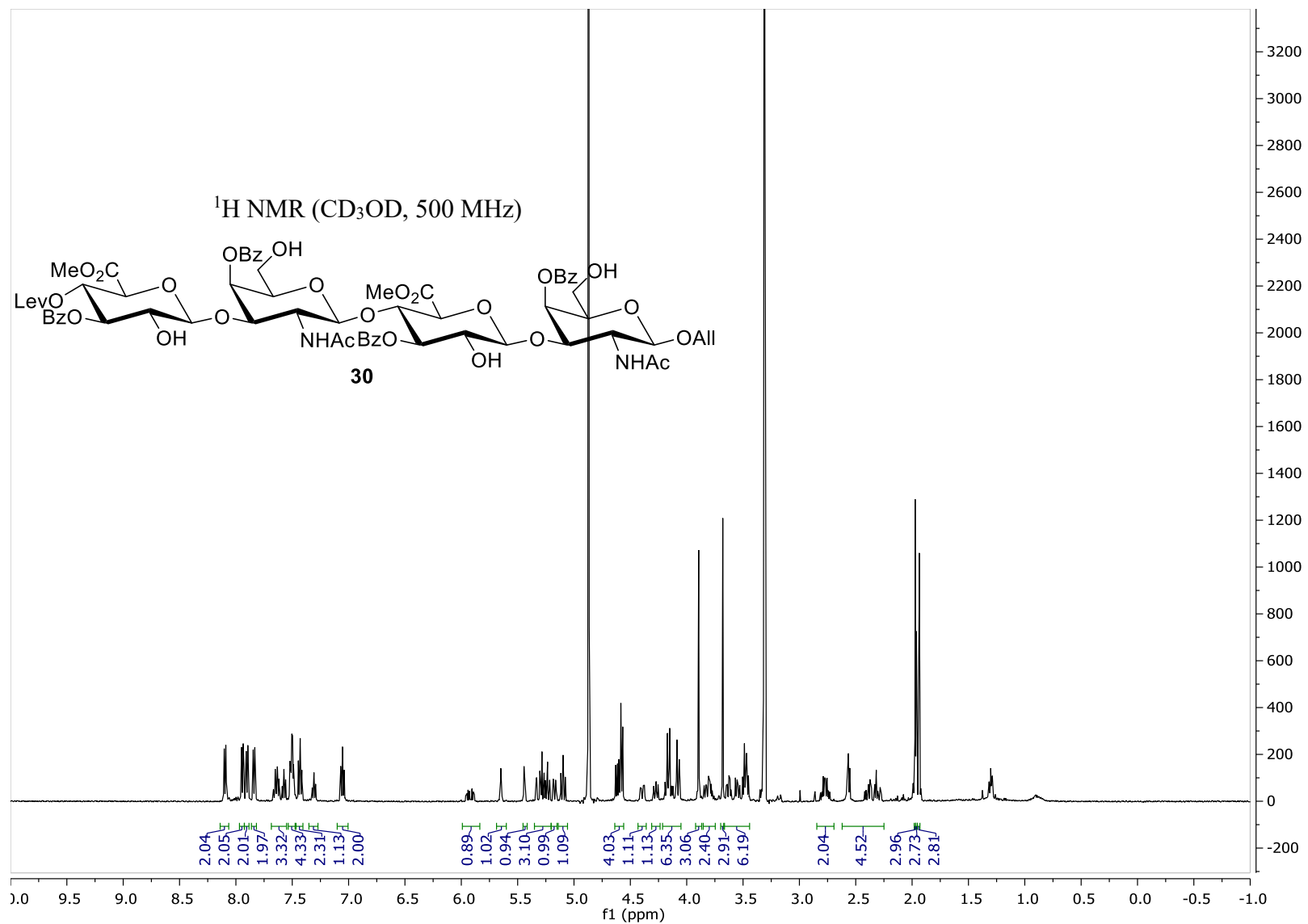


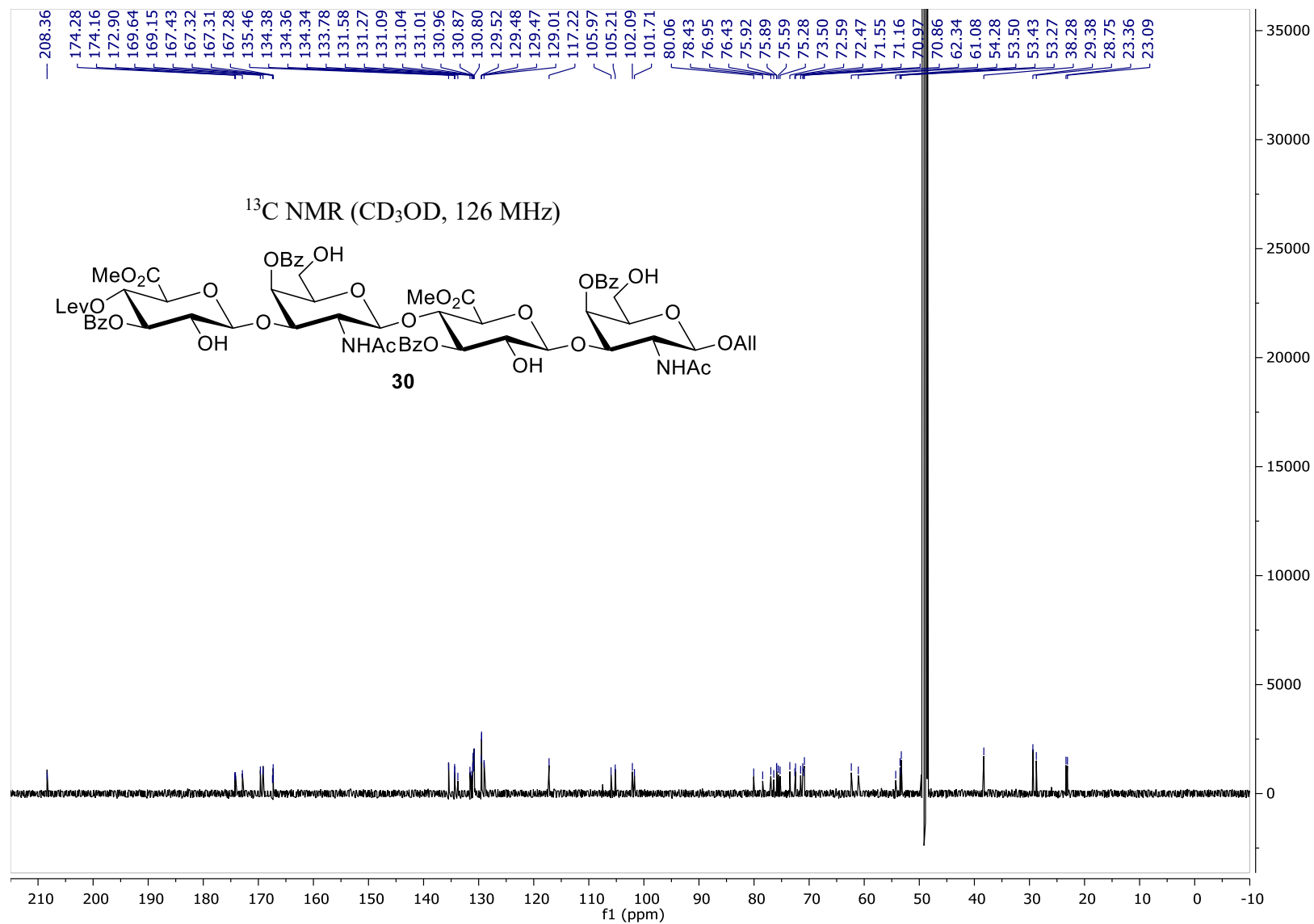


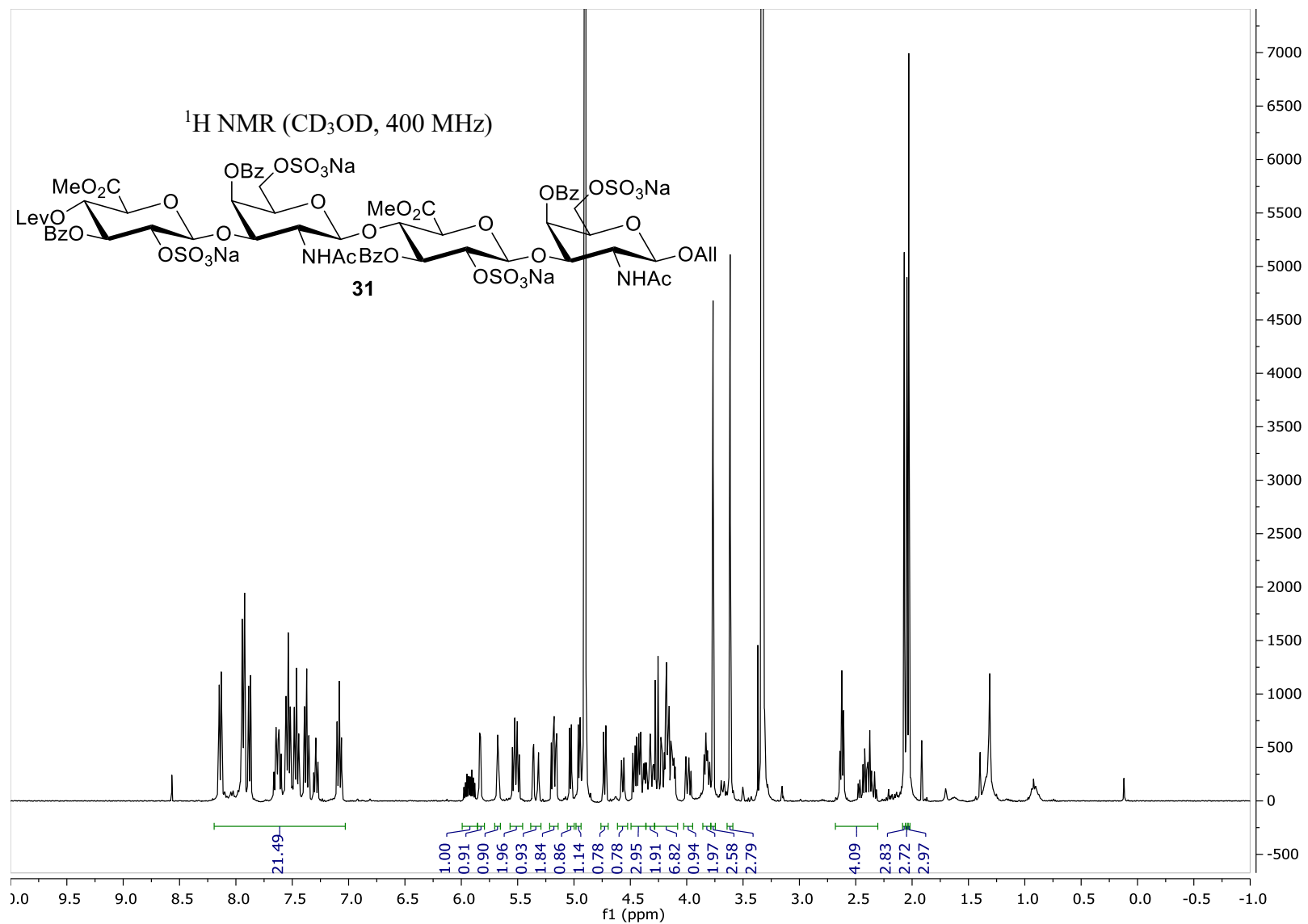


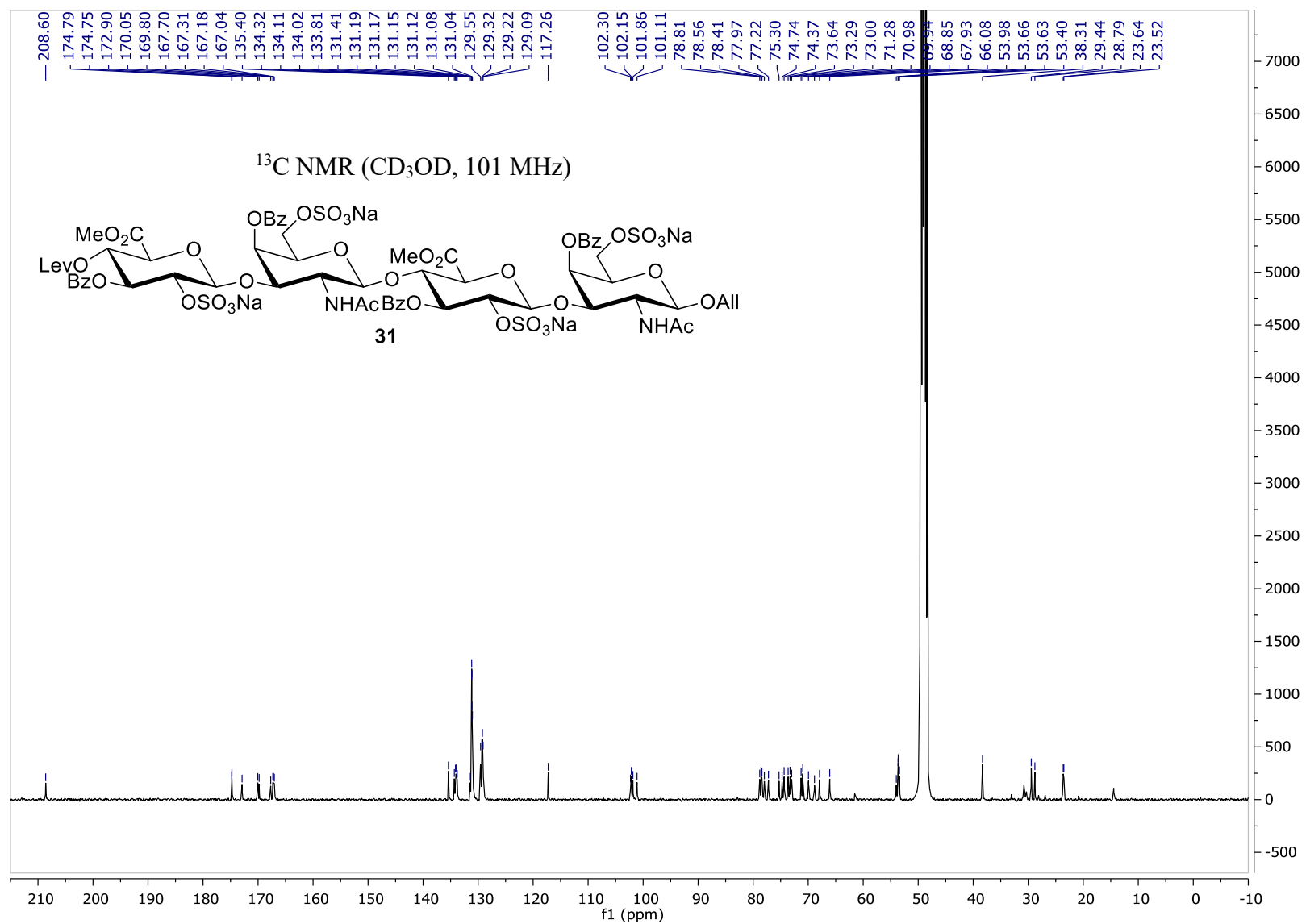


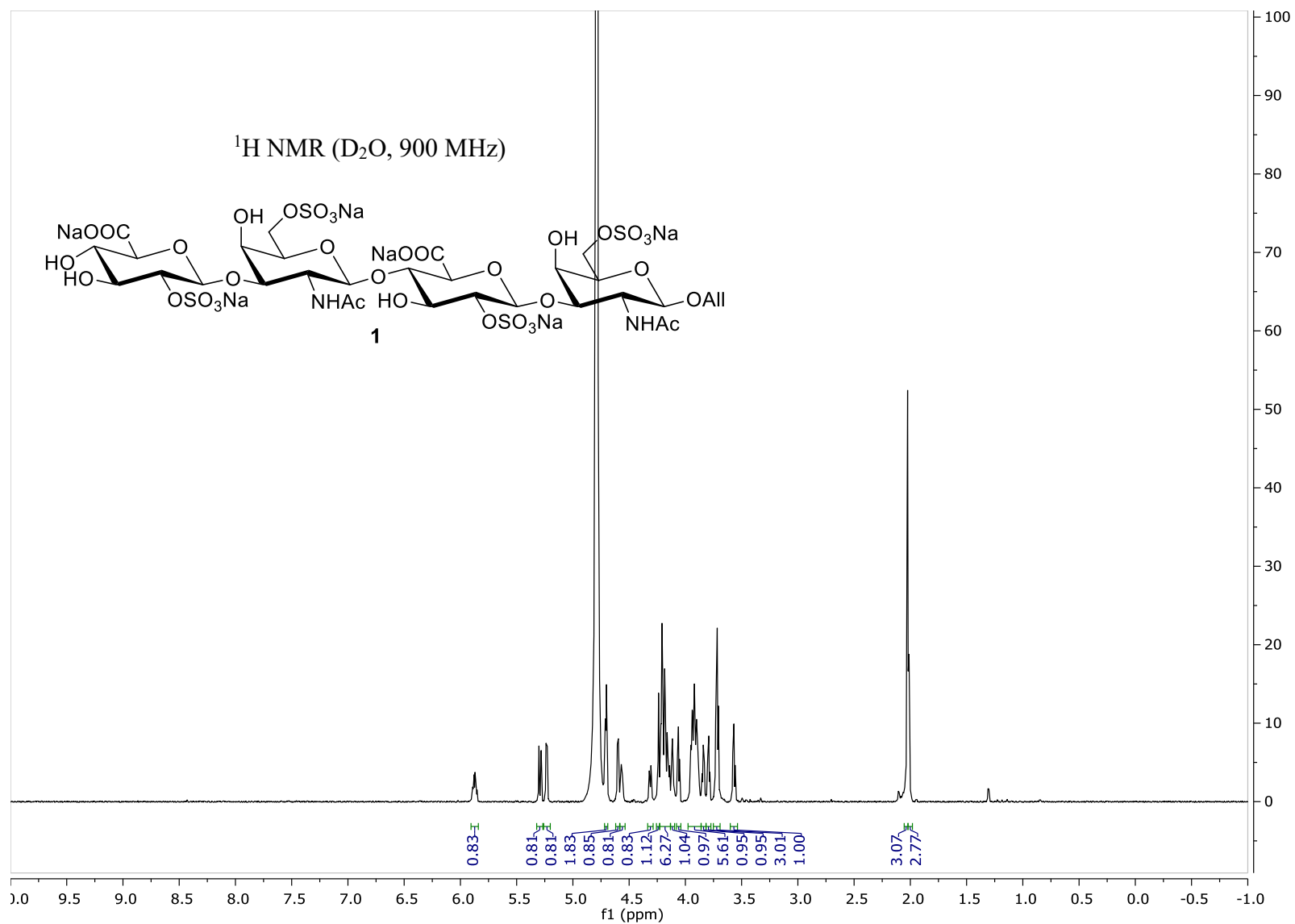


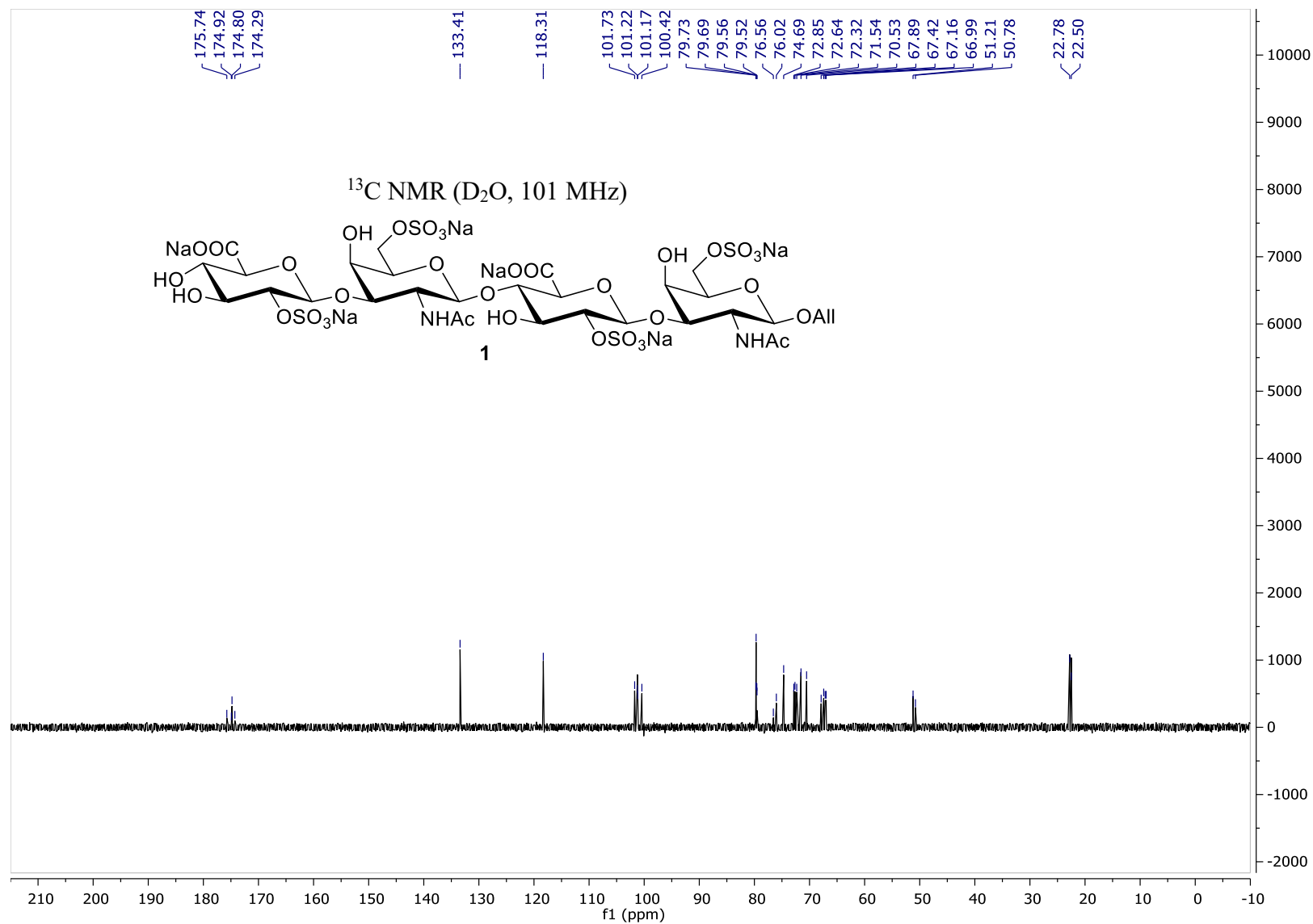




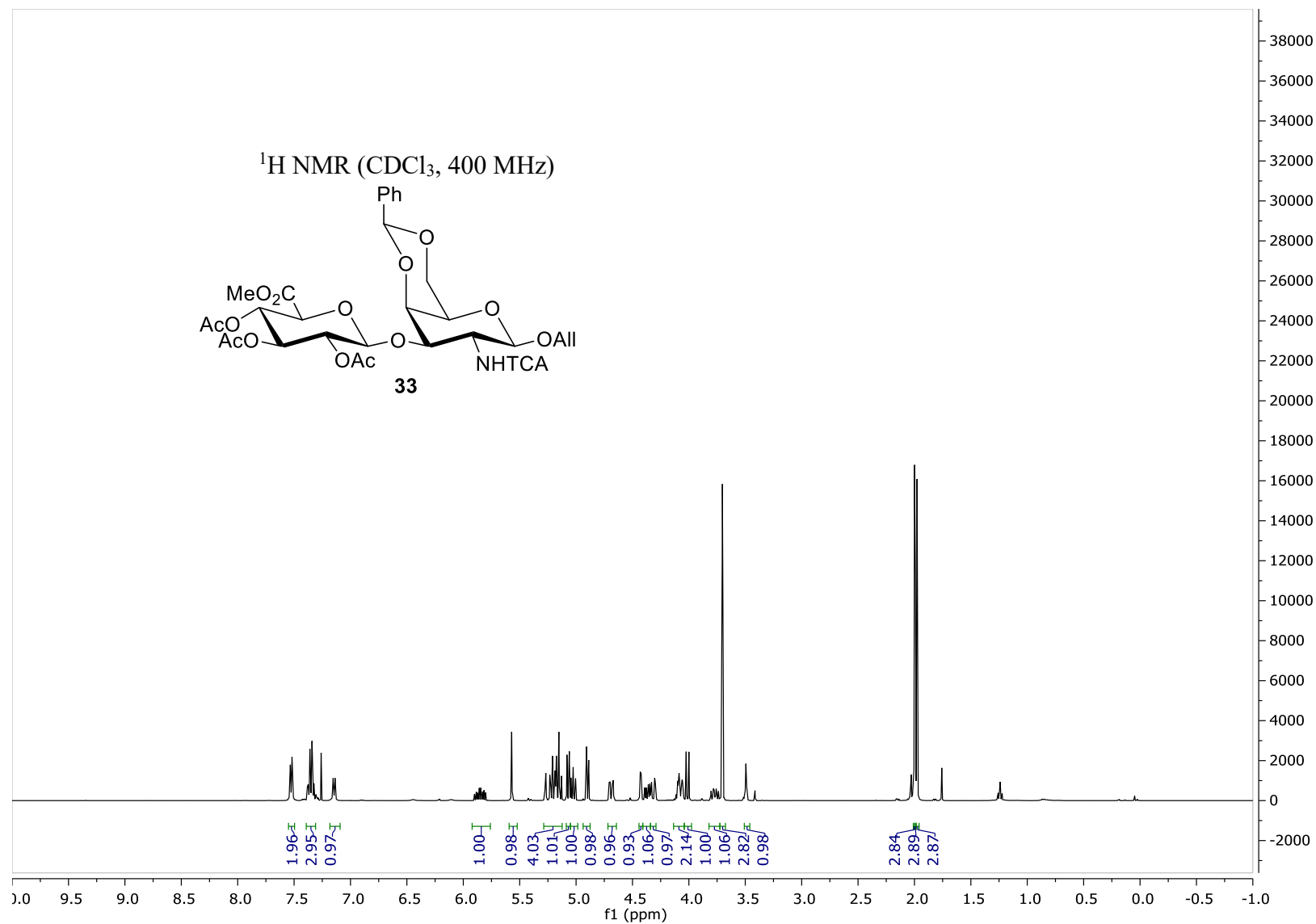


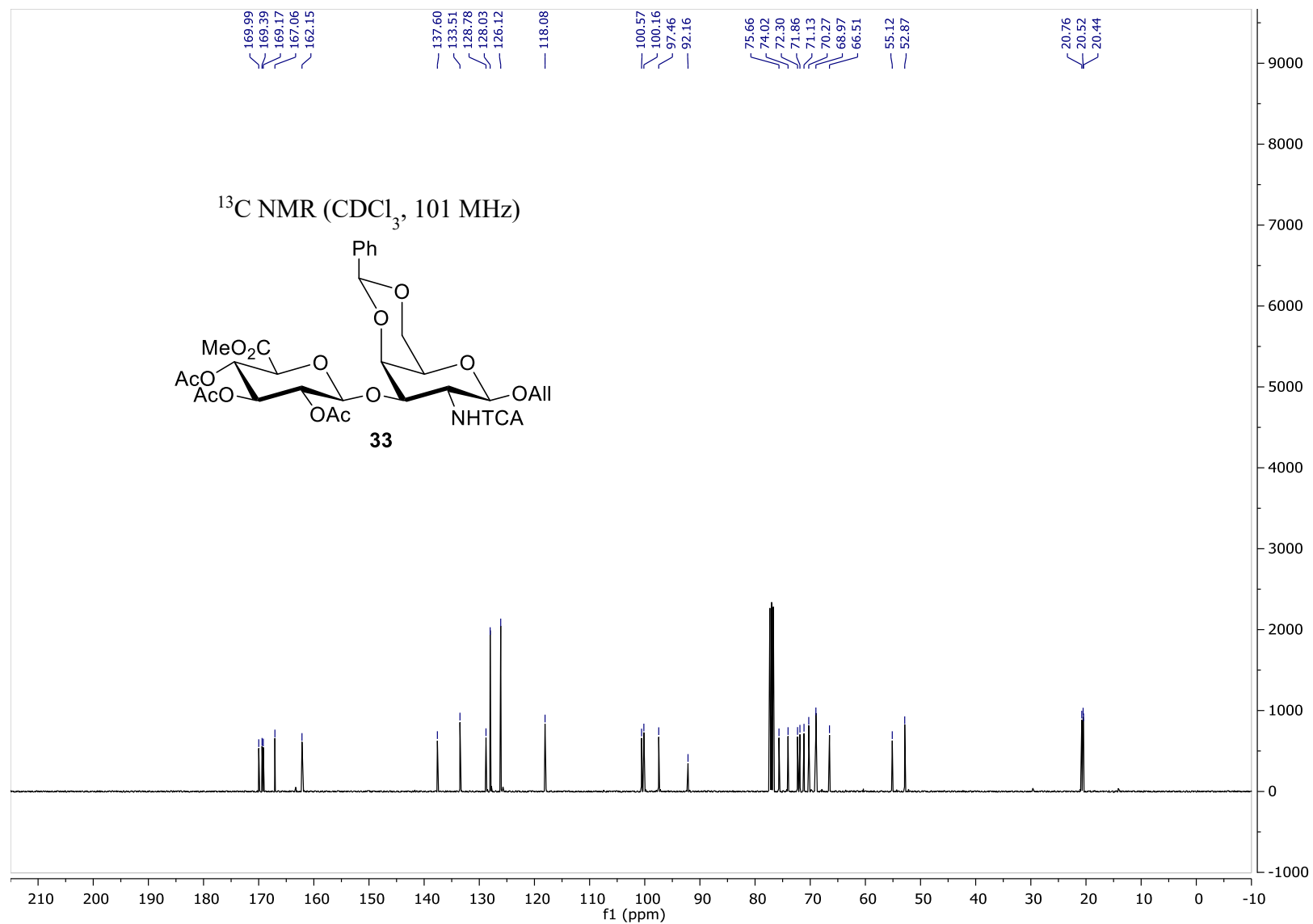


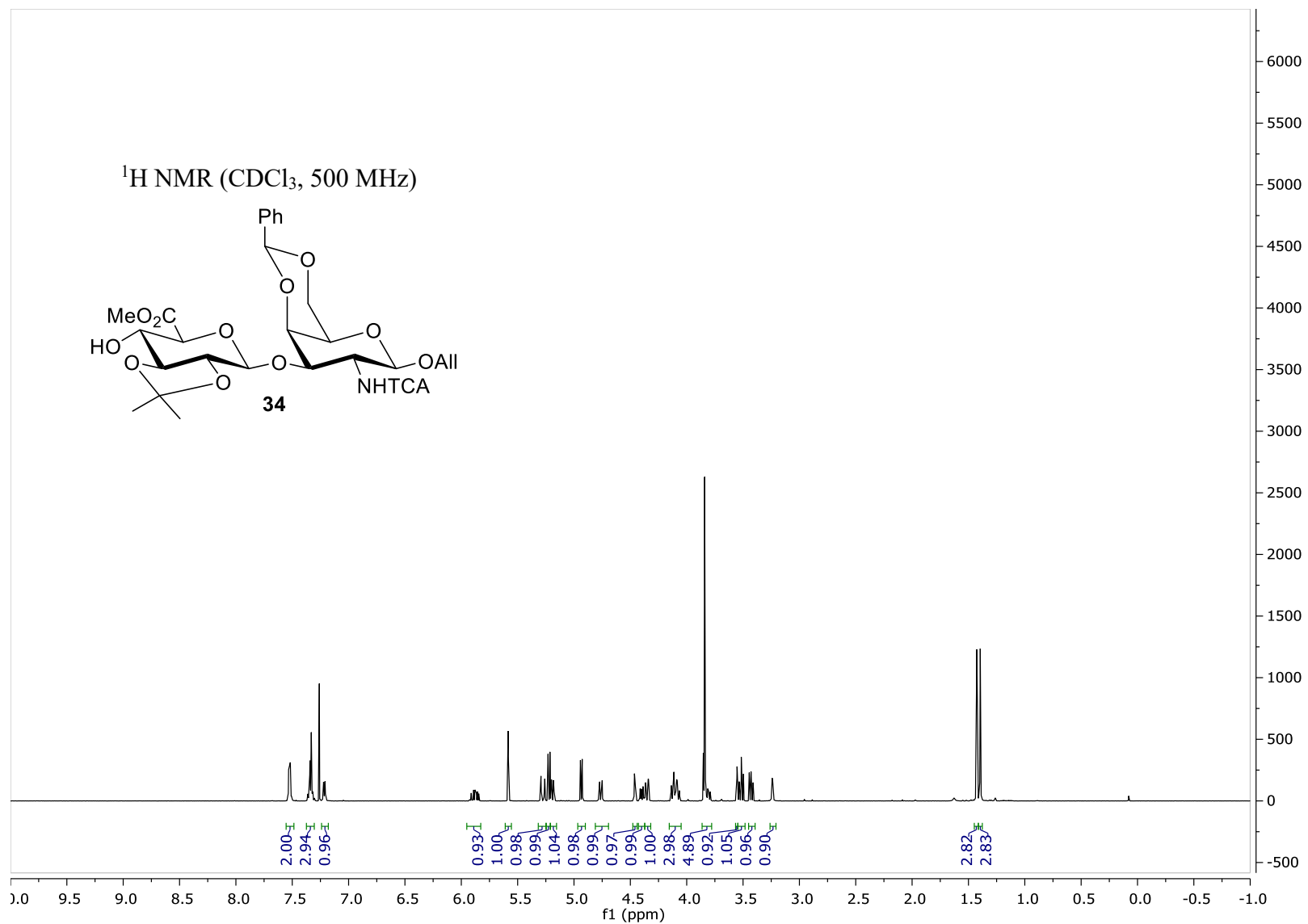


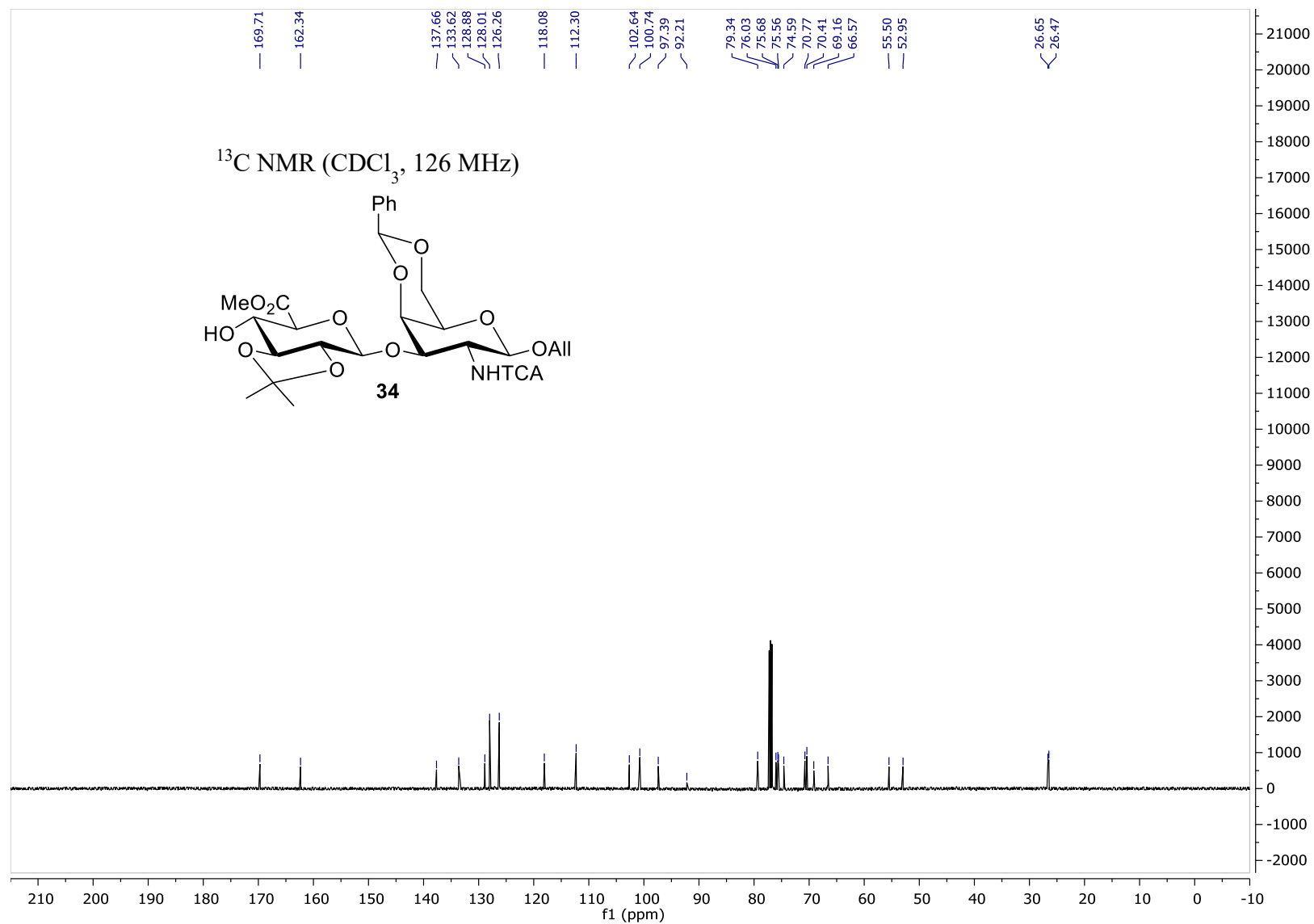


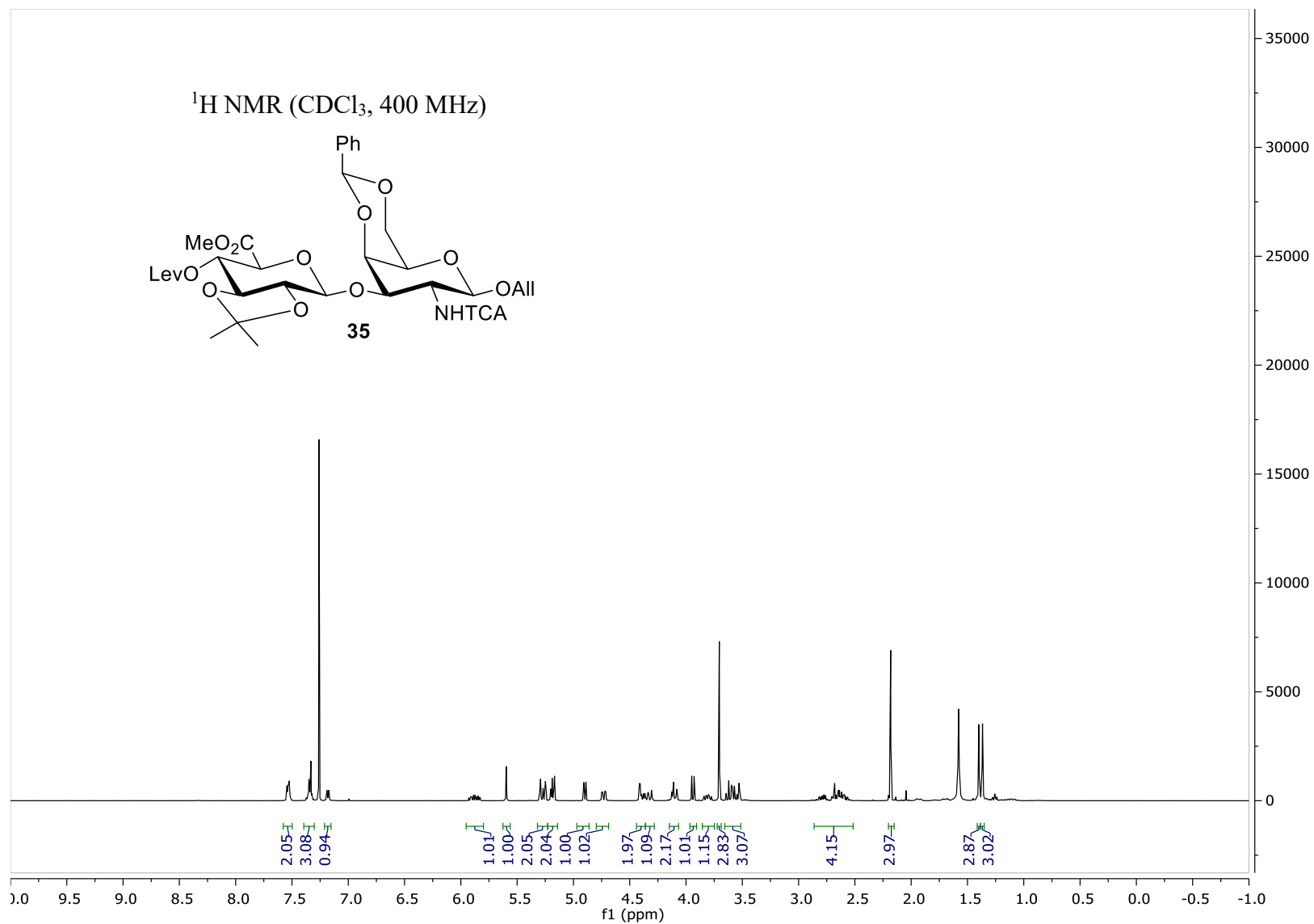


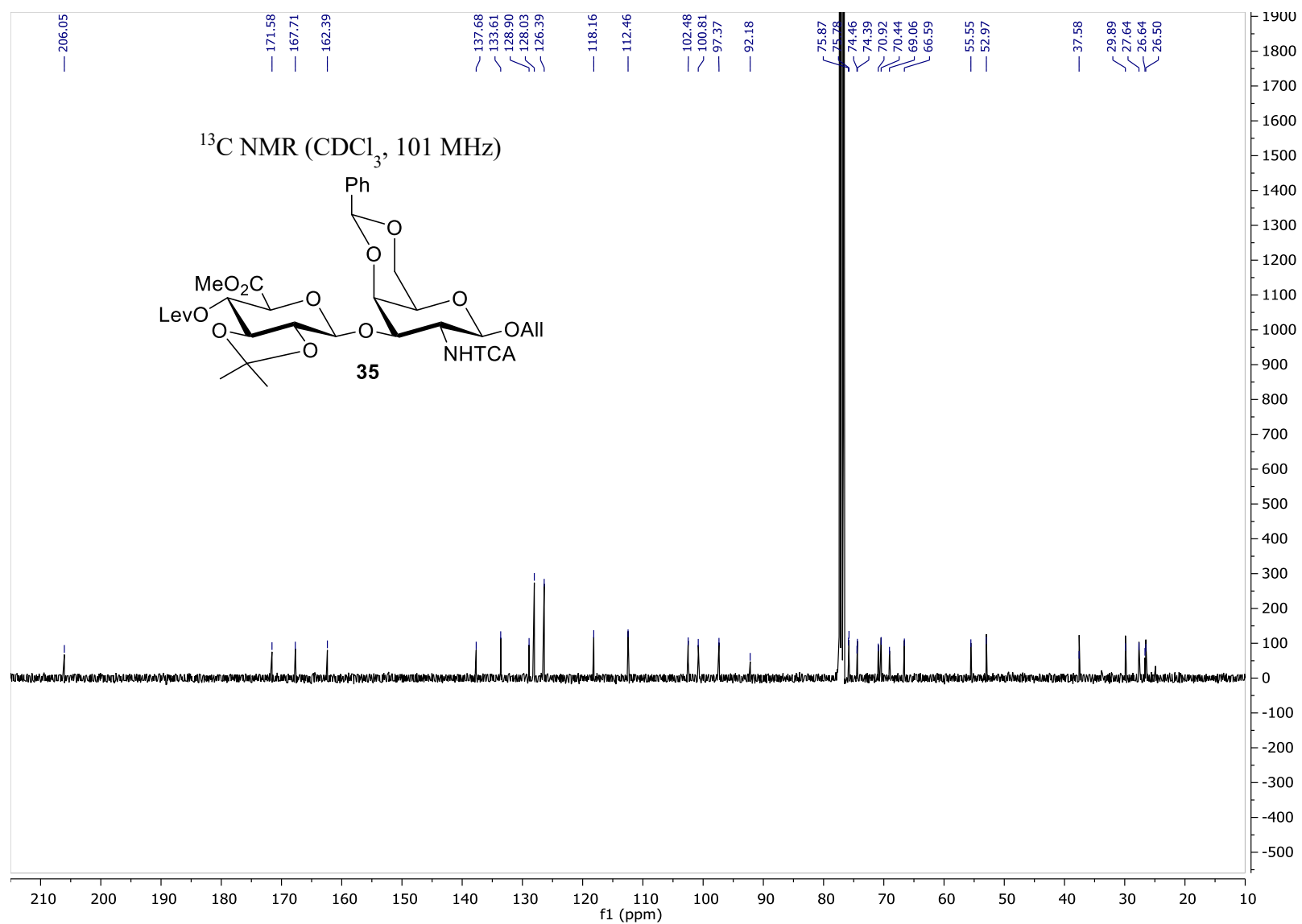


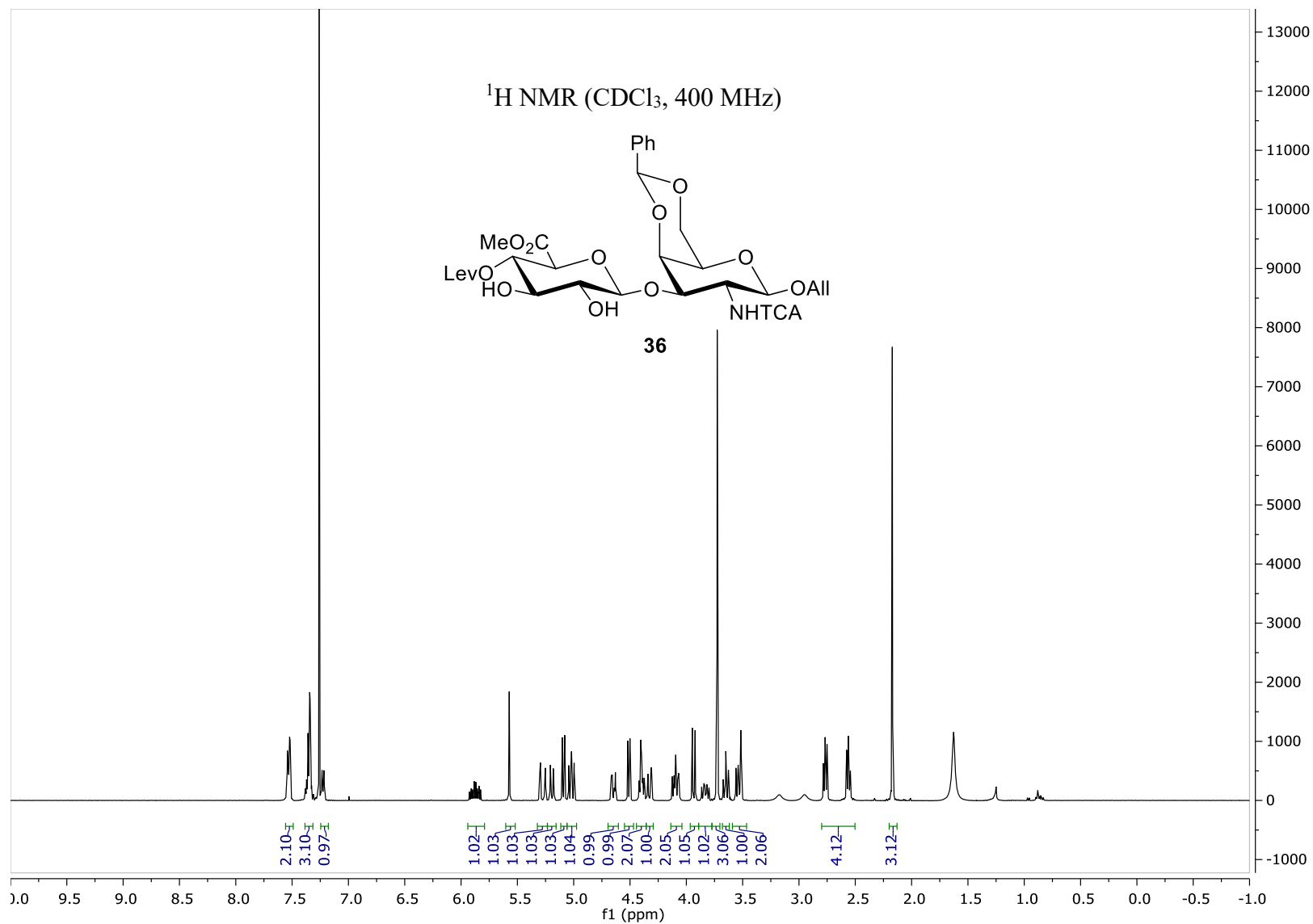


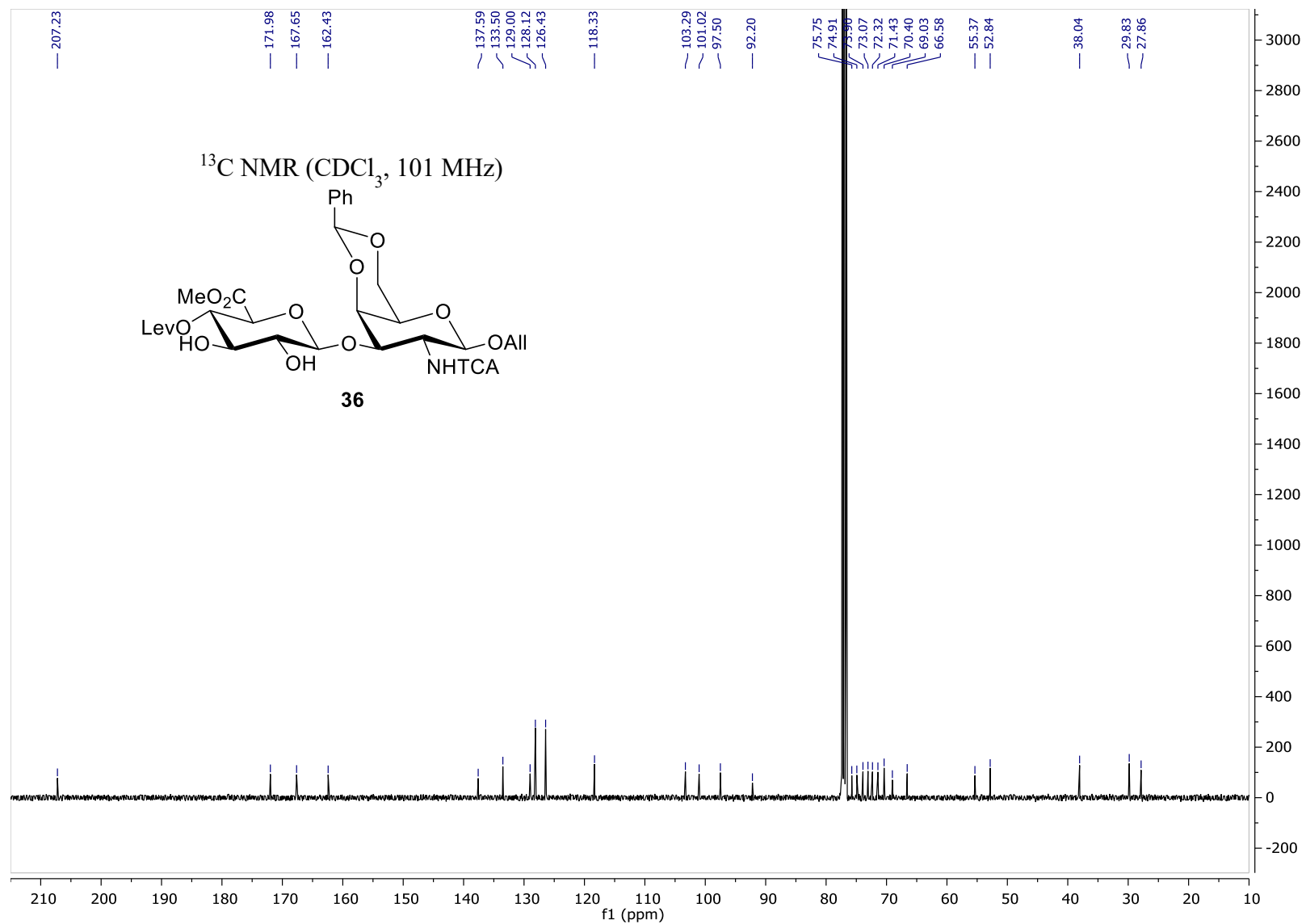




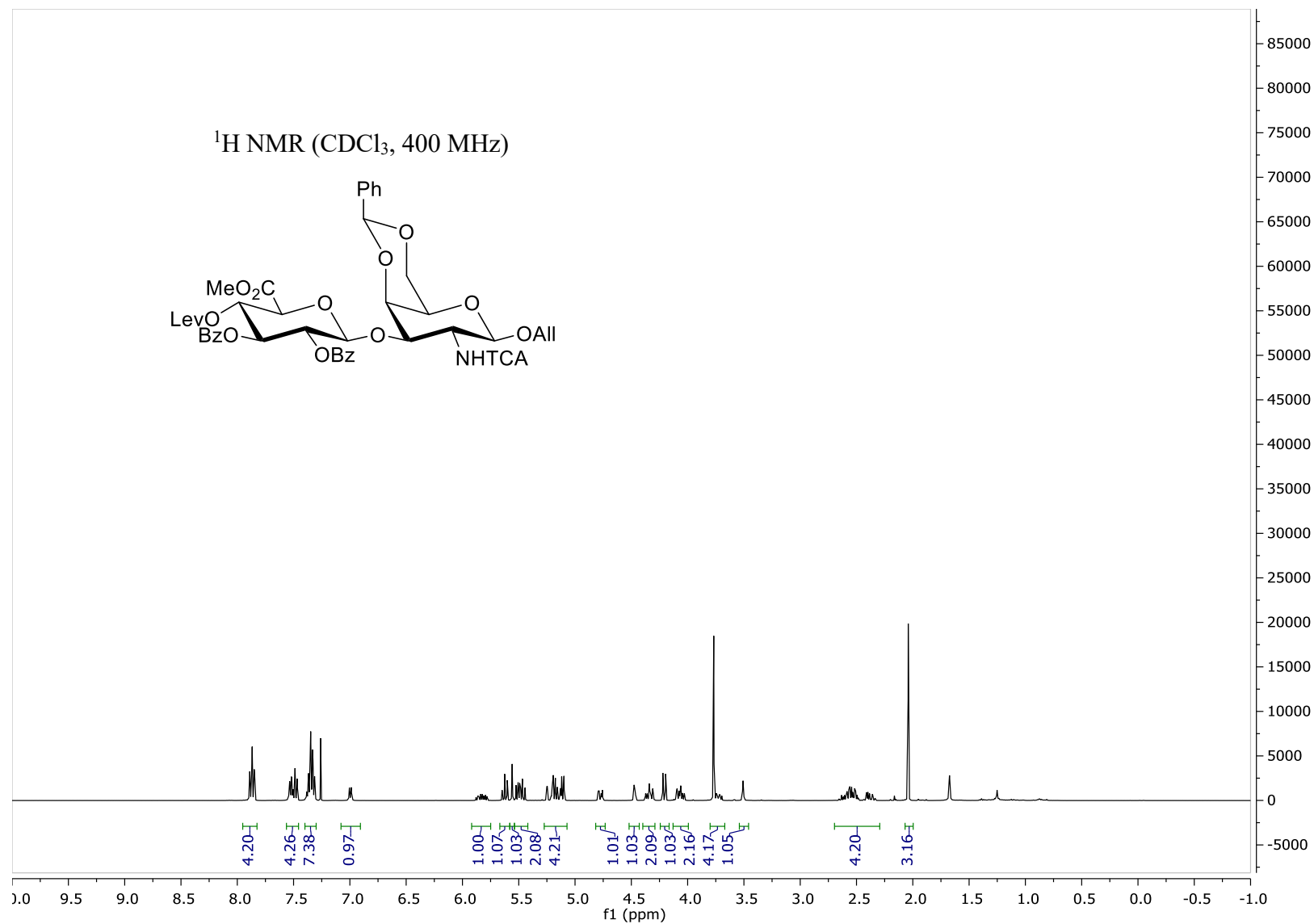




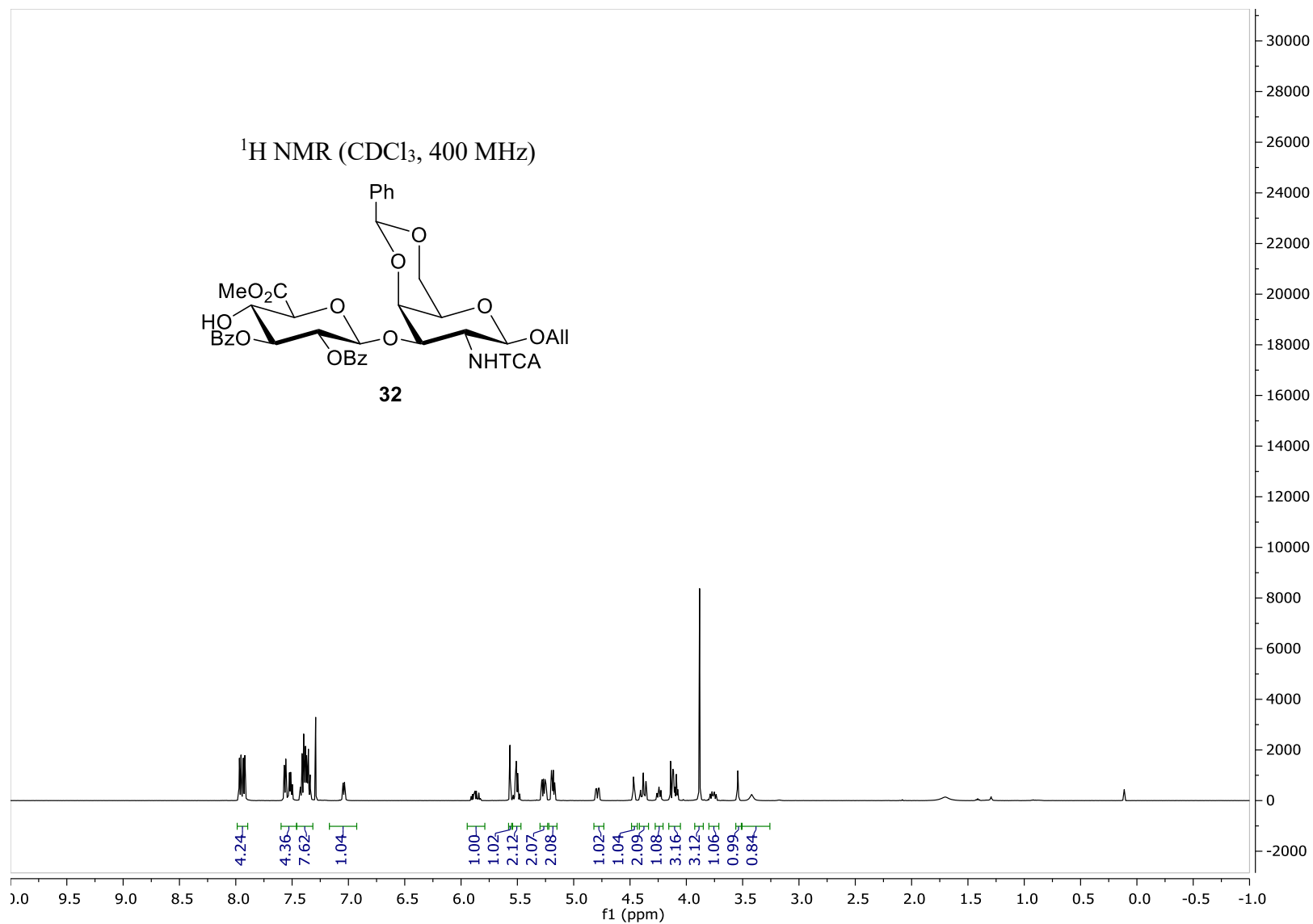




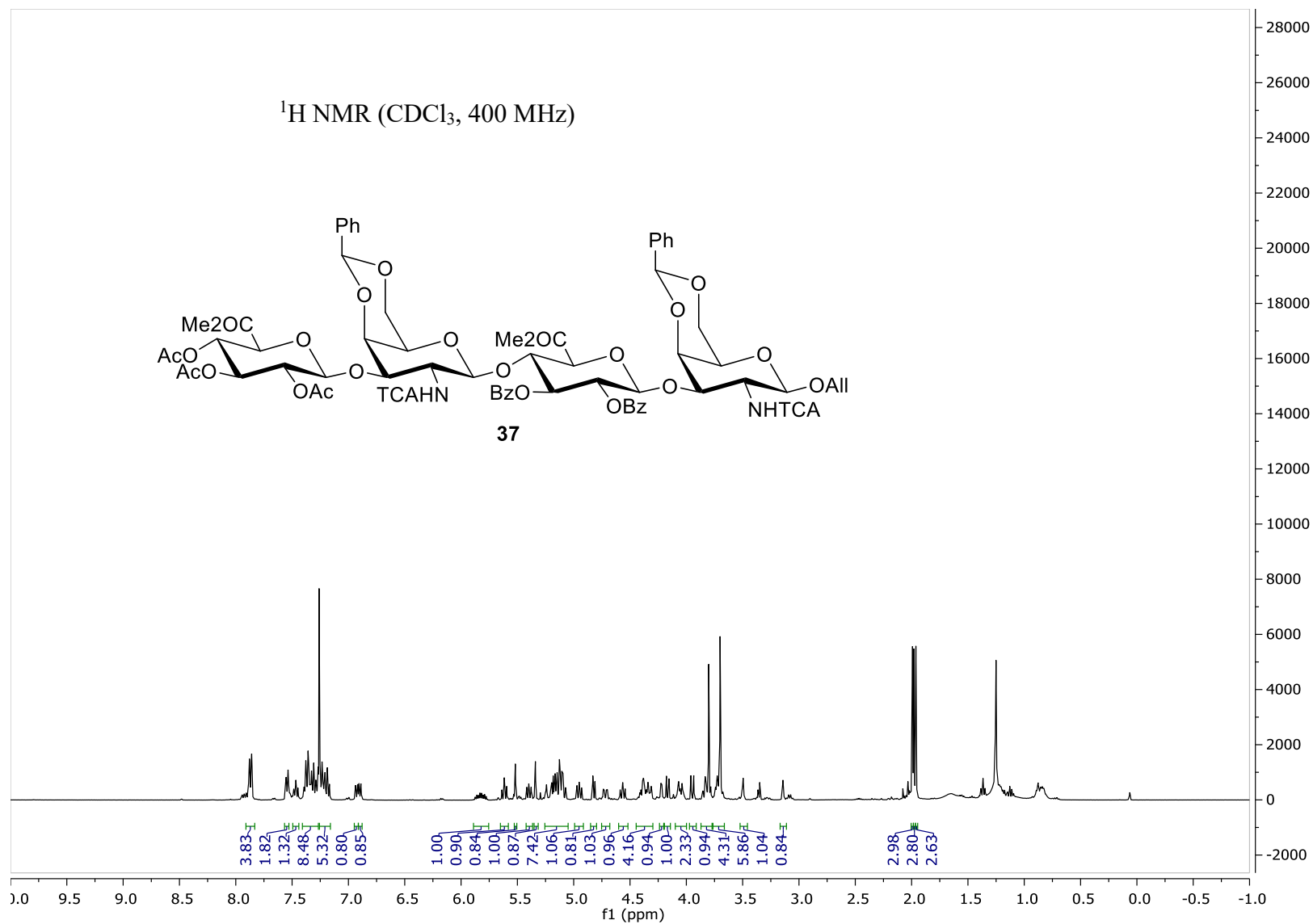


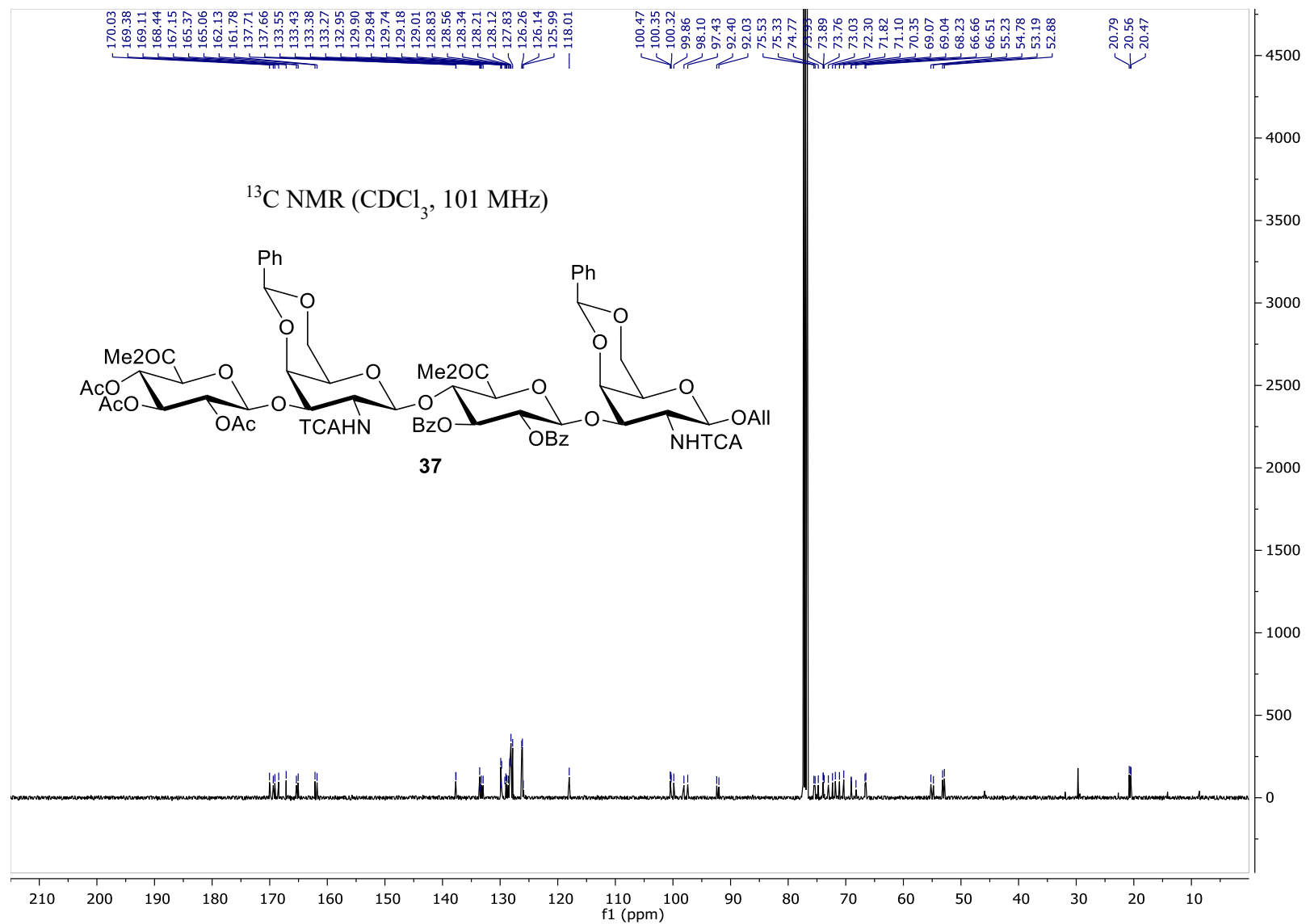


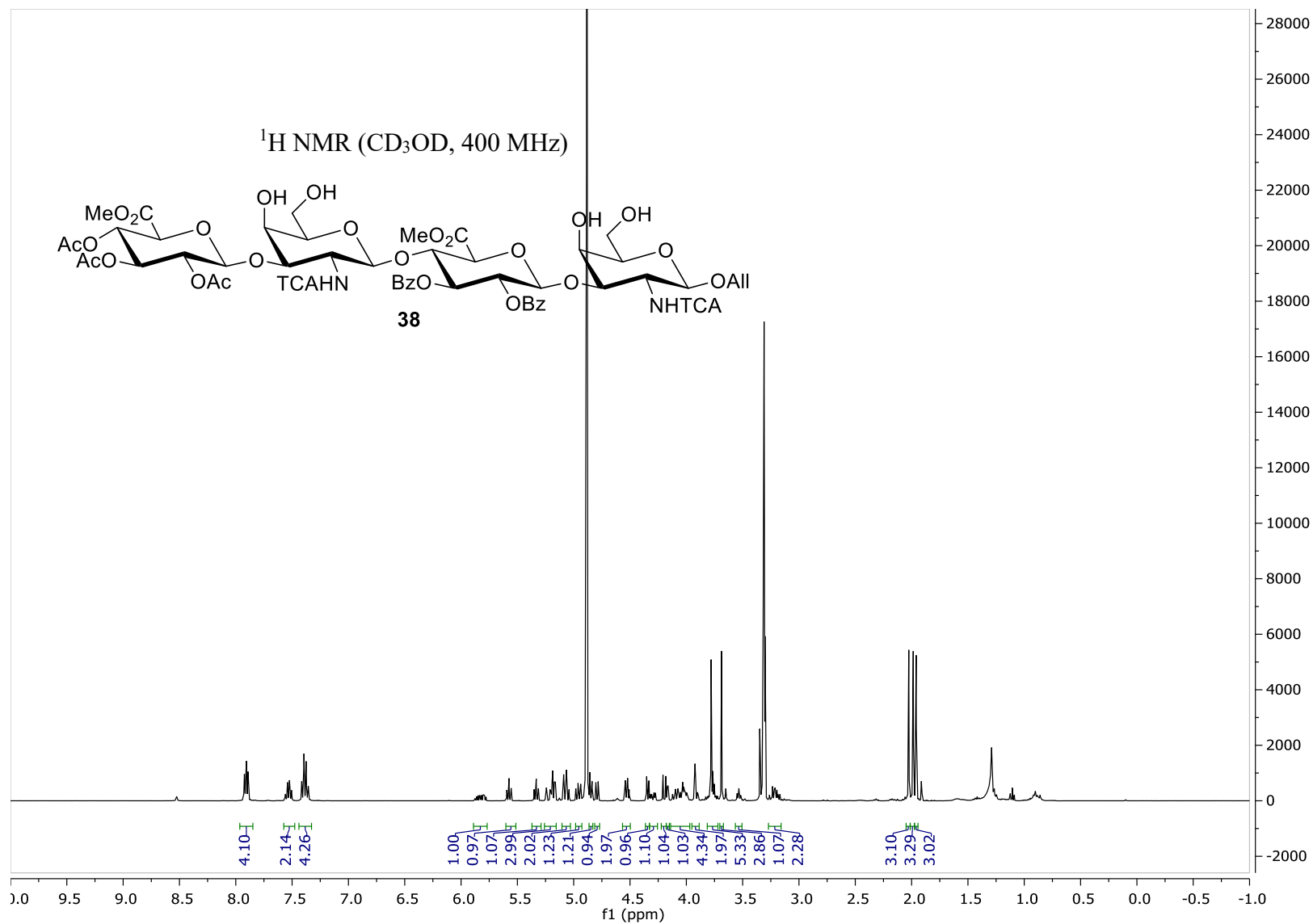


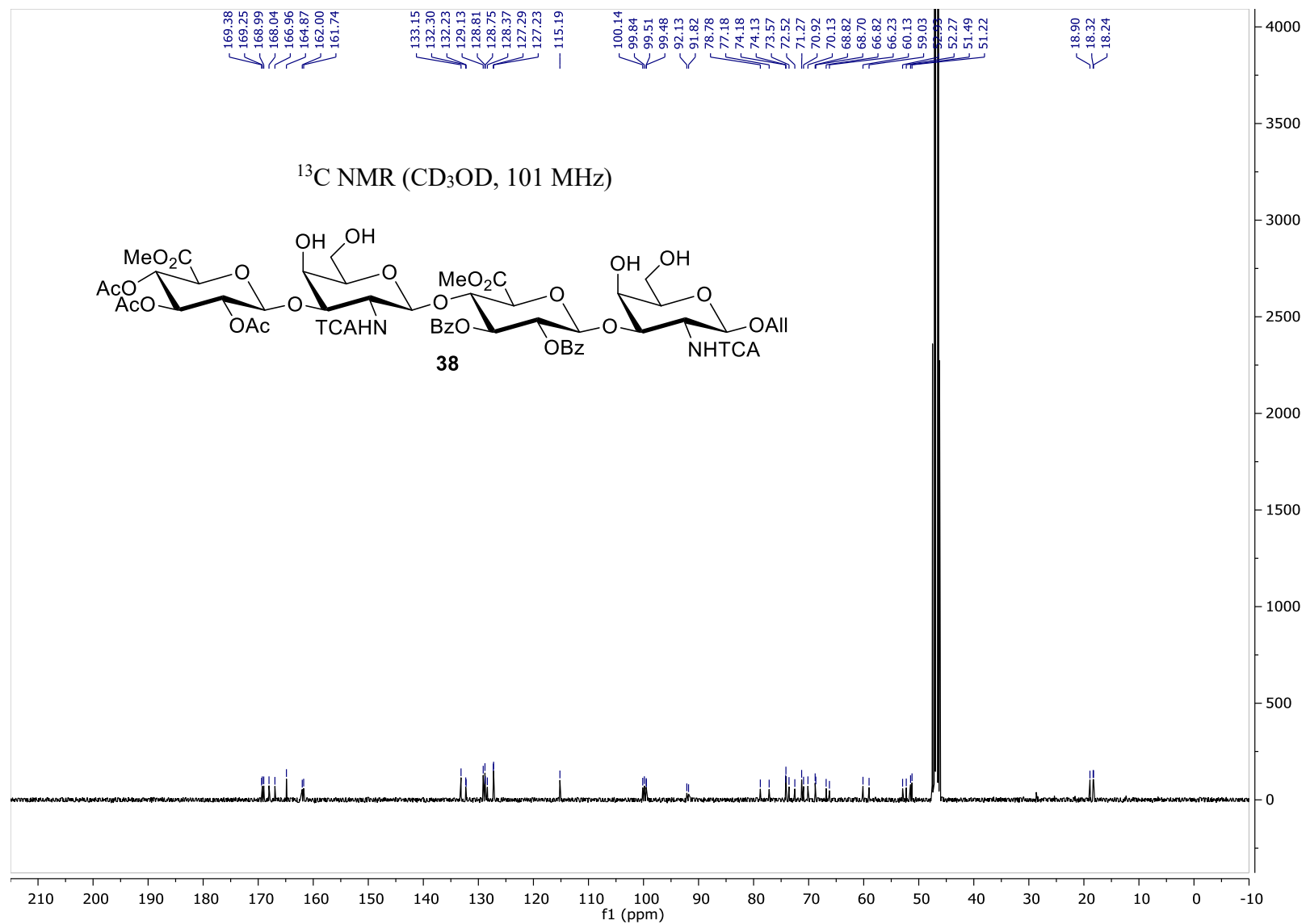




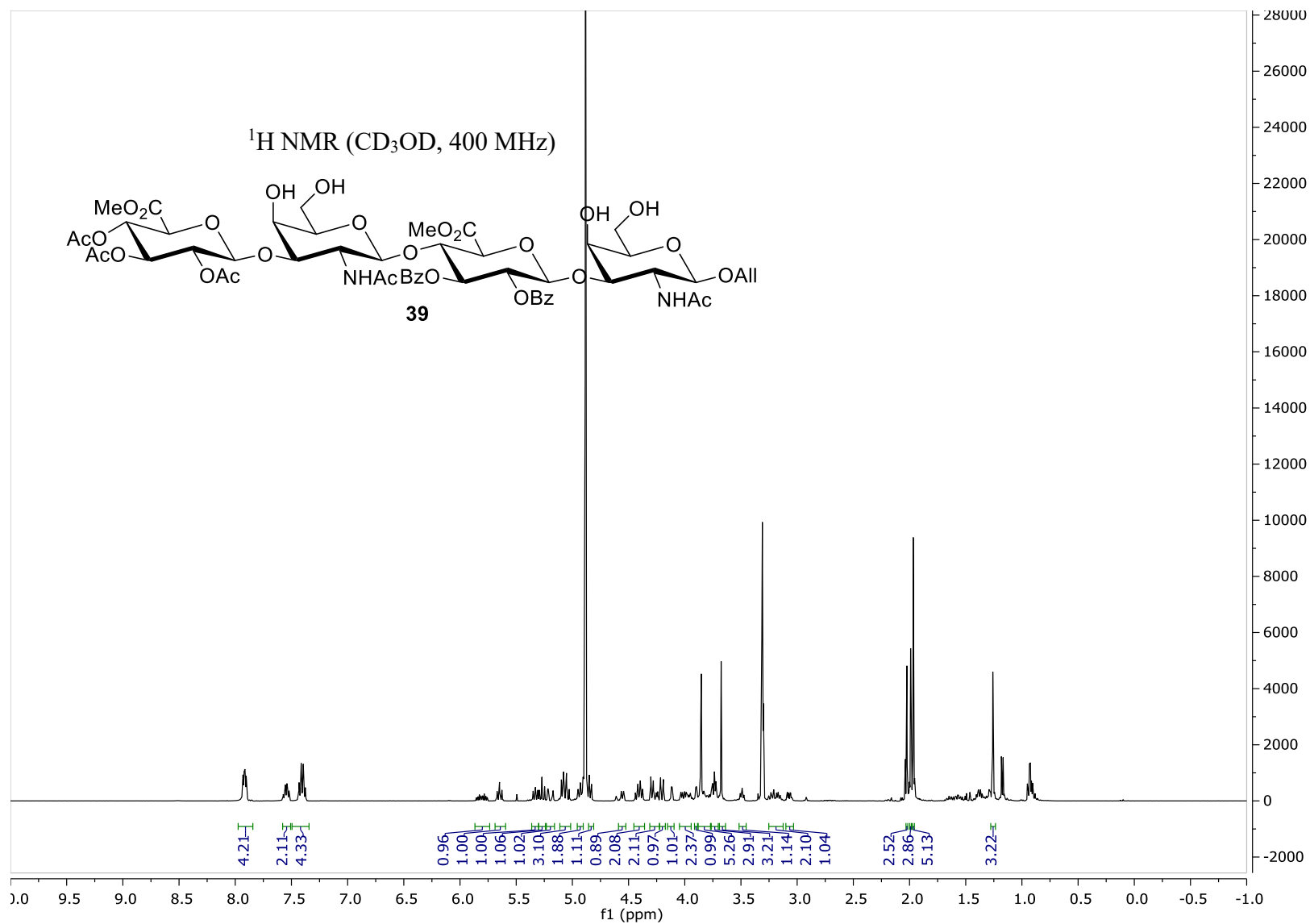


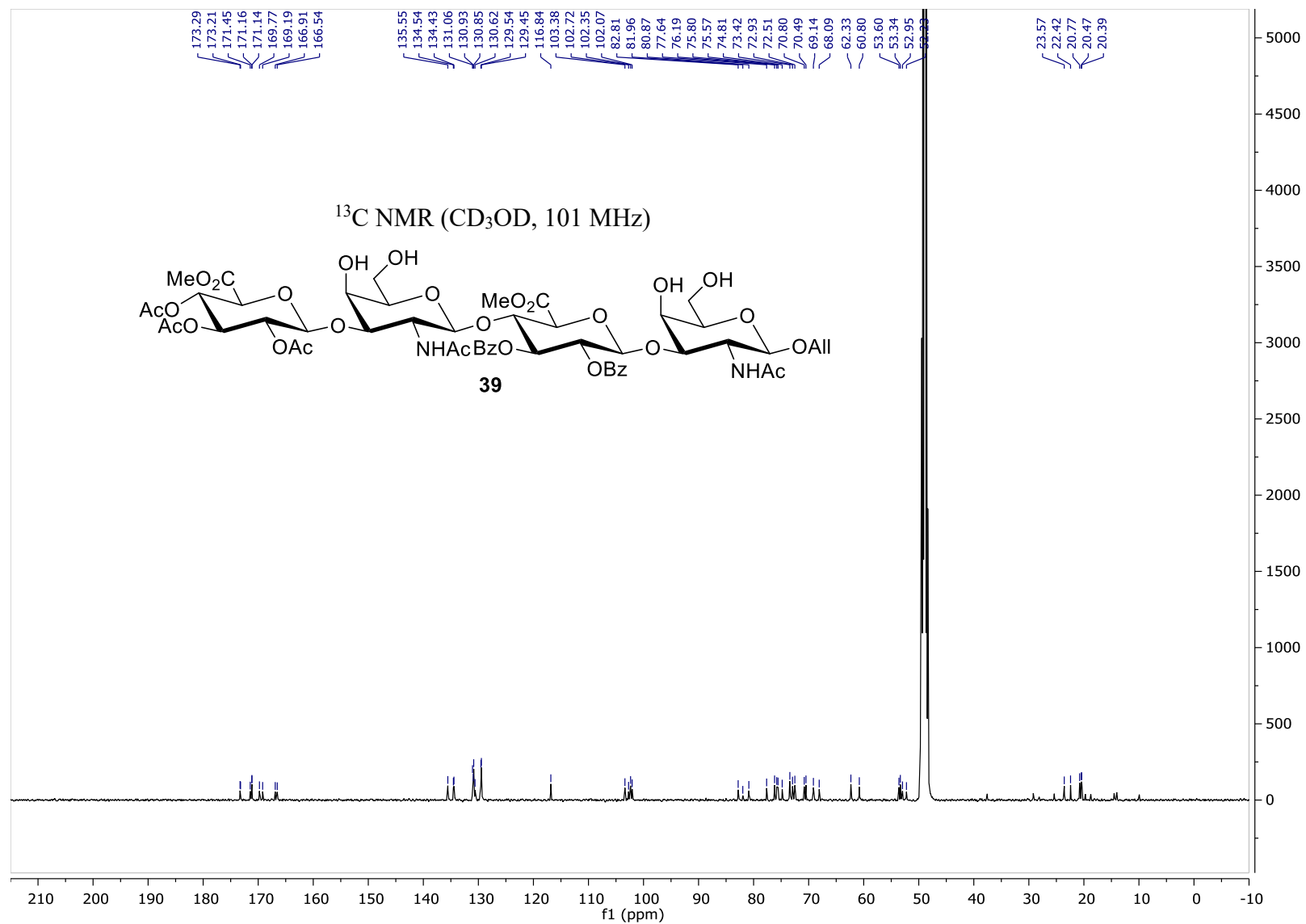


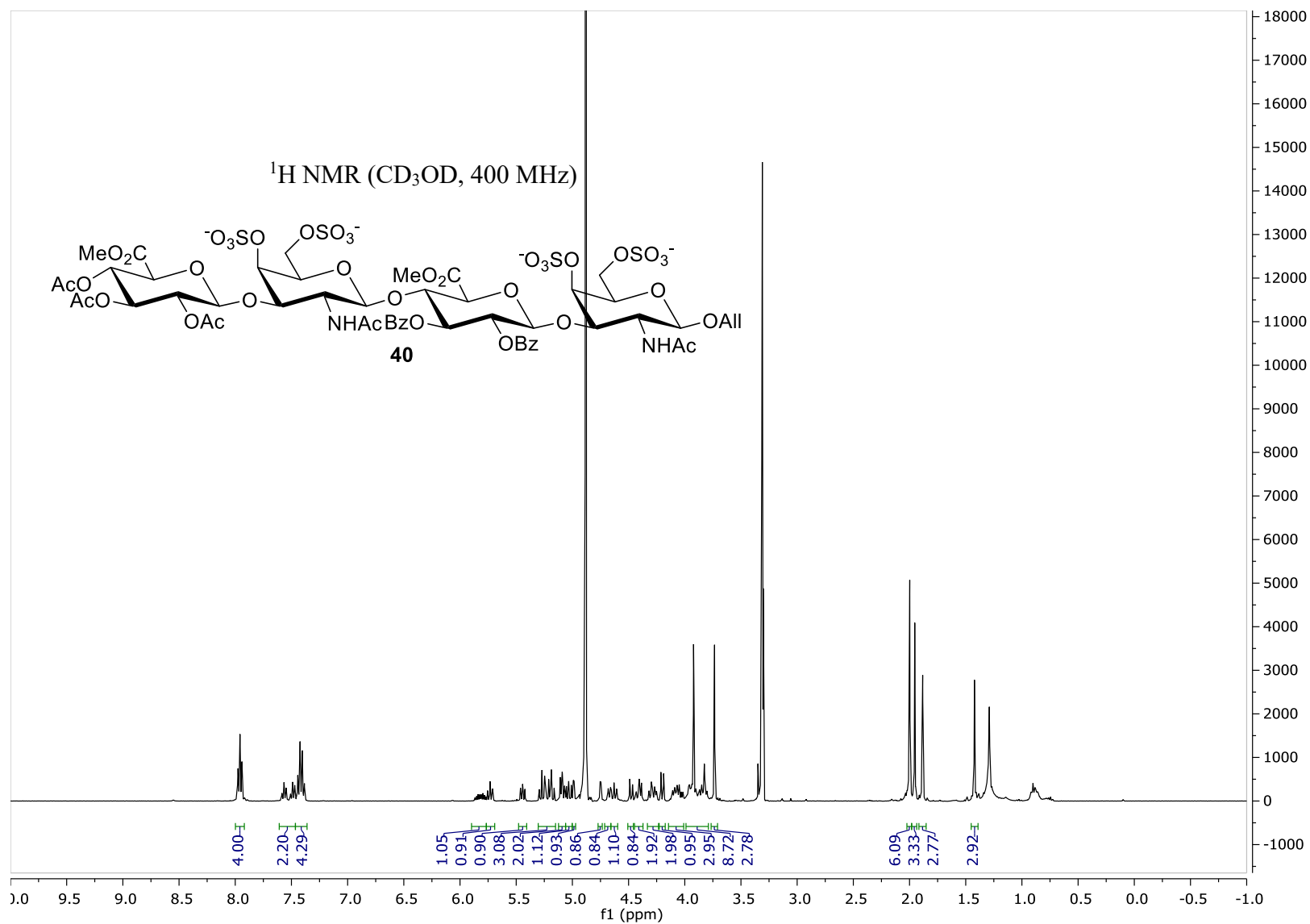


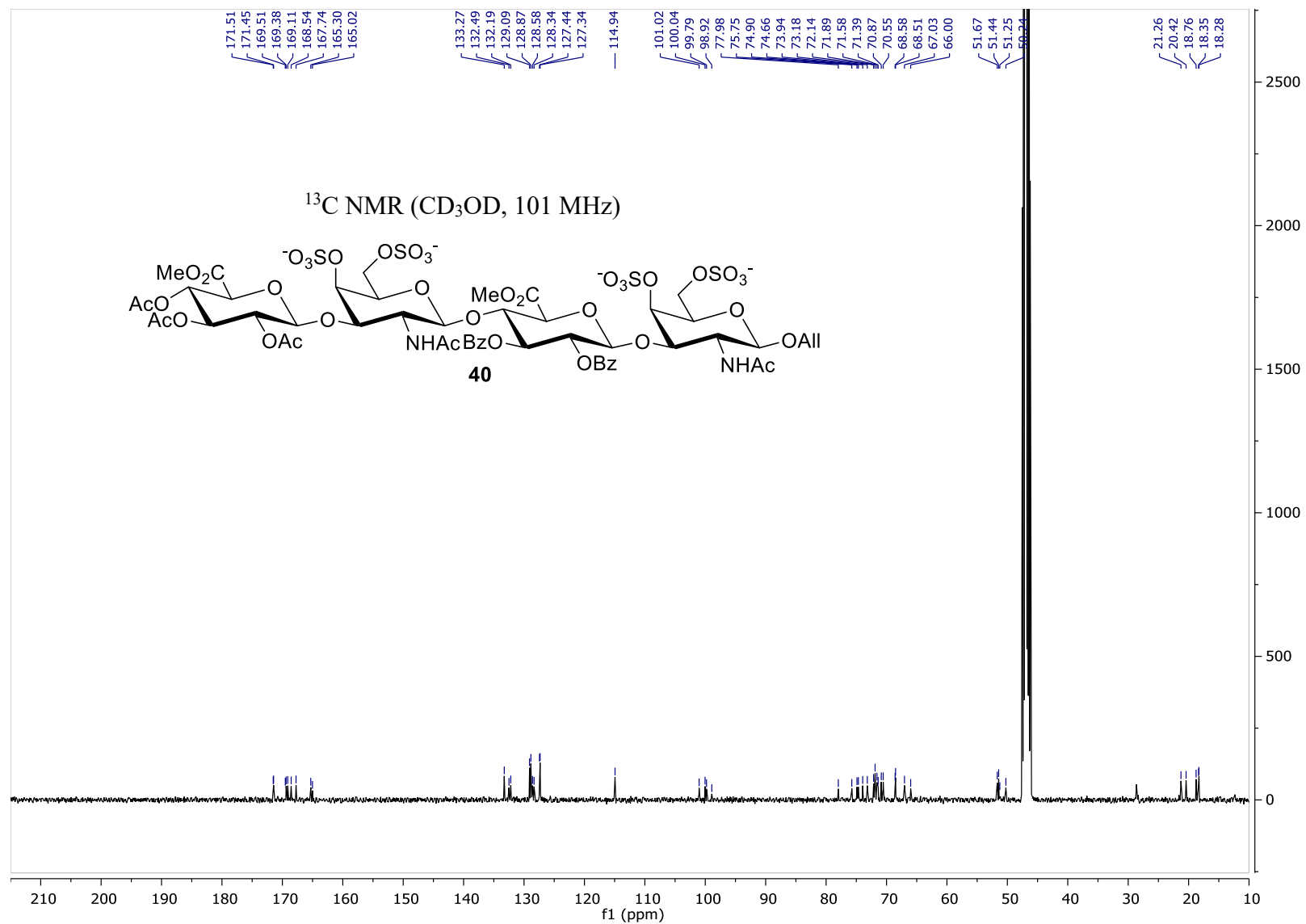


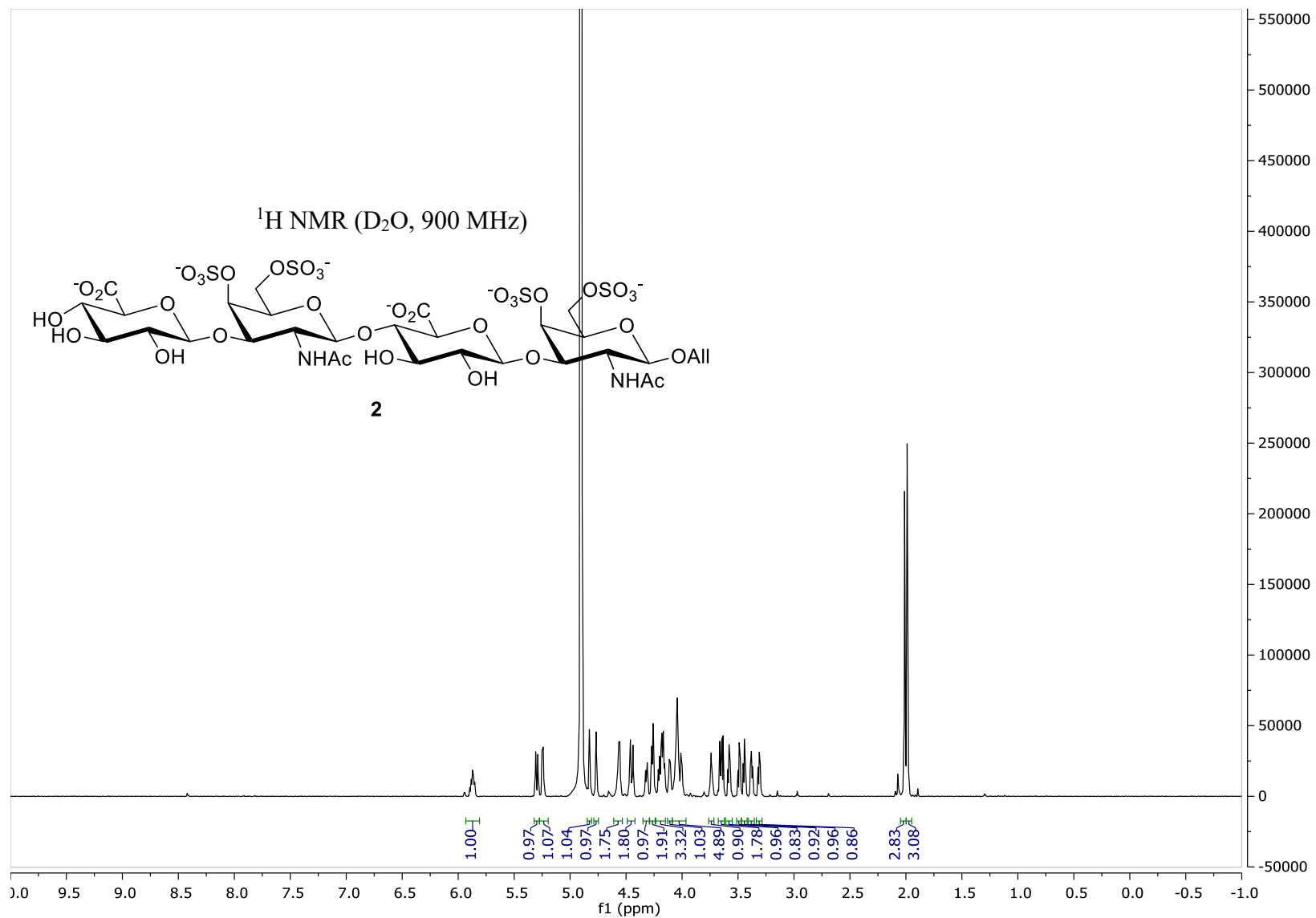


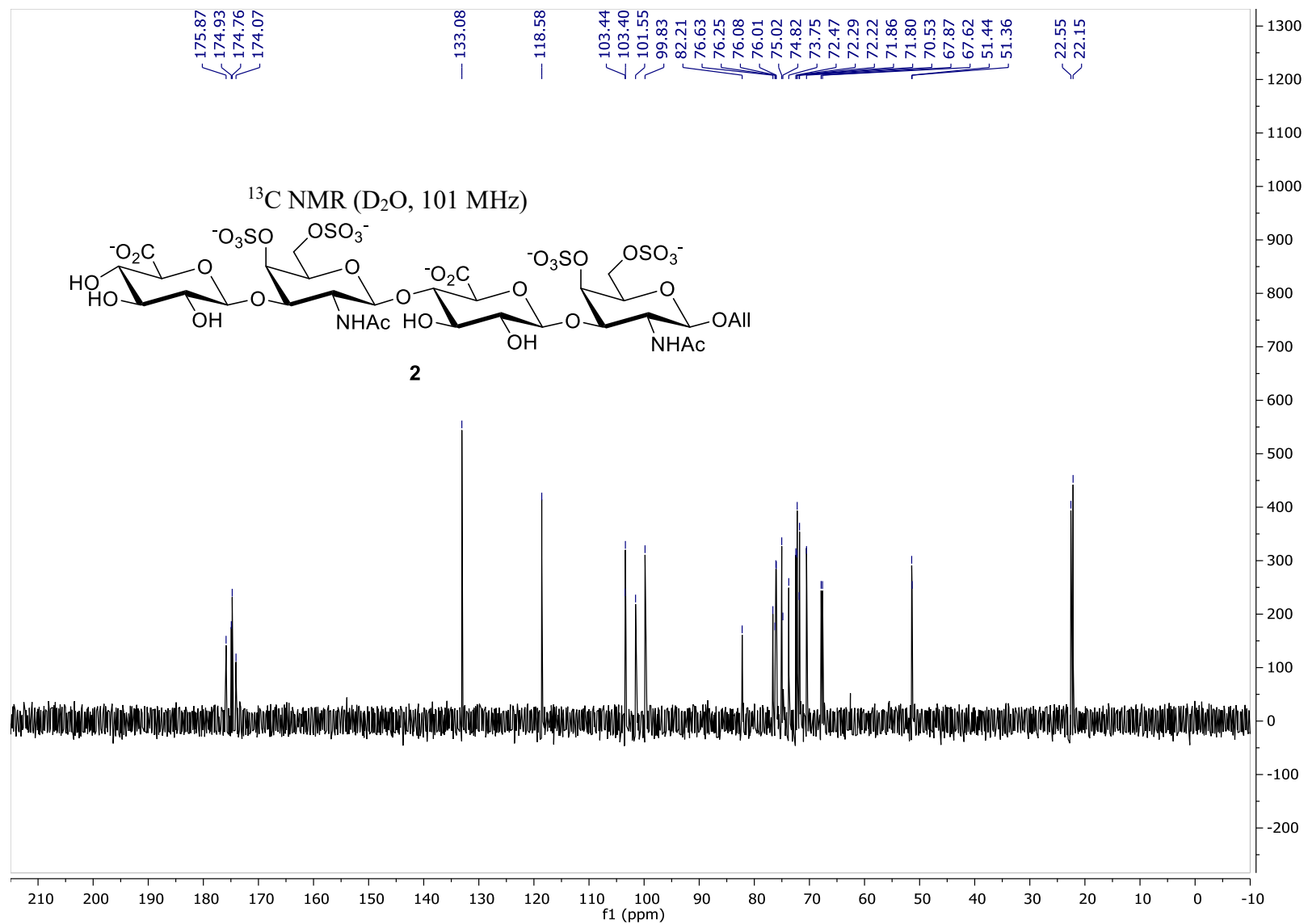












*Chapter 3*

## NMR SOLUTION STRUCTURE OF THE CS-D AND CS-E TETRASACCHARIDE

### *Chapter 3*

## NMR SOLUTION STRUCTURE OF THE CS-D AND CS-E TETRASACCHARIDE\*

### *Introduction*

To better elucidate the complex interactions between GAGs and proteins, understanding of the solution structure of GAGs is of great importance. The 3D-structural properties of GAGs can be examined by three major analytical methods: (1) molecular dynamics that utilizes a self-consistent set of physically realistic equations and semi-empirical parameters to simulate atomic and molecular motions; (2) X-ray crystallography of free or protein-bound GAG oligosaccharides; and (3) liquid or media-aligned NMR spectroscopy. While the majority of protein structures are determined by X-ray crystallography, most of the recent structural studies on GAGs are NMR based. NMR as a technique for conformational studies has several advantages: (1) samples can be studied under physiologically relevant environment with adjustable pH, salt concentration and temperature; (2) a wide range of structural properties can be obtained with different NMR experiments; and (3) samples are easy to prepare and usually recoverable. Although NMR methods are usually unfeasible for structure determination of large molecules (>60 kDa), they are suitable and have been used for GAGs with the length of 2 to 12 monosaccharide units<sup>1-3</sup>.

---

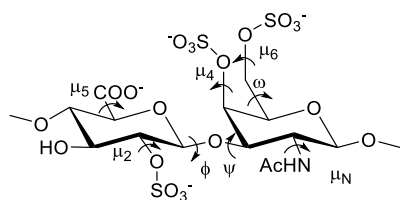
\* The NMR structure determination was done in collaboration with Dr. Andrew Almond at University of Manchester.



Typical NMR parameters utilized for structure determination include the nuclear Overhauser effect (NOE), scalar couplings ( $J$ ), residual dipolar couplings (RDC) and chemical shift anisotropy (CSA). NOE signals resulted from through-space contacts of nuclei provide inter-nuclei distance information and are the base of NMR structural determination. Inter-residual NOE signals in particular provide information regarding glycosidic linkage torsion angles. As for scalar couplings, the most commonly used type is the three-bond proton coupling. The coupling constants provide information regarding the dihedral angle between vicinal protons and therefore are useful for determining the ring pucker. RDC and CSA signals are obtained from anisotropic conditions, usually when the target molecules are partially aligned in lipid, liquid crystal or filamentous phage media. RDC and CSA provides information regarding relative orientation of the magnetic field and nucleus-nucleus vectors, and hence the relative orientation between nucleus-nucleus vectors can be obtained. Other NMR parameters such as one-bond or three bond C-H coupling can provide valuable structural information as well but are much less commonly adopted.

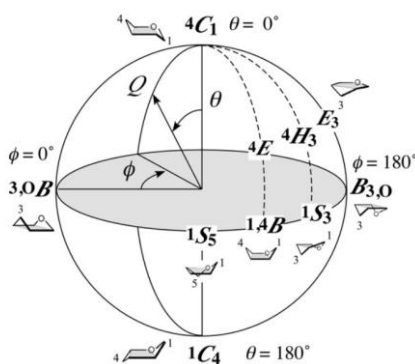
For GAGs, the important structure features include conformations around the glycosidic linkages, orientations of the exocyclic groups and ring puckering states. Conformation around a glycosidic linkages are characterized by a pair of torsion angles  $\varphi$  and  $\psi$  (figure 3.1). There are three definitions of  $\varphi$  and  $\psi$ , preferred by groups of different academic background. In this study, we will use the conventional definition unless otherwise specified. The conventional definition defines  $\varphi$  as O5-C1-O1-Cx and  $\psi$  as C1-O1-Cx-C(x+1) for a (1→x) glycosidic linkage. Orientations of the exocyclic groups are described quantitatively with torsion angles or qualitatively respect to an atom. The hydroxymethyl group in a GalNAc unit can be defined by the  $\omega$  torsion angle (O5-C5-C6-O6) and categorized

in the *gauche-gauche*, *gauche-trans*, and *trans-gauche* orientation ( $\omega$  close to  $-60^\circ$ ,  $60^\circ$  and  $180^\circ$ , respectively). Orientation of the *O*-sulfate group at position  $x$  is defined by  $\mu_x$  torsion angle ( $C(x-1)-C_x-O_x-S$ ). Since the oxygen atoms on the carboxylate groups are equivalent, we will define  $\mu_5$  as the torsion angle between the  $O6-C6-O6$  plane and the  $C4-C5-C6$  plane. Orientation of the acetamido group will be defined with respect to  $H2$  as  $\mu_N$  ( $HN-N2-C2-H2$ ).



**Figure 3.1.** Schematic representation of the torsion angles of interest.

The puckering states of a sugar ring can be described with a set of Cremer-Pople polar coordinate  $(\theta, \psi, Q)^4$ , as illustrated in Figure 3.2. For example, the  ${}^4C_1$  chair corresponds to  $\theta = 0^\circ$  and the  $B_{3,O}$  boat corresponds to  $\theta = 90^\circ$  and  $\psi = 180^\circ$ . The definitions of the 38 canonical pyranose ring puckers are listed in the appendix for Chapter 3.



**Figure 3.2.** Spherical mapping of pyranose conformations represented by the Cremer–Pople polar coordinates. The closest canonical pucker of the IUPAC name is used here to concisely specify sugar ring puckers.

Due to the repetitive nature of GAGs, resonance overlap is often serious and prevents the measurement of sufficient structurally important parameters. Direct structure determination with NMR data only is therefore usually challenging. Semi-empirical force field such as MM3<sup>5</sup>, AMBER<sup>6</sup> or GLYCAM<sup>7</sup> are often employed in the form of restrained molecular dynamics to compensate the lack of sufficient NMR data.

To the best of our knowledge, five NMR structural studies of chondroitin sulfate has been published. The four recent studies are briefly introduced below. In the work by Kamerling and coworkers<sup>8</sup>, the glycosidic linkage torsion angles of 8 disaccharide units were described by the potential energy map calculated with the MM3<sup>5</sup> force field in a grid search manner. One inter-residual NOE signal was measured for each glycosidic linkage. The NOE derived distances were consistent with the distances measured from global minimum conformation on the energy map of each disaccharide unit. The glycosidic linkage torsion angles at the global minimum were reported for each disaccharide unit. Scalar coupling constants and ring puckering were not measured or discussed.

Solera and coworkers<sup>9</sup> investigated the structure of a CS-E tetrasaccharide and a heptasulfated non-natural CS tetrasaccharide with NMR methods and molecular dynamics simulation. The AMBER<sup>10</sup> and GLYCAM<sup>11</sup> force field was used for the simulation. A total of two inter-residual NOE signals were measured for the CS-E tetrasaccharide and compared against distance measured over the trajectory of molecular dynamics simulation. Superimposition of structures along the simulation was presented, but no value of glycosidic

linkage torsion angles were reported. The experimentally derived distances were consistent with the simulated distances. Scalar coupling constants suggested  ${}^4C_1$  pucker for all the monosaccharide units, which is consistent with the molecular dynamics simulation.

Sattelle and coworkers<sup>12</sup>, studied an unsulfated chondroitin hexasaccharide with restrained molecular dynamics. Scalar coupling constants of the hexasaccharide were consistent with the  ${}^4C_1$  pucker of GalNAc and GlcA units, justifying the use of endocyclic torsional restraints to preclude the formation of non- ${}^4C_1$  pucker during the simulation. Eight inter-residual distance restraints derived from NOE signals were supplemented to the GLYCAM<sup>7</sup> force field as energy penalty functions. A molecular dynamics simulated annealing protocol was performed for 250 rounds to generate energy minimized structures. The top 10% most energetically favorable model conformers were accepted as the final structure ensemble and the ensemble was deposited in the PDB with code 2KQO, which is the only CS NMR structure available in the PDB.

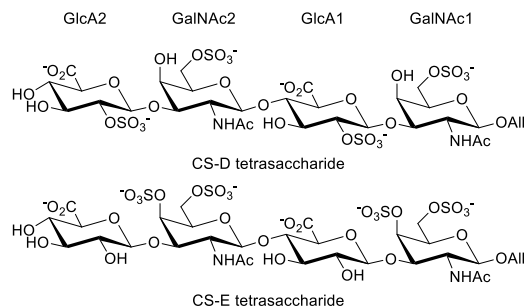
Yu and coworkers<sup>13</sup> used RDC and CSA offsets to determine the conformational preferences of a CS pentasaccharide. The inherent acetyl groups were replaced with  ${}^{13}C$  labeled acetyl groups by a series of chemical reactions to enhance sensitivity for the RDC and CSA experiments. Inter-residual NOE signals and  $J_{H2-HN}$  values were also measured to serve as distance and torsional restraints. Structure of the pentasaccharide was determined using an iterative calculation of alignment parameters from an assumed structure with the REDCAT software<sup>14</sup> and optimization of the structure using a simulated annealing protocol with the XPLOR-NIH software<sup>15-16</sup>. 10 structures with lowest energy were selected for the

final ensemble and the average glycosidic linkage torsion angles were reported. Endocyclic scalar coupling constants or the puckering states were not reported.

Here, we report the solution structure determination of the CS-D and CS-E tetrasaccharide. To the best of our knowledge, this is the first NMR structural study of a CS-D tetrasaccharide. On the other hand, our work of the CS-E tetrasaccharide is the most thorough study to date. This is also the first CS structural study that determined the glycosidic linkage torsion angles of GAGs and resolved a second conformer for CS based on NMR data and chemical information only, without molecular mechanics.

To aid the NMR structure determination, we performed free molecular dynamics of 10  $\mu$ s for the CS-D and CS-E tetrasaccharide to provide structural insights. However, the free molecular dynamics was not directly involved in our structure determination. As for the NMR experiments, we utilized the NOE and scalar coupling data to determine the glycosidic linkage torsion angles and ring puckering states. Orientation of the exocyclic moieties such as the carboxylate or hydroxymethyl groups was modeled with the GLYCAM<sup>7</sup> force field, as no definitive NMR data was available.

Structures of the CS-D and CS-E tetrasaccharides are shown in figure 3.3. From the reducing end, the monosaccharides are referred to as GalNAc1, GlcA1, GalNAc2, and GlcA2 throughout in this chapter.

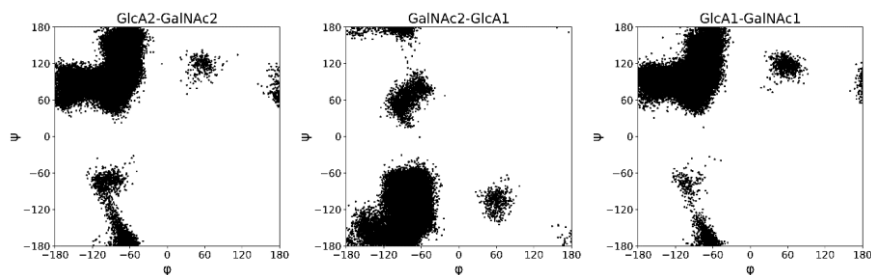


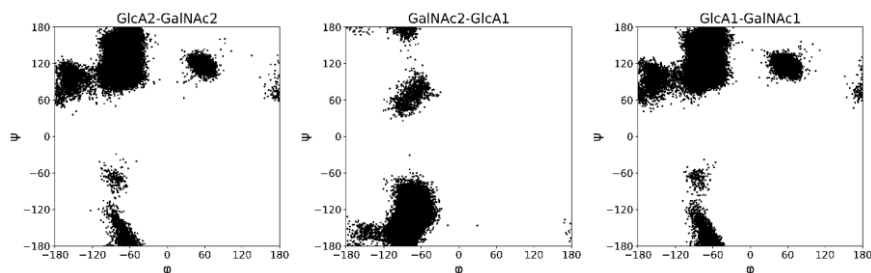
**Figure 3.3.** Structure of the CS-D and CS-E tetrasaccharides.

### Free molecular dynamics and analysis

We performed a 10  $\mu$ s molecular dynamics simulation with the GLYCAM force field<sup>7</sup> in explicit water. Molecular dynamics simulation performed at microsecond timescale is suggested to be sufficient to reproduce conformational equilibrium in monosaccharides and linear sulfated glycosaminoglycans containing up to ten pyranoses<sup>1</sup>. The glycosidic linkage torsion angles, ring puckering states, and orientation of the exocyclic groups were extracted from the trajectory. The data is briefly presented here. Further analysis will be presented and compared along with the NMR-based structures.

Distribution of glycosidic linkage torsion angles of the CS-D and CS-E tetrasaccharide are shown below as a scatter plot in figure 3.4, respectively. The average values of the angles are summarized in table 3.1 and table 3.2. These values will be further analyzed along with the NMR calculated structure.





**Figure 3.4.** Distribution of glycosidic torsion angles of the CS-D (top) and CS-E (bottom) tetrasaccharide.

	$\beta(1\rightarrow3)$	$\beta(1\rightarrow4)$	$\beta(1\rightarrow3)$
Torsion	GlcA2-GalNAc2	GalNAc2-GlcA1	GlcA1-GalNAc1
$\varphi$	-90	-80	-81
$\psi$	99	-122	113

**Table 3.1.** Average glycosidic linkage torsion angles of the CS-D tetrasaccharide.

	$\beta(1\rightarrow3)$	$\beta(1\rightarrow4)$	$\beta(1\rightarrow3)$
Torsion	GlcA2-GalNAc2	GalNAc2-GlcA1	GlcA1-GalNAc1
$\varphi$	-71	-77	-72
$\psi$	133	-132	130

**Table 3.2.** Average glycosidic linkage torsion angles of CS-E tetrasaccharide.

The ring puckering state was analyzed with the Cremer-Pople parameters and summarized as shown in table 3.3 and 3.4. For the CS-D tetrasaccharide, the simulation suggests that both GalNAc units have a dominant (>99.9%)  ${}^4C_1$  pucker. GlcA2 is mostly  ${}^4C_1$  but not as rigid as the GalNAc units. On the other hand, GlcA1 is predicted to be very flexible in that it does not have a dominant pucker. On the contrary, all monosaccharide units in the CS-E tetrasaccharide units have a dominant (>99%)  ${}^4C_1$  pucker.

GalNAc1		GlcA1		GalNAc2		GlcA2	
${}^4C_1$	99.94%	${}^1C_4$	30.81%	${}^4C_1$	99.96%	${}^4C_1$	87.46%
${}^1S_3$	0.03%	${}^4C_1$	27.04%	${}^1S_3$	0.01%	${}^1S_3$	5.54%
${}^4E$	0.01%	${}^1S_3$	20.66%	${}^4H_5$	0.01%	${}^1C_4$	2.21%
${}^4H_3$	0.01%	$B_{30}$	13.86%	${}^4E$	0.01%	${}^{14}B$	2.12%
${}^4H_5$	0.01%	${}^{14}B$	3.80%	${}^{14}B$	0.01%	$B_{30}$	1.48%

**Table 3.3.** Population of the 5 most-populated puckers of the CS-D tetrasaccharide from molecular dynamics simulation.

GalNAc1		GlcA1		GalNAc2		GlcA2	
<sup>4</sup> C <sub>1</sub>	99.69%	<sup>4</sup> C <sub>1</sub>	99.29%	<sup>4</sup> C <sub>1</sub>	99.83%	<sup>4</sup> C <sub>1</sub>	99.66%
<sup>1</sup> S <sub>3</sub>	0.204%	<sup>1</sup> C <sub>4</sub>	0.21%	<sup>1</sup> S <sub>3</sub>	0.06%	<sup>1</sup> S <sub>3</sub>	0.09%
<sup>4</sup> H <sub>3</sub>	0.03%	<sup>1</sup> S <sub>3</sub>	0.12%	<sup>4</sup> H <sub>5</sub>	0.03%	<sup>1</sup> C <sub>4</sub>	0.07%
<sup>4</sup> E	0.03%	<sup>B</sup> <sub>30</sub>	0.10%	<sup>4</sup> E	0.03%	<sup>1</sup> B	0.06%
<sup>4</sup> H <sub>5</sub>	0.01%	<sup>0</sup> S <sub>2</sub>	0.06%	<sup>1</sup> B	0.02%	<sup>B</sup> <sub>30</sub>	0.02%

**Table 3.4.** Population of the 5 most-populated puckers of the CS-E tetrasaccharide from molecular dynamics simulation.

As for orientations of the exocyclic groups, the 4-*O*-sulfate group in CS-E and the carboxylate groups in both CS-D and CS-E are rather rigid and adopt an eclipsed orientation with regard to H4 and H5 ( $\mu_4 \approx -120^\circ$ ;  $\mu_5 \approx 60^\circ$ ) respectively. The 2-*O*-sulfate groups of CS-D favors an eclipsed orientation ( $\mu_2 \approx 120^\circ$ ) over an anti orientation ( $\mu_2 \approx -60^\circ$ ) with respect to H2 with the eclipsed/anti ratio  $\approx 5$ . The acetamido groups in both tetrasaccharides favor the trans orientation with respect to H2 ( $\mu_N \approx 180^\circ$ ). The hydroxymethyl groups in both tetrasaccharides are flexible and adopt both *trans-gauche* and *gauche-trans* orientation ( $\omega \approx 180^\circ$  or  $60^\circ$ ). While the hydroxymethyl groups in the CS-D tetrasaccharide have a *trans-gauche/gauche-trans* ratio about 2.5, the *trans-gauche* orientation is less preferred (*trans-gauche/gauche-trans* ratio = 1.25) in CS-E due to the repulsive interaction with the 4-*O*-sulfate groups. The 6-*O*-sulfate groups in both the CS-D and CS-E tetrasaccharide virtually rotate freely, exploring all the conformational space except the eclipsed orientation with respect to C5 ( $\mu_6 = 0^\circ$ ).

### NMR experiments and data analysis

With the synthesized CS-D and CS-E tetrasaccharides in hand, we decide to determine the structure with liquid NMR methods. High quality NMR spectra were recorded



with a 900 MHz Bruker instrument equipped with a cryoprobe at University of California, Berkeley. 1D and 2D spectra include proton, double quantum filtered homonuclear correlation spectroscopy (DQF-COSY),  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear single quantum coherence (HSQC), total correlation spectroscopy (TOCSY), and NOESY were recorded in  $\text{D}_2\text{O}$  and/or 10%  $\text{D}_2\text{O}$  in  $\text{H}_2\text{O}$  solvent. Three-bond homonuclear coupling constants and NOE intensities were extracted from the  $^1\text{H}$ , DQF-COSY and NOESY spectra for structure determination. The rotating-frame Overhauser spectroscopy (ROESY) was also recorded for CS-D, because the ROESY experiment produced better spectra than the NOESY experiment when recorded with a local 600 MHz instrument. We suspected that the molecular weight of the tetrasaccharide may fall in the unfavorable range of NOESY with the 900 MHz instrument, too. However, the increased magnetic field moved the NOE unfavorable range down and away from the molecular weight of the CS-D tetrasaccharide. The NOESY spectra turned out to be better than the ROESY spectra, so the data measured from ROESY was not used for the CS-D tetrasaccharide.

The NOE intensities were obtained by measuring the integral of a given proton-proton cross-peak. All possible cross-peak positions were measured. If integral of a position where a correlation is possible was found to be below the noise level, a “no-NOE” was assigned. While a NOE signal provides more structural information by designating a certain distance between the source proton pair, a no-NOE signal can indicate that the proton pair is not close enough to have an identifiable peak. After careful examination of the spectra, we observed 16 and 16 coupling constants, 11 and 29 inter-residual NOE signals, 53 and 36 inter-residual “no-NOE” signals, and 16 intra-residual NOE signals for CS-D and CS-E

respectively. The measured coupling constants summarized in the table 3.5 and 3.6. NOE data are shown in the appendix.

	GalNAc1	GlcA1	GalNAc2	GlcA2
$^3J_{H1,H2}$	8.4	7.3	8.0	7.6
$^3J_{H2,H3}$	13.1 <sup>a</sup>	8.9	13.1 <sup>a</sup>	9.0
$^3J_{H3,H4}$	4.4 <sup>a</sup>	8.7	4.4 <sup>a</sup>	9.6
$^3J_{H4,H5}$	n/a <sup>b</sup>	9.6	n/a <sup>b</sup>	9.6
$^3J_{HN,H2}$	9.2		9.5	

**Table 3.5.** Experimental scalar coupling constants of CS-D. All couplings are reported in units of hertz. <sup>a</sup> Coupling constants measured from DQF-COSY spectrum. <sup>b</sup> Not available due to resonance overlap.

	GalNAc1	GlcA1	GalNAc2	GlcA2
$^3J_{H1,H2}$	8.5	7.9	7.3	7.8
$^3J_{H2,H3}$	11.0 <sup>a</sup>	9.4	12.9 <sup>a</sup>	9.4
$^3J_{H3,H4}$	3.0	8.7	4.3 <sup>a</sup>	9.1
$^3J_{H4,H5}$	n/a <sup>b</sup>	9.7	n/a <sup>b</sup>	9.8
$^3J_{HN,H2}$	10.3		10.2	

**Table 3.6.** Experimental scalar coupling constants of CS-E. All couplings are reported in units of hertz. <sup>a</sup> Coupling constants measured from DQF-COSY spectrum. <sup>b</sup> Not available due to resonance overlap.

### Pucker analysis

Since the coupling constants provides information about the torsional angle between vicinal protons, they are often employed to determine the puckering state of the monosaccharide units. Fortunately, every coupling constant on every GlcA unit of both tetrasaccharides are available, providing unambiguous evidence for pucker determination. The large coupling constants of every vicinal proton pairs in every GlcA unit suggest that the GlcA units in both tetrasaccharides are in the  $^4C_1$  form. The large NOE signals of the diaxial proton pairs H1-H3, H1-H5, H3-H5, and H2-H4 further support the  $^4C_1$  assignment.

On the other hand, while  $^3J_{H4,H5}$  are not available on the GalNAc units due to peak overlapping, the large coupling constants between the axial protons and the characteristic

small coupling constants of the axial-equatorial pair H3-H4 all indicate the  ${}^4C_1$  form. The intra-residual NOE data also points toward the  ${}^4C_1$  pucker, such as the large NOE signal of the diaxial proton pairs H1-H3, H1-H5 and H3-H5, and the small NOE signal of the axial-equatorial proton pair H1-H4. The GalNAc units in dermatan sulfate has also been reported<sup>17</sup> to be exclusively in the  ${}^4C_1$  form based on coupling constants and NOE data.

The large  ${}^3J_{HN,H2}$  in all GalNAc units suggest the  $\mu_N$  torsion angle to be close to either  $0^\circ$  or  $180^\circ$ . The large intra-residual NOE signal of the HN-H1 and HN-H3 pair further indicate the acetamido group to be *trans* oriented with respect to H2.

The puckering states of each monosaccharide unit and  $\mu_N$  of CS-E determined by NMR is consistent with the free molecular dynamics simulation. However, while puckering states of the GalNAc units and  $\mu_N$  of CS-D as determined by NMR and simulation are consistent, the simulation and NMR data suggests different puckering states for the GlcA units. Although the experimental values clearly indicate the  ${}^4C_1$  pucker, in principle it is still possible that coupling constants of a single pucker to coincidentally be similar to coupling constants averaged from a combination of puckers. Therefore, quantitative comparison necessary to reject the inconsistent prediction.

To directly and quantitatively compare the puckering states determined by experimental values and predicted by simulation, we calculated the average coupling constants of each monosaccharide over the trajectory with the Haasnoot-de Leeuw-Altona equation<sup>18</sup>. The Haasnoot-de Leeuw-Altona equation considers effect of electronegativity of atoms bonded to the proton-bearing carbon, and hence is more suitable than the original Karplus equation<sup>19</sup> for determining coupling constants in carbohydrates.

Since molecular dynamics simulation predicted virtually exclusive  ${}^4C_1$  pucker for all the monosaccharide units in the CS-E tetrasaccharide and both GalNAc units in CS-D, simulated coupling constants were fully consistent with the experimental coupling constant as expected (Table 3.7 and Table 3.8). The slightly larger deviation of the coupling constants measured from the DQF-COSY experiment should not be considered as involvement of non- ${}^4C_1$  pucker. Instead, it has been suggested that larger constants measured in DQF-COSY experiments are normal when the linewidths are somewhat broad due to the cancellation of overlapping of the antiphase peaks.

For CS-D, the large deviation between the simulated and experimental values of the GlcA units in CS-D demonstrated that the puckers observed from molecular dynamics trajectory did not produce similar experimental value as the  ${}^4C_1$  pucker (Table 3.7), suggesting that the prediction was off. To further determine whether the most abundant non- ${}^4C_1$  puckers predicted by simulation is present but in smaller quantities, we extracted the three most abundant puckers from the trajectory and calculated the coupling constants of each pucker (Table 3.9). Then we performed a linear least-square fitting of the experimental values with the three sets of coupling constants with MATLAB R2016a Optimization App. The experimental values were still best fitted by the  ${}^4C_1$  pucker without mixing with the  ${}^1C_4$  and  ${}^1S_4$  pucker. Multiple starting points were used to ensure that the optimization wasn't trapped in the local minimum. Therefore, by demonstrating that the experimental value could not be better fitted with the introduction of non- ${}^4C_1$  puckers, the GlcA units of CS-D is exclusively  ${}^4C_1$  like other monosaccharide units in CS-D and CS-E.

	GalNAc1		GlcA1		GalNAc2		GlcA2	
	Exp.	Sim.	Exp.	Sim.	Exp.	Sim.	Exp.	Sim.
$^3J_{H1,H2}$	8.4	8.3	7.5	3.2	8.0	8.3	7.6	6.6
$^3J_{H2,H3}$	13.1 <sup>a</sup>	10.5	8.9	5.8	13.1 <sup>a</sup>	10.5	9.0	7.9
$^3J_{H3,H4}$	4.4 <sup>a</sup>	2.6	8.7	6.6	4.4 <sup>a</sup>	2.6	8.8	8.0
$^3J_{H4,H5}$	n/a <sup>b</sup>	0.9	9.6	6.1	n/a <sup>b</sup>	0.9	10.5	9.9

**Table 3.7.** Comparison of coupling constants of the CS-D tetrasaccharide from experiments and simulation. All couplings are reported in units of hertz. <sup>a</sup>Coupling constants measured from DQF-COSY spectrum. <sup>b</sup>Not available due to resonance overlap.

	GalNAc1		GlcA1		GalNAc2		GlcA2	
	Exp.	Sim.	Exp.	Sim.	Exp.	Sim.	Exp.	Sim.
$^3J_{H1,H2}$	8.5	8.2	7.9	7.4	7.3	8.3	7.8	7.4
$^3J_{H2,H3}$	11.0 <sup>a</sup>	10.5	9.4	8.5	12.9 <sup>a</sup>	10.6	9.4	8.5
$^3J_{H3,H4}$	3.0	2.4	8.7	8.4	4.3 <sup>a</sup>	2.4	9.1	8.2
$^3J_{H4,H5}$	n/a <sup>b</sup>	0.8	9.7	9.9	n/a <sup>b</sup>	0.9	9.8	9.8

**Table 3.8.** Comparison of coupling constants of the CS-E tetrasaccharide from experiments and simulation. All couplings are reported in units of hertz. <sup>a</sup>Coupling constants measured from DQF-COSY spectrum. <sup>b</sup>Not available due to resonance overlap.

	GlcA1				GlcA2			
	Exp.	$^1C_4$	$^4C_1$	$^1S_3$	Exp.	$^4C_1$	$^1S_3$	$^1C_4$
$^3J_{H1,H2}$	7.5	1.5	7.4	1.3	7.6	7.3	1.3	1.5
$^3J_{H2,H3}$	8.9	1.9	8.5	6.5	9.0	8.4	6.2	2.0
$^3J_{H3,H4}$	8.7	2.5	8.4	9.3	8.8	8.2	9.2	2.7
$^3J_{H4,H5}$	9.6	1.0	9.9	8.1	10.5	9.8	8.2	1.1

**Table 3.9.** Coupling constants of the three most abundant puckers predicted in the molecular dynamics simulation. All couplings are reported in units of hertz.

Aside from our results, available X-ray structures<sup>20-22</sup> of free or protein-bound CS also agree with the above conclusions that every monosaccharide unit in CS is in the  $^4C_1$  form and that the acetamido groups adopt a *trans* orientation. Similar to our molecular dynamics simulation, previous simulation studies<sup>23-24</sup> also predicted dominant  $^4C_1$  form in the GalNAc units, the *trans* orientation of the acetamido group, and the presence of non- $^4C_1$  puckers of GlcA units.

### Structure determination of CS-E

Based on the data available, we decided to determine the solution structure of the two tetrasaccharides in two stages: (1) Determine the backbone conformation (i.e., torsion angles of the glycosidic linkages) with NOE data. (2) Model the exocyclic groups with the GLYCAM<sup>7</sup> force field. We need to utilize the GLYCAM force field to model the exocyclic groups such as the  $\text{-COO}^-$  and  $\text{-CH}_2\text{OSO}_3^-$  groups since we don't have experimental data regarding orientation of the exocyclic groups. As for the backbone, we calculated the structure without involving molecular mechanics. Although the force fields are constantly improving, they are still far from perfect and do not always give good prediction on every aspect. Therefore, we aim to determine the structure by experimental data as much as possible and model the rest with a force field.

To determine the backbone structure with the NMR data, we decided to employ a distance geometry approach without molecular mechanics. Each monosaccharide unit was treated as a rigid body and fixed in the  $^4\text{C}_1$  pucker based on the pucker analysis. Only the  $\varphi$  and  $\psi$  glycosidic linkage torsion angles were allowed to rotate. The acetamido groups were fixed at the trans orientation ( $\mu_{\text{N}} = 180^\circ$ ) in order to apply the inter-residual NOE signals from amide protons (HN). We chose not to model the monosaccharide rings and the acetamido groups as flexible units because only a moderate amount of inter-residual signals was available. If we allowed the rings and acetamido groups to be flexible, the increased degrees of freedom can very possibly lead to poorly-defined structures.

Distance geometry was implemented as a score function. The distance between an inter-residual proton pair  $m$  ( $r_{\text{NOE},m}$ ) and smallest possible distance between inter-residual

pair  $n$  ( $r_{no\_NOE,n}$ ) was derived from the NOE and “no-NOE” data. The square of difference between the distances measured from calculated structure ( $r_{calc,m}$  and  $r_{calc,n}$ ) and experimental distances ( $r_{NOE,m}$  and  $r_{no\_NOE,n}$ ) were then calculated to give the score ( $S$ ) (equation 3.1). The weighing factor  $k_{m,NOE}$  is assigned for each cross-peak based on quality of the NOE signal. Note that for the “no-NOE” signals, the square of difference was only calculated when the  $r_{calc,n}$  was smaller than  $r_{no\_NOE,n}$  (equation 3.3). A uniform weighing factor  $k_{noise}$  is used for all no-NOE signals. Basically, the score function is composed of a series of harmonic potential wells that constrain interproton distances in the calculated structure ( $r_{calc}$ ) to the experimental values.

$$S = \sum S_{NOE,m} + \sum S_{noNOE,n} \quad (\text{equation 3.1})$$

$$S_{NOE,m} = k_{m,NOE} (r_{m,NOE} - r_{m,calc.})^2 \quad (\text{equation 3.2})$$

$$S_{no\_NOE,n} = \begin{cases} k_{noise} (r_{n,noise} - r_{n,calc.})^2 & \text{if } r_{n,noise} > r_{n,calc.} \\ 0 & \text{if } r_{n,noise} < r_{n,calc.} \end{cases} \quad (\text{equation 3.3})$$

Since the NOE spectrum is measured for both the D<sub>2</sub>O and 10% D<sub>2</sub>O in 90% H<sub>2</sub>O sample, both data sets are incorporated in the score function. Generally speaking, the D<sub>2</sub>O spectrum has a better quality than the 10% D<sub>2</sub>O in 90% H<sub>2</sub>O spectrum, so the  $k_{m,NOE}$  is typically larger for the NOE signals measured in the D<sub>2</sub>O spectrum. Minimization of the score function was then performed by exploring the conformational space composed of the six torsion angles by simulated annealing with the SimTK molecular modeling API (molmodel).

It should be noted that NOE data assignment and structure calculation is an iterative process. In the beginning only high-quality and unambiguous NOE cross-peaks were used

for calculation. The calculated structure was then used to help the assignment of ambiguous cross-peaks. For ambiguous cross-peaks, the assignments were first made by assumptions and then validated (or rejected) after structure calculation. In the end, the calculated ensemble of structures must meet the following criteria to be accepted as ideal final structures. First of all, the ensemble of structures must satisfy all the experimentally derived restraints. Second, the structures must be physically allowed. Any two atoms should not be clashing into each other in space. The van der Waals radii will be used to evaluate the steric conflict. Finally, the ensemble of structures should be converged. If an ensemble of converged structures is not obtained, more structural restraints or re-assignment of NMR data will be required.

For the CS-E tetrasaccharide, the data assignment-structure calculation iterations were performed for about twenty rounds before the simulated annealing generates a converged ensemble of structures. The distance restraints were nicely represented in the ensemble except for a few larger deviations of the long-distance restraints. Interatomic distances were then compared with the van der Waals radii and no gross violation was found. Therefore, we are satisfied with this ensemble calculated from NMR restraints.

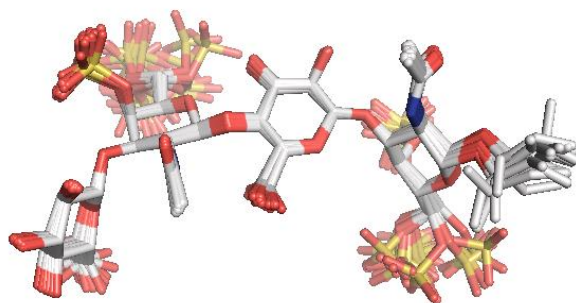
Exocyclic groups were then modeled with the GLYCAM force field with the backbone structure restrained at all times. 25 structures with the best scores were selected as the final ensemble of structures, shown in figure 3.5. Average glycosidic torsions of the ensemble is shown in table 3.10. Molecular mechanics modeling of the exocyclic groups suggested that the hydroxymethyl group of GalNAc units is flexible and adopts both the *gauche-trans* and *trans-gauche* orientation. The 6-*O*-sulfate group of GalNAc units is also flexible, as it either presents an *anti* orientation or an eclipsed orientation with respect to C6 or H6, respectively. On the other hand, the 4-*O*-sulfate group of GalNAc units is rigid and



adopts an eclipsed orientation with respect to H4. The oxygens of carboxylate group in the GlcA units are also rigid and found to be in the eclipsed and *anti* orientation with respect to H5.

Torsion	$\beta(1\rightarrow3)$ GlcA2-GalNAc2	$\beta(1\rightarrow4)$ GalNAc2-GlcA1	$\beta(1\rightarrow3)$ GlcA1-GalNAc1
$\varphi$	-71 (2)	-74 (2)	-60 (1)
$\psi$	147 (2)	-120 (3)	132 (2)

**Table 3.10.** Average glycosidic torsions in the ensemble of the top 25 CS-E conformers. Values are reported in units of degrees. Standard deviations are given in parentheses.



**Figure 3.5.** Ensemble of the top 25 structure of the CS-E tetrasaccharide.

### Structure determination of CS-D

We utilized the same distance geometry/simulated annealing approach to determine the glycosidic linkage angles of CS-D. The data assignment/structure calculation iterations were performed for about thirty rounds before the simulated annealing generates a convergent ensemble of structures. The distance restraints for the GlcA2-GalNAc2 and GlcA1-GalNAc1 linkages were nicely represented in the ensemble, however, a few restraints were violated in the GalNAc2-GlcA1 linkage. Interatomic distances were compared to the van der Waals radii and no gross violation was found between the GlcA2-GalNAc2 and

GlcA1-GalNAc1 linkages. However, the C3 and O3 atom of GlcA1 was found to be very close to the C1 atom of GalNAc2, with the distance of 2.6 and 2.4 Å, respectively (figure 3.6). The van der Waals radius of carbon and oxygen atom is 1.70 Å and 1.52 Å, respectively. This ensemble of structure was therefore not accepted due to the unsatisfied experimental restraints and violated van der Waals radii of the GalNAc2-GlcA1 linkage.

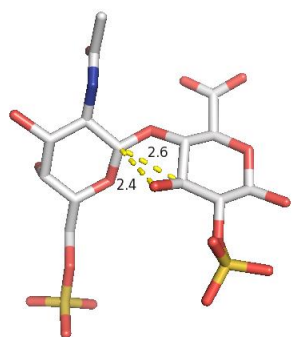


Figure 3.6. The GalNAc2-GlcA1 (1→4) linkage of the calculated structure of CS-D. The C1 of GalNAc2 is in close proximity to the C3 and O3 of GlcA1, causing steric clashes that make this structure physically unacceptable.

In order to resolve this problem, we looked into the free molecular dynamics we performed for insights. As seen in figure 3.3, a minor population around (-80,60) was clearly observed aside from the major population around (-80,-120). The two conformations are referred as the I (invert) and N (normal) conformation in this study, respectively, since the minor conformation is nearly 180° opposite from N in the  $\psi$  angle. Based on this observation in molecular dynamics, we hypothesized that an equilibrium of two conformations does exist in reality and caused a significant effect in the measured NOE signal, resulting in the unacceptable structures.

Since the intensity of a NOE signal is proportional to  $r^{-6}$ , the NOE signal will reflect the population of smaller distance more strongly than that of larger distance even with a much smaller population. For example, for a given proton pair with the distance of 2 Å in 10% of population and 4 Å in the other 90%. The NOE derived distance will not be close to 4 Å, but will be  $(2^{-6} \times 0.1 + 4^{-6} \times 0.9)^{-1/6} = 2.87$  Å. Therefore, a small population of shorter distance is possible to contribute significantly to the measured NOE signal and make the experimentally derived distance unrealistic, if the difference of distance in the two populations is large.

We then aim to reconstruct the original contribution of each conformation from the averaged experimental data and then back-calculate the NOE signals that would be observed from a sample composed of exclusive N or exclusive I conformation. The first step is to obtain the population of the two conformations, which can be calculated using two suitable NOE cross-peaks. Proton pairs of the cross-peaks used in the calculation must have interproton distances that are significantly different in the N and I conformation, otherwise the two conformations will contribute similarly to the signal, providing no information regarding the relative population. The cross-peaks also need to be strong and of high quality (i.e., have a clear peak shape and not overlapping with other peaks) to avoid measurement error. Cross-peaks of the two proton pairs GlcA1H3-GalNAc2H1 and GlcA1H4-GalNAc2H1 fit the above criteria and were used to calculate the population of N and I.

To utilize the two signals to calculate population of the N conformation ( $P_N$ ), we first assumed the N and I conformations have torsion angles  $(\varphi_N, \psi_N)$  and  $(\varphi_I, \psi_I)$  of (80, -120) and (80, 60), as observed in molecular dynamics. The distances between the proton pairs GlcA1H3-GalNAc2H1 ( $r_1$ ) and GlcA1H4-GalNAc2H1 ( $r_2$ ) were then measured.

Again, the distances ( $r_1$ ,  $r_2$ ) will depend on the N and I conformation chosen. The ratio of NOE intensity of GlcA1H3-GalNAc2H1 ( $I_1$ ) and GlcA1H4-GalNAc2H1 ( $I_2$ ) was the used to calculate  $P_N$  as shown in equation 3.4.

$$\frac{I_1}{I_2} = \frac{P_N r_{N,1}^{-6} + (1 - P_N) r_{I,1}^{-6}}{P_N r_{N,2}^{-6} + (1 - P_N) r_{I,2}^{-6}} \text{ (equation 3.4)}$$

$$I_{N,x} = \frac{I_x r_{N,x}^{-6}}{(r_{N,x}^{-6} + r_{I,x}^{-6}) P_N} \text{ (equation 3.5)}$$

With the population calculated, the signal intensity of proton pair x that would be measured from the N conformation ( $I_{N,x}$ ) without the presence of I conformation (or vice versa) can then be back-calculated from experimentally measured intensity ( $I_x$ ), by equation 3.5. The back-calculated data was then used as the input for distance geometry. Although the back-calculation of signal can be performed for any choice of two conformations, the output conformations obtained from distance geometry calculation must resemble the choice of input conformations as a validation. To our delight, when the N and I conformation observed from molecular dynamics was chosen to back-calculate the signals, distance geometry based on the back-calculated data did produce conformations resembling to the input conformations.

It should be noted that a total of 8 NOE and 9 no-NOE data points were involved in the calculation of the (1→4) linkage. Because only the ratio of two NOE signals ( $I_1/I_2$ ) was used to calculate  $P_N$ , the absolute value ( $I_1$  and  $I_2$ ) and rest of the data points still provides enough structural information for unbiased calculation and validation. Three randomly assumed test pairs of ( $\psi_N$ ,  $\phi_N$ ) and ( $\psi_I$ ,  $\phi_I$ ) were used to reconstruct the NOE data. Structural calculation based on these sets of reconstructed NOE data generated very

different output conformations (i.e.,  $(\psi_N, \varphi_N)$  and  $(\psi_I, \varphi_I)$  are inconsistent to the input values).

We then aim to find the optimal input conformations by minimizing the difference between the choice of input conformations and output conformations through exploring the search space composed of 4 variables, namely the  $\varphi$  and  $\psi$  angle of the N and I conformation. About 100 pairs of conformation were used for back-calculation and distance geometry calculation until the output  $\varphi$  and  $\psi$  angles of the N and I conformation were in good agreement with the input. The distance restraints were also nicely represented in the output conformations. Furthermore, the van der Waals radii are no longer violated in the output conformations.

We therefore accepted the backbone structure and performed the exocyclic group modeling for both the N conformer and I conformer with the GLYCAM force field. 25 structures with the best scores were selected as the final ensemble of structures, shown in figure 3.6 and 3.7. Average glycosidic torsions of the CS-D N and CS-D I ensemble is shown in table 3.10 and 3.11. The calculated population of the N and I conformer is 95.5% and 4.5%, respectively. As can be seen in figure 3.6 and 3.7, the two conformers present very distinct structural feature. While the N conformer is in an extended conformation, the I conformer presents a concaved structure that has a kink in the GalNAc2-GlcA1 linkage due to the nearly 180° flip in the  $\psi$  angle of the (1→4) linkage.

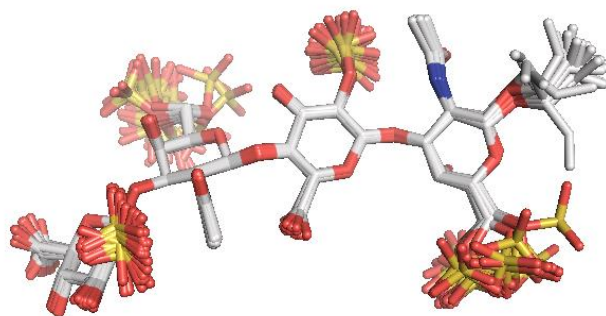
Molecular mechanics modeling suggested that exocyclic groups of the CS-D N conformer adopt similar orientations with that of the CS-E tetrasaccharide. The hydroxymethyl group of GalNAc units is flexible and adopts both the *trans-gauche* and *gauche-trans* orientation. The 6-O-sulfate group of GalNAc units is also flexible and presents

either an *anti* orientation or an eclipsed orientation with respect to C6 or H6, respectively. On the other hand, the 2-*O*-sulfate group and the carboxylate oxygens of GlcA units are more rigid and adopt an eclipsed orientation and an anti/eclipsed orientation with regard to H2 and H5, respectively.

As for the CS-D I conformer, the orientations of exocyclic groups of the terminal monosaccharide units are similar to the CS-D N conformer. However, change of the  $\psi$  angle in the (1 $\rightarrow$ 4) linkage brings the two internal units in proximity. The hydroxymethyl group and carboxylate of the internal monosaccharide units therefore exhibit different behaviors to prevent steric clashes with the neighbor internal monosaccharide unit. The hydroxymethyl group of GalNAc2 becomes more rigid and only adopts the *trans-gauche* orientation, avoiding the GlcA1 unit. The carboxylate group on GlcA1 occurs in an eclipsed orientation with respect to C4, as opposed to be eclipsed with H5. The internal 2-*O*-sulfate and 6-*O*-sulfate group are not affected by the change of  $\psi$  angle.

Torsion	$\beta(1\rightarrow3)$ GlcA2-GalNAc2	$\beta(1\rightarrow4)$ GalNAc2-GlcA1	$\beta(1\rightarrow3)$ GlcA1-GalNAc1
$\varphi$	-109(3)	-72(2)	-93(3)
$\psi$	119(2)	-120(2)	112(3)

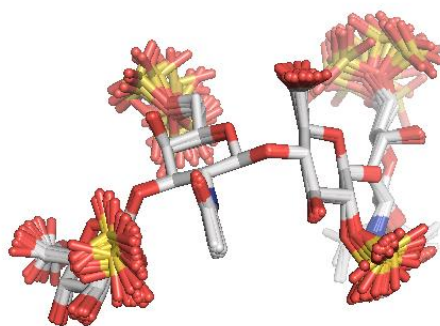
Table 3.11. Average glycosidic linkage torsion angles of the CS-D N conformer.



**Figure 3.6.** Ensemble of the top 25 structure of the CS-D tetrasaccharide N conformer.

Torsion	$\beta(1\rightarrow3)$ GlcA2-GalNAc2	$\beta(1\rightarrow4)$ GalNAc2-GlcA1	$\beta(1\rightarrow3)$ GlcA1-GalNAc1
$\varphi$	-108(3)	-81(2)	-95(3)
$\psi$	118(3)	55(1)	112(3)

Table 3.12. Average glycosidic linkage torsion angles of the CS-D I conformer.



**Figure 3.7.** Ensemble of the top 25 structure of the CS-D tetrasaccharide I conformer.

**Free energy landscape of each linkages in CS-D and CS-E**

After obtaining the NMR structures of the CS-D and CS-E tetrasaccharide, the immediate question is: Does a second conformer exist in other linkages of CS-D or CS-E as well?

We first looked into the molecular dynamics simulation for insights. As aforementioned, the I conformer of CS-D is observed in the simulation. However, the same conformational space is occupied in CS-E in the same GalNAc2-GlcA1 linkage, too, with a lower relative population than that in CS-D. This population is referred to as the CS-E I conformer and the major population is referred as the CS-E N conformer in this study. Observation of the CS-E I conformer in the simulation suggests the CS-E I conformer is not sterically forbidden and that the free energy gap between the N and I conformer in CS-E is larger than that in CS-D. To visualize the relative energy between conformations, we derived the free energy landscape for each linkage from the molecular dynamics simulation by Boltzmann's factor. We binned the population with a 5° interval and calculated the energy relative to the most populated (lowest energy) region by the following equation:

$$\Delta G = -k_B T \ln \left( \frac{P}{P_0} \right) \text{ (equation 3.6)}$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature of the simulations (298 K), and  $P$  and  $P_0$  are the number of frames with the  $\varphi$  and  $\psi$  binned in a given 5° x 5° space and in the most populated 5° x 5° space, respectively. Because the simulation trajectories contain non- $^4C_1$  puckers, especially in CS-D, only the frames with every monosaccharide units in the  $^4C_1$  form were extracted and analyzed. The derived free energy landscape of each linkage in CS-D and CS-E is shown in figure 3.8 and the local minima are summarized in table 3.13 and 3.14.

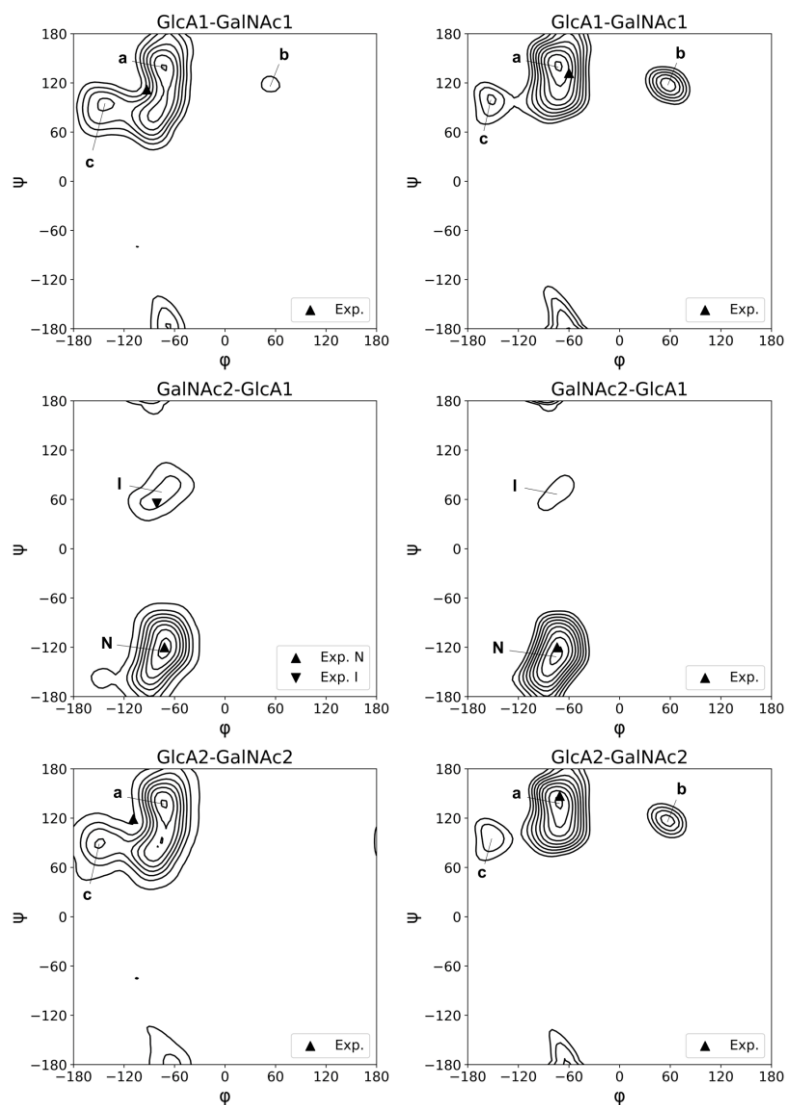


Linkage	minimum	$\Phi$	$\Psi$	$\Delta G^a$
GlcA1-GalNAc1	a	-70	135	0.0
	b	50	120	2.1
	c	-135	100	1.1
GalNAc2-GlcA1	N	-70	-125	0.0
	I	-75	70	2.0
GlcA2-GalNAc2	a	-70	135	0.0
	c	-155	80	1.6

**Table 3.13.** Low-energy conformers of each linkage in the CS-D tetrasaccharide. <sup>a</sup>Relative energy in kcal/mol with respect to global minimum.

Linkage	minimum	$\phi$	$\Psi$	$\Delta G^a$
GlcA1-GalNAc1	a	-70	140	0.0
	b	60	120	1.1
	c	-155	105	2.3
GalNAc2-GlcA1	N	-75	-130	0.0
	I	-65	80	2.8
GlcA2-GalNAc2	a	-70	140	0.0
	b	60	115	1.6
	c	-155	105	2.5

**Table 3.14.** Low-energy conformers of each linkage in the CS-E tetrasaccharide. <sup>a</sup>Relative energy in kcal/mol with respect to global minimum.



**Figure 3.8.** Free energy landscape contour maps of each linkage in the CS-D (left) and CS-E (right) tetrasaccharide. Glycosidic torsion angles of experimentally determined structures are marked with triangles. Isocontour levels are drawn at 0.4 kcal/mol increments from 0.4 kcal/mol above the global energy minimum. Codes a, b, c, N, and I represent the energy minima.

The free energy landscape suggested that the energy gap between the CS-D N and CS-D I conformer is 2.0 kcal/mol, which is close to the experimental energy gap 1.8 kcal/mol derived from the relative population 95.5% and 4.5% observed by NMR-based calculation. The larger energy gap between the CS-E conformers (2.8 kcal/mol) suggested by simulation may explain why the CS-E I conformer was not observed experimentally; however, we believe the real energy gap to be larger than 2.8 kcal/mol.

The energy gap 2.8 kcal/mol was used to calculate the expected relative population of CS-E N and CS-E I conformers, which is 99.1% and 0.9%, respectively. This population along with the distances measured from the CS-E N and CS-E I conformer ( $r_N$  and  $r_I$ ) was then used to back-calculate the expected NOE signal and expected NOE-derived interproton distance. The expected distance can next be compared with the experimentally observed distance to validate the relative population suggested by simulation. The proton pair that is the most sensitive to the relative population (i.e., the proton pair with the largest ( $r_N - r_I$ ) value) was GalNAc2H1-GlcA1H3, with the  $r_N$  and  $r_I$  of 4.5 Å and 2.2 Å, respectively. The back-calculated expected distance of GalNAc2H1-GlcA1H3 turned out to be 3.7 Å, which is inconsistent with the experimentally observed distance 4.7 Å. Since the observed distance is in very good consistency with  $r_N$ , we conclude that the CS-E I conformer is practically non-existent.

As for the GlcA1-GalNAc1 linkage, the free energy landscape suggested an isolated second conformer “b” in both linkages with a low energy gap of 1.1 kcal/mol. We then calculated the expected NOE distance of the proton pair most sensitive to the population of the “b” conformer, which is GlcA1H2-GalNAc1H3. The expected NOE distance derived

from the relative population of “b” (13.5%),  $r_a$  (4.5 Å) and  $r_b$  (2.0 Å) was calculated to be 2.8 Å. The experimentally observed distance ( $r$ ) is 3.9 Å, suggesting a potential equilibrium of the “a” and “b” conformer but with a population of “b” lower than 13.5%. If we use  $r$ ,  $r_a$ , and  $r_b$ , to calculate the relative population of “a” and “b”, we will obtain 98.9% and 1.1%, respectively. The fact that the glycosidic torsion angles of the calculated ensemble fell slightly off-center of the energy minimum may also imply that the NMR data is slightly affected by a second conformer, however, this implication relies on accuracy of the simulation.

On the other hand, while the equilibrium potentially exists, it is insufficient to fully confirm the existence of this “b” conformer with our current NMR data since we did not determine a physically allowed structure without the inclusion of a second conformer. Also, the rest of the NMR restraints still agree very well with the “a” conformer, including the restraints somewhat sensitive to the “b” conformer. Lastly, the experimentally observed distance (3.9 Å) is relatively long and may suffer from error more significantly compared to the experimentally observed distance (3.0 Å) of GalNAc2H1-GlcA1H3 in CS-D, which is the most sensitive to the I conformer and used to determine the relative population of CS-D I. To sum up, although a small population (~ 1%) of a second conformer potentially exists, our NMR data is not sufficient to fully confirm or reject the possibility.

The GlcA2-GalNAc2 linkage of CS-E is similar to the GlcA2-GalNAc2 linkage, with “b” conformer having a 1.6 kcal/mol energy gap. The expected NOE distance of GlcA2H2-GalNAc2H3 was calculated accordingly with the relative population of “b” (6.3%),  $r_a$  (4.5 Å) and  $r_b$  (2.1 Å). The calculated expected distance is 2.8 Å, which is very

inconsistent with the experimentally observed distance ( $r$ ) 4.2 Å. Since the distance measured in the “a” conformer (4.5 Å) is already in good agreement with the experimentally observed distance (4.2 Å), it is unlikely that a second conformer with a strong effect on the signal exists.

We next looked to the GlcA1-GalNAc1 and GlcA2-GalNAc2 linkage of CS-D. The free energy landscape suggested a relative broad conformational space for both linkages. No significant isolated conformer is observed in both linkages. Although a “c” local minimum can be found in the landscape, it is not an isolated conformer separated with the major conformer by a significant energy barrier like CS-D I. Instead, the “c” local minimum can be considered as a demonstration of flexibility of the “a” conformer. As previously mentioned, the NMR data represents an averaged measurement. The fact that the glycosidic torsion angles of the calculated ensembles do not overlap perfectly with the energy minimum may be accounted for by the non-symmetrical shape of the conformational space around the “a” conformer (figure 3.8).

The flexibility of the “a” conformer in CS-D may also account for the difference of glycosidic torsion angles between the GlcA-GalNAc linkages of CS-D and CS-E. Although the “c” minima are also observed in CS-E, the energy gaps between “c” and “a” are smaller in CS-D (1.1 and 1.6 kcal/mol) than in CS-E (2.3 and 2.5 kcal/mol). The energy barriers between “c” and “a” are lower in CS-D, too, as can be observed from the continuously occupied conformational space (figure 3.8). Since both the energy gaps and barriers are lower in CS-D, the simulation suggests that the GlcA-GalNAc linkages of CS-E are more rigid. As a result, the NMR determined the  $\varphi$  and  $\psi$  angles in CS-D are more affected by the “c”

minima (i.e., the flexibility of “a”) and shifted more towards the “c” minima. But again, this rationalization relies on the accuracy of the simulations.

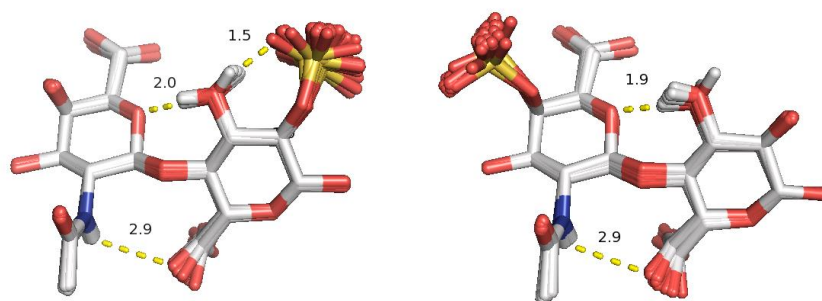
To sum up, while a minor conformer CS-D I is clearly observed and determined by NMR and calculation for the GalNAc2-Glc1 linkage of CS-D, the CS-E I conformer is practically non-existent as determined by NMR data and simulation. As for the Glc1-GalNAc1 linkage in CS-E, our NMR data is not sufficient to fully confirm or reject a potential second conformer “b” with ~1% of population. For the Glc2-GalNAc2 linkage, the combination of NMR data and simulation suggests that a second conformer does not exist. On the other hand, a second isolated conformer with low energy is not observed in free energy landscape of the GlcA-GalNAc linkages of CS-D and NMR data does not indicate a second conformer, either.

#### **Inter-residual interactions of the CS-D and CS-E tetrasaccharide**

After exploring the energy of each conformers, we then investigated the calculated structure to find potential interactions that may account for structural difference between CS-D and CS-E.

As shown in figure 3.9, there is an inter-residual hydrogen bond between O3 of GalNAc2 and the 3-hydroxyl hydrogen in the GalNAc2-GlcA1 linkage of both CS-D and CS-E. However, in CS-D the 3-hydroxyl hydrogen can also form an intra-residual hydrogen bond with the 2-*O*-sulfate group, weakening the stabilizing force of the linkage. In the top 25 structures, 11 of the 3-hydroxyl hydrogens point to the 2-*O*-sulfate group in CS-D, but only one of the 3-hydroxyl hydrogens point to the 2-position in CS-E. One other inter-residual hydrogen bond can be observed between the amide proton and the carboxylate

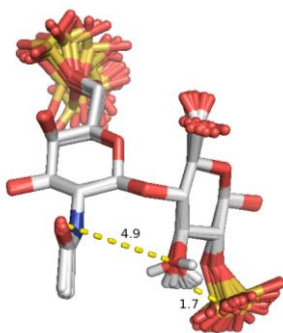
group. This hydrogen bond is similarly weak in both the CS-D N and CS-E N conformers, and therefore will not affect the difference between energy gaps of the N and I conformer in CS-D and CS-E. Both hydrogen bonds have also been proposed with previous molecular dynamics study of non-sulfated chondroitin<sup>12,24</sup>.



**Figure 3.9.** Comparison of inter-residual interaction in NMR-based GalNAc2-GlcA1 linkage structure for CS-D (left) and CS-E (right). For clarity, the 6-*O*-sulfate groups and all hydrogen atoms are hidden except for the 3-hydroxyl hydrogen of GlcA1 and the amide proton of GalNAc2. Indicated distances (Å) are measured for GalNAc2O5-GlcA1HO3, GlcA1HO3-GlcA1OS2, and GalNAc2HN-GlcAO6.

After locating a potential difference between the CS-D N and CS-E N conformers, we next examined the CS-D I conformer. As can be seen in figure, no 3-hydroxyl hydrogen points to the carbonyl oxygens, which are the closest hydrogen bond acceptor in GalNAc2, suggesting that no inter-residual hydrogen bond is formed. The distance between the 3-hydroxyl oxygen and the carbonyl oxygen (4.9 Å) is also larger than the maximum distance (4.0 Å) that allows the formation of a hydrogen bond<sup>25</sup>. Although we don't have a NMR-determined structure for the CS-E I conformer, molecular dynamics simulations suggest that the CS-D I and CS-E I conformer share very similar glycosidic linkage torsion angles. Therefore, we do not expect an inter-residual hydrogen bond for the CS-E I conformer,

either. On the other hand, a repulsive electrostatic interaction between the 6-*O*-sulfate groups and carboxylate groups can be observed in the CS-D I conformer, causing the carboxylate group to adopt an orientation unseen in other structures. Since the 6-*O*-sulfate group is also present in CS-E I conformer along with an additional 4-*O*-sulfate group, the repulsive interaction is expected to be stronger in CS-E I than in CS-D I, destabilizing the CS-E I conformer.



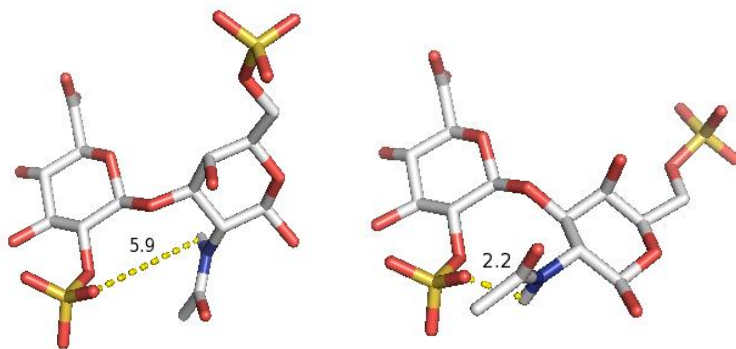
**Figure 3.10.** Structure of the NMR-based GalNAc2-GlcA1 linkage of CS-D I. For clarity, all hydrogen atoms are hidden except for the 3-hydroxyl hydrogen of GlcA1. Indicated distances (Å) are measured for GalNAc2ON2-GlcA1O3 and GlcA1HO3-GlcA1OS2.

To sum up the above observations, the CS-E N conformer stabilizes the GalNAc2-GlcA1 linkage more strongly than the CS-D N conformer, through a stronger hydrogen bond. On the other hand, the CS-E I conformer is more destabilized than the CS-D I conformer by having an extra 4-*O*-sulfate group. Therefore, the energy gap between the CS-E N and I conformers are expected to be larger than the CS-D N and I conformer. As a result, while the equilibrium of the N and I conformer can be observed in CS-D experimentally, the CS-E I conformer is practically non-existent.

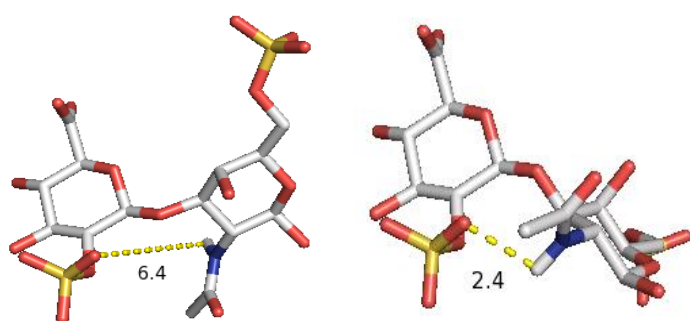


We then looked in the GlcA1-GalNAc1 and GlcA2-GalNAc2 linkages of CS-D and CS-E. For both linkages, the  $\varphi$  and  $\psi$  angles are both smaller in CS-D, which may be explained by the extended conformational space in CS-D as aforementioned (figure 3.8). We believe that the loss of the hydrogen bond donor on the 2-position of CS-D and/or the extra negatively charged group on the 4-*O*-position of CS-E contributed to this observation. An inter-residual hydrogen bond between the 2-hydroxyl hydrogen of GlcA and carbonyl oxygen of GalNAc has been proposed in previous studies for unsulfated chondroitins<sup>12,24</sup>, which will be abolished in the GlcA-GalNAc linkages of CS-D.

We then built the model of the “a” and “c” minima of both GlcA-GalNAc linkages in CS-D and CS-E based the corresponding glycosidic torsion angles (table 3.13 and 3.14) to explore the potential cause of the more flexible conformational space in CS-D. As shown in figure 3.11, and 3.12 the “c” conformer in both the GlcA1-GalNAc1 and GlcA2-GalNAc2 linkages of CS-D have an inter-residual hydrogen that can not be observed in the “a” conformer, which may explain the smaller energy gap in CS-D.



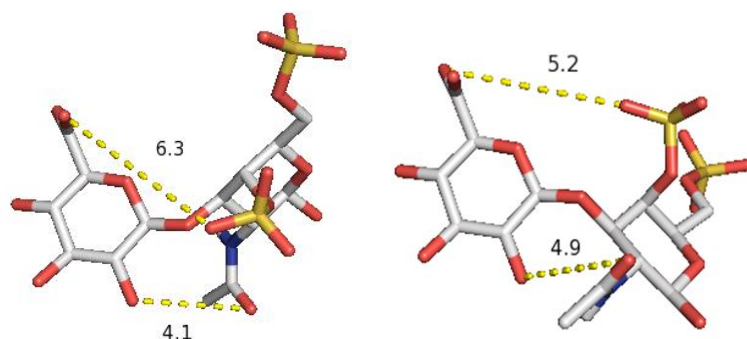
**Figure 3.11.** The “a” (left) and “c” (right) conformer of the GlcA1-GalNAc1 linkage in CS-D. All hydrogen atoms except for the amide proton are hidden for clarity. Indicated distances (Å) are measured for GlcA1OS2-GalNAc1HN.



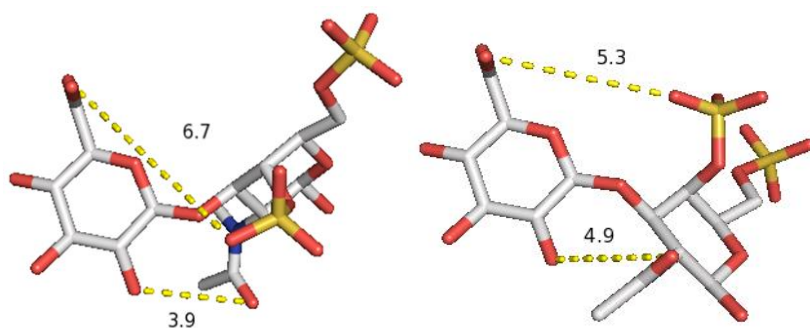
**Figure 3.12.** The “a” (left) and “c” (right) conformer of the GlcA2-GalNAc2 linkage in CS-D. All hydrogen atoms except for the amide proton are hidden for clarity. Indicated distances (Å) are measured for GlcA2OS2-GalNAc2HN.

As for CS-E, a hydrogen bond between the 2-hydroxyl hydrogen of GlcA and the carbonyl oxygen of GalNAc could be observed, albeit weak. Since the cutoff of X---A distance in a hydrogen bond X-H---A is  $4.0 \text{ \AA}^{25}$ , the “a” conformer have a borderline hydrogen bond and the “c” conformer do not, as seen in figure 3.13 and 3.14. On the other hand, the extra 4-*O*-sulfate groups in CS-E are also closer to the carboxylate groups in the “c” conformer than in the “a” conformer, which may lead to electrostatic repulsion that destabilizes the “c” conformer. Together the hydrogen bond and the electrostatic interaction may partly account for why the “c” conformer is less favored in CS-E than in CS-D.

To sum up, we identified the inter-residual interactions that may account for the higher flexibility of the GlcA-GalNAc linkage in CS-D. The higher flexibility can potentially in turn explain the difference between the NMR-based glycosidic torsion angles of the GlcA-GalNAc linkage in CS-D and CS-E.



**Figure 3.13.** The “a” (left) and “c” (right) conformer of the GlcA1-GalNAc1 linkage in CS-E. All hydrogen atoms except for the amide proton are hidden for clarity. Indicated distances (Å) are measured for GlcA1HO2-GalNAc1ON2 and GlcA1O6-GalNAc1OS6.



**Figure 3.14.** The “a” (left) and “c” (right) conformer of the GlcA2-GalNAc2 linkage in CS-E. All hydrogen atoms except for the amide proton are hidden for clarity. Indicated distances (Å) are measured for GlcA2HO2-GalNAc2ON2 and GlcA2O6-GalNAc2OS6.

#### *Comparison of the glycosidic torsion angles with published structures*

The glycosidic torsion angles of CS-D N, CS-D I and CS-E determined by the NMR data and distance geometry/simulated annealing calculation in this study are compared with selected published structures (table 3.15). Our glycosidic torsion angles of the  $\beta(1\rightarrow4)$

linkages are in general consistent with the published structures. As for the  $\beta(1\rightarrow3)$  linkages, the  $\varphi$  angles of CS-D and the  $\psi$  angles are different from the published structures. However, we must bear in mind that while our glycosidic is NMR-based without the involvement of molecular mechanics, the published structures utilized molecular mechanics along with less NMR data points. Also, the sulfation pattern is different between our molecules with the published structures. We envisioned that future study of CS-A, CS-C and unsulfated chondroitin with the current method will offer more direct comparison and methodological insights for GAG structure determination.

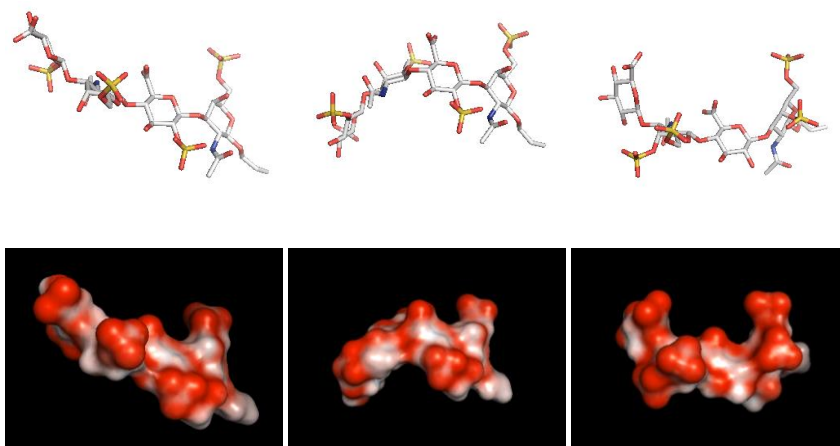
Method	Length <sub>a</sub>	Structure	torsion	$\beta(1\rightarrow3)$	$\beta(1\rightarrow4)$
NMR/DG <sup>b</sup> /SA <sup>c</sup>	4	CS-D N	$\varphi$	-101	-72
			$\psi$	116	-120
NMR/DG/SA	4	CS-D I	$\varphi$	-102	-82
			$\psi$	115	55
NMR/DG/SA	4	CS-E	$\varphi$	-66	-74
			$\psi$	140	-120
NMR/rMD <sup>d</sup> /SA <sup>12</sup>	6	CS-0 <sup>e</sup>	$\varphi$	-72	-73
			$\psi$	108	-117
NMR/rMD/SA <sup>13</sup>	5	CS-A	$\varphi$	-61	-67
			$\psi$	109	-124
NMR/MM <sup>f,8</sup>	8	CS-A	$\varphi$	-80	-80
			$\psi$	90	-110
X-ray <sup>20</sup>	8	CS-A	$\varphi$	-80	-98
			$\psi$	128	-129

**Table 3.15.** Comparison of glycosidic torsion angles in the present study with selected published CS structures. Values are reported in units of degrees. <sup>a</sup>Length is presented in the number of monosaccharide units. <sup>b</sup>Distance geometry. <sup>c</sup>Simulated annealing. <sup>d</sup>Restrained molecular dynamics. <sup>e</sup>Unsulfated chondroitin. <sup>f</sup>Molecular mechanics.

### ***Electrostatic potential map of CS-D and CS-E***

To better understand how the solution structure may interact with proteins, we chose a representative structure from the ensembles and calculated its electrostatic surface by the Adaptive Poisson-Boltzmann Solver (APBS)<sup>26</sup>. APBS is a popular free macromolecular electrostatics calculation program that has been continually updated since its release in 2001.

As can be seen in figure, each structure exhibits a characteristic electrostatic surface. The CS-D N has an extended structure with the negatively charged group in plane with the backbone. The CS-D I is concaved and presents the negatively charged groups on the convex side. The CS-E structure is also extended but has the negatively charge groups pointing to the same direction, perpendicular to the backbone. With the characteristic electrostatic surfaces, we expect these CS structures to demonstrate different protein binding specificity and have unique biological functions. The calculated surfaces may also aid the protein docking study of CS.



**Figure 3.15.** Stick representation (top) and electrostatic potential surface (bottom) of the CS-D N (left), CS-D I (middle) and CS-E (right). The charged areas are colored on the surface from blue (+20 kT/e) to red (−20 kT/e).

### **Conclusion**

The solution structure of CS-D and CS-E has been determined with high quality data obtained from a state-of-the-art 900 MHz instrument and distance geometry coupled with simulated annealing. A second conformer of the  $\beta(1\rightarrow3)$  GalNAc-GlcA linkage was also observed and determined experimentally for the first time. Together our NMR data and free molecular dynamics simulation suggests that a second conformer does not exist for other linkages in CS-D and CS-E, with the exception of GlcA1-GalNAc1 linkage of CS-E. Our current data is not sufficient to fully confirm or reject a potential second conformer with ~1% of the overall population. Potential inter-residual interactions that may account for the structural difference between the CS-D and CS-E are also identified in the calculated structures.

### **Experimental Methods**

**NMR Spectroscopy.** For NMR analysis, the CS-D and CS-E tetrasaccharide samples were dissolved in 300  $\mu$ L of solvent ( $D_2O$  or 10%  $D_2O$  in  $H_2O$ ) and adjusted to pH 6.0 before transferred into a Shigemi tube. The  $[^1H]$ -1D,  $[^1H,^{13}C]$ -HSQC,  $[^1H,^1H]$ -DQF-COSY and  $[^1H,^1H]$ -NOESY were recorded with a 21.1 Tesla Bruker Avance II NMR spectrometer (Bruker Corporation, Billerica, MA) equipped with a CP TCI multinuclear cryoprobe.  $[^1H]$ -1D,  $[^1H]$ -DQF-COSY and  $[^1H,^{13}C]$ -HSQC spectra were standard experiments. The  $[^1H,^1H]$ -NOESY spectra recorded with a  $^1H$  frequency of 900.25 MHz and a mixing time of 600 ms.

2048 and 1536 complex points were collected in the direct and indirect dimensions, respectively. Spectra were processed and analyzed with TopSpin and MestraNova. Appropriate linear prediction, window functions, and zero-filling were used to achieve the maximum possible resolution from each dataset.

**Molecular dynamics preparation.** Necessary parameters for the allyl moiety were imported from the GAFF (generalized AMBER force-field) parameter sets; partial charges were estimated for the allyl moiety by performing multi-conformer RESP fitting<sup>27-28</sup> by fitting to these electrostatic potentials with a restraint weight (qwt) of 0.0005 to electrostatic potential maps calculated using GAUSSIAN09 at the HF/6-31G(\*) level of theory for 20 separate conformers of allyl  $\beta$ -D-glucopyranoside (extracted at equal intervals over a 50 ns molecular dynamics trajectory total based on an initial guess of partial charges from a single conformer). To maintain compatibility with GLYCAM non-polar (aliphatic) hydrogen atoms were fixed at zero charge and the glucose atoms were fixed at the GLYCAM version 06j-1 partial charges of residue [name OGB].

The starting structures of the CS-D and CS-E tetrasaccharides were modeled using the standard GLYCAM (version 06j-1)<sup>7</sup> and [sulfate] parameter sets<sup>29</sup>. Adjustments to the sulfonate oxygen partial charges were made to maintain the correct overall charge as described previously<sup>29</sup>. Each molecule was minimized *in vacuo* using 1,000 steps of steepest descent. They were then solvated into cubic boxes of water leaving 1.2 nm of water between the van der Waals extent of each molecule and the box edge and charge neutralized using sodium ions. Any water molecules overlapping the van der Waals surface were deleted using the AMBER leap tool. Each carbohydrate was solvated in explicit TIP3P water molecules with a cubic box of equal side length using the leap program from AmberTools<sup>28</sup>, resulting

Commented [KY1]:

Commented [KY2]: Partial charges are given in appendix ?.

Commented [KY3]: [You might want to cross reference the exact molecular structures here – I guess they are elsewhere in your thesis]

in water boxes of  $\approx 55$  Å. Solute atoms were positioned at least 12 Å from the solvent box edge and the assemblies were neutralized by adding explicit  $\text{Na}^+$  ions.

**Molecular dynamics simulations.** Free molecular dynamics simulations of the tetrasaccharides were performed for 10  $\mu\text{s}$  each using the ~~AMBER16~~ GLYCAM force field and NVIDIA GeForce GTX 980 graphics processors (Kepler architecture) using methods described previously<sup>30</sup>. Briefly, following initial conjugate-gradient energy minimization (1000 steps in water) the assemblies were heated from 0 to 298 K and then equilibrated in the NPT ensemble for 20 ns, and the first 300 ns were discarded, prior to 10  $\mu\text{s}$  of NVT production dynamics; 3D-coordinate data were recorded at 100 ps intervals. The velocity-Verlet integration algorithm and a hydrogen mass repartitioning scheme enabled a 4 fs time-step to be used without affecting the equilibrium distribution<sup>31-32</sup>. Hydrogen atoms were constrained using M-SHAKE<sup>33</sup> and electrostatic interactions were calculated via the particle mesh Ewald method, with a grid spacing of less than or equal to 1.0 Å (in the X, Y and Z dimensions). Electrostatic and van der Waals interactions were truncated at 9 Å and the recommended scaling factor for GLYCAM carbohydrate 1–4 interactions (1.0) was employed.

**Structure calculations.** Reference intra-residual NOE signals were used to build a model relating cross-peak intensity ( $I$ ) to proton–proton distance ( $r$ ) using the relationship  $I = cr^{-6}$ . The constant of proportionality ( $c$ ) was then derived from a linear plot of  $r^{-6}$  and  $I$  of the

**Commented [KY4]:** We used GLYCAM right? Typo?



reference NOEs. This was employed to compute experimental distance restraints from the inter-residual NOEs and “no-NOEs”.

Starting conformation of the tetrasaccharides used for structure calculation was generated using online GLYCAM tools<sup>34</sup> and input into the SimTK molecular modeling API (molmodel). The H2-C2-N2-HN torsion angle was fixed at 180°. Each sugar residue was in a fixed pucker geometry (the <sup>4</sup>C<sub>1</sub> pucker) and modeled as a rigid body, while the glycosidic degrees of freedom between the sugar residue rigid bodies were modeled as pin joints (rotation about the bond axis, but no translation), corresponding to the  $\psi$  and  $\phi$  angles involved in each glycosidic linkages. 250 rounds of simulated annealing were performed with 500,000 steps for a given input of NMR data. The starting conformation of the pin joints was randomized and then subjected to simulated annealing performed by applying molecular velocities appropriate to the temperature. The simulated annealing was designed to provide minimum solutions to the score function (equation 3.1), by applying forces to the model during simulated annealing that are proportional to the squared distance deviation from a desired measured value.

After the backbone conformation was calculated and accepted, simulated annealing with the GLYCAM force field was employed to determine the orientation of the exocyclic groups. Briefly, the tetrasaccharide was prepared as described in the molecular dynamics methods (including solvent TIP3P explicit water molecules and ions). Each structure was heated to 298 K for 500 ps and then cooled to 0 K over 3 ns. During the whole 3.5 ns, the backbone conformation and the acetamido was restrained to the result of SIMTK calculation with a weighing factor of 1,000 to maintain the backbone conformation determined with NMR data. The structures are analyzed with VMD<sup>35</sup> and visualized with PYMOL<sup>36</sup>.

### Electrostatic surface calculation

The representative structure was chosen by measuring the RMSD of the ensemble. The structure that gives the lowest average heavy-atom (non-hydrogen) RMSD of the ensemble when used as the reference structure was selected be the representative structure. The PQR files required by the APBS software<sup>26</sup> were manually prepared with the atomic charges used in free molecular dynamics simulation and atomic coordinates of the representative structures. PYMOL and the PYMOL plugin ABPS tools 2.1 were used to generate and visualize the electrostatic surfaces.

### References

- (1) Zsiska, M.; Meyer, B., Influence of sulfate and carboxylate groups on the conformation of chondroitin sulfate related disaccharides. *Carbohydrate Research* **1993**, *243* (2), 225.
- (2) Mulloy, B.; Forster, M. J.; Jones, C.; Davies, D. B., N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochemical Journal* **1993**, *293* ( Pt 3), 849.
- (3) Holmbeck, S. M.; Petillo, P. A.; Lerner, L. E., The solution conformation of hyaluronan: a combined NMR and molecular dynamics study. *Biochemistry* **1994**, *33* (47), 14246.
- (4) Cremer, D.; Pople, J. A., General Definition of Ring Puckering Coordinates. *Journal of the American Chemical Society* **1975**, *97* (6), 1354.
- (5) Lii, J. H.; Allinger, N. L., Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .3. The Vanderwaals Potentials and Crystal Data for Aliphatic and Aromatic-Hydrocarbons. *Journal of the American Chemical Society* **1989**, *111* (23), 8576.
- (6) Case, D. A.; Cerutti, D. S.; T.E. Cheatham, I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; A.W. Goetz; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; P. Li, C. L.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A., *AMBER 2017*. University of California: San Francisco, 2017.

- (7) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *Journal of Computational Chemistry* **2008**, *29* (4), 622.
- (8) Blanchard, V.; Chevalier, F.; Imbert, A.; Leeftink, 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* (5), 1167.
- (9) Solera, C.; Macchione, G.; Maza, S.; Kayser, M. M.; Corzana, F.; de Paz, J. L.; Nieto, P. M., Chondroitin Sulfate Tetrasaccharides: Synthesis, Three-Dimensional Structure and Interaction with Midkine. *Chemistry-a European Journal* **2016**, *22* (7), 2356.
- (10) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossvy, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Wolf, R. M.; J. Liu, X. W.; Brozell, S. R.; Steinbrecher, T.; H.Gohlke; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., *AMBER 12*. University of California, San Francisco.: 2012.
- (11) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B., Molecular Mechanical and Molecular Dynamic Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM\_93 Parameter Development. *The Journal of Physical Chemistry* **1995**, *99* (11), 3832.
- (12) Sattelle, B. M.; Shakeri, J.; Roberts, I. S.; Almond, A., A 3D-structural model of unsulfated chondroitin from high-field NMR: 4-sulfation has little effect on backbone conformation. *Carbohydrate Research* **2010**, *345* (2), 291.
- (13) Yu, F.; Wolff, J. J.; Amster, I. J.; Prestegard, J. H., Conformational preferences of chondroitin sulfate oligomers using partially oriented NMR spectroscopy of  $^{13}\text{C}$ -labeled acetyl groups. *Journal of the American Chemical Society* **2007**, *129* (43), 13288.
- (14) Valafar, H.; Prestegard, J. H., REDCAT: a residual dipolar coupling analysis tool. *Journal of Magnetic Resonance* **2004**, *167* (2), 228.
- (15) 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* (1), 47.
- (16) 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* (1), 65.

- (17) Silipo, A.; Zhang, Z.; Canada, F. J.; Molinaro, A.; Linhardt, R. J.; Jimenez-Barbero, J., Conformational analysis of a dermatan sulfate-derived tetrasaccharide by NMR, molecular modeling, and residual dipolar couplings. *Chembiochem* **2008**, *9* (2), 240.
- (18) Haasnoot, C. A. G.; Deleeuw, F.; Altona, C., THE RELATIONSHIP BETWEEN PROTON-PROTON NMR COUPLING-CONSTANTS AND SUBSTITUENT ELECTRONEGATIVITIES .1. AN EMPIRICAL GENERALIZATION OF THE KARPLUS EQUATION. *Tetrahedron* **1980**, *36* (19), 2783.
- (19) Karplus, M., Vicinal Proton Coupling in Nuclear Magnetic Resonance. *Journal of the American Chemical Society* **1963**, *85* (18), 2870.
- (20) Winter, W. T.; Arnott, S.; Isaac, D. H.; Atkins, E. D., Chondroitin 4-sulfate: the structure of a sulfated glycosaminoglycan. *Journal of Molecular Biology* **1978**, *125* (1), 1.
- (21) Aguda, A. H.; Panwar, P.; Du, X.; Nguyen, N. T.; Brayer, G. D.; Bromme, D., Structural basis of collagen fiber degradation by cathepsin K. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111* (49), 17474.
- (22) Michel, G.; Pojasek, K.; Li, Y.; 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* (31), 32882.
- (23) Sattelle, B. M.; Shakeri, J.; Cliff, M. J.; Almond, A., Proteoglycans and their heterogeneous glycosaminoglycans at the atomic scale. *Biomacromolecules* **2015**, *16* (3), 951.
- (24) Almond, A.; Sheehan, J. K., Glycosaminoglycan conformation: do aqueous molecular dynamics simulations agree with x-ray fiber diffraction? *Glycobiology* **2000**, *10* (3), 329.
- (25) Desiraju, G. R.; Steiner, T., *The weak hydrogen bond in structural chemistry and biology*. Oxford University Press: Oxford, 1999.
- (26) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98* (18), 10037.
- (27) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993**, *97* (40), 10269.

(28) Case, D. A.; Betz, R. M.; Cerutti, D. S.; T.E. Cheatham, I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A., *AMBER 2016*. University of California: San Francisco, 2016.

(29) Singh, A.; Tessier, M. B.; Pederson, K.; Wang, X.; Venot, A. P.; Boons, G. J.; Prestegard, J. H.; Woods, R. J., Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. *Canadian Journal of Chemistry* **2016**, *94* (11), 927.

(30) Sattelle, B. M.; Almond, A., Shaping up for structural glycomics: a predictive protocol for oligosaccharide conformational analysis applied to N-linked glycans. *Carbohydrate Research* **2014**, *383*, 34.

(31) Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; De Fabritiis, G., High-throughput all-atom molecular dynamics simulations using distributed computing. *J Chem Inf Model* **2010**, *50* (3), 397.

(32) Sadiq, S. K.; Noe, F.; De Fabritiis, G., Kinetic characterization of the critical step in HIV-1 protease maturation. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109* (50), 20449.

(33) Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H., A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry* **2001**, *22* (5), 501.

(34) Woods, R. J. *GLYCAM Web*, Complex Carbohydrate Research Center, University of Georgia, Athens, GA.: 2009.

(35) Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **1996**, *14* (1), 33.

(36) PyMOL *The PyMOL Molecular Graphics System, Version 1.7.4.5 Edu*, Schrödinger, LLC.: , 2014.

*Appendix for Chapter 3*

RELEVANT DATA FOR NMR STRUCTURE DETERMINATION

4C1	0	<=	$\theta^\circ$	<	36					
OE	36	<=	$\theta^\circ$	<	72	&	-15	<=	$\phi^\circ$	< 15
OH1	36	<=	$\theta^\circ$	<	72	&	15	<=	$\phi^\circ$	< 45
E1	36	<=	$\theta^\circ$	<	72	&	45	<=	$\phi^\circ$	< 75
2H1	36	<=	$\theta^\circ$	<	72	&	75	<=	$\phi^\circ$	< 105
2E	36	<=	$\theta^\circ$	<	72	&	105	<=	$\phi^\circ$	< 135
2H3	36	<=	$\theta^\circ$	<	72	&	135	<=	$\phi^\circ$	< 165
E3	36	<=	$\theta^\circ$	<	72	&	165	<=	$\phi^\circ$	< 195
4H3	36	<=	$\theta^\circ$	<	72	&	195	<=	$\phi^\circ$	< 225
4E	36	<=	$\theta^\circ$	<	72	&	225	<=	$\phi^\circ$	< 255
4H5	36	<=	$\theta^\circ$	<	72	&	255	<=	$\phi^\circ$	< 285
E5	36	<=	$\theta^\circ$	<	72	&	285	<=	$\phi^\circ$	< 315
OH5	36	<=	$\theta^\circ$	<	72	&	315	<=	$\phi^\circ$	< 345
3OB	72	<=	$\theta^\circ$	<	108	&	345	<=	$\phi^\circ$	< 15
3S1	72	<=	$\theta^\circ$	<	108	&	15	<=	$\phi^\circ$	< 45
B14	72	<=	$\theta^\circ$	<	108	&	45	<=	$\phi^\circ$	< 75
5S1	72	<=	$\theta^\circ$	<	108	&	75	<=	$\phi^\circ$	< 105
25B	72	<=	$\theta^\circ$	<	108	&	105	<=	$\phi^\circ$	< 135
2SO	72	<=	$\theta^\circ$	<	108	&	135	<=	$\phi^\circ$	< 165
B3O	72	<=	$\theta^\circ$	<	108	&	165	<=	$\phi^\circ$	< 195
1S3	72	<=	$\theta^\circ$	<	108	&	195	<=	$\phi^\circ$	< 225
14B	72	<=	$\theta^\circ$	<	108	&	225	<=	$\phi^\circ$	< 255
1S5	72	<=	$\theta^\circ$	<	108	&	255	<=	$\phi^\circ$	< 285
B25	72	<=	$\theta^\circ$	<	108	&	285	<=	$\phi^\circ$	< 315
OB2	72	<=	$\theta^\circ$	<	108	&	315	<=	$\phi^\circ$	< 345
EO	108	<=	$\theta^\circ$	<	144	&	345	<=	$\phi^\circ$	< 15
1HO	108	<=	$\theta^\circ$	<	144	&	15	<=	$\phi^\circ$	< 45
1E	108	<=	$\theta^\circ$	<	144	&	45	<=	$\phi^\circ$	< 75
1H2	108	<=	$\theta^\circ$	<	144	&	75	<=	$\phi^\circ$	< 105
E2	108	<=	$\theta^\circ$	<	144	&	105	<=	$\phi^\circ$	< 135
3H2	108	<=	$\theta^\circ$	<	144	&	135	<=	$\phi^\circ$	< 165
3E	108	<=	$\theta^\circ$	<	144	&	165	<=	$\phi^\circ$	< 195
3H4	108	<=	$\theta^\circ$	<	144	&	195	<=	$\phi^\circ$	< 225
E4	108	<=	$\theta^\circ$	<	144	&	225	<=	$\phi^\circ$	< 255
5H4	108	<=	$\theta^\circ$	<	144	&	255	<=	$\phi^\circ$	< 285
5E	108	<=	$\theta^\circ$	<	144	&	285	<=	$\phi^\circ$	< 315
5HO	108	<=	$\theta^\circ$	<	144	&	315	<=	$\phi^\circ$	< 345
1C4	144	<=	$\theta^\circ$	<	180					

**Table A3.1.** Ring puckering definitions of the 38 canonical pyranose ring puckers.

Inter-residual pair	restraints	Result
2:H3->3:H1	0.19	0.18
2:H4->3:H1	0.39	0.36
2:H5->3:H1	0.24	0.22
2:H3->3:H1	0.17	0.18
2:H4->3:H1	0.36	0.36
2:H5->3:H1	0.21	0.22
3:HN->2:H3	0.32	0.32
3:HN->2:H4	0.46	0.5
1:HN->2:H2	0.49	0.49
1:HN->2:H1	0.33	0.32
1:H4->2:H1	0.34	0.35
3:HN->4:H2	0.45	0.45
3:HN->4:H1	0.32	0.33
3:H4->4:H1	0.33	0.34
1:H4->2:H1	0.36	0.35
3:H4->4:H1	0.35	0.34

**Table A3.2.** Inter-residual NOE restraints used for CS-D N structure calculation and result. Values are reported in nm.

Inter-residual pair	restraints	Result
2:H2->3:H4	0.24	0.86
2:H2->3:H1	0.24	0.46
2:H3->3:H4	0.25	0.59
2:H4->3:H4	0.25	0.72
2:H5->3:H4	0.23	0.51
3:H1->2:H1	0.23	0.36
3:H2->2:H2	0.23	0.68
3:H3->2:H2	0.23	0.71
3:H5->2:H2	0.23	0.67
1:H1->2:H1	0.38	0.43
1:H2->2:H2	0.38	0.5
1:H2->2:H1	0.38	0.43
1:H3->2:H2	0.38	0.44
1:H5->2:H2	0.38	0.66
1:H5->2:H1	0.38	0.43
2:H2->1:H2	0.4	0.5
2:H2->1:H5	0.4	0.66
2:H2->1:H1	0.4	0.66
2:H2->1:H4	0.4	0.5
2:H3->1:H1	0.43	0.68
2:H3->1:H2	0.43	0.65



2:H3->1:H5	0.43	0.7
2:H3->1:H4	0.43	0.61
2:H4->1:H1	0.43	0.83
2:H4->1:H2	0.43	0.73
2:H4->1:H5	0.43	0.77
2:H4->1:H4	0.38	0.59
2:H5->1:H2	0.38	0.67
2:H5->1:H5	0.38	0.58
2:H5->1:H1	0.38	0.65
2:H5->1:H4	0.38	0.47
3:H1->4:H1	0.38	0.43
3:H2->4:H2	0.38	0.48
3:H3->4:H2	0.38	0.43
3:H5->4:H2	0.38	0.66
4:H2->3:H1	0.4	0.64
4:H2->3:H4	0.4	0.51
4:H3->3:H2	0.38	0.64
4:H3->3:H1	0.38	0.66
4:H3->3:H4	0.38	0.61
4:H4->3:H1	0.43	0.83
4:H4->3:H2	0.43	0.72
4:H4->3:H3	0.43	0.58
4:H4->3:H4	0.43	0.61
4:H4->3:H5	0.43	0.78
4:H5->3:H2	0.38	0.67
4:H5->3:H1	0.38	0.66
4:H5->3:H4	0.38	0.48

**Table A3.3.** Inter-residual “no-NOE” restraints used for CS-D N structure calculation and result. Values are reported in nm.

Inter-residual pair	restraints	Result
2:H3->3:H1	0.19	0.18
2:H4->3:H1	0.39	0.36
2:H5->3:H1	0.24	0.22
1:H4->2:H1	0.36	0.35
3:H4->4:H1	0.35	0.34
2:H3->3:H1	0.17	0.18
2:H4->3:H1	0.36	0.36
2:H5->3:H1	0.21	0.22
3:HN->2:H3	0.32	0.32
3:HN->2:H4	0.46	0.5
1:HN->2:H2	0.49	0.49

1:HN->2:H1	0.33	0.32
1:H4->2:H1	0.34	0.35
3:HN->4:H2	0.45	0.46
3:HN->4:H1	0.32	0.33
3:H4->4:H1	0.33	0.34

**Table A3.4.** Inter-residual NOE restraints used for CS-D I structure calculation and result. Values are reported in nm.

Inter-residual pair	restraints	Result
2:H2->3:H4	0.24	0.86
2:H2->3:H1	0.24	0.46
2:H3->3:H4	0.25	0.59
2:H4->3:H4	0.25	0.72
2:H5->3:H4	0.23	0.51
3:H1->2:H1	0.23	0.36
3:H2->2:H2	0.23	0.68
3:H3->2:H2	0.23	0.71
3:H5->2:H2	0.23	0.67
1:H1->2:H1	0.38	0.43
1:H2->2:H2	0.38	0.5
1:H2->2:H1	0.38	0.43
1:H3->2:H2	0.38	0.44
1:H5->2:H2	0.38	0.66
1:H5->2:H1	0.38	0.43
2:H2->1:H2	0.4	0.5
2:H2->1:H5	0.4	0.66
2:H2->1:H1	0.4	0.66
2:H2->1:H4	0.4	0.5
2:H3->1:H1	0.43	0.68
2:H3->1:H2	0.43	0.65
2:H3->1:H5	0.43	0.7
2:H3->1:H4	0.43	0.61
2:H4->1:H1	0.43	0.83
2:H4->1:H2	0.43	0.73
2:H4->1:H5	0.43	0.77
2:H4->1:H4	0.38	0.59
2:H5->1:H2	0.38	0.67
2:H5->1:H5	0.38	0.58
2:H5->1:H1	0.38	0.65
2:H5->1:H4	0.38	0.47
3:H1->4:H1	0.38	0.43
3:H2->4:H2	0.38	0.48

3:H3->4:H2	0.38	0.43
3:H5->4:H2	0.38	0.66
4:H2->3:H1	0.4	0.64
4:H2->3:H4	0.4	0.51
4:H3->3:H2	0.38	0.64
4:H3->3:H1	0.38	0.66
4:H3->3:H4	0.38	0.61
4:H4->3:H1	0.43	0.83
4:H4->3:H2	0.43	0.72
4:H4->3:H3	0.43	0.58
4:H4->3:H4	0.43	0.61
4:H4->3:H5	0.43	0.78
4:H5->3:H2	0.38	0.67
4:H5->3:H1	0.38	0.66
4:H5->3:H4	0.38	0.48

**Table A3.5.** Inter-residual “no-NOE” restraints used for CS-D I structure calculation and result. Values are reported in nm.

Inter-residual pair	restraints	Result
2:H2->1:H3	0.35	0.43
2:H2->1:H4	0.36	0.44
2:H3->1:H3	0.37	0.51
2:H5->1:H3	0.34	0.38
2:H4->1:H4	0.51	0.54
1:H3->2:H1	0.25	0.26
2:H1->1:H4	0.48	0.43
2:H2->3:H1	0.41	0.49
2:H2->3:H5	0.48	0.6
2:H3->3:H1	0.39	0.44
2:H5->3:H1	0.35	0.4
2:H4->3:H5	0.35	0.39
2:H4->3:H1	0.22	0.22
2:H4->3:H4	0.55	0.58
4:H2->3:H3	0.35	0.44
4:H2->3:H2	0.35	0.52
4:H2->3:H4	0.43	0.47
3:H3->4:H1	0.27	0.25
3:H2->4:H1	0.27	0.37
4:H1->3:H4	0.54	0.44
1:HN->2:H1	0.3	0.26
1:HN->2:H2	0.5	0.55
1:HN->2:H3	0.48	0.51

1:HN->2:H5	0.43	0.43
1:HN->2:H4	0.66	0.65
3:HN->4:H1	0.27	0.23
3:HN->4:H2	0.52	0.53
3:HN->2:H2	0.76	0.72
3:HN->4:H3	0.48	0.47
3:HN->4:H4	0.61	0.62
3:HN->2:H3	0.7	0.59
3:HN->4:H5	0.41	0.4
3:HN->2:H5	0.51	0.47
3:HN->2:H4	0.41	0.44
2:H2->1:H3	0.39	0.43
2:H2->1:H4	0.42	0.44
2:H3->1:H3	0.44	0.51
2:H5->1:H3	0.35	0.38
1:H3->2:H1	0.25	0.26
1:H4->2:H1	0.42	0.43
2:H4->1:H4	0.61	0.54
2:H1->1:H4	0.45	0.43
2:H2->3:H1	0.44	0.49
2:H2->3:H5	0.51	0.6
2:H3->3:H1	0.47	0.44
2:H5->3:H1	0.42	0.4
2:H4->3:H5	0.37	0.39
2:H4->3:H1	0.25	0.22
3:H4->2:H4	0.6	0.58
4:H2->3:H3	0.42	0.44
4:H2->3:H2	0.42	0.52
3:H3->4:H1	0.26	0.25
3:H2->4:H1	0.26	0.37
3:H4->4:H1	0.46	0.44
3:H4->4:H2	0.52	0.47
3:H4->4:H5	0.57	0.56

**Table A3.6.** Inter-residual NOE restraints used for CS-E structure calculation and result. Values are reported in nm.

Inter-residual pair	restraints	Result
2:H2->1:H3	0.35	0.43
2:H2->1:H4	0.36	0.44
2:H3->1:H3	0.37	0.51
2:H5->1:H3	0.34	0.38
2:H4->1:H4	0.51	0.54

4:H2->3:H1	0.46	0.67
4:H2->3:H5	0.46	0.65
2:H2->1:H1	0.46	0.68
2:H2->3:H4	0.46	0.8
4:H3->3:H1	0.46	0.7
4:H3->3:H2	0.46	0.61
4:H3->3:H3	0.46	0.51
4:H3->3:H4	0.46	0.66
4:H3->3:H5	0.46	0.76
4:H4->3:H1	0.46	0.78
4:H4->3:H2	0.46	0.72
4:H4->3:H3	0.46	0.53
4:H4->3:H4	0.46	0.59
4:H4->3:H5	0.46	0.74
2:H3->1:H1	0.46	0.72
2:H3->1:H4	0.46	0.64
2:H3->3:H4	0.46	0.73
2:H3->3:H5	0.46	0.6
4:H5->3:H1	0.46	0.57
4:H5->3:H4	0.46	0.56
4:H5->3:H5	0.46	0.61
2:H5->1:H1	0.46	0.58
2:H5->1:H4	0.46	0.54
2:H5->3:H4	0.46	0.78
2:H5->3:H5	0.46	0.64
2:H4->1:H1	0.46	0.76
2:H4->1:H3	0.46	0.51
2:H4->1:H4	0.46	0.54
2:H4->3:H4	0.46	0.58
3:H5->2:H1	0.46	0.78
3:H5->4:H1	0.46	0.5

**Table A3.7.** Inter-residual “no-NOE” restraints used for CS-E structure calculation and result. Values are reported in nm.

*Chapter 4*

## NEURITOGENIC ACTIVITY OF CS-D AND CS-E

## Chapter 4

### NEURITOGENIC ACTIVITY OF CS-D AND CS-E

#### *Introduction*

As mentioned in chapter 1, CSs are the most abundant GAGs in brain and play critical roles in the developing CNS and in CNS injury. Paradoxically, CSs are found to exhibit both stimulatory and inhibitor effect on neuronal growth, however, the seeming conflicting results appear to be context-dependent. Factors include the neuronal lineage, neuronal age, expression of specific receptors, extracellular growth factors, and presentation of the CSs may all contribute to the observed effect.

Although CSs are traditionally thought as inhibitory cues to neuronal outgrowth or regeneration, they are also observed in developing axon pathways and in tissues that do not exclude axon entry<sup>1</sup>. Numerous studies also suggested that CSs can act as stimulatory cues to neuronal outgrowth for a range of neurons *in vitro*. Furthermore, the outgrowth promoting activity of CSs is sulfation pattern dependent. The idea of “sulfation code” was therefore proposed.

CSs in the form of CSPGs or polysaccharides enriched with a specific sulfation pattern has been shown to promote the outgrowth of embryonic neurons<sup>2-9</sup>. For example, the DSD-1-PG containing the CS epitope DSD-1 promotes neurite outgrowth from rat embryonic day 14 (E14) mesencephalic and E18 hippocampal neurons<sup>2-3</sup>. It has been further demonstrated that the promotion is DSD-1 dependent, because it can be blocked specifically by the monoclonal antibody (mAb) 473HD the only recognizes DSD-1 but not the core protein or by enzymatic digestion of the CS with chondroitinases ABC<sup>3,6</sup>. While polysaccharide enriched with the CS-D motif blocked the interaction

between mAb 473HD<sup>6,8-9</sup>, the CS-D recognizing mAb CS-56 and mAb MO-225 also bind to DSD-1<sup>10</sup>, suggesting that the DSD-1 epitope contains the D sulfation pattern. Indeed, the CS-D enriched polysaccharide also exerted neurite outgrowth activity in rat E18 hippocampal neurons that can be inhibited by mAb 473HD<sup>9</sup>. Polysaccharides enriched with the other disulfated motif, CS-E, also exhibited neurite outgrowth activity in the same setting, but the CS-A and CS-C enriched polysaccharide did not<sup>9</sup>. Importantly, the activity from CS-E enriched polysaccharide was not inhibited by mAb 473HD, suggesting that the CS-D and CS-E enriched polysaccharides function through different structural motifs<sup>9</sup>. Interestingly, although both the CS-D and CS-E enriched polysaccharide stimulated neurite outgrowth, the morphology of the neurons were different. While the CS-D enriched polysaccharides stimulated the outgrowth of dendrite-like neurite, polysaccharides enriched in the other disulfated motif, CS-E, stimulated the elongation of a prominent long axon-like neurite<sup>4</sup>, suggesting again a sulfation pattern specific effect.

While the literature suggests that the outgrowth promotion activity is sulfation pattern dependent, the polysaccharides or PGs used is far from being homogenous. For example, the CS-D and CS-E enriched polysaccharide contained only 20% and 56% of the desired CS-D and CS-E motif, respectively<sup>4</sup>, with the CS-A and CS-C motif being the major undesired component. Although the contribution from the CS-D and CS-E motif can be inferred by the fact that the CS-A and CS-C enriched polysaccharide does not promote neurite outgrowth significantly, the heterogeneity still complicated the detailed analysis.

Our group envisioned that chemically synthesized CS oligosaccharides can answer the urgent need of homogenous material and synthesized CS-A, CS-C, and CS-E tetrasaccharides. We demonstrated that the homogenous CS-E tetrasaccharide, but not the CS-A and CS-C



tetrasaccharides, promotes the neurite outgrowth of hippocampal, cortical, and dopaminergic neurons<sup>11-13</sup>, which further supported the existence of sulfation code.

In this study, we present the neuritogenic effect of the chemically synthesized CS-D and CS-E tetrasaccharides. We demonstrated that the CS-D tetrasaccharide specifically promotes dendritic outgrowth and the CS-E tetrasaccharide selectively promotes axonal outgrowth by the use of dendrite- and axon-specific markers.

### *Neuritogenic activity of the CS-D and CS-E tetrasaccharides*

To explore the abilities of the CS-D and CS-E tetrasaccharides to modulate neuronal growth, primary rat E18 hippocampal neurons were cultured on a substratum coated with poly-D-lysine (control) and the CS-D and CS-E tetrasaccharides. As opposed to previous studies where a short culture time (24 h to 48h) was used, we chose to culture the neuron for 4 days to allow axons and dendrites to be developed. After 4 days *in vitro* (DIV), the neurons were fixed and doubly immunostained with the axon-specific anti-Tau and dendrite-specific anti-microtubule associated protein 2 (MAP2) antibodies and examined by confocal microscopy.

As shown in figure 3.1 and 3.2, the CS-D and CS-E tetrasaccharide exhibited distinct neurite outgrowth promoting activity. Total dendritic length of neurons cultured on CS-D tetrasaccharide substratum was increased by  $83.0 \pm 1.2\%$ , relative to the poly-D-lysine control. The CS-D tetrasaccharide also enhanced the formation of a greater number of dendrites by  $42.8 \pm 2.0\%$ . Surprisingly, the CS-D did not exhibit significant effect on axonal outgrowth, suggesting that the CS-D sulfation motif specifically stimulated dendritic growth. On the other hand, although the CS-E tetrasaccharide stimulated dendritic growth by  $24.7 \pm 2.6\%$ , it preferentially simulated the axonal outgrowth with much greater potency, by a relative increase of  $72.1 \pm 2.2\%$ . Also, the CS-E did not

exert significant effect on dendrite number. While the CS-D and CS-E tetrasaccharides specifically promotes dendritic development and preferentially stimulated axonal growth, respectively, the CS-D disaccharide showed no appreciable neuritogenic activity, demonstrating that a minimum length of tetrasaccharide is required for activity.

This is the first study that demonstrated direct evidence of CS exhibiting dendrite- and axon-specific activity with the immunostaining of dendrite- and axon-specific marker. This is also the first time a tetrasaccharide with homogeneous D pattern has been used in biological studies, which allowed us to observe activities that cannot be observed with just a disaccharide with D motif. The synthetic CS-D tetrasaccharide is also particularly valuable, since the available CS-D enriched polysaccharide has only 20% of the CS-D motif<sup>4</sup> and may not allow exact determination of biological activities or observation of subtle property difference.

Our result clearly indicated that specific sulfate patterns are required to modulate specific activity of CS. In the previous study, we demonstrated that the sulfation pattern, not total negative charges, dictates the function of CS by showing that an artificial tetrasulfated tetrasaccharide CS-R have no neuritogenic activity while the equally charged CS-E tetrasaccharide promoted neurite outgrowth<sup>12</sup>. The current work further conformed the observation, since the CS-D and CS-E tetrasaccharides also have the same amount of charges.

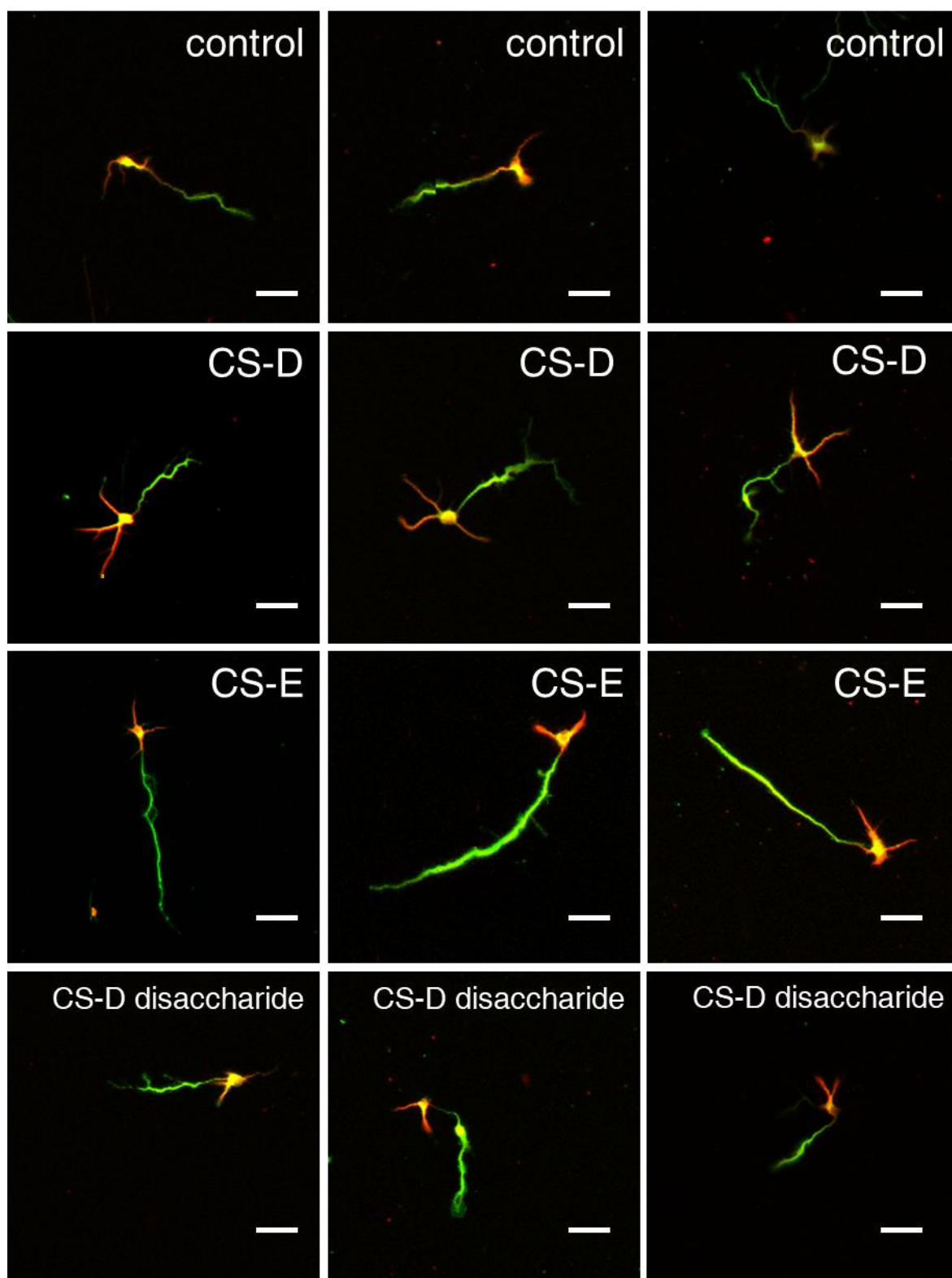


Figure 4.1. Representative immunofluorescence images of hippocampal neurons cultured on a substratum coated with poly-D-lysine (control) and the indicated synthetic tetrasaccharides (CS-D or CS-E) or synthetic CS-D disaccharide. Scale bar, 50  $\mu$ m. Neurons were fixed at 4 DIV and stained with anti-MAP-2 (red) and anti-Tau (green).

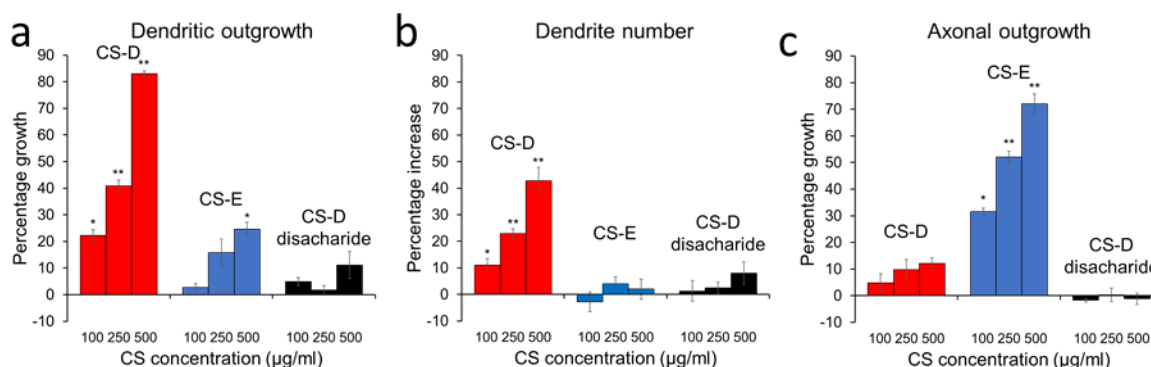


Figure 3.2. Quantitative analysis of neuritogenic activity of CS-D and CS-E. (a) The CS-E tetrasaccharide, but not the CS-D tetrasaccharide, stimulated the axonal outgrowth of embryonic hippocampal neurons. (b) The CS-D tetrasaccharide stimulated the dendritic outgrowth of hippocampal neurons, with greater potency than the CS-E tetrasaccharide. (c) The CS-D tetrasaccharide induced an increase in the number of dendrites, whereas the CS-E tetrasaccharide had no effect on dendrite number. Data represent the mean  $\pm$  SEM from three experiments and are expressed as the percentage growth or dendrite number relative to the poly-D-lysine control.  $n = 80$ – $200$  cells per experiment. Statistical analyses were performed using the one-way ANOVA test (\* $P < 0.005$ , \*\* $P < 0.0005$ .)

While having the same amount of negative charges, CSs with different sulfation patterns favor distinct structural conformations and present unique electrostatic surface for interaction with proteins, which may account for how the information encoded in the sulfation pattern is translated in cells. Indeed, our previous microarray study showed that the CS-E tetrasaccharide binds to the growth factors midkine and brain-derived neurotrophic factor and further demonstrated the abolishment of neuritogenic activity of CS-E by addition of the function-blocking antibodies against these growth factors or their cell surface receptors TrkB and receptor-type protein-tyrosine phosphatase zeta (RPTP $\beta$ ), respectively. The allyl handle on the anomeric end of the CS-D and CS-E tetrasaccharides allow the tetrasaccharides to be coated on microarrays. We expect microarrays displaying the tetrasaccharides to discover novel CS-protein interactions and account for the observed neuritogenic activity.

The ability of the CS-D and CS-E tetrasaccharide to stimulate dendritic and axonal outgrowth of hippocampal neurons suggests that CS may play critical roles in the developing CNS. Indeed, both CS-D and CS-E is expressed and spatiotemporally regulated in brains of embryonic and newborn mice<sup>14-15</sup>. The present work also supports previous studies that CS-D and CS-E studies implicate the function of CS-D and CS-E motif in the growth and development of neurons.

## **Conclusion**

With the synthetic tetrasaccharides, we demonstrated the dendrite- and axon-specific stimulatory effect of CS-D and CS-E. The lack of function of the CS-D disaccharide suggested that a minimum length of tetrasaccharide is required. The findings indicate that the CS-D and CS-E may potentially play important roles in developing nervous system *in vivo*. We envision that future protein binding studies might provide more insight to the molecular mechanism of the observed neuritogenic activities.

## ***Experimental methods***

### **Buffers and Reagents**

Cell culture media was purchased from Gibco BRL (Grand Island, NY). Reagents and poly-D-lysine coated plates were purchased from Fisher (Fairlawn, NJ). Anti-Tau antibody (MAB3420) and anti-MAP2 antibody (AB5622) were purchased from Millipore (Darmstadt, Germany). Secondary antibodies were purchased from Thermo Fisher Scientific (Waltham, MA).

### **Calibration of CS concentrations**

The relative concentrations of the CS oligosaccharides were calibrated by using the carbazole assay for uronic acid residues. Briefly, the acid borate reagent (1.5 mL of a solution of 0.80 g sodium

tetraborate, 16.6 mL H<sub>2</sub>O, and 83.3 mL sulfuric acid) was added to glass vials. The oligosaccharides were added (50  $\mu$ L of a 10 mg/mL stock in H<sub>2</sub>O) and the solution placed in a boiling H<sub>2</sub>O bath for 10 min. Following addition of the carbazole reagent (50  $\mu$ L of 0.1% w/v carbazole in 100% ethanol), the solution was boiled for 15 min. The absorbance was read at 530 nm and compared to a D-glucuronolactone standard in H<sub>2</sub>O.

### **Hippocampal Neuronal Cultures**

Hippocampal neuronal cultures were prepared using a modified version of the Goslin and Banker protocol. Embryos at the E18 stage were obtained from timed-pregnant Sprague-Dawley rats, and the hippocampus from each embryo was dissected. All the hippocampi from one preparation were transferred to a 15 mL conical tube containing 4.5 mL of ice-cold Calcium and Magnesium Free-Hank's Balanced Salt Solution (CMF-HBSS). Trypsin (2.5%, no EDTA) was added to 5 mL, and the tissue was digested for 15 min at 37 °C. The trypsin solution was removed without affecting the tissue. The tissue was then rinsed with 5 mL of CMF-HBSS three times. Next, the tissue was dissociated in 1 mL of CMF-HBSS by passing through a P1000 pipet tip twenty times. The cells were counted with a hemacytometer and plated at 20 cells/mm<sup>2</sup> on poly-D-lysine coated plates and cultivated in minimal Neurobasal supplemented with the GS21 mixture. The cultures were maintained in 5% CO<sub>2</sub> at 37 °C for 4 days. Poly-D-lysine coated plates were precoated with compounds in PBS overnight at 37 °C/ 5% CO<sub>2</sub>. The coverslips were then washed three times with PBS.

### **Immunocytochemistry of hippocampal neuronal cultures**

After 4 days in culture, hippocampal neurons were used for immunostaining. Cells were rinsed one time with PBS, fixed in 4% paraformaldehyde for 20 min at rt, washed twice with PBS, permeabilized in 0.2% Triton X-100 for 15 min at rt, and washed three times with PBS. Non-specific

binding was blocked with 2.5% NGS for 1 h at rt. The blocking solution was rinsed off one time with PBS. Cells were then incubated with anti-tau antibody (1:800) and anti-MAP2 antibody (1:1500) with 2.5% NGS in 0.02% Triton X-100 overnight at 4 °C. Excess antibody was rinsed away 5 times with PBS. Secondary antibody anti-rabbit AlexaFluor 546 (1:1000) and anti-mouse AlexaFluor 488 (1:1000) was used to fluorescently label the neurons for 1 hr at rt. Excess secondary antibody was washed off 5 times with PBS. Cells were then subjected to confocal laser microscopy.

### **Morphometric Analysis**

The neurons were imaged with Zeiss LSM 5 Pascal inverted confocal microscope (Carl Zeiss Microscopy). The images were then subjected to MetaMorph (Molecular Devices) software for quantification.

### **References**

- (1) Bicknese, A. R.; Sheppard, A. M.; Oleary, D. D. M.; Pearlman, A. L., Thalamocortical Axons Extend Along a Chondroitin Sulfate Proteoglycan-Enriched Pathway Coincident with the Neocortical Subplate and Distinct from the Efferent Path. *Journal of Neuroscience* **1994**, *14* (6), 3500.
- (2) Garwood, J.; Schnadelbach, O.; Clement, A.; Schutte, K.; Bach, A.; Faissner, A., DSD-1-proteoglycan is the mouse homolog of phosphacan and displays opposing effects on neurite outgrowth dependent on neuronal lineage. *Journal of Neuroscience* **1999**, *19* (10), 3888.
- (3) 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* (3), 783.
- (4) Hikino, M.; Mikami, T.; Faissner, A.; Vilela-Silva, A. C.; Pavao, M. S.; Sugahara, K., Oversulfated dermatan sulfate exhibits neurite outgrowth-promoting activity toward embryonic mouse hippocampal neurons: implications of dermatan sulfate in neuritogenesis in the brain. *Journal of Biological Chemistry* **2003**, *278* (44), 43744.
- (5) Nandini, C. D.; Mikami, T.; Ohta, M.; Itoh, N.; Akiyama-Nambu, F.; Sugahara, K., Structural and functional characterization of oversulfated chondroitin sulfate/dermatan sulfate hybrid chains

from the notochord of hagfish. Neuritogenic and binding activities for growth factors and neurotrophic factors. *Journal of Biological Chemistry* **2004**, 279 (49), 50799.

(6) Clement, A. M.; Nadanaka, S.; Masayama, K.; Mandl, C.; Sugahara, K.; Faissner, A., The DSD-1 carbohydrate epitope depends on sulfation, correlates with chondroitin sulfate D motifs, and is sufficient to promote neurite outgrowth. *Journal of Biological Chemistry* **1998**, 273 (43), 28444.

(7) Bao, X.; Mikami, T.; Yamada, S.; Faissner, A.; Muramatsu, T.; Sugahara, K., Heparin-binding growth factor, pleiotrophin, mediates neuritogenic activity of embryonic pig brain-derived chondroitin sulfate/dermatan sulfate hybrid chains. *Journal of Biological Chemistry* **2005**, 280 (10), 9180.

(8) Nadanaka, S.; Clement, A.; Masayama, K.; Faissner, A.; Sugahara, K., Characteristic hexasaccharide sequences in octasaccharides derived from shark cartilage chondroitin sulfate D with a neurite outgrowth promoting activity. *Journal of Biological Chemistry* **1998**, 273 (6), 3296.

(9) Clement, A. M.; Sugahara, K.; Faissner, A., Chondroitin sulfate E promotes neurite outgrowth of rat embryonic day 18 hippocampal neurons. *Neurosci Lett* **1999**, 269 (3), 125.

(10) Ito, Y.; Hikino, M.; Yajima, Y.; Mikami, T.; Sirko, S.; von Holst, A.; Faissner, A.; Fukui, S.; Sugahara, K., Structural characterization of the epitopes of the monoclonal antibodies 473HD, CS-56, and MO-225 specific for chondroitin sulfate D-type using the oligosaccharide library. *Glycobiology* **2005**, 15 (6), 593.

(11) Tully, S. E.; Mabon, R.; Gama, C. I.; Tsai, S. M.; Liu, X.; Hsieh-Wilson, L. C., A chondroitin sulfate small molecule that stimulates neuronal growth. *Journal of the American Chemical Society* **2004**, 126 (25), 7736.

(12) Gama, C. I.; Tully, S. E.; Sotogaku, N.; Clark, P. M.; Rawat, M.; Vaidehi, N.; Goddard, W. A., 3rd; Nishi, A.; Hsieh-Wilson, L. C., Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Biol* **2006**, 2 (9), 467.

(13) Sotogaku, N.; Tully, S. E.; Gama, C. I.; Higashi, H.; Tanaka, M.; Hsieh-Wilson, L. C.; Nishi, A., Activation of phospholipase C pathways by a synthetic chondroitin sulfate-E tetrasaccharide promotes neurite outgrowth of dopaminergic neurons. *Journal of Neurochemistry* **2007**, 103 (2), 749.

(14) Ishii, M.; Maeda, N., Oversulfated Chondroitin Sulfate Plays Critical Roles in the Neuronal Migration in the Cerebral Cortex. *Journal of Biological Chemistry* **2008**, 283 (47), 32610.



- (15) Ishii, M.; Maeda, N., Spatiotemporal expression of chondroitin sulfate sulfotransferases in the postnatal developing mouse cerebellum. *Glycobiology* **2008**, *18* (8), 602.