

CHEMOENZYMATIC SYNTHESIS AND ANALYSIS OF CHONDROITIN SULFATE
GLYCOPEPTIDES AND PROTEOGLYCANS

By

Po-Han Lin

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry-Doctor of Philosophy

2024

ABSTRACT

Glycosaminoglycans (GAGs) or mucopolysaccharides are a class of highly negatively charged polysaccharides, which are present in all mammalian cells and involved in various important biological events such as growth factor signaling, blood coagulation, brain development and neural stem cell migration. GAGs, except for hyaluronic acid, are commonly found covalently linked to proteoglycans (PGs). These PGs typically start with a common tetrasaccharide linkage region (GlcA β 1-3Gal β 1-3Gal β 1-4Xyl) that links to a serine residue on the core protein. The tetrasaccharide linkage region is then elongated to polysaccharides composed of disaccharide repeating units of chondroitin sulfate (CS) or heparan sulfate (HS).

In this dissertation, we first discuss the recent advances in the understanding of biosynthetic enzymes involved in the synthesis of the tetrasaccharide linkage region in chapter 1 and focus on the expression and the relationship between these enzymes. Building on the knowledge of these enzymes, in chapter 2, we report a new method using commercially available EAH Sepharose as the solid support to synthesize tetrasaccharide linkage region bearing glycopeptides. This method eliminates the tedious traditional glycosyl amino acid synthesis, greatly reducing the time need for tetrasaccharide linkage region construction. We further modified the synthesized linkage region with CS transferase and sulfotransferase to complete the synthesis of multiple CS-bearing glycopeptides. Notably, the dodecasaccharide bearing glycopeptide is the largest GAG-bearing glycopeptide synthesized to date. CSPG plays important roles in biological events, to better understand the mechanism, it is important to map the expression level and the types of CSPG. In chapter 3, an enzymatic method was developed to label and enrich CSPG in biospecimen and various fragmentation methods have been explored for quantitative CSPG mapping.

Copyright by
PO-HAN LIN
2024

ACKNOWLEDGEMENTS

As I write this acknowledgment, I realize that I am finally reaching the end of my PhD journey. Looking back on these 5.5 years, I am filled with gratitude for my PhD advisor, Dr. Xuefei Huang. I could never have completed this program without his continuous and patient guidance. Dr. Huang not only provided financial support but also set an example as an advisor. The freedom he granted me allowed me to learn various techniques and complete this thesis.

I would also like to express my gratitude to my second reader, Dr. Jetze Tepe, and committee members Dr. Jian Hu and Dr. Kevin Walker. Their strictness during my first committee meeting served as a wake-up call. Their advice and comments have been vital in my growth as an independent researcher.

Special thanks go to my collaborators, Dr. Jian Liu, Dr. Jon Amster, Dr. Angela Wilson, Dr. Jesús Jiménez Barbero, Dr. Junfeng Ma, and their students. This thesis would not be complete without their efforts. I want to acknowledge Dr. Dan Homles, Dr. Xie Li from the MSU NMR facility, Dr. Tony Schillmiller from the MS facility, and Doug Whitten from the proteomic facility. I have learned a lot about NMR and MS from them, and the data you are about to see owes much to their expertise.

I express my appreciation to my former and current lab mates, always helpful in the lab and providing valuable advice. I am also grateful that they allowed me to steal their glassware and tubes.

Thanks to my family and friends for their love and support; they are the reason I never gave up. Special thanks to my comrade Kamel Meguellati, who always encouraged me to think outside the box. We shall meet again at TSRI. Lastly, to my Miko, thanks for being cute throughout this whole journey.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	vi
Chapter 1. Recent Advance On Glycosyltransferases Involved In The Tetrasaccharide Linkage	
Region Synthesis	1
1.1 Introduction	1
1.2 Family With Sequence Similarity 20 Member B	3
1.3 β -1,3-Galactosyltransferase 6	6
1.4 Phosphoxylose Phosphatase	9
1.5 Glucuronyltransferase I	11
1.6 Outlook	15
REFERENCES	16
Chapter 2. Expedient Solid Phase Supported Chemo-Enzymatic Synthesis of Chondroitin Sulfate	
Proteoglycan Glycopeptides	22
2.1 Introduction	22
2.2 Results and Discussion.....	24
2.3 Outlook	38
2.4 Experimental Section	39
REFERENCES	68
APPENDIX A: SUPPLEMENTARY FIGURES, SCHEMES AND TABLES	74
APPENDIX B: PRODUCT CHARACTERIZATION SPECTRA	81
Chapter 3. Comprehensive Mapping of CSPG in Biological Samples Using a Chemo-Enzymatic	
Method	198
3.1 Introduction	198
3.2 Results and Discussion	201
3.3 Outlook	212
3.4 Experimental Section.....	213
REFERENCES	218
APPENDIX: SUPPLEMENTARY FIGURES, SCHEMES AND TABLES	221

LIST OF ABBREVIATIONS

ATP: Adenosine Triphosphate

BAV: Bicuspid Aortic Valve

β 3GALT6: β -1,3-Galactosyl Transferase 6

β 4GALT7: β -1,4-Galactosyl Transferase 7

β 3GAT3: β -1,3-Glucuronic Acid Transferase 3

BLI: Biolayer Interferometry

CatG: Cathepsin G

Chpf: Chondroitin Polymerizing Factor

Chpf2: Chondroitin Polymerizing Factor 2

Chsy1: Chondroitin Sulfate Synthase 1

CMP: Cytosine monophosphate

CPG: Controlled Pore Glass

CS: Chondroitin Sulfate

CS-A: Chondroitin Sulfate A

Csgalnact2: Chondroitin Sulfate N-acetylgalactosaminyltransferase 2

CSPG: Chondroitin Sulfate Proteoglycan

CuAAC: Copper-Catalyzed Azide-Alkyne Cycloaddition

CZE-FT-ICR MS: Capillary Zone Electrophoresis-Fourier Transform-Ion Cyclotron Resonance

Mass Spectrometry

DBCO: Dibenzocyclooctyne

DIPEA: *N,N*-Diisopropylethylamine

DS: Dermatan Sulfate

DTT: Dithiothreitol

Ext1: Exostosin Glycosyltransferase 1

Ext2: Exostosin Glycosyltransferase 2

Extl2: Exostosin-like Glycosyltransferase 2

Extl3: Exostosin-like Glycosyltransferase 3

FAM20B: Family With Sequence Similarity 20 Member B

ff14SB: Force Field 14SB

GAG: Glycosaminoglycan

Gal: Galactose

GalA: Galacturonic Acid

GBVA/WSA: Generalized-Born Volume Integral / Weighted Surface Area

Glc: Glucose

GlcA: Glucuronic Acid

GlcAT-I: Glucuronyltransferase I

GlcNAc: *N*-Acetylglucosamine

GlcNAc: Uridine Diphosphate *N*-Acetylglucosamine

GlcA-pNp: Glucuronic Acid-*p*-Nitrophenyl

GDP: Guanosine Diphosphate

HPLC: High-Performance Liquid Chromatography

HS: Heparan Sulfate

HSPG: Heparan Sulfate Proteoglycan

HEK293F: Human Embryonic Kidney 293F cells

HEPES: 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic Acid

IPTG: Isopropyl β -D-1-thiogalactopyranoside

KfoC: Chondroitin Polymerase

LC/MS/MS: Liquid Chromatography-Mass Spectrometry

Man: Mannose

MES: 2-(N-Morpholino)ethanesulfonic Acid

MOE: Molecular Operating Environment

MS/MS: Tandem Mass Spectrometry

NHS-LC-Biotin: *N*-Hydroxysuccinimide long-chain biotin

NMR: Nuclear Magnetic Resonance

PAPS: 3'-Phosphoadenosine-5'-Phosphosulfate

PEGA: Polyethylene Glycol Polyacrylamide Copolymer

PG: Proteoglycan

SA: Streptavidin

SAX: Strong Anion Exchange

TEABC: Triethylammonium bicarbonate

TM: Thrombomodulin

TOF: Time of Flight

UDP: Uridine Diphosphate

UDP-GalNAc: Uridine Diphosphate *N*-Acetyl Galactosamine

UDP-GalNAc: Uridine diphosphate *N*-acetylglucosamine

UDP-GalNAz: Uridine Diphosphate *N*-Acetylgalactosamine Azide

UDP-GlcA: Uridine Diphosphate Glucuronic Acid

UDP-GlcNAz: Uridine Diphosphate *N*-Acetylglucosamine Azide

Xyl: Xylose

XT-1: Xylosyl Transferase-1

XYLP: 2-Phosphoxylose Phosphatase

YPD: Yeast Extract Peptone Dextrose

Chapter 1. Recent Advance On Glycosyltransferases Involved In The Tetrasaccharide

Linkage Region Synthesis

1.1 Introduction

Glycosaminoglycans (GAGs), also known as mucopolysaccharides, are predominantly found in vertebrates. They encompass heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), keratan sulfate, and hyaluronic acid (HA). Most of the GAGs except for HA are presented on proteoglycans (PGs), which are located in the extracellular matrix and on cell surface^{1, 2}. The biological activities of PGs are diverse, including growth factor signaling, wound repair, blood coagulation, brain development, and neural stem cell migration³⁻⁷.

The structure of PGs consists of a protein with an “SA” or “SG” dipeptide motif⁸, a tetrasaccharide linkage region (glucuronic acid (GlcA)- β (1 $\rightarrow$ 3)-galactose (Gal)- β (1 $\rightarrow$ 3)-galactose (Gal)- β (1 $\rightarrow$ 4)-xylose (Xyl)- β (1 $\rightarrow$ O)-serine (Ser)), and repeating disaccharide units. The synthesis of the tetrasaccharide linkage region involves six enzymatic steps (fig.1). Xylosyltransferase-1 (XT-1) recognizes a peptide or protein substrate with an SG or SA motif and transfers a xylose in β configuration to the serine residue⁸. Subsequently, β -1,4-galactosyltransferase 7 (β 4GALT7) transfers a galactose residue to the 4-position of xylose^{9, 10}, followed by the phosphorylation of the 2-position of xylose by kinase Family with Sequence Similarity 20, Member B (FAM20B). This phosphate group significantly enhances the yield of the next enzymatic step involving β -1,3-galactosyltransferase 6 (β 3GALT6), which transfers the second galactose to the disaccharide moiety^{11, 12}. The absence of this phosphate prevents the synthesis of the tetrasaccharide linkage region, resulting in a trisaccharide linkage region instead¹³. At this point, 2-phosphoxylose phosphatase 1 (XYLP) cleaves the phosphate from xylose¹⁴. Finally, β -1,3-glucuronyltransferase

3 (β 3GAT3) adds a glucuronic acid to the trisaccharide glycopeptide, completing the tetrasaccharide linkage region¹⁵.

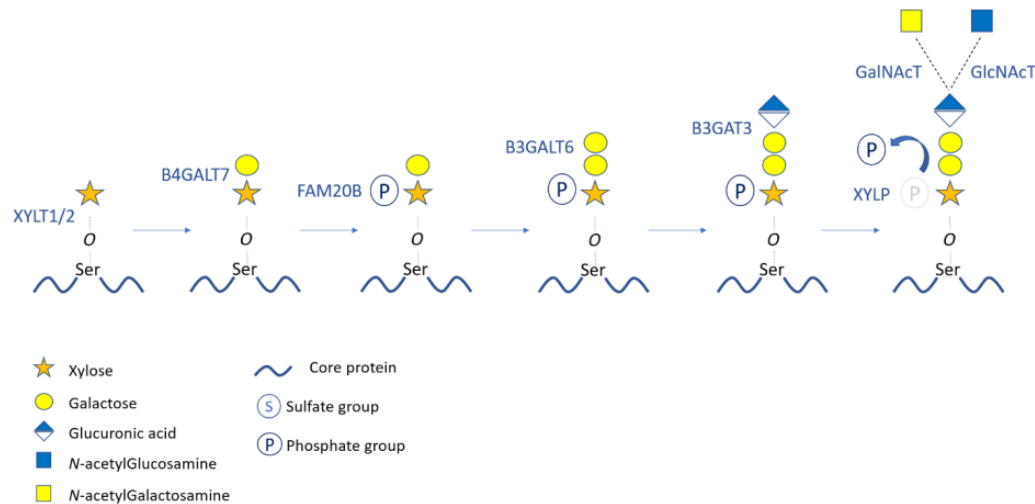


Figure 1.1. Schematic demonstration of the biosynthesis of tetrasaccharide linkage¹⁶.

The complete tetrasaccharide linkage region can be extended by chondroitin sulfate *N*-acetylgalactosaminyltransferase 1 (Csgalact2), chondroitin Sulfate *N*-acetylgalactosaminyltransferase 2 (Csgalact2), or by exostosin-like glycosyltransferase 2 (Extl2), exostosin-like glycosyltransferase 3 (Extl3)¹⁷⁻²⁰. Pentasaccharides containing *N*-acetylgalactosamine (GalNAc) are further elongated by bifunctional enzymes like chondroitin sulfate synthase 1 (Chsy1), Chondroitin Polymerizing Factor (Chpf), and Chondroitin Polymerizing Factor 2 (Chpf2), resulting in polymerized CS chain as part of CSPG¹⁸. Similarly, pentasaccharides containing *N*-acetylglucosamine (GlcNAc) are elongated by exostosin glycosyltransferase 1 (Ext1) or exostosin glycosyltransferase 2 (Ext2), leading to polymerized HS chain for HSPG²⁰. The length, sulfation, and epimerization pattern of these repeating disaccharide units can naturally vary, resulting in a highly heterogeneous glycan pattern².

While previous research primarily focused on the glycan part of PGs viewing the protein mainly as a carrier of GAGs in biological events, recent studies from our group²¹ and others²²⁻²⁴ demonstrate that both the protein and GAGs can actively participate in protein binding. Hence, it is imperative to consider both the protein and glycan in biological studies, necessitating the use of structurally well-defined homogeneous PGs or glycopeptides. However, due to the inherent heterogeneity of PGs, the isolation of homogeneous PGs or glycopeptides remains unattainable.

As tetrasaccharide linkage region is an important bridge between the core protein and GAGs, it is necessary to review the enzymes involved in its biosynthesis. Over the years, there are several reviews about these enzymes^{25, 26}. Recently, a review covered the efforts made in xylosyltransferase I (XT-1) and beta-1,4-galactosyltransferase 7 (B4GALT7)²⁷. In this review, we will focus on the recent progress made in the expression and substrate specificity of other enzymes responsible for tetrasaccharide linkage region synthesis. The relationships between these enzymes will also be discussed.

1.2 Family With Sequence Similarity 20 Member B

Family with Sequence Similarity 20 Member B (FAM20B) is a protein belonging to the FAM20 family, which plays a significant role in various biological processes²⁸⁻³⁰. The FAM20 family consists of three proteins, namely FAM20A, FAM20B, and FAM20C. FAM20C functions as a Golgi casein kinase, phosphorylating SxE/pS motifs. FAM20A acts as a pseudo kinase that interacts with FAM20C, enhancing its activity³¹. Conversely, FAM20B functions as a xylosylkinase, phosphorylating the 2-position of xylose in the tetrasaccharide linkage region, making it crucial for GAG synthesis¹¹.

Studies investigating the depletion of FAM20B have revealed its essential role in GAG biosynthesis. When FAM20B is depleted or dysfunctional, it can lead to abnormalities in cartilage matrix organization, early-stage chondrocyte development, development of supernumerary teeth, chondrosarcoma with major postnatal ossification defects, and severe craniofacial defect³²⁻³⁷. Understanding the consequences of

FAM20B depletion provides valuable insights into the mechanisms underlying these developmental disorders and underscores the significance of FAM20B in GAG synthesis and its relationship with these disorders.

1.2.1 Expression of FAM20B

The discovery of FAM20 series dated back to 2005³⁸, but the expression of human FAM20B was achieved later. In 2009, Kitagawa group expressed it with Hela cells using a stable transfection system¹¹. At that time, the relationship between FAM20B and B3GALT6 was not clear and FAM20B was only known as a xylosylkinase.

Shortly after this, in 2011, Kimmel group reported that FAM20B could be expressed *in vivo* when Tol 2 expression plasmids were injected into the embryo of zebrafish³³. In 2013, Kitagawa group again expressed FAM20B but using the African green monkey kidney fibroblast-like cell line (COS-1 cells) with the plasmid inserted with a cleavable (Immunoglobulin G) IgG domain. Upon harvest, the culture medium was incubated with IgG-Sepharose for further purification³⁹.

In 2014, Dixon group expressed FAM20B using Hi5 cells¹². The desired gene containing truncated FAM20B(aa 42-409) was fused to a maltose binding protein (MBP)-6Xhis tag and was then generated as a baculovirus plasmid. Hi5 cells were infected and then incubated for 2 days. The medium was collected and purified with Ni-NTA resin. When necessary, the MBP fusion protein can be cleaved by Tobacco Etch Virus (TEV) protease to obtain pure FAM20B¹². Interestingly, in 2018, Dixon and Xiao group together reported another expression of FAM20B⁴⁰. The expression system was very similar to their 2014 publication with the exception that the protein sequence was changed to 55-402. This change in the sequence does not alter the protein function.

1.2.2 Acceptor Specificity of FAM20B

The substrate specificity of FAM20B was initially documented in 2009 by the Kitagawa group¹¹. Their study employed α -TM (α -thrombomodulin) as the substrate for FAM20B. Notably, α -TM is a glycoprotein found on the surface of endothelial cells⁴¹. Unlike β -TM, the CS variant of thrombomodulin, α -TM only possesses the tetrasaccharide linkage region, which can potentially serve as a substrate for FAM20B. FAM20B recognizes and accepts tetrasaccharide-bearing PG α -TM with a k_{cat} value of 102 pmol/h per mL of medium. When the acceptor was switched to a trisaccharide with serine as the aglycon (Gal β 1-3Gal β 1-4Xyl-Ser), similar kinase activity of 128 pmol/h per mL of medium was observed. This indicates that FAM20B does not require a protein/peptide aglycon for acceptor binding and can readily phosphorylate trisaccharides or tetrasaccharides.

In 2014, the Dixon group conducted a kinetic assay of FAM20B using [γ -³²P]ATP to evaluate its kinetics with various substrates¹². Three substrates were employed: tetra-Bn (GlcA1-3Gal β 1-3Gal β 1-4Xyl β 1-Bn), Gal-Xyl-Bn (Gal β 1-4Xyl β 1-Bn), and Xyl-Bn (Xyl β 1-Bn), all containing a β -benzyl (Bn) group as the aglycon of the substrate. Both tetra-Bn and the disaccharide Gal-Xyl-Bn served as the substrate with K_m values of 40 μ M and 42 μ M, respectively. In contrast, the enzyme did not have much activity toward the monosaccharide Xyl-Bn.

By combining the results from these two studies, it is clear that FAM20B can carry out phosphorylation without a stringent requirement on the aglycon. Instead, it is primarily the number of saccharides that influences its activity. The findings indicate that the substrate must possess a minimum disaccharide motif (Gal-Xyl) for the reaction to occur, implying that galactose potentially plays a crucial role in binding.

1.3 β -1,3-Galactosyltransferase 6

B3GALT6, also referred to as β -1,3-glucuronyltransferase 6, is an enzyme of significant importance in the biosynthesis of the tetrasaccharide linkage region. The discovery and comprehension of B3GALT6 have significantly advanced our understanding of PG biosynthesis and its profound impact on various biological processes.

In 2013, the Ikegawa group conducted a study that reported the connection between B3GALT6 mutations and several types of connective tissue disorders, such as lax skin, muscle hypotonia, joint dislocation, skeletal dysplasia, and deformities⁴²⁻⁴⁴. This investigation highlighted the crucial role of B3GALT6 in the development and maintenance of various tissues, including the skin, bones, cartilage, tendons, and ligaments. The findings from this study shed light on the vital contributions of B3GALT6 to the intricate processes involved in tissue growth, organization, and overall physiological homeostasis. Further research continues to explore the precise mechanisms by which B3GALT6 operates and its implications for understanding and potentially treating connective tissue disorders⁴⁴.

The knowledge on the molecular mechanisms and significance of B3GALT6 has broader implications for areas such as developmental biology, skeletal formation, and tissue homeostasis. Ongoing research continues to explore the precise functions and regulatory mechanisms of B3GALT6, aiming to deepen our knowledge of its contribution to physiological and pathological processes.

1.3.1 Expression of B3GALT6

Although the initial discovery of B3GALT6 occurred in 2001,⁴⁵ the expression of B3GALT6 was achieved much later. In 2014, the Dixon group published a study describing the expression of B3GALT6 utilizing a baculovirus expression system¹². A truncated version of B3GALT6

consisting of amino acids 31 to 329 was inserted into a bacmid along with a His-MBP affinity tag. The constructed bacmid was then transfected into Hi5 cells, and the resulting medium was collected after a 2-day period. The fusion protein was subsequently purified using a Ni-NTA purification method.

To date, the Dixon report remains the sole existing publication detailing the expression of B3GALT6¹². In 2018, the Moremen lab outlined their approach to construct a library of glycosyltransferases and glycoside hydrolases⁴⁶. In their study, a truncated form of B3GALT6 comprising amino acids 35 to 329 was cloned into the mammalian expression vector pGen2-DEST. To enhance solubility, a His tag and avidin tag were introduced at the N-terminus of B3GALT6, and a super-folded green fluorescent protein (GFP) was added between the tags and the desired protein sequence. While there has been no literature reporting the successful expression of this particular plasmid, our research group has successfully expressed this enzyme using the aforementioned construct, which will be elaborated upon in Chapter 2 of this thesis.

1.3.2 Acceptor Specificity of B3GALT6

The acceptor specificity of B3GALT6 exhibited a strong dependence on FAM20B. In Dixon's *in vitro* investigation conducted in 2014¹², it was observed that B3GALT6 displayed minimal activity when tested with Gal-Xyl-Bn as a substrate. Interestingly, when the disaccharide was phosphorylated by FAM20B prior to B3GALT6 activity, the K_m value for this reaction decreased by approximately 230-fold, suggesting the phosphorylated xylose is critical for binding.

Previously, the role of FAM20B remained unclear, with its recognition limited to its kinase activity in phosphorylating xylose. However, in 2014, the study conducted by the Dixon group provided insights into the relationship between FAM20B and B3GALT6¹². They generated FAM20B knockout (KO) human bone osteosarcoma epithelial cells (U2OS). Two distinct methods

were employed to assess the overall GAGs content in the cell lysate. The first method involved the widely used 3G10 antibody, which specifically detects HS. The second method utilized the isotope [^{35}S] to quantify the sulfate content within GAGs. Comparative analysis with wild-type (WT) U2OS cells against the FAM20B KO U2OS cells revealed a significant 95% reduction in GAG levels as detected by the 3G10 antibody, and a similar outcome was observed using the ^{35}S isotope measurement approach. Furthermore, the researchers investigated the linkage region of glypican 1 (GPC1) in the FAM20B KO cells^{12, 47}. Through β -elimination and mass spectrometry analysis, they discovered that instead of the expected tetrasaccharide linkage region, a trisaccharide linkage region comprising Sia α 2–3Gal β 1–4Xyl β 1 was observed. Intriguingly, in 2018, the Yang group conducted a similar experiment by knocking out FAM20B in Chinese hamster ovary cells (CHO)¹³. High-performance liquid chromatography (HPLC) analysis demonstrated a roughly 3-fold decrease in CS and a 6-fold decrease in HS levels compared to those from WT CHO cells. This finding suggests that the impact on GAG levels may vary in different cell lines when specific GAG-associated enzymes are altered. Subsequent experiments involving the KO of B3GALT6 in CHO cells revealed that CS and HS were still detectable, although the overall content of GAG was significantly reduced compared to WT CHO cells. Similarly, in 2018, the Malfait group also reported the decrease of GAGs in patients with mutated B3GALT6 genes⁴⁸. Additionally, in 2019, the Larson group reported the identification of a trisaccharide linkage region (GlcA-Gal-Xyl) within GAGs, indicating that cells are capable of elongating GAGs with an incomplete linkage region⁴⁹. Collectively, these findings suggest that B3GALT6 can only function when the disaccharide (Gal-Xyl) moiety is phosphorylated. However, even in the absence of B3GALT6, GAG synthesis is not completely halted but instead occurring in a lower yield, leading to the formation of a trisaccharide linkage region instead.

1.4 Phosphoxylose Phosphatase

2-Phosphoxylose phosphatase, also known as XYLP, is a phosphatase responsible for dephosphorylation of xylose. It is one of the least studied enzymes among those involved in the linkage region synthesis. To the best of my knowledge, the only systematic study of this enzyme was reported in 2014 by Kitagawa group¹⁴.

1.4.1 Expression of XYLP

In the 2014 study reported by Kitagawa group, a truncated version of XYLP(aa 38-480) was fused to pEF-BOS vector. Transfection reagent FuGENE 6 was used to transfect XYLP into COS-1 cells. Two days post-transfection, medium was collected and purified with IgG-Sepharose beads.

1.4.2 Substrate Specificity of XYLP

The substrate specificity of XYLP was reported in the aforementioned publication¹⁴. The Kitagawa group conducted comparisons using different substrates: a phosphorylated trisaccharide (Gal-Gal-Xyl(2P)-TM) and a tetrasaccharide (GlcA-Gal-Gal-Xyl(2P)-TM), both bearing α -TM. The results showed that only the phosphorylated trisaccharide with α -TM was accepted as a substrate by XYLP with little activities for the tetrasaccharide with α -TM. Furthermore, when comparing GalNAc-type (GalNAc-GlcA-Gal-Gal-Xyl(2P)-TM, a potential precursor to CSPG) and GlcNAc-type (GlcNAc-GlcA-Gal-Gal-Xyl(2P)-TM, a potential precursor to HSPG) compounds, both bearing a phosphorylated pentasaccharide with α -TM, XYLP only dephosphorylated the GalNAc-type compound, while no reaction was observed when GlcNAc-type was used as a substrate.

It is important to note that XYLP does not accept other phosphorylated glycans on glycoproteins such as osteopontin or matrix extracellular phosphoglycoprotein. This observation

suggests that XYL's substrate specificity is limited to 2-phosphoxylose. Interestingly, when alkaline phosphatase (ALP), a commercially available phosphatase, was used, all the substrates mentioned above could be dephosphorylated. In particular, co-expression of XYLP and B3GAT3 led to rapid dephosphorylation of the linkage region. This is because B3GAT3, a transferase responsible for GlcA addition to the trisaccharide, can form oligomers with XYLP to enhance the dephosphorylation activity over 16-folds.

1.4.3 Relationship between FAM20B, B3GALT6 and XYLP

Following the synthesis of the phosphorylated linkage region trisaccharide, β -1,3-glucuronyltransferase 3 (B3GAT3) transfers GlcA to the phosphorylated trisaccharide (Gal-Gal-Xyl(2P)-Ser) linked to a serine residue. Simultaneously, xylose dephosphorylation is initiated by XYLP. During this phase, chondroitin (Chn) or heparan sulfate (HS) polymerases facilitate the polymerization of disaccharide chains onto the tetrasaccharide linkage region. Excessive acceleration of linkage region phosphorylation by FAM20B and/or attenuated Xyl dephosphorylation by XYLP may lead to the accumulation of biosynthetic intermediates, specifically phosphorylated linkage tetrasaccharides. Notably, EXTL2 is a negative regulator of HS synthesis, which is inhibited when the linkage region is phosphorylated, therefore terminating HS synthesis⁵⁰.

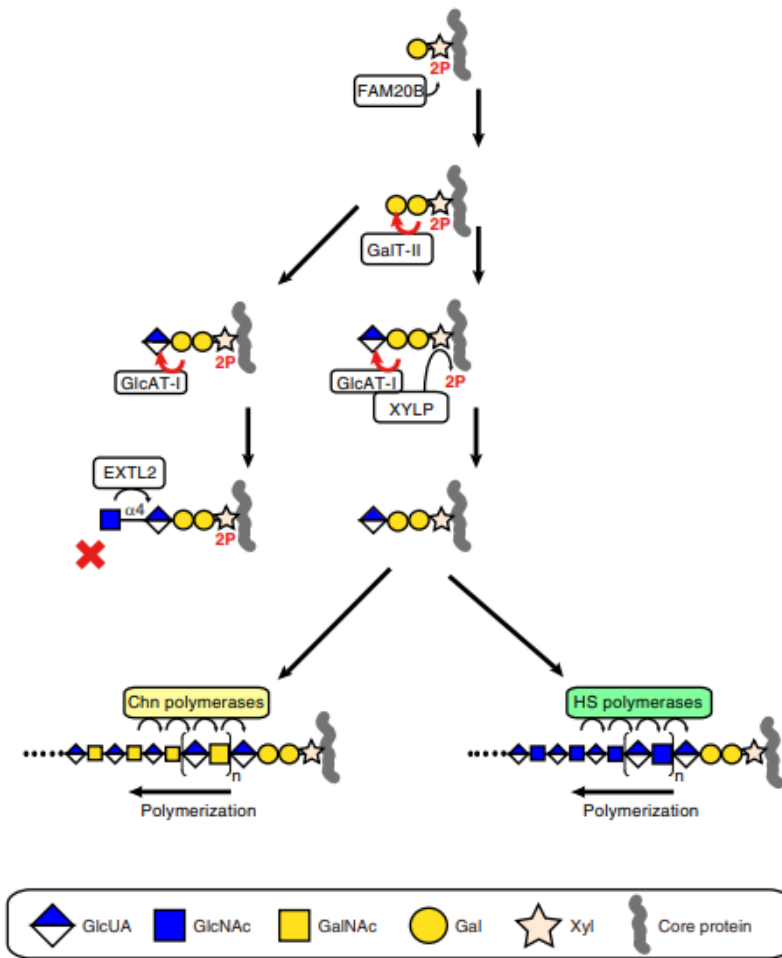


Figure 1.2. Phosphorylation and dephosphorylation of Xyl residues regulate the formation of the linkage region and GAG biosynthesis¹⁴.

1.5 Glucuronyltransferase I

Human glucuronyltransferase I (GlcAT-I), also known as B3GAT3, is an enzyme that plays a crucial role in assembling tetrasaccharide linkage region. GlcAT-I specifically catalyzes the transfer of a GlcA to trisaccharide (Gal-Gal-Xyl) linkage region. GlcA is an important component of GAGs such as HS and CS. The addition of GlcA to the PG chain by GlcAT-I is a key step in the modification and maturation of GAG molecules. The synthesis of HS and CS is completely abolished when GlcAT-I is knocked out in cells¹³, which subsequently affects the

integrity and function of tissues and contributes to the development of certain diseases such as recessive joint dislocations and congenital heart defects, including bicuspid aortic valve (BAV) and aortic root dilatation⁵¹.

Overall, GlcAT-I is an important enzyme for PG biosynthesis, contributing to the structural diversity and functional versatility of these complex carbohydrates in the body. Its study has implications in understanding developmental processes, disease mechanisms, and potentially identifying therapeutic targets for GAG-related disorders.

1.5.1 Expression of GlcAT-I

In 1998, Sugahara group reported the expression of the first expression of GlcAT-1, a truncated version of GlcAT-1, lacking the 43 amino acids from *N*-terminal of the transferase. This enzyme was cloned into vector pSVL containing an insulin signal sequence and a protein A sequence⁵², and then transfected into COS-1 cells using LipofectAMINE. Two days after the transfection, medium was separated and purified with IgG bearing Sepharose.

In 1999, Esko group reported a similar method using stable transfection⁵³. In this study, cDNA encoding amino acid 30-335 was cloned into pCDNA1, then further fused into vector pRK5-F10-PROTA with C-terminal protein A. This plasmid was then transfected into COS-7 cells using LipofectAMINE, and stable transfectants were selected using 0.2 mg/ml active G418. Supernatants after transfection were purified with rabbit IgG beads.

During the same year, Esko group reported another expression of GlcAT-I using a stable transfection method⁵⁴. Gene contains sequence of GlcAT-1 was cloned from CHO cells and inserted into vector pCDNA3. Plasmid was then transfected into mutated CHO cells with Lipofectin with appropriated colony selected.

In 2000, to study the role of cysteine in GlcAT-1 function, Fournel-Gigleux group reported the expression of multiple truncated or mutated human GlcAT-1 using yeast as the expression host⁵⁵. Constructs include one lacking the predicted N-terminal cytoplasmic tail (GlcAT-I Δ NT) or one further fused with the yeast prepro- α -factor secretion leader peptide (GlcAT-I Δ NT/TMD). Similar sequences but lacking the first 25 N-terminal amino acids was fused with yeast prepro- α -factor secretion leader peptide and an antisense primer corresponding to the coding sequence for the last six amino acids (α F-GlcAT-I Δ NT/TMD). Another 2 mutants were constructed with full sequence of GlcAT-I and C33 and C301 mutated to alanine respectively. These enzymes were cloned into yeast vector pPICZB plasmids and transformed into *P. pastoris* SMD 1168 by lithium chloride. Colony was then selected on yeast extract peptone dextrose (YPD) plates then grown in Buffered Glycerol-complex Medium (BMGY). After induction and incubation in a rotary shaker, cells were resuspended in cold breaking buffer (50 mM sodium phosphate, pH 7.4, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 5% (v/v) glycerol) then lysed by vortexing with glass beads with the resulting mixture pelleted by centrifugation. Pellets were resuspended by Dounce homogenization in sucrose- 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer. GlcAT-I was recovered by ammonium sulfate precipitation. Downstream analysis shows that GlcAT-I, fused with the yeast prepro- α -factor secretion leader peptide, is successfully cleaved during expression, and the glycosylation state of GlcAT-I Δ NT/TMD is similar to the WT membrane-bounded GlcAT-I. The result also shows that without the N-terminal cytoplasmic tail, GlcAT-I Δ NT does not alter the enzyme's ability to target and associate with the yeast membrane, as well as its activity. On the other hand, deletion of the cytoplasmic tail allows GlcAT-1 to enter the secretory pathway, suggesting the lack of a retention signal in the stem. Furthermore, the mutation of C33A ($K_m = 67.23 \mu\text{M}$) suppresses dimer formation, eventually leading to impaired

activity compared to WT GlcAT-1 ($K_m = 37.03 \mu\text{M}$), possibly due to the abolished monomer-monomer interaction. Mutation of C301A completely abolishes its activity, suggesting that Cys is either involved in acid-base catalysis or binding with the acceptor or donor.

At the same year, Negeshi group reported another expression of GlcAT-1 using *E.coli* as the host⁵⁶. Truncated human GlcAT-1 with amino acids from 76-335 was cloned into PET-28a with an *N*-terminal 6X His tag. Plasmid was then transformed into BL21(DE3), and the transformed colony was selected on LB agar plates. Upon induction and incubation, cells were lysed and the desired protein was purified by Ni-NTA column. This GlcAT-1 was further concentrated to 30.8 mg/mL and used for crystallization. It was found that the aforementioned C301 is not involved in binding with acceptor and donor, and E281 acts as the base to deprotonate the Gal.

1.5.2 Substrate Specificity of GlcAT-1

In 1998, the Sugahara group discussed the acceptor specificity of GlcAT-1⁵². They compared several acceptor molecules: Gal β 1-3Gal β 1-4Xyl β 1-O-Ser, GalNAc β 1-4GlcA β 1-3Gal β 1-3Gal β 1-4Xyl β 1-O-Ser, GalNAc β 1-4GlcA β 1-3GalNAc β 1-4GlcA β 1-3Gal β 1-3Gal β 1-4Xyl β 1-O-Ser, chondroitin, and asialoorosomucoid (Gal β 1-4GlcNAc-R). Reactions were only observed when the trisaccharide linkage region Gal β 1-3Gal β 1-4Xyl β 1-O-Ser was used as an acceptor. Penta- and septa-saccharides with GalNAc at the non-reducing end showed no reactions, demonstrating that GlcAT-1 can only transfer GlcA to the linkage region, rather than elongating the chondroitin-bearing backbone. Similarly, chondroitin and asialoorosomucoid were not active as acceptors of GlcAT-1.

In 2002, the Fournel-Gigleux group reported on the donor substrate specificity of GlcAT-1. They utilized UDP-GlcA, UDP-GalA, UDP-Glc, UDP-Gal, UDP-GlcNAc, UDP-Man, and

GDP-Man as donors, with Gal-Gal as the acceptor. Only UDP-GlcA and UDP-GalA exhibited reactions, while other donors gave minimal to no reactivities. To further elucidate the reaction mechanism, two mutants, H308A and H308R, were generated. Results revealed that mutating histidine 308 to alanine completely abolished reactions for all donors, suggesting that the histidine residue is crucial for catalytic activity. Interestingly, when histidine 308 was mutated to arginine, all donors displayed activity except for UDP-Gal. When the natural donor UDP-GlcA was used, the mutant exhibited a decreased $V_{\max}$ of $23.78 \text{ mol}\cdot\text{min}^{-1} \text{ mgP}^{-1}$, significantly lower than the native GlcAT-1, which exhibited a $V_{\max}$ of $68.03 \text{ mol}\cdot\text{min}^{-1} \text{ mgP}^{-1}$.

1.6 Outlook

The tetrasaccharide linkage region plays a pivotal role in the intricate machinery of PG biosynthesis. Although the biosynthetic enzymes have been identified, the enzymatic synthesis of the tetrasaccharide linkage region has yet to be reported. Further research needs to be done to utilize these enzymes for assembly of tetrasaccharide on peptides or proteins. Numerous questions remain unanswered, particularly regarding the relationship between the phosphorylated linkage region and downstream HS or CS elongation. While the aforementioned phosphorylated linkage region functions as a negative regulator of HS synthesis⁵⁰, the mechanism by which nature regulates CS synthesis remains unexplained. Further research is essential to comprehend how nature orchestrates the level of glycosylation, particularly in CS and HS, which share the same linkage region. For instance, the case of bikunin, one of the simplest CSPGs⁵⁷, it remains unknown the reason this PG bears a CS chain instead of HS. In summary, the future of research in the tetrasaccharide linkage region holds the promise of unlocking new frontiers in our understanding of glycosaminoglycan biology. This offers potential solutions to longstanding medical challenges and opens up exciting opportunities for therapeutic innovation.

REFERENCES

1. DeAngelis, P. L., Evolution of glycosaminoglycans and their glycosyltransferases: Implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. *Anat. Rec.* **2002**, 268 (3), 317-326.
2. Gandhi, N. S.; Mancera, R. L., The structure of glycosaminoglycans and their interactions with proteins. *Chem. Biol. Drug Des.* **2008**, 72 (6), 455-482.
3. Bishop, J. R.; Schuksz, M.; Esko, J. D., Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* **2007**, 446 (7139), 1030-1037.
4. Galindo, L. T.; Mundim, M.; Pinto, A. S.; Chiarantin, G. M. D.; Almeida, M. E. S.; Lamers, M. L.; Horwitz, A. R.; Santos, M. F.; Porcionatto, M., Chondroitin sulfate impairs neural stem cell migration through ROCK activation. *Mol. Neurobiol.* **2018**, 55 (4), 3185-3195.
5. Li, L.; Ly, M.; Linhardt, R. J., Proteoglycan sequence. *Mol. BioSyst.* **2012**, 8, 1613-1625.
6. Oohira, A.; Matsui, F.; Tokita, Y.; Yamauchi, S.; Aono, S., Molecular interactions of neural chondroitin sulfate proteoglycans in the brain development. *Arch. Biochem. Biophys.* **2000**, 374 (1), 24-34.
7. Xu, D.; Esko, J. D., Demystifying heparan sulfate–protein interactions. *Annu. Rev. Biochem.* **2014**, 83, 129-157.
8. Briggs, D. C.; Hohenester, E., Structural basis for the initiation of glycosaminoglycan biosynthesis by human xylosyltransferase 1. *Structure* **2018**, 26 (6), 801-809.
9. Siegbahn, A.; Manner, S.; Persson, A.; Tykesson, E.; Holmqvist, K.; Ochocinska, A.; Rönnols, J.; Sundin, A.; Mani, K.; Westergren-Thorsson, G.; Widmalm, G.; Ellervik, U., Rules for priming and inhibition of glycosaminoglycan biosynthesis; probing the β 4GalT7 active site. *Chem. Sci.* **2014**, 5 (9), 3501-3508.
10. Gao, J.; Lin, P.-h.; Nick, S. T.; Huang, J.; Tykesson, E.; Ellervik, U.; Li, L.; Huang, X., Chemoenzymatic synthesis of glycopeptides bearing galactose–xylose disaccharide from the proteoglycan linkage region. *Org. Lett.* **2021**, 23 (5), 1738-1741.
11. Koike, T.; Izumikawa, T.; Tamura, J.; Kitagawa, H., FAM20B is a kinase that phosphorylates xylose in the glycosaminoglycan-protein linkage region. *Biochem. J.* **2009**, 421 (2), 157-162.
12. Wen, J.; Xiao, J.; Rahdar, M.; Choudhury, B. P.; Cui, J.; Taylor, G. S.; Esko, J. D.; Dixon, J. E., Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, 111 (44), 15723-15728.

13. Chen, Y.-H.; Narimatsu, Y.; Clausen, T. M.; Gomes, C.; Karlsson, R.; Steentoft, C.; Spliid, C. B.; Gustavsson, T.; Salanti, A.; Persson, A.; Malmström, A.; Willén, D.; Ellervik, U.; Bennett, E. P.; Mao, Y.; Clausen, H.; Yang, Z., The GAGome: a cell-based library of displayed glycosaminoglycans. *Nat. Methods* **2018**, *15* (11), 881-888.
14. Koike, T.; Izumikawa, T.; Sato, B.; Kitagawa, H., Identification of phosphatase that dephosphorylates xylose in the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol.Chem.* **2014**, *289* (10), 6695-6708.
15. Pedersen, L. C.; Tsuchida, K.; Kitagawa, H.; Sugahara, K.; Darden, T. A.; Negishi, M., Heparan/chondroitin sulfate biosynthesis: Structure and mechanism of human glucuronyltransferase I. *J. Biol.Chem.* **2000**, *275* (44), 34580-34585.
16. Haouari, W.; Dubail, J.; Poüs, C.; Cormier-Daire, V.; Bruneel, A., Inherited proteoglycan biosynthesis defects—current laboratory tools and bikunin as a promising blood biomarker. *Genes* **2021**, *12* (11), 10.3390/genes12111654.
17. Uyama, T.; Kitagawa, H.; Tamura, J.-i.; Sugahara, K., Molecular cloning and expression of human chondroitinN-Acetylgalactosaminyltransferase: The key enzyme for chain initiation and elongation of chondroitin/dermatan sulfate on the protein linkage region tetrasaccharide shared by heparin/heparan sulfate. *J. Biol.Chem.* **2002**, *277* (11), 8841-8846.
18. Izumikawa, T.; Uyama, T.; Okuura, Y.; Sugahara, K.; Kitagawa, H., Involvement of chondroitin sulfate synthase-3 (chondroitin synthase-2) in chondroitin polymerization through its interaction with chondroitin synthase-1 or chondroitin-polymerizing factor. *Carbohydr. Polym.* **2007**, *403* (3), 545-552.
19. Kitagawa, H.; Shimakawa, H.; Sugahara, K., The tumor suppressor EXT-like Gene EXTL2 encodes an α 1, 4-N-Acetylhexosaminyltransferase that transfers N-Acetylgalactosamine and N-Acetylglucosamine to the common glycosaminoglycan-protein linkage region: The key enzyme for the chain initiation of heparan sulfate. *J. Biol.Chem.* **1999**, *274* (20), 13933-13937.
20. Lind, T.; Tufaro, F.; McCormick, C.; Lindahl, U.; Lidholt, K., The putative tumor suppressors ext1 and ext2 are glycosyltransferases required for the biosynthesis of heparan sulfate. *J. Biol. Chem.* **1998**, *273* (41), 26265-26268.
21. Yang, W.; Eken, Y.; Zhang, J.; Cole, L. E.; Ramadan, S.; Xu, Y.; Zhang, Z.; Liu, J.; Wilson, A. K.; Huang, X., Chemical synthesis of human syndecan-4 glycopeptide bearing O-, N-sulfation and multiple aspartic acids for probing impacts of the glycan chain and the core peptide on biological functions. *Chem. Sci.* **2020**, *11* (25), 6393-6404.
22. Kim, S. K.; Henen, M. A.; Hinck, A. P., Structural biology of betaglycan and endoglin, membrane-bound co-receptors of the TGF-beta family. *Exp. Biol. Med.* **2019**, *244* (17), 1547-1558.
23. Iozzo, R. V., Series Introduction: Heparan sulfate proteoglycans: Intricate molecules with intriguing functions. *J. Clin. Invest.* **2001**, *108* (2), 165-167.

24. Herndon, M. E.; Stipp, C. S.; Lander, A. D., Interactions of neural glycosaminoglycans and proteoglycans with protein ligands: assessment of selectivity, heterogeneity and the participation of core proteins in binding. *Glycobiology* **1999**, 9 (2), 143-55.
25. Breton, C.; Fournel-Gigleux, S.; Palcic, M. M., Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.* **2012**, 22 (5), 540-549.
26. Mizumoto, S.; Ikegawa, S.; Sugahara, K., Human genetic disorders caused by mutations in genes encoding biosynthetic enzymes for sulfated glycosaminoglycans *J. Biol. Chem.* **2013**, 288 (16), 10953-10961.
27. Gao, J.; Huang, X., Chapter Three - Recent advances on glycosyltransferases involved in the biosynthesis of the proteoglycan linkage region. In *Advances in Carbohydrate Chemistry and Biochemistry*, Baker, D. C., Ed. Academic Press: 2021; Vol. 80, pp 95-119.
28. Eames, B. F.; Yan, Y.-L.; Swartz, M. E.; Levic, D. S.; Knapik, E. W.; Postlethwait, J. H.; Kimmel, C. B., Mutations in fam20b and xylt1 reveal that cartilage matrix controls timing of endochondral ossification by inhibiting chondrocyte maturation. *PLoS Genet.* **2011**, 7 (8), 10.1371/journal.pgen.1002246.
29. Vogel, P.; Hansen, G.; Read, R.; Vance, R.; Thiel, M.; Liu, J.; Wronski, T.; Smith, D.; Jeter-Jones, S.; Brommage, R., Amelogenesis imperfecta and other biomineralization defects in Fam20a and Fam20c null mice. *Vet. Pathol.* **2012**, 49 (6), 998-1017.
30. Ma, P.; Yan, W.; Tian, Y.; Wang, J.; Feng, J. Q.; Qin, C.; Cheng, Y.-S. L.; Wang, X., Inactivation of Fam20B in joint cartilage leads to chondrosarcoma and postnatal ossification defects. *Sci. Rep.* **2016**, 6 (1), 10.1038/srep29814.
31. Worby, C. A.; Mayfield, J. E.; Pollak, A. J.; Dixon, J. E.; Banerjee, S., The ABCs of the atypical Fam20 secretory pathway kinases. *J. Biol.Chem.* **2021**, 296, 10.1016/j.jbc.2021.100267.
32. Vogel, P.; Hansen, G. M.; Read, R. W.; Vance, R. B.; Thiel, M.; Liu, J.; Wronski, T. J.; Smith, D. D.; Jeter-Jones, S.; Brommage, R., Amelogenesis imperfecta and other biomineralization defects in Fam20a and Fam20c null mice. *Vet. Pathol.* **2012**, 49 (6), 998-1017.
33. Eames, B. F.; Yan, Y.-L.; Swartz, M. E.; Levic, D. S.; Knapik, E. W.; Postlethwait, J. H.; Kimmel, C. B., Mutations in fam20b and xylt1 Reveal That Cartilage Matrix Controls Timing of Endochondral Ossification by Inhibiting Chondrocyte Maturation. *PLOS Genetics* **2011**, 7 (8), e1002246.
34. Tian, Y.; Ma, P.; Liu, C.; Yang, X.; Crawford, D. M.; Yan, W.; Bai, D.; Qin, C.; Wang, X., Inactivation of Fam20B in the dental epithelium of mice leads to supernumerary incisors. *Eur. J. Oral Sci.* **2015**, 123 (6), 396-402.

35. Wu, J.; Bollinger, A. T.; He, X.; Gu, G. D.; Miao, H.; Dean, M. P. M.; Robinson, I. K.; Božović, I., Angle-resolved transport measurements reveal electronic nematicity in cuprate superconductors. *J. Supercond. Nov. Magn.* **2020**, *33* (1), 87-92.
36. Ma, P.; Yan, W.; Tian, Y.; Wang, J.; Feng, J. Q.; Qin, C.; Cheng, Y. S. L.; Wang, X., Inactivation of Fam20B in joint cartilage leads to chondrosarcoma and postnatal ossification defects. *Sci. Rep.* **2016**, *6* (1), 10.1038/srep29814.
37. Liu, X.; Li, N.; Zhang, H.; Liu, J.; Zhou, N.; Ran, C.; Chen, X.; Lu, Y.; Wang, X.; Qin, C.; Xiao, J.; Liu, C., Inactivation of Fam20b in the neural crest-derived mesenchyme of mouse causes multiple craniofacial defects. *Eur. J. Oral Sci.* **2018**, *126* (5), 433-436.
38. Nalbant, D.; Youn, H.; Nalbant, S. I.; Sharma, S.; Cobos, E.; Beale, E. G.; Du, Y.; Williams, S. C., FAM20: an evolutionarily conserved family of secreted proteins expressed in hematopoietic cells. *BMC Genom.* **2005**, *6* (1), 11-32.
39. Nadanaka, S.; Zhou, S.; Kagiya, S.; Shoji, N.; Sugahara, K.; Sugihara, K.; Asano, M.; Kitagawa, H., EXTL2, a member of the EXT family of tumor suppressors, controls glycosaminoglycan biosynthesis in a xylose kinase-dependent manner. *J. Biol.Chem.* **2013**, *288* (13), 9321-9333.
40. Zhang, H.; Zhu, Q. Y.; Cui, J. X.; Wang, Y. X.; Chen, M. J.; Guo, X.; Tagliabracci, V. S.; Dixon, J. E.; Xiao, J. Y., Structure and evolution of the Fam20 kinases. *Nat. Comm.* **2018**, *9*, 10.1038/s41467-018-03615-z.
41. Nadanaka, S.; Kitagawa, H.; Sugahara, K., Demonstration of the immature glycosaminoglycan tetrasaccharide sequence GlcAbeta1-3Galbeta1-3Galbeta1-4Xyl on recombinant soluble human alpha-thrombomodulin. An oligosaccharide structure on a "part-time" proteoglycan. *J. Biol. Chem.* **1998**, *273* (50), 33728-33734.
42. Trejo, P.; Rauch, F.; Glorieux, F. H.; Ouellet, J.; Benaroch, T.; Campeau, P. M., Spondyloepimetaphysal dysplasia with joint laxity in three siblings with B3GALT6 mutations. *Mol. Syndromol.* **2017**, *8* (6), 303-307.
43. Han, S.; Xu, X.; Wen, J.; Wang, J.; Xiao, S.; Pan, L.; Wang, J., New genetic mutations in a Chinese child with Ehlers-Danlos syndrome-like spondyloepimetaphyseal dysplasia: A case report. *Front. Pediatr.* **2022**, *10*, 10.3389/fped.2022.1073748.
44. Nakajima, M.; Mizumoto, S.; Miyake, N.; Kogawa, R.; Iida, A.; Ito, H.; Kitoh, H.; Hirayama, A.; Mitsubuchi, H.; Miyazaki, O.; Kosaki, R.; Horikawa, R.; Lai, A.; Mendoza-Londono, R.; Dupuis, L.; Chitayat, D.; Howard, A.; Leal, Gabriela F.; Cavalcanti, D.; Tsurusaki, Y.; Saitsu, H.; Watanabe, S.; Lausch, E.; Unger, S.; Bonafé, L.; Ohashi, H.; Superti-Furga, A.; Matsumoto, N.; Sugahara, K.; Nishimura, G.; Ikegawa, S., Mutations in B3GALT6, which encodes a glycosaminoglycan linker region enzyme, cause a spectrum of skeletal and connective tissue disorders. *Am. J. Hum. Genet.* **2013**, *92* (6), 927-934.

45. Cole, S. E.; Mao, M. S.; Johnston, S. H.; Vogt, T. F., Identification, expression analysis, and mapping of B3galt6, a putative galactosyl transferase gene with similarity to *Drosophila* brainiac. *Mamm. Genome* **2001**, *12* (2), 177-179.
46. Moremen, K. W.; Ramiah, A.; Stuart, M.; Steel, J.; Meng, L.; Forouhar, F.; Moniz, H. A.; Gahlay, G.; Gao, Z.; Chapla, D.; Wang, S.; Yang, J.-Y.; Prabhakar, P. K.; Johnson, R.; Rosa, M. d.; Geisler, C.; Nairn, A. V.; Seetharaman, J.; Wu, S.-C.; Tong, L.; Gilbert, H. J.; LaBaer, J.; Jarvis, D. L., Expression system for structural and functional studies of human glycosylation enzymes. *Nat. Chem. Biol.* **2018**, *14* (2), 156-162.
47. Wen, J.; Xiao, J.; Rahdar, M.; Choudhury, B. P.; Cui, J.; Taylor, G. S.; Esko, J. D.; Dixon, J. E., Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (44), 15723-15728.
48. Van Damme, T.; Pang, X.; Guillemin, B.; Gulberti, S.; Syx, D.; De Rycke, R.; Kaye, O.; de Die-Smulders, C. E. M.; Pfundt, R.; Kariminejad, A.; Nampoothiri, S.; Pierquin, G.; Bulk, S.; Larson, A. A.; Chatfield, K. C.; Simon, M.; Legrand, A.; Gerard, M.; Symoens, S.; Fournel-Gigleux, S.; Malfait, F., Biallelic B3GALT6 mutations cause spondylodysplastic Ehlers–Danlos syndrome. *Hum. Mol. Genet.* **2018**, *27* (20), 3475-3487.
49. Persson, A.; Nilsson, J.; Vorontsov, E.; Noborn, F.; Larson, G., Identification of a non-canonical chondroitin sulfate linkage region trisaccharide. *Glycobiology* **2019**, *29* (5), 366-371.
50. Nadanaka, S.; Zhou, S.; Kagiya, S.; Shoji, N.; Sugahara, K.; Sugihara, K.; Asano, M.; Kitagawa, H., EXTL2, a member of the EXT family of tumor suppressors, controls glycosaminoglycan biosynthesis in a xylose kinase-dependent manner. *J. Biol. Chem.* **2013**, *288* (13), 9321-33.
51. Baasanjav, S.; Al-Gazali, L.; Hashiguchi, T.; Mizumoto, S.; Fischer, B.; Horn, D.; Seelow, D.; Ali, Bassam R.; Aziz, Samir A. A.; Langer, R.; Saleh, Ahmed A. H.; Becker, C.; Nürnberg, G.; Cantagrel, V.; Gleeson, Joseph G.; Gomez, D.; Michel, J.-B.; Stricker, S.; Lindner, Tom H.; Nürnberg, P.; Sugahara, K.; Mundlos, S.; Hoffmann, K., Faulty initiation of proteoglycan synthesis causes cardiac and joint defects. *Am. J. Med. Genet.* **2011**, *89* (1), 15-27.
52. Kitagawa, H.; Tone, Y.; Tamura, J.-i.; Neumann, K. W.; Ogawa, T.; Oka, S.; Kawasaki, T.; Sugahara, K., Molecular cloning and expression of glucuronyltransferase I involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol. Chem.* **1998**, *273* (12), 6615-6618.
53. Wei, G.; Bai, X.; Sarkar, A. K.; Esko, J. D., Formation of HNK-1 determinants and the glycosaminoglycan tetrasaccharide linkage region by UDP-GlcUA:Galactose beta1, 3-glucuronosyltransferases. *J. Biol. Chem.* **1999**, *274* (12), 7857-64.
54. Bai, X.; Wei, G.; Sinha, A.; Esko, J. D., Chinese hamster ovary cell mutants defective in glycosaminoglycan assembly and glucuronosyltransferase i. *J. Biol. Chem.* **1999**, *274* (19), 13017-13024.

55. Ouzzine, M.; Gulberti, S.; Netter, P.; Magdalou, J.; Fournel-Gigleux, S., Structure/Function of the Human Gal β 1,3-glucuronosyltransferase: Dimerization and functional activity are mediated by two crucial cysteine residues. *J. Biol.Chem.* **2000**, 275 (36), 28254-28260.
56. Pedersen, L. C.; Tsuchida, K.; Kitagawa, H.; Sugahara, K.; Darden, T. A.; Negishi, M., Heparan/chondroitin sulfate biosynthesis - Structure and mechanism of human glucuronyltransferase I. *J. Biol.Chem.* **2000**, 275 (44), 34580-34585.
57. Ly, M.; Leach, F. E., 3rd; Laremore, T. N.; Toida, T.; Amster, I. J.; Linhardt, R. J., The proteoglycan bikunin has a defined sequence. *Nat. Chem. Biol.* **2011**, 7 (11), 827-833.

Chapter 2. Expedient Solid Phase Supported Chemo-Enzymatic Synthesis of Chondroitin Sulfate Proteoglycan Glycopeptides

2.1 Introduction

Proteoglycans (PGs), a family of glycoproteins, are commonly found on cell surfaces and within the extracellular matrix. They play pivotal roles in many biological events including cell proliferation, inflammation, and viral infection¹⁻⁵. PGs contain one or more glycosaminoglycan (GAG) chains covalently conjugated to serine (Ser) residues on a core protein backbone through a typical tetrasaccharide linkage with the sequence of glucuronic acid (GlcA)- β 1-3-galactose (Gal)- β 1-3-galactose (Gal)- β 1-4-xylose (Xyl)- β 1-Ser (GlcA β 1-3Gal β 1-3Gal β 1-4Xyl β 1-Ser). The GAG chain of proteoglycans can be chondroitin sulfate (CS) or heparan sulfate (HS) with sulfations at various hydroxyl groups forming CS proteoglycan (CSPG) or HS proteoglycan⁶. The sulfation patterns of PGs are highly heterogeneous due to incomplete enzymatic sulfations of the glycan chains, resulting in large structural diversities of naturally existing PGs.

Traditionally, the biological functions of PGs are thought to be generally directed by the glycan chains attached. However, an increasing body of research has suggested that the core protein can be important as well⁷⁻⁹, with the core protein and the glycan chain potentially exhibiting synergistic effects for the biological functions of PGs¹⁰⁻¹². To gain deeper insights into the functions of PGs and decipher the respective roles of the glycan and the core protein, it is imperative to obtain structurally well-defined and homogeneous glycopeptides and PGs.

With the high heterogeneity of PGs, it is almost impossible to purify homogeneous PG structures from natural sources. Several strategies have been developed for the syntheses of HS and CS bearing glycopeptides with the native tetrasaccharide linkage region¹²⁻¹⁴. The chemical synthesis of GAG-bearing glycopeptides is a formidable challenge, stemming from the intricate

series of protecting group manipulation required, glycosylation reactions, chemical sulfation, and the incompatibilities between typical peptide and GAG synthesis conditions¹⁵⁻²⁰. The longest HS and CS glycopeptides prepared to date bear octasaccharides on the peptide backbones^{21, 22}. These syntheses were tedious with the total number of synthetic steps needed well over 100 for some of the targets.²¹ Recently, the Huisgen alkyne-azide cycloaddition reactions were utilized to conjugate HS with protein backbones bearing alkynyl tyrosine¹⁰. While ground-breaking, these PG mimetics contain heterogeneous glycans and the glycan chain was linked through an unnatural triazole moiety to tyrosine in the core protein.

To greatly expedite the synthesis of the native tetrasaccharide linkage region and CS-bearing glycopeptide, in this study, we introduce a new chemoenzymatic method facilitated by solid phase support. The underlying principle involves the cloning and expression of the enzymes required for PG synthesis. This is followed by conjugation of the peptide backbone onto Sepharose beads, with subsequent successive rounds of enzymatic extensions and modifications. The native tetrasaccharide linkage region and CS-bearing glycopeptides formed were then released under a mild reaction condition without affecting the sensitive glycopeptides. Leveraging this powerful strategy, we successfully generated multiple tetrasaccharide linkage-bearing glycopeptides bearing diverse amino acid sequences in the backbone, as well as CS-bearing glycopeptides of varying lengths in mg scales. The affinities of the bikunin glycopeptides with a potential receptor, cathepsin G (CatG) were investigated and rationalized with computational modeling, demonstrating the important role of glycan sulfation for CatG binding.

2.2 Results and Discussion

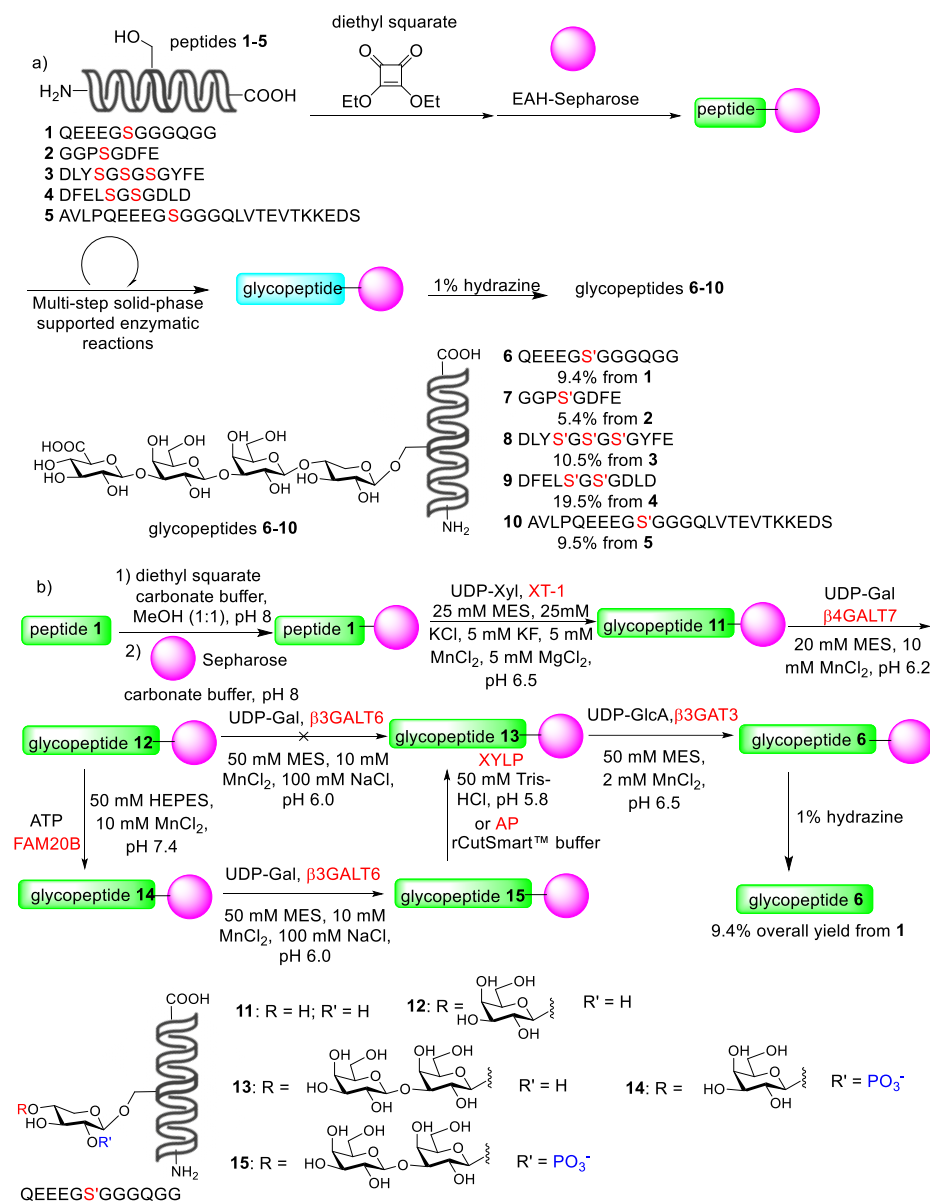
2.2.1 Construction of tetrasaccharide linkage region-bearing glycopeptide

There are multiple challenges in establishing a viable enzymatic route for PG synthesis, which include the identification of suitable enzymes to catalyze the synthesis, the production of the enzymes, and the time-consuming process of isolating the highly polar product from the aqueous reaction media. In order to expedite the synthesis and reduce the time needed to purify the highly polar glycopeptides, we investigated the possibility of performing enzymatic synthesis of PG glycopeptides on solid phase²³⁻²⁵. Various solid supports have been reported for enzymatic synthesis of glycans or glycopeptides, which include polyethylene glycol polyacrylamide copolymer (PEGA)²⁶, amine-functionalized silica²⁵, controlled pore glass (CPG)²⁷, and thermo-responsive water-soluble polymers^{28, 29}. While they have demonstrated compatibility with enzymes, each of the solid supports needs unique consideration. For instance, swelling resins like PEGA, due to their limited pore size, may prove insufficient for reactions requiring enzymes with molecular weight higher than 50 kDa²⁹. Conversely, non-swelling solid supports like amine-functionalized silica and CPG may exhibit less compatibility with certain enzymes compared to swelling solid supports²⁹. Thermo-responsive water-soluble polymers can provide solution-like environment for enzymatic synthesis, while allowing precipitation of the product from the reaction media through heating after the reaction^{30, 31}. However, when the glycan becomes charged such as after sialylation, the polymer can no longer be precipitated from the solution upon heating³¹. Thus, it may not be suitable for PG synthesis due to the highly negative charged nature of GAGs on PG.

To enable the chemo-enzymatic synthesis of PG glycopeptides, we explored Sepharose³² as a potential solid phase support. Sepharose possesses commendable swelling properties in aqueous buffers with large pore sizes (~20,000 kDa), and is commonly employed for protein

purification³³. PG glycopeptides exhibit limited stability under strongly acidic or basic conditions, primarily due to the susceptibility of glycopeptides to undergo glycan elimination under a basic condition^{16, 20}, and the potential for *O*-sulfate loss under an acid condition. Consequently, a linker that can be cleaved under a mild condition, is imperative to conjugate the precursor peptides to solid phase. We opted to utilize diethyl squarate as a traceless linker³⁴ as it can yield the native glycopeptide after cleavage from the resin.

To establish the feasibility of solid phase supported enzymatic synthesis of PG glycopeptide, the peptide acceptor was functionalized through its *N*-terminal amine with diethyl squarate in a mixed solvent of carbonate buffer and methanol (**Scheme 2.1a**). After 6 hours, liquid chromatography mass spectrometry (LCMS) analysis confirmed the formation of the desired squarate-modified peptide with the complete consumption of the free peptide. Subsequently, the reaction mixture was incubated with EAH Sepharose, a commercially available Sepharose resin with an 11-atom hydrophilic spacer arm from the surface (2 equiv. based on free amine on the Sepharose to the peptide), in a carbonate buffer. The resulting slurry was kept in a frit-fitted syringe and agitated with end-to-end rotation for 24 hours, when LCMS analysis indicated no free squarate-modified peptide remained in solution.



Scheme 2.1. a) Schematic demonstration of the solid phase supported enzymatic synthesis of tetrasaccharide linkage region bearing glycopeptides **6-10**. The serine glycosylation sites are indicated in red; b) Sepharose supported enzymatic synthesis of glycopeptide **6** from peptide **1**.

In order to establish the cleavage conditions, we first treated the peptide **1** loaded Sepharose with boric acid in combination with concentrated ammonia³². This yielded some of the desired

peptide product along with squarate-modified peptide as indicated by LCMS analysis (data not shown). However, this method also generated a notable amount of ammonium borate salt, adversely affecting subsequent high performance liquid chromatography (HPLC) purification. The second approach investigated utilized 5% aqueous hydrazine³⁰. Interestingly, while it cleaved the peptide from Sepharose, it also led to an undesired side product with the *N*-terminal amino acid residue removed. Reducing the concentration of hydrazine to 1% completely mitigated this side reaction.

With the solid phase immobilization and cleavage conditions established, we moved on to express the requisite enzymes to form the glycosyl bonds in the linkage region, which included the xylosyl transferase-1 (XT-1)^{35, 36}, β -1,4-galactosyl transferase 7 (β 4GALT7)³⁷⁻³⁹, β -1,4-galactosyl transferase 6 (β 3GALT6)⁴⁰, and β -1,3-glucuronic acid transferase 3 (β 3GAT3)⁴¹. We found that β 4GALT7 and β 3GAT3 could be expressed well in *E. coli* with the yields of 16.7 and 5 mg/L respectively, while XT-1, FAM20B and β 3GALT6 should be expressed in the HEK293F cells with the desired enzymes isolated from the supernatant in 10, 2.3 and 16.7 mg/L respectively.

Solid-phase enzymatic syntheses have traditionally been performed using a glycosylated peptide as the substrate to initiate enzymatic reactions^{23, 25, 26, 29}. This strategy commences from chemical synthesis of the glycosylated amino acid cassette followed by its incorporation into the glycopeptide chain, which requires the usage of an excess of the valuable glycosyl amino acid building blocks. As an alternative, we opted to explore direct glycosylation of the peptide on solid phase. The Sepharose resin loaded with peptide **1** was treated with uridine diphosphate (UDP)-Xyl and XT-1 (MW: 87 kDa) in a reaction buffer comprising 25 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 25 mM KCl, 5 mM KF, 5 mM MgCl₂, 5 mM MnCl₂, at pH 6.5 over a 12-hour period with end-over-end rotation at 4°C, which was repeated once to ensure complete conversion

(**Scheme 2.1 b**). The crude product was subjected to cleavage using 1% hydrazine solution in water followed by analysis via LCMS, which showed the desired target glycopeptide **11** as the sole product. With the confirmation of the successful xylosylation, the xylosylated glycopeptide on Sepharose was incubated with UDP-Gal and β 4GALT7 in 20 mM MES and 10 mM MnCl_2 , at pH 6.2 at 4°C (**Scheme 2.1 b**). Subsequent treatment of the Sepharose with 1% hydrazine gave glycopeptide **12** as the desired product suggesting the successful transfer of the Gal unit to xylosyl peptide **12**. Next, we tested the transfer of a second Gal to the linkage region. Unfortunately, upon incubation of the Gal-Xyl glycopeptide **12** bearing Sepharose with UDP-Gal and β 3GALT6, no desired trisaccharide Gal-Gal-Xyl glycopeptide **13** was obtained. This failure was not due to the solid phase support as incubation of free glycopeptide **12** with UDP-Gal and β 3GALT6 in solution also failed to yield the desired trisaccharide glycopeptide **13**.

The enzyme Family With Sequence Similarity 20 Member B (FAM20B) is a kinase capable of phosphorylating the 2-OH of xylose in the tetrasaccharide linkage region⁴². It can act as a molecular switch regulating the functions of β 3GALT6⁴³. To test whether the 2-O phosphorylation of the xylose can enhance the glycopeptide synthesis yield, disaccharide glycopeptide **12** on Sepharose was subjected to adenosine triphosphate (ATP) and FAM20B in 50 mM *N*-(2-hydroxyethyl)piperazine-*N'*-(2-ethanesulfonic acid) (HEPES) and 10 mM MnCl_2 buffer at pH 7.4. This was followed by the treatment with UDP-Gal and β 3GALT6 (**Scheme 2.1 b**). Gratifyingly, cleavage of the glycopeptide from the Sepharose following this sequence of reactions showed the successful formation of the phosphorylated Gal-Gal-Xyl bearing glycopeptide **15**. This suggests FAM20B can phosphorylate the glycopeptide attached on solid phase and the 2-O phosphorylation significantly enhanced the yield for the Gal transfer by β 3GALT6. The

β 3GALT6⁴⁴ expressed is a truncated form of the protein comprising amino acids 35 to 329. This is the first demonstration that such a construct is enzymatically active.

To regenerate the non-phosphorylated glycan, we expressed 2-phosphoxylose phosphatase (XYLP)⁴⁵, which dephosphorylated glycopeptide **15** to glycopeptide **13**. As XYLP was expressed in HEK293F cells, we explored the potential replacement of XYLP with the commercially available alkaline phosphatase (AP) to reduce the costs associated with mammalian cell expression. Interestingly, AP could also efficiently dephosphorylate glycopeptide **15** on Sepharose. Subsequently, trisaccharide glycopeptide **13** was extended on Sepharose with β 3GAT3 with UDP-GlcA as the donor in 50 mM MES, 2 mM MnCl₂, at pH 6.5 producing glycopeptide **6** bearing the full tetrasaccharide linkage region. The overall yield from peptide **1** to glycopeptide **6** was 9.6%, which is an average of 77% yield per synthetic step. NMR and MS data of **9** were fully consistent with its structure.

With the successful enzymatic synthesis of glycopeptide **6** on solid phase, we tested the generality of the approach with several other representative peptide substrates **2-5** (**Scheme 2.1 a**), which contain a variety of amino acid residues including acidic, basic, aromatic, and aliphatic amino acids flanking the glycosylation sites. These peptide sequences were derived from common proteoglycans in nature, which include syndecan 3 (peptides **2** and **3**), syndecan 4 (peptide **4**) and bikunin (peptides **1** and **5**). Peptides **2** and **3** have two glycosylation sites each, while peptide **4** has three glycosylation sites. Gratifyingly, following the same reaction protocol on Sepharose for the synthesis of glycopeptide **6**, peptides **2-5** were successfully converted to glycopeptides **7-10**, each bearing the full tetrasaccharide linkage regions in 5.4%, 10.5%, 19.5%, and 9.5% yields respectively demonstrating the robustness of the synthetic protocol. For each glycopeptide synthesis, it took nine steps from the peptide backbone. In comparison, chemical synthesis of a

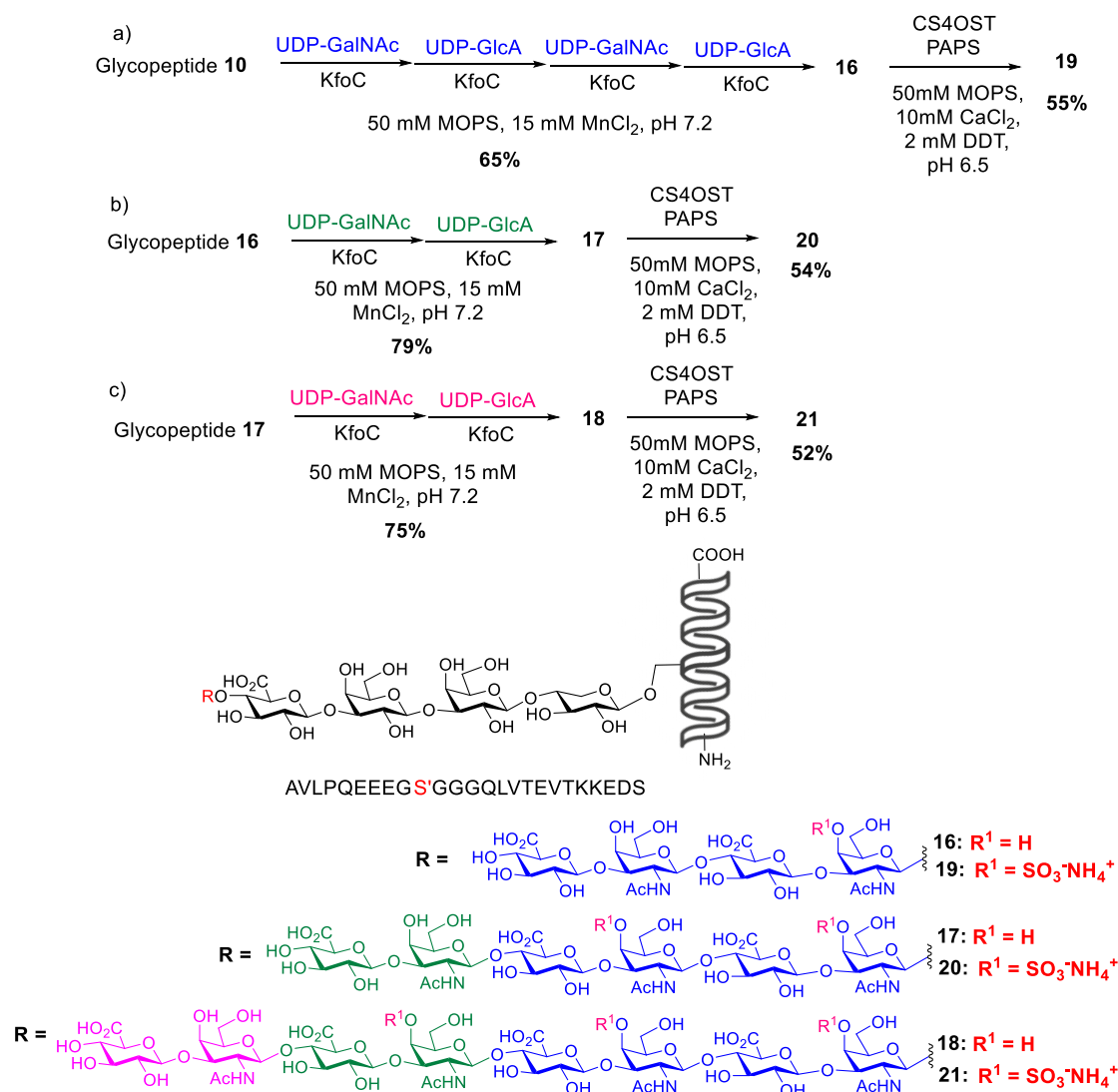
tetrasaccharide linkage region bearing peptide took 39 total synthetic steps and 1.1% yield for the longest linear steps (27 steps) from the peptide and commercially available carbohydrate building blocks¹⁴.

2.2.2 Synthesis of CS-bearing glycopeptides

Bikunin, also known as inter- α -trypsin inhibitor or trypstatin, is a naturally existing CSPG⁴⁶. Initially discovered in urine and human plasma, it is implicated in various biological activities for anti-inflammation and cancer⁴⁷⁻⁵⁰, and has been utilized to treat acute inflammatory disorders including sepsis^{51, 52}. With the tetrasaccharide linkage region bearing glycopeptide **10** in hand, we proceed to synthesize homogenous bikunin glycopeptide.

In nature, the synthesis of CSPG is directed by the immediate sugar residue added to the linkage region, with the transfer of an *N*-acetyl galactosamine (GalNAc) residue to the tetrasaccharide linkage region by the GalNAc transferase-1 (GalNAcT-I)⁵³ initiating CS synthesis. The glycan chain is further extended by the GlcA transferase⁵⁴ and the GalNAc transferase⁵⁵ forming the CS backbone. Subsequently, various sulfo-transferases will selectively install *O*-sulfates onto the CS backbone forming CSPG. KfoC is a bacterial enzyme involved in the synthesis of the capsular chondroitin backbone of *Escherichia coli* K4, which is bifunctional capable of transferring both GlcA and GalNAc to a chondroitin chain.⁵⁶ Rather than expressing GalNAcT-I, we tested KfoC's ability to direct the synthesis of CS glycopeptide. As illustrated in **Scheme 2.2 a**, glycopeptide **10** was first treated with UDP-GalNAc and KfoC in 50 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS) and 15 mM MnCl₂ buffer at pH 7.2, followed by solid phase extraction through C18 silica gel. The resulting fractions containing the desired product were lyophilized and then subjected to treatment with UDP-GlcA and KfoC. This process was repeated two more times with the alternating usage of UDP-GalNAc and UDP-GlcA, producing

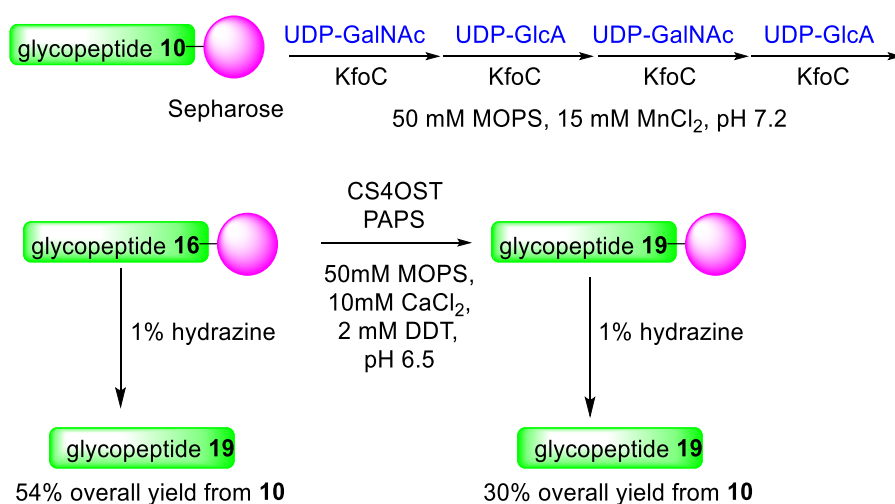
octasaccharide chondroitin glycopeptide **16** in 65% yield from tetrasaccharide glycopeptide **10**. This suggests that KfoC can be useful to not only extend the chondroitin backbone, but also initiate the formation of chondroitin backbone from the linkage region to enable CS glycopeptide synthesis.



Scheme 2.2. Enzymatic synthesis of a) CS octasaccharide glycopeptide **19**; b) CS decahexasaccharide glycopeptide **20**; c) CS dodecasaccharide glycopeptide **21**.

With the octasaccharide glycopeptide **16** in hand, its glycan chain was further extended by alternating UDP-GalNAc and UDP-GlcA in the presence of KfoC, generating chondroitin glycopeptides **17** and **18** bearing decasaccharide and dodecasaccharide chain in 79% and 75% yields respectively (**Schemes 2.2 b** and **2.2c**). To synthesize CS glycopeptides, glycopeptides **16**, **17** and **18** were subjected to 4-*O* sulfation using 3'-phosphoadenosine-5'-phosphosulfate (PAPS) and CS 4-*O* sulfotransferase (CS4OST)⁵⁷ (**Scheme 2.2**). The corresponding *O*-sulfated glycopeptides **19**, **20** and **21** were isolated using C18 reverse phase HPLC in 55%, 54%, and 52% yields respectively. The dodecasaccharide glycopeptide **21** bears three *O*-sulfates, which is the longest GAG-bearing glycopeptide synthesized to date.

To test the possibility of synthesizing the CS chain on solid phase support, we incubated the glycopeptide **10** attached Sepharose with KfoC and alternating UDP-GalNAc, UDP-GlcA, UDP-GalNAc, and UDP-GlcA followed by cleavage from the solid phase with 1% hydrazine (**Scheme 2.3**). The glycopeptide **16** was obtained in 54% overall yield from **10** in 4 days. The sulfation reactions could be performed on solid phase as well. Treatment of Sepharose bearing **16** with PAPS and CS4OST followed by 1% hydrazine cleavage led to the CS octasaccharide glycopeptide **19** from **10** in 30% overall yield from glycopeptide **10**. While the overall yield of the solid phase supported synthesis of **19** was similar to that from solution based synthesis (**Scheme 2.2**), solid phase synthesis cut down the amount of time needed for synthesis by about 50% as it reduced the need for the time consuming intermediate purification and lyophilization for water removal and sample concentration.



Scheme 2.3. Solid phase supported synthesis of CS glycopeptide 19.

Previously, a CS octasaccharide bearing glycopeptide was synthesized via chemical synthesis using a convergent strategy.²² From commercially available carbohydrate building blocks, it took more than 80 synthetic steps in total to complete with an overall yield of 0.73% for the longest linear synthetic sequence of 26 steps. In comparison, the enzymatic synthesis of CS glycopeptide 19 took a total of 14 synthetic steps from peptide 5 in 3.4% overall yield.

2.2.3 Determination of the location of sulfations in CS glycopeptides by mass spectrometry (MS) (Collaboration with Dr. Jon Amster)

As there are multiple GalNAc residues thus potential sulfation sites within glycopeptide 19-21, a MS based methodology was applied to determine the position(s) of GalNAc sulfated. The glycopeptides were digested with actinase E, and the resulting serine glycans were fragmented and sequenced using capillary zone electrophoresis Fourier transform ion cyclotron resonance (CZE-FT-ICR) MS. For example, for dodecasaccharide glycopeptide 21, fragment Y₇ reveals that the initial two sulfates are situated on GalNAc 5 and GalNAc 7 (**Figure 2.1**). The fragments B₂ and B₇ suggest that the GalNAc closest to the non-reducing end (GalNAc 11) was not sulfated, and

GalNAc 9 was sulfated. This cumulative evidence demonstrates a preference for sulfation on GalNAc residues toward the reducing end. The locations of sulfates on glycopeptides **19** and **20** were determined analogously (Figure 2.4).

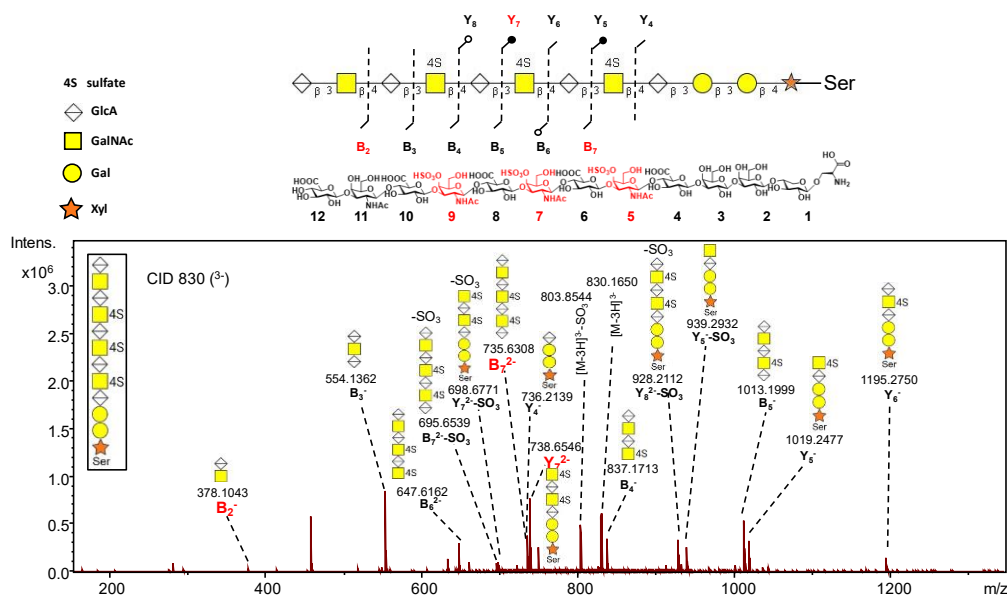


Figure 2.1. CZE-FT-ICR MS fragmentation pattern of glycopeptide **21**. The dashed lines on the structure indicate fragments with no sulfate loss observed. Black filled circles on the sequence indicate both fragment ions with sulfate loss and without sulfate loss were observed. The empty circle indicates a fragment ion with sulfate loss was observed.

2.2.4 Determination of the dissociation constant of CS-bearing glycopeptide with Cathepsin G by Bio-layer interferometry (BLI)

With structurally defined glycopeptides in hand, we aim to better understand the structural features needed for bikunin glycopeptide binding. Cathepsin G (CatG), a neutrophil serine protease, plays a crucial role in regulating various physiological processes, including inflammation, digestion, smooth muscle contraction, and tissue remodeling.⁵⁸ While CatG has been reported to interact with CSPG, the structural requirement of CSPG binding has not been established.⁵⁹ To measure the binding affinity towards CatG, peptides **5** and glycopeptides **10**, **18-21**, as well as

biotinylated 50 kDa CS and CS-A were biotinylated and immobilized on streptavidin coated sensors for biolayer interferometry (BLI) studies. As shown in **Table 2.1**, biotinylated peptide **5**, the tetrasaccharide linkage region bearing glycopeptide **10**, and the non-sulfated chondroitin dodecasaccharide glycopeptide **18** exhibited similar dissociation constants (K_D) of 400, 650, and 490 nM, respectively. Interestingly, sulfated 8-mer CS-bearing glycopeptide **19**, sulfated 10-mer CS-bearing glycopeptide **20**, and sulfated 12-mer CS-bearing glycopeptide **21** showed 10-fold lower K_D values at 49, 74, and 56 nM, indicating that sulfation on the glycopeptide could significantly enhance binding. The K_D values of commercially available CS and CS-A polymers with CatG were 180 and 220 nM, respectively. Collectively, these results demonstrate that both the sulfated glycan and the peptide backbone can participate in interaction with CatG.

Compound	K_D (nM) with Cathepsin G
5	400 ± 40
10	650 ± 80
18	490 ± 18
19	49 ± 5
20	74 ± 7
21	56 ± 8
50 kDa CS	180 ± 10
50 kDa CS-A	220 ± 20

Table 2.1. BLI experiments determined dissociation constants of biotinylated compounds **5**, **10**, **18-21**, 50 kDa CS and 50 kDa CS-A.

2.2.5 Modeling of bikunin glycopeptide binding with cathepsin G (collaboration with Dr. Angela Wilson)

To better understand how sulfates can enhance the binding with CatG, molecular modeling was performed. We prepared the structures of peptide **5**, non-sulfated dodecasaccharide glycopeptide **18** and three sulfated glycopeptides **19-21** and CatG (PDB: 1T32) using Molecular Operating Environment (MOE). CatG has a highly charged surface with multiple arginine residues (**Figure 2a**). All three sulfates of glycopeptide **21** were observed to interact with residues from CatG. The *O*-sulfate on GalNAc 5 residue interacts with R147 and R188 through hydrogen bonding, similarly, the *O*-sulfates on GalNAc 7 and 9 residues also form hydrogen bonds with the R148 residue (**Figure 2.3**). In the case of octasaccharide CSPG **19**, its *O*-sulfate forms hydrogen bonds with R185 and R186 (**Figure 2.6 b**). For the decasaccharide CSPG **20**, and its *O*-sulfate on GalNAc 7 forms hydrogen bonds with R239 (**Figure 2.7 c**). For the unsulfated glycopeptide **18**, the surface arginine residues of CatG do not directly form hydrogen bonding with the glycopeptide.

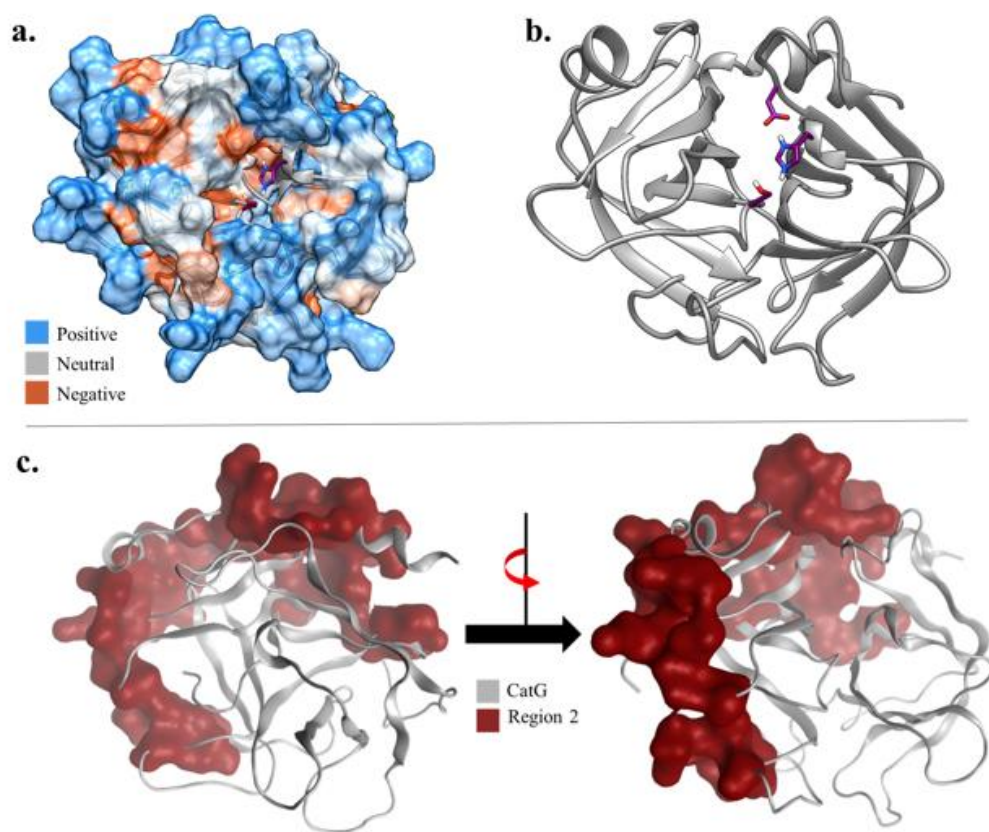


Figure 2.2. The surface and the binding surface of CatG protein. (a) Overall surface of CatG is shown in electrostatic representation with predominant positive charged residues. Blue color represents the positive charge and red color indicates the negative charge. (b) The catalytic residues, His57, Asp102, Ser195, of CatG are shown. (c) The selected region surface of CatG protein for glycopeptide binding. CatG protein has the same orientation in all figures. The numbering of the residues follows the 1T32 PDB numbering.

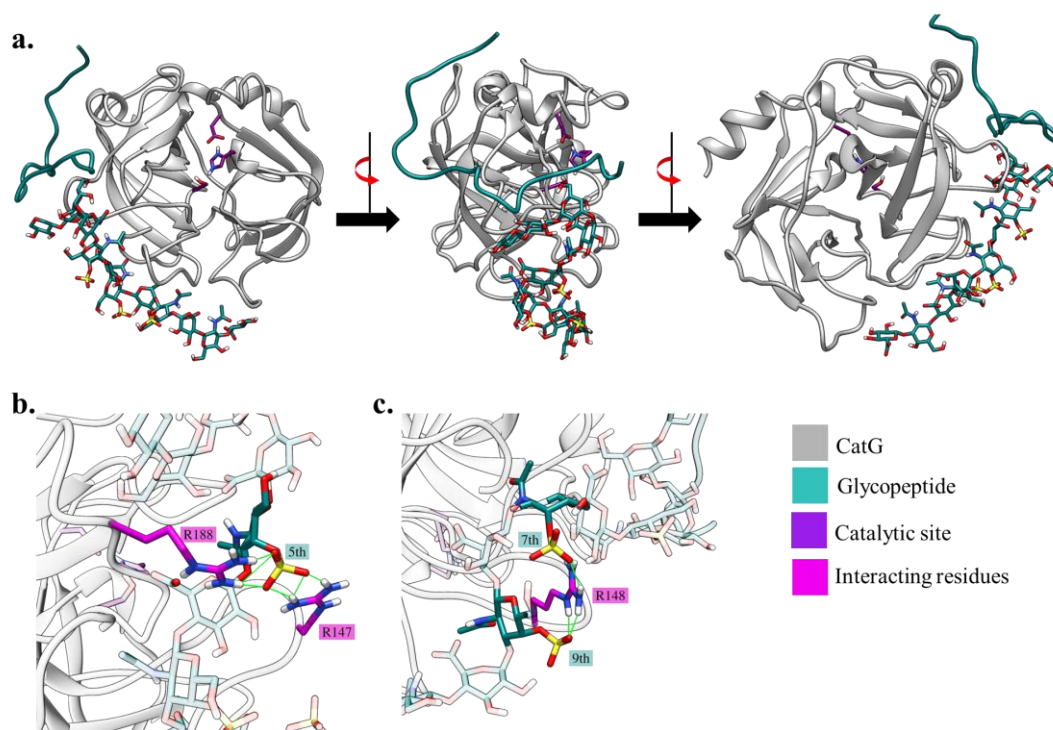


Figure 2.3. (a) The highest scoring pose of **21** with CatG. (b) The direct hydrogen bond interactions of sulfate on GalNAc 5. (c) The direct hydrogen bond interactions of sulfates on GalNAc 7 & 9. The numbering of the residues follows the 1T32 PDB numbering.

2.3 Outlook

Chemoenzymatic synthesis of CSPG glycopeptides has been successfully developed for the first time. This approach comprises several key elements. The commercially available Sepharose beads were utilized for multi-step enzymatic reactions and proved robust under the enzymatic reaction toward CSPG and hydrazine mediated cleavage conditions. The diethyl squarate selected was as a traceless linker, which is compatible with the CS glycopeptide, yielding the native glycopeptide as the final product after cleavage from Sepharose. The GAG-initiating transferase XT-1 transferred the first saccharide, xylose, to the peptide substrate, obviating the need to synthesize the glycosylated amino acid module for glycopeptide formation. To the best of

our knowledge, this represents the first example of direct glycosylation of peptides on solid phase rather than using a glycopeptide as an enzymatic substrate to initiate the synthesis. All requisite enzymes have been obtained and the necessary reaction sequence has been identified enabling the synthesis of PGs. With this approach, we have successfully synthesized five distinct tetrasaccharide linkage region-bearing glycopeptides, varying in peptide length, number of glycosylation sites, and polarity of amino acid residues flanking the glycosylation sites. Furthermore, this method is efficacious in producing CS-bearing glycopeptides including the longest GAG-bearing glycopeptide synthesized to date. Compared to chemical synthesis, this new chemoenzymatic strategy reduced the total number of synthetic steps required by more than 80%.

The availability of the various well-defined synthetic bikunin glycopeptides enabled the binding study with CatG. The presence of sulfate on the glycopeptide significantly enhances its affinity towards CatG, implying that CatG may potentially interact with bikunin *in vivo*. Docking studies provide further insights into the interactions between sulfates and residues on CatG, thereby shedding light on the mechanism by which bikunin may engage with its respective protein partner. Therefore, the efficient chemoenzymatic strategy developed opens new avenues to synthesize and investigate the biological functions of PGs.

2.4 Experimental Section

2.4.1 Materials

Plasmids for XT-1 and B4GALT7 were previously documented,^{36, 39} while FAM20B and B3GALT6 plasmids were graciously provided by Dr. Jack Dixon and Dr. Kelley Moremen. The XYLP and B3GAT3 plasmids were constructed following established literature methods.^{45, 60} Enzymes and substrates, including KfoC, CS4OST, UDP-GalNAc, and PAPS, were generously gifted by Dr. Jian Liu. The Expi 293 Expression system, along with Coomassie Brilliant Blue G-

250, DTT, and EAH Sepharose, were purchased from Thermo Fischer Scientific (Waltham, MA). Nickel columns and Nickel resins, SDS-PAGE gels, 10x Tris/Glycine/SDS electrophoresis buffer, prestained protein ladder, sample loading buffer, and Coomassie Blue R-250 were obtained from Bio-rad (Hercules, California). Shrimp alkaline phosphatase (rSAP) was acquired from NEB (Ipswich, MA). Diethyl squarate, UDP-galactose, UDP-glucuronic acid, and ATP were sourced from Sigma Aldrich (St. Louis, MO). UDP-xylose was purchased from the Complex Carbohydrate Research Center (Athens, Georgia). The peptides were synthesized by Synpeptide (China), and syringes with frit were procured from Torviq (Tucson, AZ). The 50 kDa CS and 50 kDa CS-A were purchased from HAworks (Bedminster, NJ). Cathepsin G, Human Neutrophil was purchased from Athens Research & Technology, Inc. (Athens, Georgia). All other chemicals were purchased from commercial sources and used without additional purifications unless otherwise noted.

2.4.2 General Information

High-performance liquid chromatography was carried out with two systems: LC-8A Solvent Pumps, DGU-14A Degasser, SPD-10A UV-Vis Detector, SCL-10A System Controller (Shimadzu Corporation, JP); G7111B 1260 quat pump, G7129A 1260 vial sampler, G7114A 1260 VWD, G1364F 1260 FC-AS, G1328C 1260 Man. Inj. (Agilent Technologies, CA) and Vydac 218TP 10 μ m C18 Preparative HPLC column (HICHRON Limited, VWR, UK) or 20RBAX 300SB-C18 Analytical HPLC column (Agilent Technologies, CA) using HPLC-grade acetonitrile (EMD Millipore Corporation, MA) and Milli-Q water (EMD Millipore Corporation, MA). A variety of eluting gradients were set up on LabSolutions software (Shimadzu Corporation, JP) and Agilent Open lab control panel (Agilent Technologies, CA). The dual wavelength UV detector was set at 220 nm and 254 nm for monitoring the absorbance from the amide and aromatic region.

NMR data were obtained with Bruker 600 and 800 MHz NMR (Bruker, MA) at ambient temperature.

2.4.3 General procedure of peptide conjugation to EAH Sepharose

A solution of 10 mg of peptide in carbonate buffer (25 μ L, 0.1 M, pH 8) was added to a 1.5 mL Eppendorf. Diethyl squarate (3 equiv.) was diluted with MeOH (25 μ L) and then added to the peptide mixture. The pH value of the mixture was adjusted to 8 and incubated for 6h at RT until no starting material was observed by LCMS. Upon completion, mixtures were lyophilized, the resulting white solid dissolved in carbonate buffer (1 mL, 0.1 M, pH 8). EAH Sepharose (9 μ mol amine per mL of drained Sepharose, 2 equiv. of peptide) was washed with water (20 mL) twice, carbonate buffer (20 mL, 0.1 M, pH 8) once and transferred to a syringe (10 mL) with frit. The peptide carbonate mixture was added to the syringe and agitated with end-to-end rotation for 1 day at room temperature (RT) until no squarate conjugated peptide was observed by LCMS in the supernatant.

2.4.4 β 3GAT3 expression, purification and characterization

β 3GAT3-expressing BL21 competent cells were cultured onto a kanamycin containing petri dish, which was incubated at 37 °C overnight. One colony of BL21 cells was picked and inoculated into 10 mL Luria-Bertani (LB) starter culture containing kanamycin (50 μ g/mL). The cell culture was incubated at 37 °C overnight. The starter culture was then transferred into autoclaved 1L LB medium (with 30 mg/L kanamycin) and incubated at 37 °C with shaking at 250 rpm. When the OD₆₀₀ reached 1.0. IPTG (1 mM) was added to induce protein expression at 23 °C for 16 hours. Cells were centrifuged at 4 °C, 10,000 g for 10 min. Cell pellet was lysed using 1X Cellytic in buffer (10 mL), 50 U/mL benzonase, 0.2 mg/mL lysozyme and 1 tablet of cComplete™ Protease Inhibitor Cocktail EDTA-free for 20 min at ambient temperature. Clarified lysate was

purified by a nickel column (a. washing buffer: 20 mM phosphate, 0.5 M NaCl and 40 mM imidazole; b. eluting buffer: 20 mM phosphate, 0.5 M NaCl and 40-250 mM imidazole). Protein purity was confirmed with SDS-PAGE gel electrophoresis and the concentration and expression yield were determined by the standard Bradford assay.

2.4.5 FAM20B, B3GALT6 and XYLP expression, purification and characterization

Expi293F cells were grown in FreeStyle™ 293 Expression Medium on a platform shaker in humidified 37 °C CO₂ (8%) incubator with rotation at 150 rpm. When the cell density reached between 4×10^5 and 3×10^6 cells/ml, cells were split to a density of 1.5×10^6 cells/ml and cultured 1 day with fresh medium. Desired plasmid (1 µg plasmid per ml medium) was diluted with Opti-MEM I Reduced-Serum Medium then mixed with ExpiFectamine™ 293 following the manufacturing protocol. This mixture was incubated for 15 min at RT then dropwise into the cells. At this point, cell density should be around 3×10^6 cells/ml. The flask was returned to the shaker platform in the incubator. After 1 day, transfection enhancer was added. Six days after the transfection, medium was harvested. Clarified medium was purified by nickel column (a. washing buffer: 20 mM Tris, 0.5 M NaCl and 40 mM imidazole; b. eluting buffer: 20 mM Tris, 0.5 M NaCl and 250 mM imidazole). Protein purity was confirmed with SDS-PAGE gel electrophoresis and the concentration and expression yield were determined by standard Bradford assay.

2.4.6 General procedure of enzymatic glycosylation on peptide-conjugated Sepharose

Step 1:

Peptide conjugated Sepharose (1 mL, 50% loading) was drained and then resuspended in 4 mL of XT-1 reaction buffer. This buffer contained the following components: 25 mM MES, 25 mM KCl, 5 mM KF, 5 mM MgCl₂, 5 mM MnCl₂, pH 6.5. To this resuspended mixture, 100 µg of XT-1 and UDP-xylose (2 equiv. relative to the number of reactive sites) were added. The resulting

reaction mixture was agitated with end-to-end rotation at 4 °C for a duration of 12 h. It is important to perform the reaction at 4 °C as higher reaction temperature tends to lead to precipitation of the enzyme during the agitated reaction. Subsequently, the mixture was filtered, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated one time.

Step 2:

The drained xylosylated peptide conjugated Sepharose from step 1 was resuspended in 4 mL of a B4GALT7 reaction buffer. This buffer consisted of 20 mM MES and 10 mM MnCl₂, pH 6.2. To this resuspended mixture, 250 µg of B4GALT7 and UDP-Gal (2 equiv. per SG sites) were added. The resulting reaction mixture was agitated with end-to-end rotation at 4 °C for 12 h. Subsequently, the mixture was subjected to filtration, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated.

Step 3:

Drained disaccharide glycopeptide conjugated Sepharose from step 2 (1 mL, 50% loading) was resuspended in 4 mL of a FAM20B reaction buffer. This buffer solution consisted of 50 mM HEPES and 10 mM MnCl₂, pH 7.4. To this resuspended mixture, 200 µg of FAM20B and ATP (3 equiv. per SG sites) were added. The resulting reaction mixture was subjected to end-to-end rotation at 4 °C for a duration of 12 h. Subsequently, the mixture was filtered, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated.

Step 4:

Drained phosphorylated disaccharide glycopeptide conjugated Sepharose (1 mL, 50% loading) from step 3 was resuspended using 4 mL of a B3GALT6 reaction buffer. This buffer solution consisted of 50 mM MES, 10 mM MnCl₂, 100 mM NaCl, pH 6.0. To this resuspended

mixture, 200 µg B3GALT6, UDP-Gal (0.6 equiv. per SG sites) were added. The resulting reaction mixture was mixed with end-to-end rotation at 4 °C for a duration of 12 h. Subsequently, the mixture was filtered, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated.

Step 5:

Drained phosphorylated trisaccharide glycopeptide conjugated Sepharose (1 mL, 50% loading) from step 4, was resuspended in 4 mL of a XYLP reaction buffer. This buffer solution consisted of 50 mM Tris-HCl, pH 5.8. To this reaction mixture, 200 µg XYLP was added. The resulting reaction mixture was subjected to end-to-end rotation at 4 °C for a duration of 12 h. Subsequently, the mixture was filtered, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated.

Step 6:

Drained trisaccharide glycopeptide conjugated Sepharose (1 mL, 50% loading) from step 5, was resuspended in 4 mL of a B3GAT3 reaction buffer. This buffer solution consisted of 50 mM MES, 2 mM MnCl₂, pH 6.5. To this mixture, 500 µg B3GAT3, UDP-GlcA (2 equiv. per SG site) were added. The resulting reaction mixture was subjected to end-to-end rotation at 4 °C for 12 h. Subsequently, the mixture was filtered, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated.

Step 7:

Drained tetrasaccharide glycopeptide conjugated Sepharose (1 mL, 50% loading) from step 6, was resuspended in 4 mL of a KfoC reaction buffer. This buffer solution consisted of 50 mM MOPS, 15 mM MnCl₂, pH 7.2. To this resuspended mixture, 100 µg KfoC and UDP-GalNAc (12.5 mM). The resulting reaction mixture was subjected to end-to-end rotation at 4 °C for 12 h.

Subsequently, the mixture was filtered, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated.

Step 8:

Drained pentasaccharide glycopeptide conjugated Sepharose (1 mL, 50% loading) from step 7, was resuspended in 4 mL of a KfoC reaction buffer. This buffer solution consisted of 50 mM MOPS, 15 mM MnCl₂, pH 7.2. To this resuspended mixture, 100 µg KfoC and UDP-GlcA (2 equiv. per SG sites). The resulting reaction mixture was subjected to end-to-end rotation at 4 °C for a duration of 12 h. Subsequently, the mixture was subjected to filtration, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated.

Step 9:

Drained CS backbone-bearing glycopeptide conjugated Sepharose (1 mL, 50% loading) from step 8 was resuspended in 4 mL of a CS4OST reaction buffer. This buffer solution consisted of 50mM MOPS, 10mM CaCl₂, fresh 2 mM DDT, pH 6.5. To this resuspended mixture, 400 µg CS4OST and PAPS (2 equiv. per SG sites). The resulting reaction mixture was subjected to end-to-end rotation at 4 °C for 12 h. Subsequently, the mixture was filtered, followed by two washes with a total volume of 20 mL of water each time. This entire process was then repeated.

2.4.7 General procedure of glycopeptide biotinylation

To a solution of peptide or glycopeptide in DMSO/H₂O (1/1, 0.1 ml) was added NHS-LC-Biotin (4 equiv.) and DIPEA (pH ~ 8.5). The reaction was incubated at 37 °C for 2 h. Then the mixture was dried *in vacuo* and purified by HPLC.

2.4.8 CZE-FT-ICR MS Analysis

Capillary zone electrophoresis (CZE) was performed using a CMP ECE-001 CZE system (CMP scientific, Brooklyn, NY). The CZE was interfaced to the mass spectrometer with an

electrokinetically pumped sheath flow CE-MS interface (EMASS-II interface, CMP Scientific). Mass spectra were collected in negative mode on Bruker 9.4 T Solarix FT-ICR mass spectrometer (Bruker Daltonics, Bremen, Germany). Mass spectra were collected between m/z 150 – 3000 with 1M data points and a 0.5592 s transient. Ion accumulation time was set to 0.3 s and the time of flight (TOF) was set to 0.8 ms. The flow rate of the drying gas was set to 2 L/min at 180 °C. The inlet capillary voltage of the FT-ICR was set to 0 V.

CZE separations were performed using fused silica capillaries (60 cm x 360 μ m OD x 50 μ m ID) functionalized with dichlorodimethylsilane (DMS) neutral coated capillary. DMS functionalization and HF etching procedures have been reported previously⁶¹. A 130 cm long functionalized capillary was segmented into two equal length pieces to ensure the uniformity of the internal derivatization. The final 10 mm of the outlet end of each capillary was etched with hydrofluoric acid to a conical shape with an outer diameter at the terminus of < 100 μ m to reduce the mixing volume for analytes entering the sheath flow interface.

Ammonium formate solution (25 mM) in 70% (v/v) methanol/water was used as a sheath liquid (SL) and a background electrolyte (BGE). The etched ends of both functionalized capillaries were positioned 0.5 mm from the tip of a borosilicate glass emitter orifice (0.75 mm ID, 5.0 cm length and 20 μ m opening diameter of tip). The distance between the emitter opening and the inlet of MS was ca. 2.5 mm. The potential difference between the spray tip and the entrance to the mass spectrometer ESI inlet was -2.0 kV voltage. Each sample was injected into a CZE capillary using a pressure of 400 mbar for 10 s, resulting in circa 115 nL volume and 10.6% of the total capillary volume. The capillary was completely rinsed with fresh BGE after each run to remove residual carryover for the next run. In terms of the auto MS/MS mode (CID), a preferential and exclusion list was implemented for auto CID. MS1 scan was performed first followed by three MS/MS scans

under the external ion accumulation time of 0.5 s. The collision voltage was fixed between 13 and 15 V for each mass spectrum.

Mass spectra were analyzed using Compass Data Analysis v4.1 software (Bruker Scientific, Bremen, Germany), in-house software developed in MATLAB (The MathWorks, Natick, MA) as well as Glycoworkbench to annotate fragment ions.⁶²

2.4.9 General procedure for BLI binding assay

The binding assay was performed on the Octet K2 System (Pall ForteBio). The biotinylated compounds were incubated with streptavidin (SA) sensors for 2 min. The sensor was then balanced in the assay buffer (PBS containing 0.005% P20) and dipped into Cathepsin G solution in assay buffer at 2000 nM, 1000 nM, 500 nM, 250 nM concentrations. After 5 min of association, the sensor was brought back to the previous assay buffer for a 5 min dissociation step. At the end of the assay, the sensor was regenerated in 2 M NaCl to remove the bound protein. Each measurement was repeated 3 times on the same sensor. The control assay was performed with another sensor loaded with a 2 mM biotin solution.

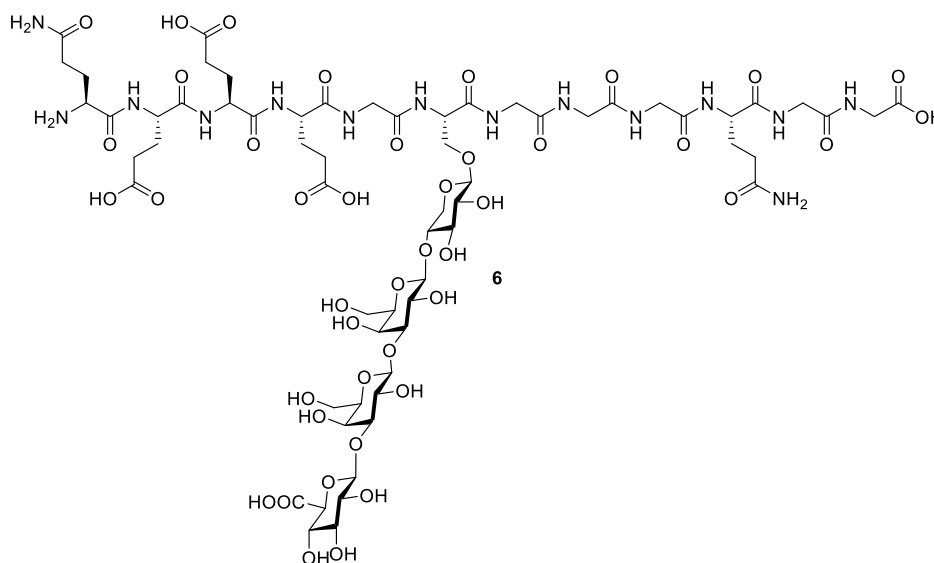
2.4.10 Docking methods

The protein structure of Cathepsin G (PDB ID: 1T32, Res.: 1.85 Å) was prepared using Molecular Operating Environment (MOE 2022.02) with structure preparation module. The protonation states of the titratable residues were determined using PropKa at pH 7. The glycopeptides were prepared on CharmmGUI, with ff14SB and GLYCAM forcefields for peptide and glycan, respectively. The glycopeptide structures were minimized in MOE and used for docking.

MOE is used for the docking procedure, which involves two different approaches. First, all CatG surface was selected as a potential binding site for the glycopeptides, and rigid docking was

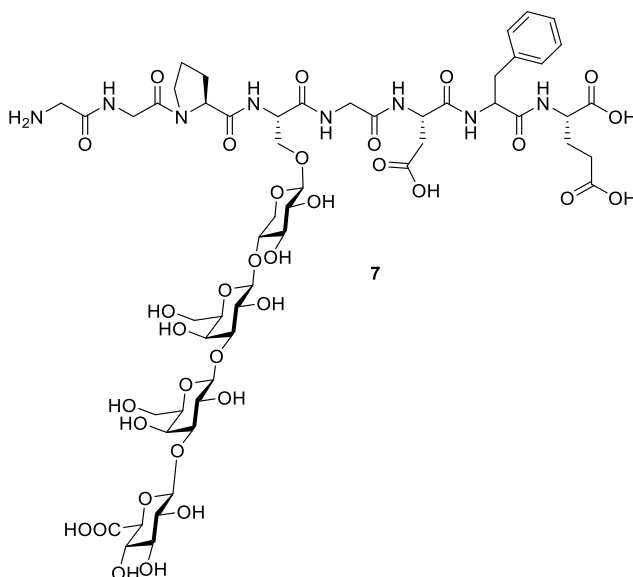
performed with GBVA/WSA scoring function, reporting 10 poses. This process was repeated three times, resulting in 30 accumulated poses for 579S glycopeptide (**Figure S1**). Region is determined by the Site Finder algorithm of MOE by only selecting predicted site with lowest hydrophobicity score (**Figure 2(c)**). For the highest scoring poses, the pose is minimized together with CatG protein and rescored.

2.4.11 Product characterization



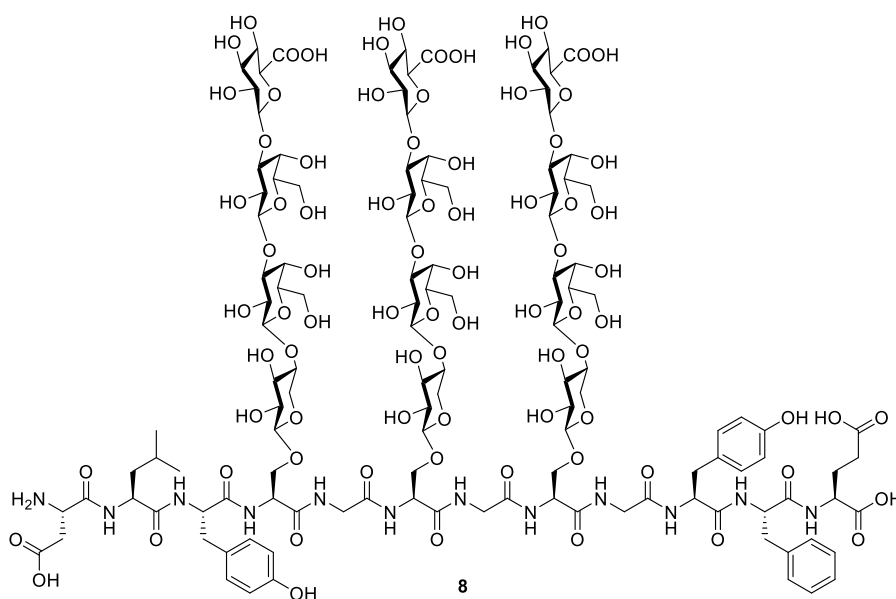
Peptide **1** (10 mg, 9.1 μmol) was conjugated to EAH Sepharose (2 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 6 of the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. Crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried *in vacuo* and purified by prep C-18 HPLC (0-10% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **9** (1.5 mg) in 9.4% yield. ^1H NMR (800 MHz, D_2O) δ 4.66 – 4.60 (m, 3H), 4.46 (d, J = 7.9 Hz, 1H), 4.40 – 4.31 (m, 6H), 4.19 – 4.15 (m, 1H), 4.14 – 4.12 (m, 2H), 4.06 – 4.02 (m, 2H), 3.98 – 3.89 (m,

11H), 3.85 (d, $J = 10.2$ Hz, 2H), 3.80 – 3.60 (m, 11H), 3.55 (t, $J = 9.0$ Hz, 1H), 3.50 – 3.45 (m, 2H), 3.39 – 3.32 (m, 2H), 3.26 (t, $J = 8.6$ Hz, 1H), 2.52 – 2.34 (m, 9H), 2.33 (t, $J = 7.9$ Hz, 1H), 2.16 – 2.00 (m, 6H), 2.00 – 1.90 (m, 4H); ^{13}C NMR (201 MHz, D_2O) δ 182.23, 177.82, 177.22, 177.07, 176.76, 175.27, 174.73, 173.75, 173.72, 173.18, 173.10, 173.00, 172.04, 172.01, 171.98, 171.59, 171.46, 171.33, 171.30, 169.25, 169.09, 103.88, 103.55, 102.79, 101.28, 82.29, 82.26, 81.91, 76.27, 75.85, 75.14, 74.88, 74.70, 73.66, 72.98, 72.64, 71.57, 70.04, 69.73, 68.63, 68.36, 67.97, 62.92, 60.99, 60.85, 56.67, 53.62, 53.14, 53.09, 53.03, 52.90, 52.16, 42.65, 42.42, 42.36, 42.33, 30.96, 30.13, 30.05, 29.85, 26.51, 26.36, 26.26, 26.18, 26.15, 26.08, 25.96. HRMS (ESI) m/z : $[\text{M} + 2\text{H}]^{2+}$ Calcd for Chemical Formula: $\text{C}_{63}\text{H}_{100}\text{N}_{14}\text{O}_{42}$ 862.3051; Found 862.3098.



Peptide **2** (10 mg, 13 μmol) was conjugated to EAH Sepharose (2.9 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 6 of the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. Crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried *in vacuo* and purified by prep C-18 HPLC (0-30%

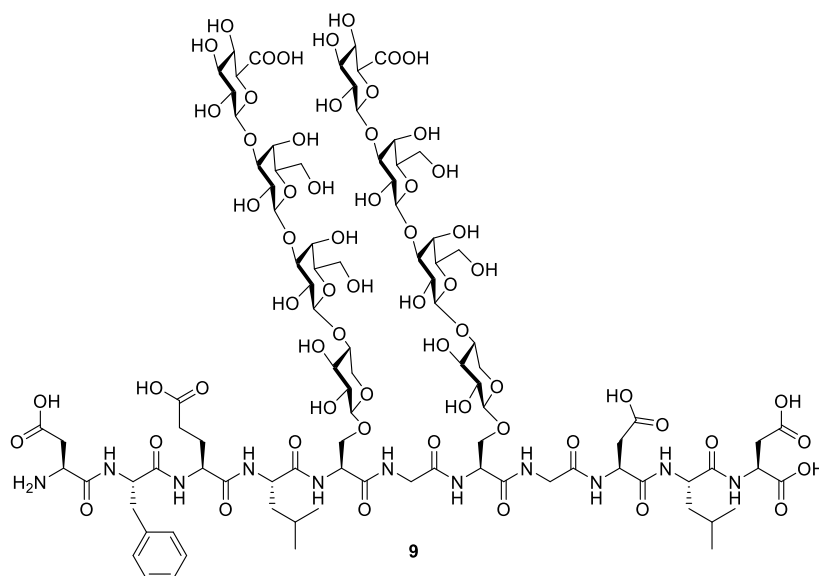
acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **6** (980 μg) in 5.4% overall yield from peptide **1**. ^1H NMR (800 MHz, D_2O) δ 7.28 – 7.15 (m, 5H), 4.63 – 4.52 (m, 2H), 4.51 – 4.45 (m, 1H), 4.43 – 4.32 (m, 3H), 4.15 – 4.07 (m, 4H), 4.05 – 4.02 (m, 1H), 4.01–3.97 (m, 1H), 3.93 – 3.83 (m, 2H), 3.87 – 3.76 (m, 3H), 3.75 – 3.49 (m, 15H), 3.44 – 3.40 (m, 2H), 3.35 – 3.26 (m, 2H), 3.26 – 3.20 (m, 1H), 3.12 (dd, J = 13.8, 5.3 Hz, 1H), 2.92 (dd, J = 13.8, 5.3 Hz, 1H), 2.52 (dd, J = 16.0, 8.7 Hz, 1H), 2.37 (dd, J = 16.0, 8.7 Hz, 1H), 2.24 – 2.19 (m, 1H), 2.13 (t, J = 7.2 Hz, 2H), 2.00 – 1.86 (m, 4H), 1.82–1.75 (m, 1H); ^{13}C NMR (201 MHz, D_2O) δ 180.56, 177.77, 177.12, 175.96, 174.56, 172.80, 172.00, 171.34, 170.62, 169.46, 167.49, 164.98, 136.48, 129.28, 128.66, 127.05, 103.91, 103.57, 102.95, 101.33, 82.40, 81.97, 76.34, 76.18, 75.26, 74.89, 74.75, 73.72, 73.10, 72.68, 71.73, 70.10, 69.78, 68.60, 68.41, 67.97, 62.94, 61.01, 60.92, 54.75, 53.49, 51.37, 47.06, 42.44, 41.74, 40.35, 38.02, 36.87, 32.70, 29.38, 27.99, 24.43. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for Chemical Formula: $\text{C}_{55}\text{H}_{81}\text{N}_8\text{O}_{34}$ 1397.4847; Found 1397.4796.



Peptide **3** (10 mg, 7.8 μmol) was conjugated to EAH Sepharose (1.73 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 6 of the general procedures of enzymatic

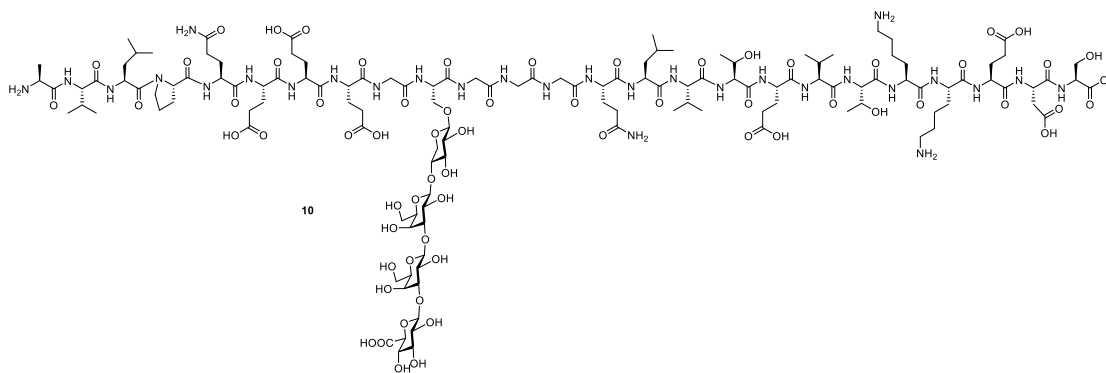
glycosylation on peptide-conjugated Sepharose. Crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried *in vacuo* and purified by prep C-18 HPLC (0-30% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound 7 (2.63 mg) in 10.5% overall yield.

^1H NMR (800 MHz, D_2O) δ 7.30 – 7.27 (m, 2H), 7.25 – 7.22 (m, 1H), 7.18 – 7.16 (m, 2H), 7.09 – 7.07 (m, 2H), 6.95 – 6.92 (m, 2H), 6.78 – 6.72 (m, 4H), 4.65 – 4.58 (m, 7H), 4.58 – 4.53 (m, 3H), 4.51 (t, J = 8.1 Hz, 1H), 4.45 (d, J = 7.9 Hz, 1H), 4.41 (t, J = 7.7 Hz, 1H), 4.39 – 4.32 (m, 3H), 4.32 – 4.28 (m, 2H), 4.27 – 4.24 (m, 1H), 4.22 – 4.17 (m, 2H), 4.16 – 4.07 (m, 8H), 4.05 – 3.93 (m, 5H), 3.92 – 3.78 (m, 7H), 3.78 – 3.57 (m, 36H), 3.56 – 3.50 (m, 4H), 3.50 – 3.46 (m, 6H), 3.39 – 3.35 (m, 3H), 3.34 – 3.21 (m, 6H), 3.07 (dd, J = 13.9, 6.4 Hz, 1H), 3.01 (dd, J = 14.0, 7.5 Hz, 1H), 2.94 – 2.87 (m, 2H), 2.85 – 2.73 (m, 4H), 2.33 – 2.27 (m, 2H), 2.09 – 2.02 (m, 1H), 1.86 (sext, J = 6.7 Hz, 1H), 1.48 – 1.40 (m, 2H), 1.40 – 1.34 (m, 1H), 0.83 (d, J = 6.1 Hz, 3H), 0.79 (d, J = 6.1 Hz, 3H); ^{13}C NMR (201 MHz, D_2O) δ 177.34, 175.62, 175.19, 174.87, 174.76, 173.79, 173.15, 172.28, 171.97, 171.54, 171.11, 170.68, 170.57, 169.17, 154.51, 136.30, 130.59, 130.39, 129.28, 128.61, 127.92, 127.04, 115.39, 103.89, 103.55, 102.90, 102.80, 101.44, 101.41, 101.25, 82.30, 76.72, 76.27, 75.78, 75.27, 75.14, 74.91, 74.84, 74.75, 73.63, 73.14, 72.99, 72.66, 71.69, 70.08, 69.76, 68.63, 68.15, 67.96, 62.87, 62.90, 60.85, 55.16, 54.59, 53.61, 53.48, 53.03, 52.70, 50.21, 42.51, 39.62, 37.02, 36.22, 35.93, 30.12, 26.23, 24.13, 21.86, 20.92. HRMS (ESI) m/z : $[\text{M} - 2\text{H}]^{2-}$ Calcd for Chemical Formula: $\text{C}_{126}\text{H}_{182}\text{N}_{12}\text{O}_{82}$ 1587.5228; Found 1587.5127.



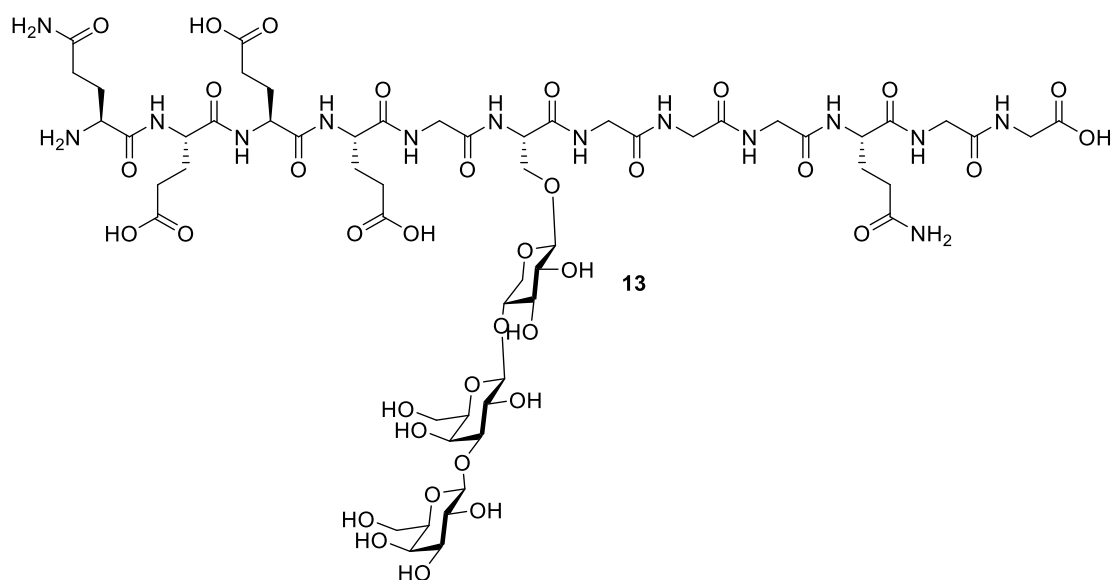
Peptide **4** (10 mg, 8.6 μ mol) was conjugated to EAH Sepharose (1.92 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 6 of the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. Crude products were obtained after incubating the glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried *in vacuo* and purified by prep C-18 HPLC (0-30% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **8** (4 mg) in 19.5% yield. ^1H NMR (800 MHz, D_2O) δ 7.34 – 7.30 (m, 2H), 7.29 – 7.26 (m, 1H), 7.23 – 7.20 (m, 2H), 4.68 – 4.56 (m, 9H), 4.37 (dd, J = 10.2, 8.0 Hz, 2H), 4.38 (t, J = 8.0 Hz, 2H), 4.34 – 4.31 (m, 1H), 4.30 – 4.27 (m, 2H), 4.20 – 4.16 (m, 1H), 4.16 – 4.08 (m, 6H), 4.08 – 4.00 (m, 3H), 4.00 – 3.94 (m, 2H), 3.93 – 3.87 (m, 4H), 3.81 – 3.59 (m, 23H), 3.55 (t, J = 8 Hz, 2H), 3.50 – 3.45 (m, 4H), 3.39 – 3.31 (m, 4H), 3.28 (q, J = 8.5 Hz, 2H), 3.07 (dd, J = 13.9, 8.0 Hz, 1H), 3.01 (dd, J = 13.9, 8.0 Hz, 1H), 2.91 – 2.72 (m, 6H), 2.36 – 2.31 (m, 2H), 2.02 – 1.96 (m, 1H), 1.87 – 1.82 (m, 1H), 1.65 – 1.52 (m, 6H), 0.83 (d, J = 5.5 Hz, 3H), 0.90 – 0.86 (m, 6H), 0.81 (d, J = 5.5 Hz, 3H); ^{13}C NMR (201 MHz, D_2O) δ 177.25, 175.03, 174.86, 174.52, 173.82, 172.59, 172.28, 172.21,

171.70, 171.45, 171.35, 170.95, 168.87, 135.95, 129.07, 128.73, 127.22, 103.88, 103.55, 102.92, 102.88, 101.28, 82.28, 81.92, 76.26, 75.68, 75.12, 74.86, 74.69, 73.65, 73.02, 72.97, 72.61, 71.55, 70.04, 69.72, 68.56, 68.37, 68.07, 67.98, 62.91, 60.97, 60.85, 55.23, 53.65, 53.48, 52.53, 52.43, 52.34, 50.16, 49.97, 42.58, 42.51, 39.53, 39.38, 36.64, 36.18, 36.06, 35.52, 29.95, 26.35, 24.23, 24.12, 22.20, 22.09, 20.88, 20.47. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for Chemical Formula: $C_{94}H_{145}N_{11}O_{62}$ 1209.9257; Found 1209.9203.

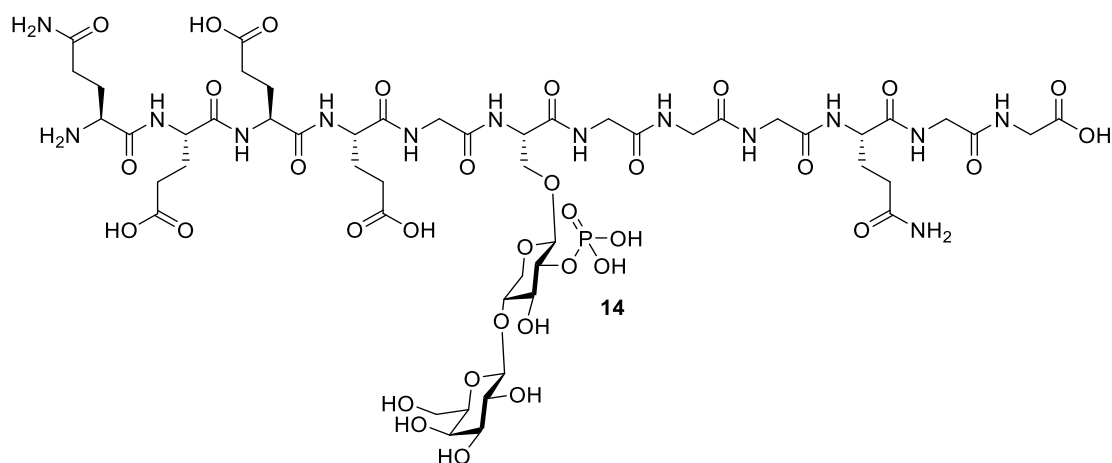


Peptide **5** (10 mg, 3.86 μ mol) was conjugated to EAH Sepharose (0.86 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 6 of the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. Crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried *in vacuo* and purified by prep C-18 HPLC (0-50% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **10** (1.2 mg) in 9.5% yield. 1H NMR (800 MHz, D_2O) δ 4.73 – 4.70 (m, 1H), 4.66 (d, J = 7.9 Hz, 1H), 4.63 – 4.58 (m, 3H), 4.46 (d, J = 7.9 Hz, 1H), 4.40 – 4.20 (m, 14H), 4.20 – 4.16 (m, 1H), 4.16 – 4.00 (m, 9H), 4.00 – 3.86 (m, 10H), 3.86 – 3.59 (m, 14H), 3.55 (t, J = 9.2 Hz, 1H), 3.51 –

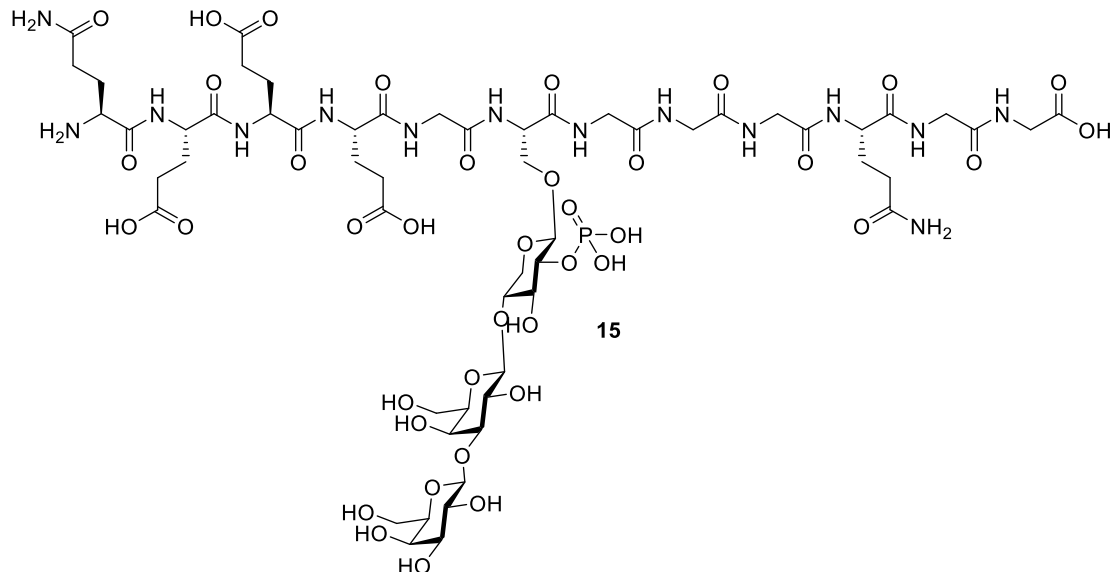
3.45 (m, 2H), 3.39 – 3.31 (m, 4H), 3.25 (t, $J = 8.5$ Hz, 1H), 2.96 – 2.87 (m, 5H), 2.79 (dd, $J = 16.9, 8.2$ Hz, 1H), 2.49 – 2.37 (m, 10H), 2.37 – 2.24 (m, 5H), 2.12 – 1.90 (m, 15H), 1.89 – 1.82 (m, 1H), 1.80 – 1.74 (m, 2H), 1.74 – 1.67 (m, 2H), 1.67 – 1.60 (m, 5H), 1.60 – 1.51 (m, 5H), 1.46 (d, $J = 7.1$ Hz, 3H), 1.43 – 1.30 (m, 4H), 1.17 – 1.12 (m, 6H), 0.98 – 0.84 (m, 27H), 0.82 (d, $J = 5.7$ Hz, 3H); ^{13}C NMR (201 MHz, D_2O) δ 177.77, 177.70, 177.11, 177.02, 176.96, 176.93, 176.89, 174.57, 174.44, 174.18, 173.72, 173.58, 173.44, 173.35, 173.10, 173.07, 172.98, 172.91, 172.46, 172.02, 171.84, 171.52, 171.46, 171.30, 170.60, 103.89, 103.56, 102.78, 101.29, 82.24, 81.91, 76.28, 75.39, 75.09, 74.88, 74.69, 73.67, 72.93, 72.65, 71.46, 70.05, 69.73, 68.61, 68.37, 68.01, 66.88, 62.93, 61.34, 60.99, 60.84, 60.45, 59.74, 59.61, 59.33, 59.18, 58.97, 55.80, 53.67, 53.53, 53.41, 53.30, 53.02, 53.00, 52.92, 52.86, 52.47, 50.28, 50.04, 48.72, 47.82, 42.47, 42.34, 39.06, 38.84, 35.70, 30.96, 30.94, 30.30, 30.21, 29.98, 29.92, 29.25, 26.20, 26.14, 26.06, 26.00, 24.66, 24.28, 24.24, 22.30, 21.95, 21.91, 20.80, 20.44, 18.78, 18.76, 18.32, 18.30, 18.28, 17.82, 17.66, 16.57. HRMS (ESI) m/z : $[\text{M} + 3\text{H}]^{3+}$ Calcd for Chemical Formula: $\text{C}_{131}\text{H}_{216}\text{N}_{29}\text{O}_{64}$ 1073.8223; Found 1073.8217.



Peptide **1** (10 mg, 9.1 μmol) was conjugated to EAH Sepharose (2 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 3 of the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. To monitor the reaction, crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried in vacuo and purified by prep C-18 HPLC (0-10% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **13**. ^1H NMR (800 MHz, D_2O) δ 4.62 (t, J = 4.8 Hz, 1H), 4.56 (d, J = 7.8 Hz, 1H), 4.47 (d, J = 7.9 Hz, 1H), 4.40 – 4.32 (m, 4H), 4.18 (dd, J = 6.9, 5.2 Hz, 1H), 4.14 (d, J = 3.3 Hz, 1H), 4.08 – 4.01 (m, 2H), 4.01 – 3.88 (m, 14H), 3.87 (d, J = 3.4 Hz, 1H), 3.80 – 3.65 (m, 7H), 3.65 – 3.59 (m, 3H), 3.57 – 3.53 (m, 2H), 3.35 (t, J = 11.1 Hz, 1H), 3.26 (t, J = 8.7 Hz, 1H), 2.54 – 2.36 (m, 9H), 2.34 – 2.31 (m, 1H), 2.18 – 2.00 (m, 6H), 2.00 – 1.89 (m, 4H); ^{13}C NMR (201 MHz, D_2O) δ 182.22, 177.81, 176.94, 176.89, 176.88, 176.75, 174.95, 173.76, 173.65, 173.52, 173.14, 173.05, 172.92, 172.01, 171.60, 171.55, 171.42, 169.09, 116.96, 115.51, 104.25, 102.79, 102.77, 101.28, 81.98, 76.27, 74.98, 74.87, 73.66, 72.64, 72.39, 70.91, 69.72, 68.64, 68.46, 62.92, 60.99, 60.86, 56.66, 53.61, 53.15, 53.04, 53.02, 52.97, 52.93, 52.81, 52.77, 52.14, 42.65, 42.42, 42.36, 42.33, 42.29, 41.20, 41.14, 30.95, 29.89, 29.85, 29.75, 29.71, 29.15, 26.50, 26.37, 26.07, 26.06, 26.01, 25.84, 25.82. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for Chemical Formula: $\text{C}_{57}\text{H}_{91}\text{N}_{14}\text{O}_{36}$ 1547.5712; Found 1547.5770.

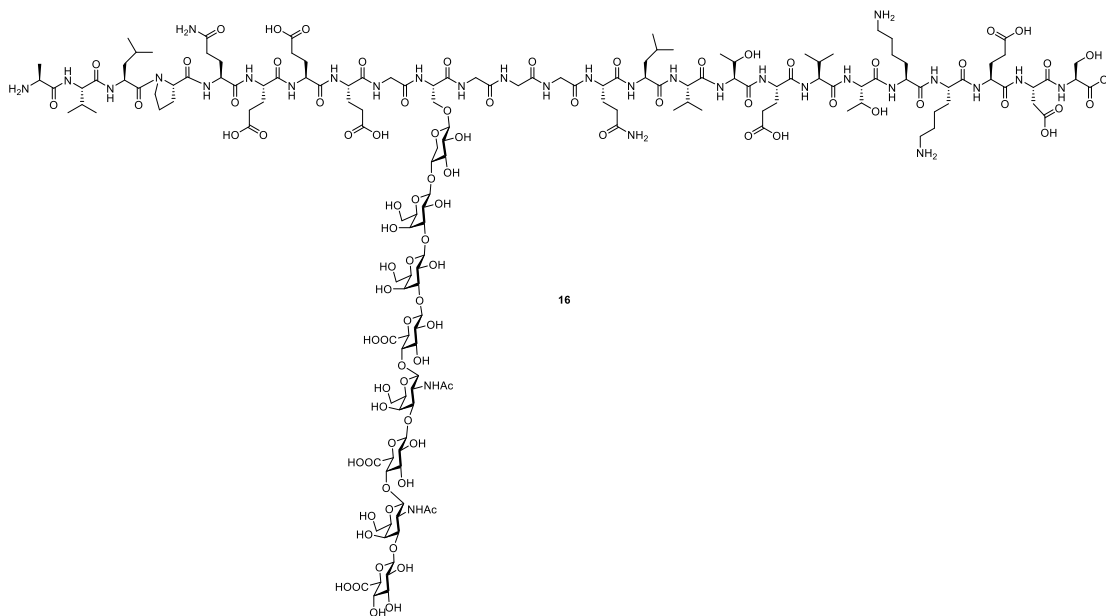


Peptide **1** (10 mg, 9.1 μmol) was conjugated to EAH Sepharose (2 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 3 of the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. To monitor the reaction, crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried in vacuo and purified by prep C-18 HPLC (0-10% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **13**. ^1H NMR (800 MHz, D_2O) δ 4.53 – 4.48 (m, 2H), 4.42 – 4.31 (m, 5H), 4.27 (d, J = 9.8 Hz, 1H), 4.07 – 3.90 (m, 13H), 3.87 – 3.83 (m, 2H), 3.80 – 3.74 (m, 3H), 3.70 – 3.66 (m, 3H), 3.66 – 3.63 (m, 1H), 3.61 – 3.58 (m, 1H), 3.46 (dd, J = 9.7, 7.9 Hz, 1H), 3.36 (t, J = 10.8 Hz, 1H), 2.52 – 2.30 (m, 10H), 2.17 – 2.03 (m, 6H), 1.99 – 1.92 (m, 4H); ^{13}C NMR (201 MHz, D_2O) δ 177.83, 176.90, 176.76, 173.78, 173.45, 172.87, 172.81, 172.67, 172.29, 172.12, 171.65, 171.58, 171.54, 169.08, 101.82, 101.58, 77.03, 75.71, 75.26, 73.13, 72.49, 70.50, 68.95, 68.55, 62.75, 62.36, 61.05, 59.19, 56.67, 54.09, 53.15, 53.06, 52.85, 52.80, 52.15, 42.93, 42.51, 42.38, 42.31, 42.26, 30.95, 29.84, 29.75, 26.49, 26.35, 26.24, 26.07, 26.00, 25.87. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for Chemical Formula: $\text{C}_{51}\text{H}_{82}\text{N}_{14}\text{O}_{34}\text{P}$ 1465.4847; Found 1465.4823.



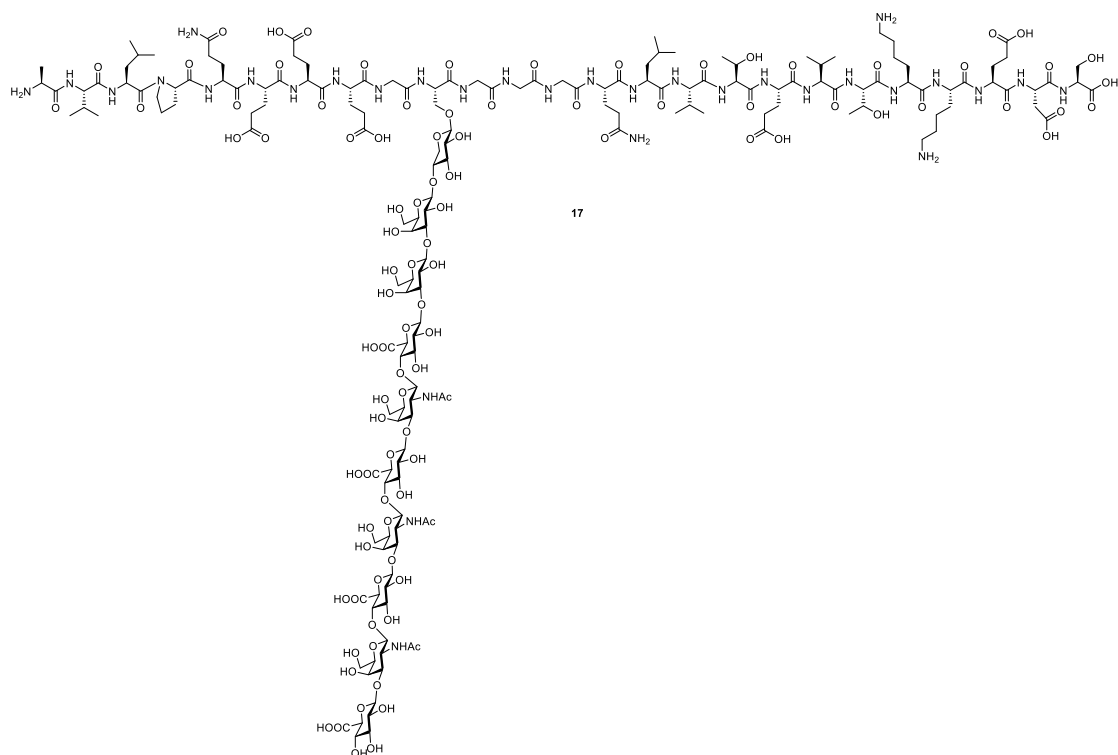
Peptide **1** (10 mg, 9.1 μmol) was conjugated to EAH Sepharose (2 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 3 of the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. To monitor the reaction, crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried *in vacuo* and purified by prep C-18 HPLC (0-10% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **15**. ^1H NMR (800 MHz, D_2O) δ 4.56 (d, $J = 7.8$ Hz, 1H), 4.52 – 4.48 (m, 2H), 4.47 (d, $J = 7.9$ Hz, 1H), 4.42 – 4.31 (m, 4H), 4.29 – 4.26 (m, 1H), 4.14 (d, $J = 3.3$ Hz, 1H), 4.06 – 4.00 (m, 4H), 3.99 – 3.90 (m, 9H), 3.88 – 3.82 (m, 2H), 3.81 – 3.74 (m, 4H), 3.74 – 3.65 (m, 6H), 3.65 – 3.60 (m, 3H), 3.56 – 3.53 (m, 1H), 3.37 (t, $J = 11.4$ Hz, 1H), 2.49 – 2.41 (m, 6H), 2.41 – 2.35 (m, 2H), 2.35 – 2.31 (m, 2H), 2.16 – 2.04 (m, 6H), 2.00 – 1.92 (m, 4H); ^{13}C NMR (201 MHz, D_2O) δ 182.25, 177.80, 176.90, 176.87, 176.86, 176.82, 176.76, 176.75, 173.78, 173.44, 173.19, 173.06, 172.85, 172.79, 172.65, 172.12, 171.65, 169.09, 104.25, 101.83, 101.18, 81.95, 77.08, 77.05, 75.63,

74.97, 74.89, 73.08, 72.39, 70.91, 69.70, 68.97, 68.46, 68.35, 62.72, 60.99, 60.85, 59.19, 56.66, 54.08, 53.16, 53.03, 52.97, 52.83, 52.81, 52.78, 52.15, 42.93, 42.51, 42.38, 42.29, 42.26, 41.00, 40.98, 30.95, 29.84, 29.73, 29.68, 29.65, 26.48, 26.36, 26.21, 26.18, 26.02, 25.97, 25.85. HRMS (ESI) m/z: $[M + 2H]^{2+}$ Calcd for Chemical Formula: $C_{57}H_{93}N_{14}O_{39}P$ 814.2722; Found 814.2753.



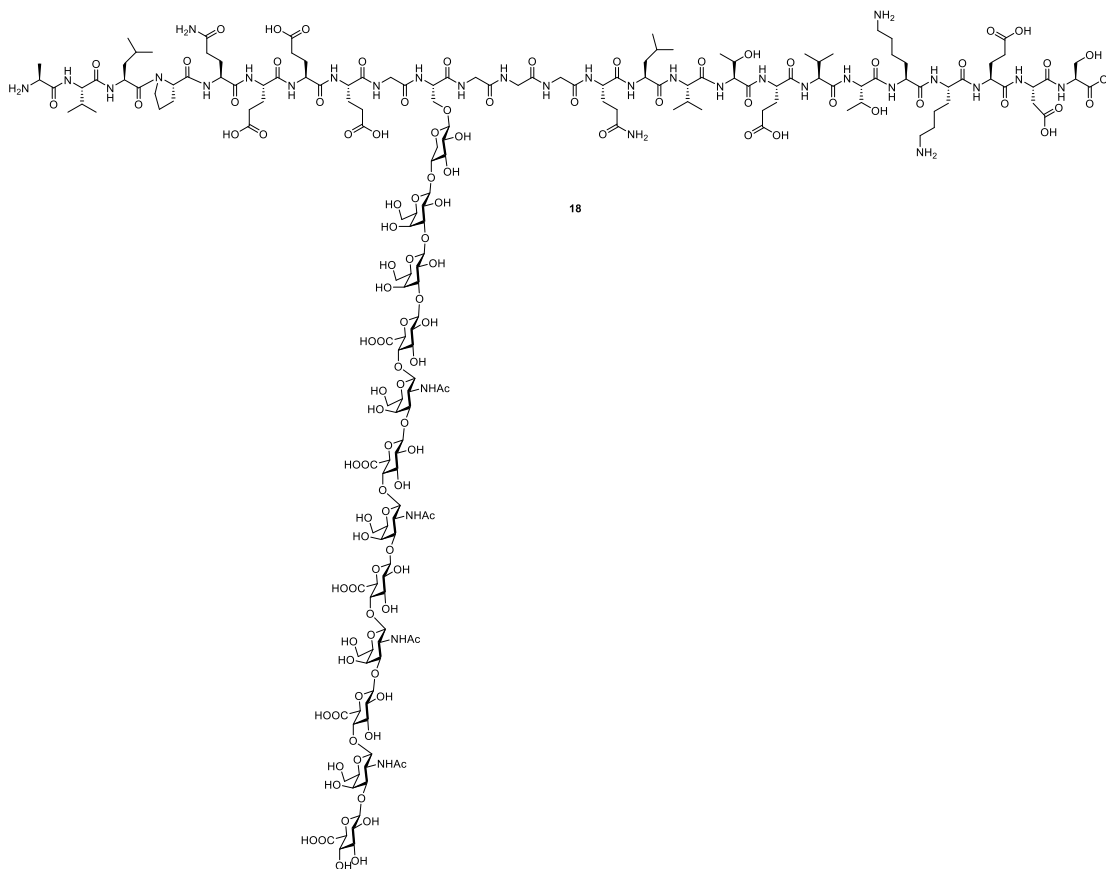
4Peptide **4** (10 mg, 9.1 μ mol) was conjugated to EAH Sepharose (2 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 8 from the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. Steps 7 and 8 were repeated twice to afford the octasaccharide bearing glycopeptide on Sepharose. Crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried *in vacuo* and purified by prep C-18 HPLC (0-50% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **16** (1.84 mg) in 5.1% yield. 1H NMR (600 MHz, D_2O) δ 4.77 – 4.74 (m, 1H), 4.70 – 4.65 (m, 3H), 4.59 –

4.49 (m, 6H), 4.46 – 4.37 (m, 8H), 4.37 – 4.26 (m, 6H), 4.24 – 4.07 (m, 12H), 4.04 – 3.92 (m, 15H), 3.91 – 3.66 (m, 24H), 3.66 – 3.45 (m, 5H), 3.42 – 3.29 (m, 4H), 3.03 – 2.93 (m, 5H), 2.91 – 2.84 (m, 1H), 2.56 – 2.43 (m, 10H), 2.43 – 2.30 (m, 5H), 2.19 – 1.96 (m, 25H), 1.96 – 1.88 (m, 1H), 1.88 – 1.80 (m, 2H), 1.79 – 1.63 (m, 2H), 1.73 – 1.65 (m, 5H), 1.66 – 1.56 (m, 5H), 1.52 (d, $J = 7.1$ Hz, 3H), 1.48 – 1.36 (m, 4H), 1.21 (t, $J = 6.6$ Hz, 6H), 1.00 – 0.91 (m, 27H), 0.88 (d, $J = 5.8$ Hz, 3H); ^{13}C NMR (151 MHz, D₂O) δ 177.83, 177.75, 176.91, 176.84, 176.80, 174.88, 174.62, 174.26, 174.01, 173.77, 173.73, 173.64, 173.53, 173.42, 173.36, 173.16, 173.10, 172.98, 172.93, 172.53, 172.14, 172.06, 171.94, 171.62, 171.51, 171.37, 170.66, 117.29, 115.35, 104.31, 104.18, 103.93, 103.80, 102.86, 101.38, 101.23, 82.32, 81.98, 80.11, 80.05, 79.89, 79.81, 76.39, 75.05, 74.93, 74.72, 74.68, 74.08, 73.76, 73.67, 72.72, 72.56, 72.45, 72.06, 71.33, 70.08, 69.80, 68.43, 68.19, 67.60, 66.93, 63.01, 61.13, 61.05, 60.92, 60.88, 60.56, 59.85, 59.70, 59.40, 59.27, 59.04, 55.21, 53.72, 53.62, 53.53, 53.41, 53.05, 52.93, 52.57, 51.05, 50.14, 50.11, 48.81, 47.88, 42.78, 42.55, 42.43, 39.52, 39.14, 38.96, 35.46, 31.03, 30.33, 30.27, 29.98, 29.93, 29.90, 29.86, 29.81, 29.30, 26.71, 26.54, 26.25, 26.20, 26.05, 24.72, 24.35, 24.31, 22.36, 22.01, 20.89, 20.53, 18.85, 18.83, 18.39, 18.37, 18.34, 17.89, 17.88, 17.72, 16.64. HRMS (ESI) m/z : $[\text{M} + 3\text{H}]^{3+}$ Calcd for Chemical Formula: C₁₅₉H₂₅₈N₃₁O₈₆ 1326.5633; Found 1326.5579.



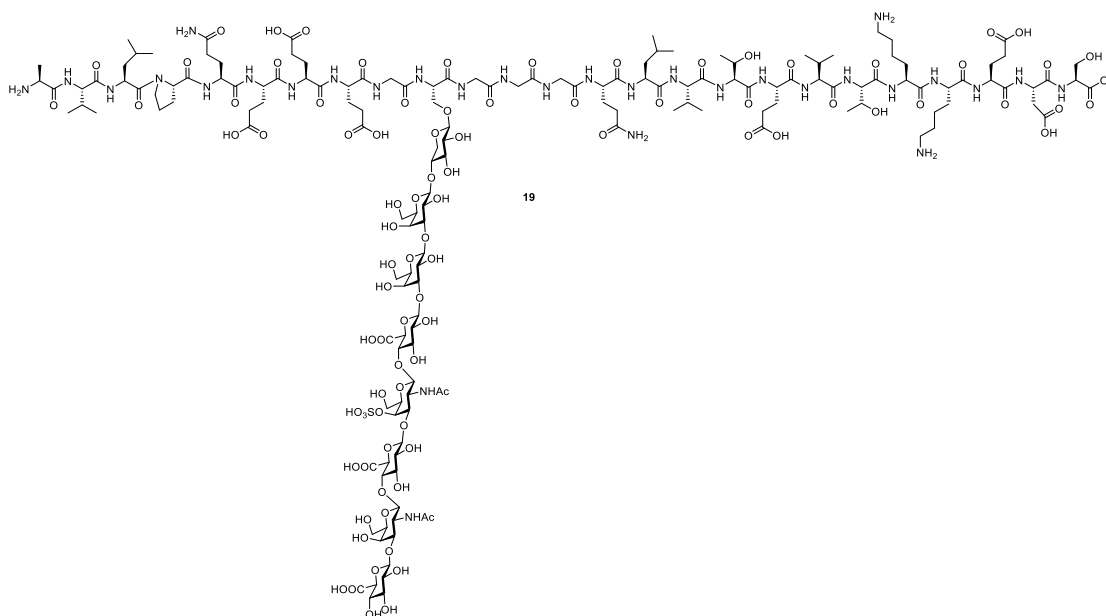
Glycopeptide **10** (5 mg, 1.5 μmol) was dissolved in 1 mL of KfoC reaction buffer (25 mM MOPS, 15 mM MnCl_2 , pH 7.2) containing 100 μg KfoC, UDP-GalNAc (2.9 mg, 4.7 μmol). The reaction mixture was incubated for 3 h at 37 $^\circ\text{C}$. Upon completion, 1 mL of MeOH was added and the mixture was centrifuged under 10,000 g for 10 min and dried *in vacuo*. The mixture was loaded to a Biotage® Sfar C18 column to perform a solid phase extraction. Fractions containing the desired product were lyophilized and redissolved in another 1 mL of KfoC reaction buffer containing 100 μg KfoC, UDP-GlcA (3 mg, 5.16 μmol). The reaction was again incubated for 3 h at 37 $^\circ\text{C}$ and was purified as aforementioned. These two reactions were repeated three times to obtain the desired decasaccharide glycopeptide. The final product was purified by prep C18 HPLC (0-50% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **17** (4.5 mg) in 33% yield. ^1H NMR (800 MHz, D_2O) δ 4.73 – 4.66 (m, 4H), 4.57 – 4.48 (m, 7H), 4.48 – 4.31 (m, 16H), 4.30 – 4.24 (m, 2H), 4.25 – 4.09 (m, 11H), 4.09 – 3.95 (m, 10H), 3.93 – 3.68

(m, 35H), 3.68 – 3.58 (m, 5H), 3.53 – 3.46 (m, 3H), 3.45 – 3.40 (m, 1H), 3.38 (t, $J = 8.9$ Hz, 1H), 3.35 – 3.31 (m, 1H), 3.03 – 2.98 (m, 4H), 2.79 (dd, $J = 16.4, 5.0$ Hz, 1H), 2.68 (dd, $J = 16.4, 4.9$ Hz, 1H), 2.44 – 2.27 (m, 15H), 2.18 – 1.94 (m, 29H), 1.90 – 1.82 (m, 2H), 1.82 – 1.75 (m, 2H), 1.75 – 1.68 (m, 5H), 1.69 – 1.59 (m, 5H), 1.55 (d, $J = 7.1$ Hz, 3H), 1.51 – 1.38 (m, 3H), 1.31 (br, 1H), 1.23 (dd, $J = 12.3, 6.4$ Hz, 6H), 1.01 – 0.92 (m, 27H), 0.90 (d, $J = 5.9$ Hz, 3H). ^{13}C NMR (201 MHz, D_2O) δ 175.15, 170.78, 104.47, 103.96, 101.08, 80.48, 79.88, 76.51, 76.32, 75.44, 75.12, 73.85, 72.93, 72.83, 72.75, 72.69, 71.90, 70.08, 69.96, 68.54, 67.94, 67.83, 67.20, 67.17, 62.20, 61.23, 61.07, 60.75, 59.62, 59.29, 58.97, 53.98, 53.82, 53.65, 52.53, 51.72, 51.08, 50.11, 48.98, 42.70, 42.65, 42.54, 39.15, 33.35, 31.09, 30.61, 30.45, 30.13, 27.55, 27.39, 26.26, 24.81, 24.32, 22.55, 22.07, 21.91, 20.94, 20.62, 18.85, 18.52, 18.04, 16.75. HRMS (ESI) m/z : $[\text{M} - 3\text{H}]^{3-}$ Calcd for Chemical Formula: $\text{C}_{173}\text{H}_{275}\text{N}_{32}\text{O}_{97}$ 1450.9198; Found 1450.9186.



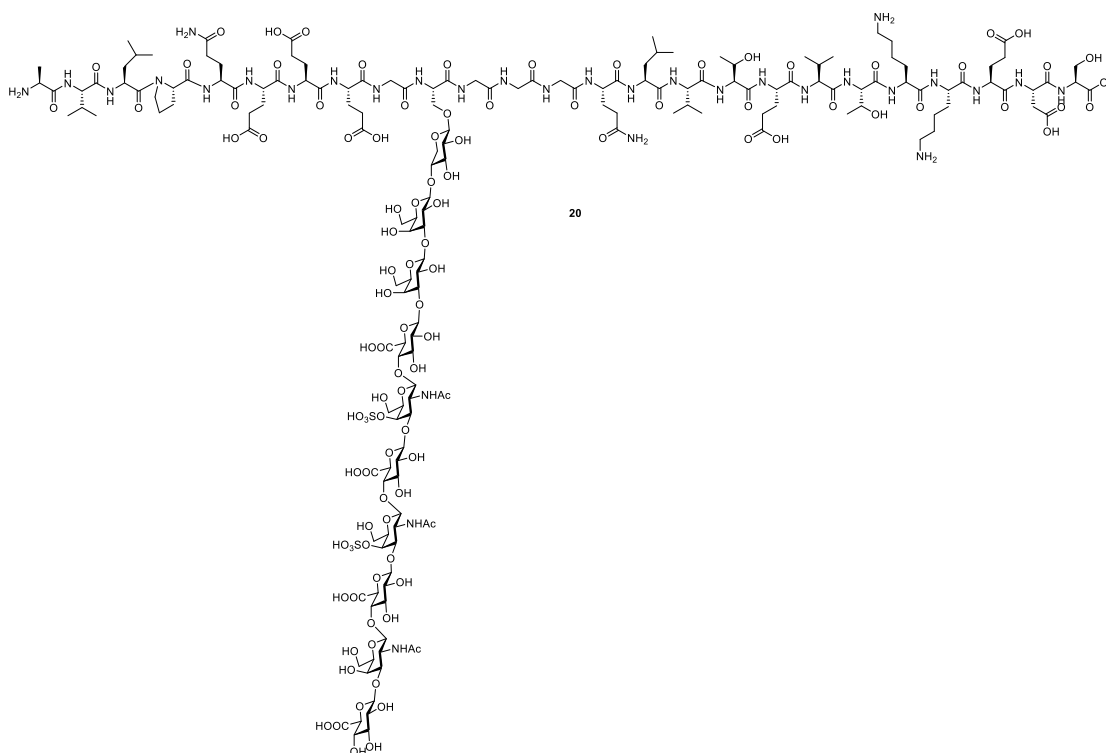
Glycopeptide **10** (5 mg, 1.5 μ mol) was dissolved in 1 mL of KfoC reaction buffer (25 mM MOPS, 15 mM MnCl_2 , pH 7.2) containing 100 μ g KfoC, UDP-GalNAc (2.9 mg, 4.7 μ mol). The reaction mixture was incubated for 3 h at 37 °C. Upon completion, 1 mL of MeOH was added and the mixture was centrifuged under 10,000 g for 10 min and dried *in vacuo*. The mixture was loaded to a Biotage® Sfar C18 column to perform a solid phase extraction. Fractions containing the desired product were lyophilized and redissolved in another 1 mL of KfoC reaction buffer containing 100 μ g KfoC, UDP-GlcA (3 mg, 5.16 μ mol). The reaction was again incubated for 3 h at 37 °C and was purified as aforementioned. These two reactions were repeated three times to obtain the desired dodecasaccharide glycopeptide. The final product was purified by prep C18 HPLC (0-50% acetonitrile/water; 0.1% trifluoroacetic acid) to obtain a white amorphous solid compound **18** (2.7 mg) in 38% yield. ^1H NMR (800 MHz, D_2O) δ 4.73 – 4.65 (m, 6H), 4.58 – 4.47 (m, 9H), 4.45 – 4.43 (m, 1H), 4.43 – 4.31 (m, 14H), 4.31 – 4.27 (m, 2H), 4.26 – 4.22 (m, 1H), 4.21 – 4.08 (m, 14H), 4.05 – 3.93 (m, 14H), 3.89 – 3.65 (m, 41H), 3.65 – 3.56 (m, 3H), 3.51 – 3.44 (m, 3H), 3.43 – 3.38 (m, 1H), 3.36 (t, J = 8.6 Hz, 3H), 3.33 – 3.29 (m, 2H), 3.03 – 2.98 (m, 4H), 2.81 (dd, J = 16.3, 4.8 Hz, 1H), 2.70 (dd, J = 16.3, 4.8 Hz, 1H), 2.44 – 2.29 (m, 15H), 2.18 – 1.88 (m, 31H), 1.87 – 1.80 (m, 3H), 1.80 – 1.73 (m, 2H), 1.73 – 1.67 (m, 5H), 1.66 – 1.55 (m, 5H), 1.52 (d, J = 7.1 Hz, 3H), 1.48 – 1.37 (m, 4H), 1.21 (dd, J = 11.0, 6.4 Hz, 6H), 1.00 – 0.91 (m, 27H), 0.88 (d, J = 5.9 Hz, 3H). ^{13}C NMR (201 MHz, D_2O) δ 179.39, 178.70, 177.77, 177.66, 176.80, 175.73, 174.97, 174.33, 173.58, 173.47, 173.25, 173.04, 172.45, 172.07, 170.75, 104.33, 104.15, 103.00, 100.84, 80.38, 79.67, 76.51, 76.19, 75.39, 75.28, 74.95, 73.94, 73.69, 72.81, 72.71, 72.45, 71.79, 70.07, 69.91, 68.46, 67.72, 66.99, 62.02, 61.06, 60.63, 59.76, 59.48, 59.27, 59.20, 59.11, 58.97, 57.18, 53.86, 53.47, 52.50, 53.34, 51.33, 50.97, 50.08, 48.83, 47.99, 47.88, 42.81, 42.54, 39.56, 39.14, 32.01, 31.06, 31.03, 30.37, 29.98, 29.32, 26.96, 26.76, 26.26, 26.19, 24.76, 24.37, 24.31,

22.47, 22.41, 22.03, 21.92, 21.89, 20.91, 20.53, 18.88, 18.81, 18.42, 18.40, 18.37, 17.86, 17.81, 17.74, 16.65. HRMS (ESI) m/z : $[M - 3H]^{3-}$ Calcd for Chemical Formula: $C_{187}H_{296}N_{33}O_{108}$ 1577.2903; Found 1577.2984.



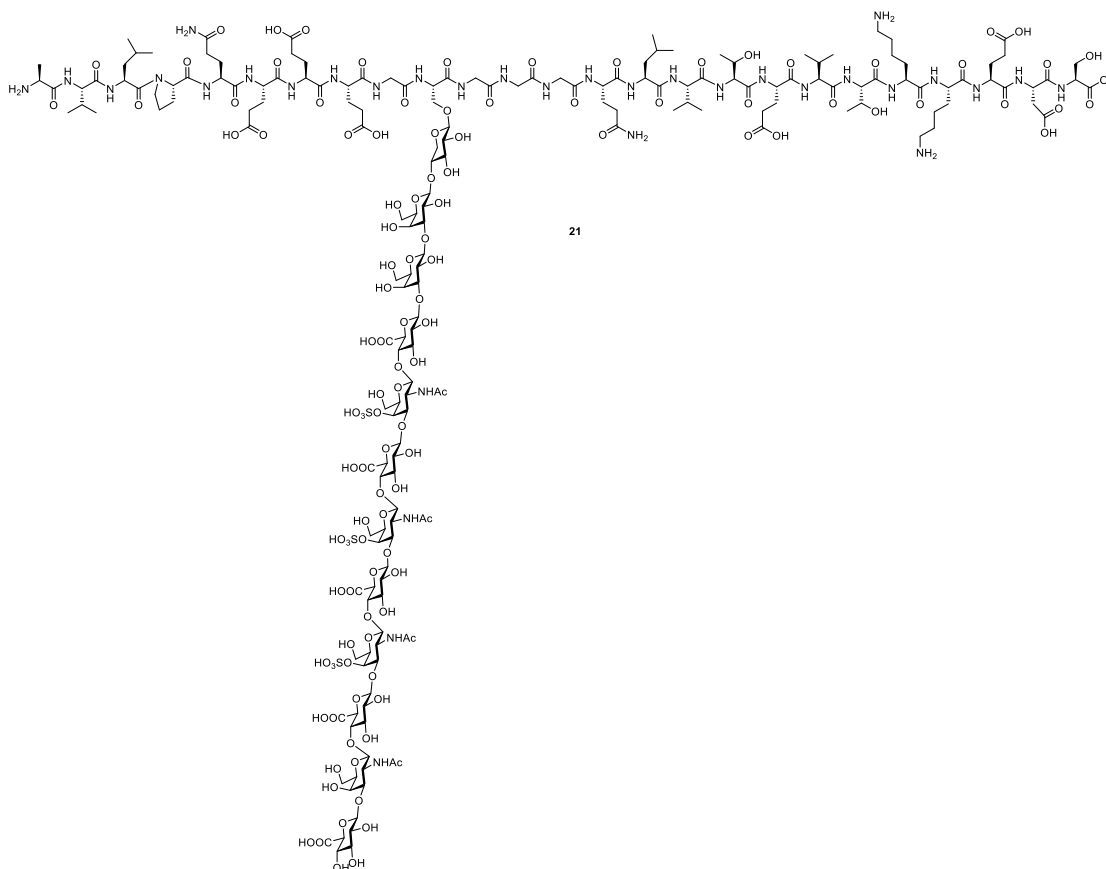
Peptide **5** (10 mg, 9.1 μ mol) was conjugated to EAH Sepharose (2 mL, drained volume) following the general procedure of peptide conjugation to EAH Sepharose. The resulting Sepharose was resuspended in buffer following steps 1 to 6 from the general procedure of enzymatic glycosylation on peptide-conjugated Sepharose. Steps 7 and 8 were repeated twice to afford octasaccharide-bearing glycopeptide. Finally, glycopeptide-conjugated Sepharose was sulfated following step 9. Crude products were obtained after incubating glycosylated Sepharose with 1% hydrazine (10 mL) three times, 12 h each. Basic solutions containing glycopeptides were dried *in vacuo* and purified by prep C-18 HPLC (0-50% acetonitrile/water; 50 mM ammonium formate) to obtain a white amorphous solid compound **19** (1.03 mg) in an overall yield of 3.4% from **5**. 1H NMR (800 MHz, D_2O) δ 4.72 – 4.64 (m, 5H), 4.57 (d, J = 8.8 Hz, 1H),

4.54 – 4.51 (m, 2H), 4.50 (d, $J = 7.6$ Hz, 1H), 4.47 (d, $J = 7.5$ Hz, 1H), 4.45 (d, $J = 8$ Hz, 1H), 4.43 – 4.23 (m, 12H), 4.23 – 4.09 (m, 9H), 4.08 – 3.93 (m, 10H), 3.91 – 3.54 (m, 35H), 3.51 – 3.35 (m, 5H), 3.35 – 3.30 (m, 2H), 3.03 – 2.98 (m, 4H), 2.81 – 2.76 (m, 1H), 2.67 (dd, $J = 16.3$, 4.7 Hz, 1H), 2.44 – 2.25 (m, 15H), 2.16 – 1.92 (m, 26H), 1.90 – 1.81 (m, 2H), 1.81 – 1.73 (m, 2H), 1.73 – 1.67 (m, 5H), 1.67 – 1.57 (m, 5H), 1.53 (d, $J = 7.1$ Hz, 3H), 1.49 – 1.38 (m, 4H), 1.21 (dd, $J = 6.4$, 5.4 Hz, 6H), 1.01 – 0.90 (m, 27H), 0.89 (d, $J = 5.9$ Hz, 3H); ^{13}C NMR (201 MHz, D_2O) δ 180.48, 175.00, 174.24, 173.70, 173.58, 173.39, 172.33, 172.09, 171.88, 171.74, 171.71, 171.55, 170.93, 170.74, 104.16, 103.86, 103.83, 103.71, 102.95, 101.27, 101.00, 82.14, 80.43, 80.09, 76.61, 76.27, 75.31, 75.29, 74.91, 74.73, 74.63, 73.92, 73.77, 73.72, 73.53, 72.84, 72.73, 72.29, 71.81, 70.02, 69.79, 69.57, 68.47, 68.05, 67.02, 62.47, 62.07, 61.07, 61.00, 60.65, 59.66, 59.51, 59.19, 58.94, 57.25, 53.76, 53.46, 53.33, 52.47, 51.50, 51.00, 50.07, 48.84, 47.88, 42.82, 42.41, 39.14, 32.74, 31.03, 30.39, 30.03, 29.98, 29.32, 27.39, 27.26, 27.07, 26.75, 26.58, 26.26, 26.18, 24.77, 24.37, 24.31, 22.42, 22.02, 21.90, 21.87, 20.92, 20.53, 18.89, 18.80, 18.42, 18.40, 17.86, 17.80, 17.76, 16.66. HRMS (ESI) m/z : $[\text{M} - 3\text{H}]^{3-}$ Calcd for Chemical Formula: $\text{C}_{159}\text{H}_{254}\text{N}_{31}\text{O}_{89}\text{S}$ 1351.2016; Found 1351.2030.



Glycopeptide **17** (1 mg, 0.22 μmol) was dissolved in 0.2 mL of CS4OST reaction buffer (50 mM MOPS, 10 mM CaCl_2 , fresh 2 mM DDT, pH 6.5) containing CS4OST (50 μg), PAPS (0.25 mg, 0.5 μmol). The reaction mixture was incubated for 6 h at 37 $^\circ\text{C}$, then another 0.2 mL of reaction buffer with enzymes and PAPS was added to the mixture again and was incubated for another 6 h at 37 $^\circ\text{C}$. Upon completion, 0.4 mL of MeOH was added and the mixture was centrifuged under 10,000 g for 10 min and dried *in vacuo*. The mixture was purified by prep C18 HPLC (0-50% acetonitrile/water; 50 mM ammonium formate) to obtain a white amorphous compound **20** (0.56 mg) in 54 % yield. ^1H NMR (800 MHz, D_2O) δ 4.72 – 4.65 (m, 4H), 4.61 – 4.27 (m, 25H), 4.25 (d, J = 5.9 Hz, 2H), 4.23 – 4.08 (m, 11H), 4.08 – 3.95 (m, 10H), 3.93 – 3.55 (m, 36H), 3.52 – 3.44 (m, 3H), 3.44 – 3.36 (m, 3H), 3.35 – 3.30 (m, 2H), 3.02 (dt, J = 12.5, 7.5 Hz, 4H), 2.80 (dd, J = 16.2, 4.7 Hz, 1H), 2.69 (dd, J = 16.2, 4.6 Hz, 1H), 2.47 – 2.29 (m, 15H), 2.17 – 1.92 (m, 29H), 1.88 – 1.81 (m, 2H), 1.81 – 1.74 (m, 2H), 1.73 – 1.64 (m, 4H), 1.68 – 1.57 (m, 4H), 1.53 (d, J = 7.1 Hz, 2H), 1.50 – 1.38 (m, 2H), 1.38 – 1.25 (m, 5H), 1.21 (dd, J = 11.6,

6.4 Hz, 6H), 1.03 – 0.91 (m, 27H), 0.89 (d, $J = 5.9$ Hz, 3H); ^{13}C NMR (201 MHz, D_2O) δ 175.08, 173.36, 173.25, 172.40, 172.07, 171.97, 171.54, 171.43, 170.68, 104.23, 103.94, 103.87, 103.04, 101.20, 101.00, 82.81, 82.16, 80.55, 80.39, 76.52, 76.36, 75.56, 75.39, 74.91, 74.75, 73.78, 72.98, 72.82, 72.33, 71.85, 70.08, 69.92, 69.59, 68.47, 67.98, 67.18, 67.01, 62.18, 61.05, 60.73, 59.76, 59.28, 58.96, 57.35, 53.80, 53.48, 52.51, 51.54, 51.22, 50.09, 48.97, 47.84, 42.68, 42.52, 39.62, 39.15, 32.69, 31.08, 30.43, 30.11, 24.79, 27.37, 27.05, 26.24, 24.47, 24.31, 22.54, 22.03, 20.98, 20.60, 20.12, 18.99, 18.35, 17.86, 17.06, 16.74. HRMS (ESI) m/z : $[\text{M} - 3\text{H}]^{3-}$ Calcd for Chemical Formula: $\text{C}_{173}\text{H}_{275}\text{N}_{32}\text{O}_{103}\text{S}_2$ 1504.2234; Found 1504.2235.



Glycopeptide **18** (1 mg, 0.21 μmol) was dissolved in 0.2 mL of CS4OST reaction buffer (50 mM MOPS, 10 mM CaCl_2 , fresh 2 mM DDT, pH 6.5) containing CS4OST (50 μg) and donor PAPS (0.25 mg, 0.5 μmol). The reaction mixture was incubated for 6 h at 37 $^\circ\text{C}$, then another 0.2 mL of reaction buffer with enzymes and PAPS was added to the mixture again and was incubated for another 6 h at 37 $^\circ\text{C}$. Upon completion, 0.4 mL of MeOH was added and the mixture was centrifuged under 10,000 g for 10 min and dried *in vacuo*. The mixture was purified by prep C18 HPLC (0-50% acetonitrile/water; 50 mM ammonium formate) to obtain a white amorphous compound **21** (0.56 mg) in 52% yield. ^1H NMR (800 MHz, D_2O) δ 4.76 – 4.73 (m, 4H), 4.73 – 4.66 (m, 5H), 4.61 – 4.55 (br, 3H), 4.53 (d, J = 8.1 Hz, 2H), 4.52 – 4.26 (m, 21H), 4.26 – 4.22 (br, 1H), 4.22 – 4.13 (m, 9H), 4.13 – 4.08 (br, 1H), 4.08 – 3.96 (m, 15H), 3.91 – 3.54 (m, 44H), 3.50 – 3.43 (m, 3H), 3.43 – 3.35 (m, 4H), 3.35 – 3.30 (m, 2H), 3.03 – 2.98 (m, 4H), 2.81 (dd, J = 16.2, 4.7 Hz, 1H), 2.70 (dd, J = 16.2, 4.6 Hz, 1H), 2.48 – 2.29 (m, 15H), 2.17 – 1.91 (m, 31H), 1.88 – 1.80 (m, 3H), 1.80 – 1.74 (m, 2H), 1.73 – 1.67 (m, 5H), 1.67 – 1.54 (m, 5H), 1.52 (d, J = 7.1 Hz, 3H), 1.50 – 1.38 (m, 4H), 1.21 (dd, J = 11.0, 6.4 Hz, 6H), 1.00 – 0.91 (m, 27H), 0.89 (d, J = 5.9 Hz, 3H); ^{13}C NMR (201 MHz, D_2O) δ 178.70, 175.05, 173.65, 173.44, 173.33, 173.48, 172.37, 172.26, 172.07, 171.55, 171.36, 170.74, 104.17, 103.88, 103.82, 103.77, 103.71, 102.92, 101.32, 101.00, 100.84, 82.65, 82.16, 80.42, 76.46, 76.32, 75.57, 75.28, 74.93, 74.57, 73.77, 73.51, 72.83, 72.72, 72.27, 71.79, 70.03, 69.80, 68.69, 68.44, 68.07, 67.66, 66.99, 62.47, 62.01, 61.07, 60.99, 60.62, 59.71, 59.47, 59.20, 58.97, 57.14, 53.77, 53.57, 53.44, 53.34, 52.99, 52.49, 51.53, 51.32, 51.01, 50.08, 48.84, 47.89, 42.71, 42.55, 39.57, 39.15, 38.99, 37.85, 32.00, 31.03, 30.36, 29.99, 29.32, 26.96, 26.75, 26.57, 26.26, 26.19, 24.76, 24.37, 24.32, 22.50, 22.41, 22.03, 21.93, 21.90, 20.91, 20.54, 18.89, 18.82, 18.42, 18.40, 17.86, 17.82, 17.76, 16.67. HRMS (ESI) m/z : $[\text{M} - 3\text{H}]^{3-}$ Calcd for Chemical Formula: $\text{C}_{187}\text{H}_{296}\text{N}_{33}\text{O}_{117}\text{S}_3$ 1657.2471; Found 1657.2493.

REFERENCES

1. De Pasquale, V.; Pavone, L. M., Heparan Sulfate Proteoglycan Signaling in Tumor Microenvironment. *Int. J. Mol. Sci.* **2020**, *21* (18), 6588.
2. Galtrey, C. M.; Fawcett, J. W., The Role of Chondroitin Sulfate Proteoglycans in Regeneration and Plasticity in the Central Nervous System. *Brain Res. Rev.* **2007**, *54*, 1-18.
3. Avram, S.; Shaposhnikov, S.; Buiu, C.; Mernea, M., Chondroitin sulfate proteoglycans: structure-function relationship with implication in neural development and brain disorders. *Biomed. Res. Int.* **2014**, *2014*, 642798.
4. Mencia, C. P.; Hussein, R. K.; Yu, P.; Geller, H. M., The Role of Chondroitin Sulfate Proteoglycans in Nervous System Development. *J. Histochem. Cytochem.* **2021**, *69* (1), 61-80.
5. Karamanos, N. K.; Piperigkou, Z.; Theocharis, A. D.; Watanabe, H.; Franchi, M.; Baud, S.; Brézillon, S.; Götte, M.; Passi, A.; Vigetti, D.; Ricard-Blum, S.; Sanderson, R. D.; Neill, T.; Iozzo, R. V., Proteoglycan Chemical Diversity Drives Multifunctional Cell Regulation and Therapeutics. *Chem. Rev.* **2018**, *118* (18), 9152-9232.
6. Li, L.; Ly, M.; Linhardt, R. J., Proteoglycan sequence. *Mol. BioSyst.* **2012**, *8*, 1613-1525.
7. Kim, S. K.; Henen, M. A.; Hinck, A. P., Structural biology of betaglycan and endoglin, membrane-bound co-receptors of the TGF-beta family. *Exp. Biol. Med.* **2019**, *244* (17), 1547-1558.
8. Iozzo, R. V., Heparan sulfate proteoglycans: intricate molecules with intriguing functions. *J. Clin. Invest.* **2001**, *108*, 165-167.
9. Herndon, M. E.; Stipp, C. S.; Lander, A. D., Interactions of neural glycosaminoglycans and proteoglycans with protein ligands: assessment of selectivity, heterogeneity and the participation of core proteins in binding. *Glycobiology* **1999**, *9*, 143-155.
10. O'Leary, T. R.; Critcher, M.; Stephenson, T. N.; Yang, X.; Hassan, A. A.; Bartfield, N. M.; Hawkins, R.; Huang, M. L., Chemical editing of proteoglycan architecture. *Nat. Chem. Biol.* **2022**, *18* (6), 634-642.
11. Zhang, Y.; Wang, N.; Raab, R. W.; McKown, R. L.; Irwin, J. A.; Kwon, I.; van Kuppevelt, T. H.; Laurie, G. W., Targeting of heparanase-modified syndecan-1 by prosecretory mitogen lacritin requires conserved core GAGAL plus heparan and chondroitin sulfate as a novel hybrid binding site that enhances selectivity. *J. Biol. Chem.* **2013**, *288* (17), 12090-12101.
12. Yang, W.; Eken, Y.; Zhang, J.; Cole, L. E.; Ramadan, S.; Xu, Y.; Zhang, Z.; Liu, J.; Wilson, A. K.; Huang, X., Chemical synthesis of human syndecan-4 glycopeptide bearing O-, N-sulfation and multiple aspartic acids for probing impacts of the glycan chain and the core peptide on biological functions. *Chem. Sci.* **2020**, *11* (25), 6393-6404.

13. Ramadan, S.; Yang, W.; Huang, X., Chapter 8. Synthesis of chondroitin sulfate oligosaccharides and chondroitin sulfate glycopeptides. In *Synthetic Glycomes*, The Royal Society of Chemistry: **2019**; pp 172-206 and references cited therein.
14. Shimawaki, K.; Fujisawa, Y.; Sato, F.; Fujitani, N.; Kuroguchi, M.; Hoshi, H.; Hinou, H.; Nishimura, S.-I., Highly efficient and versatile synthesis of proteoglycan core structures from 1,6-anhydro- β -lactose as a key starting material. *Angew. Chem., Int. Ed.* **2007**, *46* (17), 3074-3079.
15. Mende, M.; Bednarek, C.; Wawrystyn, M.; Sauter, P.; Biskup, M. B.; Schepers, U.; Bräse, S., Chemical Synthesis of Glycosaminoglycans. *Chem. Rev.* **2016**, *116* (14), 8193-8255.
16. Yang, B.; Yoshida, K.; Yin, Z.; Dai, H.; Kavunja, H.; El-Dakdouki, M. H.; Sungsuwan, S.; Dulaney, S. B.; Huang, X., Chemical Synthesis of a Heparan Sulfate Glycopeptide: Syndecan-1. *Angew. Chem. Int. Ed.* **2012**, *51* (40), 10185-10189.
17. Nilsson, M.; Westman, J.; Svahn, C.-M., Synthesis of Tri-And Tetrasaccharides Present in the Linkage Region of Heparin and Heparan Sulphate. *J. Carbohydr. Chem.* **1993**, *12* (1), 23-37.
18. Huang, T.-Y.; Zulueta, M. M. L.; Hung, S.-C., One-Pot Strategies for the Synthesis of the Tetrasaccharide Linkage Region of Proteoglycans. *Org. Lett.* **2011**, *13* (6), 1506-1509.
19. Yang, W.; Ramadan, S.; Yang, B.; Yoshida, K.; Huang, X., Homoserine as an aspartic acid precursor for synthesis of proteoglycan glycopeptide containing aspartic acid and a sulfated glycan chain. *J. Org. Chem.* **2016**, *81*, 12052-12059.
20. Yang, W.; Yoshida, K.; Yang, B.; Huang, X., Obstacles and solutions for chemical synthesis of syndecan-3 (53–62) glycopeptides with two heparan sulfate chains. *Carbohydr. Res.* **2016**, *435*, 180-194.
21. Yoshida, K.; Yang, B.; Yang, W.; Zhang, Z.; Zhang, J.; Huang, X., Chemical Synthesis of Syndecan-3 Glycopeptides Bearing Two Heparan Sulfate Glycan Chains. *Angew. Chem. Int. Ed.* **2014**, *53* (34), 9051-9058.
22. Ramadan, S.; Yang, W.; Zhang, Z.; Huang, X., Synthesis of chondroitin sulfate A bearing syndecan-1 glycopeptide. *Org. Lett.* **2017**, *19*, 4838-4841.
23. Zhang, J.; Liu, D.; Saikam, V.; Gadi, M. R.; Gibbons, C.; Fu, X.; Song, H.; Yu, J.; Kondengaden, S. M.; Wang, P. G.; Wen, L., Machine-Driven Chemoenzymatic Synthesis of Glycopeptide. *Angew. Chem. Int. Ed.* **2020**, *59* (45), 19825-19829.
24. Matsushita, T.; Handa, S.; Naruchi, K.; Garcia-Martin, F.; Hinou, H.; Nishimura, S.-I., A novel approach for the parallel synthesis of glycopeptides by combining solid-phase peptide synthesis and dendrimer-supported enzymatic modifications. *Polym. J.* **2013**, *45* (8), 854-862.

25. Schuster, M.; Wang, P.; Paulson, J. C.; Wong, C.-H., Solid-Phase Chemical-Enzymic Synthesis of Glycopeptides and Oligosaccharides. *J. Am. Chem. Soc.* **1994**, *116* (3), 1135-1136.
26. Meldal, M.; Auzanneau, F.-I.; Hindsgaul, O.; Palcic, M. M., A PEGA resin for use in the solid-phase chemical–enzymatic synthesis of glycopeptides. *J. Chem. Soc., Chem. Commun.* **1994**, (16), 1849-1850.
27. Halcomb, R. L.; Huang, H.; Wong, C.-H., Solution-and solid-phase synthesis of inhibitors of *H. pylori* attachment and E-selectin-mediated leukocyte adhesion. *J. Am. Chem. Soc.* **1994**, *116* (25), 11315-11322.
28. Fallows, T. W.; McGrath, A. J.; Silva, J.; McAdams, S. G.; Marchesi, A.; Tuna, F.; Flitsch, S. L.; Tilley, R. D.; Webb, S. J., High-throughput chemical and chemoenzymatic approaches to saccharide-coated magnetic nanoparticles for MRI. *Nanoscale Adv.* **2019**, *1* (9), 3597-3606.
29. Wen, L.; Edmunds, G.; Gibbons, C.; Zhang, J.; Gadi, M. R.; Zhu, H.; Fang, J.; Liu, X.; Kong, Y.; Wang, P. G., Toward Automated Enzymatic Synthesis of Oligosaccharides. *Chem. Rev.* **2018**, *118* (17), 8151-8187.
30. Zhang, J.; Chen, C.; Gadi, M. R.; Gibbons, C.; Guo, Y.; Cao, X.; Edmunds, G.; Wang, S.; Liu, D.; Yu, J.; Wen, L.; Wang, P. G., Machine-Driven Enzymatic Oligosaccharide Synthesis by Using a Peptide Synthesizer. *Angew. Chem., Int. Ed.* **2018**, *57* (51), 16638-16642.
31. Huang, X.; Witte, Krista L.; Bergbreiter, David E.; Wong, C.-H., Homogenous Enzymatic Synthesis Using a Thermo-Responsive Water-Soluble Polymer Support. *Adv. Synth. Cat.* **2001**, *343* (6-7), 675-681.
32. Blixt, O.; Norberg, T., Solid-Phase Enzymatic Synthesis of a Sialyl Lewis X Tetrasaccharide on a Sepharose Matrix. *J. Org. Chem.* **1998**, *63* (8), 2705-2710.
33. Kent, U. M., Purification of antibodies using protein A-sepharose and FPLC. *Immunocytochemical Methods and Protocols* **1999**, 29-33.
34. Blixt, O.; Norberg, T., Enzymatic glycosylation of reducing oligosaccharides linked to a solid phase or a lipid via a cleavable squarate linker. *Carbohydr. Res.* **1999**, *319* (1), 80-91.
35. Hwang, H. Y.; Olson, S. K.; Brown, J. R.; Esko, J. D.; Horvitz, H. R., The *Caenorhabditis elegans* genes *sqv-2* and *sqv-6*, which are required for vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase. *J. Biol. Chem.* **2003**, *278*, 11735-11738.
36. Gao, J.; Lin, P.-h.; Nick, S. T.; Liu, K.; Yu, K.; Hohenester, E.; Huang, X., Exploration of human xylosyltransferase for chemoenzymatic synthesis of proteoglycan linkage region. *Org. Biomol. Chem.* **2021**, *19* (15), 3374-3378.

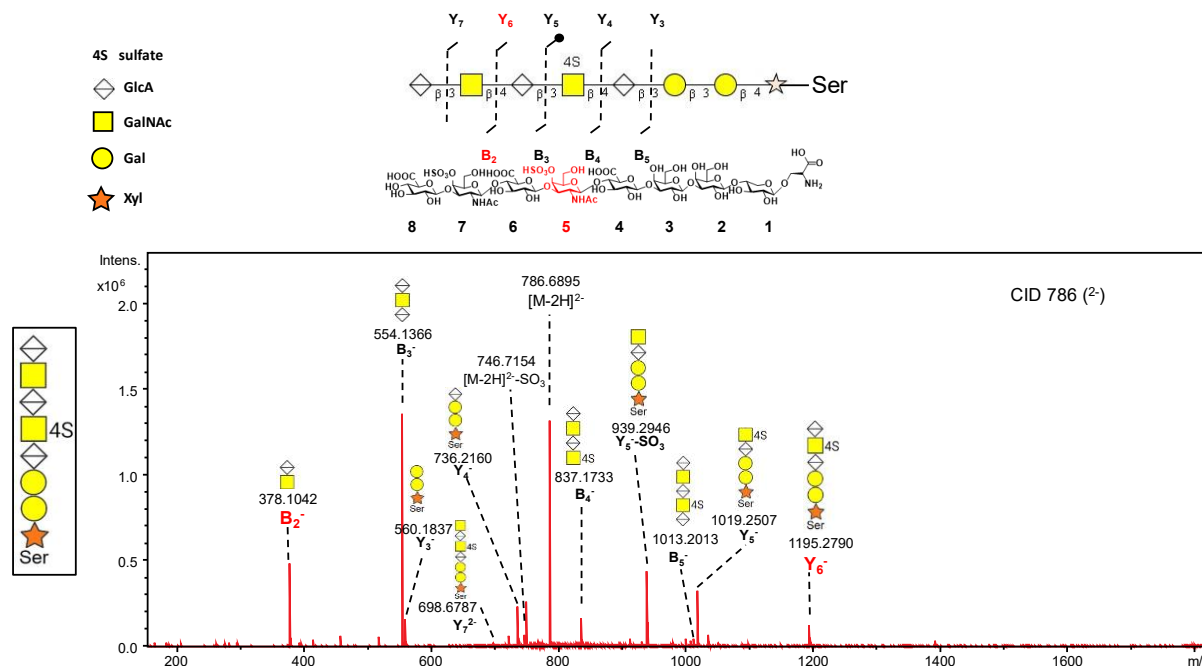
37. Talhaoui, I.; Bui, C.; Oiol, R.; Mulliert, G.; Gulberti, S.; Netter, P.; Coughtrie, M. W. H.; Ouzzine, M.; Fournel-Gigleux, S., Identification of key functional residues in the active site of human β 1,4-galactosyltransferase 7. *J. Biol. Chem.* **2010**, *285*, 37342–37358.
38. Almeida, R.; Levery, S. B.; Mandel, U.; Kresse, H.; Schwientek, T.; Bennett, E. P.; Clausen, H., Cloning and expression of a proteoglycan UDP-galactose:beta-xylose beta1,4-galactosyltransferase I. A seventh member of the human beta4-galactosyltransferase gene family. *J. Biol. Chem.* **1999**, *274*, 26165-26171.
39. Gao, J.; Lin, P.-h.; Nick, S. T.; Huang, J.; Tykesson, E.; Ellervik, U.; Li, L.; Huang, X., Chemoenzymatic synthesis of glycopeptides bearing galactose–xylose disaccharide from the proteoglycan linkage region. *Org. Lett.* **2021**, *23* (5), 1738-1741.
40. Bai, X.; Zhou, D.; Brown, J. R.; Crawford, B. E.; Hennen, T.; Esko, J. D., Biosynthesis of the linkage region of glycosaminoglycans. *J. Biol. Chem.* **2001**, *276*, 48189–48195.
41. Tone, Y.; Kitagawa, H.; Imiyab, K.; Okab, S.; Kawasaki, T.; Sugahara, K., Characterization of recombinant human glucuronyltransferase I involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *FEBS Lett.* **1999**, *459*, 415-420.
42. Koike, T.; Izumikawa, T.; Tamura, J.; Kitagawa, H., FAM20B is a kinase that phosphorylates xylose in the glycosaminoglycan-protein linkage region. *Biochem. J.* **2009**, *421* (2), 157-62.
43. Wen, J.; Xiao, J.; Rahdar, M.; Choudhury, B. P.; Cui, J.; Taylor, G. S.; Esko, J. D.; Dixon, J. E., Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (44), 15723-15728.
44. Moremen, K. W.; Ramiah, A.; Stuart, M.; Steel, J.; Meng, L.; Forouhar, F.; Moniz, H. A.; Gahlay, G.; Gao, Z.; Chapla, D.; Wang, S.; Yang, J.-Y.; Prabhakar, P. K.; Johnson, R.; Rosa, M. d.; Geisler, C.; Nairn, A. V.; Seetharaman, J.; Wu, S.-C.; Tong, L.; Gilbert, H. J.; LaBaer, J.; Jarvis, D. L., Expression system for structural and functional studies of human glycosylation enzymes. *Nat. Chem. Biol.* **2018**, *14* (2), 156-162.
45. Koike, T.; Izumikawa, T.; Sato, B.; Kitagawa, H., Identification of phosphatase that dephosphorylates xylose in the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol. Chem.* **2014**, *289* (10), 6695-6708.
46. Pugia, M. J.; Valdes, R.; Jortani, S. A., Bikunin (Urinary Trypsin Inhibitor): Structure, Biological Relevance, And Measurement. In *Adv. Clin. Chem.*, Elsevier: 2007; Vol. 44, pp 223-245.
47. Fries, E.; Blom, A. M., Bikunin — not just a plasma proteinase inhibitor. *Int. J. Biochem. Cell Biol.* **2000**, *32* (2), 125-137.

48. Matsuzaki, H.; Kobayashi, H.; Yagyu, T.; Wakahara, K.; Kondo, T.; Kurita, N.; Sekino, H.; Inagaki, K.; Suzuki, M.; Kanayama, N.; Terao, T., Plasma Bikunin As a Favorable Prognostic Factor in Ovarian Cancer. *J. Clin. Oncol.* **2005**, 23 (7), 1463-1472.
49. Ly, M.; Leach, F. E., 3rd; Laremore, T. N.; Toida, T.; Amster, I. J.; Linhardt, R. J., The proteoglycan bikunin has a defined sequence. *Nat. Chem. Biol.* **2011**, 7 (11), 827-833.
50. Ramadan, S.; Li, T.; Yang, W.; Zhang, J.; Parameswaran, N.; Huang, X., Chemical synthesis and anti-inflammatory activity of bikunin associated chondroitin sulfate 24-mer. *ACS Cent. Sci.* **2020**, 6, 913–920.
51. Karnad, D. R.; Bhadade, R.; Verma, P. K.; Moulick, N. D.; Daga, M. K.; Chafekar, N. D.; Iyer, S., Intravenous administration of ulinastatin (human urinary trypsin inhibitor) in severe sepsis: a multicenter randomized controlled study. *Intensive Care Med.* **2014**, 40 (6), 830-8.
52. Stober, V. P.; Lim, Y. P.; Opal, S.; Zhuo, L.; Kimata, K.; Garantziotis, S., Inter- α -inhibitor Ameliorates Endothelial Inflammation in Sepsis. *Lung* **2019**, 197 (3), 361-369.
53. Sato, T.; Narimatsu, H., Chondroitin Sulfate N-Acetylgalactosaminyltransferase 1,2 (CSGALNACT1,2). In *Handbook of Glycosyltransferases and Related Genes*, Taniguchi, N.; Honke, K.; Fukuda, M.; Narimatsu, H.; Yamaguchi, Y.; Angata, T., Eds. Springer Japan: Tokyo, 2014; pp 925-933.
54. Gundlach, M. W.; Conrad, H. E., Glycosyl transferases in chondroitin sulphate biosynthesis. Effect of acceptor structure on activity. *Biochem. J.* **1985**, 226 (3), 705-14.
55. Watanabe, Y.; Takeuchi, K.; Higa Onaga, S.; Sato, M.; Tsujita, M.; Abe, M.; Natsume, R.; Li, M.; Furuichi, T.; Saeki, M.; Izumikawa, T.; Hasegawa, A.; Yokoyama, M.; Ikegawa, S.; Sakimura, K.; Amizuka, N.; Kitagawa, H.; Igarashi, M., Chondroitin sulfate N-acetylgalactosaminyltransferase-1 is required for normal cartilage development. *Biochem. J.* **2010**, 432 (1), 47-55.
56. Xue, J.; Jin, L.; Zhang, X.; Wang, F.; Ling, P.; Sheng, J., Impact of donor binding on polymerization catalyzed by KfoC by regulating the affinity of enzyme for acceptor. *Biochim. Biophys. Acta. Gen. Subj.* **2016**, 1860 (4), 844-855.
57. Li, J.; Su, W.; Liu, J., Enzymatic synthesis of homogeneous chondroitin sulfate oligosaccharides. *Angew. Chem. Int. Ed* **2017**, 56, 11784-11787.
58. Zamolodchikova, T. S.; Tolpygo, S. M.; Svirshchevskaya, E. V., Cathepsin G—Not Only Inflammation: The Immune Protease Can Regulate Normal Physiological Processes. *Front. immunol.* **2020**, 11, 411 doi: 10.3389/fimmu.2020.00411.
59. Campbell, E. J.; Owen, C. A., The Sulfate Groups of Chondroitin Sulfate- and Heparan Sulfate-containing Proteoglycans in Neutrophil Plasma Membranes Are Novel Binding Sites for Human Leukocyte Elastase and Cathepsin G*. *J. Biol. Chem.* **2007**, 282 (19), 14645-14654.

60. Tone, Y.; Kitagawa, H.; Imiya, K.; Oka, S.; Kawasaki, T.; Sugahara, K., Characterization of recombinant human glucuronyltransferase I involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *FEBS Letters* **1999**, 459 (3), 415-420.
61. Sanderson, P.; Stickney, M.; Leach, F. E., 3rd; Xia, Q.; Yu, Y.; Zhang, F.; Linhardt, R. J.; Amster, I. J., Heparin/heparan sulfate analysis by covalently modified reverse polarity capillary zone electrophoresis-mass spectrometry. *J. Chromatogr. A*. **2018**, 1545, 75-83.
62. Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M., GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. *Journal of Proteome Research* **2008**, 7 (4), 1650-1659.

APPENDIX A: SUPPLEMENTARY FIGURES, SCHEMES AND TABLES

a.



b.

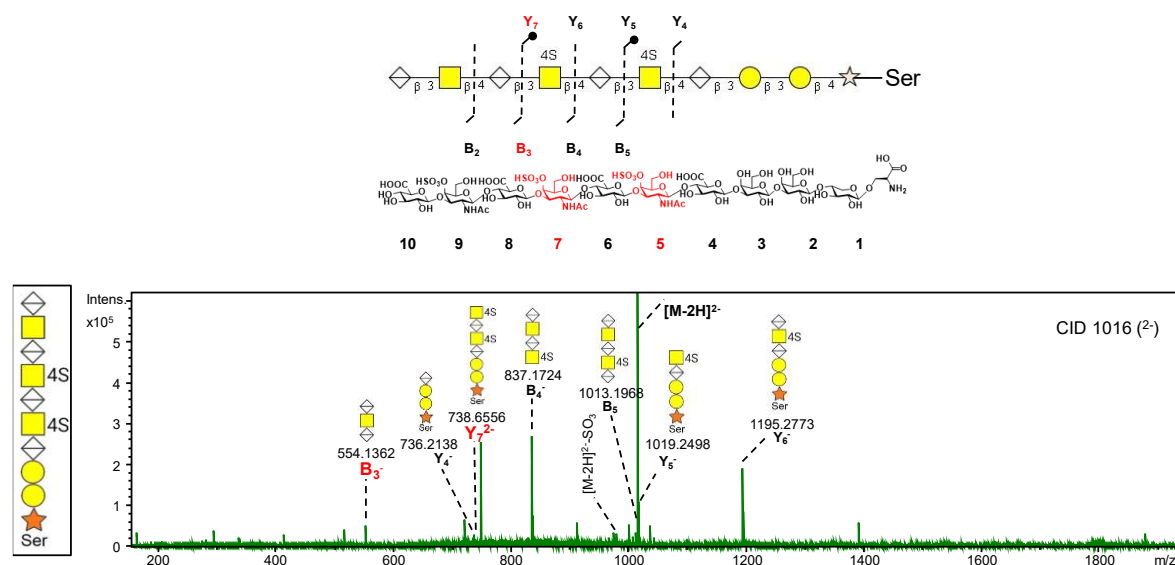


Figure 2.4. The dashed lines on the structure indicate fragments with no sulfate loss observed.

Black filled circles on the sequence indicate both fragment ions with sulfate loss and without

Figure 2.4. (cont'd)

sulfate loss were observed. The empty circle indicates a fragment ion with sulfate loss was observed. **a.** Sulfation pattern analysis of compound **19**, fragment ions B₂ and Y₆ indicate that sulfate is on GalNAc 5. **b.** Sulfation pattern analysis of compound **20**, fragment ions B₃ and Y₇ indicate that sulfate is on GalNAc 5 and GalNAc 7.

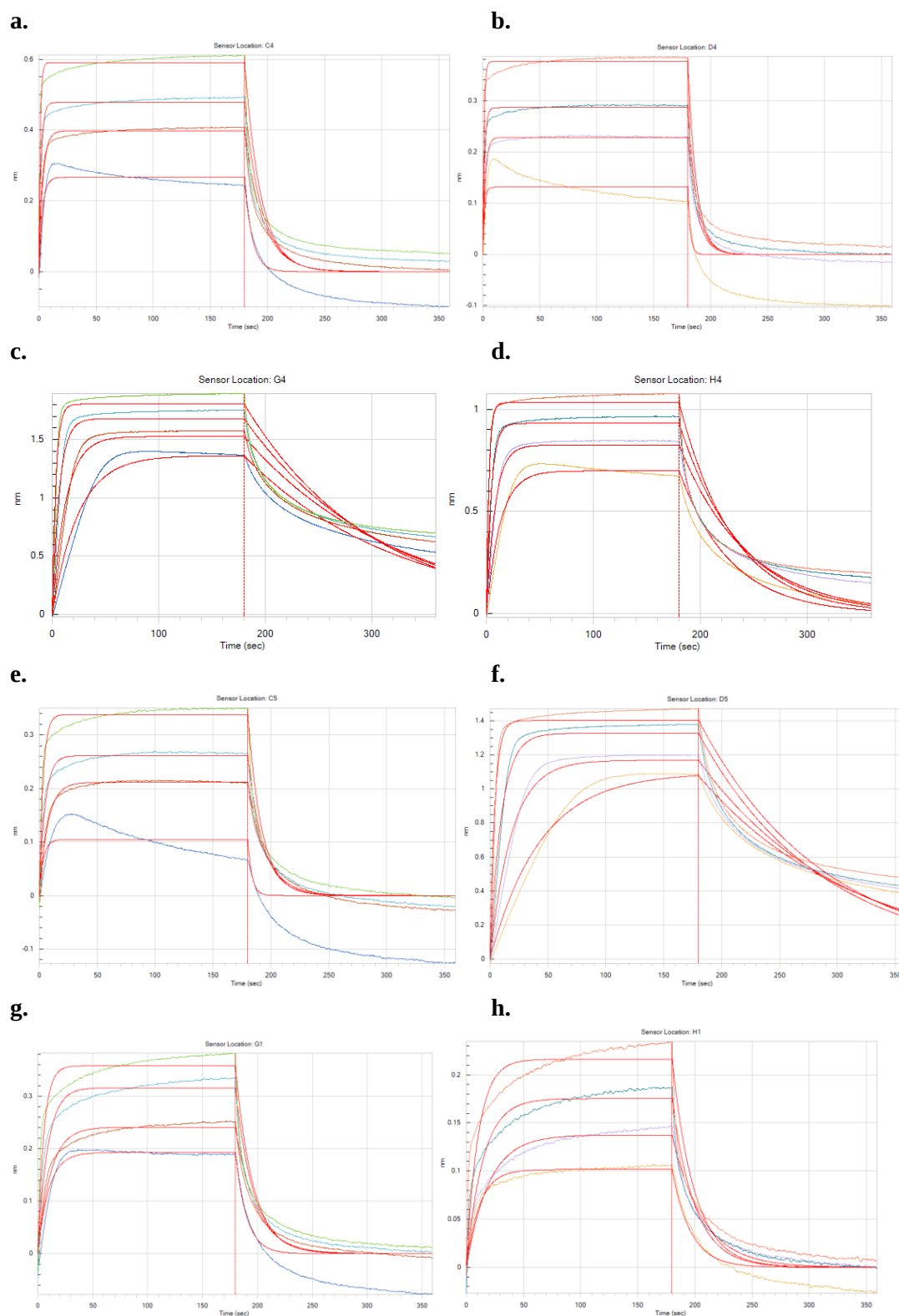


Figure 2.5. Binding of peptide and glycopeptides 5, 10, 18, 19, 20, 21 and commercially available biotinylated 50 kDa CS, 50 kDa CS-A to neutrophil Cathepsin G as measured by BLI

Figure 2.5. (cont'd)

(**a-h** respectively). The biotinylated compounds (50 nM) were immobilized on streptavidin coated biosensors, and human neutrophil Cathepsin G was captured on biosensors with four concentrations at 2000 nM, 1000 nM, 500 nM, 250 nM. Fitting curves were shown in red lines.

System	Region Scores (<i>kcal mol⁻¹</i>)
5	-17.63
12-mer-0S	-21.34
12-mer-5S	-24.81
12-mer-7S	-26.46
12-mer-9S	-26.63
12-mer-57S	-24.68
12-mer-79S	-25.64
21	-31.81
19	-26.98
20	-26.88

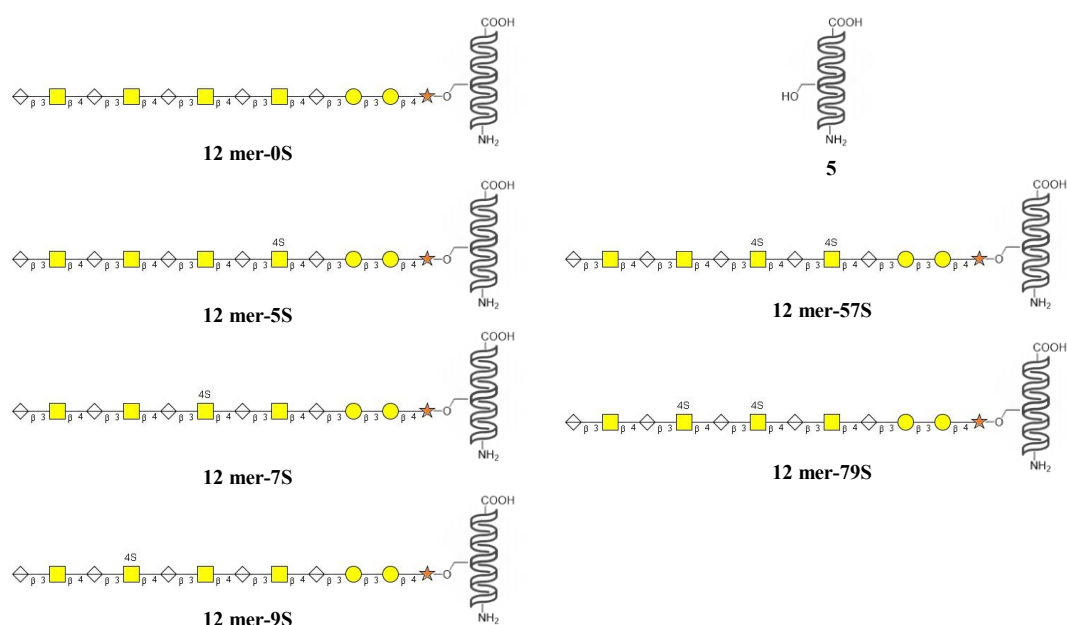


Table 2.2. The docking scores of the investigated glycopeptides, as described in the Methods section. Annotation of glycopeptide structures are shown.

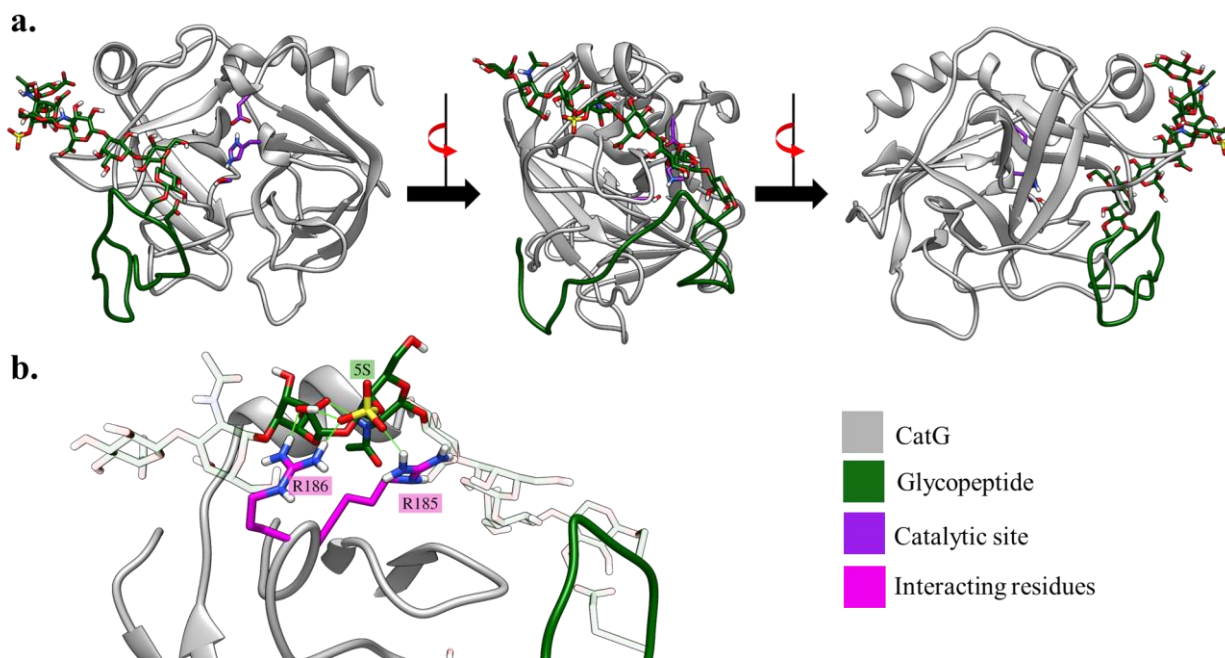


Figure 2.6. (a) The highest scoring poses for **19**. (b) The direct hydrogen bond interactions of sulfate on GalNAc 5. The numbering of the residues follows the 1T32 PDB numbering.

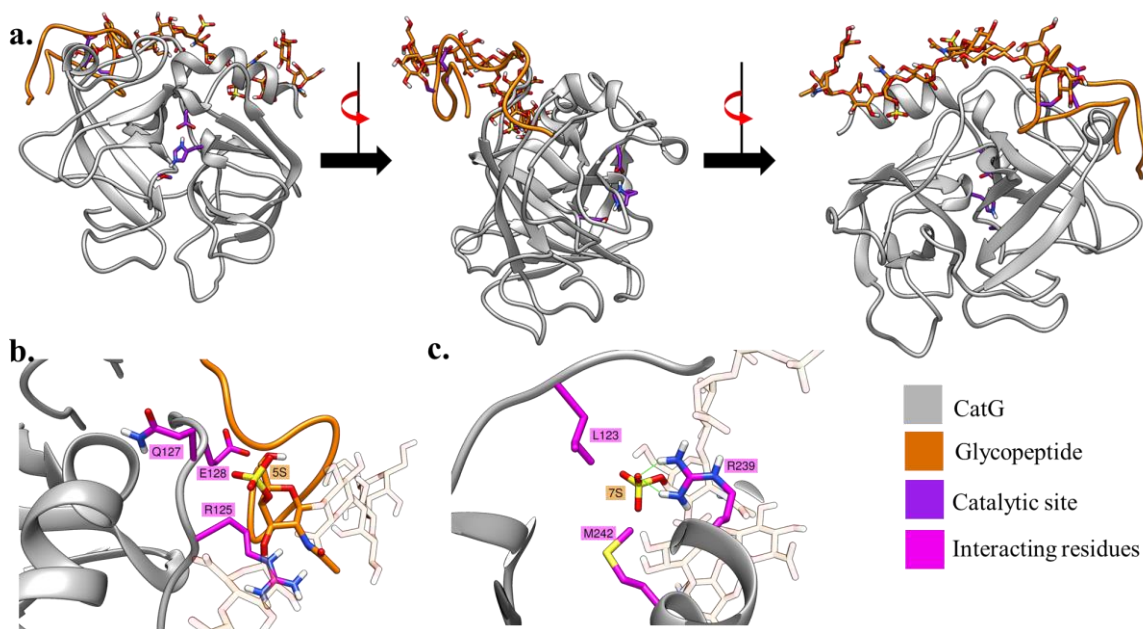


Figure 2.7. (a) The highest scoring poses for **20**. (b) The direct hydrogen bond interactions of

Figure 2.7. (cont'd)

sulfate on GalNAc 5. (c) The direct hydrogen bond interactions of sulfate on GalNAc 7. The numbering of the residues follows the 1T32 PDB numbering.

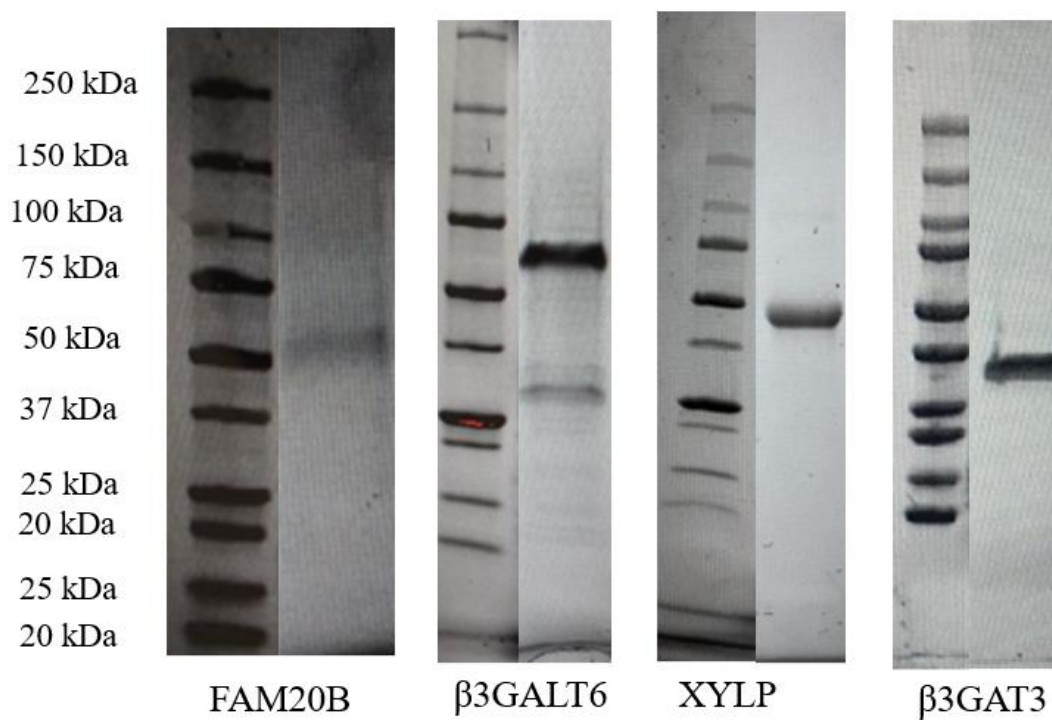


Figure 2.8. SDS page gel of purified β 3GAT3, FAM20B, β 3GALT6 and XYLP.

APPENDIX B: PRODUCT CHARACTERIZATION SPECTRA

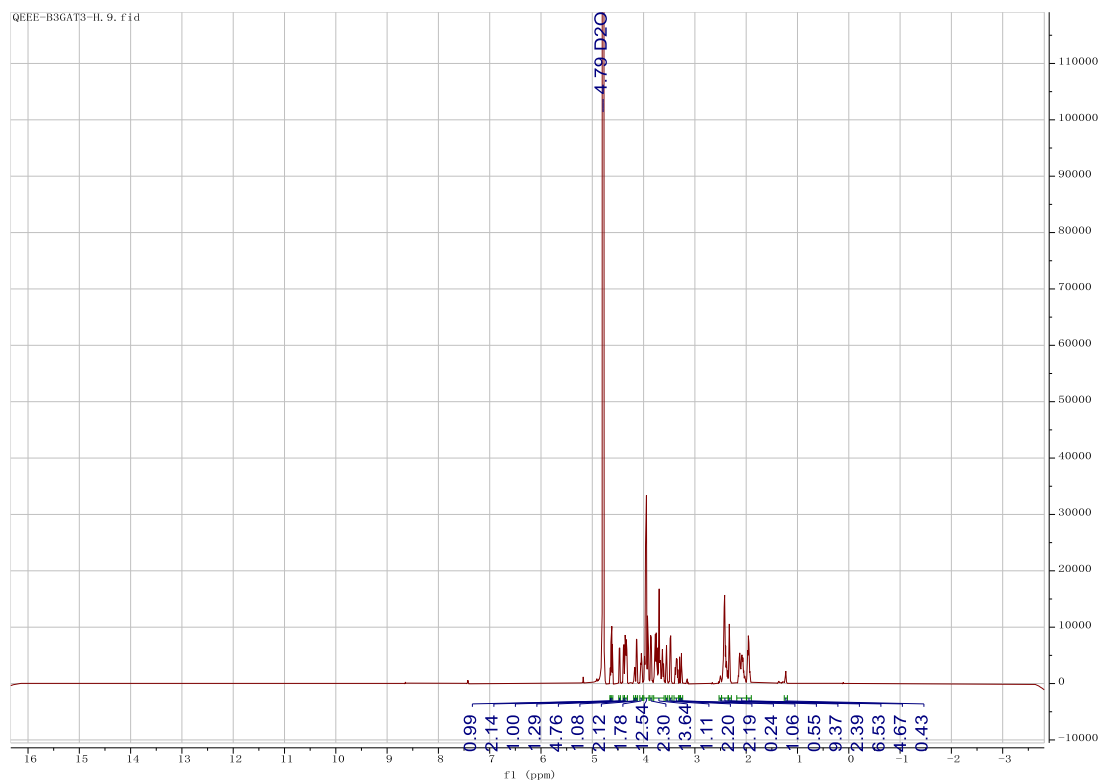
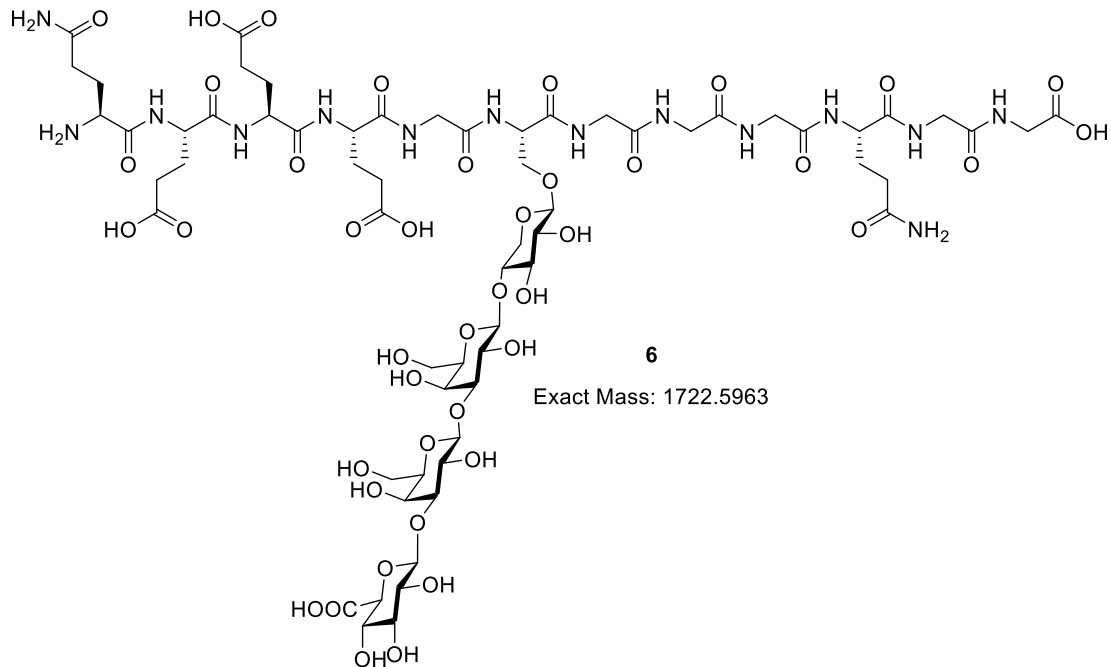


Figure 2.9. ¹H NMR of **6** (800 MHz, D₂O).

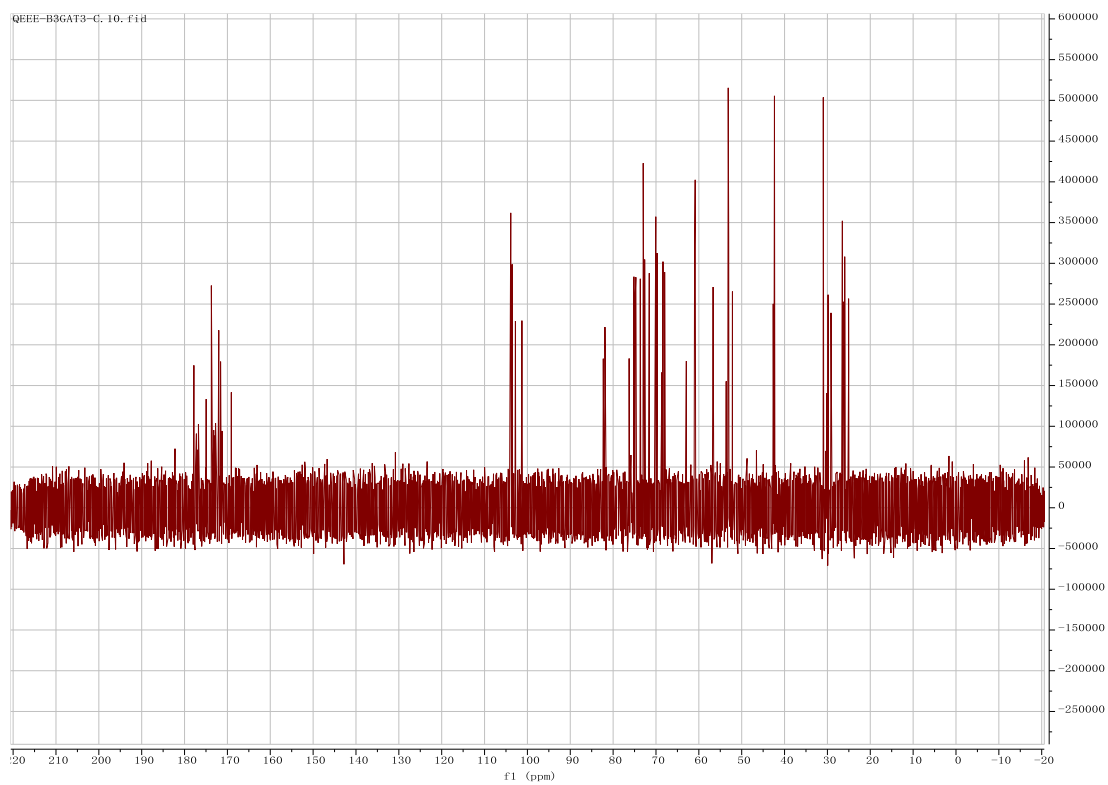
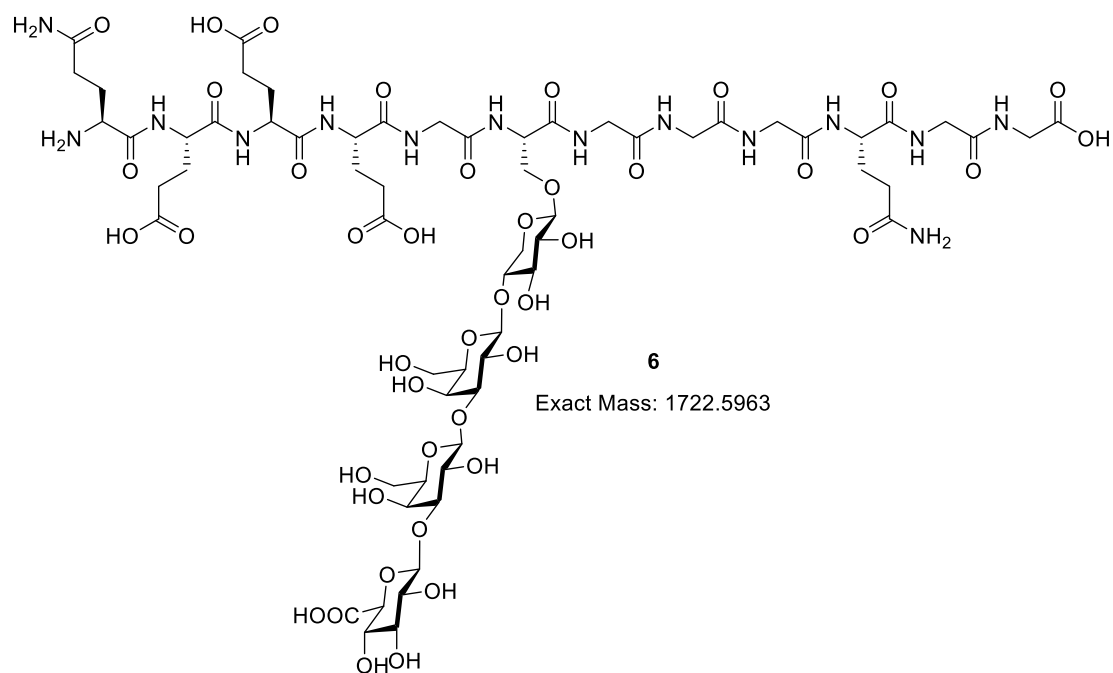


Figure 2.10. ^{13}C NMR of **6** (201 MHz, D_2O).

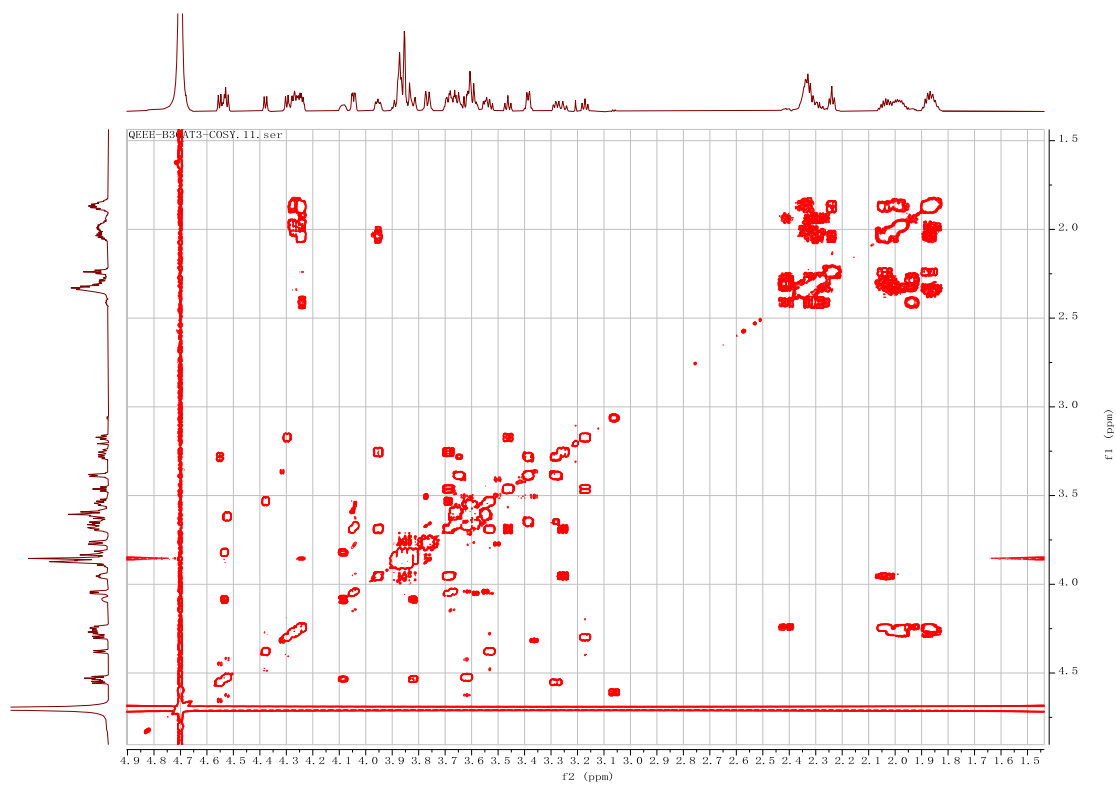
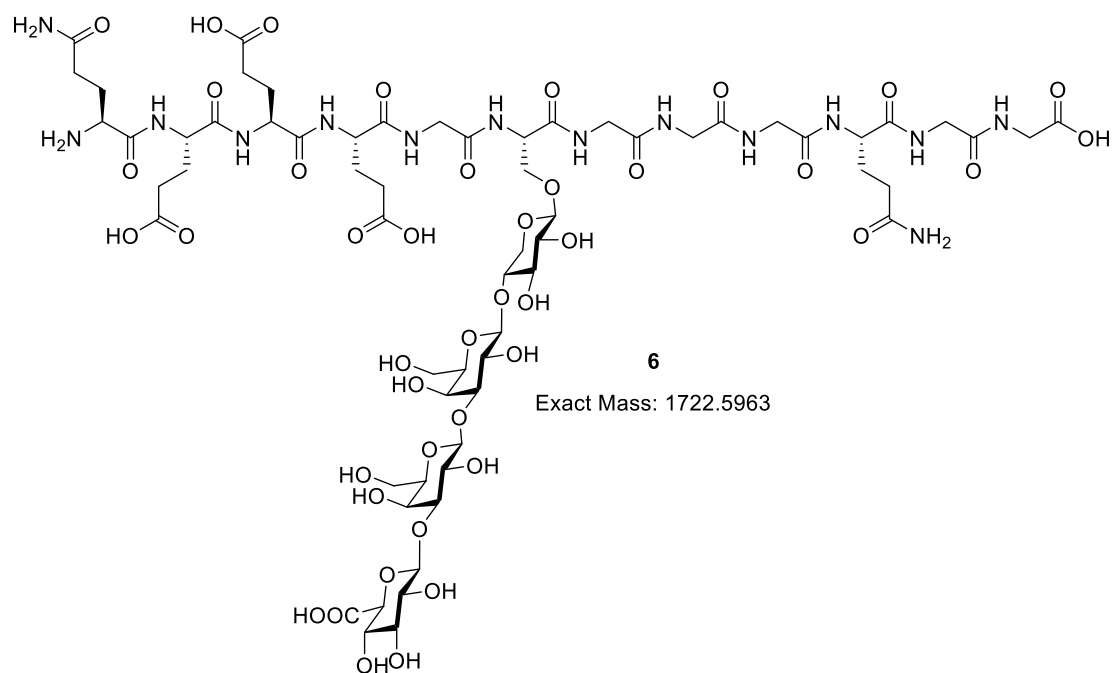


Figure 2.11. COSY of **6** (800 MHz, D₂O).

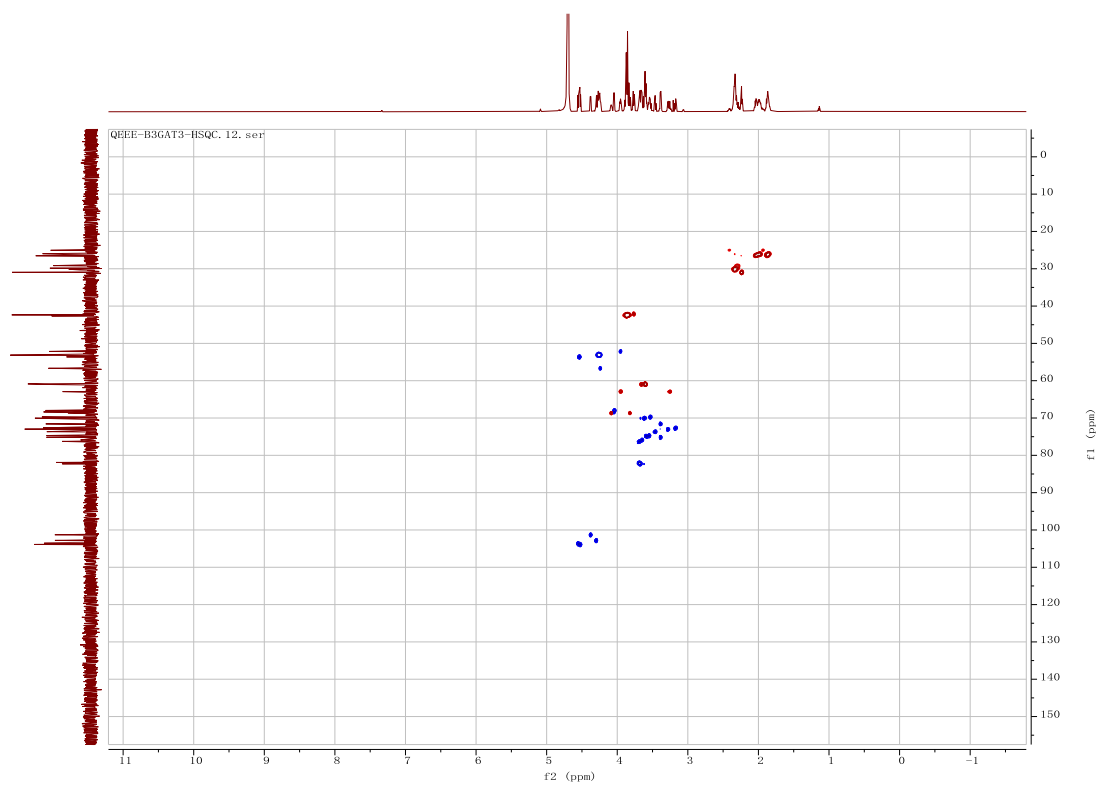
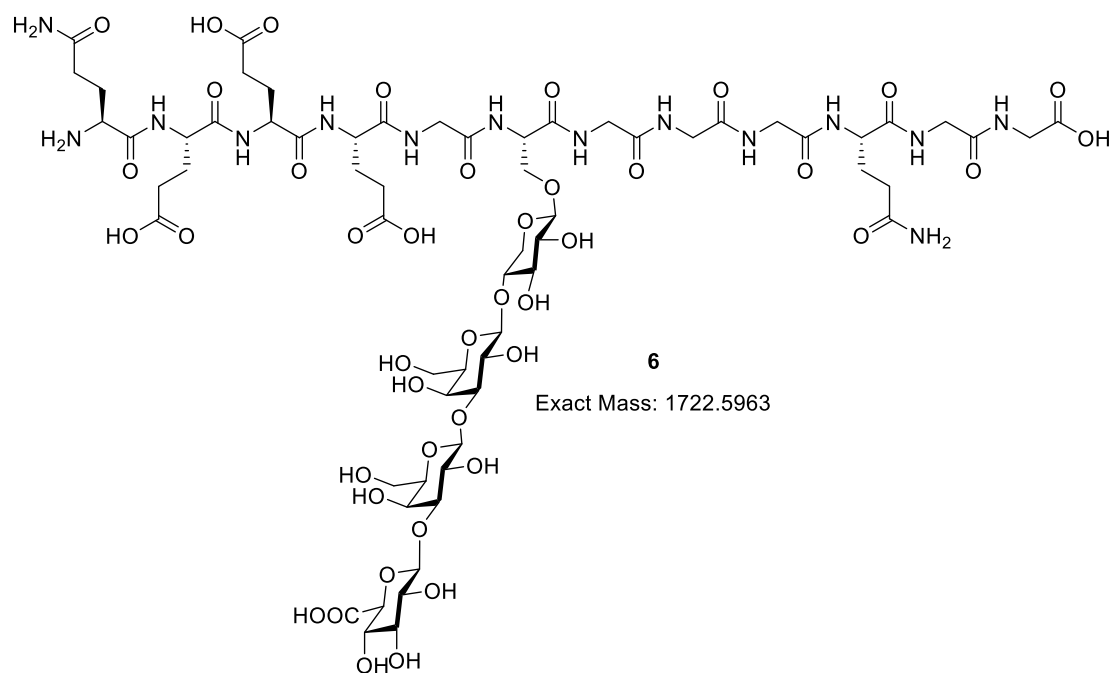


Figure 2.12. HSQC of **6** (800 MHz, D_2O).

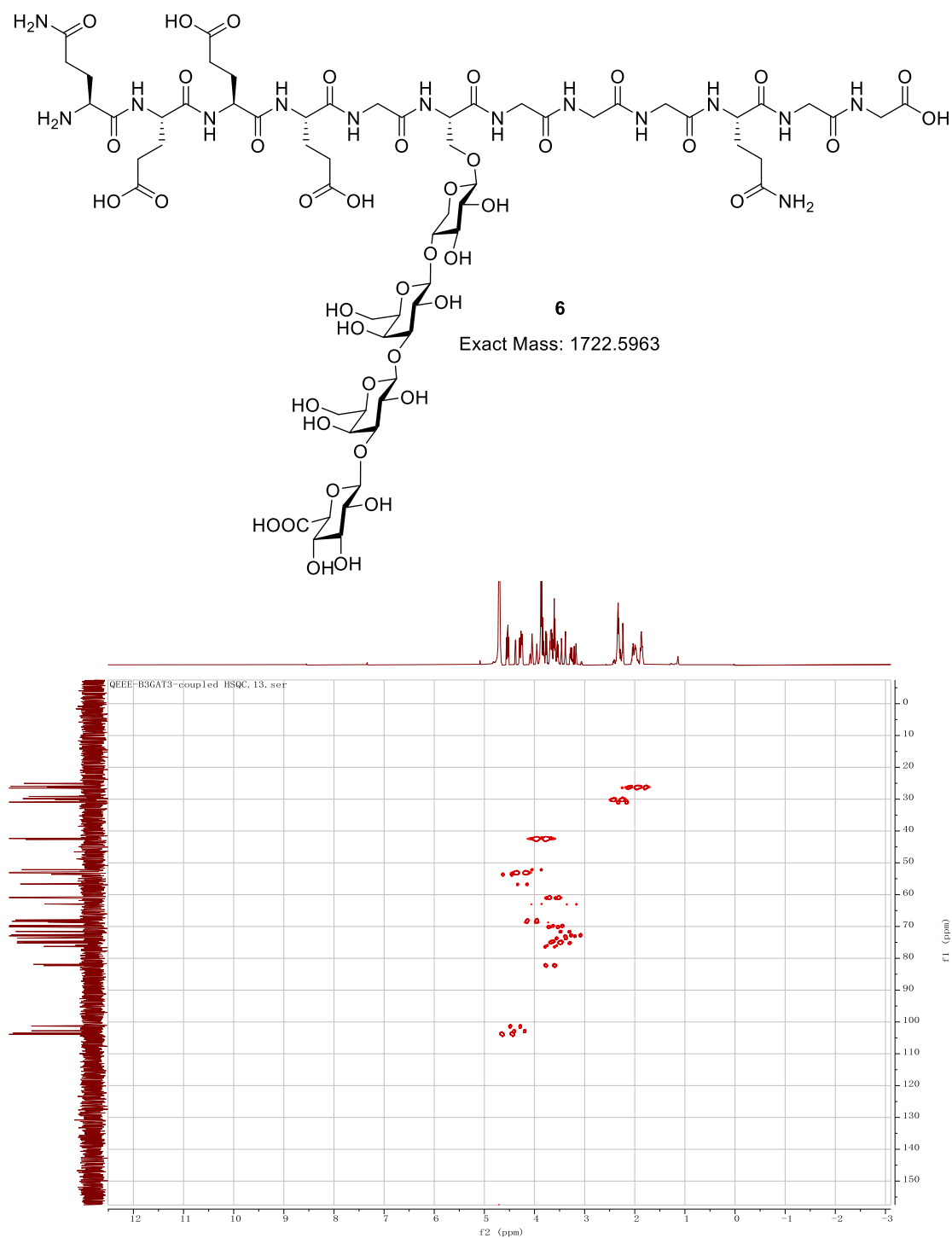


Figure 2.13. Coupled HSQC of **6** (800 MHz, D₂O).

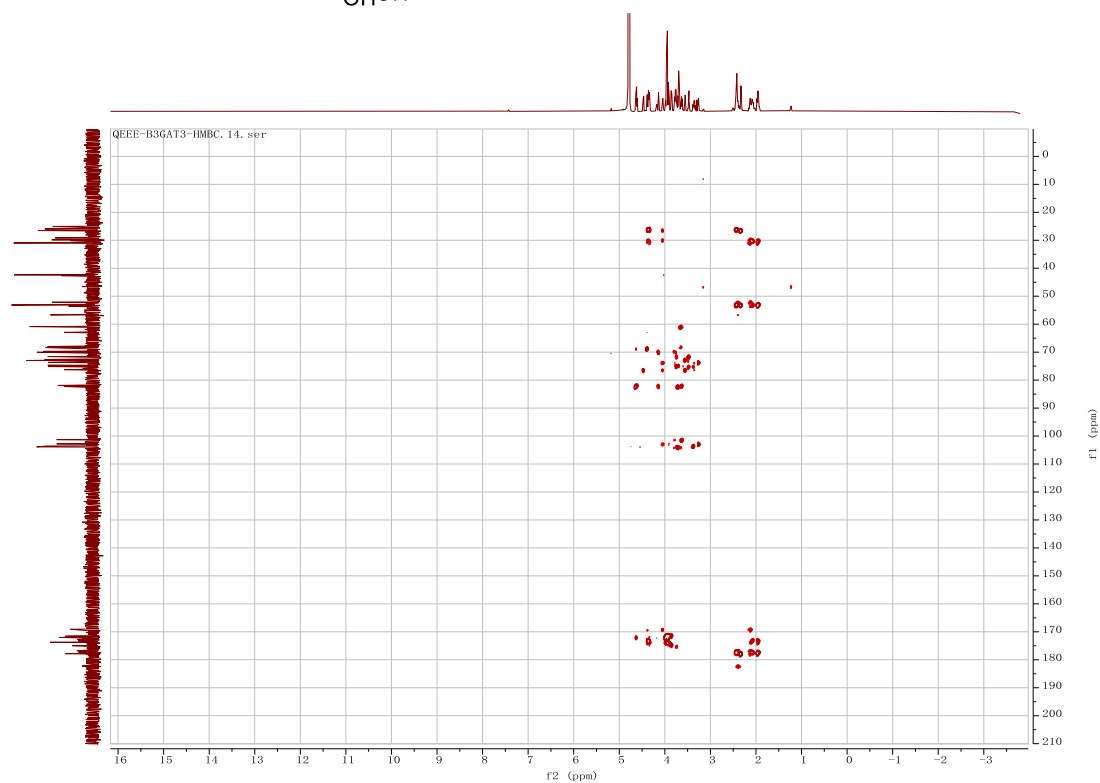
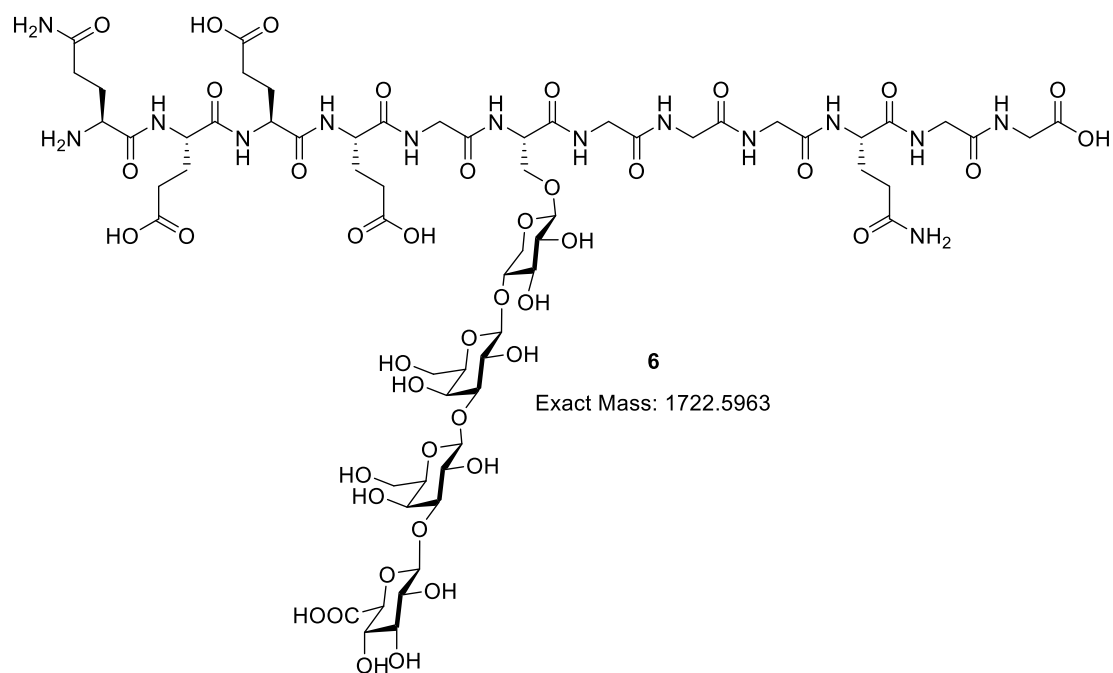


Figure 2.14. HMBC of **6** (800 MHz, D₂O).

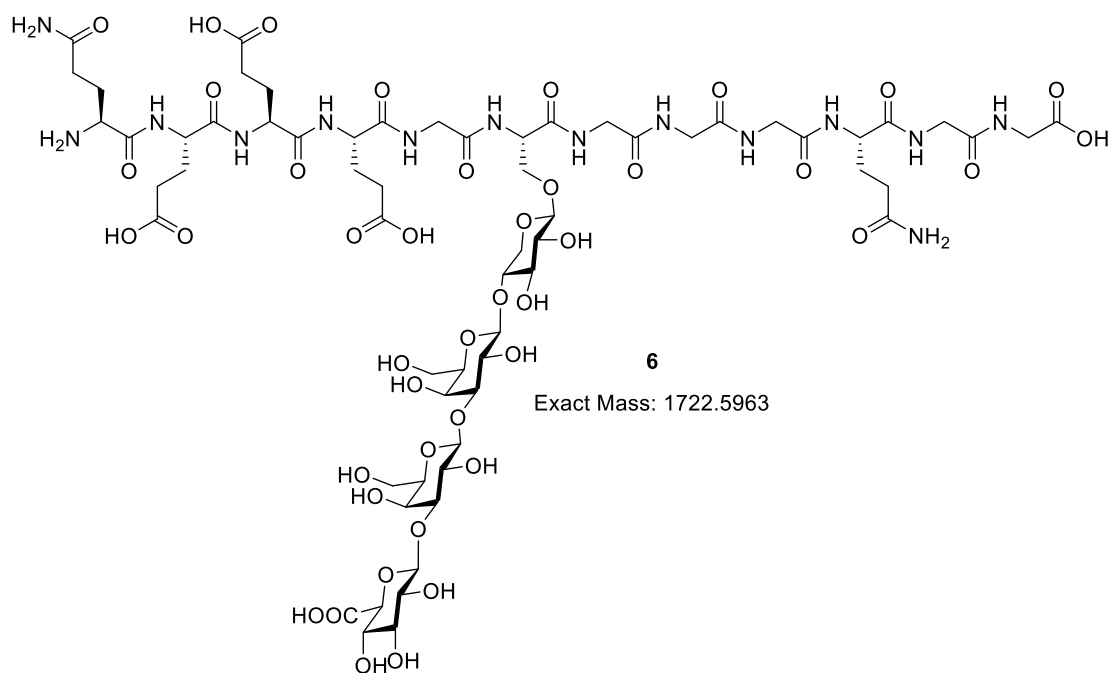
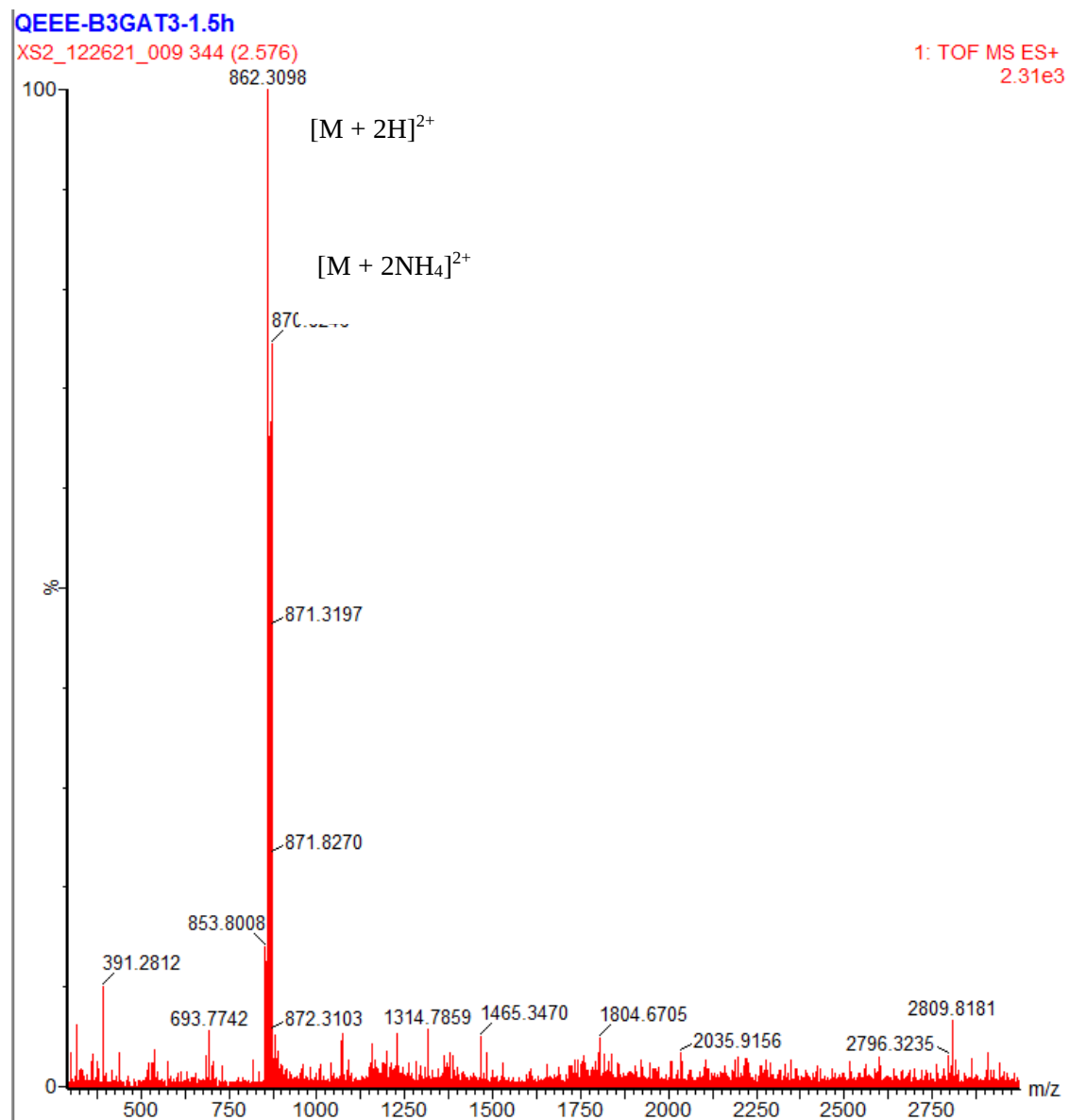


Figure 2.15. LCMS chromatogram of **6**.

Figure 2.15. (cont'd)



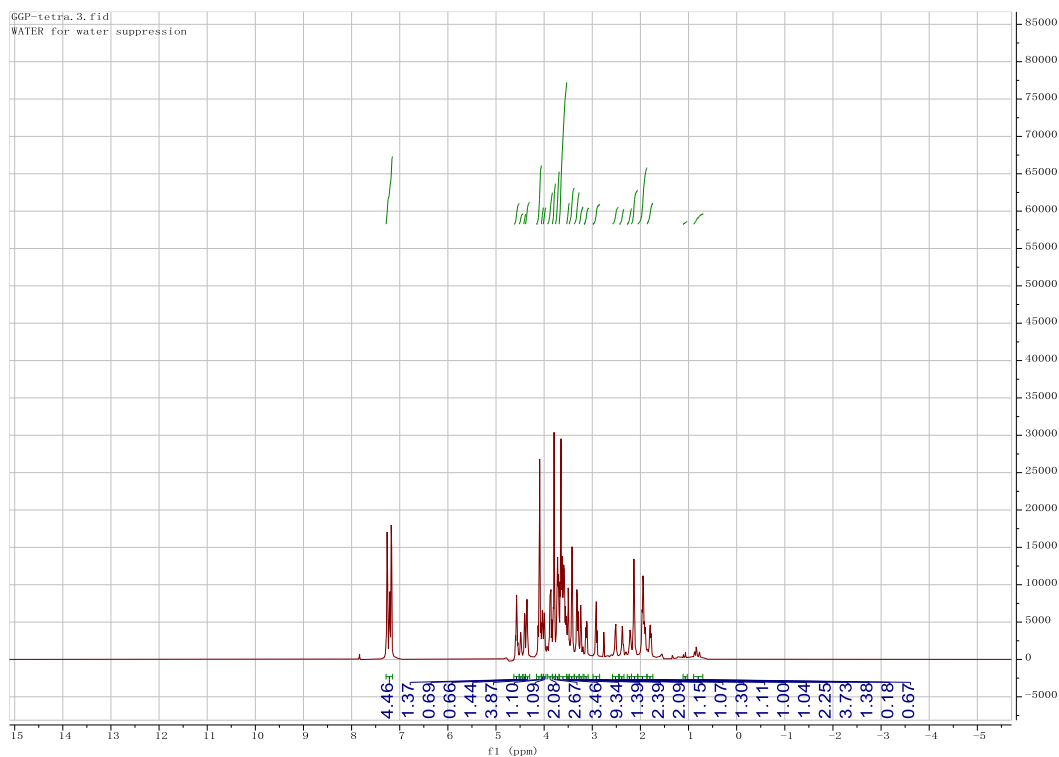
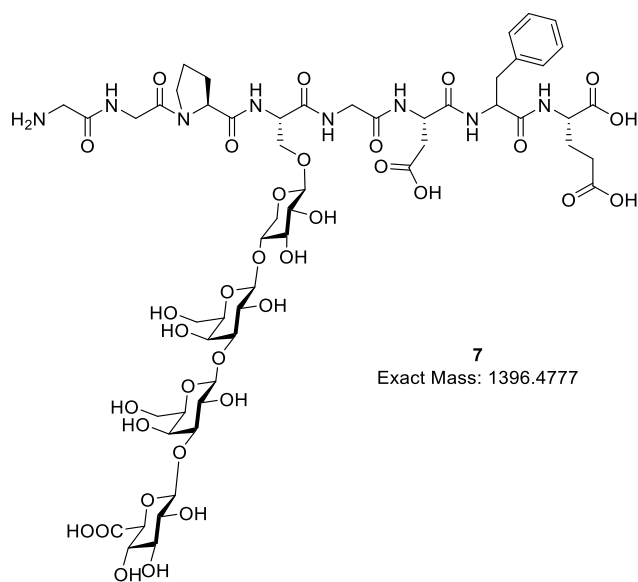


Figure 2.16. ^1H -NMR of **7** (800 MHz, D_2O).

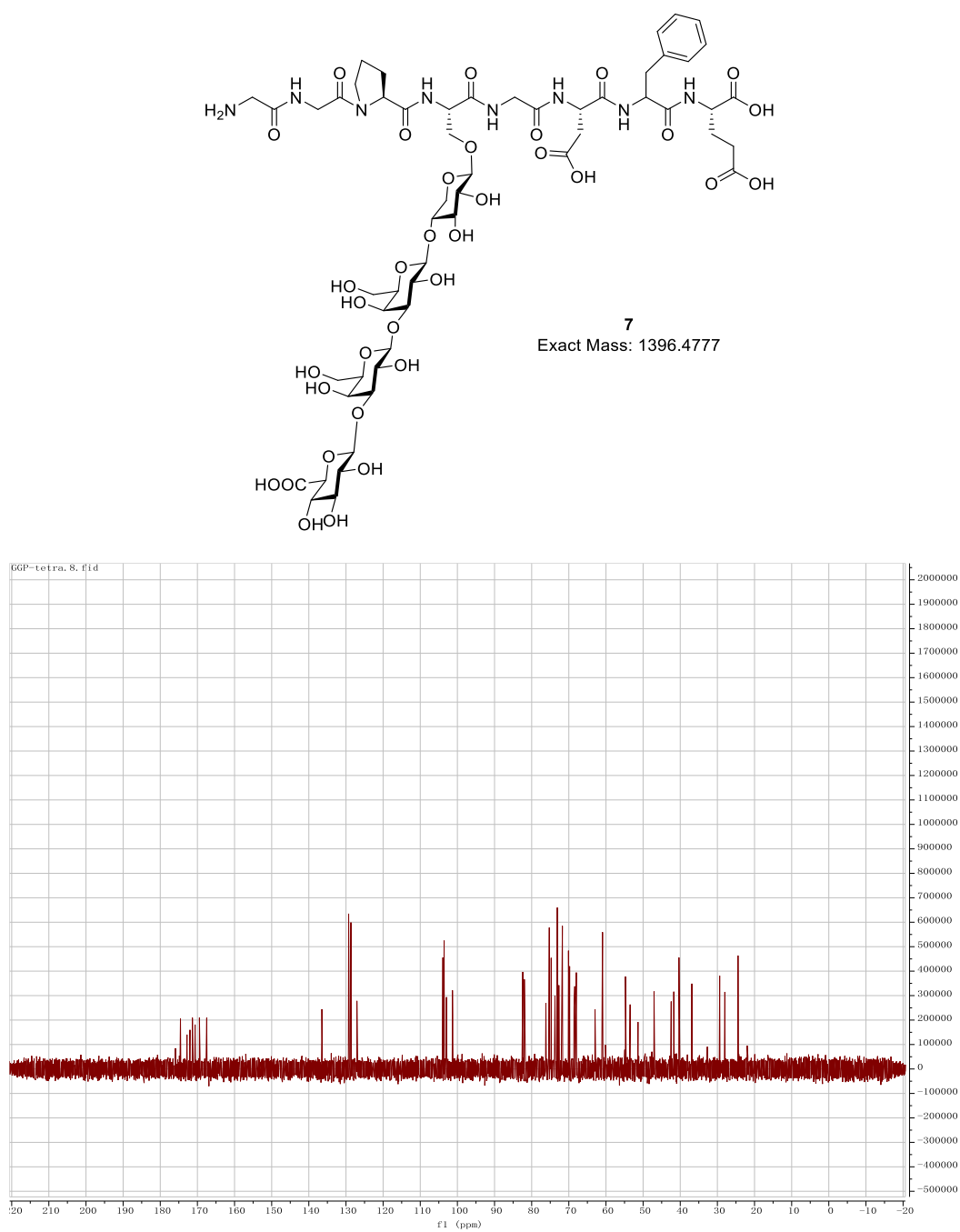


Figure 2.17. ^{13}C NMR of **7** (201 MHz, D_2O).

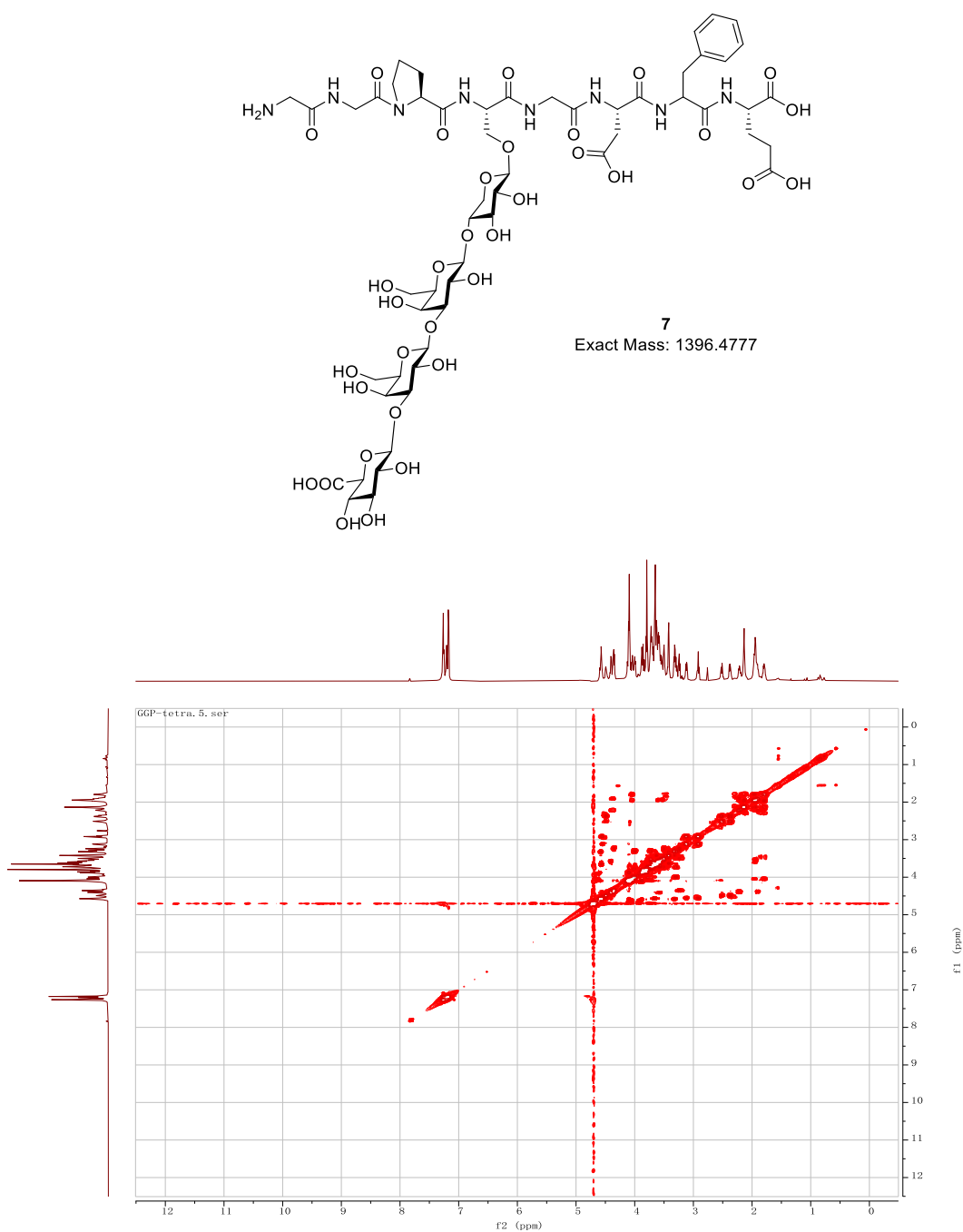


Figure 2.18. COSY of **7** (800 MHz, D₂O).

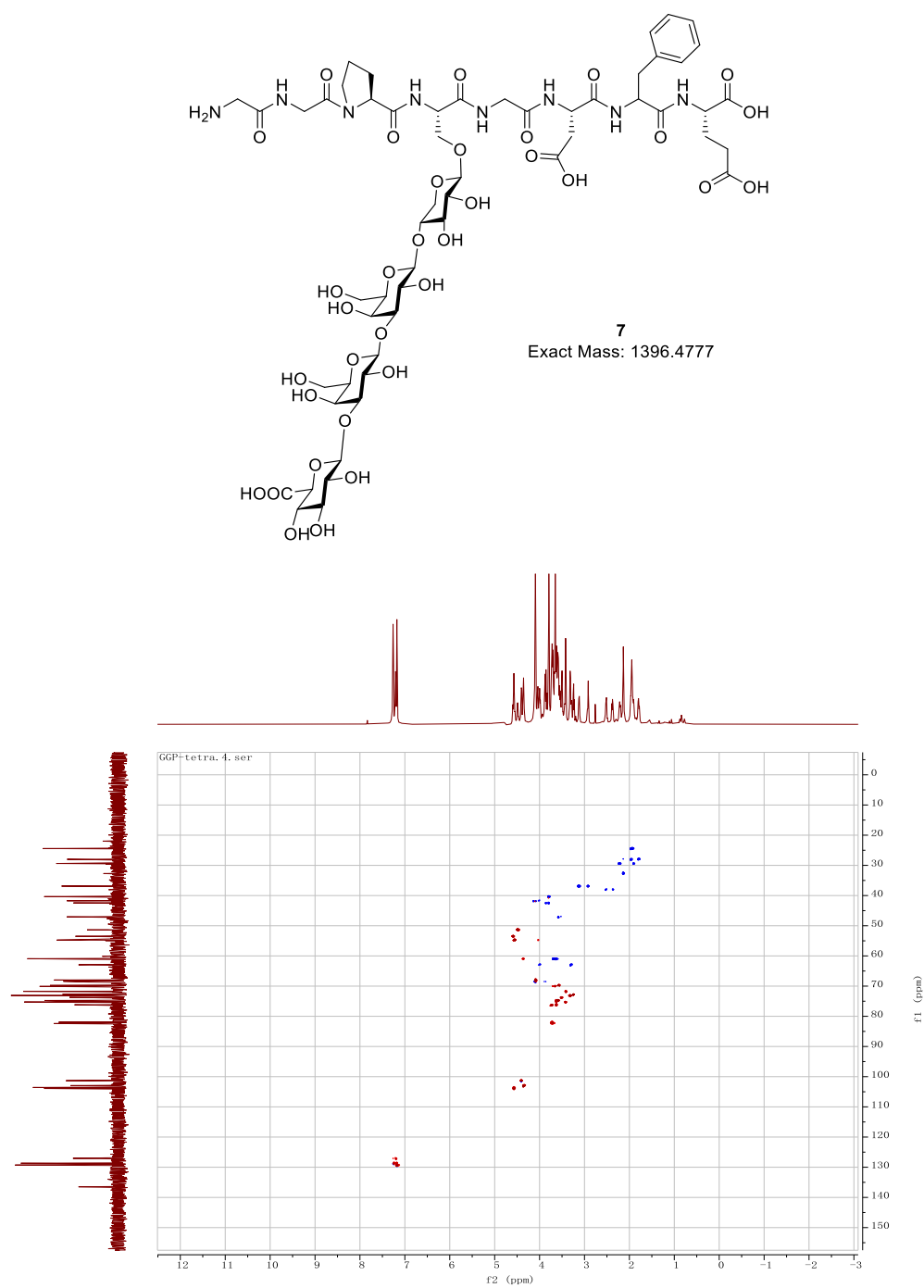


Figure 2.19. HSQC of **7** (800 MHz, D_2O).

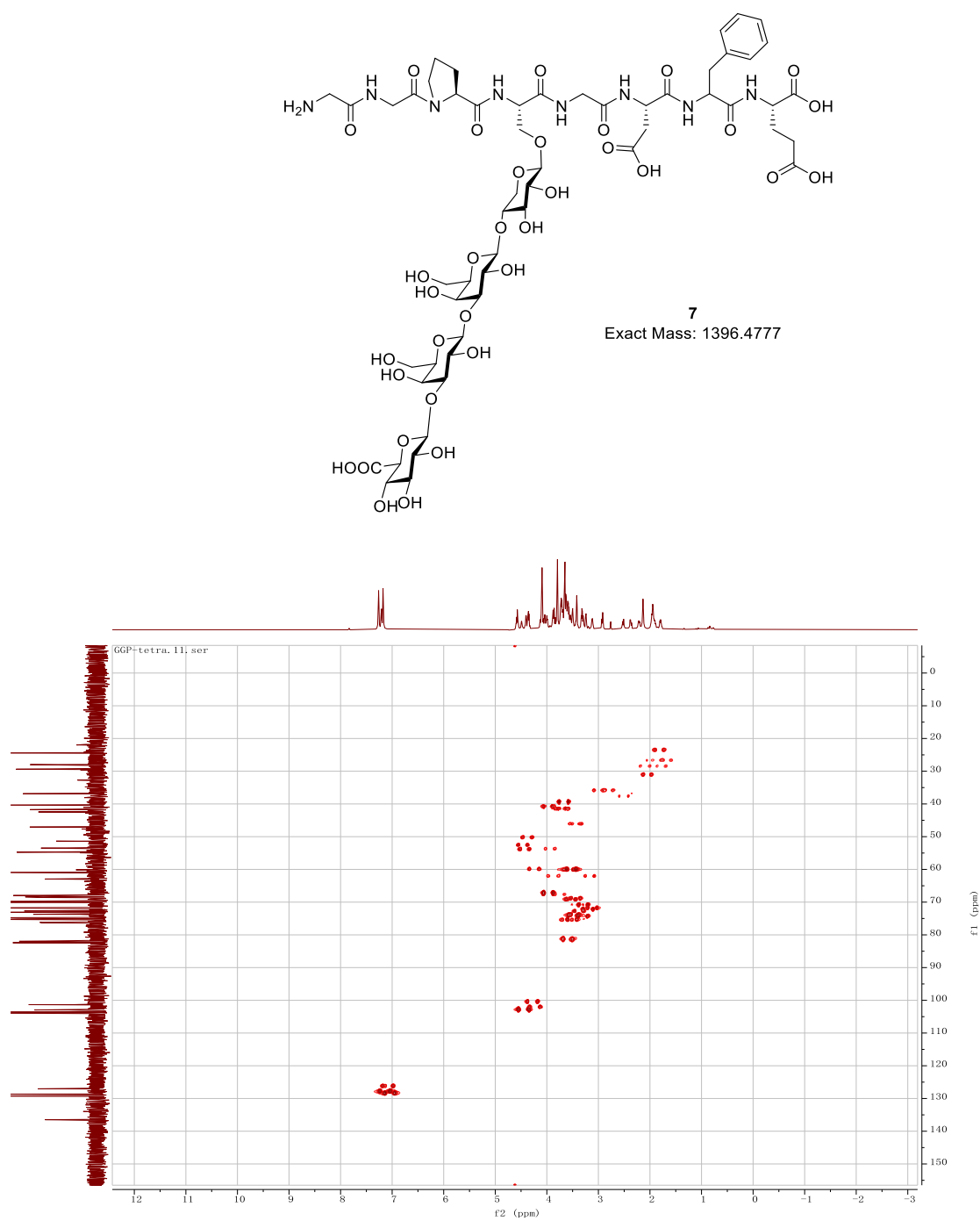


Figure 2.20. Coupled HSQC of **7** (800 MHz, D₂O).

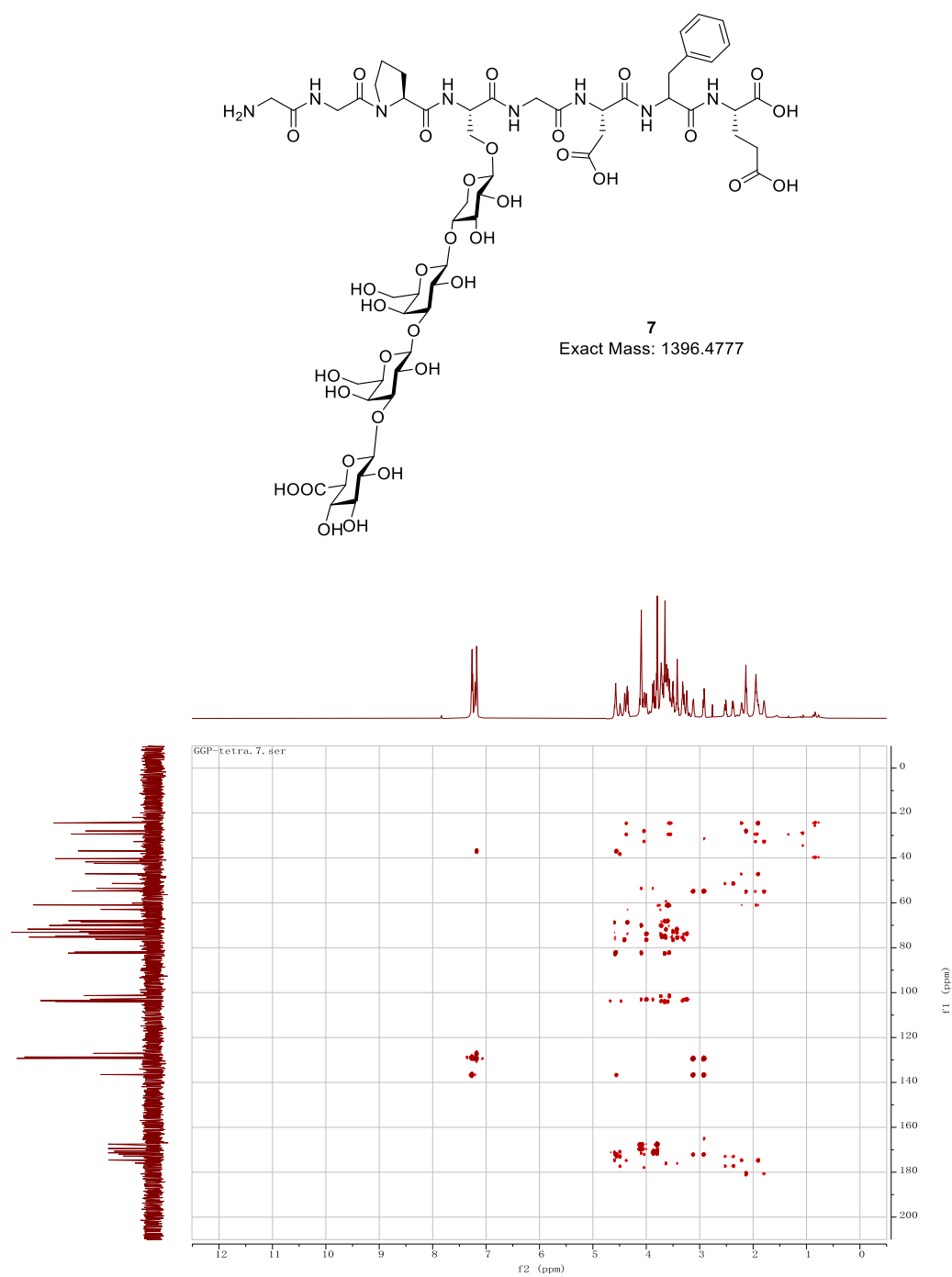


Figure 2.21. HMBC of **7** (800 MHz, D₂O).

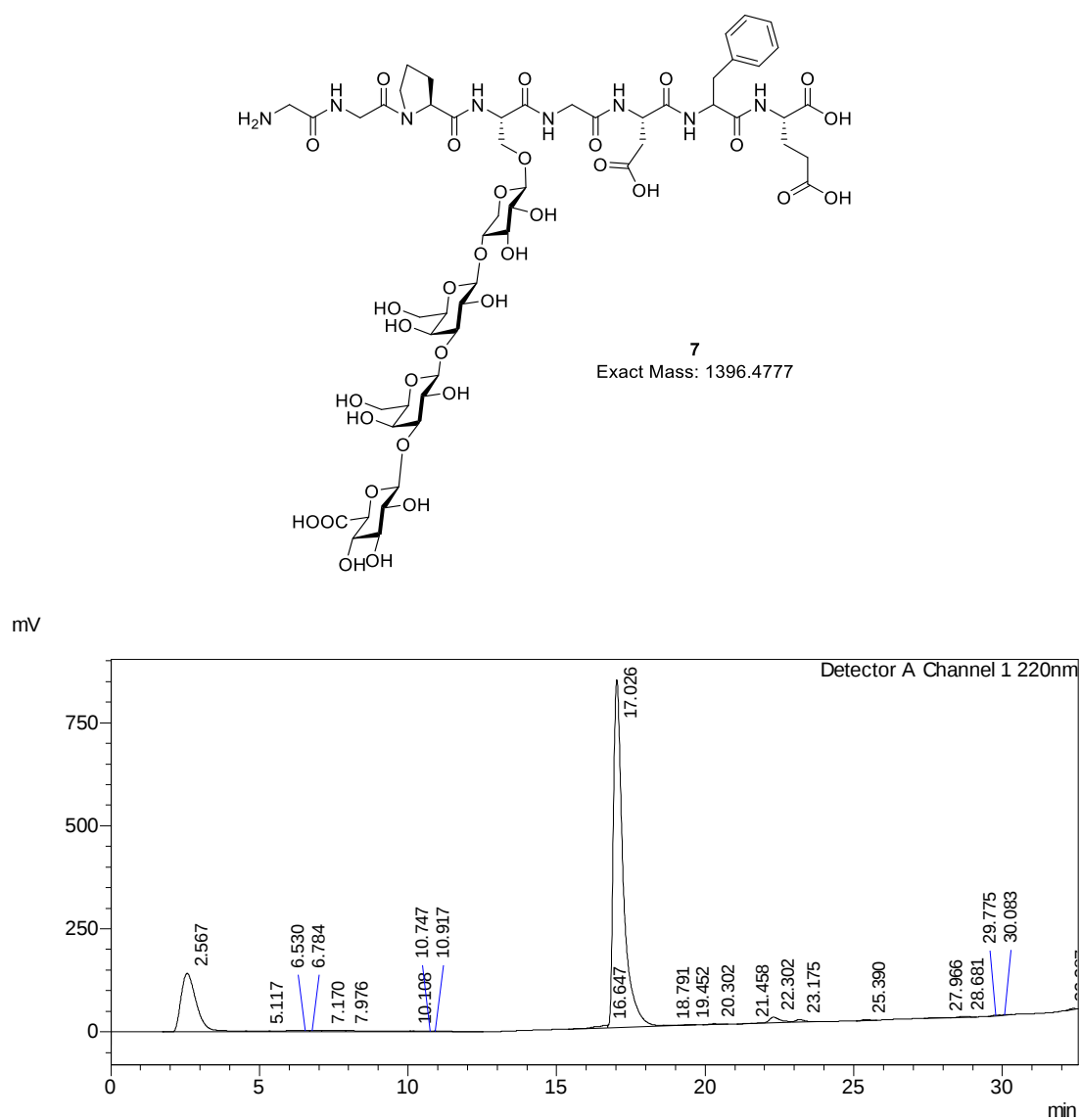


Figure 2.22. HPLC Chromatogram of **7**.

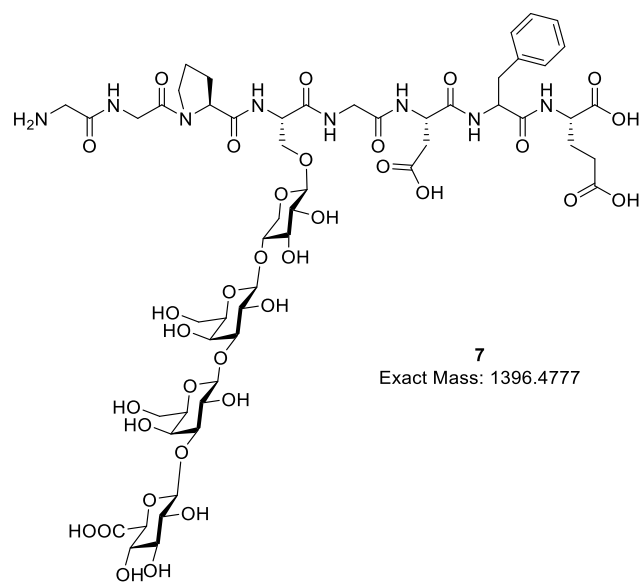


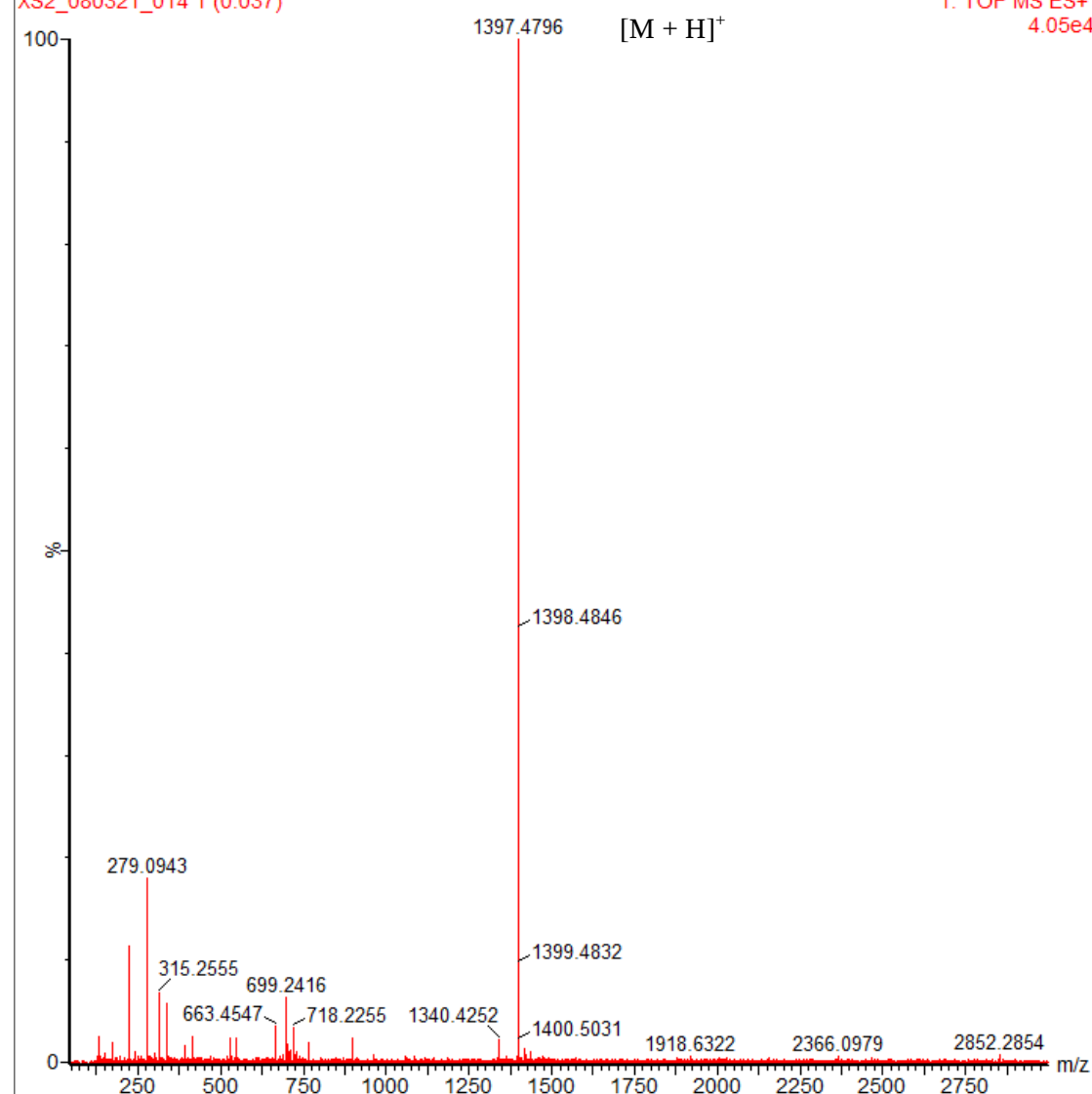
Figure 2.23. LCMS Chromatogram of **7**.

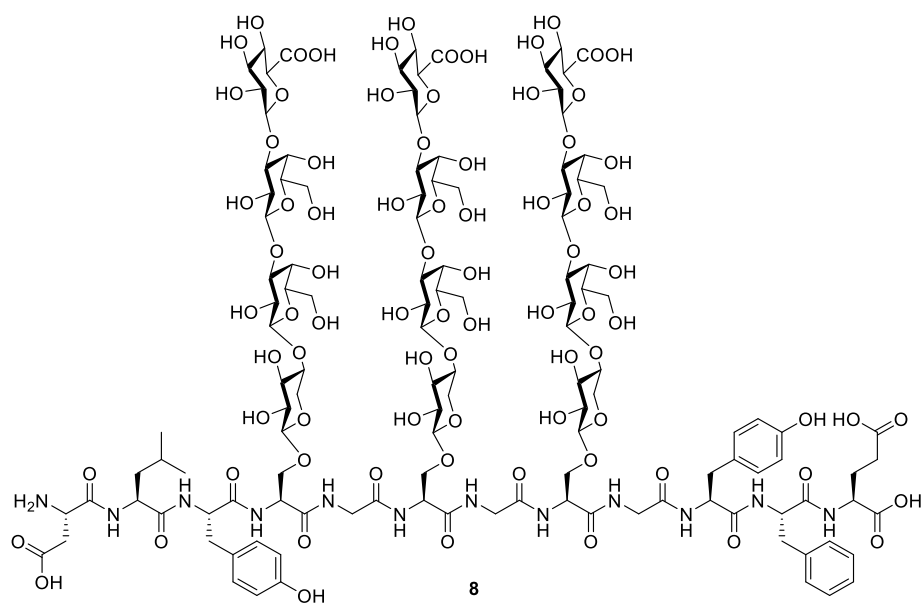
Figure 2.23. (cont'd)

GGP- HILIC B3GAT3 repurify-119

XS2_080321_014 1 (0.037)

1: TOF MS ES+
4.05e4





8

Exact Mass: 3177.0597

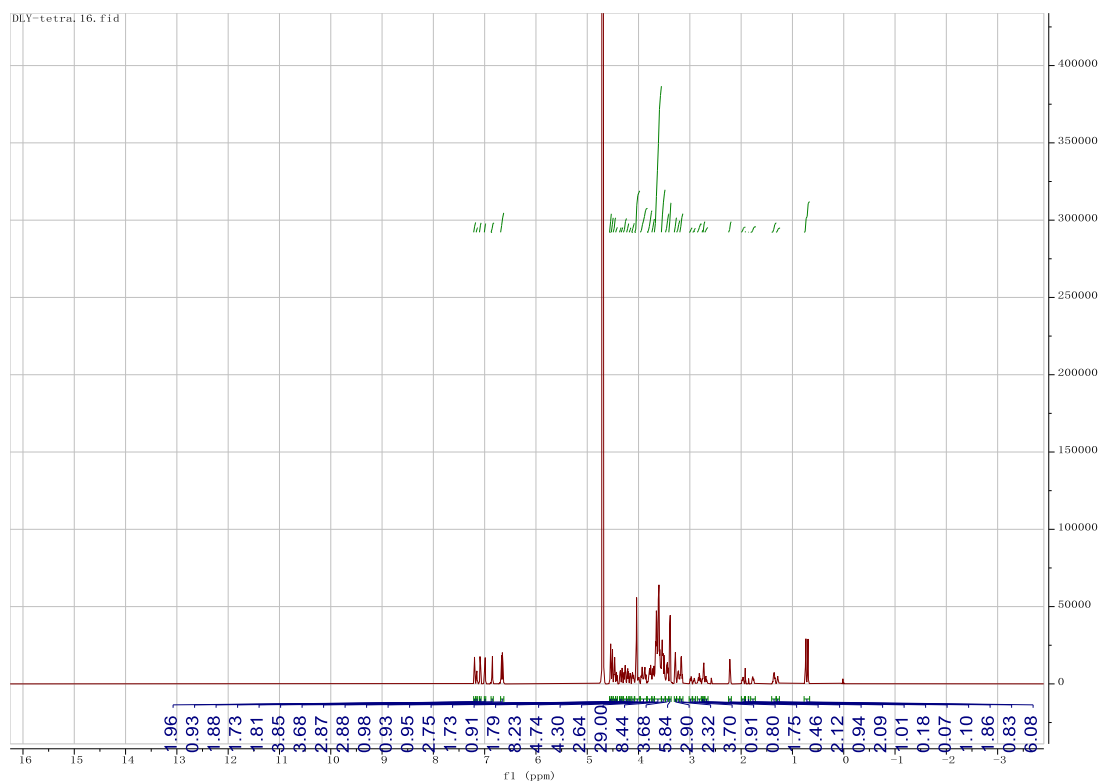
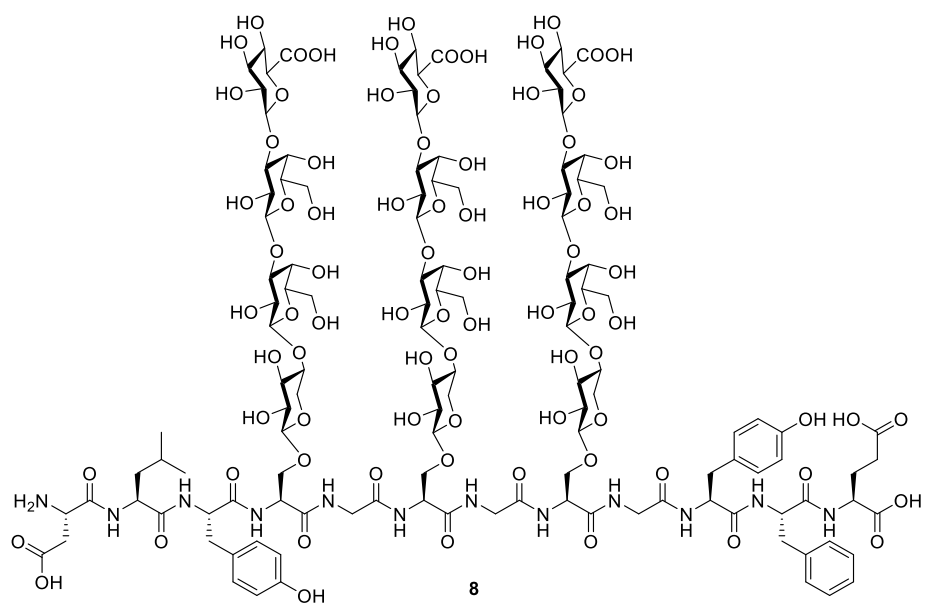


Figure 2.24. ¹H-NMR of **8** (800 MHz, D₂O).



Exact Mass: 3177.0597

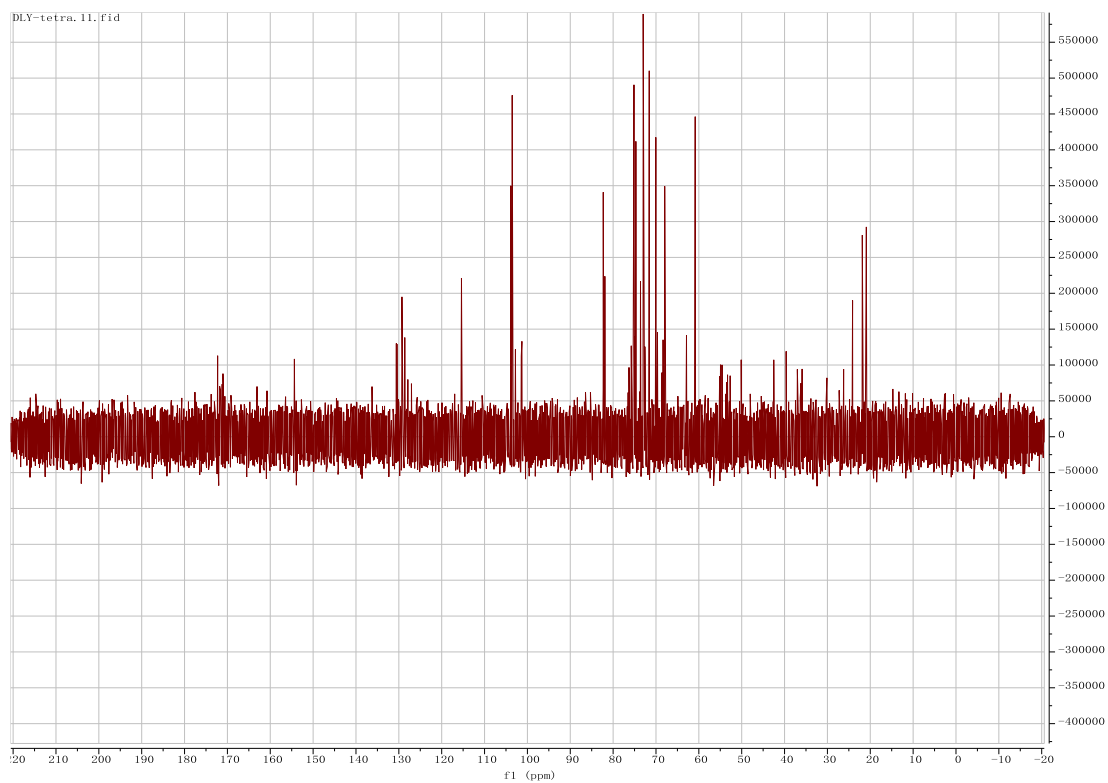
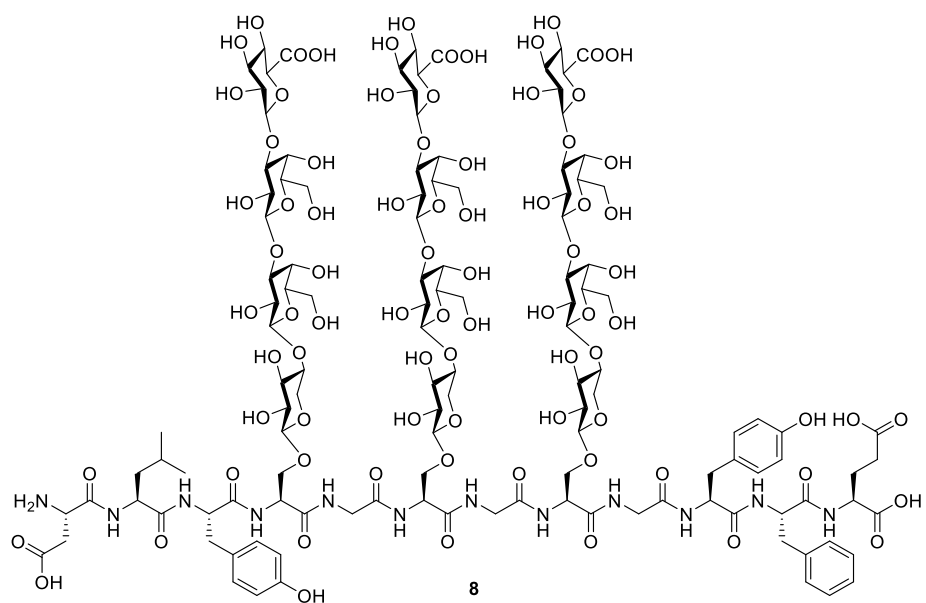


Figure 2.25. ^{13}C NMR of **8** (201 MHz, D_2O).



Exact Mass: 3177.0597



Figure 2.26. COSY of **8** (800 MHz, D₂O).

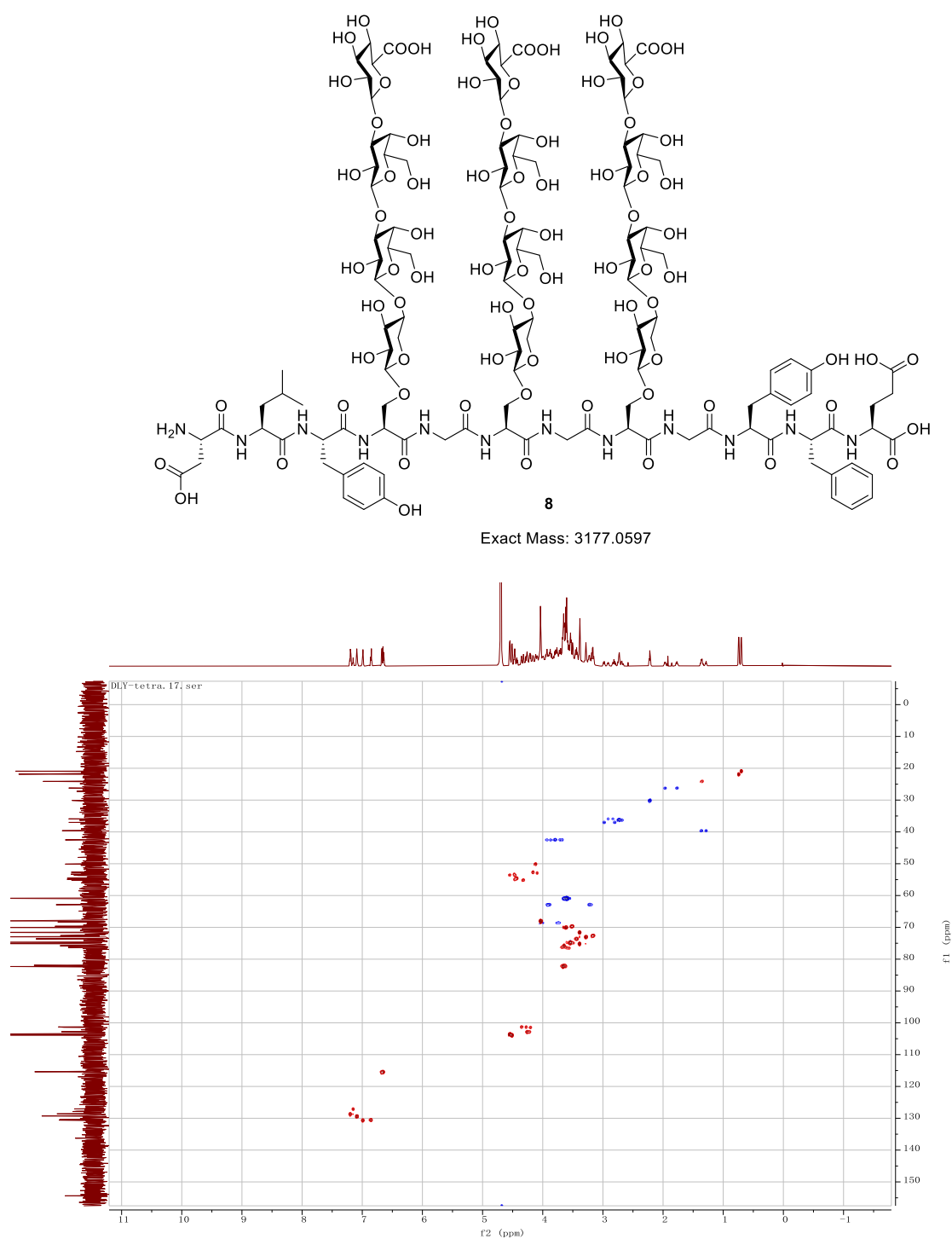
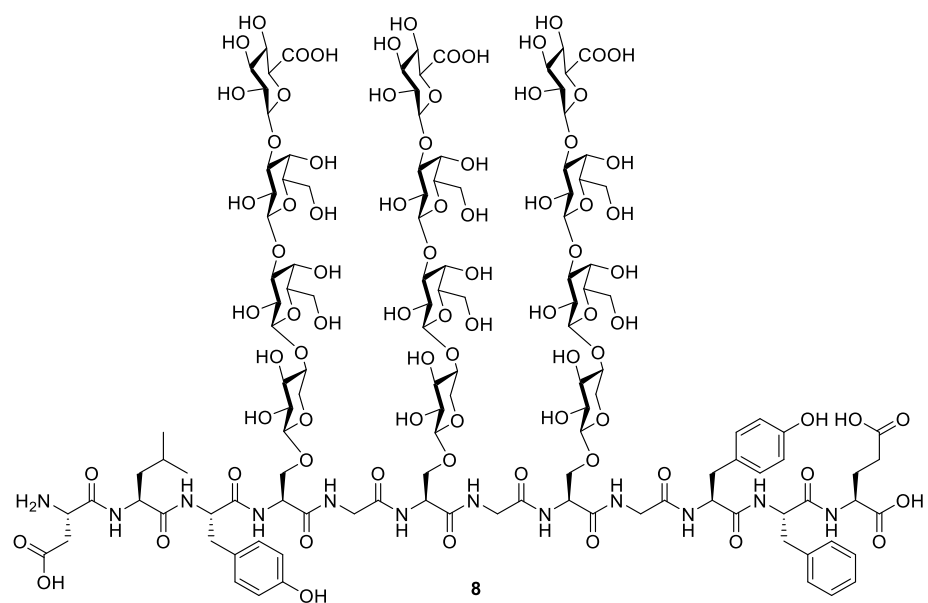


Figure 2.27. HSQC of **8** (800 MHz, D_2O).



Exact Mass: 3177.0597

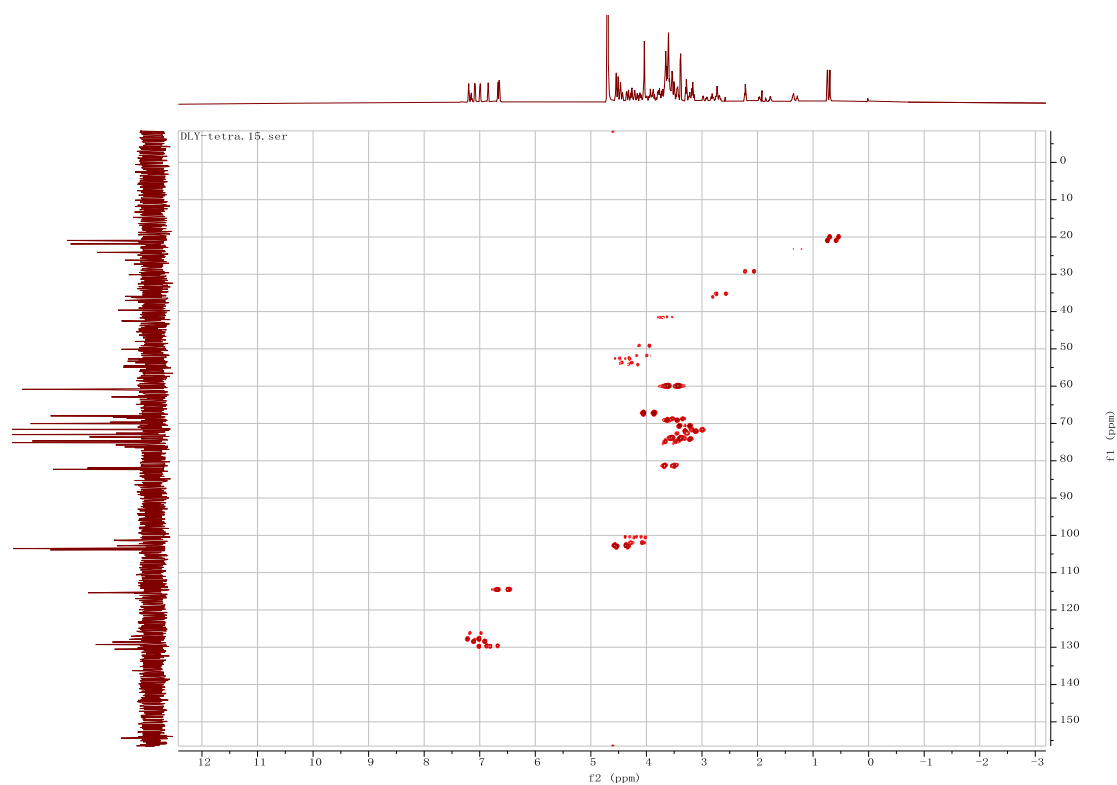
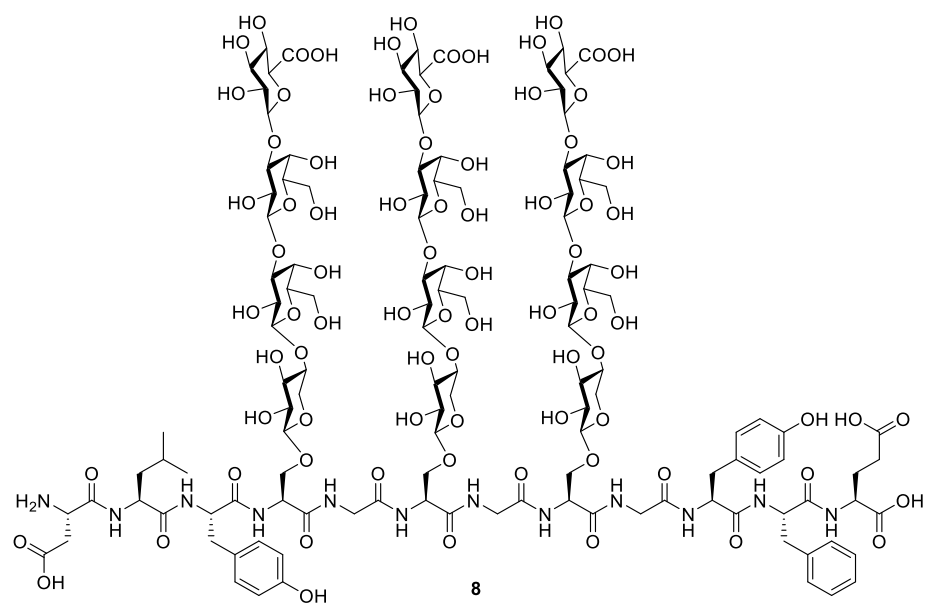


Figure 2.28. Coupled HSQC of **8** (800 MHz, D₂O).



Exact Mass: 3177.0597

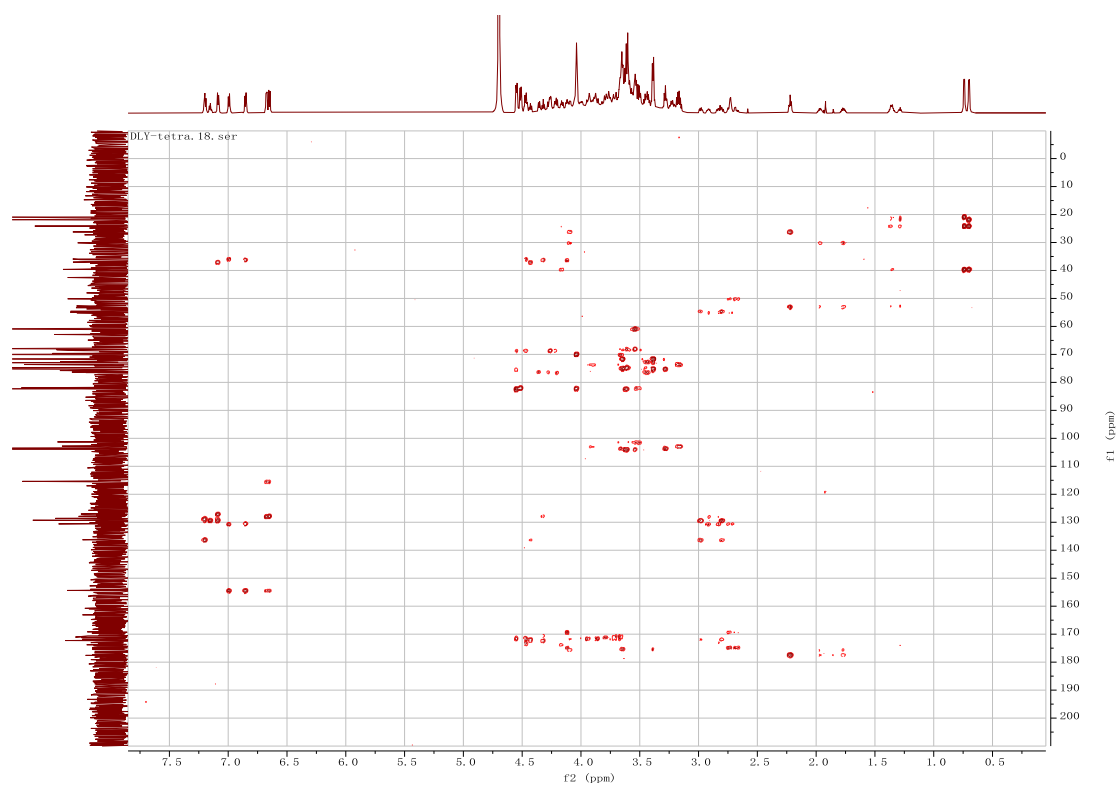
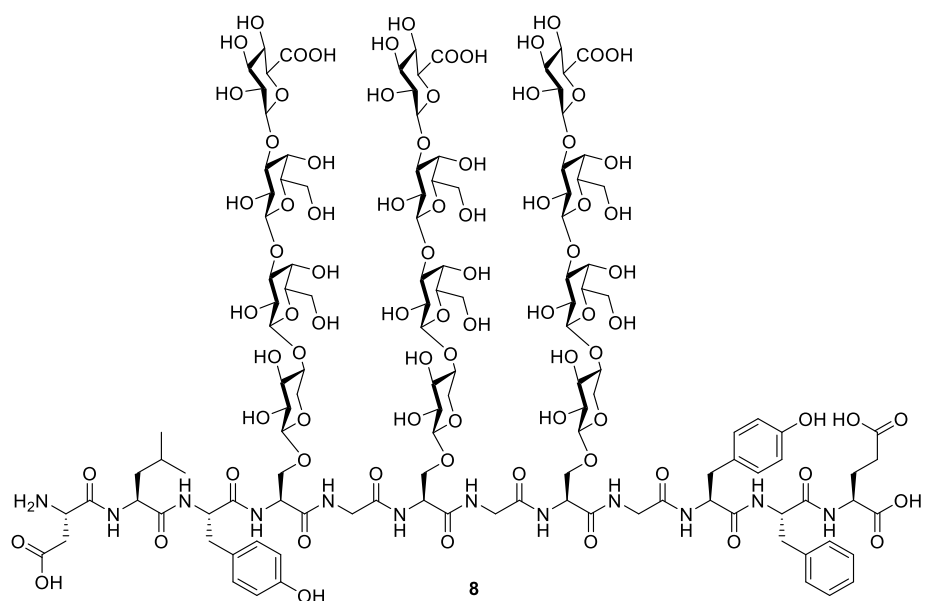


Figure 2.29. HMBC of **8** (800 MHz, D₂O).



8

Exact Mass: 3177.0597

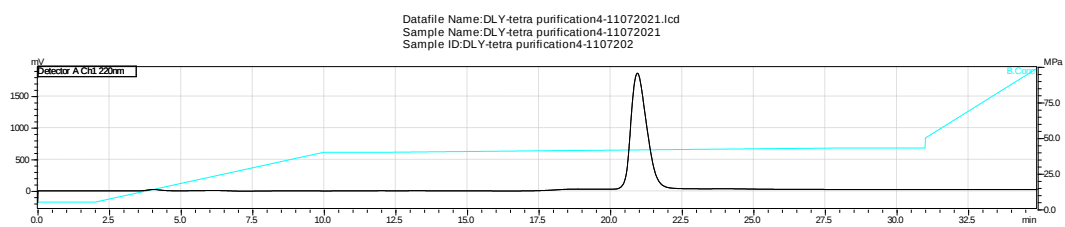


Figure 2.30. HPLC Chromatogram of **8**.

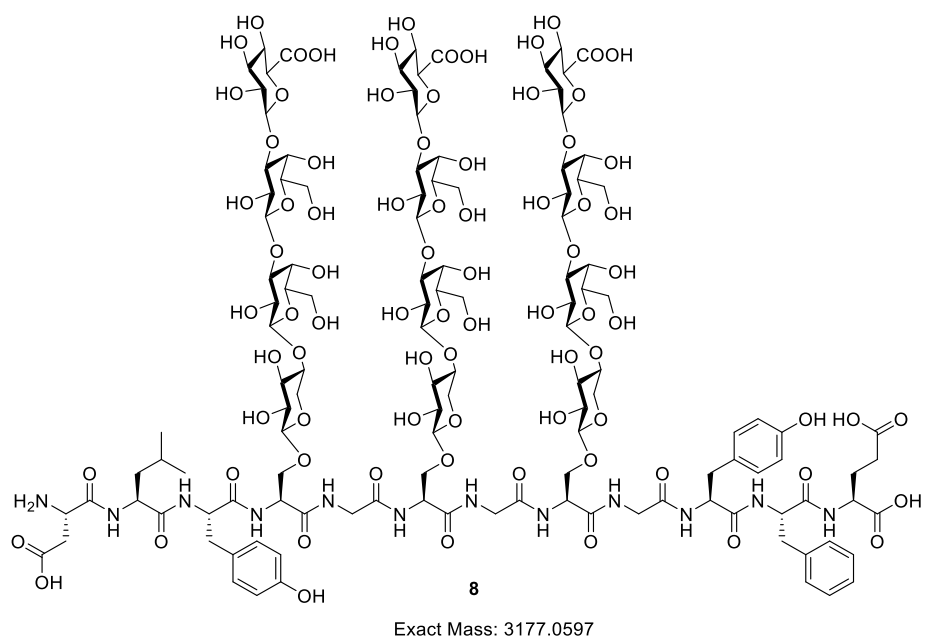
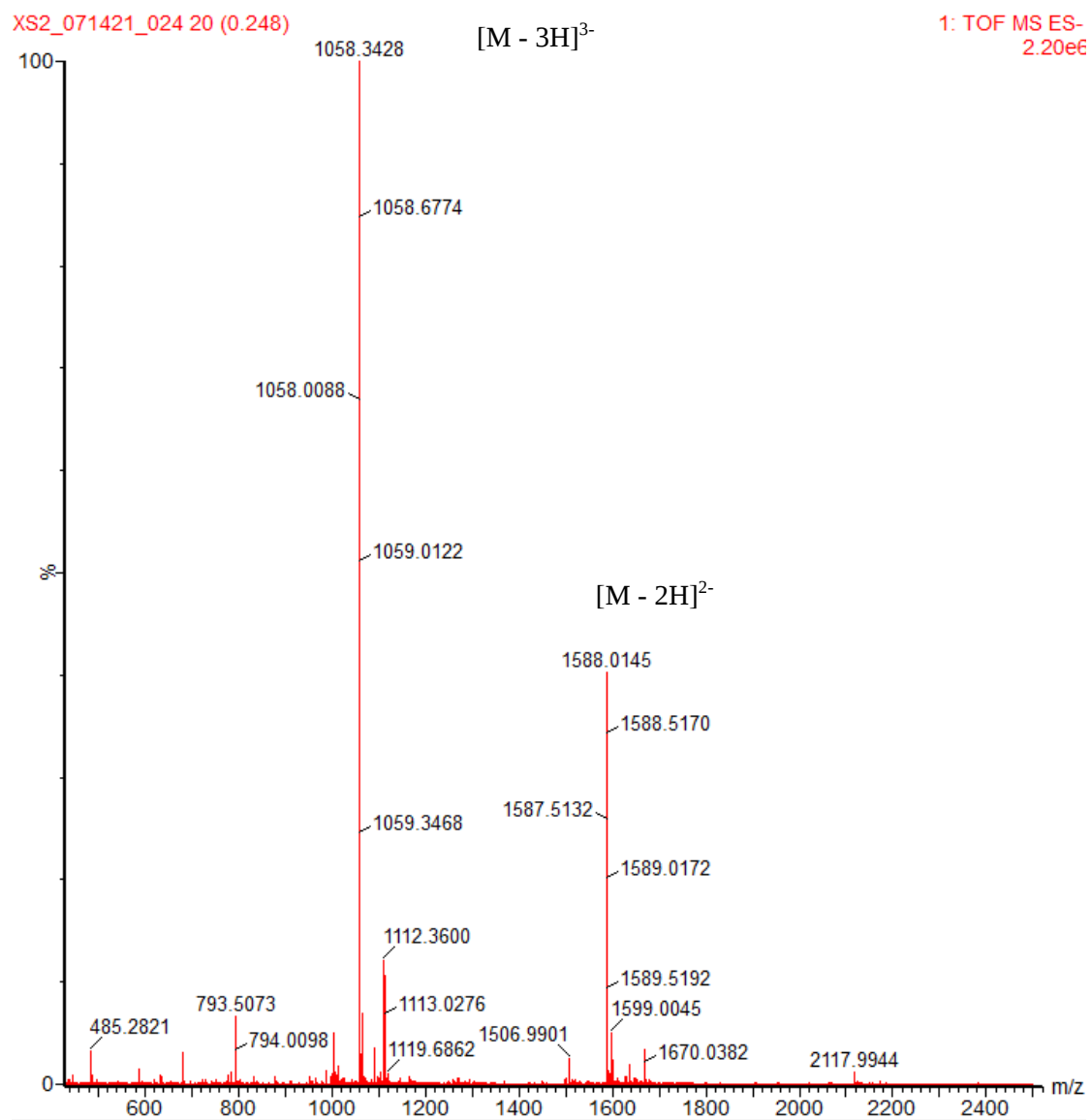


Figure 2.31. LCMS Chromatogram of **8**.

Figure 2.31. (cont'd)



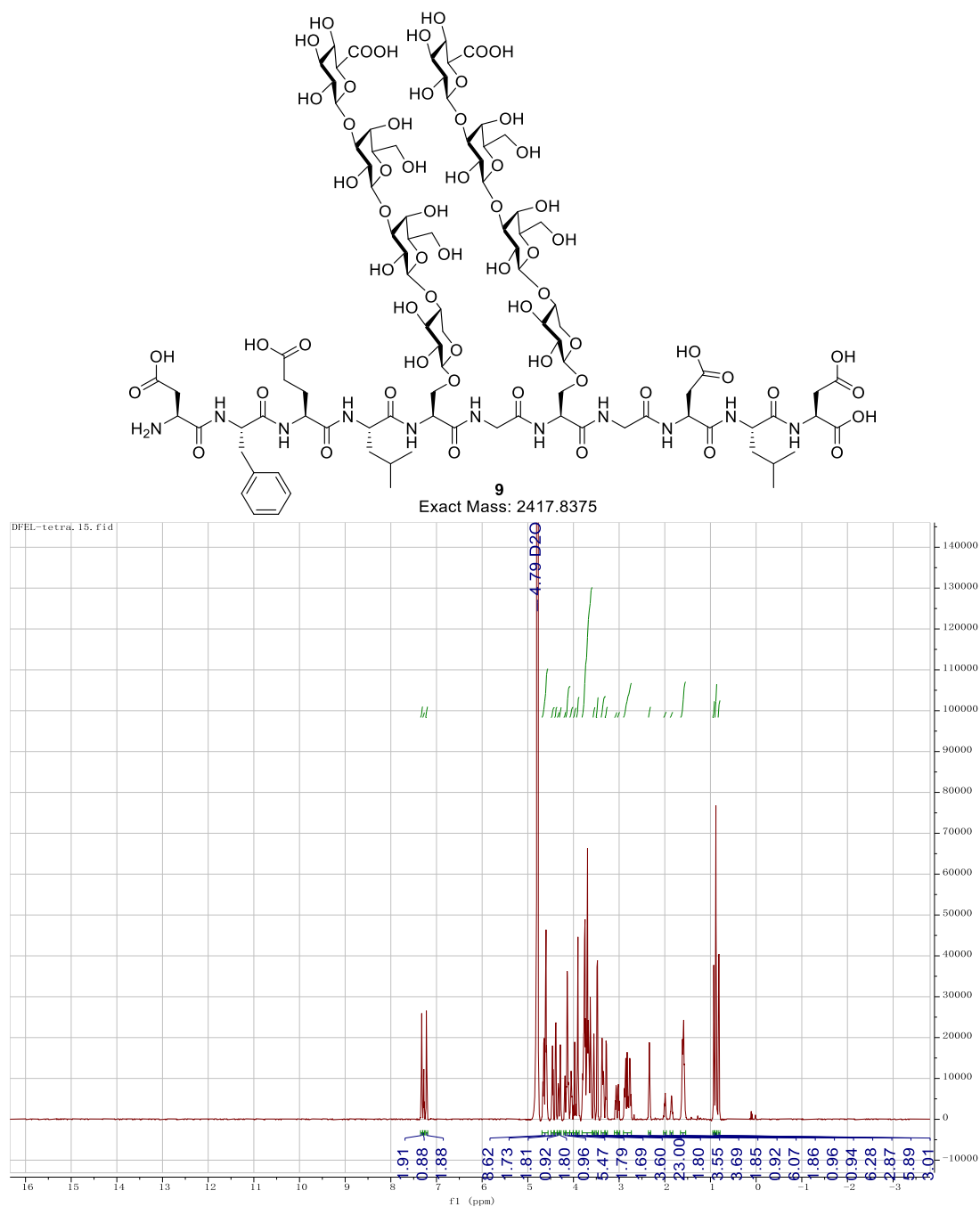


Figure 2.32. ^1H -NMR of **9** (800 MHz, D_2O).

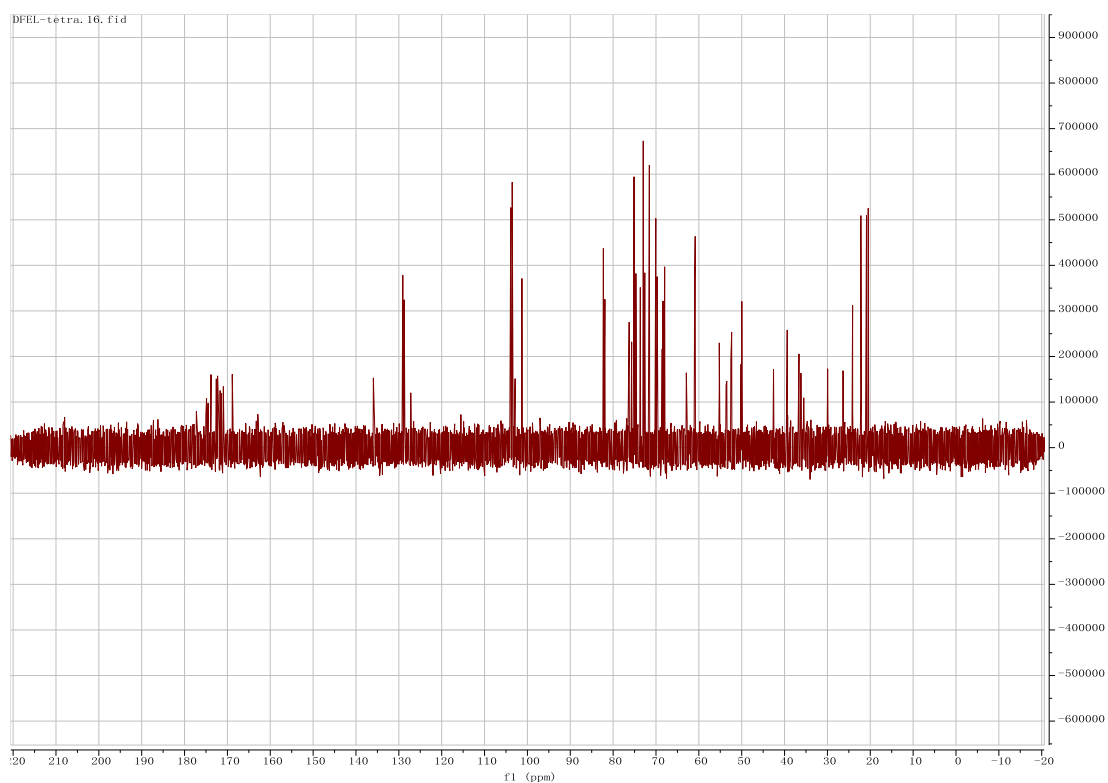
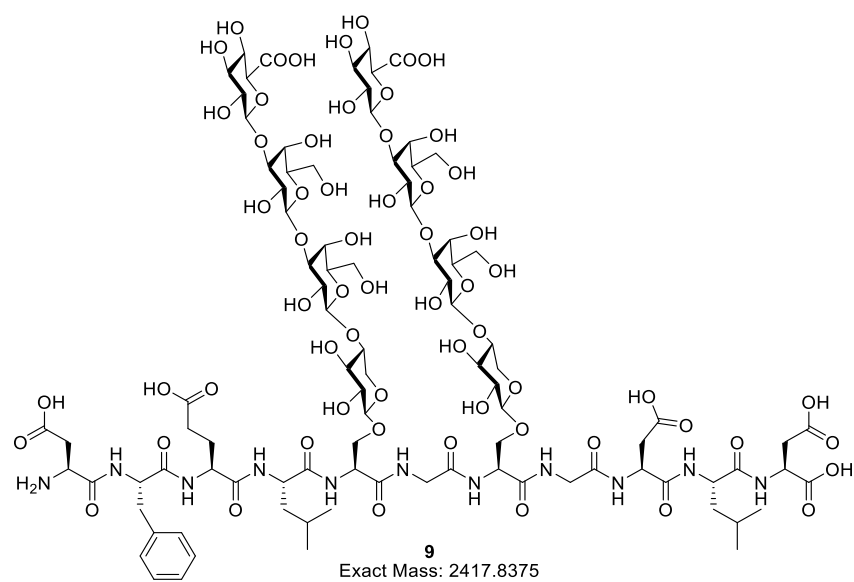


Figure 2.33. ^{13}C NMR of **9** (201 MHz, D_2O).

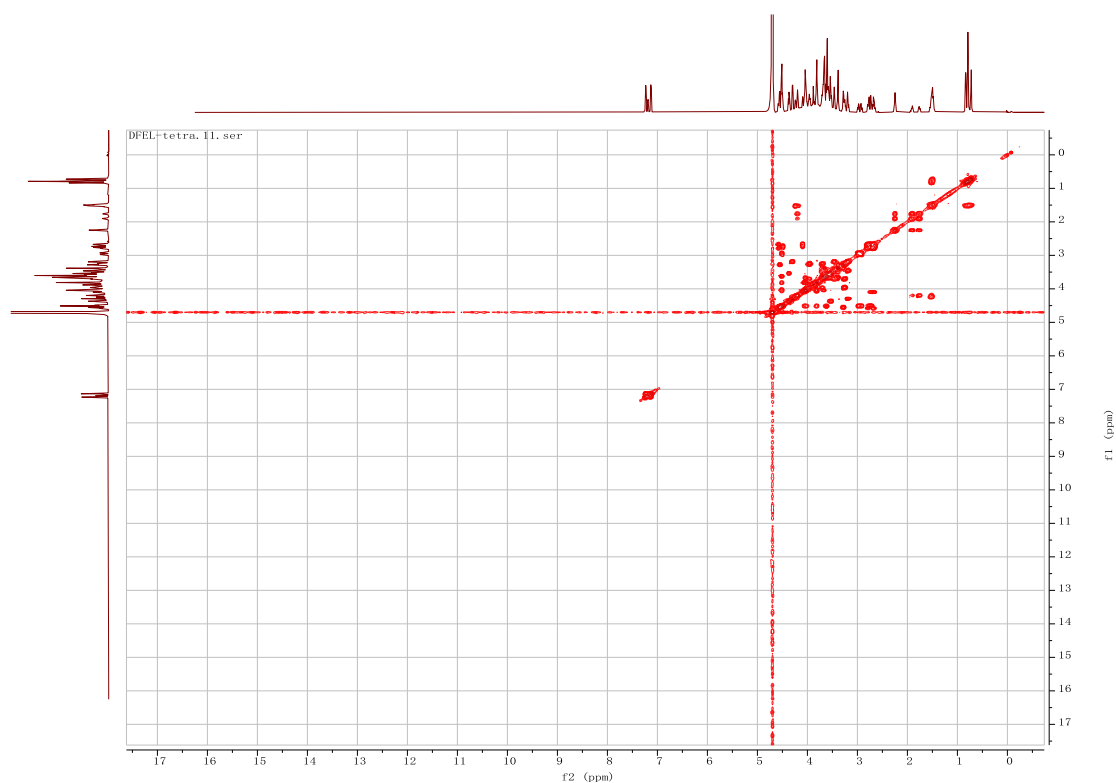
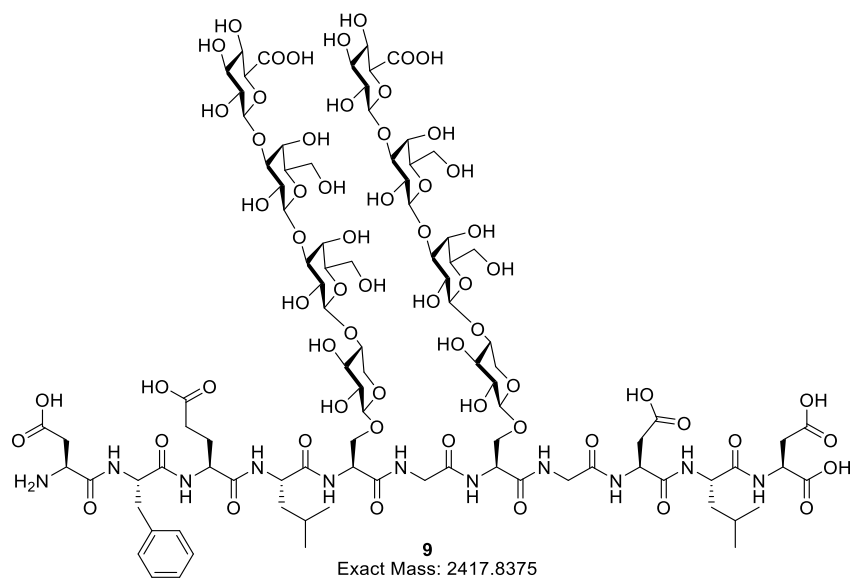


Figure 2.34. COSY of **9** (800 MHz, D₂O).

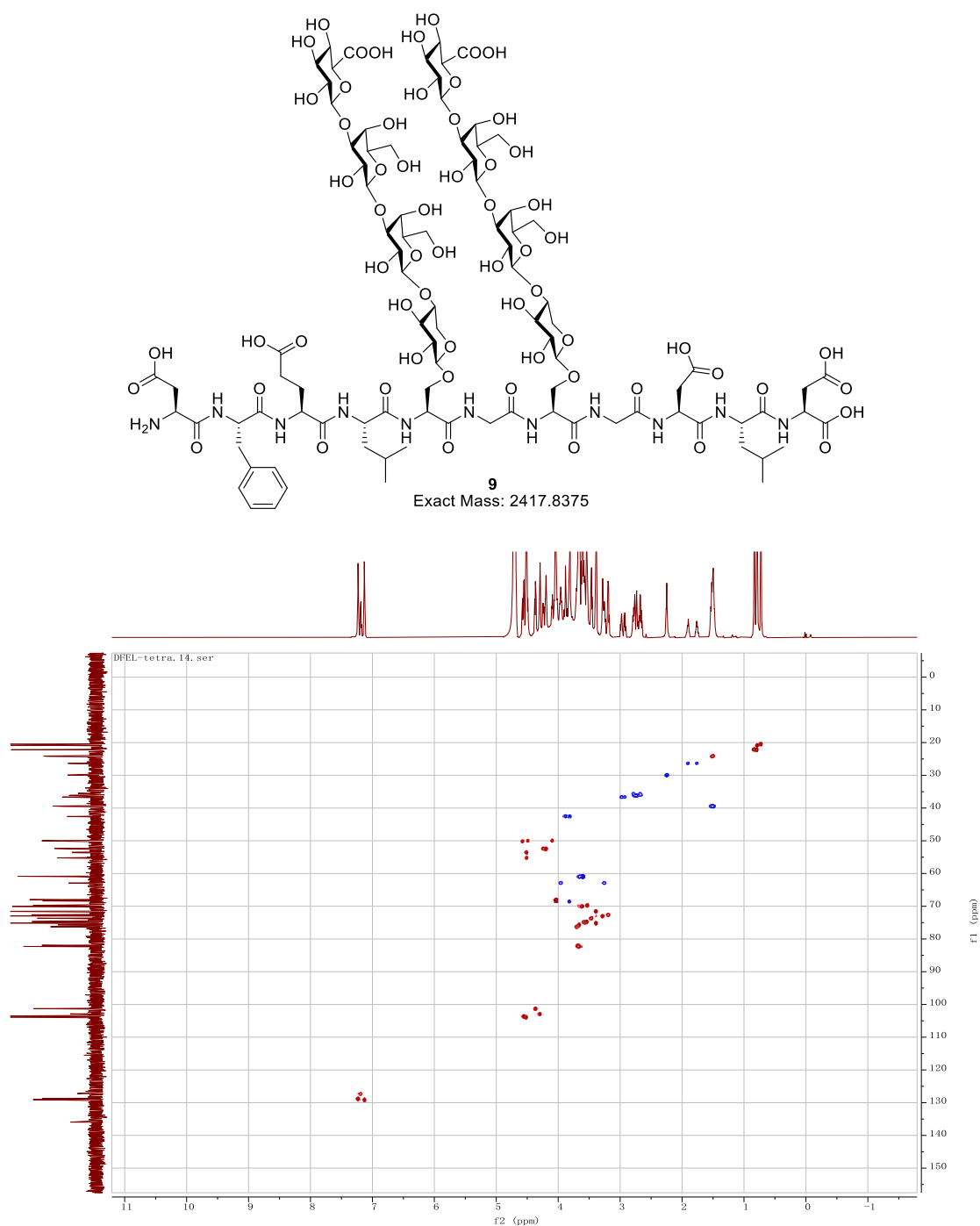


Figure 2.35. HSQC of **9**(800 MHz, D₂O).

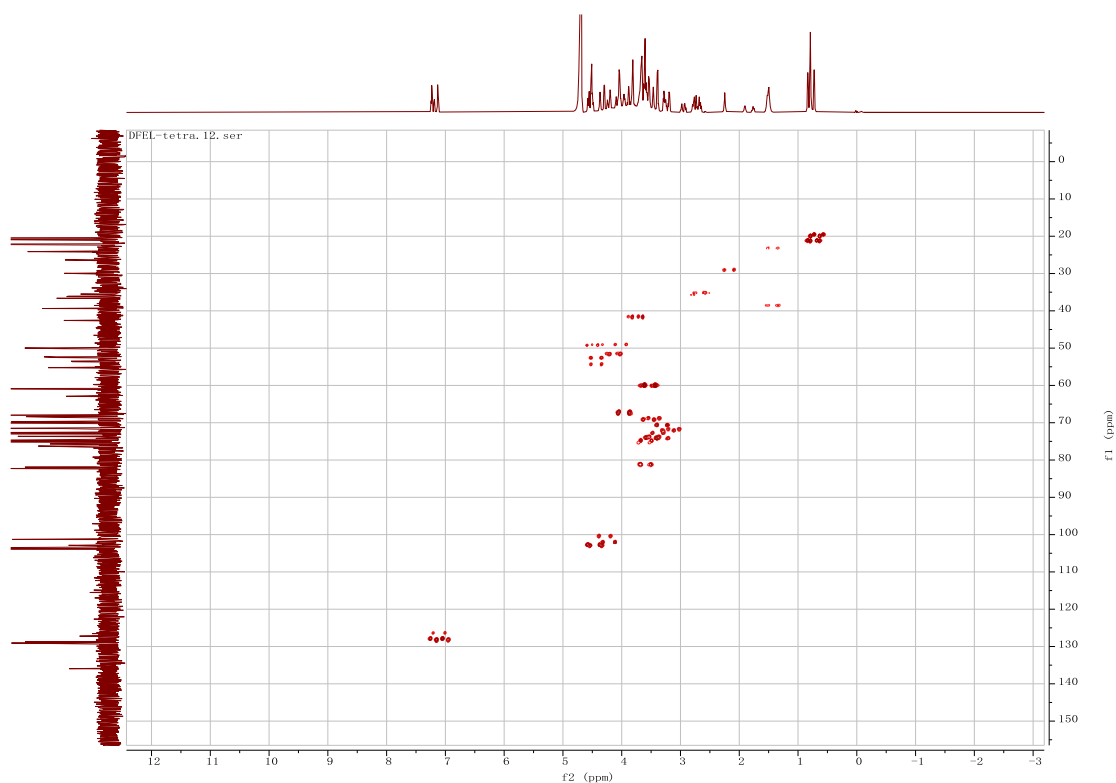
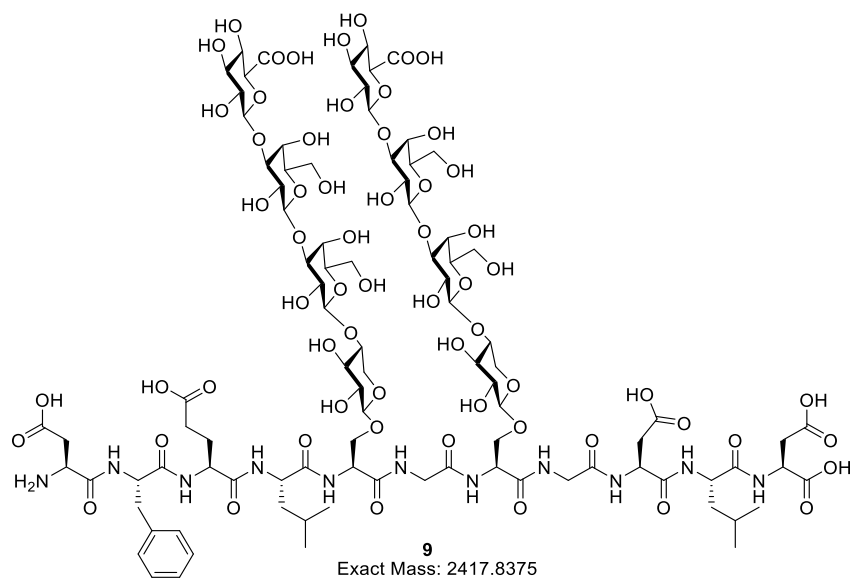


Figure 2.36. Coupled HSQC of **9** (800 MHz, D₂O).

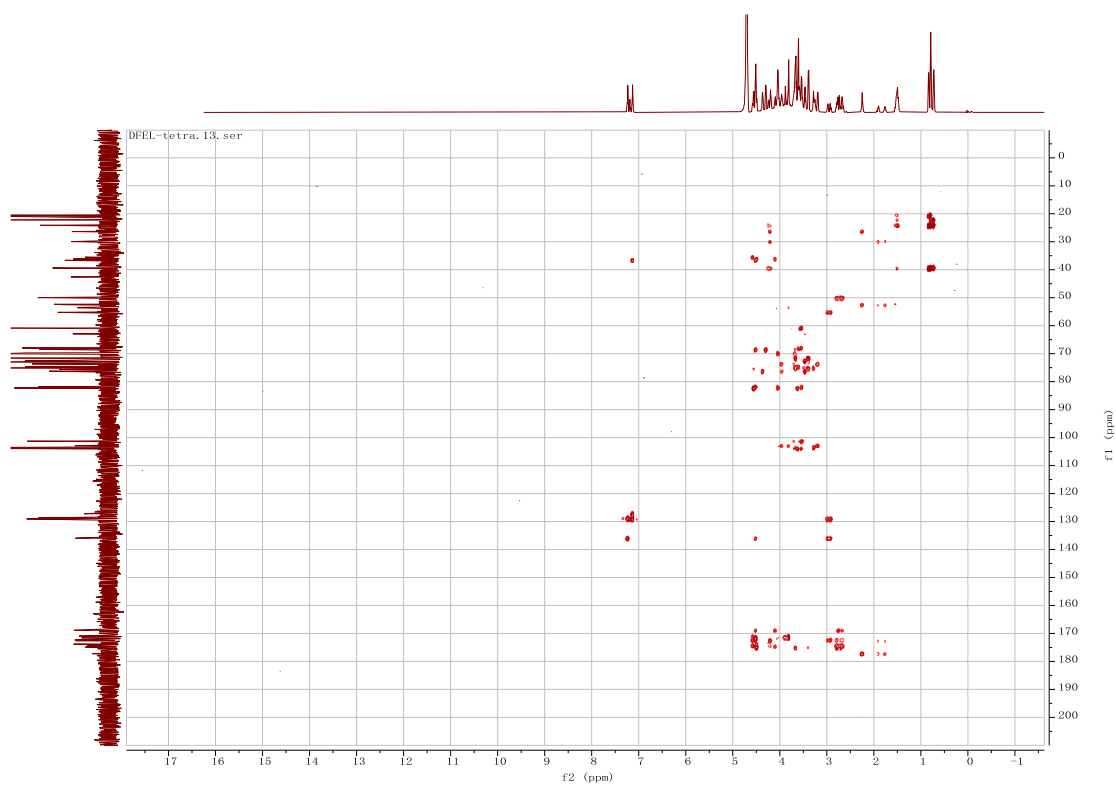
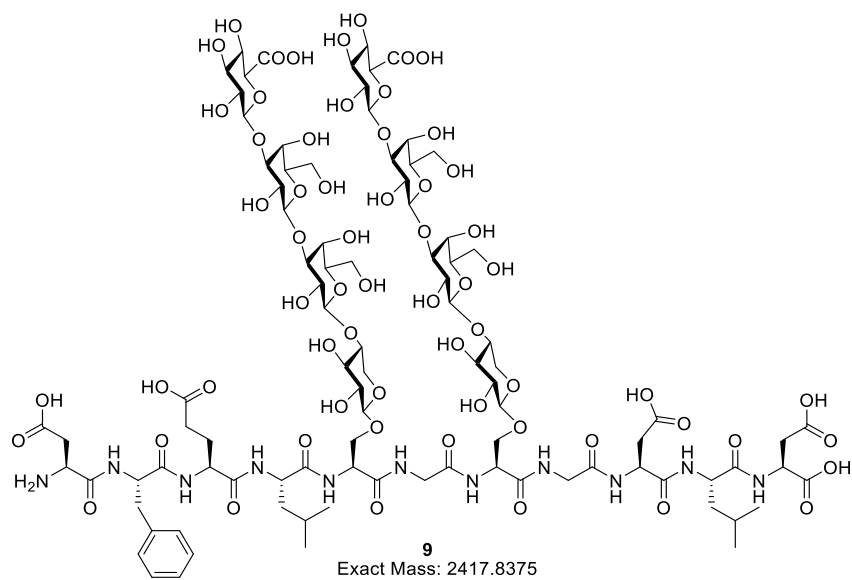


Figure 2.37. HMBC of **9** (800 MHz, D₂O).

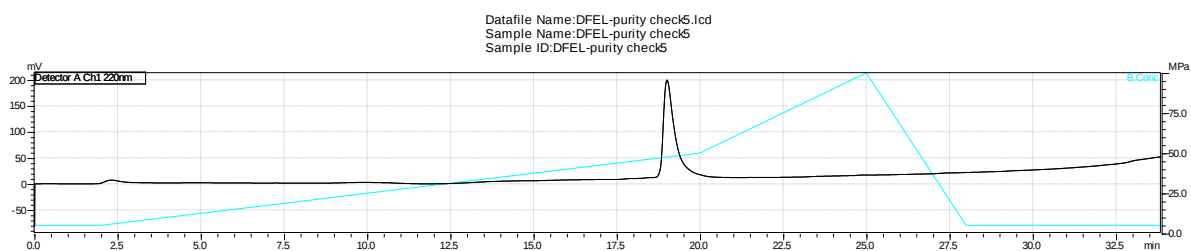
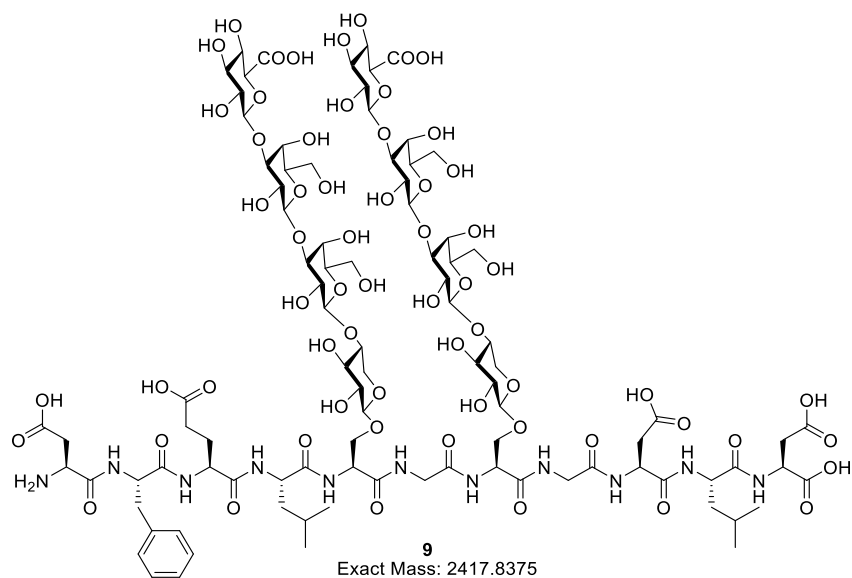


Figure 2.38. HPLC Chromatogram of **9**.

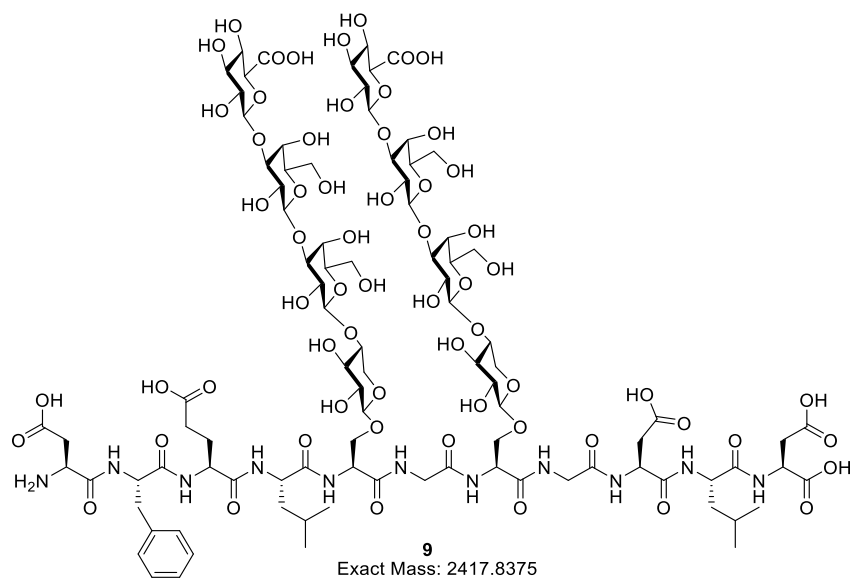
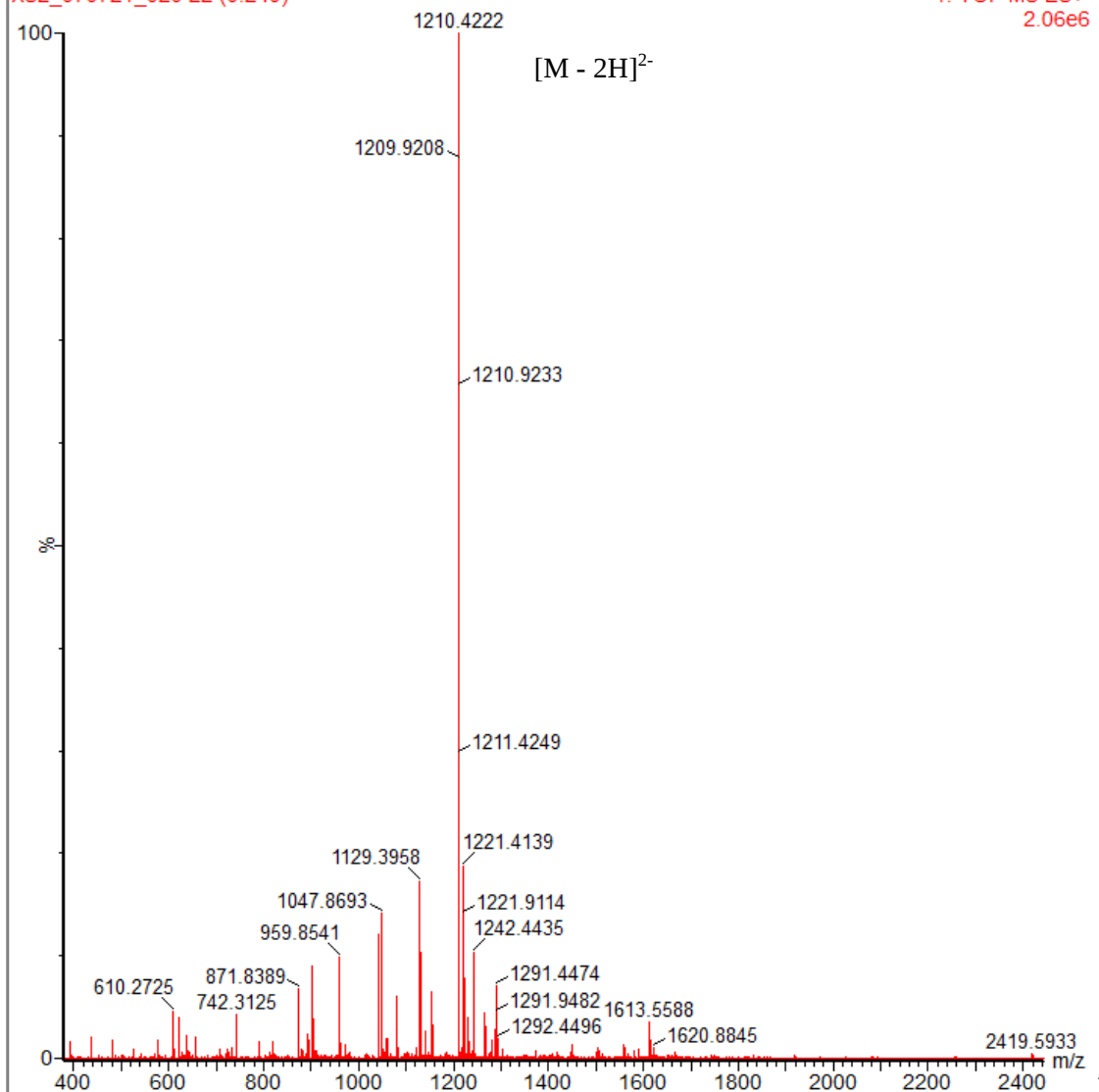


Figure 2.39. LCMS Chromatogram of **9**

Figure 2.39. (cont'd)

XS2_070721_020 22 (0.240)

1: TOF MS ES+
2.06e6



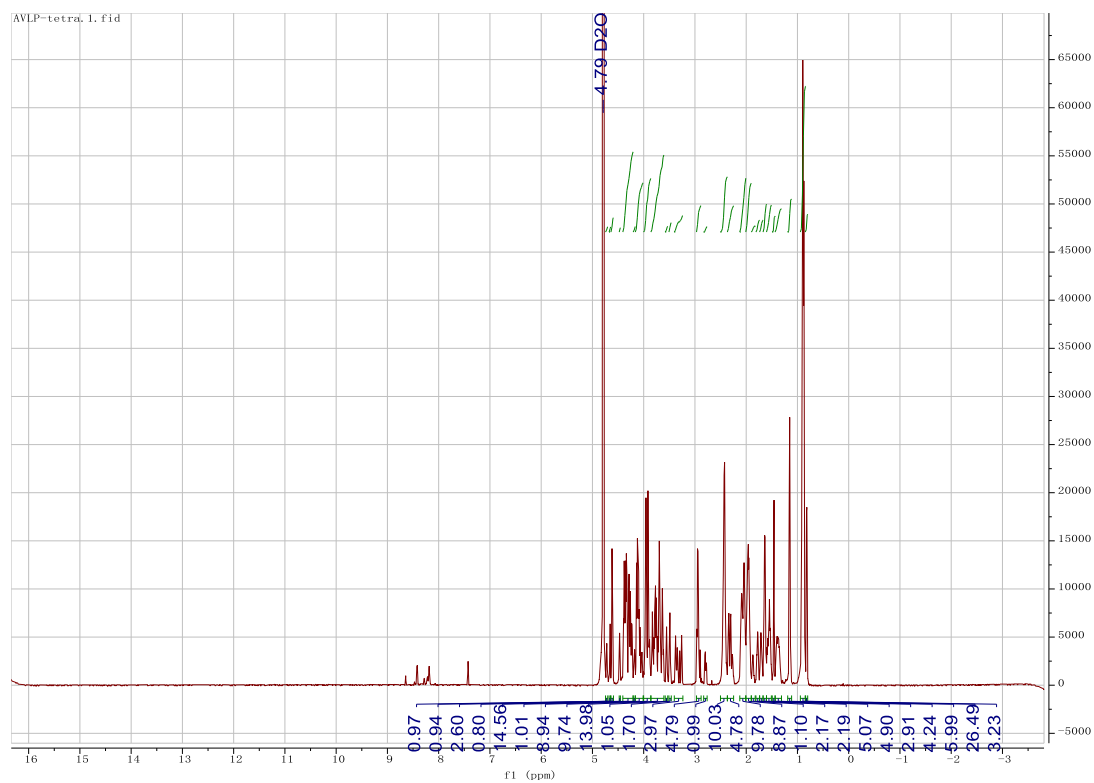
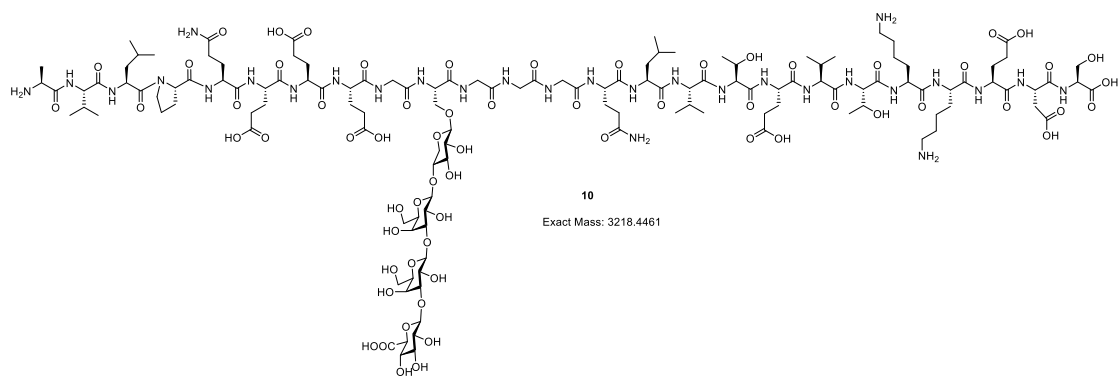


Figure 2.40. ¹H-NMR of **10** (800 MHz, D₂O).

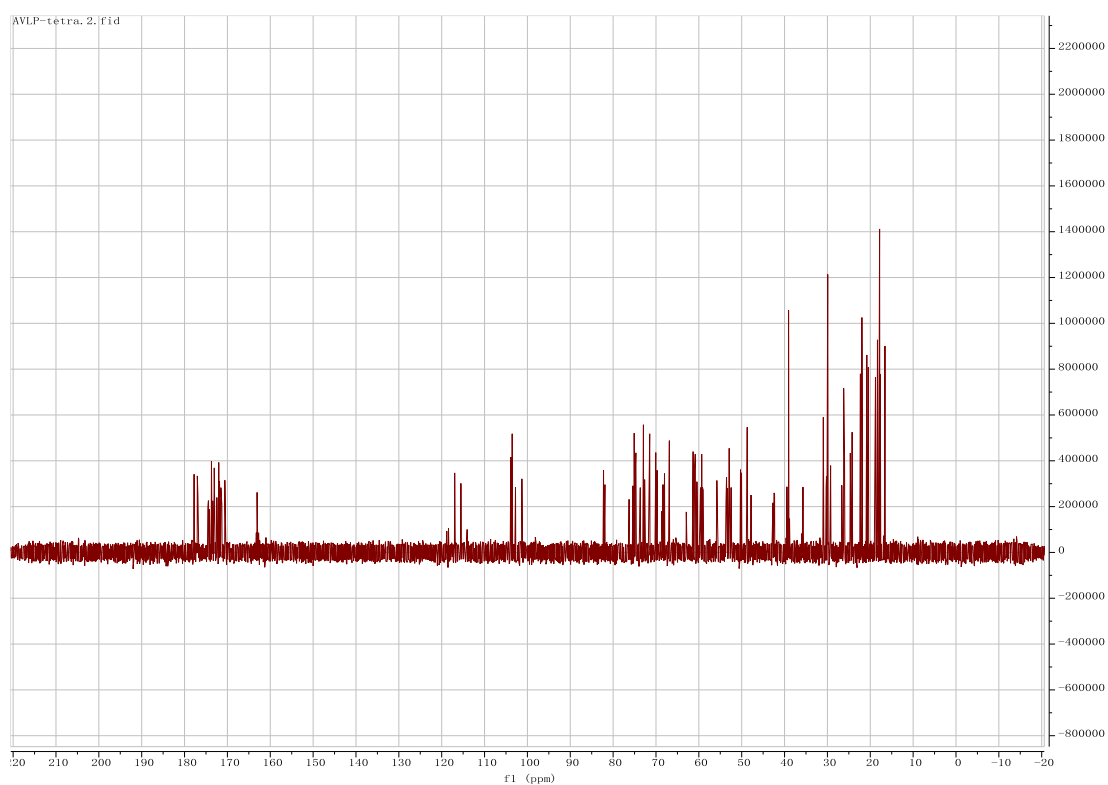
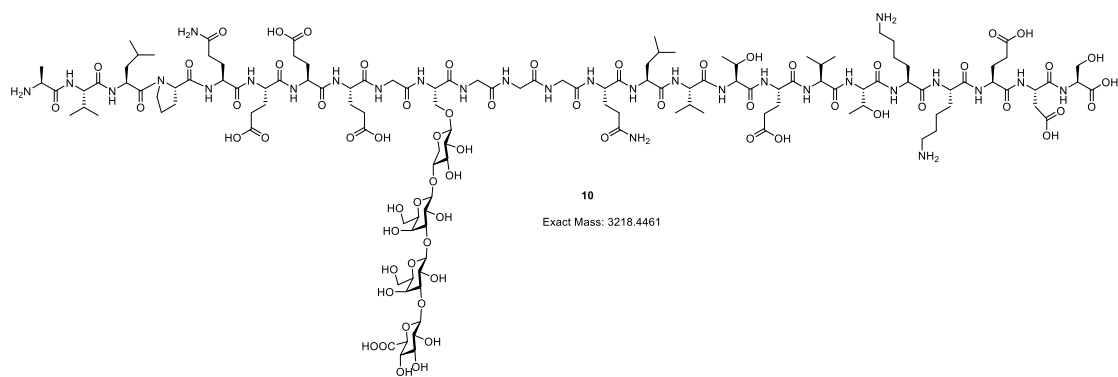


Figure 2.41. ¹³C NMR of **10** (201 MHz, D₂O).

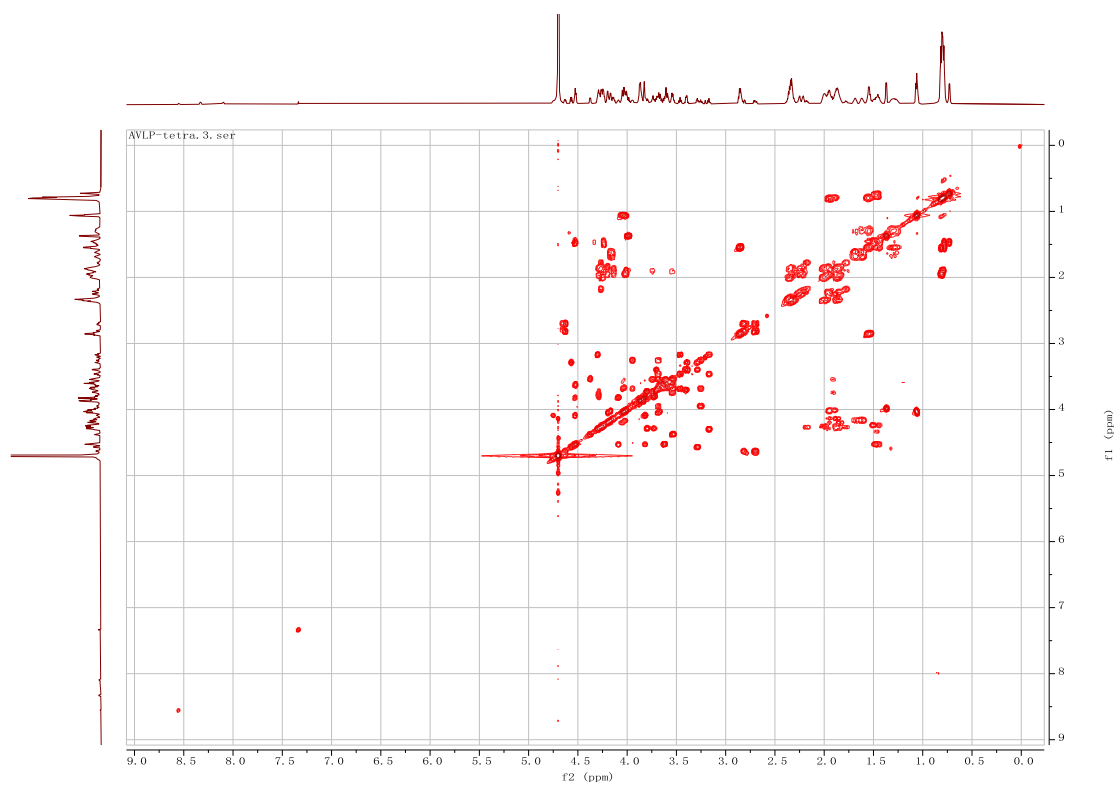
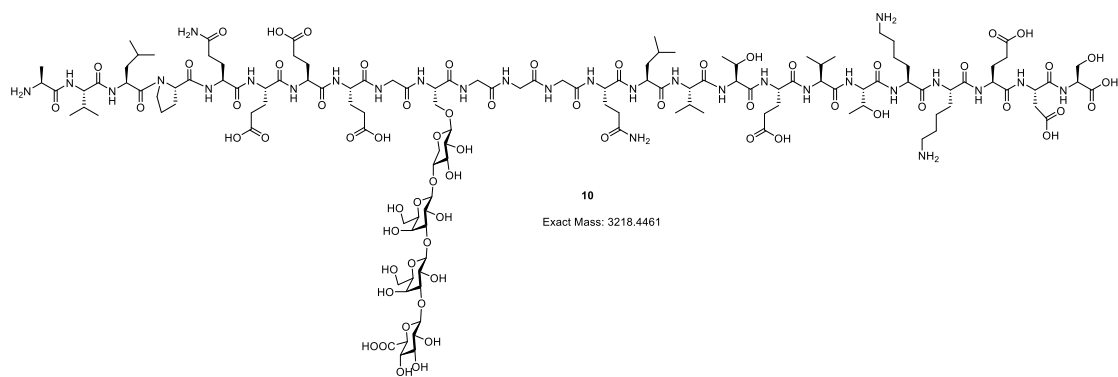


Figure 2.42. COSY of **10** (800 MHz, D₂O).

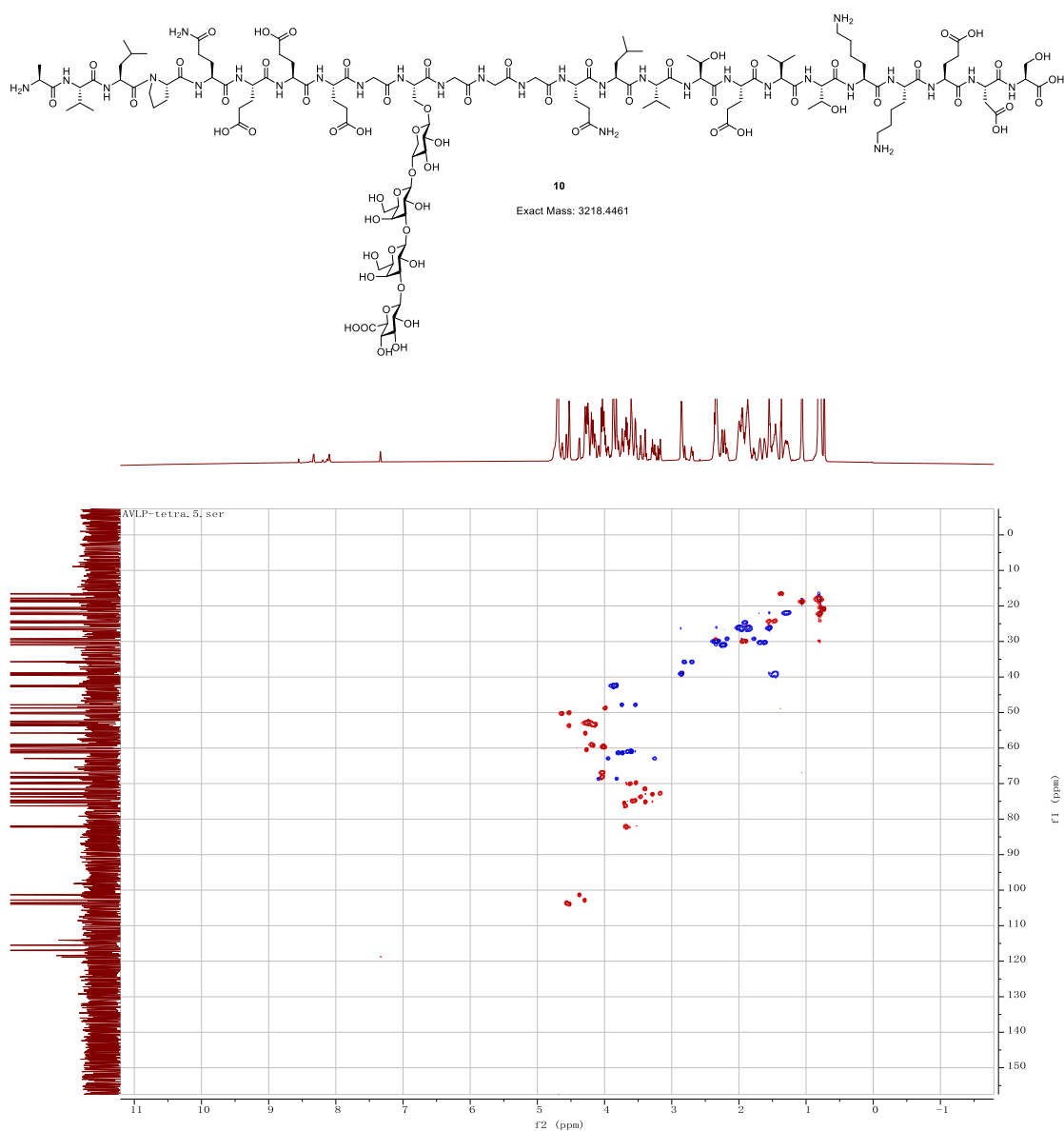


Figure 2.43. HSQC of **10** (800 MHz, D₂O).

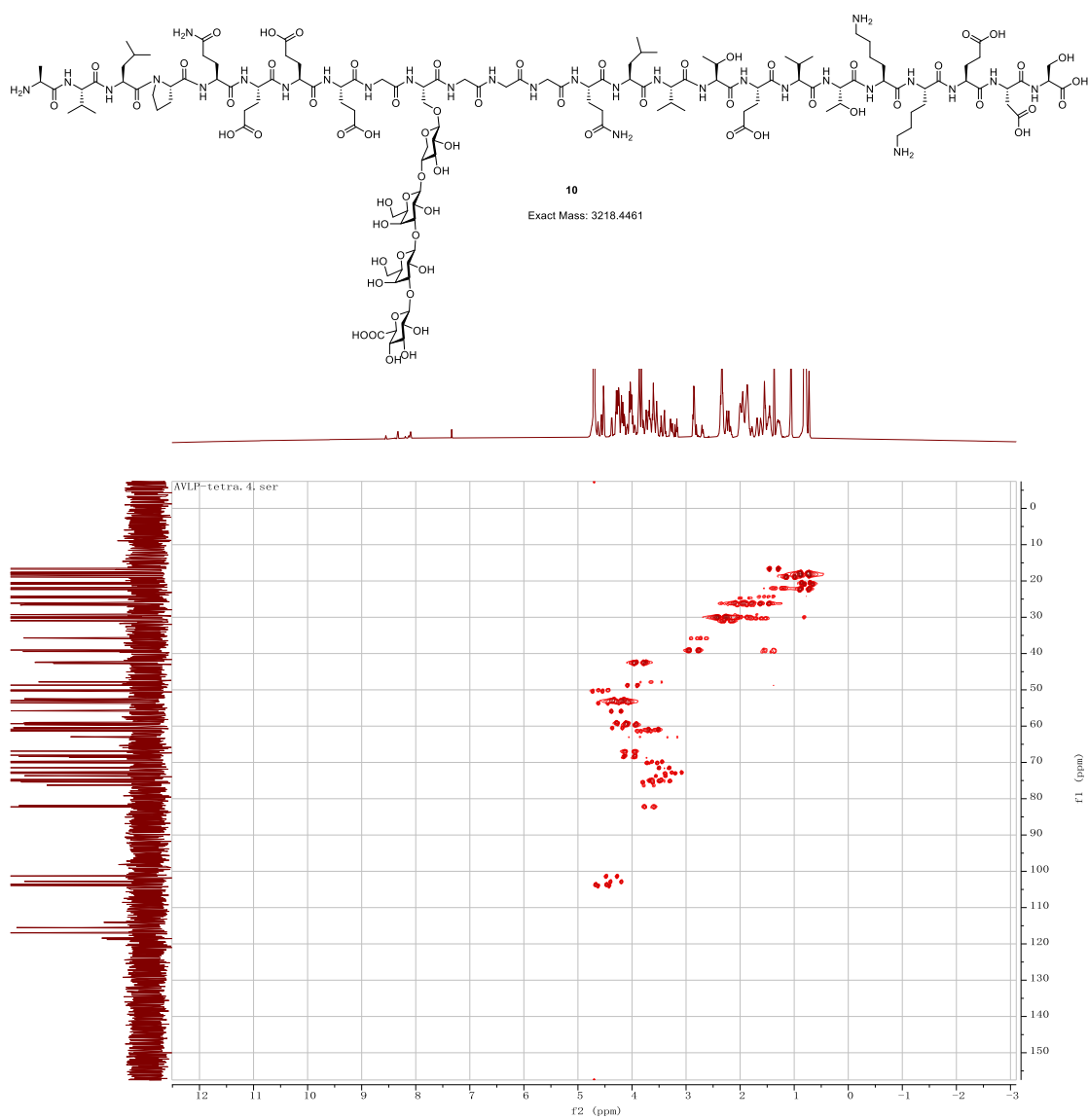


Figure 2.44. Coupled HSQC of **10** (800 MHz, D₂O).

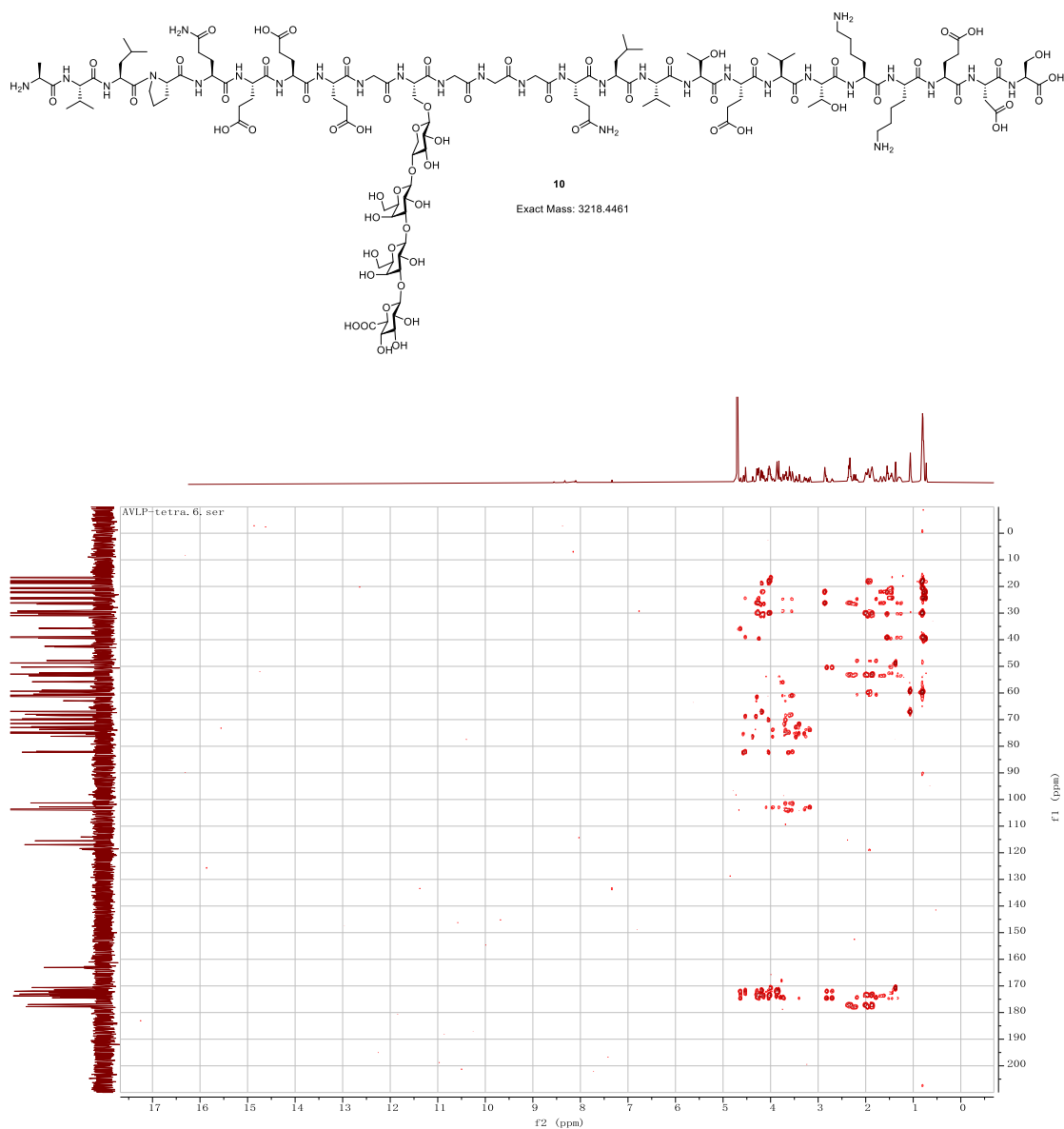


Figure 2.45. HMBC of **10** (800 MHz, D₂O).

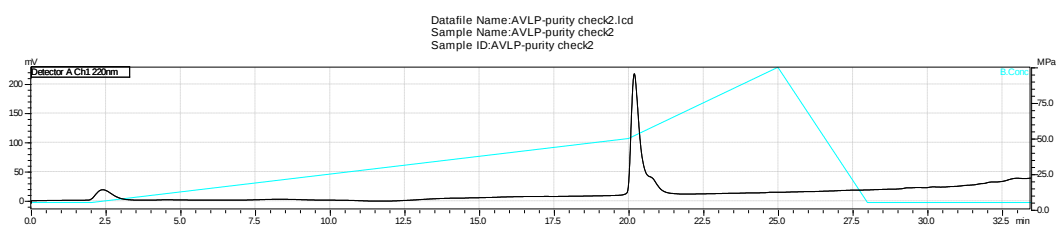
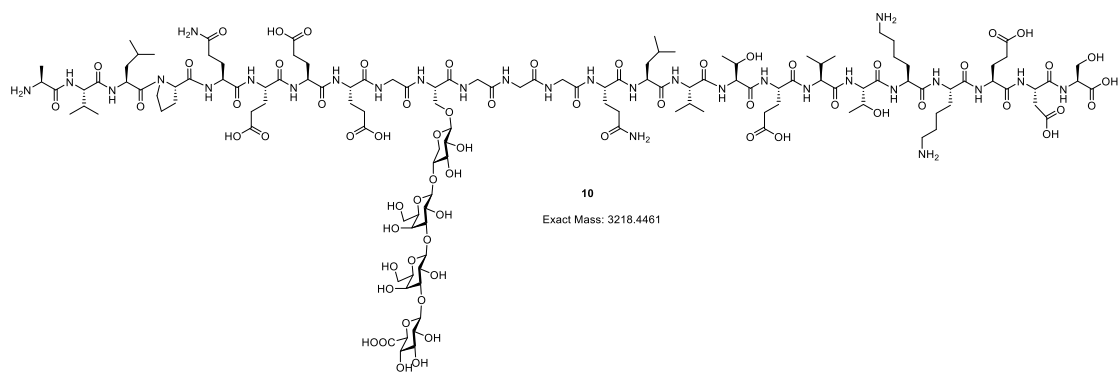
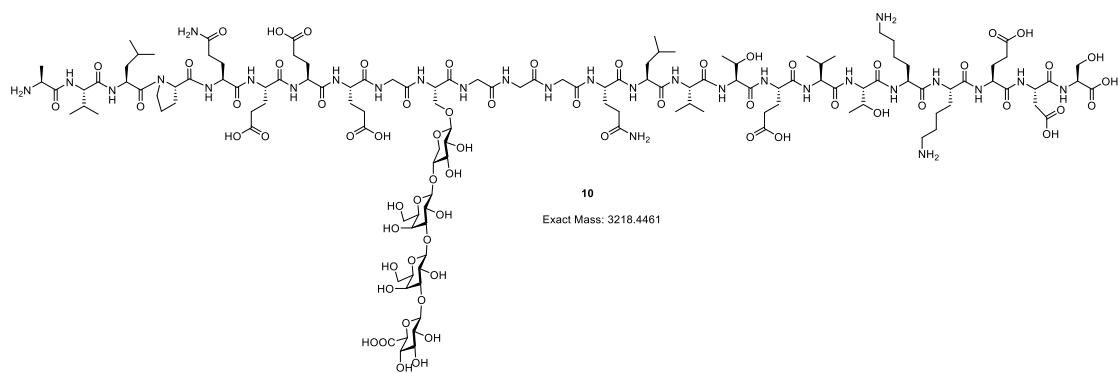


Figure 2.46. HPLC Chromatogram of **10**.



AVLP-C18-96

XS2_111621_046 29 (0.311)

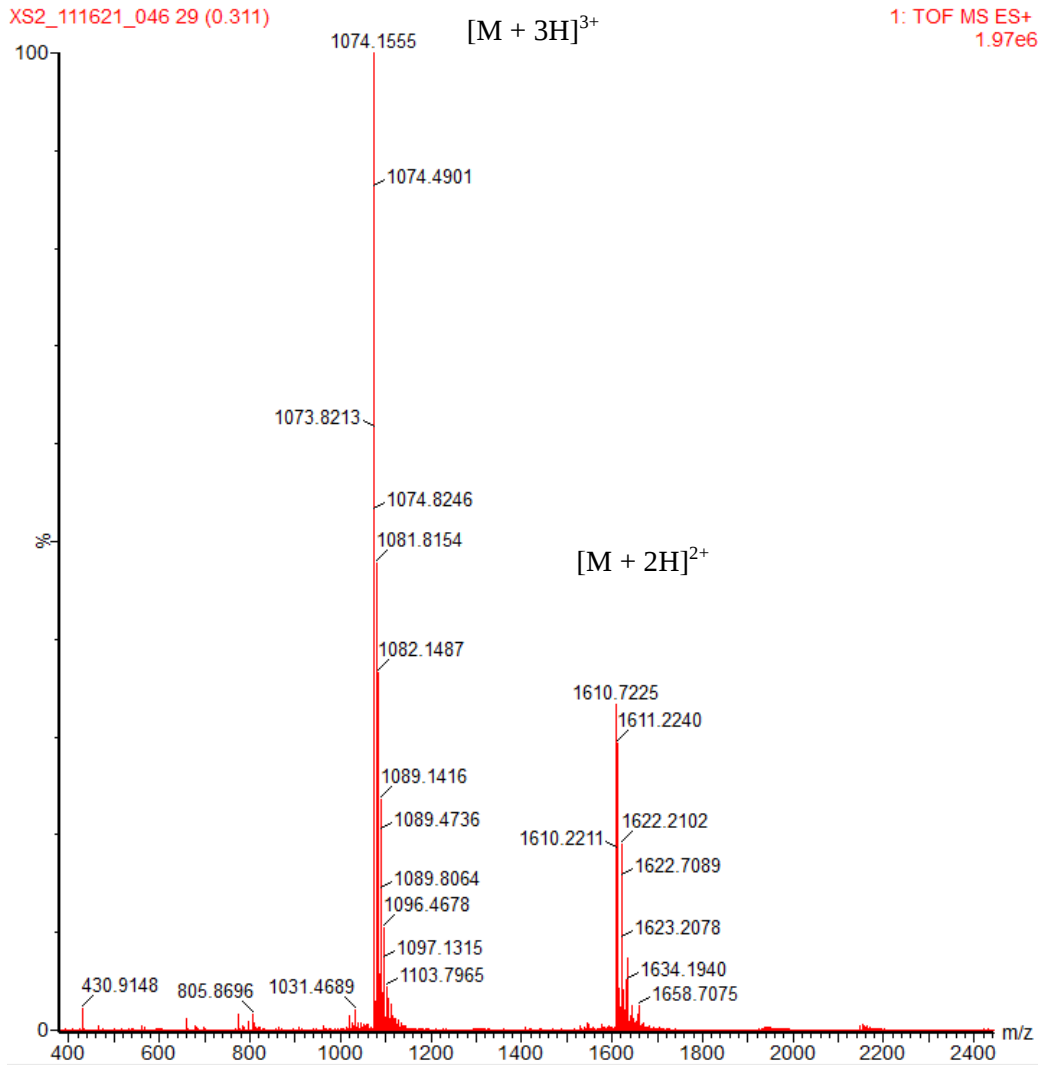


Figure 2.47. LCMS Chromatogram of **10**.

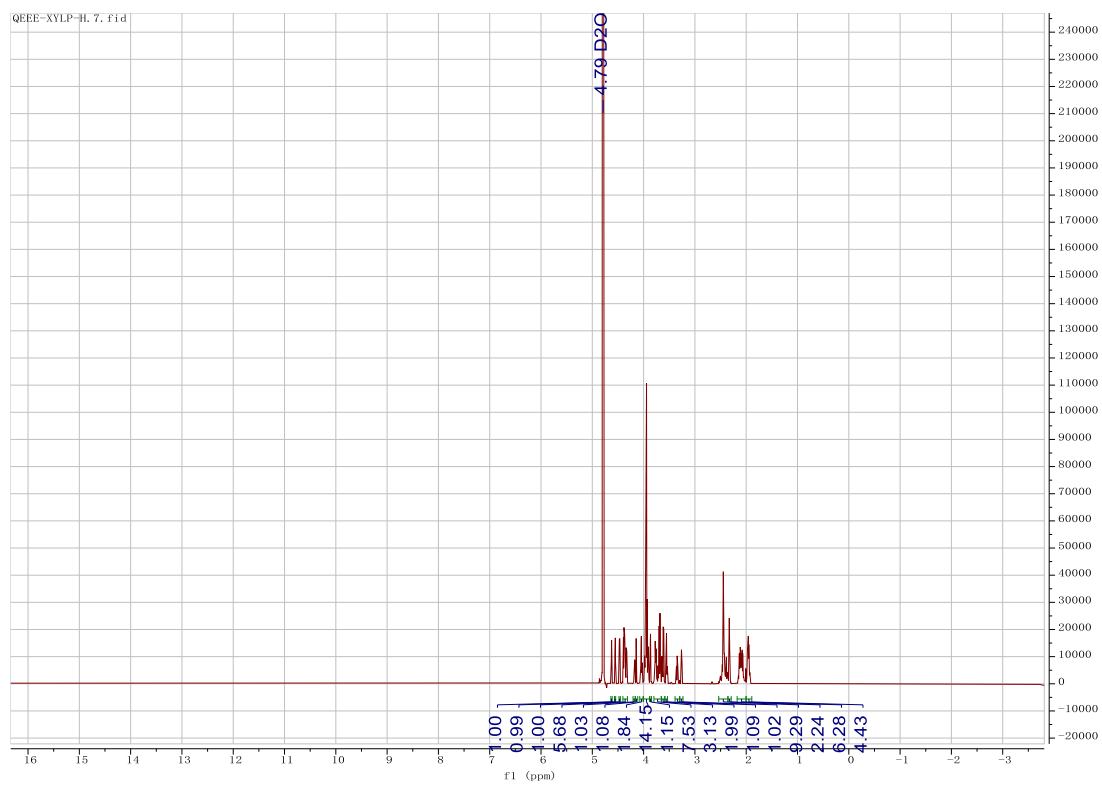
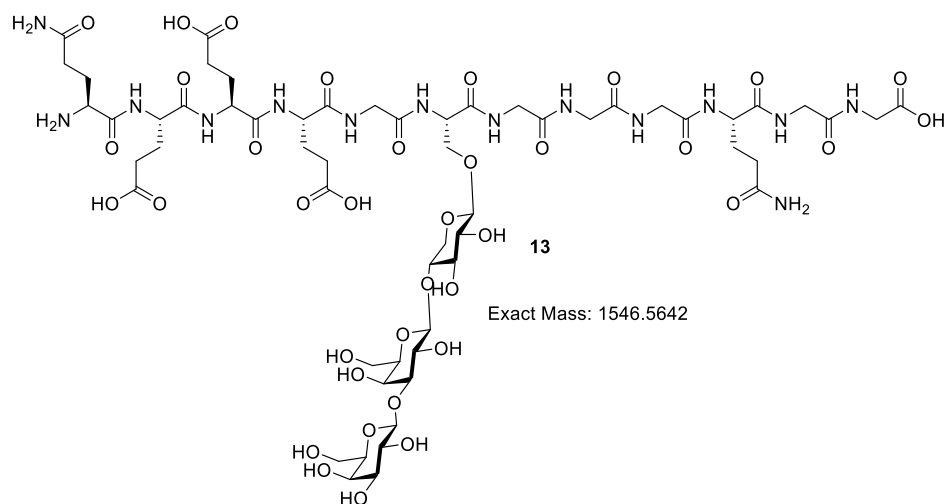


Figure 2.48. ¹H-NMR of **13** (800 MHz, D₂O).

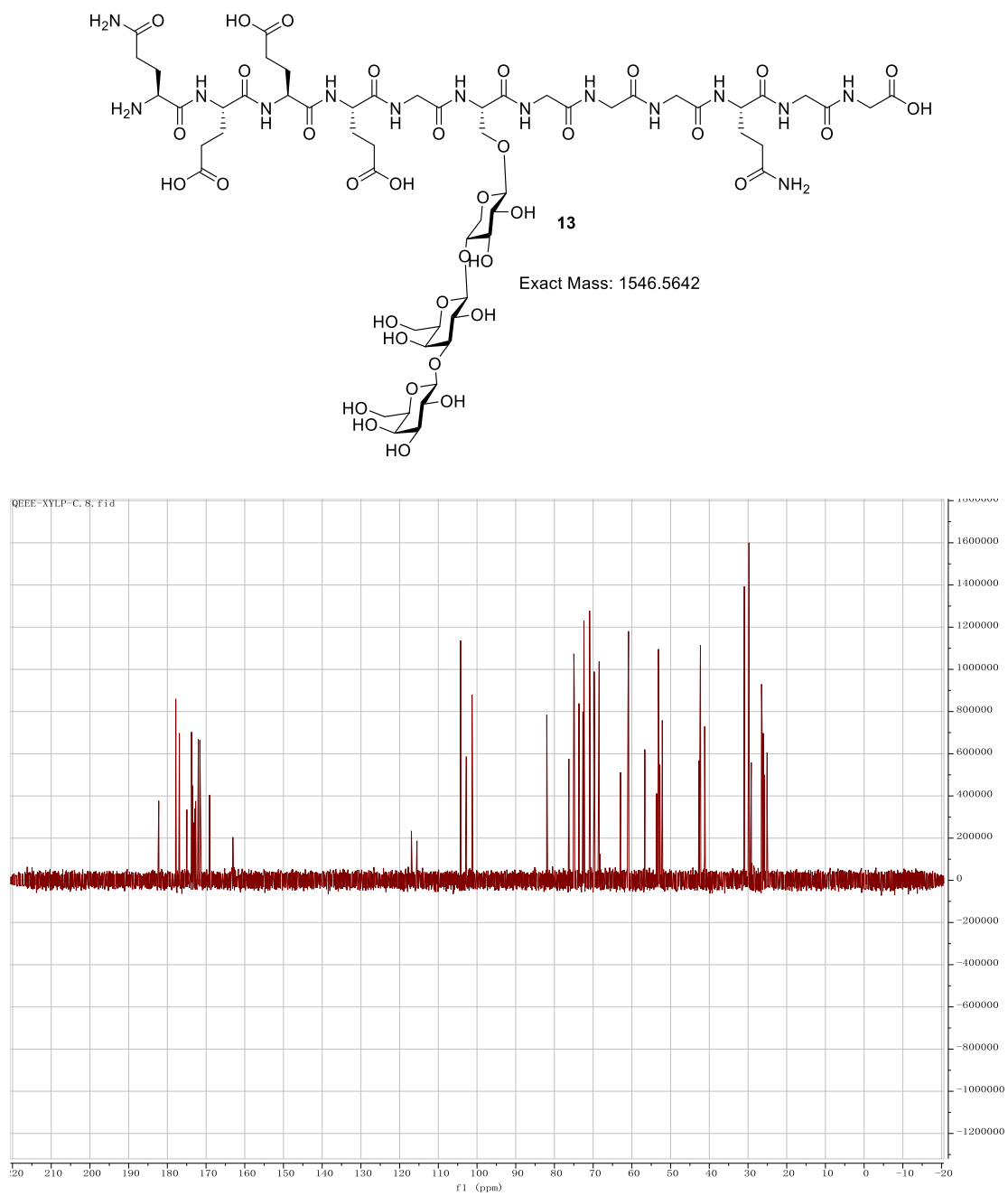


Figure 2.49. ^{13}C NMR of **13** (201 MHz, D_2O).

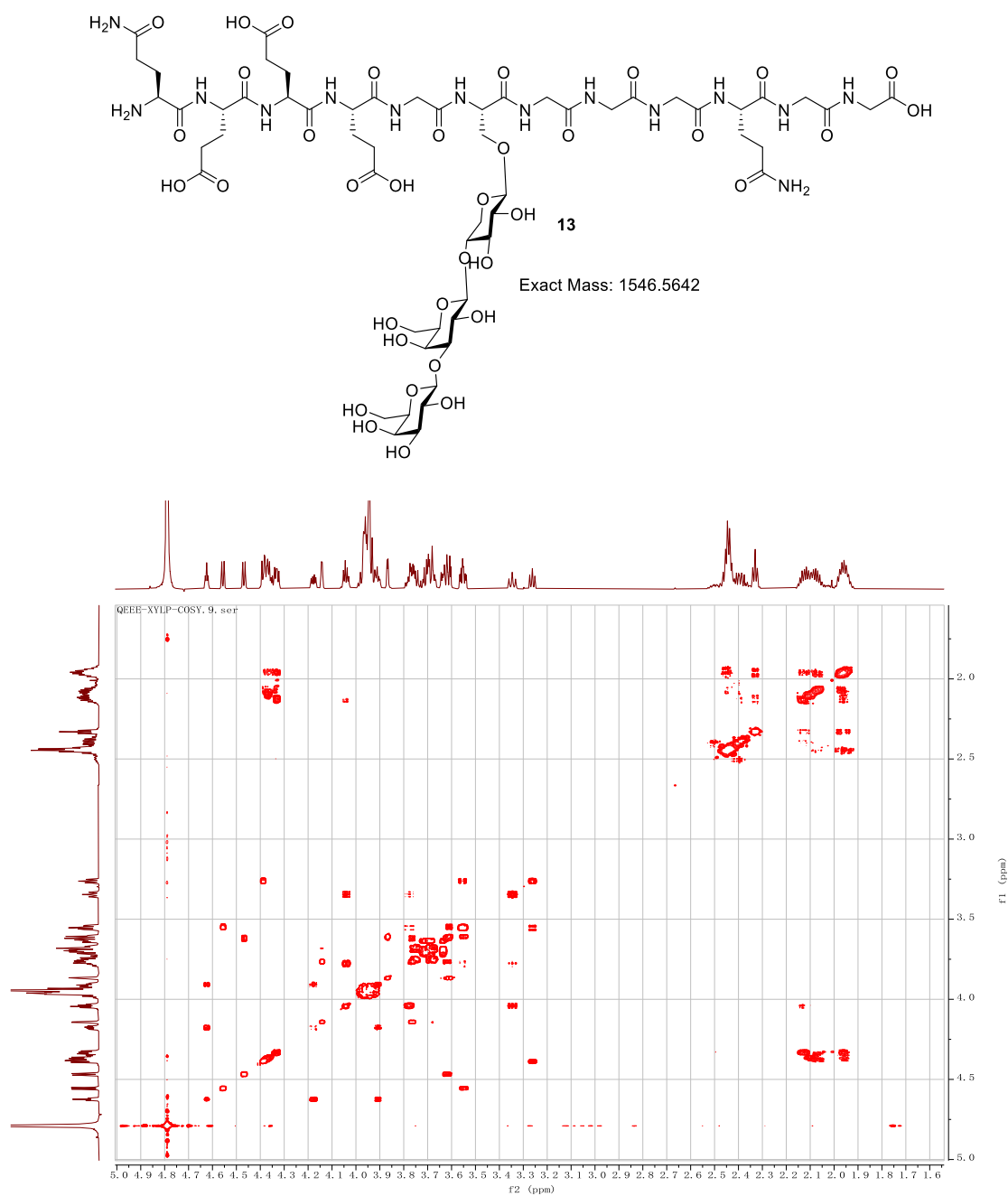


Figure 2.50. COSY of **13** (800 MHz, D₂O).

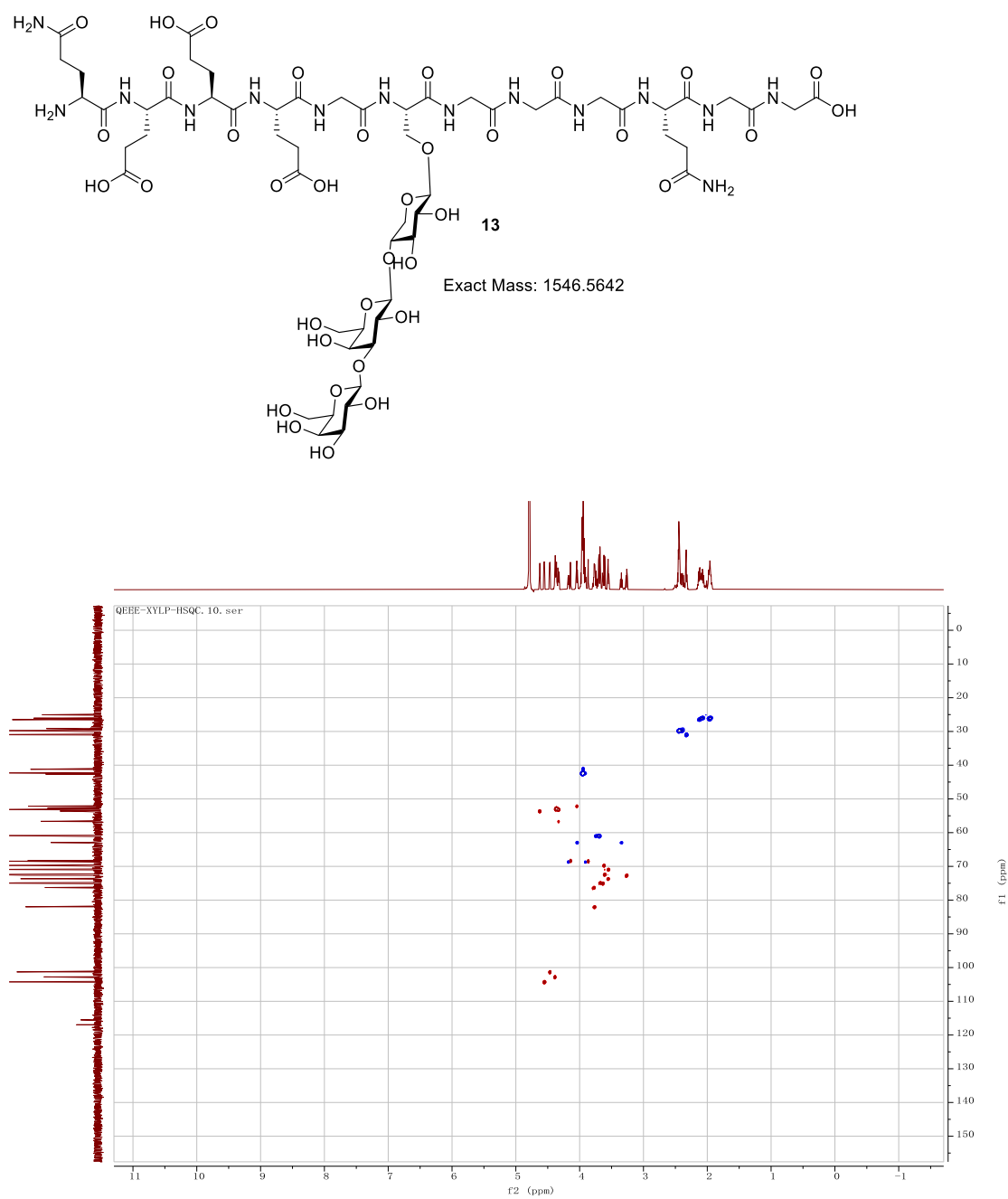


Figure 2.51. HSQC of **13** (800 MHz, D₂O).

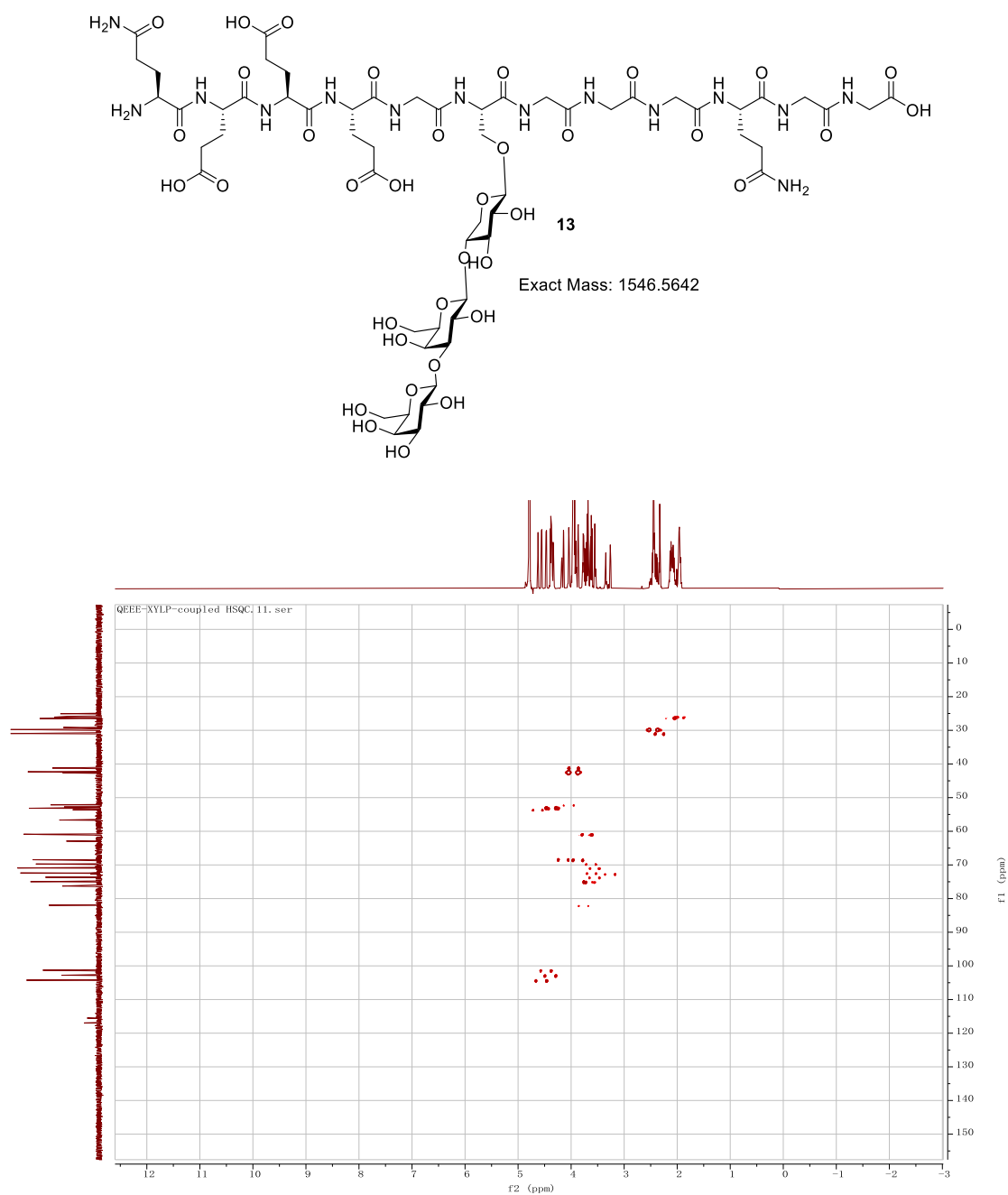


Figure 2.52. Coupled HSQC of **13** (800 MHz, D_2O).

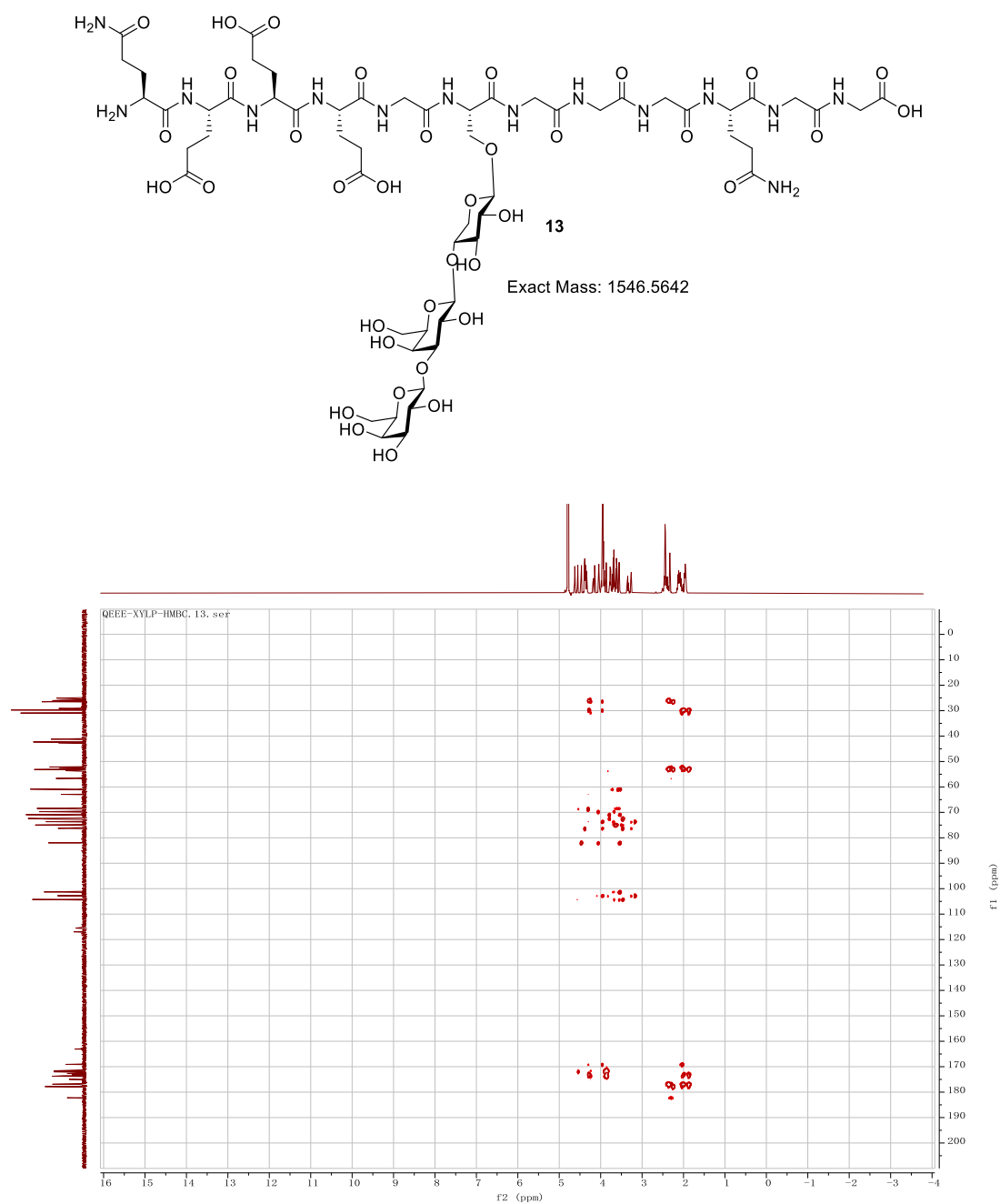


Figure 2.53. HMBC of **13** (800 MHz, D₂O).

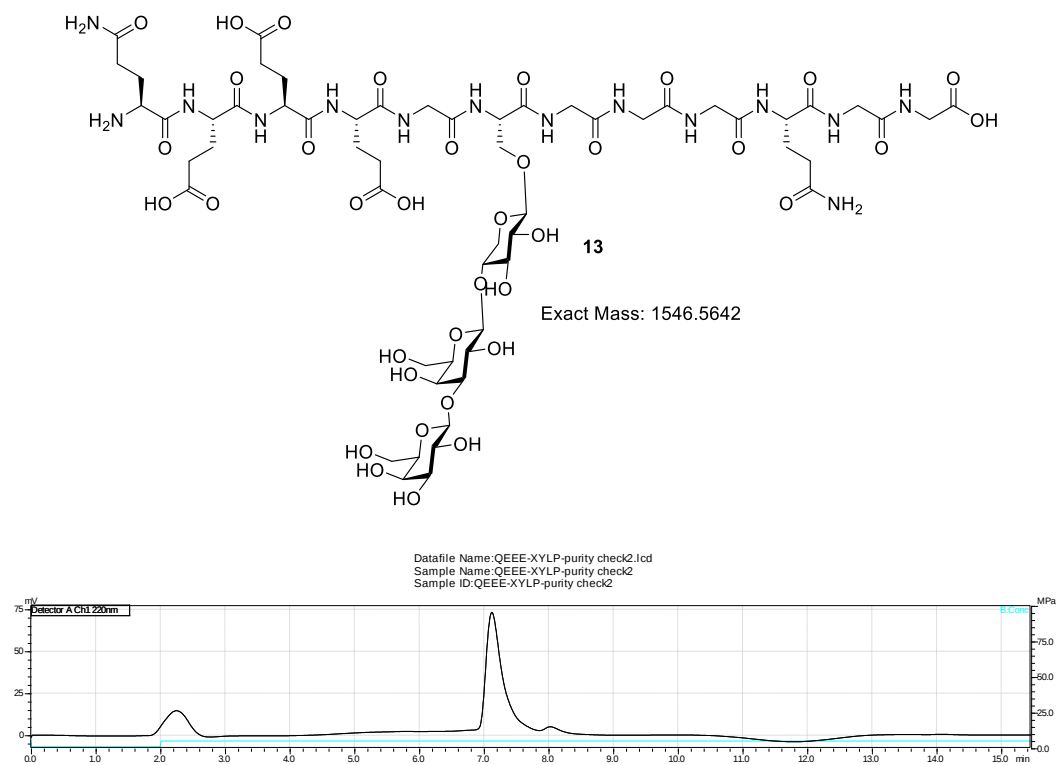


Figure 2.54. HPLC Chromatogram of **13**.

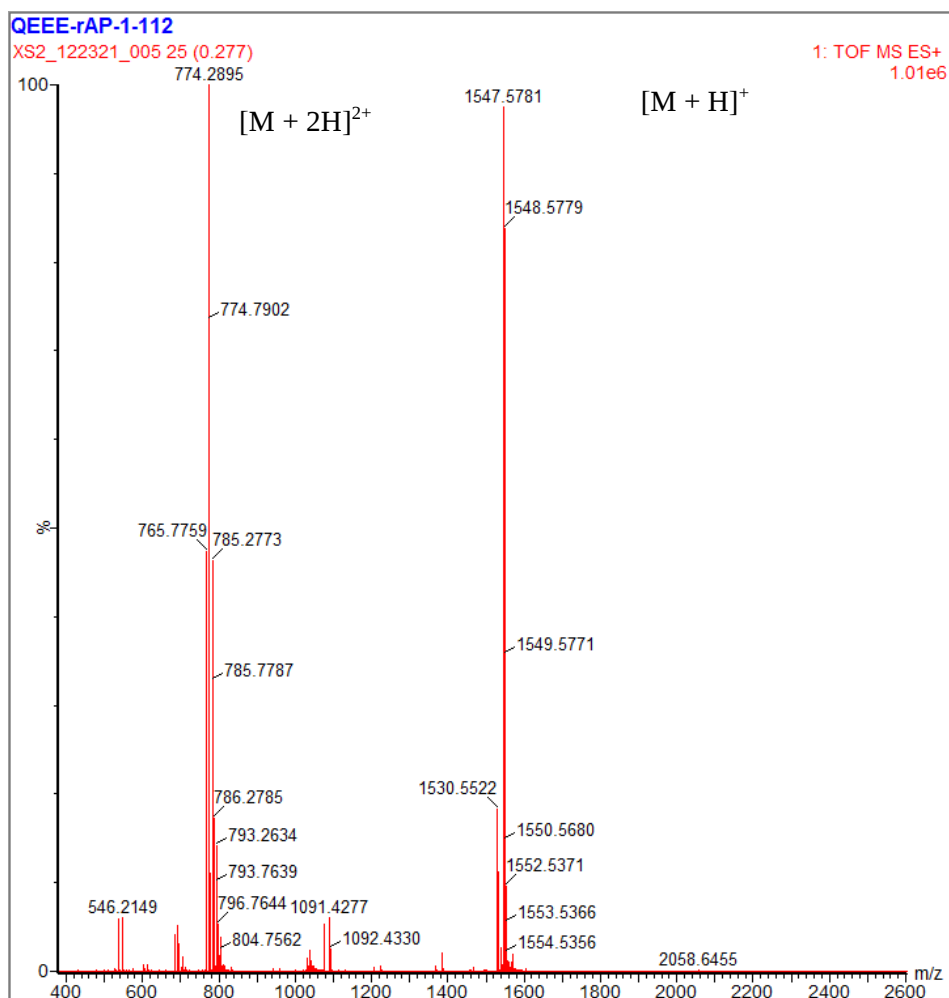
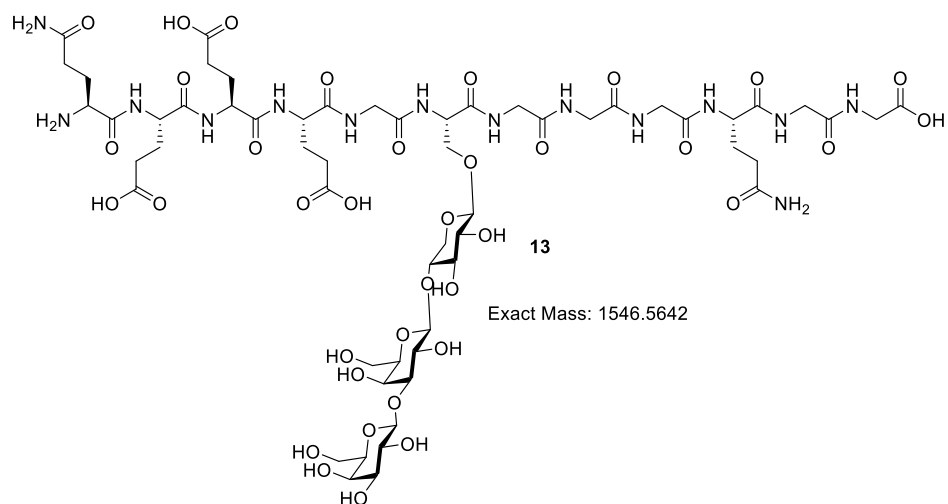
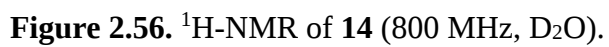
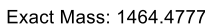


Figure 2.55. LCMS Chromatogram of **13**.



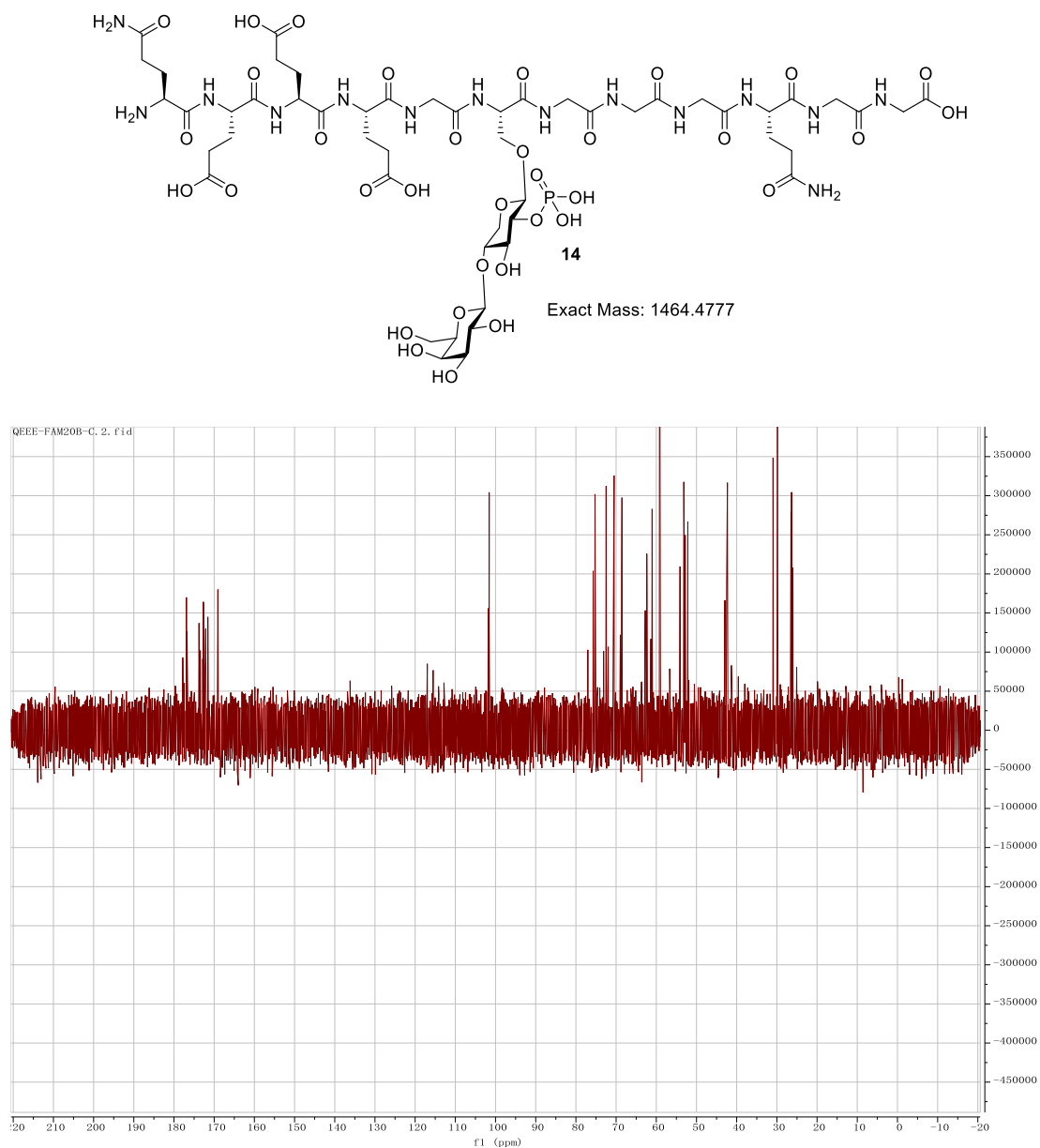


Figure 2.57. ^{13}C NMR of **14** (201 MHz, D_2O).

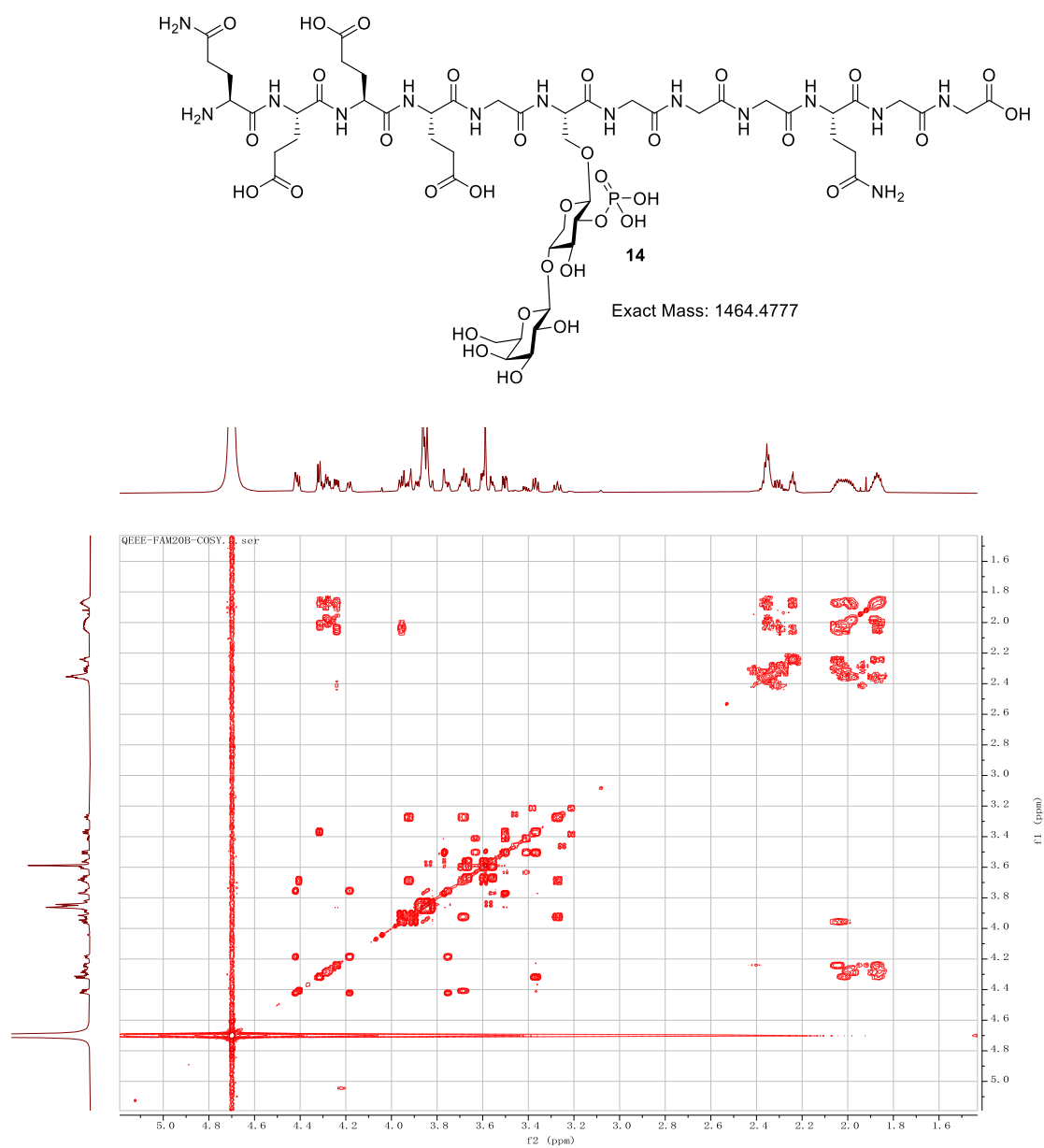


Figure 2.58. COSY of **14** (800 MHz, D₂O).

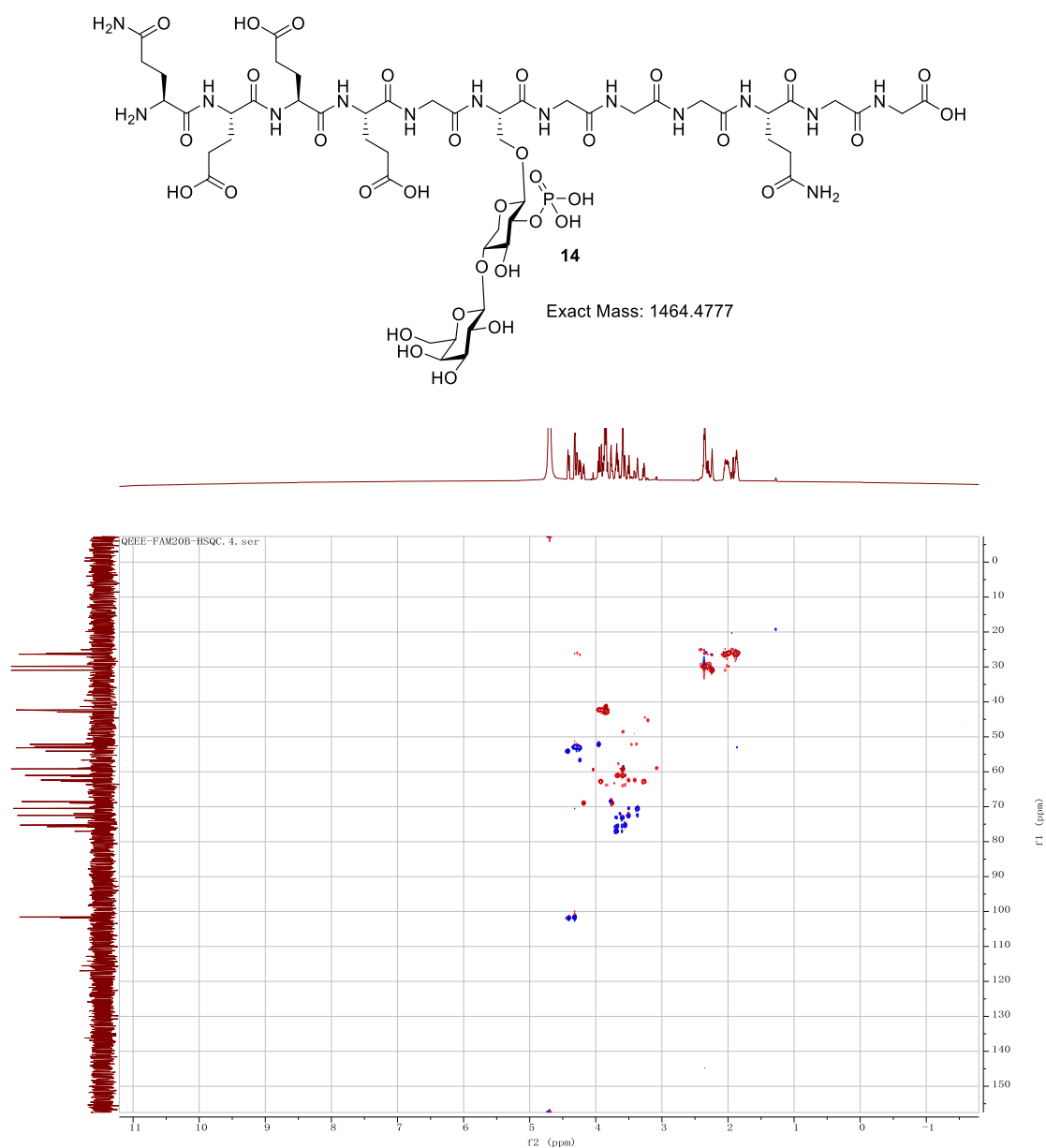


Figure 2.59. HSQC of **14** (800 MHz, D₂O).

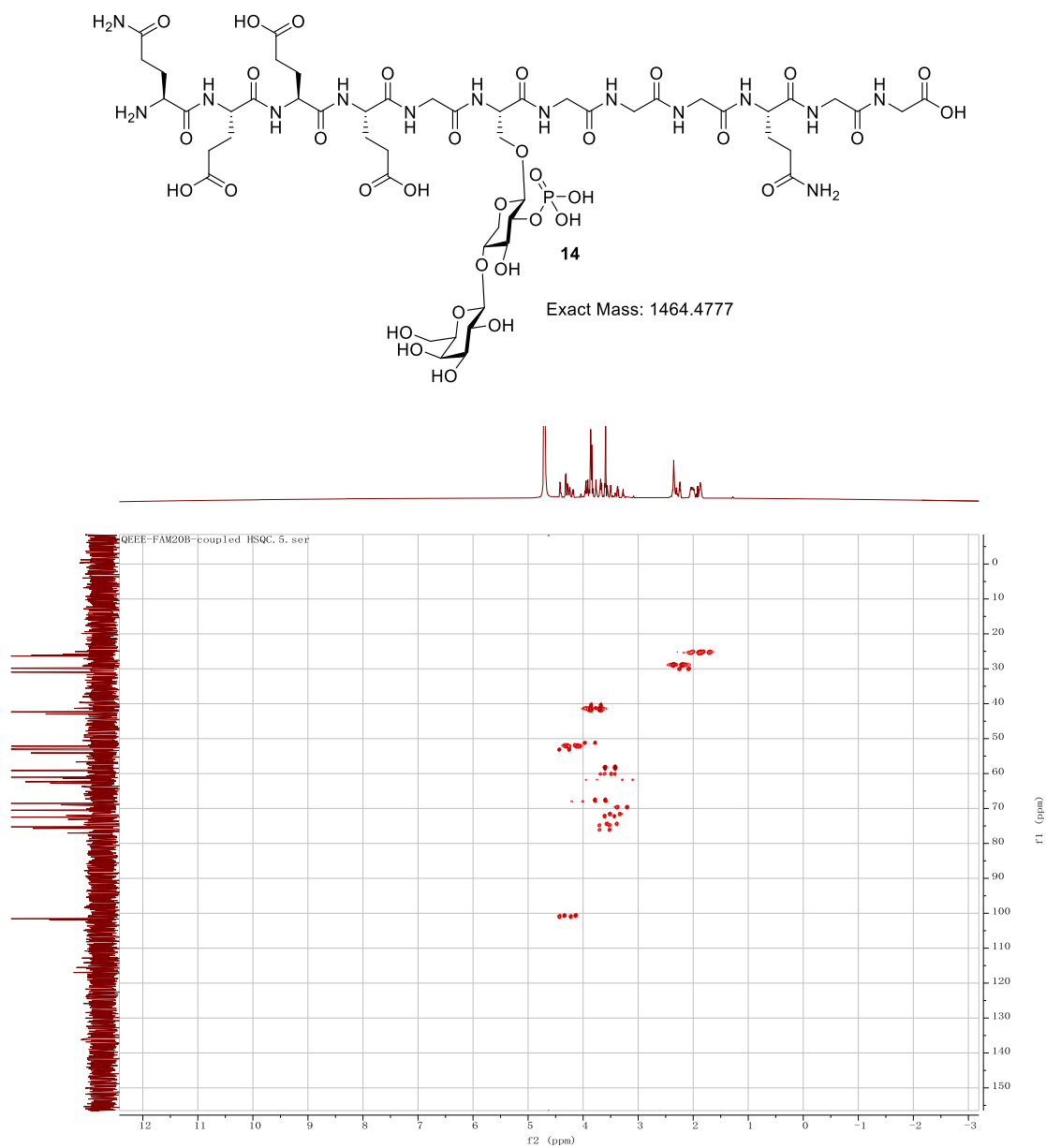


Figure 2.60. Coupled HSQC of **14** (800 MHz, D₂O).

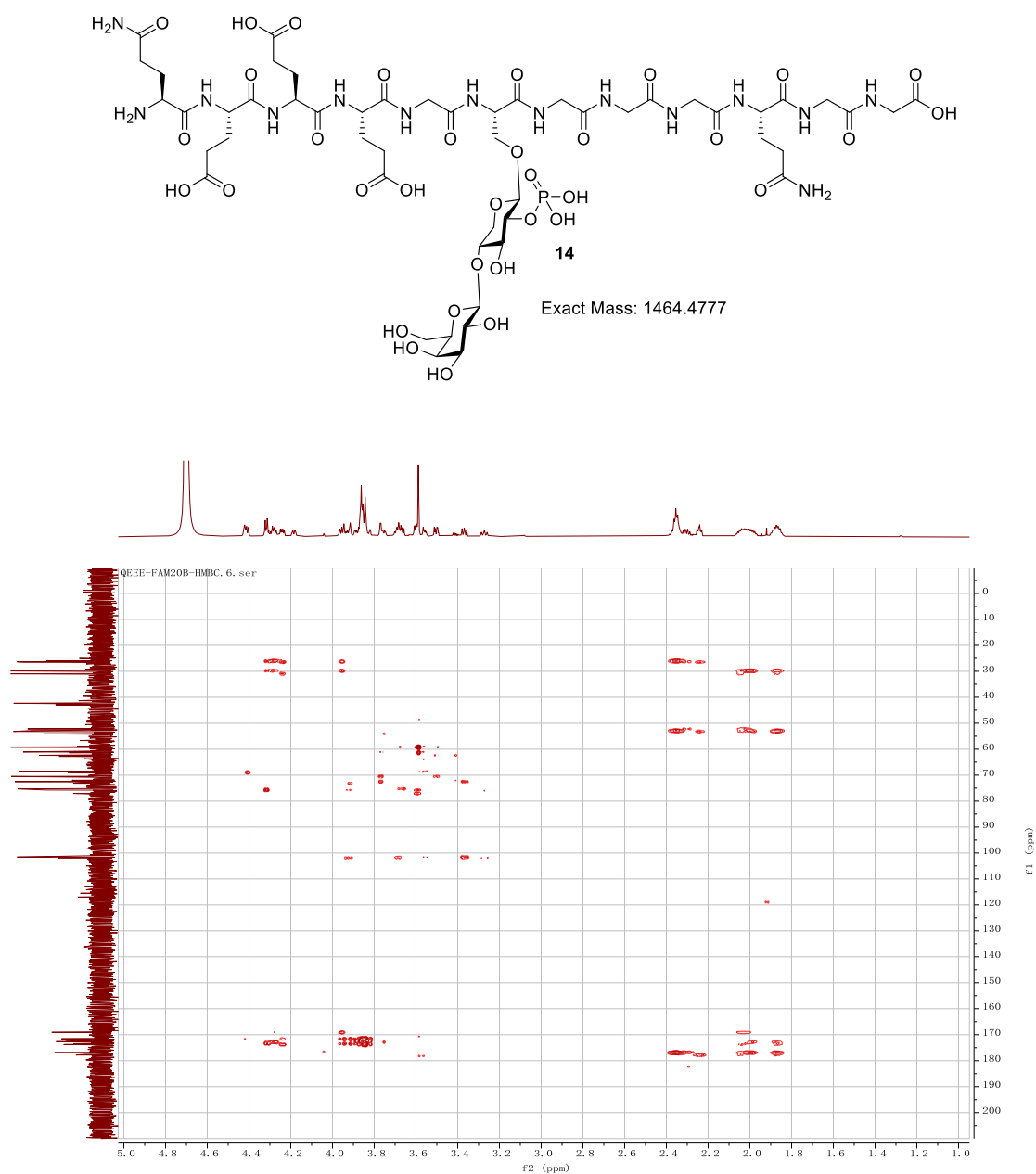


Figure 2.61. HMBC of **14** (800 MHz, D₂O).

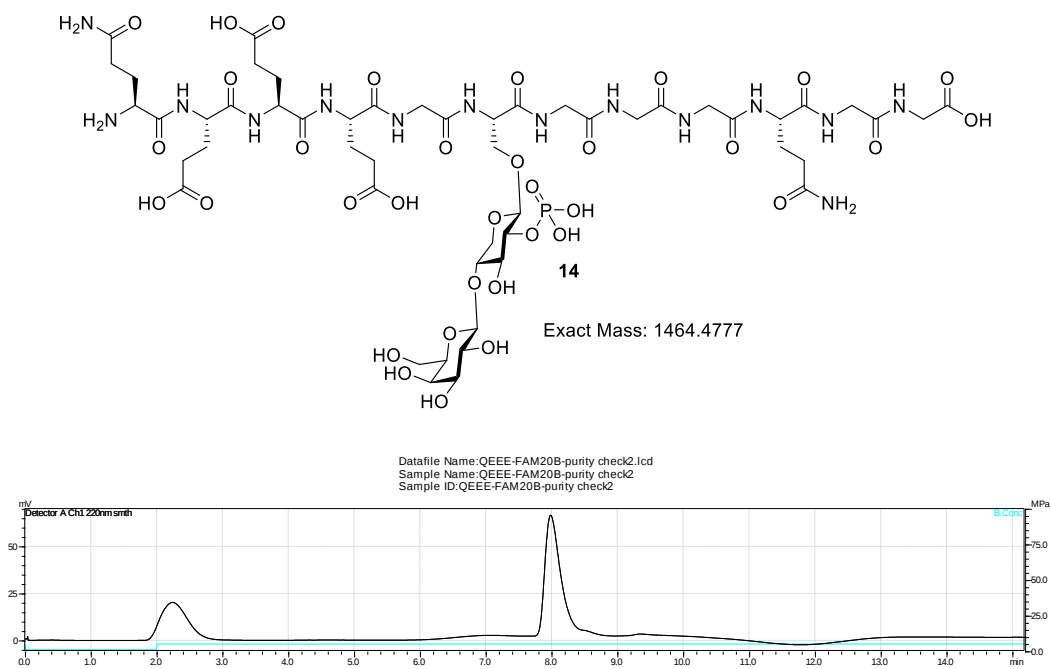
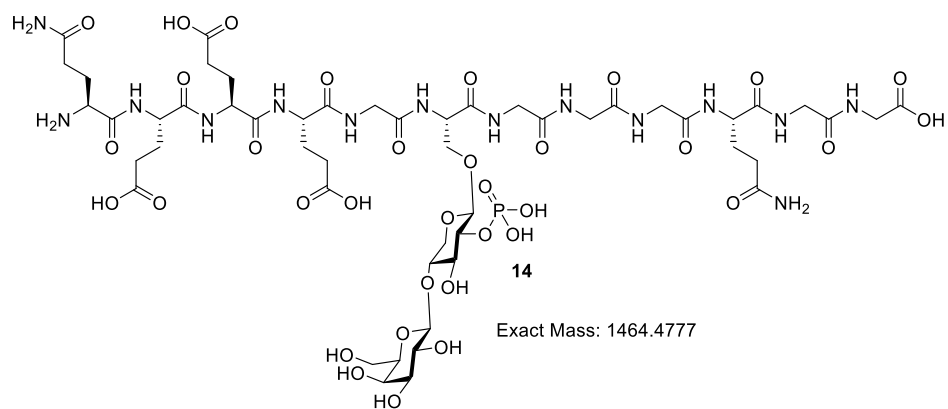


Figure 2.62. HPLC Chromatogram of **14**.



QEEE-FAM20B-quality check-134

XS2_120221_009 35 (0.374)

1: TOF MS ES+
1.12e5

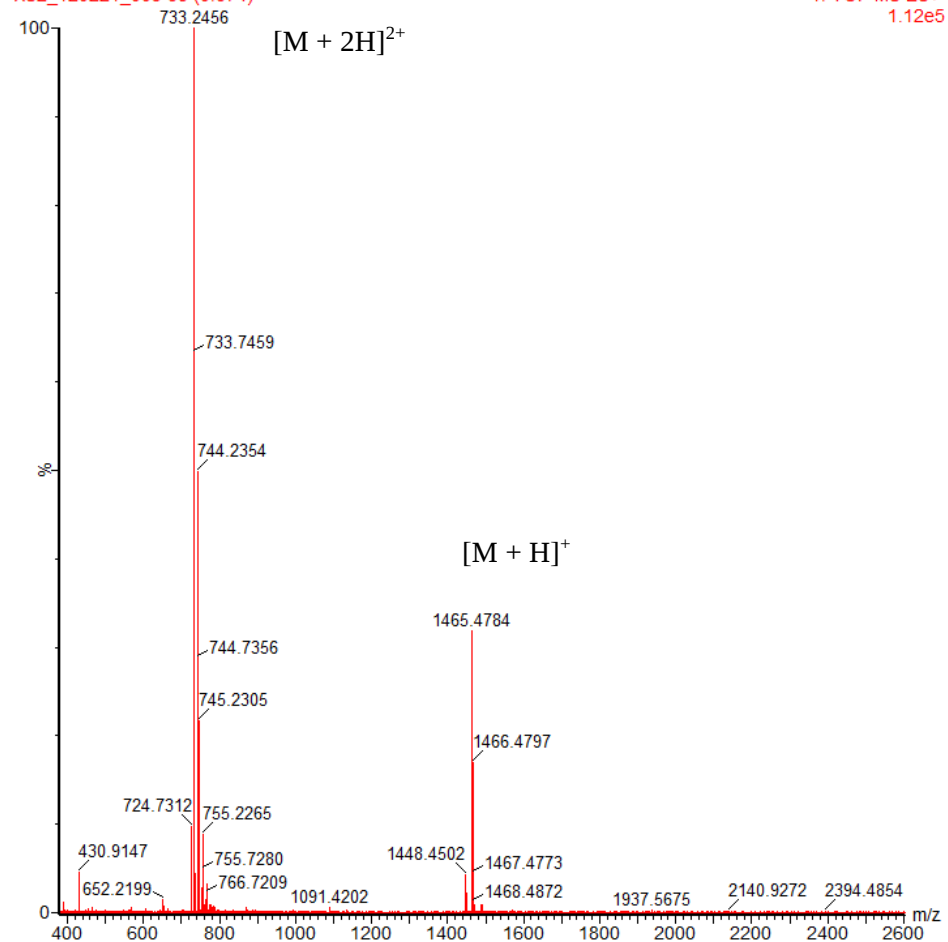


Figure 2.63. LCMS Chromatogram of **14**.

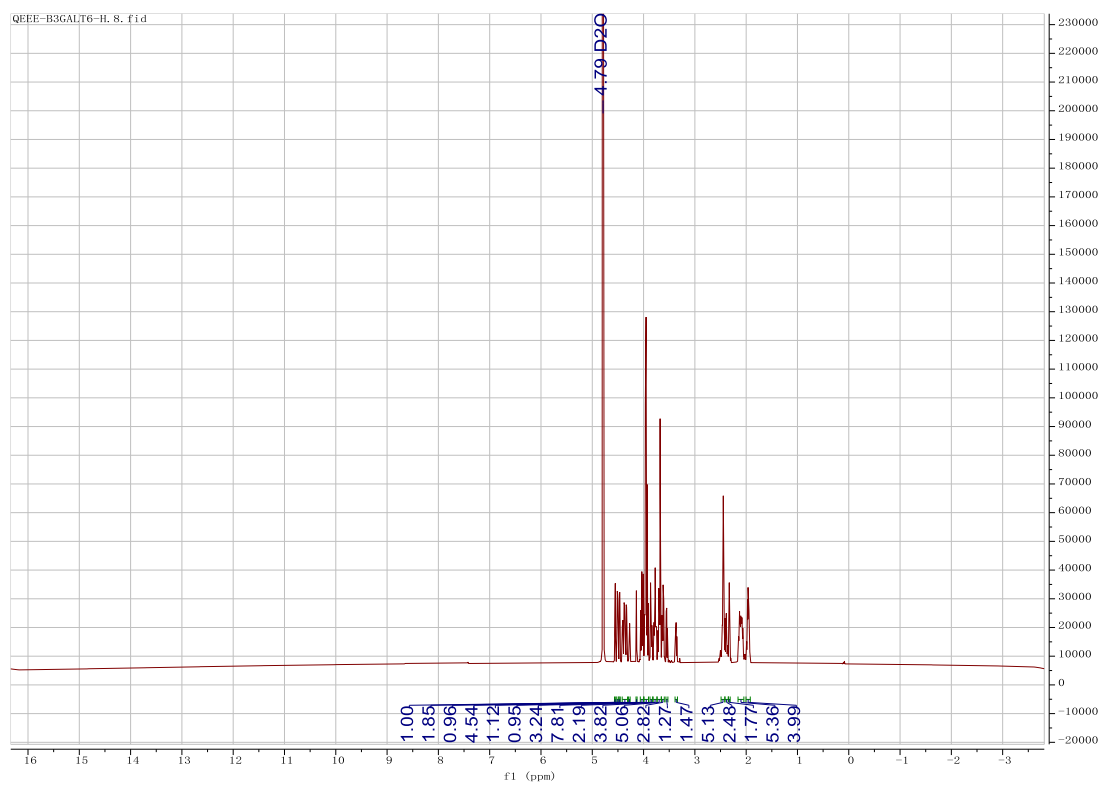
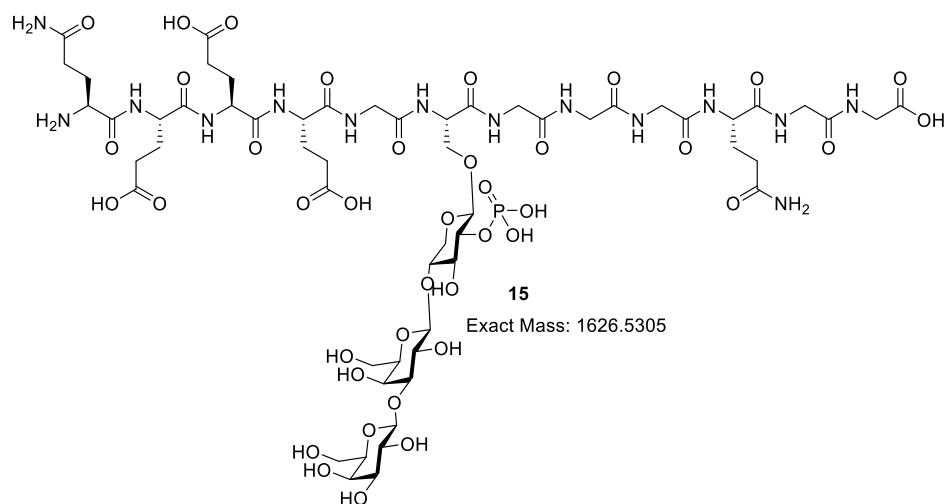


Figure 2.64. ¹H-NMR of **15** (800 MHz, D₂O).

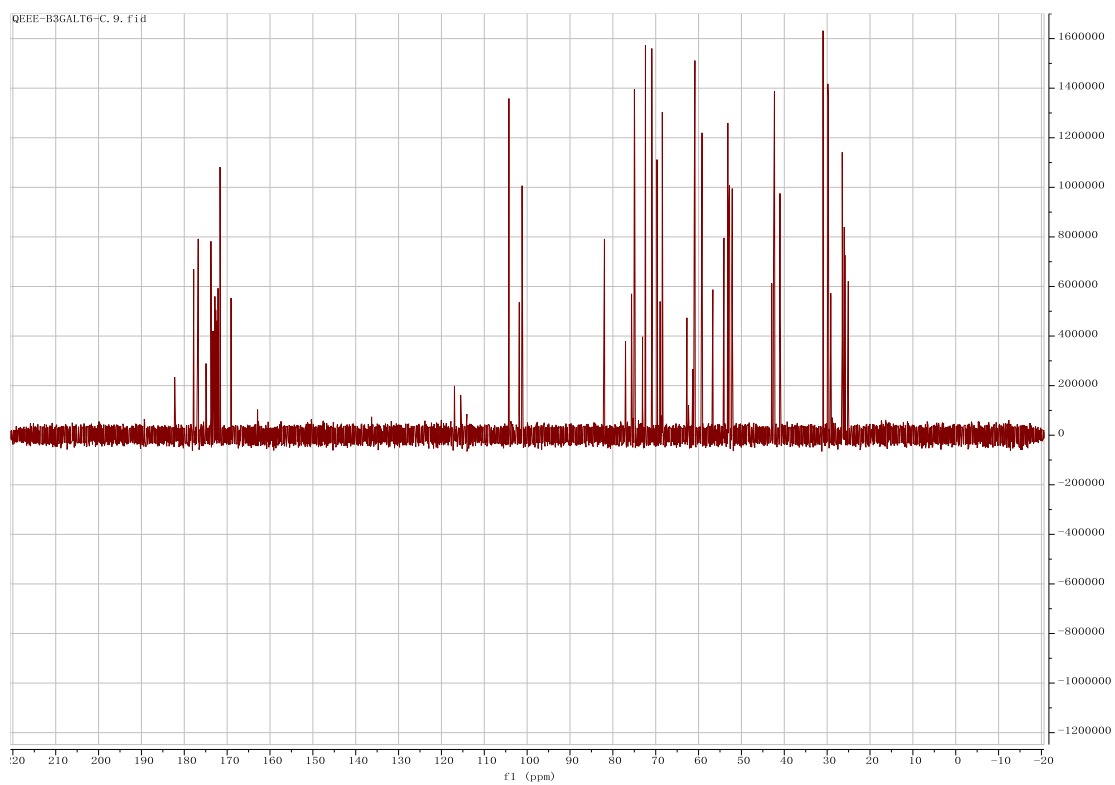
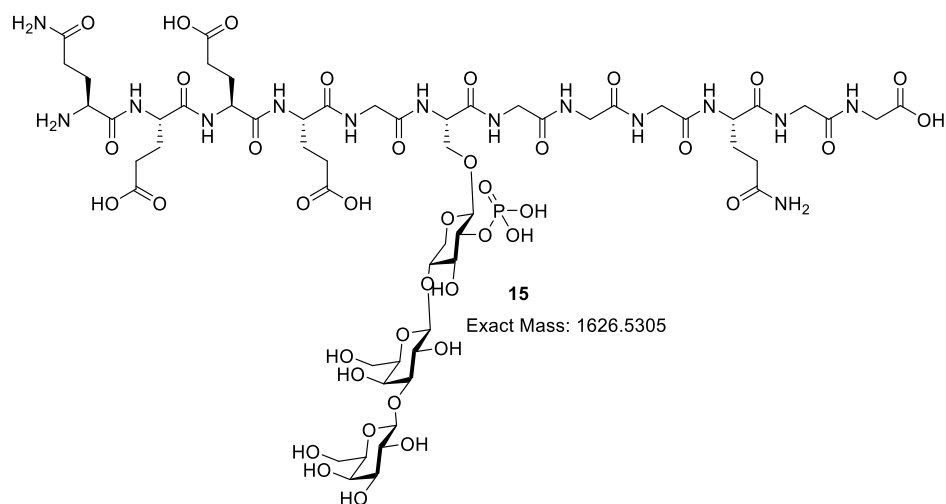


Figure 2.65. ^{13}C NMR of **15** (201 MHz, D_2O).

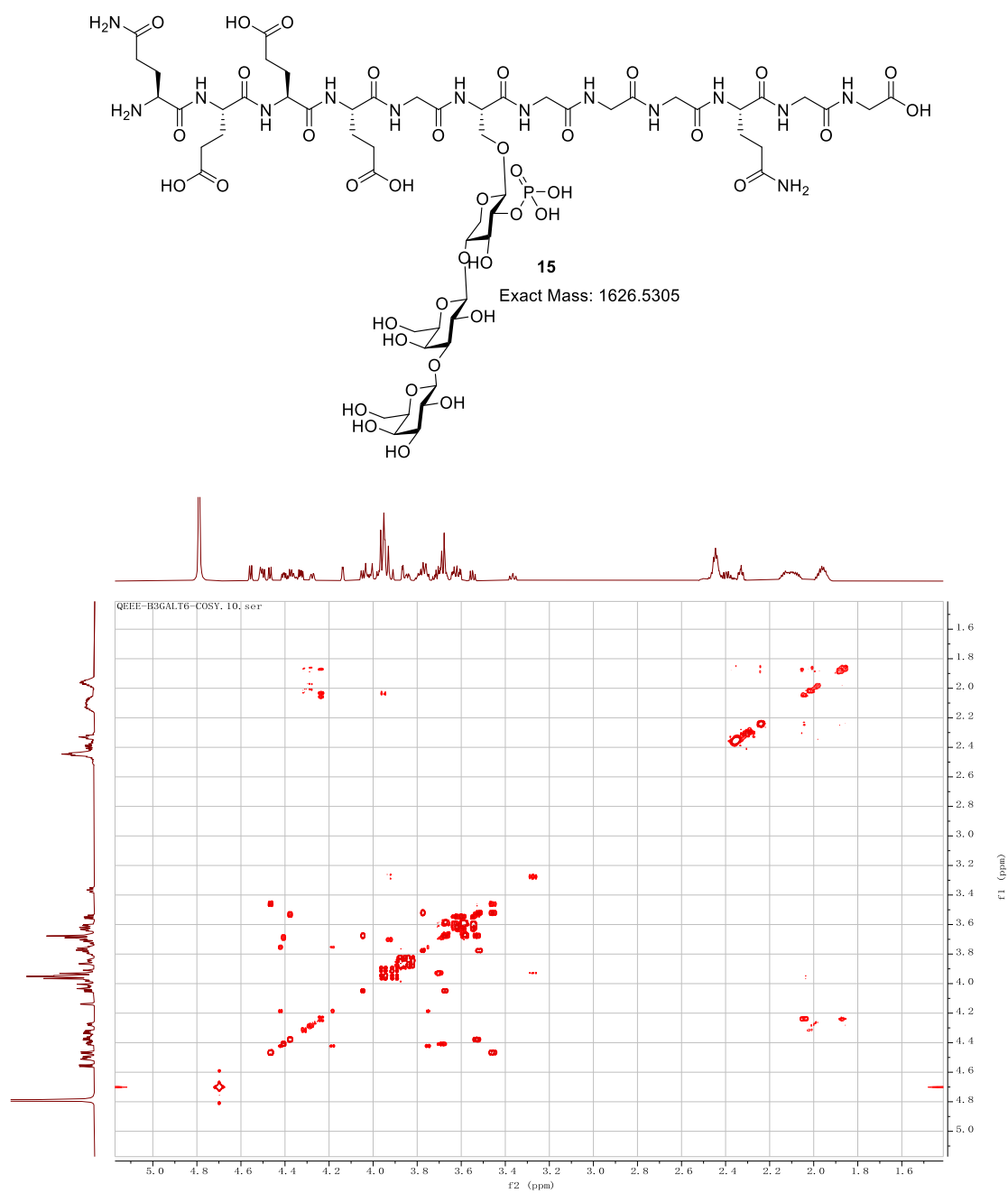


Figure 2.66. COSY of **15** (800 MHz, D₂O).

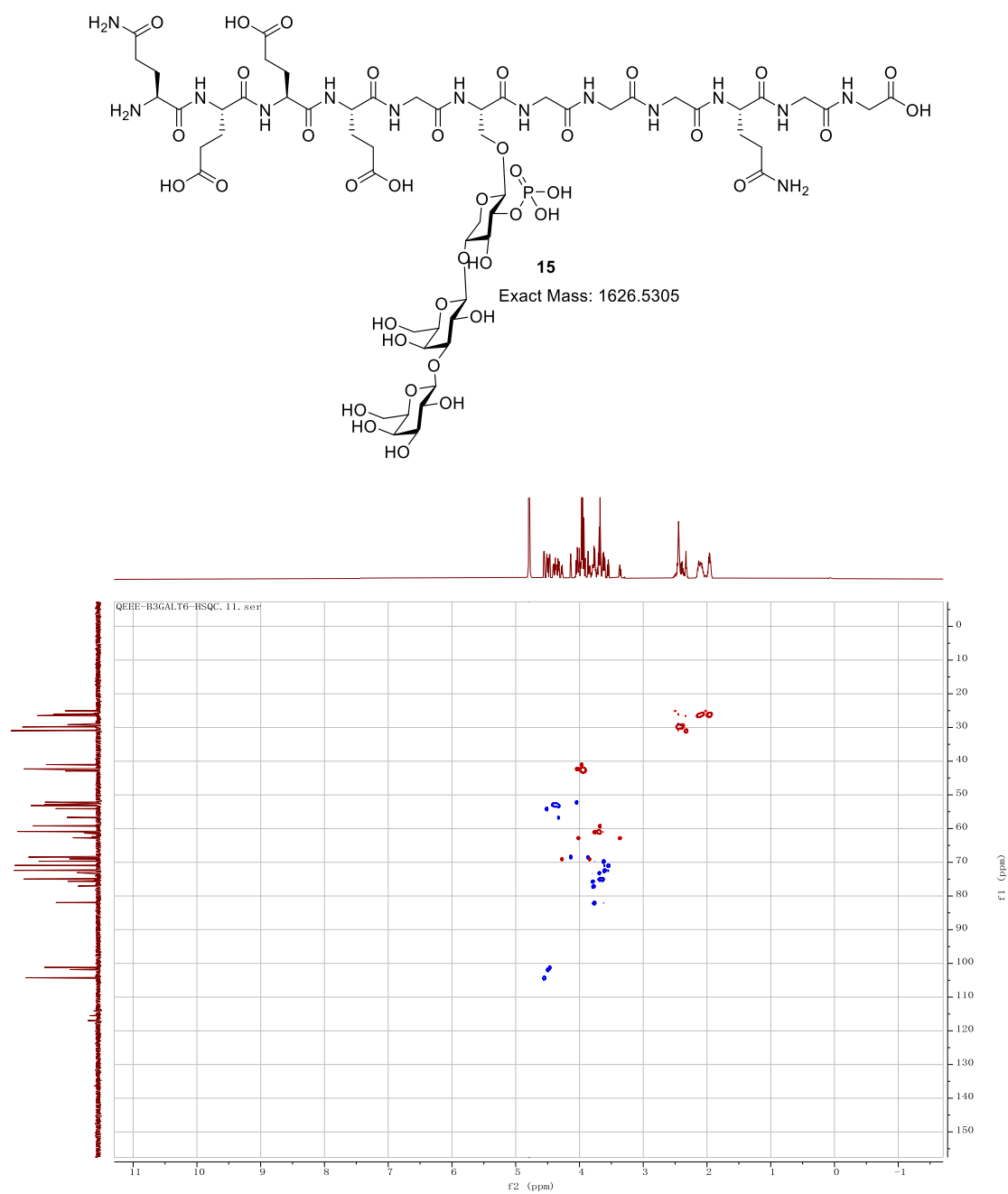


Figure 2.67. HSQC of **15** (800 MHz, D₂O).

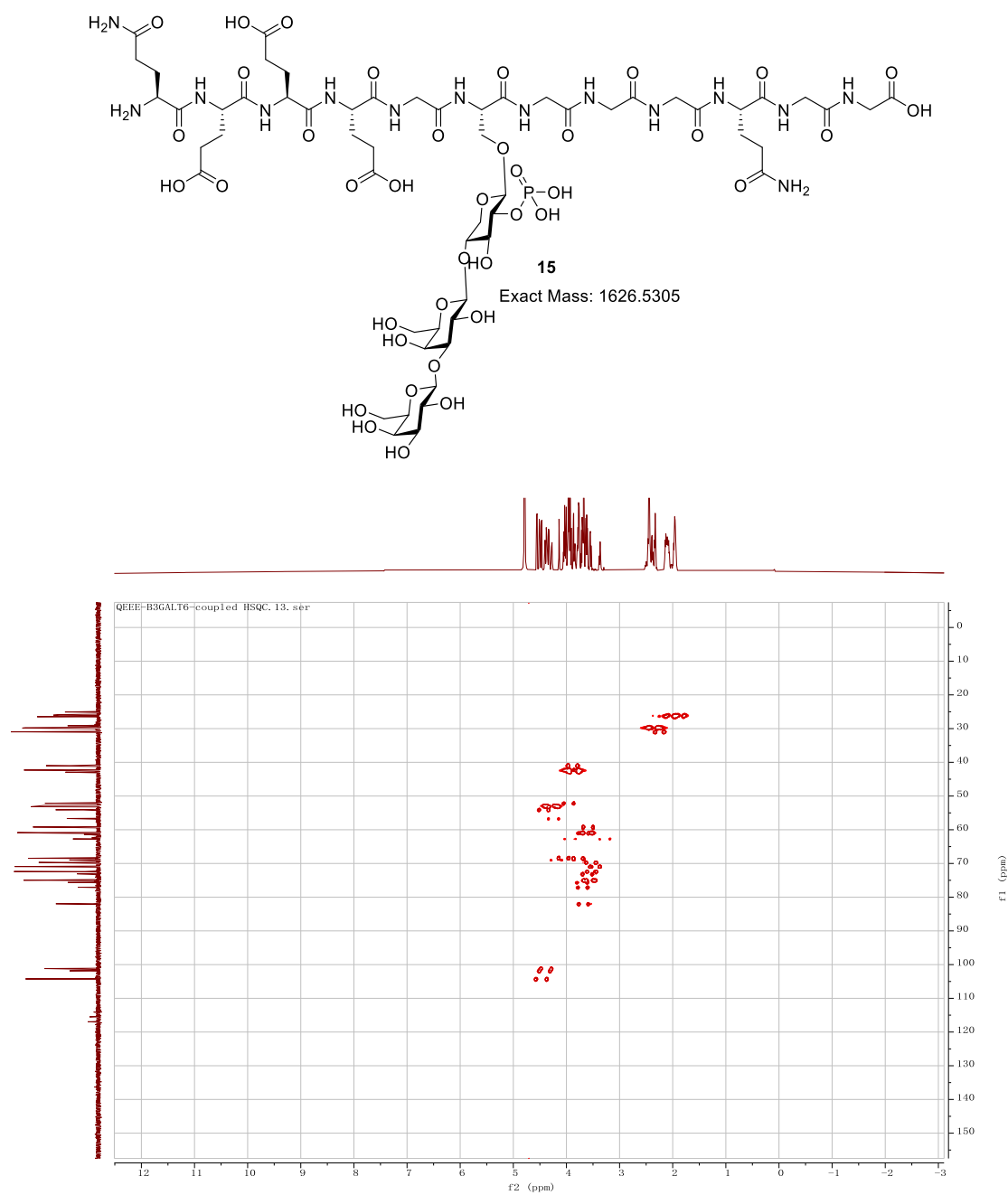


Figure 2.68. Coupled HSQC of **15** (800 MHz, D₂O).

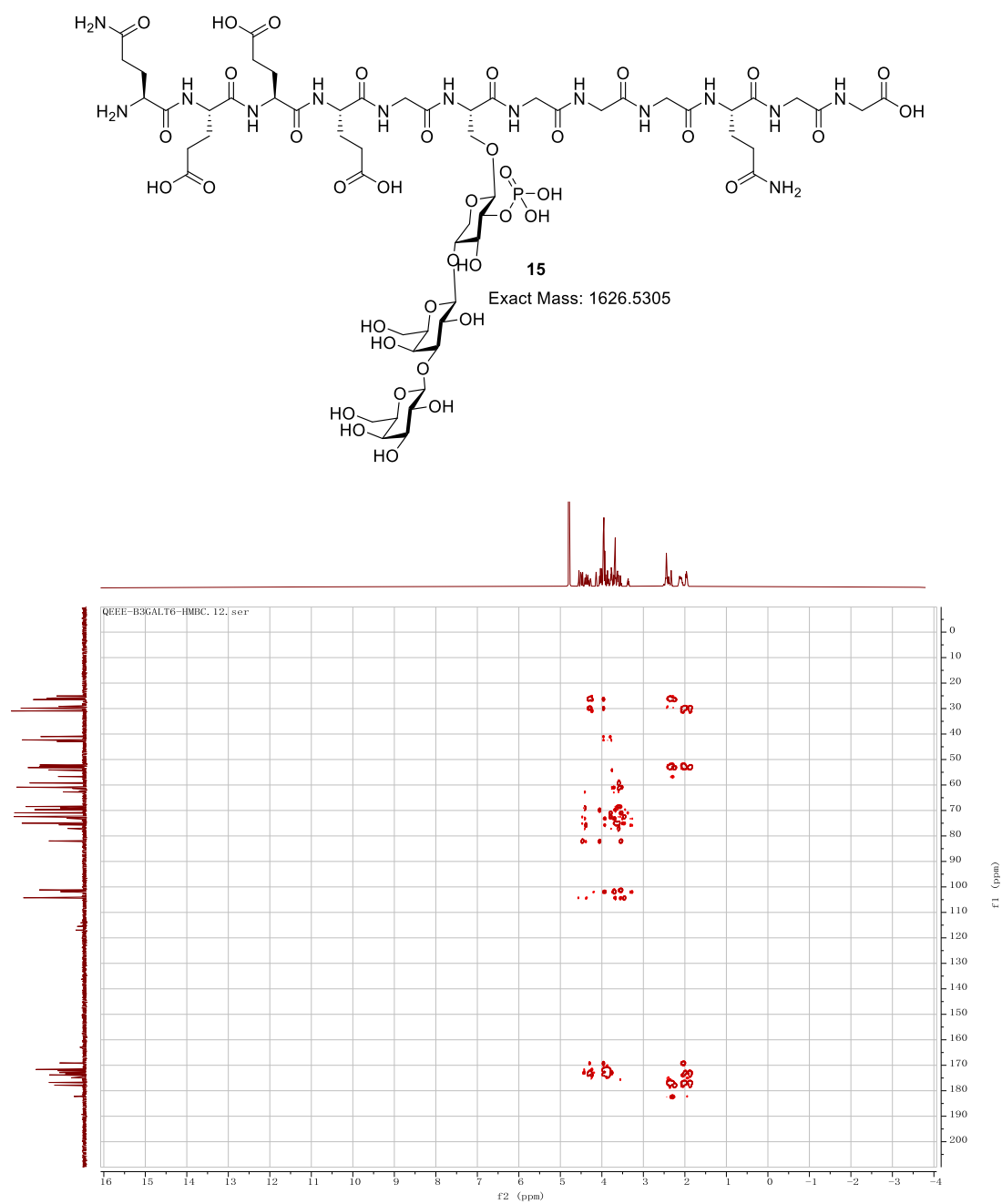


Figure 2.69. HMBC of **15** (800 MHz, D₂O).

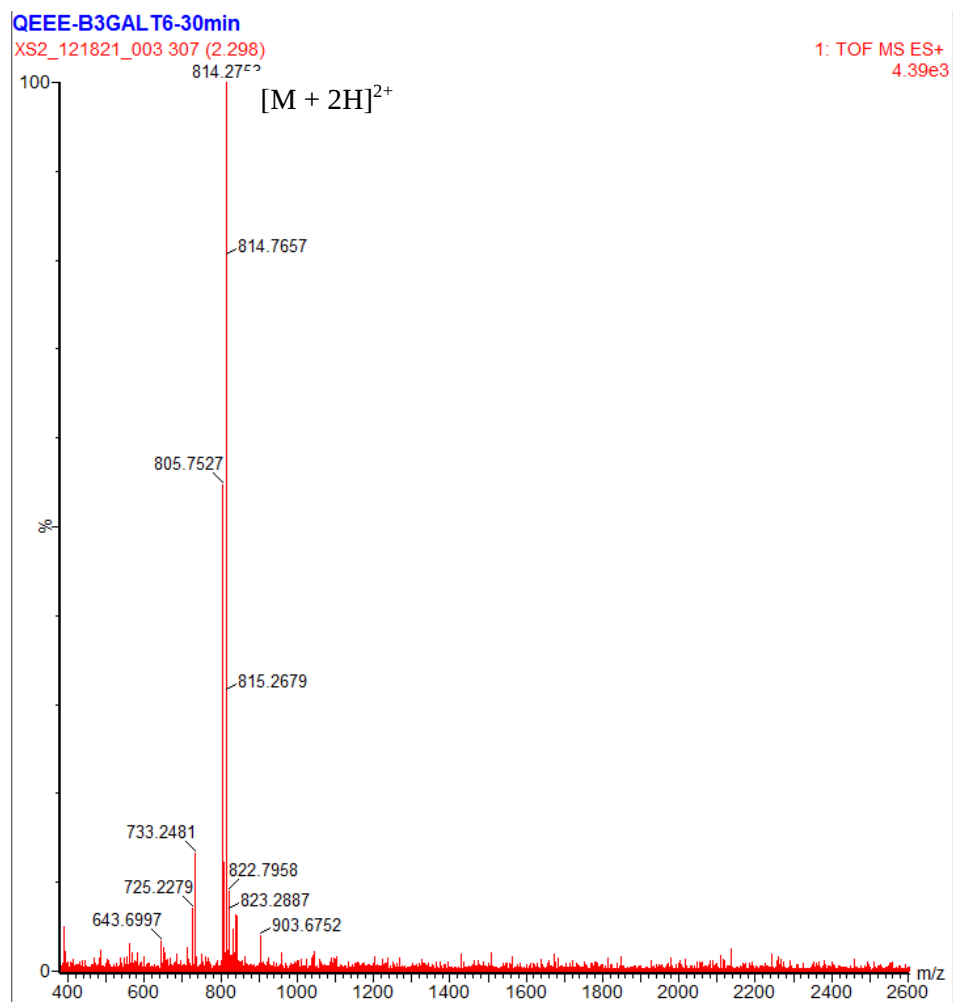
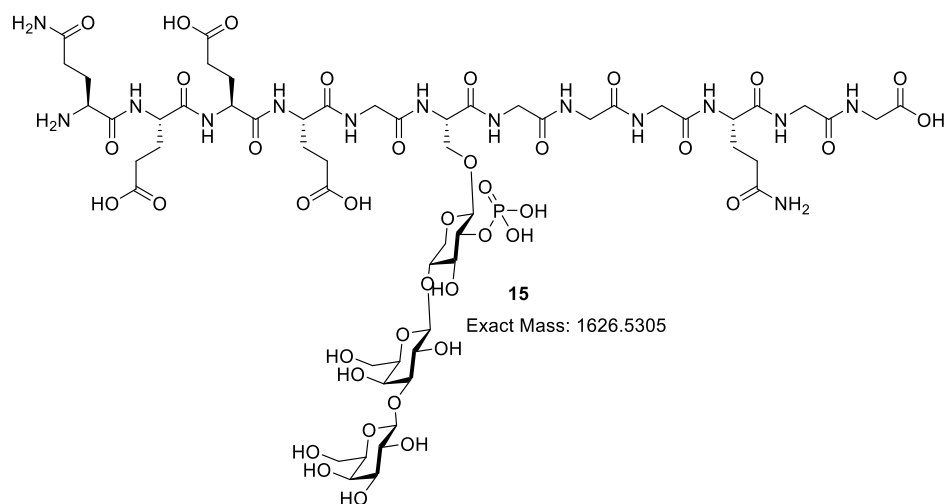


Figure 2.70. LCMS Chromatogram of **15**.

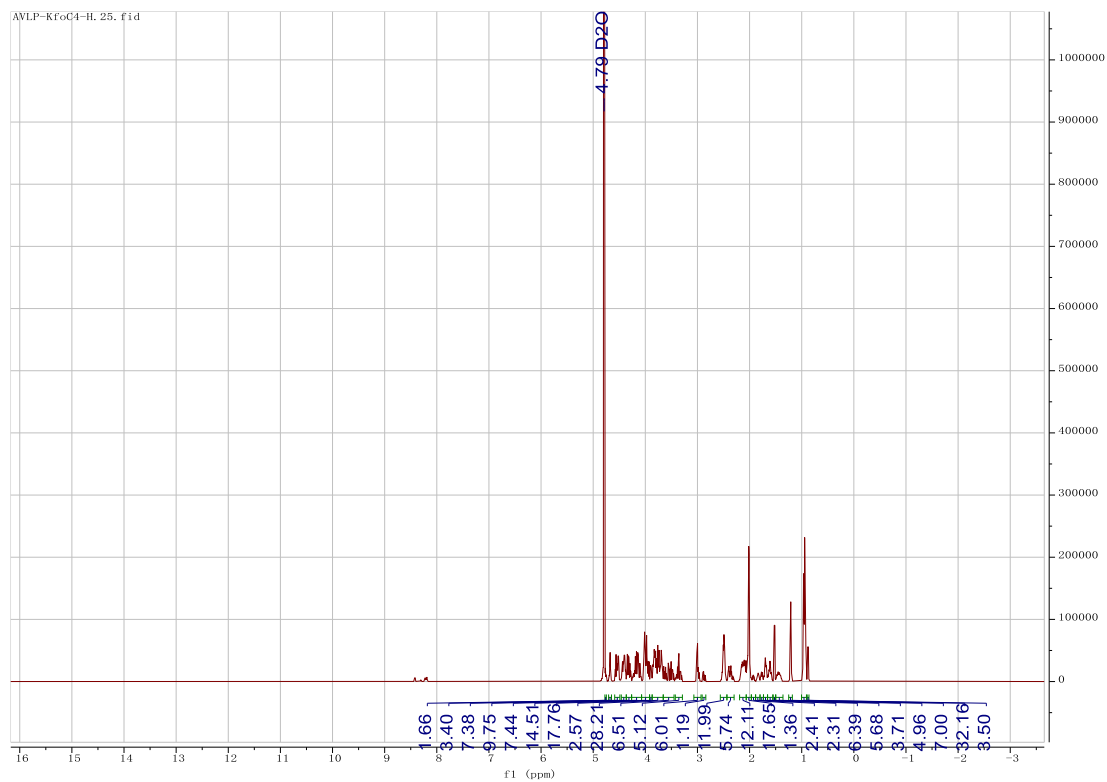
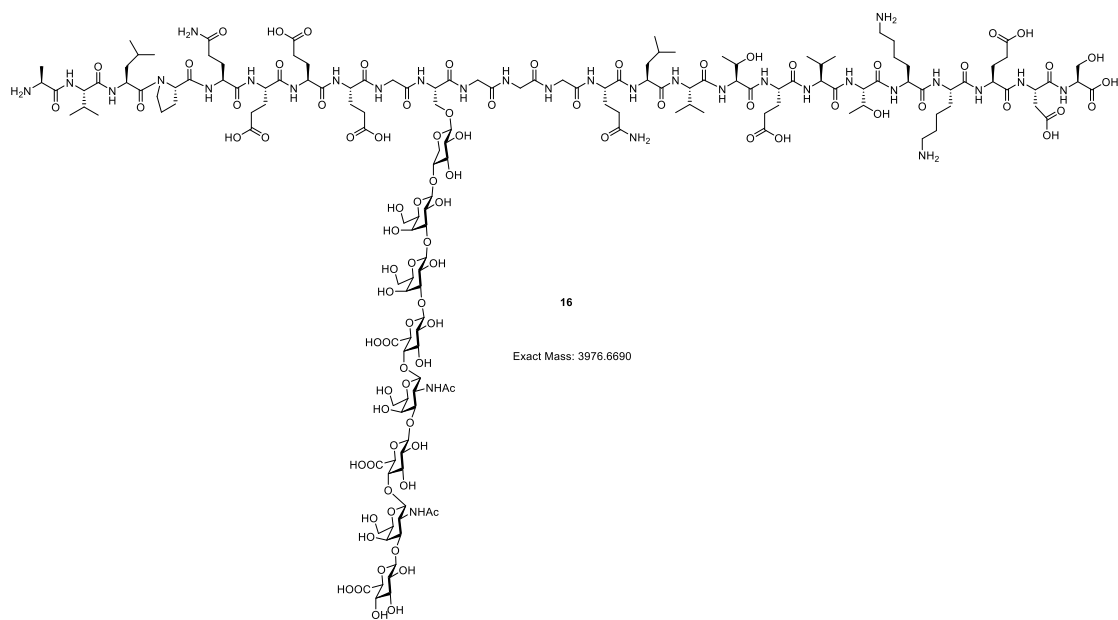


Figure 2.71. ^1H -NMR of **16** (600 MHz, D_2O).

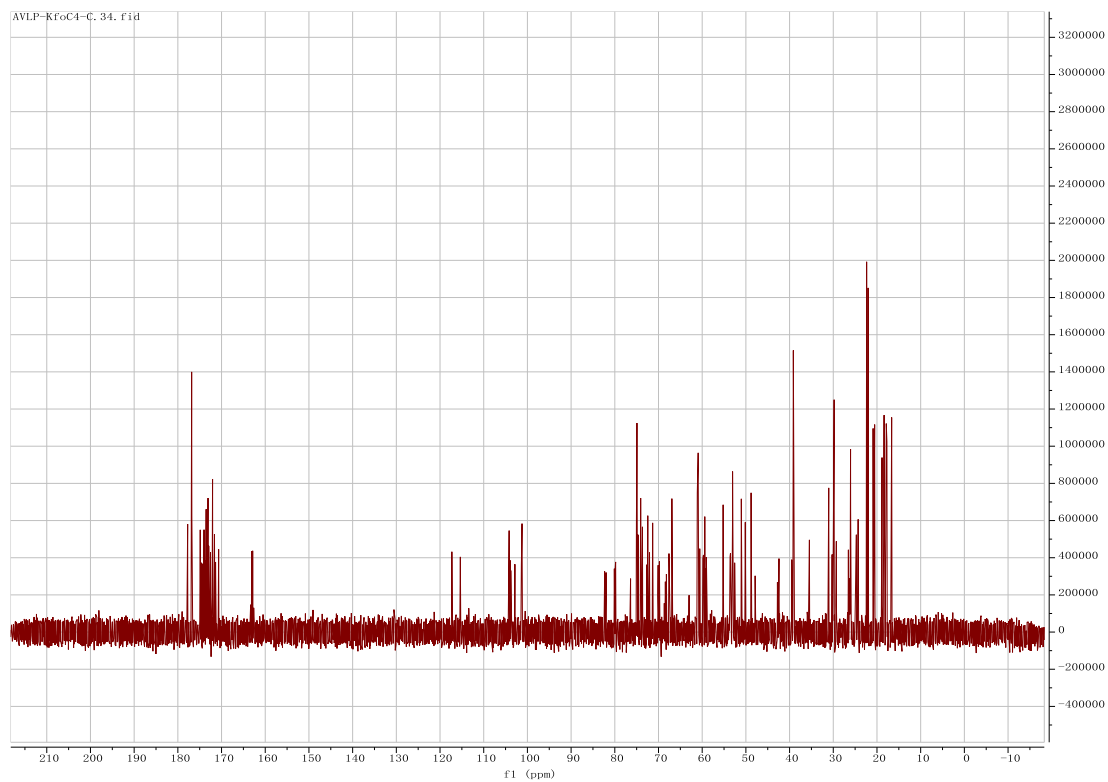
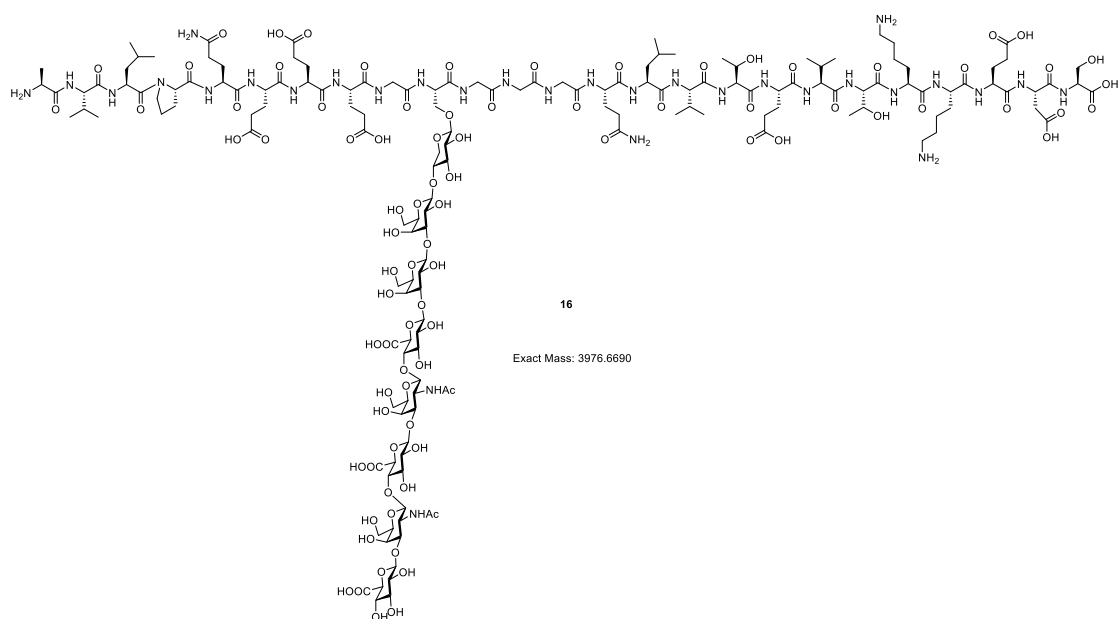


Figure 2.72. ^{13}C NMR of **16** (151 MHz, D_2O).

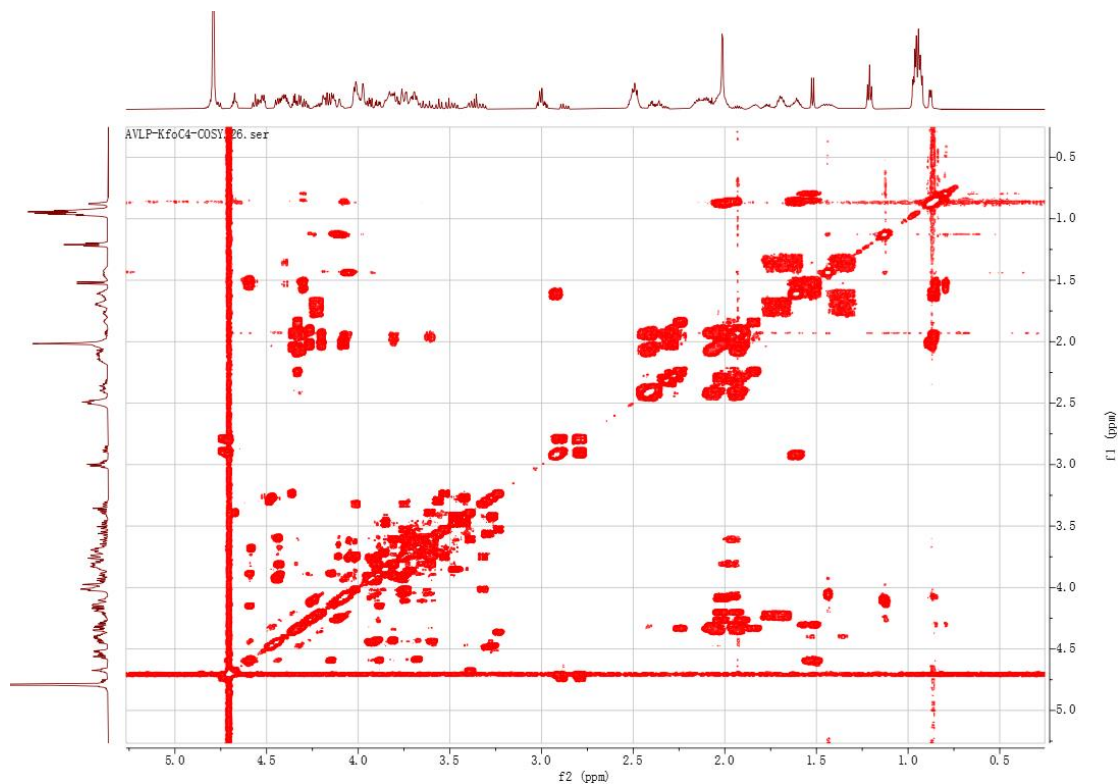
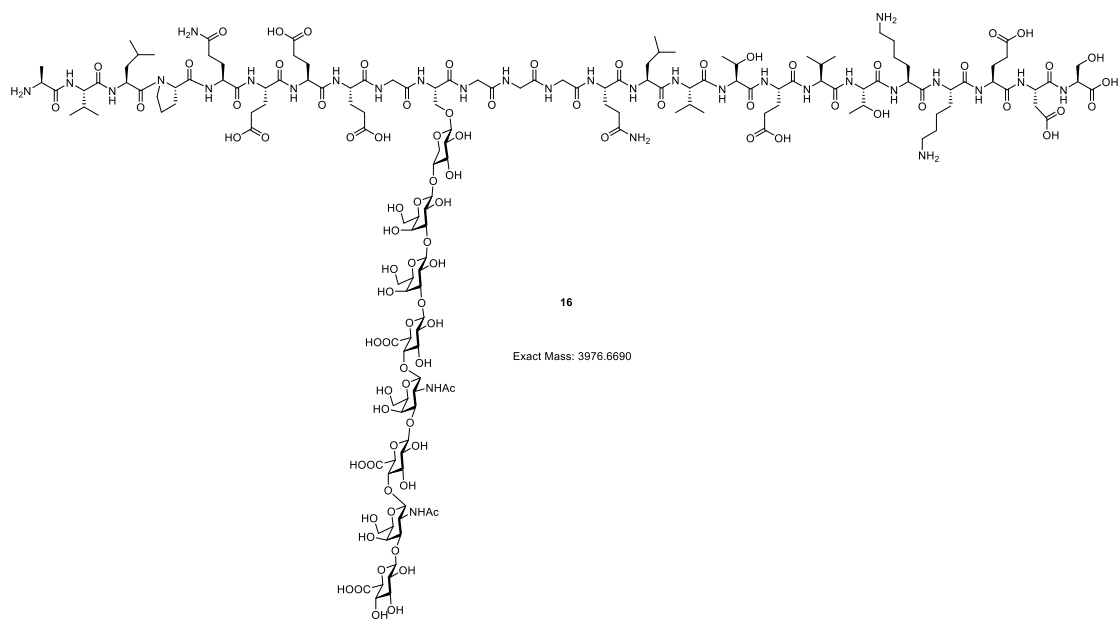


Figure 2.73. COSY of **16** (600 MHz, D₂O).

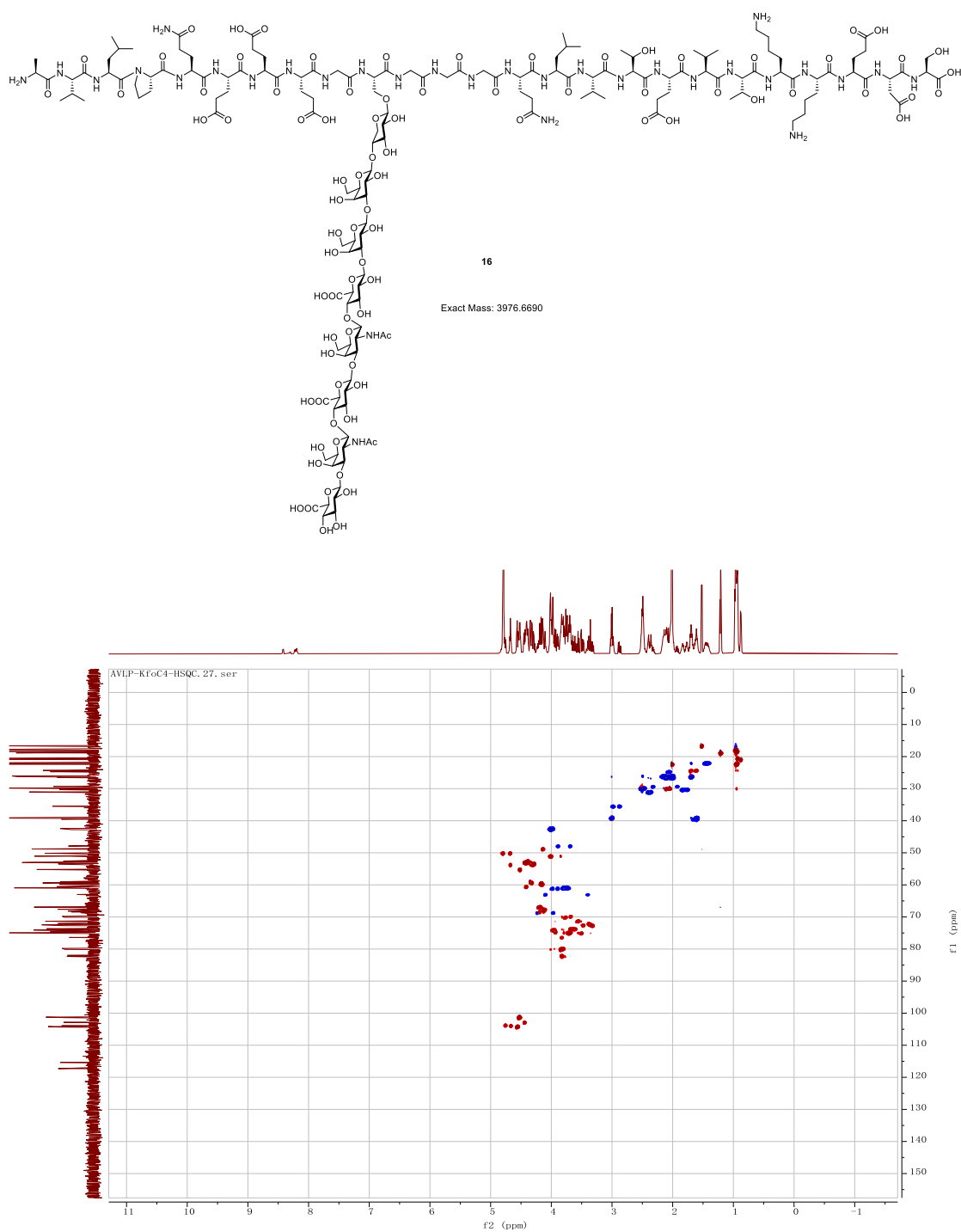


Figure 2.74. HSQC of **16** (600 MHz, D₂O).

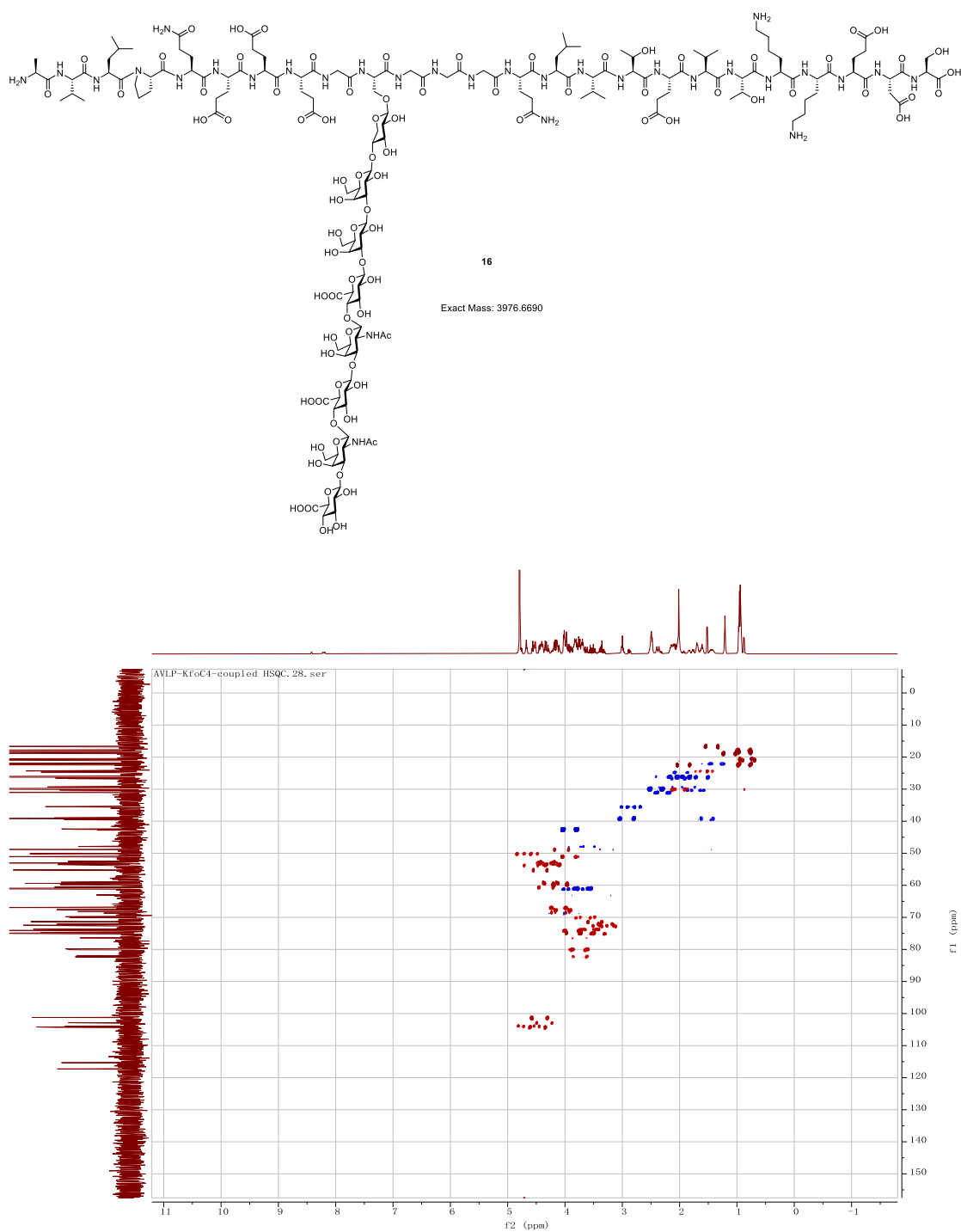


Figure 2.75. Coupled HSQC of **16** (600 MHz, D₂O).

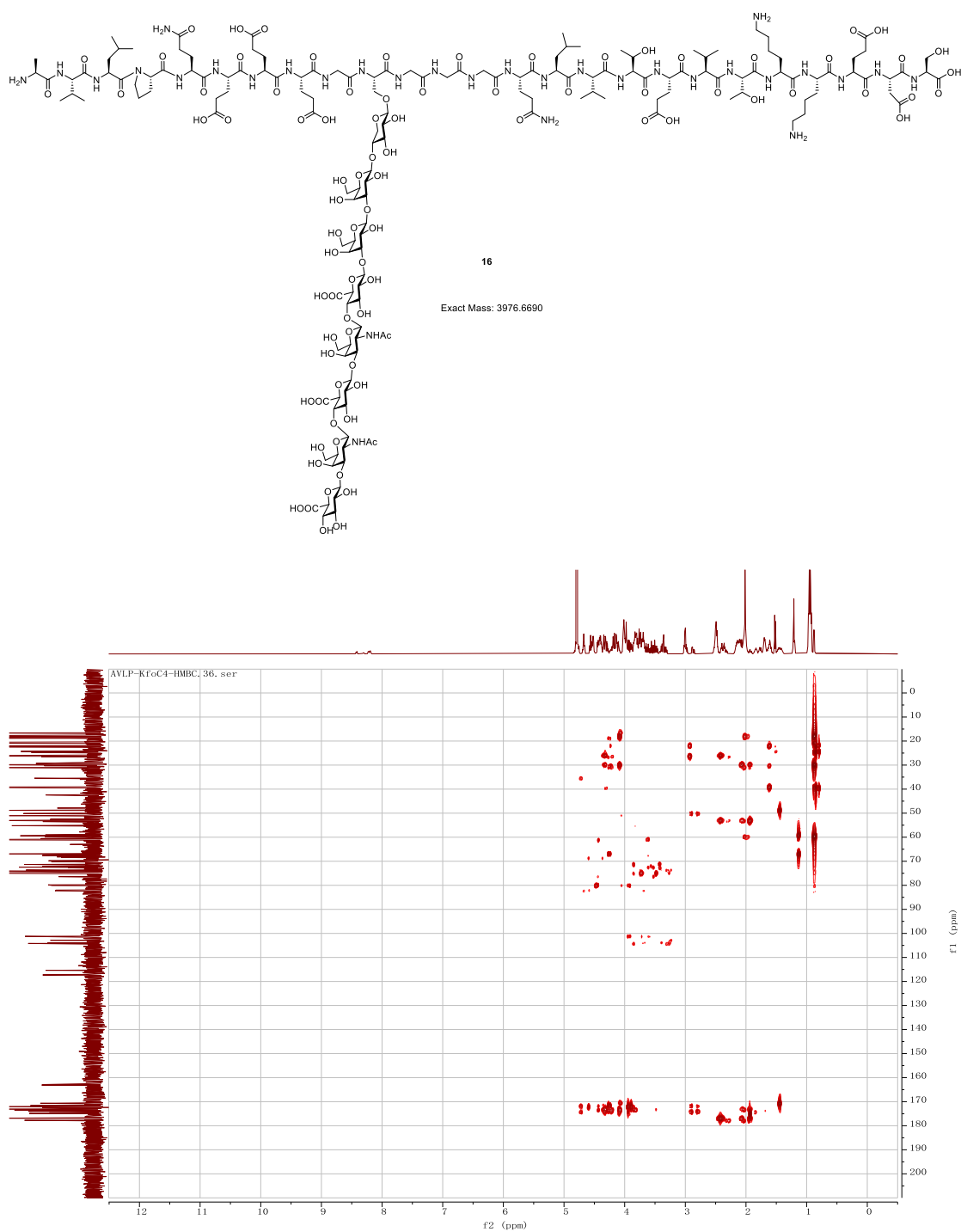


Figure 2.76. HMBC of **16** (600 MHz, D₂O).

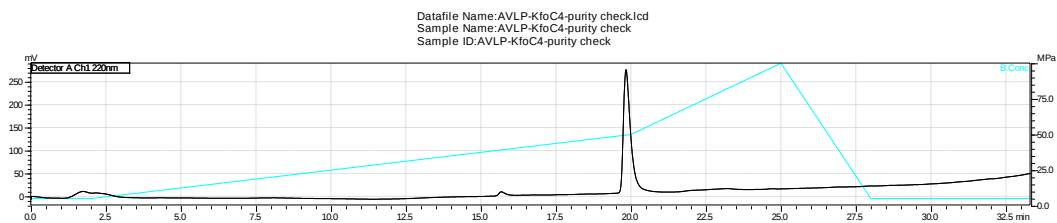
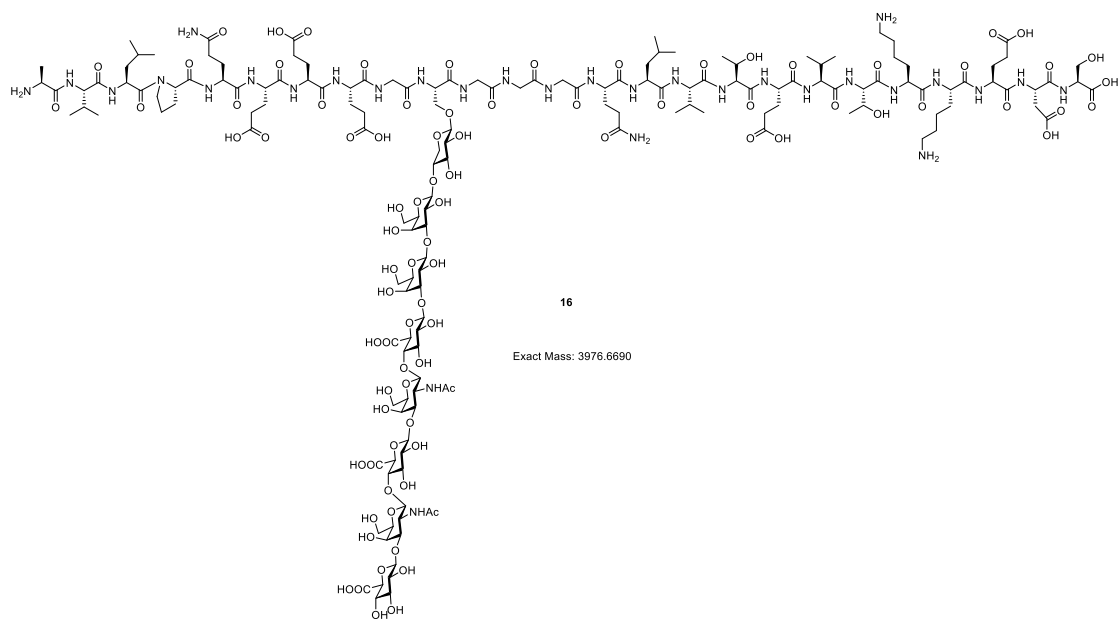


Figure 2.77. HPLC Chromatogram of **16**.

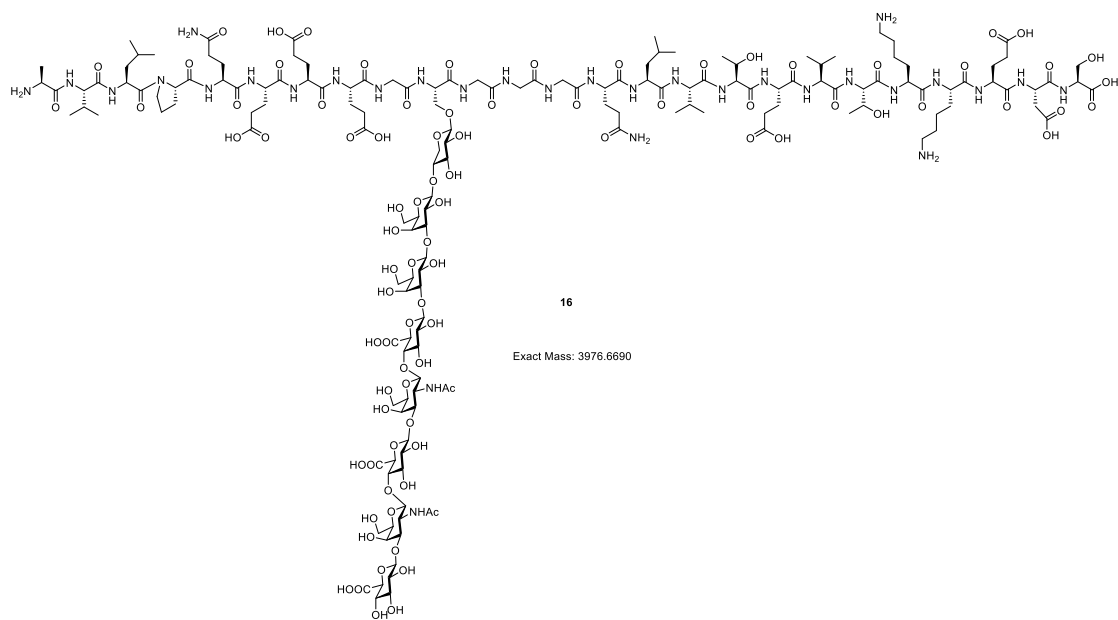
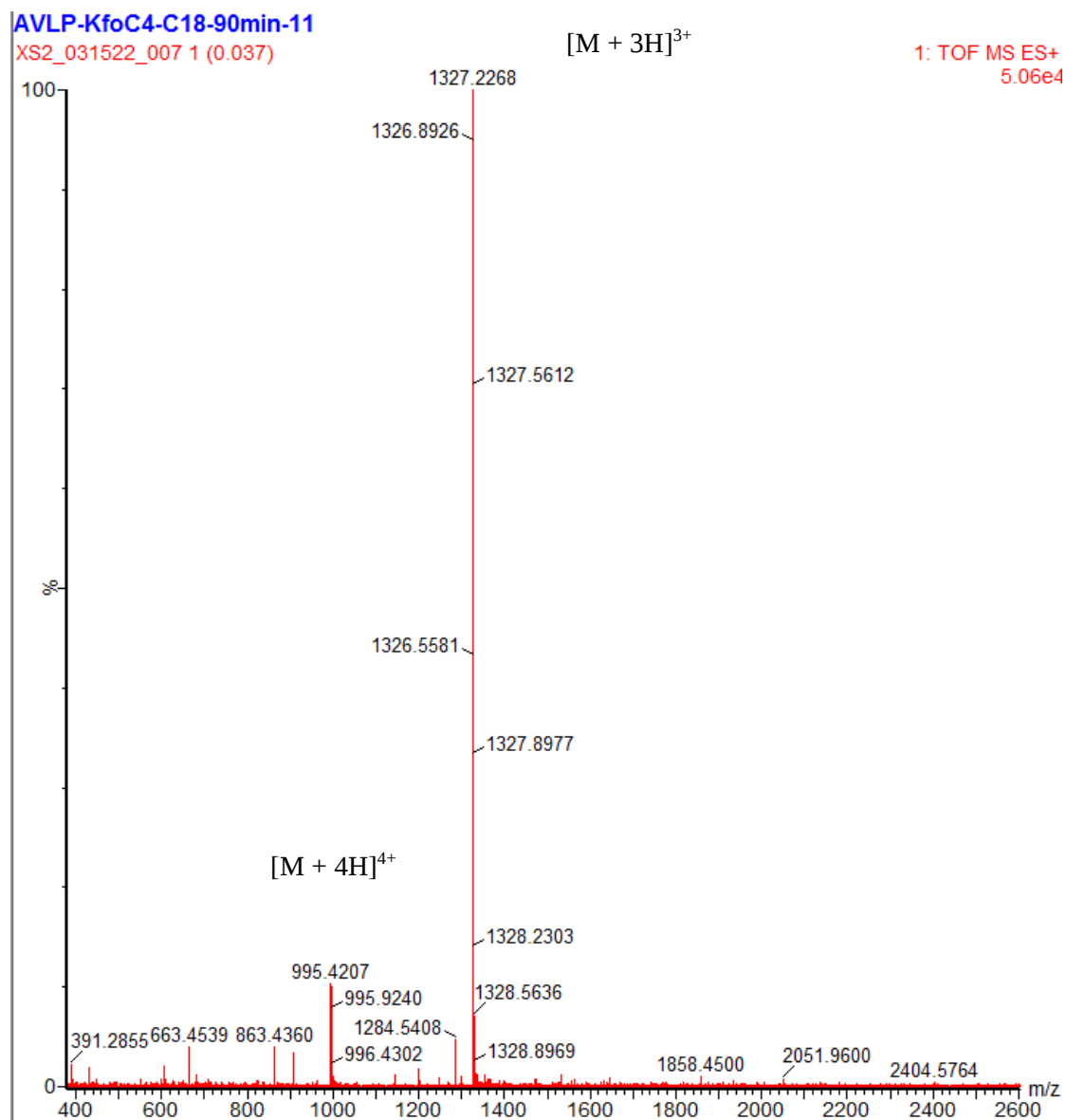


Figure 2.78. LCMS Chromatogram of **16**.

Figure 2.78. (cont'd)



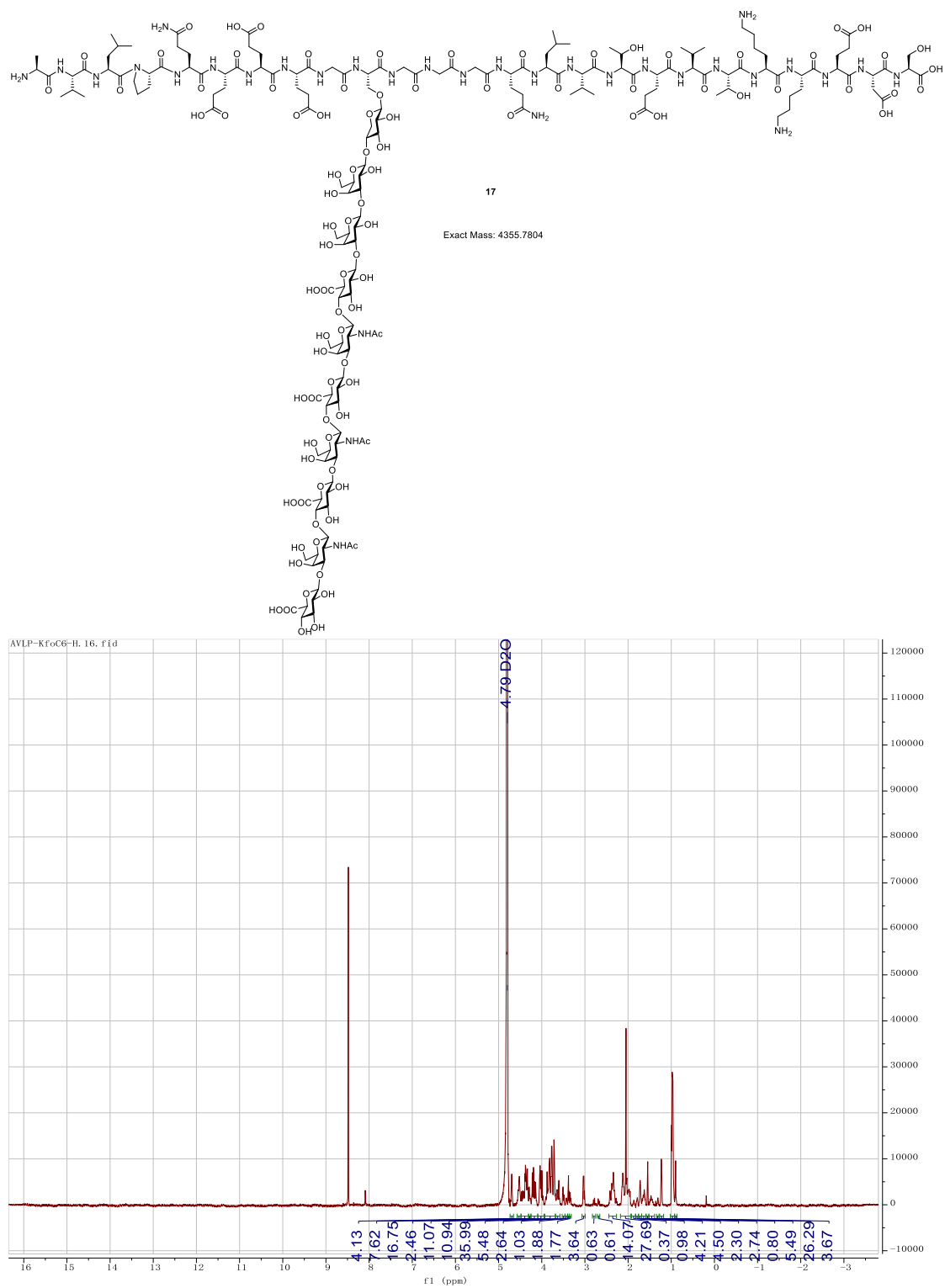


Figure 2.79. ¹H-NMR of **17** (800 MHz, D₂O).

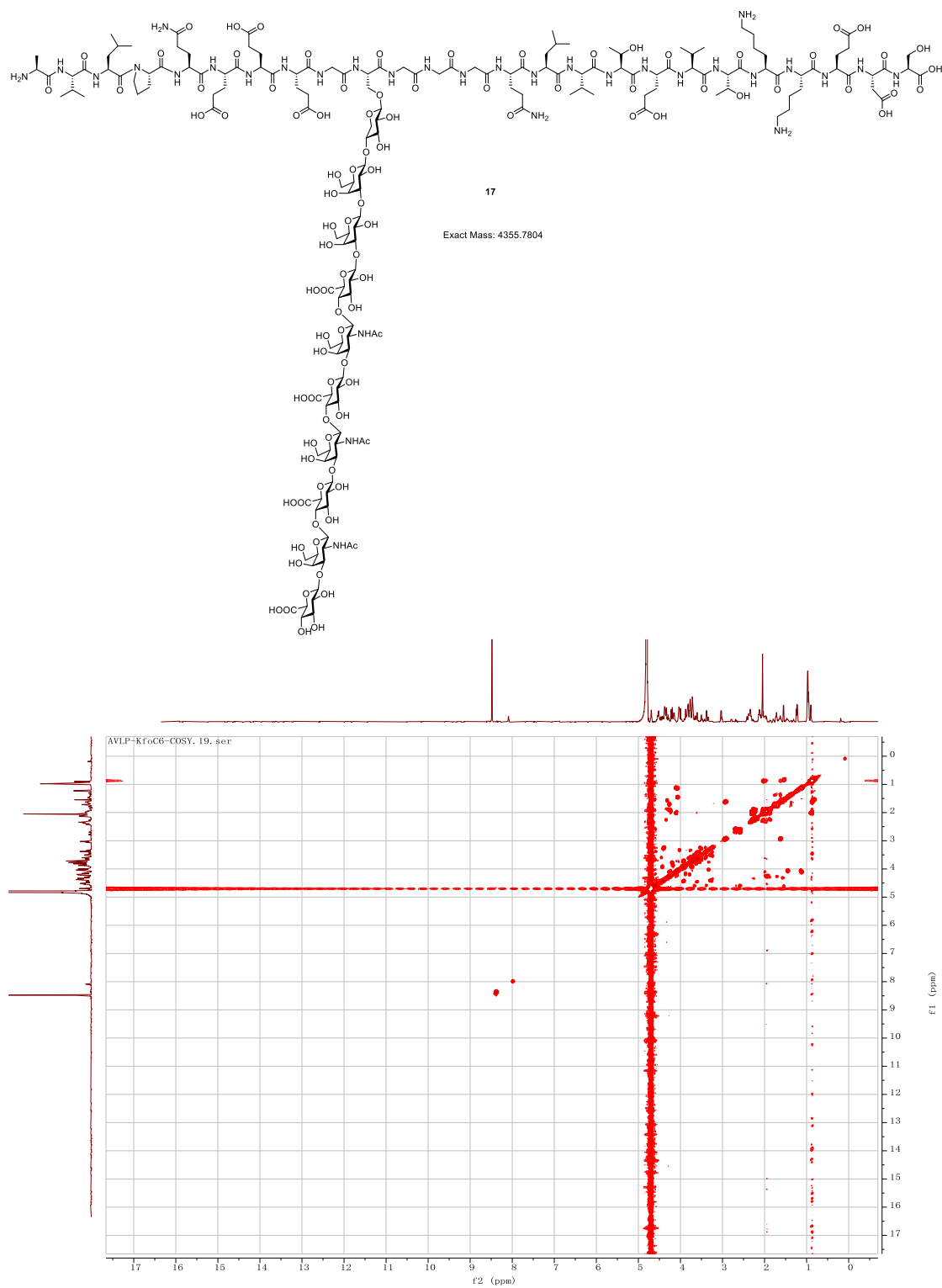


Figure 2.80. COSY of **17** (800 MHz, D₂O).

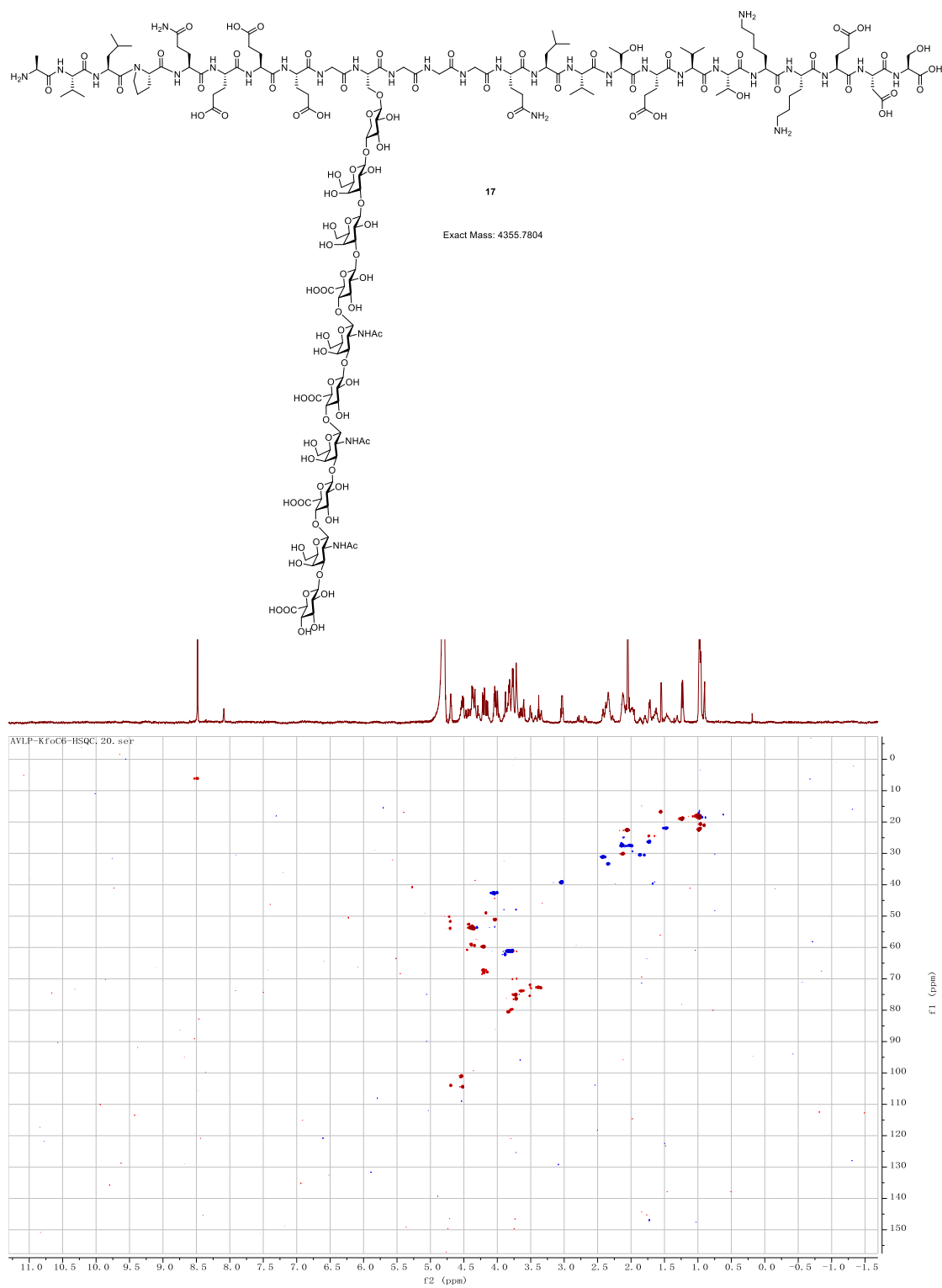


Figure 2.81. HSQC of **17** (800 MHz, D₂O).

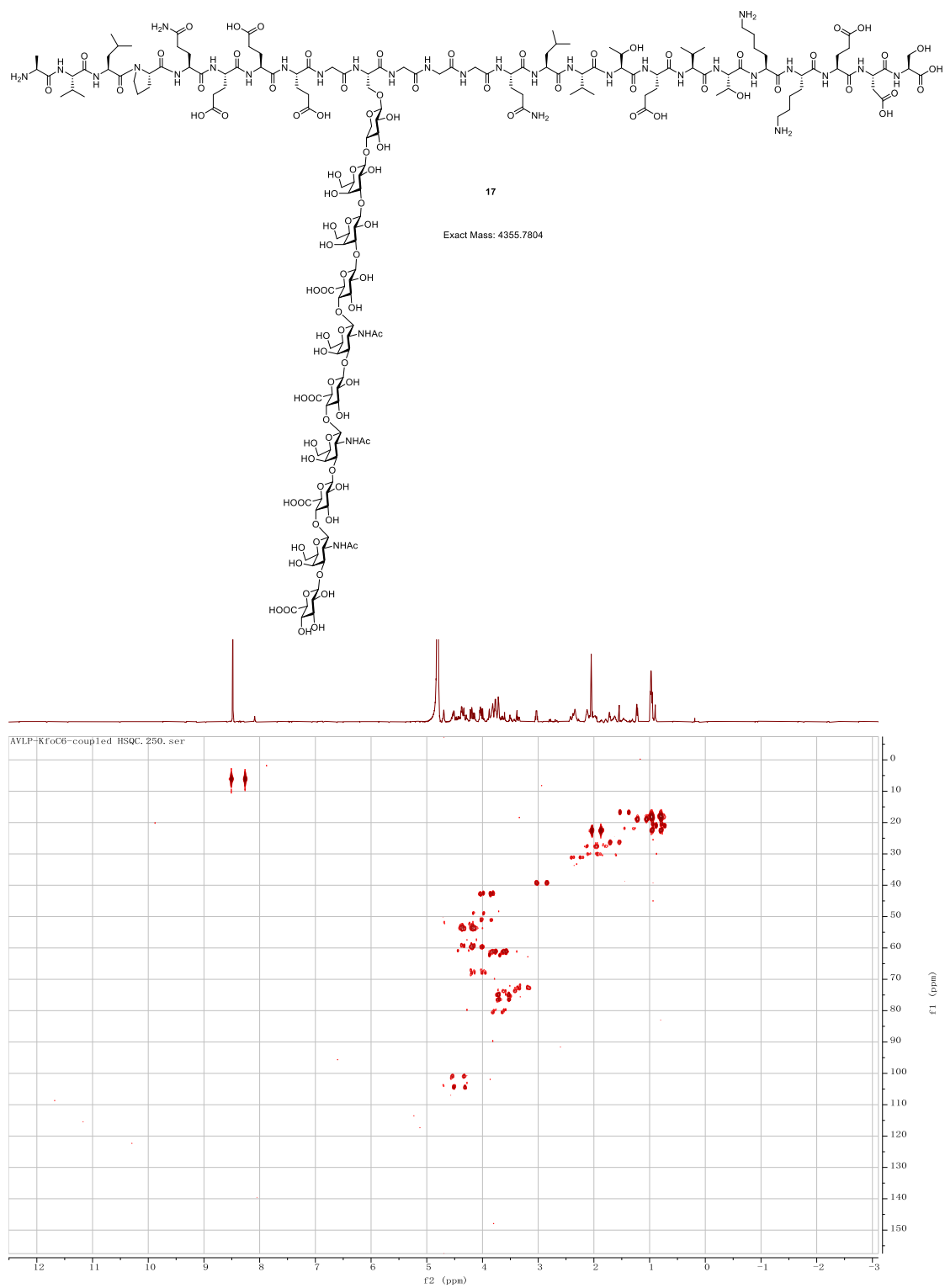


Figure 2.82. Coupled HSQC of 17 (800 MHz, D₂O).

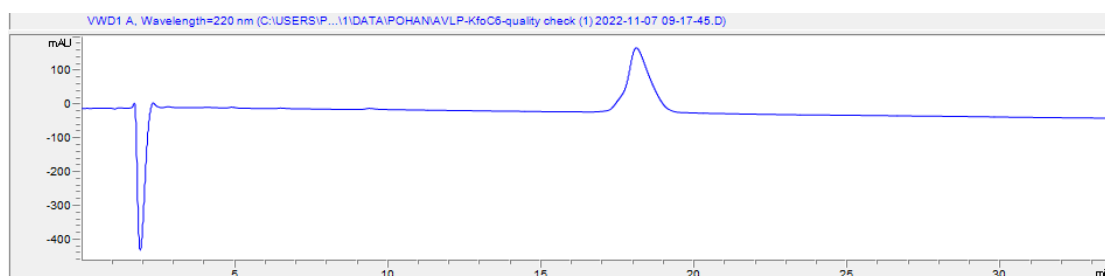
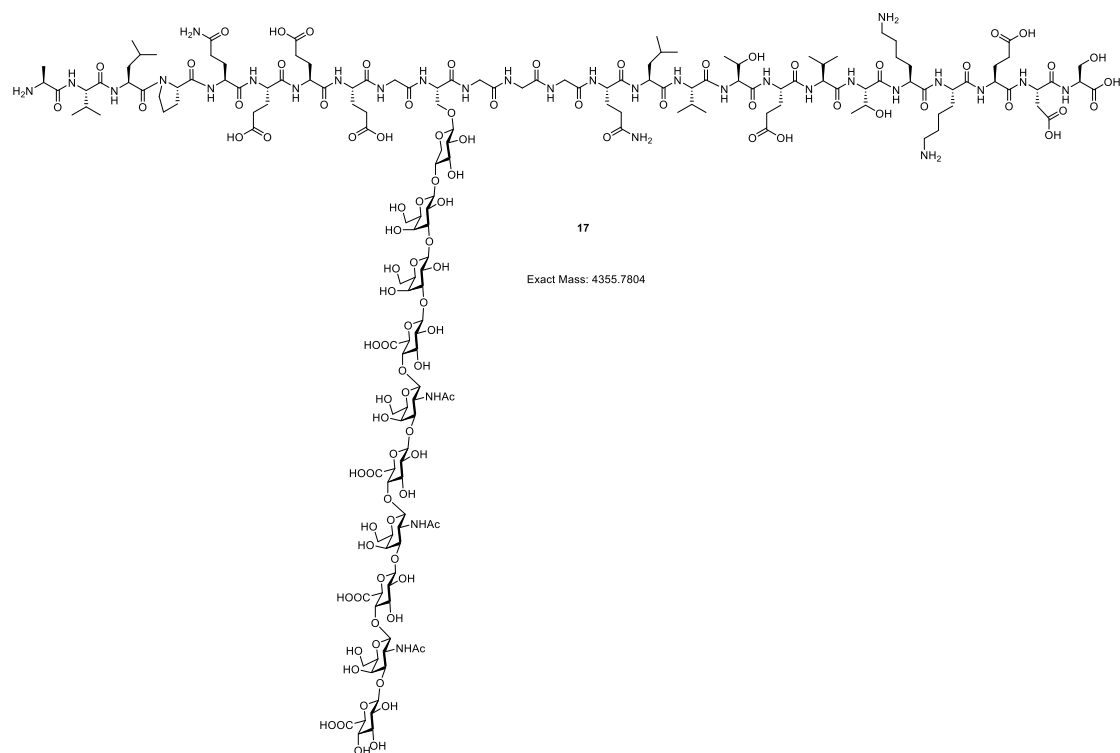
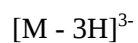
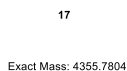


Figure 2.83. HPLC Chromatogram of 17.



161

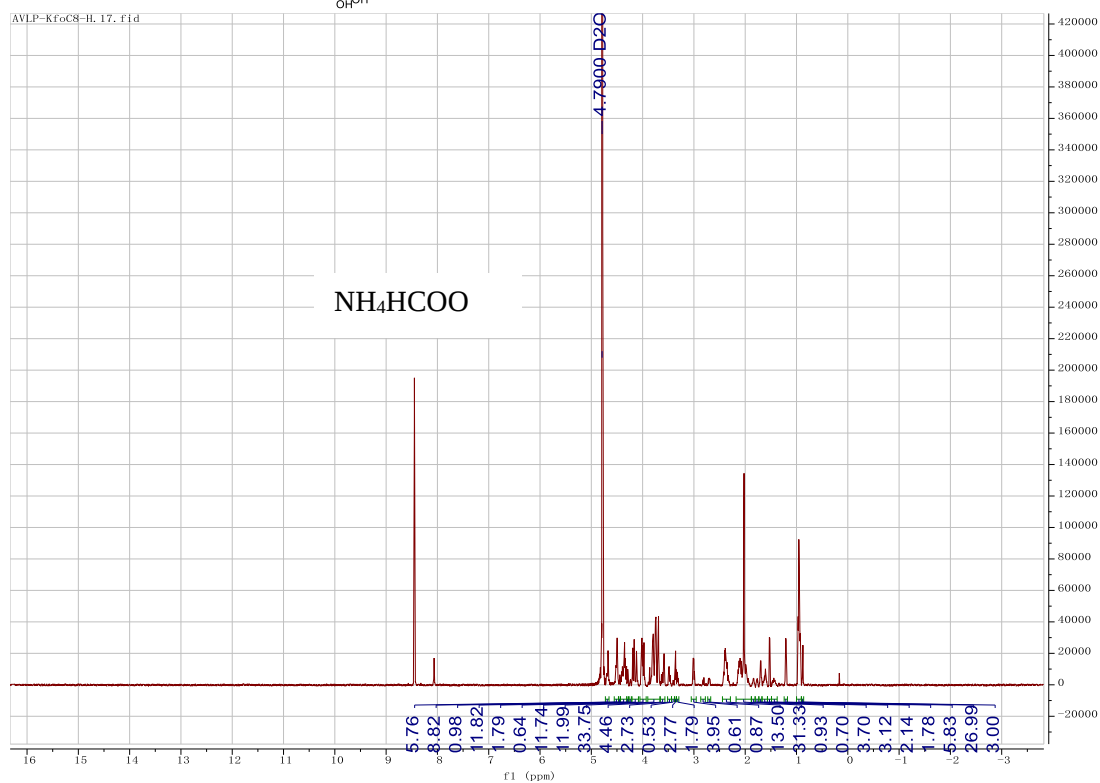
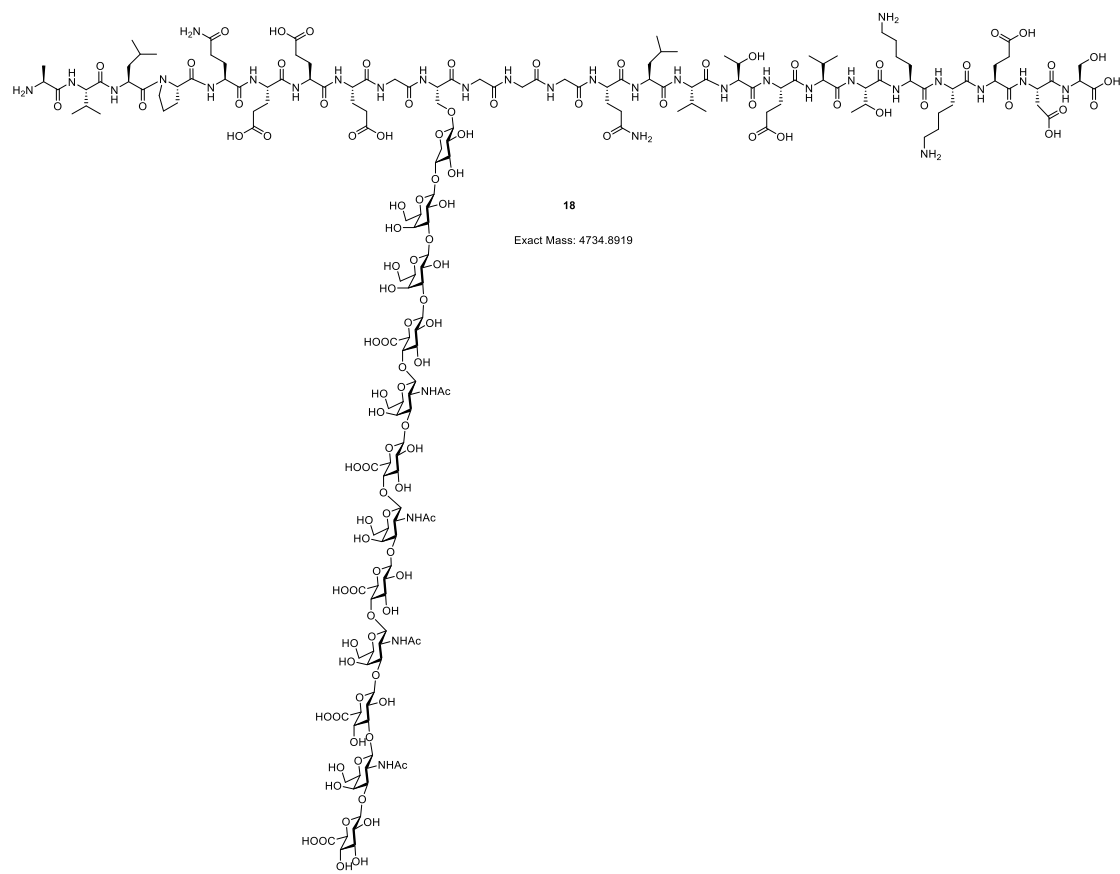


Figure 2.85. ¹H-NMR of **18** (800 MHz, D₂O).

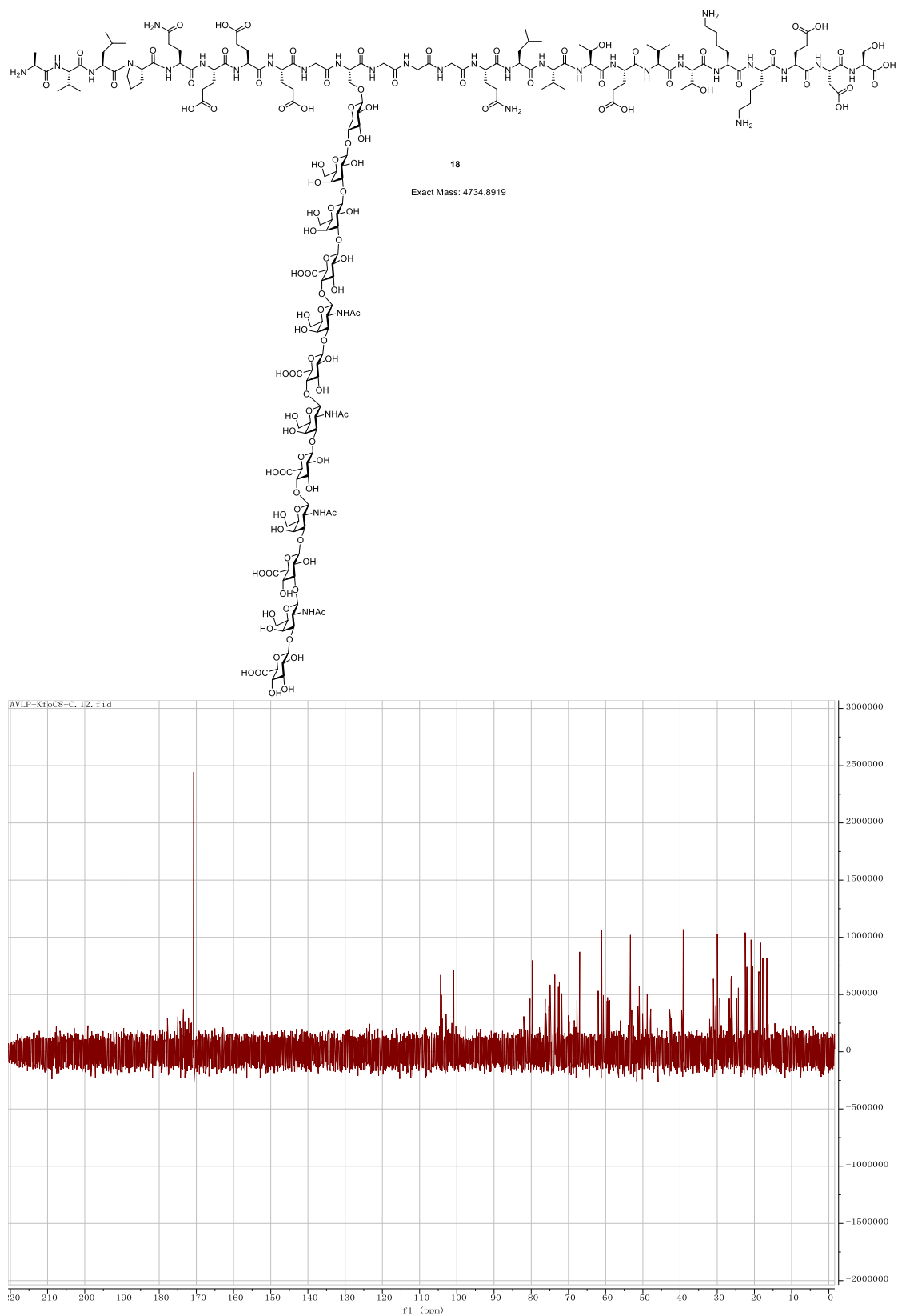


Figure 2.86. ^{13}C NMR of **18** (201 MHz, D_2O).

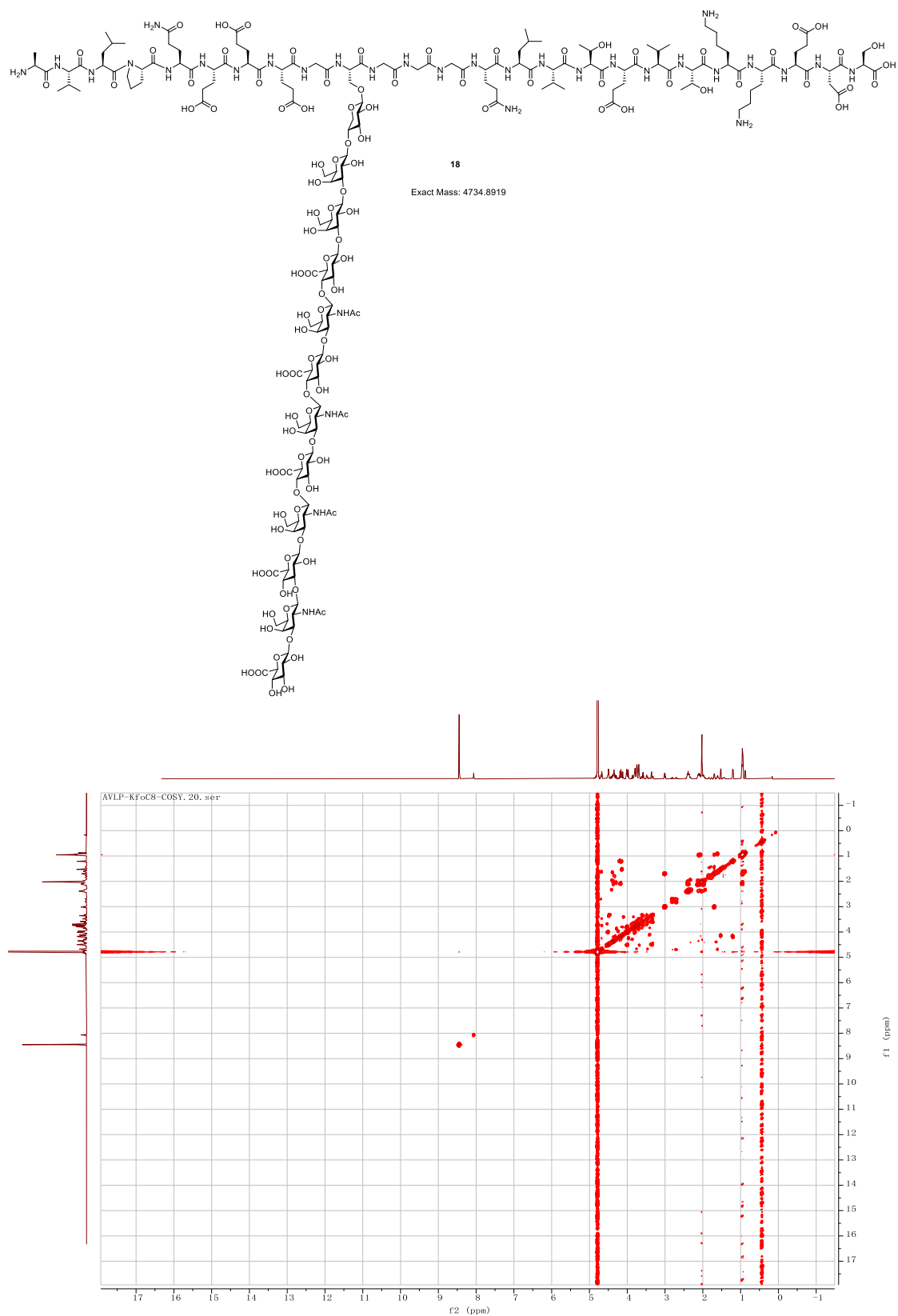


Figure 2.87. COSY of **18** (800 MHz, D₂O).

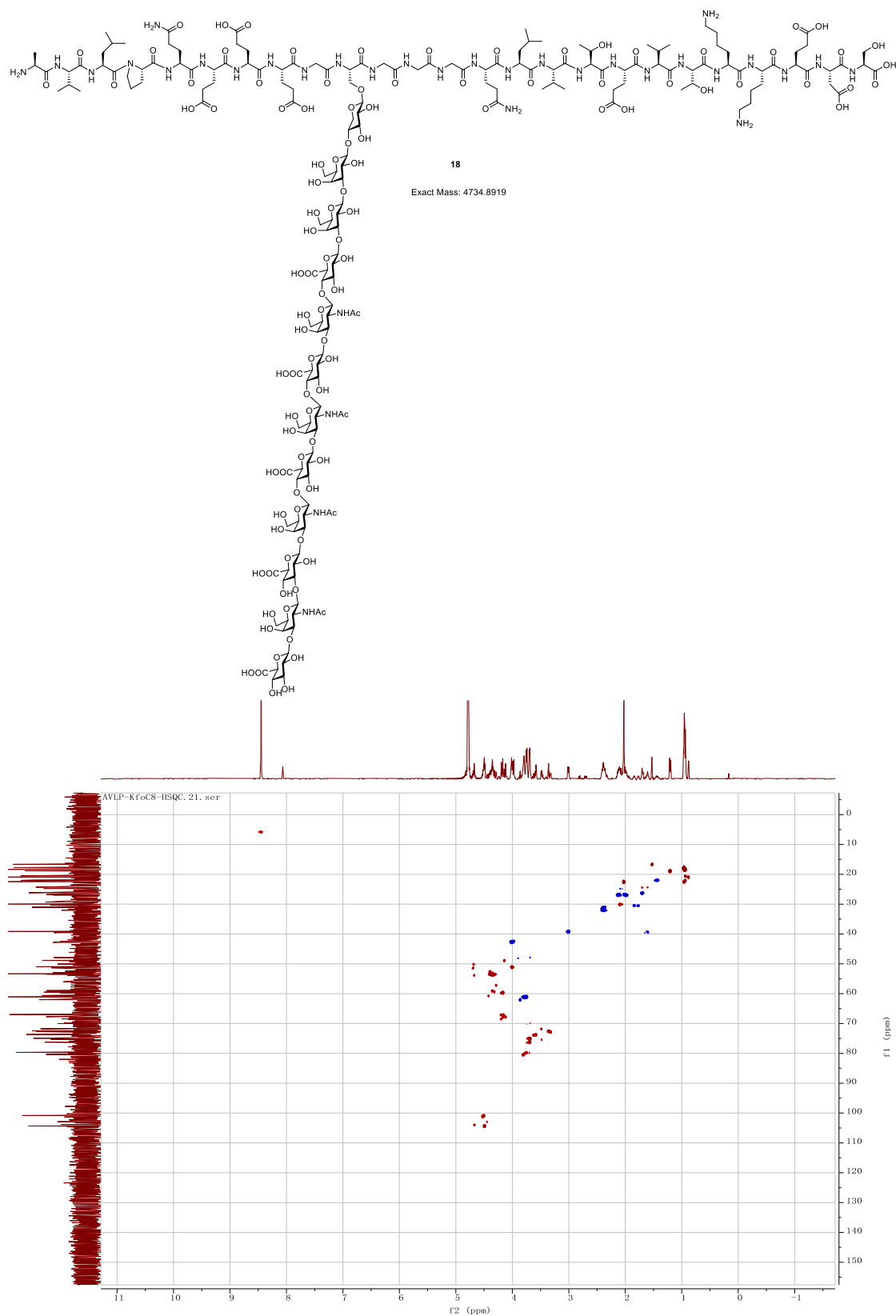


Figure 2.88. HSQC of **18** (800 MHz, D₂O).

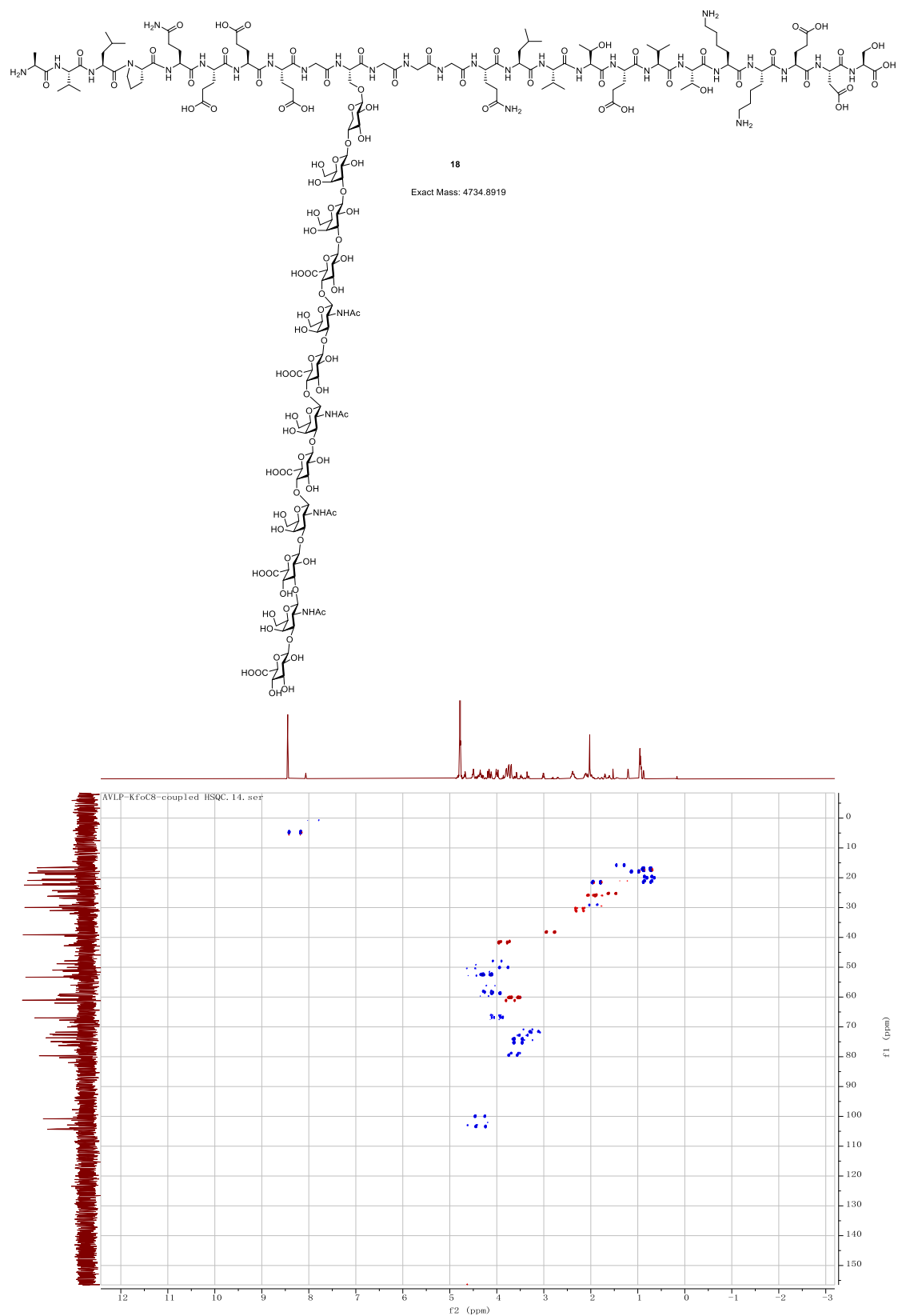


Figure 2.89. Coupled HSQC of **18** (800 MHz, D₂O).

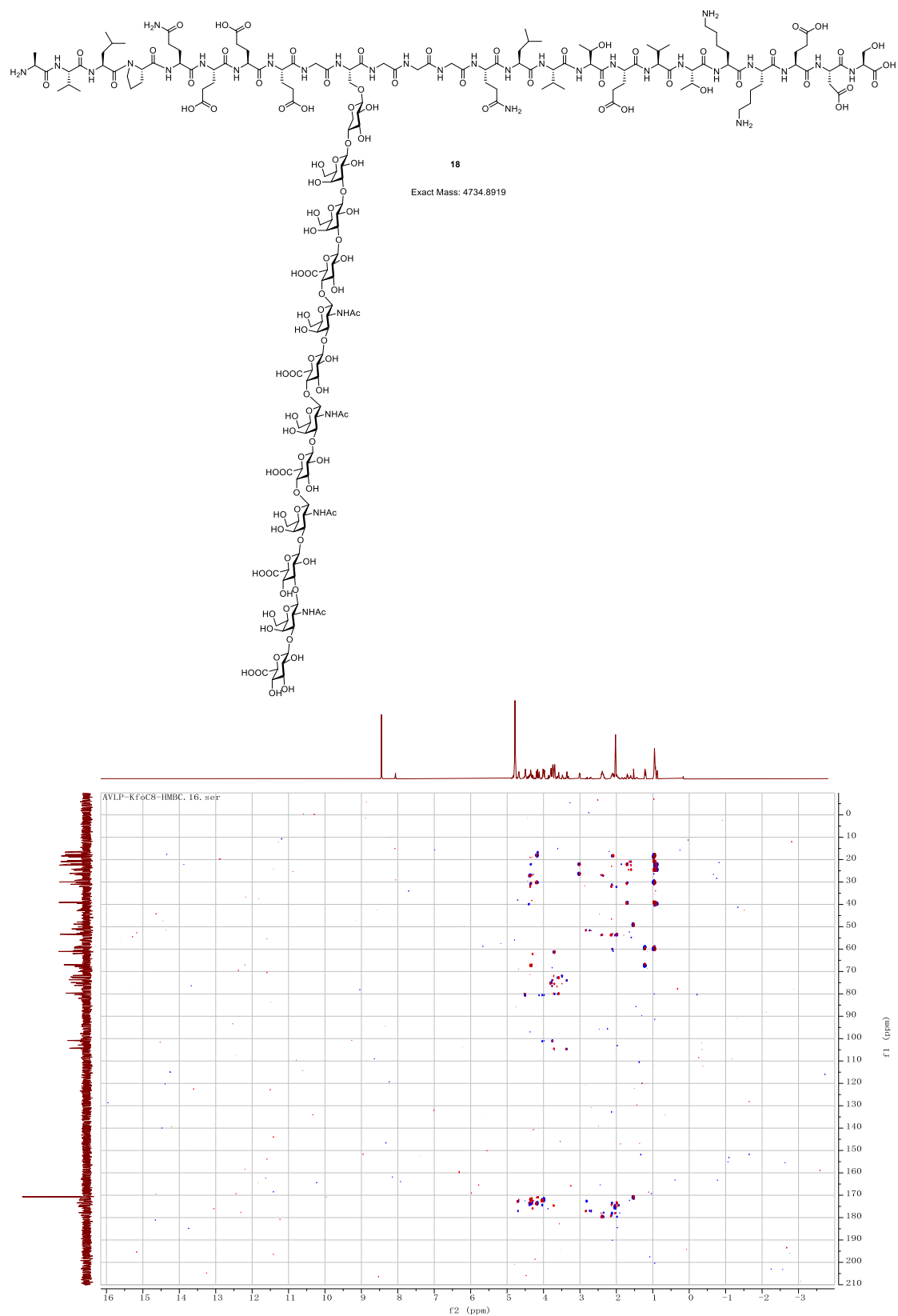


Figure 2.90. HMBC of **18** (800 MHz, D₂O).

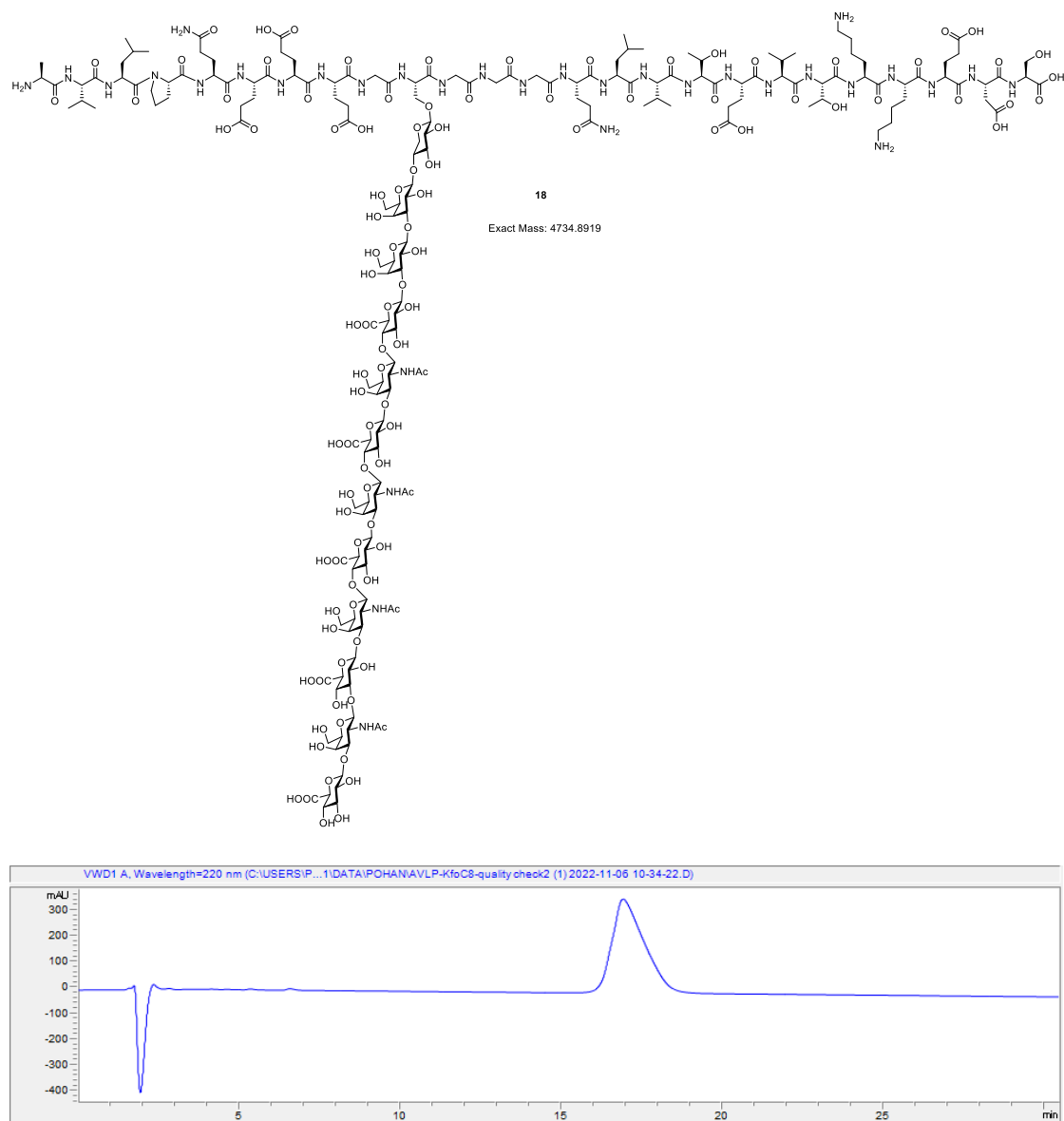
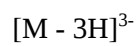
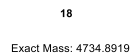


Figure 2.91. HPLC Chromatogram of **18**.



169

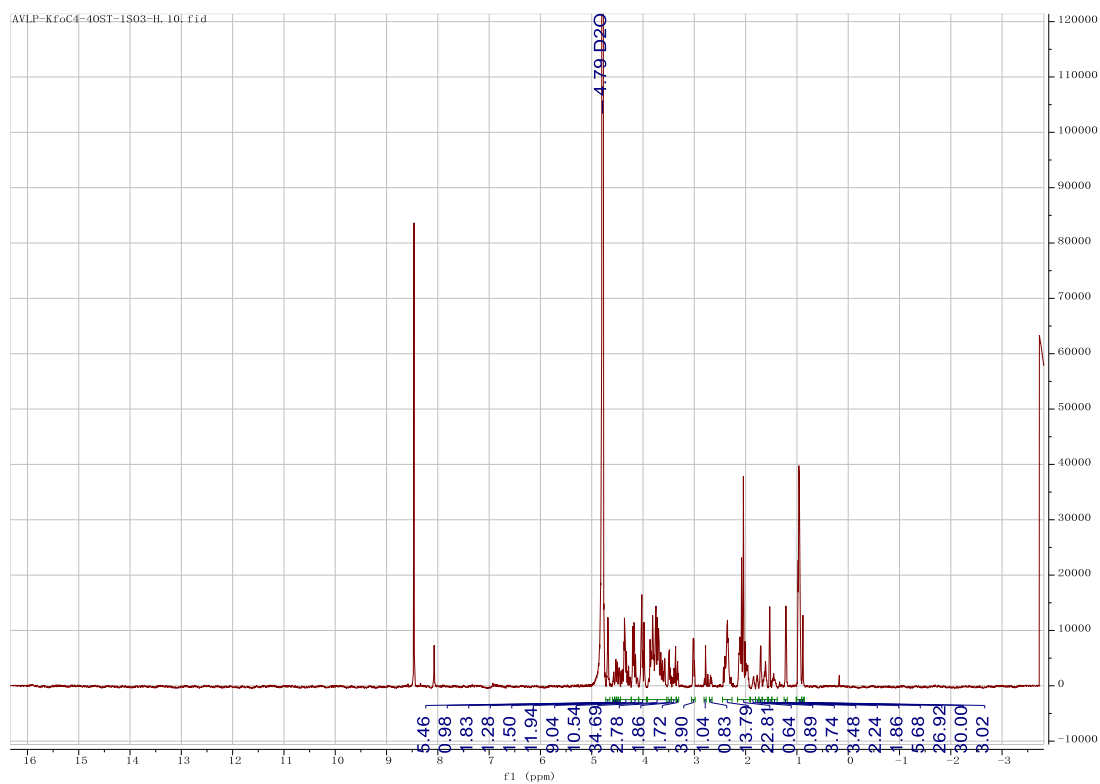
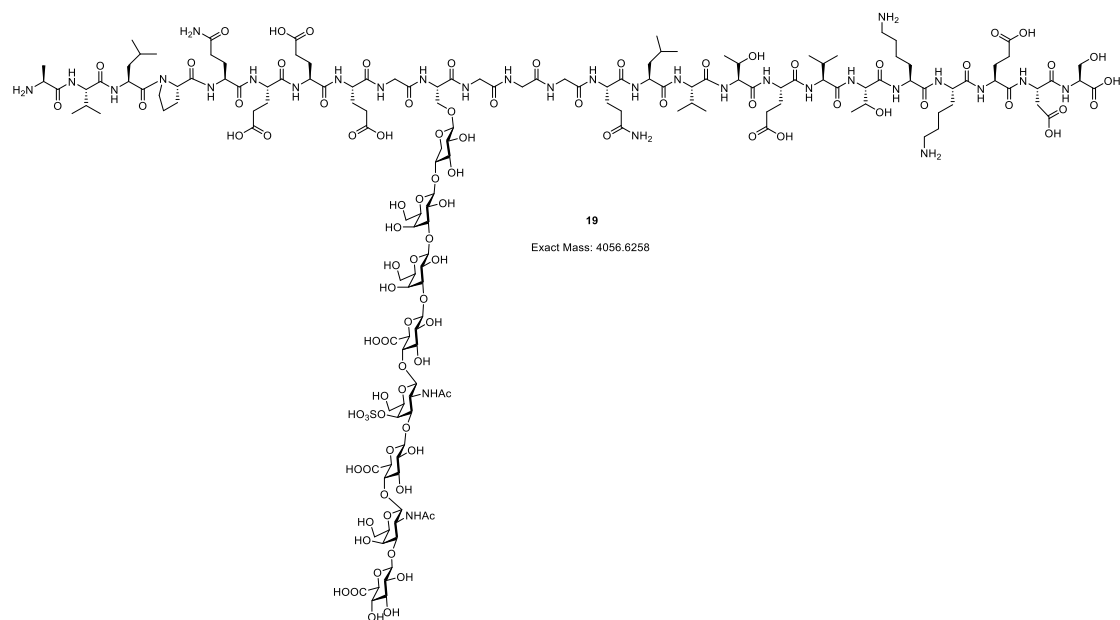


Figure 2.93. ^1H -NMR of **19** (800 MHz, D_2O).

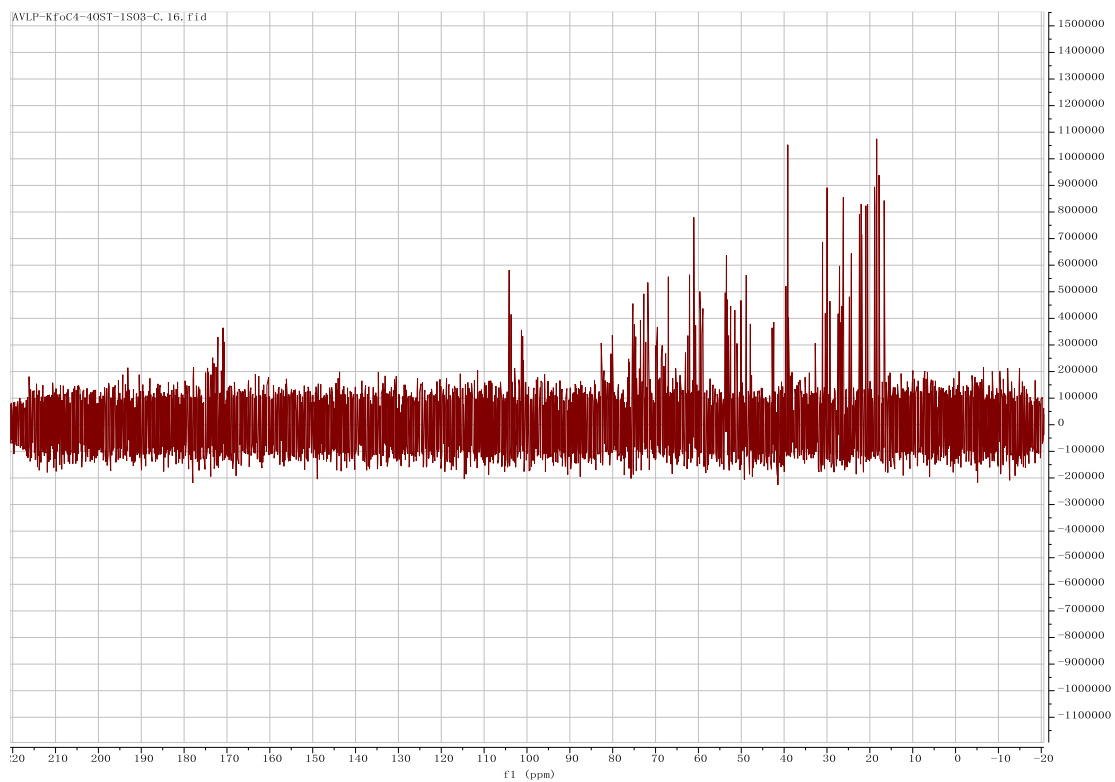
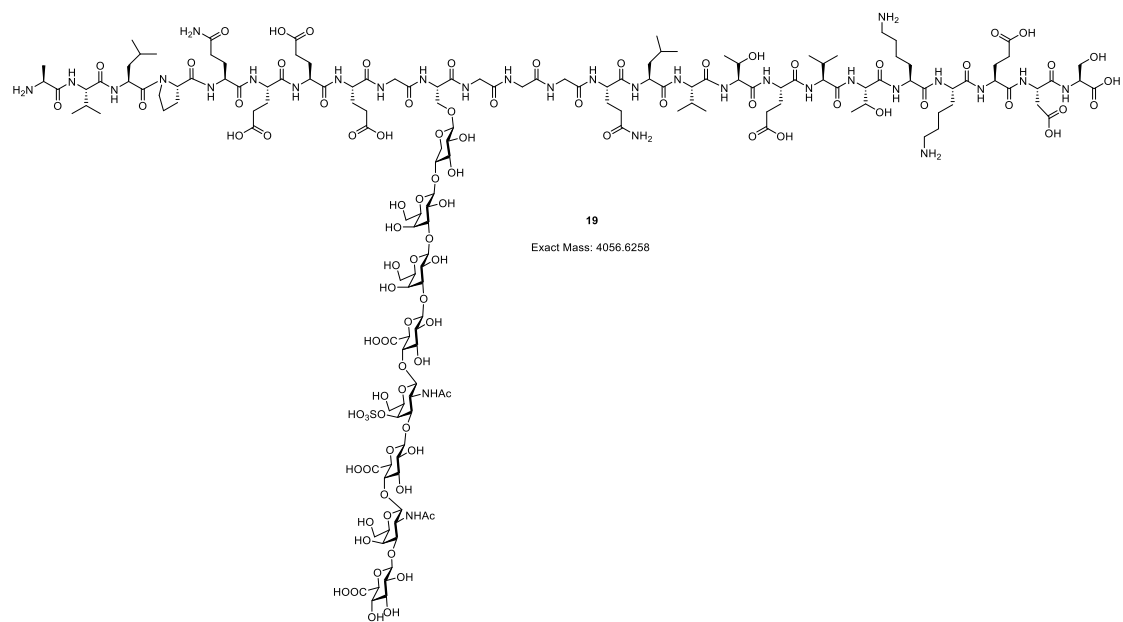


Figure 2.94. ^{13}C NMR of **19** (201 MHz, D_2O).

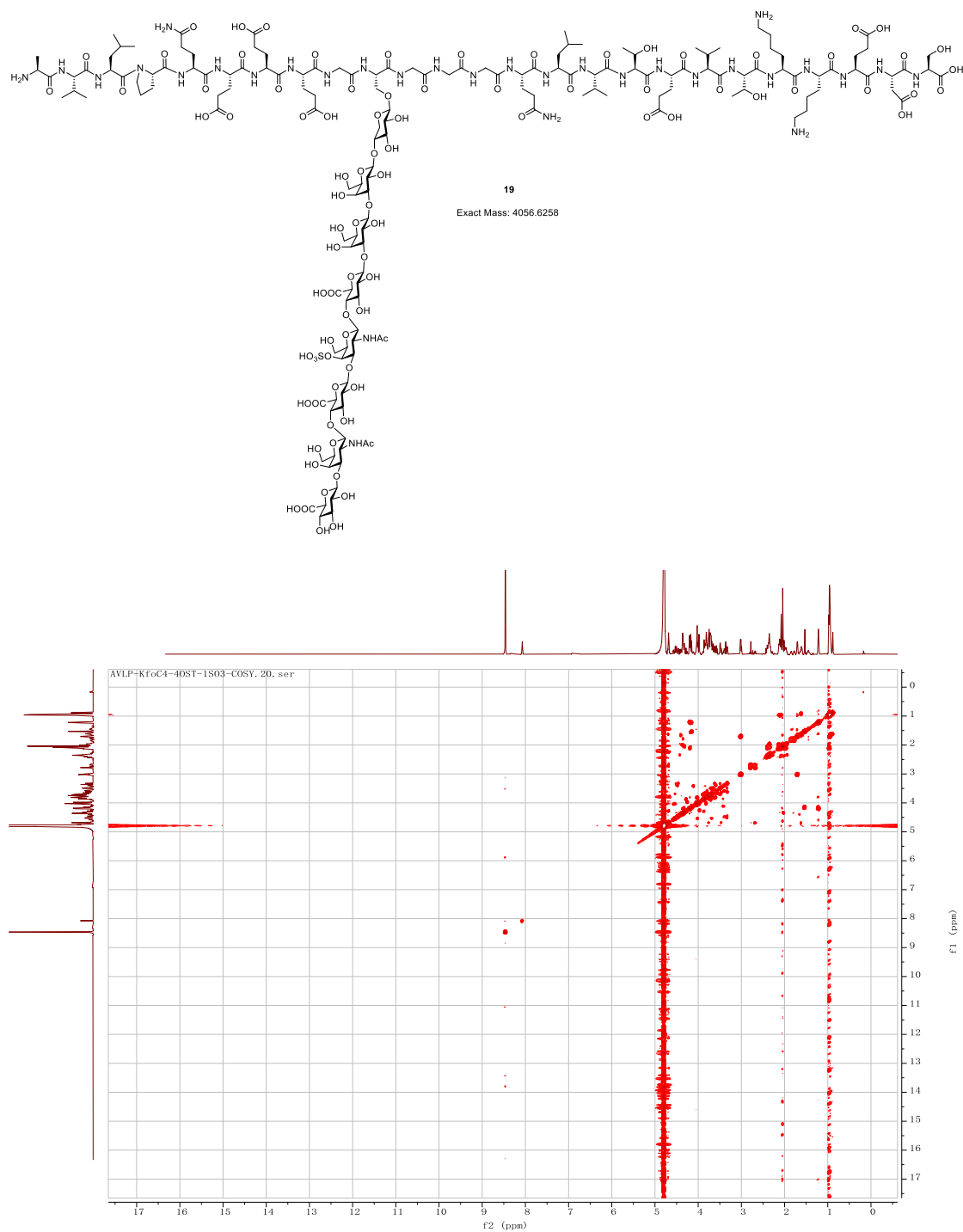


Figure 2.95. COSY of **19** (800 MHz, D_2O).

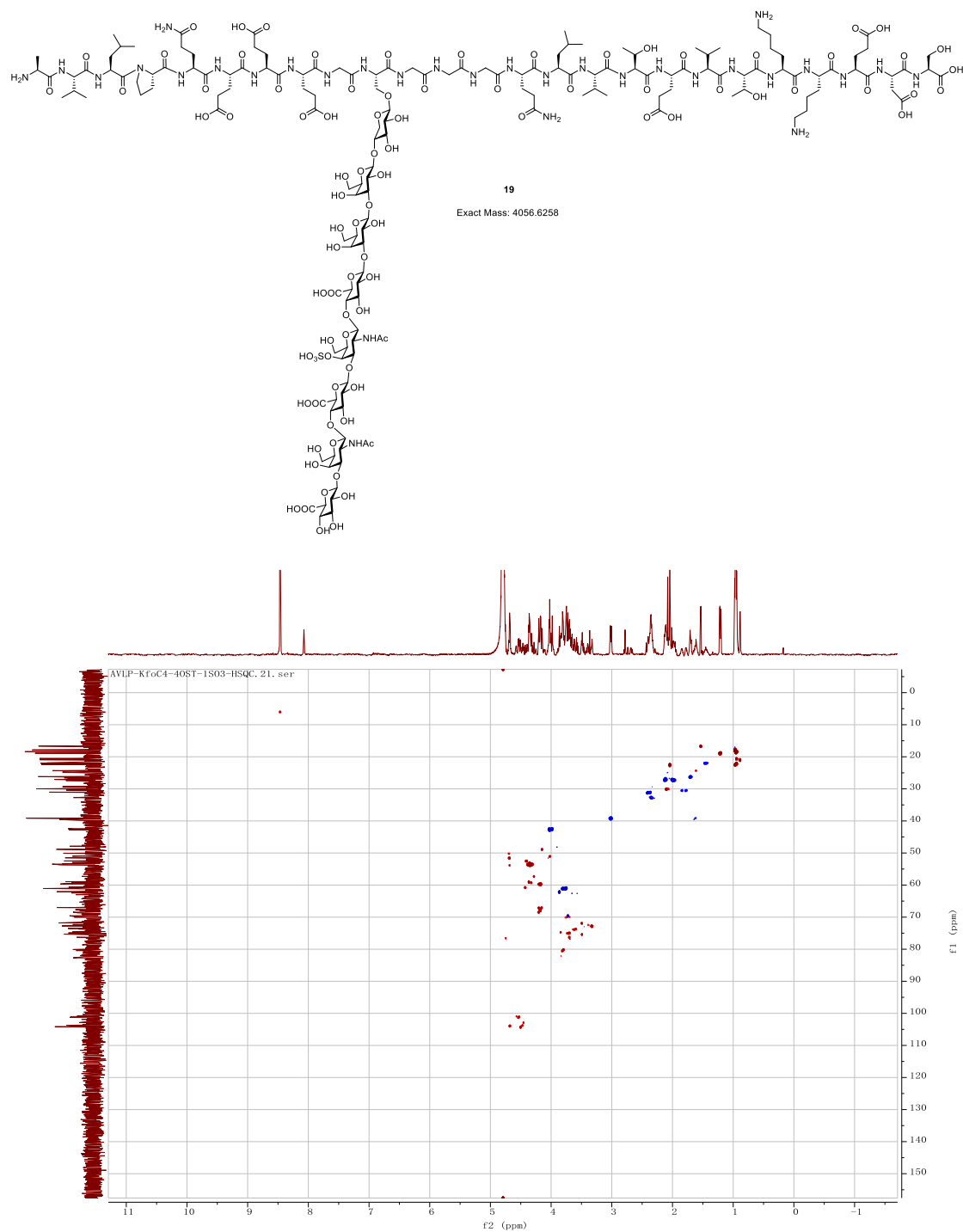


Figure 2.96. HSQC of **19** (800 MHz, D₂O).

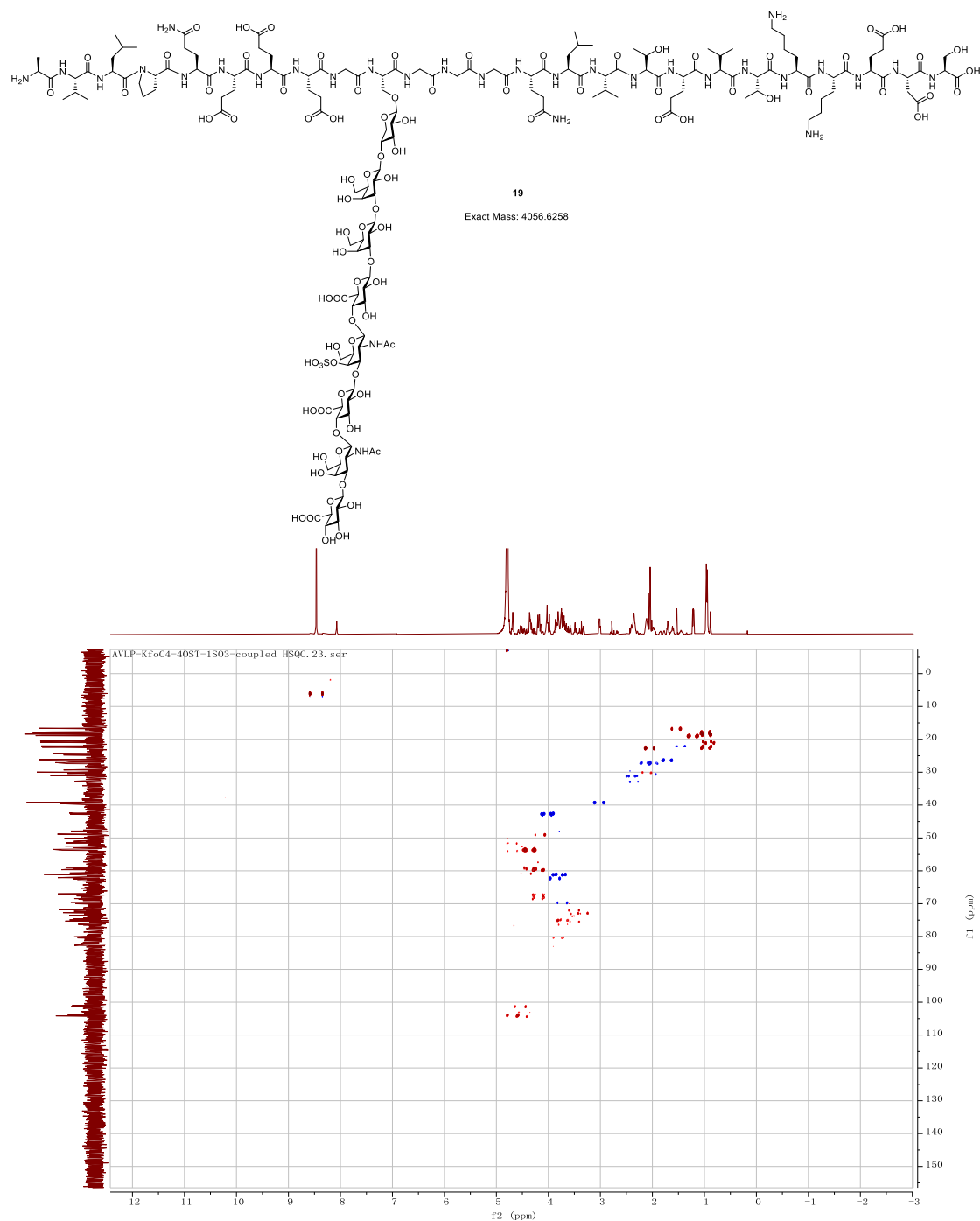


Figure2.97. Coupled HSQC of **19** (800 MHz, D₂O).

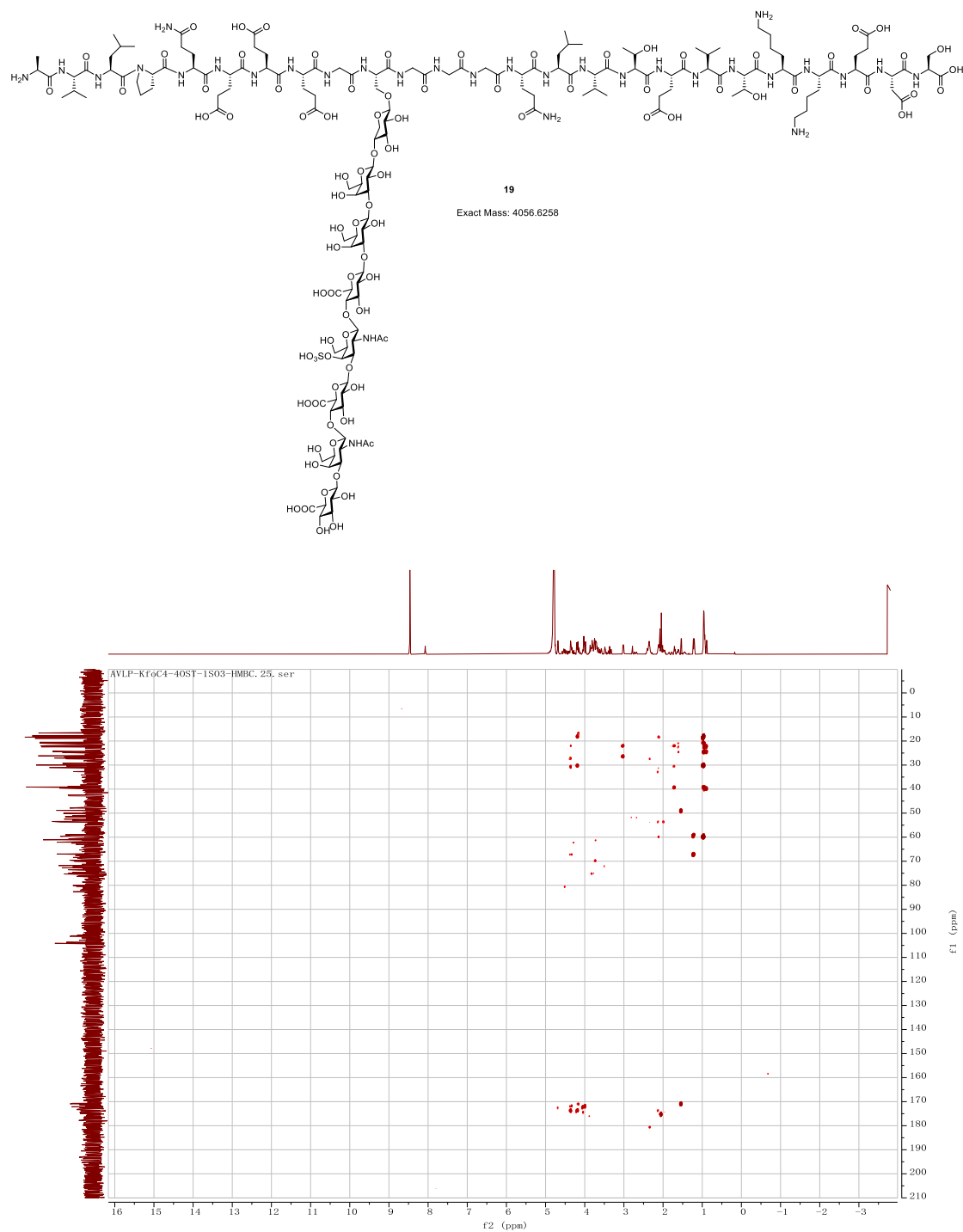


Figure 2.98. HMBC of **19** (800 MHz, D₂O).

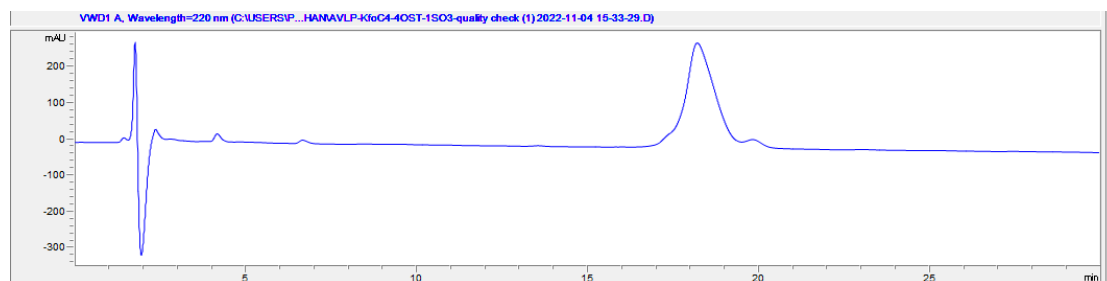
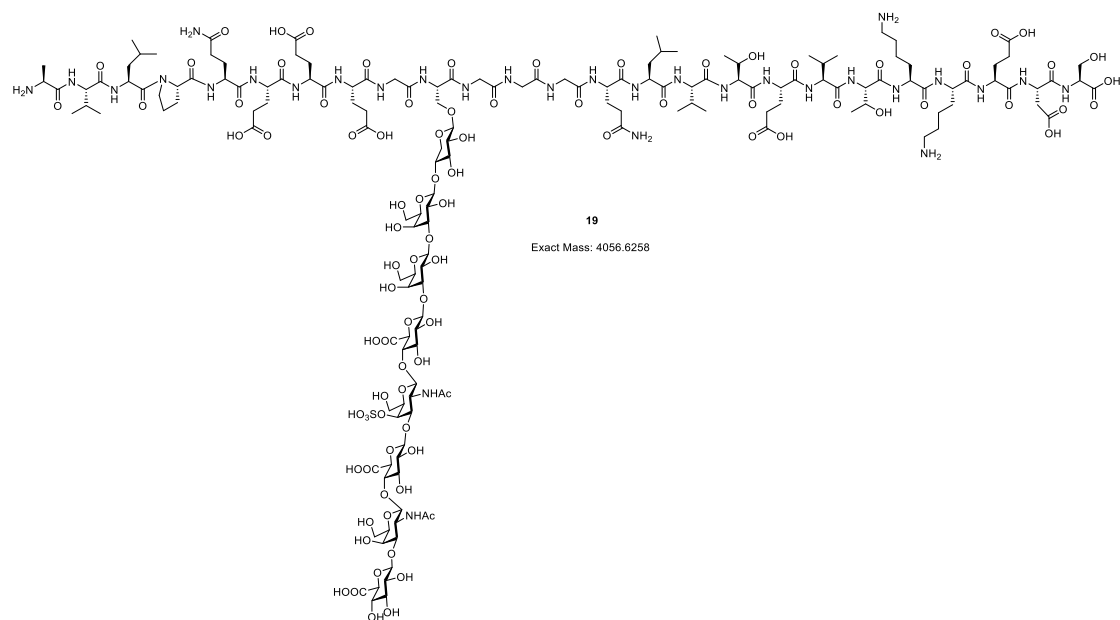


Figure 2.99. HPLC Chromatogram of **19**.

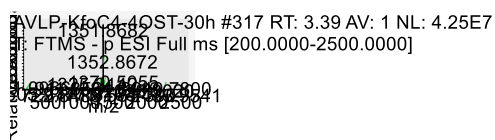
 $[M - 3H]^{3-}$

Figure 2.100. LCMS Chromatogram of 19.

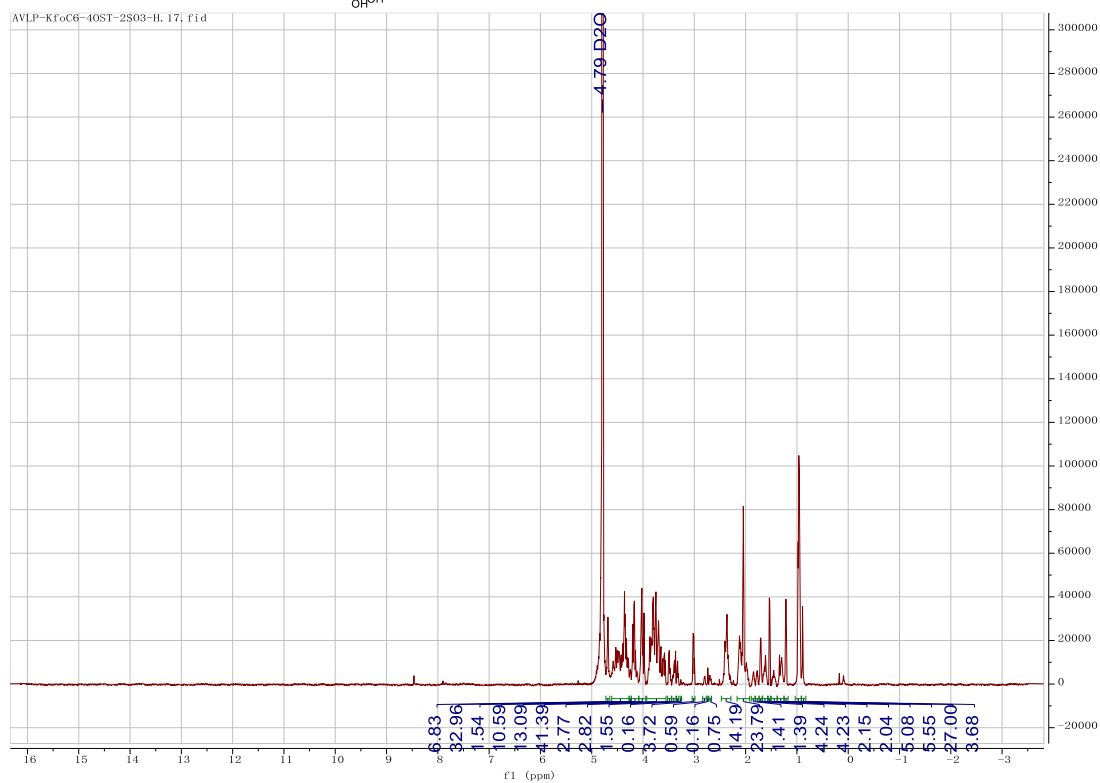
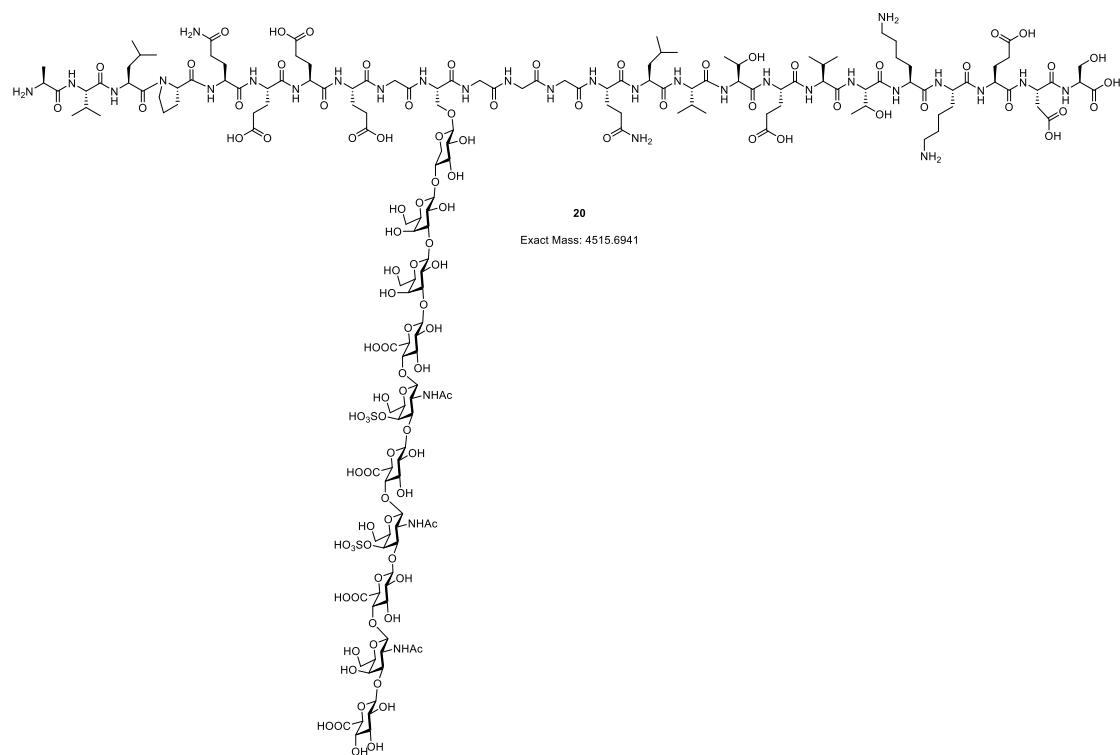


Figure 2.101. ^1H -NMR of **20** (800 MHz, D_2O).

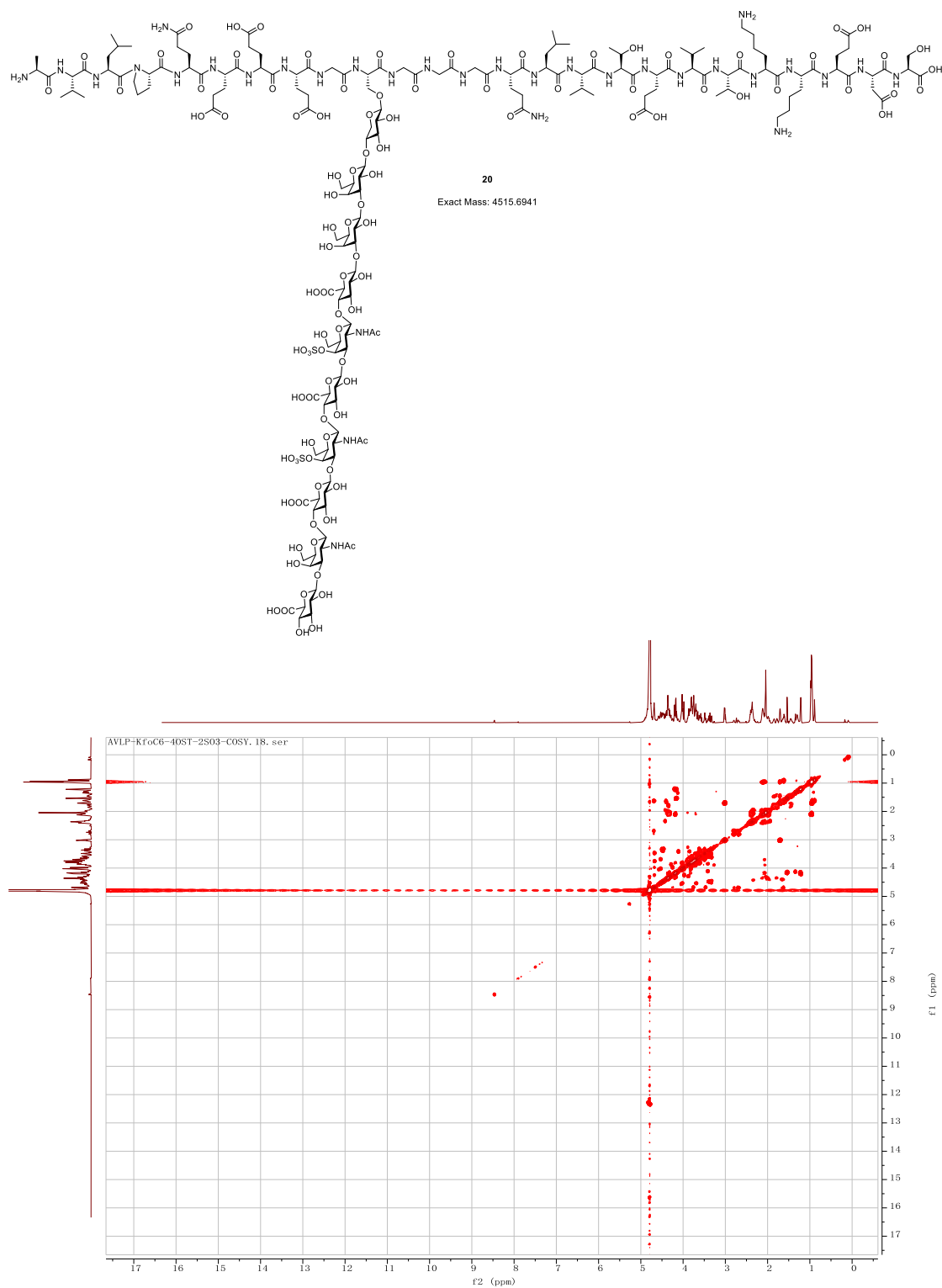


Figure 2.102. COSY of **20** (800 MHz, D₂O).

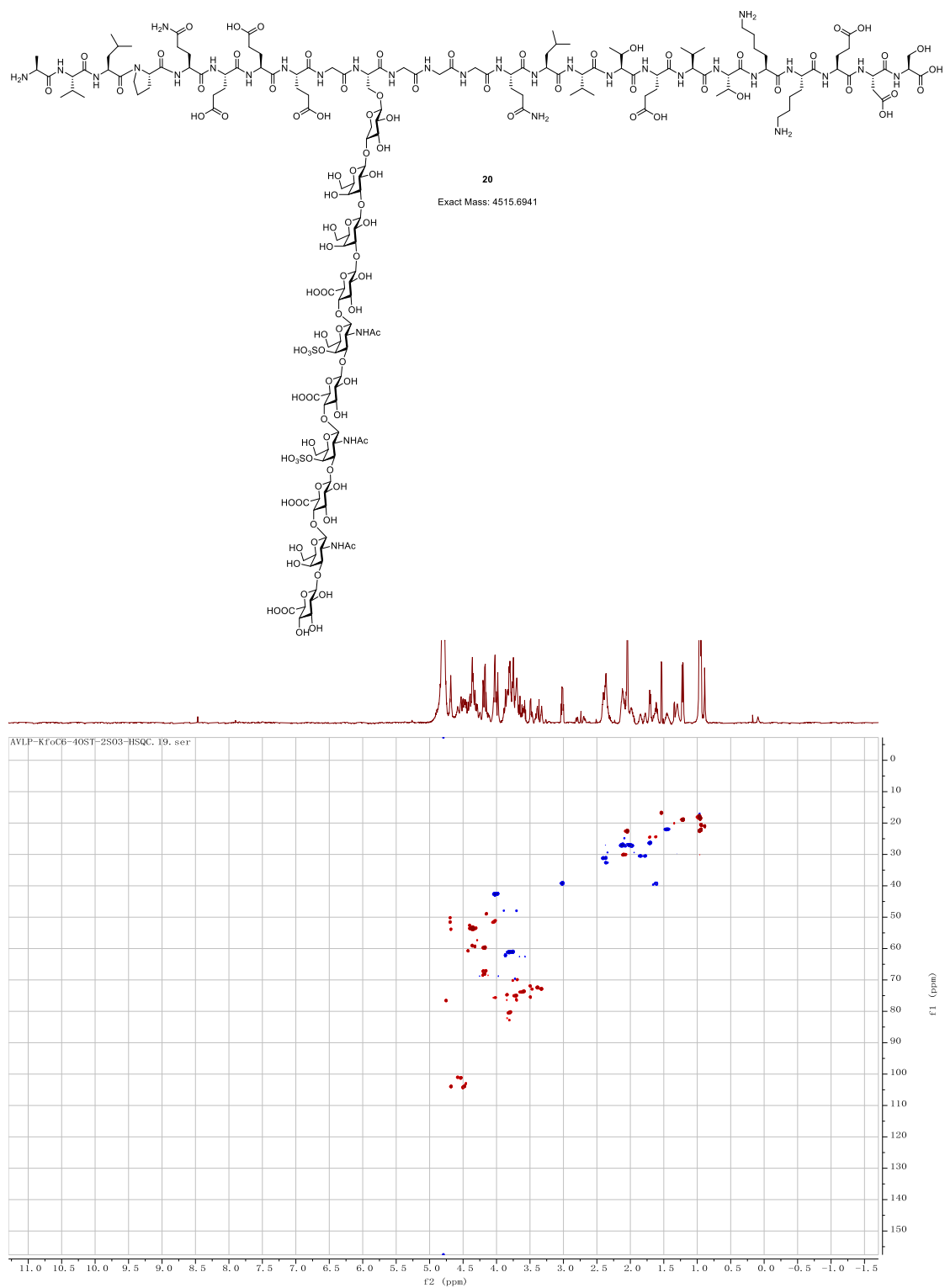


Figure 2.103. HSQC of **20** (800 MHz, D₂O).

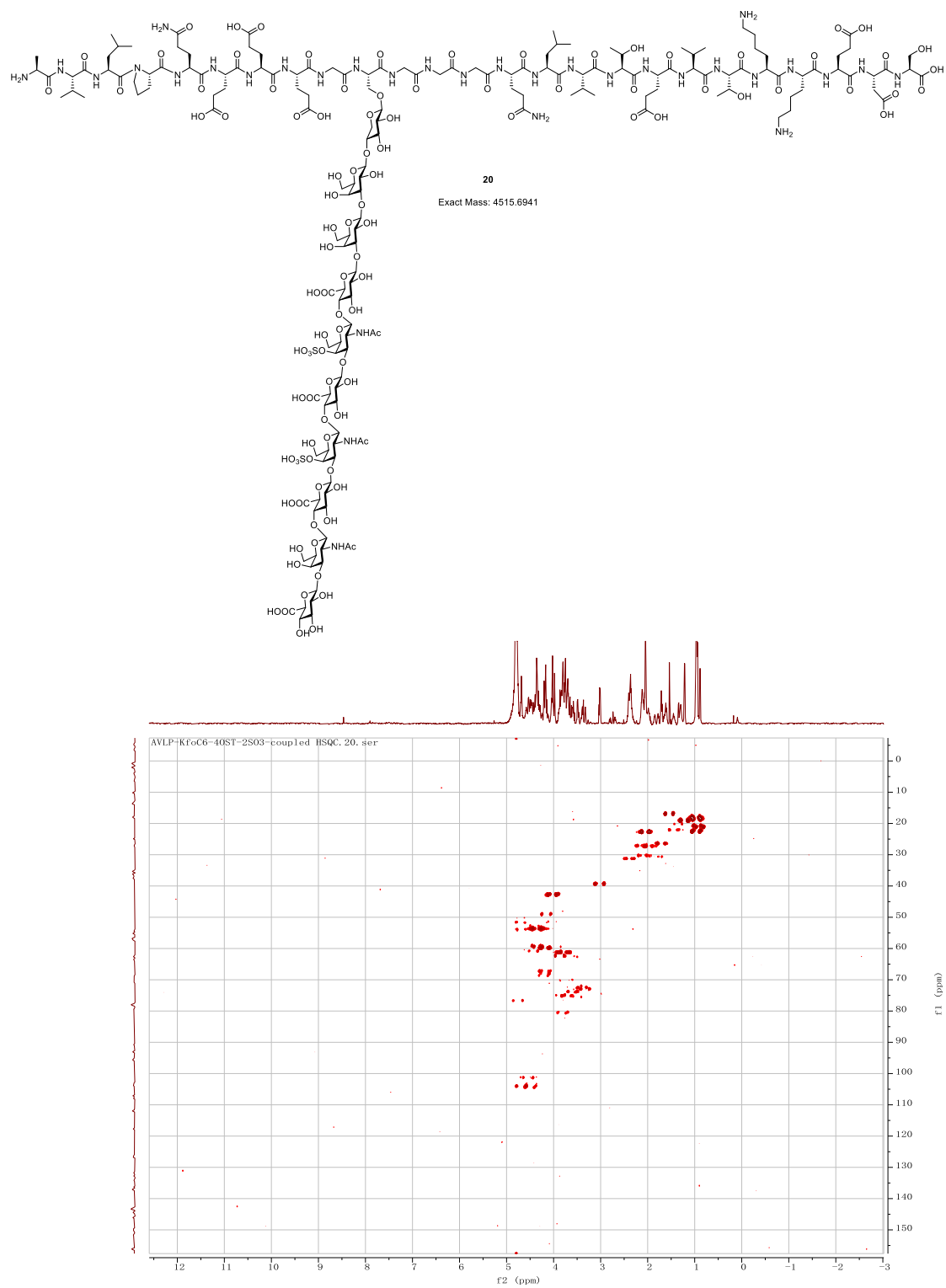


Figure 2.104. Coupled HSQC of **20** (800 MHz, D₂O).

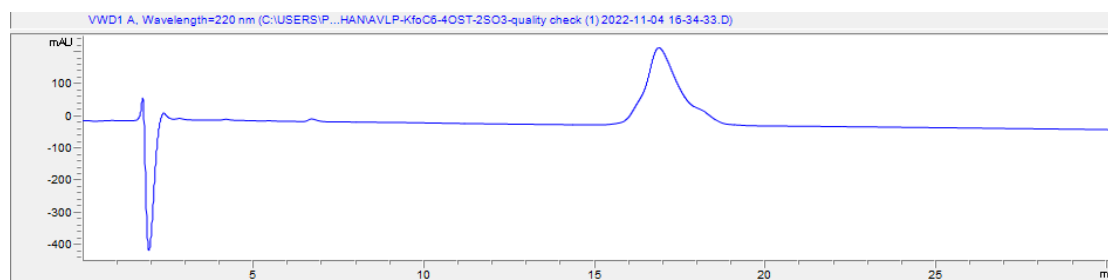
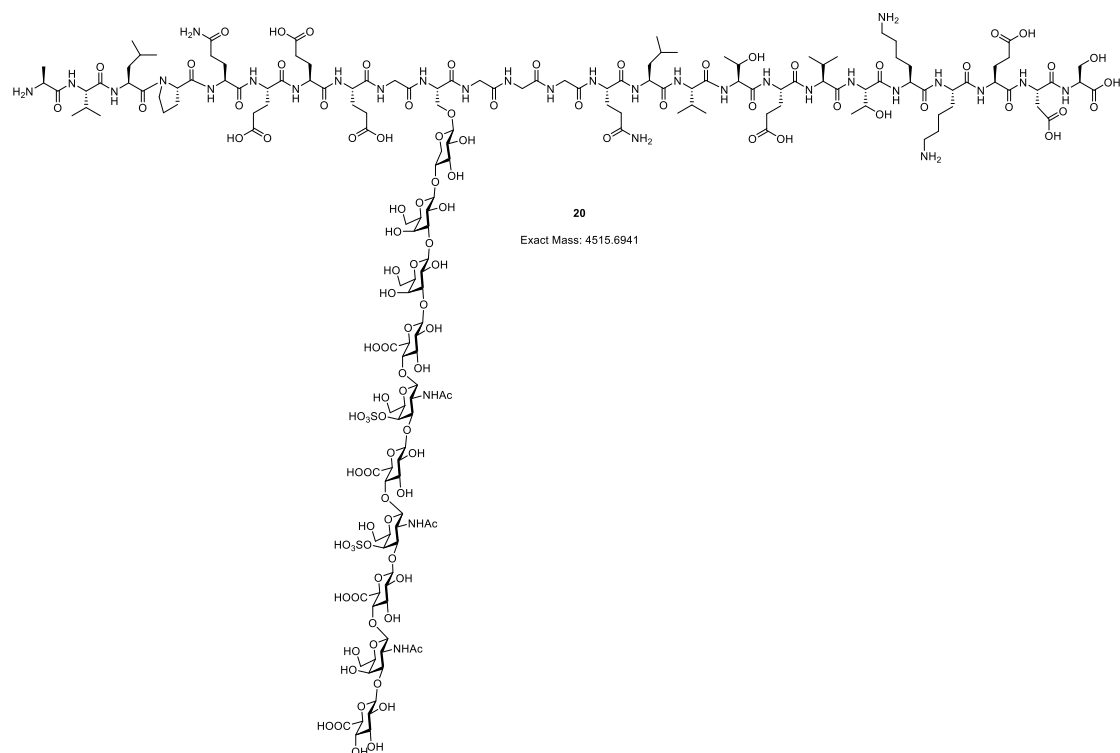
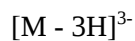
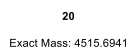


Figure 2.105. HPLC Chromatogram of **20**.



183

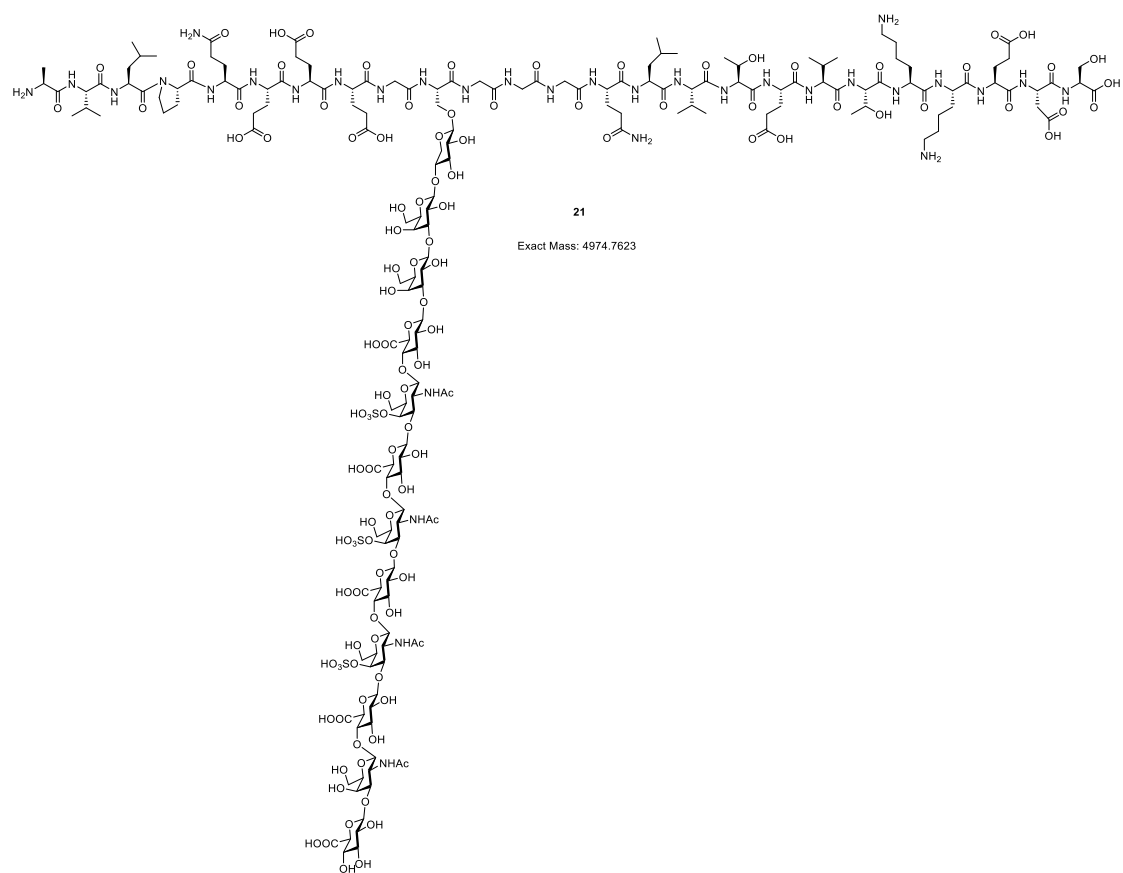
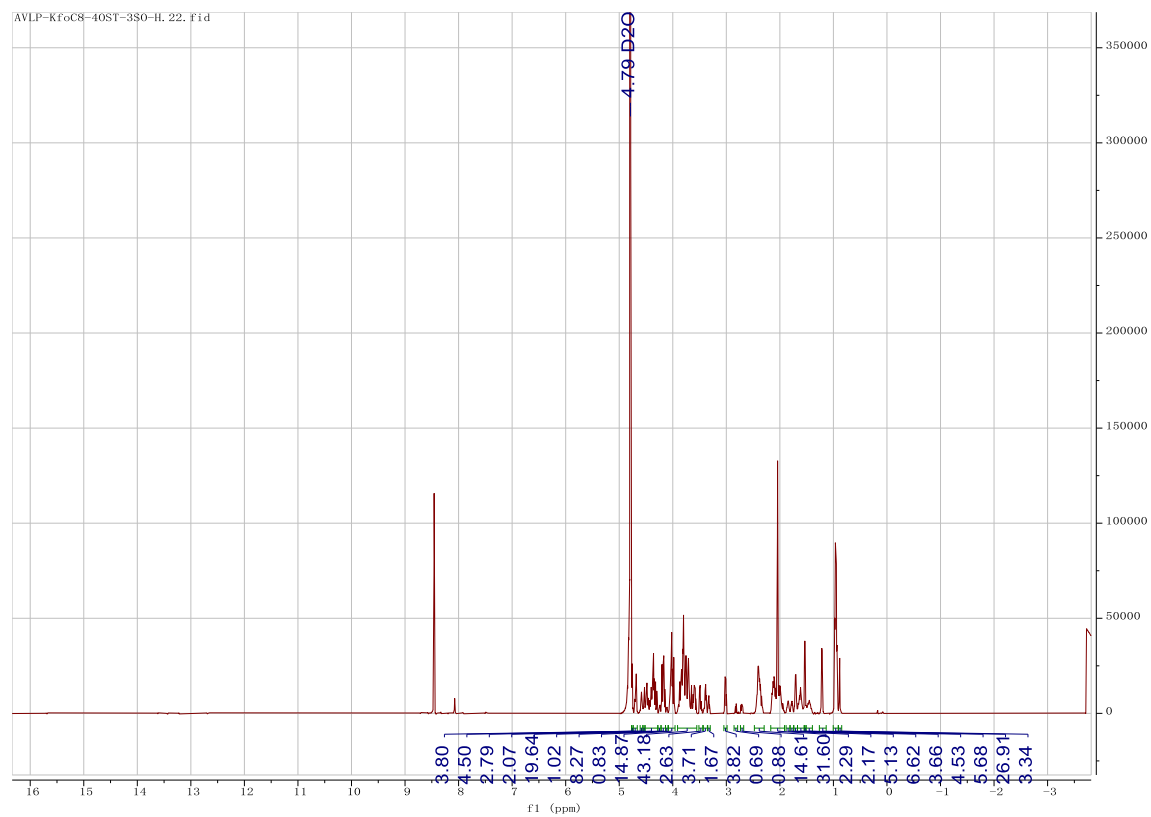


Figure 2.107. ^1H -NMR of **21** (800 MHz, D_2O).

Figure 2.107. (cont'd)



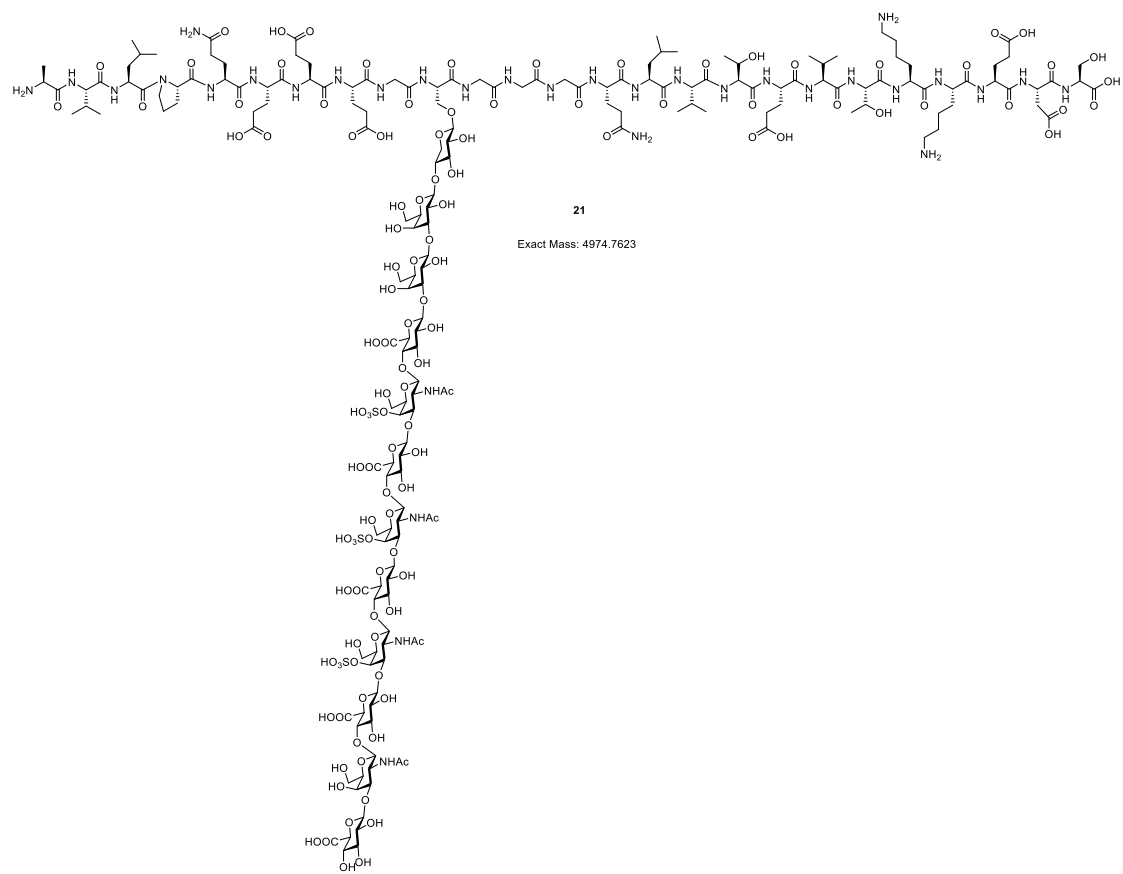
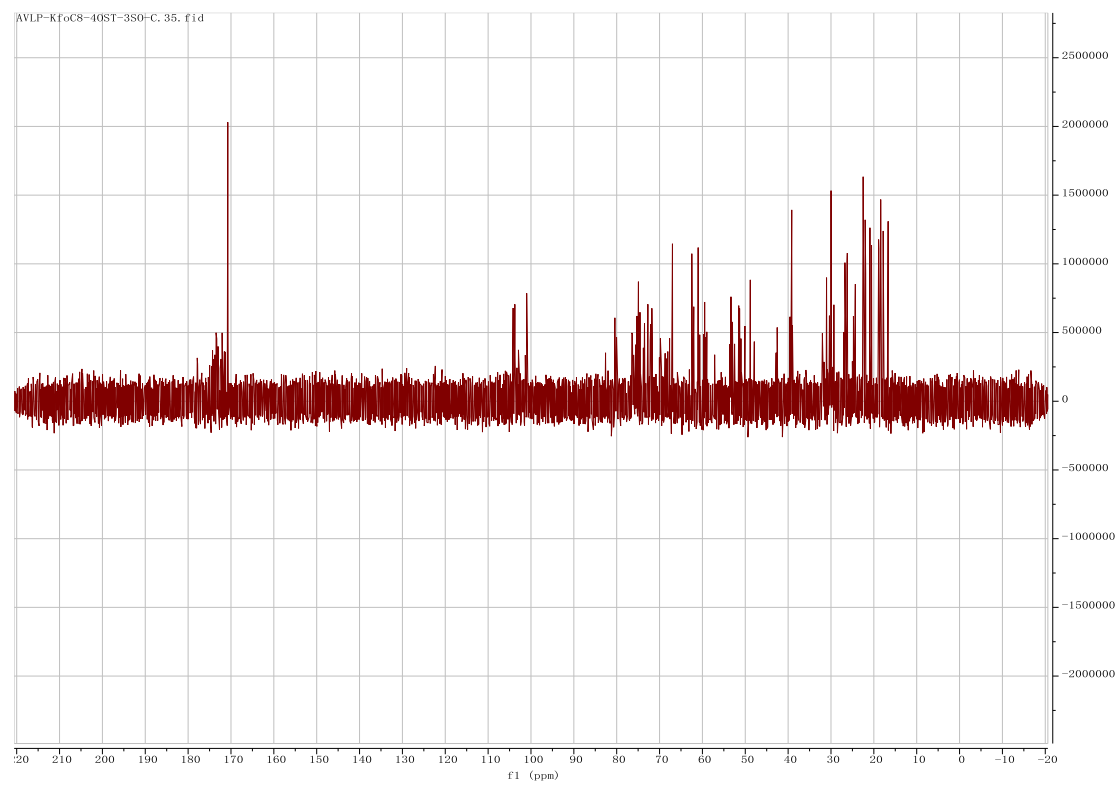


Figure 2.108. ^{13}C NMR of **21** (201 MHz, D_2O).

Figure 2.108. (cont'd)



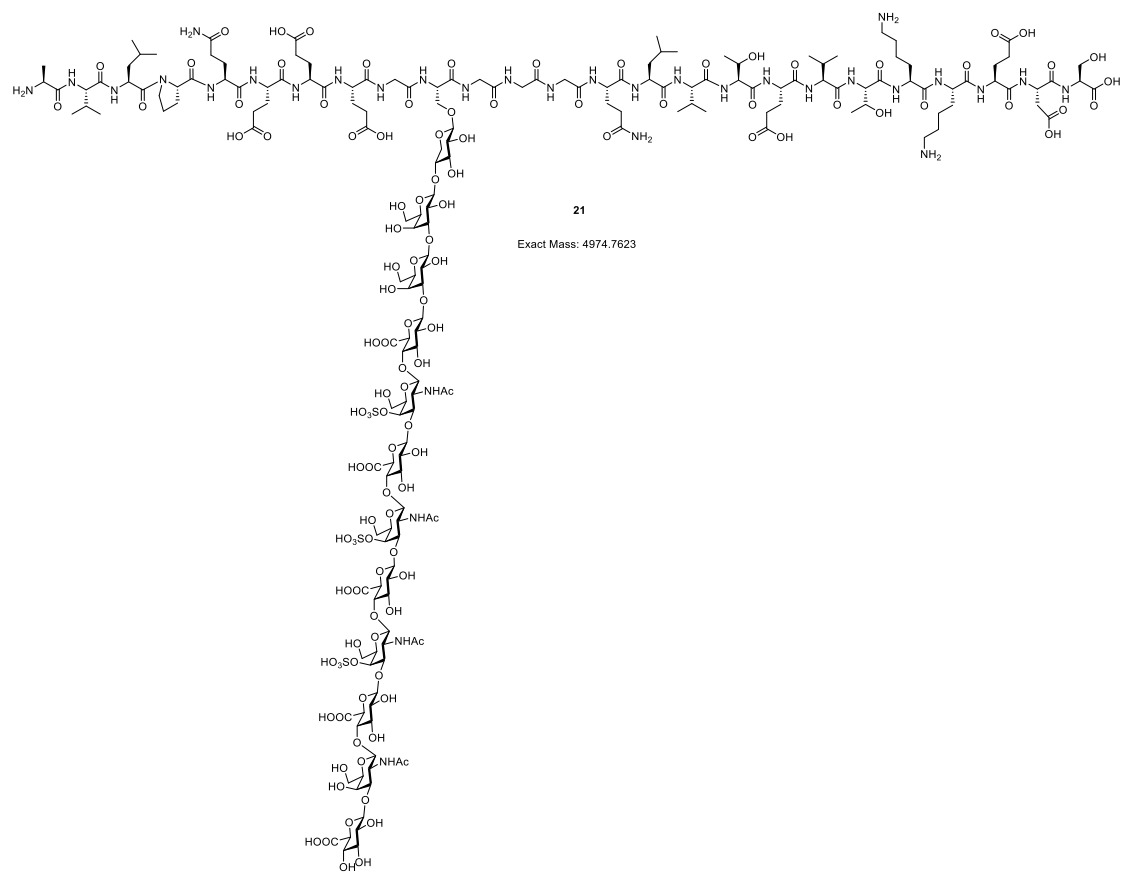
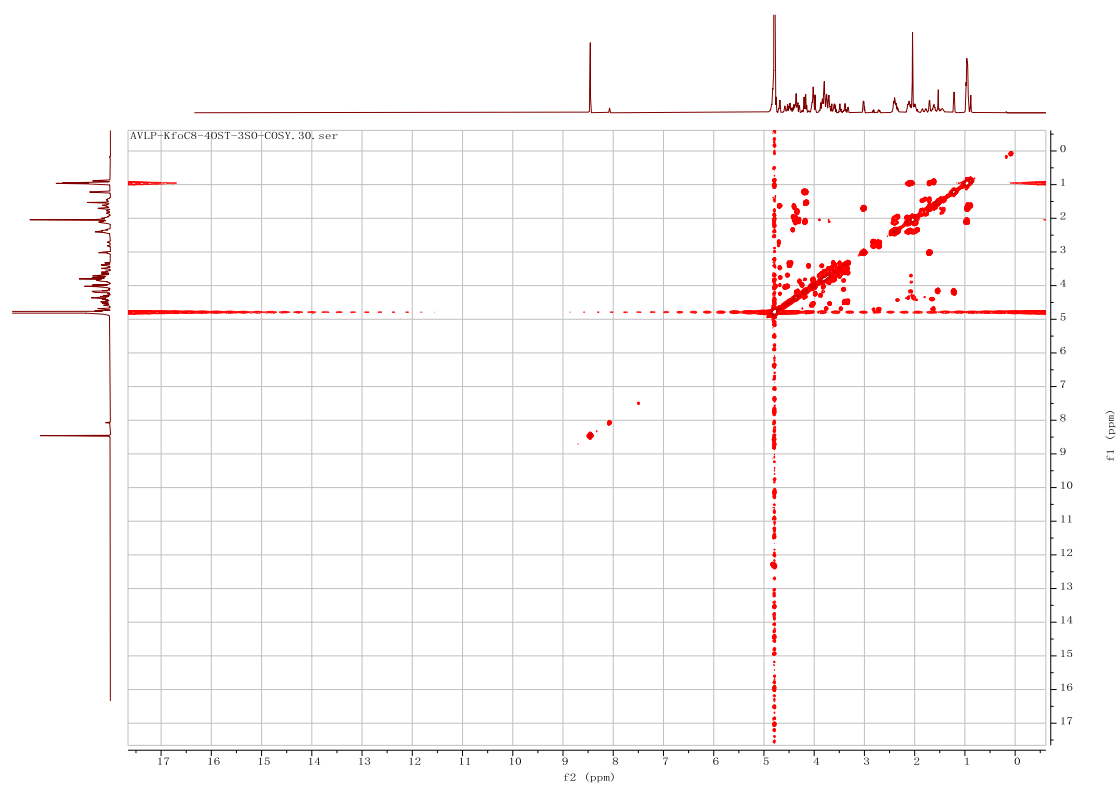


Figure 2.109. COSY of **21** (800 MHz, D₂O).

Figure 2.109. (cont'd)



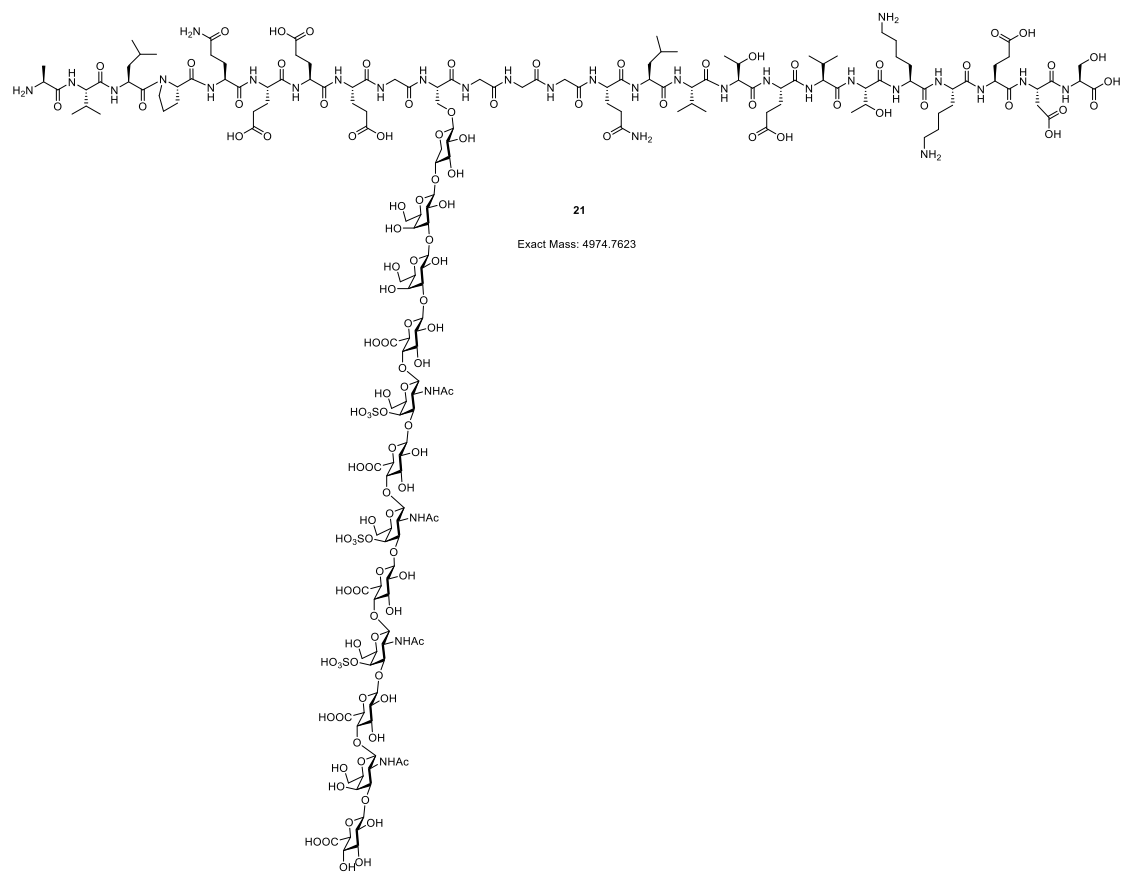
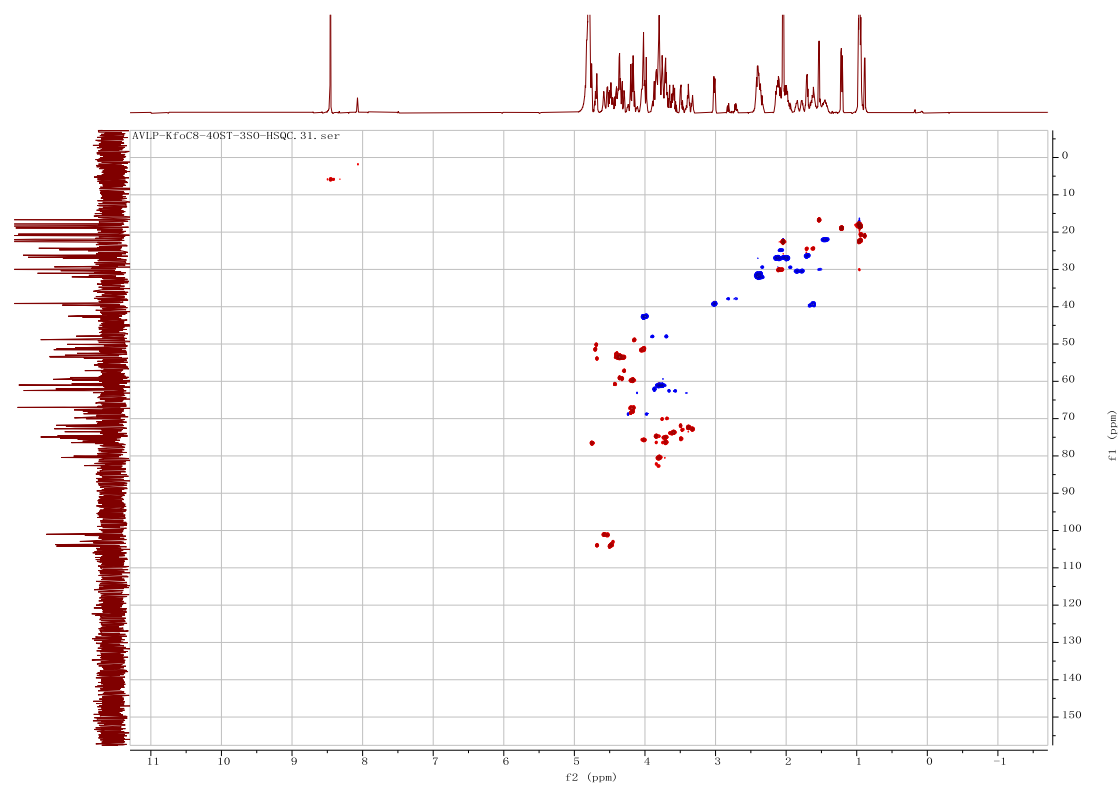


Figure 2.110. HSQC of **21** (800 MHz, D₂O).

Figure 2.110. (cont'd)



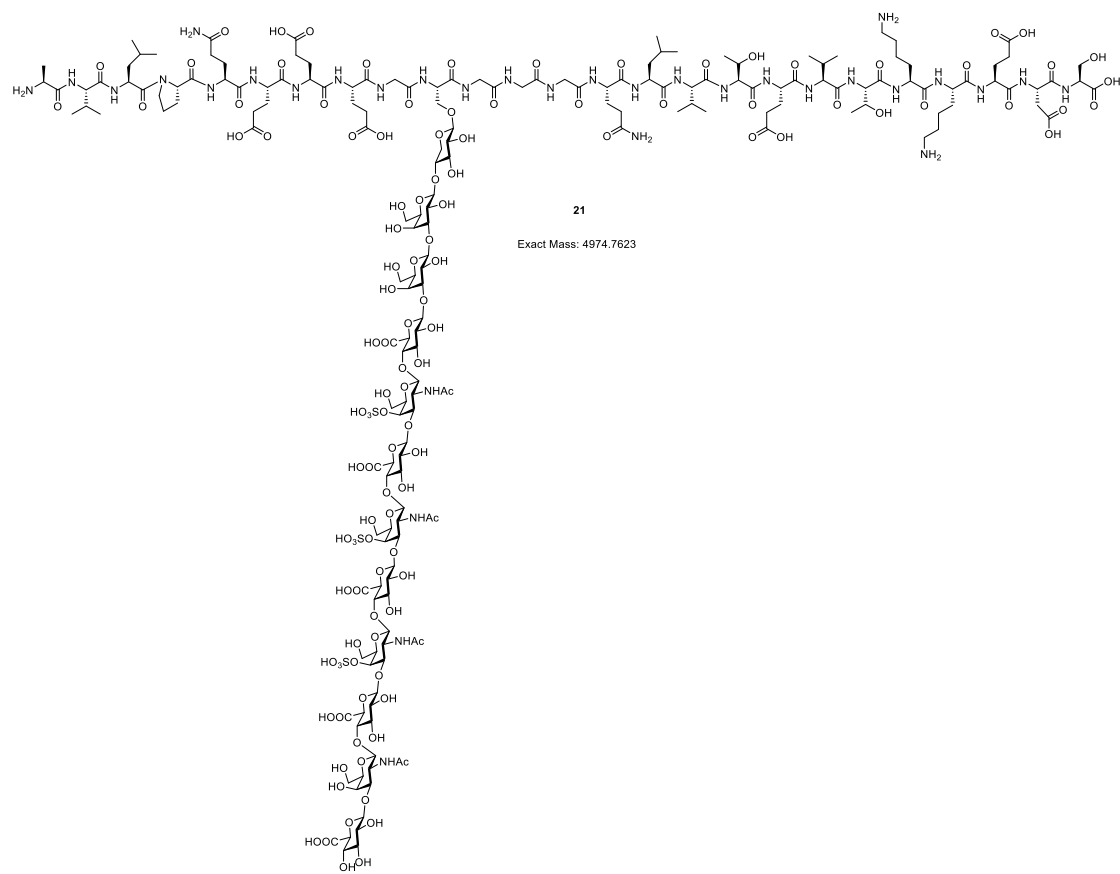
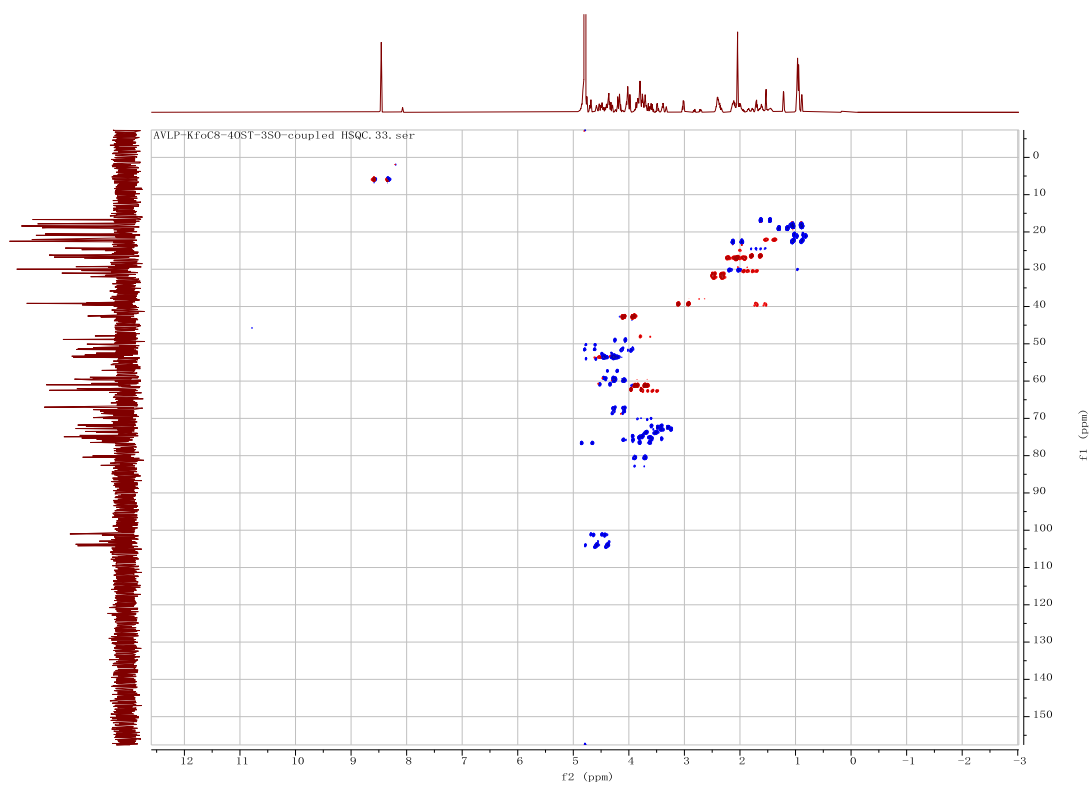


Figure 2.111. Coupled HSQC of **21** (800 MHz, D₂O).

Figure 2.111. (cont'd)



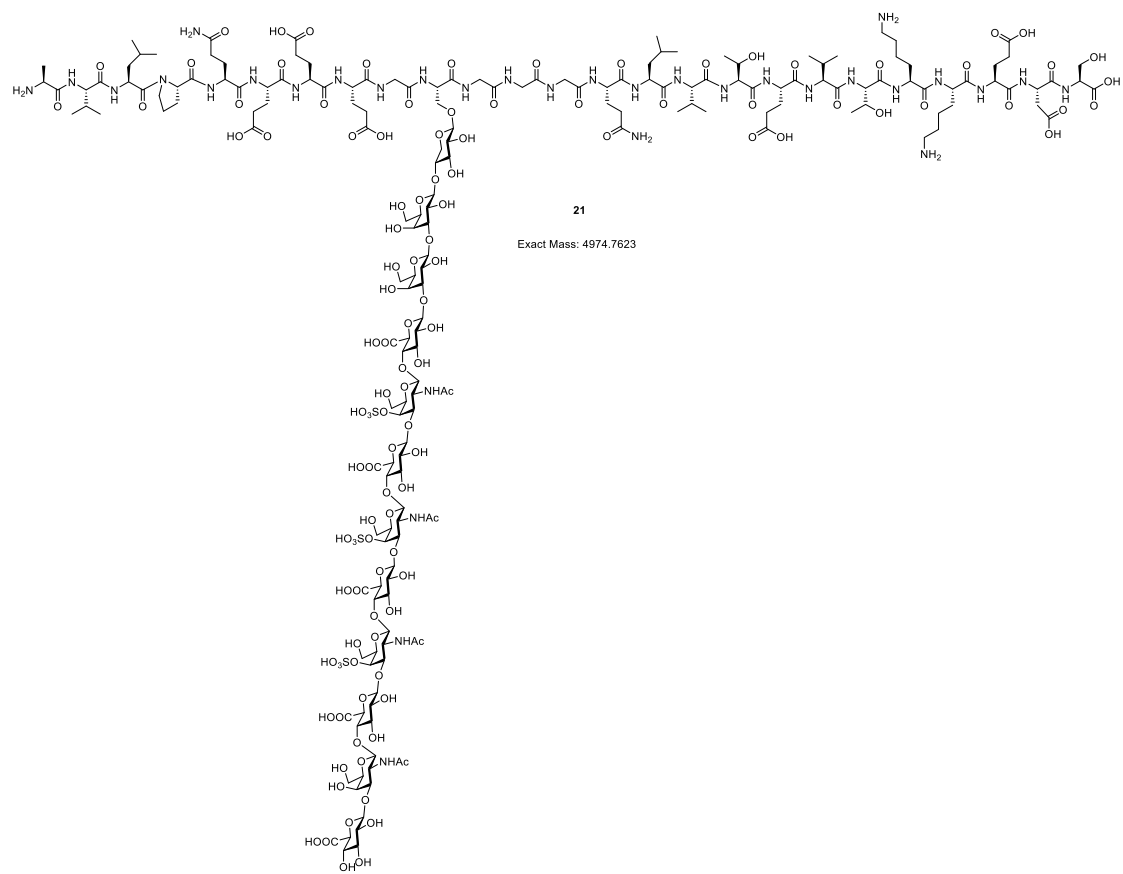
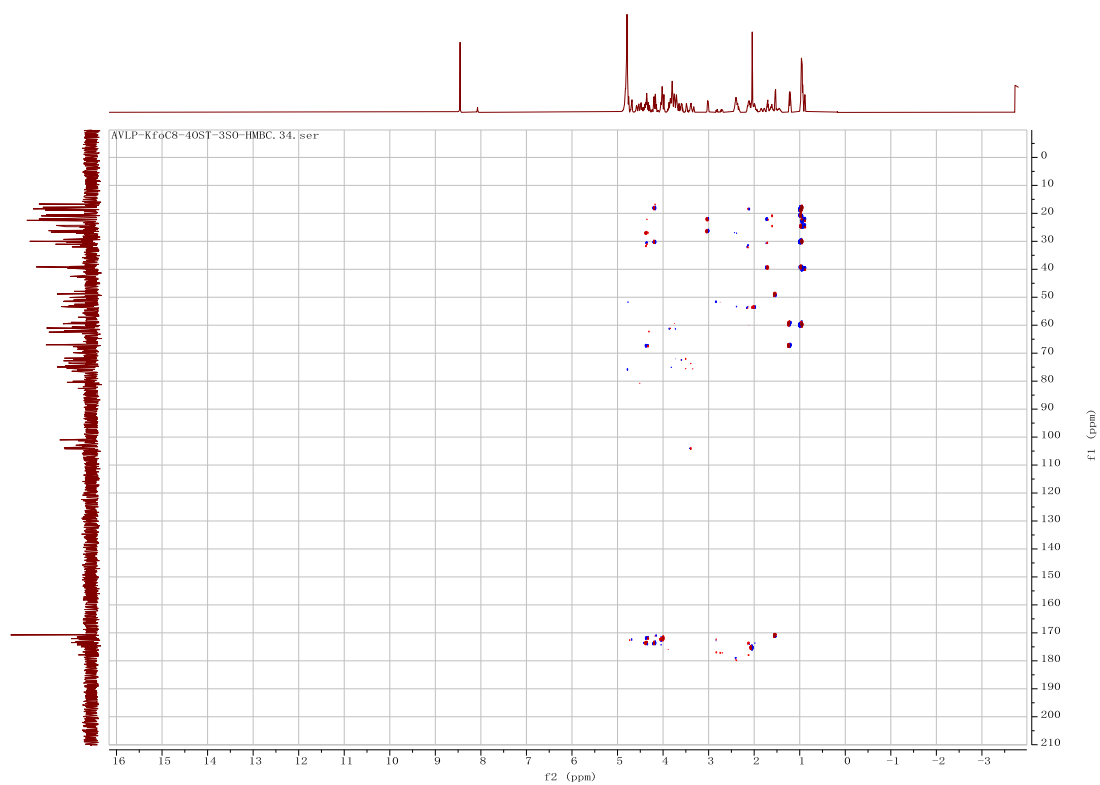


Figure 2.112. HMBC of **21** (800 MHz, D₂O).

Figure 2.112. (cont'd)



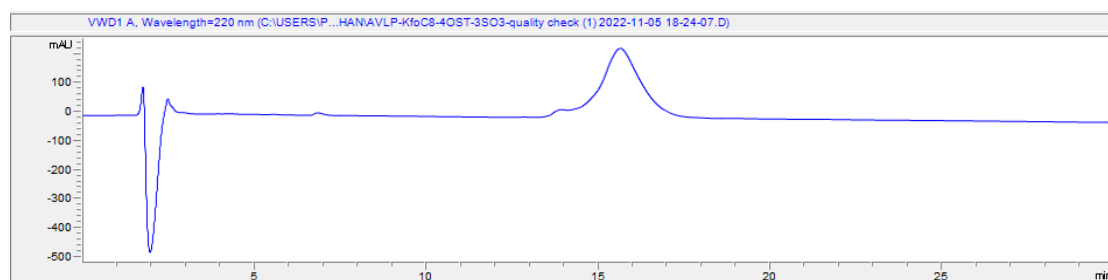
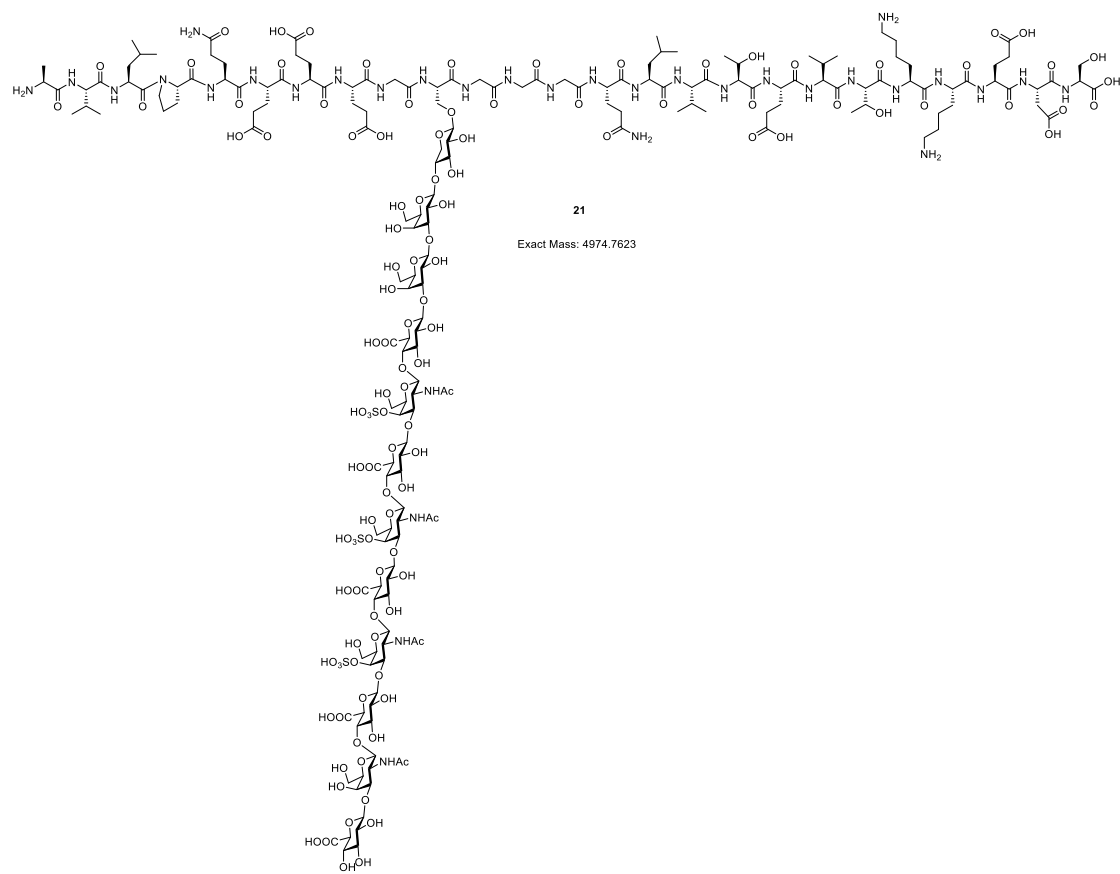
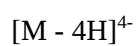
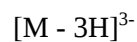
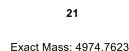


Figure 2.113. HPLC Chromatogram of **21**.



197

Chapter 3. Comprehensive Mapping of CSPG in Biological Samples Using a Chemo-Enzymatic Method

3.1 Introduction

CSPGs are a diverse class of complex molecules found abundantly in the extracellular matrix of vertebrate tissues¹⁻³. They wield significant influence over a wide array of biological processes, making them a focal point of research in the fields of carbohydrate and proteomic studies. The importance of delving into CSPGs is underscored by their pivotal roles in both health and disease. For instance, in the central nervous system, CSPGs exert substantial influence over neural development and plasticity, as well as axon guidance^{4,5}. In pathological conditions, such as spinal cord injuries and neurodegenerative diseases like Alzheimer's, altered CSPG expression and function have been implicated in inhibiting neural regeneration and repair⁶. Moreover, CSPGs play a critical role in cancer progression, influencing tumor cell behavior and metastasis⁷.

At the molecular level, CSPGs contain a central protein core, to which long glycan chains are attached. The glycan chains consist of chondroitin sulfate (CS) glycosaminoglycan (GAG) linked through the tetrasaccharide linkage region ($\text{Glc}\alpha\beta 1\text{--}3\text{Gal}\beta 1\text{--}3\text{Gal}\beta 1\text{--}4\text{Xyl}\beta 1$) with the core protein⁸⁻¹⁰. Importantly, these disaccharide units ($4\text{Glc}\alpha\beta 1\text{--}3\text{GalNAc}\beta 1$) of the CS chains are modified at varying positions through *O*-sulfation, a process that confers additional complexity to the structures of CSPG glycan chains¹¹.

Current research on CSPGs has been primarily focused on the glycan component, considering it as the primary element responsible for binding, while the protein part has been traditionally regarded merely as a glycan carrier. However, recent studies have increasingly revealed that the protein segment also plays a significant role in binding with its receptor¹². As a result, a thorough investigation of CSPGs necessitates a meticulous analysis of the associated core

proteins. This involves not only identifying the diverse isoforms of CSPGs but also characterizing their post-translational modifications and elucidating their interactions with other proteins within the extracellular matrix. This deeper understanding of the proteomic profile of CSPGs can provide invaluable insights into their specific functions in various tissues and disease contexts¹³.

In recent years, considerable efforts have been dedicated to the elucidation of the spatial distribution of CSPGs within biological specimens¹³. A primary challenge in this endeavor arises from the presence of the large, heterogeneous, negatively charged polysaccharide chains of CS within the sample matrix, which poses a formidable barrier to effective proteomic inquiry. Given the relatively low abundance of CSPGs in their natural milieu, it becomes imperative to implement sample enrichment strategies. A widely adopted approach involves the initial enzymatic digestion of the protein sample of interest using trypsin, whereby the resultant mixture of peptides is subsequently subjected to a 10 kDa centrifuge filter. This process selectively retains peptide fractions bearing substantial modifications. Further refinement is achieved through the utilization of a Strong Anion Exchange (SAX) column, targeting high negative charge fractions. Following an exhaustive digestion step employing chondroitinase ABC, the modifications on the peptide chain are trimmed to yield a distinctive hexasaccharide bearing a 4,5-unsaturated uronic acid moiety at the non-reducing end. This unique modification on trypsin-digested peptides provides a discerning handle for the identification of glycopeptides bearing CS moieties¹⁴⁻²⁰.

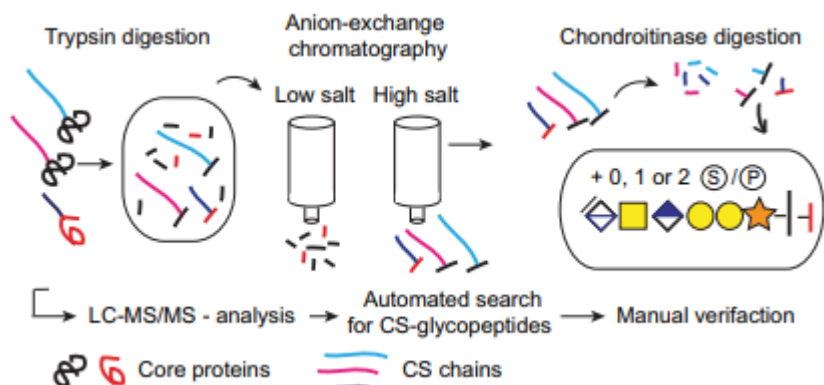


Figure 3.1. General scheme of CS-bearing glycopeptide enrichment using trypsin digestion and SAX column¹⁴.

In 2020, the Clausen group introduced a refined approach for the enrichment of CSPGs, building upon prior methodologies. Notably, they initiated the process by enriching protein samples using a column conjugated with VAR2CSA, a malaria protein prominently expressed on the surface of infected erythrocytes during *Plasmodium falciparum* infection²¹. This protein exhibits specific affinity for CS-A. This innovative methodology achieved a breakthrough in the identification of hitherto unrecognized CSPGs in both cancer cells and placental tissues²².

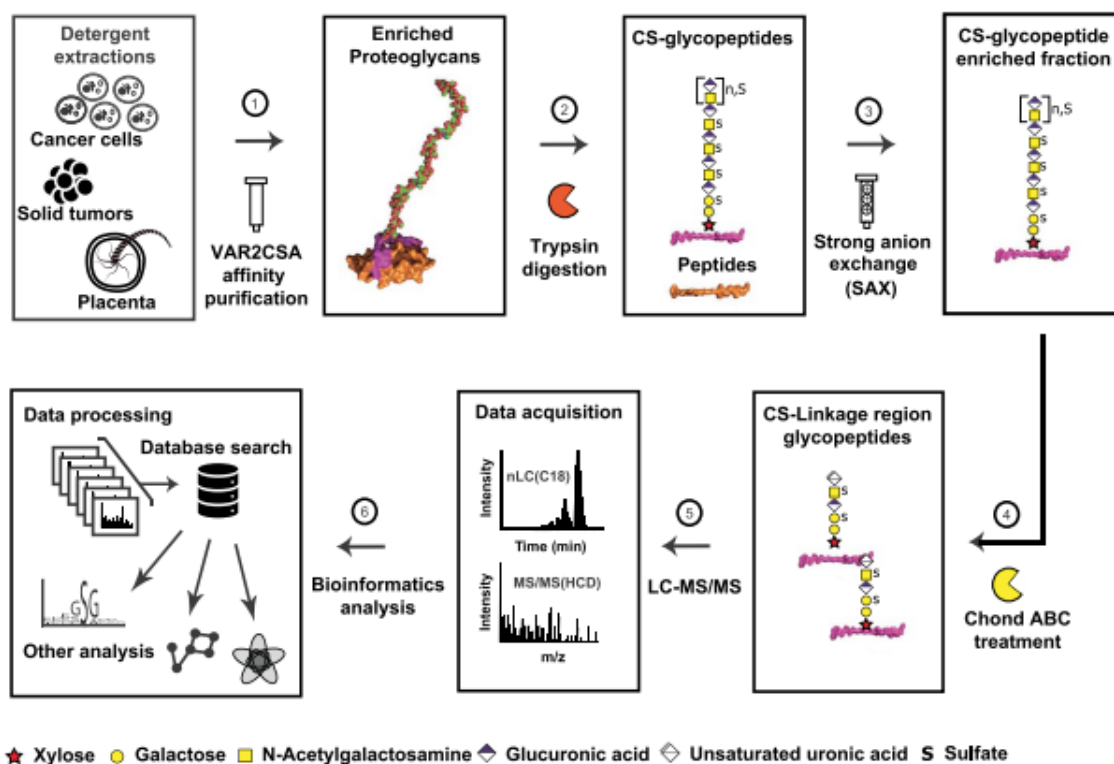


Figure 3.2. General scheme of CS-bearing glycopeptide enrichment using VAR2CSA affinity column²². Reproduced with permission from Oxford University Press.

In our current research, we are actively working to develop a comprehensive method for mapping CSPGs in biological samples using chemo-enzymatic labeling. This innovative approach holds great promise in allowing us to gain a more thorough and detailed understanding of the distribution and function of CSPGs in complex biological systems.

3.2 Results and Discussion

While the methodologies for probing CSPGs have made significant strides, they still present certain inherent limitations. Specifically, the utilization of VAR2CSA exhibits a marked specificity towards CS-A glycans, rendering it challenging to detect CSPGs bearing alternative

sulfation patterns. To address this issue, we would like to develop a general approach for CSPG mapping in biological samples.

In this chapter, we aim to develop a strategy for comprehensive evaluation and profiling CSPGs in biological samples. Our initial focus in processing CSPGs involved harnessing the product generated by chondroitinase ABC. This enzymatic reaction yields a distinctive hexasaccharide structure ($\Delta\text{GlcA}\beta 1\text{-3GalNAc}\beta 1\text{-3GlcA}\beta 1\text{-3Gal}\beta 1\text{-3Gal}\beta 1\text{-4Xyl}\beta 1$) characterized by a 4,5-unsaturated uronic acid moiety (ΔGlcA) at the non-reducing end. Subsequently, the 4,5-unsaturated uronic acid was cleaved by adding mercuric acetate, yielding a pentasaccharide product ($\text{GalNAc}\beta 1\text{-3GlcA}\beta 1\text{-3Gal}\beta 1\text{-3Gal}\beta 1\text{-4Xyl}\beta 1$)²³. Through a stepwise enzymatic reaction employing either HS-synthase PmHS2 or CS-synthase KfoC, we can accomplish the transfer of GlcA and azido-containing GalNAz onto proteins. This unique motif was subsequently conjugated with a biotinylated alkyne via copper-catalyzed Azide-Alkyne Cycloaddition (CuAAC). Consequently, samples containing CSPGs could be effectively enriched utilizing streptavidin-coated agarose beads (**Figure 3.3.**)

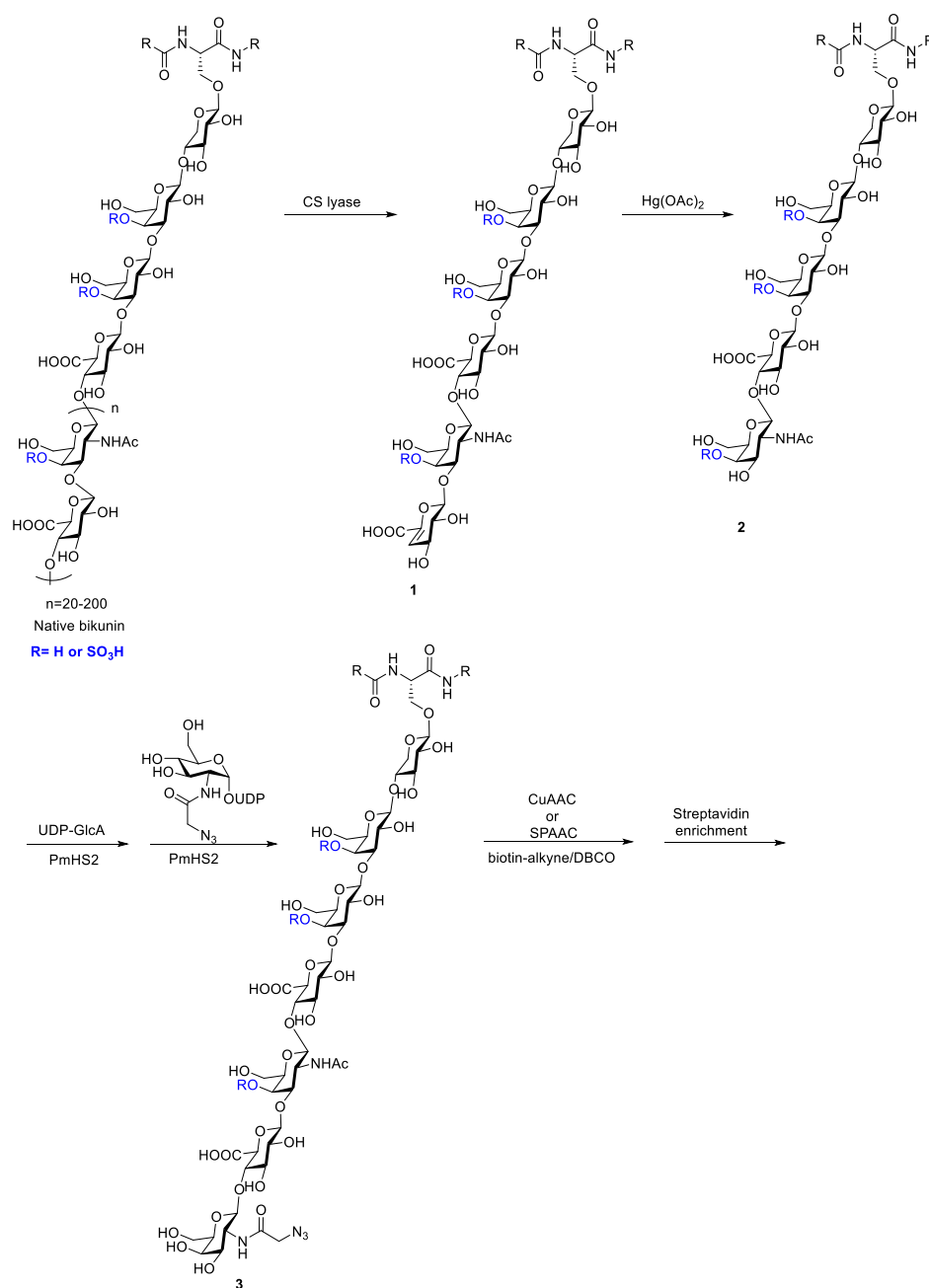
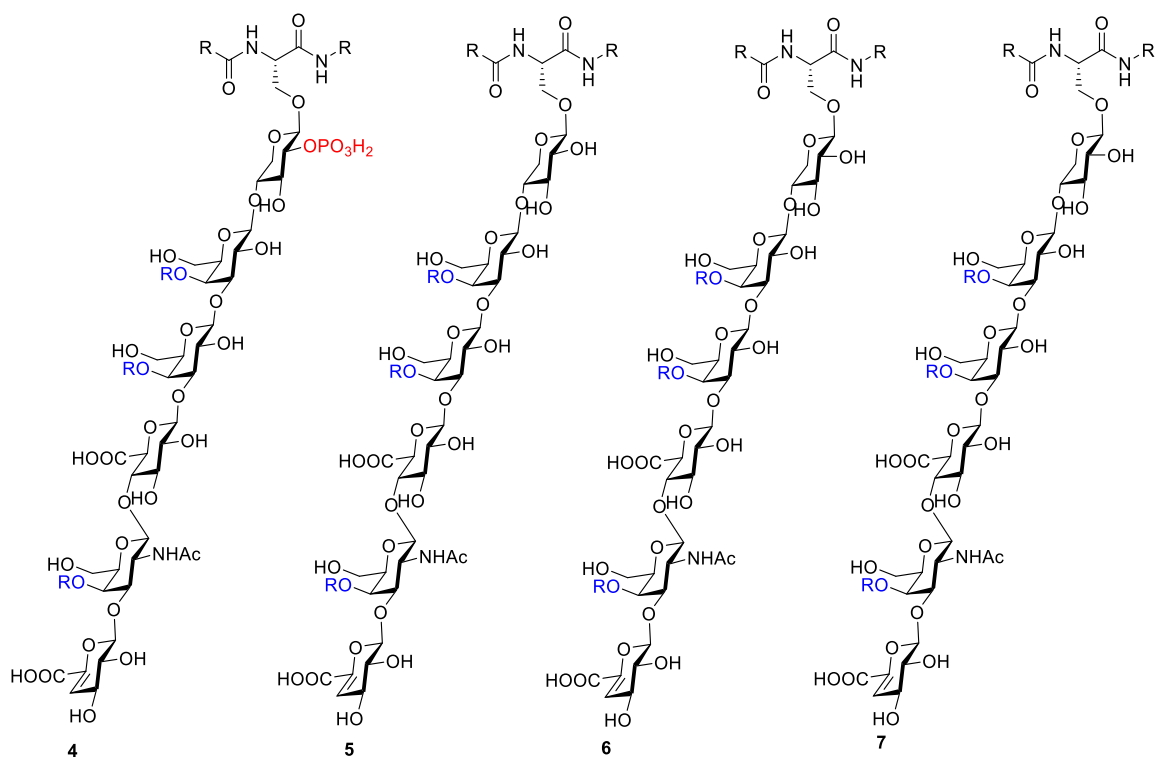


Figure 3.3. General scheme of chemo-enzymatic enrichment of CSPG in biological sample, bikunin as an example CSPG.

Bikunin, also known as inter- α -trypsin inhibitor or trypstatin, stands as one of the simplest CSPGs²⁴. Bikunin bears a CS chain on S10, making it an ideal target as the proteomic standard. The lyophilized form of bikunin underwent an initial treatment with chondroitinase ABC, followed

by protein sequencing via MS/MS. This analysis revealed four distinct modifications (glycopeptide **4-7**) within protein **1**, denoted as serine (S) S+1233, S+1232, S+1153, and S+1073 (**Figures 3.4., 3.9.**). These corresponded to four unique hexasaccharide structures characterized by varying sulfate and phosphate moieties on the glycan. Subsequently, protein **1** underwent further treatment with mercuric acetate. Subsequently, four modifications (glycopeptide **8-11**) were identified in protein **2**, designated as S+1075, S+1074, S+995, and S+915 (**Figures 3.5., 3.10.**). This consistent sulfation/phosphorylation pattern mirrored that observed in protein **1**, affirming the successful completion of this step. Unfortunately, attempts to extend the pentasaccharide to hexa- and septasaccharide with transferase KfoC or PmHS2 yielded no desirable products. From our experiment, it is possible that neither PmHS2 nor KfoC could accommodate sulfated GalNAc as the acceptor.



R = H or SO₃H

2 SO₃H on 4
+1233

3 SO₃H on 5
+1232

2 SO₃H on 6
+1153

1 SO₃H on 7
+1073

Figure 3.4. Sulfation/phosphorylation pattern of bikunin S10 found on protein **1** after chondroitinase and trypsin digestion, S+ 1233, S+ 1232, S+1153 and S+1073 were found on peptide AVL PQEEEGSGGGQLVTEVTK, respectively.

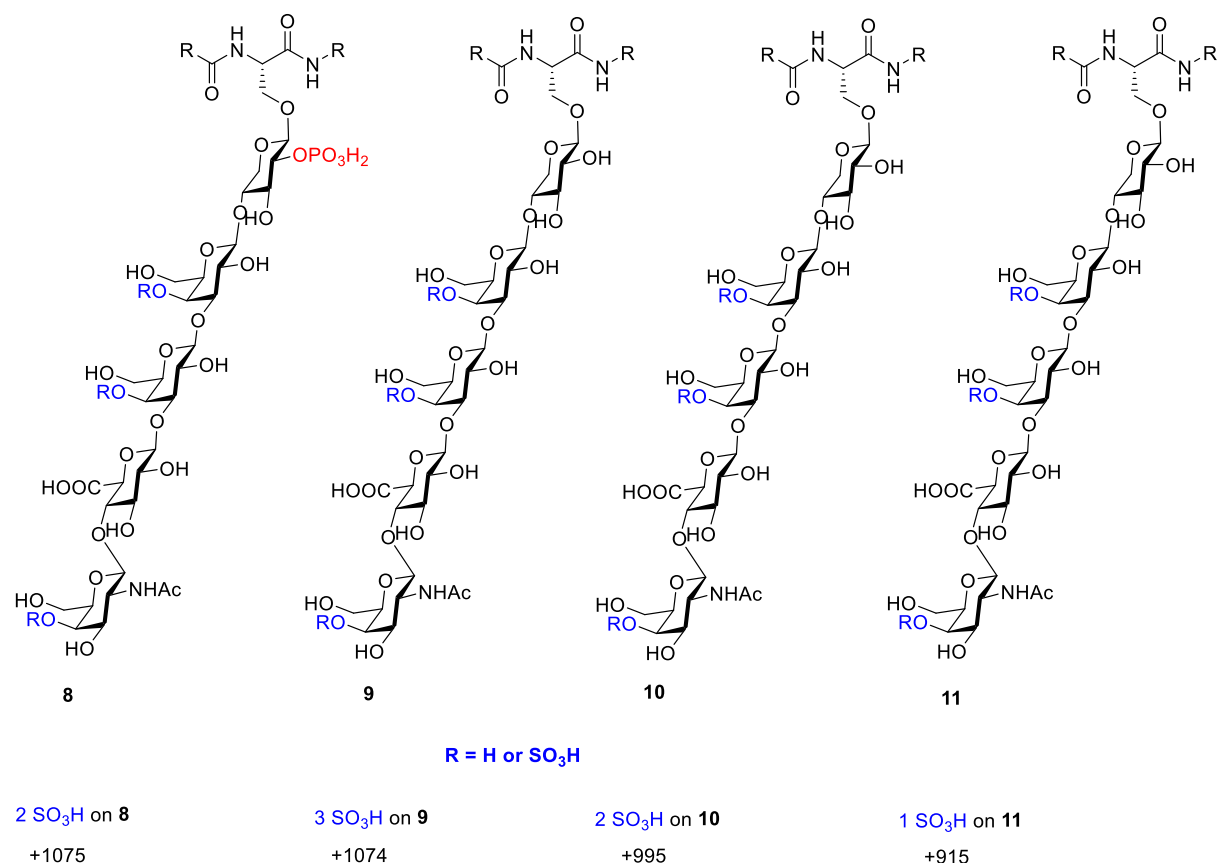


Figure 3.5. Sulfation/phosphorylation pattern of bikunin S10 found on protein 2 after mercuric acetate cleavage. S+ 1075, S+ 1074, S+995 and S+915 were found on peptide AVLPEEEGSGGGQLVTEVTK, respectively.

In order to overcome the difficulties, we explored both chemical and enzymatic approaches for glycan desulfation. In the case of chemical desulfation, prior literature suggested employing a pyridine-DMSO mixture with heat treatment²⁵. In our hands, we observed that bikunin underwent precipitation and could not be redissolved, thus rendering this approach impractical. Subsequently, we investigated desulfation using arylsulfatase B (ARSB), a known enzyme for CS-A desulfation²⁶. However, subjecting native bikunin or bikunin pre-treated with chondroitinase and mercuric acetate to ARSB did not yield the desired unsulfated pentasaccharide. This outcome is likely

attributed to the proximity of the sulfate group on the fifth sugar, GalNAc, to the linkage region and peptide, making it challenging for ARSB to access and promote desulfation.

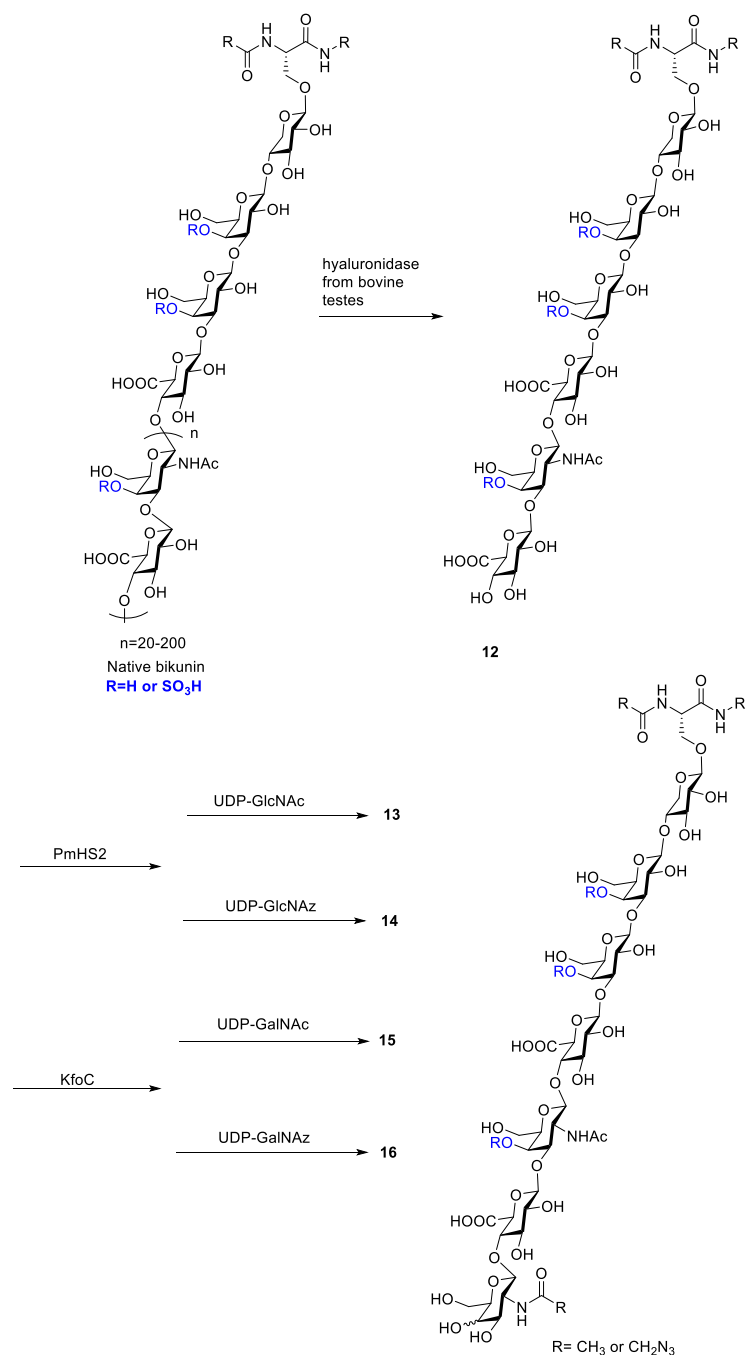


Figure 3.6. General scheme of improved chemo-enzymatic enrichment of CSPG in biological sample, bikunin as an example CSPG.

In addressing the primary cause of failure attributed to sulfate residues on the non-reducing end sugar, our investigation focused next on the enzymatic properties of hyaluronidase-bovine testes^{27, 28}. While this enzyme recognizes its primary substrate, hyaluronic acid, it has also demonstrated the ability to cleave chondroitin sulfate²⁷. Notably, exhaustive enzymatic digestion of PGs led to the formation of a hexasaccharide containing a glucuronic acid moiety at the non-reducing end²⁸, effectively circumventing the challenges posed by sulfated GalNAc as a substrate. To experimentally validate this concept, native bikunin was subjected to hyaluronidase treatment, resulting in three distinct modifications (glycopeptide **17-19**) on protein **12**, exhibiting 0, 1, or 2 sulfates in the linkage region respectively (**Figures 3.7. and 3.12.**). Furthermore, to ascertain the capability of KfoC to facilitate the transfer of additional sugar units onto the existing hexasaccharide, we conducted four distinct reactions. Employing PmHS2 to transfer GlcNAc or GlcNAz and KfoC to transfer GalNAc or GalNAz, the outcomes revealed the successful generation of the desired products (protein **13-16**) with various sulfation patterns (**Figures 3.13. and 3.20.**).

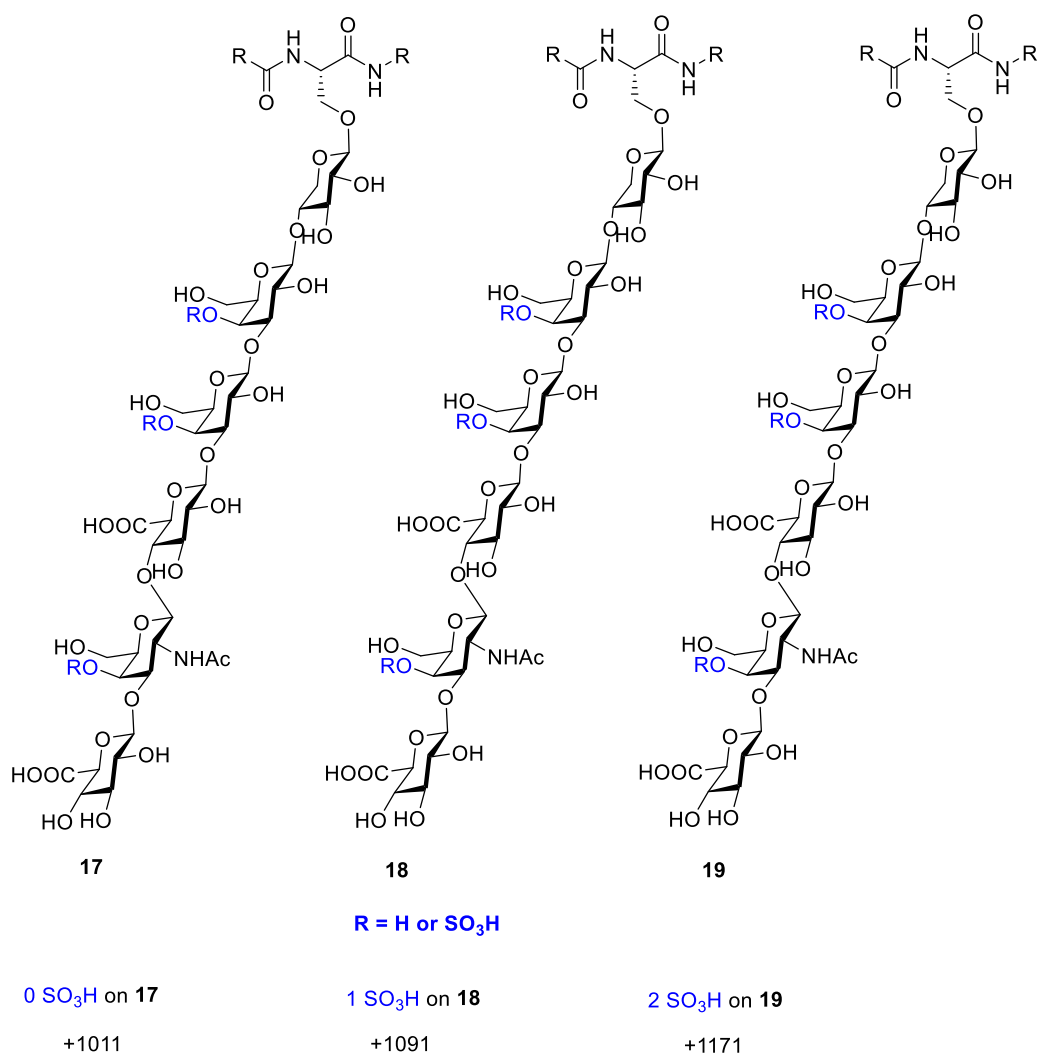


Figure 3.7. Sulfation pattern of bikunin S10 found on protein **12** after hyaluronidase digestion. S+ 1011, S+ 1091 and S +1171 were found on peptide AVL PQEEEGSGGGQLVTEVTK, respectively.

Building upon the successful utilization of native bikunin as a substrate, our research endeavors now extend towards more intricate biological samples, particularly plasma. The exploration of CS within protein samples necessitates a sequential application of two enzymatic reactions, followed by a downstream click reaction. Each of these three steps demands a distinct buffer for reaction conditions. In our prior work with bikunin samples, we employed dialysis

between steps to facilitate buffer exchange. However, the transition to lyophilized plasma samples revealed an unforeseen challenge: during each buffer exchange, proteins precipitated from the plasma, potentially attributable to the abrupt pH shift from the hyaluronidase buffer (0.1 M NaOAc, 0.15 M NaCl pH 5) to the KfoC buffer (25 mM MOPS, 10 mM MnCl₂, pH 7.2). To circumvent this issue, the development of a novel method tailored for more complex biological samples becomes imperative.

A commonly reported approach involves the precipitation of proteins using the MeOH-CHCl₃ method, forming a pellet subsequently redissolved in a buffer containing 1% sodium dodecyl sulfate (SDS)²⁹. In our investigation, we assessed the applicability of this method by initially testing whether CS on bikunin could be digested under conditions involving 1% SDS. As depicted in **Figure 3.8.**, in the absence of SDS, digestion readily occurred at room temperature, reaching completion after 24 hours at room temperature and within 6 hours at 37°C. Conversely, the introduction of SDS into the buffer led to a complete abrogation of hyaluronidase activity. This observation underscores the need for a tailored methodology to address the intricacies of biological samples, with careful consideration given to preserving enzymatic activity in the presence of specific buffer conditions.

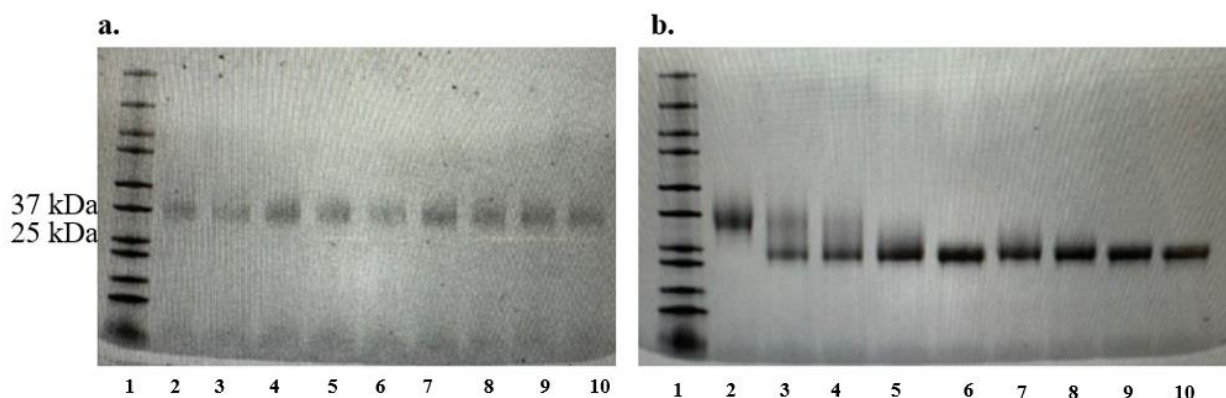


Figure 3.8. SDS PAGE gel of bikunin digested by hyaluronidase (a) with or (b) without 1% SDS.

Figure 3.8. (cont'd)

(a). Lane 1, molecular weight ladder; Lane 2, native bikunin sample; Lane 3-6, bikunin incubated at RT for 1, 2, 7, 19 h, respectively; Lane 7 -10, incubation at 37 °C for 1, 2, 7, 19 h. (b). Lane 1, ladder; Lane 2, native bikunin sample; Lane 3-6, bikunin incubated at RT for 1, 2, 6, 24 h, respectively; Lane 7 -10, incubation at 37 °C for 1, 2, 6, 24 h.

We extended our investigation to evaluate the KfoC reaction using UDP-GalNAz as the glycosyl donor and GlcA-pNp (compound **20**) as the acceptor substrate. Under native conditions, the LCMS analysis revealed the successful formation of the desired disaccharide product. However, when the reaction was conducted in the presence of 1% or 0.5% SDS, a complete abrogation of the reaction was observed (**Figure 3.11.**). In an effort to overcome this challenge, we explored alternative systems by incorporating various detergents, including Triton X-100, Tween 20, and 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS). Intriguingly, none of these detergents inhibited the enzymatic activity. Notably, only buffer containing 10% CHAPS demonstrated the capability to dissolve the protein pellet after sonication, offering a potential solution to SDS denaturation.

To finalize the protocol, several sample processing steps were improved. In MS analysis, a pivotal consideration revolves around the alkylation of cysteine residues in protein samples, achieved through the utilization of dithiothreitol and iodoacetamide. This critical step is essential for mitigating background, particularly when employing a dibenzocyclooctyne (DBCO) group-conjugated biotin reagent for enrichment. The use of DBCO introduces the potential for side reactions with thiol on cysteine, thereby amplifying background signals³⁰. Consequently, using CuAAC method would be a better choice. To optimize the enrichment method, it becomes apparent

that the incorporation of a cleavable linker is imperative. This necessity arises from the inherent challenges associated with stripping biotinylated proteins from Streptavidin agarose. Cleavable biotin linkers bearing *N*-1-(4,4-dimethyl-2,6-dioxocyclohexylidene) ethyl (Dde) was explored (Figure 3.21. – 3.22.).

3.3 Outlook

CSPGs stand as a pivotal class of proteoglycans within biological systems, underscoring their significance in cellular processes. A comprehensive exploration of CSPGs becomes imperative, necessitating the development of methods for the quantitative detection of specific CSPGs under defined conditions. The innovations in these areas will not only provide a nuanced understanding of CSPG dynamics but also open new avenues for investigating potential pathways implicated in CSPG-related diseases. The ongoing pursuit of advanced proteomic methods is crucial to enhance our analytical capabilities, offering a more nuanced and comprehensive view of CSPG interactions and functions.

As we delve into the intricacies of CSPGs, our focus extends beyond their specific detection to broader implications within the realm of proteoglycans. This adaptability underscores the broader impact of our methodologies, promising a versatile toolkit for researchers exploring the diverse landscape of proteoglycan biology. As we continue to refine and expand these techniques, we anticipate their widespread applicability, offering novel insights into the roles of proteoglycans across various biological contexts. In MS perspective, we first explored methods using in-agarose-gel digestion, then further employed higher energy collision induced dissociation (HCD). With this method, we successfully determined the glycan patterns on bikunin, but we did not get reasonable results for complex biological samples such as plasma. We are currently collaborating with Dr. Junfeng Ma, using stepped collision energy/higher energy collisional

dissociation (sceHCD) and electron-transfer/higher-energy collision dissociation (EThcD) to solve this issue. This ongoing endeavor not only advances our understanding of CSPGs but also contributes to the broader landscape of glycoscience, paving the way for transformative discoveries in the intricate interplay between proteoglycans and cellular processes.

3.4 Experimental Section

3.4.1 Materials

Chondroitinase ABC from *Proteus vulgaris*, hyaluronidase from Bovine Testes, human plasma, SDS, Triton X-100, Tween 20 and CHAPS were purchased from Millipore sigma. KfoC were a kind gift from our collaborator Dr. Jian Liu. Tris/Glycine/SDS electrophoresis buffer, prestained protein ladder, sample loading buffer, and Coomassie Blue R-250 were purchased from Bio-rad (Hercules, CA). Pierce™ NeutrAvidin™ Agarose was purchased from Thermo fisher (Waltham, MA). UDP-GlcNAz and UDP-GalNAz were purchased from Accela chembio (China). Dde-Biotin-Alkyne (CCT-1137) was purchased from vector laboratories (Newark, CA). Native bikunin were purchased from BOC Sciences (New York, NY). ARSB were purchased from R&D systems (Minneapolis, MN). All other chemical reagents were purchased from commercial sources and used without additional purifications unless otherwise noted.

3.4.2 Proteolytic Digestion

Gel bands were digested in-gel according to Shevchenko, et. al. with modifications³¹. Briefly, gel bands were dehydrated using 100% acetonitrile and incubated with 10 mM dithiothreitol in 100 mM ammonium bicarbonate, pH~8, at 56 °C for 45 min, dehydrated again and incubated in the dark with 50 mM chloroacetamide in 100 mM ammonium bicarbonate for 20 min. Gel bands were then washed with ammonium bicarbonate and dehydrated again. Sequencing grade modified trypsin was prepared to 0.005 µg/µL in 50mM ammonium bicarbonate and ~100

μL of this was added to each gel band so that the gel was completely submerged. Bands were then incubated at 37 °C overnight. Peptides were extracted from the gel by water bath sonication in a solution of 60% acetonitrile (ACN) /1% trifluoroacetic acid (TFA) and vacuum dried to ~2 μL.

3.4.3 LC/MS/MS Analysis

Peptide samples were re-suspended in 2% ACN/0.1% TFA to 20 μL and an injection of 5 μL was automatically made using a Thermo (www.thermo.com) EASYnLC 1200 onto a Thermo Acclaim PepMap RSLC 0.1 mm x 20 mm C18 trapping column and washed for 5 min using Buffer A. Bound peptides were then eluted onto a Thermo Acclaim PepMap RSLC 0.075 mm x 500 mm C18 resolving column with a gradient of 8% B to 25% B in 19 min and raising from 25% B to 40% B at 24 min at a constant flow rate of 300 nl/min. Following the gradient, the solvent mixture was raised to 90% B and held for the duration of the run (Buffer A = 99.9% Water/0.1% Formic Acid (FA), Buffer B = 80% ACN/0.1% FA/19.9% Water). Column temperature was maintained at 50 °C using an integrated column heater (PRSO-V2, Sonation GmbH, Biberach, Germany). Eluted peptides were sprayed into a ThermoScientific Q-Exactive HF-X mass spectrometer (www.thermo.com) using a FlexSpray spray ion source. Survey scans were taken in the Orbitrap (45000 resolution, determined at m/z 200) and the top 15 ions in each survey scan are then subjected to automatic HCD with fragment spectra acquired at a resolution of 7500.

3.4.4 Data Analysis

The MS/MS spectra were converted to peak lists using Mascot Distiller, v2.8.3 (www.matrixscience.com) and searched against a reference database containing all protein sequences available from Uniprot (www.uniprot.org, downloaded 2023-04-18) using the Mascot searching algorithm, v2.8.3³². The Mascot output was then analyzed using Scaffold, v5.3.0

(www.proteomesoftware.com) to probabilistically validate protein identifications. Assignments validated using the Scaffold 1% FDR confidence filter are considered true.

Mascot parameters for all databases were as follows:

- allow up to 2 missed tryptic sites
- Fixed modification of Carbamidomethyl Cysteine,
- variable modification of Oxidation of Methionine,
- peptide tolerance of +/- 10ppm
- MS/MS tolerance of 0.02 Da
- FDR calculated using randomized database search

3.4.5 General procedure of the chondroitinase ABC digestion and mercuric acetate treatment

Bikunin (4 mg) was dissolved in a reaction buffer consisting of 0.1 U of chondroitinase ABC, 0.1 M Tris-HCl, 30 mM sodium acetate, and adjusted to pH 8.0 in a total volume of 2 mL. The mixture was incubated at 37°C, and the reaction was monitored by SDS-PAGE. After overnight incubation, 0.1 M acetic acid was slowly added to the mixture until pH 5, then a stock solution of 70 mM mercuric acetate was added to the mixture (final concentration 35 mM) and incubated for 10 min at RT. The resulting mixture was dialyzed to remove the excess of mercuric acetate.

3.4.6 General procedure of hyaluronidase digestion and PmHS2/KfoC transferase reaction

Bikunin (4 mg) was dissolved in a 2 mL reaction buffer, which included 40 µg of hyaluronidase and was composed of 0.1 M NaOAc, 0.15 M NaCl, pH 5. The mixture underwent incubation at 37°C for 2 hours, and the reaction progress was monitored by SDS-PAGE. Following the reaction, protein was precipitated with the MeOH-CHCl₃ (For each 100 µL sample, 100 µL

H₂O, 600 μ L MeOH, 200 μ L CHCl₃ and 450 μ L H₂O were added sequentially with vortex). Following this, the pellet of bikunin was treated with KfoC. Specifically, 2 mL of KfoC buffer (containing 25 mM MOPS, 10 mM MnCl₂, 1 mM UDP-GalNAz or UDP-GalNAc, pH 7.2) with 10% CHAPS was added to the pellet. The mixture underwent sonication for 10 min until all precipitates were dissolved. Subsequently, 200 μ g of KfoC was added to the mixture, which was then incubated overnight at 4°C with end-over-end rotation. Upon completion of the incubation, the mixture was subjected to precipitation once again using the MeOH-CHCl₃ precipitation method. The resulting pellet was washed three times with cold MeOH to eliminate excess UDP-GalNAz.

3.4.7 General procedure of CuAAC reaction

The GalNAz-transferred pellet (from a 200 μ g reaction) was redissolved in 100 μ L PBS with 1% SDS. The pellet was sonicated until it was completely dissolved. To the SDS-containing buffer, reagents were added in the following order: 1 mM (final concentration) Dde-Biotin-Alkyne, 0.3 mM (final concentration) CuSO₄·2-(4-((Bis((1-(tert-butyl)-1H-1,2,3-triazol-4-yl)methyl)amino)methyl)-1H-1,2,3-triazol-1-yl)acetic acid (BTAA) (molar ratio 1:2), and 0.2 mM (final concentration) sodium ascorbate. The mixture was vortexed and incubated at room temperature for 2 h. Upon completion, the protein was pelleted again using the MeOH-CHCl₃ precipitation method. It was then washed five times with cold MeOH to remove excess Dde-Biotin-Alkyne.

3.4.8 General procedure of biotin enrichment and cleavage

The biotinylated pellet (from a 200 μ g reaction) was resuspended in 50 μ L of 8 M urea in 50 mM TEABC. The solution was sonicated until the entire pellet was dissolved. The sample was then diluted to 1 M urea with 50 mM TEABC, and 50 μ g trypsin was added. The solution was

shaken at 70 rpm at 37 °C overnight. High-capacity Neutravidin agarose beads (120 µL slurry) were washed with 1 mL cold PBS six times. Beads were then added to the urea solution, and the mixture was incubated with end-over-end rotation at room temperature for 3 h. The mixtures were spun down (500 g × 2 min), and the supernatant was removed. The beads were washed with cold PBS (1 mL) six times, followed by water (1 mL) six times, then washed once with 20% MeOH (1 mL), and finally once with 70% MeOH (1 mL). Upon completion of the washing steps, the beads were incubated with freshly prepared 2% hydrazine (1 mL) for 1 hour with end-over-end rotation. They were then washed with PBS (1 mL) for 5 minutes. Both fractions were collected for MS analysis.

REFERENCES

1. Francos-Quijorna, I.; Sánchez-Petidier, M.; Burnside, E. R.; Badea, S. R.; Torres-Espin, A.; Marshall, L.; de Winter, F.; Verhaagen, J.; Moreno-Manzano, V.; Bradbury, E. J., Chondroitin sulfate proteoglycans prevent immune cell phenotypic conversion and inflammation resolution via TLR4 in rodent models of spinal cord injury. *Nat. Comm.* **2022**, *13* (1), 10.1038/s41467-022-30467-5.
2. Siebert, J. R.; Conta Steencken, A.; Osterhout, D. J., Chondroitin sulfate proteoglycans in the nervous system: Inhibitors to repair. *Biomed. Res. Int.* **2014**, *2014*, 10.1155/2014/845323.
3. Mencio, C. P.; Hussein, R. K.; Yu, P.; Geller, H. M., The role of chondroitin sulfate proteoglycans in nervous system development. *J. Histochem. Cytochem.* **2020**, *69* (1), 61-80.
4. Avram, S.; Shaposhnikov, S.; Buiu, C.; Mernea, M., Chondroitin sulfate proteoglycans: structure-function relationship with implication in neural development and brain disorders. *Biomed. Res. Int.* **2014**, *2014*, 10.1155/2014/642798.
5. Yang, X., Chondroitin sulfate proteoglycans: key modulators of neuronal plasticity, long-term memory, neurodegenerative, and psychiatric disorders. *Rev. Neurosci.* **2020**, *31* (5), 555-568.
6. Sun, Y.; Xu, S.; Jiang, M.; Liu, X.; Yang, L.; Bai, Z.; Yang, Q., Role of the extracellular matrix in alzheimer's disease. *Front. aging neurosci.* **2021**, *13*, 10.3389/fnagi.2021.707466.
7. Price, M. A.; Colvin Wanshura, L. E.; Yang, J.; Carlson, J.; Xiang, B.; Li, G.; Ferrone, S.; Dudek, A. Z.; Turley, E. A.; McCarthy, J. B., CSPG4, a potential therapeutic target, facilitates malignant progression of melanoma. *Pigment Cell Melanoma Res.* **2011**, *24* (6), 1148-57.
8. Kitagawa, H.; Uyama, T.; Sugahara, K., Molecular cloning and expression of a human chondroitin synthase. *J. Biol.Chem.* **2001**, *276* (42), 38721-38726.
9. Yada, T.; Sato, T.; Kaseyama, H.; Gotoh, M.; Iwasaki, H.; Kikuchi, N.; Kwon, Y. D.; Togayachi, A.; Kudo, T.; Watanabe, H.; Narimatsu, H.; Kimata, K., Chondroitin sulfate synthase-3. Molecular cloning and characterization. *J. Biol.Chem.* **2003**, *278* (41), 39711-25.
10. Yada, T.; Gotoh, M.; Sato, T.; Shionyu, M.; Go, M.; Kaseyama, H.; Iwasaki, H.; Kikuchi, N.; Kwon, Y. D.; Togayachi, A.; Kudo, T.; Watanabe, H.; Narimatsu, H.; Kimata, K., Chondroitin sulfate synthase-2. Molecular cloning and characterization of a novel human glycosyltransferase homologous to chondroitin sulfate glucuronyltransferase, which has dual enzymatic activities. *J. Biol.Chem.* **2003**, *278* (32), 30235-47.
11. Pearson, C. S.; Mencio, C. P.; Barber, A. C.; Martin, K. R.; Geller, H. M., Identification of a critical sulfation in chondroitin that inhibits axonal regeneration. *Elife* **2018**, *7*, 10.7554/eLife.37139.

12. Yang, W.; Eken, Y.; Zhang, J.; Cole, L. E.; Ramadan, S.; Xu, Y.; Zhang, Z.; Liu, J.; Wilson, A. K.; Huang, X., Chemical synthesis of human syndecan-4 glycopeptide bearing O-, N-sulfation and multiple aspartic acids for probing impacts of the glycan chain and the core peptide on biological functions. *Chem. Sci.* **2020**, *11* (25), 6393-6404.
13. Noborn, F.; Nikpour, M.; Persson, A.; Nilsson, J.; Larson, G., Expanding the chondroitin sulfate glycoproteome — but how far? *Front. Cell Dev. Biol.* **2021**, *9*, 10.3389/fcell.2021.695970.
14. Noborn, F.; Gomez Toledo, A.; Sihlbom, C.; Lengqvist, J.; Fries, E.; Kjellén, L.; Nilsson, J.; Larson, G., Identification of chondroitin sulfate linkage region glycopeptides reveals prohormones as a novel class of proteoglycans. *Mol. Cell. Proteom.* **2015**, *14* (1), 41-49.
15. Delbaere, S.; De Clercq, A.; Mizumoto, S.; Noborn, F.; Bek, J. W.; Alluyn, L.; Gistelink, C.; Syx, D.; Salmon, P. L.; Coucke, P. J.; Larson, G.; Yamada, S.; Willaert, A.; Malfait, F., B3GALT6 knock-out zebrafish recapitulate β 3galt6-deficiency disorders in human and reveal a trisaccharide proteoglycan linkage region. *Front. Cell Dev. Biol.* **2020**, *8*, 10.3389/fcell.2020.597857.
16. Nasir, W.; Toledo, A. G.; Noborn, F.; Nilsson, J.; Wang, M.; Bandeira, N.; Larson, G., SweetNET: A bioinformatics workflow for glycopeptide MS/MS spectral analysis. *J. Proteome Res.* **2016**, *15* (8), 2826-2840.
17. Zhang, P.; Lu, H.; Peixoto, R. T.; Pines, M. K.; Ge, Y.; Oku, S.; Siddiqui, T. J.; Xie, Y.; Wu, W.; Archer-Hartmann, S.; Yoshida, K.; Tanaka, K. F.; Aricescu, A. R.; Azadi, P.; Gordon, M. D.; Sabatini, B. L.; Wong, R. O. L.; Craig, A. M., Heparan sulfate organizes neuronal synapses through neuroligin partnerships. *Cell* **2018**, *174* (6), 1450-1464.e23.
18. Noborn, F.; Gomez Toledo, A.; Nasir, W.; Nilsson, J.; Dierker, T.; Kjellén, L.; Larson, G., Expanding the chondroitin glycoproteome of *Caenorhabditis elegans*. *J. Biol. Chem.* **2018**, *293* (1), 379-389.
19. Takemura, M.; Noborn, F.; Nilsson, J.; Bowden, N.; Nakato, E.; Baker, S.; Su, T.-Y.; Larson, G.; Nakato, H., Chondroitin sulfate proteoglycan windpipe modulates hedgehog signaling in drosophila. *Mol. Biol. Cell.* **2020**, *31* (8), 813-824.
20. Nikpour, M.; Nilsson, J.; Persson, A.; Noborn, F.; Vorontsov, E.; Larson, G., Proteoglycan profiling of human, rat and mouse insulin-secreting cells. *Glycobiology* **2021**, *31* (8), 916-930.
21. Tomlinson, A.; Semblat, J. P.; Gamain, B.; Chêne, A., VAR2CSA-mediated host defense evasion of plasmodium falciparum infected erythrocytes in placental malaria. *Front. Immunol.* **2020**, *11*, 10.3389/fimmu.2020.624126.
22. Toledo, A. G.; Pihl, J.; Spliid, C. B.; Persson, A.; Nilsson, J.; Pereira, M. A.; Gustavsson, T.; Choudhary, S.; Zarni Oo, H.; Black, P. C.; Daugaard, M.; Esko, J. D.; Larson, G.; Salanti, A.; Clausen, T. M., An affinity chromatography and glycoproteomics workflow to profile the

chondroitin sulfate proteoglycans that interact with malarial VAR2CSA in the placenta and in cancer. *Glycobiology* **2020**, *30* (12), 989-1002.

23. Ludwigs, U.; Elgavish, A.; Esko, J. D.; Meezan, E.; Rodén, L., Reaction of unsaturated uronic acid residues with mercuric salts. Cleavage of the hyaluronic acid disaccharide 2-acetamido-2-deoxy-3-O-(β -D-glucopyranosyluronic acid)-D-glucose. *Biochem. J.* **1987**, *245* (3), 795-804.

24. Fries, E.; Blom, A. M., Bikunin--not just a plasma proteinase inhibitor. *Int. J. Biochem. Cell Biol.* **2000**, *32* (2), 125-37.

25. Bedini, E.; Laezza, A.; Parrilli, M.; Iadonisi, A., A review of chemical methods for the selective sulfation and desulfation of polysaccharides. *Carbohydr. Polym.* **2017**, *174*, 1224-1239.

26. Hanson, S. R.; Best, M. D.; Wong, C.-H., Sulfatases: Structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem., Int. Ed.* **2004**, *43* (43), 5736-5763.

27. Endo, M.; Kakizaki, I., Synthesis of neoproteoglycans using the transglycosylation reaction as a reverse reaction of endo-glycosidases. *Proc. Jpn. Acad., Ser. B* **2012**, *88* (7), 327-344.

28. Saitoh, H.; Takagaki, K.; Majima, M.; Nakamura, T.; Matsuki, A.; Kasai, M.; Narita, H.; Endo, M., Enzymic reconstruction of glycosaminoglycan oligosaccharide chains using the transglycosylation reaction of bovine testicular hyaluronidase. *J. Biol. Chem.* **1995**, *270* (8), 3741-7.

29. Griffin, M. E.; Jensen, E. H.; Mason, D. E.; Jenkins, C. L.; Stone, S. E.; Peters, E. C.; Hsieh-Wilson, L. C., Comprehensive mapping of O-GlcNAc modification sites using a chemically cleavable tag. *Mol. BioSyst.* **2016**, *12* (6), 1756-1759.

30. Zhang, C.; Dai, P.; Vinogradov, A. A.; Gates, Z. P.; Pentelute, B. L., Site-selective cysteine-cyclooctyne conjugation. *Angew. Chem., Int. Ed.* **2018**, *57* (22), 6459-6463.

31. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **1996**, *68* (5), 850-8.

32. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophor.* **1999**, *20* (18), 3551-67.

APPENDIX: SUPPLEMENTARY FIGURES, SCHEMES AND TABLES

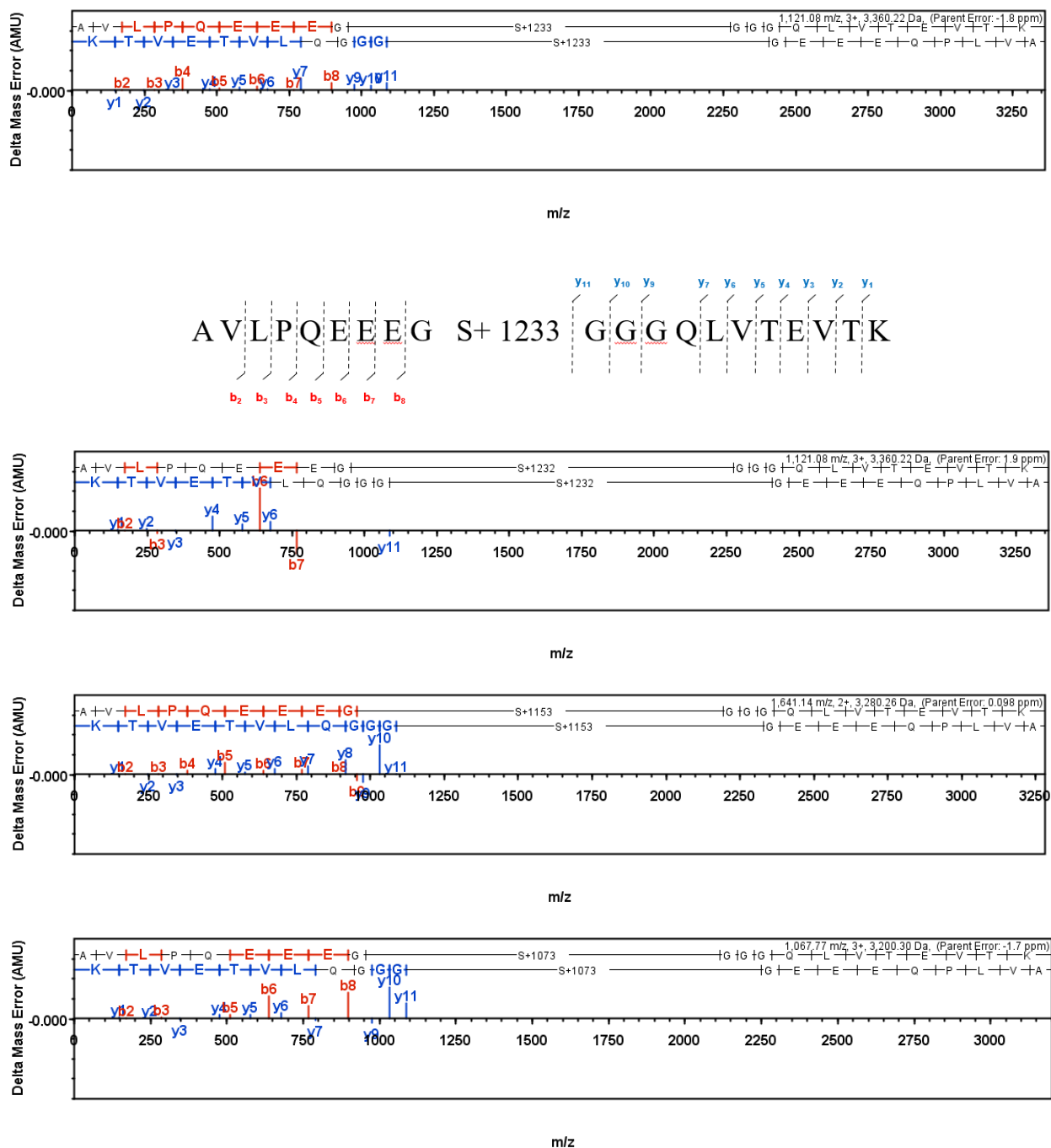


Figure 3.9. Modification of bikunin S10 found on protein 1 after chondroitinase and trypsin digestion, S+1233, S+1232, S+1153, S+1073 were found on peptide

Figure 3.9. (cont'd)

AVLPQEEEGSGGGQLVTEVTK, respectively. Blue color indicates fragment that is found from N to C terminal; red color indicates fragment that is found from C to N terminal. MS fragmentation shows the parent ion m/z 1067.77 (3+) was found, and the corresponding peptide/amino acid fragments, indicating S+ 1233.

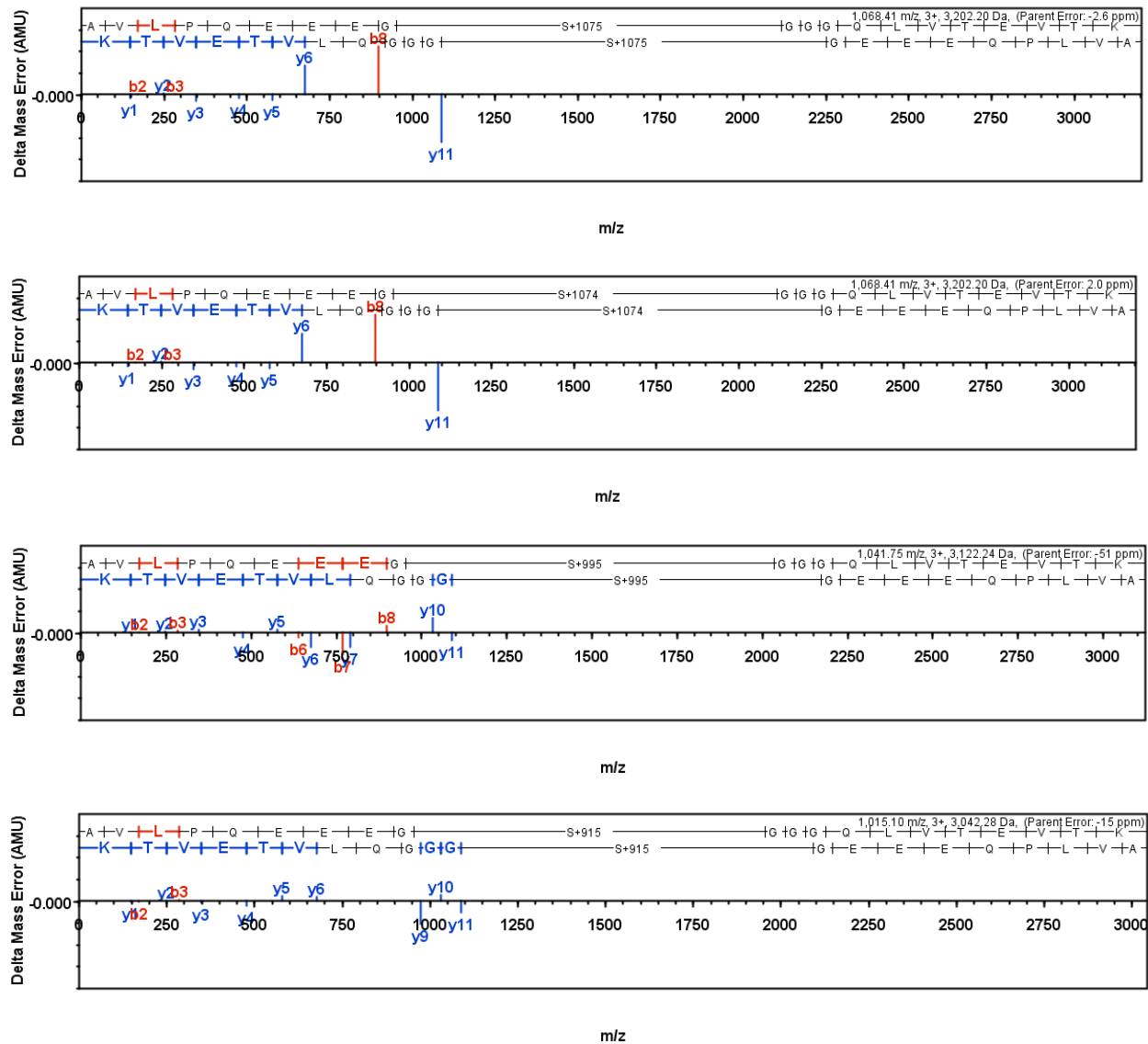


Figure 3.10. Modification of bikunin S10 found on protein 2 after mercuric acetate treatment and trypsin digestion, S+1075, S+1074, S+995, S+915 were found on peptide AVLPEEEGSGGGQLVTEVTK, respectively. Blue color indicates fragment that is found from N to C terminal; red color indicates fragment that is found from C to N terminal.

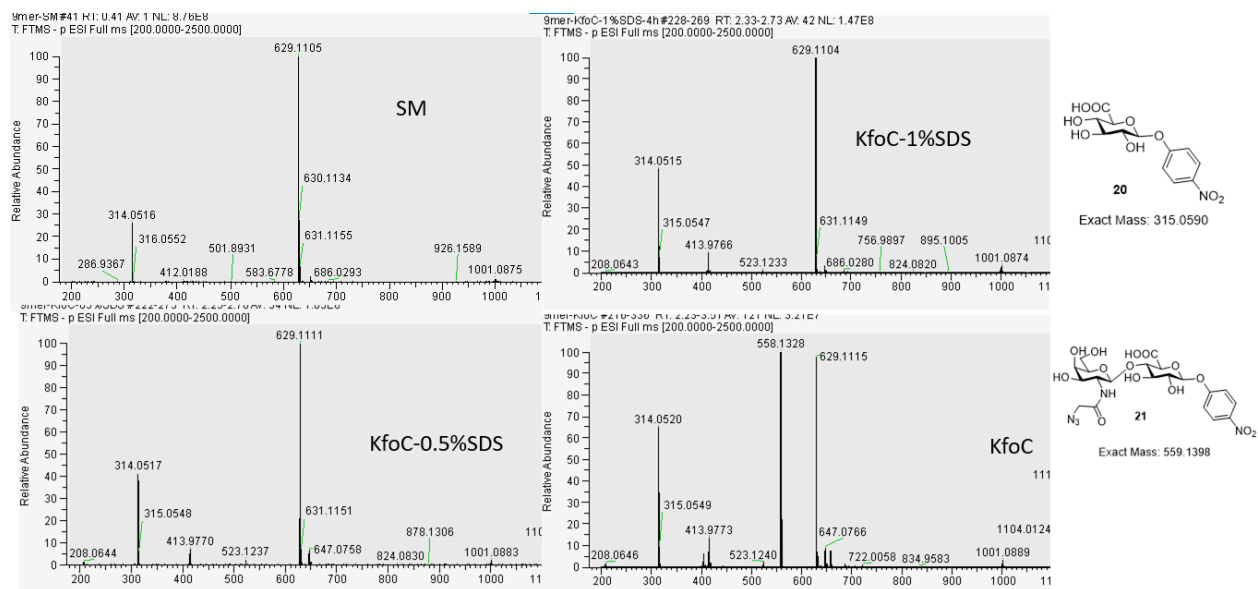


Figure 3.11. KfoC reactivity experiment with or without SDS with compound **20** as the acceptor and UDP-GalNAz as the donor. Compound **20** $[M - H]^-$ calc. 314.0520, found 314.0517; Compound **21** $[M - H]^-$ calc. 558.1328, found 558.1328.

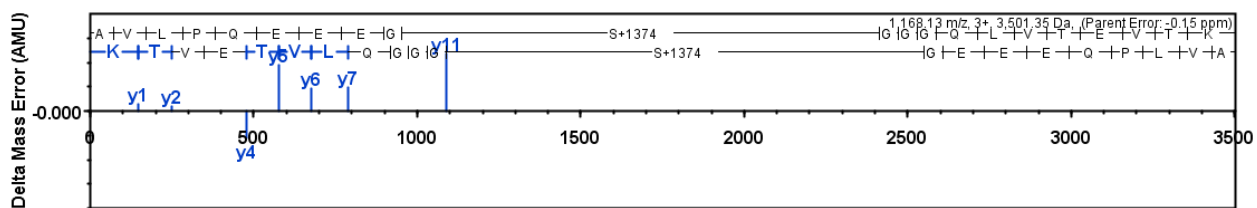
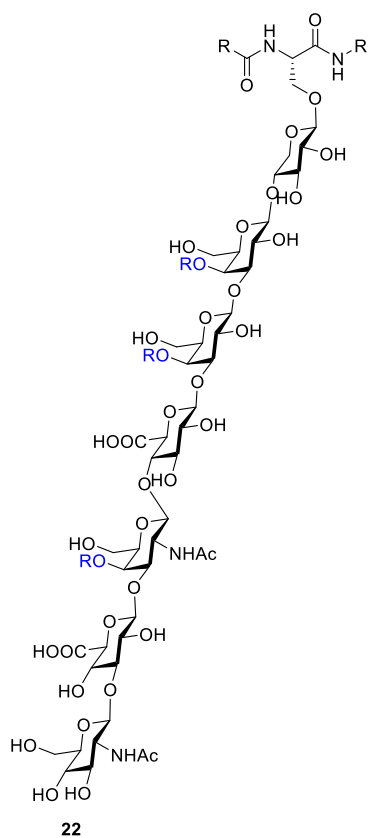


Figure 3.13. Modification of bikunin S10 found on protein **13** after hyaluronidase digestion, PmHS2/UDP-GlcNAc reaction and trypsin digestion, S+1374 was found on peptide AVLPQEEEGSGGGQLVTEVTK. Blue color indicates fragment that is found from N to C terminal; red color indicates fragment that is found from C to N terminal.



R = H or SO₃H

2 SO₃H on 22

+1374

Figure 3.14. Sulfation of bikunin S10 found on protein **13** after PmHS2/UDP-GlcNAc reaction.

S+ 1374 was found on peptide AVL PQEEEGSGGGQLVTEVTK.

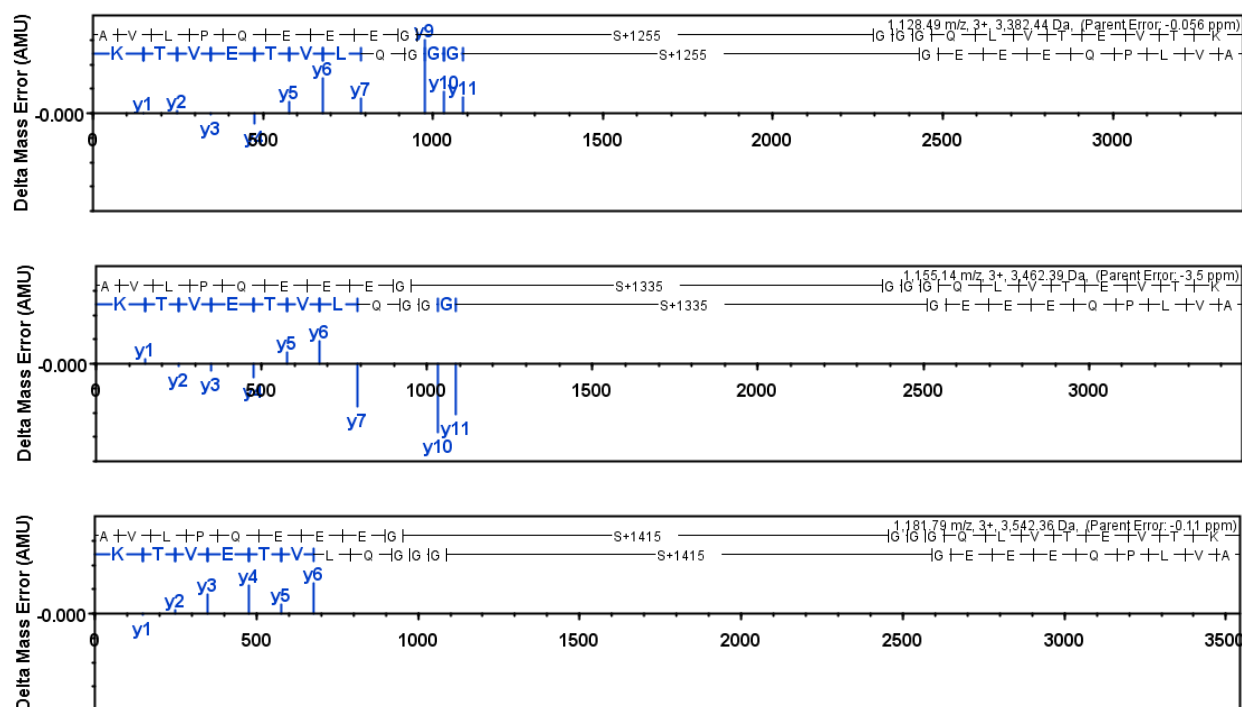
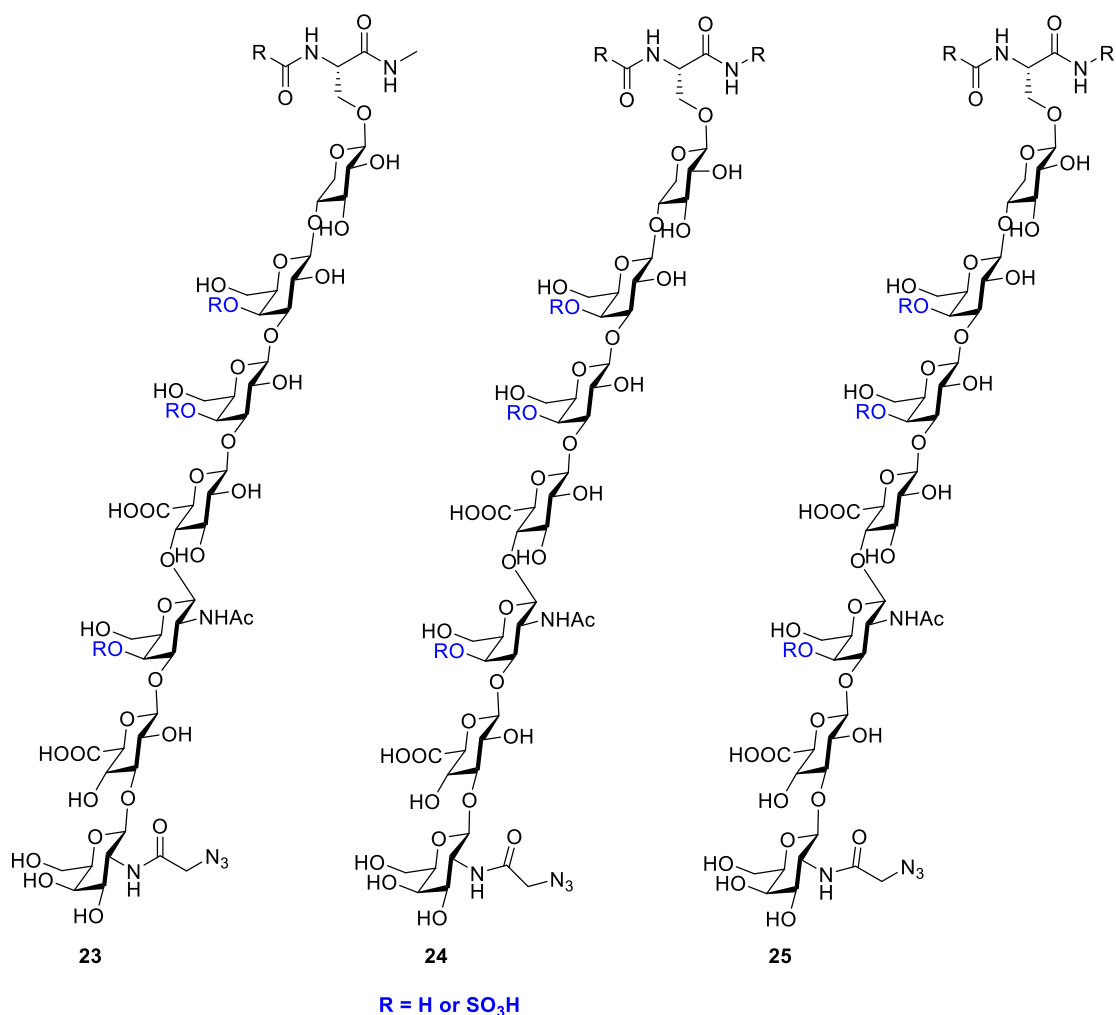


Figure 3.15. Modification of bikunin S10 found on protein **14** after hyaluronidase digestion, PmHS2/UDP-GlcNAz reaction and trypsin digestion, S+1255, S+1335, S+1415 were found on peptide AVL PQEEEGSGGGQLVTEVTK, respectively. Blue color indicates fragment that is found from N to C terminal; red color indicates fragment that is found from C to N terminal.



0 SO₃H on **23**

+1255

1 SO₃H on **24**

+1335

2 SO₃H on **25**

+1415

Figure 3.16. Sulfation of bikunin S10 found on protein **14** after PmHS2/UDP-GlcNAz reaction. S + 1255, S +1335, S +995 and S +1415 were found on peptide AVL PQEEEGSGGGQLVTEVTK, respectively.

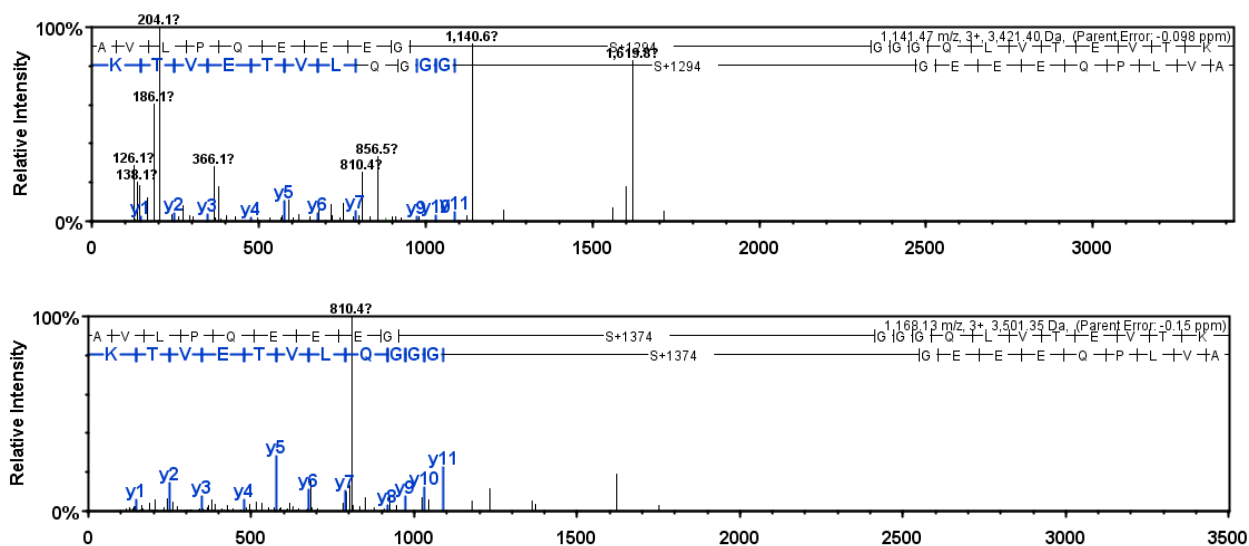
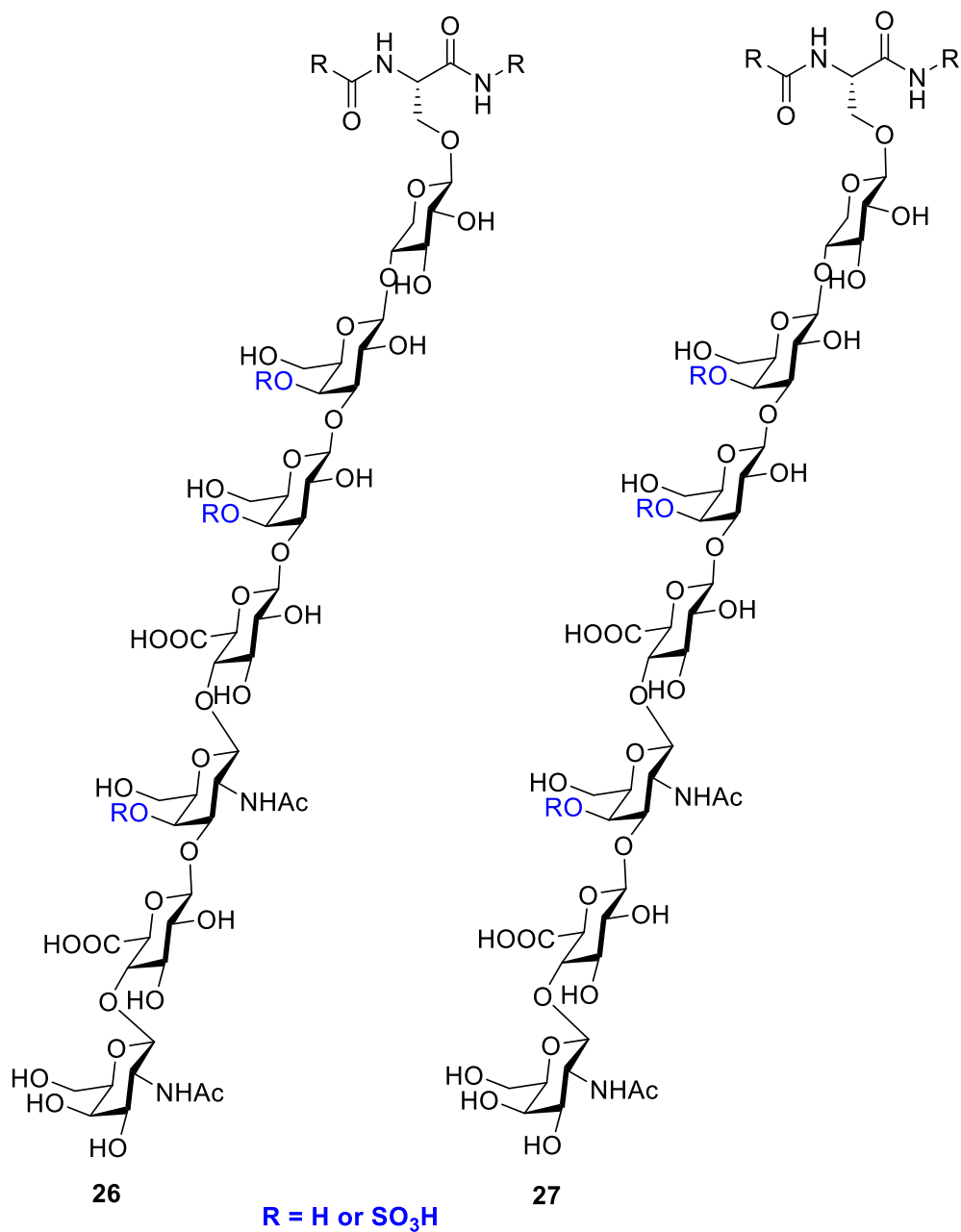


Figure 3.17. Modification of bikunin S10 found on protein **15** after hyaluronidase digestion, KfoC/UDP-GalNAc reaction and trypsin digestion, S+1294, S+1374 were found on peptide AVLPPQEEEGSGGGQLVTEVTK, respectively. Blue color indicates fragment that is found from N to C terminal; red color indicates fragment that is found from C to N terminal.



1 SO₃H on **26**

+1294

2 SO₃H on **27**

+1374

Figure 3.18. Sulfation of bikunin S10 found on protein **15** after KfoC/UDP-GalNAc reaction. S+ 1294 and S+ 1374 were found on peptide AVL PQEEEEGSGGGQLVTEVTK, respectively.

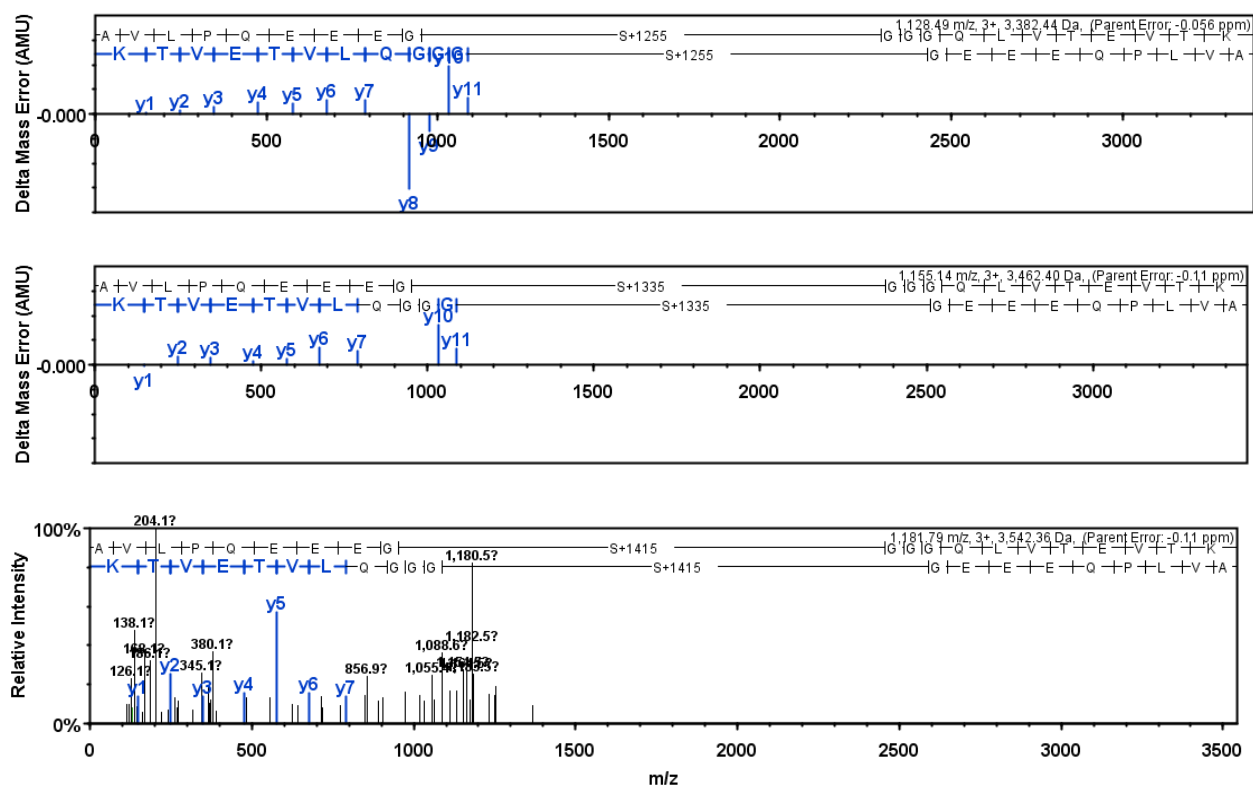
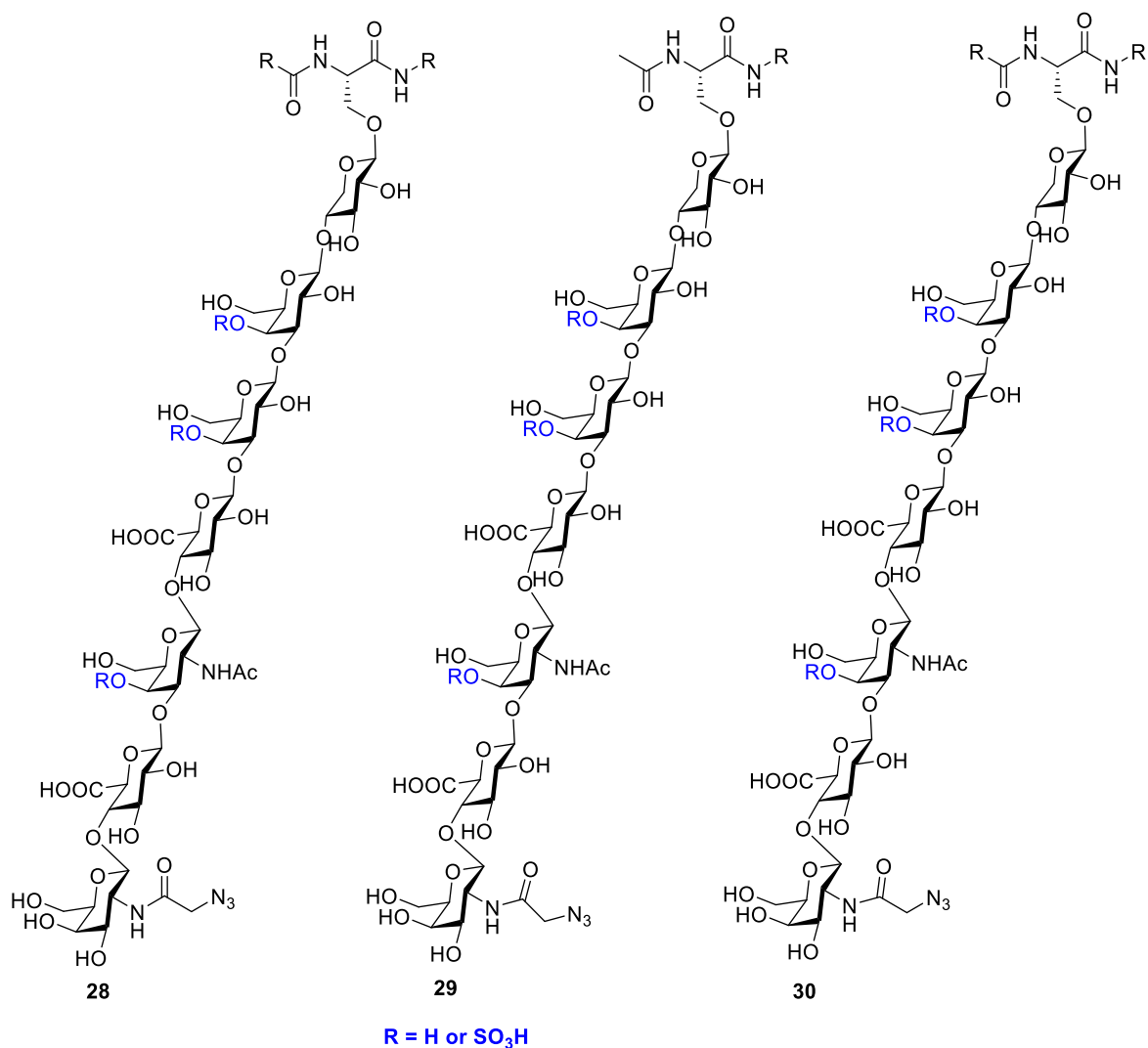


Figure 3.19. Modification of bikunin S10 found on protein **16** after hyaluronidase digestion, KfoC/UDP-GalNAz reaction and trypsin digestion, S+1255, S+1335, S+1415 were found on peptide AVLPLQEEEGSGGGQLVTEVTK, respectively. Blue color indicates fragment that is found from N to C terminal.



0 SO₃H on 28
+1255

1 SO₃H on 29
+1335

2 SO₃H on 30
+1415

Figure 3.20. Sulfation of bikunin S10 found on protein **16** after KfoC/UDP-GalNAz reaction. S+ 1255, S+ 1335 and S+1415 were found on peptide AVL PQEE GQLVTEVTK, respectively.

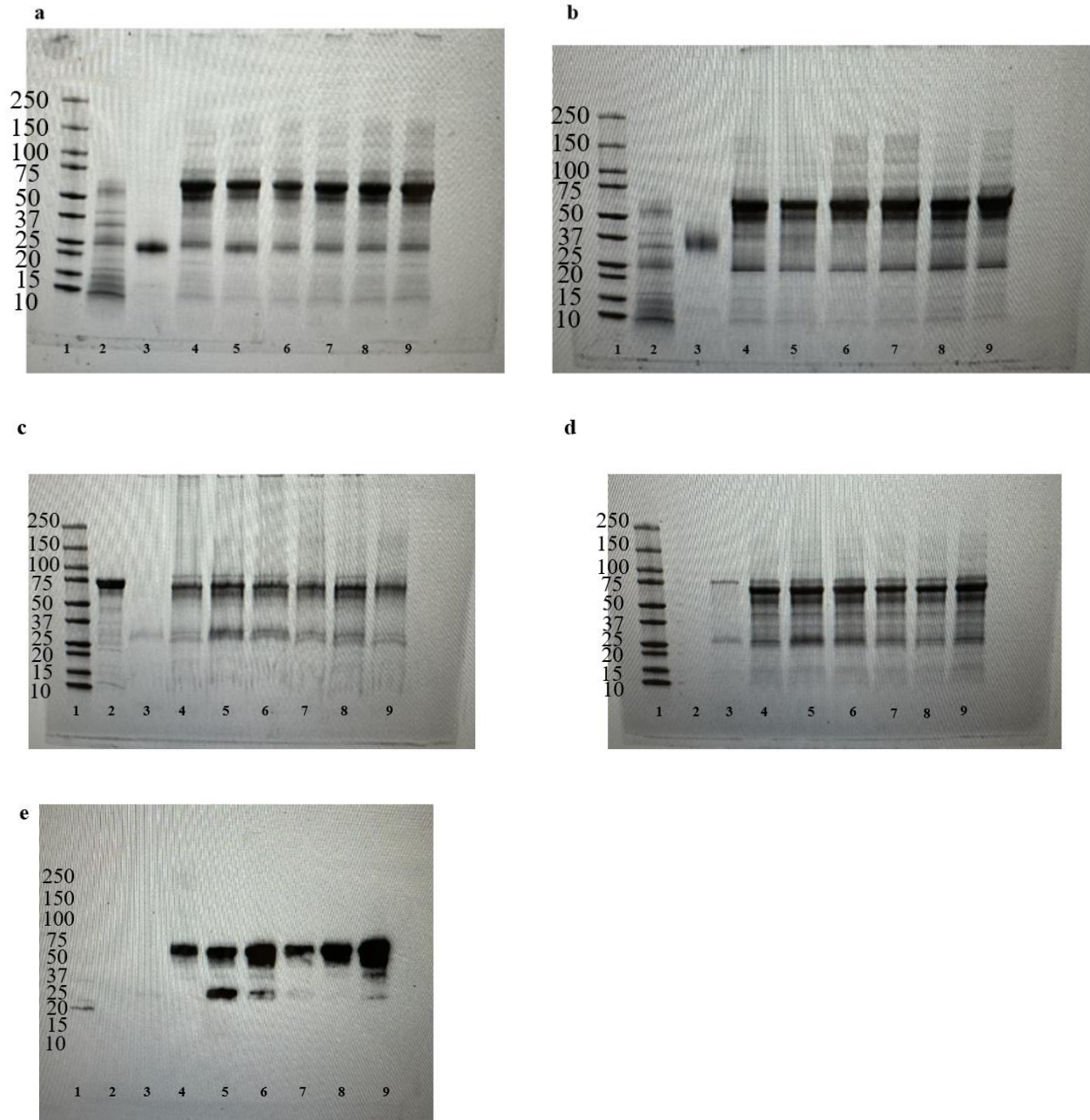


Figure 3.21. SDS PAGE and Western blot of bikunin, human plasma, and bikunin spiked human plasma incubated with hyaluronidase, KfoC/UDP-GalNAz and Dde-alkyne-biotin. Western blot (e, Lane 4 - 9) shows the signal of bikunin (around 25 kDa marker) and other biotinylated proteins pulled from human plasma, suggesting the existence of CSPG.

Figure 3.21. (cont'd)

(a) SDS page of starting materials. Lane 1, ladder; Lane 2, hyaluronidase-bovine tests; Lane 3, native bikunin; Lane 4, 200 μ g human plasma; Lane 5, 200 μ g human plasma + 50 μ g native bikunin; Lane 6, 200 μ g human plasma + 20 μ g native bikunin; Lane 7, 200 μ g human plasma + 10 μ g native bikunin; Lane 8, 200 μ g human plasma + 1 μ g native bikunin; Lane 9, 200 μ g human plasma + 0.1 μ g native bikunin.

(b) SDS page of starting materials incubated with hyaluronidase at 37 °C for 2h. Lane 1, ladder; Lane 2, hyaluronidase-bovine tests; Lane 3, native bikunin; Lane 4, 200 μ g human plasma; Lane 5, 200 μ g human plasma + 50 μ g native bikunin; Lane 6, 200 μ g human plasma + 20 μ g native bikunin; Lane 7, 200 μ g human plasma + 10 μ g native bikunin; Lane 8, 200 μ g human plasma + 1 μ g native bikunin; Lane 9, 200 μ g human plasma + 0.1 μ g native bikunin.

(c) SDS page of hyaluronidase treated materials incubated with KfoC/UDP-GalNAz at 4 °C for overnight. Lane 1, ladder; Lane 2, KfoC; Lane 3, native bikunin; Lane 4, 200 μ g human plasma; Lane 5, 200 μ g human plasma + 50 μ g native bikunin; Lane 6, 200 μ g human plasma + 20 μ g native bikunin; Lane 7, 200 μ g human plasma + 10 μ g native bikunin; Lane 8, 200 μ g human plasma + 1 μ g native bikunin; Lane 9, 200 μ g human plasma + 0.1 μ g native bikunin.

(d) SDS page of KfoC treated materials incubated with Dde-alkyne-biotin at RT overnight. Lane 1, ladder; Lane 2, KfoC; Lane 3, native bikunin; Lane 4, 200 μ g human plasma; Lane 5, 200 μ g human plasma + 50 μ g native bikunin; Lane 6, 200 μ g human plasma + 20 μ g native bikunin; Lane 7, 200 μ g human plasma + 10 μ g native bikunin; Lane 8, 200 μ g human plasma + 1 μ g native bikunin; Lane 9, 200 μ g human plasma + 0.1 μ g native bikunin.

(e) Western blot of KfoC treated materials incubated with Dde-alkyne-biotin at RT overnight, Streptavidin-HRP as antibody. Lane 1, ladder; Lane 2, KfoC; Lane 3, native bikunin; Lane 4,

Figure 3.21. (cont'd)

200 μg human plasma; Lane 5, 200 μg human plasma + 50 μg native bikunin; Lane 6, 200 μg human plasma + 20 μg native bikunin; Lane 7, 200 μg human plasma + 10 μg native bikunin; Lane 8, 200 μg human plasma + 1 μg native bikunin; Lane 9, 200 μg human plasma + 0.1 μg native bikunin.

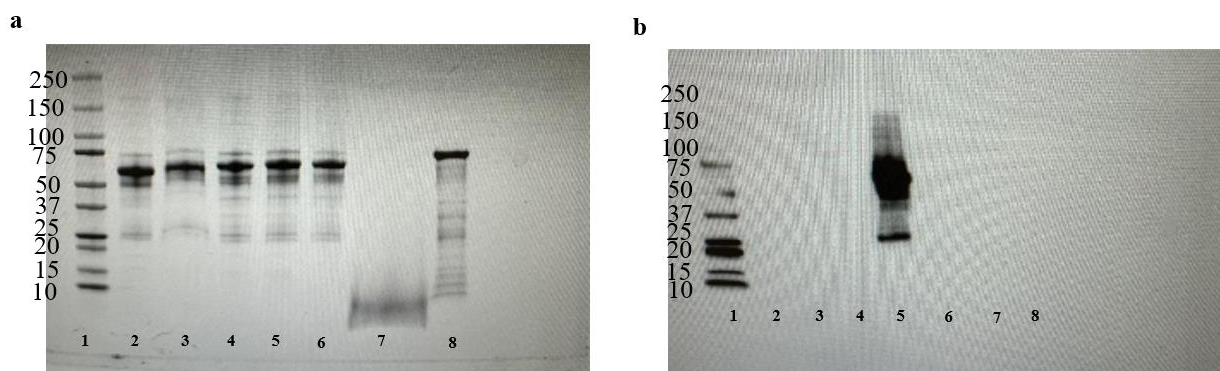


Figure 3.22. SDS page (a) and western blot (b) of human plasma incubated with hyaluronidase, KfoC/UDP-GalNAz and Dde-alkyne-biotin. Western blot (b, Lane 5) shows the signal of biotinylated proteins pulled from human plasma, suggesting the existence of CSPG in human plasma sample.

(a) SDS page of the CSPG enrichment experiment. Lane 1, ladder; Lane 2, human plasma; Lane 3, human plasma incubated with hyaluronidase at 37 °C for 2h; Lane 4, human plasma incubated with KfoC/UDP-GalNAz at 4 °C overnight; Lane 5, 2 human plasma incubated with Dde-alkyne-biotin at RT overnight; Lane 6, Flow through of Lane 5 after incubated with neutravidin beads ; Lane 7, Cleavage product after treating the neutravidin beads with 2% hydrazine; Lane 8, Mixture of KfoC and hyaluronidase.

(b) Western blot of the CSPG enrichment experiment, Streptavidin-HRP as antibody. Lane 1, ladder; Lane 2, human plasma; Lane 3, human plasma incubated with hyaluronidase at 37 °C for 2h; Lane 4, human plasma incubated with KfoC/UDP-GalNAz at 4 °C overnight; Lane 5, 2 human plasma incubated with Dde-alkyne-biotin at RT overnight; Lane 6, Flow through of Lane 5 after incubated with neutravidin beads ; Lane 7, Cleavage product after treating the neutravidin beads with 2% hydrazine; Lane 8, Mixture of KfoC and hyaluronidase.