CHEMOENZYMATIC SYNTHESIS OF HEPARAN SULFATE PROTEOGLYCAN AND MIMETICS

By

Jia Gao

A DISSERTATION

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

Chemistry–Doctor of Philosophy

2021

ABSTRACT

CHEMOENZYMATIC SYNTHESIS OF HEPARAN SULFATE PROTEOGLYCAN AND MIMETICS

By

Jia Gao

Proteoglycans (PGs) are an important class of glycoproteins widely distributed in mammals. They are involved in numerous biological events, including tumor progression, inflammation, and cellular communication. Generally, a PG is composed of a core protein and one or more glycosaminoglycan (GAG) polysaccharide chains. The GAG chain is covalently attached to the core protein via a serine residue in the consensus sequence -Ser-Gly-X-Gly- (X being any natural amino acid residue but proline) by a common tetrasaccharide linkage. Heparan sulfate proteoglycans (HSPGs), along with chondroitin sulfate proteoglycans (CSPGs) and keratan sulfate proteoglycans (KSPGs), are main subtypes of the PG family. Naturally existing HSPGs, due to complex post-translational modifications (PTMs) on the GAG chains, are highly heterogeneous. That makes direct isolation of homogeneous HSPGs from natural sources almost impossible. To date, preparing structurally defined HSPGs solely relies on formidable and tedious chemical synthesis.

In this dissertation, two novel approaches have been investigated to expedite the synthesis of HSPGs. The convergent chemoenzymatic approach takes advantage of efficient enzymatic synthesis of heparan sulfate (HS) oligosaccharides and well-developed solid phase supported peptide synthesis (SPPS). By substituting the non-functional tetrasaccharide linkage, the GAG chain and peptide were conjugated through a flexible artificial linker to make a syndecan-1 mimetic, which mimics the natural structures of syndecan-1, an important member of HSPG family. The mimetic binds strongly to integrin αvβ3, a key cell-surface protein that plays an active

role in tumor proliferation process. Furthermore, the mimetic compound is able to inhibit the migration of breast cancer cells MDA-MB-231.

In the native form of PGs, the core protein and GAG chains are connected through a common tetrasaccharide linkage consisted of GlcA-β(1→3)-Gal-β(1→3)-Gal-β(1→4)-Xyl-β(1→O)-Ser to efficiently prepare native heparan sulfate glycopeptides and glycoproteins, enzymes involved in the PG linkage biosynthesis were investigated and developed as synthetic tools. Human β-1,4-galactosyltransferase 7 (β4GalT7) was used to catalyze the transfer of galactose units and synthesize galactose-xylose (Gal-Xyl) bearing PG glycopeptides. Human xylosyltransferase I (XT-I), the enzyme that initiates PG biosynthesis in nature, was then studied and applied towards the synthesis of PG linkage region.

ACKNOWLEDGEMENTS

As I am reaching the final milestone of my PhD journey, it is time for me to express my gratitude. First of all, I would like to say thank you to my academic advisor Prof. Xuefei Huang, for his prodigious patience throughout my eight-year's adventure. Without the research freedom allowed by him, I would not be able to have the opportunity to develop an interdisciplinary skill set. The diversified research projects challenged me to become better and eventually qualified for graduation. I would also like to appreciate all my guidance committee members, Prof. William Wulff, Prof. Babak Borhan and Prof. Kevin Walker. Their thoughtful advice and comments have inspired me to grow myself as an independent research scientist.

The technical support I have received from research staff in the department is vital. It is their help that enabled me to make all the progress towards graduation. I appreciate Dr. Daniel Holmes and Dr. Li Xie for their generous support on NMR training, data collection, and all the free NMR tubes as gifts. My appreciation also goes to Prof. Daniel Jones and Dr. Tony Schilmiller for high-resolution mass spectrometer training and routine technical support. Their technical insights have helped me overcome obstacles in my research.

Besides, I am grateful to all the help from research collaborators. The contributions from Prof. Jian Liu, Dr. Yongmei Xu, Prof. Ulf Ellervik, Dr. Emil Tykesson, Prof. Lingjun Li, Dr. Junfeng Huang, Prof. Kefei Yu, and Prof. Erhard Hohenester are essential to the successes of my research projects.

In addition, I would like to express my appreciation to Huang group members, including former lab members, Dr. Bo Yang, Dr. Steven Dulaney, Dr. Hovig Kouyoumdjian, Dr. Herbert Kavunja, Dr. Zhaojun Yin, Dr. Suttipun Sungsuwan, Dr. Qian Qin, Dr. Mehdi Hossaini Nasr, Dr.

Peng Wang, Dr. Xianwu Wang, Dr. Xuanjun Wu, Dr. Yuetao Zhao, Dr. Tianlu Li, Dr. Changxin Huo, Dr. Setare Nick, Dr. Jicheng Zhang, Dr. Kedar Baryal, Dr. Shuyao Lang, Hui Li and Zeren Zhang, and current lab members, Dr. Weizhun Yang, Dr. Sherif Ramadan, Dr. Vincent Shaw, Zahra Rashidijahanabad, Hunter McFall-Boegeman, Mengxia Sun, Zibin Tan, Shivangi Chugh, Kunli Liu, Po-han Lin, Cameron Talbot, Chia-wei Yang. Without your presence and efforts, the journey would not be the same. And I see my own growth. Alongside, I want to take the chance to thank my wellness coach, Kristin Traskie, friends at MSU, peers in student groups, and collaborators in scientific community. You all have made a difference in my life and I genuinely appreciate it.

Finally, I would like to thank all the support and love from my family. You are always there for me. And I know how proud you are to see I am ready to start a new chapter of my career and life. Thank you very much for what you have offered me. I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

KEY TO ABBREVIATIONS

6AzGlc	6-azidoglucose
AA	amino acid
ACN	acetonitrile
AcOH	acetic acid
Asn	asparagine
Asp	aspartic acid
β3GalT6	β-1,3-galactosyltransferase 6
β3GAT3	β-1,3-glucuronyltransferase 3
β4GalT7	β-1,4-galactosyltransferase 7
BLI	biolayer interferometry
$CaCl_2$	calcium(II) chloride
CF	conversion factor
CHO	Chinese hamster ovary
COSY	correlation spectroscopy
CSPG	chondroitin sulfate proteoglycan
CuAAC	copper(I)-catalyzed alkyne-azide cycloaddition
$CuSO_4$	copper(II) sulfate
Cy5	Cyanine5
D_2O	deuterium oxide
DCM	dichloromethane
DEAE	diethylaminoethyl cellulose
DIC	*N,N'*-diisopropylcarbodiimide

DIPEA	diisopropylethylamine
DMEM	Dulbecco's Modified Eagle Medium
DMF	dimethylformamide
DNA	deoxyribonucleic acid
E. coli	Escherichia coli
EGF	epidermal growth factor
ESI-MS	electrospray-ionization mass spectrometry
ETD	electron-transfer dissociation
EThcD	electron-transfer/higher-energy collision dissociation
FBS	fetal bovine serum
FDR	false discovery rate
GAG	glycosaminoglycan
Gal	galactose
Gal-Xyl	galactose-xylose
Glc	glucose
GlcA	glucuronic acid
GlcN	glucosamine
GlcNAc	*N*-acetamidoglucose
Glu	glucose
Gly	glycine
H_2O	water
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HBTU	*N,N,N′,N′*-tetramethyl-O-(1H-benzotriazol-1-yl) uronium hexafluorophosphate

HCD	higher-energy collision dissociation
HCl	hydrochloric acid
$HCOONH_4$	ammonium formate
HEK-293	human embryonic kidney 293
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HLB cartridge	hydrophilic-lipophilic-balanced cartridge
HMBC	heteronuclear multiple bond correlation
HOBt	*N*-hydroxybenzotriazole
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HS	heparan sulfate
HSPG	heparan sulfate proteoglycan
HSQC	heteronuclear single quantum correlation
IdoA	iduronic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
k_{cat}	catalyst rate constant
KCl	potassium chloride
kDa	kilo-dalton
KF	potassium fluoride
KI	potassium iodide
Km	Michaelis constant
k_{on}	on-rate of ligand-receptor binding
KPB	potassium phosphate buffer
KSPG	keratan sulfate proteoglycan

LB	Luria-Bertani
LC-MS	liquid chromatography-mass spectrometry
mAb	monoclonal antibody
MALDI-TOF	matrix assisted laser desorption ionization-time of flight
MBP	maltose-binding protein
MD	molecular dynamics
MeOH	methanol
MES	2-(*N*-morpholino)ethanesulfonic acid
$MgCl_2$	magnesium(II) chloride
$MnCl_2$	manganese(II) chloride
MS	mass spectrometry
MU	4-methylumbelliferyl
NaCl	sodium chloride
$NaHCO_3$	sodium bicarbonate
NaOH	sodium hydroxide
Nap	2-naphthyl
NHS	*N*-hydroxysuccinimide
$NiSO_4$	nickel(II) sulfate
NMR	nuclear magnetic resonance spectroscopy
OD	optical density
OD_{600}	optical density at 600 nm
OP2E	one-pot two-enzyme
OXT	*Drosophila* peptide *O*-xylosyltransferase
PBS	phosphate buffered saline

$Pd(OH)_2/C$	palladium hydroxide on carbon
PEI	polyethylenimine
PGs	proteoglycans
PMSF	phenylmethylsulfonyl fluoride
PSM	peptide spectrum match
PTFE	polytetrafluoroethylene
PTM	post-translational modification
RP-HPLC	reverse phase - high performance liquid chromatography
Rpm	round-per-minute
SA	streptavidin
SaOS-2	osteosarcoma
sat.	saturated
S.D.	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
Ser	serine
SorTag	the 'LPETG' sorting signal
SPPS	solid phase supported peptide synthesis
SQV-6	*Caenorhabditis* peptide *O*-xylosyltransferase
SrtA	sortase A
$SrtA_{staph}$	sortase A *from Staphylococcus aureus*
SSTN	synstatin
t-Bu	*t*-butyl
TEV	tobacco etch virus
TFA	trifluoroacetic acid

THPTA	tris-hydroxypropyltriazolylmethylamine
TIPS	triisopropylsilane
Tris	tris(hydroxymethyl)aminomethane
UDP	uridine diphosphate
UDP-XylAz	UDP-4-azido-4-deoxyxylose
UPLC	ultra performance liquid chromatography
UV	ultraviolet
μW	microwave
V_{max}	maximum velocity
VPA	valproic acid
XT-I/II	xylosyltransferase-I/II
Xyl	xylose
Xylo_C domain	*C*-terminal domain

Chapter 1 Recent Advances on Glycosyltransferases Involved in the Biosynthesis of Proteoglycan Linker Region

1.1 Introduction

Proteoglycans are an essential family of glycoproteins consisting of a core protein with one or multiple glycosaminoglycan (GAG) chains, which are covalently attached to the protein through a common tetrasaccharide linkage consisted of GlcA-β(1→3)-Gal-β(1→3)-Gal-β(1→4)-Xyl-β(1→O)-Ser (**Figure 1.1**). PGs are widely present on cell surface and extracellular matrix. Their functions are critically important to numerous biological events, including cell adhesions, cellular signaling and interactions with growth factors.[1-4]

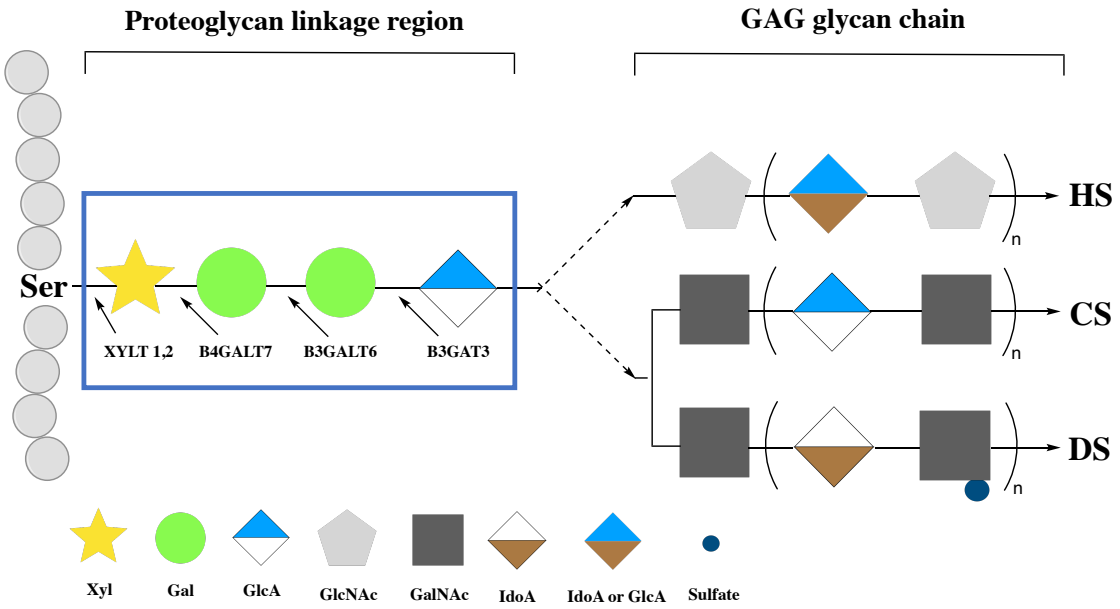


Figure 1.1 Schematic demonstration of the structure of proteoglycans. The tetrasaccharide linkage is highlighted.[5]

The biosynthesis of the PG linkage tetrasaccharide involves the deployment of four glycosyl transferases: xylosyltransferase-I/II (XT-I/II), β-1,4-galactosyltransferase 7 (β4GalT7), β-1,3-galactosyltransferase 6 (β3GalT6) and β-1,3-glucuronyltransferase 3 (β3GAT3) (**Figure**

1.2). The first successful expressions and characterizations of β3GalT6 were reported by the Furukawa and Esko groups two decades ago.[6, 7] The Sugahara group reported the first molecular cloning and expression of β3GAT3, and subsequent characterizations of this enzyme in 1990s.[8, 9] The follow-up investigations on β3GalT6 and β3GAT3 have been rather limited.[10, 11] Therefore, this current review will focus on the recent progress made on the expression, characterization and applications of the PG linkage glycosyltransferases XT-I/II and β4GalT7.

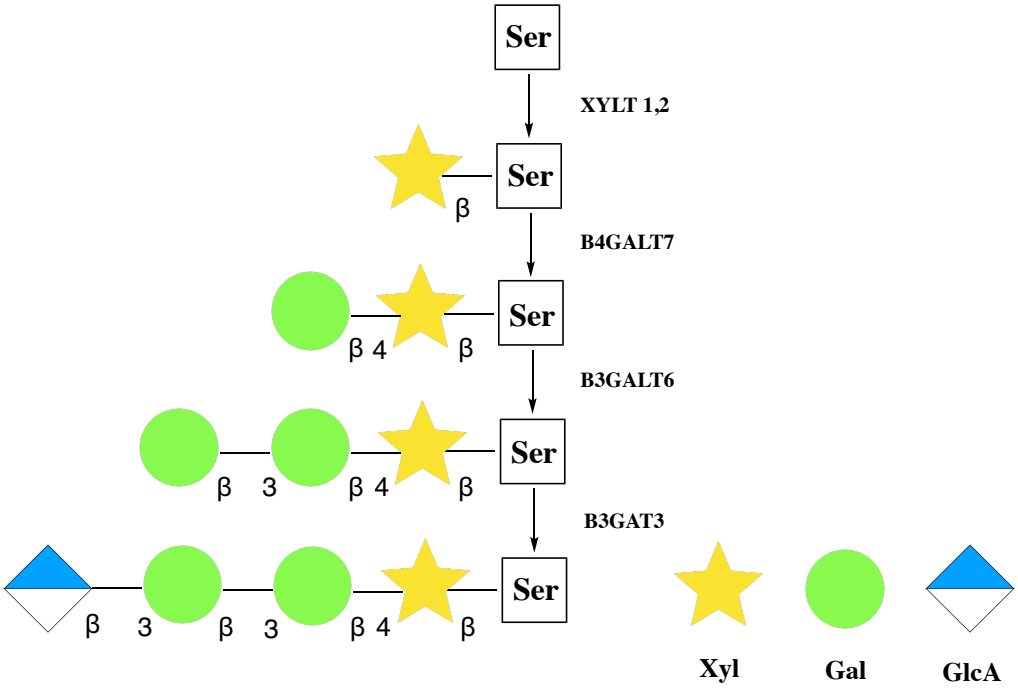


Figure 1.2 Biosynthetic assembly of the PG linkage region.[12]

1.2 Xylosyltransferase-I/II (XT-I/II)

To the best of my knowledge, the review article published by Wilson in 2004 is the first to comprehensively summarize the contemporary understandings towards UDP-α-D-xylose:proteoglycan core protein β-D-xylosyltransferases (XT-I and XT-II).[13] In 2007, Götting, Kuhn, and Kleesiek published a review emphasizing the impact of mammalian xylosyltransferases

on PG-related diseases and human health.[14] Since then, significant amounts of progress have been made to gain insights on this key enzyme.

1.2.1 Expression and Purification of XT-I/II

The discovery of peptide *O*-xylosyltransferase dates back to the 1960s.[15-19] Afterwards, this GAG-synthesis-initiating enzyme has been isolated from multiple sources.[15-24] In 2000, Götting and co-workers reported the first molecular cloning and expression of XT-I and its isoform.[25] In their study, the recombinant XT-I proteins from humans, mice and rats were successfully expressed in Chinese Hamster Ovary (CHO-K1) cells.

In 2003, the Kleesiek group described high-level expression of a soluble histidine-tagged recombinant XT-I using the High Five/pCG255-1 insect cell expression system.[26] Stable clones that express XT-I-V5-His (rXT-I-His) were generated. The human XT-I was purified by heparin affinity chromatography using a POROS 20 HE2 column followed by Nickel affinity column. The purified protein was verified by Western blot using polyclonal anti-XT-I antibodies.

Shortly after, Götting and co-workers prepared a series of XT-I enzymes with point mutations on the aspartate-any residue-aspartate (DXD) motifs by transient expression in High Five insect cells.[27] A stable clone of High Five/pCG255-1 that expresses the soluble form of histidine- and V5-tagged recombinant human XT-I with *N*-terminal 1-148 sequence truncated, rXT-I-(Δ1–148)-V5-His, was also made in this study.

Müller et al., in 2005, carried out individual site-directed mutagenesis of all 14 cysteine residues into alanine.[28] The recombinant wild-type human XT-I and the single mutants were successfully expressed in High Five insect cells to assist the structure-activity study of XT-I. A year later, in the work published by the same group, multiple *N*-terminal truncated human XT-I enzymes were smoothly produced with the same insect cell expression system.

With the successes from CHO mammalian cell and High Five insect cell expression system, the expressions of xylosyltransferases were extended to the human embryonic kidney 293 (HEK-293), human osteosarcoma (SaOS-2) mammalian system, and *Pichia pastoris* yeast system.[29, 30] In 2006, Götting group reported the first recombinant expressions of GFP-fused human XT-I and multiple GFP-tagged XT-I/II mutants using mammalian HEK-293 and SaOS-2 cells.[29] In the same year, Brunner et al. expressed two invertebrate and two vertebrate xylosyltransferases, *Drosophila* peptide *O*-xylosyltransferase (OXT), *Caenorhabditis* peptide *O*-xylosyltransferase (SQV-6), and human xylosyltransferase I/II (XT-I/II), with *Pichia pastoris* expression system.[30] Two years later, another successful story with *Pichia pastoris* expression system was reported by the Götting group.[31]

1.2.2 Acceptor Specificity of XT-I/II

The first description of the acceptors for XT-I dates back to roughly five decades ago.[15-17, 32-34] In the pioneering studies, various uncharacterized exogenous or endogenous proteins were validated to be acceptors of xylosyltransferases. Since then, understandings on the acceptor specificity of XT-I/II have been significantly expanded.

In addition to acceptor proteins, diverse peptide acceptors have been derived from the amino acid sequence around glycosaminoglycan attachment sites of different proteoglycans.[13, 20, 21, 30, 32, 34-39] Among the reported acceptors of XT-I/II, bikunin protein is known to be one of the best acceptors based on the Michaelis-Menten constants (K_m). The bikunin peptide sequence derived from the bikunin GAG-attachment site has later on been extensively used to study the acceptor recognition properties of XT-I/II.[20, 25, 30, 31, 35, 37, 40-42]

As the acceptor scope of XT-I expands, considerable effort has been put to determine its minimal binding motif, Gly-Ser-Gly or Ser-Gly-x-Gly, where x = any amino acid.[14, 38, 43-46]

Meanwhile, some evidence indicates that the presence of serine residue may not be absolutely required.[35, 47] Beyond the minimal motif of acceptor binding, a consensus favored acceptor sequence for XT-I, a-a-a-a-Gly-Ser-Gly-a-b-a, where 'a' being Glu or Asp and 'b' being Gly, Glu or Asp, was deduced by Brinkmann and co-workers in 1997, based on the peptide sequence of reported acceptors of xylosyltransferases.[20] Shortly thereafter, the common sequence was refined by the same research group to a-a-a-x-Ser-Gly-x-Gly, where a = Glu or Asp and x = any amino acid.[21]

With the successful expression of XT-1, research focus was subsequently extended to XT-II. Roch and co-workers discovered that XT-II possesses a consensus sequence analogous to that for XT-I, a-a-a-a-Gly-Ser-Gly-a-a/Gly-a, where a = Asp or Glu.[42]

Lately, to investigate the acceptor recognition property of XT-I, Briggs and Hohenester performed detailed analysis using a comprehensive bikunin-derived 12-amino-acid peptide acceptor library in which the amino acid residue at each position had been mutated to one of all the 20 common natural amino acids.[41] Although a serine residue is highly preferred at the xylosylation site, peptides with a threonine residue at position 0 also show noticeable activity levels. The -1 position, originally a glycine, can accept a wide variety of uncharged amino acids. While the -2, -3 and -4 sites generally favor acidic amino acids, individual replacement of the glutamic acid residues does not exert strong influence on the enzymatic activity. The preference for the acidic amino acids at positions preceding the xylosylation site has been attributed to non-specific charge-charge interactions with the positively charged residues around the binding pocket. For the +1 position, small amino acids including glycine, alanine, serine and threonine are strongly favored. Surprisingly, a valine residue at +2 site enhances the activity level considerably, as opposed to the native glycine. Overall, XT-I does not strictly require a certain acceptor peptide

sequence for the activity and exhibits a greater structure tolerance than previously described (**Figure 1.3**). This recent discovery furthers contemporary understanding towards XT-I acceptor recognition properties and implies vast application potentials attributing to the relaxed acceptor requirements.

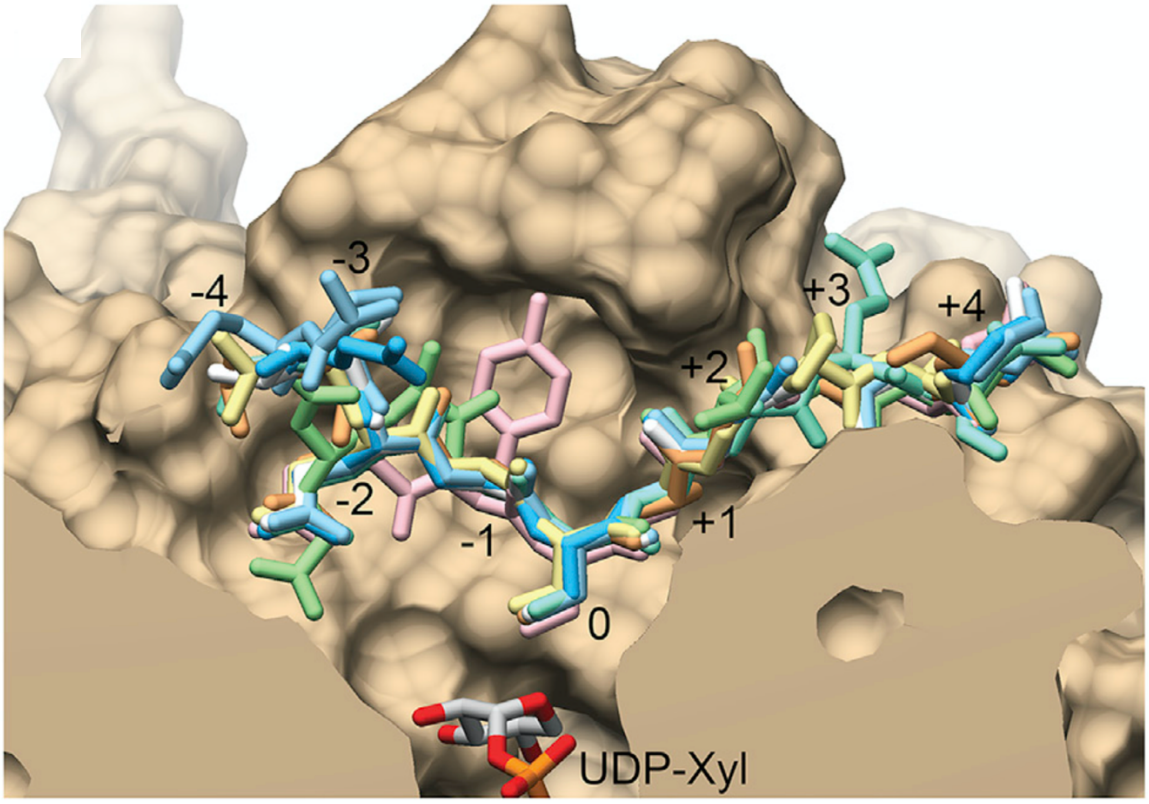


Figure 1.3 XT-I acceptor specificity. Eight peptides complexed with XT-I are superimposed.[41]

1.2.3 Donor Specificity of XT-I/II

Unlike the extensive study of acceptor promiscuity, investigations on the donor specificity of XT-I/II are rather limited and, until recently, both xylosyltransferases were considered monofunctional to UDP-xylose. In a study done by the Götting group, various non-native UDP-sugars, including UDP-glucose, UDP-galactose, UDP-glucuronic acid, and UDP-*N*-acetyl-

glucosamine were examined with a soluble XT-II to test its donor promiscuity.[31] However, there were no observable transfers of the non-native sugar to the selected peptide acceptors under testing. It suggests that the donor substrate scope of human XT-II is rather limited and may be restricted to UDP-xylose.

In 2018, Briggs and Hohenester provided an in-depth structural investigation of XT-I with high-resolution crystal structures.[41] In the crystal structure of the ternary complex of XT-I with both UDP-xylose and a peptide substrate, the presence of residue W392 in the UDP-xylose binding site restricts the available space around the C5 of xylose, which potentially restricts the donor scope of XT-I (**Figure 1.4**). This finding further supports the belief that XT-I/II could be monofunctional to UDP-xylose.

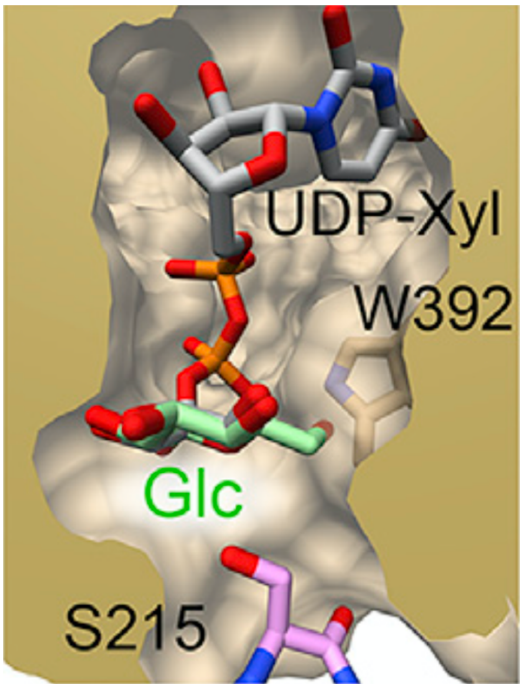


Figure 1.4 UDP-xylose binding pocket of XT-I. Residue W392 is in close proximity to the C5 of xylose.[41]

Nevertheless, a contradictory outcome was reported by Hendig group in 2015.[40] In their

work, they discovered that XT-I was able to recognize the UDP-4-azido-4-deoxyxylose (UDP-XylAz) and transferred the 4-azido-4-deoxy-xylose to the bikunin-like peptide QEEEGSGGGQKK. In comparison, the glycosylation activity from XT-II using UDP-XylAz was not observed. This is the first reported differentiation of XT-I/II activity and also, to the best of our knowledge, the only example showing that XT-I could accept non-native UDP-sugar as a donor substrate.

Since XT-I could tolerate the azido-modification on the C4 position, other small alterations on xylose may potentially be accepted by the enzyme. To better understand the donor profile of XT-I, more follow-up investigations are in great need.

1.2.4 Determination of XT-I/II Activity and Product Characterization

In the past decades, a variety of tools has been developed or applied to determine the XT-I/II activity. Dating back to 1960s, the Neufeld group and Dorfman group documented the first measurements of the XT-I activity with ^{14}C radioactive-labelled UDP-xylose sugar donor substrate.[15-17] In 2006, Brunner and co-workers applied matrix-assisted laser desorption ionization – time of flight mass spectrometry (MALDI-TOF MS) and reverse-phase high performance liquid chromatography (RP-HPLC) to analyze products of xylosyltransferase reactions.[30] To obtain detailed structural information, electrospray ionization (ESI) tandem mass spectrometry was applied for the first time to pinpoint the location of the xylose unit.[30]

To confirm the β-glycosylated linkage, Götting and co-workers examined the XT-I glycosylated products with linkage-specific cleavage by α- and β-xylosidase and base promoted release of the glycan from the glycopeptide.[25] The results clearly indicated a β-linkage between xylose and serine. This method was later extended to XT-II-catalyzed reactions by Casanova and

co-workers.[31] In their study, the linkage-specific digestion of the reaction products reveals that XT-II is also a β-xylosyltransferase.

Recently, Briggs and Hohenester utilized a commercialized glycosyltransferase kit to quantify the XT-I activity by monitoring the release of UDP from the sugar donor. The luminescence was then measured to correlate the readout with the enzymatic activity.[41]

Until now, the involvement of modern nuclear magnetic resonance (NMR) technique to characterize the product structures has yet been reported. Likely in the near future, with improvements on reaction scale and sample preparation, the conformation of the linkage would be decisively defined by NMR experiments.

1.2.5 Structure-Activity Relationships (SAR)

With advances on efficient expression and purification of XT-I, substantial progress on the structure-activity relationships of this important enzyme has been achieved during the past two decades. Especially, the high-quality crystal structures of XT-I and its ternary complex with UDP-xylose and peptide acceptors have drastically enhanced the current understanding of how XT-I interacts with the substrates and offer valuable insights on the catalytic mechanism.[41]

In 2004, Götting et al. first investigated the functions of XT-I DXD motifs with mutants that carried point mutations on the two short segments, ^{314}DED^{316} and ^{745}DWD^{747}.[27] Mutations on the first ^{314}DED^{316} motif do not affect the XT-I function. In contrast, the D745G mutation abolishes the catalytic function of XT-I, even though the alterations on ^{745}DWD^{747} do not strongly affect the donor substrate bindings.

A year later, with 14 mutants carrying individual point mutations of cysteine into alanine, Müller and co-workers investigated the importance of available cysteine residues to XT-I functions.[28] In terms of enzymatic activity, mutations on 5 of the 14 cysteine residues resulted in

over 90% loss of XT-I function. These findings imply the importance of the 5 Cys residues to the XT-I activity. Interestingly, alanine replacement of the cysteine residues close to the C-terminus did not exhibit any considerable effects on XT-I catalysis. The treatment of the cysteine-targeting *N*-phenylmaleimide reagent induced concentration-dependent inhibitions on all enzymatically active cysteine-to-alanine mutants but not the wild-type XT-I. These results indicate that all the 14 cysteine residues may exist in form of cystine and there are no free thiol groups available in wild-type XT-I. In addition, the enzymatic activity of wt XT-I and its single mutants could also be effectively reduced under the treatment of high-dose UDP or glycosaminoglycans. Meanwhile, all the mutants demonstrated comparable binding to the immobilized UDP and heparin as the wild-type XT-I. Taken together, it is likely that the cysteine residues present in XT-I do not directly participate in UDP or GAG bindings and mutations on them triggered no drastic conformational changes in the corresponding binding sites.

Shortly after, Müller and co-workers furthered their investigations with a series of *N*-terminal truncated forms of human XT-I.[48] According to their results, the first 260 amino acids at the *N*-terminus of the wild type are not required for the enzymatic activity. However, the XT-I catalytic function would be abolished with an additional deletion of 12 amino acids, $G^{261}KEAISALSRAK^{272}$, from the *N*-terminus. Since the individual replacement of each non-aliphatic residue in the 12 amino-acid sequence by alanine did not exert substantial influence on the enzyme activity in their study, it was suggested that this motif could be crucial to maintain the proper conformation of the enzyme. Interestingly, the truncation of $P^{721}KKVFKI^{727}$ motif, which is similar to the heparin-binding consensus sequence identified by Cardin and Weintraub,[49] does not affect the heparin binding of XT-I but dramatically impairs the proper enzymatic function, implying the necessity of this motif to the protein conformation.[48]

Over a decade later, in 2018, Briggs and Hohenester provided an in-depth structural investigation of XT-I with high-resolution crystal structures.[41] The structures in complex with UDP-xylose and peptide acceptors offer valuable insights on how the enzyme recognizes and interacts with the substrates. To obtain the ternary complex of XT-I with both UDP-xylose and a peptide substrate, the serine residue originally in the acceptor peptide sequence was replaced by alanine to abolish its acceptor function. The UDP diphosphate moiety of the donor binds with positively charged amino acid residues R598 and K599, instead of a divalent metal ion. The presence of residue W392 in the UDP-xylose binding site restricts the available space around the C5 of xylose, providing an explanation for the limited donor scope of XT-I (**Figure 1.5**).

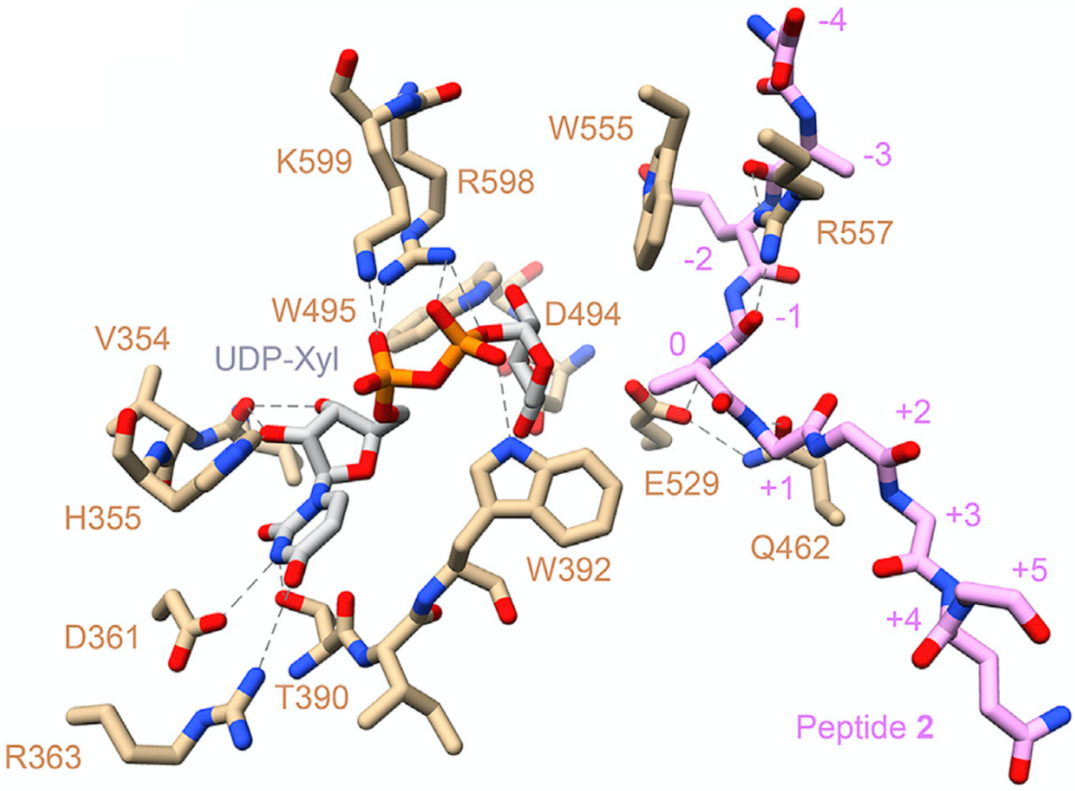


Figure 1.5 Active site of XT-I in complex with UDP-xylose donor and a peptide acceptor.[41]

The crystal structure around the peptide-binding site suggests that the network of hydrogen bonds is not sequence specific. Ten out of the eleven hydrogen bonds between the acceptor peptide and the catalytic domain occur on the carbonyl and amide groups along the peptide backbone. To gain insights into the characteristic C-terminal domain of XT-I (Xylo_C domain), a variety of single mutants was expressed. Results demonstrated that point mutations on the Xylo_C structure in contact with the catalytic GT_A domain did not impede the XT-I enzymatic functions. Briggs and Hohenester suggest that the presence of the Xylo_C domain, instead of being directly required for xylosylation activity, likely facilitates the recruitment of enzymes involved in subsequent GAG biosynthesis.

1.3 β-1,4-Galactosyltransferase 7 (β4GalT7)

1.3.1 Expression and Purification of β4GalT7

The β4GalT7 enzyme represents the seventh member of human β-1,4-galactosyltransferase family. Its molecular cloning and expression were first achieved by the Clausen group in 1999.[50] The full-length β-1,4-galactosyltransferase and a truncated version containing amino acid residues 63-327 were prepared using the Sf9 and High Five insect cell expression systems. The purification of β4GalT7 was then accomplished by sequential DEAE/Amberlite and S-Sepharose chromatography.[51]

The Lattard group, in 2009, successfully expressed the membrane form of β4GalT7 in HeLa cells and a soluble maltose-binding protein (MBP)-β4GalT7 fusion protein with an *N*-terminal truncation in *E*. Coli BL21 cells.[52] The MBP-fused β4GalT7 was purified by an amylose column. The desired protein was eluted out with 20 mM maltose in buffer A (20mM MOPS containing 150 mM NaCl at pH 7.0), and further dialyzed against the same buffer.

In a research work published by Ramakrishnan and Qasba in 2010, the catalytic domain of

Drosophila melanogaster β4GalT7, in its native form or with a variety of modifications, was individually prepared crystallization studies.[53] The variants included an enzyme with an 11-amino acid truncation from the *C*-terminus (Cd7ΔC) and ones carrying additional bovine β4GalT1 peptide fragments at the *N*-terminus (P-Cd7ΔC and P1-Cd7ΔC).

Since the MBP-β4GalT7 fusion protein produced in previous work only exhibited modest solubility and was prone to aggregation after the release of MBP fusion partner by protease, in 2010, the Qasba group designed a soluble form of human β4GalT7 using galectin-1 as the fusion partner to facilitate the folding and improve its stability and solubility.[54] This fusion form of β4GalT7 was expressed with an *E. coli* expression system. The initial purification was achieved with an alpha-lactose column and the target protein, galectin-1-human-β4GalT7, was eluted out with 100 mM lactose. Subsequently, the galectin-1 was cleaved off the protein with the Tobacco Etch Virus (TEV) protease. In this study, another MBP-fusion form of human β4GalT7 plasmid, pmal-2x-hum-β4GalT7, was constructed, and the enzyme, MBP-human-β4GalT7, was expressed effectively in *E. Coli*. The MBP-tag assisted the purification with an amylose column as previously reported.[55] Factor Xa protease cleaved off the MBP tag. The soluble form of human β4GalT7 was eventually purified with UDP-agarose columns.

In direct comparison to the two MBP-fusion forms, the galectin-1-human-β4GalT7 created exhibits great solubility and is less prone to aggregation, displaying its superior stability. It is the first documented success of galectin-1 as a fusion partner acting as a chaperone for the preparation of human β4GalT7 in *E. Coli* cells.

Meanwhile, in a study reported by Talhaoui and co-workers, HeLa cells or CHO pgsB-618 cells were transfected with either wild-type human β4GalT7 plasmid or single-mutant plasmids, individually, to aid the determination of catalytically active residues.[56] In addition, *E. coli*

BL21(DE3) cells were also used to prepare a soluble GST-fusion form of β4GalT7. Its purification was attained *via* the GST tag with glutathione-Sepharose 4B packed affinity column.

In 2013, the Qasba group unveiled the crystal structures of *Drosophila* β4GalT7 and a single mutant D211N β4GalT7 in complex with UDP-galactose as the donor and xylobiose as the acceptor, respectively.[57] In this study, the plasmid of an *N*-terminally truncated human β4GalT7 (β4GalT7Δ81) was constructed and the preparation of this truncated protein was carried out following previously reported conditions.[54]

The Fournel-Gigleux group, in 2015, constructed multiple vectors for different forms of human β4GalT7 and successfully expressed *N*-terminus truncated GST-tagged human β4GalT7 (β4GalT7ΔNt60) using *E.coli* BL21 (DE3) cells.[58] This is the most recent report of unique expression of human β4GalT7.

1.3.2 Acceptor Specificity of β4GalT7

The early report on β4GalT7 acceptor specificity dates back to 1994.[59] Esko and co-workers examined the priming of heparan sulfate using a variety of xylosides carrying non-native aglycones. This is the first demonstration that certain galactosyltransferase accepts xylosides as its substrates to enable heparan sulfate biosynthesis. In the following years, an increasing number of chemically modified xylosides were tested and the β4GalT7 acceptor scope expanded as investigations continued.[60-62]

In 2007, a library of thio-xylosides was prepared by the Ellervik group to examine the effect on GAG chain priming.[63] In the study, for the first time, they demonstrated that thio-xylosides could be tolerated by the enzymes for GAG biosynthesis. Shortly after, Abrahamsson et al. assessed GAG priming capability of various xylosylated naphthoic acid-amino acid conjugates.[64] Only the most nonpolar analog initiated the GAG biosynthesis in T24 cells. Two

years later, Victor and co-workers built a library of metabolically stable click-xylosides with hydrophobic groups attached. Priming activities were observed with this novel group of xylosides using CHO cell line.[65] The *in vitro* studies unveiled that aglycone moieties of xylosides affect sulfation, GAG chain composition and length. These results demonstrated that multiple *O*-, *S*-, and *C*-xylosides could be processed by β4GalT7 *in vitro*.

In a research work published by the Fernandez-Mayoralas and Garcia-Junceda groups in 2011, a collection of decoy xyloside acceptors was chemically synthesized and tested with a recombinant soluble β4GalT7. This was the first demonstration that recombinantly expressed β4GalT7 is promiscuous in the aglycon moieties of the xylose acceptor.[66]

Three years later, the Ellervik group further explored the substrate promiscuity of the enzyme with a truncated GST-β4GalT7 and chemically modified xyloside analogs.[67] In contrast to the great tolerance on aglycones, the truncated GST-β4GalT7 failed to process most of the xyloside analogs to any significant extent. Only a few xyloside analogs carrying modifications on C2 or C5 positions were galactosylated. Subsequent molecular modeling revealed that the binding pocket of β4GalT7 is narrow. Xylose, as the optimal substrate, is required to match with the precise set of hydrogen bond acceptors in the pocket.

In 2015, more in-depth investigations were carried out to gain understandings on acceptor structure requirements.[68] In this study, xylosides with varied aglycon size, anomeric configuration, linker length and electronic properties were carefully examined and compared. In general, only xylosides with the β-anomeric configuration would be smoothly converted by β4GalT7. The galactosylation capability of substrate can be enhanced by replacing the anomeric oxygen with sulfur. Substituting it with carbon reduces the enzymatic activity. In line with prior findings, bulky aglycons could be accepted.

Recently, a variety of xylosides and xyloside analogues carrying 2-naphthyl (Nap) or 4-methylumbelliferyl (MU) aglycone was synthesized by the Ellervik group and the Wagner group.[69-73] From the assay results, xyloside analog 2-naphthyl β-D-GlcNAc functioned as an acceptor substrate.[70] And analogs having an endocyclic sulfur atom proved to be great substrates for the enzyme.[72]

1.3.3 Donor Specificity of β4GalT7

In comparison with acceptor specificity, investigations on β4GalT7 donors are limited.[52, 56] The first detailed examination on β4GalT7 donor scope was reported in 2009 by the Lattard group. Several non-native UDP-sugars, including UDP-Xyl, UDP-Glc, UDP-Man, UDP-GlcA, UDP-GalNAc and UDP-GlcNAc, were individually incubated with purified MBP-β4GalT7. Among them, UDP-Xyl and UDP-Glc were accepted by the enzyme, although with much lower activities with 27-fold and 11-fold decreases as opposed to UDP-Gal, respectively.[52]

Fournel-Gigleux group reported similar results a year later.[56] Using 4-MU xyloside as acceptor, wild-type β4GalT7 was able to process UDP-Xyl and UDP-Glc, even though the observed activity levels were low. The W224H mutant failed to retain the donor promiscuity.

1.3.4 Determination of β4GalT7 Activity and Product Characterization

Back to 1990s, *in cellulo* GAG priming with β-D-xylosides was probed using radioactive [^{35}S]SO_4^{2-} and [6-^{3}H] D-glucosamine.[59] Later, UDP-[^{14}C]-Gal was used to track the activity of secreted β4GalT7 enzyme.[60, 74, 75] Almeida and co-workers performed one-dimensional ^{1}H NMR, two-dimensional ^{1}H-^{1}H TOCSY, and ^{13}C-decoupled ^{1}H-^{13}C HSQC and HMBC experiments to analyze the product structure in details. The NMR data confirmed the newly formed Galβ1→4Xylβ linkage.[60] In 2009, the Lattard group applied NMR techniques, including ^{1}H, ^{13}C, HSQC, TOCSY, COSY and NOESY, to thoroughly characterize the reaction products.[52] The

significant chemical shift changes on H-4 and C-4, together with a large $^{3}J_{H1',H2'}$ value, supported the desired β1→4 linkage.

In 2005, RP-HPLC equipped with a C18 column was for the first time applied to monitor the β4GalT7 reactions by Gulberti and co-workers.[76] This analytical method was then optimized and more routinely used to assess β4GalT7 enzymatic activity.[52, 66-68] A phosphatase-coupled glycosyltransferase assay, in which a phosphatase is used to convert the released UDP into inorganic phosphate for subsequent colorimetric quantification, was lately developed and applied to kinetic studies of β4GalT7.[70, 77]

1.3.5 Structure-Activity Relationships

Pioneering investigations into the β4GalT7 catalytic domain trace back to 2010.[53, 56] With the first high-resolution crystal structure of *Drosophila* β4GalT7 catalytic domain resolved, Boopathy and Pradman discovered a new Mn^{2+}-binding motif (^{241}H*X*H^{243}), in addition to the D*X*D motif common in β4GalT family.[53] Based on the molecular docking result, the O4 hydroxyl group in xylose is expected to form a strong hydrogen bond with the Asp^{211} side-chain carboxylate oxygen atom for acceptor activation. The presence of Tyr^{177} greatly limits the space in the binding pocket (**Figure 1.6**). The steric hindrance imposed by this bulky residue may explain why β4GalT7 rejects most of the chemically modified xyloside analogs as acceptors.

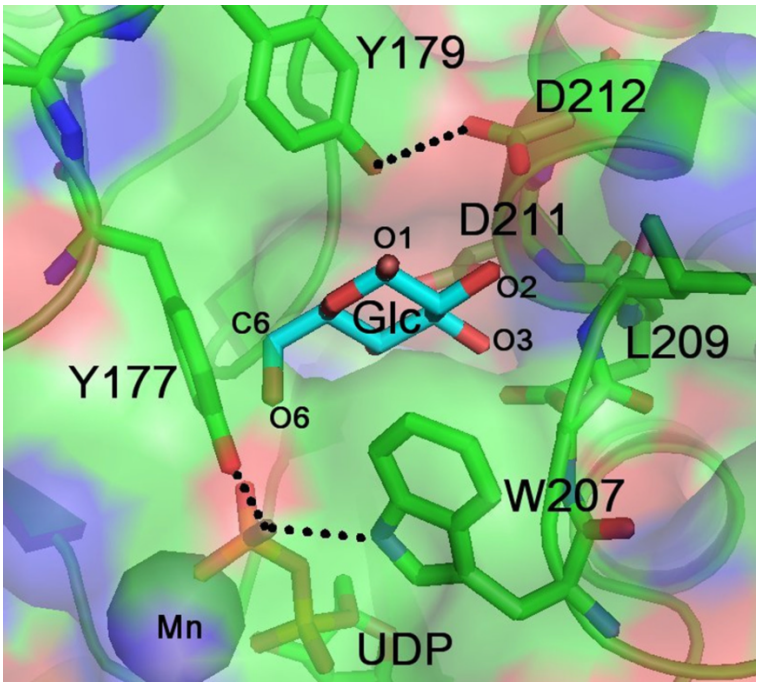


Figure 1.6 Molecular docking of glucose into the binding pocket of *Drosophila* β4GalT7. O2, O3 and O4 hydroxyl groups of docked glucose molecule are in close proximity to catalytic residue D211/D212. Residue Y177 imposes steric hindrance on the C6/O6 atom of the glucose molecule, implying only xylose would be accommodated by the enzyme.[53]

In the same year, the Fournel-Gigleux group reported the first detailed SAR investigation on the active site of human β4GalT7.[56] Canonical motifs ^{163}DVD^{165} and ^{221}FWGWGEDDE^{230} were identified in hβ4GalT7 (**Figure 1.7**). D163A or D165A point mutation completely abolished the enzyme activity. In comparison, replacement of D165 with glutamic acid retained, albeit reduced the hβ4GalT7 activity. For the *N*-terminus of conserved ^{221}FWGWGEDDE^{230} region, F221A mutation may affect the conformation of acceptor-binding site, as reflected by a 13-fold in the *K*m value of 4-MU-xylose. Meanwhile, W222F mutation did not show apparent effects on the affinity of either the donor or the acceptor. W224F and G225A mutants failed to demonstrate any observable enzyme activities, while G223A mutant maintained roughly 40% of the enzyme function. Further investigations suggested residue W224 plays a critical role in the donor and acceptor substrate binding. For the *C*-terminus of the peptide region, E227D/E230A did not impact

the donor or acceptor binding. In contrast, E227A/D228A/D229A/D229E mutants abolished the catalytic activity.

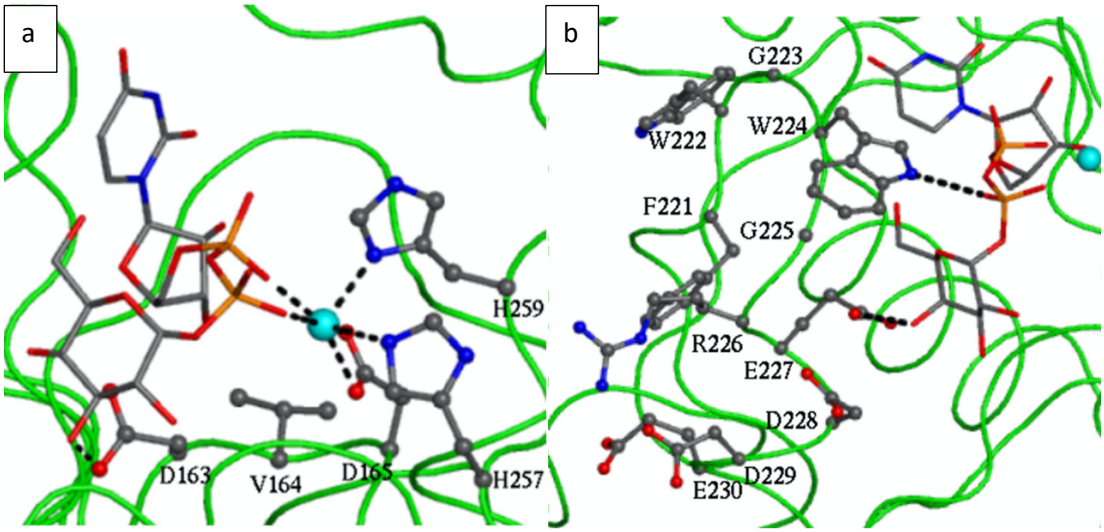


Figure 1.7 Molecular modeling of human β4GalT7 in complex with UDP-Gal. a) Predicted complex formed with UDP-Gal, Mn^{2+}, and $^{163}DVD^{165/257}HLH^{259}$; b) Predicted interaction between β-phosphate of UDP-Gal and residue W224. The protein α-carbon backbone is colored in green. Key residues in the active sites, UDP-Gal, and Mn^{2+} are highlighted.[56]

In 2013, the co-crystal structure of *Drosophila* D211N β4GalT7 mutant in the closed conformation with donor UDP-Gal and acceptor xylobiose was published by Tsutsui and co-workers.[57] In their study, an additional hydrogen bond is observed between Tyr177 side-chain -OH group and the β-phosphate oxygen atom of the UDP-Gal donor. The catalytic base Asp211 interacts with O3 and O4 atoms of the bound xylose acceptor *via* hydrogen bonds (**Figure 1.8**). Although the acceptor binding site is hydrophobic due to the presence of Tyr194, Tyr196, Tyr199 and Trp224, its neighboring region is highly positively charged to provide a high affinity to the acidic-residue-rich xylose attachment sites of native proteoglycans.

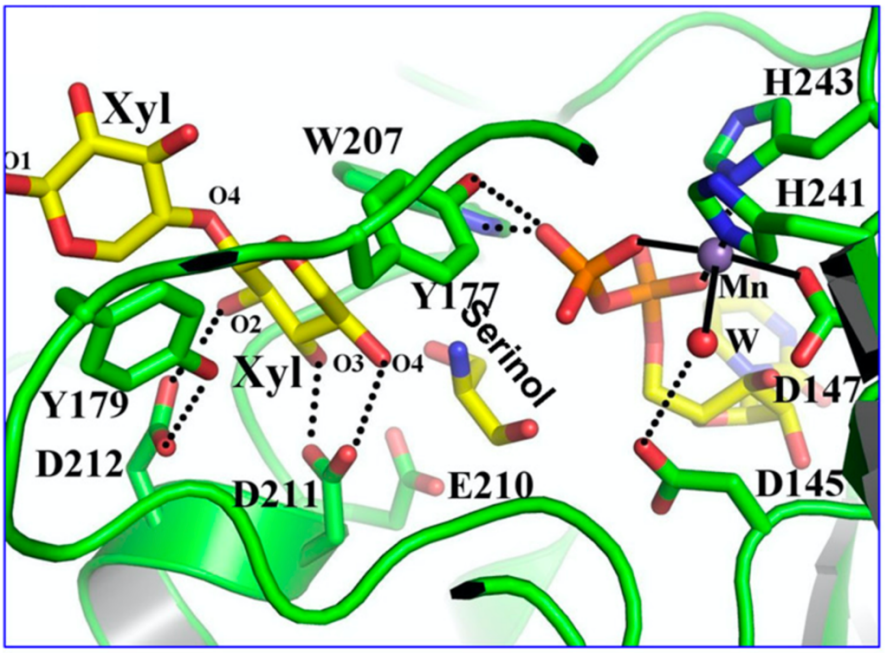


Figure 1.8 Xylobiose binding to *Drosophila* β4GalT7 in a closed conformation. The active site is colored in green.[57]

The Ellvervik group later studied the enzyme-substrate interactions with their synthesized xyloside analogs.[68] Despite the steric effect imposed by the chemical modifications of the aglycon, O2, O3, and O4 from the xylosides form a hydrogen bonding network with the catalytic residues N211 and D212 (**Figure 1.9**).

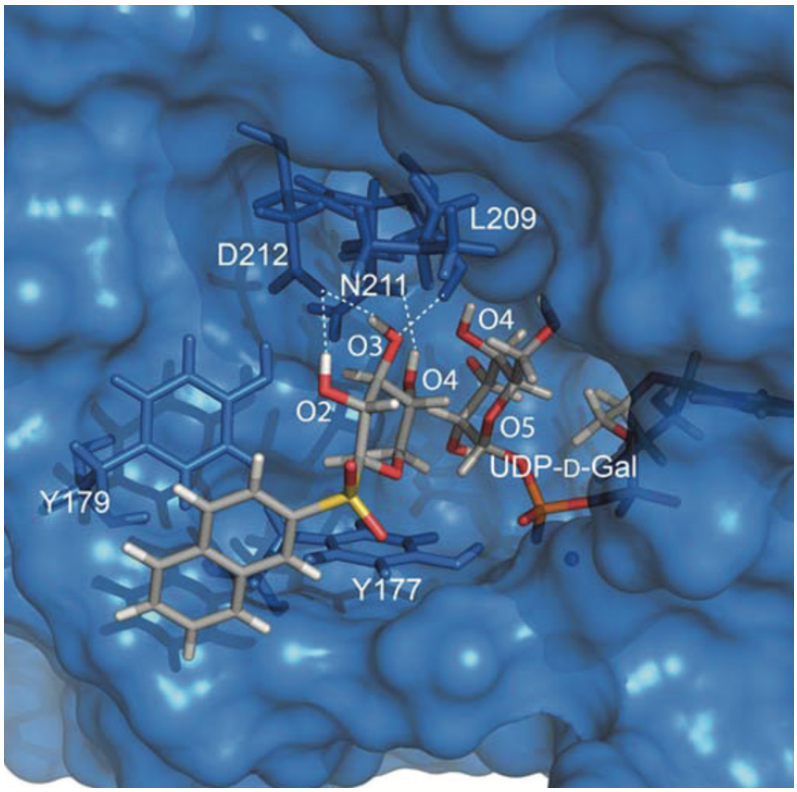


Figure 1.9 D211N β4GalT7 in complex with UDP-Gal, Mn^{2+} and a xyloside analog. The protein is colored in blue. UDP-Gal and the xyloside analog are highlighted in grey.[68]

Recently, Fournel-Gigleux group extended the computational analysis to human β4GalT7.[58] Their docking simulation results identified a hydrophobic region, formed by Tyr194, Tyr196 and Tyr199, that provides stacking interactions with the aglycone and the xylopyranoside sugar ring. The acceptor xyloside is oriented and activated through a hydrogen bond network with Asp228, Asp229 and Arg226 (**Figure 1.10**).

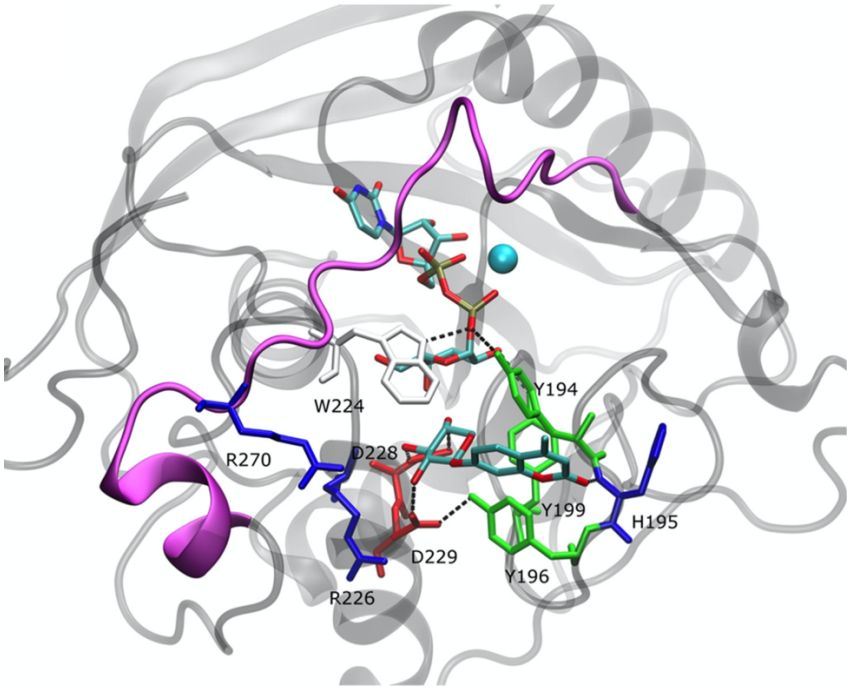


Figure 1.10 The active site of human β4GalT7 in complex with UDP-Gal, Mn^{2+} and 4-MUX. The protein α-carbon backbone is colored in grey. Key residues in the active site and substrates are highlighted.[58]

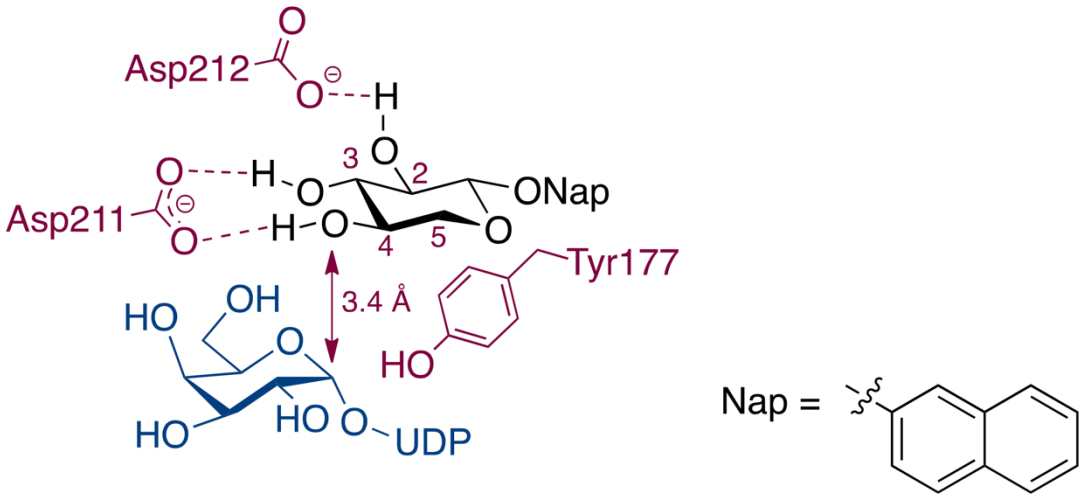


Figure 1.11 Overview of proposed binding pattern of xylosides and UDP-Gal in the β4GalT7 binding pocket.[71]

1.4. Future Outlook

While significant progress has been made on the key glycosyltransferases involved in proteoglycan linkage region synthesis, application of these biocatalysts is in its infancy. From the perspective of synthesis, deploying the four enzymes may lead to a highly efficient chemoenzymatic preparation of the PG linkage bearing glycopeptides. Together with well-developed GAG synthesis enzymes,[78, 79] it may pave the road towards native homogeneous PG glycopeptides and glycoproteins. A library as such would be highly valuable for in-depth structure-activity relationship investigations. As traditional chemical synthesis can be highly tedious and labor intensive, PG enzymatic synthesis would serve as a disruptive approach to dramatically reduce the time, effort, and materials required to prepare PG compounds, making the process faster, easier, and 'greener'.

In addition, enabled by advanced computational technology, biocatalytic enzymes could be re-designed or re-purposed to tailor specific research needs. Among the four enzymes needed to make the PG linkage, XT-I is a particularly promising target. With its ability to recognize certain binding motifs, a properly engineered XT-I variant could potentially transfer non-native sugars, for instance, an azido-sugar, to a wide range of biological proteins. The labelled proteins may then be functionalized with a variety of fluorescent probes or affinity tags to support diverse research aspirations. If the other enzymes involved in PG linkage assembly could tolerate the chemically modified glycoproteins as their substrates, they would become a highly valuable biocatalytic toolkit to facilitate investigation of the multifaceted biological functions of PGs.

REFERENCES

REFERENCES

1. Bernfield, M.; Götte, M.; Park, P. W.; Reizes, O.; Fitzgerald, M. L.; Lincecum, J.; Zako, M., Functions of Cell Surface Heparan Sulfate Proteoglycans. *Annu. Rev. Biochem.* **1999,** *68*, 729-777.

2. Lin, X., Functions of Heparan Sulfate Proteoglycans in Cell Signaling during Development. *Development* **2004,** *131*, 6009-6021.

3. Nikitovic, D.; Berdiaki, A.; Spyridaki, I.; Krasanakis, T.; Tsatsakis, A.; Tzanakakis, G. N., Proteoglycans-Biomarkers and Targets in Cancer Therapy. *Front Endocrinol.* **2018,** *9*, 69.

4. Tzanakakis, G.; Neagu, M.; Tsatsakis, A.; Nikitovic, D., Proteoglycans and Immunobiology of Cancer-Therapeutic Implications. *Front Immunol.* **2019,** *10*, 875.

5. Ritelli, M.; Cinquina, V.; Giacopuzzi, E.; Venturini, M.; Chiarelli, N.; Colombi, M., Further Defining the Phenotypic Spectrum of B3GAT3 Mutations and Literature Review on Linkeropathy Syndromes. *Genes* **2019,** *10*, 631-650.

6. Okajima, T.; Yoshida, K.; Kondo, T.; Rurukawa, K., Human Homolog of Caenorhabditis elegans sqv-3Gene Is Galactosyltransferase Involved in the Biosynthesis of the Glycosaminoglycan-Protein Linkage Region of Proteoglycans. *J. Biol. Chem.* **1999,** *274*, 22915-22918.

7. Bai, X.; Zhou, D.; Brown, J. R.; Crawford, B. E.; Hennet, T.; Esko, J. D., Biosynthesis of the Linkage Region of Glycosaminoglycans: Cloning and Activity of Galactosyltransferase II, the Sixth Member of the Beta 1,3-Galactosyltransferase Family (beta 3GalT6). *J. Biol. Chem.* **2001,** *276*, 48189-48195.

8. Kitagawa, H.; Tone, Y.; Tamura, J.; 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*, 6615-6618.

9. Tone, Y.; Kitagawa, H.; Imiya, K.; Oka, S.; Kawasaki, T.; Sugahara, K., Characterization of Recombinant Human Glucuronyltransferase I involvedin the Bosynthesis of the Glycosaminoglycan-Protein Linkage Region of Proteoglycans. *FEBS Lett.* **1999,** *459*, 415-420.

10. Kitagawa, H.; Nadanaka, S., Beta-1,3-Glucuronyltransferase 3 (Glucuronosyltransferase I) (B3GAT3). In *Handbook of Glycosyltransferases and Related Genes*, 2014; pp 849-861.

11. B., V.-C. M.; Hansen, L.; Clausen, H., UDP-Gal BetaGal Beta1,3-Galactosyltransferase Polypeptide 6 (B3GALT6). *Handb. Glycosyltransferases Relat. Genes* **2014**, 101-108.

12. Hull, E. E.; Montgomery, M. R.; Leyva, K. J., Epigenetic Regulation of the Biosynthesis & Enzymatic Modification of Heparan Sulfate Proteoglycans: Implications for Tumorigenesis and Cancer Biomarkers. *Int. J. Mol. Sci.* **2017,** *18*, 1361-1385.

13. Wilson, I. B. H., The Never-Ending Story of Peptide *O*-Xylosyltransferase. *Cell. Mol. Life Sci.* **2004,** *61*, 794-809.

14. Gotting, C.; Kuhn, J.; Kleesiek, K., Human Xylosyltransferases in Health and Disease. *Cell. Mol. Life Sci.* **2007,** *64*, 1498-1517.

15. Grebner, E. E.; Hall, C. W.; Neufeld, E. F., Glycosylation of Serine Residues by a Uridine Diphosphate-Xylose Protein Xylosyltransferase from Mouse Mastocytoma. *Arch. Biochem. Biophys.* **1966,** *116*, 391-398.

16. Grebner, E. E.; Hall, C. W.; Neufeld, E. F., Incorporation of D-xylose-C14 into Glycoprotein by Particles from Hen Oviduct. *Biochem. Biophys. Res. Commun.* **1966,** *22*, 672-677.

17. Robinson, H. C.; Telser, A.; Dorfman, A., Studies on Biosynthesis of the Linkage Region of Chondroitin Sulfate-Protein Complex. *Proc. Natl. Acad. Sci. U. S. A.* **1966,** *56*, 1859-1866.

18. Gregory, J. D.; Laurent, T. C.; Roden, L., Enzymatic Degradation of Chondromucoprotein. *J. Biol. Chem.* **1964,** *239*, 3312-3320.

19. Lindahl, U.; Roden, L., The Role of Galactose and Xylose in the Linkage of Heparin to Protein. *J. Biol. Chem.* **1965,** *240*, 2821-2826.

20. Brinkmann, T.; Weilke, C.; Kleesiek, K., Recognition of Acceptor Proteins by UDP-D-xylose Proteoglycan Core Proteinb-D-Xylosyltransferase. *J. Biol. Chem.* **1997,** *272*, 11171-11175.

21. Weilke, C.; Brinkmann, T.; Kleesiek, K., Determination of Xylosyltransferase Activity in Serum with Recombinant Human Bikuninas Acceptor. *Clin. Chem.* **1997,** *43*, 45-51.

22. Gotting, C.; Sollberg, S.; Kuhn, J.; Weilke, C.; Huerkamp, C.; Brinkmann, T.; Krieg, T.; Kleesiek, K., Serum Xylosyltransferase: A New Biochemical Marker of the Sclerotic Process in Systemic Sclerosis. *J. Invest. Dermatol.* **1999,** *112*, 919-924.

23. Kuhn, J.; Gotting, C.; Schnolzer, M.; Kempf, T.; Brinkmann, T.; Kleesiek, K., First Isolation of Human UDP-D-Xylose: Proteoglycan Core Protein Beta-D-Xylosyltransferase Secreted from Cultured JAR Choriocarcinoma Cells. *J. Biol. Chem.* **2001,** *276*, 4940-4947.

24. Gotting, C.; Kuhn, J.; Tinneberg, H.; Brinkmann, T.; Kleesiek, K., High Xylosyltransferase Activities in Human Follicular Fluidand Cultured Granulosa–Lutein Cells. *Mol. Hum. Reprod.* **2002,** *8*, 1079-1086.

25. Gotting, C.; Kuhn, J.; Zahn, R.; Brinkmann, T.; Kleesiek, K., Molecular Cloning and Expression of Human UDP-D-Xylose:Proteoglycan Core Protein Beta-D-Xylosyltransferase and its First Isoform XT-II. *J. Mol. Biol.* **2000,** *304*, 517-528.

26. Kuhn, J.; Muller, S.; Schnolzer, M.; Kempf, T.; Schon, S.; Brinkmann, T.; Schottler, M.; Gotting, C.; Kleesiek, K., High-Level Expression and Purification of Human Xylosyltransferase I in High Five Insect Cells as Biochemically Active Form. *Biochem. Biophys. Res. Commun.* **2003,** *312*, 537-544.

27. Gotting, C.; Muller, S.; Schottler, M.; Schon, S.; Prante, C.; Brinkmann, T.; Kuhn, J.; Kleesiek, K., Analysis of the DXD Motifs in Human Xylosyltransferase I Required for Enzyme Activity. *J. Biol. Chem.* **2004,** *279*, 42566-42573.

28. Muller, S.; Schottler, M.; Schon, S.; Prante, C.; Brinkmann, T.; Kuhn, J.; Gotting, C.; Kleesiek, K., Human Xylosyltransferase I: Functional and Biochemical Characterization of Cysteine Residues Required for Enzymic Activity. *Biochem. J.* **2005,** *386*, 227-236.

29. Schon, S.; Prante, C.; Bahr, C.; Kuhn, J.; Kleesiek, K.; Gotting, C., Cloning and Recombinant Expression of Active Full-Length Xylosyltransferase I (XT-I) and Characterization of Subcellular Localization of XT-I and XT-II. *J. Biol. Chem.* **2006,** *281*, 14224-14231.

30. Brunner, A.; Kolarich, D.; Voglmeir, J.; Paschinger, K.; Wilson, I. B., Comparative Characterisation of Recombinant Invertebrate and Vertebrate Peptide *O*-Xylosyltransferases. *Glycoconj. J.* **2006,** *23*, 543-554.

31. Casanova, J. C.; Kuhn, J.; Kleesiek, K.; Gotting, C., Heterologous Expression and Biochemical Characterization of Soluble Human Xylosyltransferase II. *Biochem. Biophys. Res. Commun.* **2008,** *365*, 678-684.

32. Stoolmiller, A. C.; Horwitz, A. L.; Dorfman, A., Biosynthesis of the Chondroitin Sulfate Proteoglycan. *J. Biol. Chem.* **1972,** *247*, 3525-3532.

33. Coudron, C.; Ellis, K.; Phillipson, L.; Schwartz, N. B., Preliminary Characterization of a Xylose Acceptor Prepared by Hydrogen Fluoride Treatment of Proteoglycan Core Protein. *Biochem. Biophys. Res. Commun.* **1980,** *92*, 618-623.

34. Campbell, P.; Jacobsson, I.; Benzing-Purdie, L.; Roden, L.; Fessler, J. H., Silk — A New Substrate for UDP-D-Xylose:Proteoglycan Core Protein B-D-Xylosyltransferase. *Anal. Biochem.* **1984,** *137*, 505-516.

35. Pfeil, U.; Wenzel, K., Purification and Some Properties of UDP-Xylosyltransferase of Rat Ear Cartilage. *Glycobiology* **2000,** *10*, 803-807.

36. Wilson, I. B. H., Functional Characterization of Drosophila Melanogaster Peptide *O*-Xylosyltransferase. The Key Enzyme for Proteoglycan Chain Initiation and Member of the Core *N*-Acetylglucosaminyltransferase Family. *J. Biol. Chem.* **2002,** *277*, 21207-21212.

37. Gotting, C.; Kuhn, J.; Brinkmann, T.; Kleesiek, K., Xylosylation of Alternatively Spliced Isoforms of Alzheimer APP by Xylosyltransferase. *J. Protein Chem.* **1998,** *17*, 295-302.

38. Bourdon, M. A.; Krusius, T.; Campbell, S.; Schwartz, N. B.; Ruoslahti, E., Identification and Synthesis of Cognition Signal for the Attachment of Glycosaminoglycans to Proteins. *Proc Natl Acad Sci U S A* **1987,** *84*, 3194-3198.

39. Kearns, A. E.; Campbell, S. C.; Westley, J.; Schwartz, N. B., Initiation of Chondroitin Sulfate Biosynthesis: A Kinetic Analysis of UDP-D-Xylose:Core Protein β-D-Xylosyltransferase. *Biochemistry* **1991,** *30*, 7477-7483.

40. Kuhn, J.; Gotting, C.; Beahm, B. J.; Bertozzi, C. R.; Faust, I.; Kuzaj, P.; Knabbe, C.; Hendig, D., Xylosyltransferase II is the Predominant Isoenzyme which is Responsible for the Steady-State Level of Xylosyltransferase Activity in Human Serum. *Biochem. Biophys. Res. Commun.* **2015,** *459*, 469-474.

41. Briggs, D. C.; Hohenester, E., Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1. *Structure* **2018,** *26*, 801-809.

42. Roch, C.; Kuhn, J.; Kleesiek, K.; Gotting, C., Differences in Gene Expression of Human Xylosyltransferases and Determination of Acceptor Specificities for Various Proteoglycans. *Biochem. Biophys. Res. Commun.* **2010,** *391*, 685-691.

43. Esko, J. D.; Zhang, L., Influence of Core Protein Sequence on Glycosaminoglycan Assembly. *Curr. Opin. Struct. Biol.* **1996,** *6*, 663-670.

44. Dong, S.; Cole, G. J.; Halfter, W., Expression of Collagen XVIII and Localization of Its Glycosaminoglycan Attachment Sites. *J. Biol. Chem.* **2003,** *278*, 1700-1707.

45. Winzen, U.; Cole, G. J.; Halfter, W., Agrin is a Chimeric Proteoglycan with the Attachment Sites for Heparan Sulfate/Chondroitin Sulfate Located in Two Multiple Serine-Glycine Clusters. *J. Biol. Chem.* **2003,** *278*, 30106-30114.

46. Bourdon, M. A.; Oldberg, A.; Pierschbacher, M.; Ruoslahti, E., Molecular Cloning and Sequence Analysis of a Chondroitin Sulfate Proteoglycan cDNA. *Proc. Natl. Acad. Sci. U. S. A.* **1985,** *82*, 1321-1325.

47. Mann, D. M.; Yamaguchi, Y.; Bourdon, M. A.; Ruoslahti, E., Analysis of Glycosaminoglycan Substitution in Decorin by Site-directed Mutagenesis. *J. Biol. Chem.* **1990,** *265*, 5317-5323.

48. Muller, S.; Disse, J.; Schottler, M.; Schon, S.; Prante, C.; Brinkmann, T.; J., K.; Kleesiek, K.; Gotting, C., Human Xylosyltransferase I and *N*-Terminal Truncated Forms: Functional Characterization of the Core Enzyme. *Biochem. J.* **2006,** *394*, 163-171.

49. Cardin, A. D.; Weintraub, H. J. R., Molecular Modeling of Protein-Glycosaminoglycan Interactions. *Arteriosclerosis* **1989,** *9*, 21-32.

50. Almeida, R.; Levery, S. B.; Mandel, U.; Kresse, H.; Schwientek, T.; Bennett, E. P.; Clausen, H., Cloning and Expression of a Proteoglycan UDP-Galactose b-Xylose b1,4-Galactosyltransferase I. *J. Biol. Chem.* **1999,** *274*, 26165-26171.

51. Wandall, H. H.; Hassan, H.; Mirogorodskaya, E.; Kristensen, A. K.; Roepstorff, P.; Bennett, E. P.; Nielsen, P. A.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H., Substrate Specificities of Three Members of the Human UDP-*N*-Acetyl-D-galactosamine: Polypeptide *N*-Acetylgalactosaminyltransferase Family, GalNAc-T1, -T2, and -T3. *J. Biol. Chem.* **1997,** *272*, 23503-23514.

52. Daligault, F.; Rahuel-Clermont, S.; Gulberti, S.; Cung, M. T.; Branlant, G.; Netter, P.; Magdalou, J.; Lattard, V., Thermodynamic Insights into the Structural Basis Governing the Donor Substrate Recognition by Human Beta1,4-Galactosyltransferase 7. *Biochem. J.* **2009,** *418*, 605-614.

53. Ramakrishnan, B.; Qasba, P. K., Crystal structure of the catalytic domain of Drosophila beta1,4-Galactosyltransferase-7. *J. Biol. Chem.* **2010,** *285*, 15619-15626.

54. Pasek, M.; Boeggeman, E.; Ramakrishnan, B.; Qasba, P. K., Galectin-1 as a Fusion Partner for the Production of Soluble and Folded Human Beta-1,4-Galactosyltransferase-T7 in *E. Coli*. *Biochem. Biophys. Res. Commun.* **2010,** *394*, 679-684.

55. Nallamsetty, S.; Waugh, D. S., A Generic Protocol for the Expression and Purification of Recombinant Proteins in Escherichia Coli using a Combinatorial His6-Maltose Binding Protein Fusion Tag. *Nat. Protoc.* **2007,** *2*, 383-391.

56. Talhaoui, I.; Bui, C.; Oriol, R.; Mulliert, G.; Gulberti, S.; Netter, P.; Coughtrie, M. W.; Ouzzine, M.; Fournel-Gigleux, S., Identification of Key Functional Residues in the Active Site of Human Beta1,4-Galactosyltransferase 7: A Major Enzyme in the Glycosaminoglycan Synthesis Pathway. *J. Biol. Chem.* **2010,** *285*, 37342-37358.

57. Tsutsui, Y.; Ramakrishnan, B.; Qasba, P. K., Crystal Structures of Beta-1,4-Galactosyltransferase 7 Enzyme Reveal Conformational Changes and Substrate Binding. *J. Biol. Chem.* **2013,** *288*, 31963-31970.

58. Saliba, M.; Ramalanjaona, N.; Gulberti, S.; Bertin-Jung, I.; Thomas, A.; Dahbi, S.; Lopin-Bon, C.; Jacquinet, J. C.; Breton, C.; Ouzzine, M.; Fournel-Gigleux, S., Probing the Acceptor Active Site Organization of the Human Recombinant Beta1,4-Galactosyltransferase 7 and Design of Xyloside-Based Inhibitors. *J. Biol. Chem.* **2015,** *290*, 7658-7670.

59. Fritz, T. A.; Lugemwa, F. N.; Sarkar, A. K.; Esko, J. D., Biosynthesis of Heparan Sulfate on P-D-Xylosides Depends on Aglycone Structure. *J. Biol. Chem.* **1994,** *269*, 300-307.

60. Almeida, R.; Levery, S. B.; Mandel, U.; Kresse, H.; Schwientek, T.; Bennett, E. P.; Clausen, H., Cloning and Expression of a Proteoglycan UDP-Galactose:β-Xylose β1,4-Galactosyltransferase I. *J. Biol. Chem.* **1999,** *274*, 26165-26171.

61. Takemae, H.; Ueda, R.; Okubo, R.; Nakato, H.; Izumi, S.; Saigo, K.; Nishihara, S., Proteoglycan UDP-Galactose:Beta-Xylose Beta 1,4-Galactosyltransferase I is Essential for Viability in Drosophila Melanogaster. *J. Biol. Chem.* **2003,** *278*, 15571-15578.

62. Jacobsson, M.; Ellervik, U.; Belting, M.; Mani, K., Selective Antiproliferative Activity of Hydroxynaphthyl-β-D-xylosides. *J. Med. Chem.* **2006,** *49*, 1932-1938.

63. Jacobsson, M.; Mani, K.; Ellervik, U., Effects of Oxygen-Sulfur Substitution on Glycosaminoglycan-Priming Naphthoxylosides. *Bioorg. Med. Chem.* **2007,** *15*, 5283-5299.

64. Abrahamsson, C. O.; Ellervik, U.; Eriksson-Bajtner, J.; Jacobsson, M.; Mani, K., Xylosylated Naphthoic Acid-Amino Acid Conjugates for Investigation of Glycosaminoglycan Priming. *Carbohydr. Res.* **2008,** *343*, 1473-1477.

65. Victor, X. V.; Nguyen, T. K.; Ethirajan, M.; Tran, V. M.; Nguyen, K. V.; Kuberan, B., Investigating the Elusive Mechanism of Glycosaminoglycan Biosynthesis. *J. Biol. Chem.* **2009,** *284*, 25842-25853.

66. Garcia-Garcia, J. F.; Corrales, G.; Casas, J.; Fernandez-Mayoralas, A.; Garcia-Junceda, E., Synthesis and Evaluation of Xylopyranoside Derivatives as "Decoy Acceptors" of Human Beta-1,4-Galactosyltransferase 7. *Mol. Biosyst.* **2011,** *7*, 1312-1321.

67. 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 B4galt7 Active Site. *Chem. Sci.* **2014,** *5*, 3501-3508.

68. Siegbahn, A.; Thorsheim, K.; Stahle, J.; Manner, S.; Hamark, C.; Persson, A.; Tykesson, E.; Mani, K.; Westergren-Thorsson, G.; Widmalm, G.; Ellervik, U., Exploration of the Active Site of B4galt7: Modifications of the Aglycon of Aromatic Xylosides. *Org. Biomol. Chem.* **2015,** *13*, 3351-3362.

69. Thorsheim, K.; Persson, A.; Siegbahn, A.; Tykesson, E.; Westergren-Thorsson, G.; Mani, K.; Ellervik, U., Disubstituted Naphthyl Beta-D-Xylopyranosides: Synthesis, GAG Priming, and Histone Acetyltransferase (HAT) Inhibition. *Glycoconj. J.* **2016,** *33*, 245-257.

70. Jiang, J.; Wagner, G. K., An Acceptor Analogue of Beta-1,4-Galactosyltransferase: Substrate, Inhibitor, or Both? *Carbohydr. Res.* **2017,** *450*, 54-59.

71. Thorsheim, K.; Clementson, S.; Tykesson, E.; Bengtsson, D.; Strand, D.; Ellervik, U., Hydroxylated Oxanes as Xyloside Analogs for Determination of the Minimal Binding Requirements of B4galt7. *Tetrahedron Lett.* **2017,** *58*, 3466-3469.

72. Thorsheim, K.; Willen, D.; Tykesson, E.; Stahle, J.; Praly, J. P.; Vidal, S.; Johnson, M. T.; Widmalm, G.; Manner, S.; Ellervik, U., Naphthyl Thio- and Carba-xylopyranosides for Exploration of the Active Site of beta-1,4-Galactosyltransferase 7 (beta4GalT7). *Chemistry* **2017,** *23*, 18057-18065.

73. Willen, D.; Bengtsson, D.; Clementson, S.; Tykesson, E.; Manner, S.; Ellervik, U., Synthesis of Double-Modified Xyloside Analogues for Probing the beta4GalT7 Active Site. *J. Org. Chem.* **2018,** *83*, 1259-1277.

74. Nakamura, Y.; Haines, N.; Chen, J.; Okajima, T.; Furukawa, K.; Urano, T.; Stanley, P.; Irvine, K. D.; Furukawa, K., Identification of a Drosophila Gene Encoding Xylosylprotein Beta4-Galactosyltransferase that is Essential for the Synthesis of Glycosaminoglycans and for Morphogenesis. *J. Biol. Chem.* **2002,** *277*, 46280-46288.

75. Vadaie, N.; Hulinsky, R. S.; Jarvis, D. L., Identification and Characterization of a Drosophila Melanogasterortholog of Human B1,4-Galactosyltransferase VII. *Glycobiology* **2002,** *12*, 589-597.

76. Gulberti, S.; Lattard, V.; Fondeur, M.; Jacquinet, J. C.; Mulliert, G.; Netter, P.; Magdalou, J.; Ouzzine, M.; Fournel-Gigleux, S., Phosphorylation and Sulfation of Oligosaccharide Substrates Critically Influence the Activity of Human Beta1,4-Galactosyltransferase 7 (Galt-I) and Beta1,3-Glucuronosyltransferase I (Glcat-I) Involved in the Biosynthesis of the Glycosaminoglycan-Protein Linkage Region of Proteoglycans. *J. Biol. Chem.* **2005,** *280*, 1417-1425.

77. Wu, Z. L.; Ethen, C. M.; Prather, B.; Machacek, M.; Jiang, W., Universal Phosphatase-Coupled Glycosyltransferase Assay. *Glycobiology* **2011,** *21*, 727-733.

78. Chappell, E. P.; Liu, J., Use of Biosynthetic Enzymes in Heparin and Heparan Sulfate Synthesis. *Bioorg. Med. Chem.* **2013,** *21*, 4786-4792.

79. Fu, L.; Suflita, M.; Linhardt, R. J., Bioengineered Heparins and Heparan Sulfates. *Adv. Drug Deliv. Rev.* **2016,** *97*, 237-249.

Chapter 2 Convergent Chemoenzymatic Synthesis and Biological Evaluation of a Heparan Sulfate Proteoglycan Syndecan-1 Mimetic

2.1 Introduction

Heparan sulfate proteoglycan (HSPG) consists of one or more heparan sulfate (HS) chains linked to serine residues in the core protein.[1] Ubiquitous on mammalian cell surface and in the extracellular matrix, HSPGs are involved in a wide variety of important biological processes, including regulations of growth factors, cell adhesions and cell-cell communications.[2-5] While heparan sulfate (HS) is generally considered to be the main determinant of HSPG activities, the core protein of HSPG can have significant impacts as well.[6, 7] However, due to the extreme heterogeneity of HS structures in nature, it is highly challenging to purify homogenous HSPGs from natural sources, presenting significant hurdles to decode the roles of HS and the core protein in HSPG functions. Chemical synthesis of HS glycopeptide has been reported, which is highly challenging due to instabilities of the HS glycan under typical peptide synthesis conditions.[8-10] In this chapter, I have developed a new convergent strategy integrating chemical synthesis with enzymatic reactions to synthesize a well-defined glyco-polypeptide mimicking the complex structure of HSPG such as syndecan-1.

Syndecan-1, a prototypical HSPG on the mammalian cell surface, can bind with integrins mediating cell adhesion, signaling, and migration. Synstatin (SSTN), a 36 amino acid long polypeptide corresponding to residues 92-117 of human syndecan-1, has been identified as the binding sites of αvβ3 and αvβ5 integrins.[11] While HS is known to interact with integrins, it is not clear how displaying HS in the context of a glycoprotein impacts its function. To more closely mimic the structural complexity of syndecan-1, we designed glyco-polypeptide analog **1**, which contains a 48 amino acid residue polypeptide backbone containing the full length synstatin

sequence, as well as a HS glycan chain bearing the full structural features of HS encountered in nature, including iduronic acid, glucuronic acid, *O*-sulfation and *N*-sulfation.

2.2. Results and Discussions

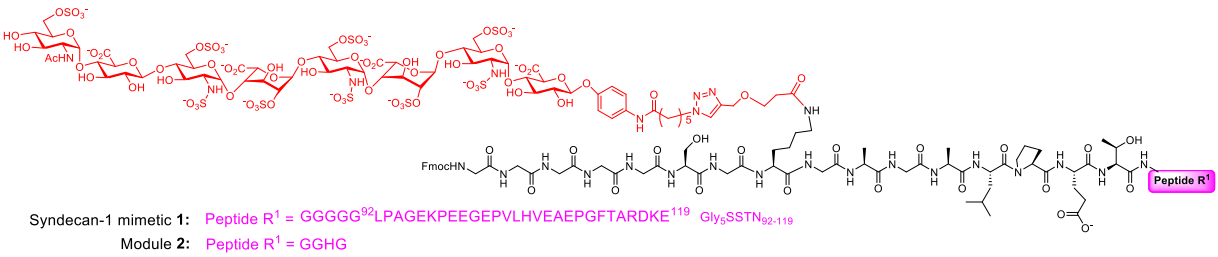
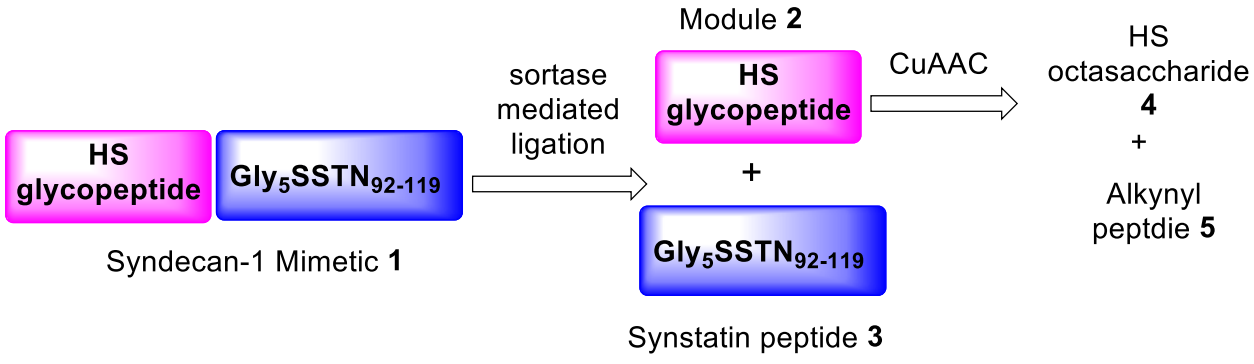


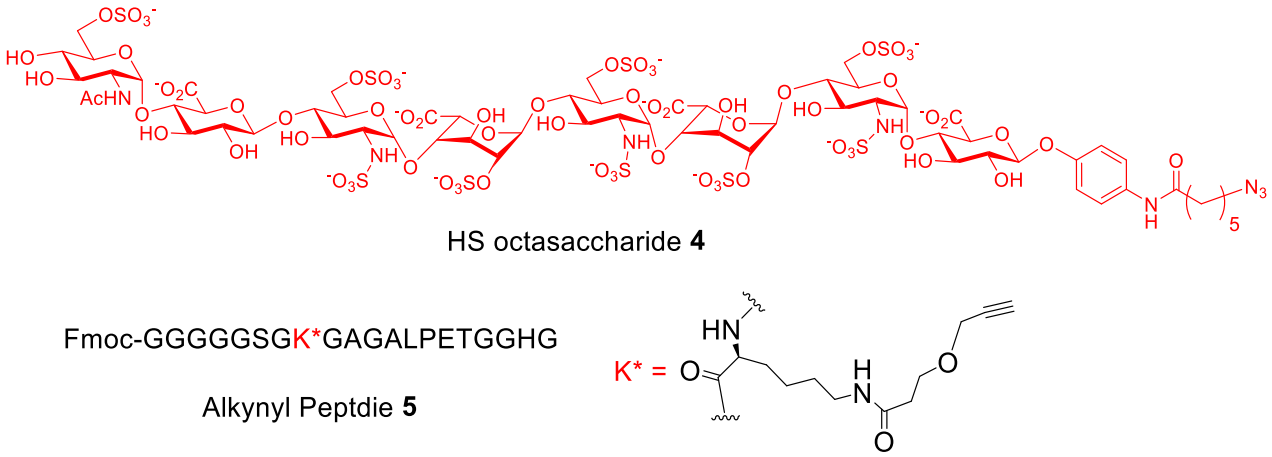
Figure 2.1 Structure of the HSPG syndecan-1 mimetic **1**.

To prepare the complex structure of HSPG mimetic **1**, retrosynthetically, the target molecule is divided into glycopeptide module **2** and synstatin$_{92\text{-}117}$ peptide **3** bearing a pentaglycine at its *N*-terminus (**Scheme 2.1**), which would be joined through an irreversible sortase A-mediated ligation. The glycopeptide **2** containing the 'LPETG' sorting sequence at the C-terminus would be assembled through the copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) of azido-oligosaccharide **4** and alkynyl peptide **5**.

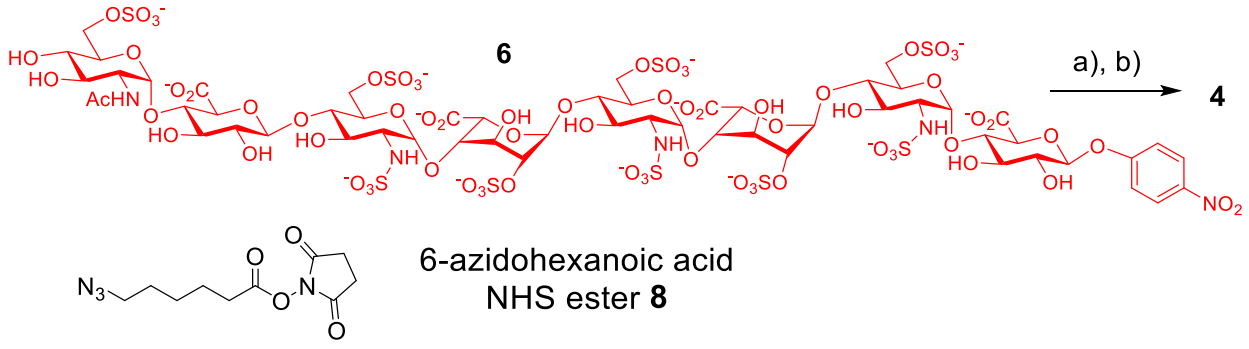


Scheme 2.1 Retrosynthetic analysis of HSPG syndecan-1 mimetic **1**.

Scheme 2.1 (cont'd)



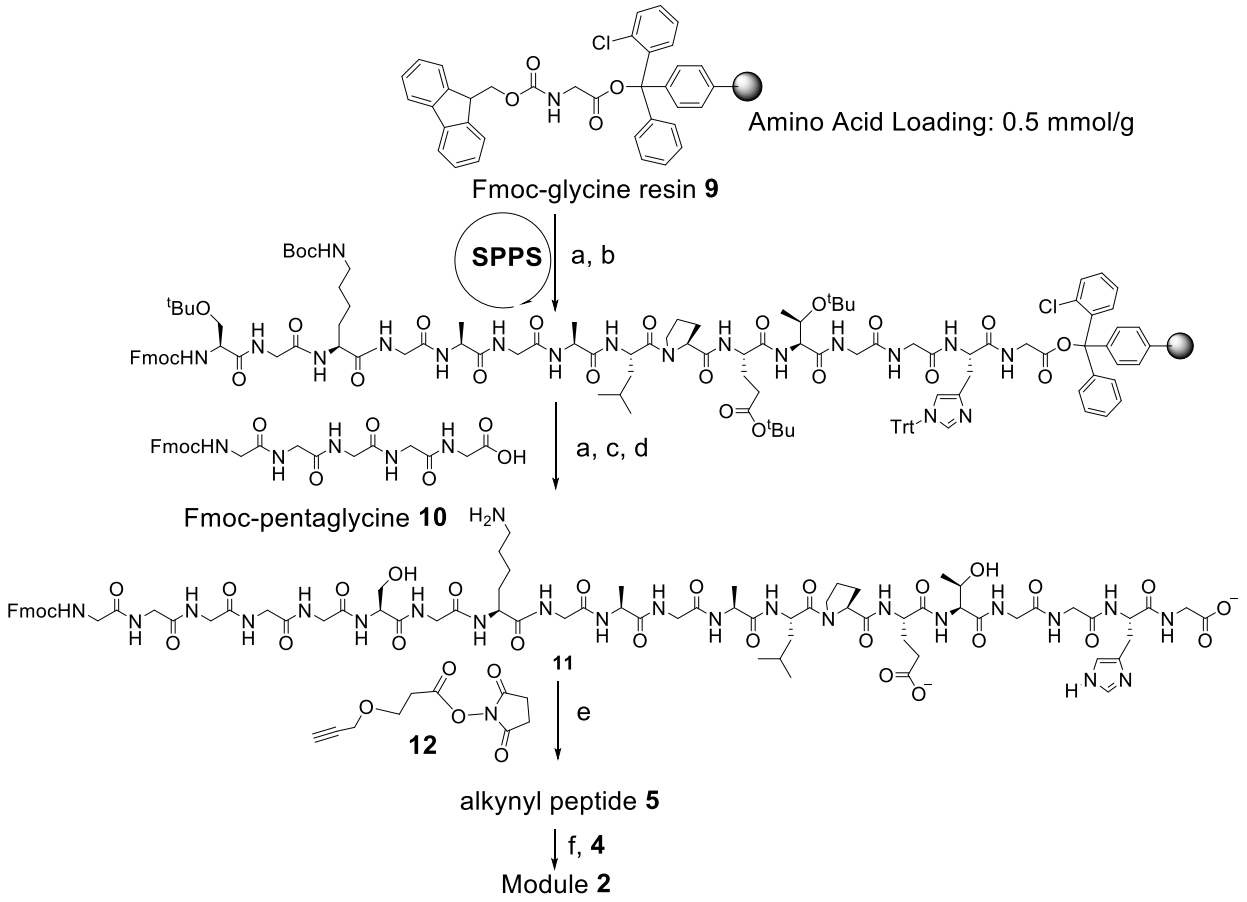
The heparin octasaccharide 6 was synthesized by the Liu group.[12] To prepare the HS oligosaccharide **4**, the nitro moiety in the aglycon of heparin octasaccharide **6** (**Scheme 2.2**) was reduced by catalytic hydrogenation.[12] This was followed by the installation of the azide linker at the reducing end with 6-azidohexanoic acid NHS ester **8** leading to azide functionalized HS octasaccharide **4** (**Scheme 2.2**).



Scheme 2.2 Synthesis of HS octasaccharide **4**. Reagents and conditions: (a) Pd/C, H_2, H_2O, 95%; (b) 6-azidohexanoic acid NHS ester **8**, aq. $NaHCO_3$, 78%.

With the glycan in hand, alkynyl peptide **5** was synthesized *via* microwave-assisted solid phase supported peptide synthesis (SPPS) starting from Fmoc-glycine loaded resin **9** (**Scheme 2.3**). The peptide **5** is terminated with pentaglycine at its *N*-terminus. Because of the synthetic difficulties of certain homooligopeptides, the Fmoc-protected pentaglycine building block Fmoc-pentaglycine **10** was prepared in a separate reaction and purified with preparative HPLC.[13, 14] **10** was then introduced to the *N*-terminus of the growing peptide chain attached to the solid phase (**Scheme 2.3**). Subsequent acidic treatment (TFA/TIPS/H_2O) cleaved the Fmoc-Gly5 terminated peptide **11** off the resin with all acid-labile protecting groups removed. After treatment of **11** with the propargyl alkyne NHS ester **12**, the target peptide **5** was obtained in 18% overall yield. In a similar manner using microwave assisted SPPS, the 33-mer synstatin peptide **3** with the *N*-terminus pentaglycine was prepared with an overall yield of 24% (Appendix **Scheme 2.5**).

To obtain the glycopeptide mimetic, azido-oligosaccharide **4** and alkynylpeptide **5** (1:1 molar ratio) were subjected to copper catalyzed alkyne azide cycloaddition (CuAAC) and the desired product module **2** was obtained in 88% yield following diethylaminoethyl cellulose (DEAE)-HPLC purification (**Scheme 2.3**). The CuAAC condition is mild, which did not affect the structural integrity of the HS glycan or the glyco-polypeptide.

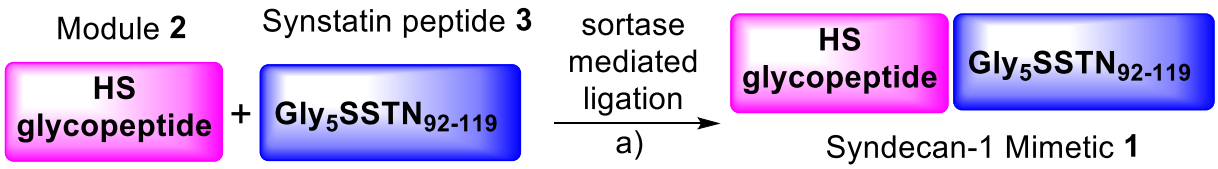


Scheme 2.3 Microwave-assisted synthesis of alkyne-functionalized SorTag-containing peptide **5** and formation of glycopeptide mimetic **2** through the CuAAC. Reagents and conditions: (a) Fmoc-deprotection: 20% piperidine/DMF, 50 °C, 2 min, microwave; (b) Amino acid coupling: 5 eq. Fmoc-AA-OH, HBTU, HOBt, DIPEA, DMF, 50 °C, 10 min, microwave; (c) Oligopeptide coupling: 5 eq. Fmoc-pentaglycine **10**, HATU, DIPEA, DMF, 50 °C, 10 min, microwave; (d) Resin cleavage: TFA/TIPS/H_2O (95:2.5:2.5, v/v/v); (e) Propargyl alkyne NHS ester **12**, aq. $NaHCO_3$, 18% overall. (f) $CuSO_4$, THPTA, Na ascorbate, H_2O, 88%.

To extend the peptide backbone, the key ligation between glycopeptide module **2** and Gly5-$SSTN_{92\text{-}119}$ **3** was carried out under the catalysis of sortase A (SrtA), a transpeptidase that crosslinks the pilin subunits to assemble pili on the surface of gram-positive bacteria (**Scheme 2.4**).[15, 16] To achieve effective ligations, SrtA from *Staphylococcus aureus* ($SrtA_{staph}$) typically requires a LPXTG-containing peptide donor (X can be any natural amino acid) and an acceptor peptide having oligoglycine fragment at its *N*-terminus.[17] $SrtA_{staph}$ is able to irreversibly couple peptide

fragments in the presence of nickel (II) sulfate if the donor peptide carries a Gly-His-Gly tripeptide at the C-terminus of the 'LPETG' sorting signal (SorTag). This results from a Nickel-peptide complex with the histidine residue at the GGHG motif, thus reducing the nucleophilicity and the reversible coupling of the cleaved peptide.[16]

The $SrtA_{staph}$-mediated ligation has been tested using GAGALPETGGHG as the donor peptide and GGGGGLPAG as the acceptor peptide. Reaction conditions, including buffer, pH, temperature, reaction time and the amount of $NiSO_4$ were carefully optimized (Appendix **Table 2.2**) to minimize the undesired hydrolytic activities and improve the coupling efficiency. Incubation of $SrtA_{staph}$ with the peptide donor at weakly acidic or neutral pH (pH 6.0 – 7.0) at 37 °C led to rapid hydrolysis of the donor. Increasing the pH of the reaction media to slightly basic (pH 8.0-8.5) and lowering the reaction temperature to 25 °C in the presence of 1.5 equivalent nickel (II) sulfate completely shut down the hydrolysis side reaction, while retaining a comparable rate of ligation reaction with the acceptor. In the presence of the donor substrate, a quantitative conversion of the substrate into the product was observed in 10 hours as monitored by LC-MS. When the optimized reaction condition was applied to the ligation of glycopeptide module **2** and the synstatin peptide **3** (**Scheme 2.4**), the desired ligation product **1** was obtained in 86% isolated yield on a milligram scale. ^{1}H NMR and HPLC analysis confirmed the product identity and purity. Glyco-polypeptide **1** has a Fmoc moiety at the *N*-terminus, which can be potentially deprotected and serve as a new acceptor for further peptide backbone extension *via* another sortase mediated ligation if necessary.



Scheme 2.4 Sortase A-Mediated Ligation. Reagents and conditions: (a) $SrtA_{staph}$ (5 mol%), 50 mM Tris-HCl buffer, 150 mM NaCl, 5 mM $CaCl_2$, 0.5 mM mercaptoethanol, $NiSO_4$ (1.5 equiv to **2**), pH 8.5, 25°C, 4 hours, 86%.

With the glyco-polypeptide mimetic **1** in hand, we investigated its binding with integrin through biolayer interferometry (BLI). The glyco-polypeptide mimic **1**, Gly_5SSTN_{92-119} **3**, and HS glycan **4** were biotinylated and immobilized onto streptavidin-coated sensors. Their bindings with soluble integrin αvβ3 were measured *via* BLI. While all three compounds were able to bind with integrin αvβ3, interestingly, little dissociation was observed in all cases under the conditions examined (**Figure 2.2**). Kinetic analysis indicated that the glyco-polypeptide mimetic **1** was able to bind integrin faster, with a k_{on} rate more than 2-fold greater than the rates of glycan or synstatin peptide (**Table 2.1**).

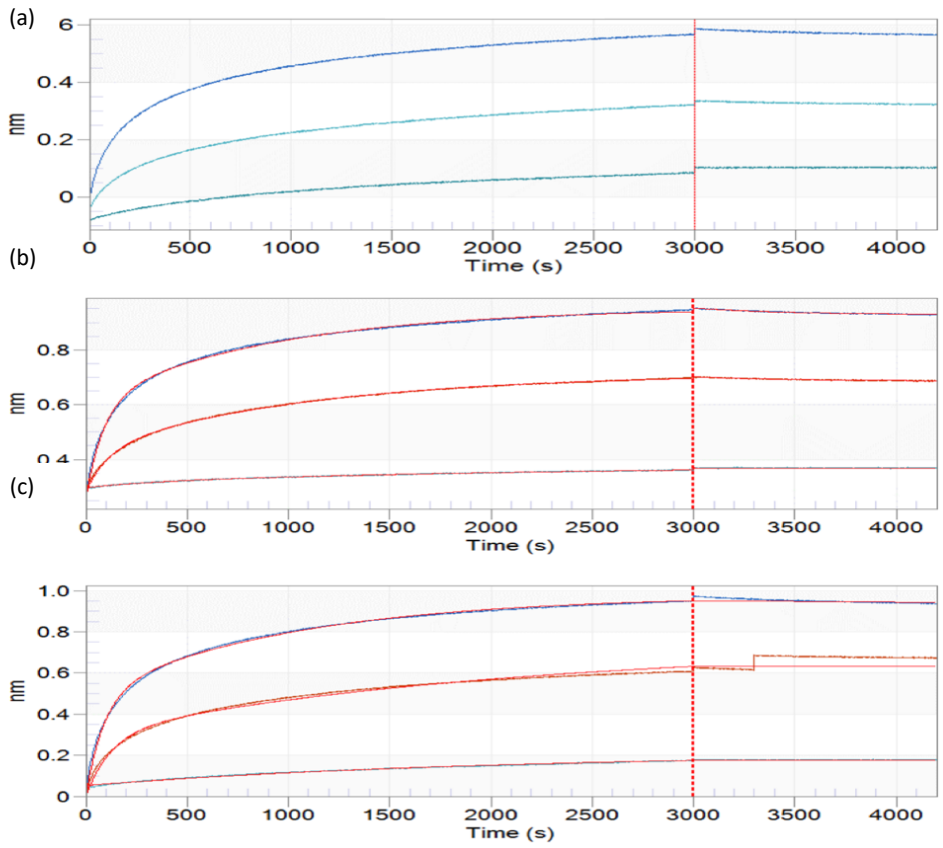


Figure 2.2 BLI sensorgrams of immobilized (a) HS octasaccharide **4**, (b) $Gly_5SSTN_{92\text{-}119}$ **3** and (c) glyco-polypeptide mimetic **1** binding with integrin αvβ3. Each set of binding curves was generated with integrin concentration 104.7 nM, 52.4 nM, and 13.1 nM, from top to bottom. Fitting curves were generated using the 2:1 binding model from Octet Data Analysis 9.0.0.12.

	Syndecan-1 Mimetic **1**	$Gly_5SSTN_{92\text{-}119}$ **3**	HS glycan **4**
k_{on} (1/Ms)	5.08 x 10^4	1.98 x 10^4	9.60 x 10^3

Table 2.1 The on-rates (k_{on}) of **1**, **3**, and **4** with integrin αvβ3.

As the glyco-polypeptide mimetic **1** can bind with integrin strongly, we next measured its effect on cancer cells. MDA-MB-231 breast carcinoma cells activate the cell-surface integrin αvβ3 through the complex formation of syndecan-1, insulin-like growth factor-1 receptor, and integrin to migrate.[18] In addition to syndecan-1 mimetic **1** and $Gly_5SSTN_{92\text{-}119}$ **3**, heparin, which binds more

tightly with integrin than heparan sulfate, was chosen to test the inhibitory effect on the migration of MDA-MB-231 using wound-healing assays (Appendix **Figure 2.4**).[3, 19] Over 20 hours, heparin and Gly5-SSTN$_{92\text{-}119}$ peptide reached the maximal inhibitory effect at the highest testing concentration of each (6 μM). The maximal inhibition from heparin is ~18% reduction in relative migration. Among the analytes, the syndecan-1 mimetic **1** at 6 μM achieved the strongest inhibition, >30% reduction in relative migration (**Figure 2.3** and Appendix **Table 2.3**).

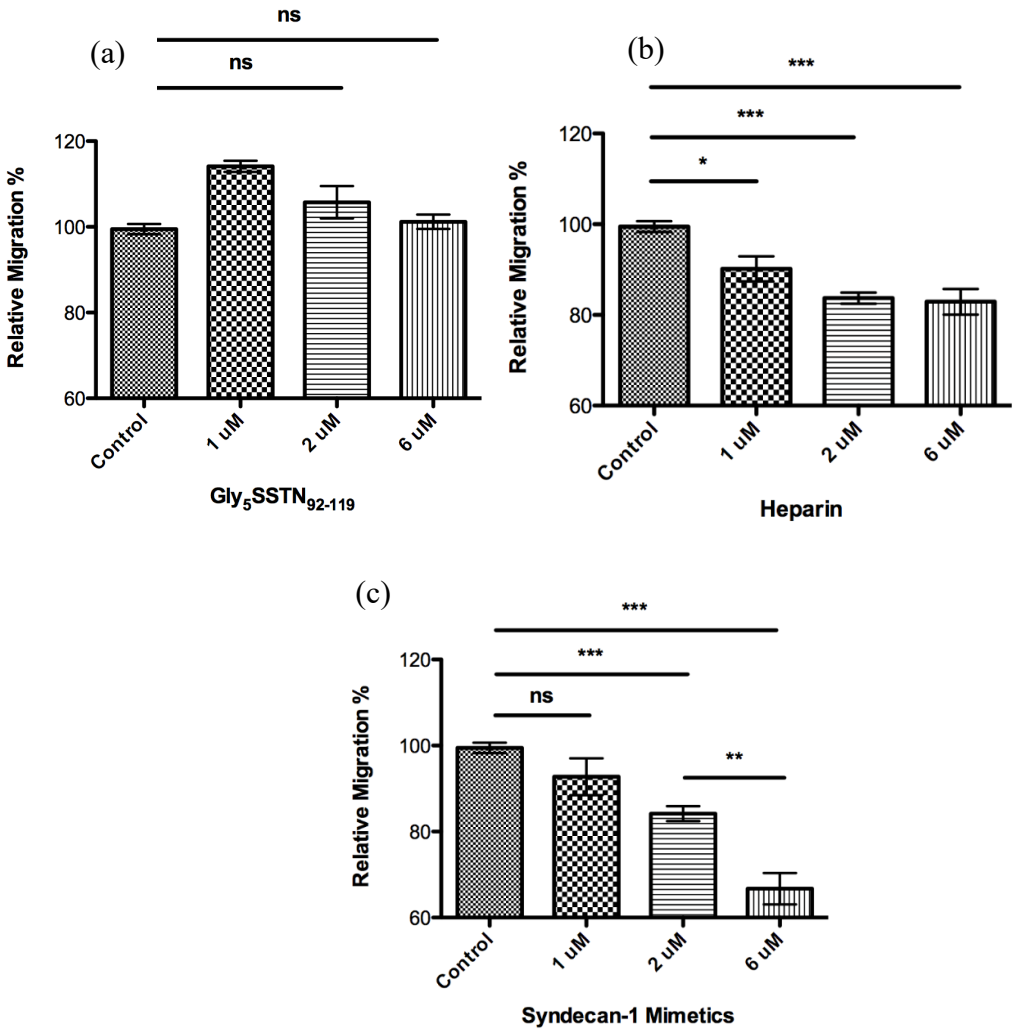


Figure 2.3 Wound-healing assay results of (a) Gly$_5$SSTN$_{92\text{-}119}$ **3**, (b) heparin, and (c) syndecan-1 mimetic **1**. Each plot is displayed as mean ± S.D. of six biological replicates. T test was used for statistical analysis. *p<0.05, **p<0.01, ***p<0.001. The p values were determined through a two-tailed unpaired t-test using GraphPad Prism.

It is possible that the enhanced efficacy of the HSPG mimic **1** compared to glycan or peptide alone was due to the ability of the mimic to simultaneously engage multiple binding sites on the integrin. To gain insights on the integrin αvβ3 binding process, *in silico* molecular docking simulations were performed, and potential integrin binding sites of $Gly_5SSTN_{92\text{-}119}$ peptide **3** and HS oligosaccharide **5** on the integrin were identified (Appendix **Figure 2.5** and **Figure 2.6**).[20, 21] The syndecan-1 mimetics **1** was found to be large enough to bridge the synstatin and HS binding sites at the same time (Appendix **Figure 2.7**, **Table 2.4** and **Table 2.5**). This finding supports a potential synergy from $SSTN_{92\text{-}119}$ and HS in integrin binding.

2.3 Conclusions

In conclusion, with the tremendous structural complexity of HSPG, access to homogeneous HS glycopeptides with defined structures is highly challenging. In this chapter, I developed an expedient approach to produce an HSPG mimetic, which contain a 48 amino acid residue polypeptide backbone and the glycan chain with the full structural features of HS in nature including iduronic acid, glucuronic acid, 2-*O*, 6-*O* and 3-*O* sulfations, and *N*-sulfation. The deployment of HS synthetic enzymes, CuAAC and sortase A-mediated ligation greatly shortens the synthetic routes and enhances the overall efficiency of the synthesis. The synthetic strategy is convergent, which can offer great potential flexibility in varying the glyco-polypeptide structures with other peptide or glycan sequences.

The interaction of the glyco-polypeptide mimic **1** with integrin was investigated. Binding study showed that the glycopeptide was able to engage integrin αvβ3 faster than either the HS glycan or synstatin peptide alone. Although, for all three ligands, dissociations are slow, the higher on-rate of HSPG mimetic suggested a cooperation of HS oligosaccharide and synstatin in integrin binding. Furthermore, the glycopeptide **1** inhibited the migration of triple negative breast cancer

cell MDA-MB-231, opening up the door to investigate the cellular functions of HSPG with structurally well-defined mimetics.

2.4 Experimental Section

2.4.1 Materials

Sortase A-expressing BL21 cells were obtained from Prof. Xue-long Sun (Cleveland State University, OH). Gibco LB broth and LB agar were purchased from Thermo Fischer Scientific (Waltham, MA). Nickel columns and Nickel resins were purchased from Bio-rad (Hercules, CA). SDS-PAGE gels and 10x Tris/Glycine/SDS electrophoresis buffer were purchased from Bio-rad (Hercules, CA). Tris-HCl buffer was purchased from MilliporeSigma (St. Louis, MO). Sephadex G-15 and G-25 were purchased from MilliporeSigma (St. Louis, MO). EZ-Link^{TM} Sulfo-NHS-LC-Biotin was purchased from Thermo Fischer Scientific (Waltham, MA). Recombinant human integrin αvβ3 was purchased from R&D Systems (Minneapolis, MN). Heparin sodium salt was purchased from MilliporeSigma (St. Louis, MO). MDA-MB-231 breast carcinoma cells were obtained from Prof. Kathy Gallo (Michigan State University, MI). Dulbecco's Modified Eagle Medium (DMEM) was purchased from MilliporeSigma (St. Louis, MO). Fetal Bovine Serum was purchased from Thermo Fischer Scientific (Waltham, MA). Human EGF was purchased from Alomone labs Ltd. (Jerusalem, Israel). Human vitronectin protein was purchased from R&D Systems (Minneapolis, MN).

2.4.2 Preparation of Oligosaccharide 7

The octasaccharide compound **6** was dissolved in H_2O (5 mg/ml), to which Pd/C (10 mg/ml) was added. The mixture was then placed under a hydrogen balloon and stirred at room temperature for 1 h. After completion of the reaction, the mixture was filtered through a PTFE

syringe filter (0.2 mm, 13 mm). The filtrate was concentrated, and the desired product was purified by a Sephadex G-10 column.

2.4.3 Preparation of Oligosaccharide 4

Compound **7** was dissolved in aqueous solution of $NaHCO_3$ at pH 8.5, after which 1.5 equivalents of 6-azidohexanoic acid NHS ester in anhydrous DMF were added. The reaction was then stirred at room temperature for 6 hours. Upon completion, the reaction mixture was directly loaded onto a Sephadex G-10 column for purification.

2.4.4 General Procedure for Automated Solid-Phase Peptide Synthesis

All the peptides reported were synthesized on a Liberty Blue™ Automated Microwave Peptide Synthesizer following standard Fmoc-based solid-phase peptide synthesis protocol. The 2-chlorotrityl resins with or without Fmoc-amino acid loaded were purchased from Chem-Impex (Wood Dale, IL). The Liberty Blue software from CEM Corporation (Matthews, NC) was used to program the synthesis, including resin swelling, amino acid loading, couplings and Fmoc-removals. Commercially available *N,N*-dimethylformamide (DMF) from Fischer Chemical (Hampton, NH) was supplied to the synthesis module as reaction and washing solvent. Peptide synthesis was enabled by sequential couplings of Fmoc-amino acid, purchased from Chem-Impex (Wood Dale, IL), which was preactivated by *N,N,N′,N′*-tetramethyl-O-(1*H*-benzotriazol-1-yl) uronium hexafluorophosphate (HBTU), *N*-hydroxybenzotriazole (HOBt), *N,N*-diisopropylethylamine (DIPEA), at 50 °C for 10 min, and the deprotections with 20% piperidine in DMF at 60 °C for 4 min. In-between each coupling/deprotection step, resin-bound peptide was thoroughly washed with DMF. For the synthesis of Fmoc-Gly5-OH peptide, Fmoc-glycine was preactivated by 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) and DIPEA instead. Resin-bound peptides were cleaved off the

solid support with a cocktail solution of trifluoroacetic acid (TFA), triisopropylsilane (TIPS) and water (TFA/TIPS/H_2O, 95:2.5:2.5). The crude peptides were then purified with reverse-phase C18 preparative HPLC. Compound purity for each peptide was confirmed with chromatograms from C18 analytical HPLC.

2.4.5 High-Performance Liquid Chromatography

LC-8A Solvent Pumps, DGU-14A Degasser, SPD-10A UV-Vis Detector, SCL-10A System Controller (Shimadzu Corporation, JP) and Vydac 218TP 10 μm C18 Preparative HPLC column (HICHROM Limited, VWR, UK) or 20RBAX 300SB-C18 Analytical HPLC column (Agilent Technologies, CA) were used for HPLC purifications with HPLC-grade acetonitrile (EMD Millipore Corporation, MA) and Milli-Q water (EMD Millipore Corporation, MA). A variety of eluting gradients were set up with the LabSolutions software. Dual-wavelength UV detector was set to 220 nm and 254 nm for monitoring the absorbance of the amide and Fmoc, respectively. The eluted compounds were checked with ESI-MS to confirm their identities. Then aqueous solutions of purified compounds were lyophilized to obtain the dry solid.

2.4.6 Sortase A Expression, Purification and Quantification

An aliquot of 2 μL sortase A-expressing BL21 competent cell culture was transferred to a kanamycin/chloramphenicol petri dish. The culture was incubated at 37 °C overnight. One colony of BL21 cells was picked to start a 10 mL culture, containing kanamycin (35 mg/L). The cell culture was incubated at 37 °C for 12-16 h until OD_{600} value reached 0.85. The starter culture was transferred into sterilized culture medium (1L containing 35 mg/L kanamycin). After roughly 5 hours, the OD_{600} reached 0.85. 0.5 mM IPTG was added to induce protein expression. The cell culture was incubated for another 4 hours at 37 °C. The cells were centrifuged at 4 °C, 5000 rpm for 10 min. The cells were then resuspended in 40 mL lysis buffer (20 mM Tris, 250 mM NaCl,

pH 8.0) and lysed by sonication. The lysate was centrifuged (20,000 *g* for 20 min). The supernatant was loaded onto a Ni-affinity column and sortase A was purified by Nickel column, using the following elution profile: a. washing buffer: 20 mM Tris, 0.5 M NaCl; b. eluting buffer: 20 mM Tris, 0.5 M NaCl and 250 mM imidazole. Dialysis was used to remove the imidazole against 2L of buffer (20 mM Tris, 150 mM NaCl, pH 8.0). Protein purity was confirmed by SDS-PAGE and the standard Bradford assay determined the concentration and expression yield of sortase A.

2.4.7 General Procedure for Sortase A-Mediated Ligation

10X Tris-HCl reaction buffer for the sortase A-mediated ligation was prepared in advance following the recipe of 500 mM Tris-HCl, 1.5 M NaCl, 50 mM $CaCl_2$, 5 mM mercaptoethanol, and 2 mM Ni(II) sulfate. The pH of the 10X reaction buffer was adjusted to 8.5 with addition of NaOH or HCl. Proper amounts of ligation substrates were dissolved and added into Tris-HCl reaction buffer, followed by the addition of sortase A. The reaction vessel was then kept at 25 °C until reaction completion. Reaction progress was monitored with LC-MS. After the reaction, enzyme was deactivated and precipitated out by addition of ethanol. The reactions were clarified by centrifugation and the supernatant was loaded onto G-15/G-25 size exclusion column for purification.

2.4.8 Size-Exclusion Purification of HS Glycopeptide

Samples were prepared in minimal amounts of distil water and then slowly transferred to a G-15/G-25 size-exclusion column. Fractions of 1 mL eluent were collected. Fractions that contain desired compounds were identified by ESI-MS analysis. Purified compounds were lyophilized to obtain the dry solid.

2.4.9 BLI Binding Experiment

BLI Octet K2 instrument (ForteBio, Molecular Devices, CA) was used for binding

experiments. Polypropylene black 96-well plates (Greiner Bio-one, Austria) and streptavidin (SA) sensor chips (ForteBio, Molecular Devices, CA) were used to assist sample preparations and detections of binding activities. The assay buffer was phosphate buffered saline (PBS) unless otherwise noted. Integrin αvβ3 protein solutions were prepared according to the assay design. To prepare biotinylated analytes, 1 mM of each amine-containing ligand compound and EZ-Link™ Sulfo-NHS-LC-Biotin (1.2 equiv.) were added to 0.1 M $NaHCO_3$ solution (pH 8.5). The reaction was proceeded overnight. Upon completion, reaction mixture was passed through G-10 column to remove the unconjugated biotin reactant. Sensors were then loaded with the biotin-labelled compounds. The binding activity (including association and dissociation) between the ligand and protein was measured by BLI monitoring. Biotin was used as the negative control for all BLI assays. The assay results were then processed by the Octet software. Various concentrations of protein were tested against each ligand to obtain the kinetic data. The curve fitting was achieved using a 2:1 heterogenous ligand binding model provided by the data-processing software.

2.4.10 Wound-Healing Assay

MDA-MB-231 breast carcinoma cells were cultivated in the 6-well plate until 90% confluent. After 24-hour starvation with serum-free medium, wounds were created by scratching the monolayer with sterile P200 pipet tips. This process was done carefully to make sure that all wounds were similar in size. A Zeiss Axionvert 200 Pred Axio Observer microscopy (Boston Industries, Inc.) was used to take microscopic images. T = 0 images were taken right after the wounding process. Then the serum-free medium was replaced by the growth medium that contained varying analyte concentrations. Growth medium without analytes was treated as the control group. Human EGF was added to stimulate cancer cell migrations. After a 20-hour incubation at 37 °C, T = 20 microscopic images were collected. Images at the same site for T = 0

and T = 20 were processed with GraphPad Prism Version 5.0c to interpret the T = 20 cell migration results.

2.4.11 Identification of Ligand Binding Sites

To initiate the search for ligand binding sites on the integrin αvβ3 protein surface, synstatin peptide $SSTN_{92\text{-}119}$ model was constructed *de novo* using an open computation platform developed by Tuffery group.[20] Integrin protein (PDB:4G1M) was used as the receptor reference to facilitate the model construction and improve the subsequent docking simulations. Independent model simulations (200 rounds) with sOPEP force field were applied to get quality peptide conformation predictions. The best candidate models were selected for the ligand-receptor molecular docking simulations.

Hot spots on the protein surface for synstatin binding were identified through examining the docking results. For the heparan sulfate binding simulations, a generic heparan sulfate tetrasaccharide structure was utilized to identify potential HS binding sites on integrin αvβ3. After uploading the integrin coordinate file to ClusPro docking platform, binding simulations were initiated under the built-in 'Heparin Ligand' mode.[22] Simulation results were then visualized and processed with UCSF Chimera software to pinpoint potential HS binding sites.[23]

2.4.12 Biomolecule Visualization

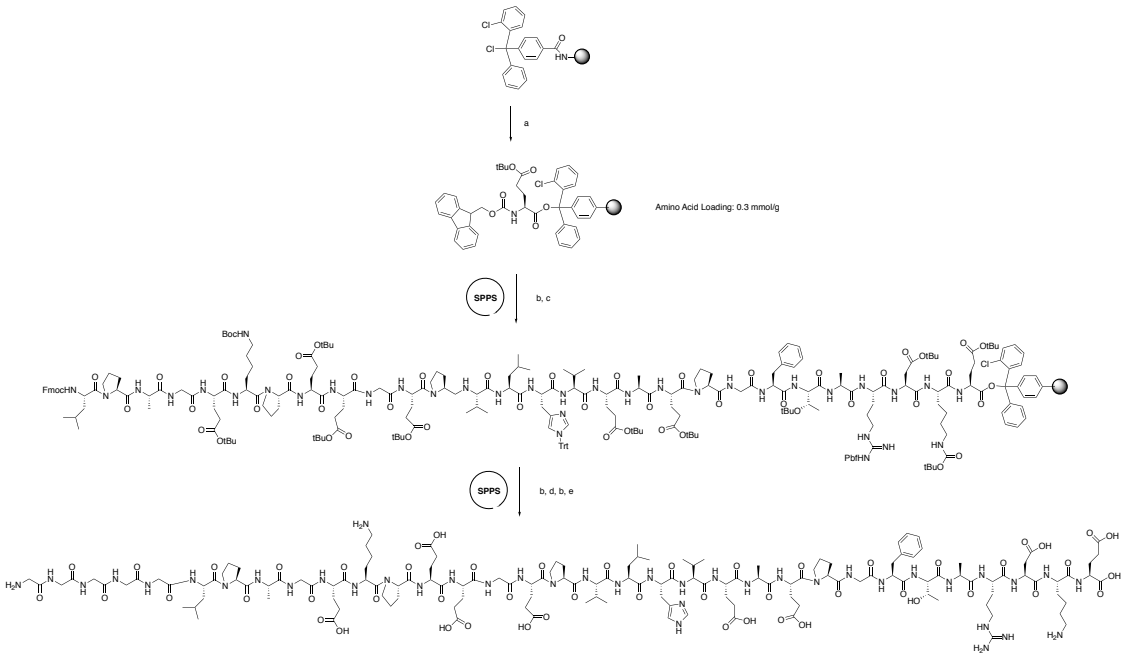
The construction of syndecan-1 mimetic started with the heparan sulfate octassachride moiety. Its structure was prepared through 'GAG Builder' program at GLYCAM-Web.[24] Counter ions were added to the negatively charged sulfate groups and the HS octassachride was solvated into a cube of water molecules. Structural optimization was accomplished with GLYCAM force field. The generated PDB file of HS octassachride was later used to construct the glycopeptide.

Ab initio modeling of the peptide backbone was achieved with QUARK program.[25] Top model that adopted a more extended conformation was selected for the syndecan-1 mimetic construction. The structure coordinates of HS octassachride and peptide backbone were input into Maestro software.[26] The artificial linkage connecting HS and peptide backbone was manually created. The resulted syndecan-1 mimetic structure was then optimized using all-atom minimization function to approximate its conformation. The dimensions of syndecan-1 mimetic and integrin αvβ3 were measured with UCSF Chimera to provide an estimation of their sizes.

APPENDICES

APPENDIX A: Supplementary Schemes, Figures and Tables

Solid-Phase Synthesis of Synstatin Peptide 3



Scheme 2.5 Solid-phase synthesis of Gly5-SSTN$_{92\text{-}117}$ peptide. Reagents and conditions: (a) Amino acid Loading: Fmoc-Glu(O-tBu)-OH, DIPEA, DMF; (b) Fmoc cleavage: 20% piperidine/DMF, 50 °C, 2 min, microwave; (c) Amino acid coupling: 5 equiv Fmoc-AA-OH, HBTU, HOBt, DIPEA, DMF, 50 °C, 10 min, microwave; (d) Oligopeptide coupling: 5 equiv Fmoc-Gly5-OH, HATU, DIPEA, DMF, 50 °C, 10 min, microwave; (e) Resin cleavage: TFA/TIPS/H_2O (95:2.5:2.5, v/v/v), 24 % overall.

MDA-MB-231 Wound Healing Assay Images

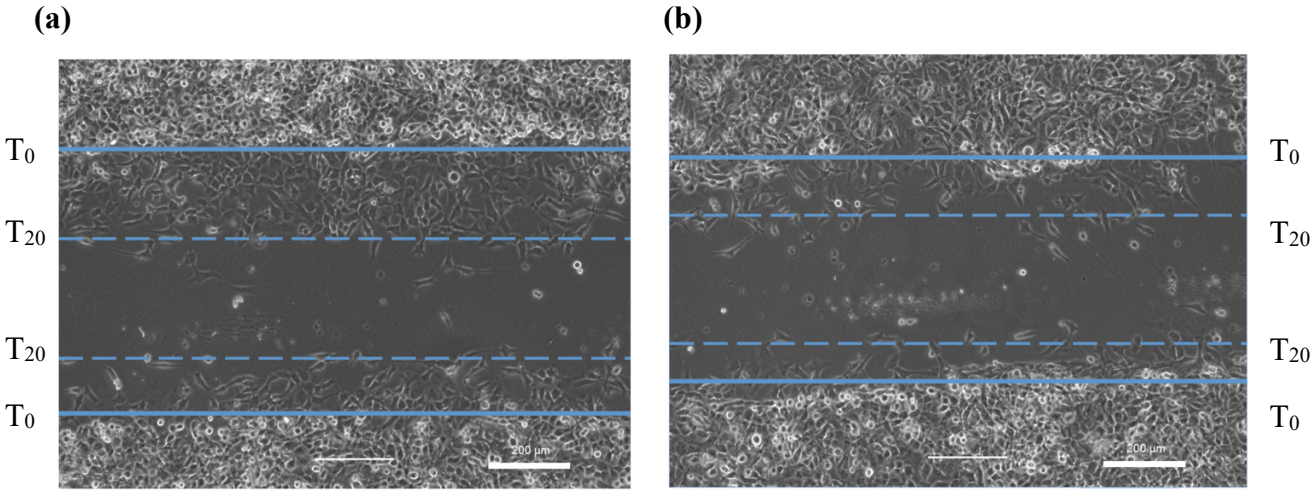


Figure 2.4 Microscopy images of MDA-MB-231 treated with (a) PBS as control and (b) synthetic HS glycopeptide (6 µM) after 20-hour incubation (solid lines for cell frontiers at T=0 and dashed lines for T=20; 10X magnification; scale bar, 200 µm).

Computer Docking Simulation Result and Biomolecule Visualization

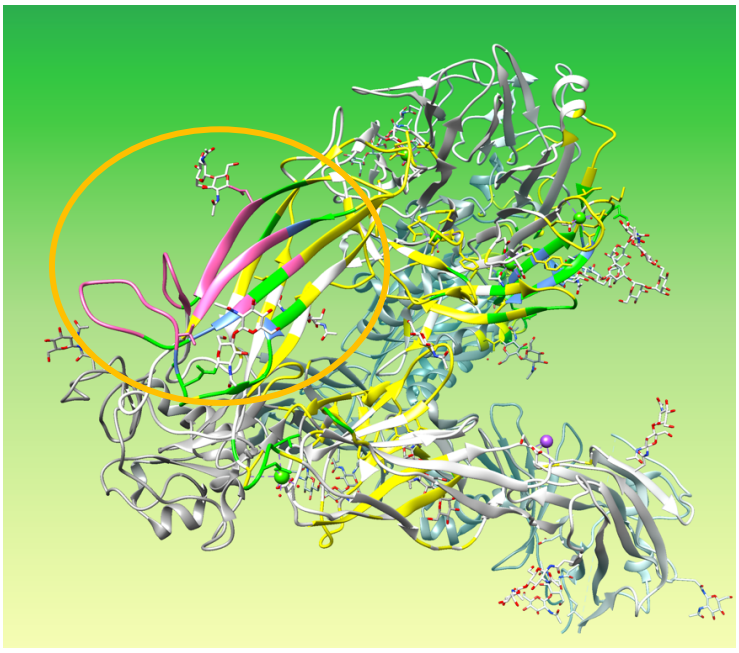


Figure 2.5 Identified synstatin peptide binding site (as circled) on the surface of integrin αvβ3 (PDB: 4G1M).

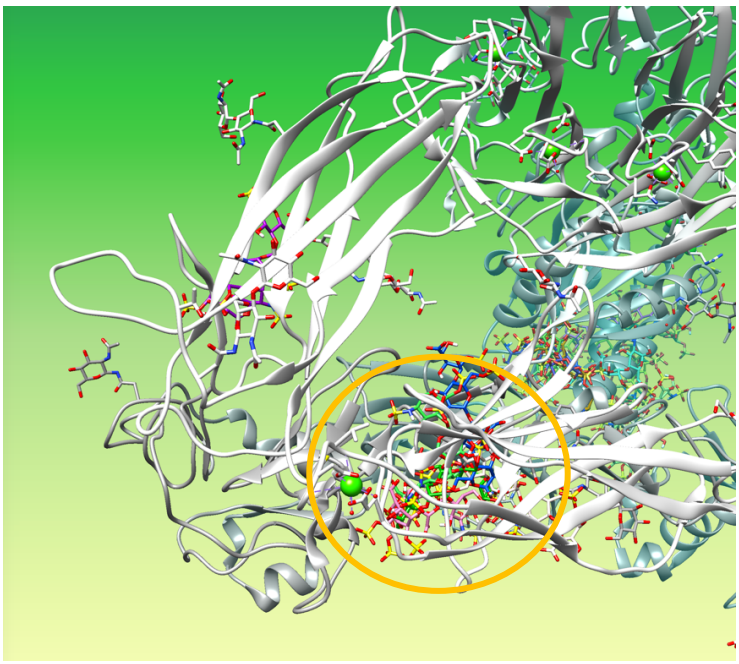


Figure 2.6 One of the identified heparan sulfate binding sites (as circled) on the surface of integrin αvβ3 (PDB: 4G1M).

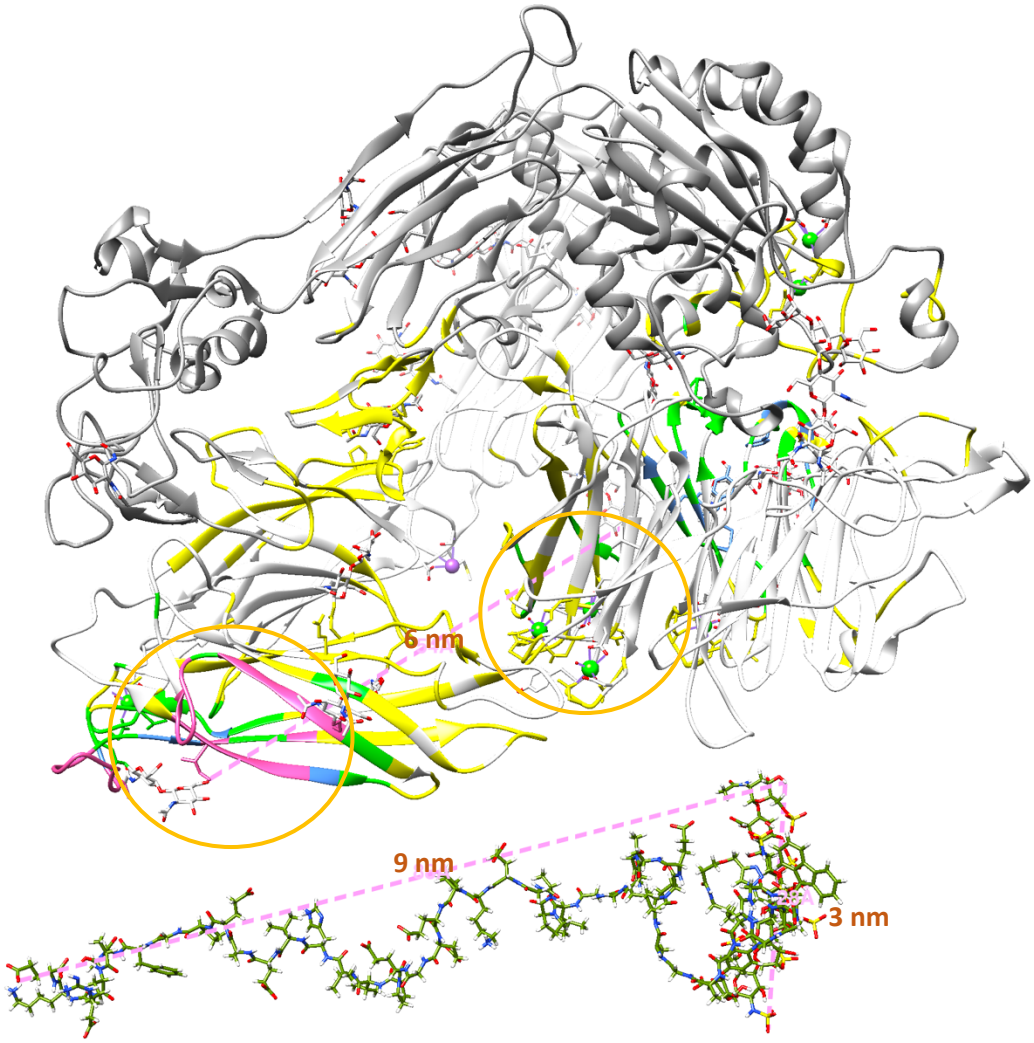


Figure 2.7 Biomolecule visualization and approximate size comparison of syndecan-1 mimetic (lower structure) and integrin αvβ3 (PDB: 4G1M). Predicted binding areas of synstatin peptide and heparan sulfate tetrasaccharide are highlighted with orange circles.

Screening Conditions for Sortase A Ligation Reaction

Substrate	**Concentration (μM)**
Peptide 1: GAGALPETGGHG	**250**
Peptide 2: GGGGGLPAG	**250**
Product	
Peptide: GAGALPETGGGGGLPAG	

Sortase Reaction Condition						
	Buffer	mol% Sortase	Reaction Time	Reaction Temperature	% conversion	Hydrolyzed Donor/Product
pH 7.0	300 mM Tris-HCl	6	10 h	37 °C	67.9	0.01:1
pH 7.5					59.8	N/A
pH 8.0					71.9	N/A
pH 8.5					71.0	N/A

Sortase Reaction Condition						
	Buffer	mol% Sortase	Reaction Time	Reaction Temperature	% conversion	Hydrolyzed Donor/Product
pH 7.0	300 mM Tris-HCl	12	10 h	37 °C	67.9	0.13:1
pH 7.5					69.0	0.01:1
pH 8.0					67.4	0.05:1
pH 8.5					N/A	N/A

Table 2.2 Screening of sortase A ligation conditions.

Table 2.2 (cont'd)

Sortase Reaction Condition						
	Buffer	mol% Sortase	Reaction Time	Reaction Temperature	% conversion	Hydrolyzed Donor/Product
pH 7.0	300 mM Tris-HCl	24	10 h	37 °C	60.5	0.26:1
pH 7.5					62.6	0.14:1
pH 8.0					66.7	0.11:1
pH 8.5					68.7	0.13:1

Sortase Reaction Condition						
	Buffer	mol% Sortase	Reaction Time	Reaction Temperature	% conversion	Hydrolyzed Donor/Product
pH 7.0	50 mM Tris-HCl	24	10 h	37 °C	51.5	0.39:1
pH 7.5					59.4	0.15:1
pH 8.0					64.6	0.15:1
pH 8.5					62.7	0.19:1

Table 2.2 (cont'd)

Sortase Reaction Condition						
	Buffer	mol% Sortase	Reaction Time	Reaction Temperature	% conversion	Hydrolyzed Donor/Product
pH 7.0	300 mM Tris-HCl	12	45 h	37 °C	26.8	1.23:1
pH 7.5					46.1	0.44:1
pH 8.0					44.9	0.18:1
pH 8.5					45.7	0.16:1

Wound Healing Assay Result

Gly5-SSTN						
Concentration (μM)	Relative Migration Area (Unit)					
6	103.92	99.61	100.60	107.91	98.96	96.18
2	102.01	102.77	99.73	97.85	122.89	109.22
1	111.56	109.67	118.42	115.84	115.74	113.54
0	100.70	98.40	94.45	99.88	103.13	100.35

Heparin						
Concentration (μM)	Relative Migration Area (Unit)					
6	82.62	95.3	83.28	80.77	73.85	81.61
2	88.3	83.49	85.63	83.17	82.23	79.42
1	97.95	98.89	88.91	83.87	88	83.28
0	100.70	98.40	94.45	99.88	103.13	100.35

Syndecan-1 Mimetic						
Concentration (μM)	Relative Migration Area (Unit)					
6	65.04	65.37	80.11	52.96	65.27	71.59
2	82.81	84.61	92.50	81.69	82.47	80.88
1	82.26	85.77	103.42	94.08	106.81	84.19
0	100.70	98.40	94.45	99.88	103.13	100.35

Table 2.3 Summary of wound-healing assay results.

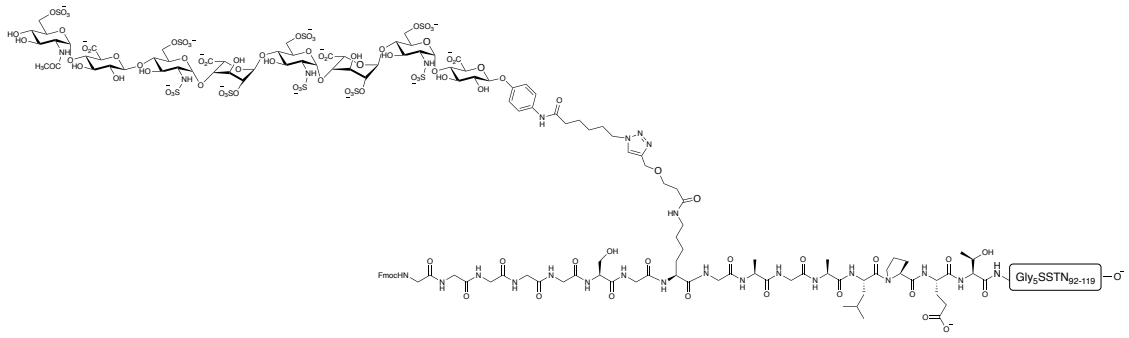
	Estimated Center-to-Center Distance
The Spotted Binding Sites	**6 nm**

Table 2.4 Measured estimated distance of the spotted synstatin and heparan sulfate binding sites.

	Longitudinal	Transversal
Syndecan-1 Mimetics	**9 nm**	**3 nm**

Table 2.5 Measured approximate dimensions of integrin αvβ3 and syndecan-1 mimetic.

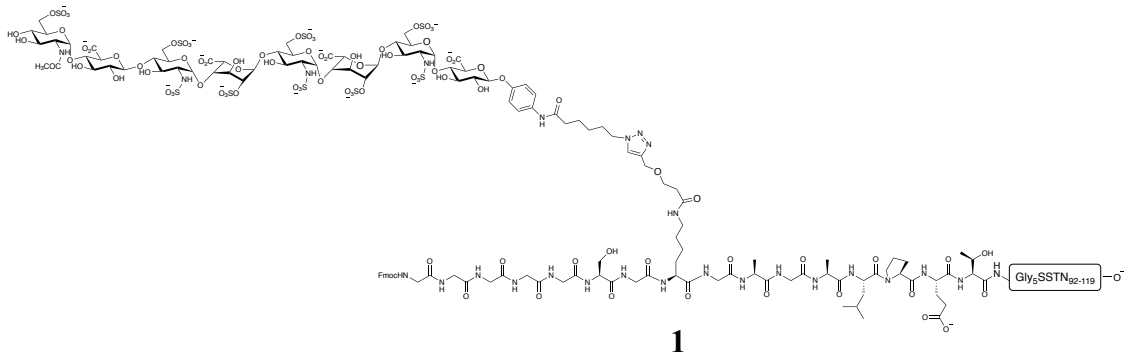
APPENDIX B: Product Characterization Spectra



1

The purity of glycopeptide **1** was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). ^{1}H-NMR (900 MHz, D_2O), δ 8.35 (m, 2H), 7.88 – 7.67 (m, 1H), 7.64 – 7.52 (m, 1H), 7.42 – 7.30 (m, 1H), 7.30 – 7.23 (m, 1H), 7.18 – 7.07 (m, 1H), 5.37-5.27 (m, 1H), 5.18-5.11 (m, 1H), 5.11-5.03 (m, 1H), 5.01-4.94 (m, 1H), 4.92-4.88 (m, 4H), 4.86-4.82 (m, 3H), 4.82-4.77 (m, 16H), 4.61– 4.55 (m, 3H), 4.55 – 4.42 (m, 3H), 4.41-4.39 (m, 1H), 4.39 – 4.22 (m, 6H), 4.22 – 4.03 (m, 8H), 4.02-3.97 (m, 3H), 3.97 – 3.92 (m, 4H), 3.92-3.88 (m, 6H), 3.87 – 3.83 (m, 6H), 3.78-3.76 (m, 4H), 3.75-3.72 (m, 5H), 3.72 – 3.65 (m, 14H), 3.65 – 3.61 (m, 12H), 3.61-3.59 (m, 7H), 3.59-3.56 (m, 10H), 3.56 – 3.53 (m, 13H), 3.53-3.51 (m, 51H), 3.50 – 3.45 (m, 12H), 3.45 – 3.38 (m, 5H), 3.28-3.22 (m, 1H), 3.22-3.16 (m, 1H), 3.13 – 3.09 (m, 6H), 3.08 – 3.04 (m, 1H), 3.04-2.99 (m, 1H), 2.98 – 2.86 (m, 2H), 2.64 – 2.60 (m, 1H), 2.57-2.49 (m, 1H), 2.34 – 2.28 (m, 1H), 2.24 – 2.12 (m, 6H), 2.10-2.03 (m, 2H), 2.00 – 1.91 (m, 7H), 1.90 – 1.83 (m, 16H), 1.81-1.76 (m, 71H), 1.76 – 1.65 (m, 4H), 1.64 – 1.54 (m, 3H), 1.53 – 1.42 (m, 4H), 1.42-1.33 (m, 4H), 1.33 – 1.31 (m, 2H), 1.28-1.27 (m, 2H), 1.27 – 1.22 (m, 43H), 1.22 – 1.15 (m, 3H), 1.13 – 1.02 (m, 3H), 0.88-0.70 (m, 8H). ^{13}C-NMR (225 MHz, D_2O), δ 152.3, 152.3, 152.2, 148.0, 147.7, 147.6, 147.5, 147.3, 147.2, 147.1, 147.0, 146.9, 146.8, 146.5, 145.9, 129.1, 128.6, 128.6, 128.5, 128.0, 127.4, 127.2, 127.1, 125.0, 123.5, 123.4, 120.2, 116.8, 101.6, 101.6, 100.2, 100.1, 99.0, 98.4, 98.3,

97.6, 96.9, 96.5, 96.4, 95.6, 95.3, 95.2, 93.1, 79.9, 76.5, 76.4, 76.2, 75.9, 75.8, 75.7, 75.4, 75.3, 75.1, 74.0, 73.4, 73.3, 73.2, 72.5, 72.1, 71.8, 71.7, 71.4, 71.3, 71.1, 70.7, 70.2, 70.1, 69.8, 69.6, 69.5, 69.4, 69.1, 69.0, 68.9, 68.8, 68.6, 67.5, 67.4, 67.3, 67.2, 67.1, 66.8, 66.5, 66.2, 66.1, 66.0, 65.6, 65.4, 64.3, 64.0, 62.9, 62.8, 62.5, 62.4, 62.0, 61.2, 61.1, 60.5, 60.4, 60.3, 59.6, 59.4, 59.2, 59.2, 59.1, 59.0, 58.8, 58.5, 58.3, 58.2, 58.0, 57.7, 57.5, 55.8, 55.7, 55.2, 55.1, 55.0, 54.4, 54.3, 54.2, 53.8, 53.7, 53.6, 53.5, 53.4, 53.3, 52.0, 51.9, 51.7, 51.6, 50.3, 50.2, 50.1, 49.8, 49.6, 49.5, 47.9, 47.4, 46.8, 46.7, 44.5, 44.4, 43.6, 43.6, 43.2, 43.2, 43.6, 42.6, 42.4, 42.3, 40.4, 39.9, 39.8, 39.1, 39.0, 38.9, 38.9, 38.7, 38.6, 37.1, 37.0, 36.0, 35.9, 34.4, 34.0, 33.8, 33.7, 33.5, 33.4, 30.5, 30.2, 30.0, 29.3, 29.2, 28.9, 28.6, 27.1, 26.2, 24.7, 24.6, 24.4, 24.3, 24.2., 24.1, 24.0, 23.3, 23.0, 22.7, 22.4, 22.4, 22.0, 22.0, 21.9, 21.7, 21.4, 20.9, 20.6, 20.0, 19.3, 18.8, 18.7, 18.6, 18.4, 18.2, 18.2, 18.0, 17.9, 17.8, 17.7, 17.2, 16.7, 16.6, 16.5, 16.5, 16.4, 15.7, 13.3, 12.9, 12.8, 12.3, 12.2, 12.1. ESI-MS: $C_{277}H_{411}N_{65}O_{145}S_9$ $[M+10H]^{4-}$ calcd: 1451.0859, obsd: 1451.0818 (2.84 ppm).



1

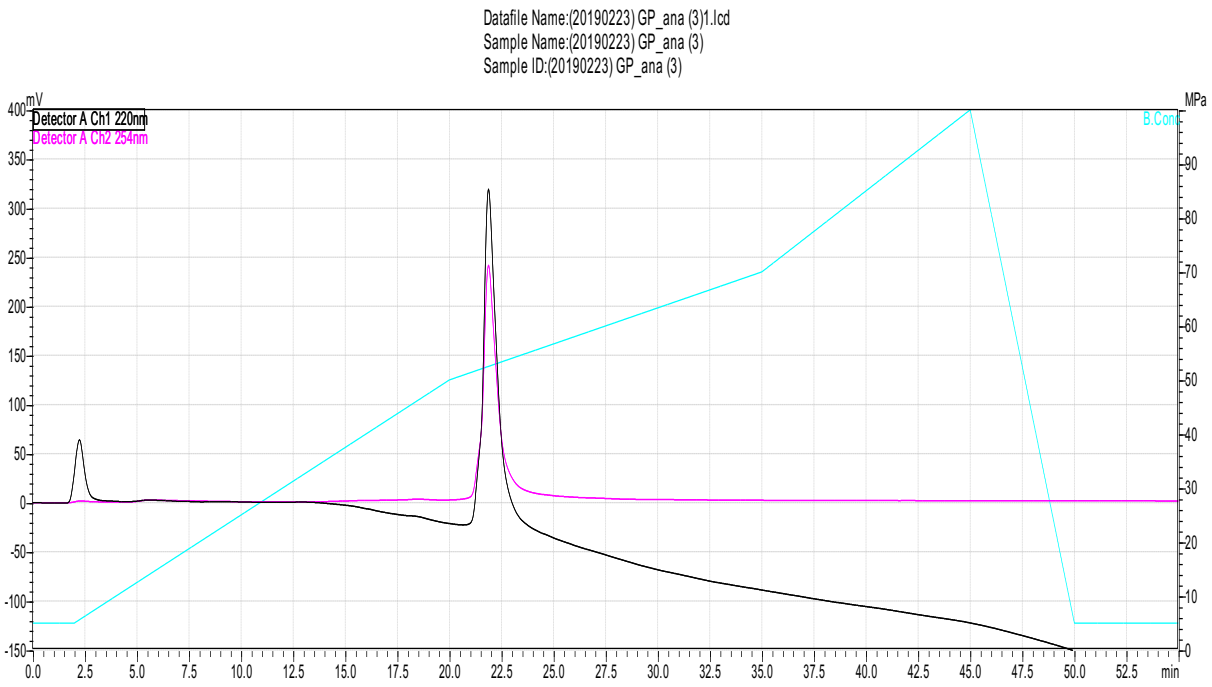
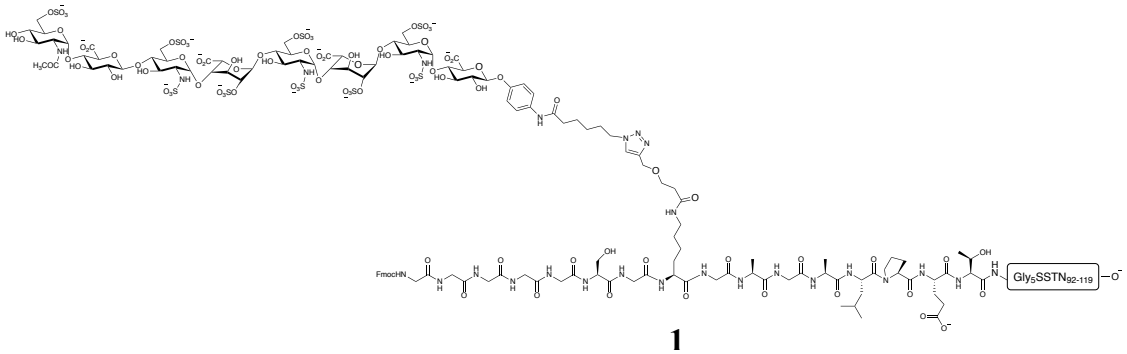


Figure 2.8 HPLC chromatogram of **1**.



1

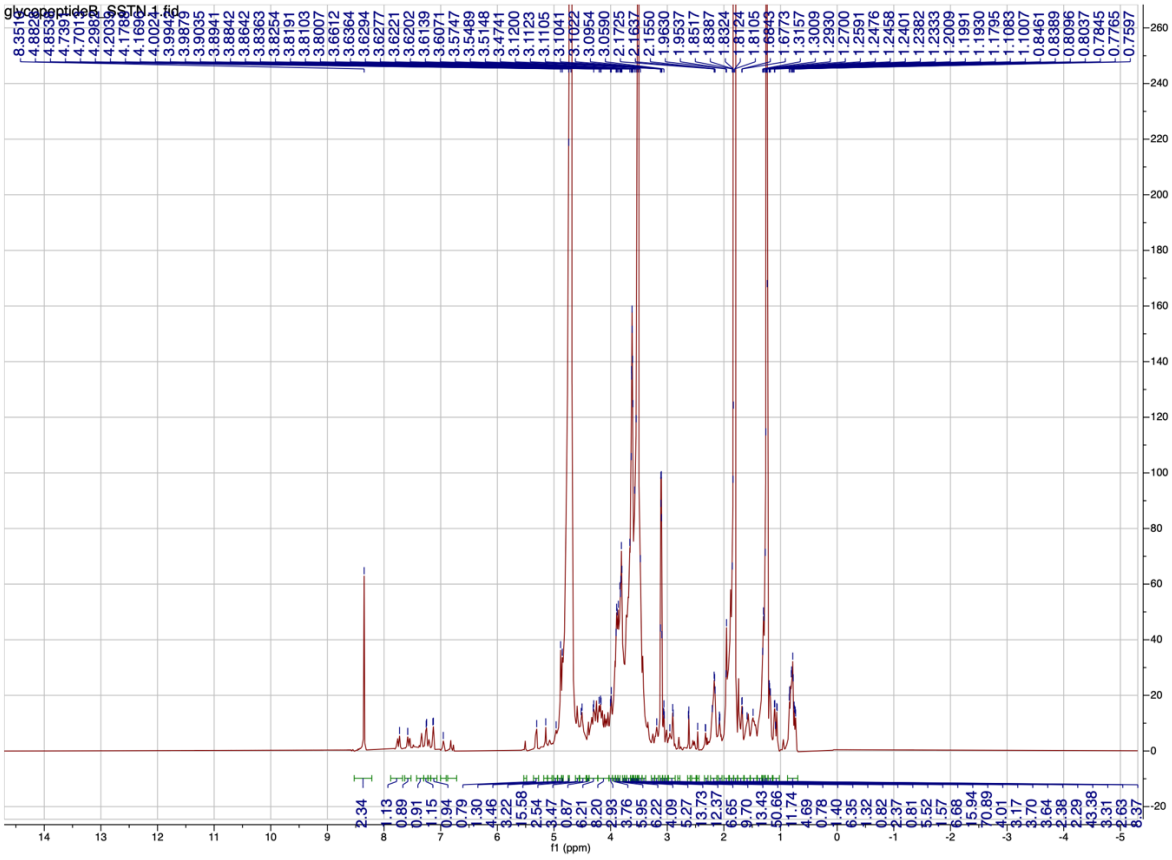
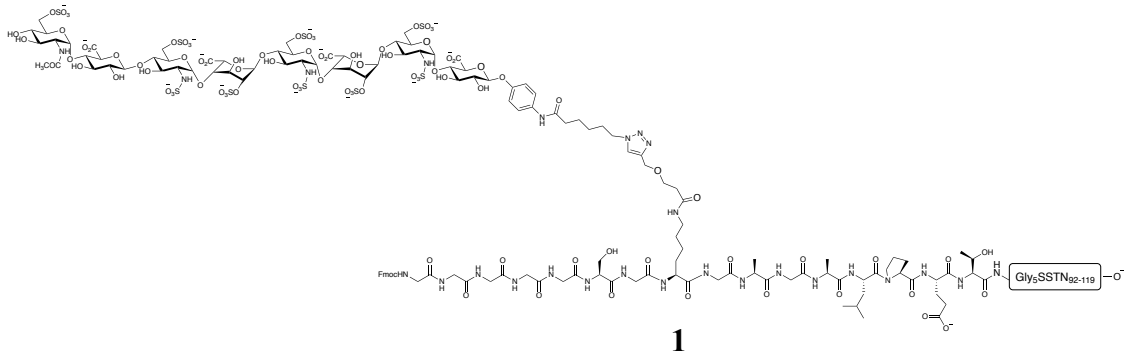


Figure 2.9 ^{1}H-NMR of **1** (900 MHz D_2O).



1

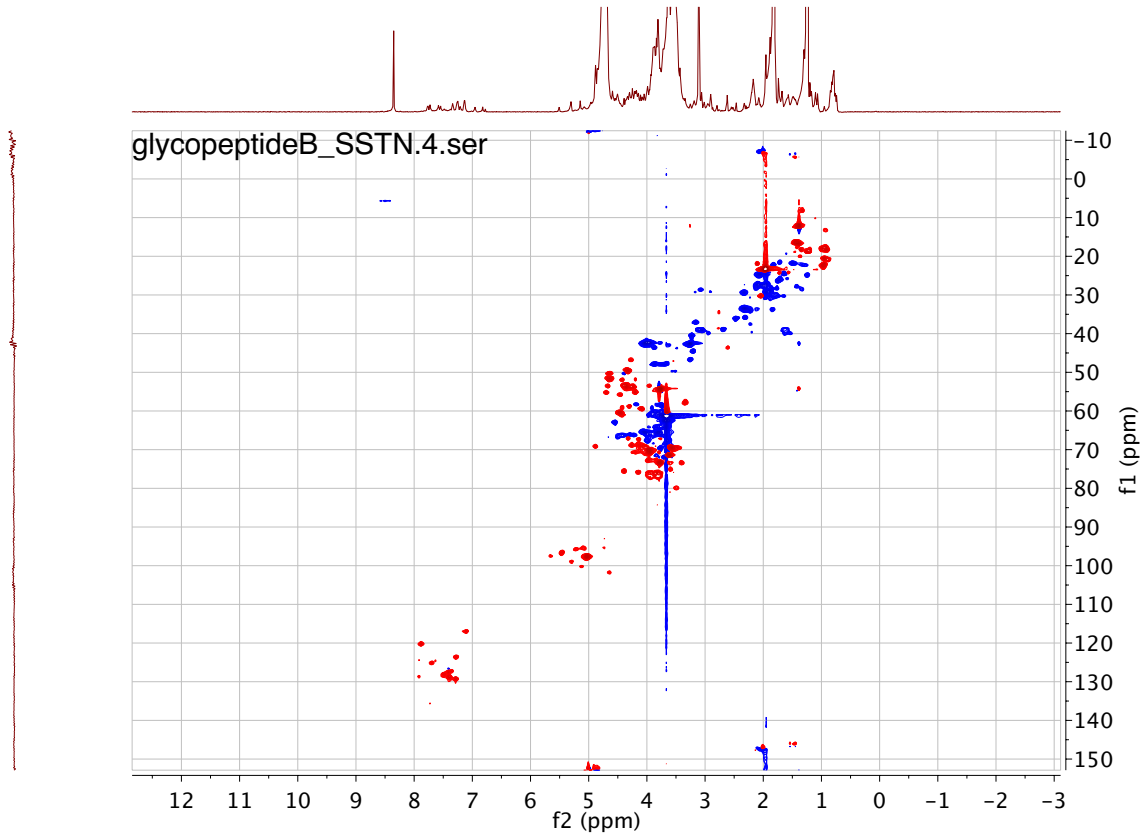
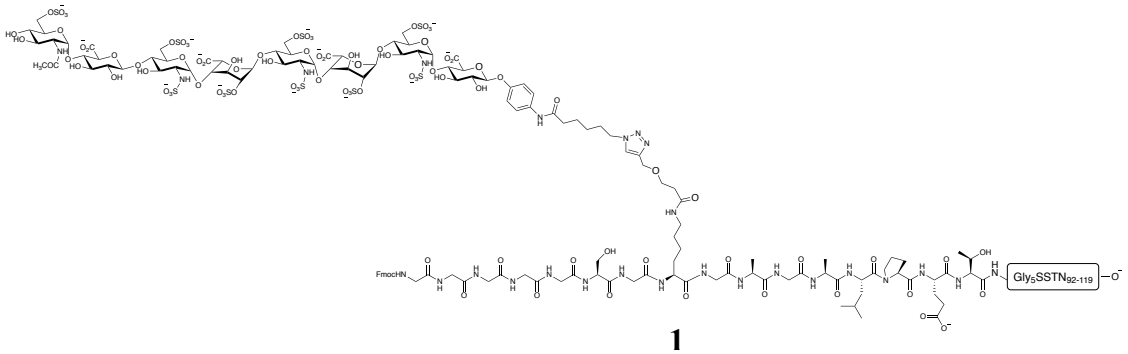


Figure 2.10 ^{1}H-^{13}C gHSQCAD of **1** (900 MHz D_2O).



1

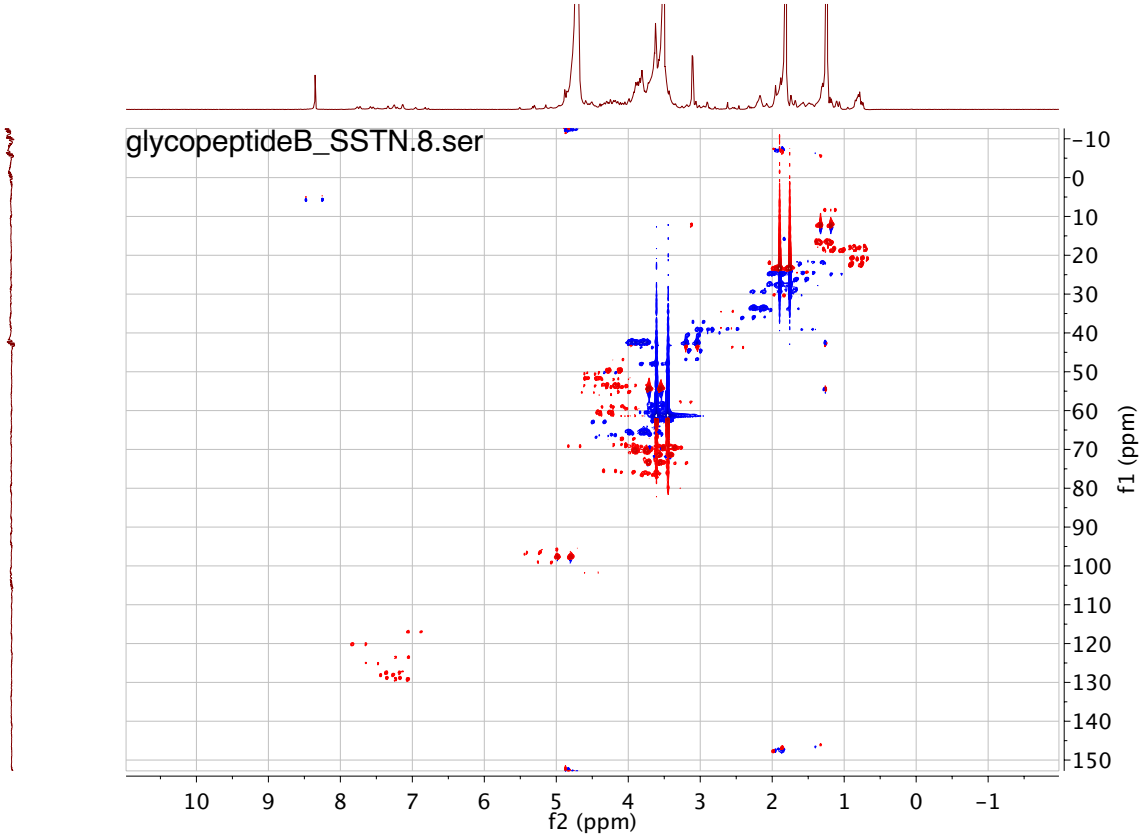
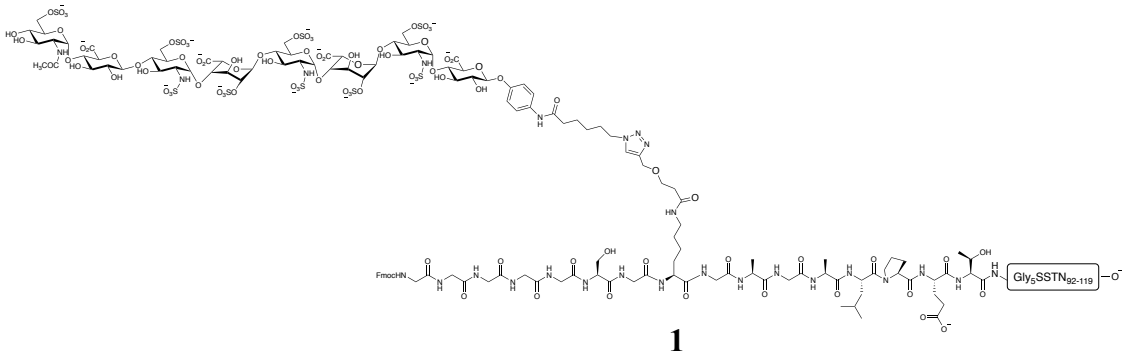


Figure 2.11 ^{1}H-^{13}C coupled gHSQCAD of **1** (900 MHz D_2O).



1

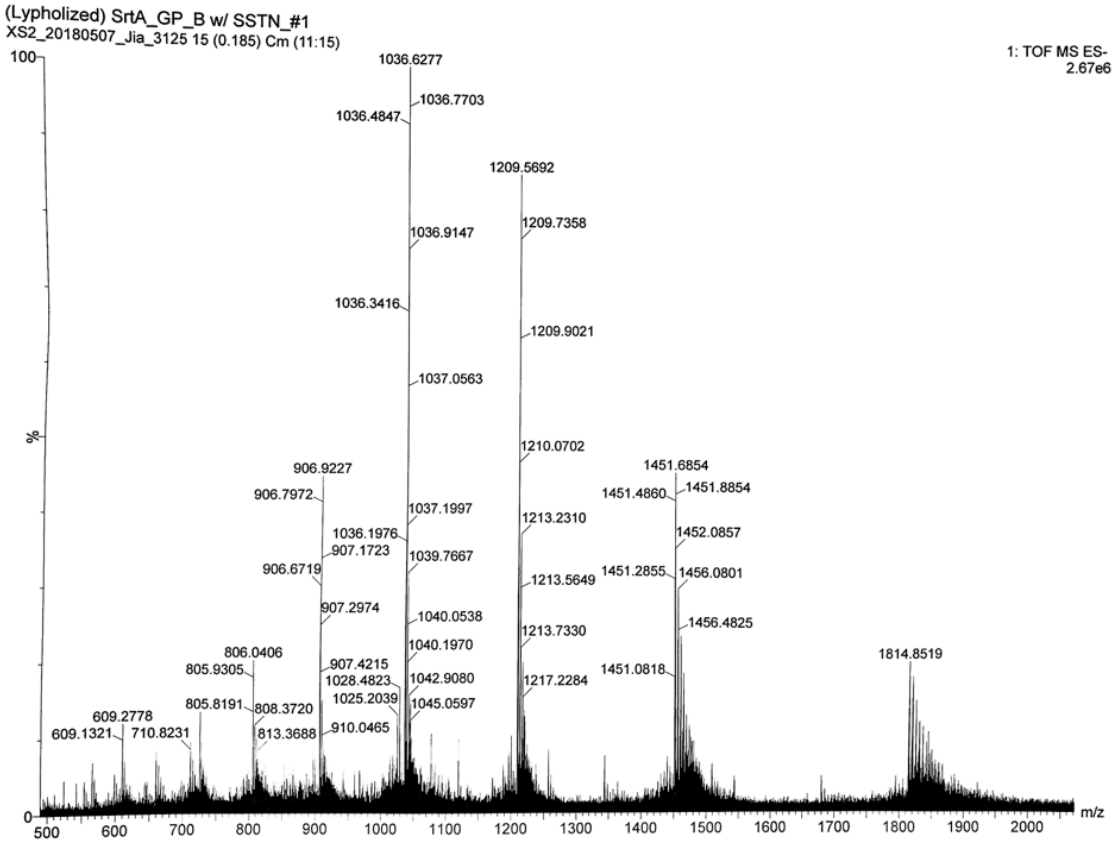
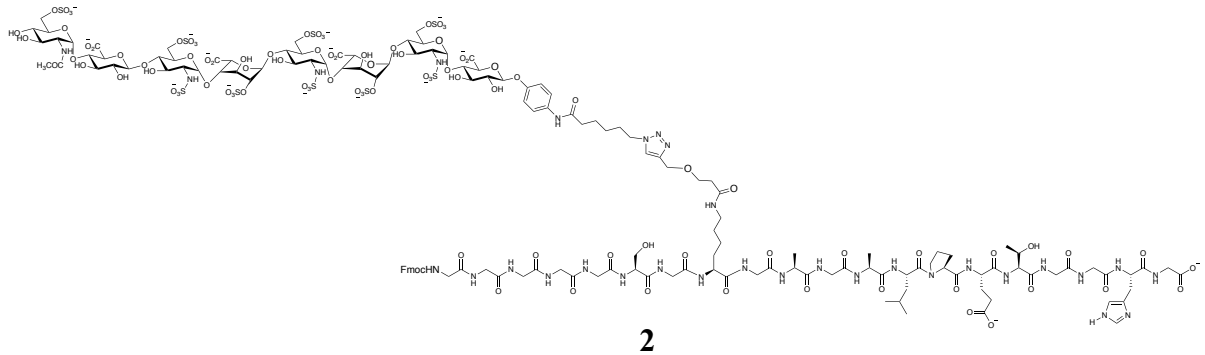
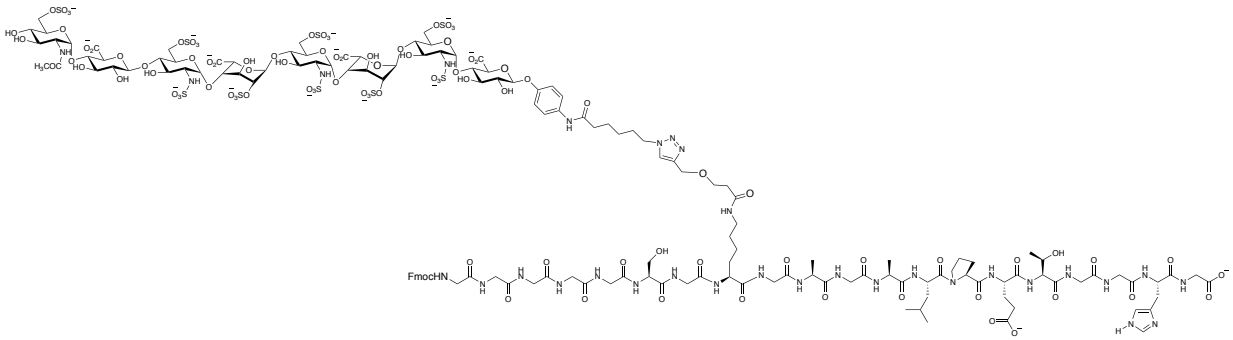


Figure 2.12 ESI-MS of **1**.



2

The purity of glycopeptide **2** was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). ^{1}H-NMR (900 MHz, D_2O), δ 8.61-8.57 (m, 2H), 7.91 – 7.80 (m, 6H), 7.67-7.63 (m, 3H), 7.62-7.55 (m, 1H), 7.46-7.42 (m, 4H), 7.39-7.35 (m, 4H), 7.34-7.28 (m, 3H), 7.27-7.19 (m, 4H), 7.17 – 6.99 (m, 4H), 5.63-5.59 (m, 1H), 5.45-5.41 (m, 1H), 4.40-4.34 (m, 2H), 4.32 – 4.12 (m, 3H), 4.05 – 3.95 (m, 10H), 3.95-3.89 (m, 24H), 3.87-3.81 (m, 8H), 3.80 – 3.67 (m, 14H), 3.66-3.62 (m, 5H), 3.60-3.56 (m, 3H), 3.46-3.42 (m, 1H), 3.41 – 3.23 (m, 5H), 3.22-3.18 (m, 2H), 3.08-3.04 (m, 4H), 2.53 – 2.40 (m, 4H), 2.38-2.34 (m, 4H), 2.29-2.25 (m, 5H), 2.13-2.09 (m, 2H), 2.02-1.98 (m, 6H), 1.91-1.87 (m, 2H), 1.86-1.82 (m, 3H), 1.77 – 1.65 (m, 4H), 1.64 – 1.51 (m, 9H), 1.44 – 1.30 (m, 17H), 1.30 – 1.12 (m, 14H), 0.91 (m, 12H). ^{13}C-NMR (225 MHz, D_2O), δ 174.6, 174.2, 174.0, 172.7, 171.8, 171.3, 171.1, 170.8, 153.8, 143.8, 143.6, 140.8, 133.5, 131.8, 128.1, 128.0, 127.4, 125.0, 124.7, 124.5, 123.4, 120.1, 117.5, 116.9, 1018, 100.2, 97.6, 97.0, 77.3, 77.0, 76.3, 76.2, 75.8, 73.4, 72.4, 70.5, 70.0, 69.3, 69.0, 68.8, 67.0, 66.8, 66.2, 66.1, 65.8, 62.8, 61.0, 60.4, 59.0, 58.0, 57.9, 57.5, 55.8, 53.9, 53.6, 53.5, 52.2, 50.2, 50.1, 49.9, 49.3, 47.7, 46.7, 43.6, 43.2, 42.5, 42.3, 42.3, 39.0, 38.9, 36.0, 35.8, 32.3, 30.3, 29.2, 28.8, 27.7, 26.9, 26.6, 24.8, 24.6, 24.4, 24.3, 22.4, 22.3, 21.9, 20.5, 18.7, 16.6, 16.3. ESI-MS: $C_{146}H_{207}N_{31}O_{99}S_9$ $[M+11H]^{4-}$ calcd: 1066.4901, obsd: 1066.4874 (2.51 ppm).



2

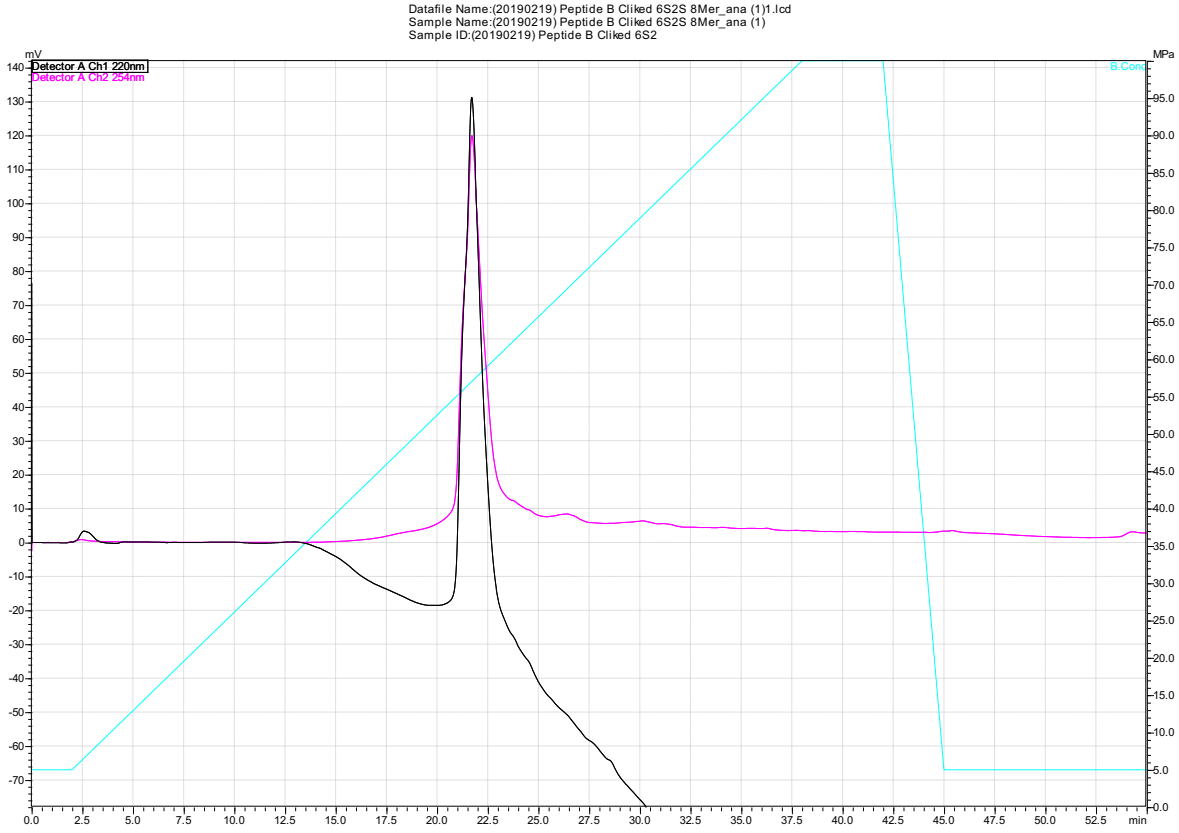
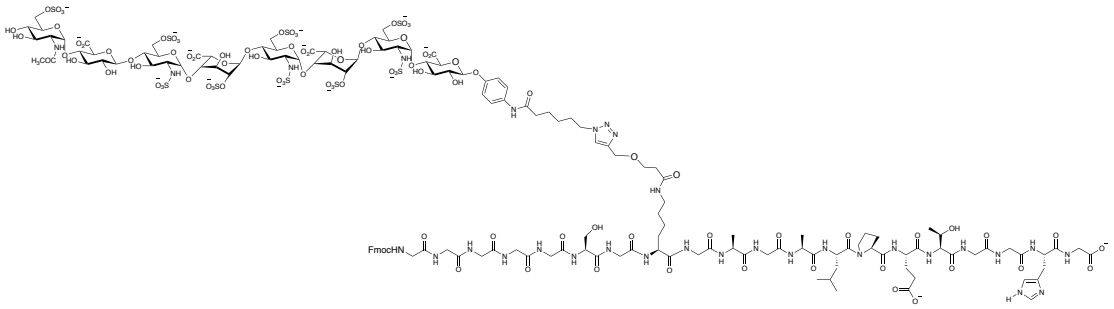


Figure 2.13 HPLC chromatogram of **2**.



2

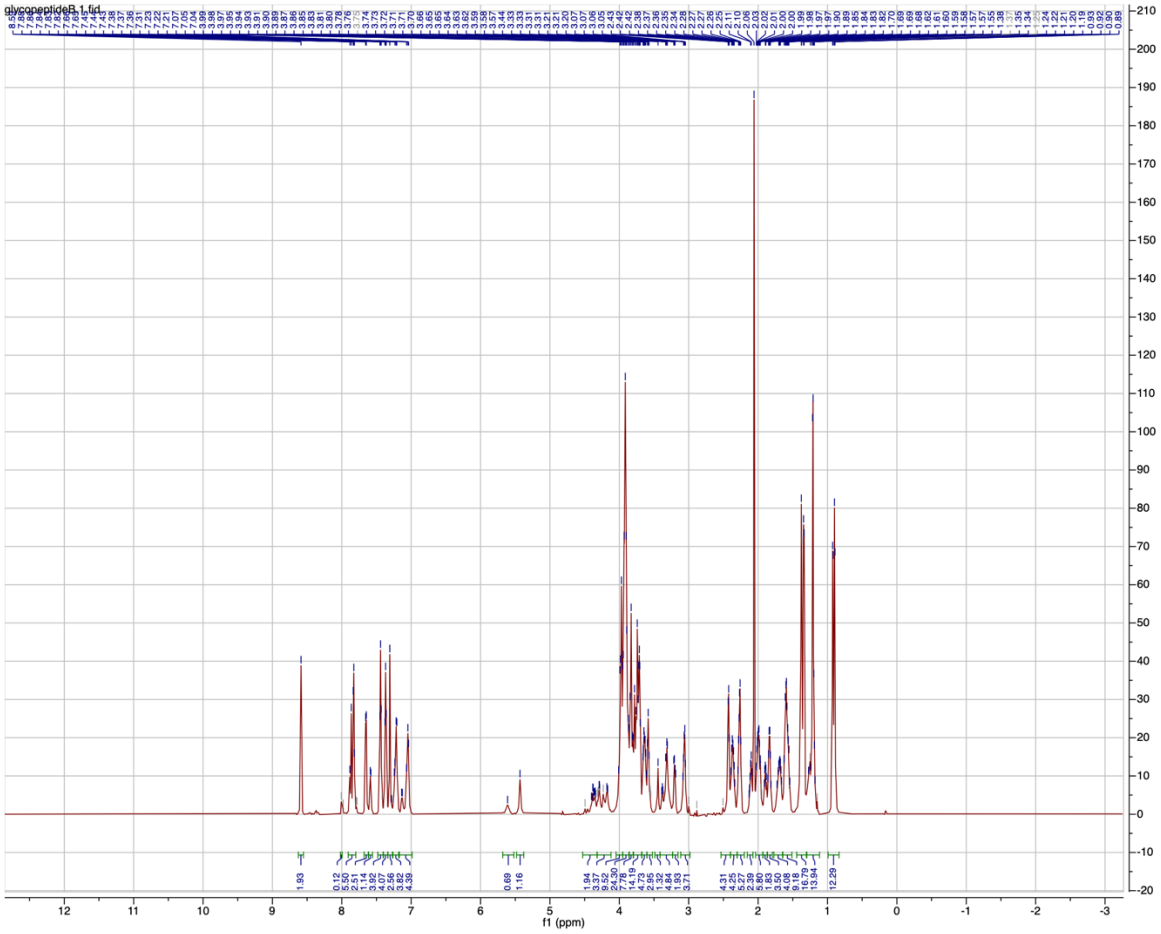
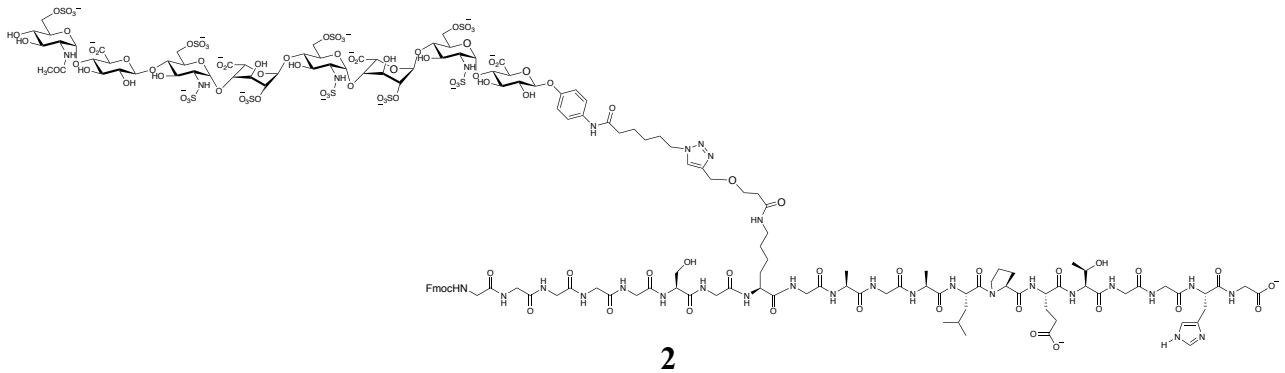


Figure 2.14 ^{1}H-NMR of **2** (900 MHz D_2O).



2

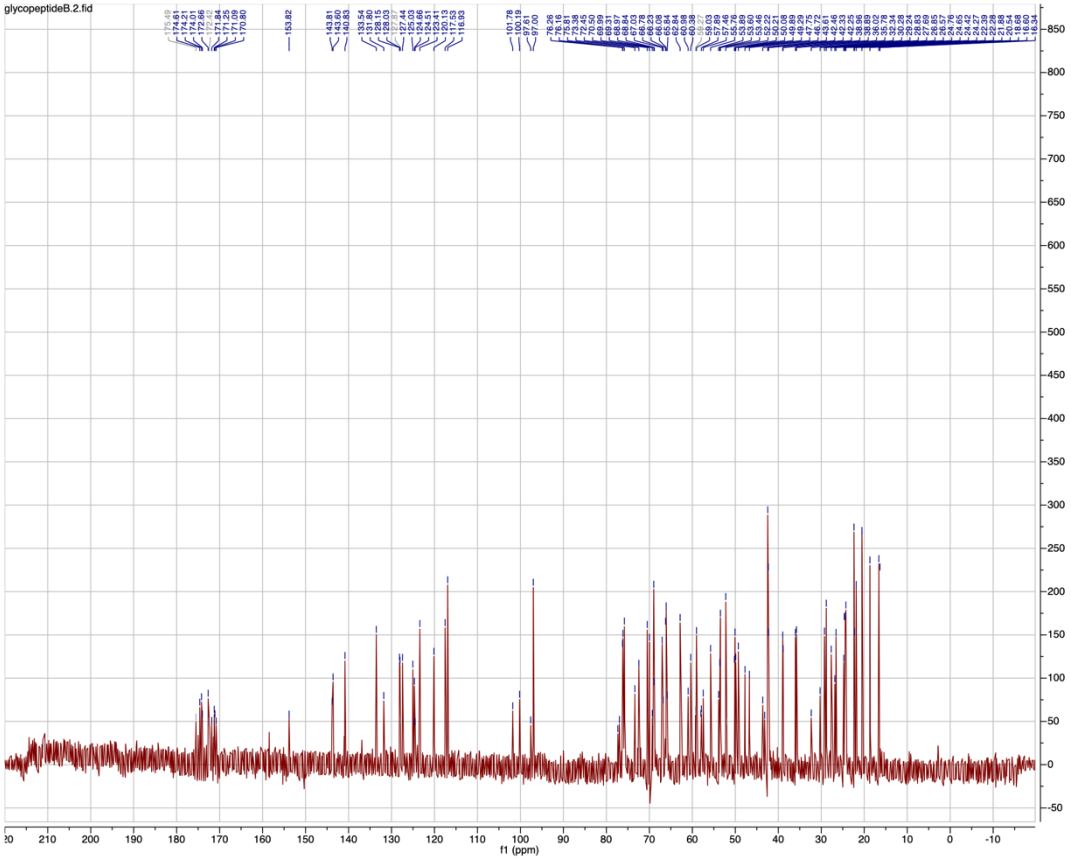
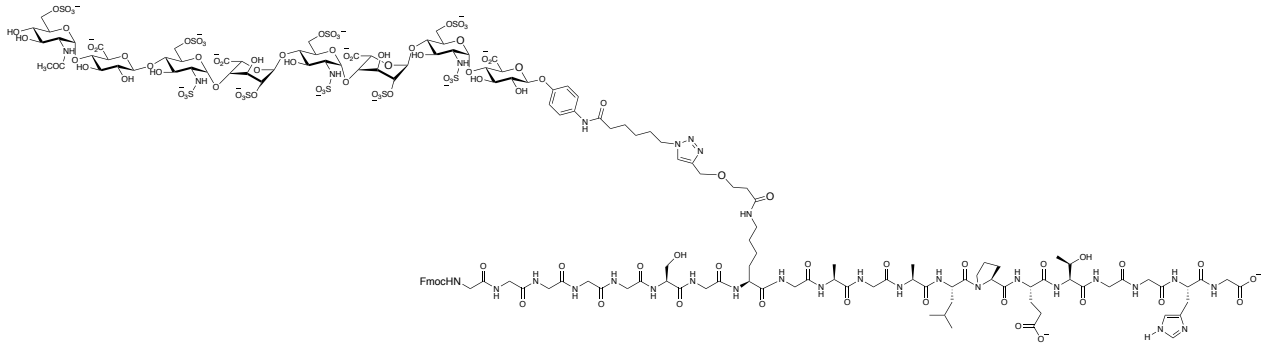


Figure 2.15 ^{13}C-NMR of **2** (225 MHz D_2O).



2

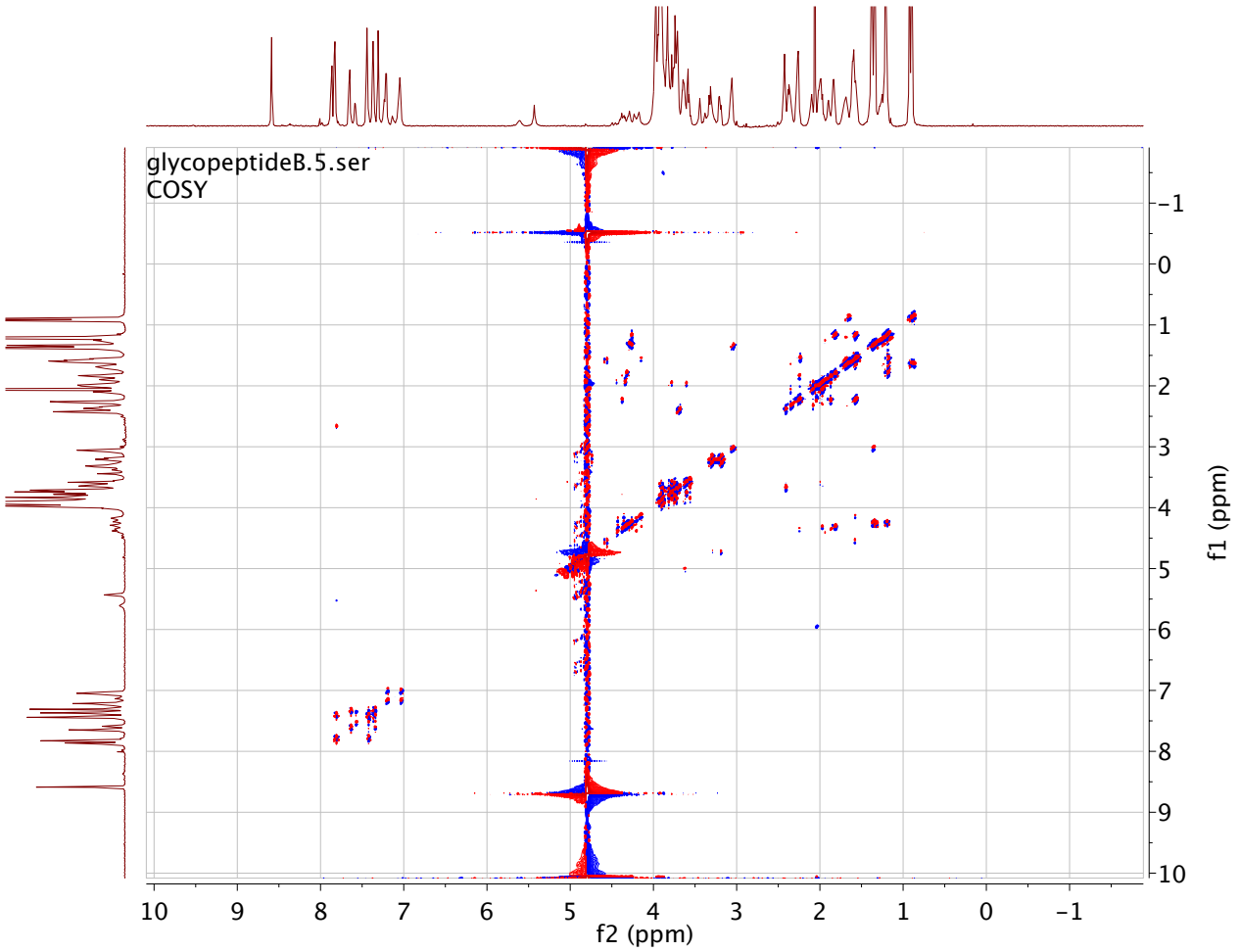
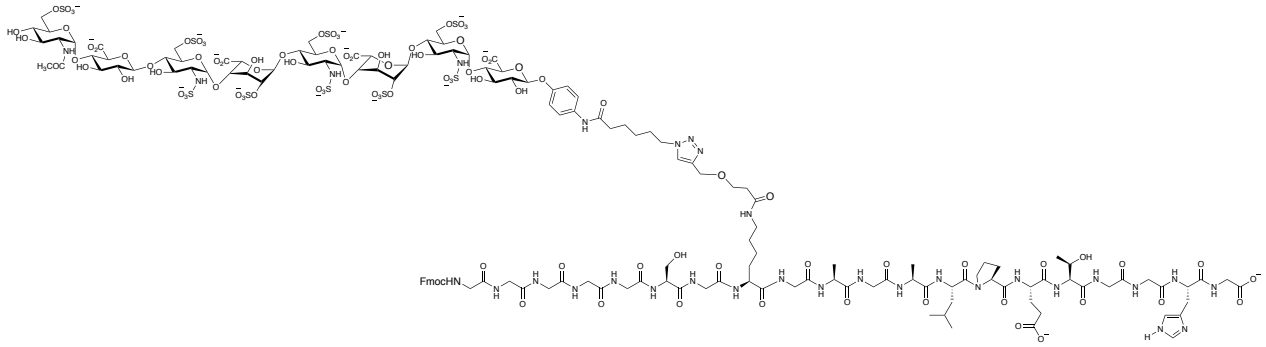


Figure 2.16 ^{1}H-^{1}H gCOSY of **2** (900 MHz D_2O).



2

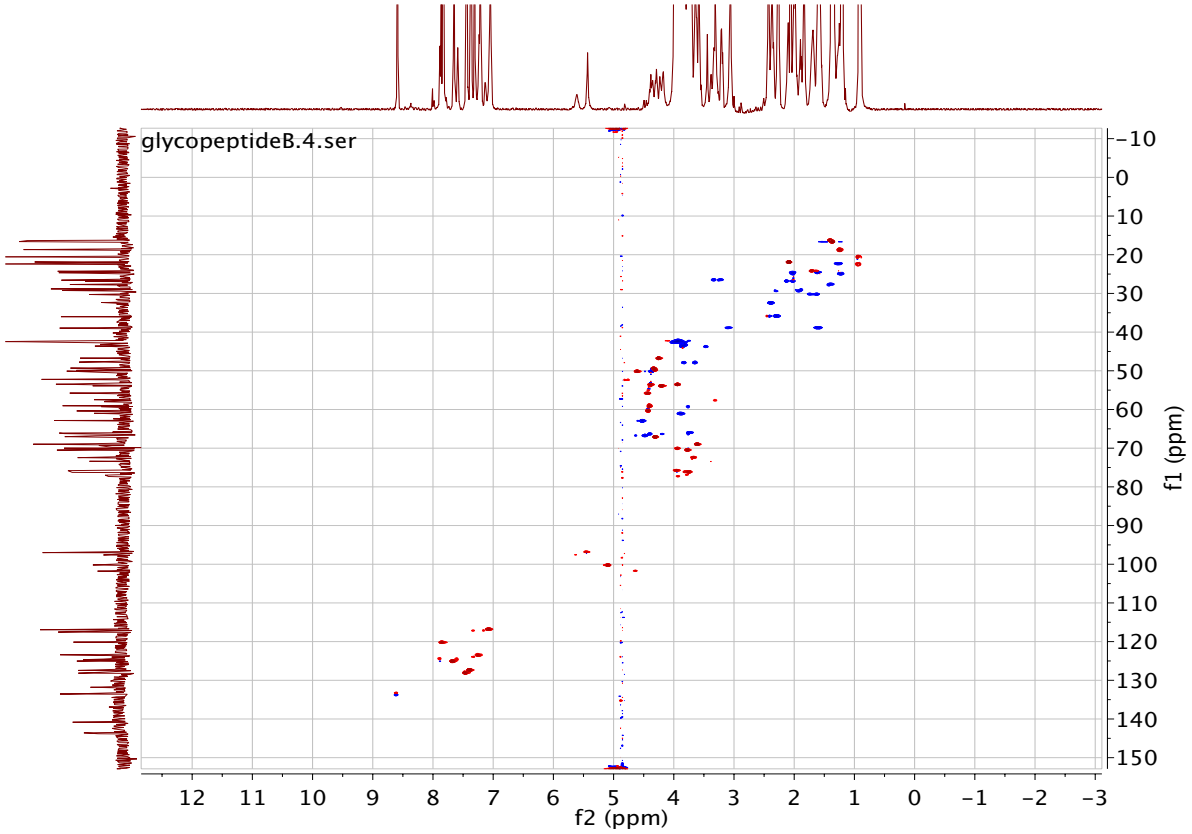
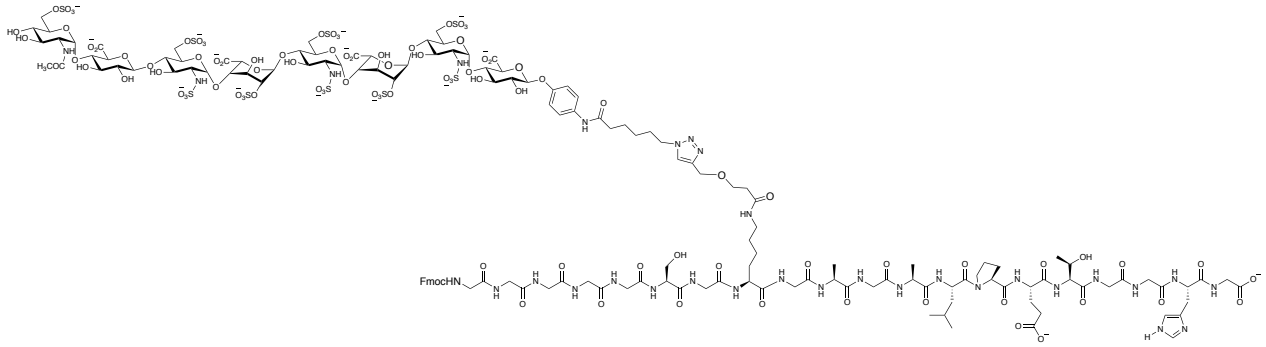


Figure 2.17 ^{1}H-^{13}C gHSQCAD of **2** (900 MHz D_2O).



2

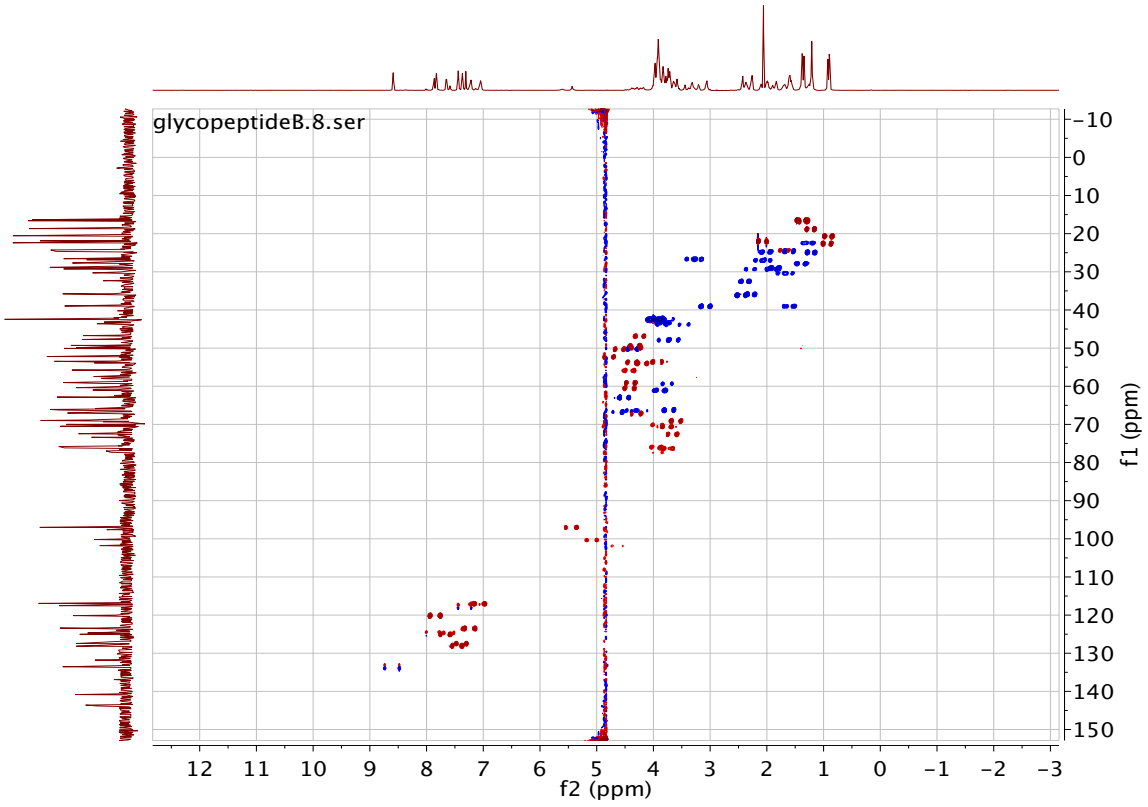
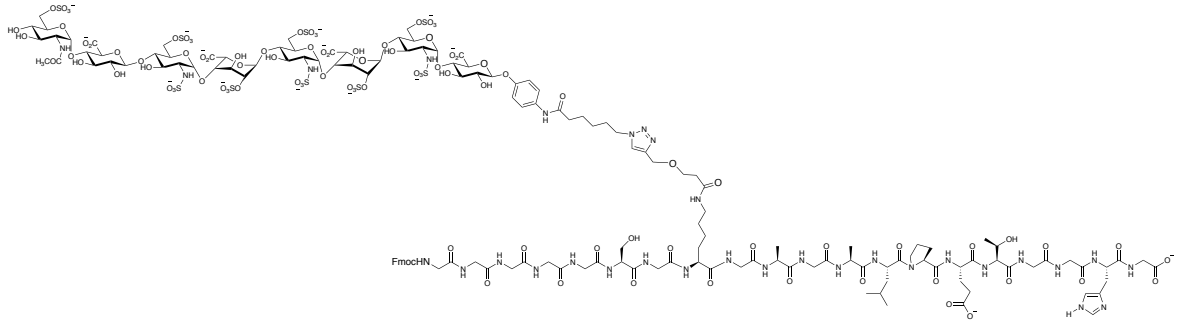


Figure 2.18 ^{1}H-^{13}C coupled gHSQCAD of **2** (900 MHz D_2O).



2

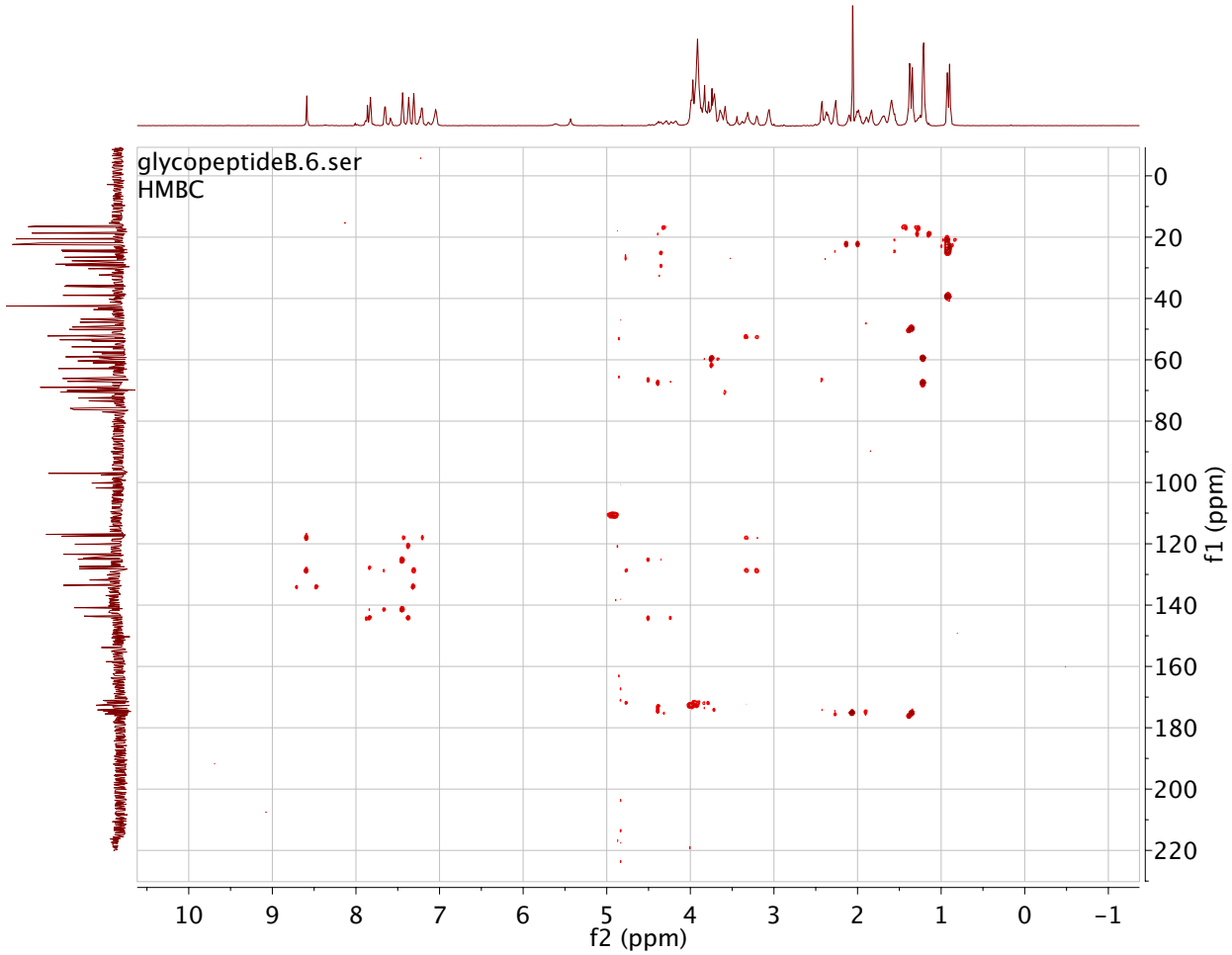
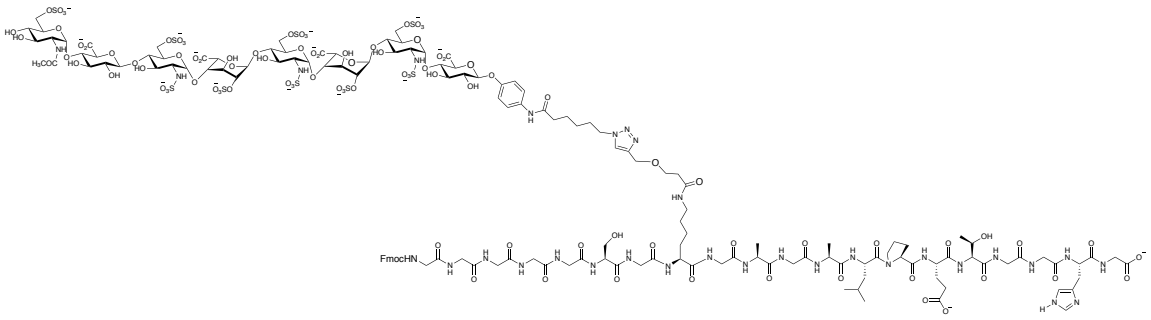


Figure 2.19 ^{1}H-^{13}C gHMBC of **2** (900 MHz D_2O).



2

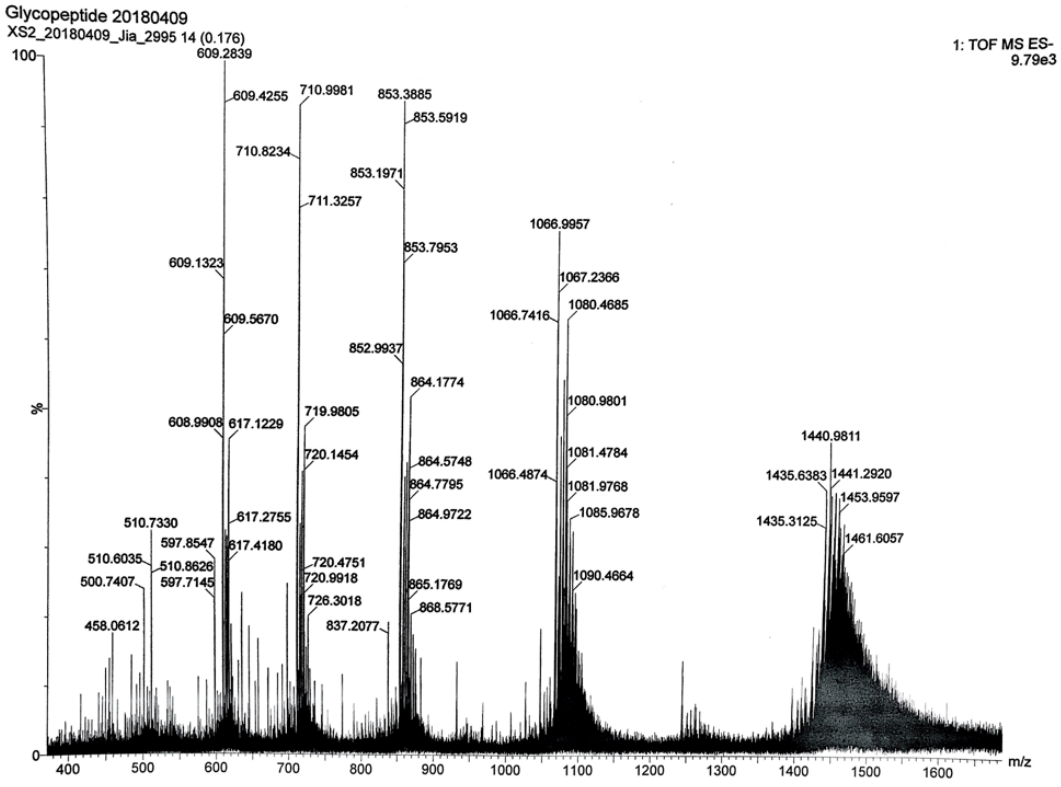
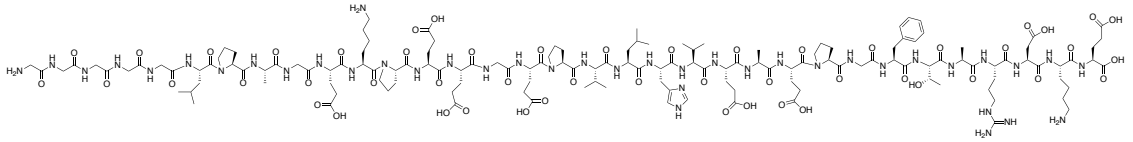
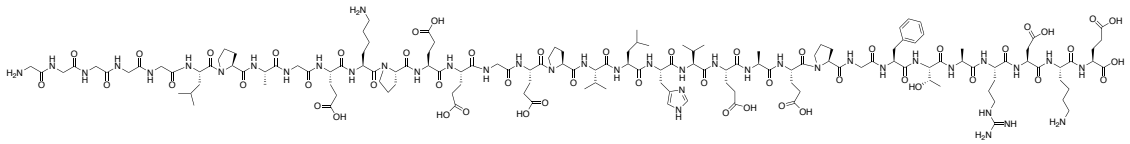


Figure 2.20 ESI-MS of **2**.



3

The purity of peptide **3** was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). ESI-MS: $C_{143}H_{226}N_{40}O_{51}$ $[M+4H]^{4+}$ calcd: 829.9075, obsd: 829.9052 (2.77 ppm).



3

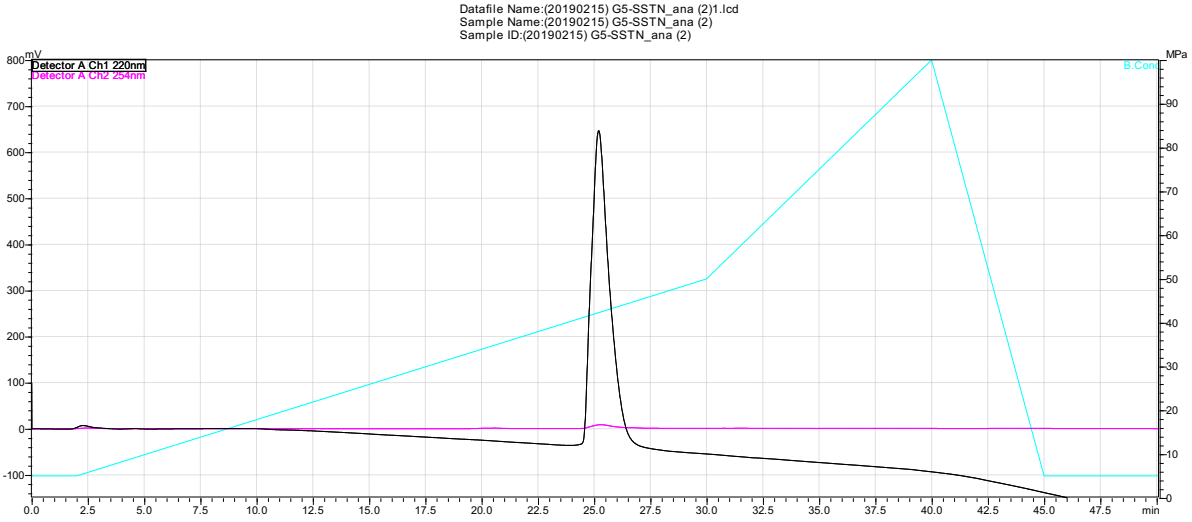
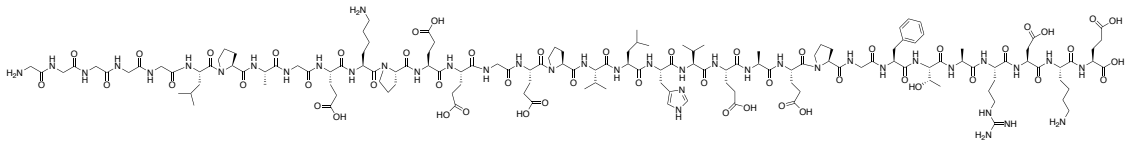


Figure 2.21 HPLC chromatogram of **3**.



3

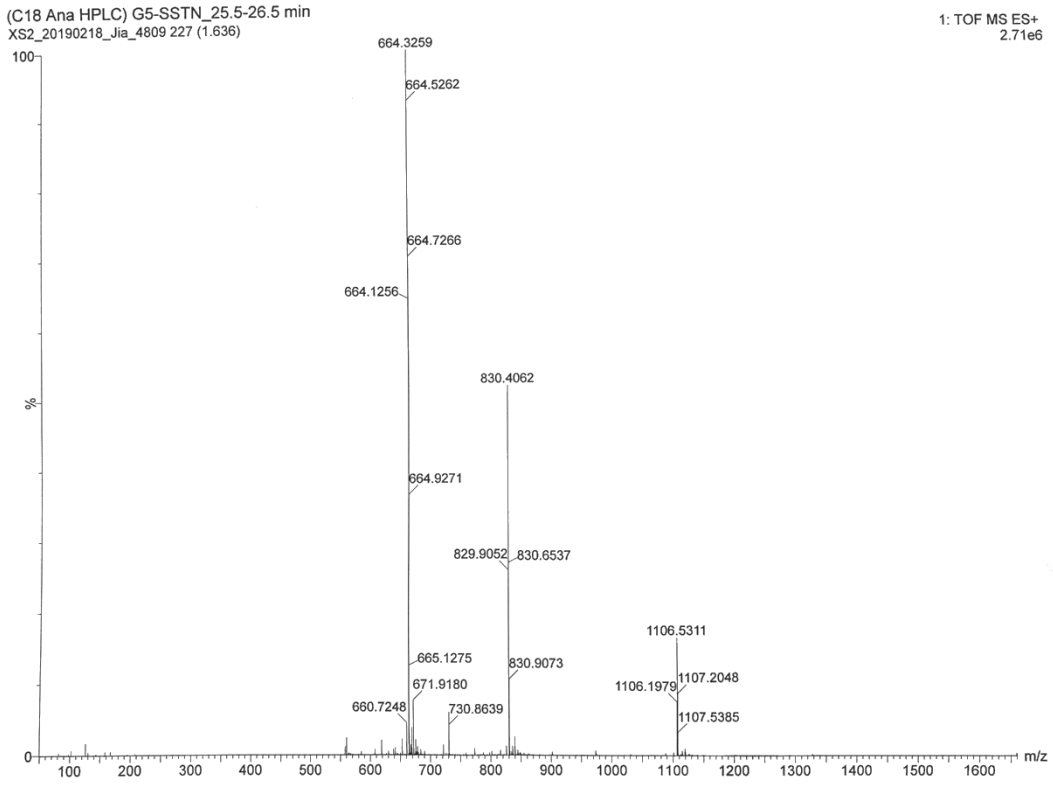
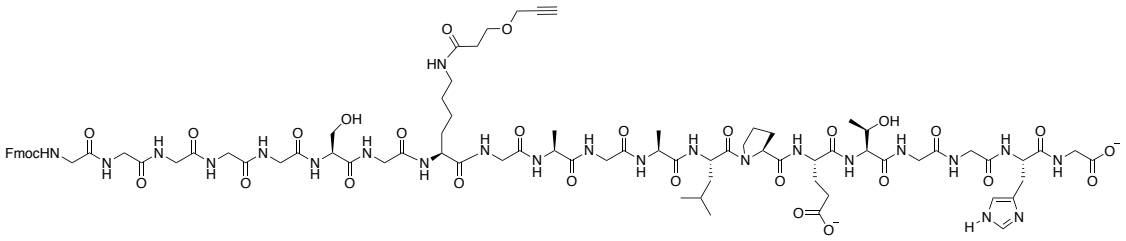
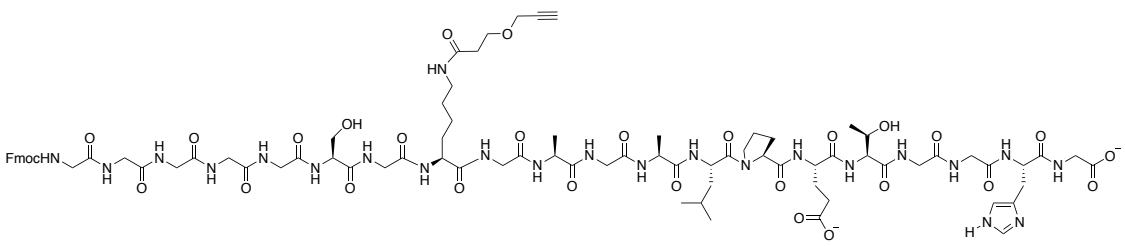


Figure 2.22 ESI-MS of **3**.



5

The purity of peptide **5** was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). ESI-MS: $C_{84}H_{118}N_{23}O_{29}$ $[M+4H]^{4+}$ calcd: 1912.8461, obsd: 1912.8388 (3.82ppm).



5

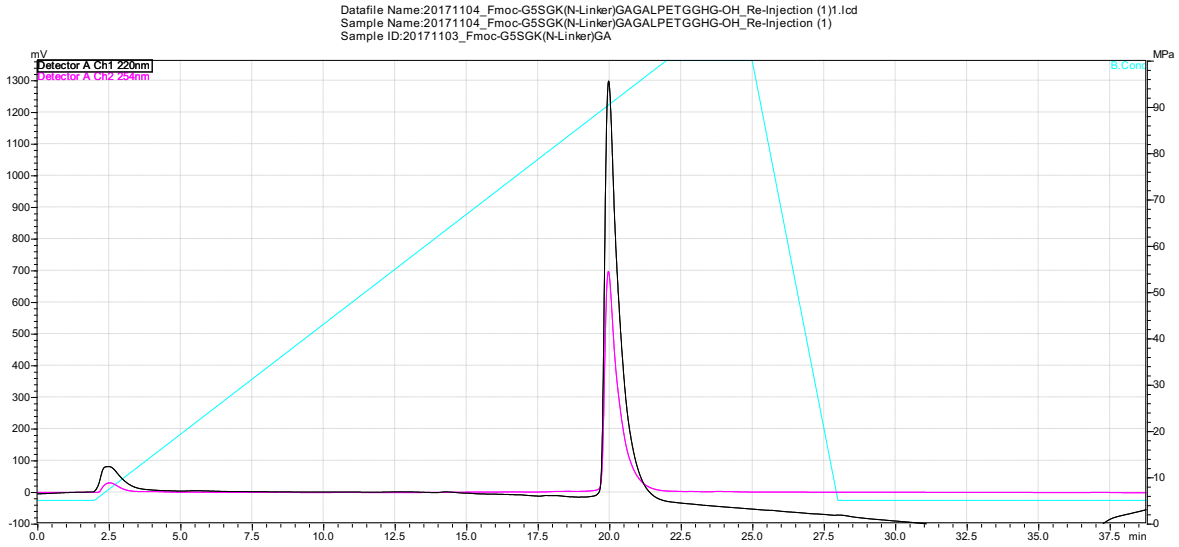
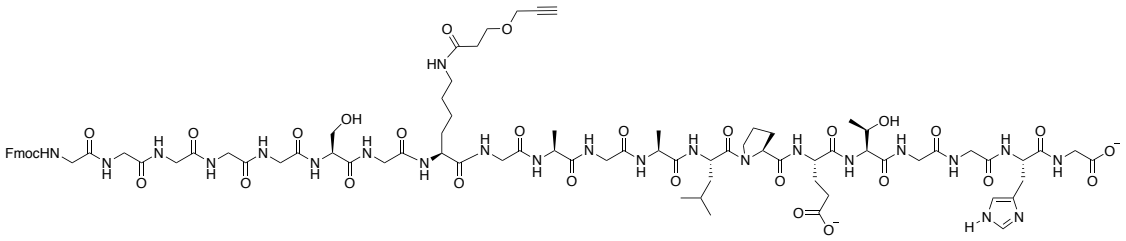


Figure 2.23 HPLC chromatogram of **5**.



5

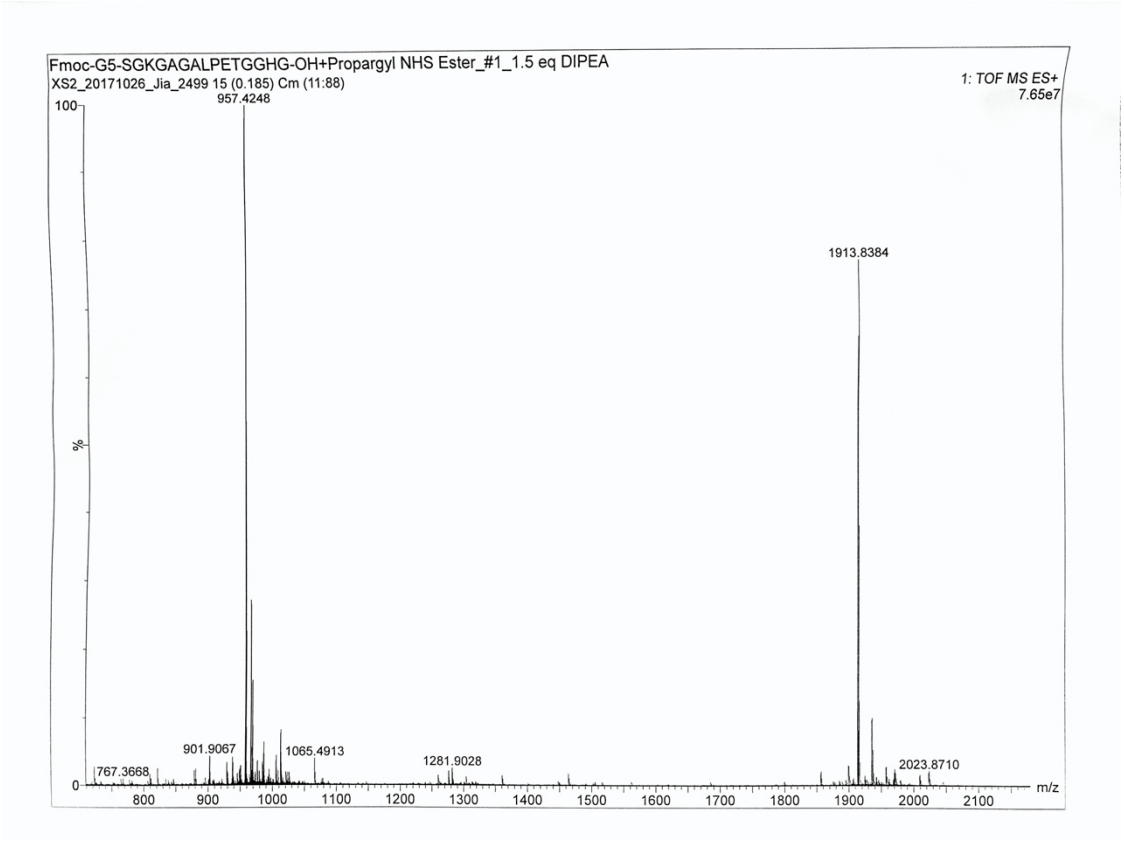


Figure 2.24 ESI-MS of **5**.

REFERENCES

REFERENCES

1. Sugahara, K.; Kitagawa, H., Heparin and Heparan Sulfate Biosynthesis. *IUBMB Life* **2002,** *54*, 163-175.

2. Sarrazin, S.; Lamanna, W. C.; Esko, J. D., Heparan Sulfate Proteoglycans. *Cold Spring Harb. Perspect. Biol.* **2011,** *3*, 1-33.

3. Peysselon, F.; Richard-Blum, S., Heparin–Protein Interactions: From Affinity and Kinetics to Biological Roles. Application to an Interaction Network Regulating Angiogenesis. *Matrix Biol.* **2014,** *35*, 73-81.

4. Ori, A.; Wilkinson, M. C.; Fernig, D. G., A Systems Biology Approach for the Investigation of the Heparin/Heparan Sulfate Interactome. *J. Biol. Chem.* **2011,** *286*, 19892-19904.

5. Harburger, D. S.; Calderwood, D. A., Integrin Signalling at a Glance. *J. Cell. Sci.* **2009,** *122*, 159-163.

6. Kirkpatrick, C. A.; Knox, S. M.; Staatz, W. D.; Fox, B.; Lercher, D. M.; Selleck, S. B., The Function of a Drosophila Glypican Does Not Depend Entirely on Heparan Sulfate Modification. *Dev. Biol.* **2006,** *300*, 570-582.

7. Capurro, M. I.; Xu, P.; Shi, W.; Li, F.; Jia, A.; Filmus, J., Glypican-3 Inhibits Hedgehog Signaling during Development by Competing with Patched for Hedgehog Binding. *Dev. Cell.* **2008,** *14*, 700-711.

8. 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*, 10185-10189.

9. 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*, 9051-9058.

10. Yang, W.; Eken, Y.; Zhang, J.; Cole, L.; 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*, 6393-6404.

11. Beauvais, D. M.; Ell, B. J.; McWhorter, A. R.; Rapraeger, A. C., Syndecan-1 Regulates Alphavbeta3 and Alphavbeta5 Integrin Activation during Angiogenesis and is Blocked by Synstatin, A Novel Peptide Inhibitor. *J. Exp. Med.* **2009,** *206*, 691-705.

12. Xu, Y.; Cai, C.; Chandarajoti, K.; Hsieh, P. H.; Li, L.; Pham, T. Q.; Sparkenbaugh, E. M.; Sheng, J.; Key, N. S.; Pawlinski, R.; Harris, E. N.; Linhardt, R. J.; Liu, J., Homogeneous Low-

Molecular-Weight Heparins with Reversible Anticoagulant Activity. *Nat Chem Biol* **2014,** *10*, 248-250.

13. Adams, D. J.; Atkins, D.; Cooper, A. I.; Furzeland, S.; Trewin, A.; Young, L., Vesicles from Peptidic Side-Chain Polymers Synthesized by Atom Transfer Radical Polymerization. *Biomacromolecules* **2008,** *9*, 2997-3003.

14. Huang, Y. C.; Guan, C. J.; Tan, X. L.; Chen, C. C.; Guo, Q. X.; Li, Y. M., Accelerated Fmoc Solid-Phase Synthesis of Peptides with Aggregation-Disrupting Backbones. *Org. Biomol. Chem.* **2015,** *13*, 1500-1506.

15. Hendrickx, A. P.; Budzik, J. M.; Oh, S. Y.; Schneewind, O., Architects at the Bacterial Surface - Sortases and the Assembly of Pili with Isopeptide Bonds. *Nat. Rev. Microbiol.* **2011,** *9*, 166-176.

16. David Row, R.; Roark, T. J.; Philip, M. C.; Perkins, L. L.; Antos, J. M., Enhancing the Efficiency of Sortase-Mediated Ligations through Nickel-Peptide Complex Formation. *Chem. Commun.* **2015,** *51*, 12548-12551.

17. Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A., Sortase-Mediated Protein Ligation a New Method for Protein Engineering. *J. Am. Chem. Soc.* **2004,** *126*, 2670-2671.

18. Beauvais, D. M.; Rapraeger, A. C., Syndecan-1 Couples the Insulin-Like Growth Factor-1 Receptor to Inside-Out Integrin Activation. *J. Cell. Sci.* **2010,** *123*, 3796-3807.

19. Faye, C.; Moreau, C.; Chautard, E.; Jetne, R.; Fukai, N.; Ruggiero, F.; Humphries, M. J.; Olsen, B. R.; Ricard-Blum, S., Molecular Interplay between Endostatin, Integrins, and Heparan Sulfate. *J Biol Chem* **2009,** *284*, 22029-22040.

20. Saladin, A.; Rey, J.; Thevenet, P.; Zacharias, M.; Moroy, G.; Tuffery, P., PEP-Sitefinder: A Tool for the Blind Identification of Peptide Binding Sites on Protein Surfaces. *Nucleic Acids Res.* **2014,** *42*, 221-226.

21. Mottarella, S. E.; Beglov, D.; Beglova, N.; Nugent, M. A.; Kozakov, D.; Vajda, S., Docking Server for the Identification of Heparin Binding Sites on Proteins. *J. Chem. Inf. Model.* **2014,** *54*, 2068-2078.

22. Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S., The Cluspro Web Server for Protein-Protein Docking. *Nat. Protoc.* **2017,** *12*, 255-278.

23. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25*, 1605-1612.

24. Woods Group. (2005-2021) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. (http://glycam.org)

25. Xu, D.; Zhang, Y., Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **2012**, *80*, 1715.

26. Schrodinger Release 2020-4: Maestro, Schrodinger, LLC, New York, NY, 2020.

Chapter 3 Chemoenzymatic Synthesis of Glycopeptides bearing Galactose-Xylose Disaccharide from the Proteoglycan Linkage Region

3.1 Introduction

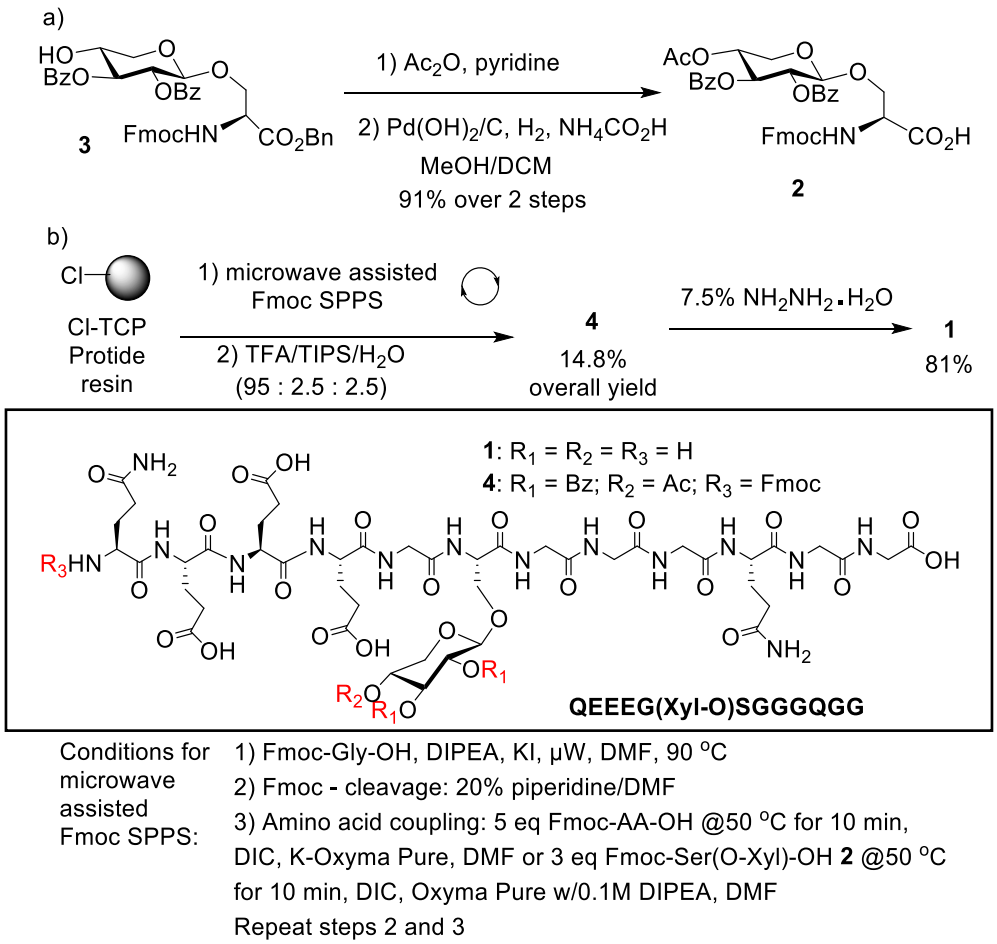
Proteoglycans (PGs) are ubiquitous in the mammalian system with important roles in many biological events including cancer development, inflammation, and immune modulation.[1-5] PGs are composed of a core protein linked with one or more glycosaminoglycan (GAG) chains through a tetrasaccharide linkage of glucuronic acid (GlcA)-β-1,3-galactose (Gal)-β-1,3-Gal-β-1,4-xylose (Xyl) covalently conjugated with serine residues of serine-glycine dipeptides.[6] As naturally existing PGs are highly heterogeneous due to complex enzymatic post-translational modification of the GAG chain, synthesis becomes important to provide the much needed, well-defined PGs to expedite their characterization in biological studies. Recently, chemical syntheses of several PG glycopeptides have been reported, which have opened up the possibilities of accessing homogeneous glycopeptides.[7, 8] However, the overall synthesis is tedious due to the need for multistep chemical manipulations. We have become interested in developing a chemoenzymatic strategy to access these glycopeptides by taking the first step to investigate the utility of human β-1,4-galactosyltransferase 7 (β4GalT7) in synthesis of Gal-Xyl bearing glycopeptides.

β4GalT7 can transfer a Gal unit from the uridine diphosphate (UDP)-Gal donor to the 4-OH of a Xyl acceptor.[9-11] Xylosides bearing hydrophobic aglycons have been shown to be competent acceptors for β4GalT7.[12] This knowledge has led to the fascinating utility of xylosides as a tool to prime cellular synthesis of glycosaminoglycans and modulate cellular functions. In addition, various xyloside analogs have been synthesized to probe the catalytic sites of β4GalT7.[13-17] However, to the best of our knowledge, β4GalT7 has not been explored for glycopeptide synthesis. Herein, for the first time we report that human β4GalT7 enzyme can be utilized to catalyze the formation of native glycopeptides bearing Gal-Xyl disaccharide on a milligram scale,

enhancing the understanding of substrate selectivities of β4GalT7 and expediting the synthesis toward structurally well-defined PGs.

3.2 Results and Discussions

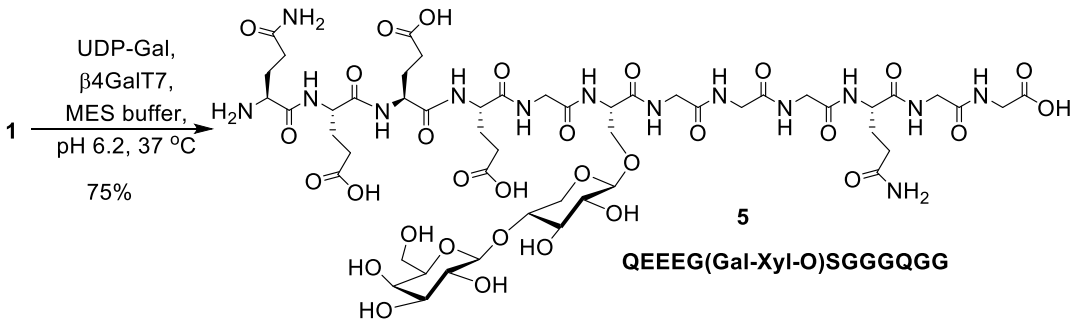
To establish the feasibility of β4GalT7 promoted glycopeptide synthesis, I first synthesized the glycopeptide **1** QEEEG(Xyl-O)SGGGQGG bearing a xylose as a potential acceptor corresponding to bikunin amino acid residues 5-14.[18] The key building block Fmoc-Ser(O-Xyl)-OH **2** was prepared from xylosyl serine **3**[19, 20] through protecting group manipulations with a 91% overall yield for the two steps (**Scheme 3.1a**). With Xyl-*O*-Ser carboxylic acid **2** in hand, automated solid phase peptide synthesis (SPPS) was carried out following Fmoc-based peptide chemistry on a chlorotrityl (Cl-TCP) ProTide resin under microwave heating at 50 ºC (**Scheme 3.1b**). The protected glycopeptide **4** was obtained in 14.8% overall yield. Following cleavage from the resin, the ester protective groups on the xylose and the *N*-terminus Fmoc moiety were removed giving the xylosylated bikunin glycopeptide **1**.



Scheme 3.1 a) Synthesis of Fmoc-Xyl-serine **2;** b) SPPS synthesis of xylosylated bikunin glycopeptide (aa: 5-14) **1**.

With the glycopeptide acceptor **1** prepared, we moved to express the polyhistidine-tagged human β4GalT7 (EC 2.4.1.133, Appendix **Figure 3.3**),[15] which was cloned into a pET plasmid and expressed in *E. coli* BL21 cells. The protein was purified by a Ni Sepharose column (Appendix **Figure 3.4**) with an expression yield of 5 mg/L. A solution of bikunin glycopeptide **1** and UDP-Gal was incubated with β4GalT7 at 37 ºC overnight (**Scheme 3.2**). High performance liquid chromatography analysis of the product mixture showed that the acceptor **1** was completely consumed. The desired Gal-Xyl disaccharide bearing glycopeptide **5** was obtained in 75% yield at milligram scales following purification by size exclusion chromatography. The product structure

was validated by nuclear magnetic resonance (NMR) and mass spectrometry (MS). Heteronuclear NMR analysis showed a coupling constant of $^{1}J_{C1,\,H1}$ = 161.6 Hz from anomeric position of Gal unit, which confirmed the newly formed β-glycosyl linkage between the Gal and Xyl units.[21]



Scheme 3.2 β4GalT7-catalyzed galactosylation of glycopeptide **1** to Gal-Xyl bearing glycopeptide **5**.

To test the scope of the galactosylation reaction catalyzed by hβ4GalT7, xylosylated glycopeptides **6** - **11** from several other naturally existing PGs were prepared via SPPS (**Figure 3.1**). These substrates include sequences from bikunin as well as members of the syndecan family PGs, representing common PGs from nature including glycopeptides with multiple Xyl moieties (glycopeptides **8** – **11**). These glycopeptides contain aromatic, hydrophobic and also hydrophilic amino acid residues adjacent to the glycosylation sites, which enhanced the structural diversity of the acceptors for hβ4GalT7. To prepare the glycopeptides, I first followed the same SPPS protocol used to make glycopeptide **1**, starting from the chlorotrityl ProTide resin. However, several glycopeptides were obtained in low overall yields (<10%) (Appendix **Table 3.3**). As the chlorotrityl resin can be unstable under heating,[22] we tested an alternative of using the more heat stable Cl-MPA ProTide resins. Together with a lowered reaction temperature from 50 °C to 30 °C for amino acid coupling, yields of the glycopeptides were significantly improved (Appendix **Table 3.3**).

β4GalT7-catalyzed galactosylation reactions were carried out on glycopeptides **6-11** to examine the scope of this transferase. Inspiringly, all enzymatic reactions successfully produced the desired products. Glycopeptides **8-11** bearing multiple Xyl units could be successfully galactosylated in all Xyl sites when 2 equiv of UDP-Gal donor per Xyl was added to the reaction mixture (**Table 3.1**). This suggests with an excess of UDP-Gal donor, β4GalT7 can drive the reaction to completion including on substrates with multiple glycosylation sites in close proximity to each other.

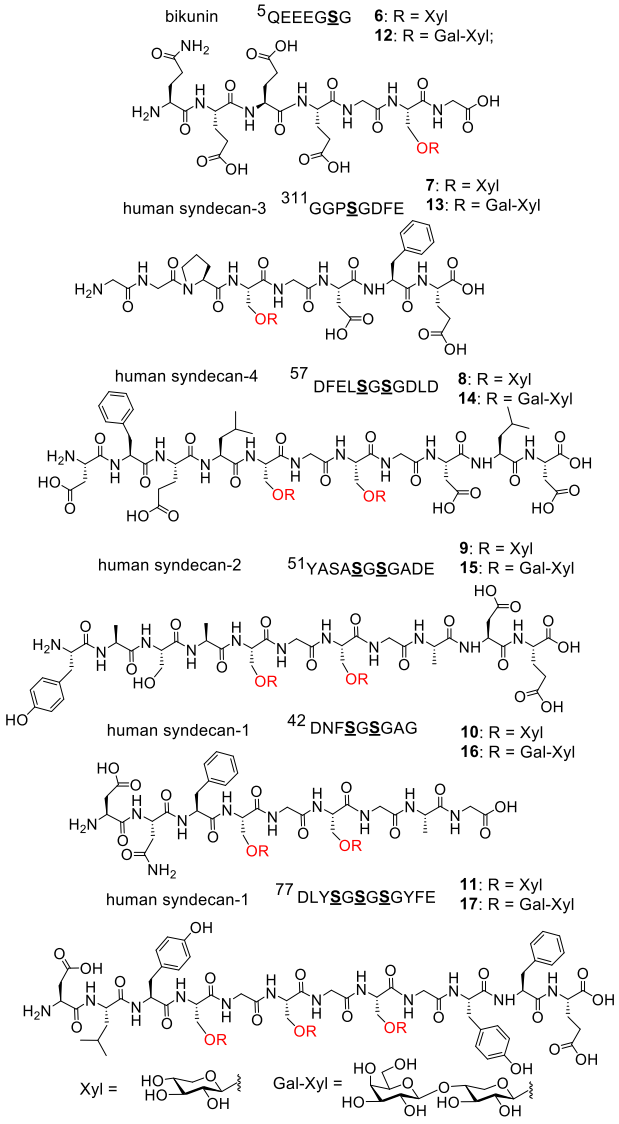


Figure 3.1 Structures of glycopeptides **6-11** with the serine glycosylation sites underlined.

Acceptor	Product	Yield (%)
6	**12**	82
7	**13**	91
8	**14**	81
9	**15**	81
10	**16**	77
11	**17**	78

Table 3.1 Yield summary of β4GalT7-catalyzed galactosylation.

How β4GalT7 interacts with the native glycopeptide substrates is not yet well understood. To gain deeper insights, we performed kinetics analysis of the enzyme on selected substrates using a modified phosphatase-coupled transferase assay.[23] The K_m value of hβ4GalT7 for UDP-Gal was calculated to be 0.04 mM (Appendix **Figure 3.7**). For glycopeptides **1** and **7** containing a single Xyl, the K_m values were about 0.1 mM. Glycopeptides **8** and **9** have two Xyl per chain, which have higher K_m values, approximating a weaker binding by the enzyme (**Table 3.2, Appendix Figures 3.8-3.11**).

Substrate	K_m (mM)	V_{max} (pmol/min/μg)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}mM^{-1}$)
1	0.07 ± 0.01	158	10	144
7	0.10 ± 0.01	460	28	281
8	0.39 ± 0.09	70	4	11
9	0.28 ± 0.06	159	9	34

Table 3.2 Summary of kinetic results from glycopeptide substrates.

For substrates with two Xyl units, we next investigated whether there were site preferences by the enzyme when the reaction was performed with sub-stoichiometric quantities of the donor UDP-Gal. Glycopeptide **9** was subjected to hβ4GalT7-catalyzed galactosylation in the presence of 1 equiv of UDP-Gal. The glycopeptides bearing only one Gal-Xyl disaccharide were observed with electrospray ionization (ESI)-MS. To determine the site of galactosylation, analysis of the glycopeptides was performed by tandem MS fragmentation of the glycopeptides. Successes in this analysis critically depended on retaining the glycan during peptide fragmentation in MS^2, which was challenging due to the lability of the glycosidic linkage with the peptide backbone. Through a collaboration with Dr. Lingjun Li (University of Wisconsin)'s laboratory, after exploring multiple fragmentation methods, the electron-transfer/higher-energy collision dissociation (EThcD) hybrid fragmentation technique, an integrated dissociation method combining electron-transfer dissociation (ETD) and higher-energy collision dissociation (HCD), was found suitable.[24] Following fragmentation of the peptide backbone in MS^2, fragment ions corresponding to glycopeptide fragments with the Gal-Xyl disaccharide at either Ser5 or Ser7 site were identified. The cumulative total ion count values of the respective peaks exhibited a 1:3 ratio of these two regio-isomers (Appendix **Table 3.4**), suggesting a preference for Ser7 galactosylation by β4GalT7.

To better understand the site preference, computational studies were performed by docking the glycopeptide **1** into the crystal structure of the complex of D211N mutant of β4GalT7 with the donor and the acceptor (PDB: 4M4K).[25] Earlier studies showed that D211 is a key catalytic residue. D211N mutation enabled a catalytically stalled ternary complex to form. The docking structure obtained showed that the glycopeptides with 4-OH of xylose pointing towards the center of the active site and being oriented by N211 explaining the preference for glycosylation at the 4-OH (**Figure 3.2a**). For glycopeptide **9** with two Xyl reaction sites, the Xyl at Ser7 preferentially forms

hydrogen bonds with Asn211/Asp212 in the active site and orients itself for the galactosylation (**Figure 3.2b**). The energy difference between Ser7 and Ser5 in the reactive site was calculated to be ≥ 0.3 kcal/mol, providing a potential explanation for higher reactivity of the Xyl unit on Ser 7 over that on Ser5 for β4GalT7 promoted galactosylation.

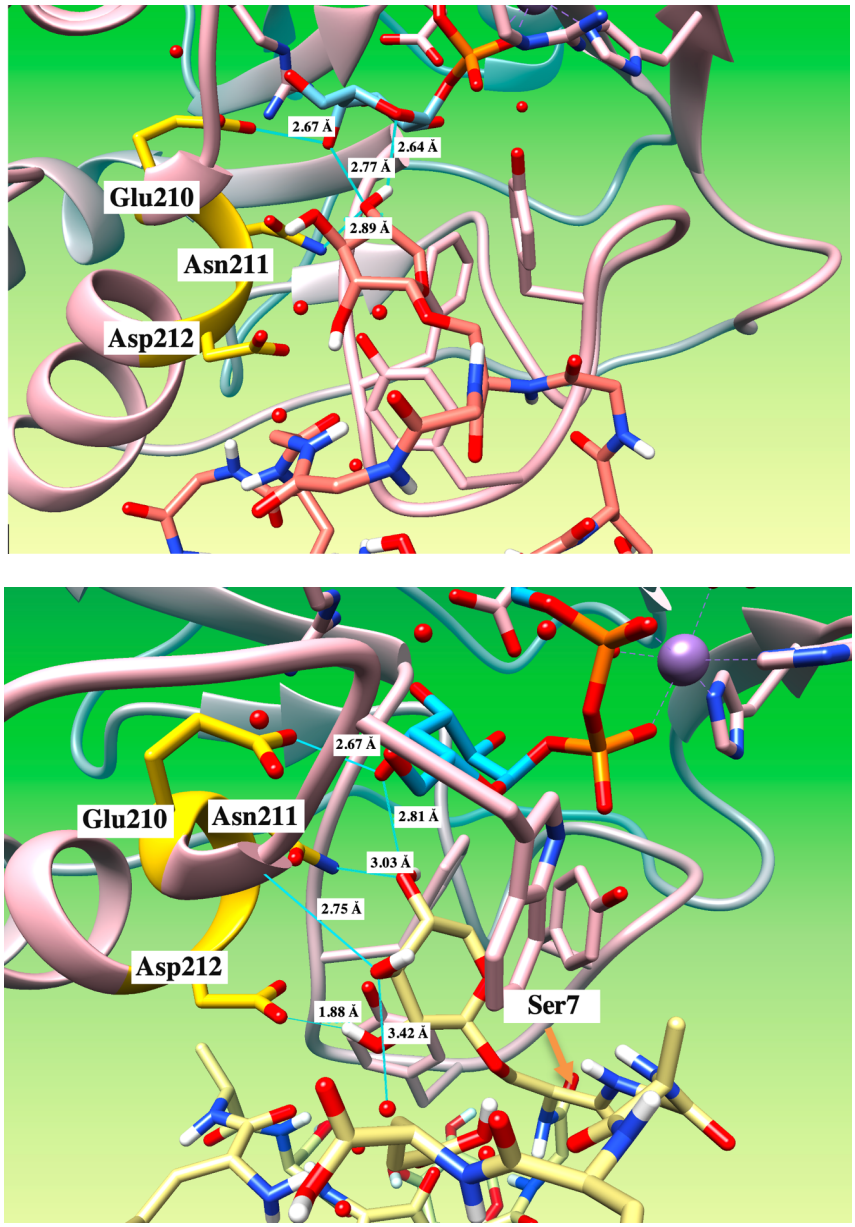


Figure 3.2 a) Docking structure of QEEEG(Xyl-O)SGGGQGG **1** with D211N mutant of β4GalT7 (PDB: 4M4K). (Catalytic residues Glu210/Asn211/Asp212 are highlighted in the protein backbone; Xylose unit is centered and colored in orange red; Galactose unit is colored in light blue; Heteroatoms are colored differently as H in white, O in red and N in deep blue; Hydrogen

Figure 3.2 (cont'd) bonds potentially involved in the catalytic process are labeled with corresponding inter-atomic distance. b) Docking structure of YASA(Xyl-O)SG(Xyl-O)SGADE **9** with β4GalT7 suggests a preference toward Ser7 site by the enzyme (Xylose unit on Ser7 site is centered and colored in khaki).

3.3 Conclusion

In conclusion, human β4GalT7 (EC 2.4.1.133) has been found to be able to transfer the Gal unit to a xylosylated glycopeptide acceptor. Diverse native glycopeptides bearing Gal-Xyl disaccharides have been prepared *via* β4GalT7 catalysis at milligram scale in good yields for the first time. Glycopeptides with multiple Xyl units can be effectively galactosylated as well. The high efficiency, broad substrate scope, and operational simplicity of β4GalT7 render it a useful tool toward the synthesis of homogeneous PGs.

3.4 Experimental Section

3.4.1 Materials

β4GalT7-expressing BL21 cells were obtained from Prof. Ulf Ellervik (Lund University, Sweden). Gibco LB broth, LB agar and Coomassie Brilliant Blue G-250 were purchased from Thermo Fischer Scientific (Waltham, MA). Nickel columns and Nickel resins were purchased from Bio-rad (Hercules, CA). SDS-PAGE gels, 10x Tris/Glycine/SDS electrophoresis buffer, prestained protein ladder, sample loading buffer, and Coomassie Blue R-250 were purchased from Bio-rad (Hercules, CA). Tris-HCl buffer was purchased from MilliporeSigma (St. Louis, MO). UDP-galactose was purchased from Complex Carbohydrate Research Center (Athens, GA). Amino acid building blocks were purchased from Chem-Impex International, Inc (Wood Dale, IL). Glycosyltransferase Activity Kit was purchased from R&D Systems. All other chemical reagents were purchased from commercial sources and used without additional purifications unless otherwise noted.

3.4.2 General Information

High-performance liquid chromatography was carried out with LC-8A Solvent Pumps, DGU-14A Degasser, SPD-10A UV-Vis Detector, SCL-10A System Controller (Shimadzu Corporation, JP) and Vydac 218TP 10 μm C18 Preparative HPLC column (HICHROM 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)). The dual-wavelength UV detector was set at 220 nm and 254 nm for monitoring the absorbance from amide and Fmoc-, correspondingly. 3D structure of glycopeptide compounds was prepared with Maestro software. Docking simulations were acquired with AutoDock Vina and UCSF Chimera (UCSF, CA). Enzymatic activity was quantified by absorbance at 620 nm using a SpectraMax M3 96-well plate reader (Molecular Devices, CA). Enzymatic glycosylation sites were analyzed by Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Thermo Fischer Scientific, MA). LC-MS2 data was processed with Byonic™ search engine (Protein Metrics, CA). NMR data were obtained with DirectDrive2 500 MHz and Varian 900 MHz NMR spectrometer (Agilent, CA) at ambient temperature.

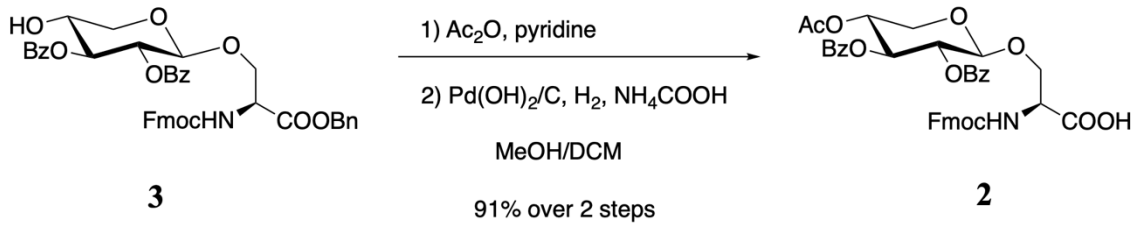
3.4.3 β4GalT7 Expression, Purification and Characterization

hβ4GalT7-expressing BL21 competent cell were cultured onto kanamycin/chloramphenicol containing petri dish, which was incubated at 37 °C overnight. One colony of BL21 cells was picked and inoculated into 10 mL starter culture containing kanamycin at concentration of 30 mg/L. The cell culture was incubated at 37 °C overnight. The starter culture was then transferred into autoclaved 1L culture medium (with 30 mg/L kanamycin) and incubated at 37 °C with shaking at 250 rpm. After roughly 3-4 hours, the OD_{600} reached 0.5. IPTG (0.56 mM,

MilliporeSigma, MO) was added to induce protein expression at 32 °C for 20 hours. Cells were centrifuged at 4 °C, 5,000 g for 10 min. Cell pellet was lysed using Cellytic in 20 mM Tris buffer, pH 7.6, 50 U/mL benzonase, 0.2 mg/mL lysozyme and 1mM PMSF (MilliporeSigma, MO) for 20 min at ambient temperature. Clarified lysate was purified by nickel column (Cytiva, MA) (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 standard Bradford assay.

3.4.4 Glycosyl Amino Acid Building Block Preparation

The glycosyl amino acid building block **3** was prepared following the previously reported conditions.[26, 27]



3 **2**

N-Fluorenylmethyloxycarbonyl-O-(2,3-di-O-benzoyl-4-O-acetyl-β-D-xylopyranosyl)-L-serine (**2**). Compound **3** (227 mg, 0.3 mmol) was dissolved into pyridine (2 mL), followed by the addition of acetic anhydride (61 μL, 0.6 mml). The reaction mixture was stirred at room temperature overnight. It was then diluted with DCM and washed against dilute HCl solution. The reaction intermediate was concentrated and dissolved into MeOH/DCM (1:1, v/v, 10 mL), followed by $Pd(OH)_2/C$ (50 mg) and $HCOONH_4$ (21.2 mg, 0.898 mmol). The mixture was stirred under H_2 at ambient temperature for 30 min and then filtered via a PTFE membrane (pore size 0.22 μm). The filtrate was concentrated under vacuum without further purification to afford compound **2 (**193.6

mg, 91%). $[\alpha]_D^{20}$ = - 3.7 ° (c = 14.18, methanol). ^{1}H-NMR (500 MHz, CD_3OD), 1.96-2.00 (s, 3 H), 3.50-3.59 (m, 1 H), 3.85-3.92 (m, 1 H), 4.02-4.11 (m, 2 H), 4.11-4.21 (m, 2 H), 4.24-4.31 (m, 1 H), 4.46-4.51 (m, 1 H), 4.83-4.87 (m, 1 H), 5.02-5.16 (m, 3 H), 5.21-5.26 (m, 1 H), 5.51-5.57 (m, 1 H), 7.22-7.30 (m, 5 H), 7.31-7.39 (m, 3 H), 7.39-7.45 (m, 2 H), 7.45-7.54 (m, 1 H), 7.55-7.62 (m, 3 H), 7.76-7.81 (m, 1 H), 7.90-7.95 (m, 3 H); ^{13}C-NMR (500 MHz, CD_3OD), 19.0, 54.1, 60.9, 66.6, 66.7, 68.0, 68.4, 70.7, 71.1, 99.9, 119.5, 124.8, 124.9, 126.7, 127.3, 127.4, 127.9, 127.9, 128.1, 128.1, 128.2, 128.9, 129.1, 129.4, 133.1, 133.2, 135.6, 141.1, 141.1, 143.6, 143.8, 156.8, 165.2, 165.4, 169.7, 170.0. ESI-MS: $C_{39}H_{36}NO_{12}$ $[M+H]^+$ calcd: 710.2232, obsd: 710.2243 (1.55 ppm).

3.4.5 General Procedure for Automated Solid-Phase Glycopeptide Substrate Synthesis

All the glycopeptides were synthesized on a Liberty Blue™ Automated Microwave Peptide Synthesizer following the standard Fmoc-based solid-phase peptide synthesis protocol. The Cl-TCP(Cl) ProTide resins were purchased from CEM Corporation. The Liberty Blue software (CEM Corporation, NC) was used to program the synthesis, including resin swelling, amino acid loading, couplings and Fmoc- removal. Commercially available *N,N*-dimethylformamide (DMF) from Fischer Chemical was supplied to the synthesis module as a reaction and washing solvent. Peptide synthesis was enabled by sequential couplings of Fmoc-amino acid (purchased from Chem-Impex, Wood Dale, IL), which was preactivated by DIC, Oxyma Pure and DIPEA, at 50 °C for 10 min, and deprotections with 20% piperidine in DMF at 60 °C for 4 min. In-between each coupling/deprotection step, resin-bound peptide was thoroughly washed with DMF. For the incorporation of the glycosyl amino acid **2**, double coupling was applied by recycling the unreacted glycosyl amino acid building block. Resin-bound peptides were cleaved off the solid support with a cocktail solution of trifluoroacetic acid (TFA), triisopropylsilane (TIPS) and water

(TFA/TIPS/H_2O, 95:2.5:2.5). The crude peptides were then purified with reverse-phase C18 preparative HPLC. Compound purity was confirmed by C18 analytical HPLC analysis.

3.4.6 General Procedure for Glycopeptide Deprotection

Partially protected glycopeptide was first dissolved in H_2O (0.85 mL). An 80% hydrazine hydrate solution (hydrazine, 51%, 0.15 mL) was then added slowly to initiate the reaction. The resulting mixture was stirred at ambient temperature overnight. The desired fully deprotected glycopeptide product was purified with a Sephadex G-10 column.

3.4.7 General Procedure for β4GalT7-Catalyzed Glycosylation

10x MES reaction buffer for β4GalT7-catalyzed glycosylation was prepared in advance following the recipe of 200 mM MES, 100 mM $MnCl_2$. The pH of the 10x reaction buffer was adjusted to 6.2 by adding concentrated NaOH solution. A solution of 1 mM glycopeptide substrate and 1.5 mM UDP-galactose (1.5 equiv per glycosylation site) was made with the reaction buffer. The addition of β4GalT7 enzyme (0.5 mol%) initiated the glycosylation. The reaction solution was kept at 37 °C overnight. The reaction progress was monitored with LC-MS. After the reaction, the enzyme was deactivated and precipitated out of the reaction mixture by adding ethanol. The mixture was centrifuged, and the supernatant was loaded onto a G-10 size exclusion column for purification.

3.4.8 General Procedure for Enzyme-Substrate Docking

3D structure of the substrate was prepared with ChemDraw 16.0 and Schrodinger Maestro software. After importing the substrate structure from ChemDraw into Maestro, it was energetically optimized via the built-in function “Minimize-All Atoms”. The optimized structure was then output as a mol2 file for the subsequent molecular dynamic docking. To initiate the docking experiments, a high-resolution enzyme crystal structure as a PDB file, along with the

substrate structure as a mol2 file, was imported into UCSF Chimera software. The enzyme-substrate molecular docking was achieved with AutoDock Vina, an integrated program in UCSF Chimera.[28, 29] For the docking set-up, the enzyme was chosen as the "Receptor" and the substrate was selected as "Ligand". The "Receptor search volume" was defined to ensure that space around the catalytic binding pocket was included for a proper docking simulation, while balancing the demand towards computation resource. Default settings of "Receptor options" and "Ligand options" were used. "Number of binding modes", "Exhaustiveness of search" and "Maximum energy difference (kcal/mol)" options were adjusted to the maximum level to ensure the quality of the simulation. The docking experiment was then executed via Opal web service. Computation results were available upon completion of the experiment.

3.4.9 Phosphatase-Coupled Enzymatic Kinetic Assay

The kinetic assay protocol follows the general assay conditions reported by R&D Systems Inc. with modifications.[23]

30 µL reaction solutions of UDP-galactose, glycopeptide acceptor and β4GalT7 enzyme were prepared in the 96-well plate. The plate was covered with a plate sealer and incubated at 37 °C for 20 min. 12 µL 10x phosphatase assay buffer, 3 µL $MnCl_2$ solution (100 mM), 3 µL MilliQ water and 2 µL coupling phosphatase 1 (20 ng/µL), were quickly added to a total volume of 50 µL. The plate was covered with a plate sealer again and incubated at 37 °C for 20 min. After the incubation, 30 µL of Malachite Green Reagent A was quickly added to each well. The solutions were gently mixed by tapping the plate. 100 µL of deionized or distilled water was added to each well. 30 µL of Malachite Green Reagent B was then added to each well. Solutions were mixed gently by tapping the plate. The plate was incubated for 5 minutes at room temperature to have consistent color development. The optical density of each well was determined using a microplate

reader set to 620 nm, and the OD was adjusted by subtracting the reading of the negative control. Product formation was calculated using the conversion factor determined from the phosphate standard curve.

3.4.10 LC/ESI-MS/MS Analysis and Data Processing

The glycopeptide sample was first desalted using a Hydrophilic-Lipophilic-Balanced (HLB) cartridge (Waters, Milford, MA). The desalted sample was dissolved in 0.1% FA and analyzed on the Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a Dionex UPLC system. A binary solvent system composed of 0.1% formic acid in H_2O (A) and 0.1% formic acid in ACN (B) was used for all analyses. Samples were loaded and separated on a 75 μm x 15 cm homemade column packed with 1.7 μm, 150 Å, BEH C18 material obtained from a Waters UPLC column (part no. 186004661). The LC gradient for intact glycopeptides was set as the following: 3%-30% A (18-33 min), 85% A (33-43 min), and 3% A (43-53 min). The mass spectrometer was operated in data dependent mode using a top-speed approach (cycle time of 3 s). HCD triggered EThcD was employed. MS1 scan was acquired from *m/z* 300–2000 (120,000 resolution, $4e^5$ AGC, 50 ms injection time) followed by EThcD MS/MS acquisition of the selected precursors in the Orbitrap (60,000 resolution, $2e^5$ AGC, 250 ms injection time) with an optimized user-defined charge-dependent reaction time (+2 50 ms; +3 25 ms; +4-5 15 ms; +6-8 10 ms) supplemented by 25% HCD activation.

All raw data files were searched against the known peptide sequence using PTM-centric search engine Byonic (version 3.3, Protein Metrics, San Carlos, CA). Searches were performed with a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.03 Da. Xylose(Pent(1)) and Xylose-Galactose(Hex(1)Pent(1)) were embedded in Byonic as the glycan database. Only these *O*-glycopeptides with PSMs with an FDR ≤ 1% and Byonic score over 150

were considered as a reliable identification. The ratio of coeluted glycopeptides with different glycoforms (regio-isomers) was calculated by manually checking their MS2 spectra and cumulatively counting the intensities of c, z ions bearing specific glycans.

APPENDICES

APPENDIX A: Supplementary Schemes, Figures and Tables

Sequence Analysis

Start codon and stop codon shown in Red
NcoI, NotI were avoided during optimization

1. The original sequence of B4GALT7-e.coli

ATGFPSRRKAAQLPWEDGRSGLLSGGLPRKCSVFHLFVACLSLGFFSLLWLQLSCSGDVARAVRGQGQETSGPPRACPPEPPPEHWEEDA
SWGPHRLAVLVPFRERFEELLVFVPHMRRFLSRKKIRHHIYVLNQVDHFRFNRAALINVGFLESSNSTDYIAMHDVDLLPLNEELDYGFP
EAGPFHVASPELHPLYHYKTYVGGILLLSKQHYRLCNGMSNRFWGWGREDDEFYRRIKGAGLQLFRPSGITTGYKTFRHLHDPAWRKRDQ
KRIAAQKQEQFKVDREGGLNTVKYHVASRTALSVGGAPCTVLNIMLDCDKTATPWCTFSTAA

The optimized (for *Escherichia coli*) sequence of B4GALT7-e.coli
A 196 T 218 C 279 G 291 | GC%: 57.93% | Length: 984

ATGTTTCCGAGTCGCCGTAAAGCAGCCCAGCTGCCTTGGGAGGATGGTCGCAGCGGTTTATTAAGCGGCGGTCTGCCGCGTAAGTGCAGT
GTTTTCCACCTGTTCGTGGCCTGTCTGAGCCTGGGTTTCTTTAGCCTGCTGTGGCTGCAGCTGAGCTGTAGCGGTGATGTGGCCCGCGCA
GTTCGTGGTCAAGGCCAGGAGACCAGTGGTCCGCCTCGTGCATGCCCTCCGGAACCGCCTCCGGAGCATTGGGAGGAAGATGCCAGTTGG
GGTCCGCATCGCCTGGCAGTGCTGGTTCCGTTCCGCGAACGCTTTGAGGAACTGCTGGTTTTCGTGCCGCACATGCGTCGCTTTCTGAGC
CGCAAAAAGATCCGCCACCACATTTACGTGCTGAACCAGGTGGACCACTTCCGCTTTAATCGTGCCGCCCTGATTAATGTGGGCTTCCTG
GAGAGCAGCAACAGCACCGATTACATCGCCATGCACGATGTGGATCTGCTGCCGCTGAATGAAGAGCTGGATTACGGCTTTCCGGAAGCC
GGCCCGTTTCATGTGGCCAGCCCGGAACTGCACCCGCTGTACCACTATAAAACCTACGTGGGCGGCATCCTGCTGCTGAGCAAGCAGCAT
TATCGCCTGTGTAACGGCATGAGCAACCGTTTCTGGGGCTGGGGCCGTGAAGATGACGAGTTCTATCGCCGCATCAAAGGTGCCGGTCTG
CAGTTATTTCGCCCGAGCGGCATCACCACCGGCTACAAGACATTTCGCCATCTGCACGATCCGGCATGGCGTAAACGTGATCAGAAACGC
ATCGCCGCCCAGAAACAGGAACAGTTCAAGGTGGATCGCGAGGGCGGCCTGAACACCGTTAAGTATCACGTGGCAAGCCGTACCGCACTG
AGTGTGGGTGGCGCACCTTGTACCGTGCTGAACATCATGCTGGACTGCGATAAAACCGCAACCCCTTGGTGTACATTTAGCTAA

Figure 3.3 β4GalT7 amino acid and gene sequence.

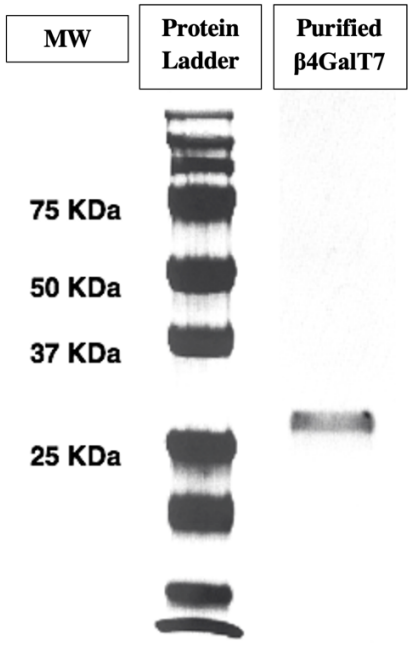


Figure 3.4 SDS-PAGE gel of purified β4GalT7.

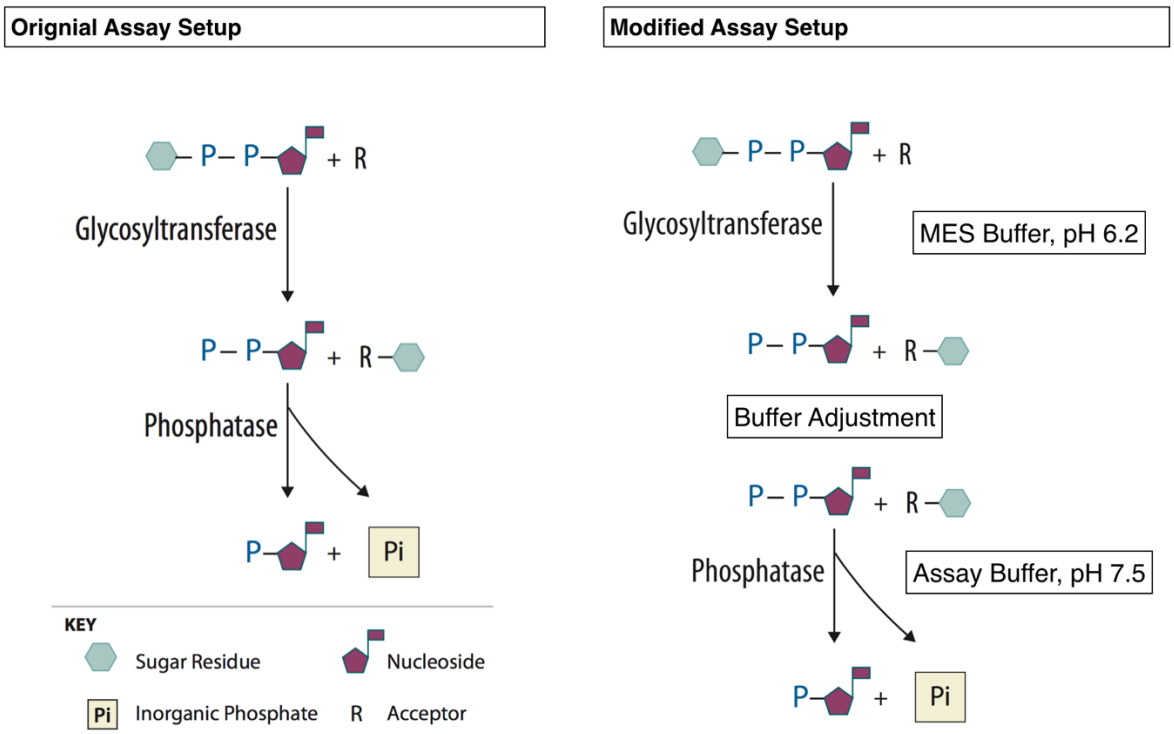


Figure 3.5 Schematic demonstrations of the original and the modified kinetic assay set-up.[3]

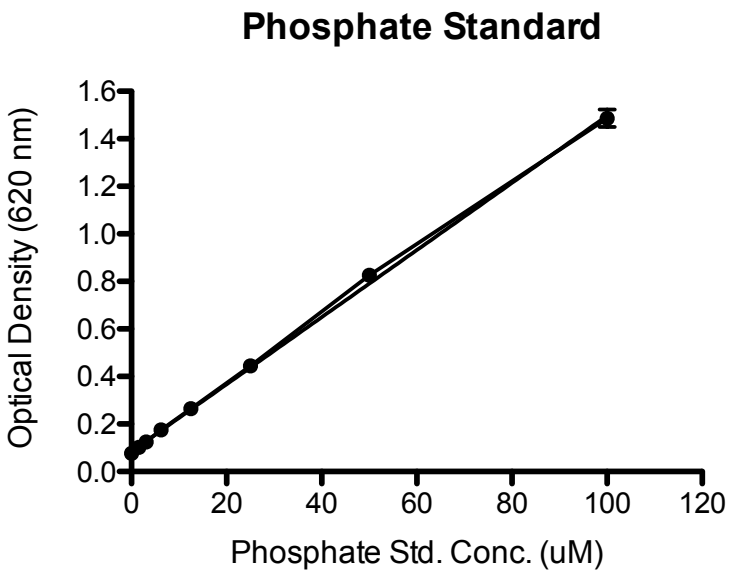


Figure 3.6 Phosphate conversion factor measurement. Conversion factor was calculated as 3541 pmol/OD (Plot is displayed as mean ± S.D. of two replicates, phosphate standard concentration = 50 µL).

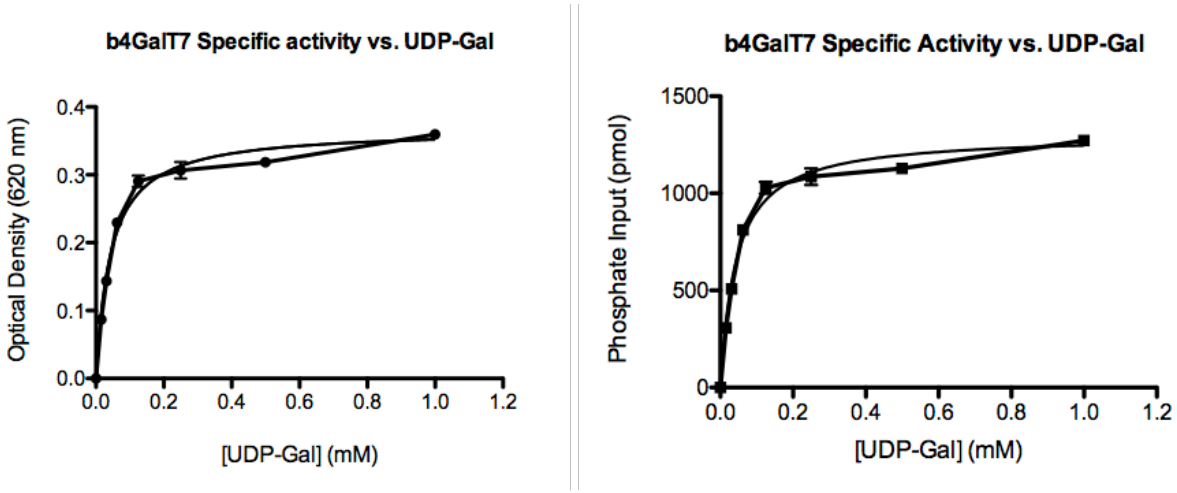


Figure 3.7 Phosphatase-coupled assay result of UDP-Gal. k_{cat} = 27.5 min^{-1}, K_m = 0.04 mM, k_{cat}/K_m = 635 $mM^{-1}min^{-1}$.

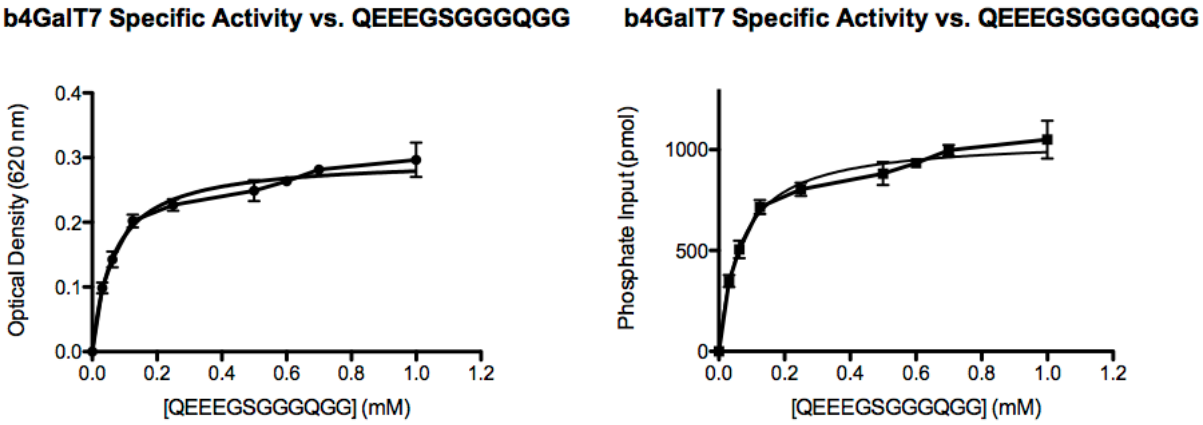


Figure 3.8 Phosphatase-coupled assay result of QEEEGSGGGQGG **1**. k_{cat} = 10 min^{-1}, K_m = 0.07 mM, k_{cat}/K_m = 144 $mM^{-1}min^{-1}$.

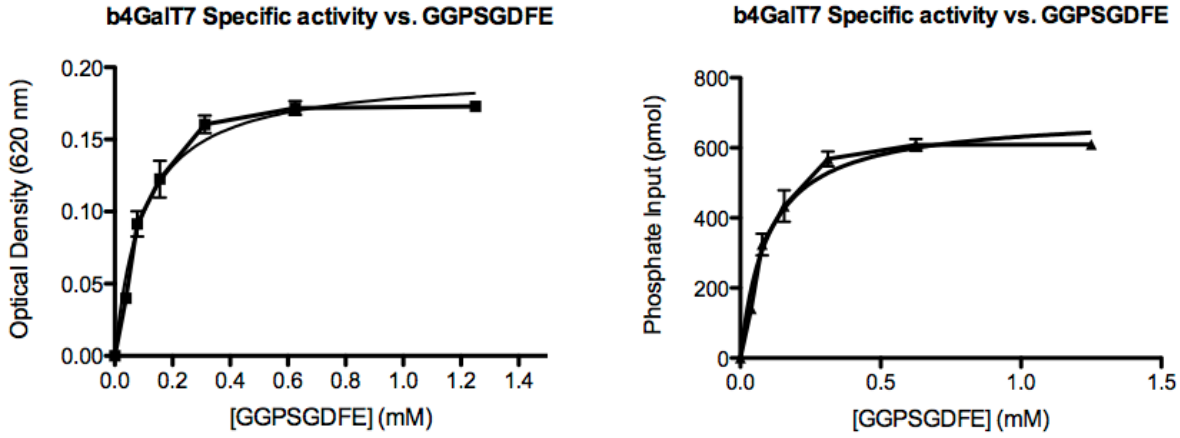


Figure 3.9 Phosphatase-coupled assay result of GGPSGDFE **7**. k_{cat} = 28 min^{-1}, K_m = 0.10 mM, k_{cat}/K_m = 281 $mM^{-1}min^{-1}$.

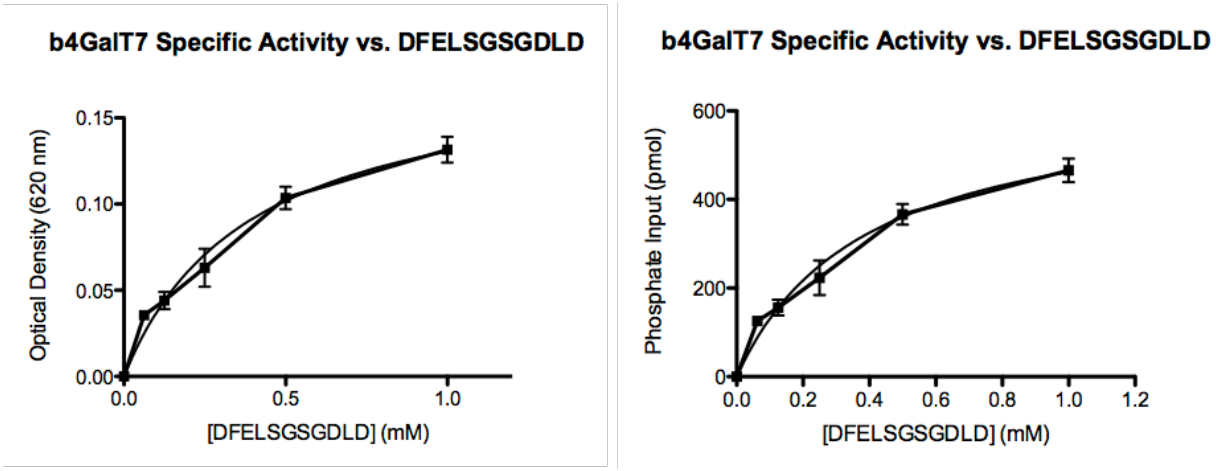


Figure 3.10 Phosphatase-coupled assay result of DFELSGSGDLD **8**. k_{cat} = 4 min^{-1}, K_m = 0.39 mM, k_{cat}/K_m = 11 $mM^{-1}min^{-1}$.

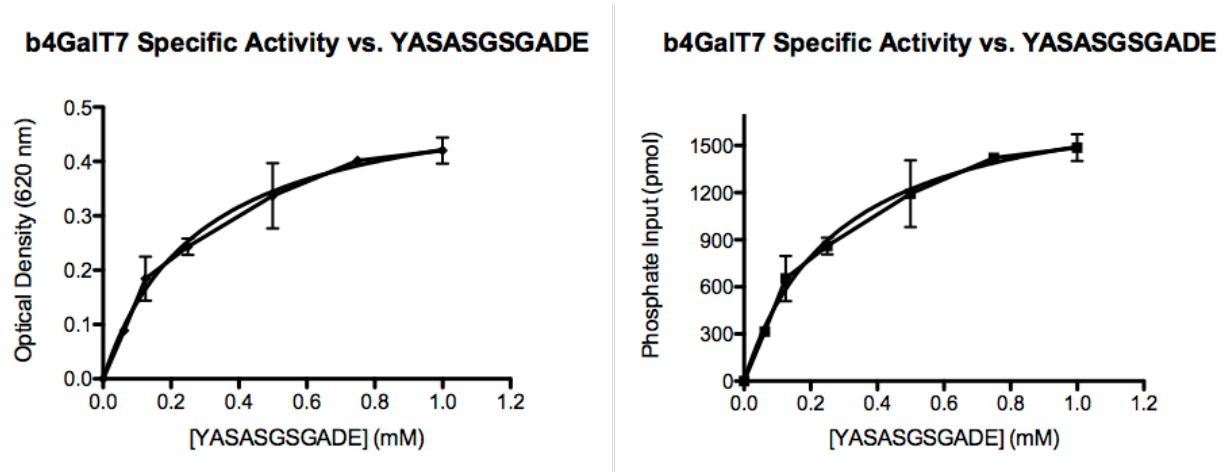
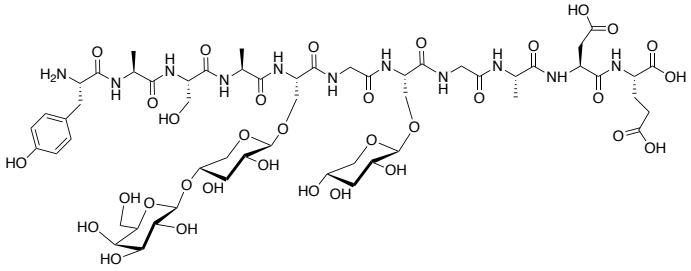


Figure 3.11 Phosphatase-coupled assay result of YASASGSGADE **9**. k_{cat} = 9 min^{-1}, K_m = 0.28 mM, k_{cat}/K_m = 34 $mM^{-1}min^{-1}$.

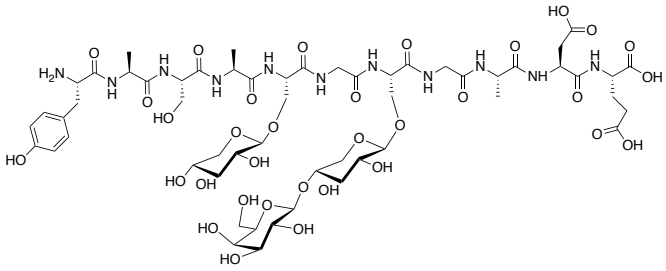
Sequence	SPPS Yield (%) (Cl-TCP Resin, 50 °C)	SPPS Yield (%) (Cl-MPA Resin)*	Deprotection Yield (%)
QEEEGS(O-Xyl)G **6**	14.6	N/A	75
GGPS(O-Xyl)GDFE **7**	12.8	N/A	87
DFELS(O-Xyl)GS(O-Xyl)GDLD **8**	5.6	30	78
YASAS(O-Xyl)GS(O-Xyl)GADE **9**	7.4	25	82
DNFS(O-Xyl)GS(O-Xyl)GAG **10**	11.5	26	75
DLYS(O-Xyl)GS(O-Xyl)GS(O-Xyl)GYFE **11**	2.6	13	63

Table 3.3 Summary of synthesized glycopeptides and the corresponding yields. (N/A: not performed) (*Coupling of the glycosyl amino acid was performed at 50 ºC and the couplings of non-glycosylated amino acids were performed at 30 ºC)

SDC2_Human
51
YASASGSGADE



Exact Mass: 1439.5311



Exact Mass: 1439.5311

Fragmentation: EThcD; Data Searching: Byonic software

LC-MS2 Result Summary

Analysis #1

	MS2 ions of YASAS[+294.09508]GS[+132.04226]GADE							MS2 ions of YASAS[+132.04226]GS[+294.09508]GADE							
Ions	m/z	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5	Scan 6	Ions	m/z	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5	Scan 6
y6	667.2417	1.53E+04	3.72E+04	1.20E+05	6.50E+02	0.00E+00	0.00E+00	y6+Gal	829.2945	4.65E+04	7.53E+05	7.72E+04	1.27E+04	2.89E+03	9.12E+02
y5	610.2202	0.00E+00	0.00E+00	1.27E+04	0.00E+00	0.00E+00	0.00E+00	y5+Gal	772.273	8.16E+03	8.79E+04	0.00E+00	1.02E+03	0.00E+00	0.00E+00
b6	831.3254	5.15E+03	6.48E+04	3.14E+04	0.00E+00	0.00E+00	0.00E+00	b6-Gal	669.2726	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
c6	848.352	4.36E+04	1.16E+05	5.65E+05	1.35E+03	5.82E+02	0.00E+00	c6-Gal	686.2992	6.23E+04	9.15E+05	1.24E+05	2.12E+04	3.13E+03	1.29E+03
c5	791.3305	0.00E+00	0.00E+00	5.18E+04	0.00E+00	0.00E+00	0.00E+00	c5-Gal	629.2777	0.00E+00	7.17E+04	3.34E+04	0.00E+00	0.00E+00	0.00E+00
							Auto-annotated MS2 ions intensity	**Auto-annotated MS2 ions intensity**						**Total intensity**	
							1.07E+06	**2.22E+06**						**3.29E+06**	
							Relative Ratio of YASAS[+294.09508]GS[+132.04226]GADE	**Relative Ratio of YASAS[+132.04226]GS[+294.09508]GADE**							
							32.41%	**67.59%**							

Table 3.4 LC-MS2 characterization of glycosylation intermediates.

Table 3.4 (cont'd)

Analysis #2

MS2 ions of YASAS[+294.09508]GS[+132.04226]GADE							MS2 ions of YASAS[+132.04226]GS[+294.09508]GADE						
ions	m/z	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5	ions	m/z	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5
y6	667.2417	3.07E+04	1.19E+05	1.07E+05	4.65E+04	2.03E+03	y6+Gal	829.2945	9.18E+04	1.31E+06	2.23E+05	2.84E+04	4.64E+04
y5	610.2202	0.00E+00	3.30E+04	0.00E+00	0.00E+00	0.00E+00	y5+Gal	772.273	1.40E+04	2.05E+05	9.25E+04	0.00E+00	7.26E+03
b6	831.3254	1.41E+04	8.95E+04	5.58E+04	1.39E+04	2.12E+03	b6-Gal	669.2726	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.54E+03
c6	848.352	7.24E+04	2.81E+05	5.75E+05	7.99E+04	6.09E+03	c6-Gal	686.2992	1.21E+05	1.88E+06	8.18E+05	3.47E+04	6.23E+04
c5	791.3305	2.66E+04	7.13E+04	6.45E+04	1.20E+05	1.40E+03	c5-Gal	629.2777	8.71E+03	1.59E+05	2.00E+05	1.15E+04	5.81E+03

Auto-annotated MS2 ions intensity	Auto-annotated MS2 ions intensity	Total intensity
1.81E+06	**5.09E+06**	**6.90E+06**

Relative Ratio of YASAS[+294.09508]GS[+132.04226]GADE	Relative Ratio of YASAS[+132.04226]GS[+294.09508]GADE
26.27%	**73.73%**

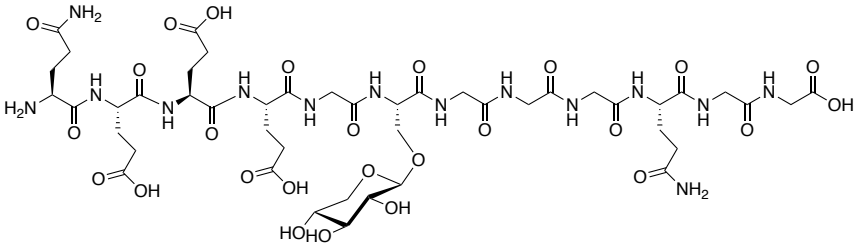
Analysis #3

MS2 ions of YASAS[+294.09508]GS[+132.04226]GADE								MS2 ions of YASAS[+132.04226]GS[+294.09508]GADE							
ions	m/z	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5	Scan 6	ions	m/z	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5	Scan 6
y6	667.241	1.89E+04	6.48E+04	2.70E+05	0.00E+00	3.28E+03	5.22E+03	y6+Gal	829.2945	5.67E+04	1.09E+06	2.26E+05	2.24E+04	4.99E+04	1.68E+04
y5	610.2202	0.00E+00	1.22E+04	1.70E+04	0.00E+00	0.00E+00	0.00E+00	y5+Gal	772.273	9.08E+03	1.77E+05	5.02E+04	0.00E+00	8.53E+03	1.17E+03
b6	831.3254	1.47E+04	7.13E+04	5.71E+04	0.00E+00	1.63E+03	0.00E+00	b6-Gal	669.2726	0.00E+00	1.25E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00
c6	848.352	5.49E+04	1.21E+05	1.01E+06	0.00E+00	5.50E+03	7.63E+03	c6-Gal	686.2992	8.37E+04	1.47E+06	6.70E+05	6.12E+04	7.83E+04	1.24E+04
c5	791.3305	1.62E+04	2.37E+04	8.80E+04	9.02E+03	1.85E+03	1.84E+03	c5-Gal	629.2777	1.22E+04	1.12E+05	1.34E+05	0.00E+00	4.68E+03	1.66E+03

Auto-annotated MS2 ions intensity	Auto-annotated MS2 ions intensity	Total intensity
1.88E+06	**4.36E+06**	**6.24E+06**

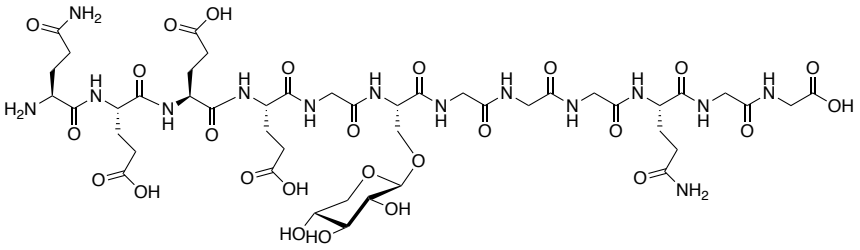
Relative Ratio of YASAS[+294.09508]GS[+132.04226]GADE	Relative Ratio of YASAS[+132.04226]GS[+294.09508]GADE
30.08%	**69.92%**

APPENDIX B: Product Characterization Spectra



1

The purity of glycopeptide was verified with analytical C-18 HPLC (0-10% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = - 1.53 ^{o} (c = 0.08, H_2O). ^{1}H-NMR (500 MHz, D_2O), δ 4.50 (t, *J* = 4.8 Hz, 1H), 4.30 – 4.17 (m, 3H), 4.17 – 3.98 (m, 4H), 3.91 – 3.67 (m, 15H), 3.60 (d, *J* = 2.0 Hz, 4H), 3.42 (td, *J* = 9.9, 5.5 Hz, 2H), 3.34 – 3.20 (m, 3H), 3.19 – 3.01 (m, 3H), 2.21 (t, *J* = 7.5 Hz, 3H), 2.18 – 2.05 (m, 8H), 2.07 – 1.95 (m, 3H), 1.93-1.87 (m, 4H), 1.87 – 1.74 (m, 5H), 1.74 (s, 8H). ^{13}C-NMR (225 MHz, D_2O), δ 177.9, 177.0, 174.3, 173.8, 173.4, 172.1, 171.6, 170.9, 103.0, 75.5, 72.9, 72.8, 69.5, 69.1, 68.7, 65.2, 53.9, 53.8, 53.7, 53.7, 53.2, 52.4, 52.2, 44.5, 43.2, 42.7, 42.4, 42.4, 33.6, 31.0, 30.1, 27.7, 27.6, 27.5, 26.7, 26.6, 26.6, 25.1, 23.2, 22.2, 21.4. ESI-MS: $C_{45}H_{70}N_{14}O_{26}$ $[M+H]^+$ calcd: 1223.4659, obsd: 1223.4637 (1.8 ppm).



1

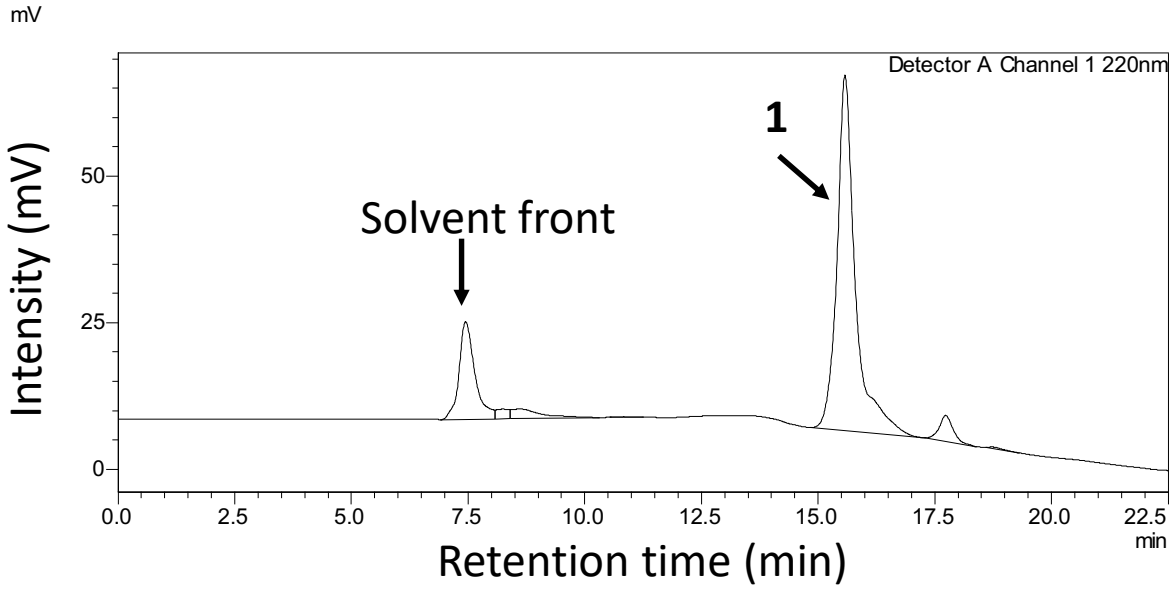
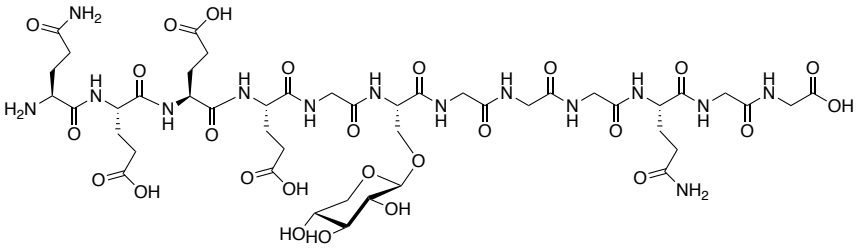


Figure 3.12 HPLC chromatogram of **1**.



1

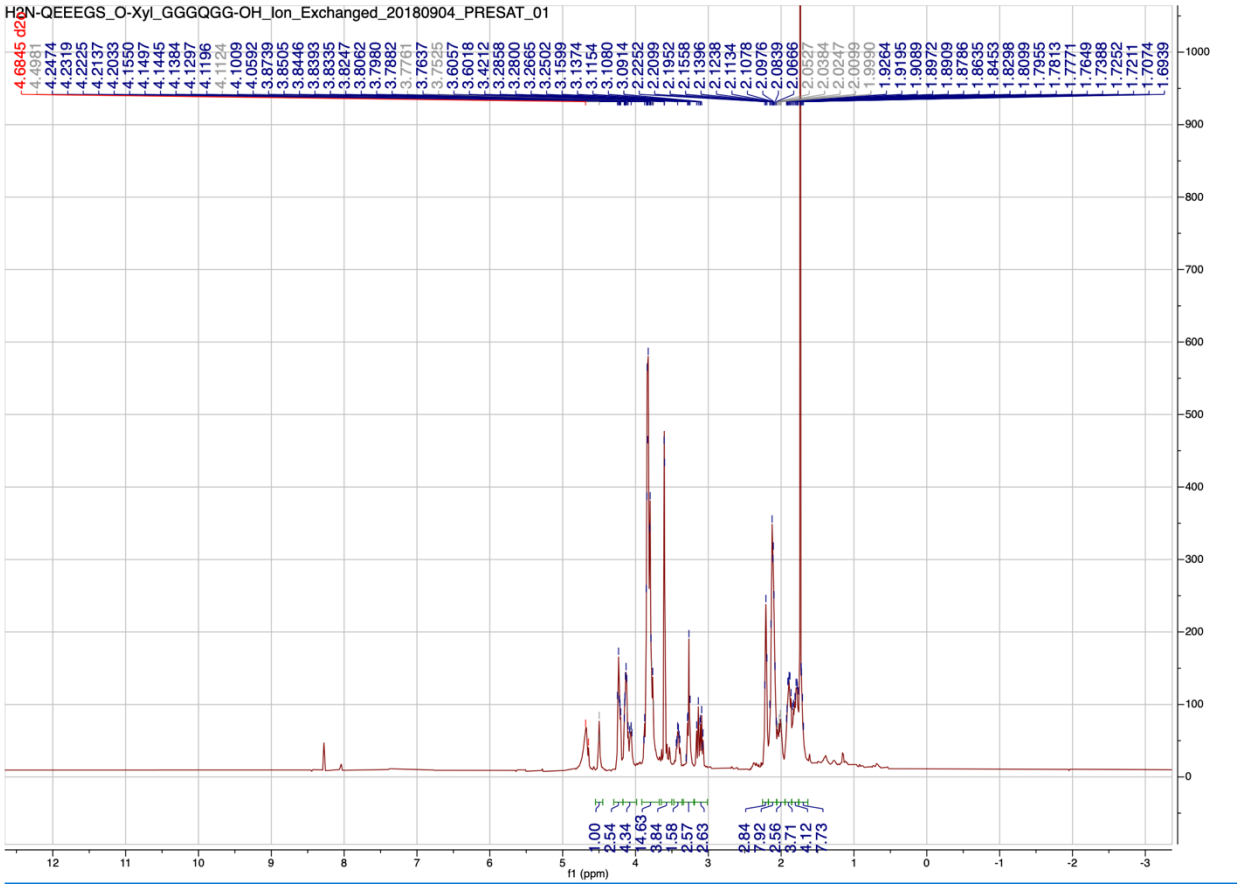
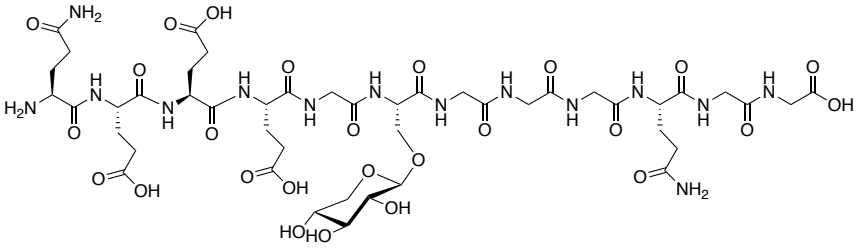


Figure 3.13 ^{1}H-NMR of **1** (500 MHz, D_2O).



1

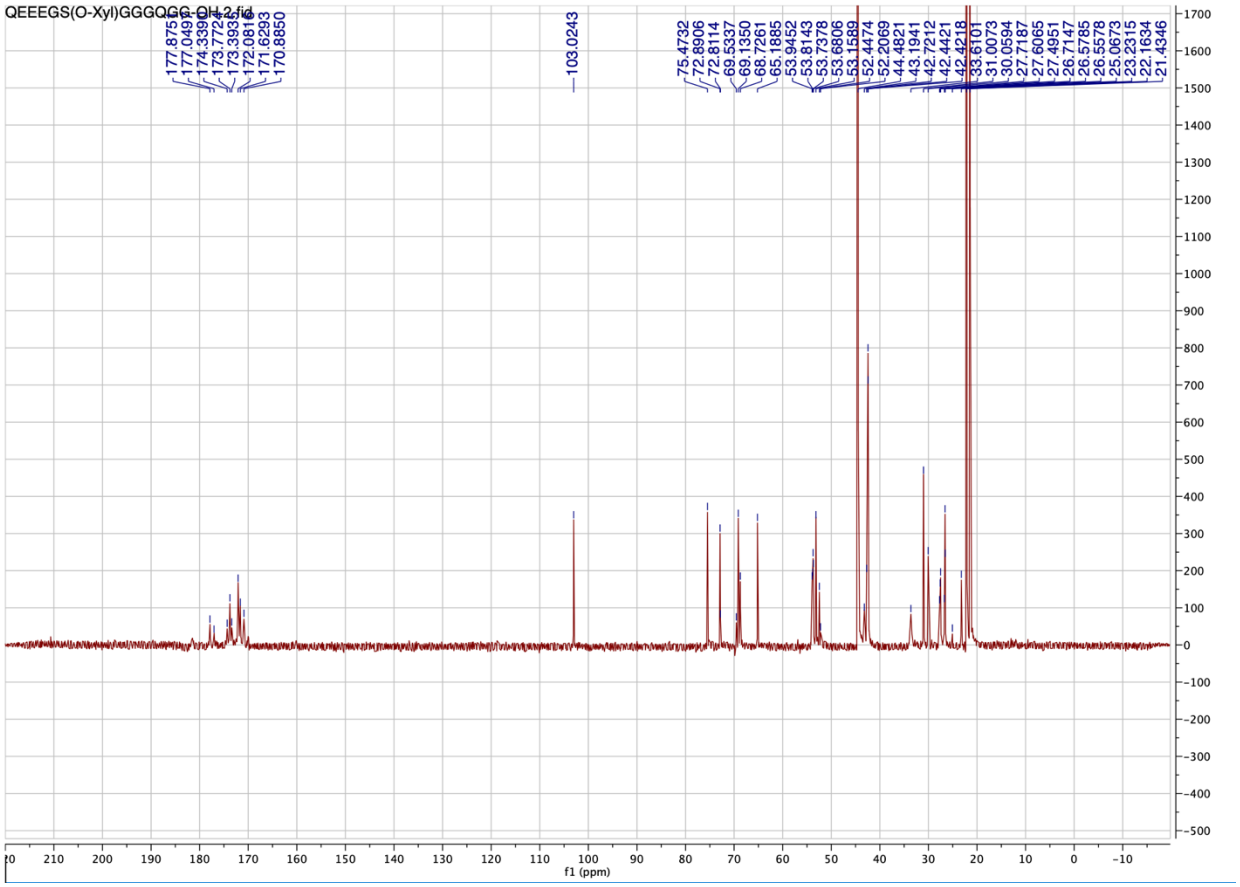
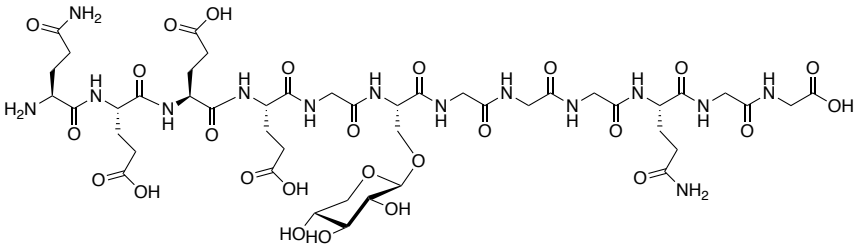


Figure 3.14 ^{13}C-NMR of **1** (225MHz, D_2O).



1

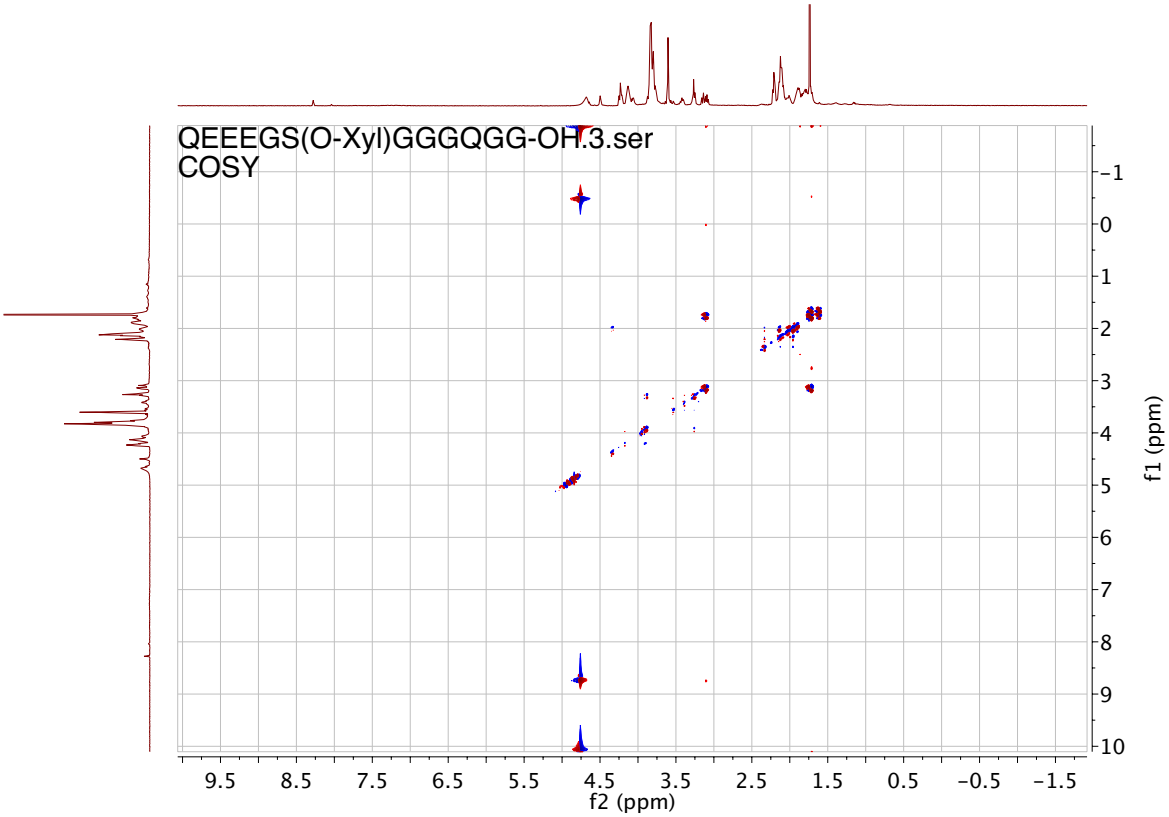
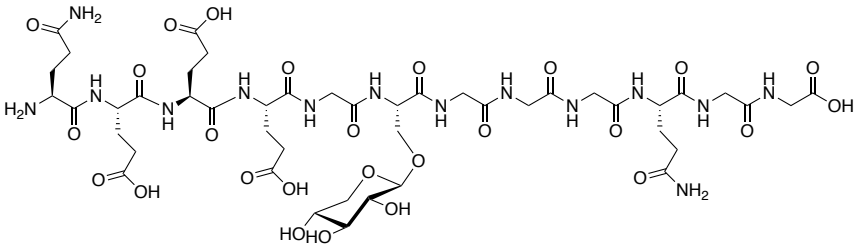


Figure 3.15 COSY NMR of **1** (900MHz, D_2O).



1

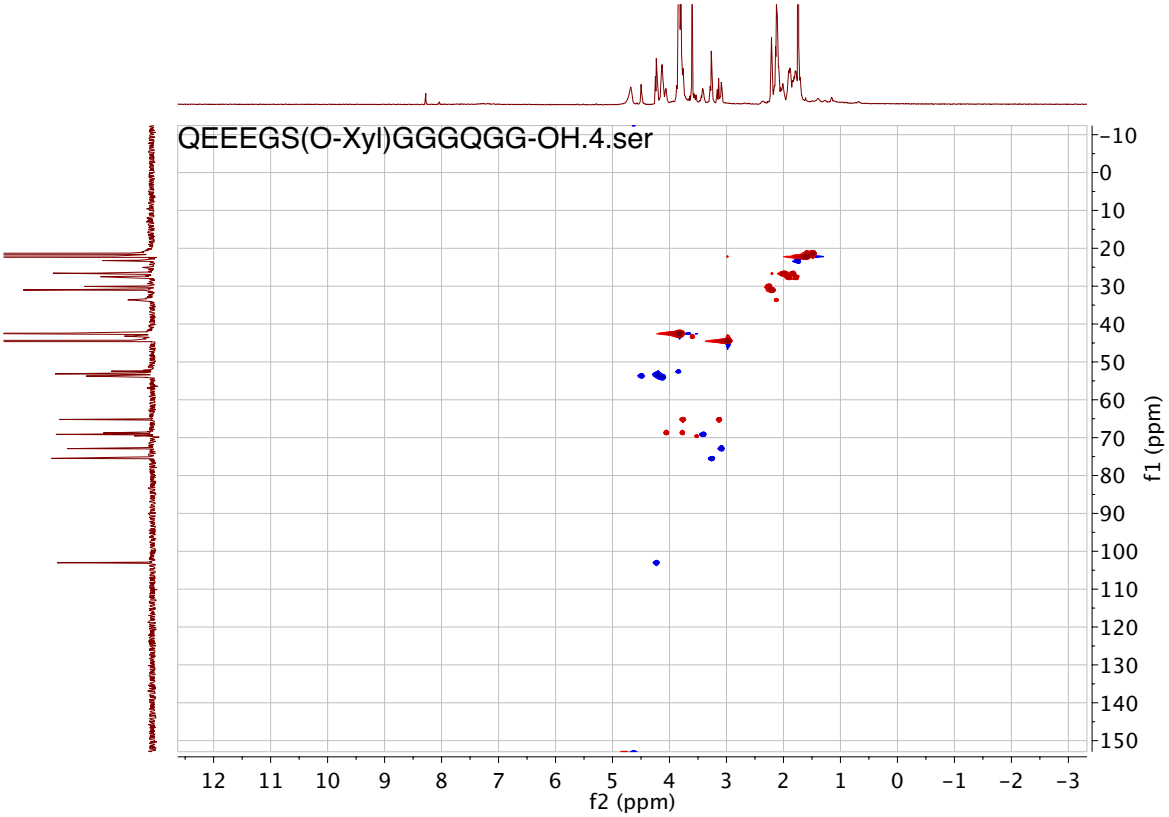
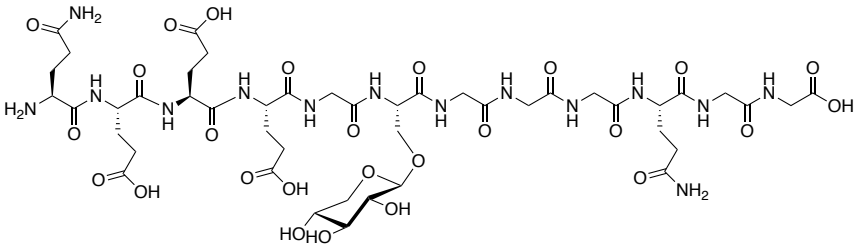
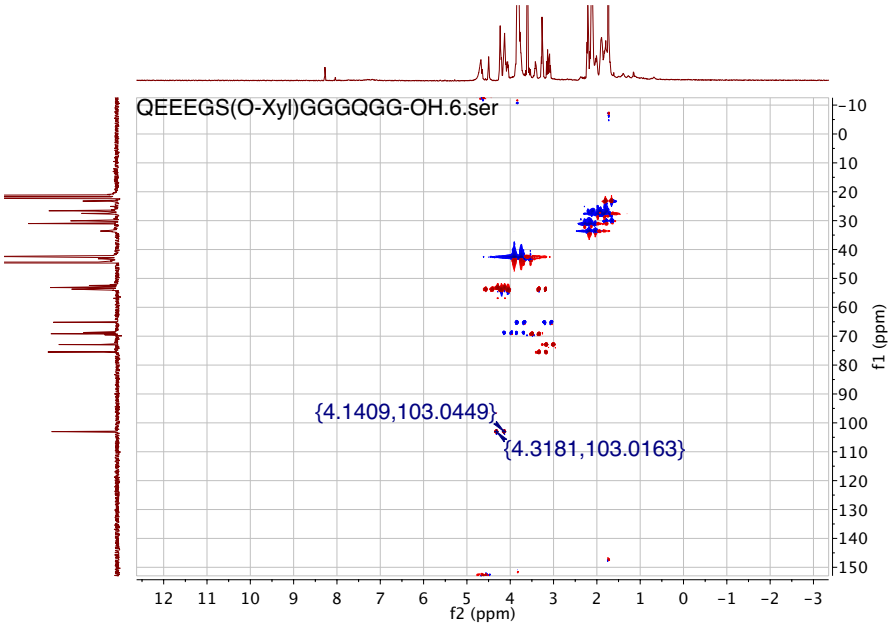


Figure 3.16 HSQC NMR of **1** (900MHz, D_2O).



1



$^{1}J_{C1,\,H1}$ = 159.5 Hz

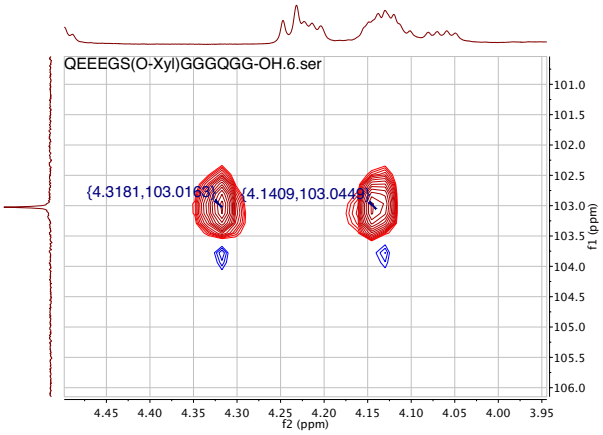
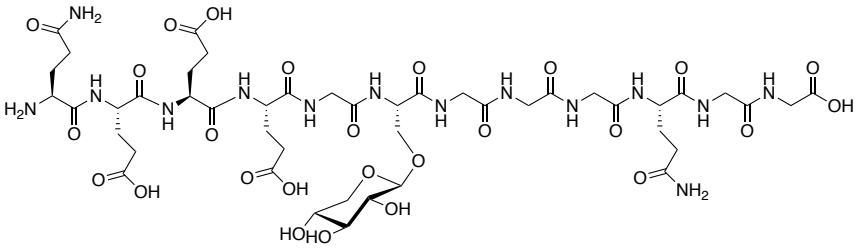


Figure 3.17 HSQC-coupled NMR of **1** (900MHz, D_2O).



1

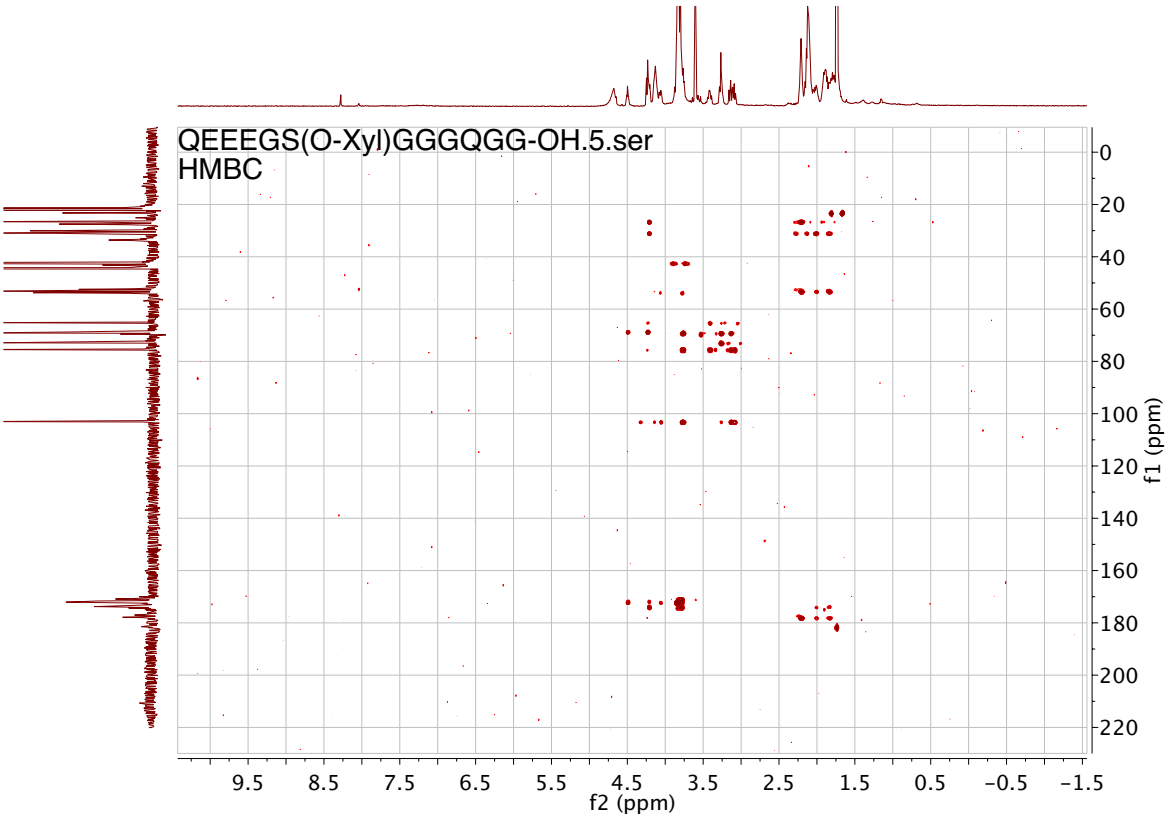
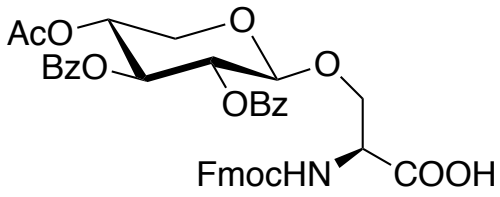
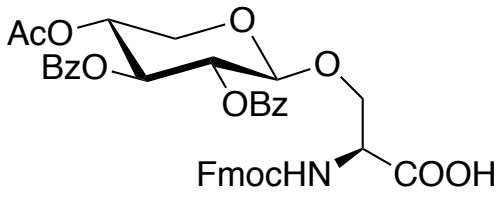


Figure 3.18 HMBC NMR of **1** (900MHz, D_2O).



2

$[\alpha]_D^{20}$ = - 3.7 ° (c = 0.20, methanol). ^{1}H-NMR (500 MHz, CD_3OD), δ 7.95 – 7.90 (m, 3H), 7.78 (dd, *J* = 7.6, 3.4 Hz, 2H), 7.62 – 7.52 (m, 3H), 7.52 – 7.45 (m, 1H), 7.42 (t, *J* = 7.8 Hz, 2H), 7.36 (dq, *J* = 15.5, 7.5 Hz, 3H), 7.30 – 7.22 (m, 5H), 5.55 (q, *J* = 6.8, 5.7 Hz, 1H), 5.24 (dd, *J* = 8.0, 6.0 Hz, 1H), 5.13 (d, *J* = 12.3 Hz, 1H), 5.09 – 5.01 (m, 2H), 4.85 (d, *J* = 6.0 Hz, 1H), 4.49 (t, *J* = 4.6 Hz, 1H), 4.28 (dd, *J* = 10.4, 6.8 Hz, 1H), 4.21 – 4.11 (m, 2H), 4.11 – 4.01 (m, 2H), 3.89 (dd, *J* = 10.4, 4.3 Hz, 1H), 3.54 (dd, *J* = 12.1, 7.8 Hz, 1H), 1.98 (s, 3H). ^{13}C-NMR (125 MHz, CD_3OD), δ 170.0, 169.8, 165.4, 165.2, 156.8, 143.8, 143.6, 141.1, 141.1, 135.6, 133.2, 133.2, 129.4, 129.1, 128.9, 128.2, 128.1, 128.1, 127.9, 127.9, 127.4, 127.4, 126.8, 124.9, 124.8, 119.5, 99.9, 71.1, 70.7, 68.4, 68.0, 66.7, 66.7, 61.1, 54.2, 48.2, 48.1, 47.9, 47.7, 47.6, 47.4, 47.2, 47.1, 46.8, 19.2. ESI-MS: $C_{39}H_{36}NO_{12}$ $[M+H]^+$ calcd: 710.2232, obsd: 710.2243 (1.55 ppm).



2

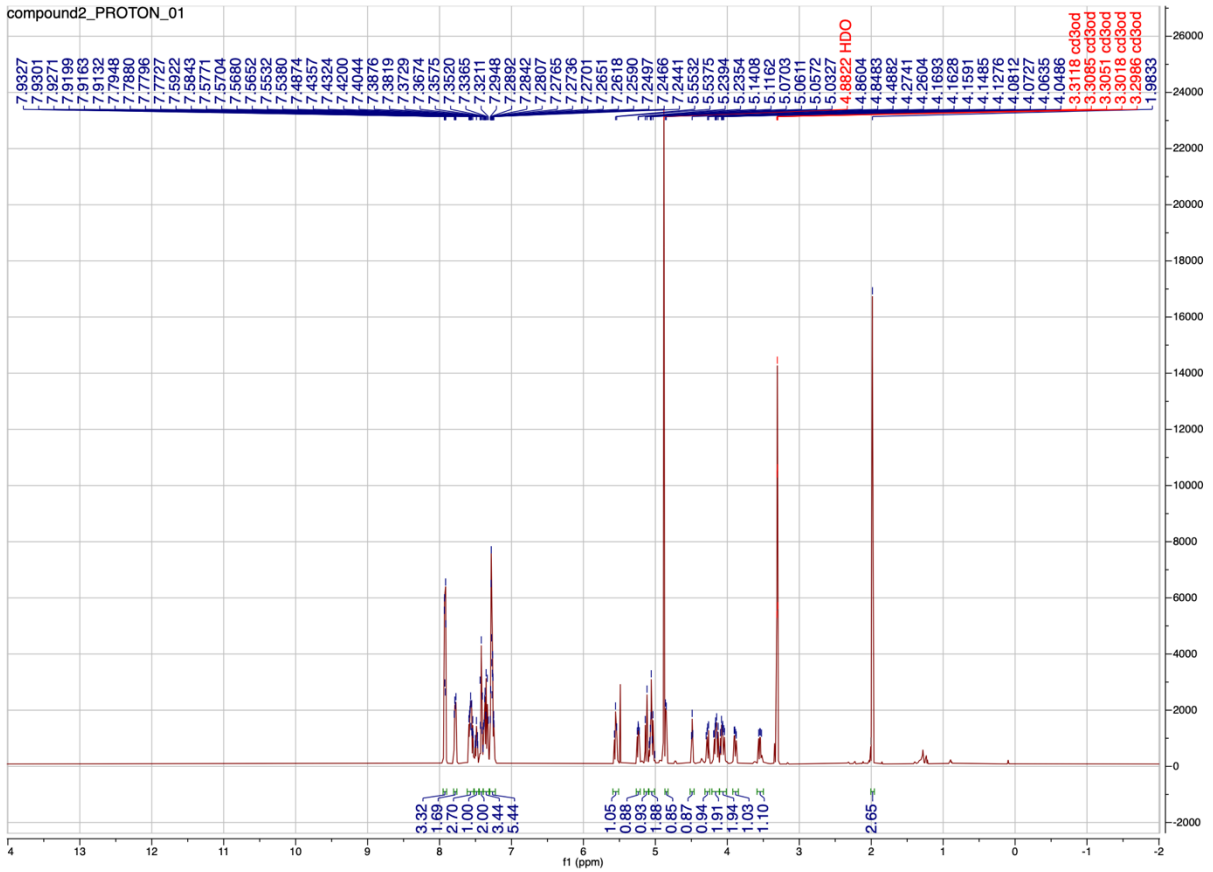
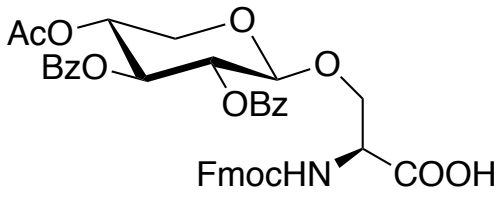


Figure 3.19 ^{1}H NMR (500 MHz, CD_3OD).



2

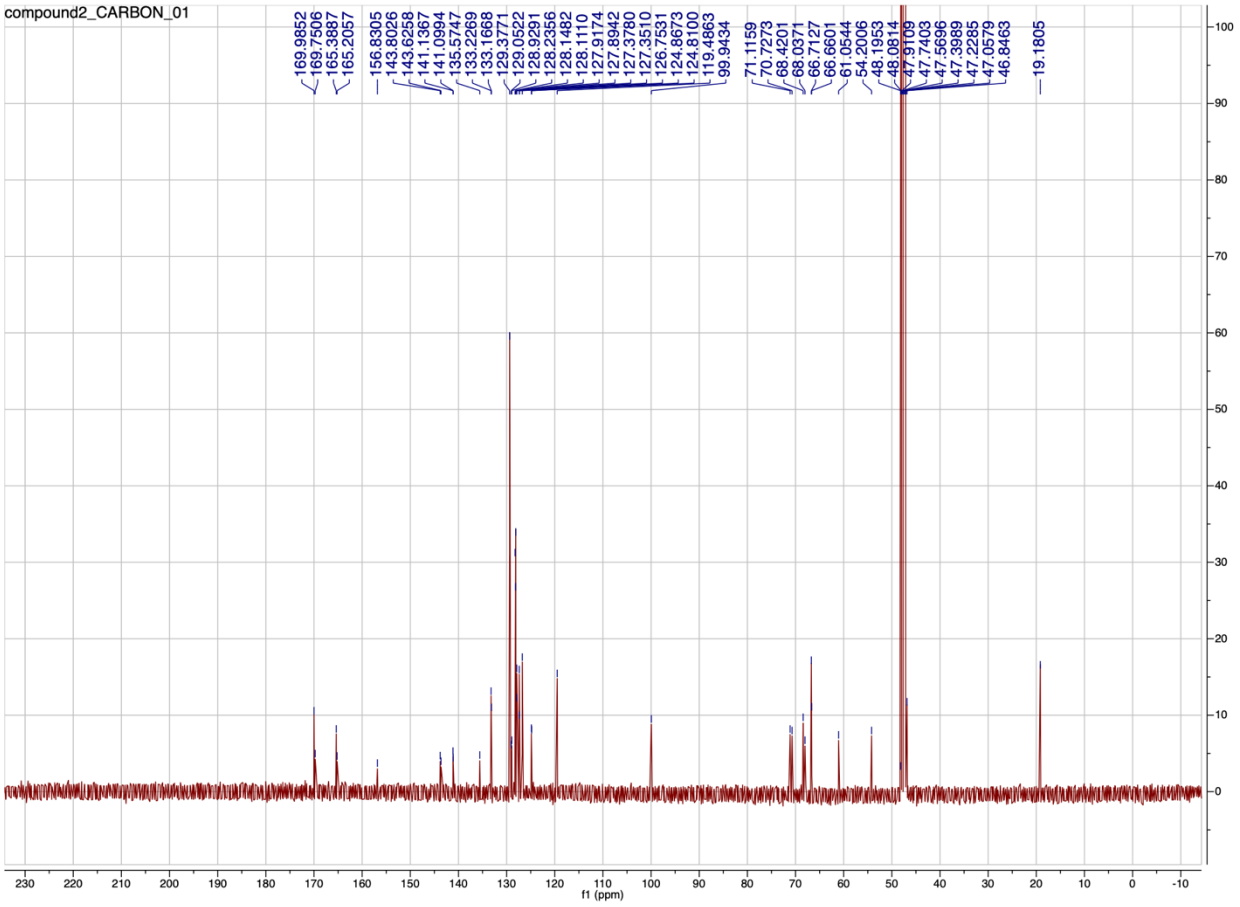
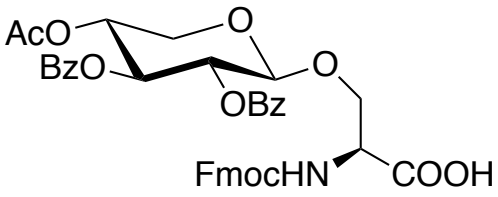


Figure 3.20 ^{13}C NMR (125 MHz, CD_3OD).



2

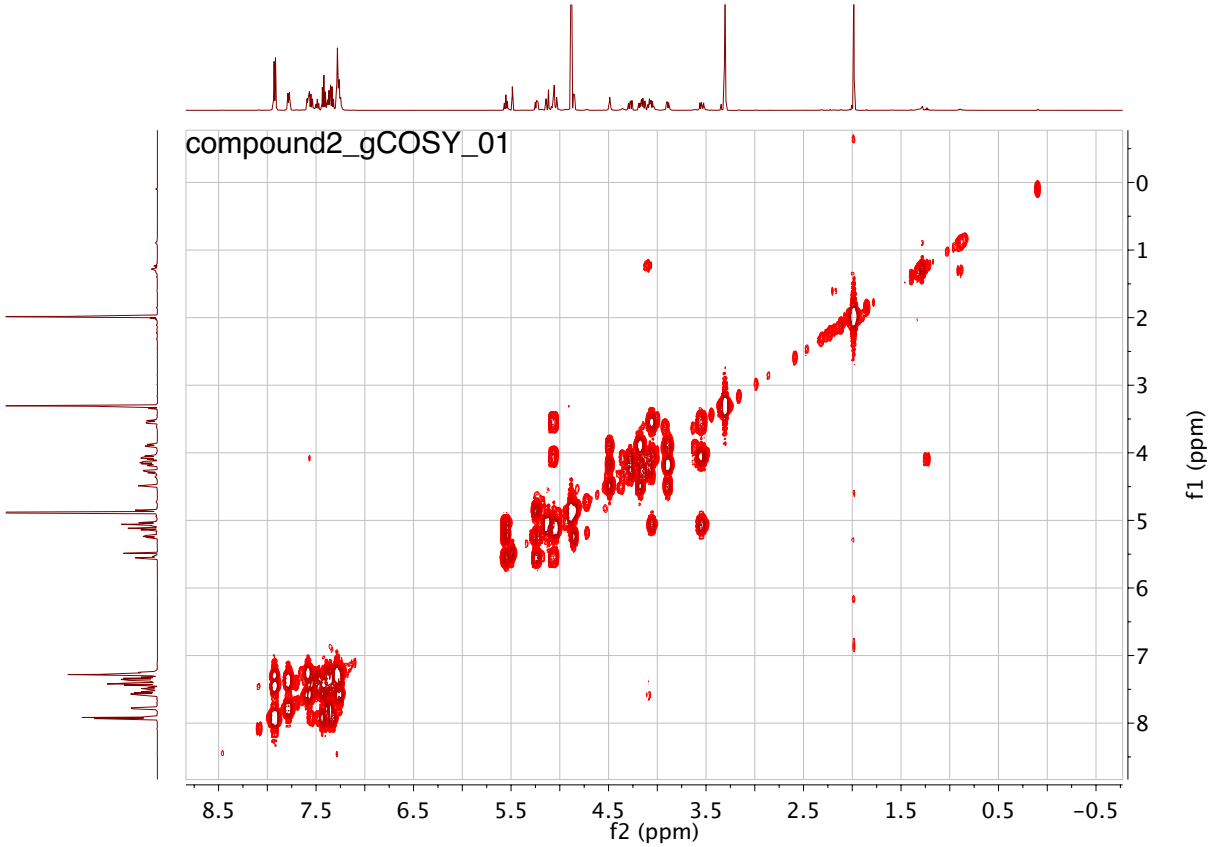
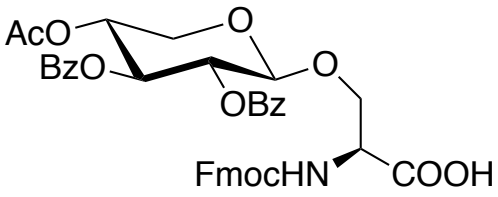


Figure 3.21 COSY NMR of **2** (500MHz, CD_3OD).



2

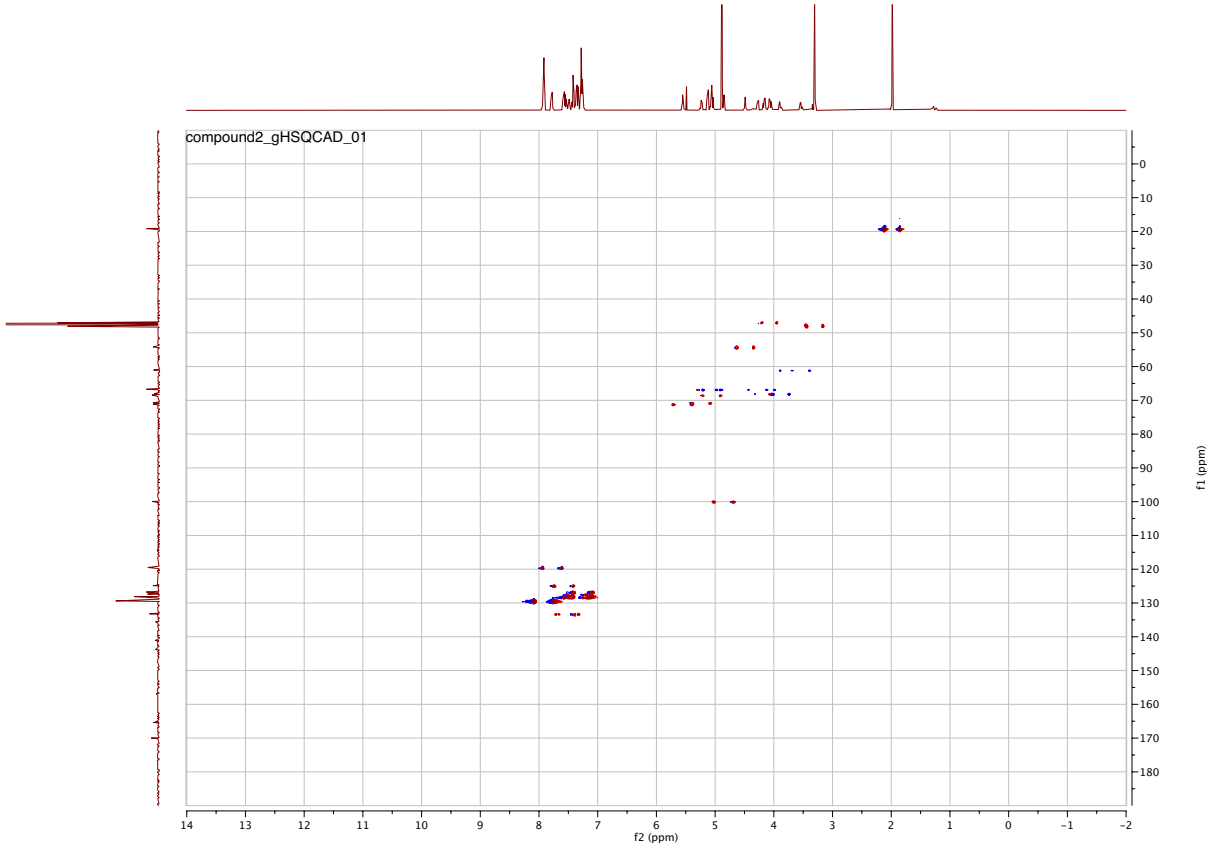
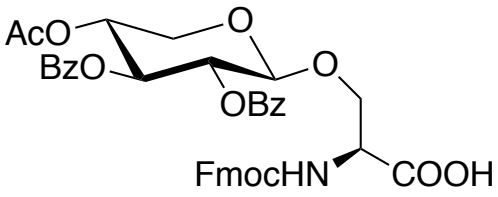


Figure 3.22 HSQC NMR of **2** (500MHz, CD_3OD).



2

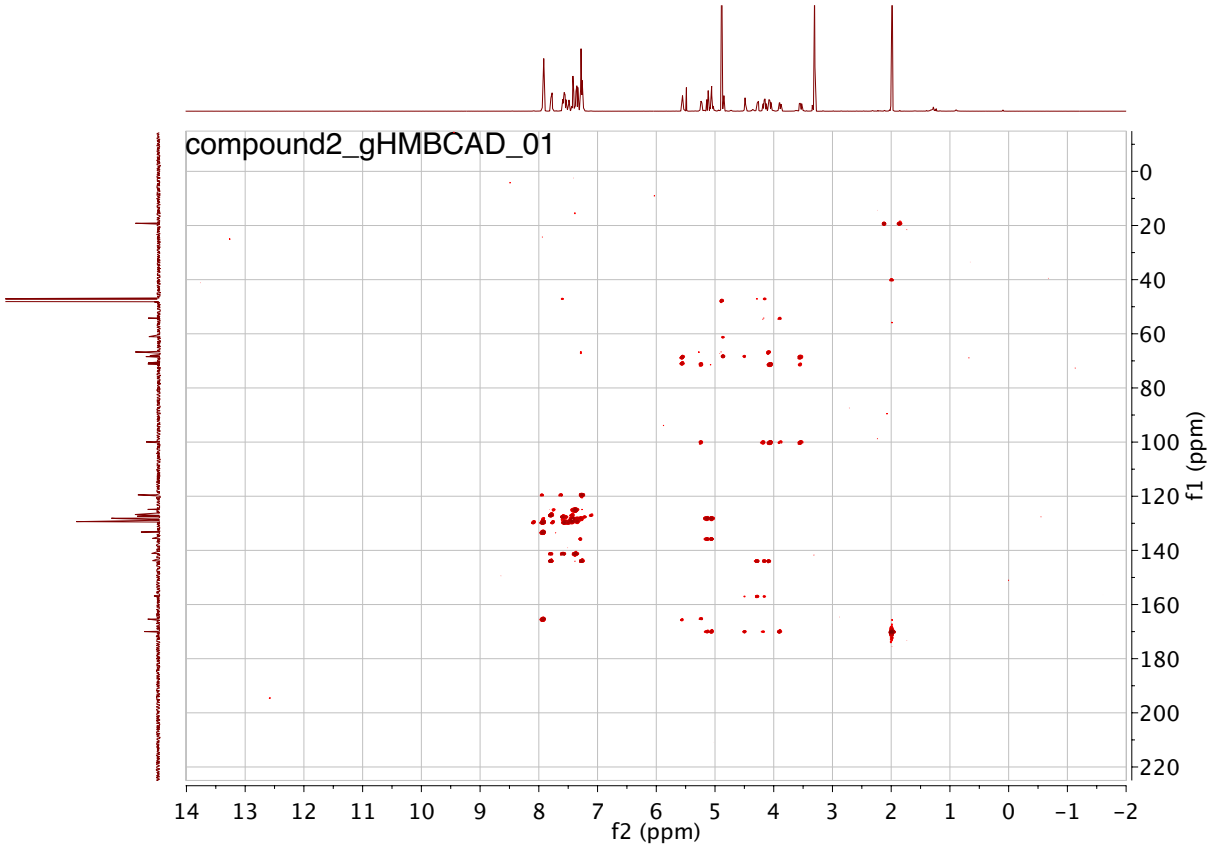
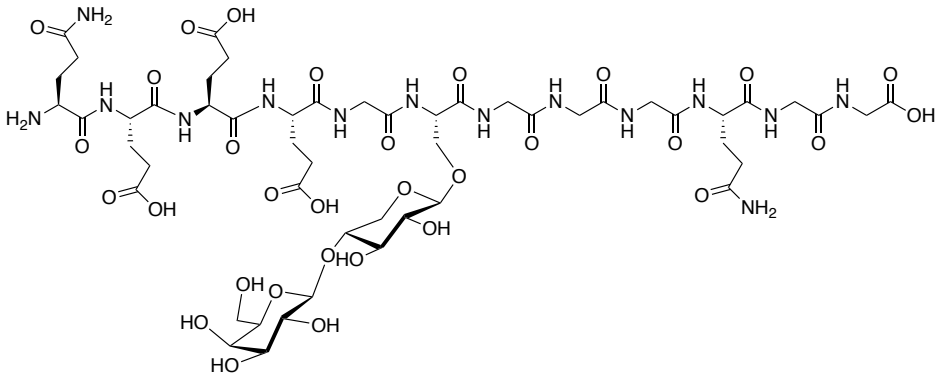
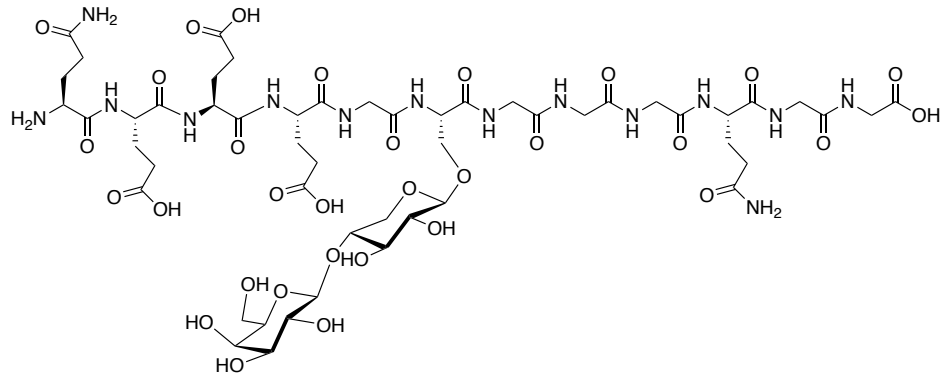


Figure 3.23 HMBC NMR of **2** (500MHz, CD_3OD).



5

The purity of glycopeptide was verified with analytical C-18 HPLC (0-10% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = -1.57 ° (c = 0.14, H_2O). ^{1}H NMR (900 MHz, D_2O), δ 5.83 (s, 2H), 4.29 (m, 3H), 4.23 (s, 4H), 4.08 (s, 2H), 3.85 (s, 10H), 3.76 (s, 4H), 3.70 – 3.62 (m, 4H), 3.59 (m, 3H), 3.55 (s, 3H), 3.49 (dd, *J* = 11.2, 6.0 Hz, 5H), 3.45 (s, 2H), 3.35 (t, *J* = 8.9 Hz, 2H), 3.24 (t, *J* = 12.4 Hz, 1H), 3.16 (s, 1H), 2.30 (s, 1H), 2.24 (s, 1H), 2.03 (s, 1H), 1.86 (s, 1H), 1.25 – 0.95 (m, 3H). ^{13}C NMR (225 MHz, D_2O) δ 102.4, 92.1, 88.4, 72.5, 70.6, 68.8, 62.9, 61.0, 53.5, 42.6, 30.9, 26.6. ESI-MS: $C_{51}H_{80}N_{14}O_{31}$ $[M+H]^+$ calcd: 1385.5187, obsd: 1385.5135 (3.75 ppm).



5

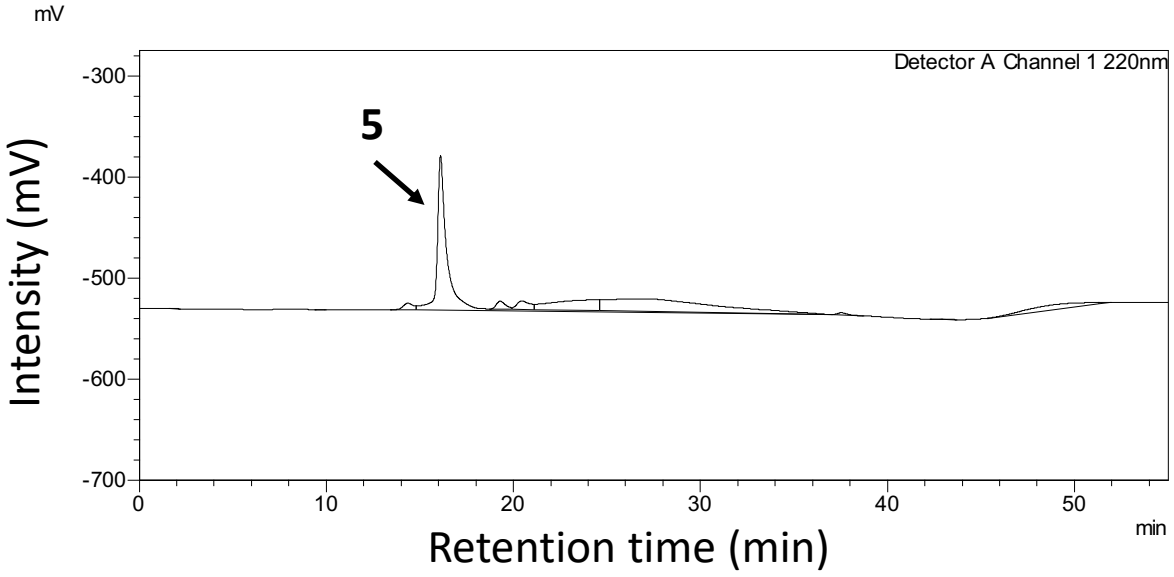
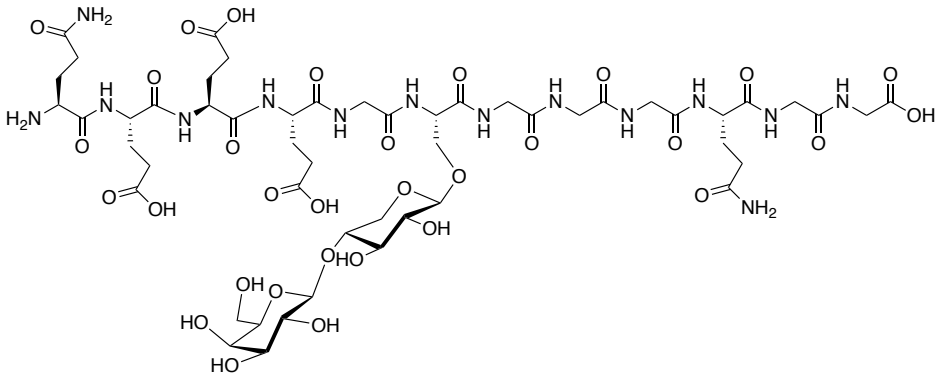


Figure 3.24 HPLC chromatogram of **5**.



5

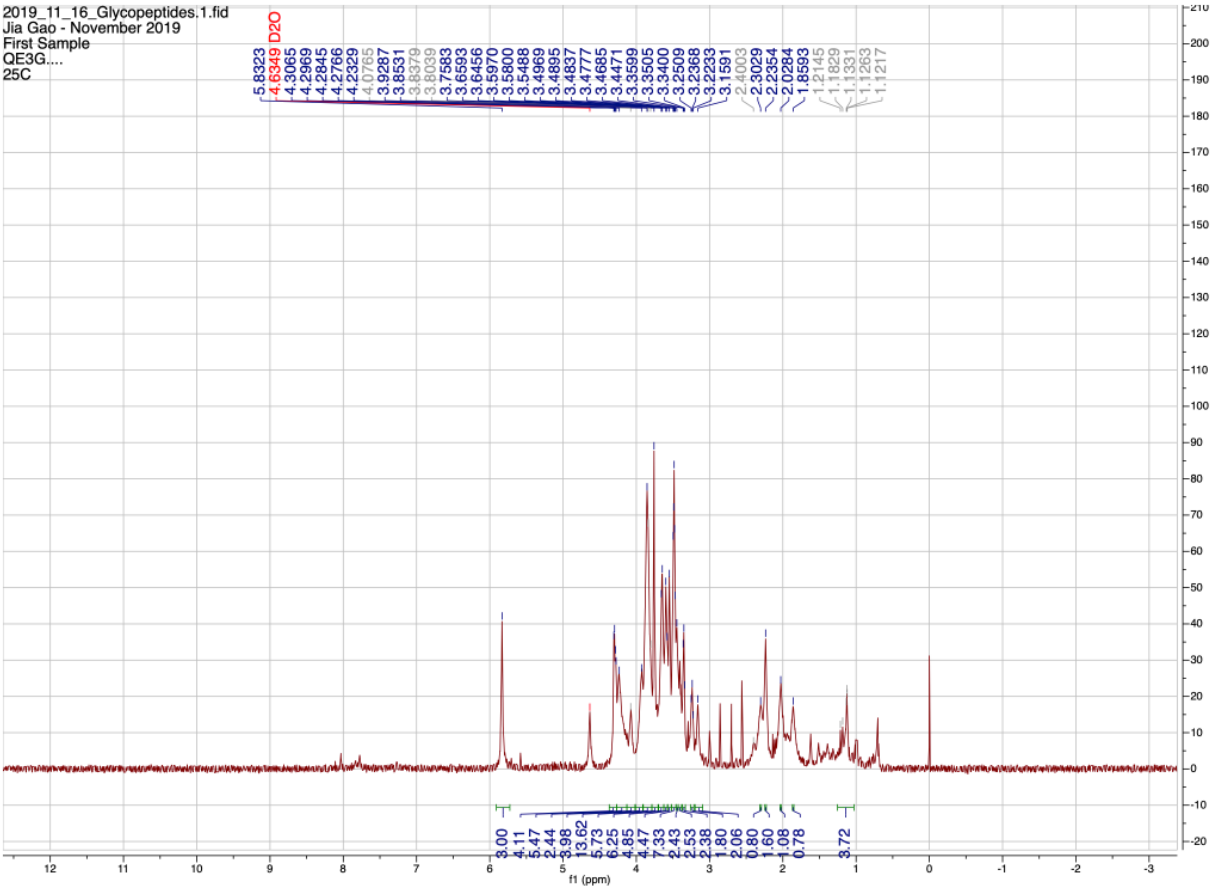
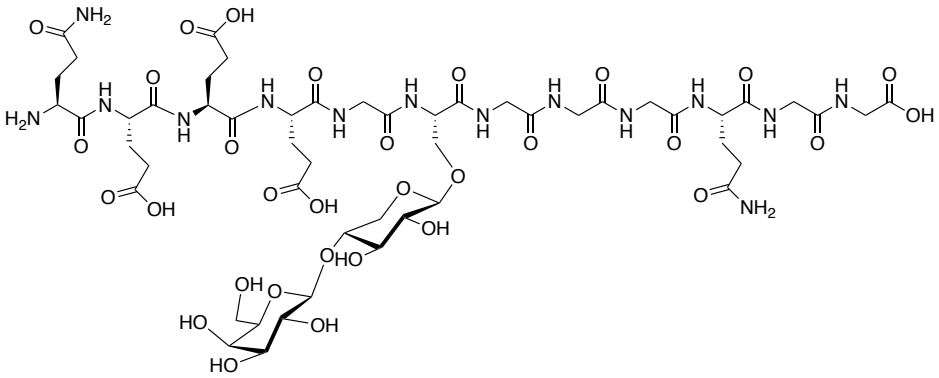


Figure 3.25 ^{1}H NMR of **5** (900 MHz, D_2O).



5

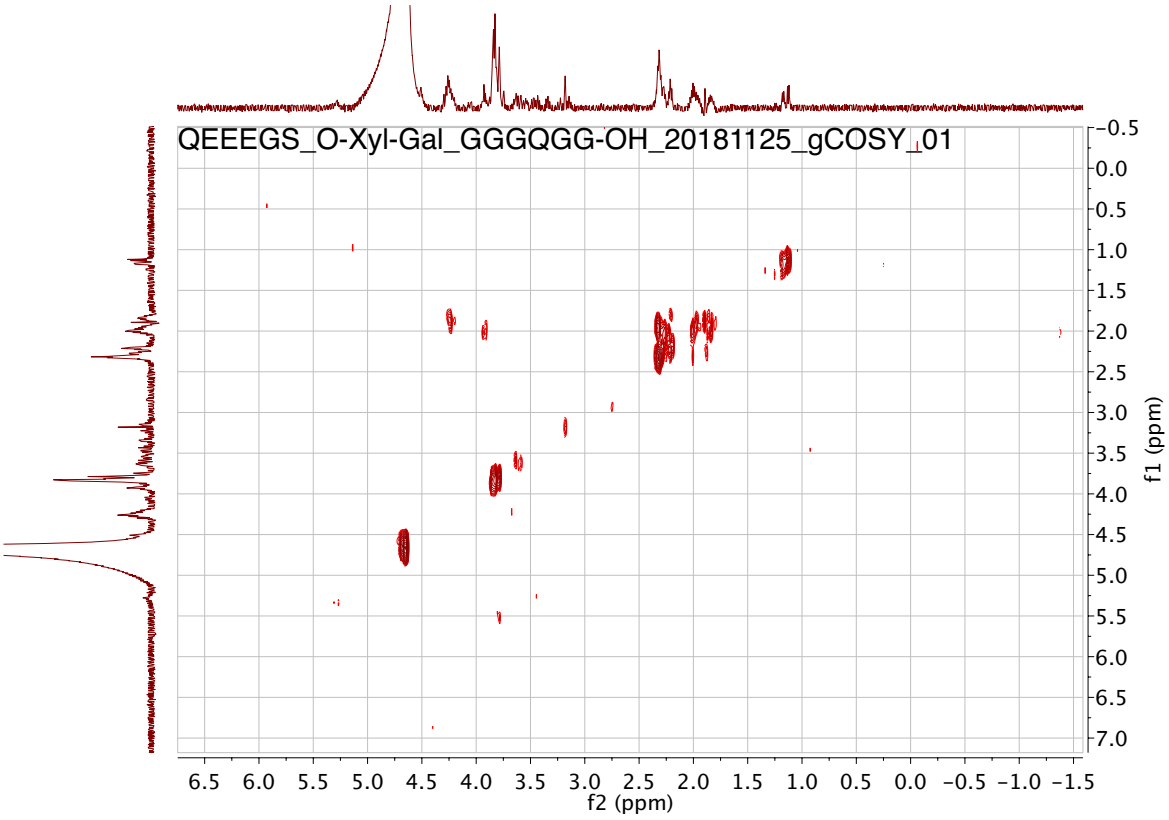
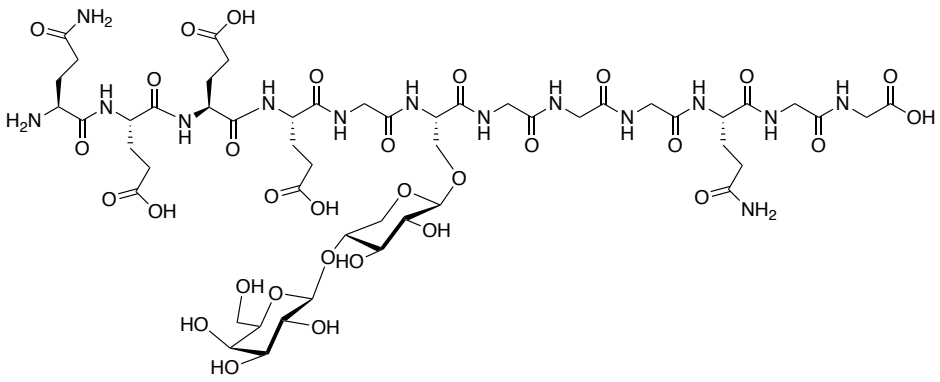


Figure 3.26 COSY NMR of **5** (500 MHz, D_2O).



5

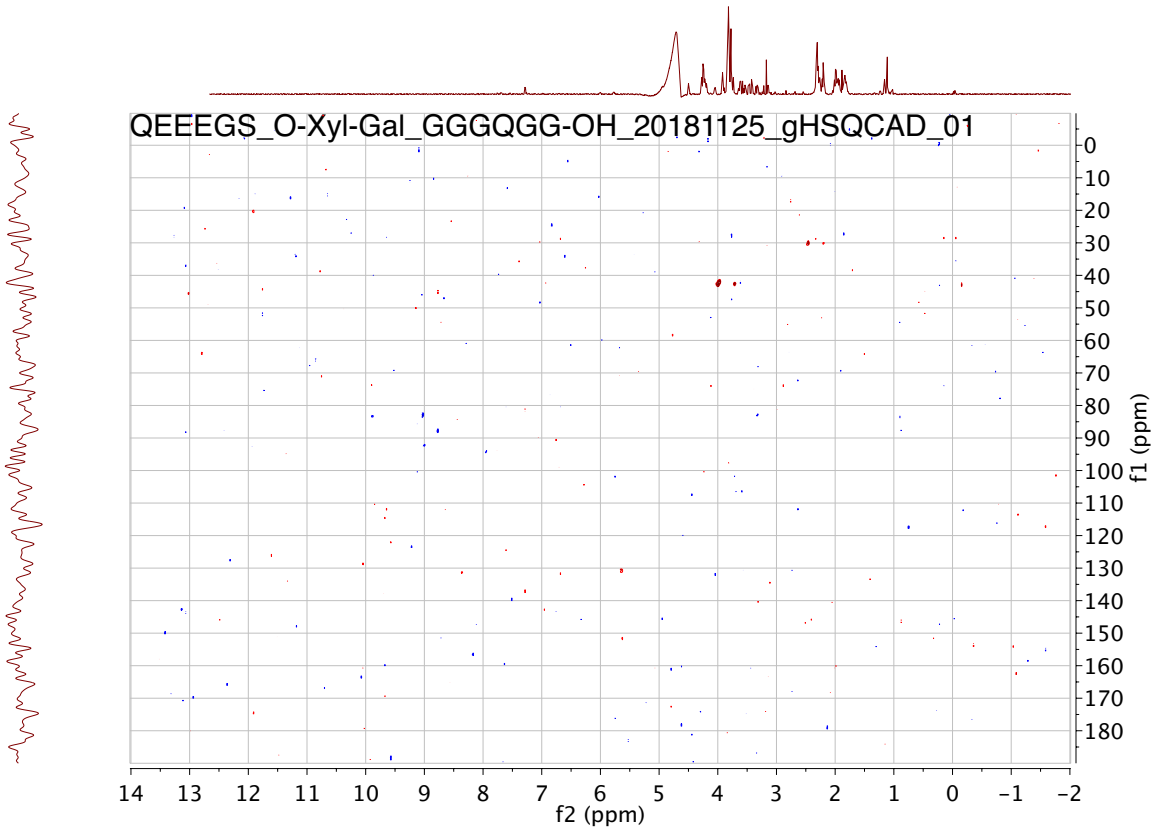
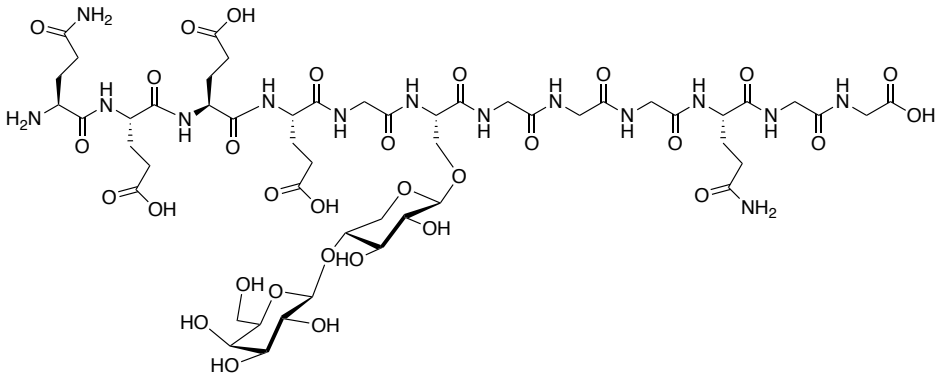
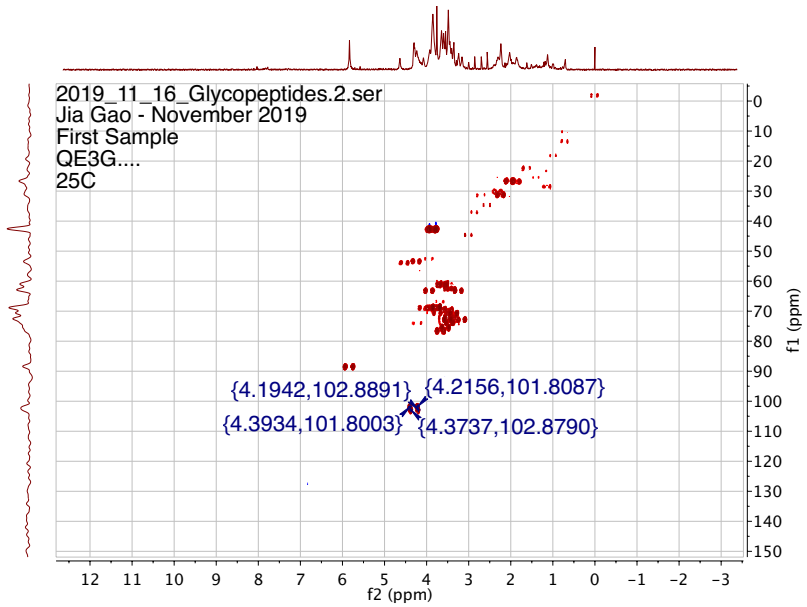


Figure 3.27 HSQC NMR of **5** (500 MHz, D_2O).



5



$^{1}J_{C1,\,H1}$ = 160.0 Hz, 161.6 Hz

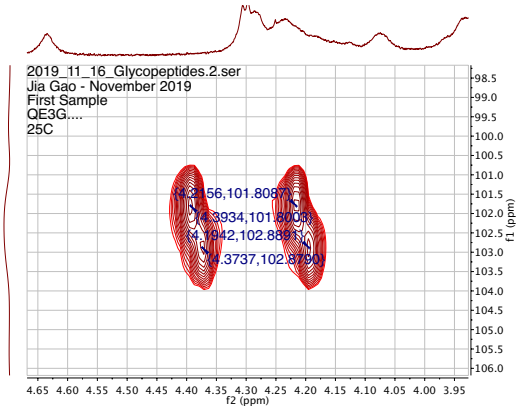
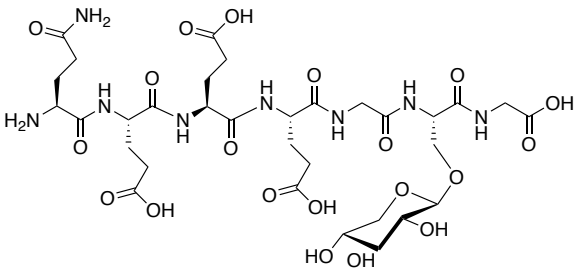
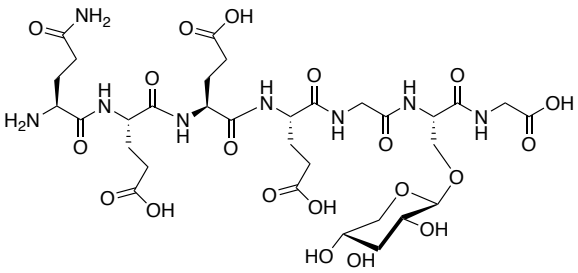


Figure 3.28 HSQC-coupled NMR of **5** (900 MHz, D_2O).



6

The purity of glycopeptide was verified with analytical C-18 HPLC (0-10% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = - 1.570 ° (c = 0.14, H_2O). ^{1}H-NMR (900 MHz, D_2O), δ 4.60 (t, *J* = 5.1 Hz, 1H), 4.30 (d, *J* = 7.8 Hz, 2H), 4.23 – 4.18 (m, 3H), 4.10 (m, 1H), 3.97 – 3.88 (m, 3H), 3.87 – 3.80 (m, 3H), 3.72 – 3.63 (m, 3H), 3.51 (m, 2H), 3.49 – 3.44 (m, 1H), 3.33 (m, 1H), 3.23 – 3.15 (m, 3H), 3.05 (t, *J* = 5.8 Hz, 1H), 2.29 – 2.12 (m, 10H), 2.02 – 1.93 (m, 4H), 1.90 – 1.79 (m, 6H), 1.67 (p, *J* = 5.8 Hz, 1H), 1.56 (m, 1H). ^{13}C-NMR (225 MHz, D_2O), δ 181.4, 178.2, 176.3, 174.3, 173.8, 173.7, 171.3, 170.7, 164.3, 160.3, 103.2, 102.8, 75.4, 72.8, 72.8, 69.1, 69.0, 68.9, 66.0, 65.1, 65.1, 53.9, 53.8, 53.7, 53.4, 53.3, 44.5, 43.4, 42.5, 33.6, 33.5, 30.9, 27.7, 27.6, 27.6, 22.2, 21.5. ESI-MS: $C_{32}H_{50}N_8O_{20}$ $[M+H]^+$ calcd: 867.3214, obsd: 867.3209 (0.58 ppm).



6

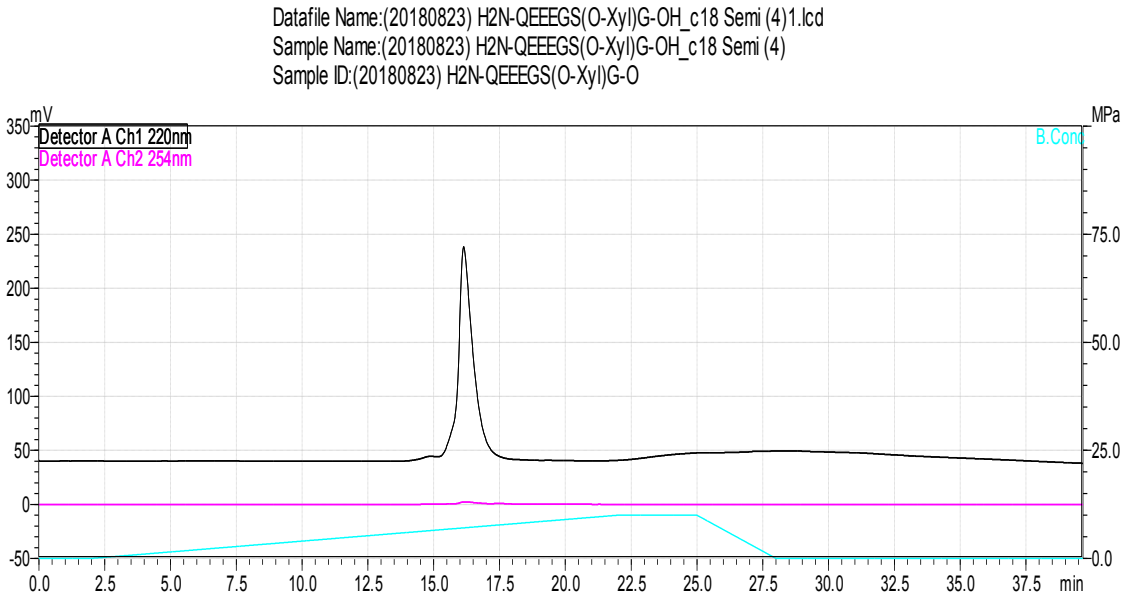
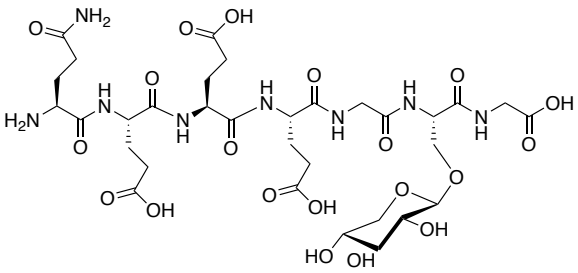


Figure 3.29 HPLC chromatogram of **6**.



6

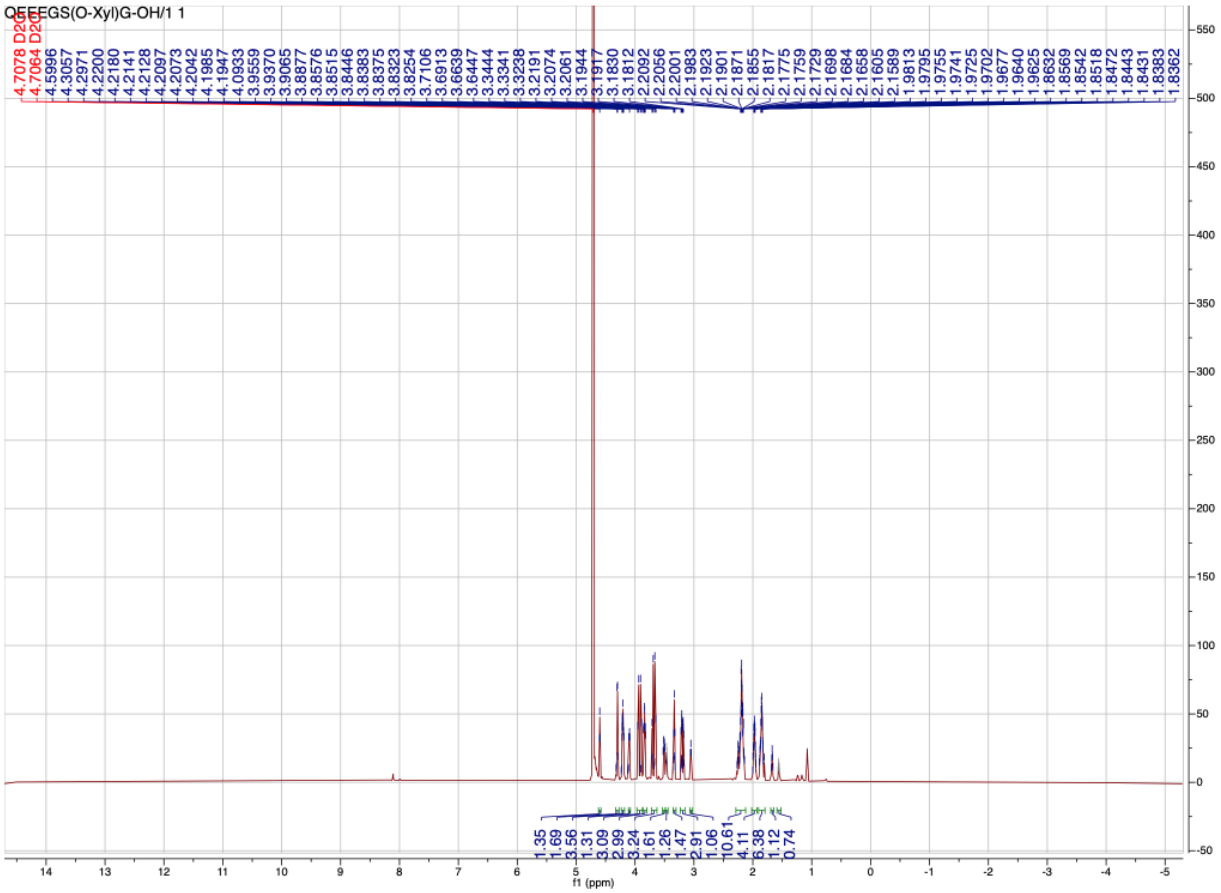
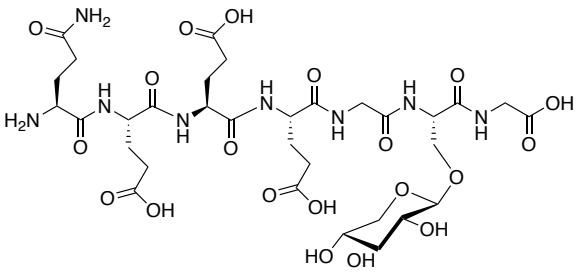


Figure 3.30 ^{1}H NMR of **6** (900 MHz, D_2O).



6

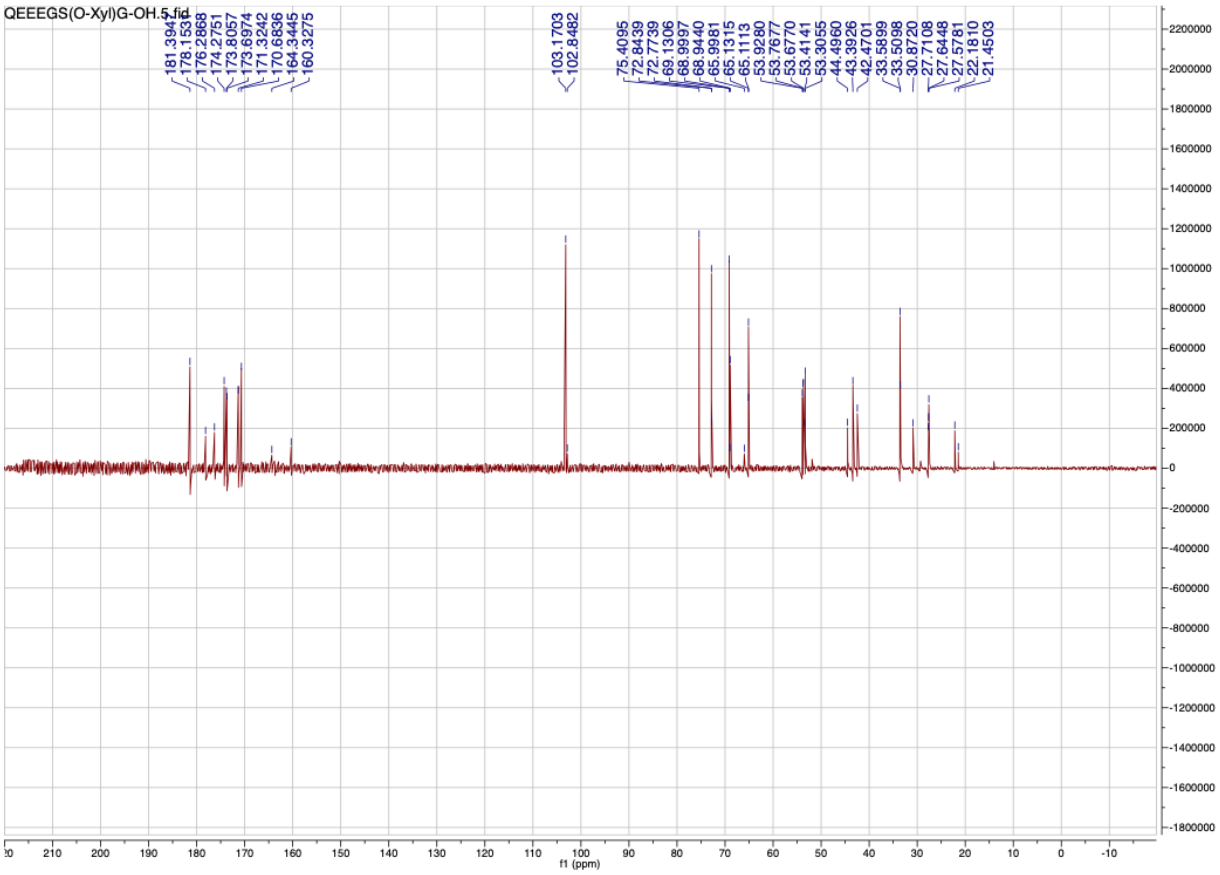
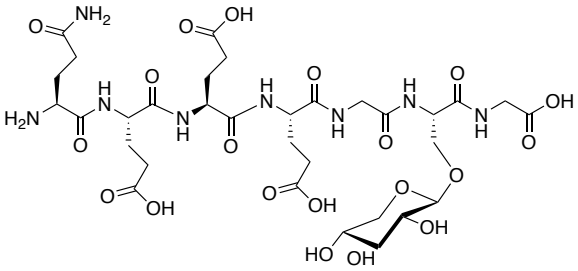


Figure 3.31 ^{13}C NMR of **6** (225 MHz, D_2O).



6

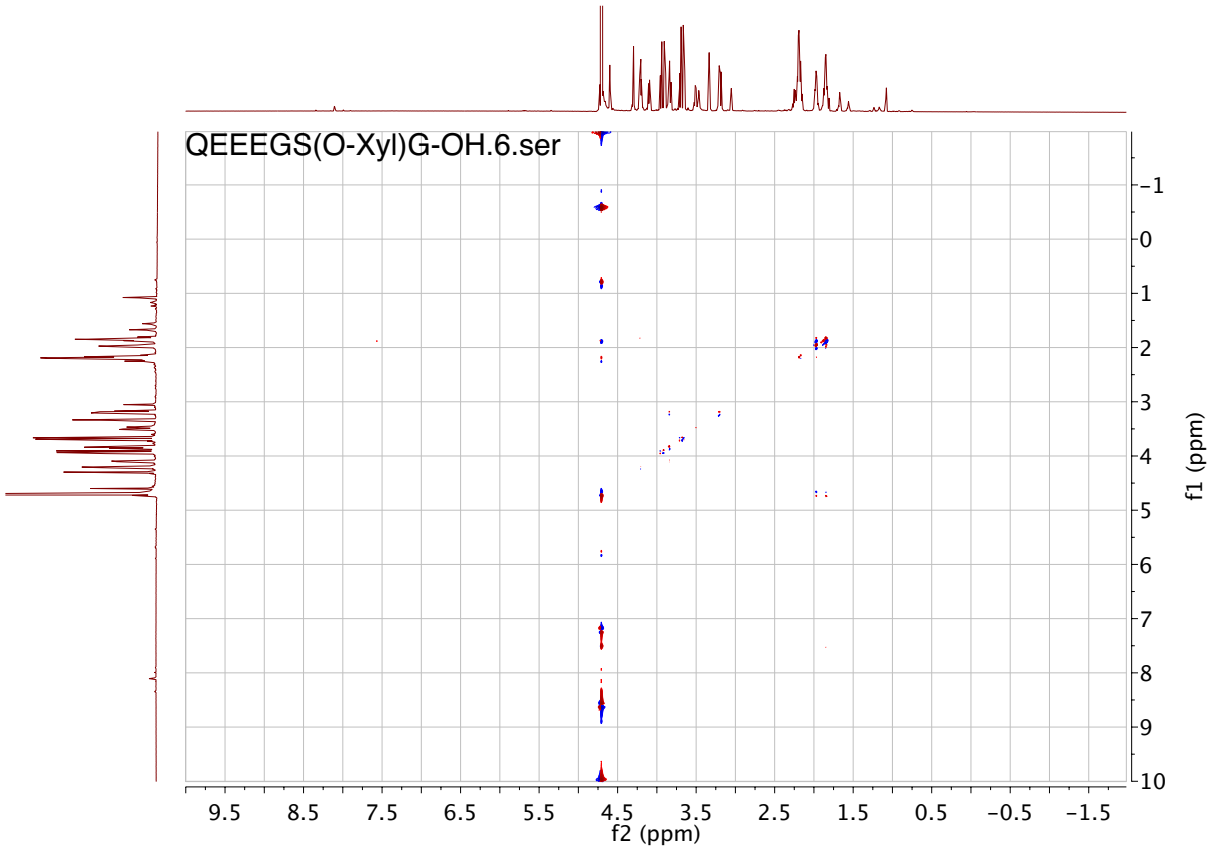
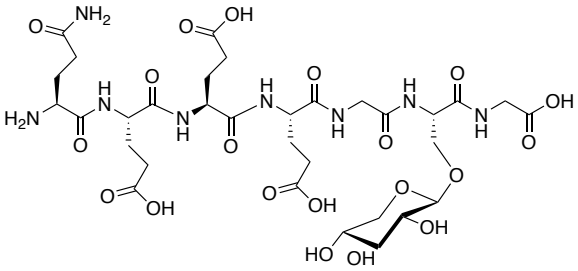


Figure 3.32 COSY NMR of **6** (900 MHz, D_2O).



6

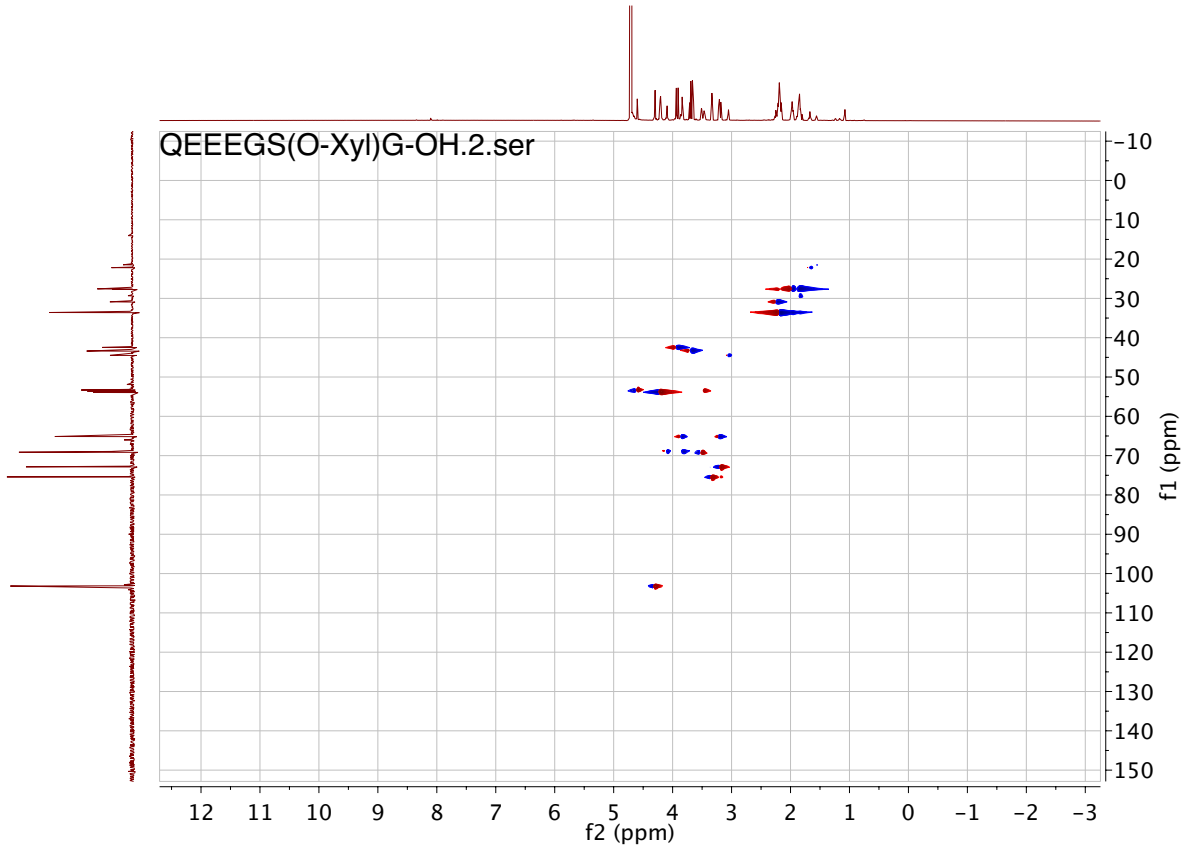
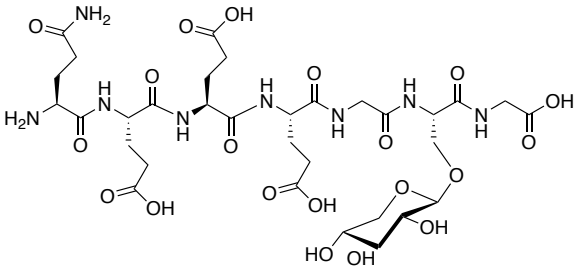
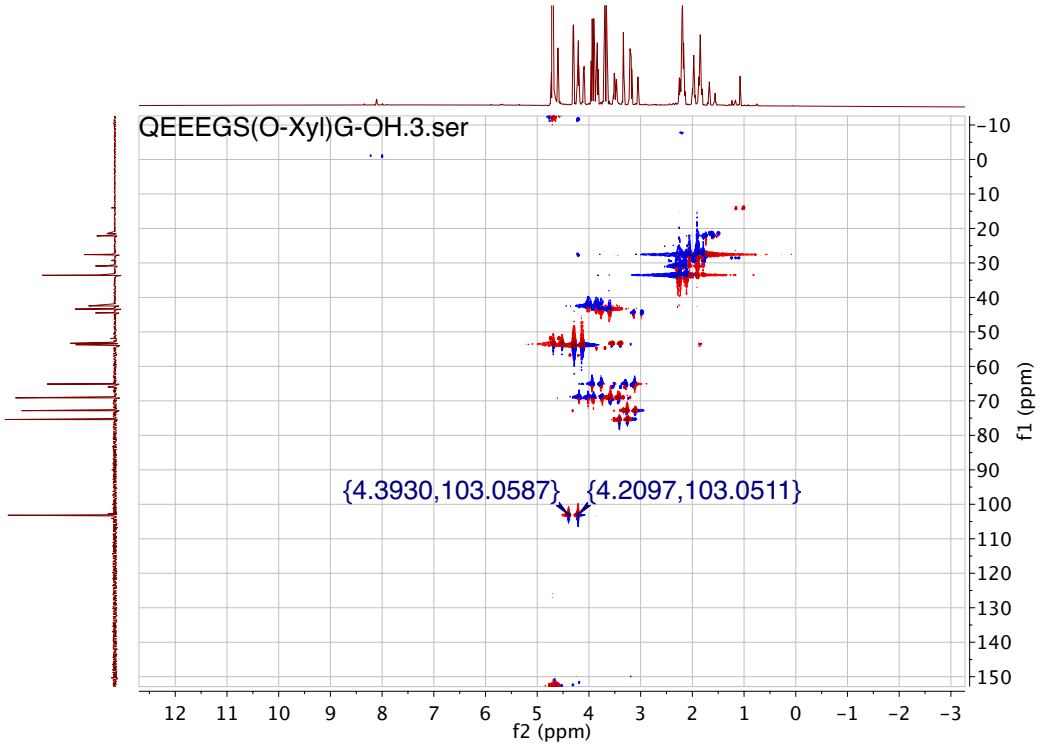


Figure 3.33 HSQC NMR of **6** (900 MHz, D_2O).



6



$^{1}J_{C1,\,H1}$ = 164.9 Hz

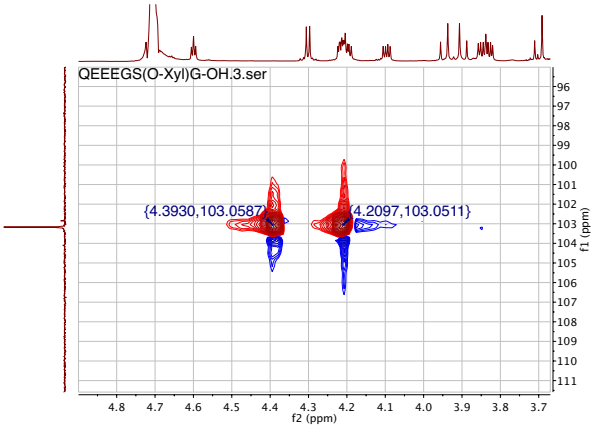
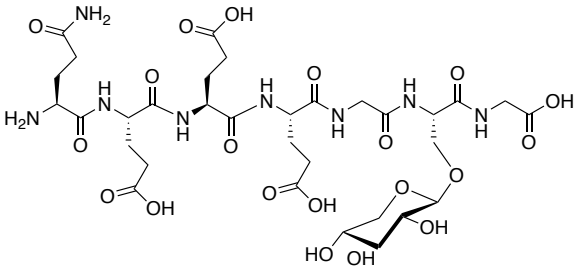


Figure 3.34 HSQC-coupled NMR of **6** (900 MHz, D_2O).



6

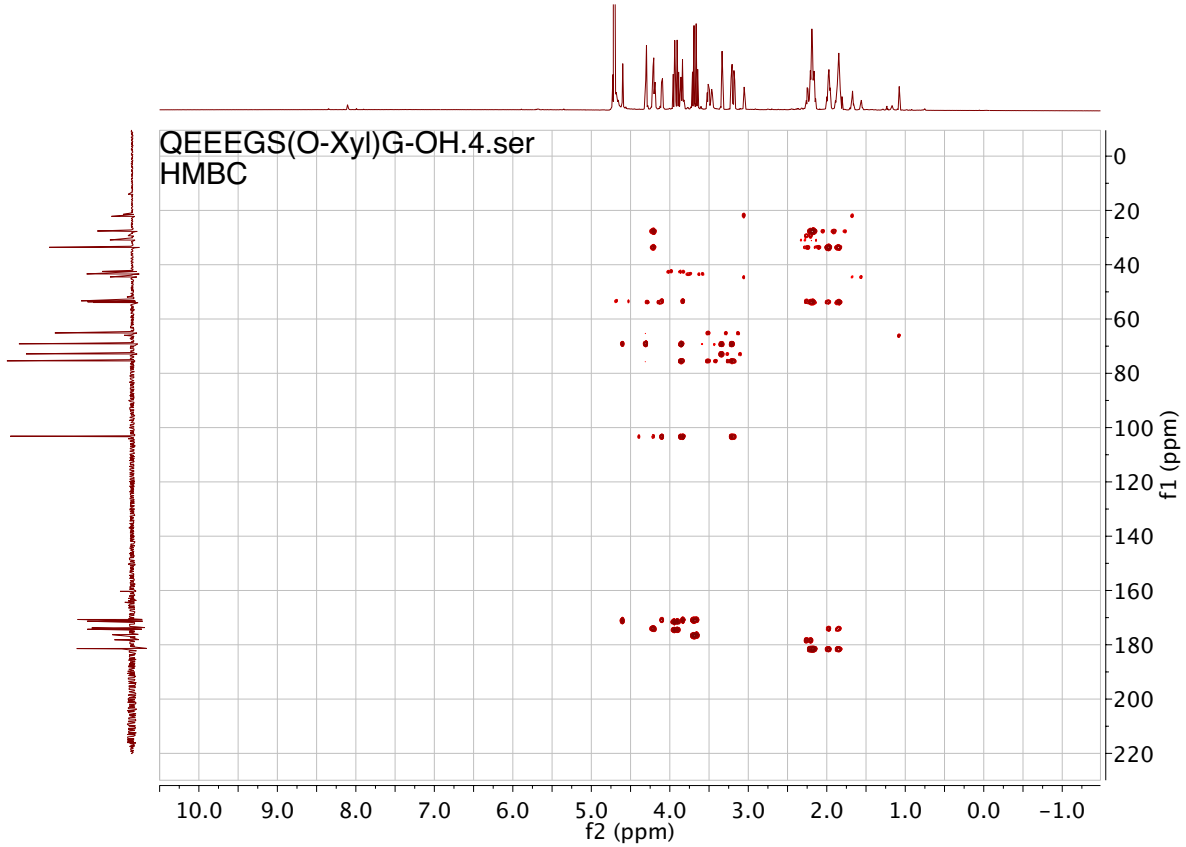
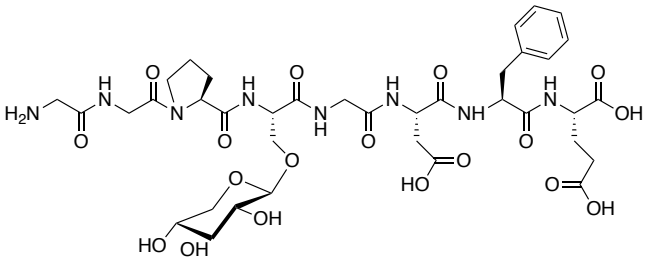
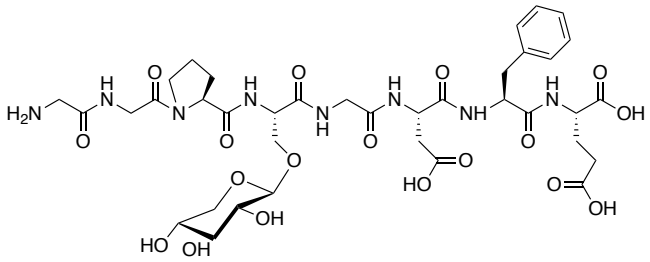


Figure 3.35 HMBC NMR of **6** (900 MHz, D_2O).



7

The purity of glycopeptide was verified with analytical C-18 HPLC (5-30% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = - 3.4 ^{o} (c = 0.16, H_2O). ^{1}H-NMR (900 MHz, D2O), δ 7.28 – 7.12 (m, 5H), 4.69 – 4.61 (m, 2H), 4.61 – 4.51 (m, 3H), 4.46 (dt, *J* = 8.7, 4.4 Hz, 1H), 4.37 (dd, *J* = 8.7, 5.1 Hz, 1H), 4.31 (m, 1H), 4.12 – 4.05 (m, 2H), 4.06 – 3.94 (m, 3H), 3.88 – 3.75 (m, 5H), 3.60 – 3.39 (m, 4H), 3.42 – 3.26 (m, 3H), 3.22-3.16 (m, 2H), 3.16 – 3.07 (m, 1H), 2.91 (m, 1H), 2.48 (m, 1H), 2.36 – 2.26 (m, 1H), 2.24 – 2.16 (m, 1H), 2.05 (t, *J* = 8.3 Hz, 3H), 1.99 – 1.85 (m, 4H), 1.85 – 1.72 (m, 2H); ^{13}C-NMR (225 MHz, D2O) δ 182.0, 178.0, 177.4, 174.5, 172.8, 171.9, 171.3, 170.4, 169.4, 163.1, 163.0, 162.8, 162.7, 160.5, 136.3, 129.2, 128.5, 126.9, 118.1, 116.8, 115.5, 114.2, 103.0, 75.4, 72.7, 69.0, 68.6, 65.0, 60.6, 55.1, 54.6, 53.4, 51.4, 46.9, 42.9, 42.3, 41.5, 38.2, 36.9, 33.8, 29.3, 28.5, 24.3. ESI-MS: $C_{37}H_{52}N_8O_{16}$ $[M+H]^+$ calcd: 897.3473, obsd: 897.3443 (3.34 ppm).



7

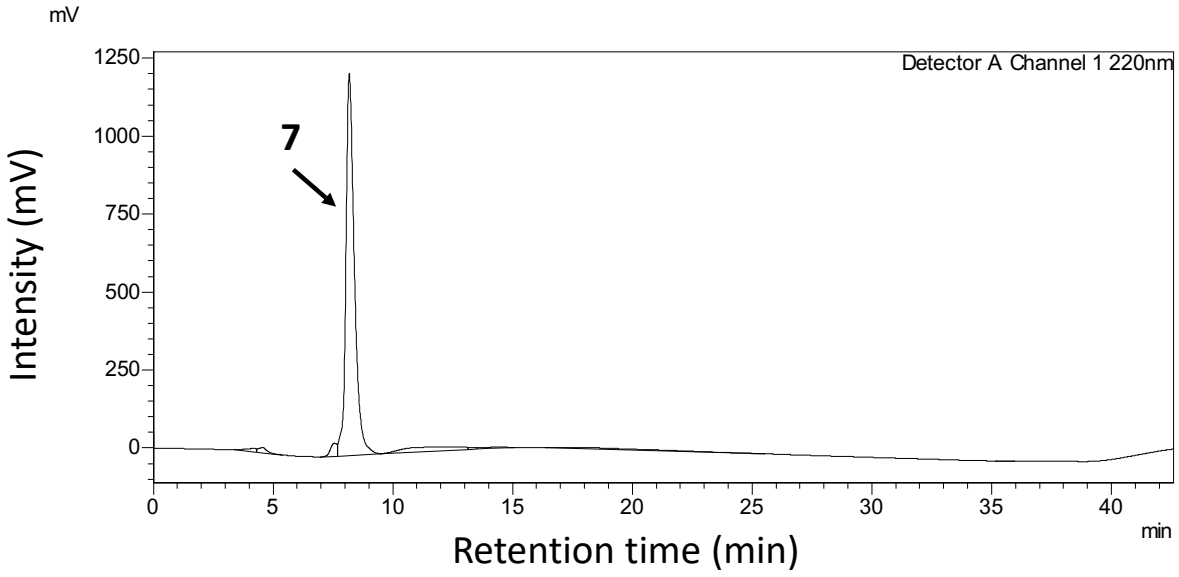
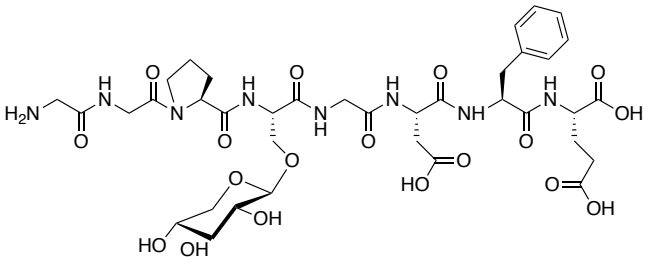


Figure 3.36 HPLC chromatogram of **7**.



7

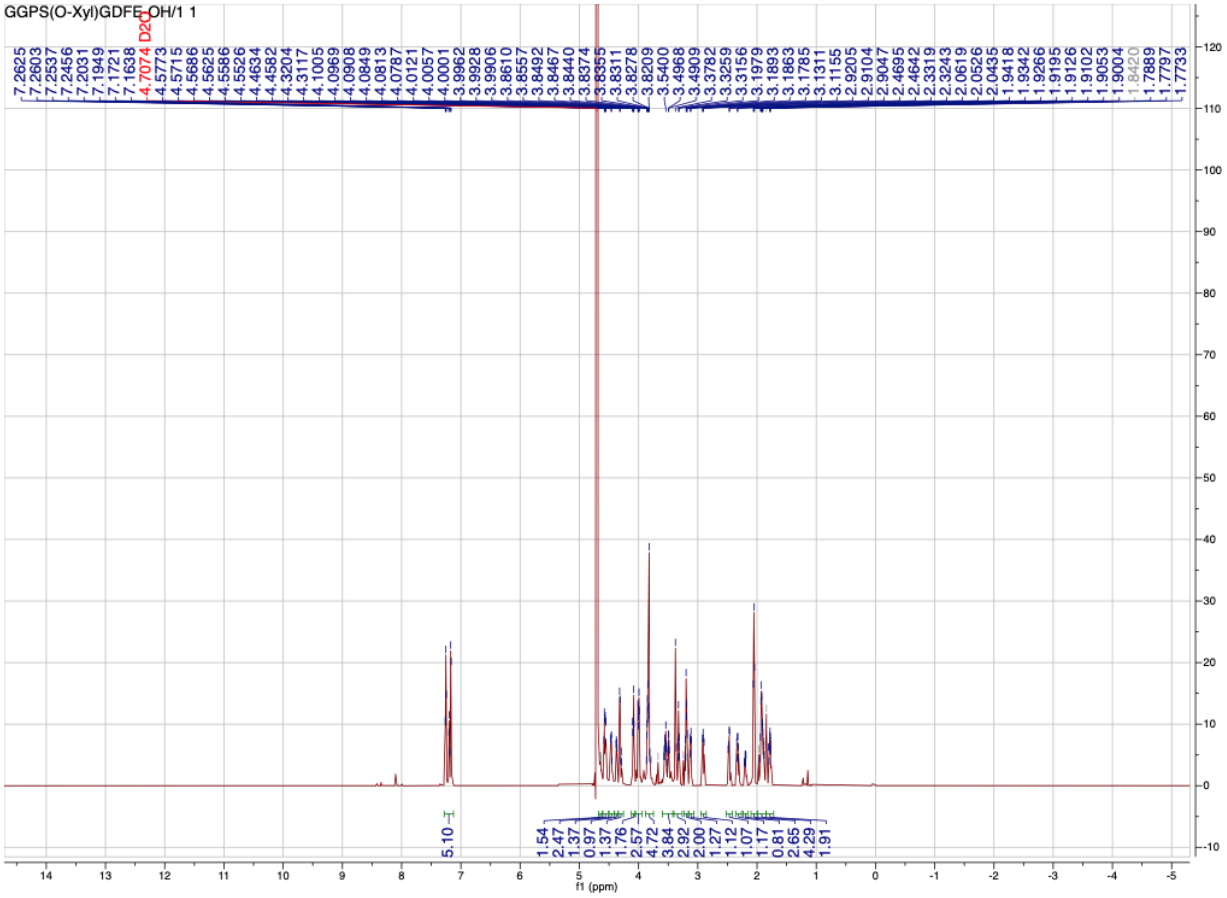
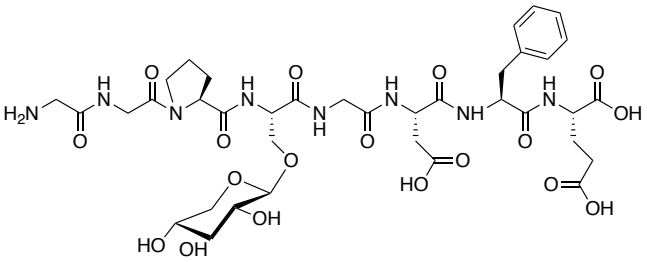


Figure 3.37 ^{1}H NMR of **7** (900 MHz, D_2O).



7

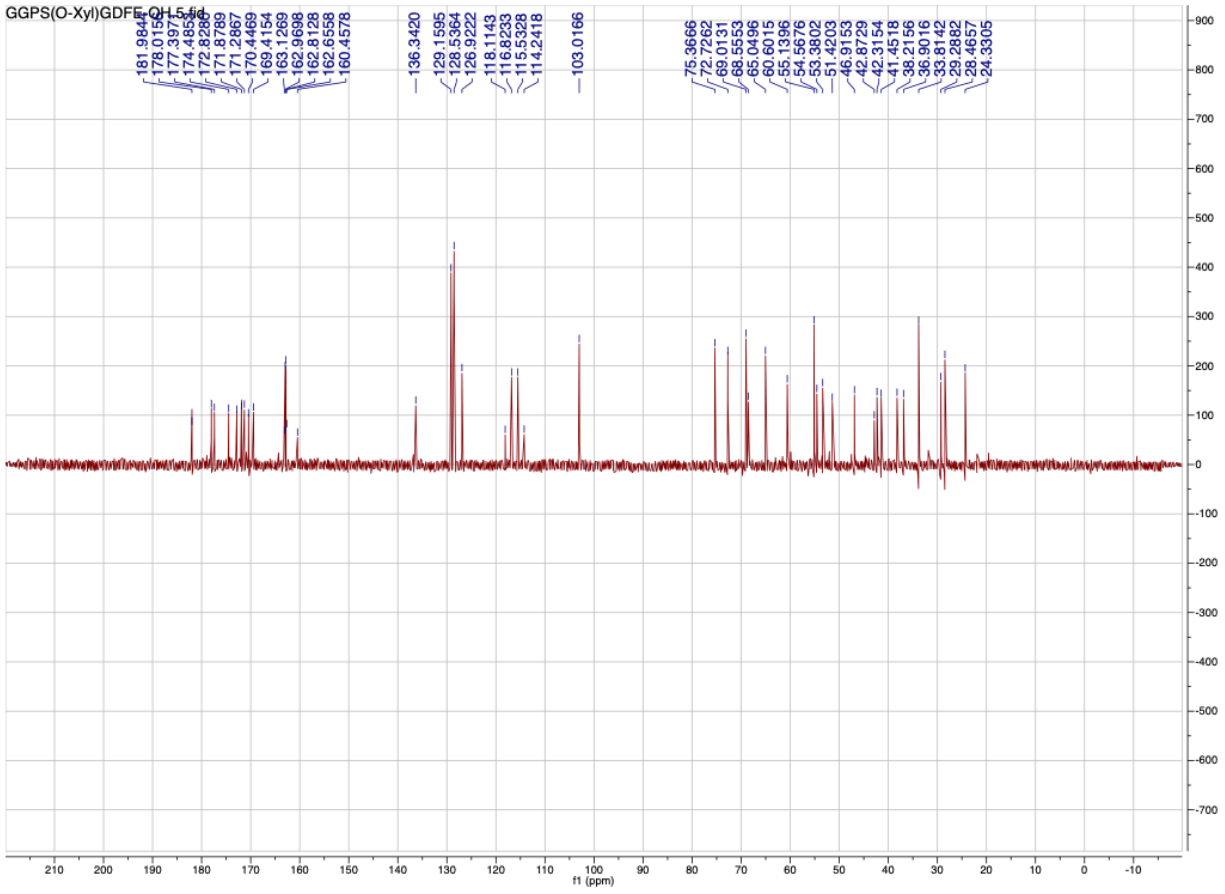
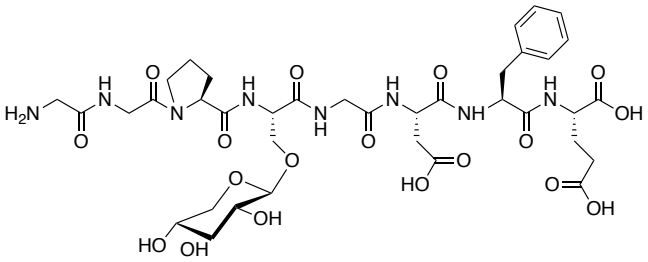
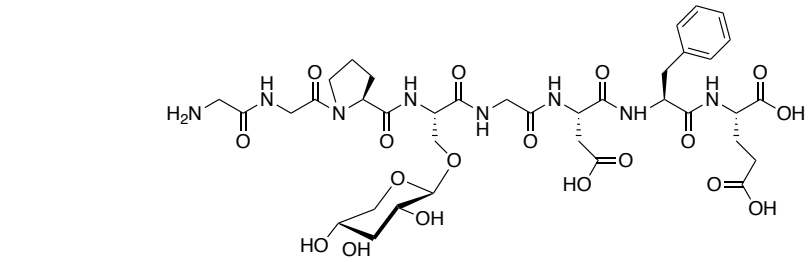


Figure 3.38 ^{13}C-NMR of **7** (225 MHz, D_2O).



7

Figure 3.39 COSY NMR of **7** (500 MHz, D_2O).



7

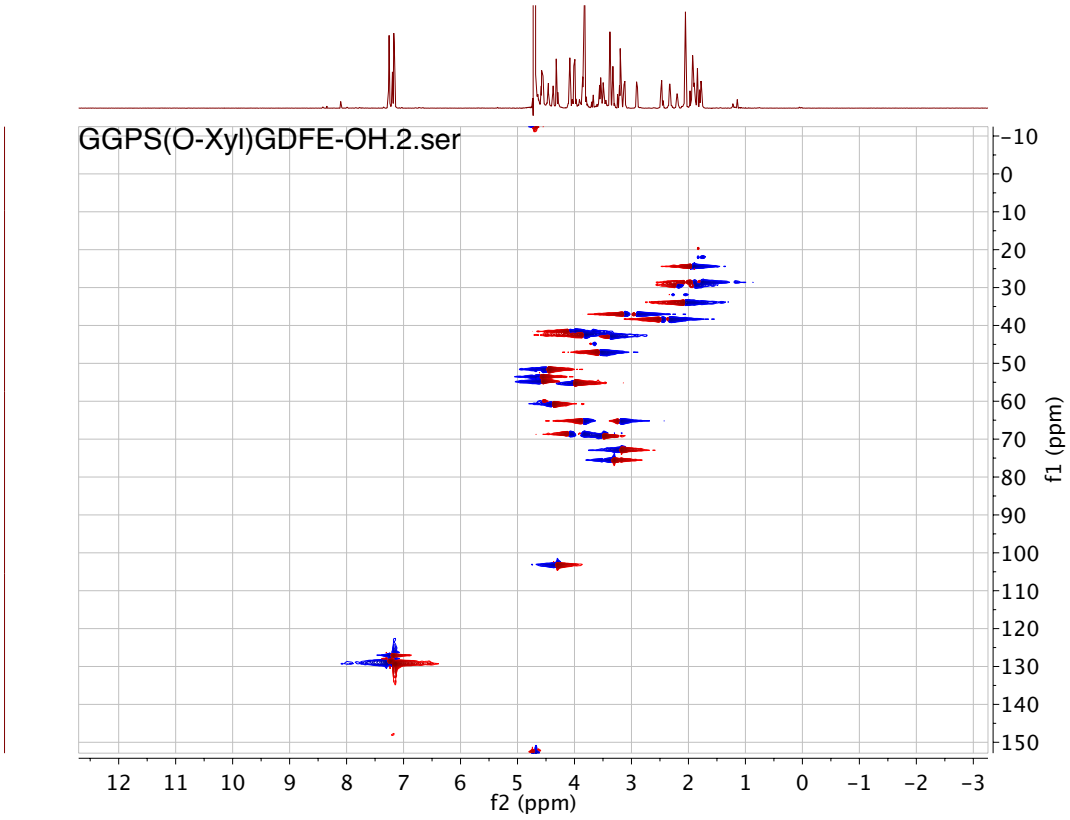
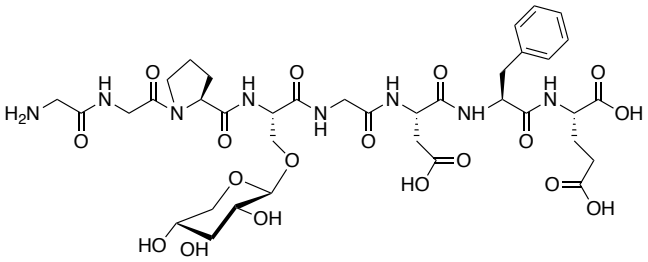
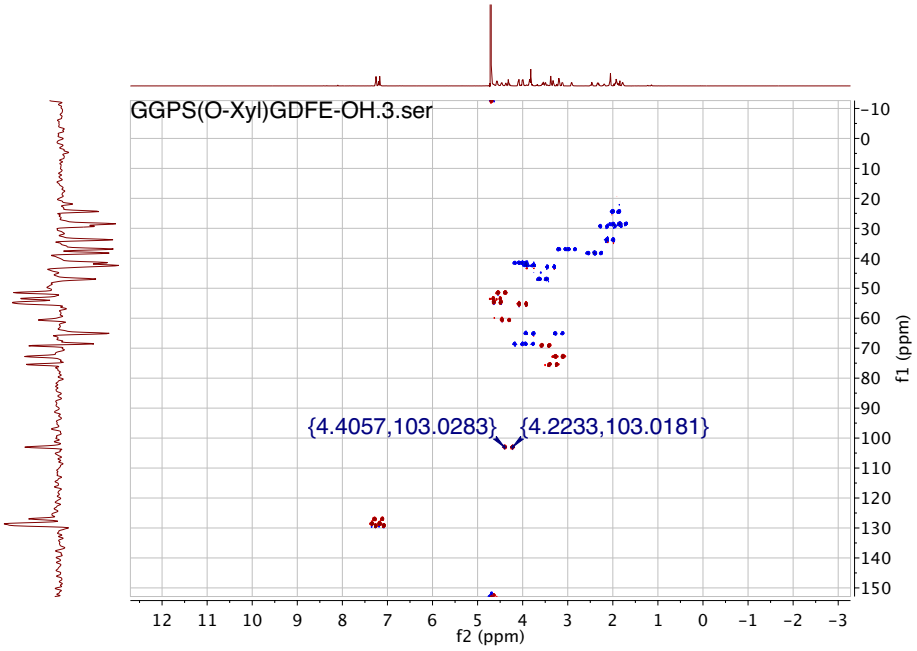


Figure 3.40 HSQC NMR of **7** (900 MHz, D_2O).



7



$^{1}J_{C1,\,H1}$ = 164.2 Hz

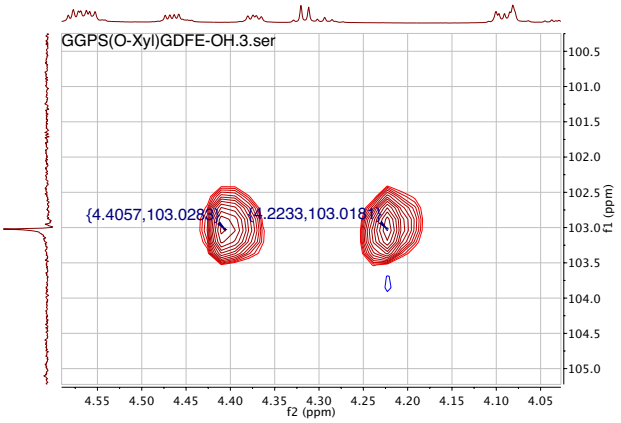
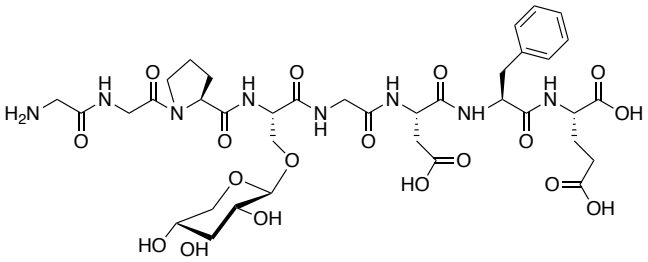


Figure 3.41 HSQC-coupled NMR of **7** (900 MHz, D_2O).



7

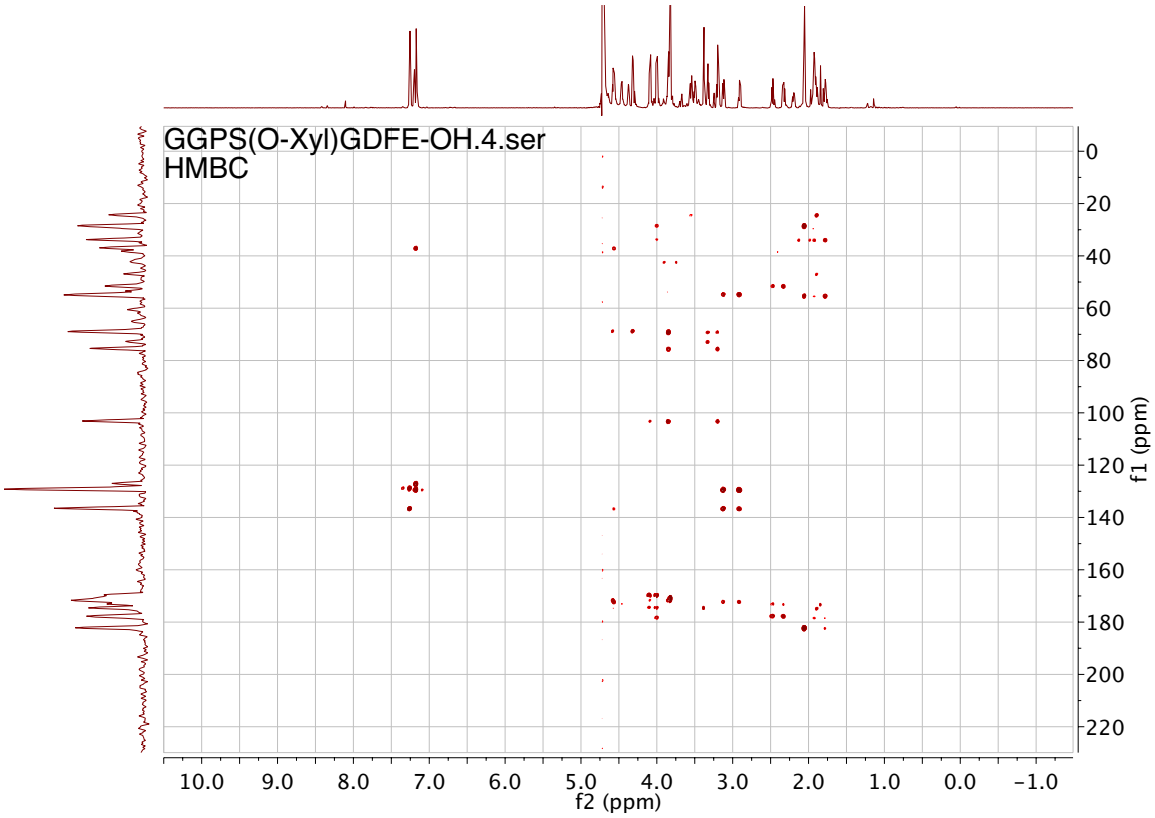
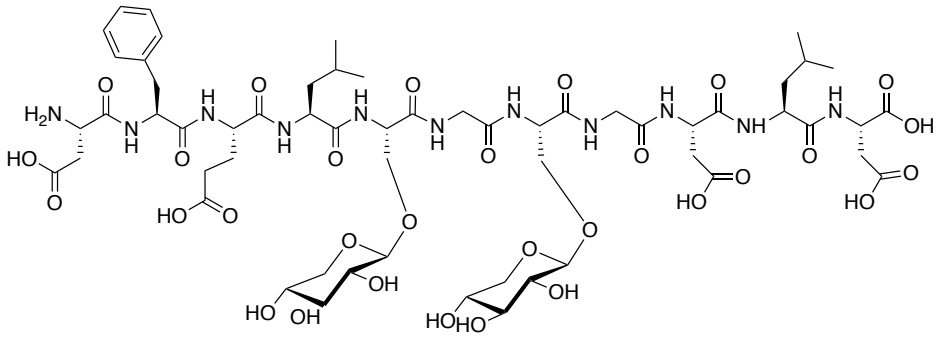
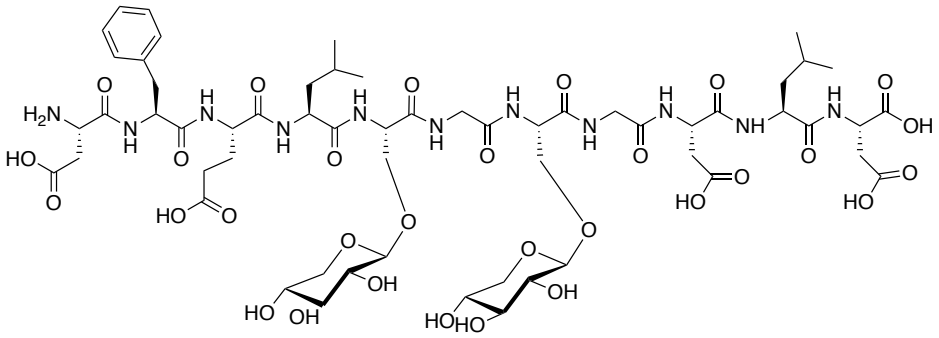


Figure 3.42 HMBC NMR of **7** (900 MHz, D_2O).



8

The purity of glycopeptide was verified with analytical C-18 HPLC (0-30-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = + 0.01 ° (c = 0.10, H_2O). ^{1}H-NMR (500 MHz, D_2O) δ 7.26 – 7.01 (m, 5H), 4.51-4.46 (m, 3H), 4.30 (t, *J* = 6.3 Hz, 1H), 4.27 – 4.15 (m, 3H), 4.03-3.98 (m, 2H), 3.92 – 3.67 (m, 7H), 3.27 (t, *J* = 9.2 Hz, 1H), 3.20 – 3.06 (m, 3H), 2.58-2.54 (m, 7H), 2.15 (t, *J* = 7.4 Hz, 2H), 1.50-1.46 (m, 7H), 0.92 – 0.40 (m, 14H); ^{13}C-NMR (125 MHz, D_2O) δ 175.6, 163.1, 162.8, 162.8, 129.0, 129.0, 129.0, 128.7, 128.7, 128.7, 119.7, 117.4, 115.1, 115.1, 103.0, 75.4, 72.8, 72.7, 69.1, 69.0, 65.1, 50.4, 24.2, 24.1, 22.3, 22.2, 20.8, 20.8, 20.4. ESI-MS: $C_{58}H_{87}N_{11}O_{30}$ $[M+2H]^{2+}$ calcd: 1418.5693, obsd: 1418.5635 (4.09 ppm).



8

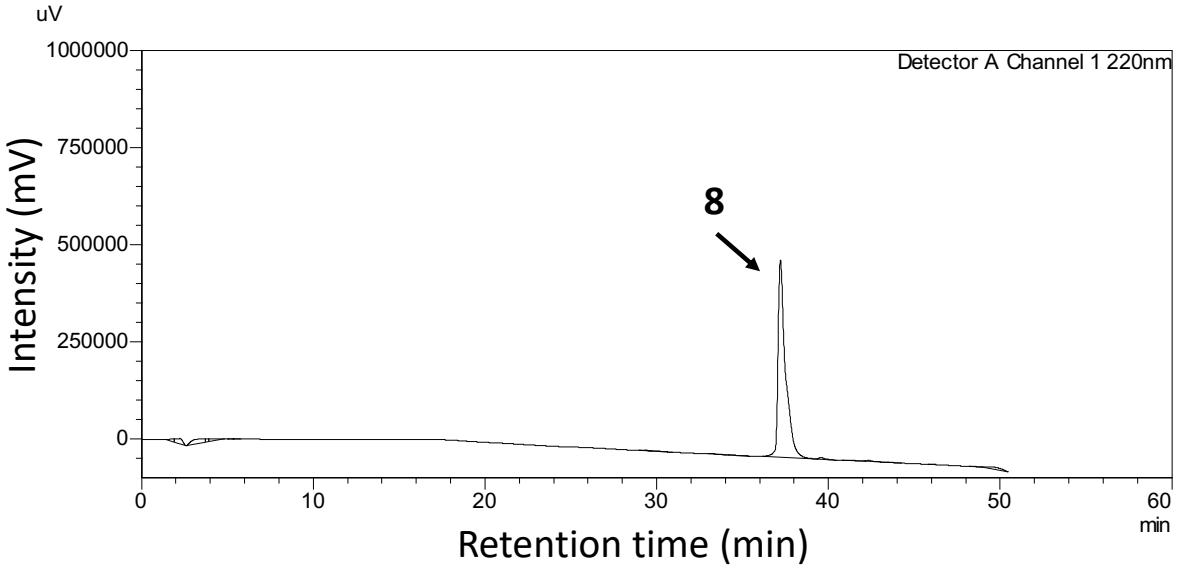
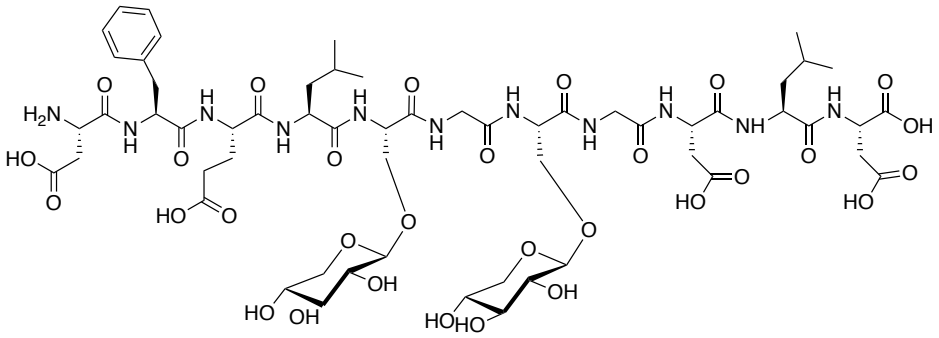


Figure 3.43 HPLC chromatogram of **8**.



8

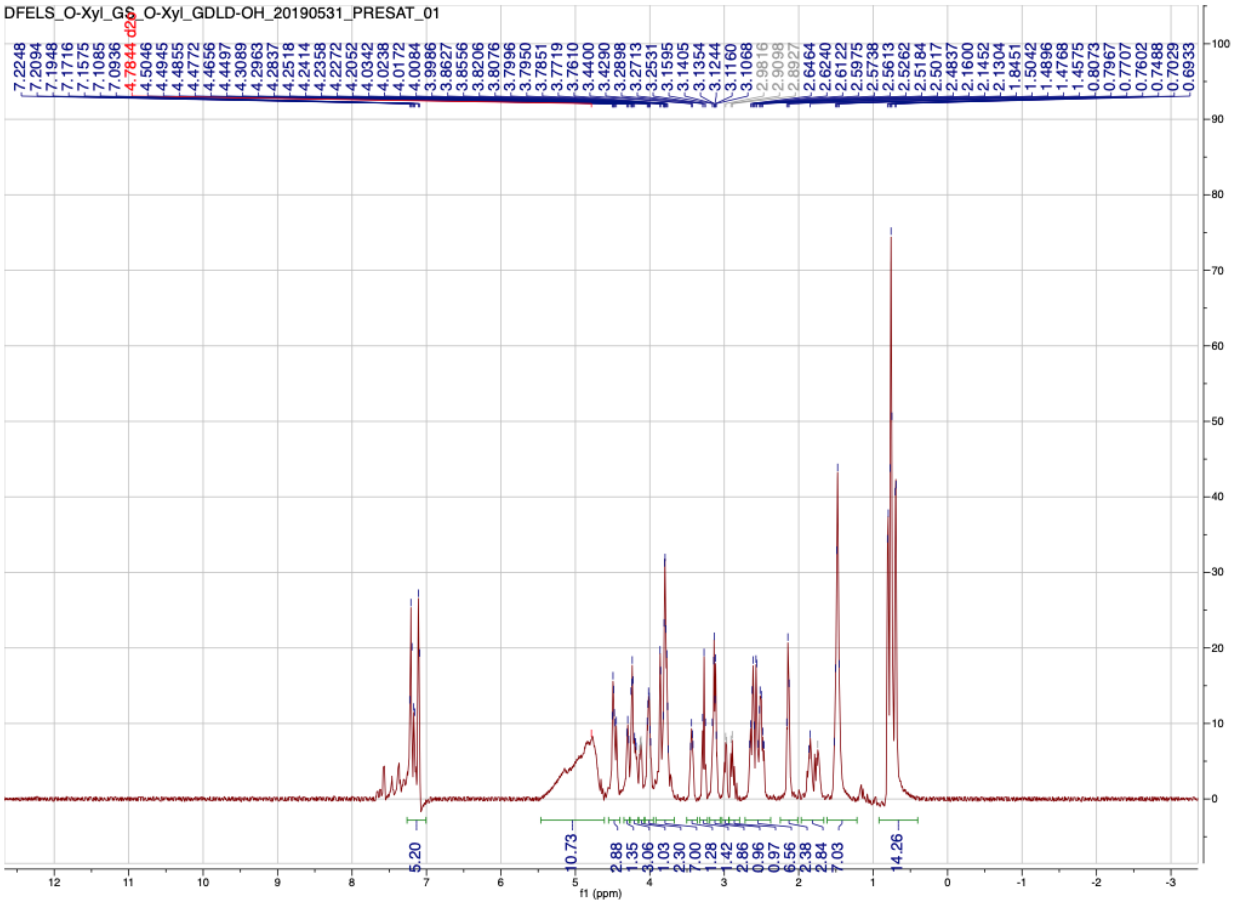
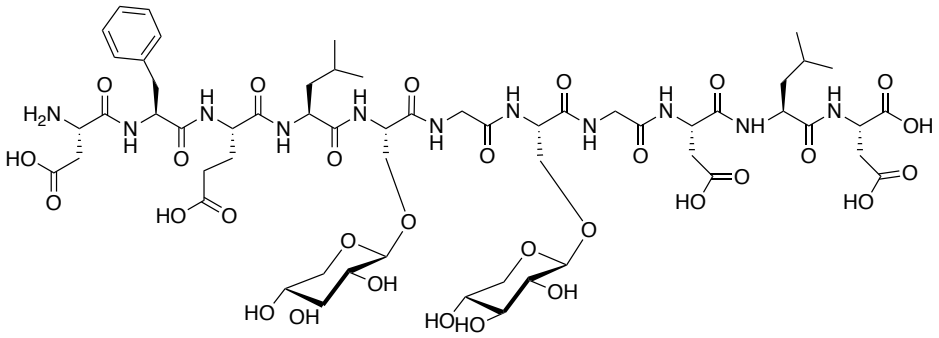


Figure 3.44 ^{1}H-NMR of **8** (500 MHz, D_2O).



8

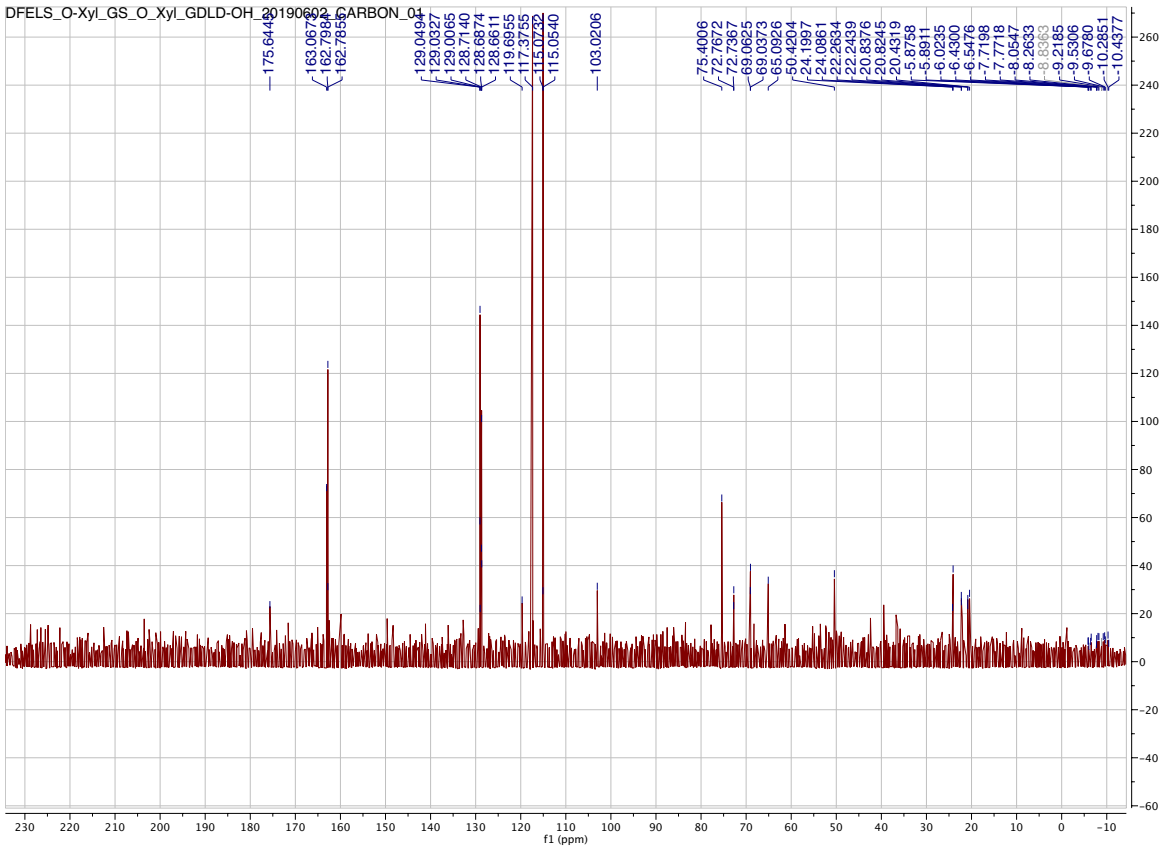
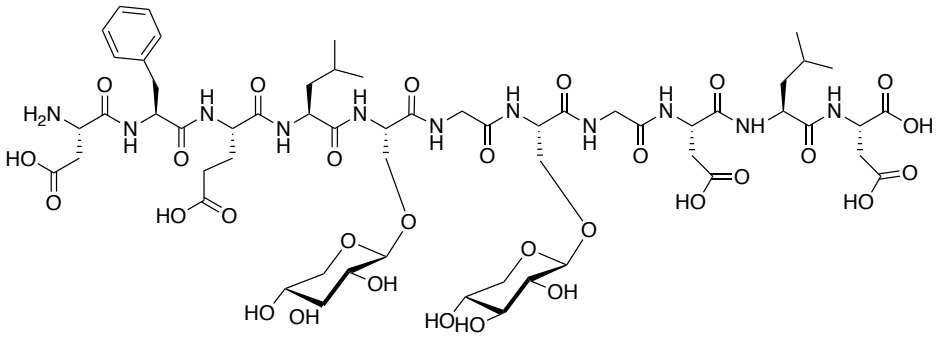


Figure 3.45 ^{13}C-NMR of **8** (125 MHz, D_2O).



8

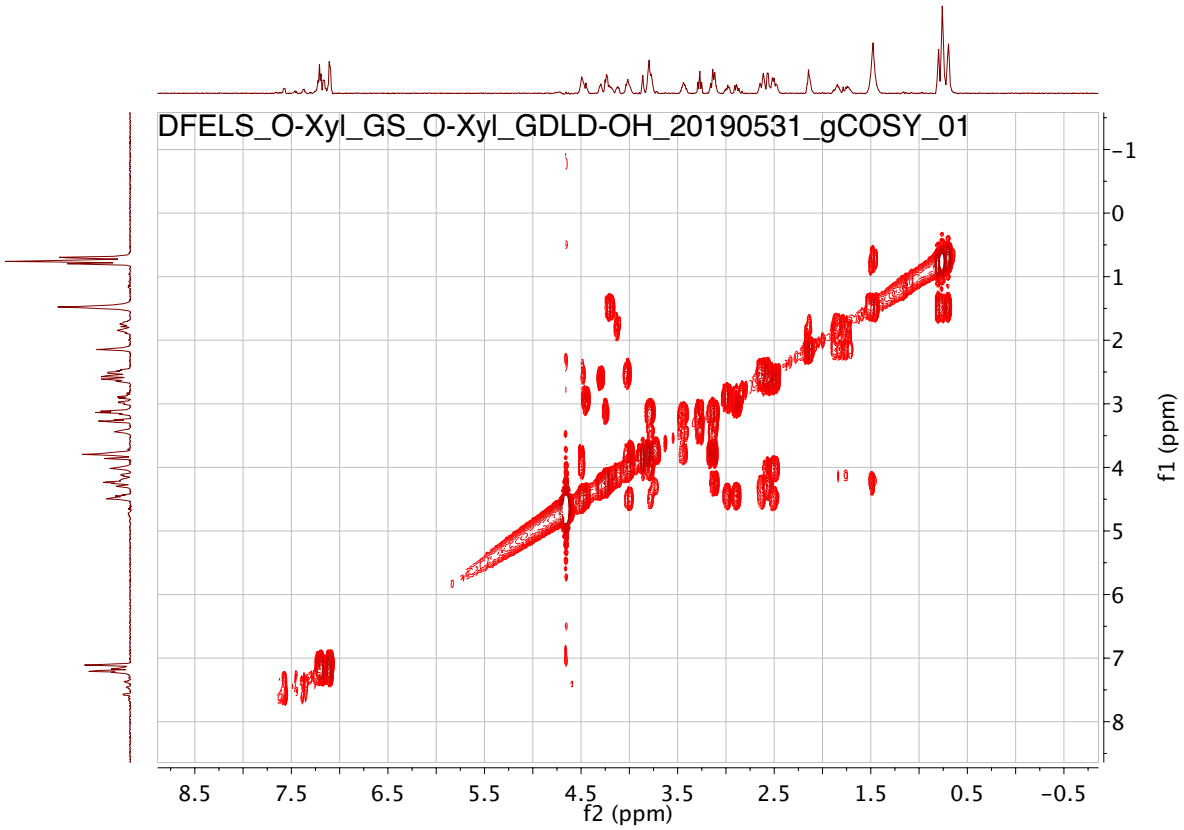
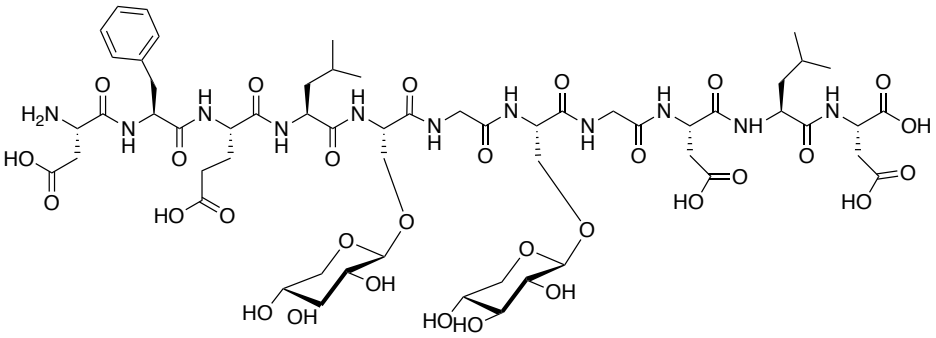


Figure 3.46 COSY NMR of **8** (500 MHz, D_2O).



8

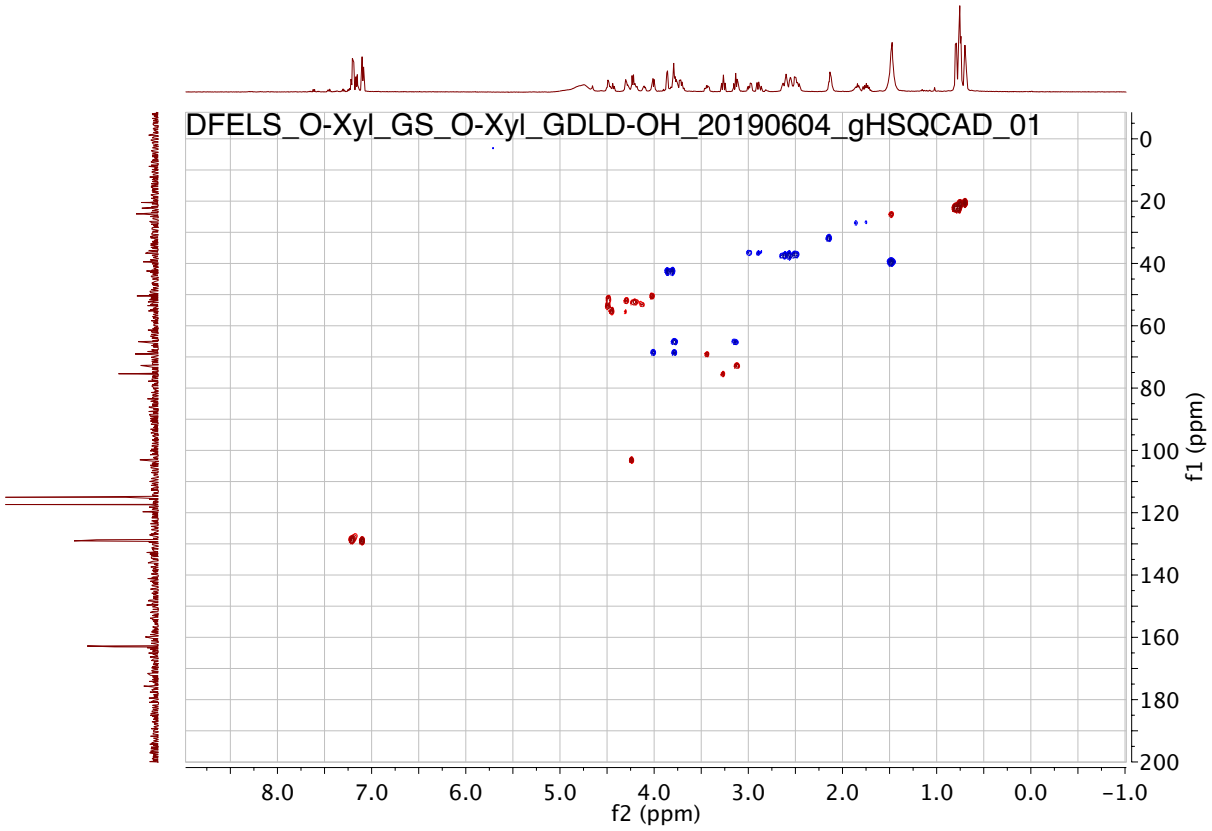
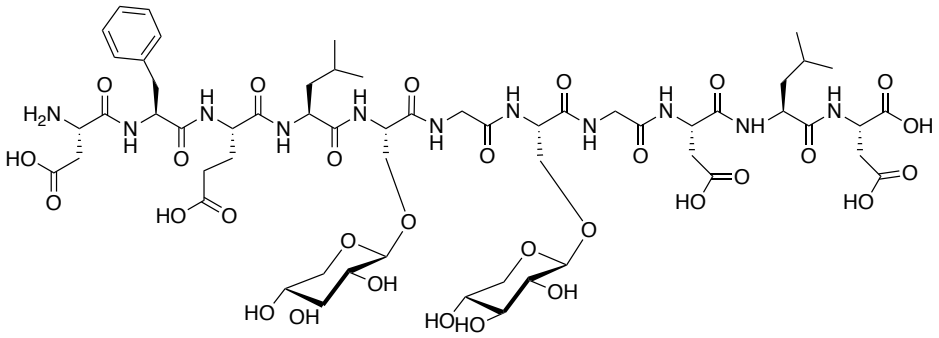
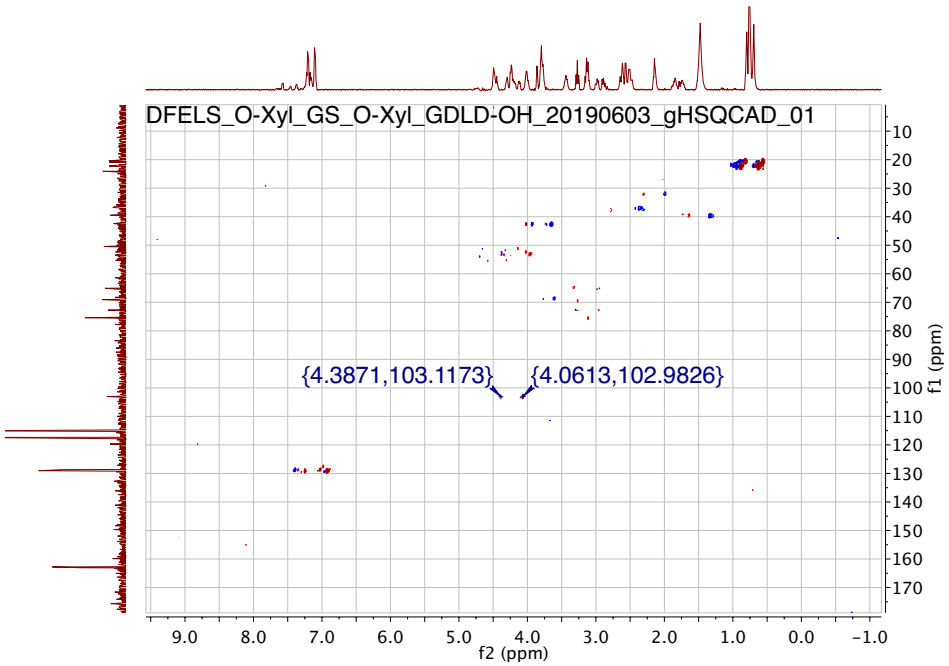


Figure 3.47 HSQC NMR of **8** (500 MHz, D_2O).



8



$^{1}J_{C1,H1}$ = 162.9 Hz

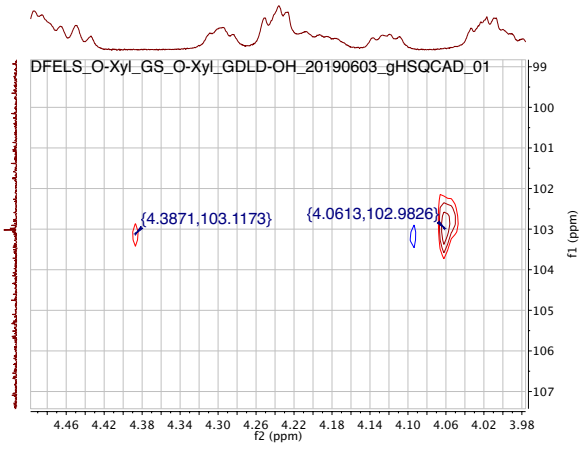
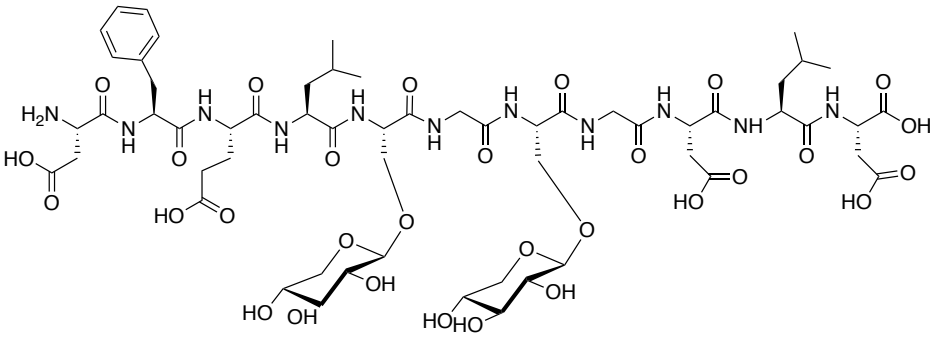


Figure 3.48 HSQC NMR of **8** (500 MHz, D_2O).



8

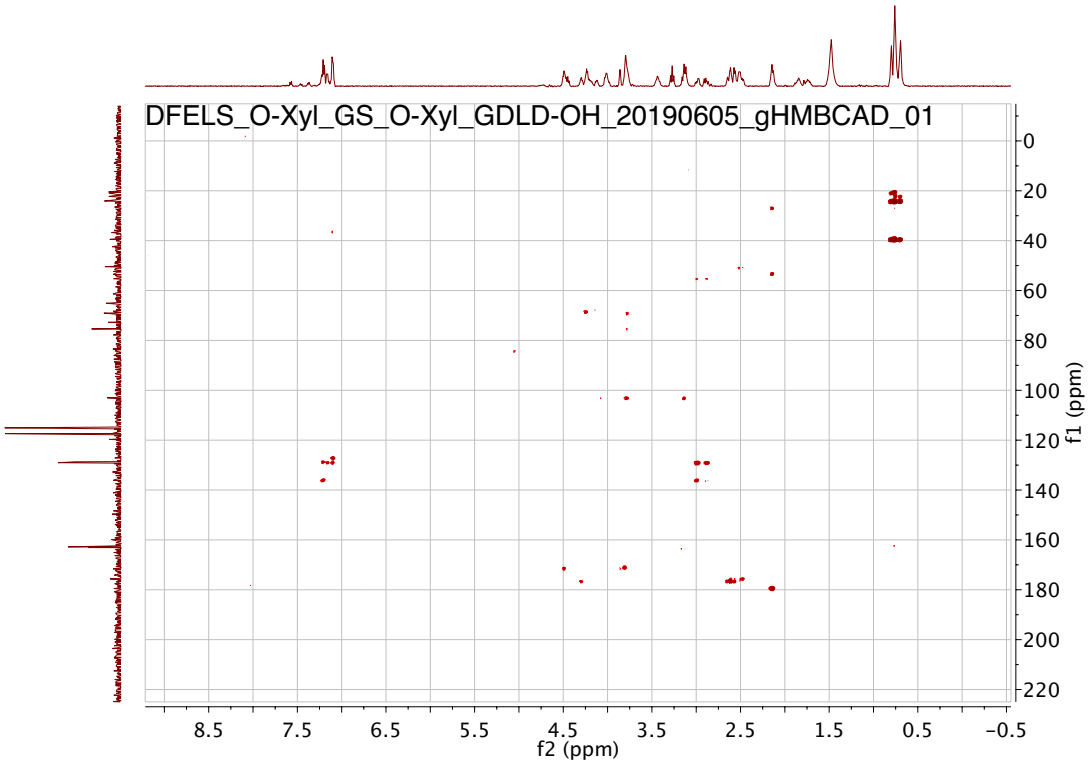
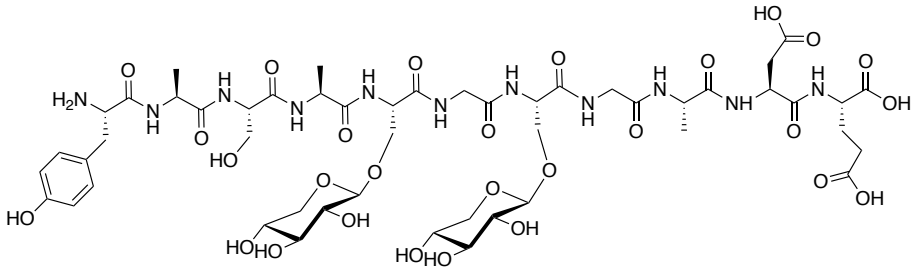
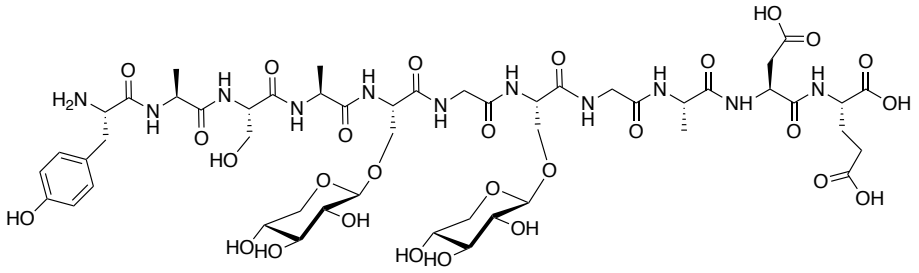


Figure 3.49 HMBC NMR of **8** (500 MHz, D_2O).



9

The purity of glycopeptide was verified with analytical C-18 HPLC (0-30-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20} = -0.003\ ^{o}$ (c = 0.06, H_2O). ^{1}H-NMR (500 MHz, D_2O), δ 7.06 – 6.93 (m, 2H), 6.73-6.68 (m, 2H), 4.58 – 4.40 (m, 3H), 4.35 – 4.12 (m, 6H), 4.11 – 3.94 (m, 4H), 3.93 – 3.67 (m, 12H), 3.43 (d, *J* = 8.8 Hz, 2H), 3.27 (m, 2H), 3.20 – 3.08 (m, 5H), 3.08 – 2.87 (m, 3H), 2.64 – 2.59 (m, 2H), 2.48 – 2.44 (m, 2H), 2.11 (t, *J* = 7.9 Hz, 3H), 1.98 – 1.86 (m, 2H), 1.82 – 1.67 (m, 2H), 1.30 – 1.18 (m, 14H); ^{13}C-NMR (125 MHz, D_2O) δ 174.6, 172.3, 159.9, 142.0, 131.8, 130.8, 130.8, 122.3, 115.7, 108.9, 82.7, 74.3, 69.0, 65.1, 60.5, 49.4, 42.7, 30.1, 18.0, 17.4, 16.6. ESI-MS: $C_{55}H_{83}N_{11}O_{33}$ $[M+H]^{2+}$ calcd: 1278.4856, obsd: 1278.4772 (6.57 ppm).



9

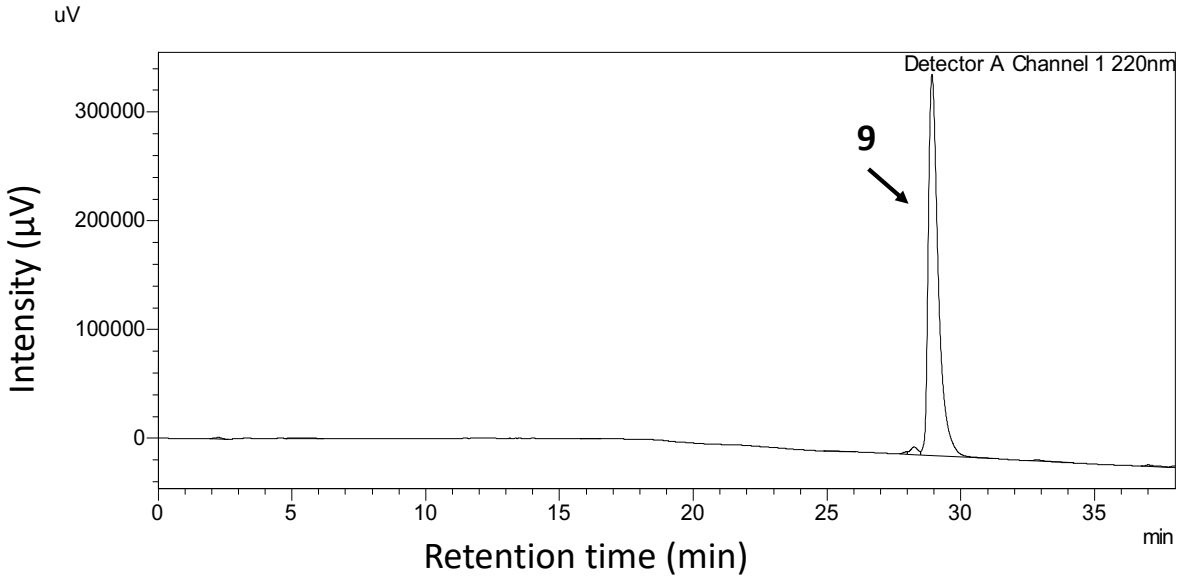
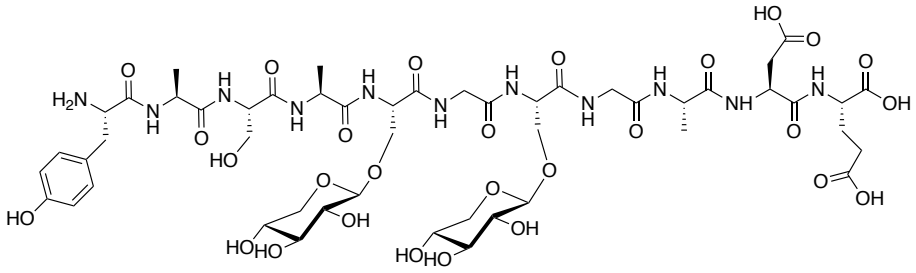


Figure 3.50 HPLC chromatogram of **9**.



9

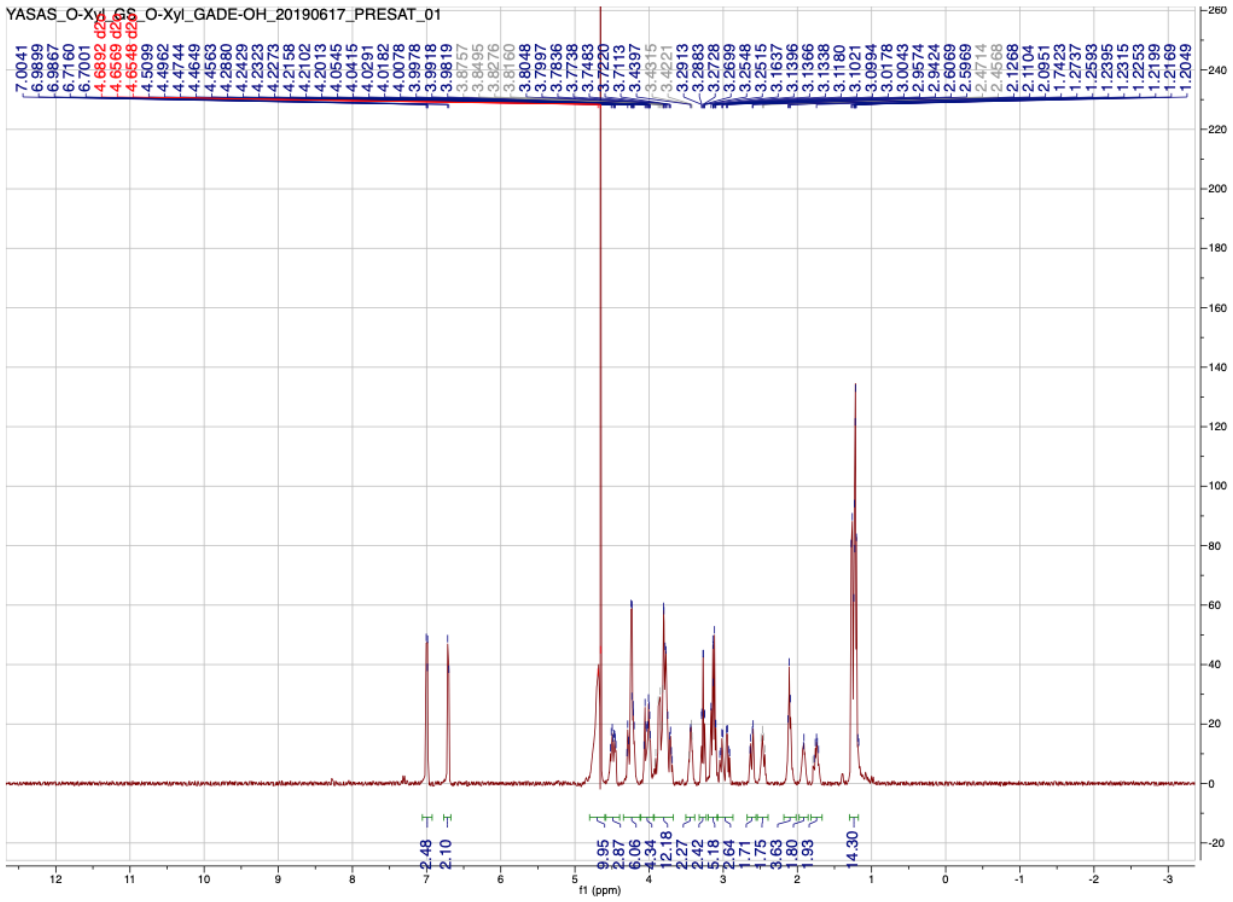
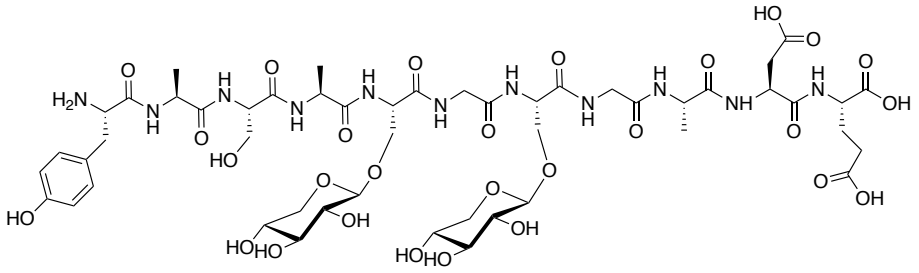


Figure 3.51 ^{1}H-NMR of **9** (500 MHz, D_2O).



9

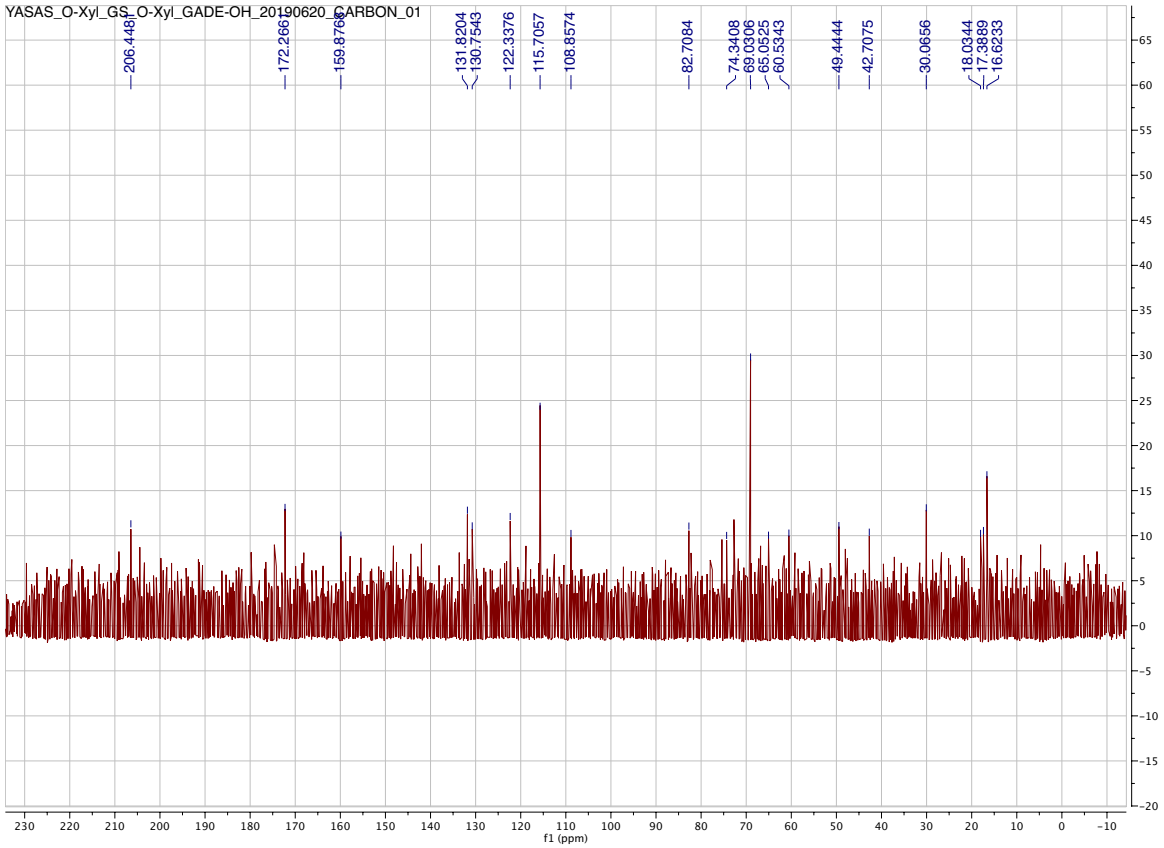
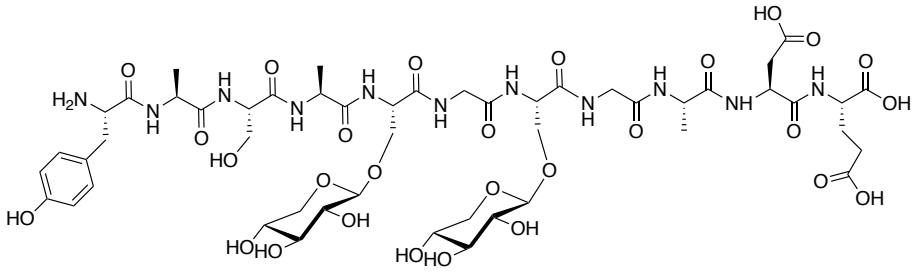


Figure 3.52 ^{13}C-NMR of **9** (125 MHz, D_2O).



9

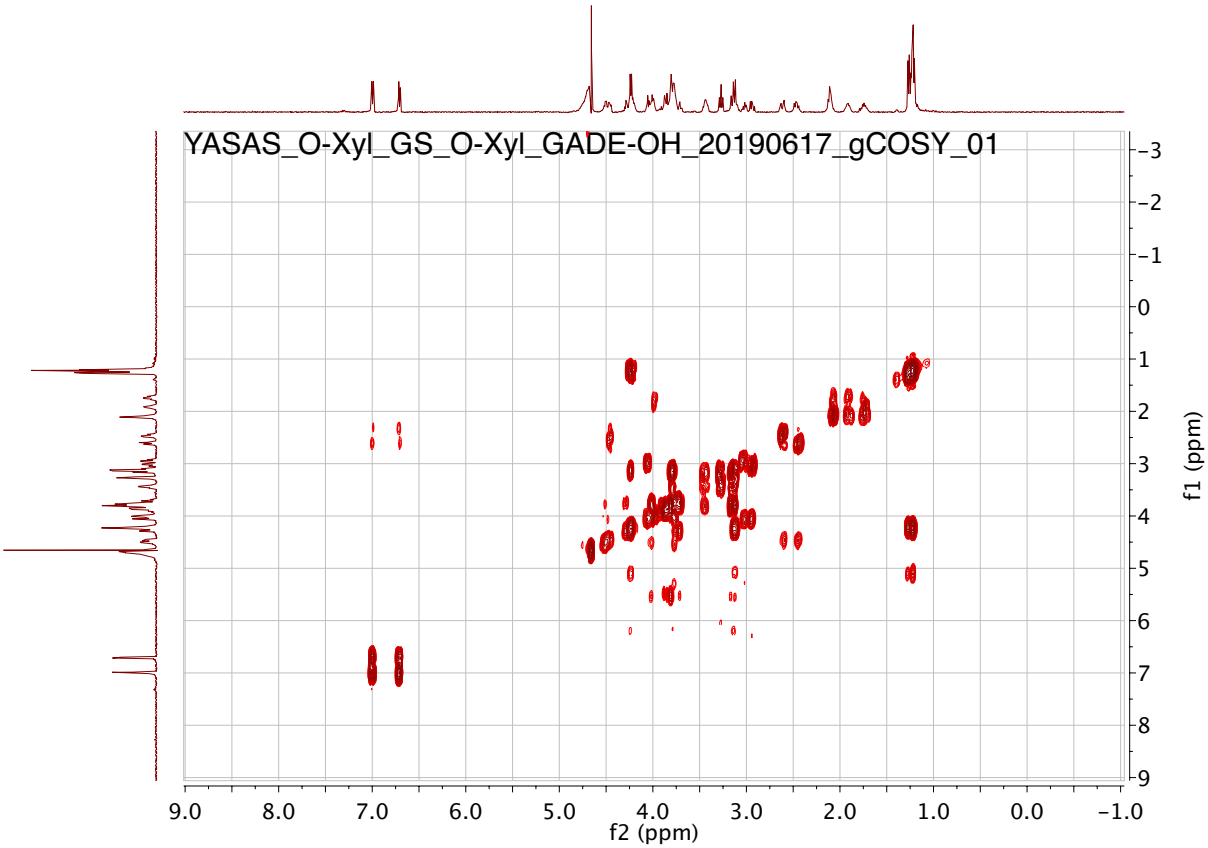
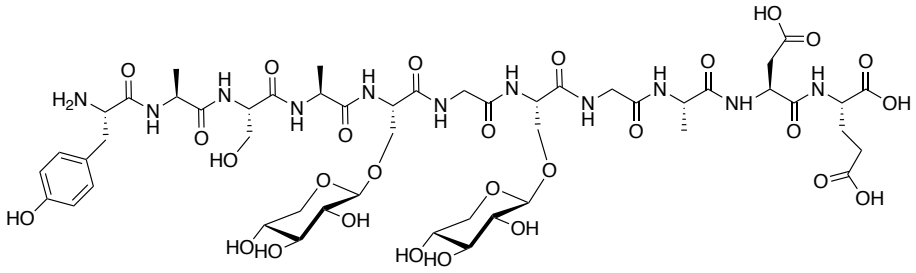


Figure 3.53 COSY NMR of **9** (500 MHz, D_2O).



9

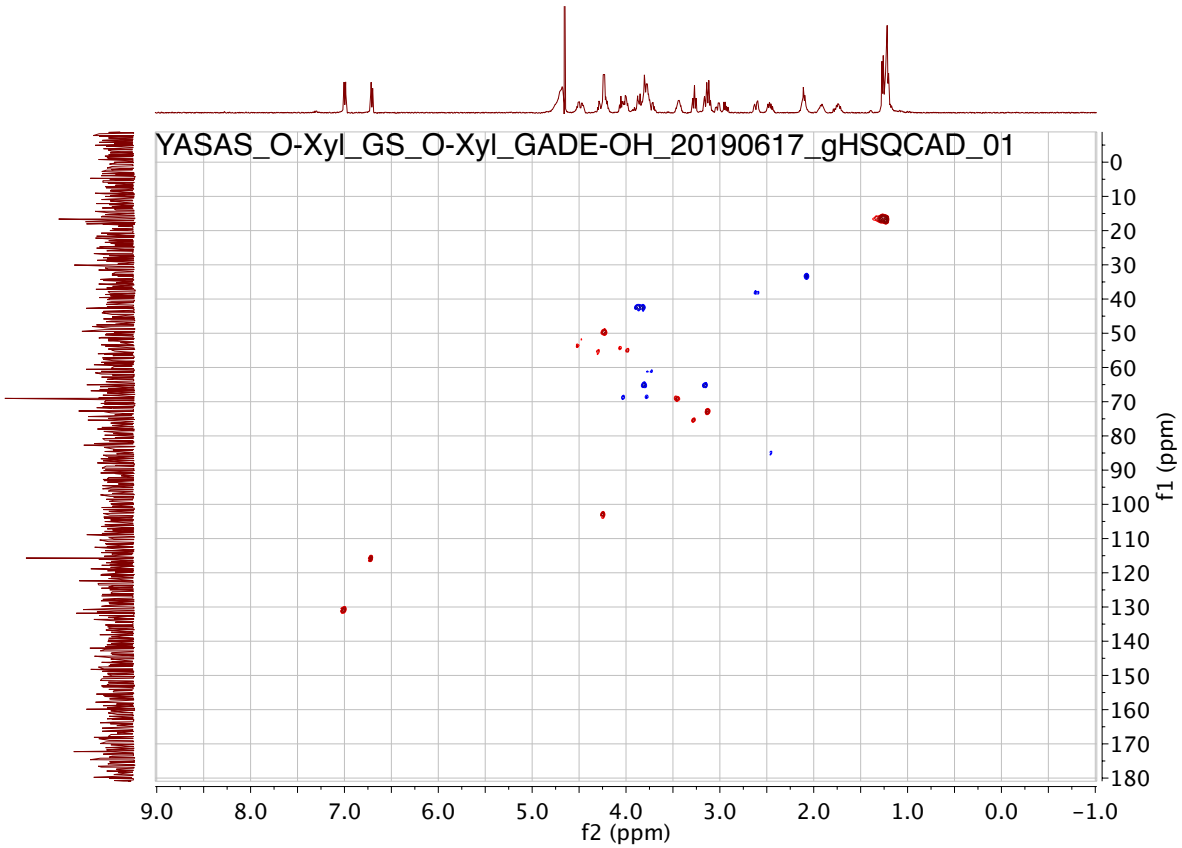
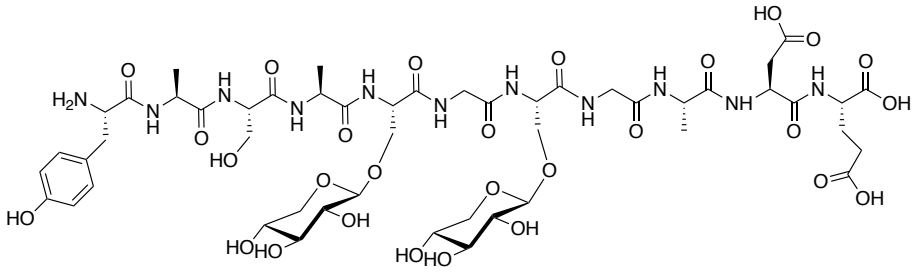
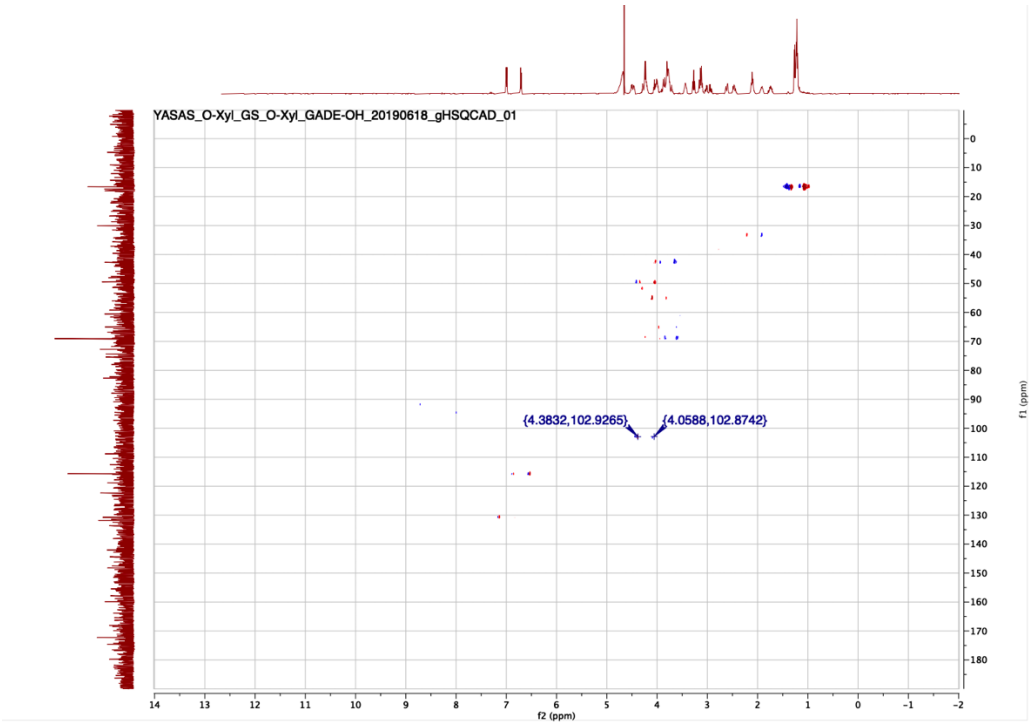


Figure 3.54 HSQC NMR of **9** (500 MHz, D_2O).



9



$^{1}J_{C1,H1}$ = 162.2 Hz

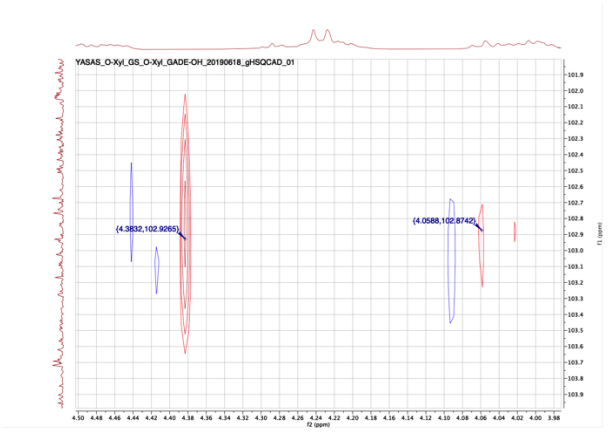
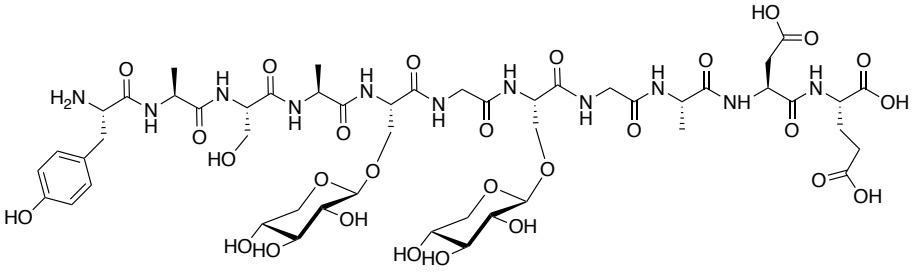


Figure 3.55 HSQC-coupled NMR of **9** (500 MHz, D_2O).



9

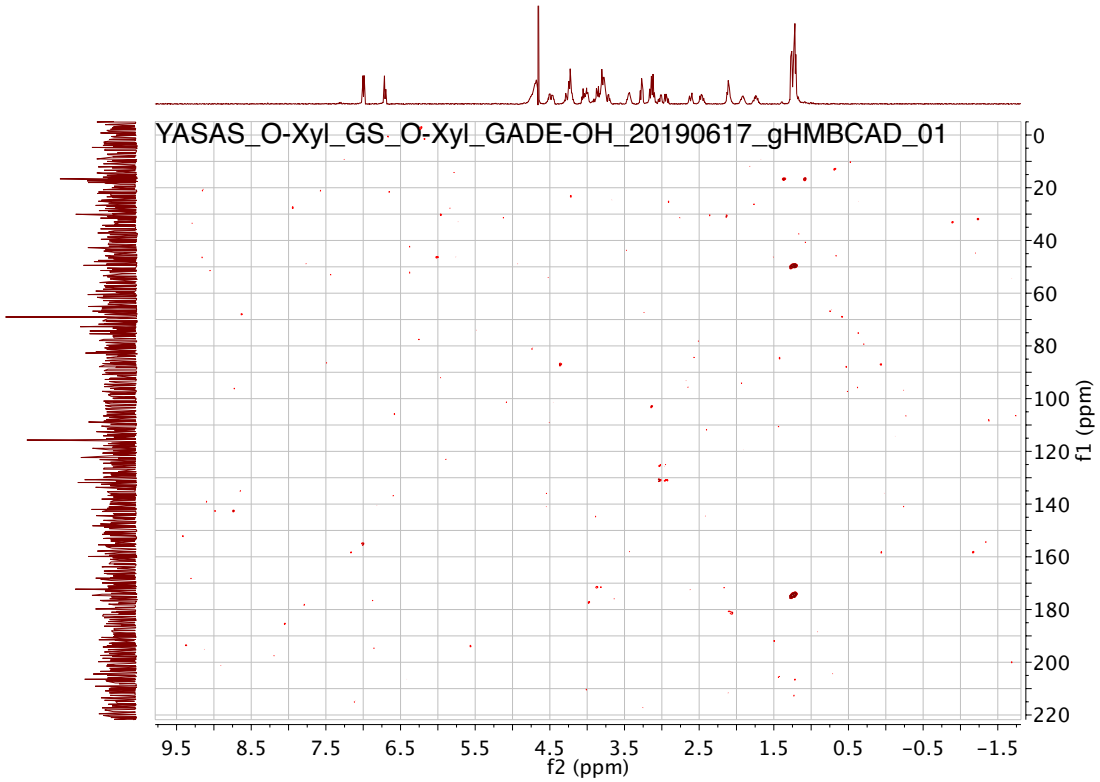
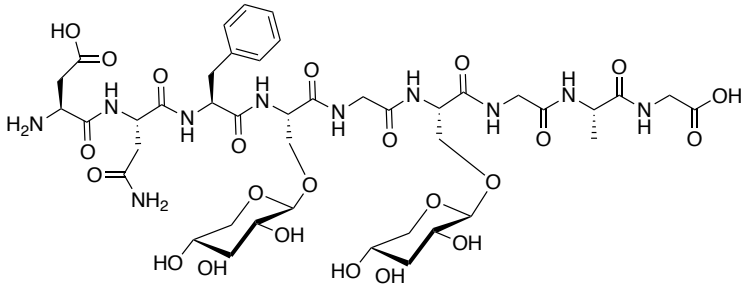
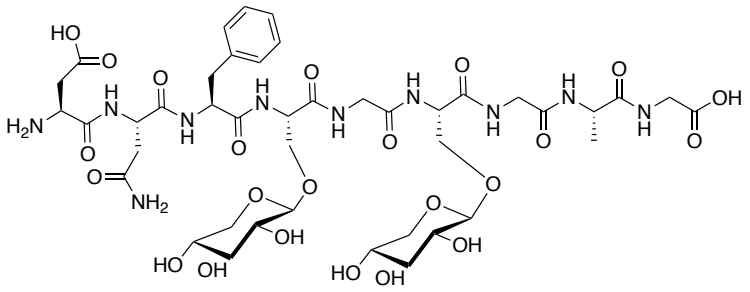


Figure 3.56 HMBC NMR of **9** (500 MHz, D_2O).



10

The purity of glycopeptide was verified with analytical C-18 HPLC (0-30-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20} = +0.020^{\circ}$ (c = 0.03, H_2O). ^{1}H-NMR (500 MHz, D_2O) δ 7.19 - 7.13 (m, 5H), 4.66 (s, 5H), 4.58 – 4.39 (m, 3H), 4.24 – 4.20 (m, 3H), 4.01 (m, 2H), 3.91 – 3.67 (m, 11H), 3.68 – 3.51 (m, 4H), 3.46 3.41 (m, 5H), 3.32 – 3.19 (m, 3H), 3.18 – 3.05 (m, 4H), 2.99 – 2.94 (m, 2H), 2.63 – 2.37 (m, 3H), 1.23 (t, *J* = 7.3 Hz, 4H); ^{13}C-NMR (125 MHz, D_2O) δ 207.4, 174.2, 170.7, 166.6, 151.8, 129.1, 128.7, 110.0, 103.0, 101.5, 72.8, 69.0, 65.1, 43.2, 24.0. ESI-MS: $C_{42}H_{62}N_{10}O_{23}$ $[M+H]^+$ calcd: 1075.4062, obsd: 1075.4014 (4.46 ppm).



10

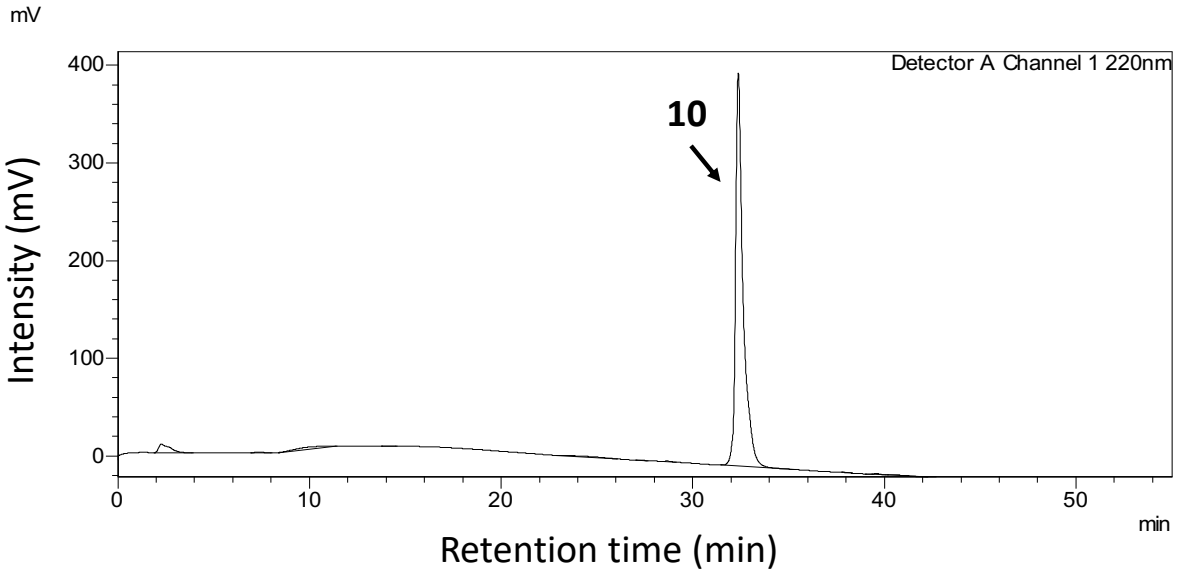
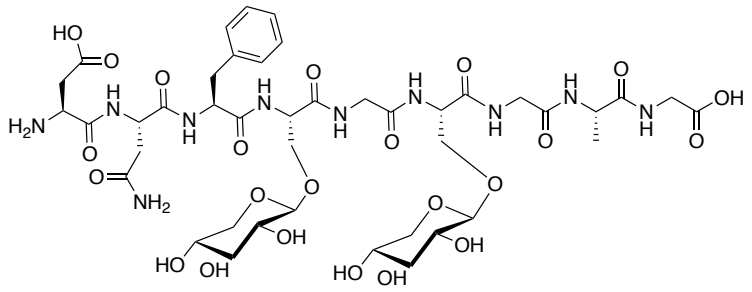


Figure 3.57 HPLC chromatogram of **9**.



10

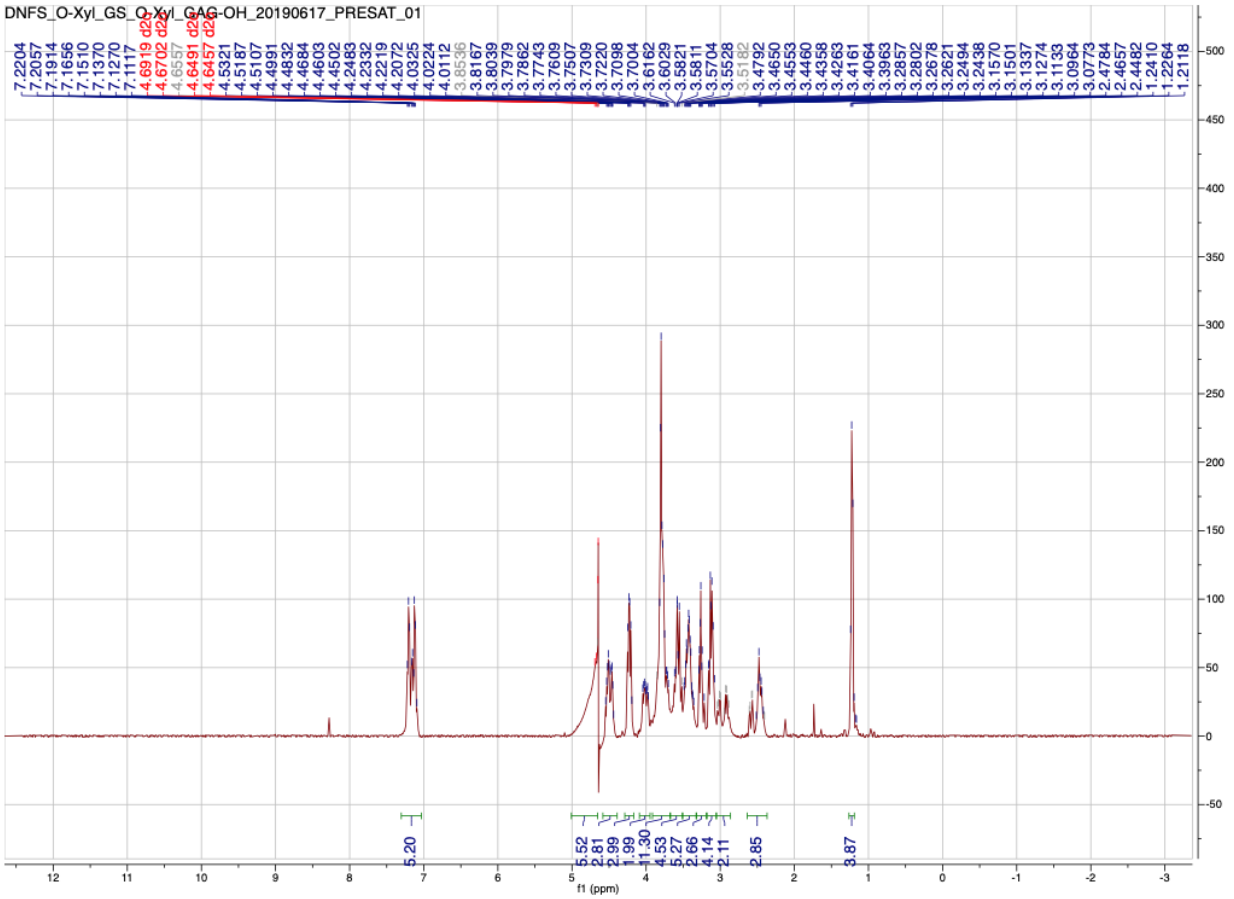
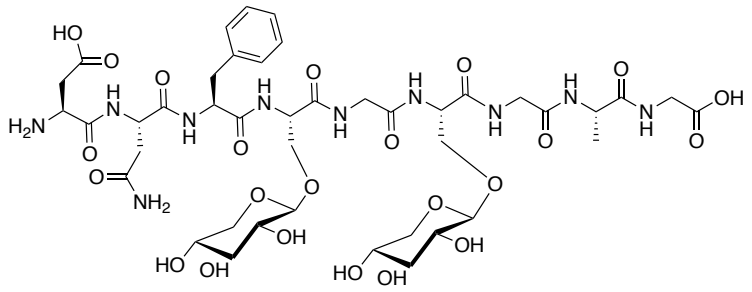


Figure 3.58 ^{1}H-NMR of **10** (500 MHz, D_2O).



10

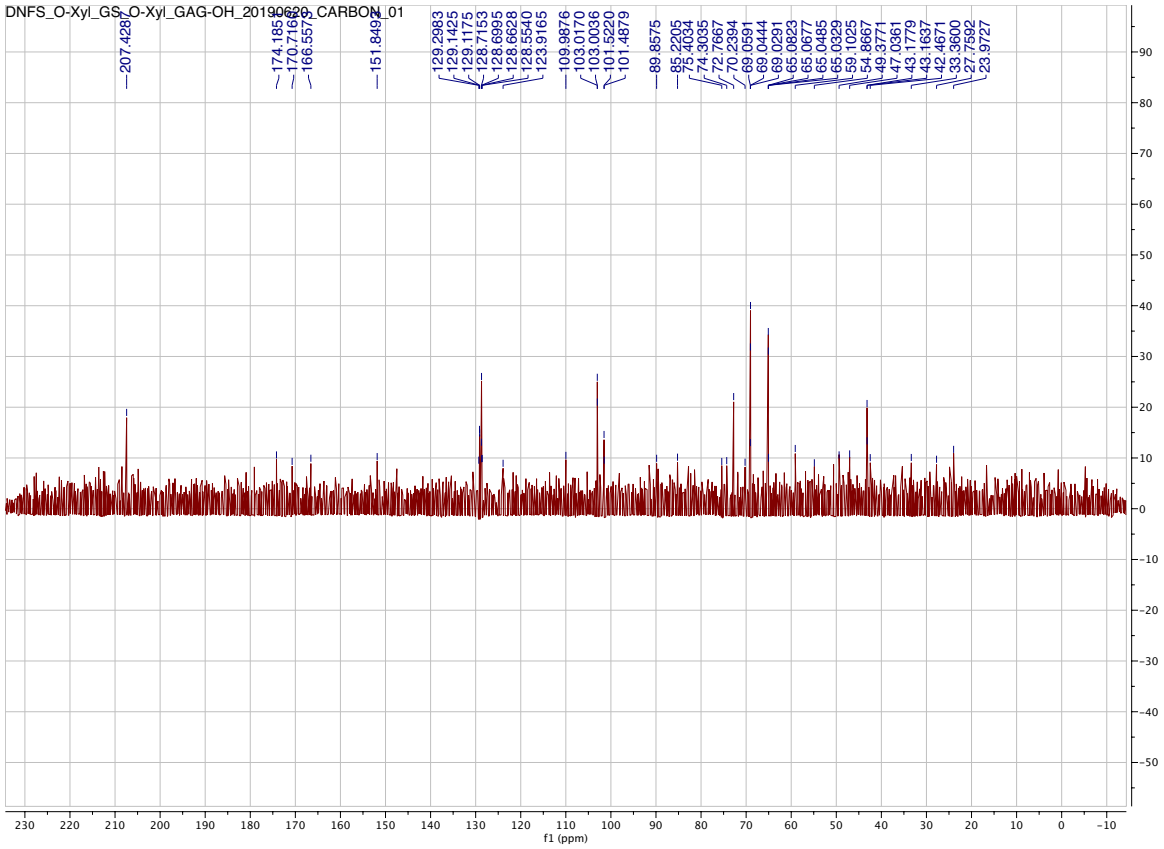
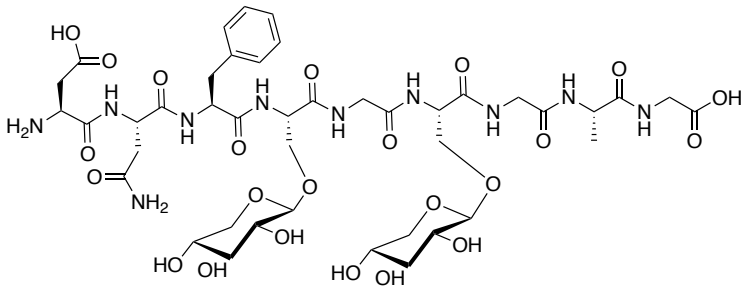


Figure 3.59 ^{13}C-NMR of **10** (125 MHz, D_2O).



10

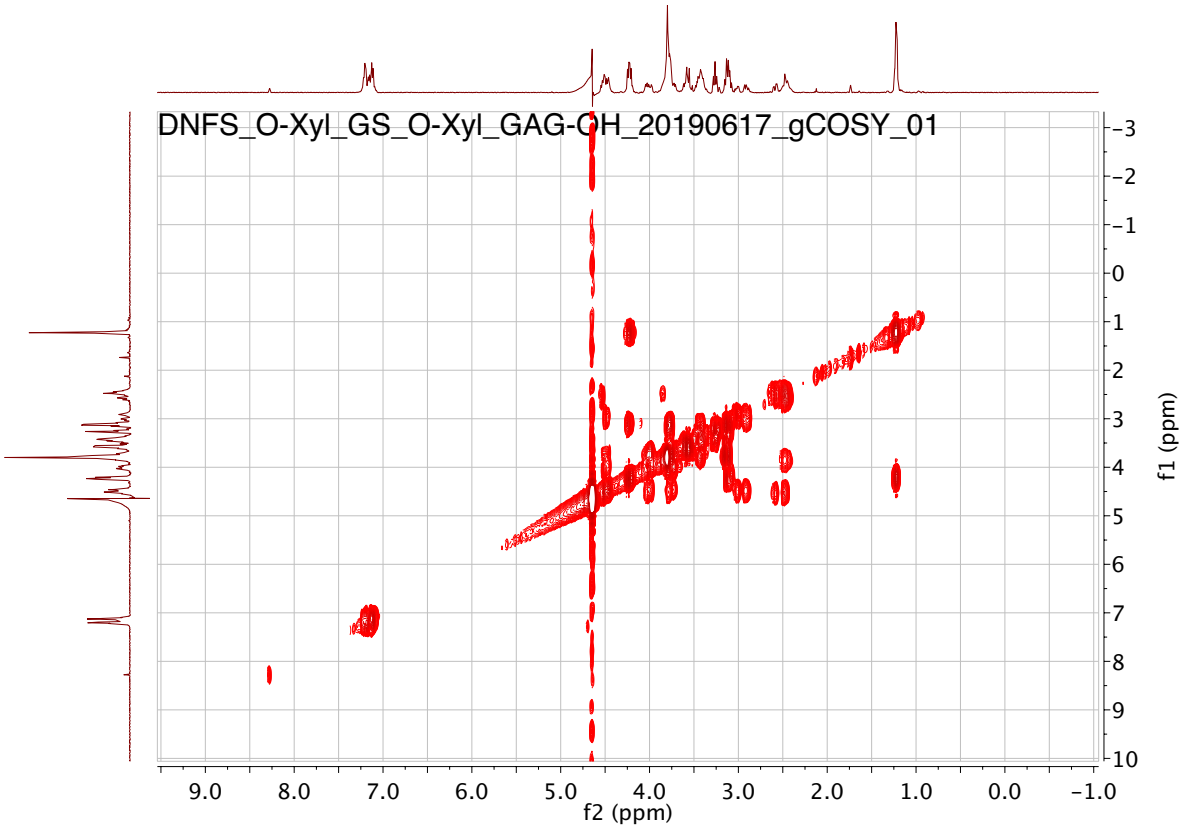
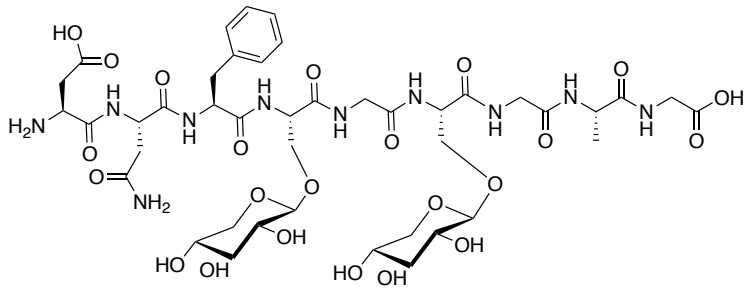


Figure 3.60 COSY NMR of **10** (500 MHz, D_2O).



10

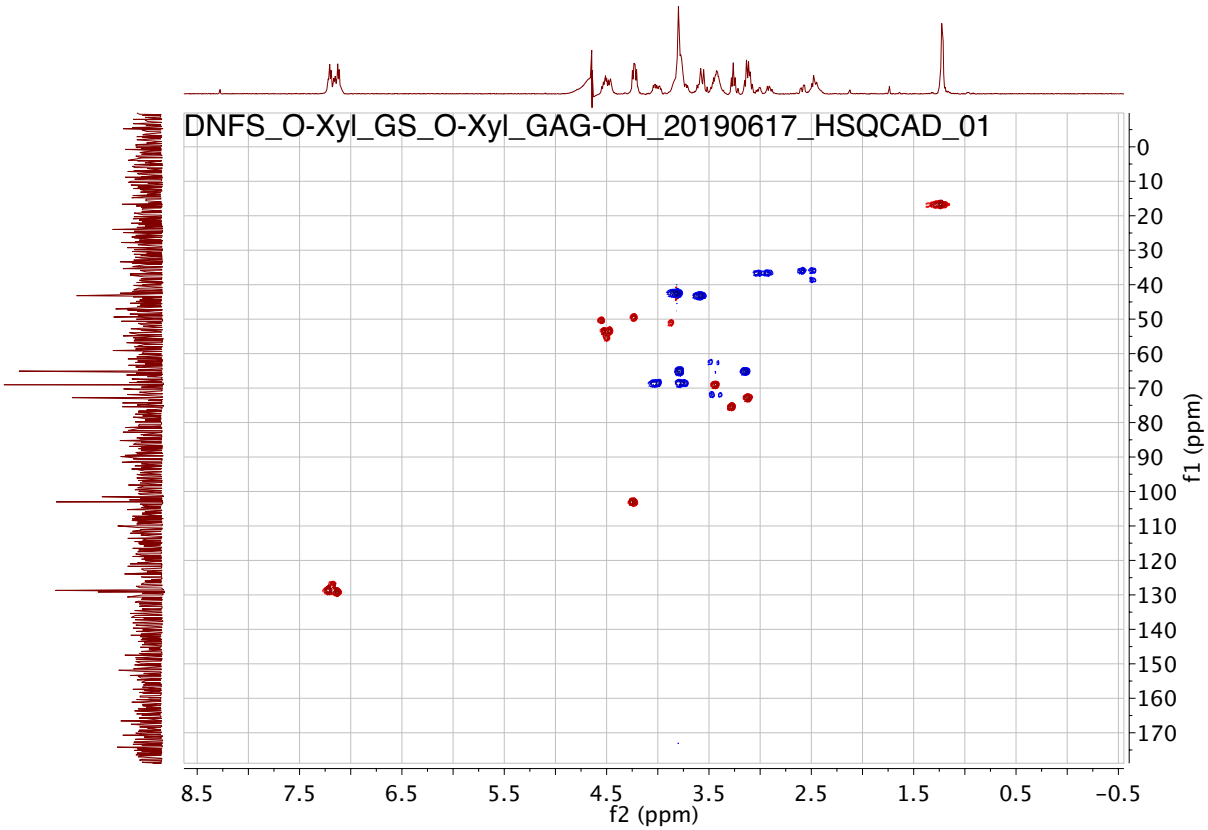
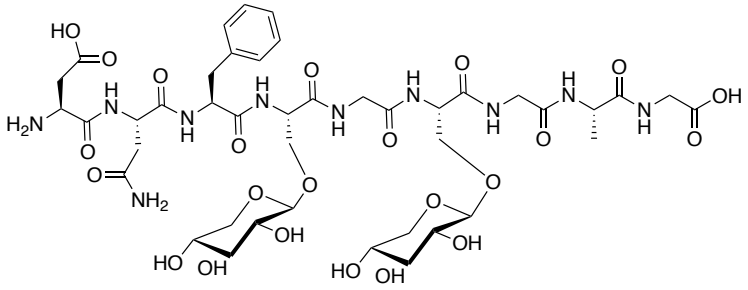


Figure 3.61 HSQC NMR of **10** (500 MHz, D_2O).



10

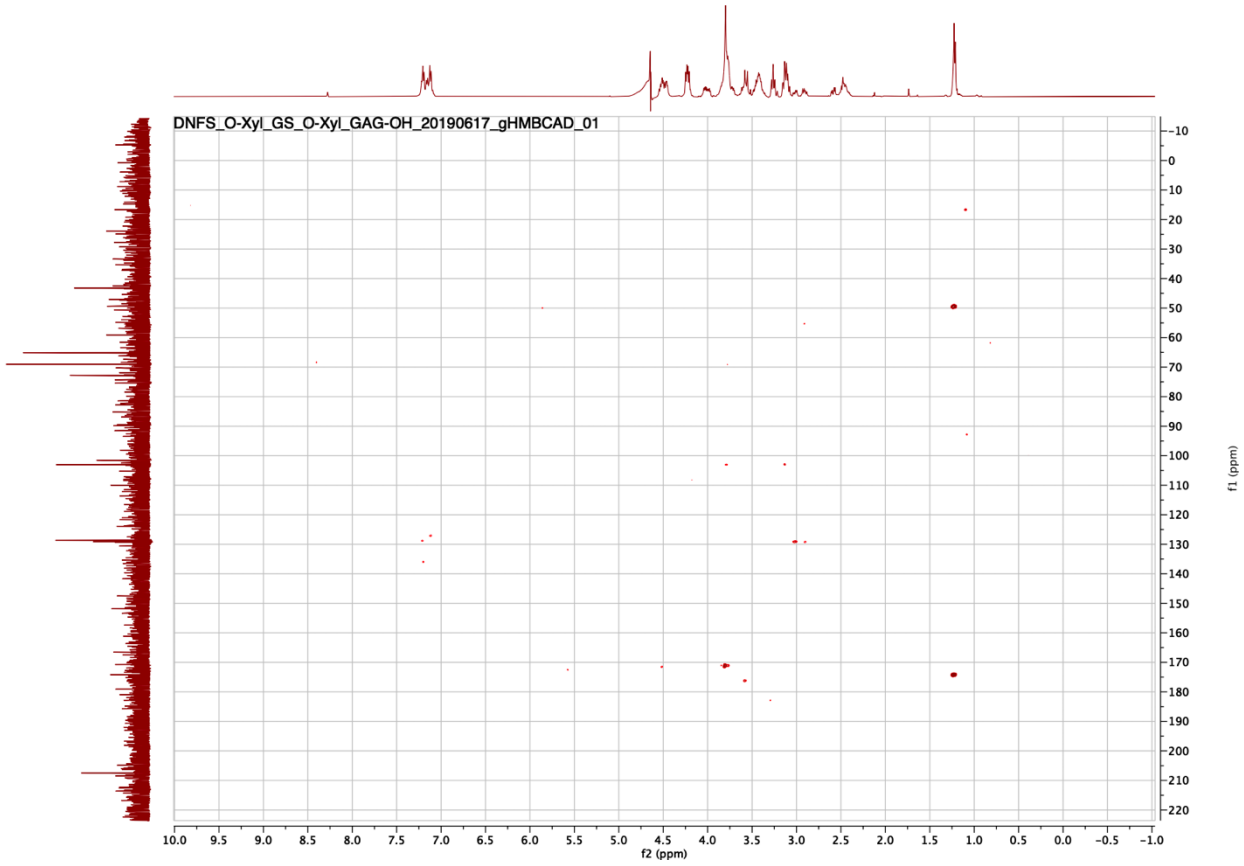
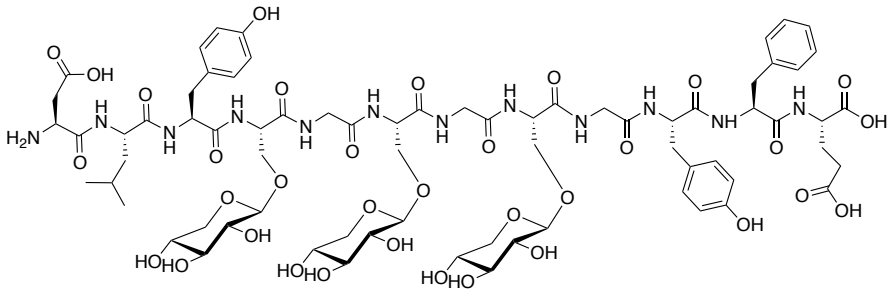
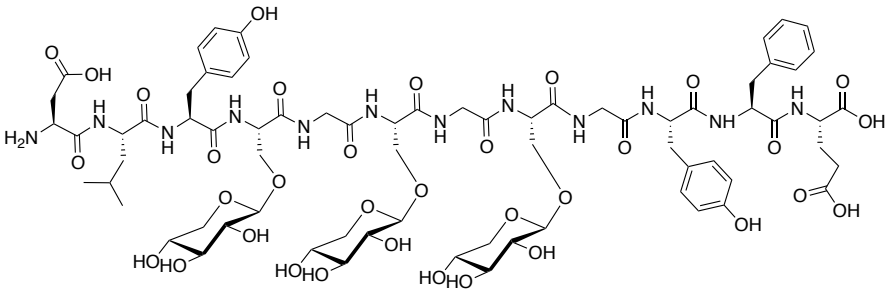


Figure 3.62 HMBC NMR of **10** (500 MHz, D_2O).



11

The purity of glycopeptide was verified with analytical C-18 HPLC (0-30-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20} = +0.013^{\circ}$ (c = 0.02, H_2O). ^{1}H-NMR (500 MHz, D_2O), δ 7.21 – 7.09 (m, 3H), 7.05 (d, *J* = 7.2 Hz, 2H), 6.97 (d, *J* = 7.0 Hz, 2H), 6.84 (d, *J* = 8.1 Hz, 2H), 6.63 (m, 4H), 4.93 – 4.62 (m, 32H), 4.56 – 4.39 (m, 3H), 4.31 (t, *J* = 7.6 Hz, 1H), 4.25 – 4.12 (m, 3H), 4.04 – 3.94 (m, 2H), 3.93 – 3.62 (m, 14H), 3.42 (d, *J* = 11.5 Hz, 3H), 3.30 – 3.18 (m, 3H), 3.09 (m, 5H), 3.03 – 2.87 (m, 1H), 2.85 – 2.61 (m, 2H), 2.61 – 2.40 (m, 1H), 1.99 (t, *J* = 8.3 Hz, 2H), 1.85 (d, *J* = 9.6 Hz, 1H), 1.74 (s, 1H), 1.33 (s, 1H), 0.69 (m, 6H); ^{13}C-NMR (125 MHz, D_2O) δ 160.3, 153.8, 136.8, 128.5, 128.4, 128.1, 117.9, 117.2, 115.3, 99.1, 95.2, 89.7, 24.5, 21.3. ESI-MS: $C_{72}H_{101}N_{12}O_{34}$ $[M+H]^+$ calcd: 1677.6538, obsd: 1677.6555 (1.0 ppm).



11

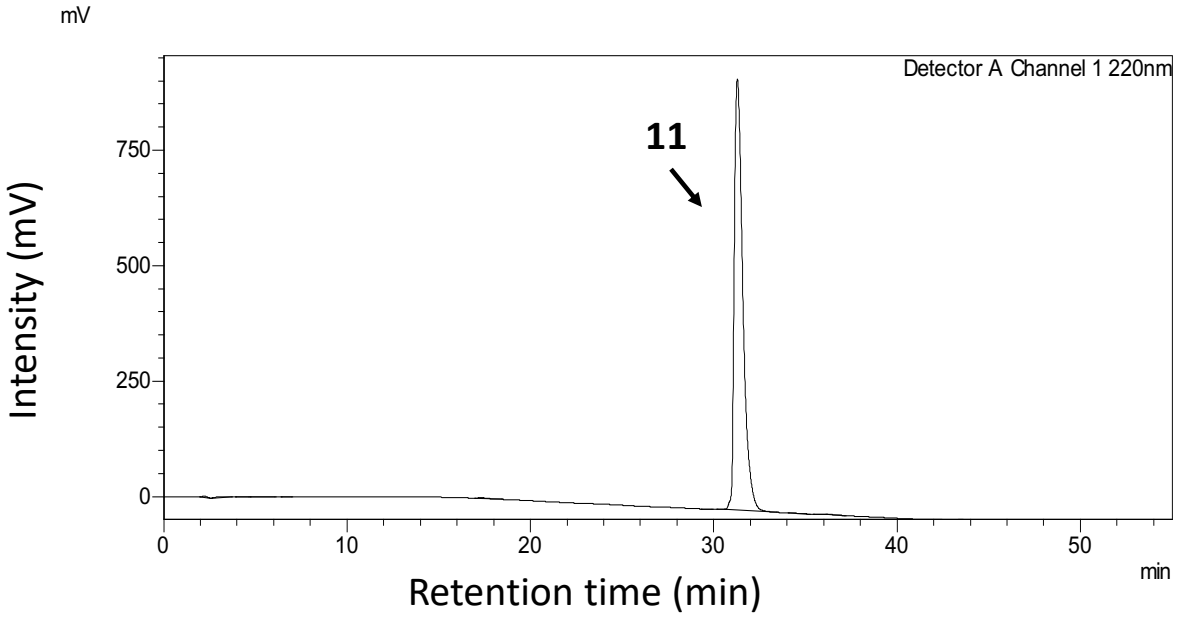
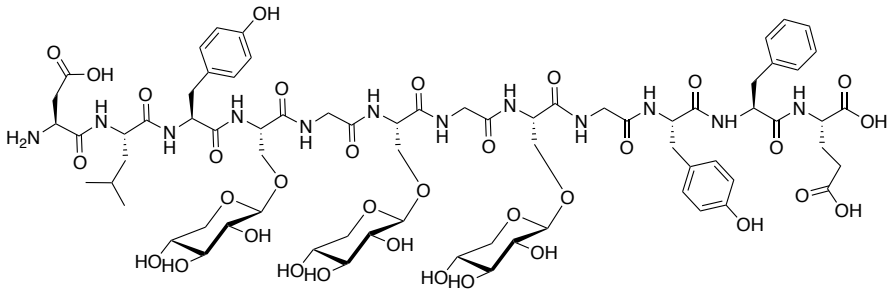


Figure 3.63 HPLC chromatogram of **11**.



11

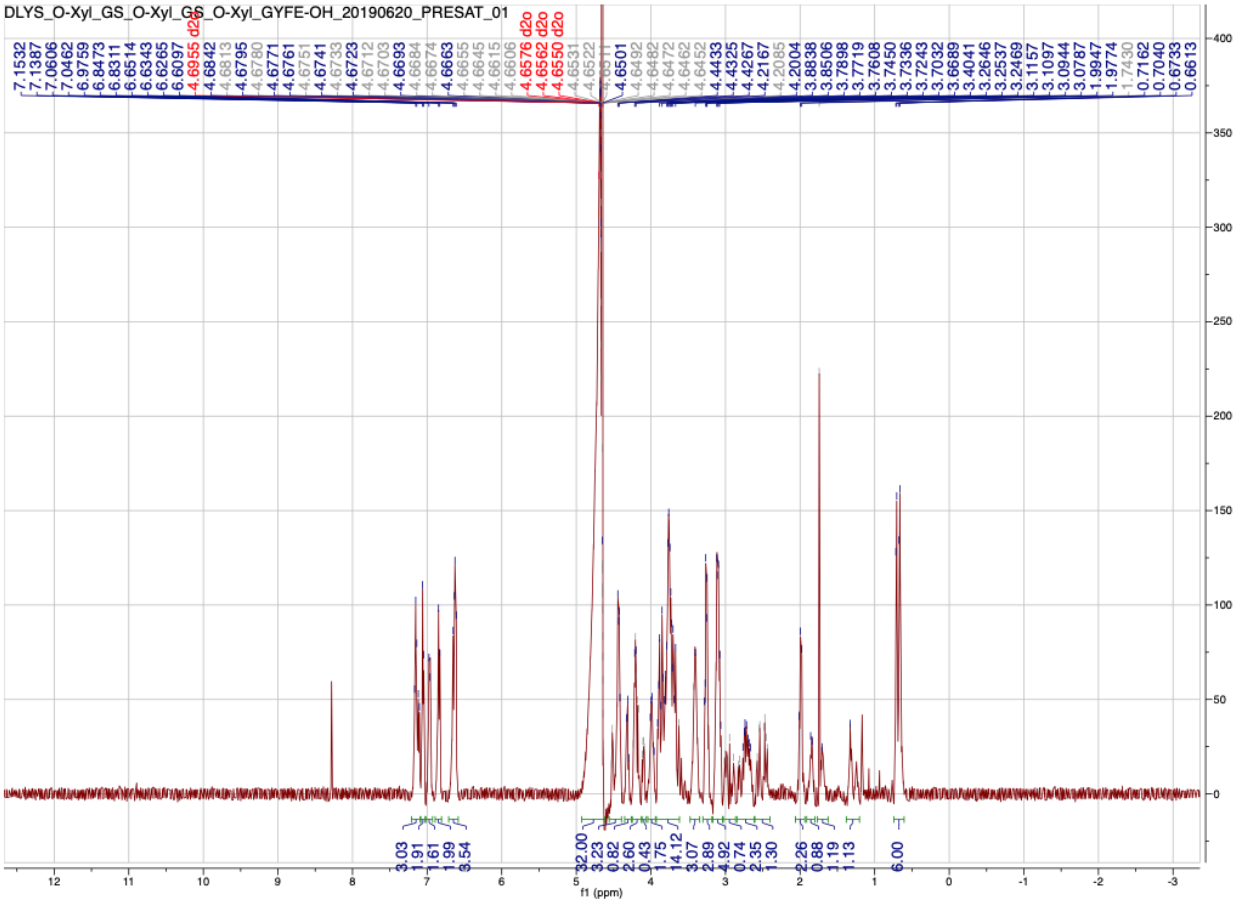
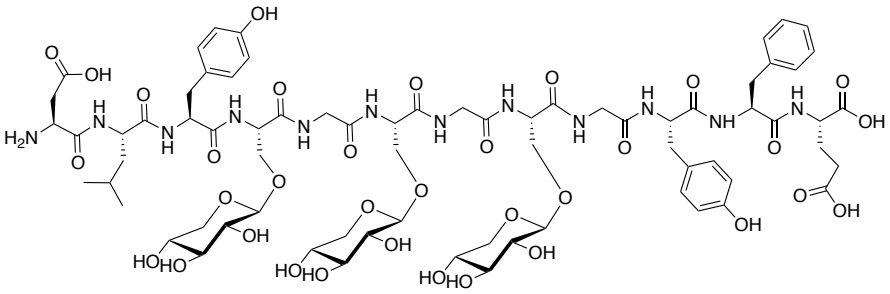


Figure 3.64 ^{1}H-NMR of **11** (500 MHz, D_2O).



11

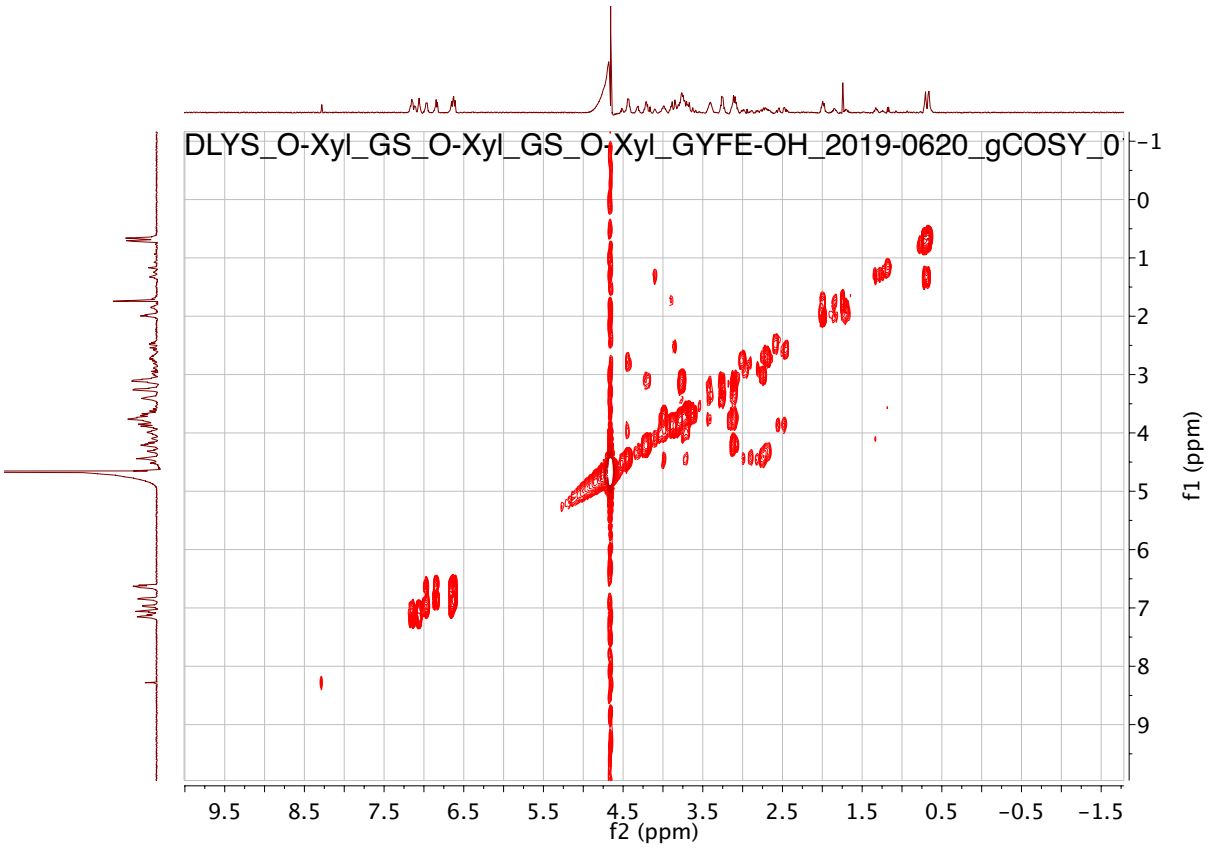
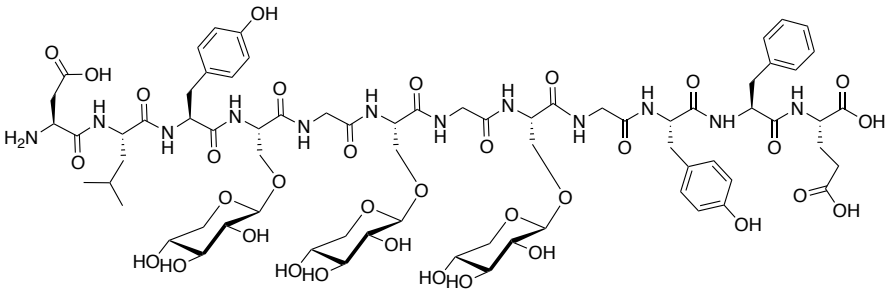


Figure 3.65 COSY NMR of **11** (500 MHz, D_2O).



11

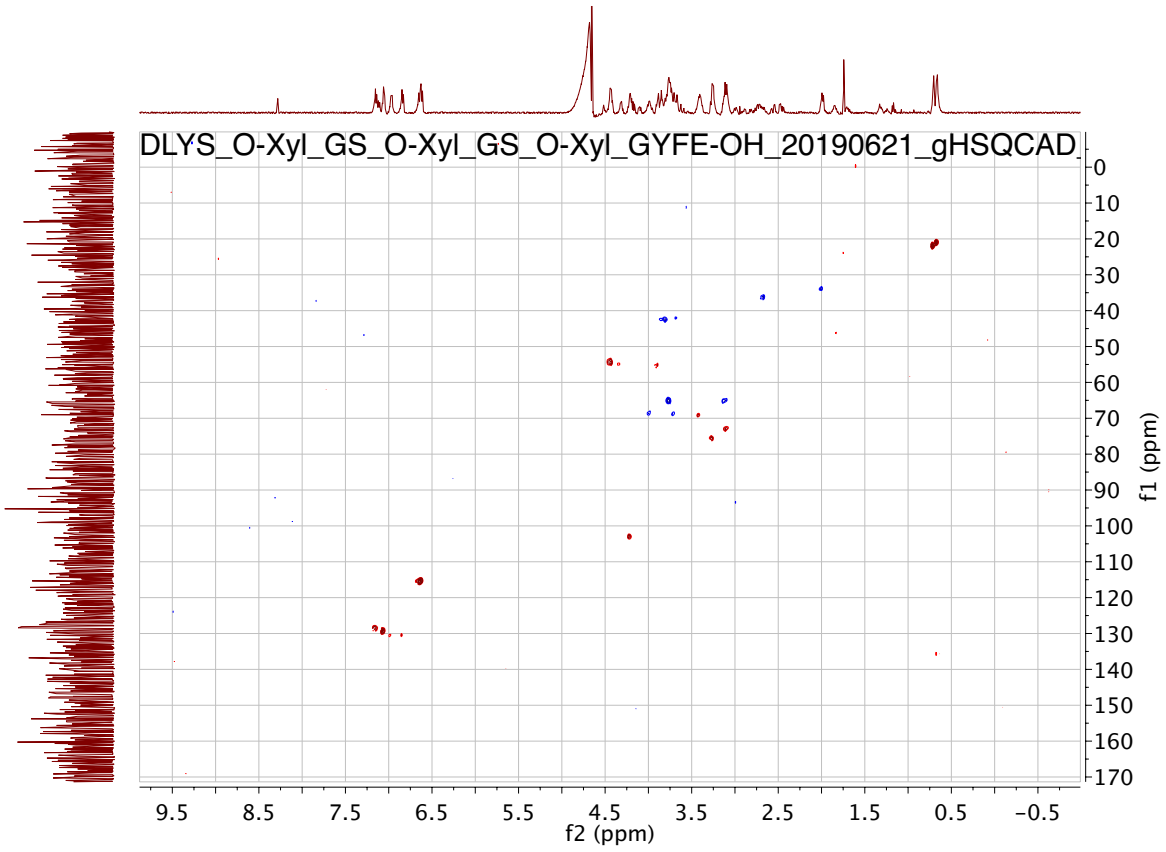
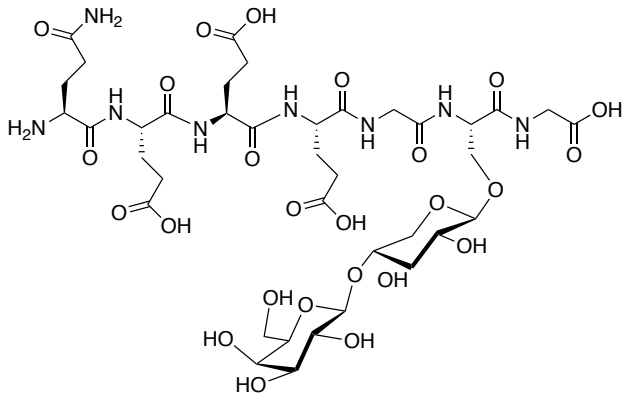
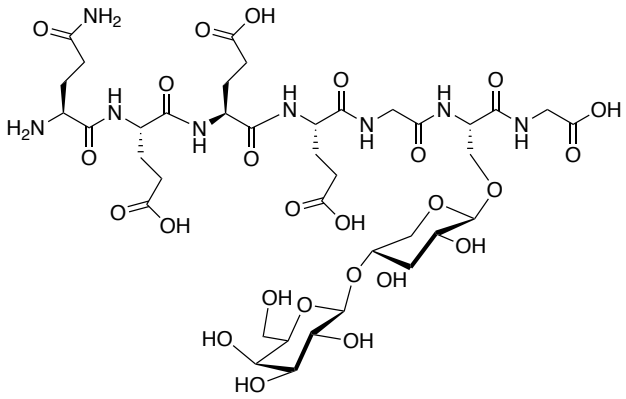


Figure 3.66 HSQC NMR of **11** (500 MHz, D_2O).



12

The purity of glycopeptide was verified with analytical C-18 HPLC (water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20} = -1.530^{\circ}$ (c = 0.08, H_2O). ^{1}H-NMR (500 MHz, D_2O), δ 4.53 (t, *J* = 5.0 Hz, 1H), 4.34 – 4.14 (m, 6H), 4.05 – 4.01 (m, 2H), 3.99 – 3.85 (m, 3H), 3.86 – 3.71 (m, 7H), 3.72 – 3.60 (m, 3H), 3.61 – 3.50 (m, 3H), 3.51 – 3.38 (m, 3H), 3.26 – 3.07 (m, 3H), 2.41 – 2.19 (m, 10H), 2.07 – 1.90 (m, 6H), 1.90 – 1.75 (m, 4H). ^{13}C NMR (225 MHz, D_2O) δ 102.0, 76.3, 72.9, 70.5, 68.7, 63.0, 61.0, 53.0, 44.6, 42.5, 30.1, 26.4. ESI-MS: $C_{38}H_{60}N_8O_{25}$ $[M+H]^+$ calcd: 1029.3743, obsd: 1029.3717 (2.53 ppm).



12

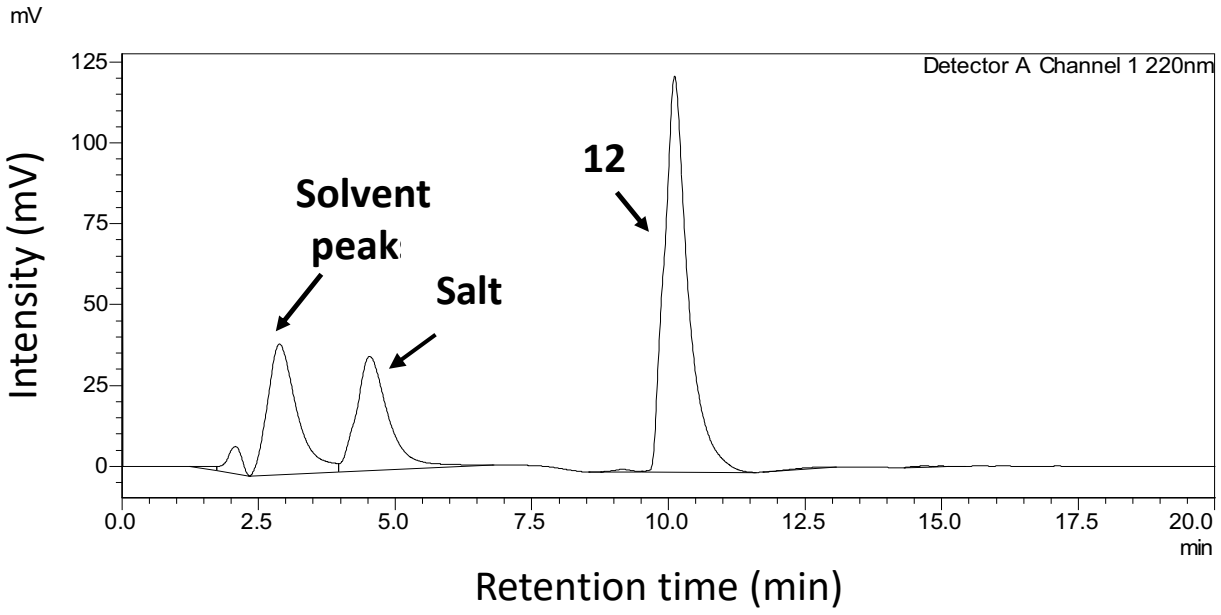
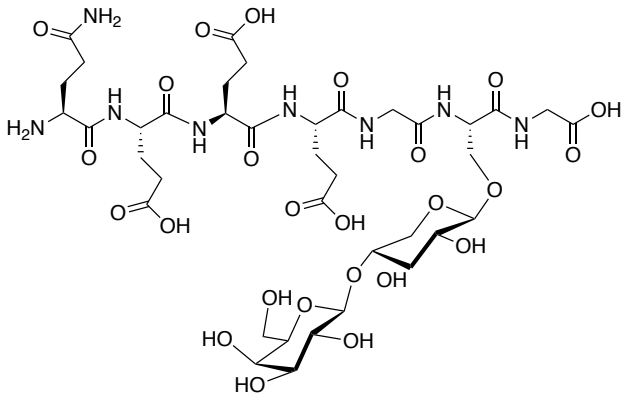


Figure 3.67 HPLC chromatogram of **12**.



12

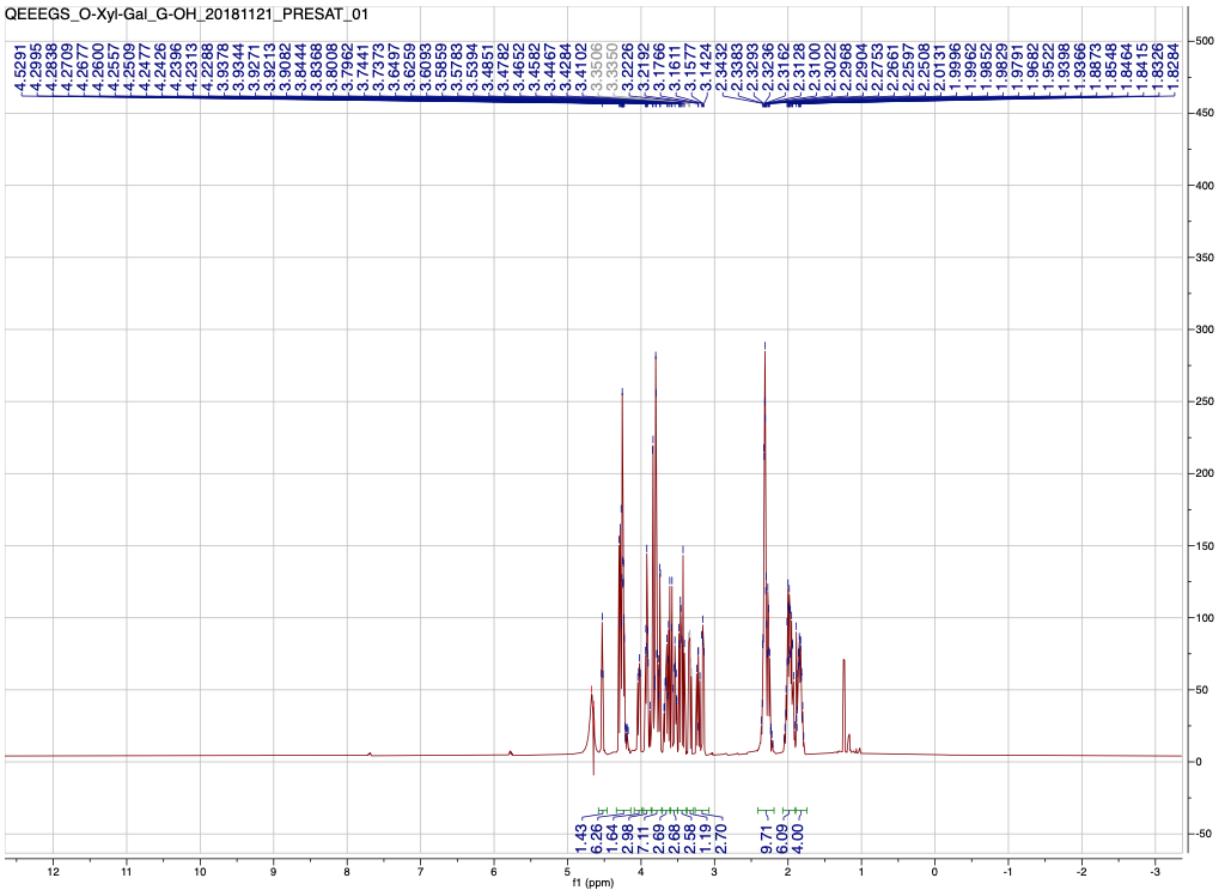
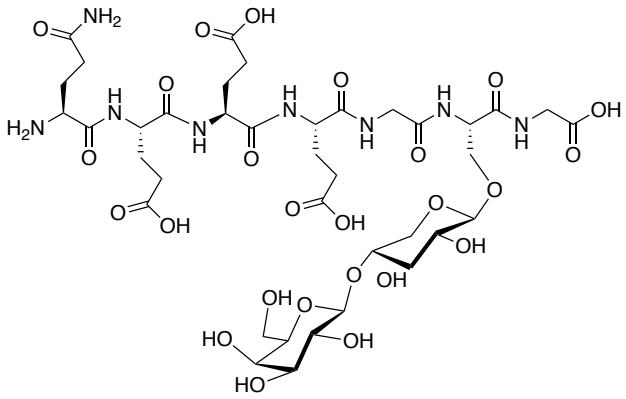


Figure 3.68 ^{1}H-NMR NMR of **12** (500 MHz, D_2O).



12

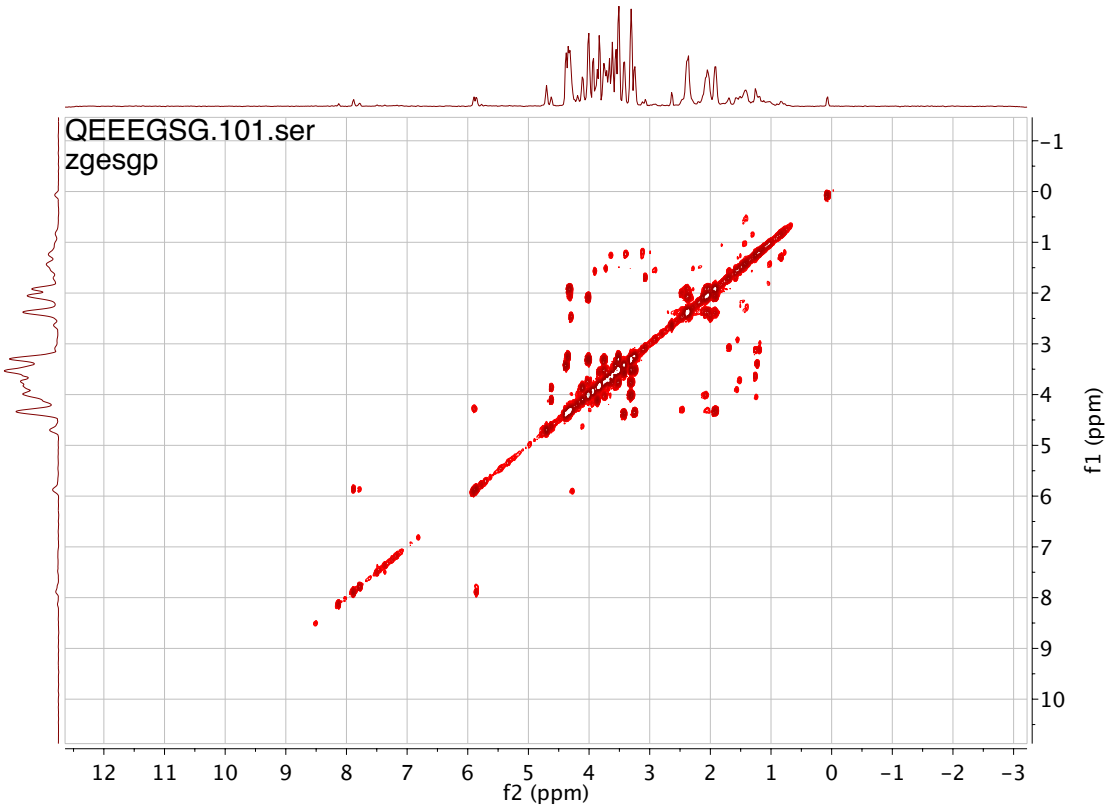
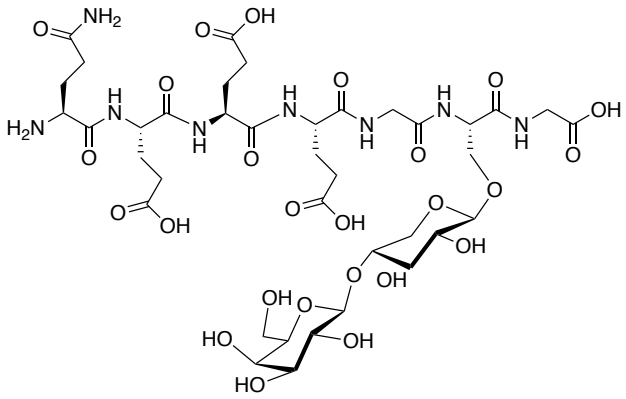


Figure 3.69 COSY NMR of **12** (900 MHz, D_2O).



12

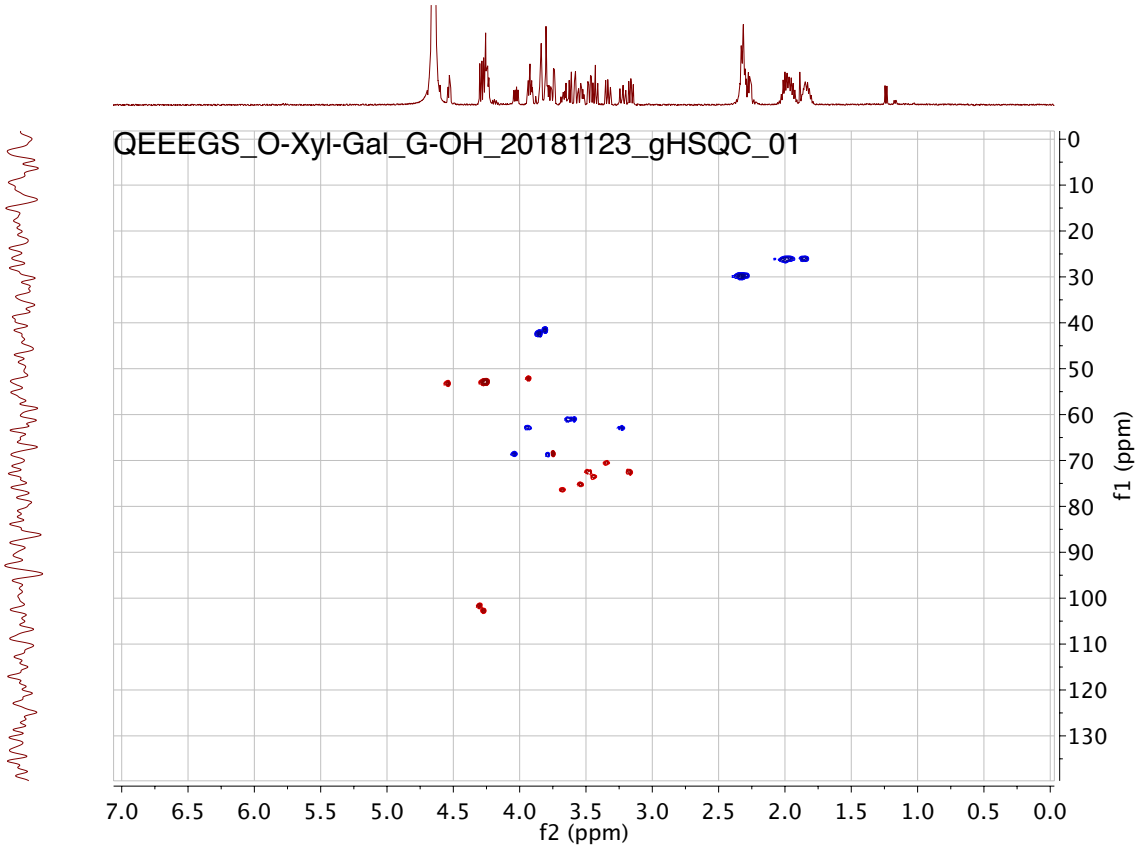
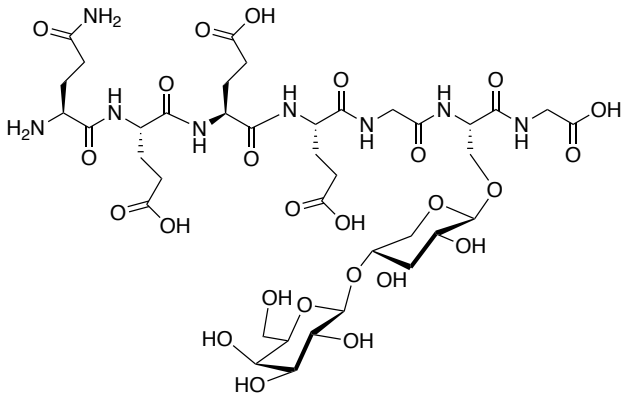
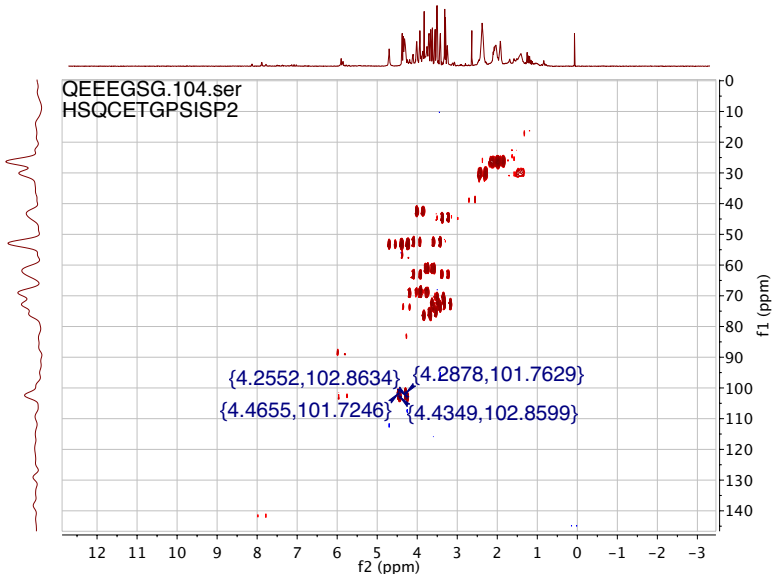


Figure 3.70 HSQC NMR of **12** (500 MHz, D_2O).



12



$^{1}J_{C1,\,H1}$ = 159.9 Hz, 161.7 Hz

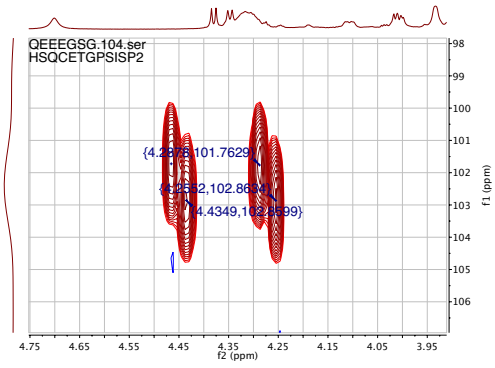
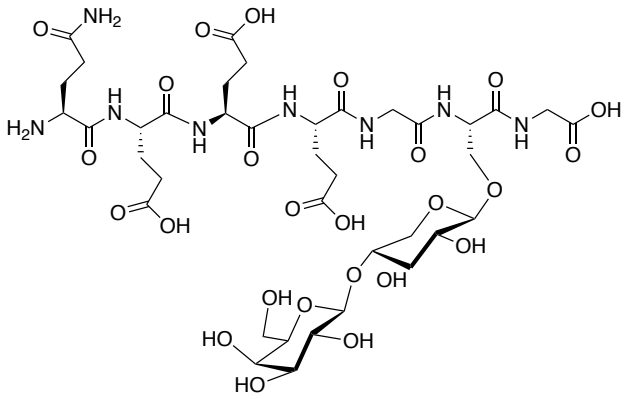


Figure 3.71 HSQC-coupled NMR of **12** (900 MHz, D_2O).



12

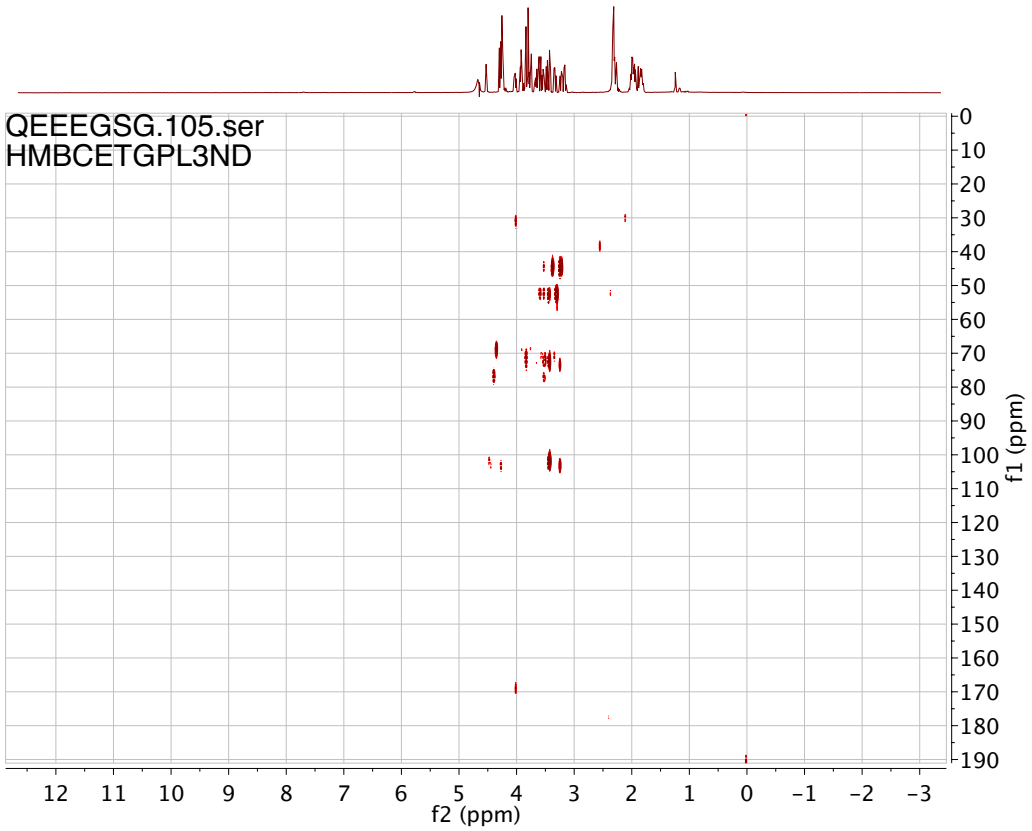
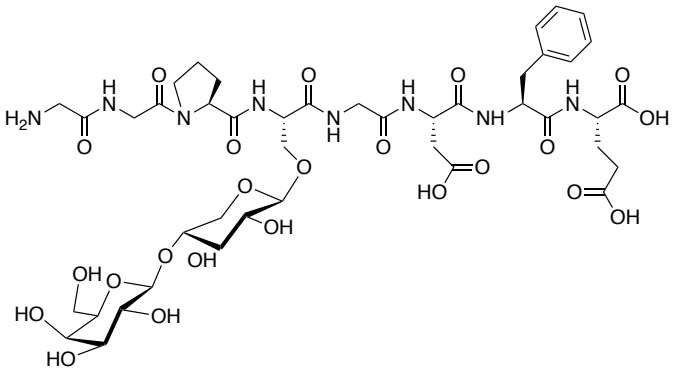
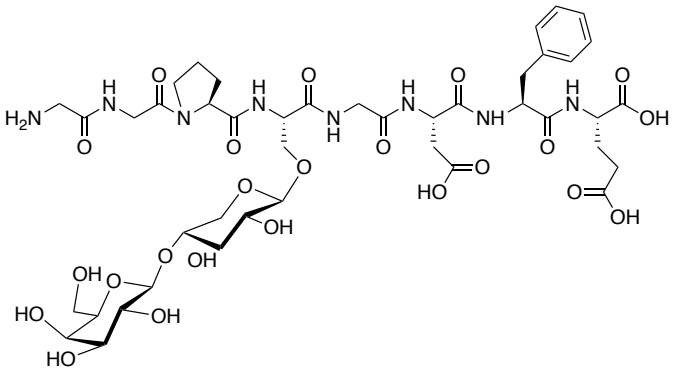


Figure 3.72 HMBC NMR of **12** (900 MHz, D_2O).



13

The purity of glycopeptide was verified with analytical C-18 HPLC (5-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = -3.400° (c = 0.16, H_2O). ^{1}H-NMR (500 MHz, D_2O) δ 7.84 – 7.75 (m, 1H), 7.26 – 7.02 (m, 5H), 5.88 – 5.71 (m, 2H), 4.65 (d, *J* = 5.0 Hz, 1H), 4.53 (s, 1H), 4.47 (d, *J* = 9.3 Hz, 1H), 4.27 (t, *J* = 8.7 Hz, 2H), 4.24 – 4.14 (m, 3H), 4.11 (s, 1H), 4.08 – 3.94 (m, 4H), 3.91 (d, *J* = 11.2 Hz, 2H), 3.87 – 3.76 (m, 3H), 3.73 (d, *J* = 7.7 Hz, 4H), 3.69 – 3.55 (m, 5H), 3.54 – 3.38 (m, 6H), 3.33 (t, *J* = 8.4 Hz, 1H), 3.26 – 3.06 (m, 3H), 3.06 – 2.95 (m, 1H), 2.94 – 2.82 (m, 1H), 2.68 (m, 1H), 2.55 (m, 1H), 2.22 (t, *J* = 6.3 Hz, 2H), 2.14 (s, 1H), 2.00 (s, 1H), 1.86 – 1.82 (m, 4H); ^{13}C-NMR (125 MHz, D_2O) δ 140.8, 129.2, 128.6, 102.8, 102.6, 95.8, 95.7, 88.6, 88.3, 82.8, 73.8, 73.6, 73.6, 71.8, 71.8, 70.0, 69.5, 69.2, 69.0, 68.9, 68.9, 68.9, 68.3, 68.3, 68.2, 66.0, 65.8, 64.9, 64.8, 60.9, 60.8, 52.4, 52.1, 52.1, 47.2, 46.7, 33.9. ESI-MS: $C_{43}H_{62}N_8O_{23}$ $[M+H]^+$ calcd: 1059.4001, obsd: 1059.3901 (3.34 ppm).



13

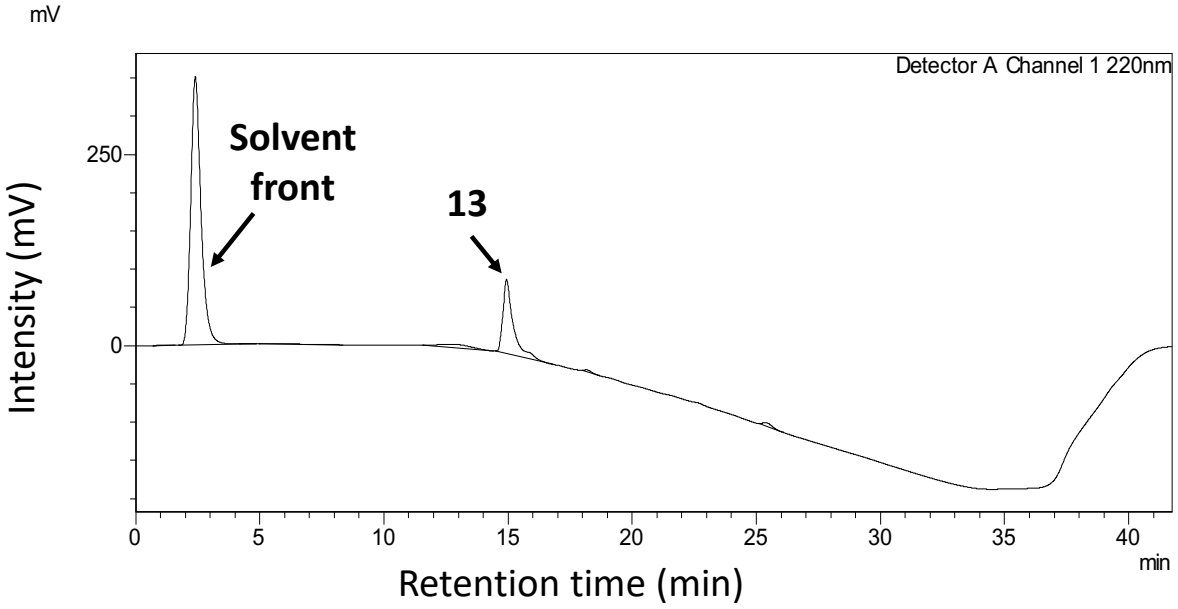
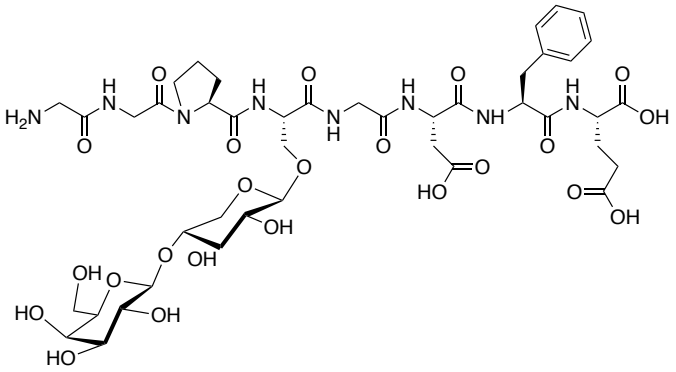


Figure 3.73 HPLC chromatogram of **13**.



13

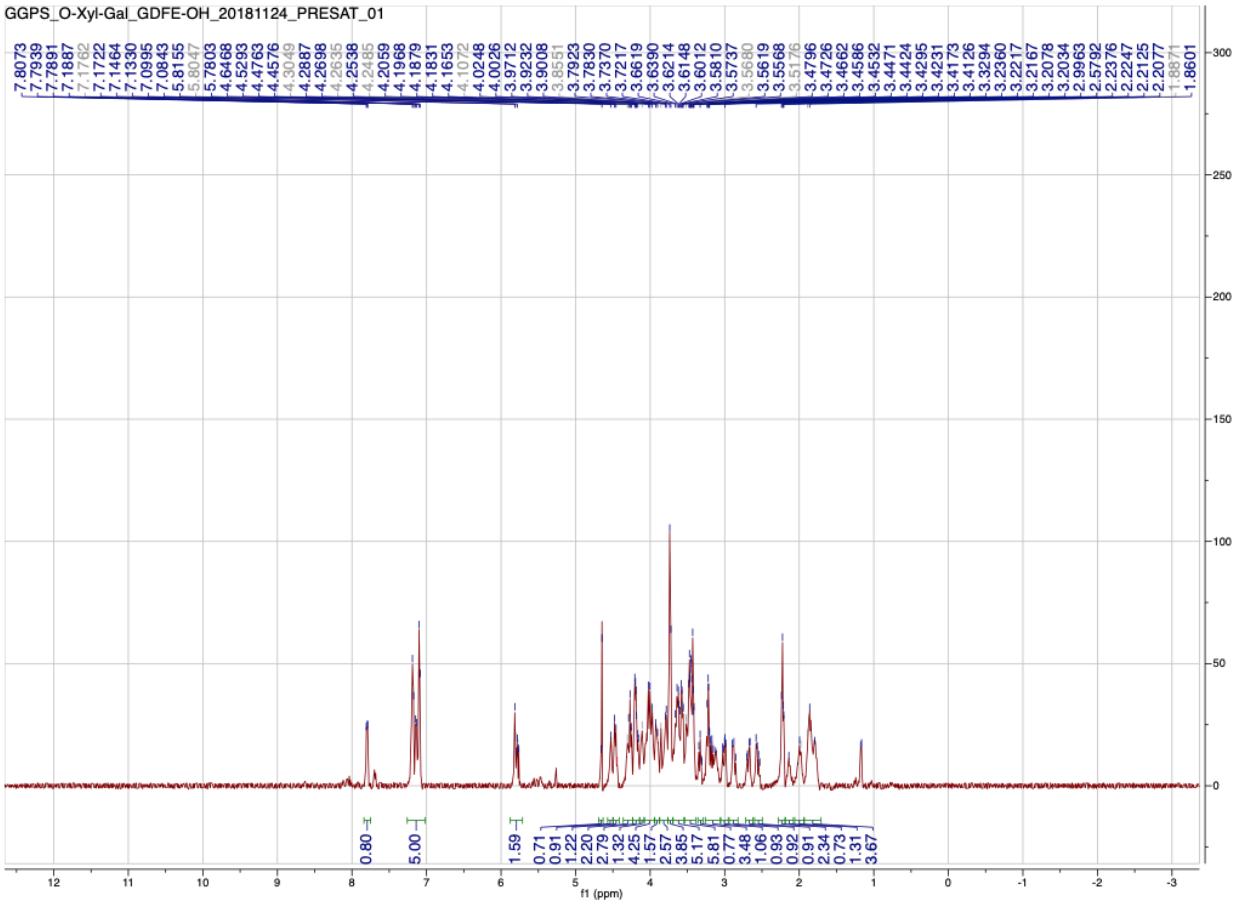
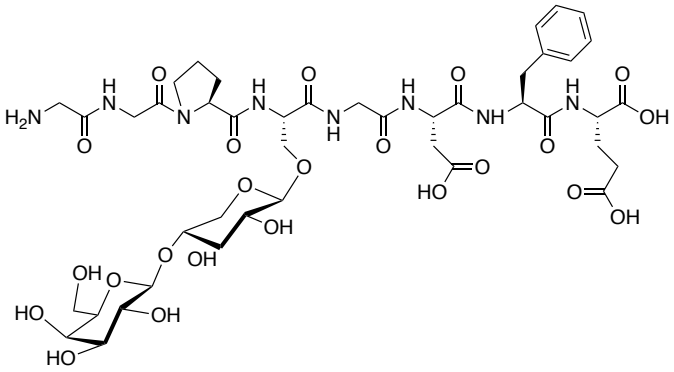


Figure 3.74 ^{1}H-NMR NMR of **13** (500 MHz, D_2O).



13

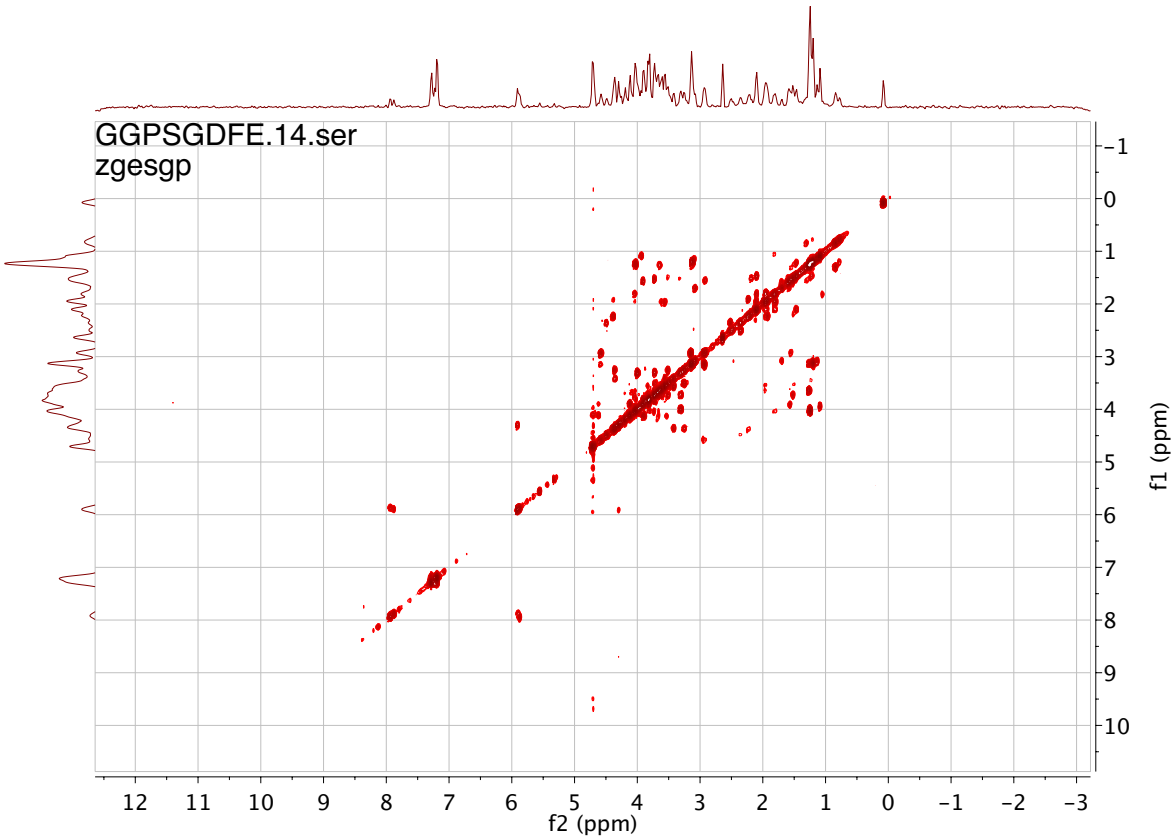
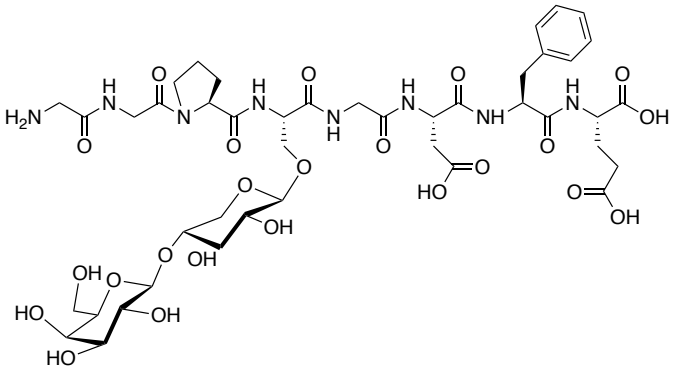


Figure 3.75 COSY NMR of **13** (900 MHz, D_2O).



13

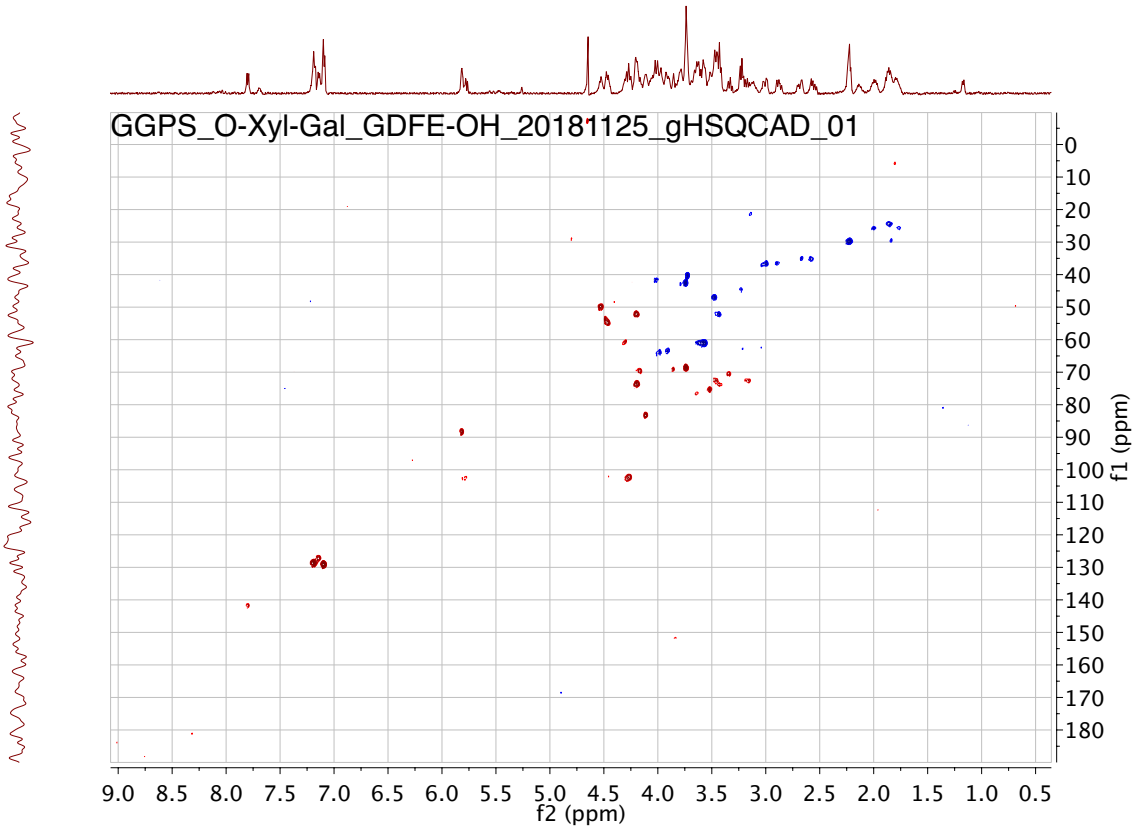
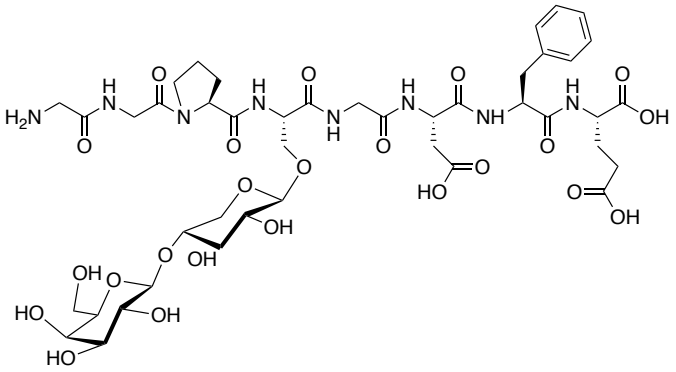
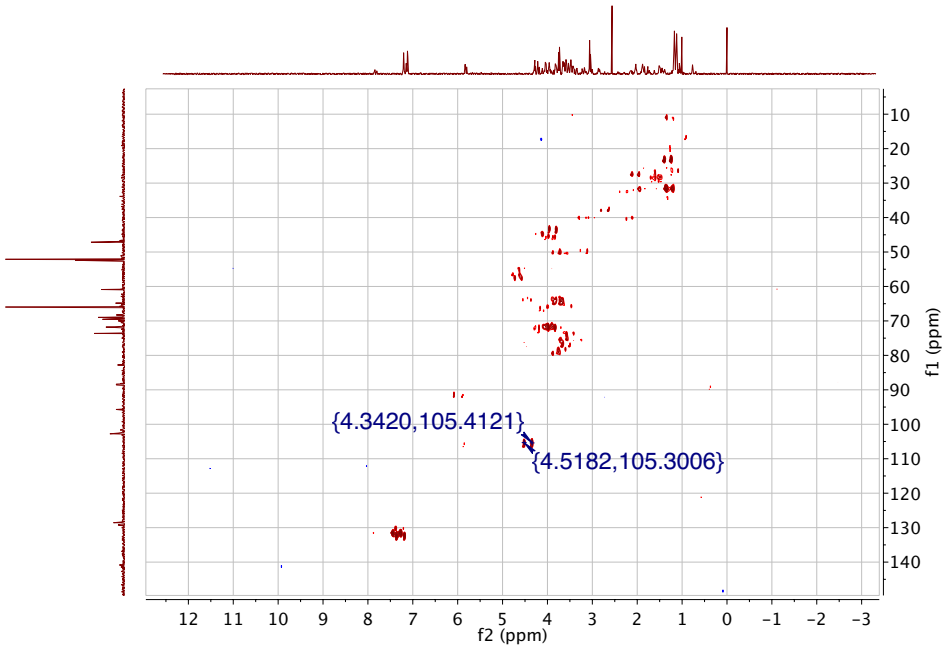


Figure 3.76 HSQC NMR of **13** (500 MHz, D_2O).



13



$^{1}J_{C1,\,H1}$ = 158.6 Hz, 158.6 Hz

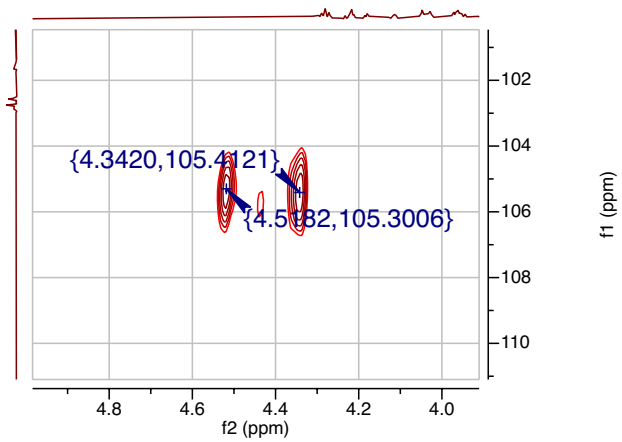
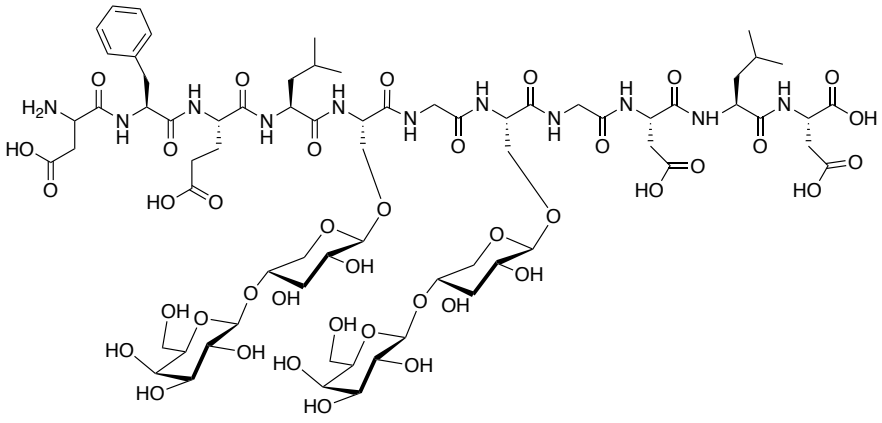
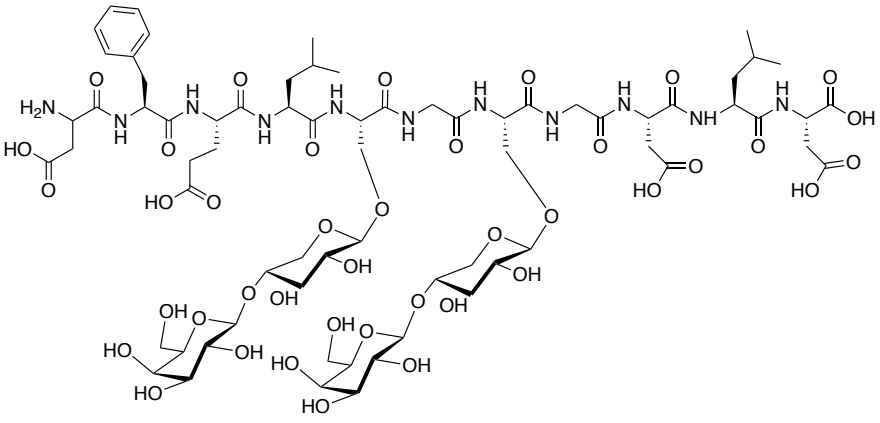


Figure 3.77 HSQC-coupled NMR of **13** (500 MHz, D_2O).



14

The purity of glycopeptide was verified with analytical C-18 HPLC (0-30-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = + 0.900 ° (c = 0.01, H_2O). ^{1}H-NMR (900 MHz, D_2O) δ 7.79 (m, 1H), 7.31 – 7.06 (m, 5H), 5.82 (m, 1H), 4.40 – 4.25 (m, 2H), 4.22 (s, 1H), 4.12 (m, 2H), 4.01 (m, 3H), 3.94 (m, 5H), 3.88 – 3.78 (m, 5H), 3.76 (d, *J* = 4.2 Hz, 4H), 3.71 – 3.62 (m, 7H), 3.59 (d, *J* = 11.9 Hz, 5H), 3.55 (t, *J* = 6.5 Hz, 5H), 3.49 – 3.45 (m, 8H), 3.38 – 3.33 (m, 4H), 3.31 – 3.20 (m, 4H), 3.18 (t, *J* = 9.3 Hz, 3H), 3.04 - 2.99 (m, 2H), 2.91 (d, *J* = 10.1 Hz, 1H), 2.11 – 2.07 (m, 2H), 1.87 – 1.84 (m, 2H), 1.78 – 1.73 (m, 1H), 1.64 – 1.60 (m, 1H), 1.52 – 1.49 (m, 6H), 1.41 – 1.37 (m, 2H), 1.20 -1.16 (m, 2H), 1.15 – 1.11 (m, 1H), 1.09 – 1.05 (m, 1H), 0.90 – 0.59 (m, 14H); ^{13}C-NMR (225 MHz, D_2O) δ 143.2, 128.9, 114.8, 111.9, 102.5, 97.2, 88.3, 82.3, 77.6, 73.5, 70.8, 68.7, 60.8, 44.9, 28.8, 22.1, 18.4, 8.7. ESI-MS: $C_{70}H_{107}N_{11}O_{40}$ $[M+2H]^{2+}$ calcd: 871.8412, obsd: 871.8452 (4.59 ppm).



14

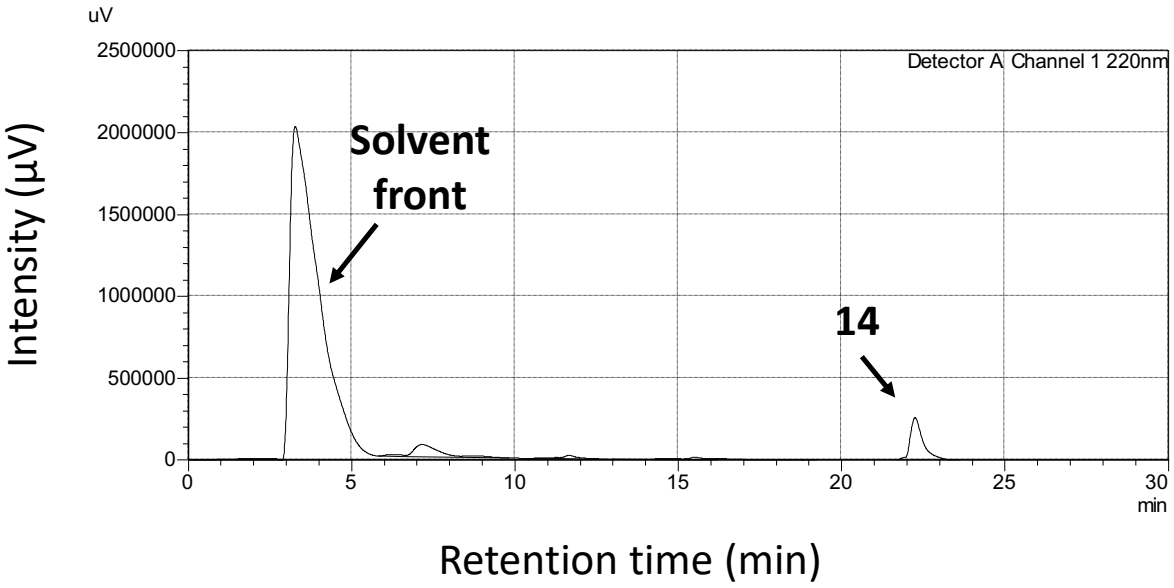
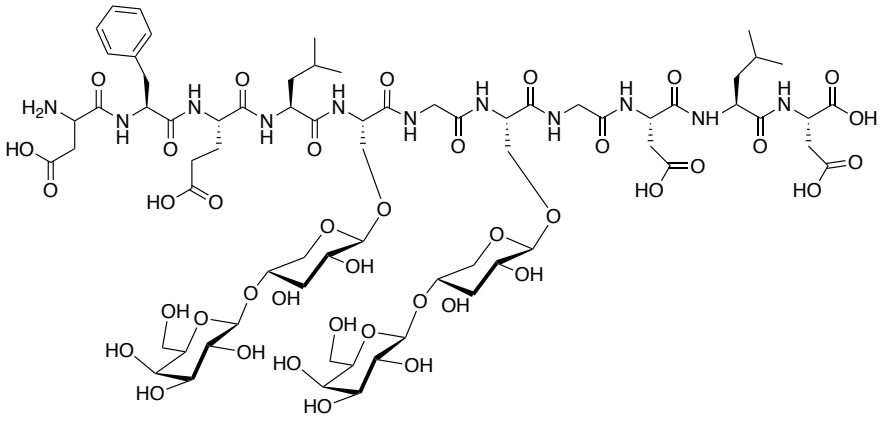


Figure 3.78 HPLC chromatogram of **14**.



14

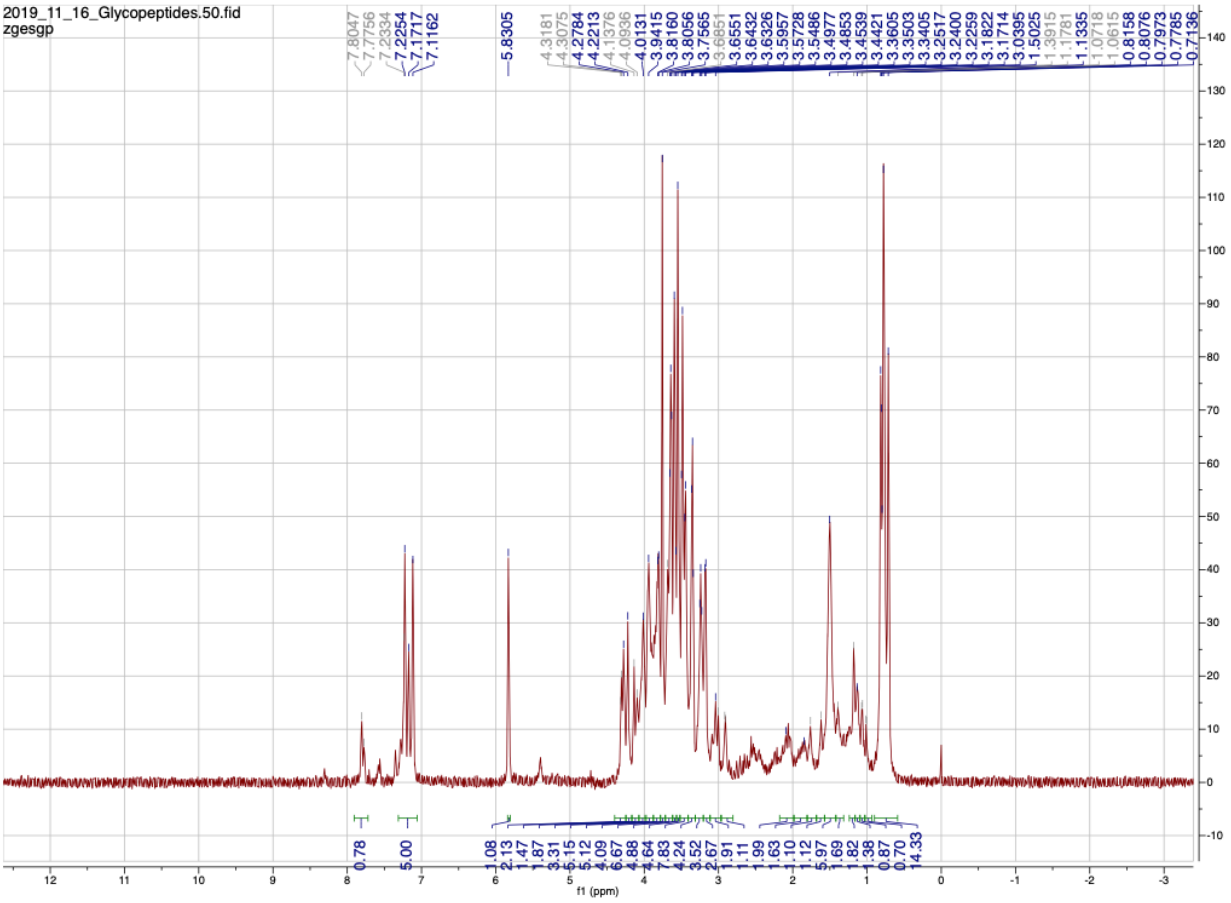
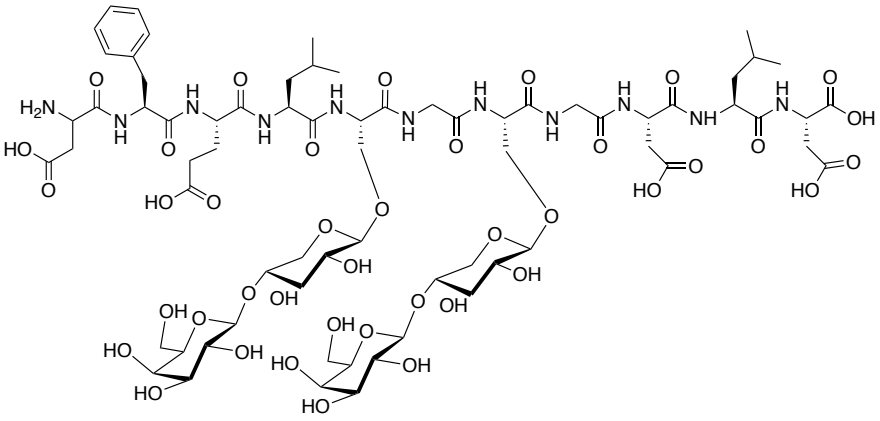


Figure 3.79 ^{1}H-NMR NMR of **14** (900 MHz, D_2O).



14

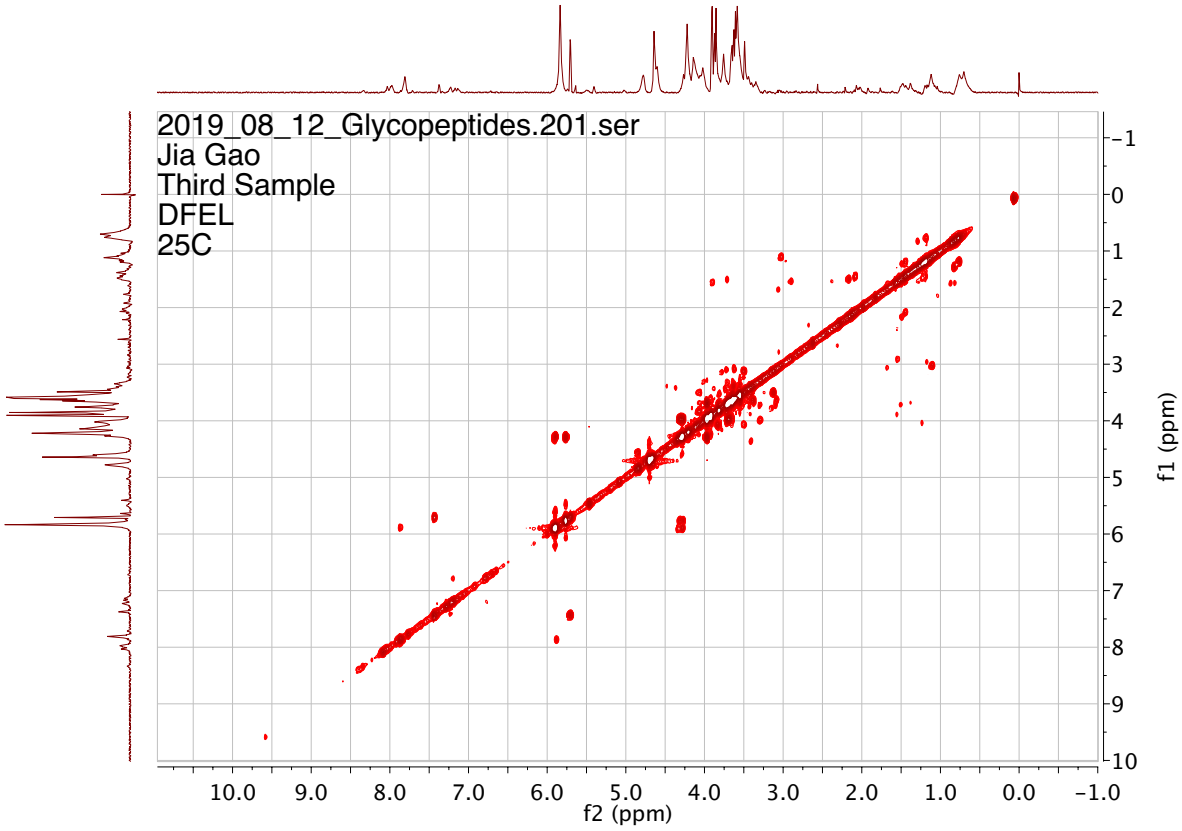
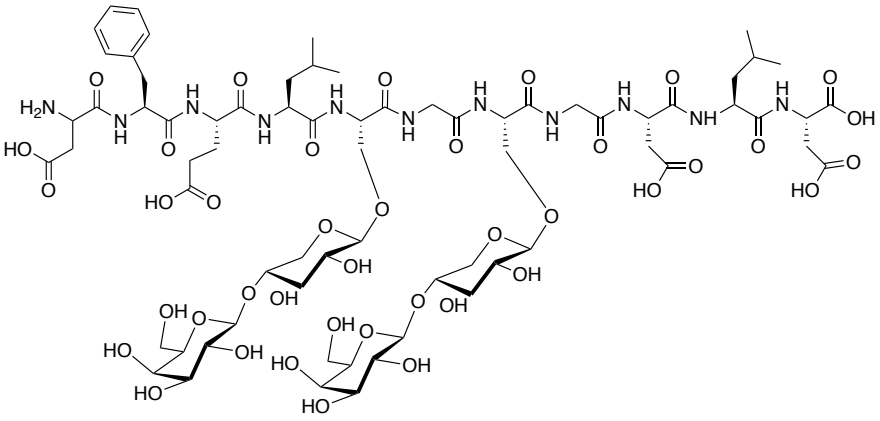


Figure 3.80 COSY NMR of **14** (900 MHz, D_2O).



14

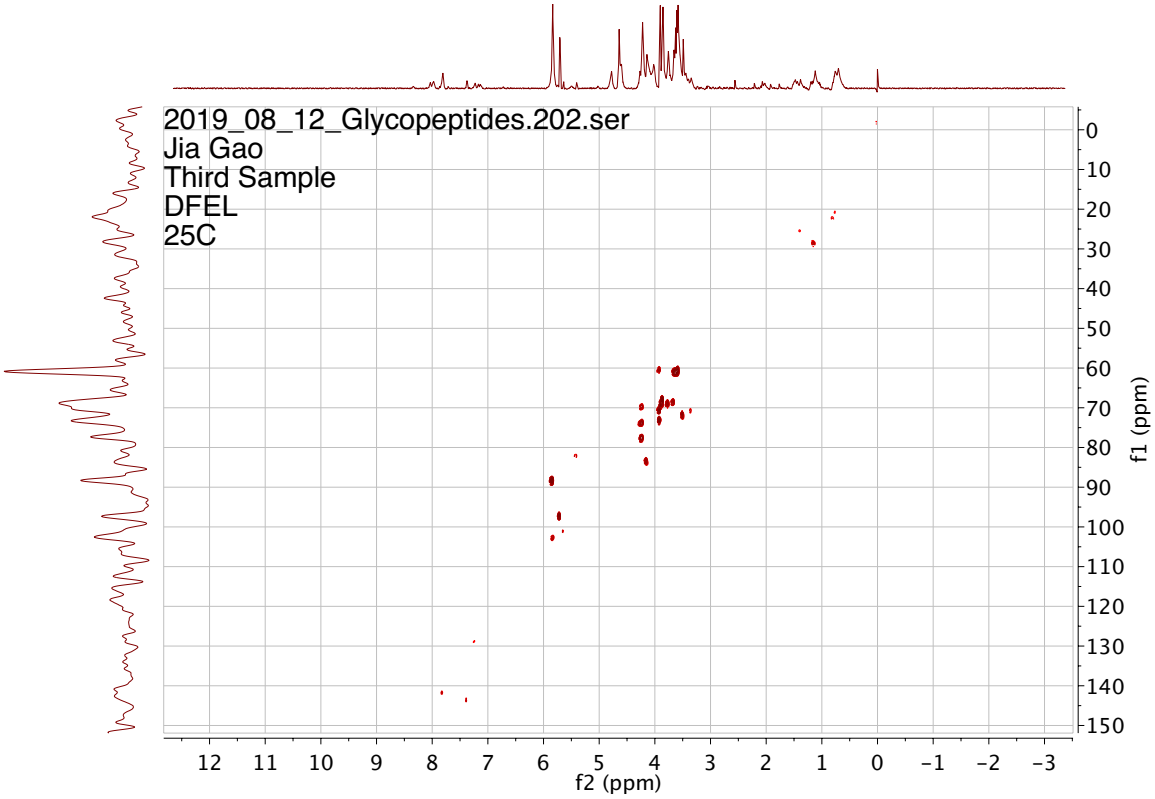
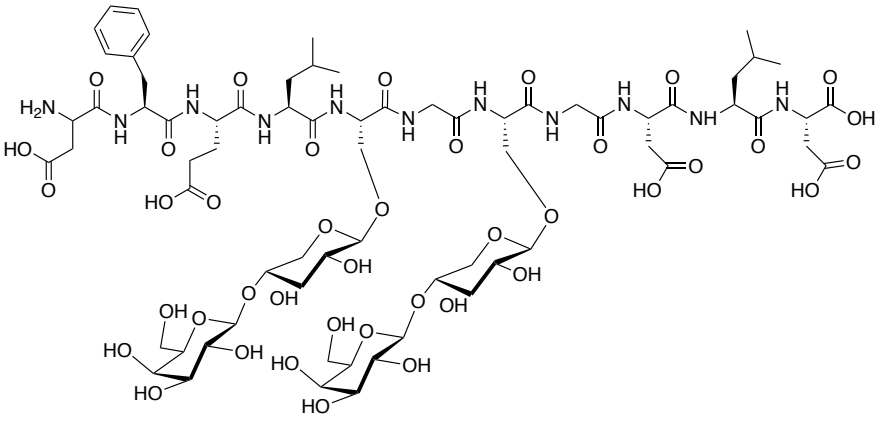
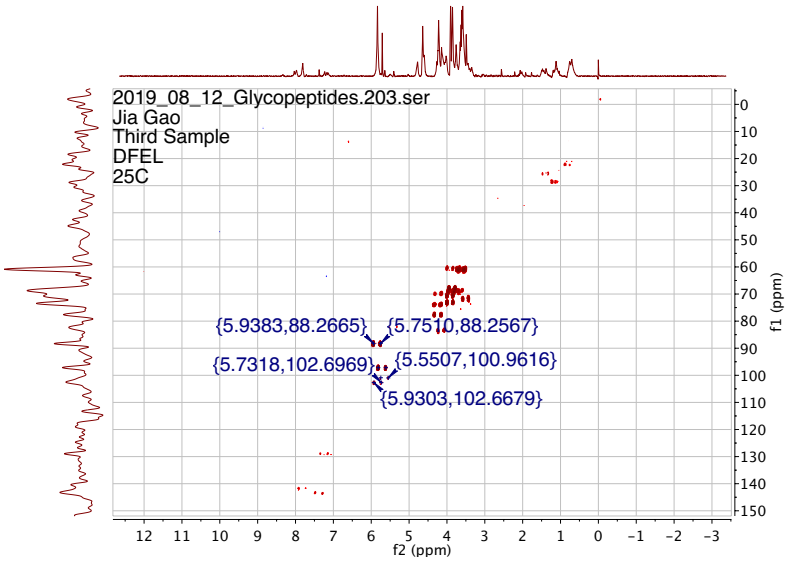


Figure 3.81 HSQC NMR of **14** (900 MHz, D_2O).



14



$^{1}J_{C1,\,H1}$ = 160.9 Hz, 156.4 Hz, 158.8 Hz, 149.8 Hz

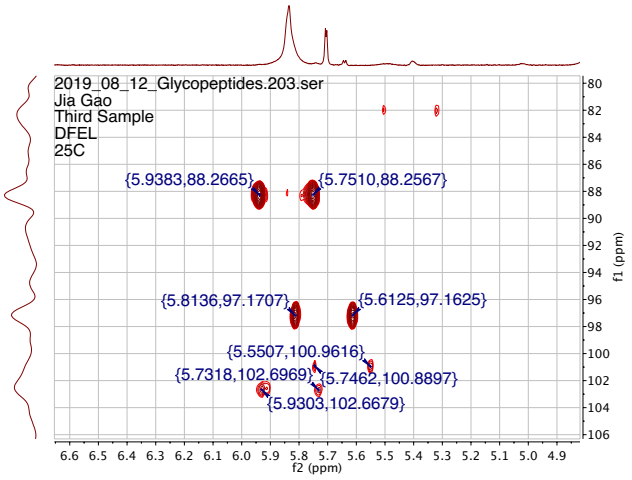
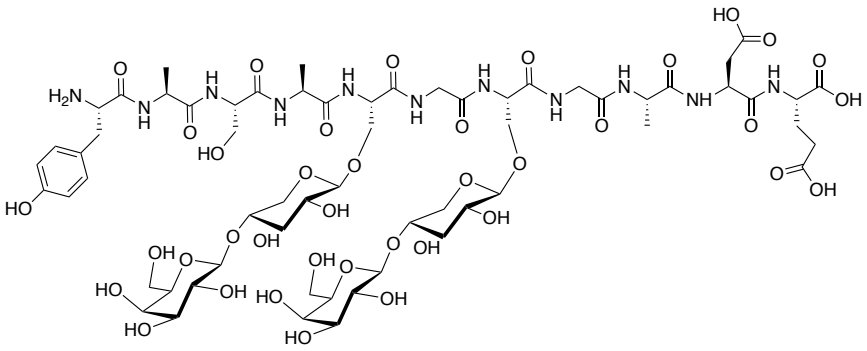
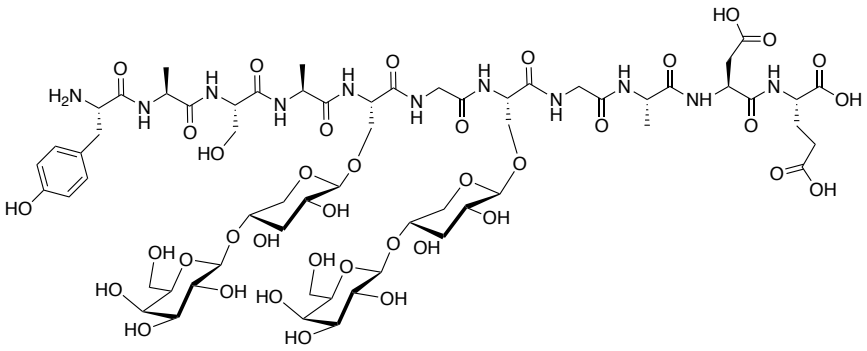


Figure 3.82 HSQC-coupled NMR of **14** (900 MHz, D_2O).



15

The purity of glycopeptide was verified with analytical C-18 HPLC (0-30-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = + 1.200 ° (c = 0.01, H_2O). ^{1}H-NMR (500 MHz, D_2O) δ 6.97 (m, 2H), 6.68 (m, 2H), 5.99 – 5.71 (m, 6H), 4.69 – 4.46 (m, 24H), 4.25 – 4.21 (m, 8H), 3.89 – 3.85 (m, 9H), 3.74 – 3.70 (m, 6H), 3.68 – 3.56 (m, 9H), 3.54 (d, *J* = 5.8 Hz, 7H), 3.42 – 3.38 (m, 7H), 1.26 – 1.22 (m, 3H), 1.21 – 1.12 (m, 3H); ^{13}C NMR (225 MHz, D_2O) δ 138.3, 130.8, 104.8, 101.9, 98.3, 95.0, 91.0, 88.4, 77.4, 72.8, 71.2, 68.6, 62.7, 60.8, 59.5, 29.2, 22.0, 18.5, 16.6. ESI-MS: $C_{62}H_{95}N_{11}O_{38}$ $[M+2H]^{2+}$ calcd: 801.7993, obsd: 801.8046 (1.62 ppm).



15

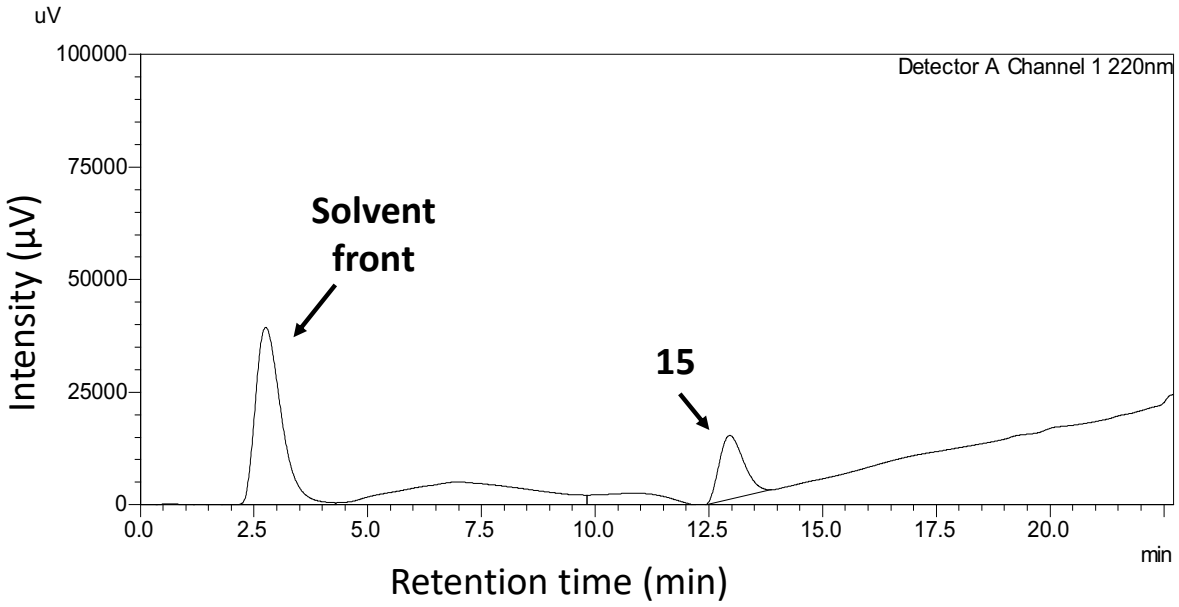
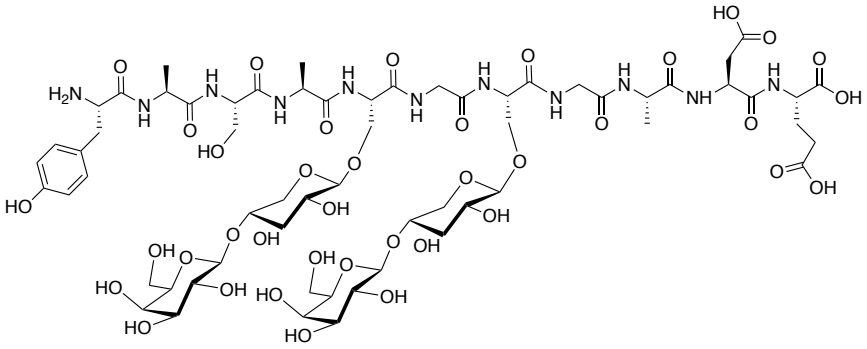


Figure 3.83 HPLC chromatogram of **15**.



15

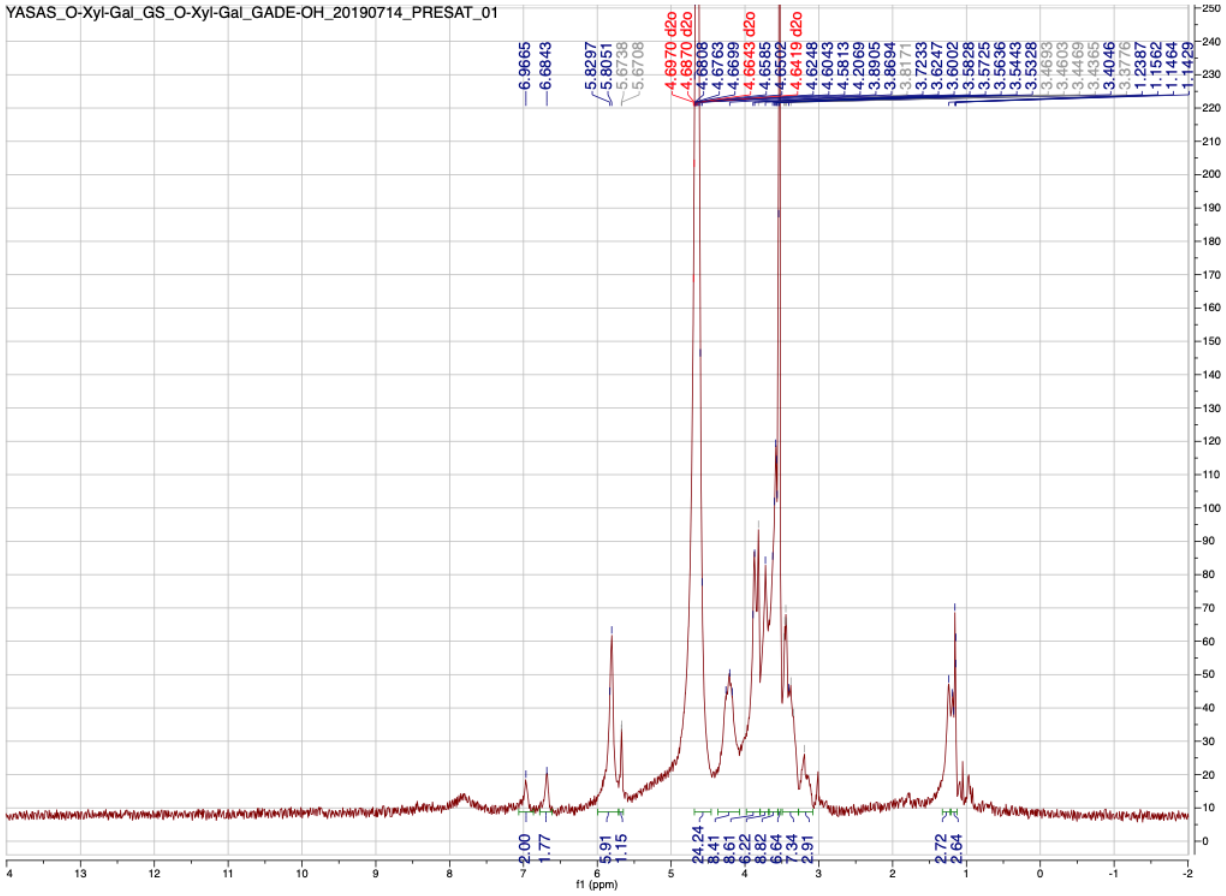
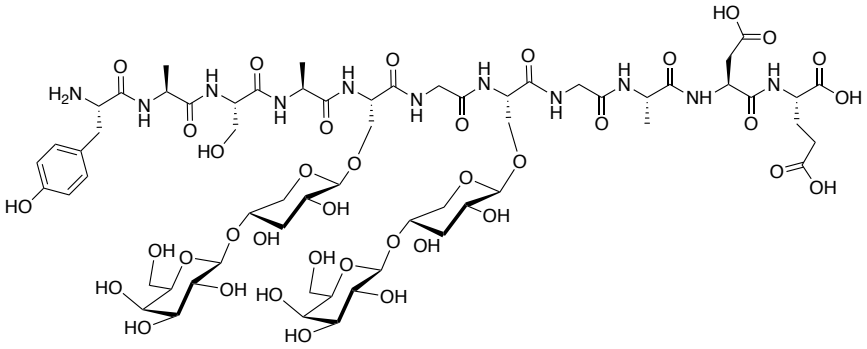


Figure 3.84 ^{1}H-NMR of **15** (500 MHz, D_2O).



15

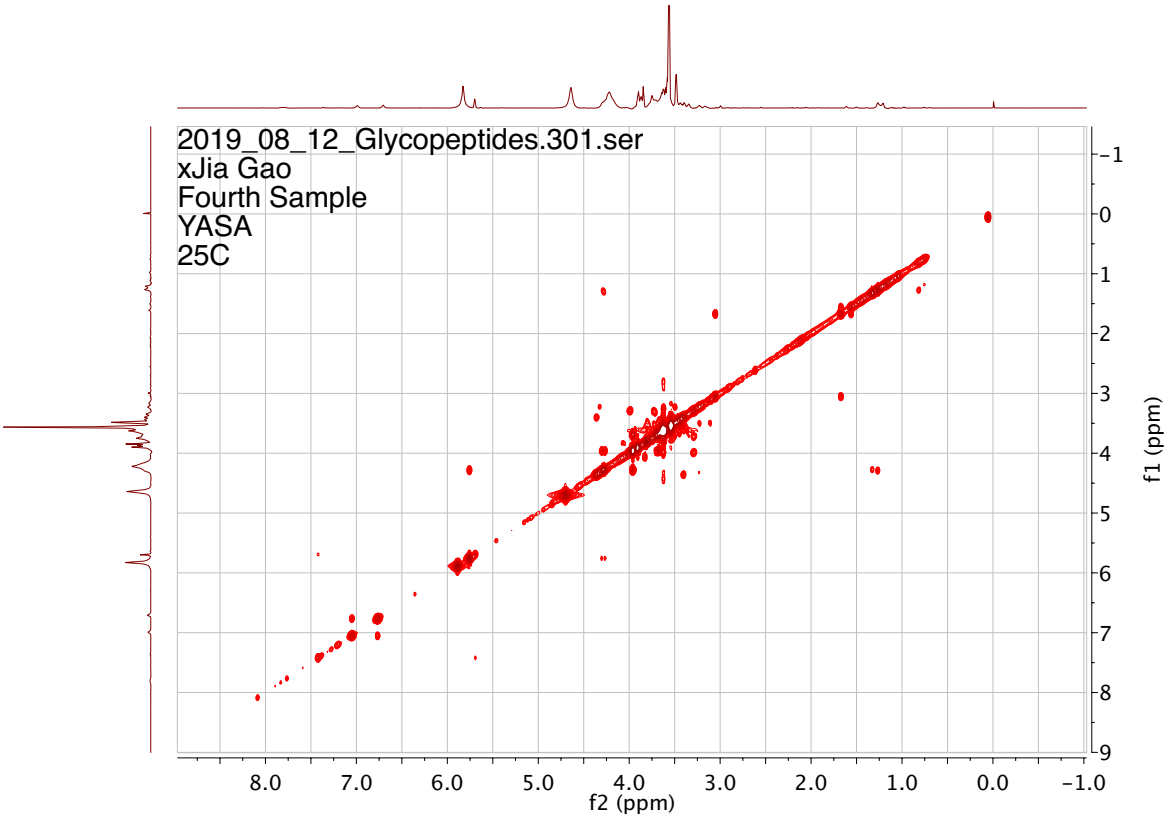
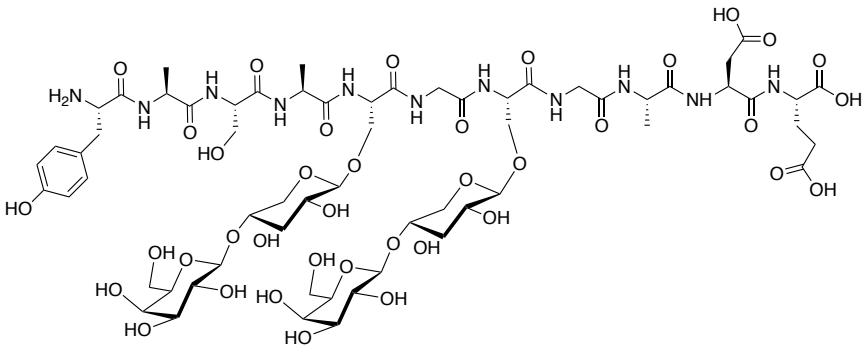


Figure 3.85 COSY NMR of **15** (900 MHz, D_2O).



15

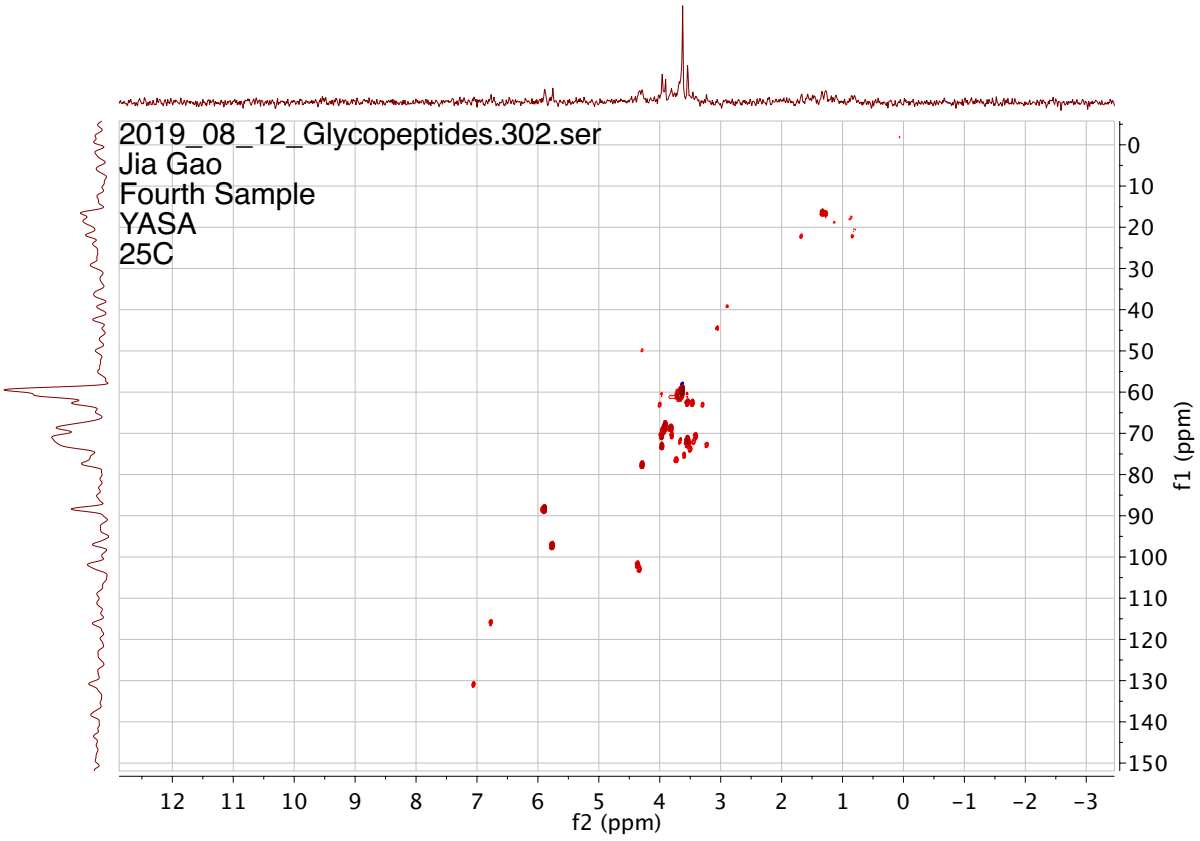
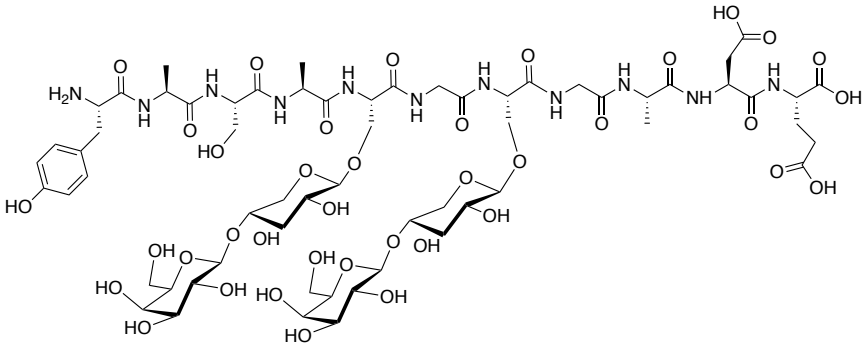
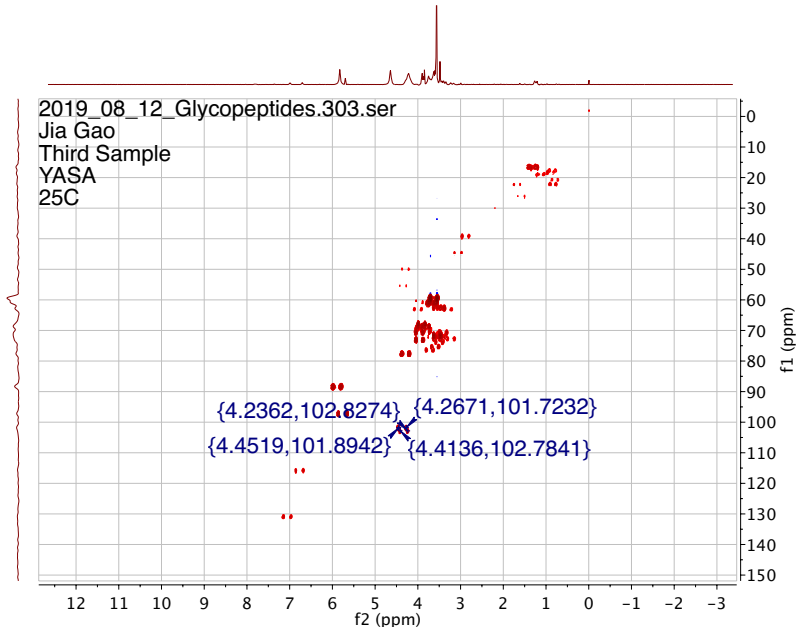


Figure 3.86 HSQC NMR of **15** (900 MHz, D_2O).



15



$^{1}J_{C1,H1}$ = 158.9 Hz,162.5 Hz; 159.7 hz,159.7 Hz

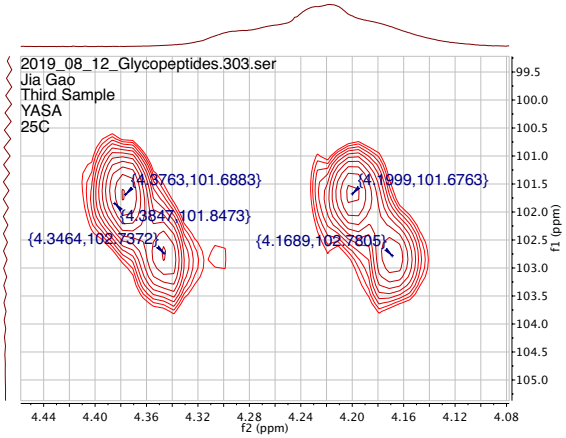
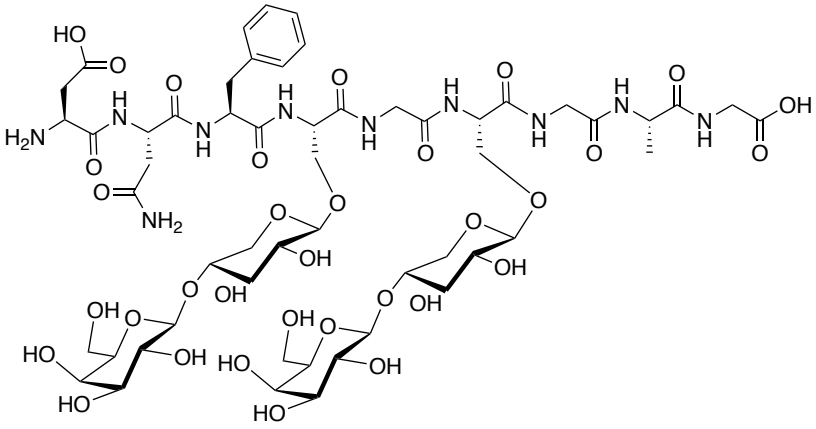
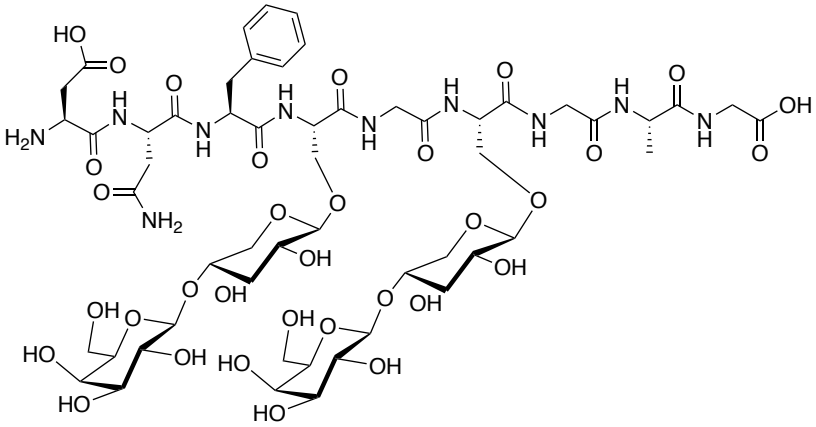


Figure 3.87 HSQC-coupled NMR of **15** (900 MHz, D_2O).



16

The purity of glycopeptide was verified with analytical C-18 HPLC (0-30-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20} = +1.100^{\circ}$ (c = 0.01, H_2O). ^{1}H-NMR (900 MHz, D_2O), δ 7.83 – 7.79 (m, 1H), 7.29 – 7.05 (m, 5H), 5.86 – 5.82 (m, 5H), 5.71 (d, *J* = 4.5 Hz, 1H), 4.77 – 4.74 (m, 3H), 4.35 – 4.08 (m, 7H), 4.06 – 4.03 (m, 1H), 3.96 – 3.84 (m, 8H), 3.84 -3.80 (m, 1H), 3.76 (d, *J* = 3.5 Hz, 2H), 3.73 – 3.61 (m, 11H), 3.61 – 3.56 (m, 11H), 3.55 (d, *J* = 9.3 Hz, 3H), 3.51 – 3.46 (m, 6H), 3.35 (q, *J* = 7.8 Hz, 2H), 3.26 – 3.22 (m, 2H), 1.26 – 1.22 (m, 4H); ^{13}C-NMR (225 MHz, D_2O) δ 128.9, 102.0, 101.7, 88.2, 72.5, 70.8, 68.7, 61.9, 60.8, 46.9, 42.6, 16.6. ESI-MS: $C_{54}H_{82}N_{10}O_{33}$ $[M+H]^+$ calcd: 1399.5119, obsd: 1399.5148 (2.07 ppm).



16

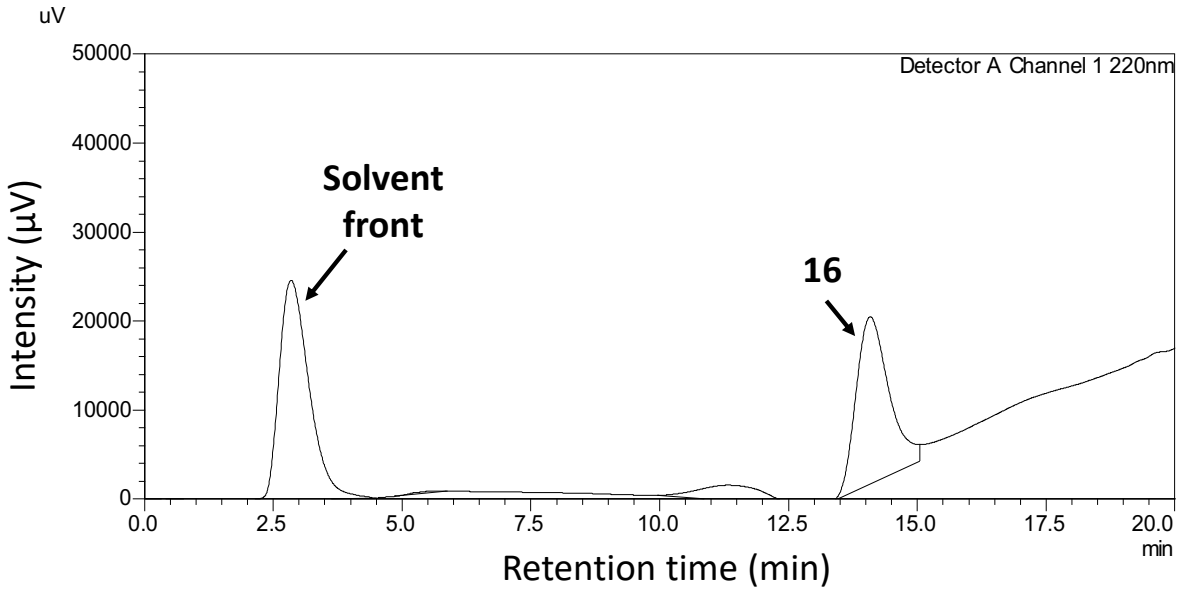
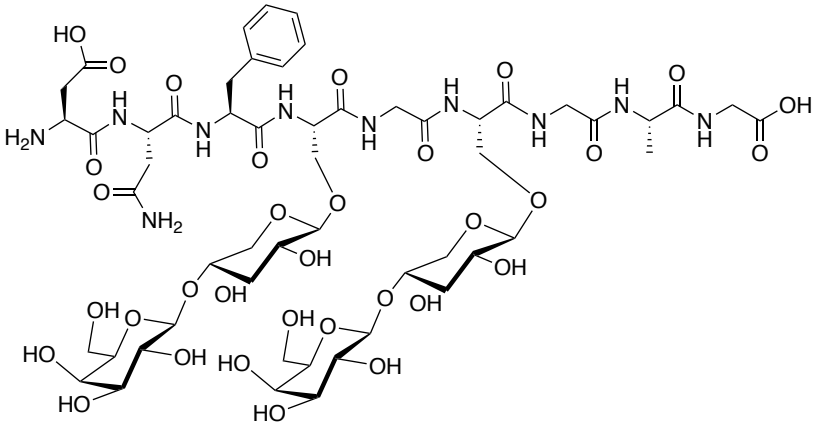


Figure 3.88 HPLC chromatogram of **16**.



16

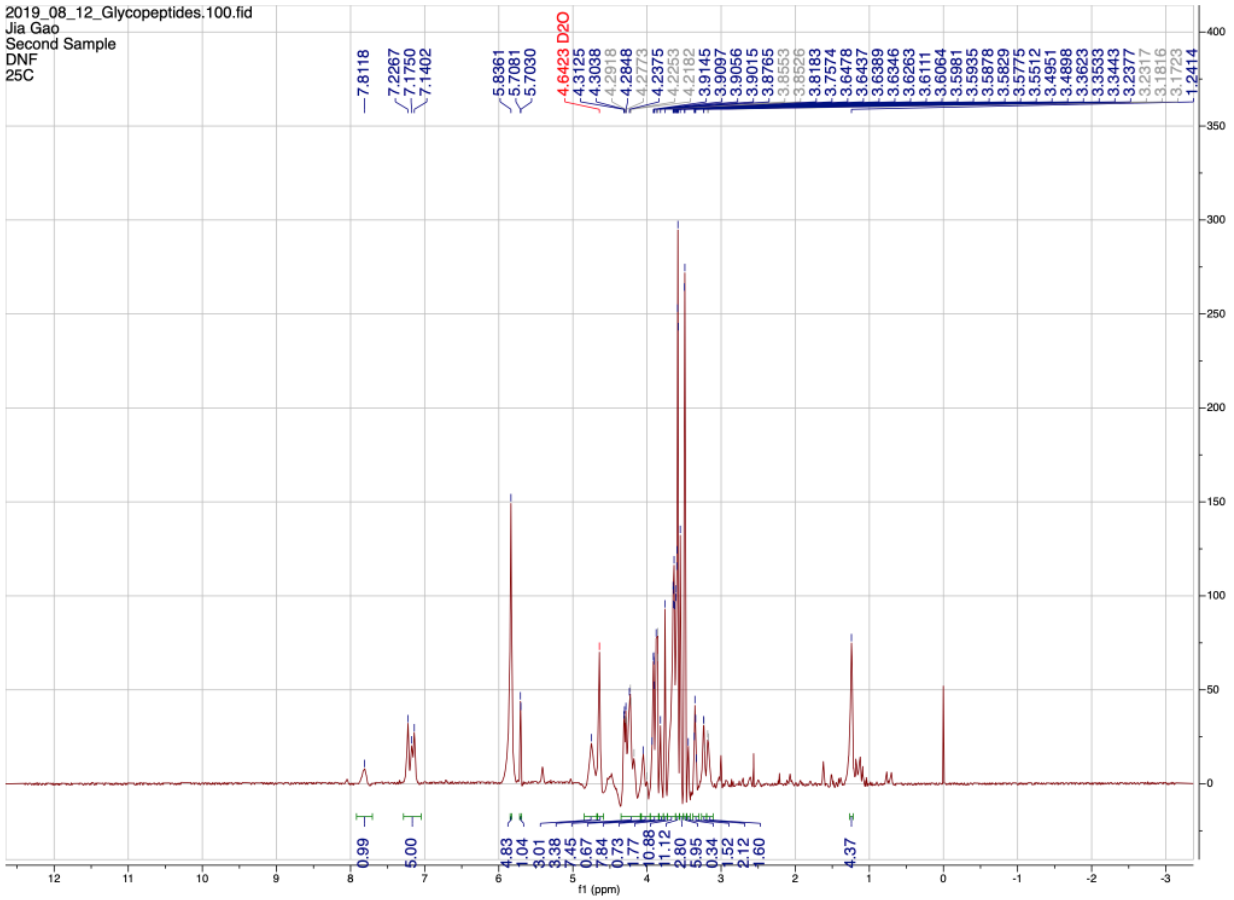
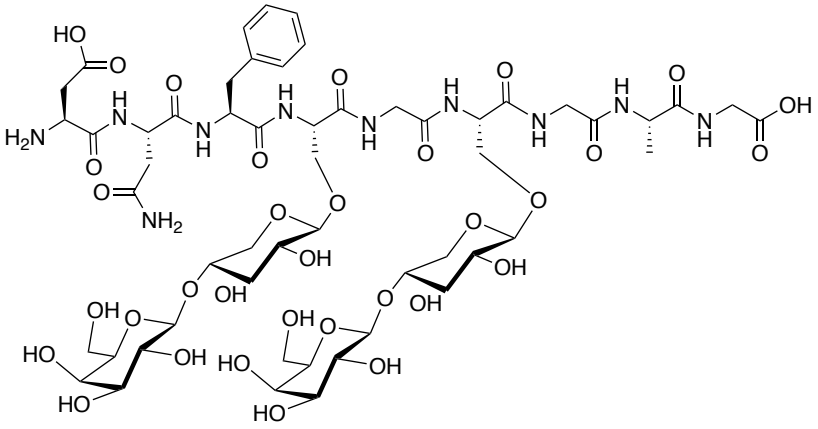


Figure 3.89 ^{1}H- NMR of **16** (900 MHz, D_2O).



16

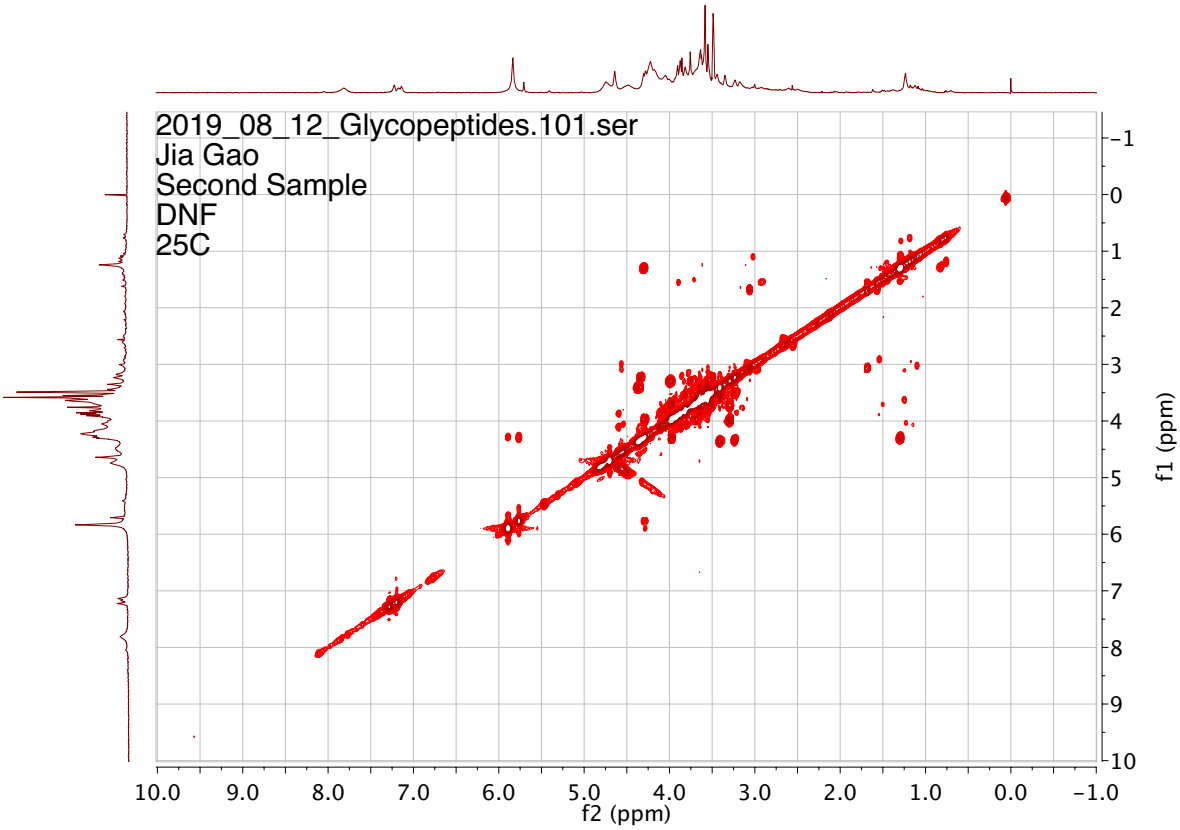
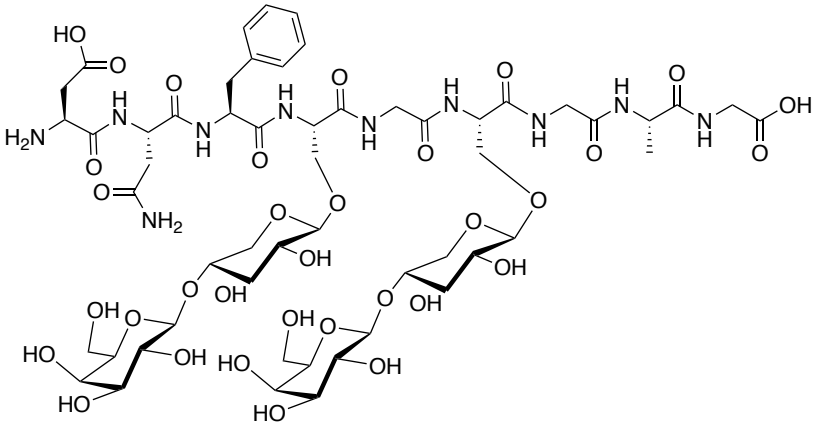


Figure 3.90 COSY NMR of **16** (900 MHz, D_2O).



16

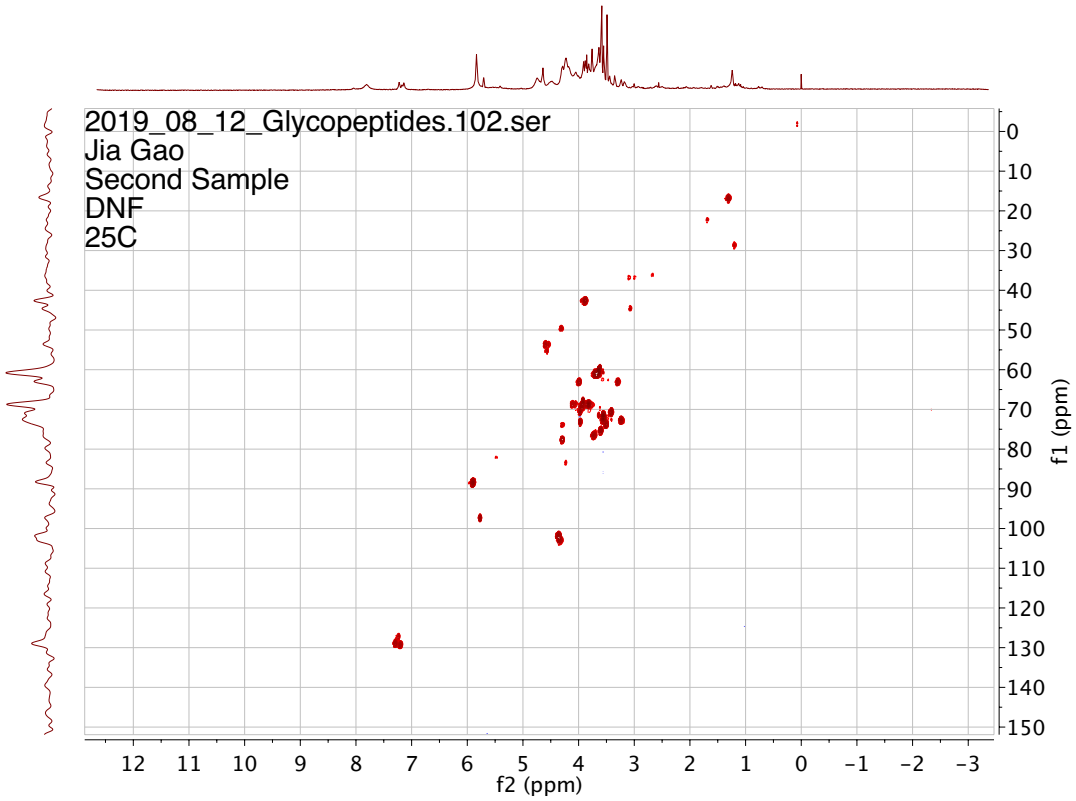
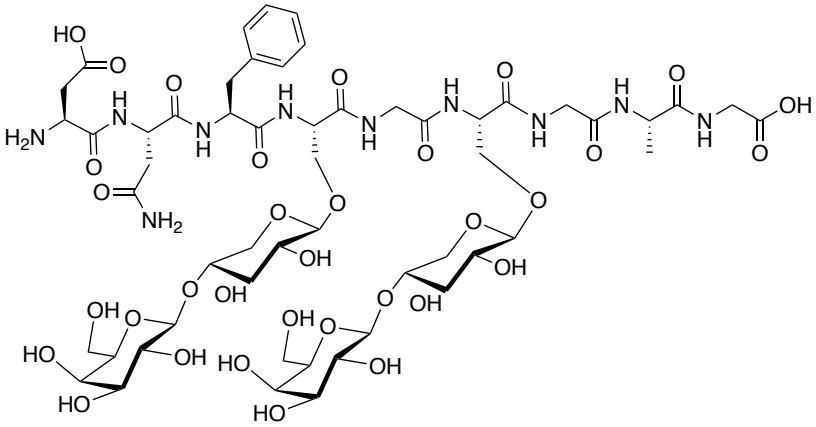
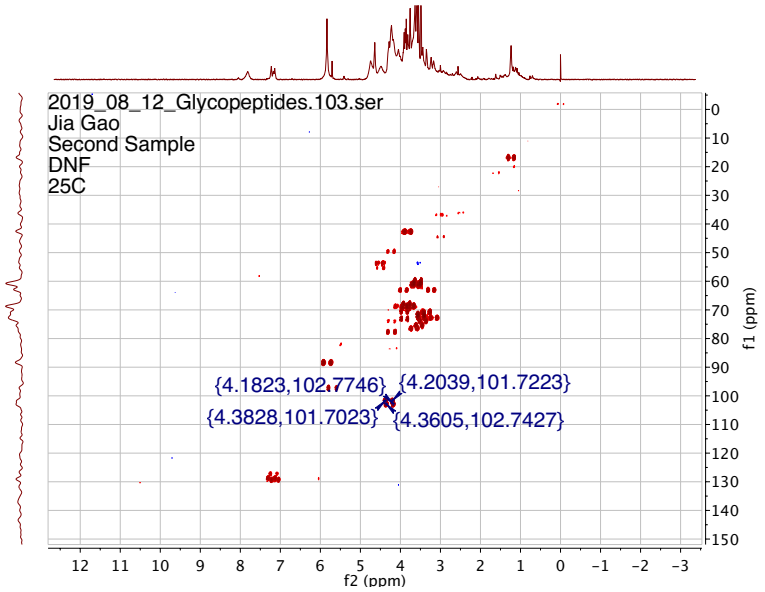


Figure 3.91 HSQC NMR of **16** (900 MHz, D_2O).



16



$^{1}J_{C1,H1}$ = 161.0 Hz,159.9 Hz; 160.4 hz,158.5 Hz

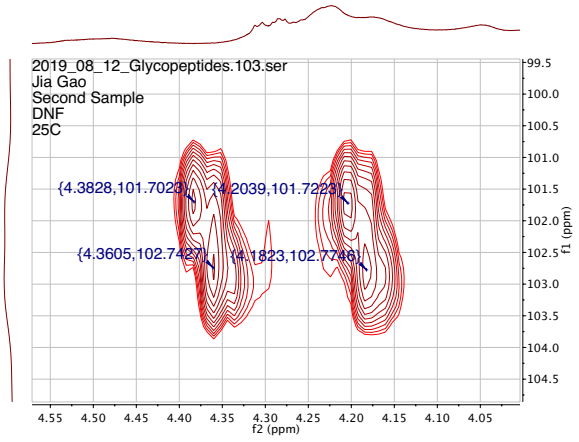
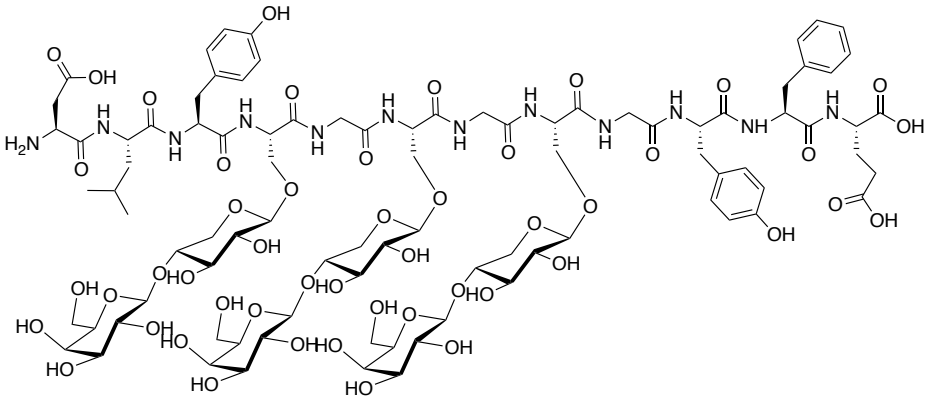
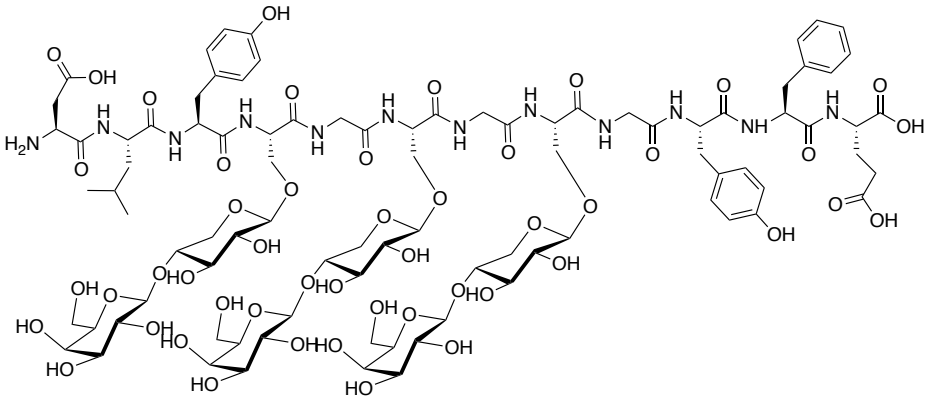


Figure 3.92 HSQC-coupled NMR of **16** (900 MHz, D_2O).



17

The purity of glycopeptide was verified with analytical C-18 HPLC (0-30-100% acetonitrile/water; 0.1% trifluoroacetic acid). $[\alpha]_D^{20} = +1.100^{\circ}$ (c = 0.01, H_2O). ^{1}H-NMR (800 MHz, D_2O), δ 7.82 – 7.78 (m, 2H), 7.18 – 7.14 (m, 3H), 7.08 – 7.04 (m, 1H), 7.00 – 6.98 (m, 2H), 6.87 – 6.83 (m, 1H), 6.67 -6.63 (m, 3H), 5.85 – 5.81 (m, 5H), 4.47 – 4.43 (m, 3H), 4.38 – 4.17 (m, 12H), 4.16 – 4.12 (m, 6H), 4.05 – 4.01 (m, 8H), 3.89 – 3.84 (m, 11H), 3.82 – 3.70 (m, 12H), 3.63 – 3.58 (m, 20H), 3.55 – 3.50 (m, 3H), 3.49 – 3.38 (m, 8H), 3.36 – 3.31 (m, 3H), 3.20 – 3.16 (m, 5H), 2.72 – 2.67 (m, 4H), 1.32 – 1.28 (m, 7H), 0.71 – 0.67 (m, 5H). ^{13}C-NMR (200 MHz, D_2O) δ 133.1, 132.0, 118.0, 105.5, 104.3, 98.5, 90.8, 79.0, 77.9, 76.3, 75.2, 73.1, 71.2, 67.9, 65.6, 63.6, 57.6, 56.1, 45.1, 31.1, 28.0, 26.5, 24.6, 23.5, 19.9, 10.9. ESI-MS: $C_{90}H_{132}N_{12}O_{49}$ $[M+2H]^{2+}$ calcd: 1082.4098 obsd: 1082.4048 (4.62 ppm).



17

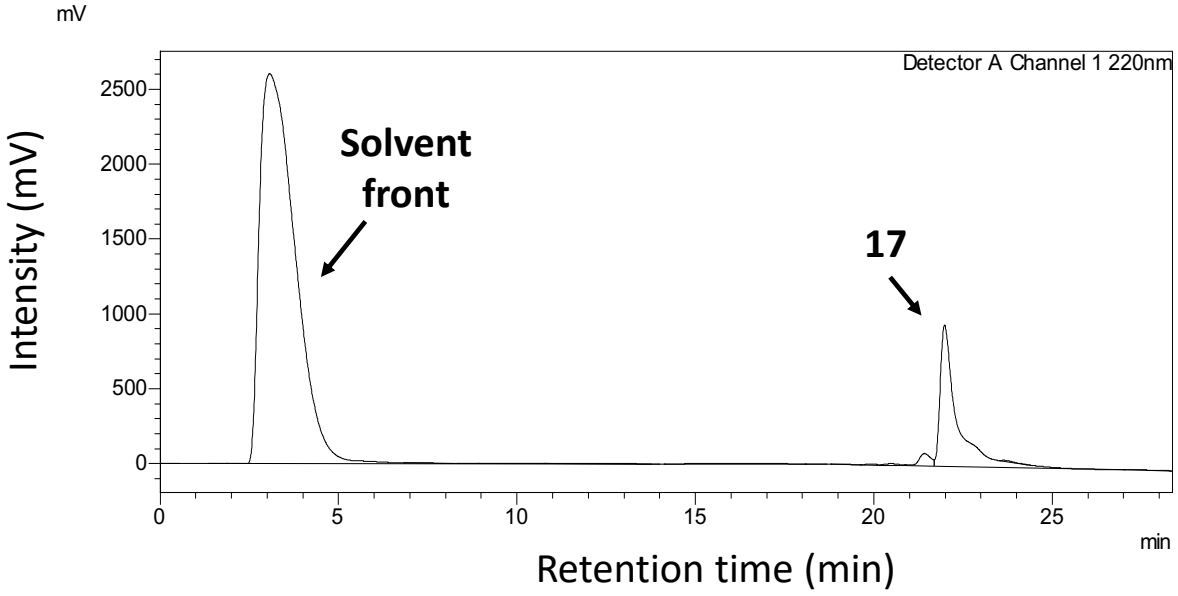
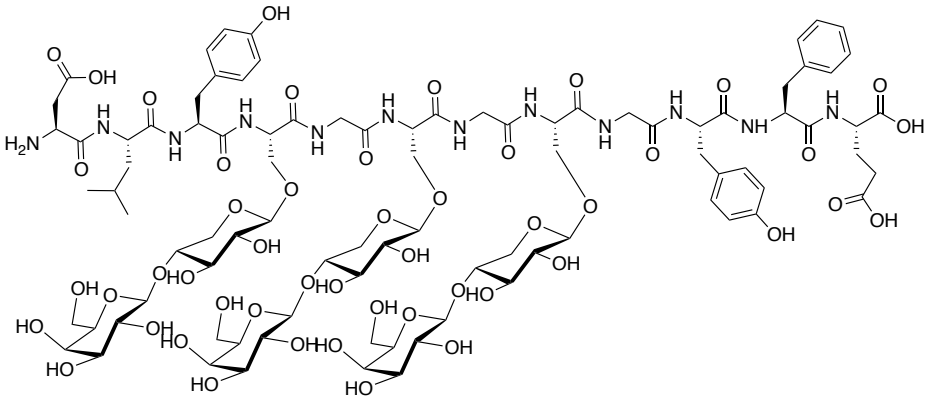


Figure 3.93 HPLC chromatogram of **17**.



17

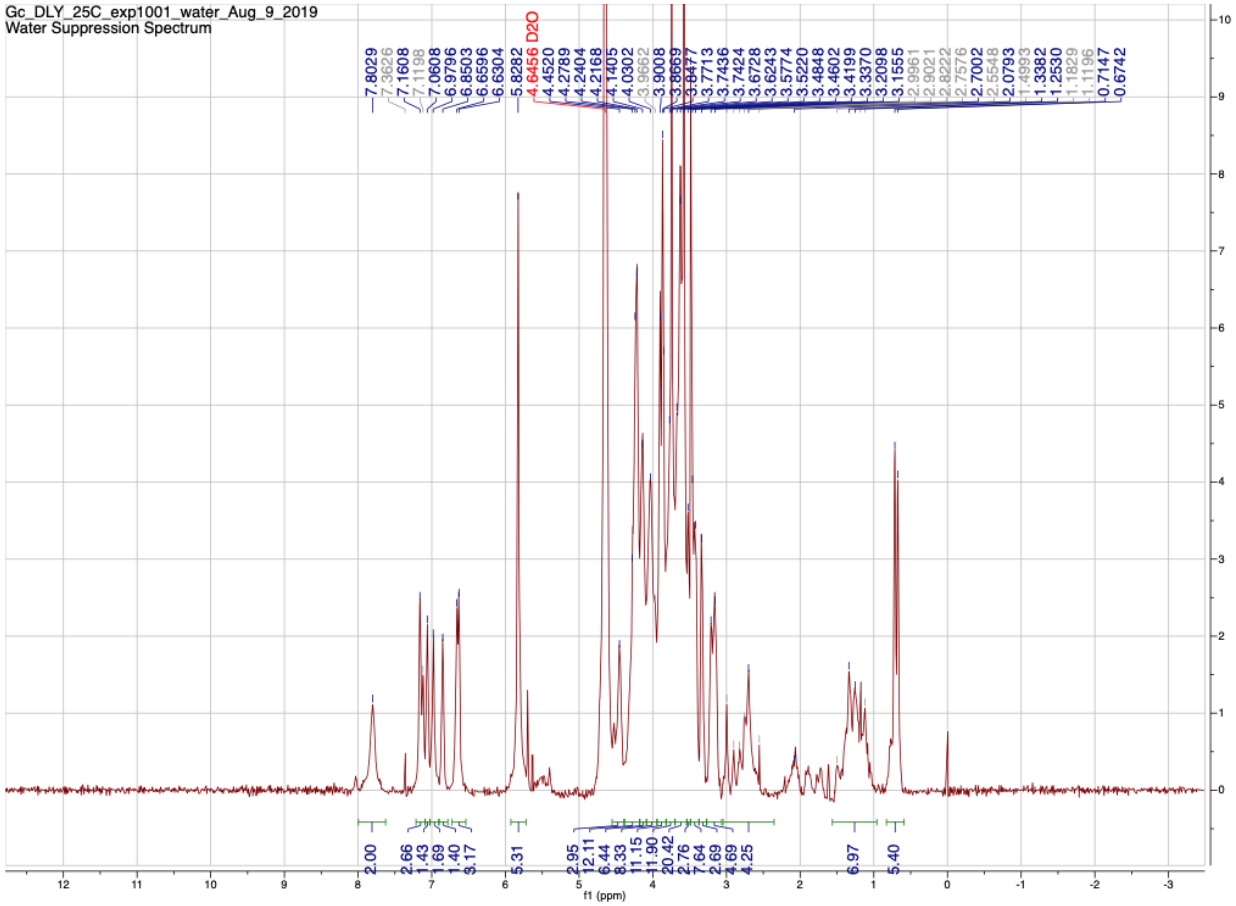
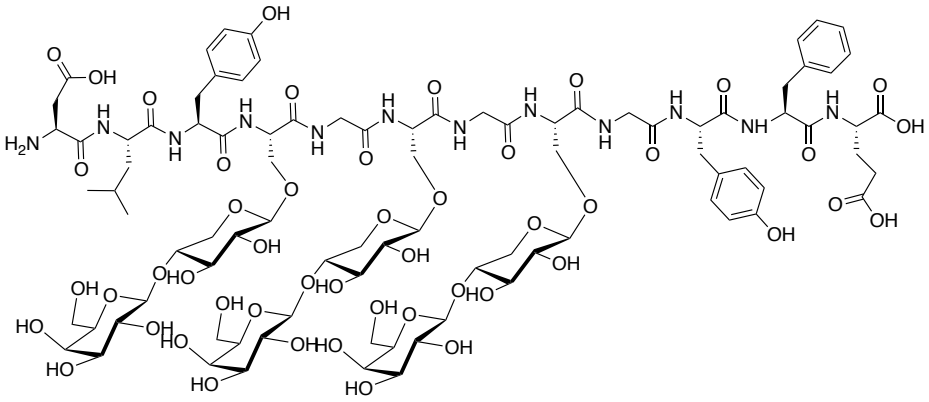


Figure 3.94 ^{1}H-NMR of **17** (800 MHz, D_2O).



17

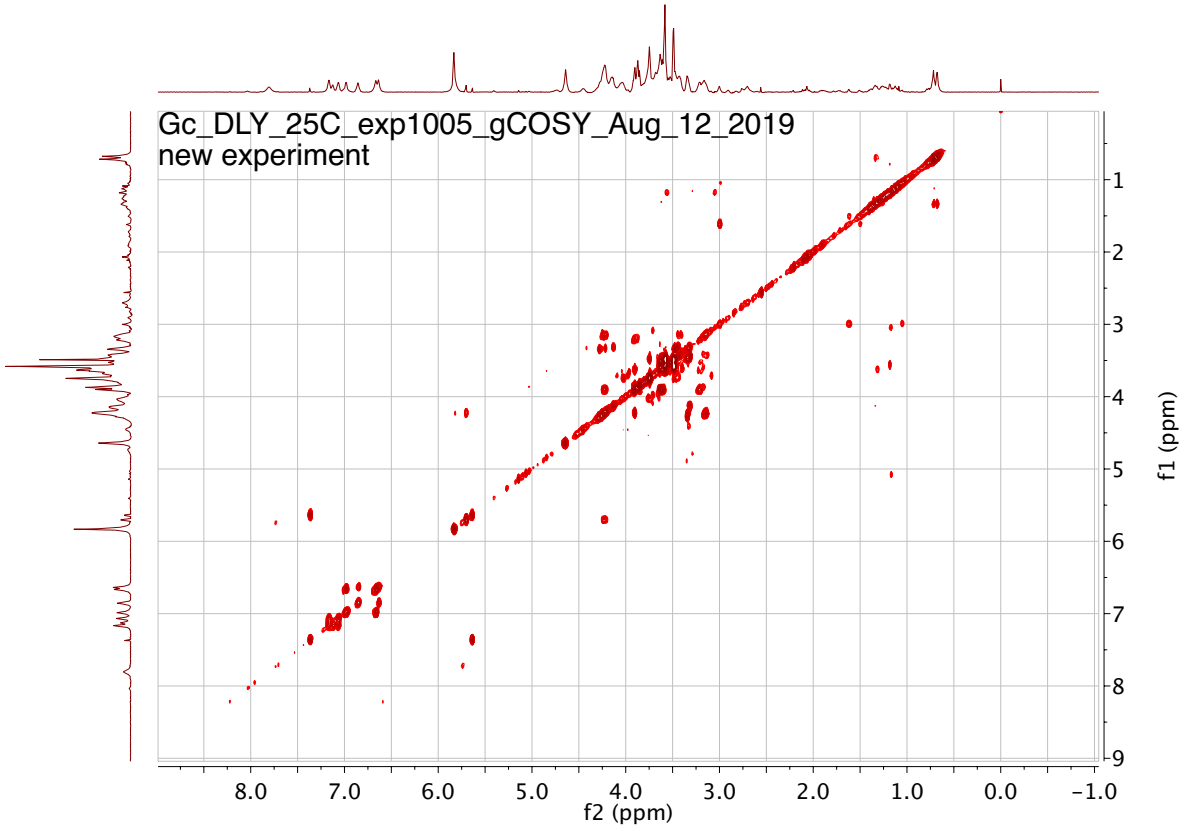
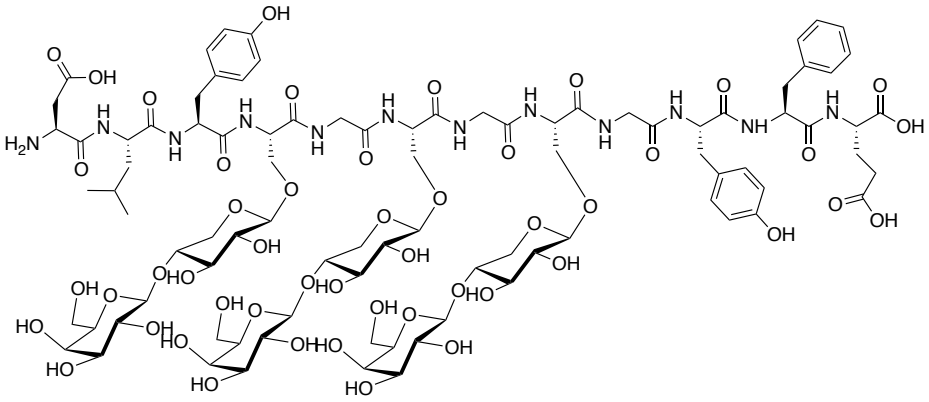


Figure 3.95 COSY NMR of **17** (800 MHz, D_2O).



17

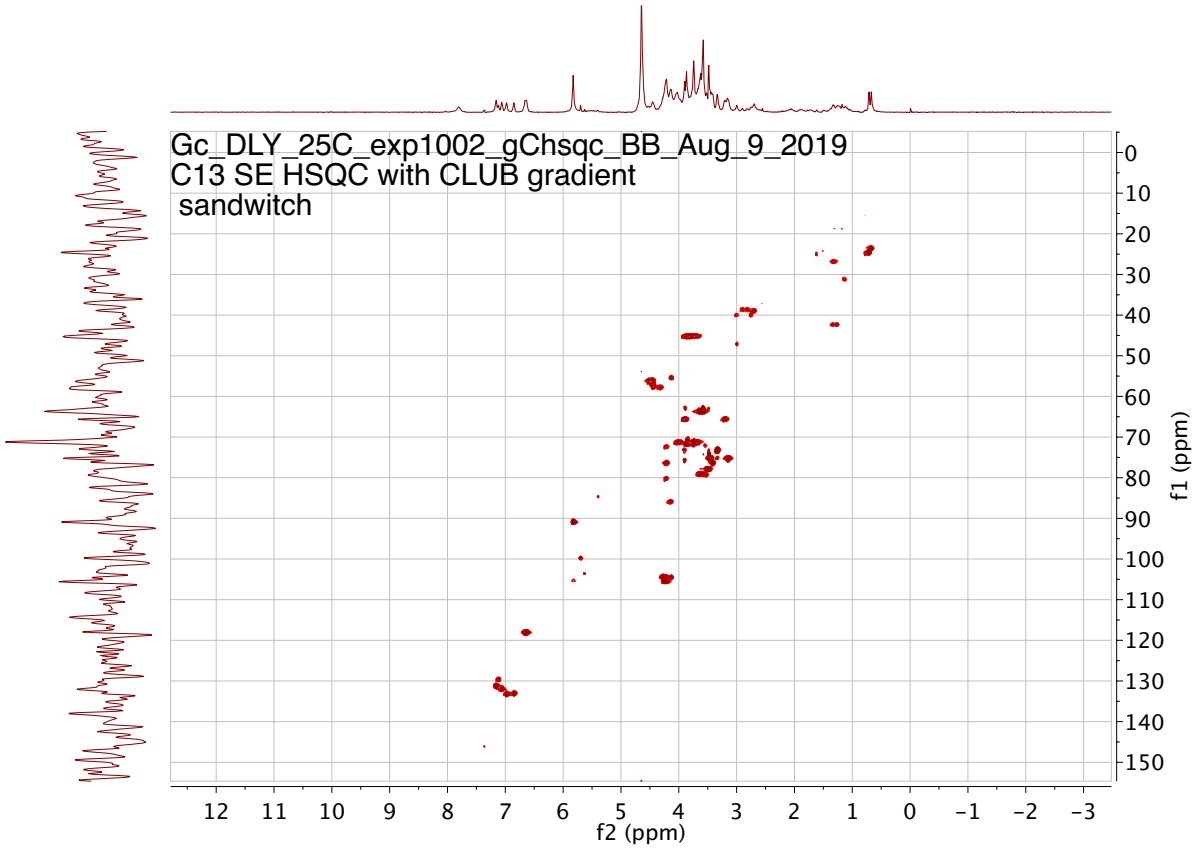
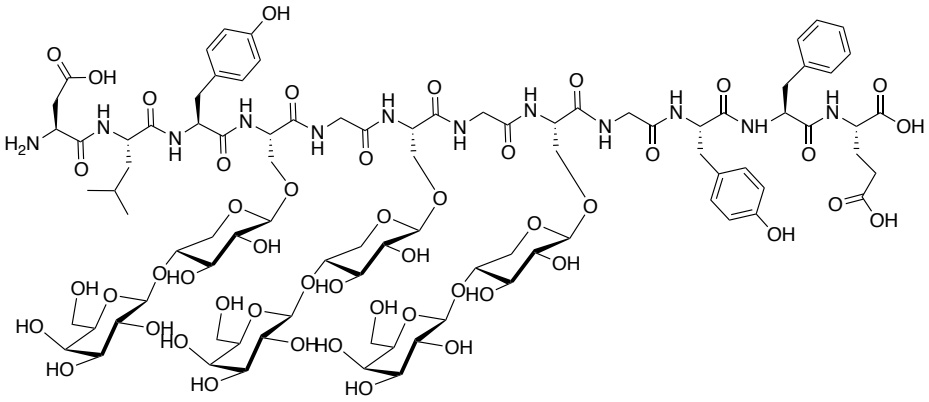
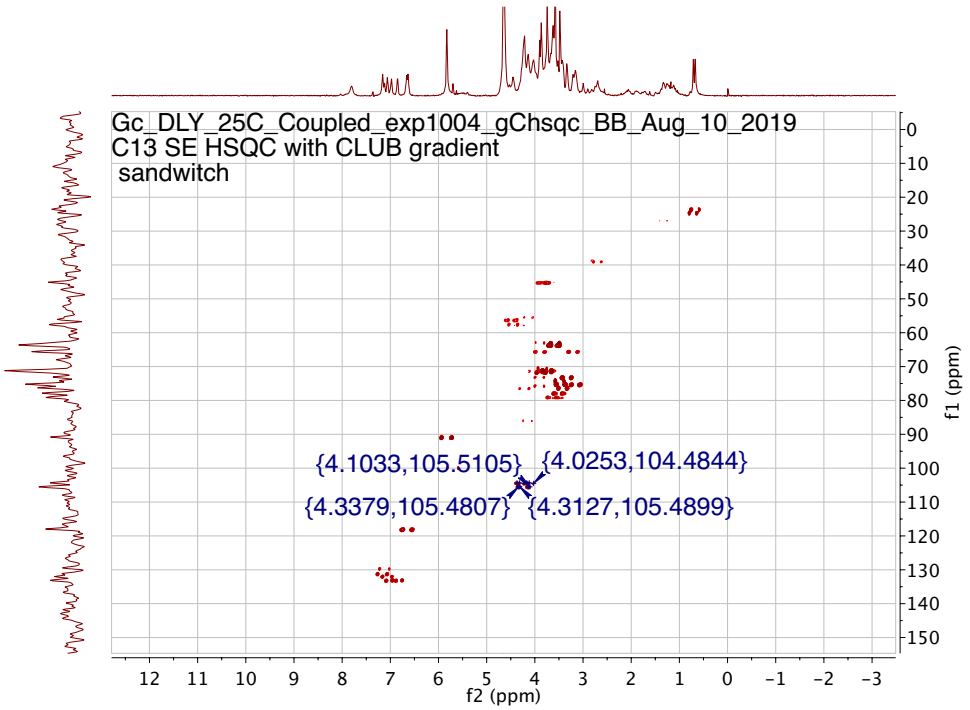


Figure 3.96 HSQC NMR of **17** (800 MHz, D_2O).



17



$^{1}J_{C1,H1}$ = 159.7 Hz, 159.7 Hz, 159.7 Hz; 162.8 Hz, 161.8 Hz, 167.5 Hz

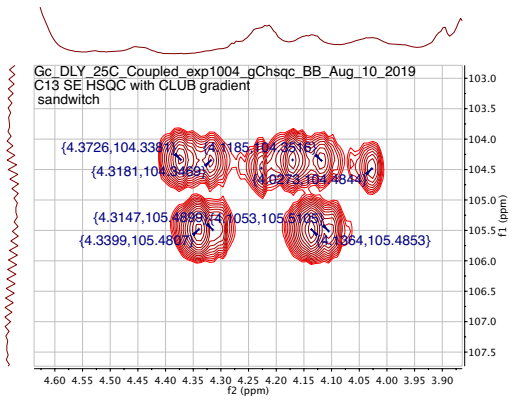


Figure 3.97 HSQC-coupled NMR of **17** (800 MHz, D_2O).

REFERENCES

REFERENCES

1. Tzanakakis, G.; Neagu, M.; Tsatsakis, A.; Nikitovic, D., Proteoglycans and Immunobiology of Cancer—Therapeutic Implications. *Front. Immunol.* **2019,** *10*, doi: 10.3389/fimmu.2019.00875.

2. Bernfield, M.; Götte, M.; Park, P. W.; Reizes, O.; Fitzgerald, M. L.; Lincecum, J.; Zako, M., Functions of Cell Surface Heparan Sulfate Proteoglycans. *Annu. Rev. Biochem.* **1999,** *68*, 729-777.

3. Lin, X., Functions of Heparan Sulfate Proteoglycans in Cell Signaling during Development. *Development* **2004,** *131*, 6009-6021.

4. Yanagishita, M., Function of Proteoglycans in the Extracellular Matrix. *Acta Pathol. Jpn.* **1993,** *43*, 283-293.

5. Nikitovic, D.; Berdiaki, A.; Spyridaki, I.; Krasanakis, T.; Tsatsakis, A.; Tzanakakis, G. N., Proteoglycans-Biomarkers and Targets in Cancer Therapy. *Front Endocrinol.* **2018,** *9*, 69.

6. Esko, J. D.; Zhang, L., Influence of Core Protein Sequence on Glycosaminoglycan Assembly. *Curr. Opin. Struc. Biol.* **1996,** *6*, 663-670.

7. Yang, W.; Eken, Y.; Zhang, J.; Cole, L. E.; Ramadan, S.; Xu, Y.; Zhang, Z.; Liu, J.; Wilson, A.; 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*, 6393-6404 and references cited therein.

8. Li, T.; Yang, W.; Ramadan, S.; Huang, X., Synthesis of *O*-Sulfated Human Syndecan-1-like Glyco-polypeptides by Incorporating Peptide Ligation and *O*-Sulfated Glycopeptide Cassette Strategies. *Org. Lett.* **2020,** *22*, 6429-6433 and references cited therein.

9. 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.

10. Daligault, F.; Rahuel-Clermont, S.; Gulberti, S.; Cung, M. T.; Branlant, G.; Netter, P.; Magdalou, J.; Lattard, V., Thermodynamic Insights into the Structural Basis Governing the Donor Substrate Recognition by Human Beta1,4-Galactosyltransferase 7. *Biochem. J.* **2009,** *418*, 605-614.

11. Talhaoui, I.; Bui, C.; Oriol, R.; Mulliert, G.; Gulberti, S.; Netter, P.; Coughtrie, M. W.; Ouzzine, M.; Fournel-Gigleux, S., Identification of Key Functional Residues in the Active Site of Human Beta1,4-Galactosyltransferase 7: A Major Enzyme in the Glycosaminoglycan Synthesis Pathway. *J. Biol. Chem.* **2010,** *285*, 37342-37358.

12. Chua, J. S.; Kuberan, B., Synthetic Xylosides: Probing the Glycosaminoglycan Biosynthetic Machinery for Biomedical Applications. *Acc. Chem. Res.* **2017,** *50*, 2693-2705.

13. Saliba, M.; Ramalanjaona, N.; Gulberti, S.; Bertin-Jung, I.; Thomas, A.; Dahbi, S.; Lopin-Bon, C.; Jacquinet, J. C.; Breton, C.; Ouzzine, M.; Fournel-Gigleux, S., Probing the Acceptor Active Site Organization of the Human Recombinant Beta1,4-Galactosyltransferase 7 and Design of Xyloside-Based Inhibitors. *J. Biol. Chem.* **2015,** *290*, 7658-7670.

14. 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 B4galt7 Active Site. *Chem. Sci.* **2014,** *5*, 3501-3508.

15. Siegbahn, A.; Thorsheim, K.; Stahle, J.; Manner, S.; Hamark, C.; Persson, A.; Tykesson, E.; Mani, K.; Westergren-Thorsson, G.; Widmalm, G.; Ellervik, U., Exploration of the Active Site of Beta4galt7: Modifications of the Aglycon of Aromatic Xylosides. *Org. Biomol. Chem.* **2015,** *13*, 3351-3362.

16. Willen, D.; Bengtsson, D.; Clementson, S.; Tykesson, E.; Manner, S.; Ellervik, U., Synthesis of Double-Modified Xyloside Analogues for Probing the beta4GalT7 Active Site. *J. Org. Chem.* **2018,** *83*, 1259-1277.

17. Thorsheim, K.; Willen, D.; Tykesson, E.; Stahle, J.; Praly, J. P.; Vidal, S.; Johnson, M. T.; Widmalm, G.; Manner, S.; Ellervik, U., Naphthyl Thio- and Carba-xylopyranosides for Exploration of the Active Site of beta-1,4-Galactosyltransferase 7 (beta4GalT7). *Chemistry* **2017,** *23*, 18057-18065.

18. Gebhard, W.; Schreititmüller, T.; Vetr, H.; Wachter, E.; Hochstrasser, K., Complementary DNA and Deduced Amino Acid Sequences of Porcine A1-Microglobulin and Bikunin. *FEBS Lett.* **1990,** *269*, 32-36.

19. 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*, 10185-10189.

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. Echalier, C.; Al-Halifa, S.; Kreiter, A.; Enjalbal, C.; Sanchez, P.; Ronga, L.; Puget, K.; Verdie´, P.; Amblard, M.; Martinez, J.; Subra, G., Heating and Microwave Assisted SPPS of C-Terminal Acid Peptides on Trityl Resin: The Truth behind the Yield. *Amino Acids* **2013,** *45*, 1395-1403.

22. Bock, K.; Pedersen, C., A Study of ^{13}C-H Coupling Constants in Hexopyranoses. *J. Chem. Soc., Perkin Tran*s. *2* **1974,** 293-297.

23. Wu, Z. L.; Ethen, C. M.; Prather, B.; Machacek, M.; Jiang, W., Universal Phosphatase-Coupled Glycosyltransferase Assay. *Glycobiology* **2011,** *21*, 727-733.

24. Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L.; Kent, K. C.; Li, L., Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)- enabled Intact Glycopeptide/Glycoproteome Characterization. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1751-1764.

25. Tsutsui, Y.; Ramakrishnan, B.; Qasba, P. K., Crystal Structures of Beta-1,4-Galactosyltransferase 7 Enzyme Reveal Conformational Changes and Substrate Binding. *J. Biol. Chem.* **2013,** *288*, 31963-31970.

26. 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*, 10185-10189.

27. Cai, D.; Guan, D.; Ding, K.; Huang, W., Regioselective Protection of Xylose for Efficient Synthesis of Arabino-Alpha-1,3-Xylosides. *Tetrahedron Letters* **2017,** *58*, 4293-4295.

28. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25*, 1605-1612.

29. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., Autodock4 And Autodocktools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009,** *30*, 2785-2791.

Chapter 4 Exploration of Human Xylosyltransferase for Chemoenzymatic Synthesis of Proteoglycan Linkage Region

4.1 Introduction

Proteoglycans (PGs) are an essential class of glycoproteins that are ubiquitous in the mammalian systems. They are directly involved in numerous biological processes including tumor progression, cell adhesion, and regulation of growth factors.[1-3] Structurally, PGs consist of a core protein and one or more glycosaminoglycan (GAG) chains, which are linked through glucuronic acid (GlcA)-β-1,3-galactose (Gal)-β-1,3-Gal-β-1,4-xylose (Xyl) tetrasaccharide linkages attached to serine residues of serine-glycine dipeptides.[4] Due to the complexity of post-translational modifications on the GAG chains, PGs from natural sources are highly heterogeneous. To date, structurally defined proteoglycan glycopeptides can only be prepared through chemical synthesis. However, the general synthetic process is highly challenging and tedious, owing to the presence of many sensitive functional groups, thus requiring meticulous designs of the protective group strategy and the synthetic route.[5-7] To expedite the PG preparations, we have become interested in developing a synthetic strategy deploying the enzymes involved in biosynthetic assembly of the tetrasaccharide linkage. Herein, I report my results on the utility of human xylosyltransferase I (XT-I), the enzyme responsible for initiating PG synthesis in humans.

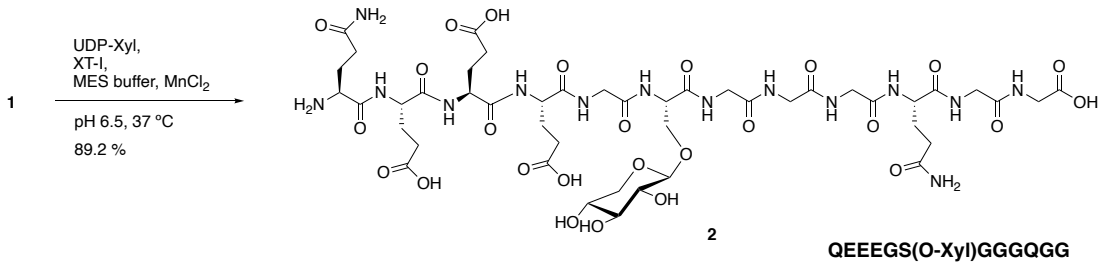
XT-I natively catalyze the transfer of the Xyl from UDP-Xyl to the side chain of certain serine residues in the PG core protein.[8-10] A consensus sequence for peptide acceptors has been deduced as Gly-Ser-Gly or Ser-Gly-x-Gly (x being any natural amino acid), with acidic residues commonly present near the GAG attachment site.[8, 11, 12] Till now, XT-I has not been utilized for synthetic purposes of the PG. I report for the first time that human XT-I enzyme can be used to efficiently synthesize native xylosylated PG glycopeptides at milligram scale, and the combination of human XT-I with human β-4-galactosyl transferase 7 (β4GalT7)[13-15] enabled one pot synthesis

of glycopeptides bearing Gal-Xyl disaccharides. Moreover, I investigated XT-I donor promiscuity. Its ability to transfer an unnatural donor such as 6-azidoglucose (6AzGlc) opens the door to introduce a biorthogonal handle to label peptide and protein substrates.

4.2 Results and Discussions

To explore the synthetic potential and capability of XT-I, we selected a bikunin-like peptide sequence QEEEGSGGGQGG as the initial peptide substrate.[16, 17] The preparation of QEEEGSGGGQGG was achieved with Fmoc-based solid-phase peptide synthesis (SPPS) using Cl-MPA ProTide resin under microwave condition. Acidic treatment of the peptide loaded resins cleaved the peptide off the resin followed by Fmoc-removal from the *N*-terminus leading to 43.2% isolated yield of bikunin peptide **1** (Appendix **Scheme 4.4**).

To express the polyhistidine-tagged human XT-I (EC 2.4.2.26),[12] plasmid encoding signal peptide-His6-XT-I was constructed and used to transfect HEK-293F cells (Appendix **Figure 4.3**). Secreted XT-I protein was purified using a Ni Sepharose affinity column with an expression yield of 5 mg/L. Xylosylation was then initiated by sequentially adding UDP-Xyl (1.2 equiv), peptide **1** (1 equiv), and XT-I (0.025 mol%) to the MES reaction buffer. After overnight incubation at 37 ºC, quantitative conversion of **1** to xylosylated glycopeptide **2** (**Scheme 4.1**) was confirmed with high-resolution mass spectrometry (HRMS) and high-performance liquid chromatography (HPLC). The desired glycopeptide product **2** was isolated via G-10 size exclusion chromatography in 89.2% yield at milligram scales. HRMS and nuclear magnetic resonance (NMR) analyses confirmed the structure of β-glycosylated product ($^1J_{C1,\,H1}$ =159.5 Hz),[18] which was identical to the chemically synthesized glycopeptide **2**.[19]



Scheme 4.1 XT-I-catalyzed xylosylation of bikunin peptide **1**.

Investigation furthered with peptide substrates **3-6** (**Figure 4.1** and **Table 4.5**),[20-23] which contain diverse residues, including hydrophilic or hydrophobic residues flanking the glycosylation site. In addition, peptides **4** and **5** have two potential sites of glycosylation, while peptide **6** has three sites. Excitingly, XT-I enzyme smoothly converted all the peptide substrates to the glycosylated products with desired stereoselectivity (**Table 4.1**). All glycopeptide structures were confirmed through HPLC, NMR, and MS comparisons with glycopeptides synthesized chemically.[19] In addition, a recombinant polyhistidine-tagged human CD44 hyaluronic acid binding domain protein ($hCD44_{20\text{-}178}$)[24] was successfully xylosylated by XT-I demonstrating that XT-I can utilize a protein as an acceptor as well (**Figure 4.39 and 4.40**).

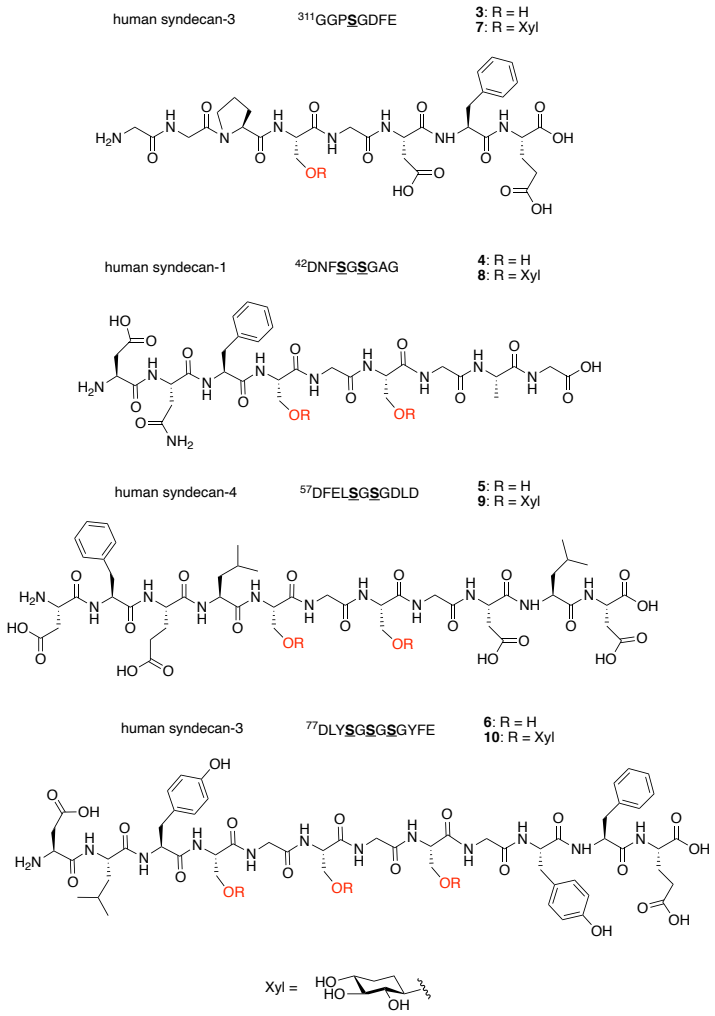


Figure 4.1 Structures of peptide **3-6** and glycopeptide **7-10** with the serine xylosylation site highlighted.

Acceptor	Product	Reaction Yield (%)
3	**7**	100
4	**8**	68.6
5	**9**	73.8
6	**10**	86.5

Table 4.1 Summary of XT-I catalyzed peptide glycosylation yields.

To attain more in-depth understandings on XT-I activity and its substrate preference, enzyme kinetics were measured for multiple peptide acceptors using a modified phosphatase-coupled glycosyltransferase assay.[25] Among the analytes, XT-I demonstrates the highest affinity and catalytic efficiency towards the bikunin peptide **1** (**Table 4.2**). The differential k_{cat}/K_m values for various peptide sequences suggest that presence of acidic residues *N*-terminal to the xylose attachment site may facilitate enzyme activities.

Substrate	K_m (μM)	V_{max} (pmol/min/μg)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}mM^{-1}$)
1	49.8 ± 4.9	350.9 ± 30.6	28	562
3	308.0 ± 69.4	45.4 ± 4.8	3	10
4	133.8 ± 27.0	196.7 ± 16.3	16	120
5	164.4 ± 23.0	183.8 ± 10.7	15	91

Table 4.2 Summary of kinetic data from peptide substrate **3-6**.

I next investigated the donor selectivity of XT-I. While XT-I was believed to be monofunctional to UDP-Xyl.[26] A variety of UDP-sugars were tested as XT-I donors, including UDP-Xyl, UDP-glucose (Glc), UDP-Galactose (Gal), UDP-*N*-acetyl glucosamine (GlcNAc) and UDP-6-azidoglucose (6AzGlc) with peptide **1** as the acceptor. UDP-Gal and UDP-GlcNAc were not transferred at detectable amounts. Examination of the crystal structure of XT-I (PDB code: 6EJ7)[12] shows that axial 4-OH of galactoside would clash with Asp494 and Glu529 (the catalytic base) in the active site of the enzyme (**Figure 4.2**). For UDP-GlcNAc, the 2-N-acetyl group of UDP-GlcNAc could be accommodated, but it could not form the hydrogen bond to Arg598 as present when UDP-Xyl was bound.

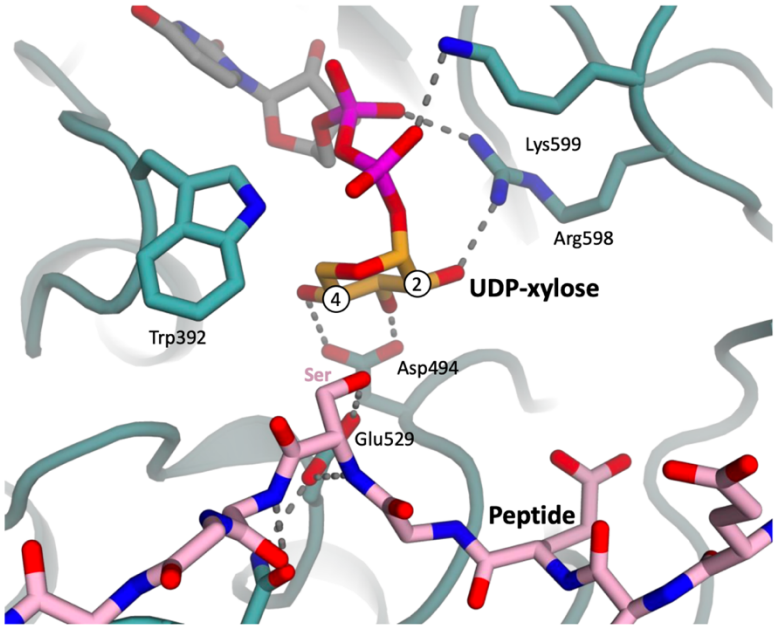
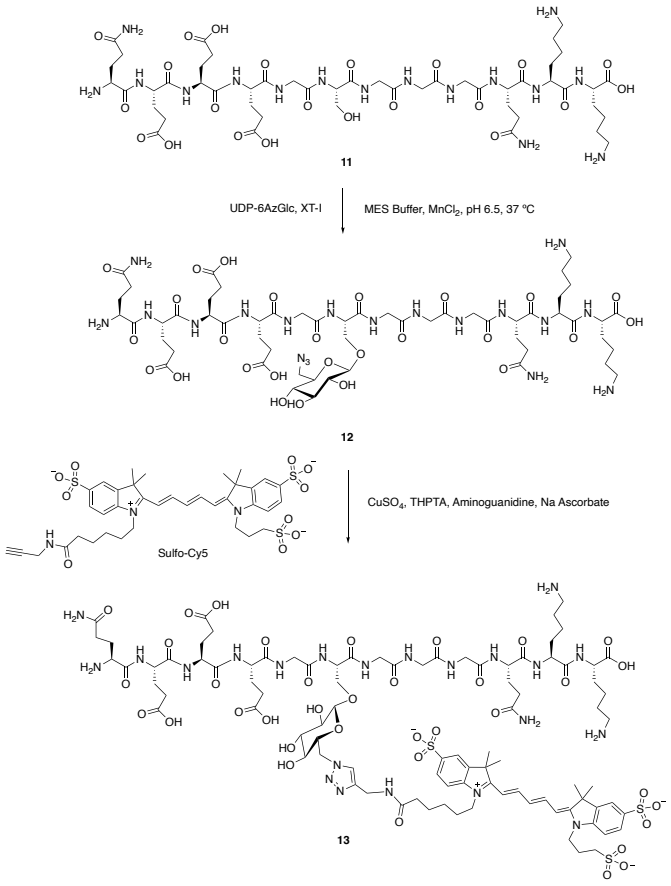


Figure 4.2 Structure of the active site of XT-I bound with UDP-Xyl and the peptide acceptor derived from the crystal structure (PDB code: 6EJ7). The 2-OH and 4-OH of UDP-Xyl have been labeled with the numbers 2 and 4 in circles. The key residues in the active site interacting with the UDP-Xyl have been highlighted. The structure 6EJ7 is a ternary complex of XT-I, UDP-Xyl and the acceptor peptide with a Ser-to-Ala mutation (to prevent Xyl transfer occurring in the crystal). To generate this figure, the serine was inserted back into the peptide acceptor to demonstrate the geometry of the acceptor complex. (Docking simulation was performed by Po-han Lin)

Interestingly, besides UDP-Xyl, noticeable enzymatic activities were observed with UDP-Glc and UDP-6AzGlc (**Table 4.3**). The successful transfer of 6AzGlc to bikunin peptide **11** by XT-I indicates its potential to be developed as a valuable biolabeling tool. As a proof of concept, azide-tagged glycopeptide **12** and alkynyl sulfo-Cy5 were subject to copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC). The desired Cy5 conjugated glycopeptide **13** was successfully produced (**Scheme 4.2**).

Substrate	K_m (μM)	V_{max} (pmol/min/μg)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}mM^{-1}$)
UDP-Xyl	43.4 ± 6.9	165.9 ± 6.4	13	266
UDP-Glc	84.0 ± 26.6	20.4 ± 1.9	2	20
UDP-6AzGlc	23.4 ± 10.5	11.0 ± 1.0	1	39

Table 4.3 Summary of kinetic results from UDP-sugar donors.



Scheme 4.2 XT-I catalyzed transfer of non-native 6AzGlc to bikunin peptide **11**, followed by incorporation of Cy5 fluorescent dye via 'Click' reaction.

To test whether the enzymatically prepared xylosyl peptide is a viable substrate for β4GalT7, xylosylated peptide **8** was treated with β4GalT7 and UDP-Gal To further the potential of XT-I to adopt UDP-6AzGlc, gatekeeper residue W392 in human XT-I (PDB: 6EJ7) was

replaced by alanine using UCSF Chimera program.[12, 27] As the result from docking simulation showed improved but sub-optimal binding between the enzyme and UDP-6AzGlc, in addition, R598 was swapped with Lysine to provide more space in the binding pocket. After local energy minimization, the resulted double mutant shows the potential to accept UDP-6AzGlc as its native donor substrate (**Figure 4.3**).

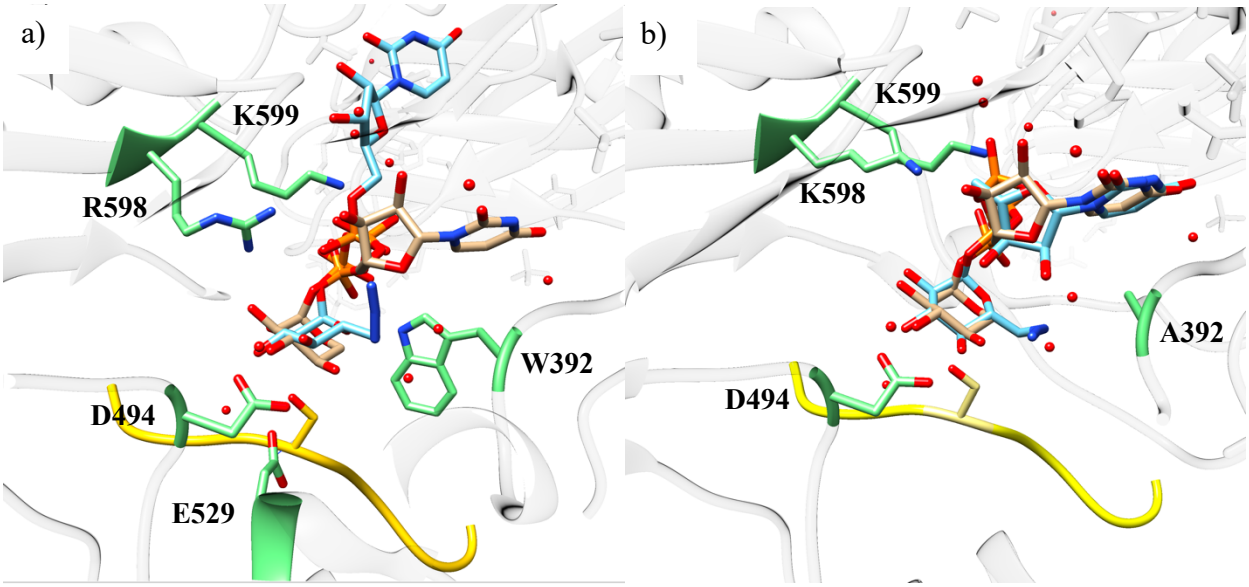
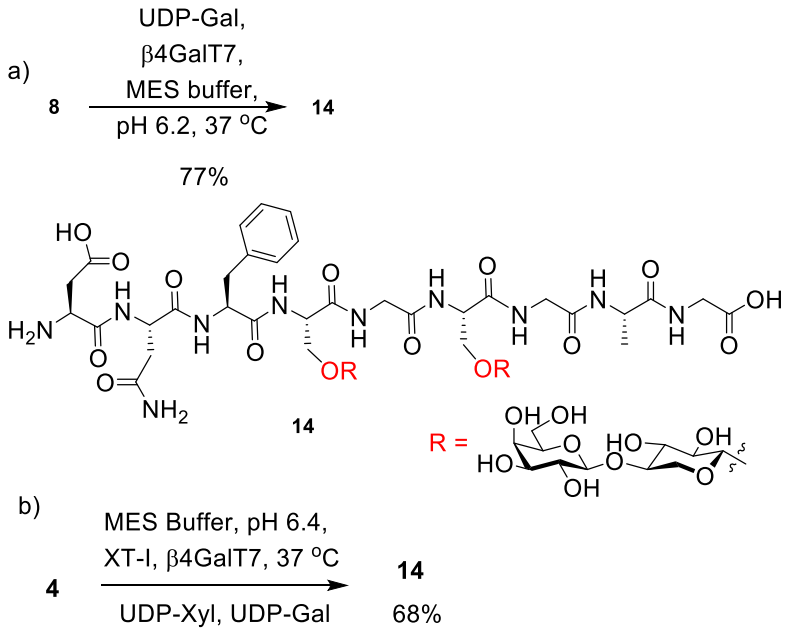


Figure 4.3 a) Wild-type human XT-I (PDB:6EJ7) in complex with UDP-Xyl (in brown color) or UDP-6AzGlc (in light blue color) and an acceptor peptide (as in **Figure 4.2**, in yellow color). C5 of xylopyranose is in close proximity with residue W392 (in green color); b) *in silico* engineered human XT-I W392A/R598K double mutant in complex with UDP-Xyl (in brown color) or UDP-6AzGlc (in light blue color) and the acceptor peptide (in yellow color).

As proteoglycans can contain long glycan chains, it is important that the glycan of the synthetic xylosyl peptides can be extended. In nature, a glycosyl transferase such as the β4GalT7 is responsible for adding a galactose unit to the xylose from the UDP galactose (UDP-Gal) donor.[13-15] Recently, β4GalT7 has been shown to be able to galactosylate chemically synthesized xylosylated peptides.[19] To test whether the enzymatically prepared xylose peptide is a viable substrate for β4GalT7, xylosylated peptide **8** was treated with β4GalT7 and UDP-Gal (**Scheme**

4.3a). The glycopeptide **14**[19] with two Gal-Xyl disaccharide was successfully produced in 77% yield. Thus, the overall yield for the stepwise conversion of **4** to **14** with XT-I glycosylation followed by β4GalT7 reaction was 53%. To further improve the synthetic efficiency, one pot synthesis was explored with XT-I and β4GalT7. Peptide **4**, UDP-Xyl, UDP-Gal, XT-I, and β4GalT7 were incubated together in the MES reaction buffer at 37 ºC overnight (**Scheme 4.3b**). Encouragingly, a full conversion of acceptor peptide **4** was observed with an isolated yield of 68% for glycopeptide **14**. Besides peptide **4**, this one-pot two-enzyme (OP2E) protocol smoothly converted peptides **3, 5**, and **6** to the corresponding glycopeptides **15-17**[19] (**Figure 4.4**) with higher yields compared to the stepwise synthesis (**Table 4.4**). The polyhistagged hCD44$_{20\text{-}178}$ protein was also glycosylated by the OP2E method to yield the Gal-Xyl modified CD44 (Appendix **Figure 4.42**).



Scheme 4.3 a) Galactosylation of glycopeptide **8** by β4GalT7 to form glycopeptide **14** bearing galactose-xylose disaccharide; b) OP2E synthesis of **14** from peptide **4** by one pot reaction with XT-I and β4GalT7.

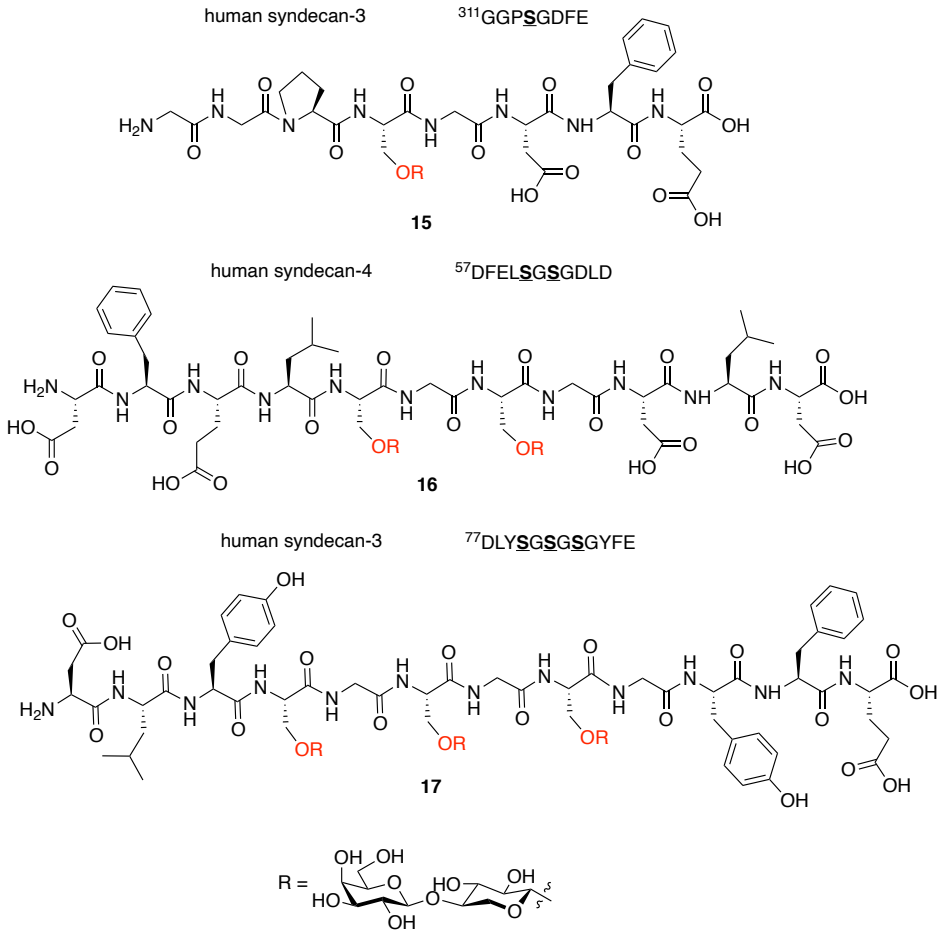


Figure 4.4 Structures of OP2E glycopeptide products **15-17**. Glycosylated serine sites are highlighted.

Acceptor	Product	Stepwise Synthesis Yield (%)	Reaction Yield (%)
3	**15**	91	94
4	**14**	53	68
5	**16**	60	89
6	**17**	68	91

Table 4.4 Yield summary of OP2E synthesis.

Enzymatic synthesis of glycopeptide such as **14** is more efficient than the corresponding chemical synthesis. Due to the need for multiple protecting group manipulation to prepare the two strategically protected monosaccharide building blocks followed by the technically challenging chemical glycosylations and deprotection reactions, it would have taken over 20 synthetic steps to access a glycopeptide such as **14** *via* chemical glycosylation from commercially available monosaccharides.[28] Thus, the OP2E protocol can significantly improve the overall synthetic efficiencies.

In the OP2E protocol for glycopeptide synthesis, XT-I presumably installed the xylose onto the peptide first, followed by β4GalT7 promoted galactosylation of the xylosylated peptide as in the case for stepwise synthesis. Alternatively, β4GalT7 may galactosylate UDP-Xyl first with subsequent transfer of the UDP disaccharide donor to the peptide acceptor catalyzed by XT-I.

However, the formation of disaccharide donor in OP2E reaction is unlikely to occur at an appreciable rate as β4GalT7 prefers β-xyloside acceptors.[29] The UDP-Xyl has an α-anomeric configuration and the UDP moiety would clash with the active site of β4GalT7 enzyme. Furthermore, the crystal structure of XT-I (PDB code: 6EJ7)[12] shows that the active site of XT-I (**Figure 4.2**) is not sufficiently large to accommodate a disaccharide donor.

4.3 Conclusions

In conclusion, for the first time, human XT-I (EC 2.4.2.26) enzyme has been utilized to efficiently synthesize structurally diverse xylosylated glycopeptides at milligram scales with a range of peptide acceptors as well as the His tag bearing $hCD44_{20\text{-}178}$ protein. XT-I was found tolerant toward several non-native UDP-sugar donors, particularly UDP-6AzGlc, rendering it potentially a valuable tool to label biological proteins. The one-pot two-enzyme method developed

further enhanced the synthetic efficiency and the overall yield, paving the way toward efficient chemoenzymatic synthesis of PG glycopeptides and glycoproteins.

4.4 Experimental Section

4.4.1 Materials

Signal peptide-His6-XT-I gBlocks gene was purchased from Integrated DNA Technologies (Coralville, IA). FreeStyle 293 Expression Medium and Coomassie Brilliant Blue G-250 were purchased from Thermo Fischer Scientific (Waltham, MA). Nickel columns and Nickel resins were purchased from Bio-rad (Hercules, CA). SDS-PAGE gels, 10x Tris/Glycine/SDS electrophoresis buffer, prestained protein ladder, sample loading buffer, and Coomassie Blue R-250 were purchased from Bio-rad (Hercules, CA). Tris-HCl buffer was purchased from MilliporeSigma (St. Louis, MO). UDP-xylose was purchased from Complex Carbohydrate Research Center (Athens, GA). Amino acid building blocks were purchased from Chem-Impex International, Inc (Wood Dale, IL). Cy5-alkyne was purchased from MilliporeSigma (St. Louis, MO). Glycosyltransferase Activity Kit was purchased from R&D Systems. All other chemical reagents were purchased from commercial sources and used without additional purifications unless otherwise noted.

4.4.2 General Information

High-performance liquid chromatography was carried out with LC-8A solvent pumps, DGU-14A degasser, SPD-10A UV-Vis detector, SCL-10A system controller (Shimadzu Corporation, JP) and Vydac 218TP 10 μm C18 preparative HPLC column (HICHROM 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)). The dual-wavelength UV detector was set at 220 nm and 254 nm for monitoring the absorbance from amide and aromatic groups correspondingly. 3D structure of glycopeptide compounds was prepared with Maestro software. Docking simulations were acquired with AutoDock Vina and UCSF Chimera (UCSF, CA). Enzymatic activity was quantified by absorbance at 620 nm using a SpectraMax M3 96-well plate reader (Molecular Devices, CA). NMR data were obtained with DirectDrive2 500 MHz (Agilent, CA) at ambient temperature.

4.4.3 XT-I Expression, Purification and Characterization

HEK-293F cells were grown in FreeStyle™ 293 Expression Medium on a platform shaker in humidified 37 °C CO_2 (5%) incubator with rotation at 150 rpm. When the cell density reached between 4 x 10^5 and 3 x 10^6 cells/ml, cells were split to a density of 1 x 10^6 cells/ml and cultured overnight in the same condition. Cells were then transfected with His6-XT-I gene 24 hours after they were split. Before transfection, cells were harvested by centrifugation at 1200 rpm for 10 min at room temperature and re-suspended in fresh pre-warmed media. To transfect the cells, a final concentration of 2-3 μg/ml of the XT-I gene and 9 μg/ml of PEI were added. PEI stock solution was prepared at the concentration of 1 mg/ml in a buffer containing 25 mM HEPES and 150 mM NaCl, pH 7.4. The flask was returned to the shaker platform in the incubator. Cells were diluted 1:1 with pre-warmed media supplemented with valproic acid (VPA) to a final concentration of 2.2 mM. Four to six days after the transfection, cells were harvested. Clarified lysate was purified by nickel column (Cytiva, MA) (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 40-250 mM imidazole). Protein purity was confirmed with SDS-PAGE gel electrophoresis and the concentration and expression yield were determined by standard Bradford assay.

4.4.4 General Procedure for Automated Solid-Phase Peptide Substrate Synthesis

All the peptides were synthesized on a Liberty Blue™ Automated Microwave Peptide Synthesizer following the standard Fmoc-based solid-phase peptide synthesis protocol. The Cl-MPA ProTide resins were purchased from CEM Corporation. The Liberty Blue software (CEM Corporation, NC) was used to program the synthesis, including resin swelling, amino acid loading, couplings and Fmoc- removal. Commercially available *N,N*-dimethylformamide (DMF) from Fischer Chemical was supplied to the synthesis module as reaction and washing solvent. Peptide synthesis was enabled by sequential couplings of Fmoc-amino acid (purchased from Chem-Impex), which was preactivated by DIC, Oxyma Pure and *N*,*N*-diisopropyl-*N*-ethyl amine at 50 °C for 10 min, and deprotections with 20% piperidine in DMF at 60 °C for 4 min. In-between each coupling/deprotection step, resin-bound peptide was thoroughly washed with DMF. Resin-bound peptides were cleaved off the solid support with a cocktail solution of trifluoroacetic acid (TFA), triisopropylsilane (TIPS) and water (TFA/TIPS/H_2O, 95:2.5:2.5). The crude peptides were then purified with reverse-phase C18 preparative HPLC. Compound purity was confirmed by C18 analytical HPLC analysis.

4.4.5 General Procedure for XT-I-Catalyzed Glycosylation

The 10x 2-(*N*-morpholino)ethanesulfonic acid (MES) reaction buffer for XT-I-catalyzed glycosylation was prepared in advance following the recipe of 250 mM MES, 250 mM KCl, 50 mM KF, 50 mM $MgCl_2$, 50 mM $MnCl_2$. The pH of the 10x reaction buffer was adjusted to 6.5 by adding concentrated NaOH solution. A solution of 1 mM peptide substrate and 1.1-3.0 mM UDP-Xyl (1.1-3.0 equiv. per glycosylation site, depending on peptide acceptors) was made with the reaction buffer. The addition of XT-I enzyme (0.02 mol%) initiated the glycosylation. The reaction solution was kept at 37 °C overnight. The reaction progress was monitored with LC-MS. After the

reaction, the enzyme was deactivated and precipitated out of the reaction mixture by addition of ethanol. The mixture was centrifuged, and the supernatant was loaded onto a G-10 size exclusion column for purification.

4.4.6 General Procedure for Enzyme-Substrate Docking and *In Silico* Enzyme Engineering

3D structure of the substrate was prepared with ChemDraw 16.0 and Schrodinger Maestro software. After importing the substrate structure from ChemDraw into Maestro, it was energetically optimized *via* the built-in function "Minimize-All Atoms". The optimized structure was then output as a mol2 file for the subsequent molecular dynamic docking. To initiate the docking experiments, a high-resolution enzyme crystal structure as a PDB file, along with the substrate structure as a mol2 file, was imported into UCSF Chimera software. The enzyme-substrate molecular docking was achieved with AutoDock Vina, an integrated program in UCSF Chimera. [22, 23] For the docking set-up, the enzyme was chosen as the "Receptor" and the substrate was selected as "Ligand". The "Receptor search volume" was defined to ensure that space around the catalytic binding pocket was included for a proper docking simulation, while balancing the demand towards computation resource. Default settings of "Receptor options" and "Ligand options" were used. "Number of binding modes", "Exhaustiveness of search" and "Maximum energy difference (kcal/mol)" options were adjusted to the maximum level to ensure the quality of the simulation. The docking experiment was then executed via Opal web service. Computation results were available upon completion of the experiment.

Human XT-I crystal (PDB:6EJ7) was selected for *in silico* enzyme engineering. R598 residue was replaced by lysine through Chimera built-in function 'Rotamers'. Lysine residue poses with highest predicted possibilities were selected to examine potential clashes/contacts. If contacts with nearby residues were detected, residues in contact, together with K598, were processed

through ‘Minimize Structure’ function. All other atoms, except the selected ones, were fixed to reduce the computation workload. The resulted clash-free XT-I mutant structure was then used to perform the enzyme-substrate docking simulation.

4.4.7 General Procedure for XT-I-Catalyzed Transfer of UDP-6-Azidoglucose

A solution of 0.5 mM peptide substrate and 2.5 mM UDP-xylose (5 equiv. per glycosylation site) was made with the MES reaction buffer. The addition of XT-I enzyme (0.1 mol%) initiated the glycosylation. The reaction solution was kept at 37 °C overnight. The reaction progress was monitored with LC-MS. After the reaction, the enzyme was deactivated and precipitated out of the reaction mixture by addition of ethanol. The mixture was centrifuged, and the supernatant was carried over without further purification.

4.4.8 General Procedure for Copper (I)-Catalyzed Azide-Alkyne Cycloaddition

To a solution of azide-tagged glycopeptide **12** (100 μM), $CuSO_4$ (20 mM), (tris-hydroxypropyltriazolylmethylamine) (THPTA) ligand (10 mM), aminoguanidine (100 mM), Cy5-alkyne (1 mM), and Na ascorbate (100 mM) were added. The reaction tube was attached to a 20 round-per-minute (rpm) end-over-end rotator. The reaction was allowed to proceed for 2 hours at room temperature. The formation of Cy5 conjugated glycopeptide **13** was confirmed using LC-MS (ESI-MS: $C_{91}H_{131}N_{22}O_{36}S_3^{2-}$, calcd: 734.6092, obsd: 734.5991 (13.7ppm))

4.4.9 General Procedure for One-Pot Two-Enzyme (OP2E) Glycosylation

The 10x MES reaction solution for XT-I and β4GalT7 OP2E glycosylation was prepared following the recipe of 225 mM MES, 125 mM KCl, 25 mM KF, 25 mM $MgCl_2$, 75 mM $MnCl_2$. A solution of 1 mM peptide and 1.5-3.0 mM UDP-xylose (1.5-3.0 equiv. per glycosylation site, depending on peptide acceptors) and 2.0 mM UDP-galactose (2.0 equiv. per glycosylation site) was made with the reaction buffer. XT-I enzyme (0.05 mol%) and β4GalT7 enzyme (0.5 mol%)

were added to initiate the glycosylation reactions. The reaction tube was kept in an incubator at 37 °C overnight. The reaction progress was monitored via LC-MS. Upon reaction completion, the reaction mixture was directly injected into HPLC, and the reaction yield was quantified from peak areas of HPLC chromatograms.

4.4.10 Phosphatase-Coupled Enzymatic Kinetic Assay

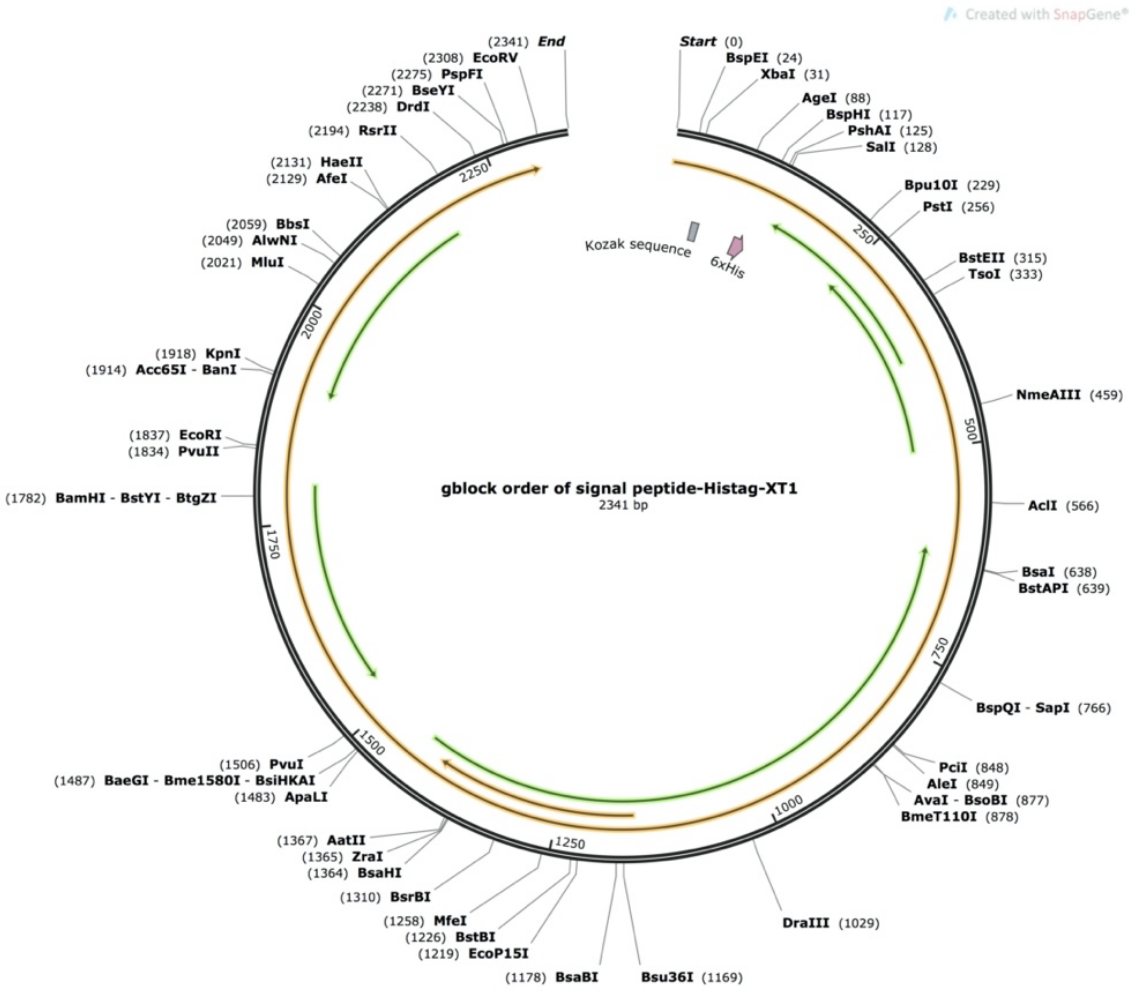
The kinetic assay protocol follows the general assay conditions reported by R&D Systems Inc. with modifications.[20]

30 μL reaction solutions of UDP-galactose, glycopeptide acceptor and β4GalT7 enzyme were prepared in the 96-well plate. The plate was covered with a plate sealer and incubated at 37 °C for 20 min. 12 μL 10x phosphatase assay buffer, 3 μL $MnCl_2$ solution (100 mM), 3 μL MilliQ water and 2 μL coupling phosphatase 1 (20 ng/μL), were quickly added to a total volume of 50 μL. The plate was covered with a plate sealer again and incubated at 37 °C for 20 min. After the incubation, 30 μL of Malachite Green Reagent A was quickly added to each well. The solutions were gently mixed by tapping the plate. 100 μL of deionized or distilled water was added to each well. 30 μL of Malachite Green Reagent B was then added to each well. Solutions were mixed gently by tapping the plate. The plate was incubated for 5 minutes at room temperature to have consistent color development. The optical density of each well was determined using a microplate reader set to 620 nm, and the OD was adjusted by subtracting the reading of the negative control. Product formation was calculated using the conversion factor determined from the phosphate standard curve.

APPENDICES

APPENDIX A: Supplementary Figures, Schemes and Tables

Signal peptide-His6-XT-I Sequence



5’AAGACACCGGGACCGATCCAGCCTCCGGACTCTAGAGCCGCCACCATGGGTTGGA
GTTGTATCATCCTTTTCCTGGTAGCTACCGCAACCGGTGTTCATTCACATCACCACCA
TCATCATGACGTAAGTCGACCTCCTCACGCAAGAAAGACGGGTGGCTCTAGCCCGG
AGACTAAGTATGACCAGCCGCCGAAGTGCGACATTAGCGGTAAAGAAGCGATCTCT
GCCCTGAGCCGGGCAAAATCAAAACACTGCAGACAGGAGATTGGTGAGACGTATTG
CCGACACAAACTGGGGCTCCTCATGCCAGAGAAGGTAACCAGATTTTGTCCGCTGG
AGGGGAAGGCCAACAAAAACGTCCAATGGGACGAGGATAGCGTCGAATACATGCCT
GCGAATCCCGTCAGGATCGCGTTTGTCCTTGTCGTTCATGGCCGAGCGAGCAGACAA
CTTCAGCGCATGTTTAAAGCAATCTACCACAAAGACCATTTCTATTATATTCATGTCG

Figure 4.5 XT-I gene sequence.

Figure 4.5 (cont'd)

ATAAGCGGTCAAACTACCTGCACCGGCAGGTACTCCAGGTTTCACGCCAATACTCCA
ACGTTCGCGTAACTCCATGGCGGATGGCCACGATCTGGGGTGGGGCTTCACTCCTCT
CAACGTATTTGCAGAGCATGCGAGACCTTCTGGAAATGACTGACTGGCCATGGGACT
TTTTCATCAATTTGAGCGCAGCCGACTATCCAATCCGAACCAATGATCAGCTTGTAG
CATTTCTGAGTCGCTATAGGGACATGAATTTCCTGAAGAGCCATGGGCGGGATAACG
CGCGGTTCATACGAAAGCAAGGGCTGGATAGGCTGTTTCTTGAATGCGACGCACAC
ATGTGGAGGCTTGGGGATAGAAGGATTCCCGAGGGGATCGCCGTGGATGGAGGAAG
CGACTGGTTCCTTCTGAATCGACGGTTTGTCGAGTATGTCACGTTCAGCACGGATGA
TTTGGTCACGAAAATGAAACAATTCTACAGTTATACGCTCCTGCCCGCTGAGAGCTT
CTTCCACACGGTGTTGGAAAACTCCCCGCATTGTGATACAATGGTTGATAATAATTT
GAGGATTACAAATTGGAATCGAAAACTTGGGTGCAAATGTCAGTATAAGCATATAG
TGGACTGGTGTGGATGTTCTCCTAATGACTTTAAACCTCAGGATTTTCATCGATTCCA
GCAGACAGCACGGCCTACTTTTTTTGCGCGAAAATTCGAAGCAGTCGTCAATCAAGA
GATTATCGGACAATTGGATTACTACCTGTATGGAAACTATCCTGCCGGTACGCCTGG
GCTCCGCTCCTATTGGGAGAATGTCTATGATGAACCTGACGGAATACATTCCCTTAG
TGACGTCACCCTCACTCTTTATCATAGTTTTGCACGCTTGGGTCTGAGACGGGCCGA
AACTTCTCTTCATACAGACGGCGAAAACAGTTGTCGCTATTACCCGATGGGCCACCC
CGCATCAGTGCACCTTTATTTCCTGGCCGATCGATTCCAGGGGTTTCTGATCAAGCAT
CATGCGACAAACCTCGCAGTGAGCAAATTGGAAACTCTTGAAACCTGGGTGATGCC
CAAAAAAGTGTTCAAAATCGCTAGTCCTCCCTCCGACTTTGGTAGGTTGCAGTTCTC
CGAAGTAGGGACAGATTGGGACGCGAAGGAGAGACTGTTTCGGAACTTCGGCGGGT
TGTTGGGACCGATGGATGAGCCAGTTGGCATGCAAAAGTGGGGCAAAGGGCCTAAC
GTCACTGTAACAGTGATCTGGGTGGATCCAGTCAACGTCATCGCCGCAACTTACGAT
ATACTGATTGAGAGTACAGCTGAATTCACCCACTATAAACCGCCCTTGAACCTTCCC
CTGCGACCTGGAGTGTGGACCGTTAAGATTCTTCACCACTGGGTACCTGTGGCGGAG
ACGAAATTTTTGGTGGCCCCGTTGACTTTTTCCAATCGACAACCTATAAAGCCTGAA
GAGGCCCTTAAACTGCACAACGGTCCACTGCGAAACGCGTATATGGAACAGTCTTTC
CAGTCTCTGAACCCTGTACTTAGTCTTCCAATAAATCCGGCCCAAGTTGAGCAAGCC
CGGCGGAATGCCGCTTCCACTGGAACAGCGCTCGAAGGATGGCTTGATAGCCTGGTT
GGAGGTATGTGGACAGCCATGGACATCTGCGCCACCGGACCGACCGCGTGTCCGGT
GATGCAAACTTGTTCTCAGACTGCGTGGTCTAGCTTCTCACCTGATCCAAAGTCCGA
GCTGGGCGCAGTGAAACCCGACGGTAGACTTAGGTGATATCTCGACAATCAACCTCT
GGATTACAAAATTT 3'

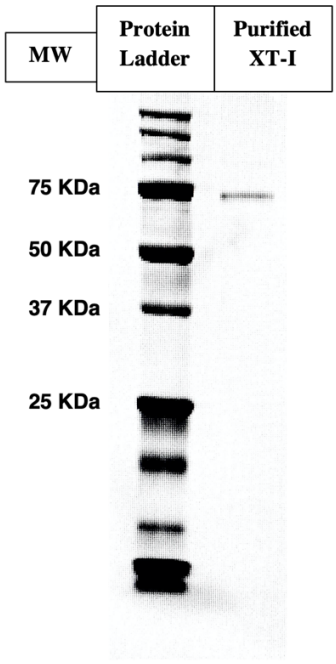


Figure 4.6 SDS-PAGE gel of purified XT-I.

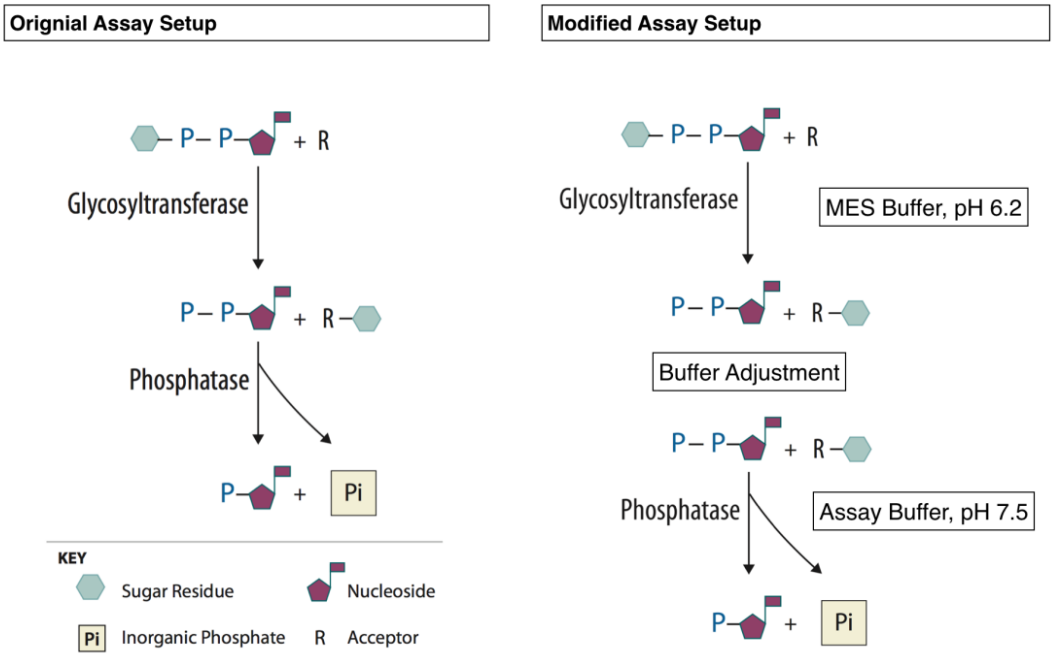


Figure 4.7 Schematic demonstrations of the original[20] and the modified kinetic assay set-up.

Phosphate Standard

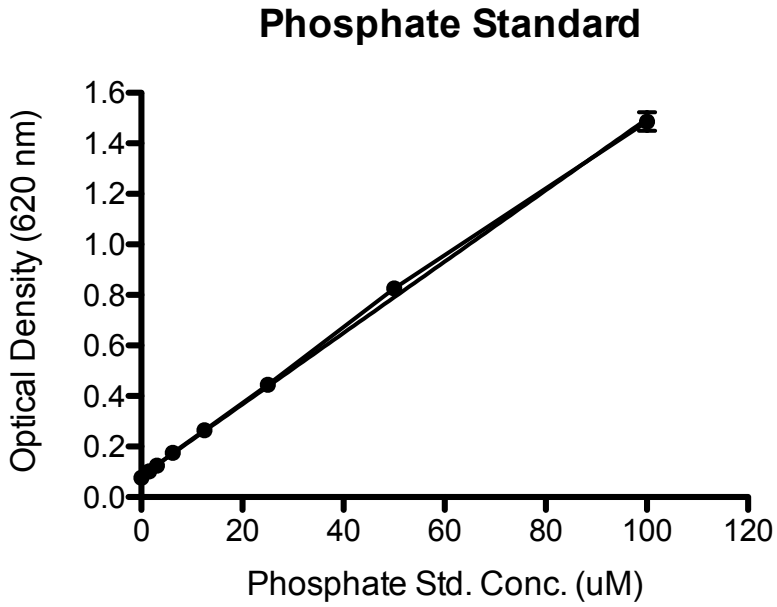
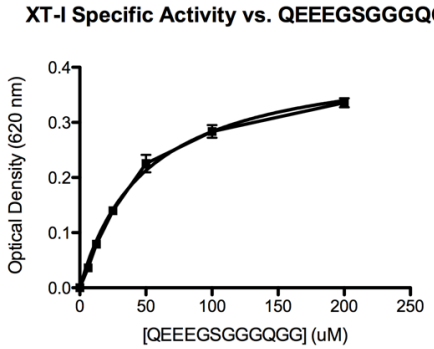


Figure 4.8 Phosphate conversion factor measurement. Conversion factor = 3541 pmol/OD (Plot is displayed as mean ± S.D. of two replicates, phosphate standard volume = 50 µL).



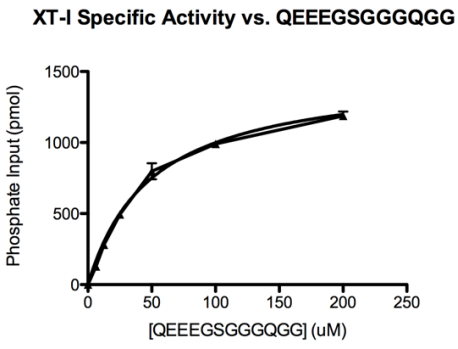


Figure 4.9 Phosphatase-coupled assay result of QEEEGSGGGQGG **1**. k_{cat} = 28 min^{-1}, K_m = 49.8 mM, k_{cat}/K_m = 562 mM^{-1} min^{-1}.

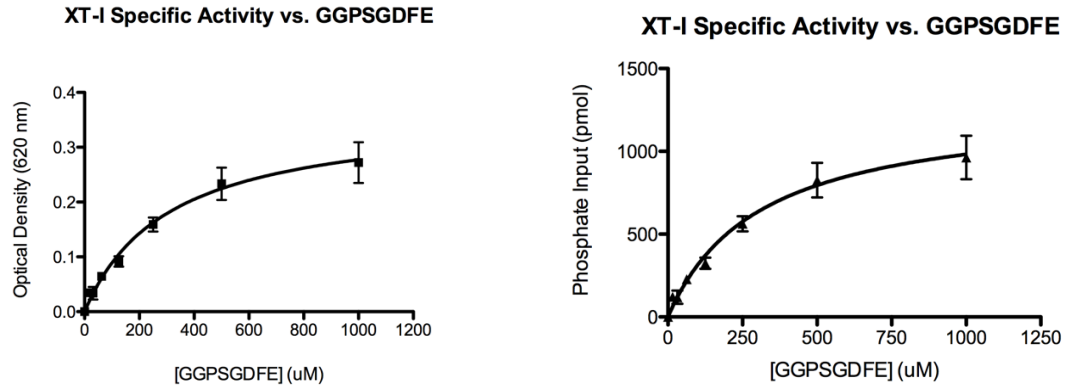


Figure 4.10 Phosphatase-coupled assay result of GGPSGDFE **3**. k_{cat} = 3 min^{-1}, K_m = 308.0 mM, k_{cat}/K_m = 10 mM^{-1} min^{-1}.

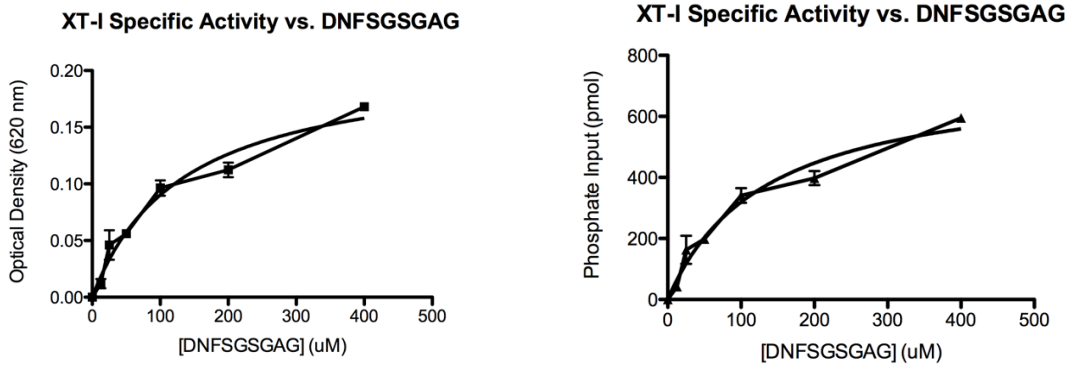


Figure 4.11 Phosphatase-coupled assay result of DNFSGSGAG **4**. k_{cat} = 16 min^{-1}, K_m = 133.8 mM, k_{cat}/K_m = 120 mM^{-1} min^{-1}.

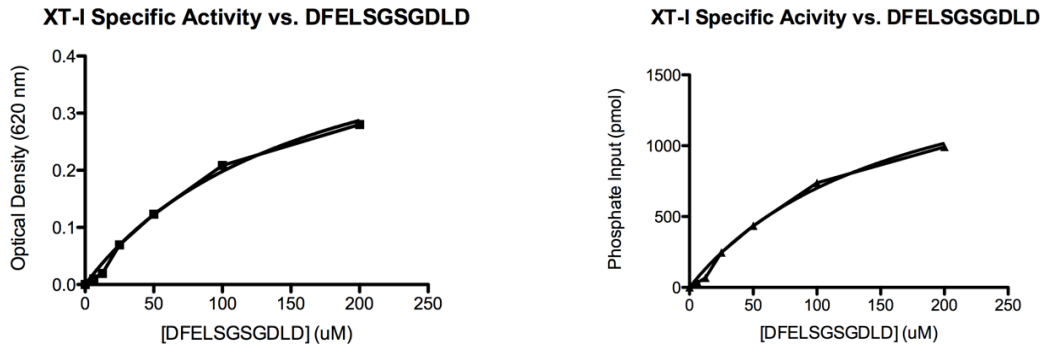


Figure 4.12 Phosphatase-coupled assay result of DFELSGSGDLD **5**. k_{cat} = 15 min^{-1}, K_m = 164.4 mM, k_{cat}/K_m = 91 mM^{-1} min^{-1}.

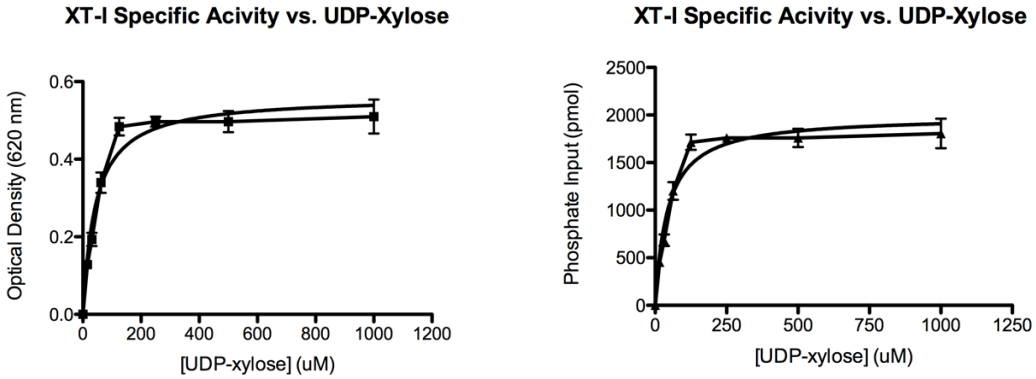


Figure 4.13 Phosphatase-coupled assay result of UDP-xylose. k_{cat} = 13 min^{-1}, K_m = 43.4 mM, k_{cat}/K_m = 266 mM^{-1} min^{-1}.

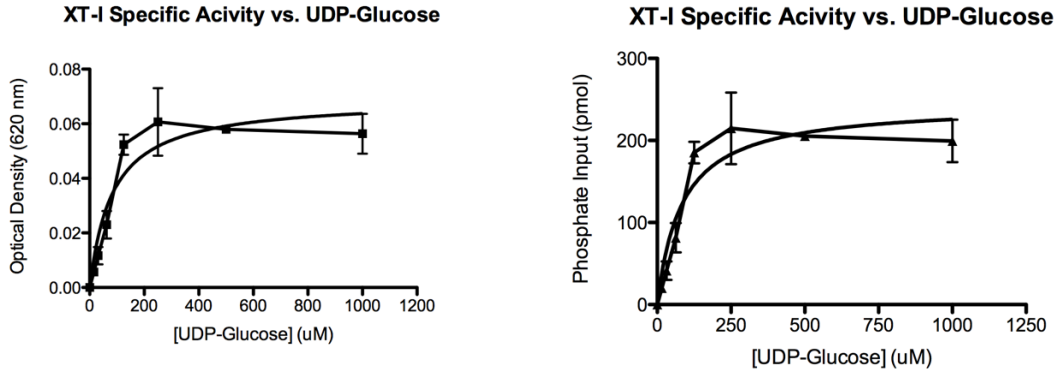


Figure 4.14 Phosphatase-coupled assay result of UDP-glucose. k_{cat} = 2 min^{-1}, K_m = 84.0 mM, k_{cat}/K_m = 33 mM^{-1} min^{-1}.

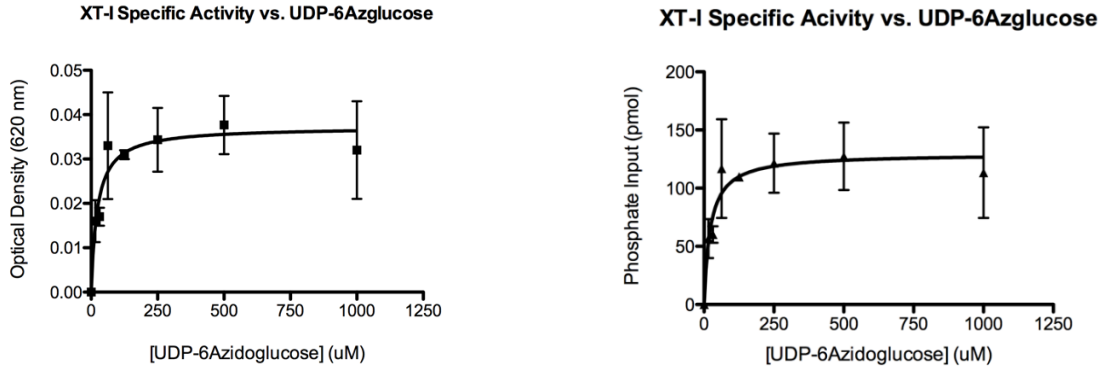
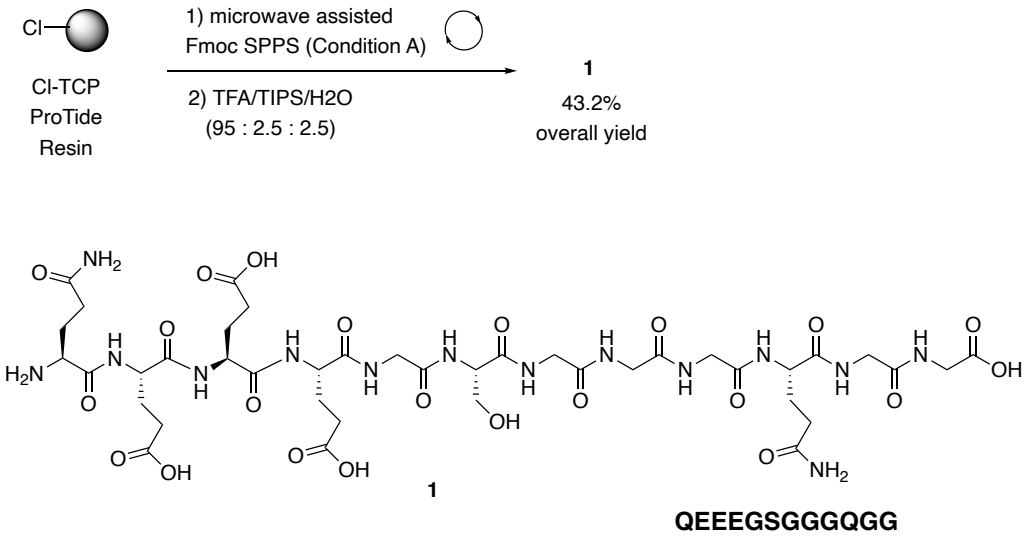


Figure 4.15 Phosphatase-coupled assay result of UDP-6-azido-glucose. k_{cat} = 1 min^{-1}, K_m = 23.4 mM, k_{cat}/K_m = 18 mM^{-1} min^{-1}.



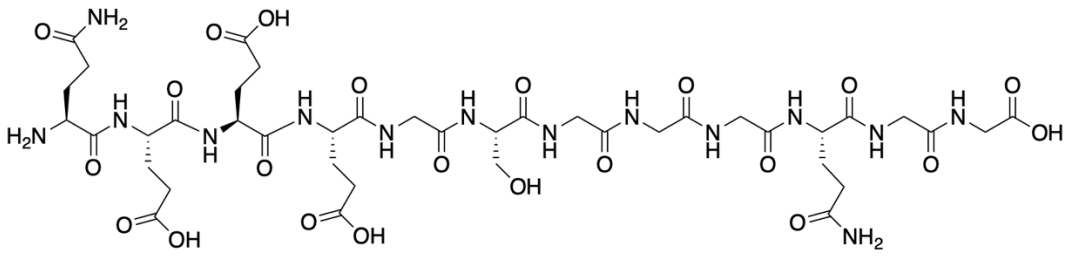
SPPS conditions : 1) Fmoc-AA-OH, DIPEA, KI, μW, DMF, 90 ºC
2) Fmoc- removal: 20% piperidine in DMF
3) Amino acid coupling: 5 eq Fmoc-AA-OH @50 ºC for 10 min,
DIC, Oxyma Pure w/0.1 M DIPEA, DMF
Repeat step 2 and 3

Scheme 4.4 SPPS synthesis of bikunin glycopeptide (QEEEGSGGGQGG) **1**.

Sequence	SPPS Yield (%)
QEEEGSGGGQGG **1**	43.2
GGPSGDFE **3**	47.7
DNFSGSGAG **4**	61.7
DFELSGSGDLD **5**	38.2
DLYSGSGSGYFE **6**	33.1
QEEEGSGGGQKK **11**	47.8

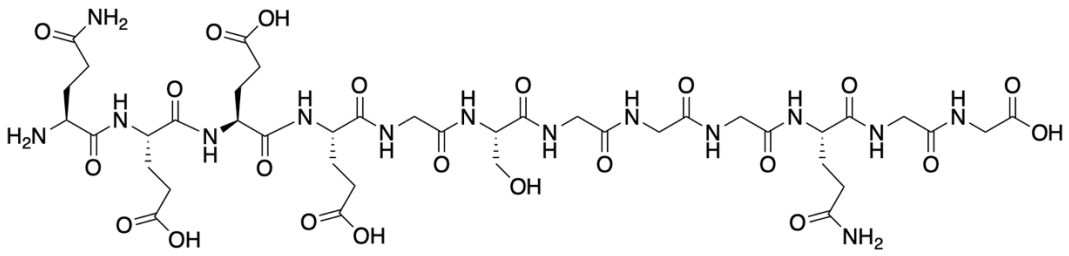
Table 4.5 Summary of synthesized peptide acceptors and the corresponding yields.

APPENDIX B: Product Characterization Spectra



1

The purity of peptide **1** was verified with analytical C-18 HPLC (water, 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$= + 28 (c 0.1, H_2O, specific rotation was collected by Po-han Lin). ^{1}H-NMR (500 MHz, D_2O) δ 4.37 – 4.15 (m, 2H), 3.99 – 3.65 (m, 5H), 3.05 – 2.93 (m, 9H), 2.47 – 2.14 (m, 5H), 2.06 – 1.89 (m, 2H), 1.90 – 1.76 (m, 2H), 1.60 (m, 10H), 1.52 – 1.45 (m, 5H); ^{13}C-NMR (125 MHz, D_2O) δ 60.9, 60.9, 56.7, 55.7, 55.6, 53.1, 52.9, 52.9, 52.1, 46.6, 45.3, 44.6, 44.3, 44.3, 43.3, 42.6, 42.3, 42.1, 42.1, 42.1, 41.3, 40.0, 31.5, 30.7, 29.9, 29.9, 29.7, 28.7, 27.4, 27.2, 26.4, 26.2, 26.0, 25.2, 24.5, 24.4, 23.2, 22.0, 22.0, 21.7, 21.2, 21.1, 20.4, 19.7, 17.7, 16.9, 14.4. ESI-MS: $C_{40}H_{63}N_{14}O_{22}$ $[M+H]^+$ calcd: 1091.4236, obsd: 1091.4216 (1.8 ppm)



1

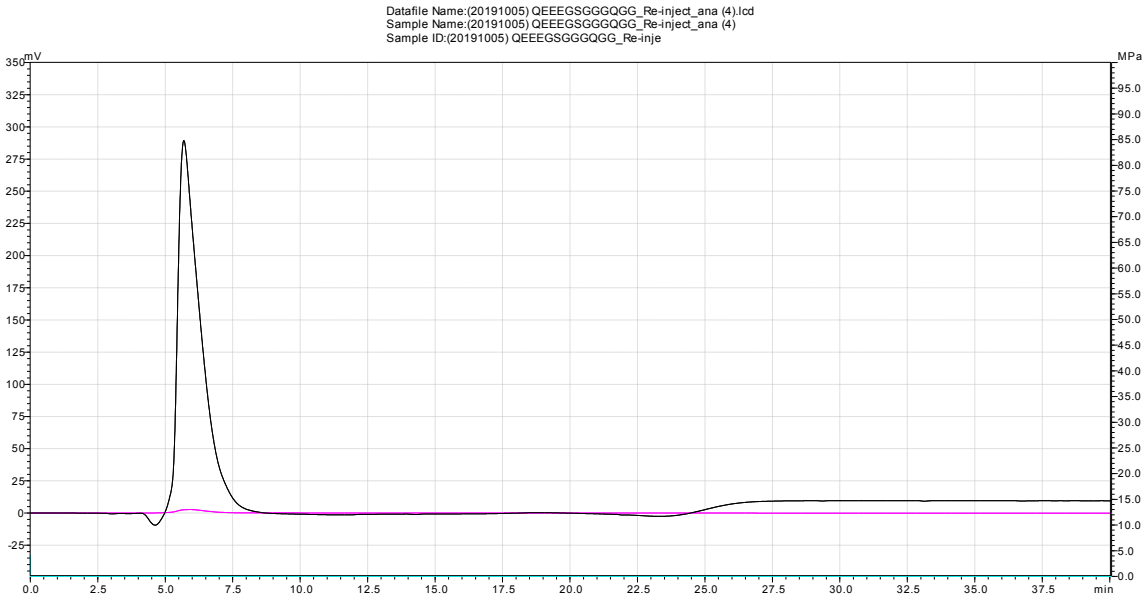
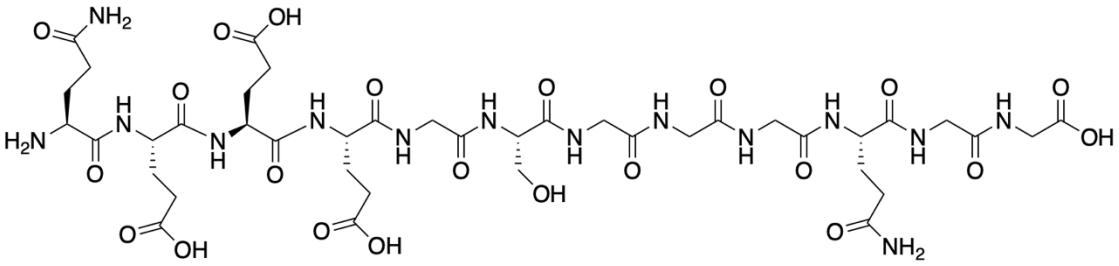


Figure 4.16 HPLC chromatogram of **1**.



1

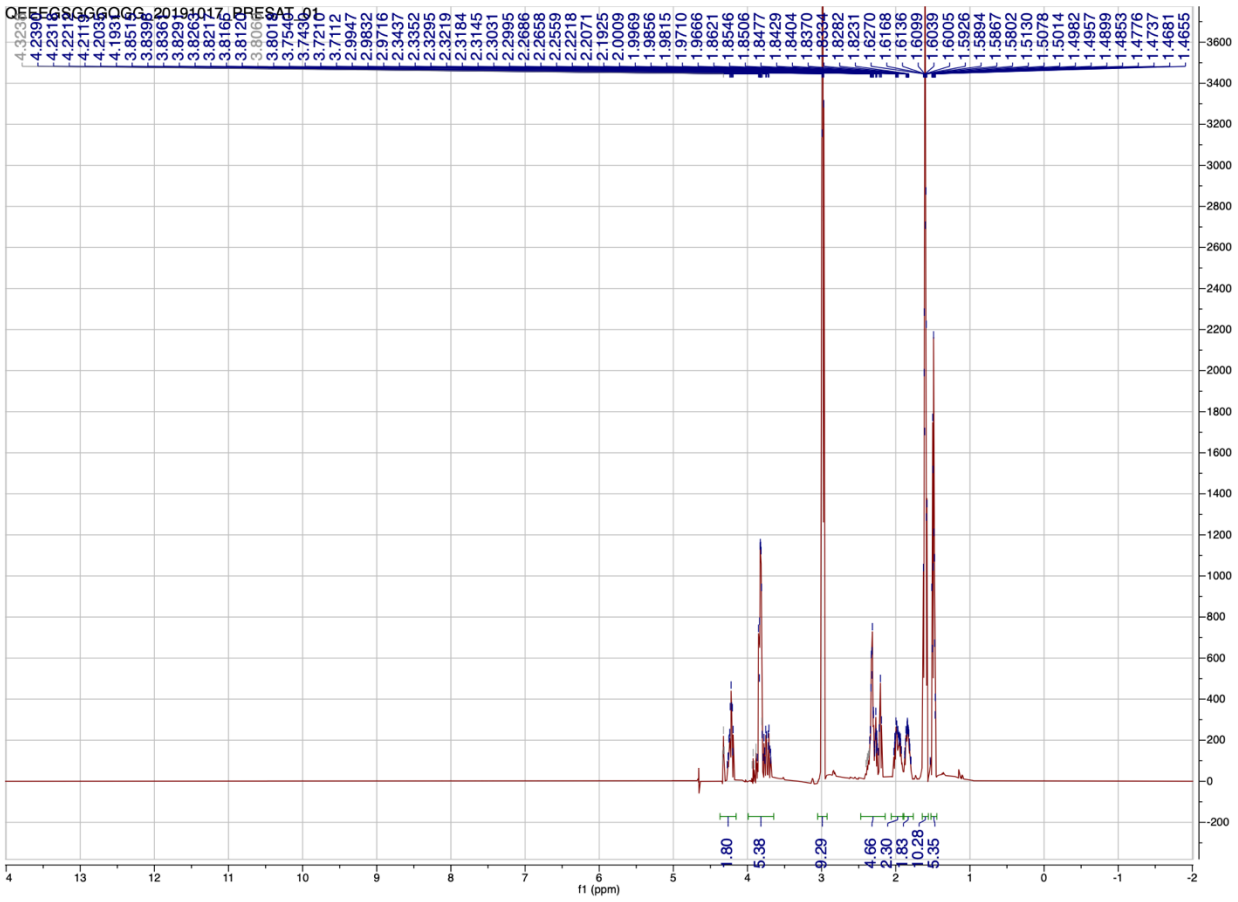
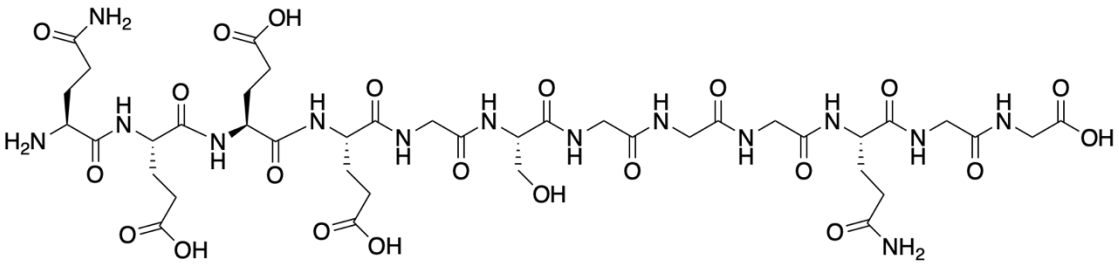


Figure 4.17 ^{1}H-NMR of **1** (500 MHz, D_2O).



1

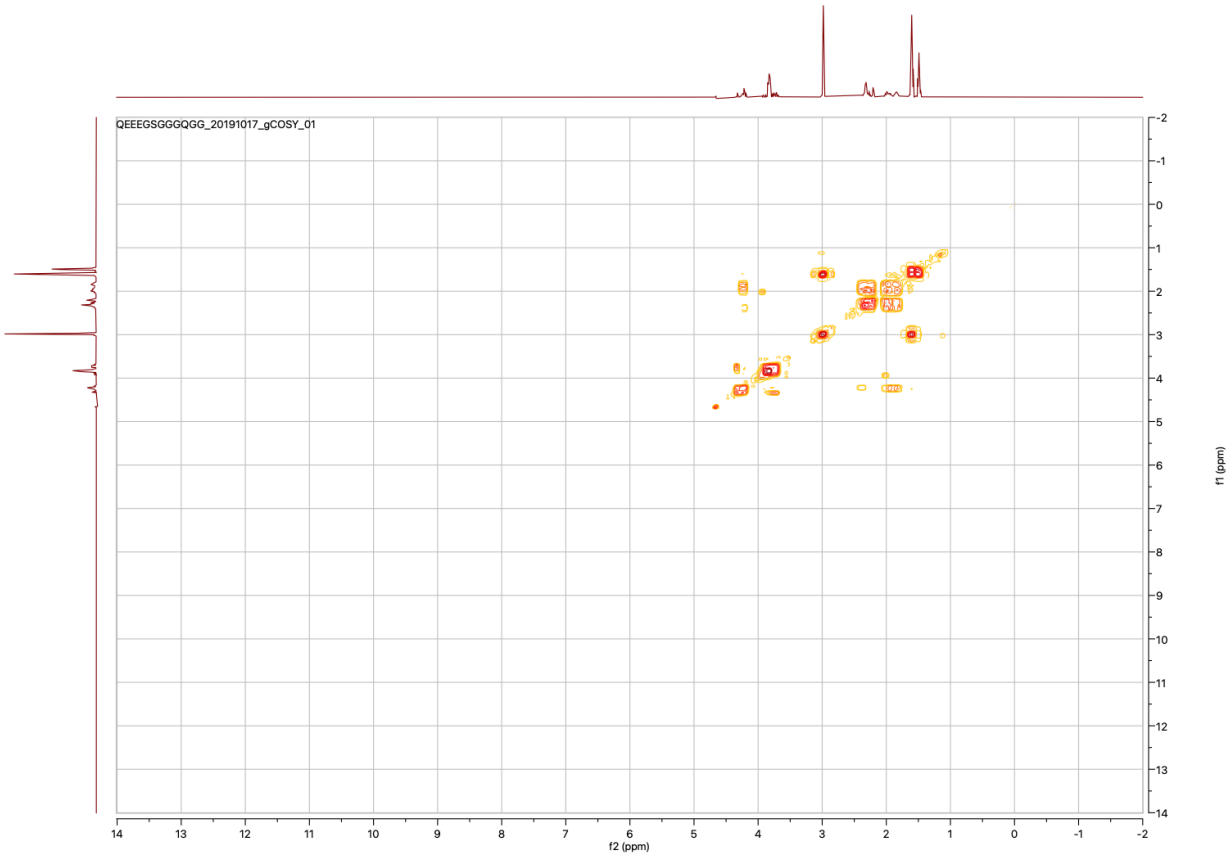
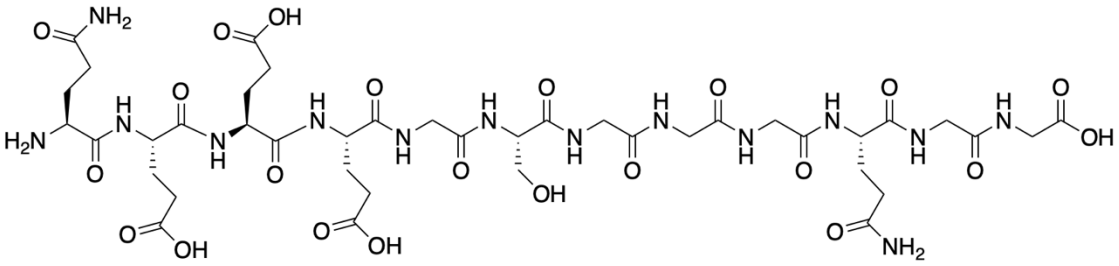


Figure 4.18 COSY NMR of **1** (500 MHz, D_2O).



1

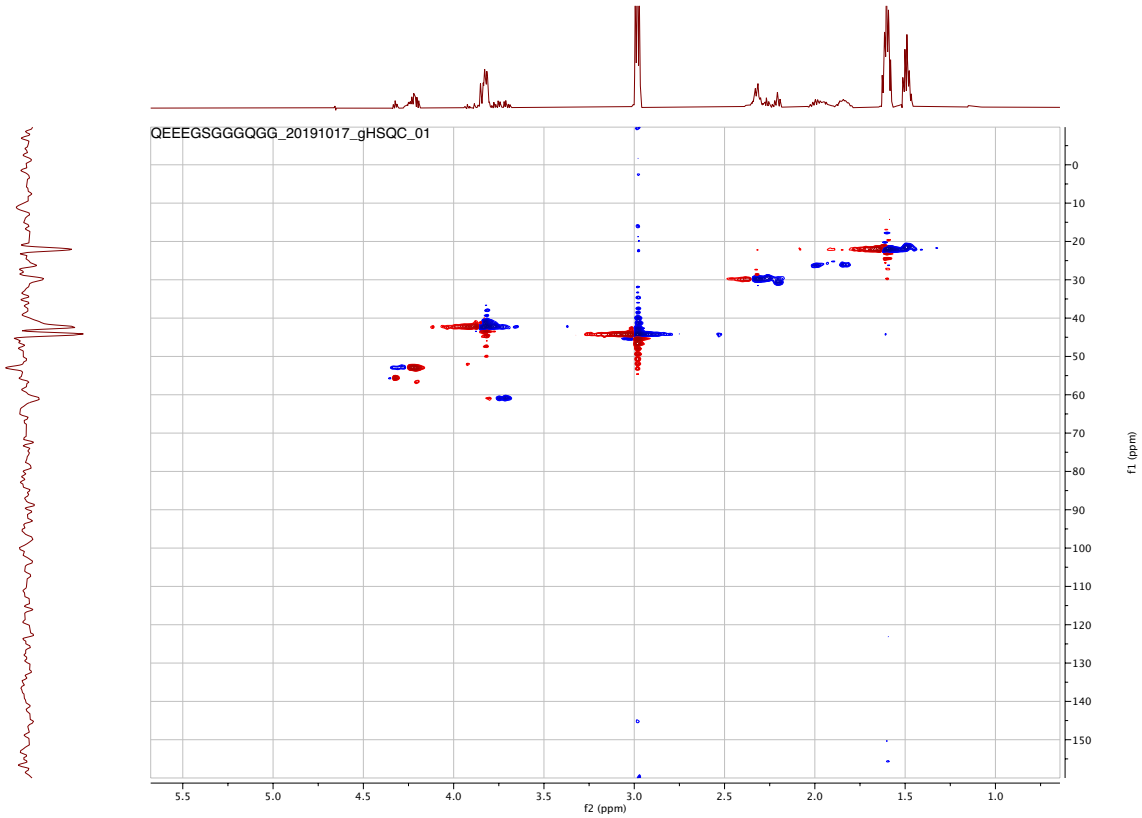
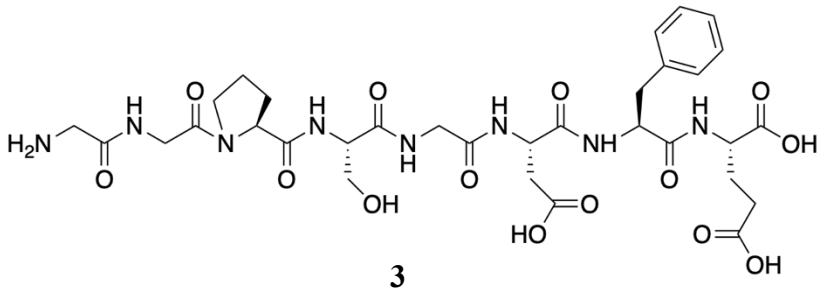
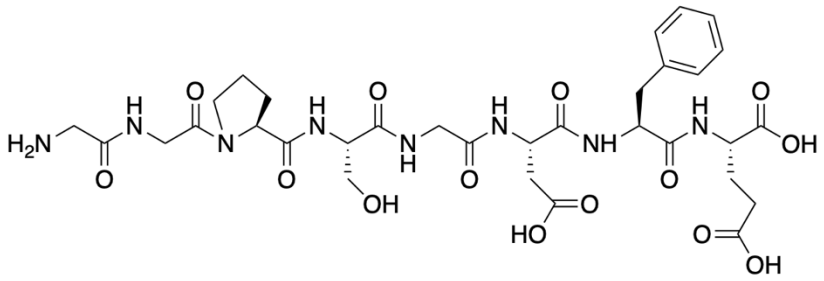


Figure 4.19 HSQC NMR of **1** (500 MHz, D_2O).



3

The purity of peptide **3** was verified with analytical C-18 HPLC (5-100% acetonitrile/water, 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$= - 111 (c 0.1, H_2O, specific rotation was collected by Po-han Lin). ^1H NMR (500 MHz, D_2O) δ 7.19 (m, 2H), 7.17 – 7.11 (m, 1H), 7.11 – 7.07 (m, 2H), 4.53 (m, 1H), 4.46 (m, 1H), 4.34 – 4.26 (m, 2H), 4.23-4.19 (m, 1H) 4.09 – 3.92 (m, 2H), 3.85 – 3.64 (m, 6H), 3.53 – 3.42 (m, 2H), 3.03-2.99 (m, 1H), 2.92 – 2.83 (m, 1H), 2.72 – 2.63 (m, 1H), 2.58-2.53 (m, 1H), 2.25-2.21 (m, 2H), 2.15-2.11 (m, 1H), 2.05 – 1.94 (m, 1H), 1.90 – 1.72 (m, 4H); ^{13}C NMR (125 MHz, D_2O) δ 177.0, 174.5, 174.5, 173.8, 172.5, 172.2, 171.7, 170.8, 169.1, 167.4, 136.1, 129.1, 128.6, 127.0, 60.9, 60.5, 55.6, 54.8, 51.9, 49.8, 47.0, 42.4, 41.6, 40.5, 40.2, 36.6, 35.0, 29.7, 29.3, 25.7, 24.3. ESI-MS: $C_{32}H_{45}N_8O_{14}$ $[M+H]^+$ calcd: 765.3050, obsd: 755.3022 (3.7 ppm)



3

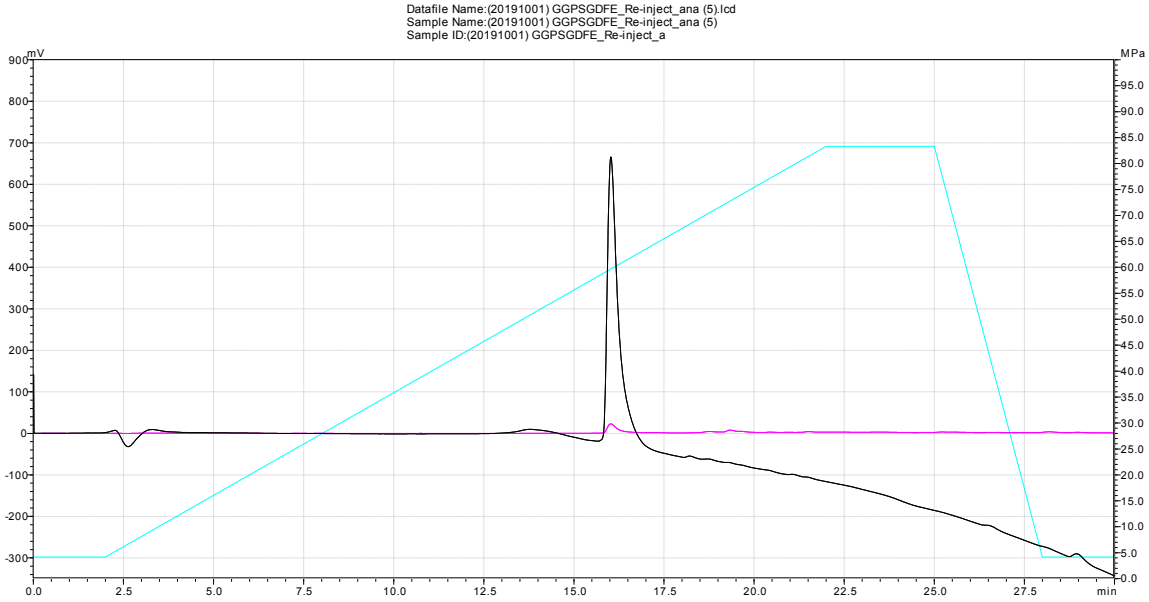
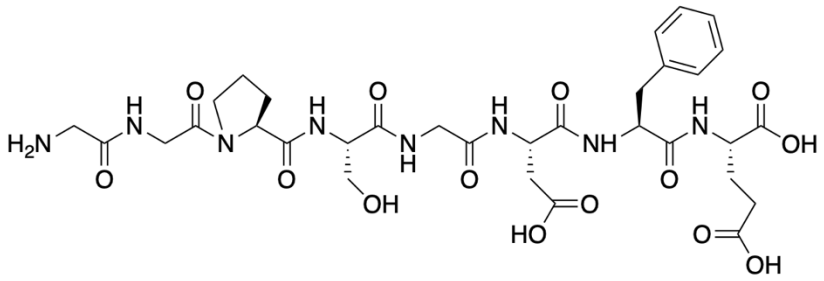


Figure 4.20 HPLC chromatogram of **3**.



3

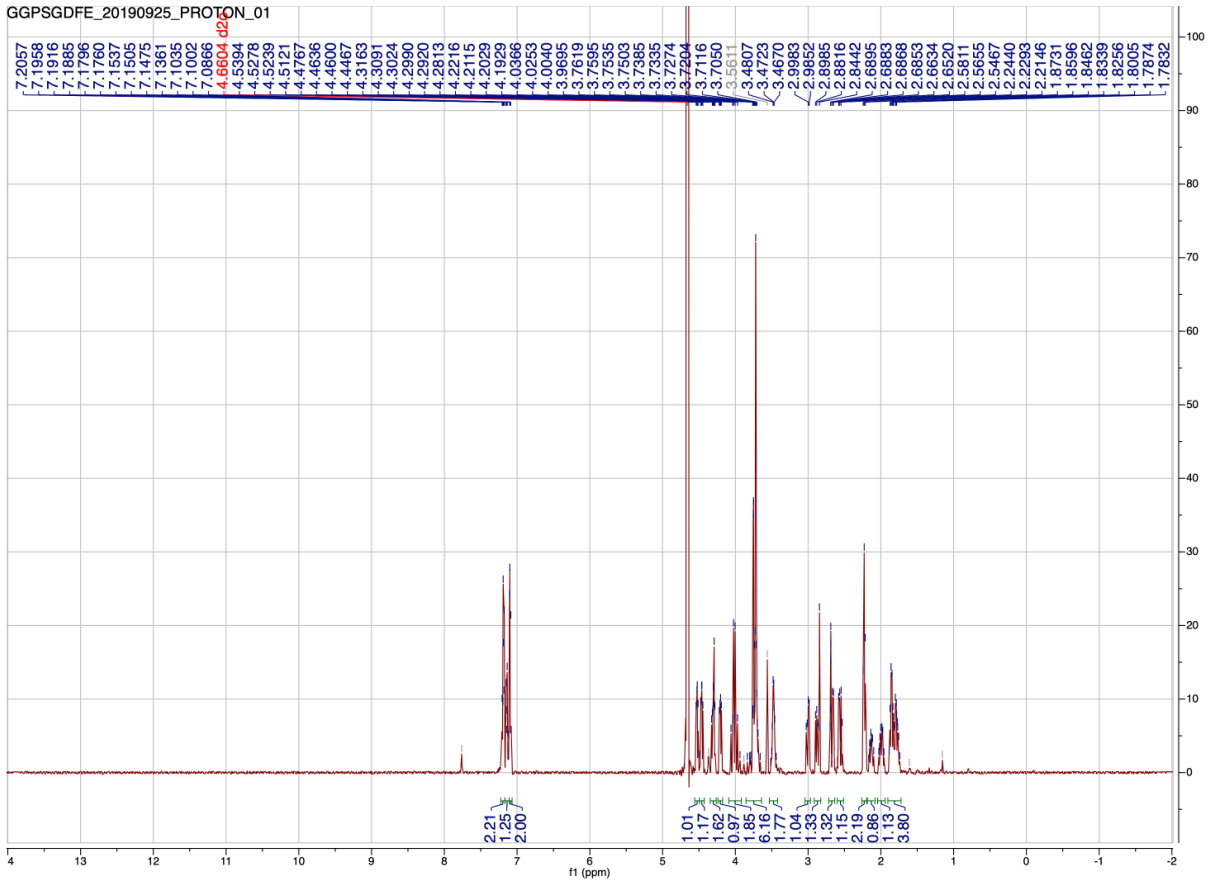
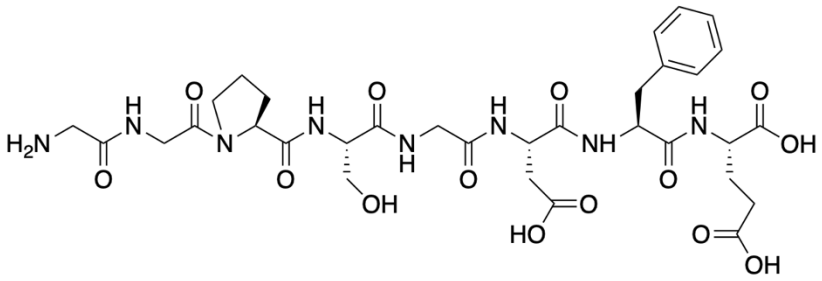
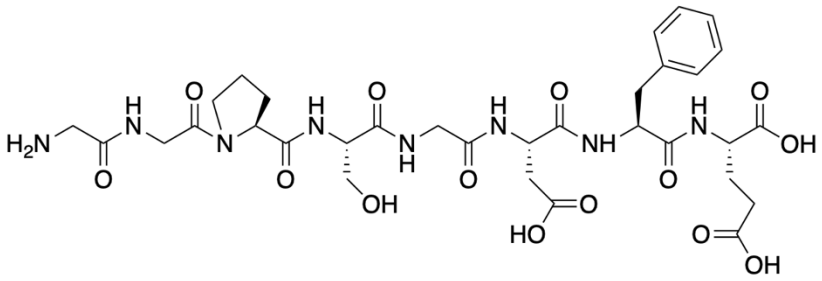


Figure 4.21 ^{1}H-NMR of **3** (500 MHz, D_2O).



3

Figure 4.22 COSY NMR of **3** (500 MHz, D_2O).



3

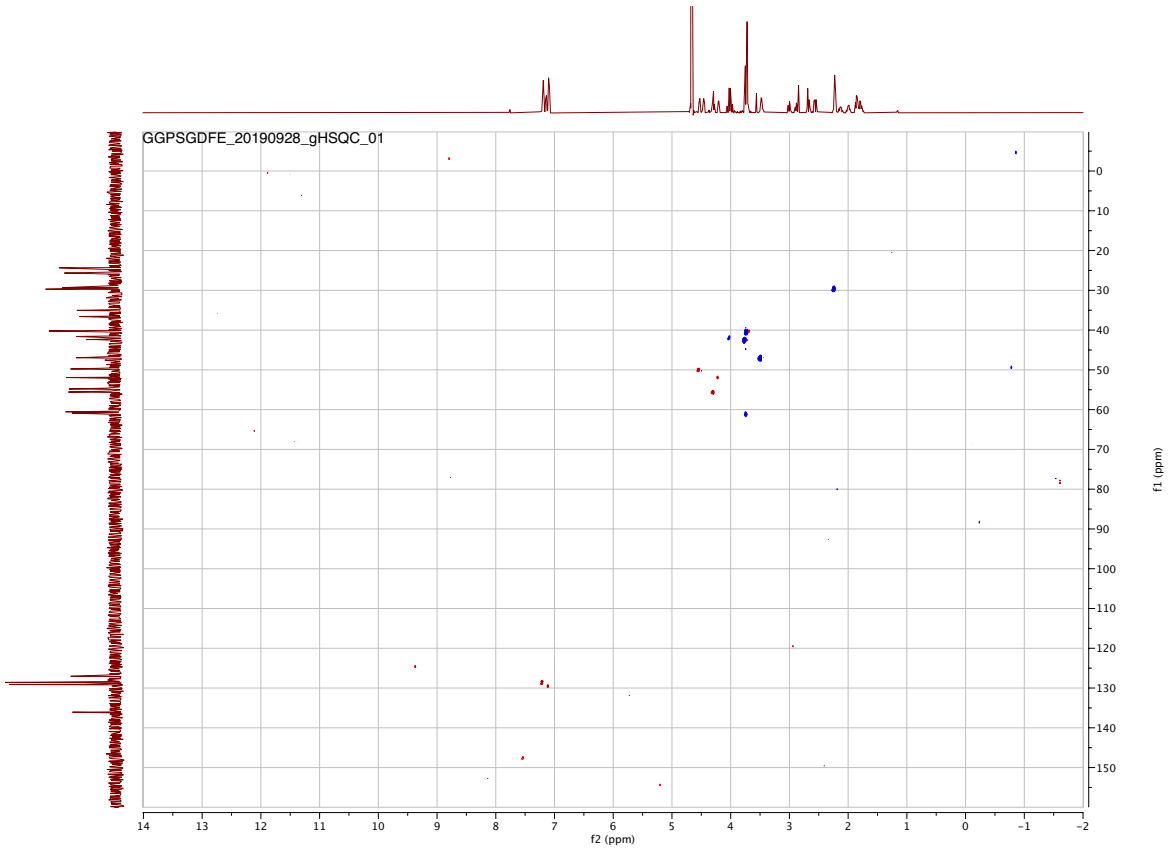
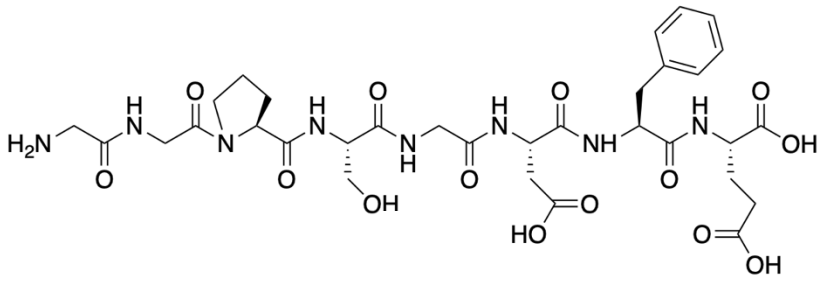
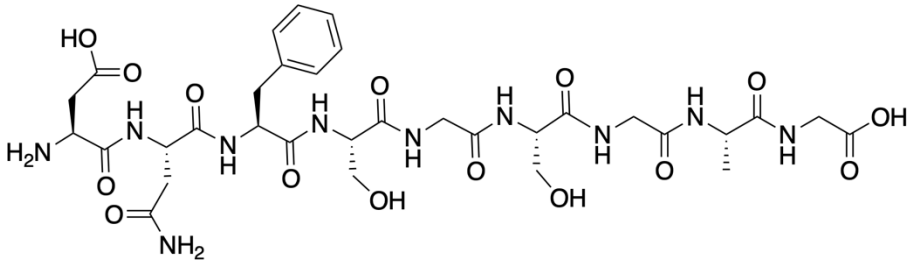


Figure 4.23 HSQC NMR of **3** (500 MHz, D_2O).



4

The purity of peptide **4** was verified with analytical C-18 HPLC (5-100% acetonitrile/water, 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$= - 329 (c 0.1, H_2O, specific rotation was collected by Po-han Lin). ^1H NMR (500 MHz, D_2O) δ 7.24 – 7.08 (m, 5H), 4.56-4.52 (m, 2H), 4.32 (t, *J* = 5.1 Hz, 1H), 4.27 (t, *J* = 5.2 Hz, 1H), 4.20 (q, *J* = 7.2 Hz, 1H), 4.13 (t, *J* = 6.3 Hz, 1H), 3.86-3.84 (m, 1H), 3.82 – 3.80 (m, 2H), 3.78 (d, *J* = 12.8 Hz, 3H), 3.75 (d, *J* = 6.1 Hz, 1H), 3.74 – 3.62 (m, 4H), 3.03-3.01 (m, 1H), 2.92-2.88 (m, 1H), 2.78 – 2.71 (m, 2H), 2.62-2.60 (m, 1H), 2.52-2.50 (m, 1H), 1.22 (d, *J* = 7.2 Hz, 3H); ^{13}C NMR (125 MHz, D_2O) δ 129.1, 129.1, 128.7, 127.1, 60.9, 60.9, 60.9, 55.6, 55.4, 55.0, 50.3, 49.6, 49.4, 42.3, 42.3, 41.9, 41.9, 41.2, 36.7, 36.7, 35.9, 35.9, 35.1, 16.5. ESI-MS: $C_{32}H_{47}N_{10}O_{15}$ $[M+H]^+$ calcd: 811.3217, obsd: 811.3207 (1.2 ppm)



4

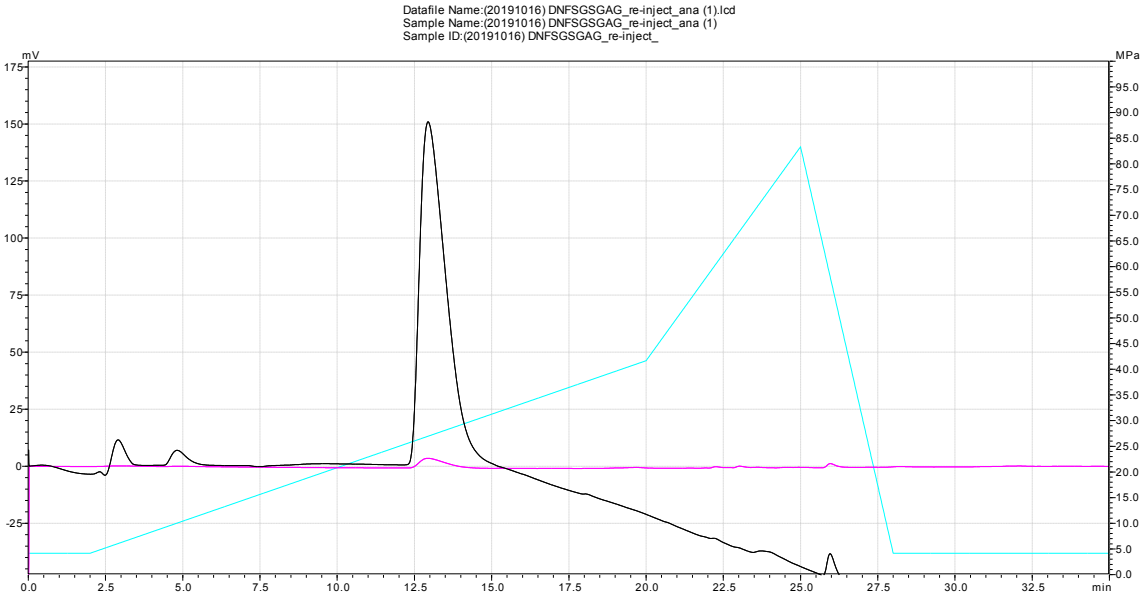
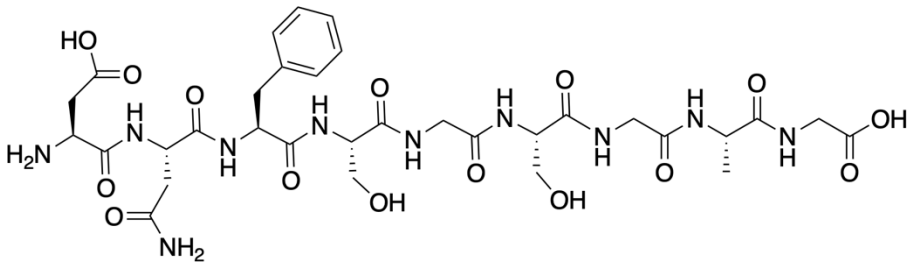


Figure 4.24 HPLC chromatogram of **4** (500 MHz, D_2O).



4

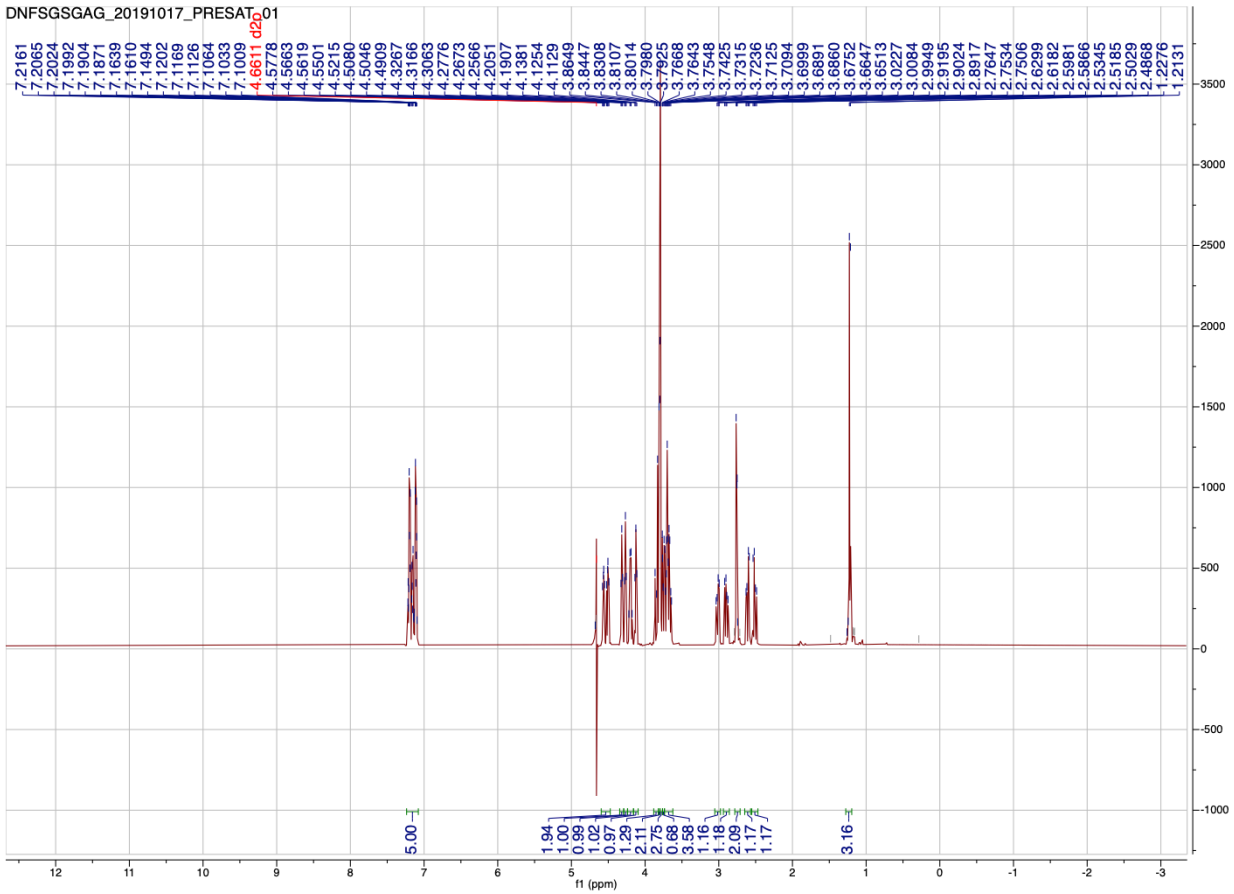
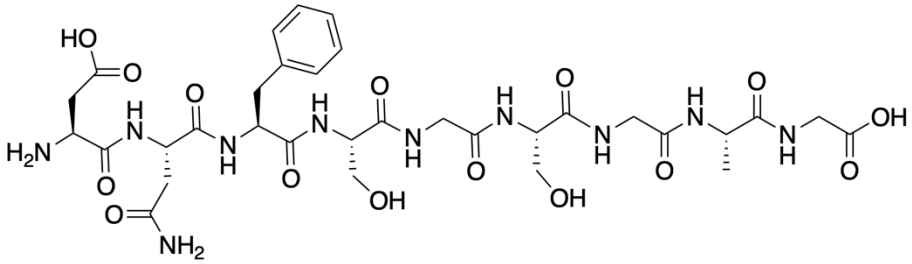


Figure 4.25 ^{1}H-NMR of **4** (500 MHz, D_2O).



4

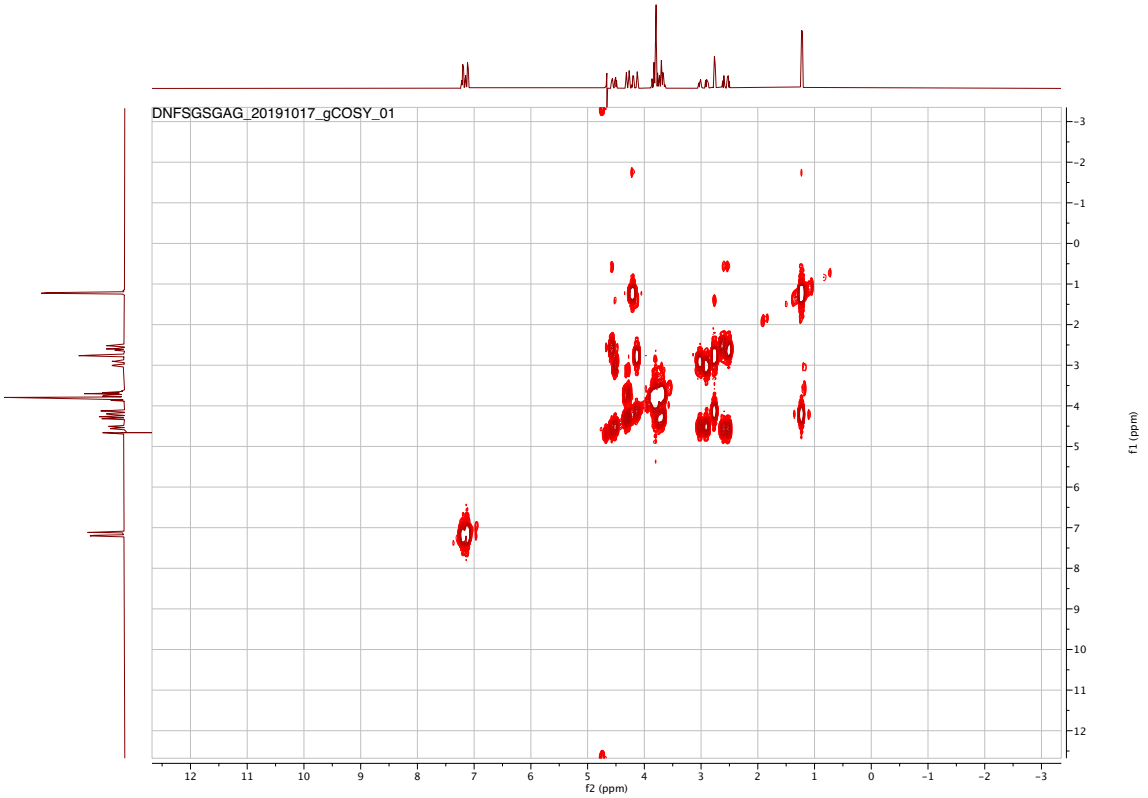
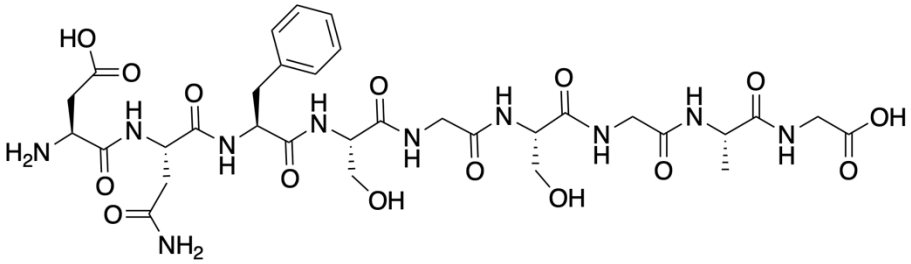


Figure 4.26 COSY NMR of **4** (500 MHz, D_2O).



4

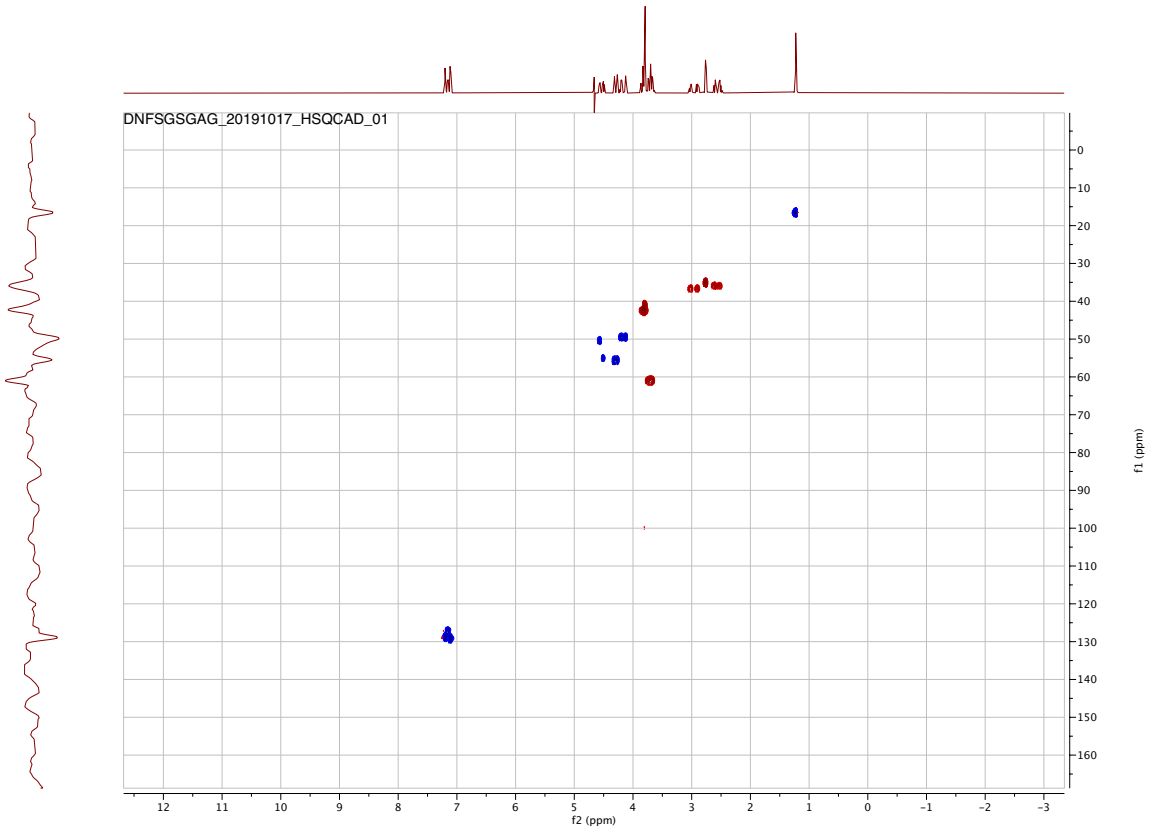
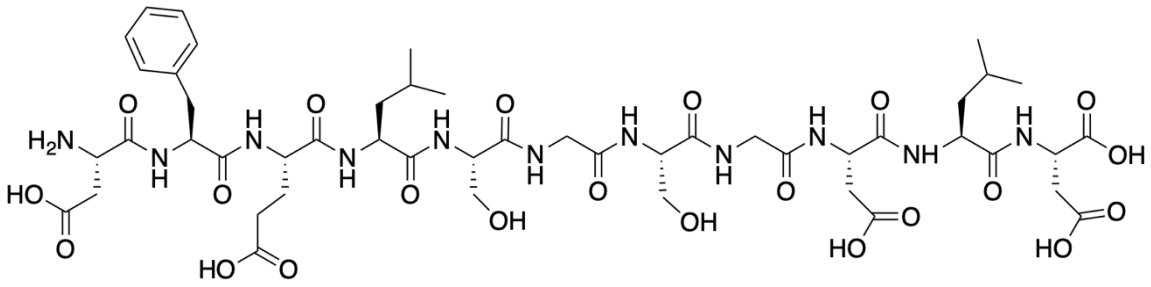
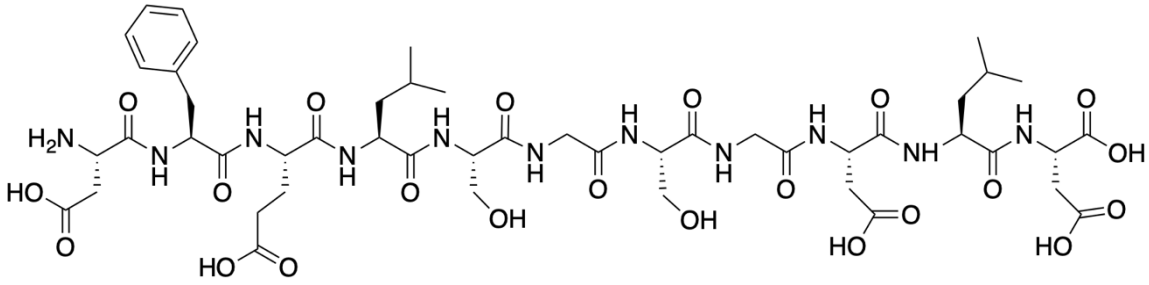


Figure 4.27 HSQC NMR of **4** (500 MHz, D_2O).



5

The purity of peptide **5** was verified with analytical C-18 HPLC (5-100% acetonitrile/water, 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$= + 76 (c 0.1, H_2O, specific rotation was collected by Po-han Lin). ^1H NMR (500 MHz, D_2O) δ 7.24 – 7.18 (m, 2H), 7.18 – 7.12 (m, 1H), 7.12 – 7.06 (m, 2H), 4.55 (s, 1H), 4.52 – 4.45 (m, 1H), 4.32 – 4.26 (m, 1H), 4.24 – 4.08 (m, 3H), 3.92 – 3.77 (m, 4H), 3.77 – 3.66 (m, 4H), 2.92 (t, *J* = 8.5 Hz, 2H), 2.86 – 2.70 (m, 5H), 2.70 – 2.61 (m, 1H), 2.26 – 2.19 (m, 2H), 1.92 – 1.81 (m, 1H), 1.74 (dt, *J* = 14.1, 7.3 Hz, 1H), 1.53 – 1.42 (m, 6H), 1.16 (d, *J* = 1.2 Hz, 1H), 0.85 – 0.74 (m, 10H), 0.72 – 0.67 (m, 4H). ESI-MS: $C_{48}H_{72}N_{11}O_{22}$ $[M+H]^+$ calcd: 1154.4848, obsd: 1154.4822 (2.3 ppm)



5

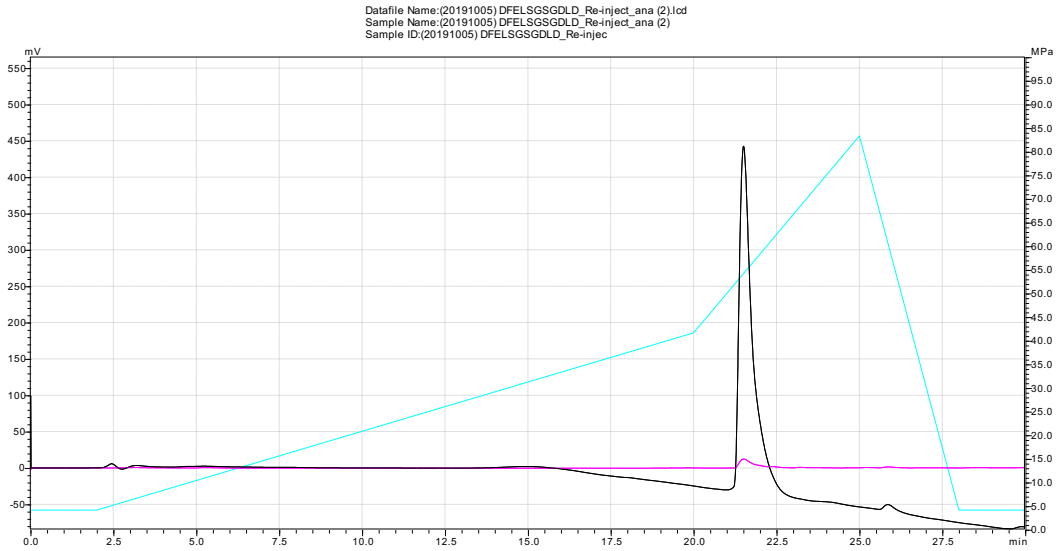
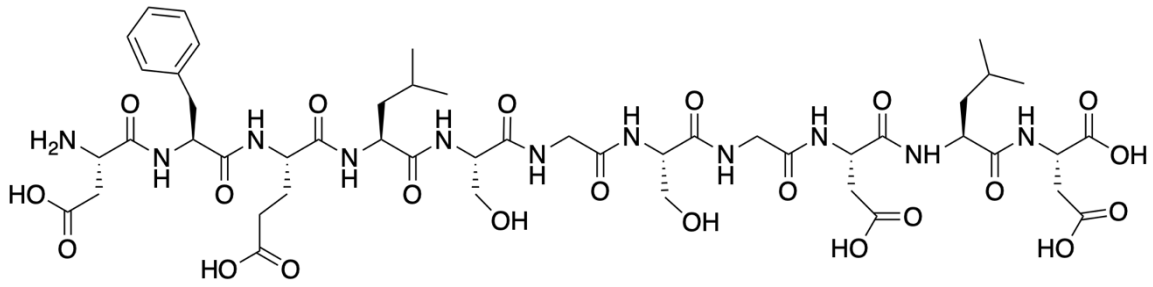


Figure 4.28 HPLC chromatogram of **5**.



5

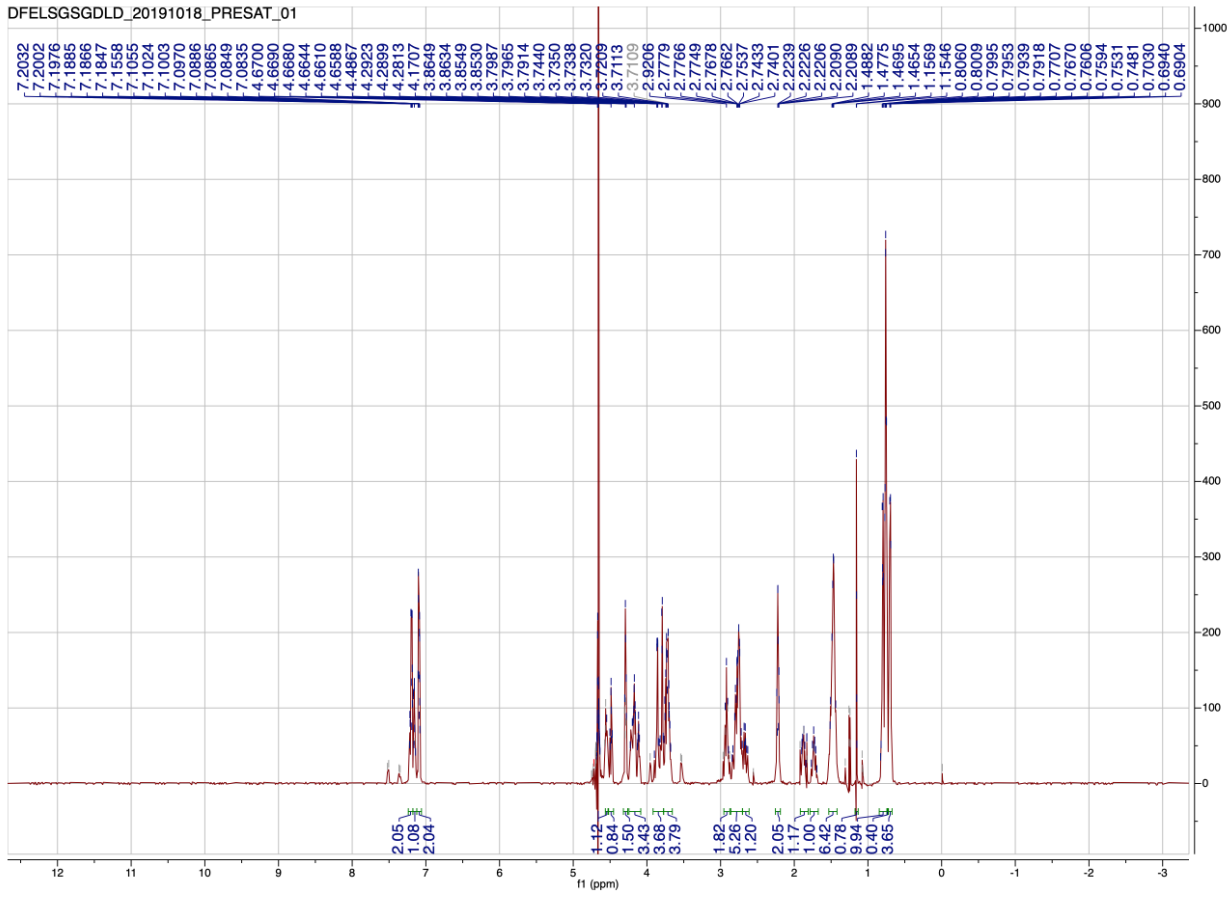
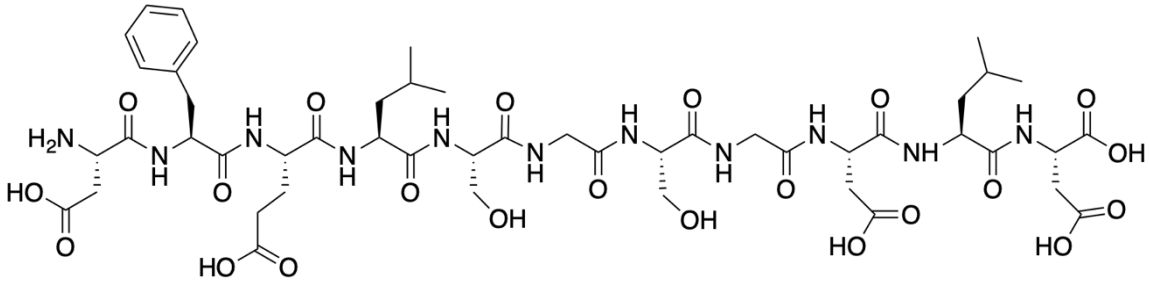


Figure 4.29 ^{1}H-NMR of **5** (500 MHz, D_2O).



5

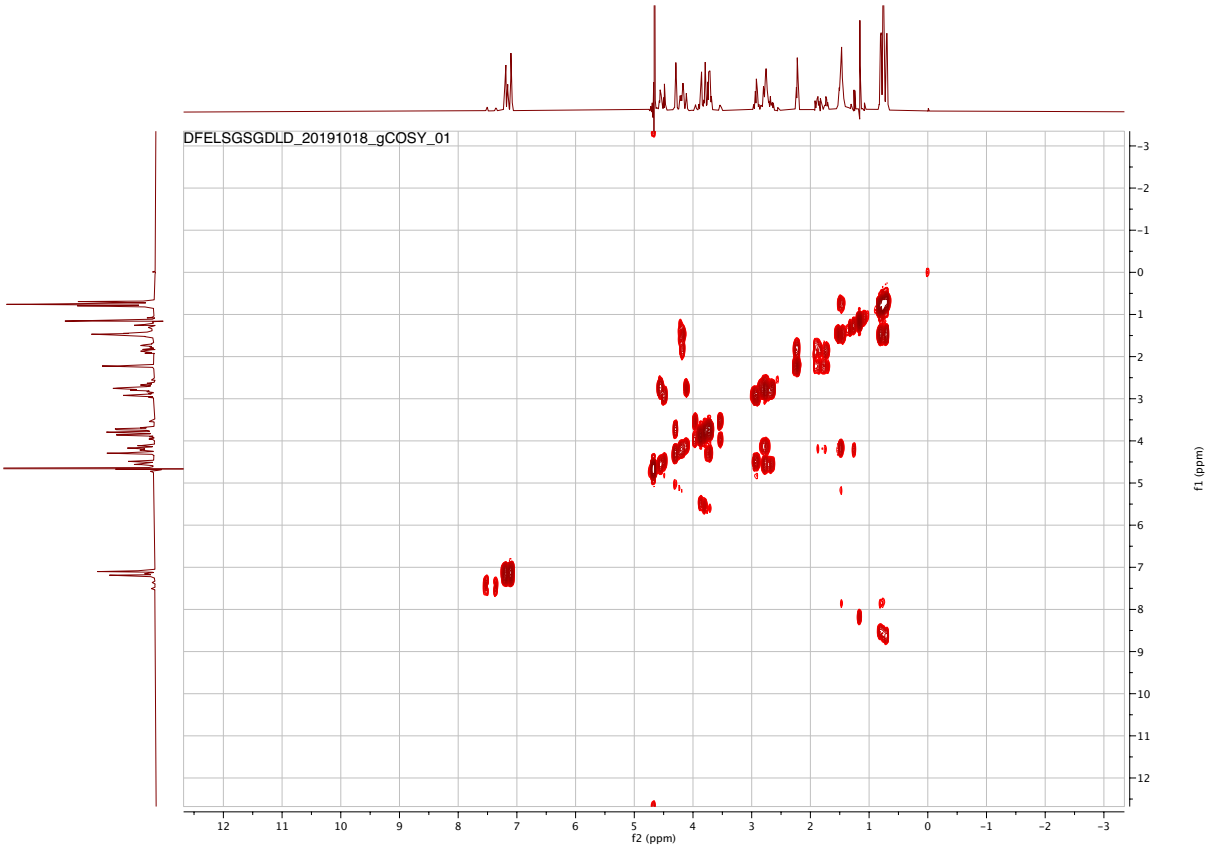
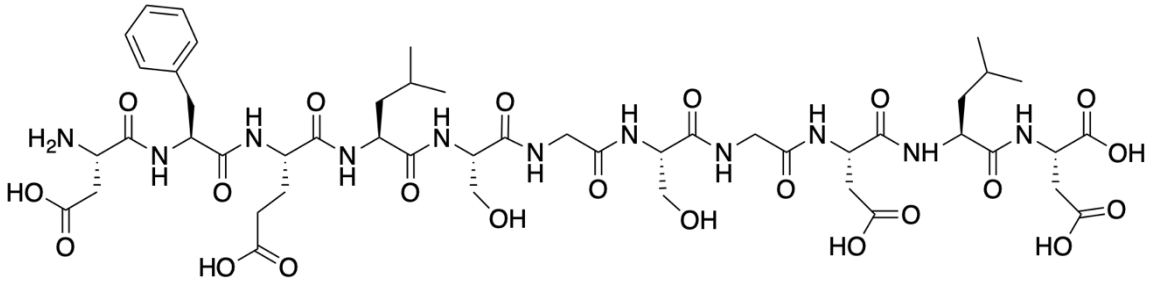


Figure 4.30 COSY NMR of **5** (500 MHz, D_2O).



5

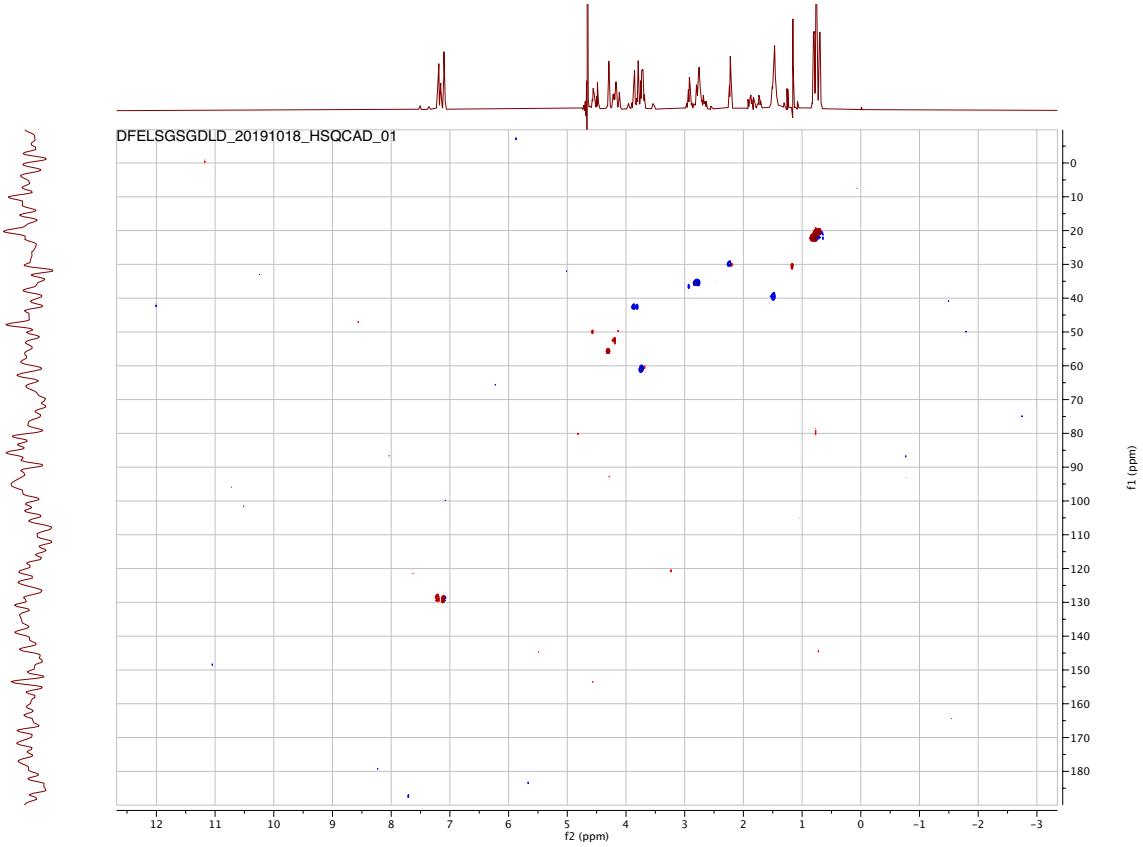
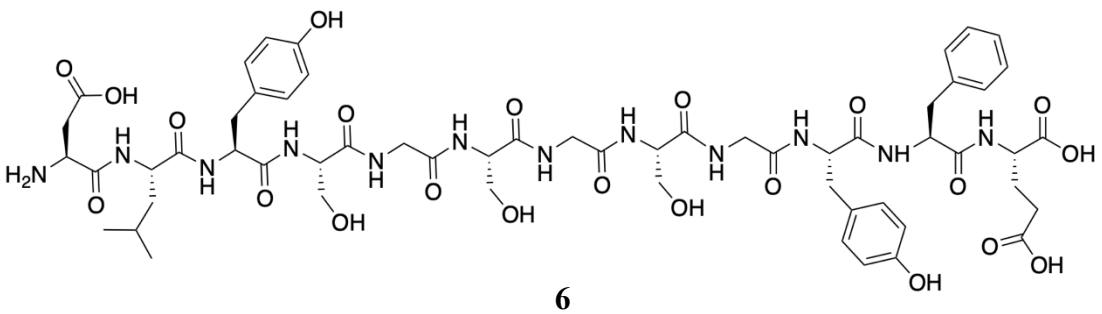
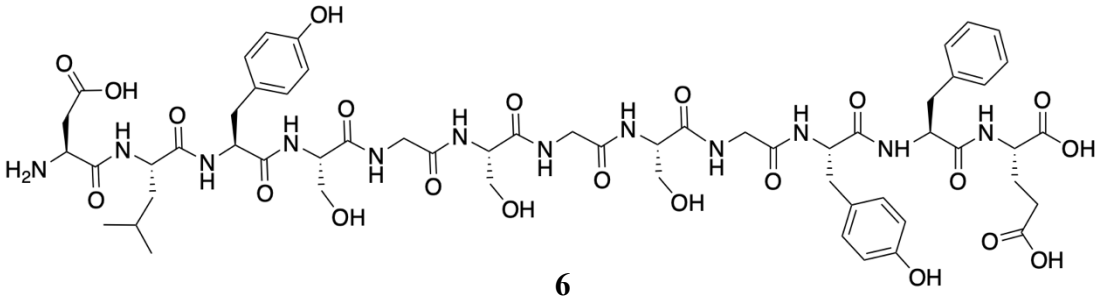


Figure 4.31 HSQC NMR of **5** (500 MHz, D_2O).



6

The purity of peptide **6** was verified with analytical C-18 HPLC (5-100% acetonitrile/water, 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = -36 (c 0.1, H_2O, specific rotation was collected by Po-han Lin). ^{1}H NMR (500 MHz, D_2O) δ 7.21 – 7.08 (m, 3H), 7.07 – 7.01 (m, 2H), 6.96 – 6.90 (m, 2H), 6.85 – 6.79 (m, 2H), 6.67 – 6.57 (m, 4H), 4.43-4.40 (m, 2H), 4.35 – 4.22 (m, 4H), 4.16 – 4.08 (m, 3H), 3.92-3.88 (m, 1H), 3.84 – 3.71 (m, 5H), 3.71 – 3.64 (m, 6H), 3.61 (dd, *J* = 11.9, 5.2 Hz, 2H), 2.92-2.88 (m, 2H), 2.84 – 2.62 (m, 7H), 2.55 (s, 1H), 2.20 (t, *J* = 7.4 Hz, 2H), 1.97-1.93 (m, 1H), 1.76-1.72 (m, 1H), 1.36 – 1.29 (m, 2H), 1.27-1.23 (m, 2H), 0.70-0.66 (m, 6H). ^{13}C NMR (125 MHz, D_2O) δ 135.9, 130.9, 130.8, 130.4, 130.3, 129.1, 128.4, 127.0, 117.0, 115.5, 115.3, 115.2, 114.3, 113.0, 111.2, 110.2, 88.8, 65.2, 62.7, 62.1, 61.1, 60.2, 59.9, 58.6, 58.1, 57.9, 55.9, 55.7, 55.6, 55.4, 55.1, 54.9, 54.7, 54.6, 52.9, 52.6, 51.3, 51.1, 49.6, 42.5, 42.3, 39.6, 39.5, 38.6, 37.0, 36.2, 36.0, 35.7, 32.2, 29.9, 27.7, 25.9, 25.7, 24.0, 21.9, 20.7, 20.7, 16.6. ESI-MS: $C_{57}H_{77}N_{12}O_{22}$ $[M+H]^+$ calcd: 1281.5270, obsd: 1281.5206 (5.0 ppm)



6

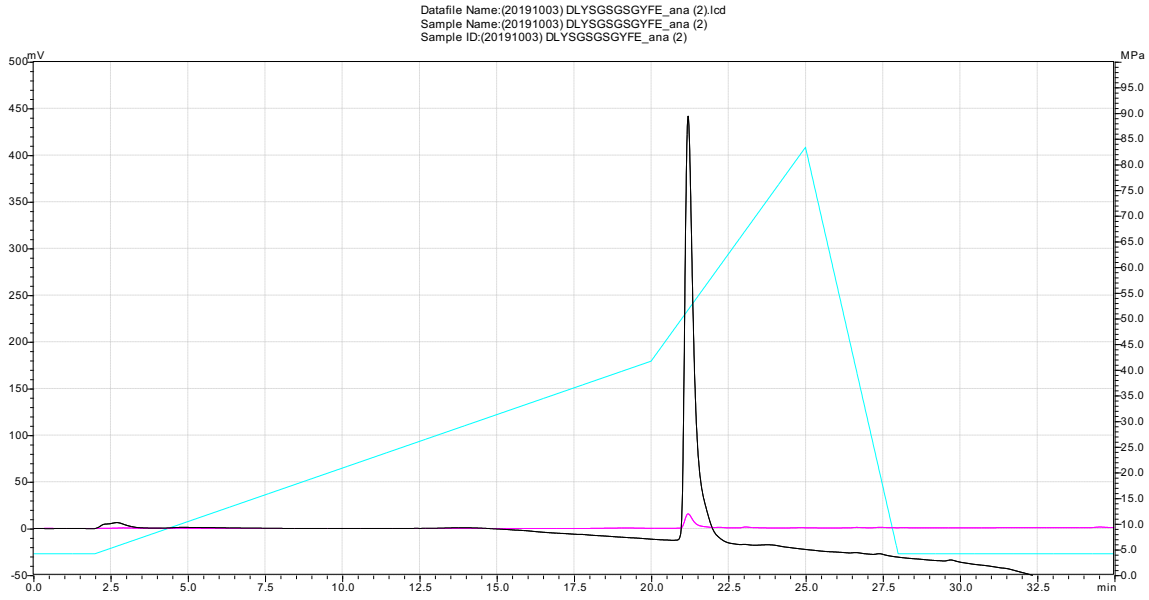
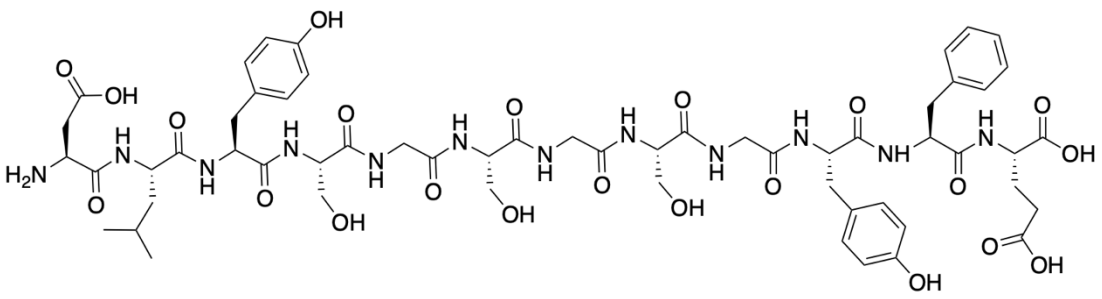


Figure 4.32 HPLC chromatogram of **6**.



6

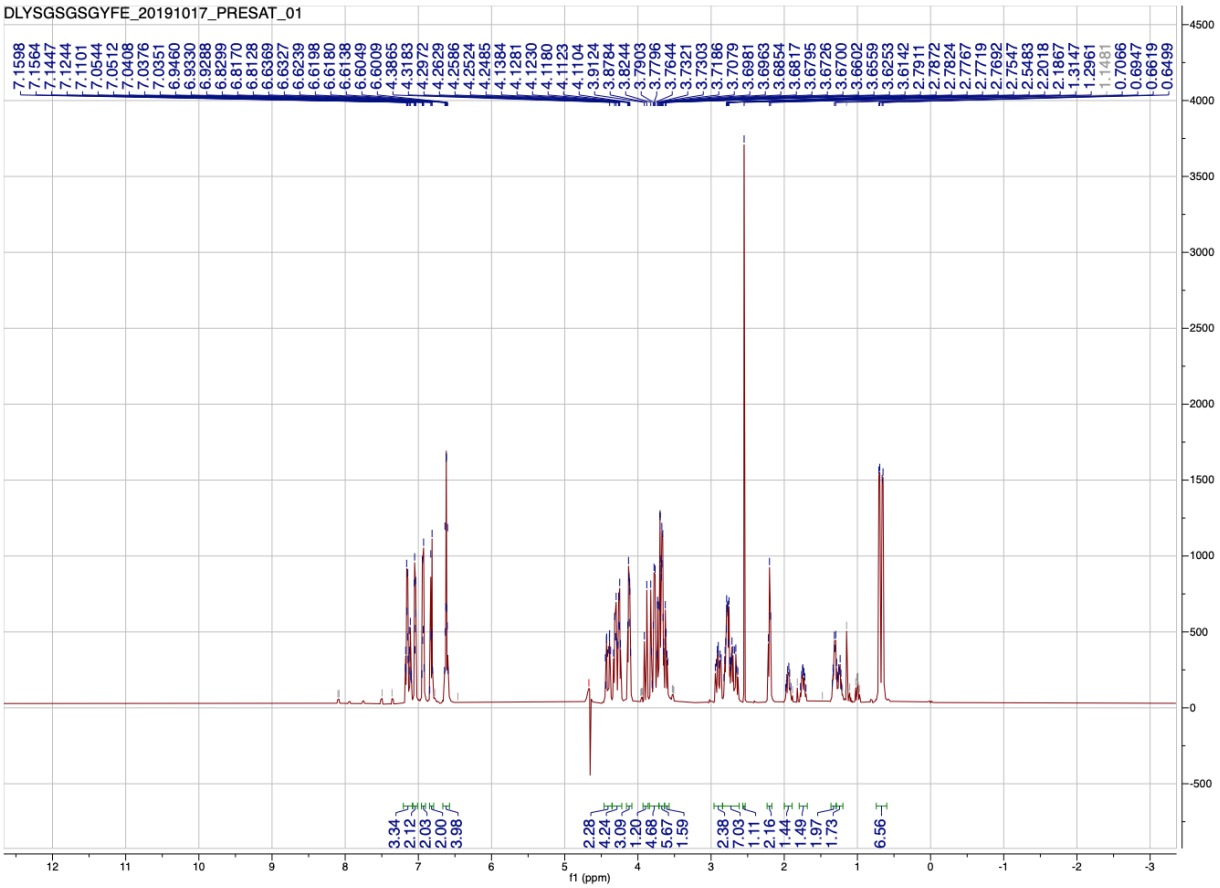
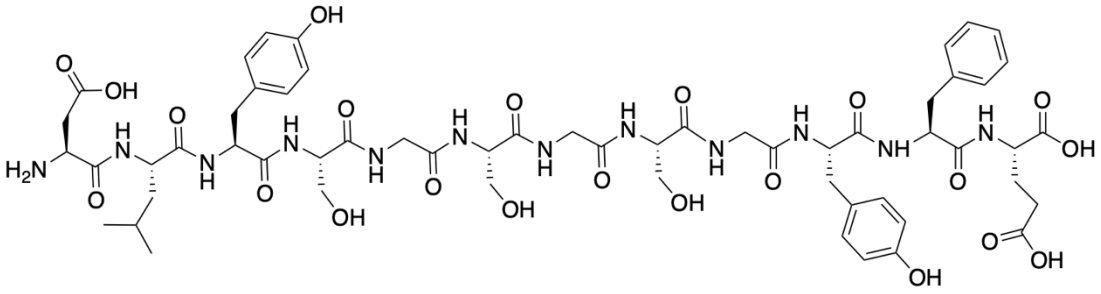


Figure 4.33 ^{1}H-NMR of **6** (500 MHz, D_2O).



6

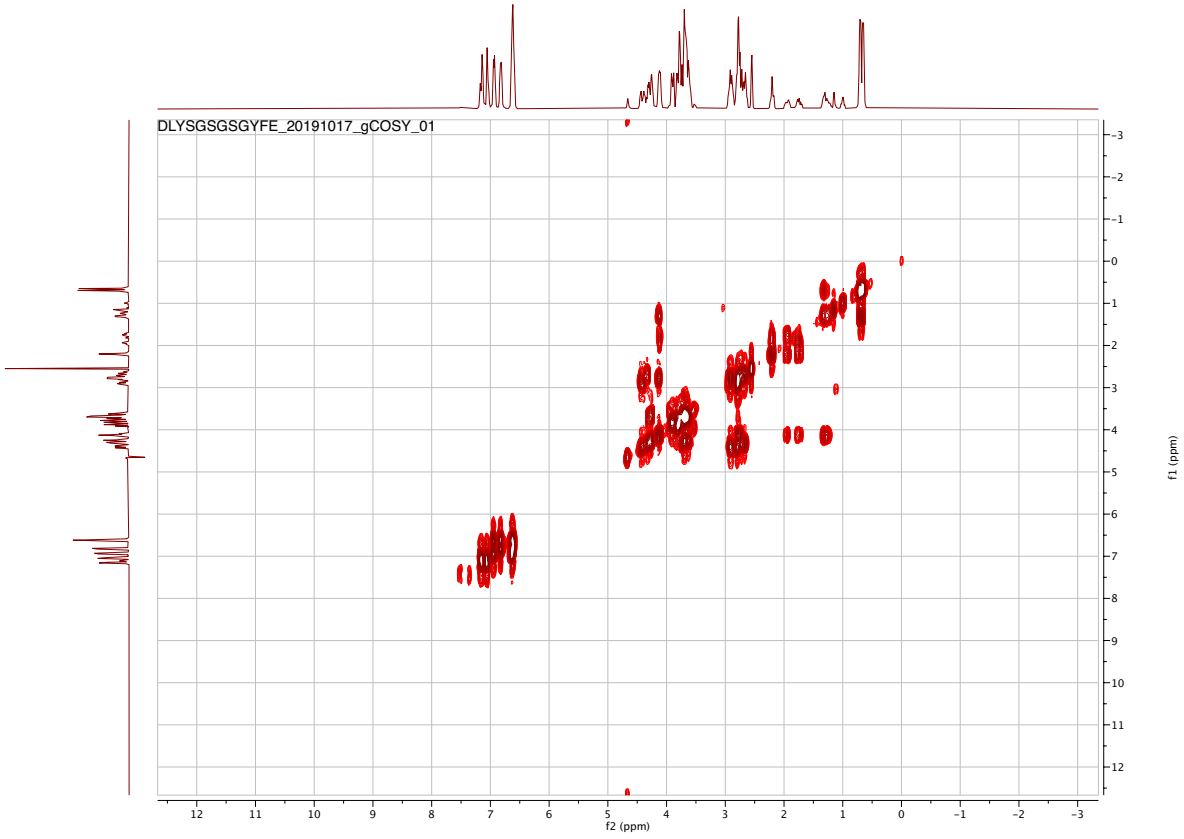
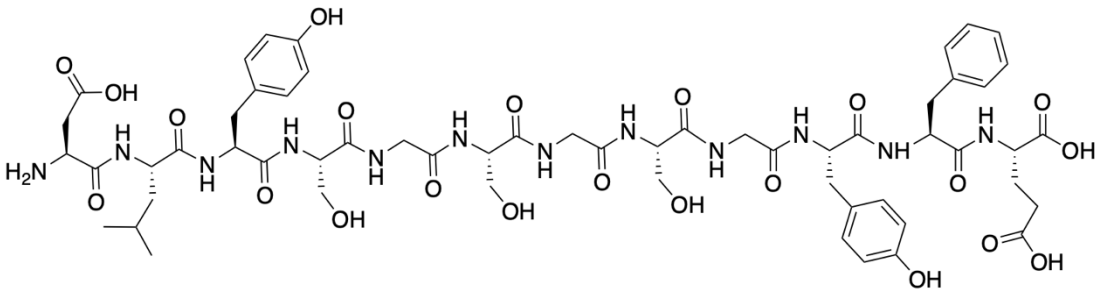


Figure 4.34 COSY NMR of **6** (500 MHz, D_2O).



6

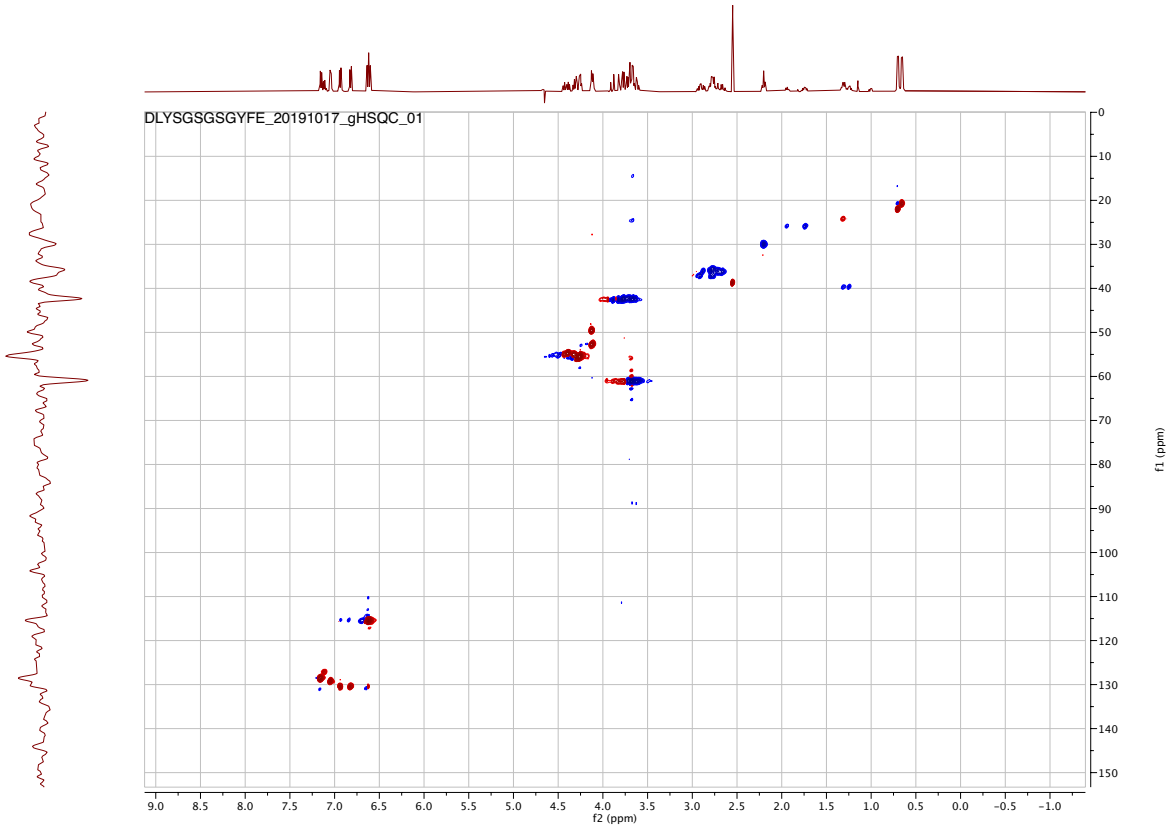
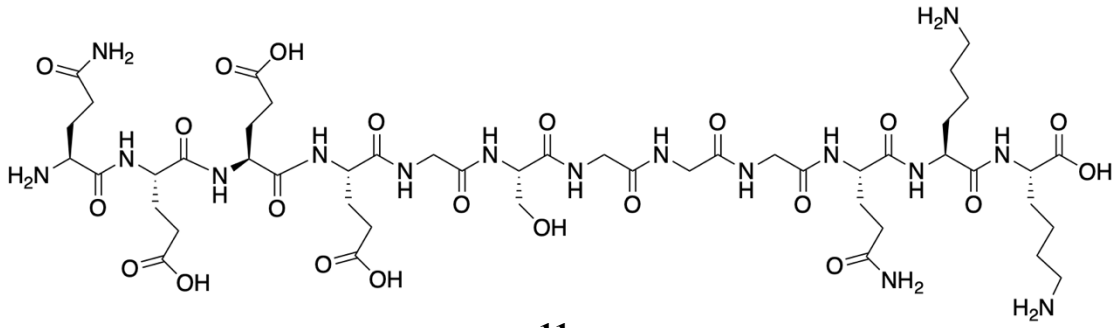
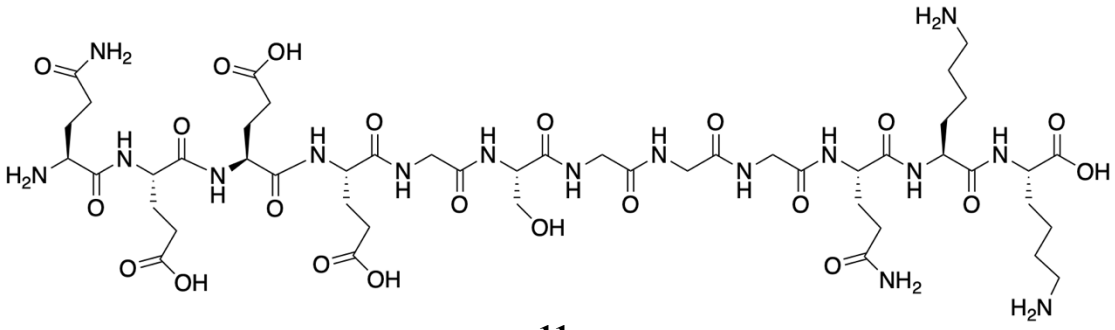


Figure 4.35 HSQC NMR of **6** (500 MHz, D_2O).



11

The purity of peptide **11** was verified with analytical C-18 HPLC (water, 0.1% trifluoroacetic acid). $[\alpha]_D^{20}$ = + 31 (c 0.1, H_2O). ^1H NMR (500 MHz, D_2O) δ 4.35 – 4.08 (m, 3H), 3.98 – 3.81 (m, 3H), 3.81 – 3.67 (m, 2H), 3.08 – 2.92 (m, 8H), 2.83 (s, 2H), 2.41 – 2.16 (m, 6H), 2.07 – 1.70 (m, 6H), 1.64-1.56 (m, 10H), 1.55 – 1.45 (m, 7H), 1.31-1.27 (m, 3H). ^{13}C NMR (125 MHz, D_2O) δ 60.9, 56.7, 55.6, 54.1, 53.3, 52.9, 52.1, 46.6, 45.5, 45.3, 44.8, 44.6, 44.3, 44.1, 43.6, 43.5, 43.3, 42.6, 42.3, 42.1, 40.8, 39.3, 39.1, 38.1, 36.8, 34.0, 32.2, 31.8, 31.3, 30.7, 30.2, 29.9, 29.7, 28.7, 27.4, 27.2, 27.0, 26.4, 26.2, 26.0, 25.9, 24.5, 23.7, 23.2, 22.9, 22.0, 21.7, 21.4, 21.2, 21.1, 20.4, 19.7, 17.7. ESI-MS: $C_{48}H_{81}N_{16}O_{22}$ $[M+H]^+$ calcd: 1233.5706, obsd: 1233.5679 (2.2 ppm)



11

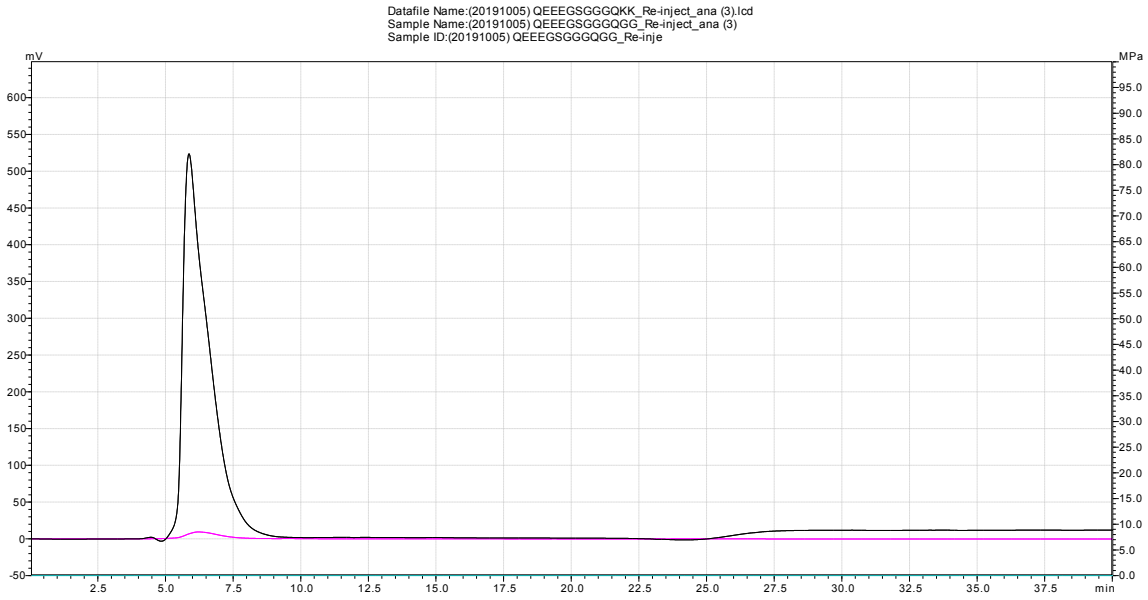
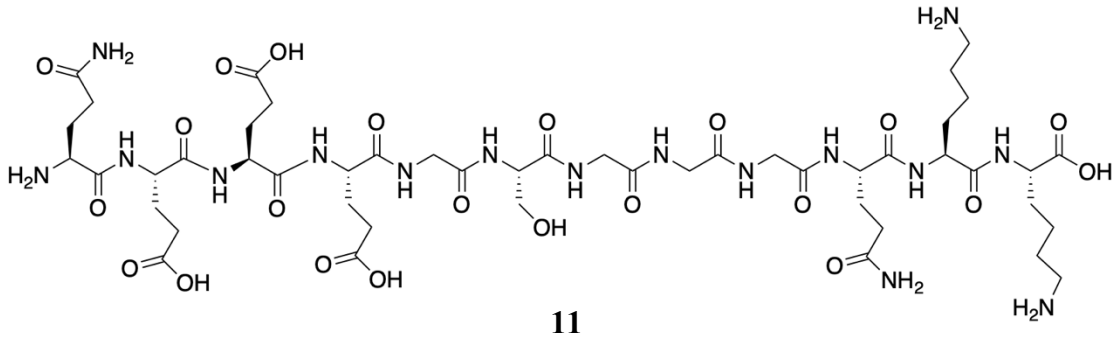


Figure 4.36 HPLC chromatogram of **11**.



11

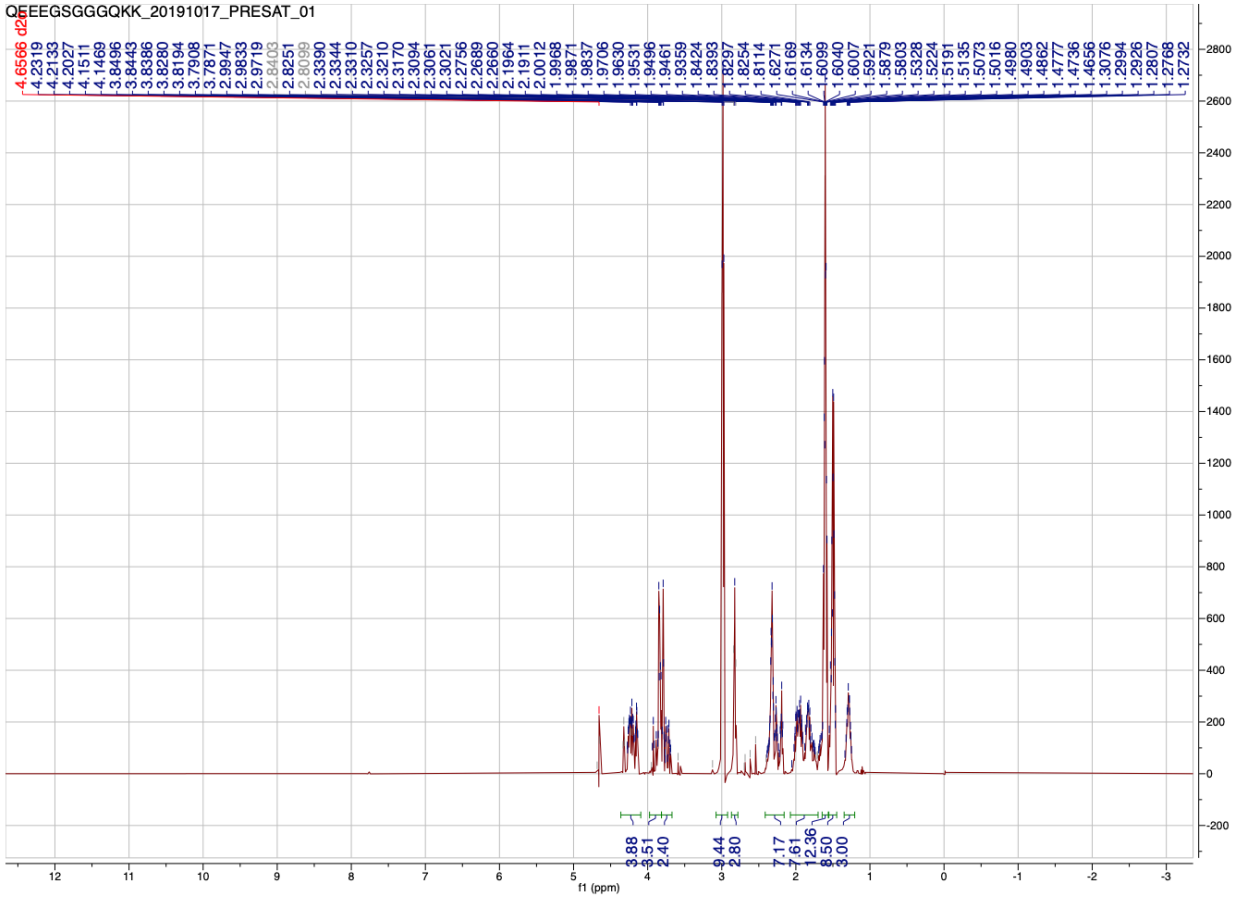
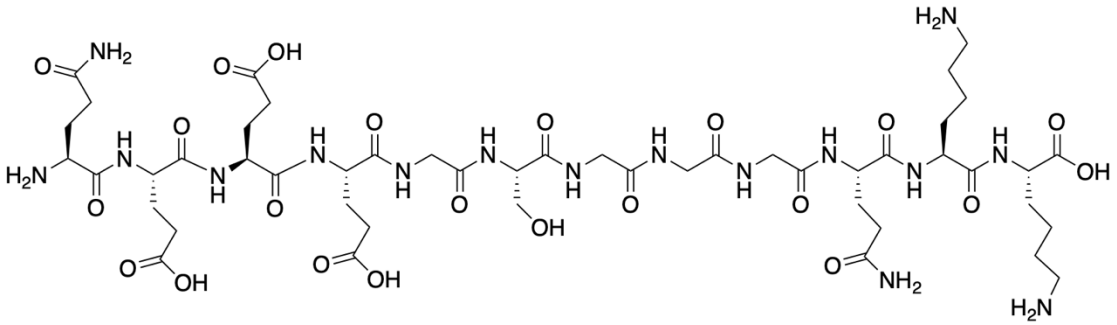


Figure 4.37 ^{1}H-NMR of **11** (500 MHz, D_2O).



11

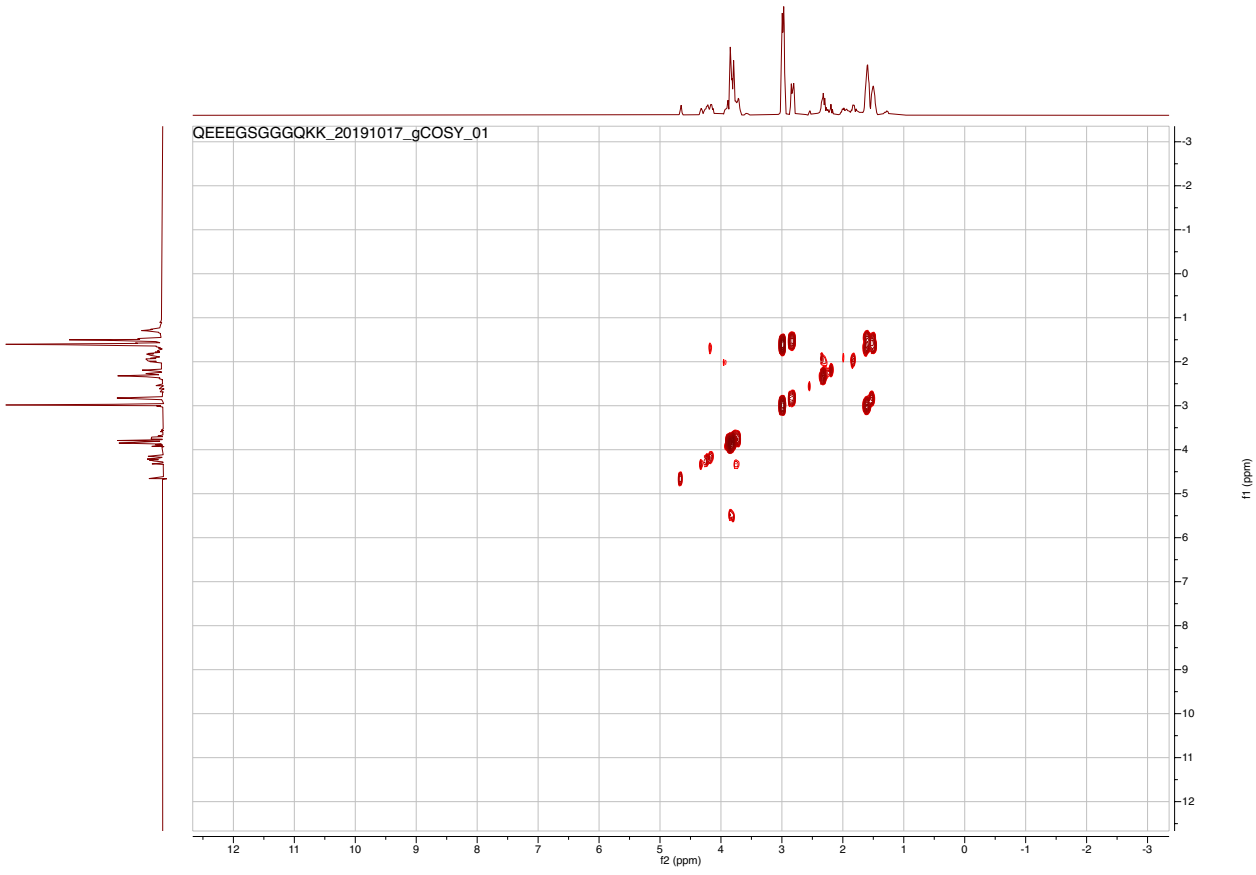
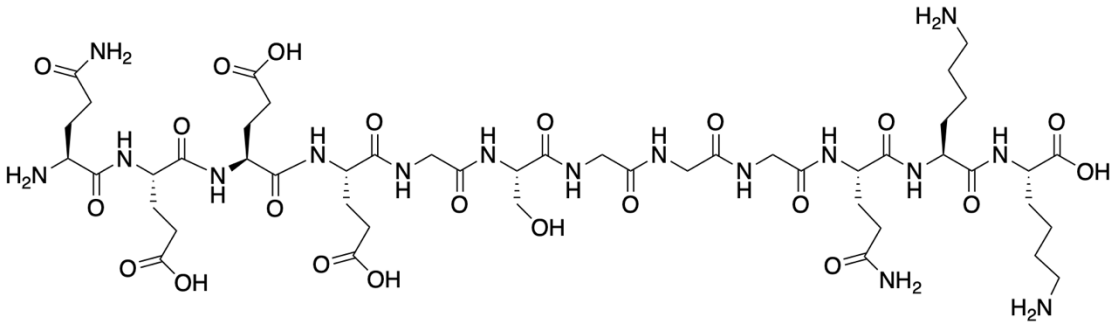


Figure 4.38 COSY NMR of **11** (500 MHz, D_2O).



11

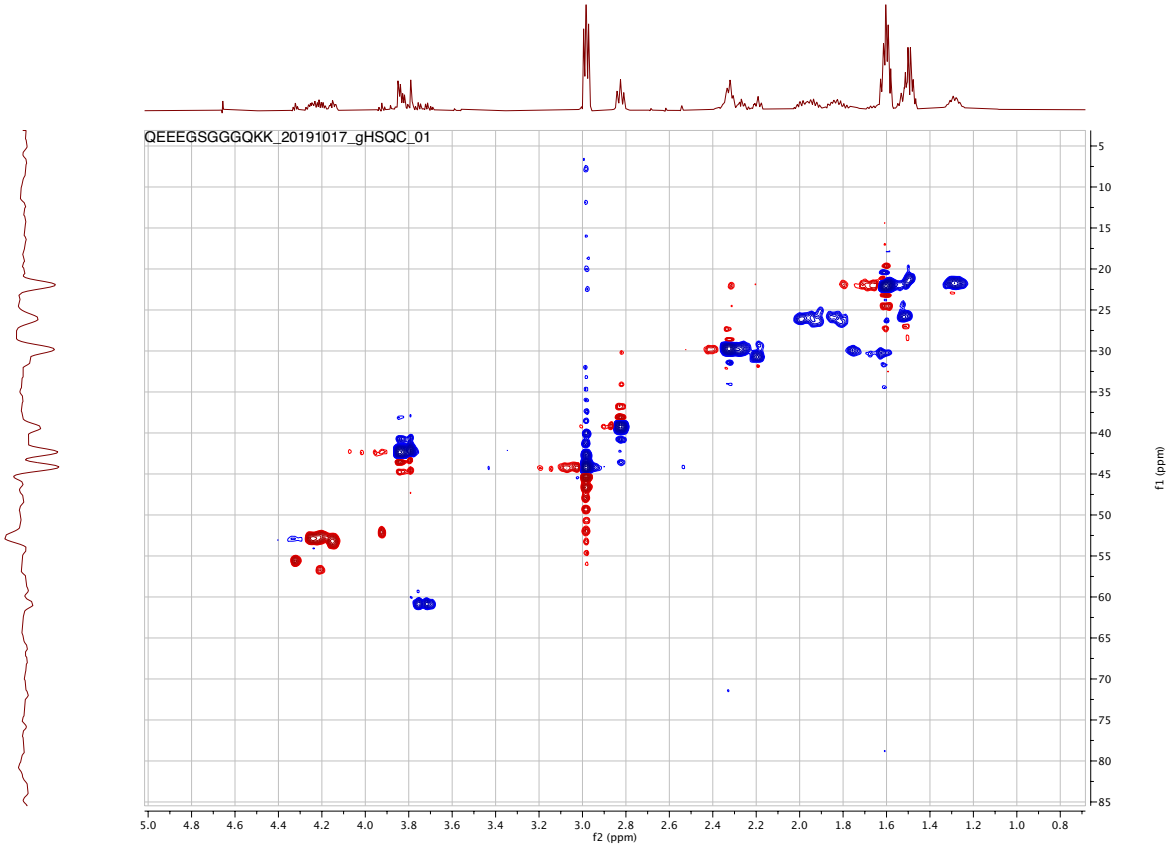


Figure 4.39 HSQC NMR of **11** (500 MHz, D_2O).

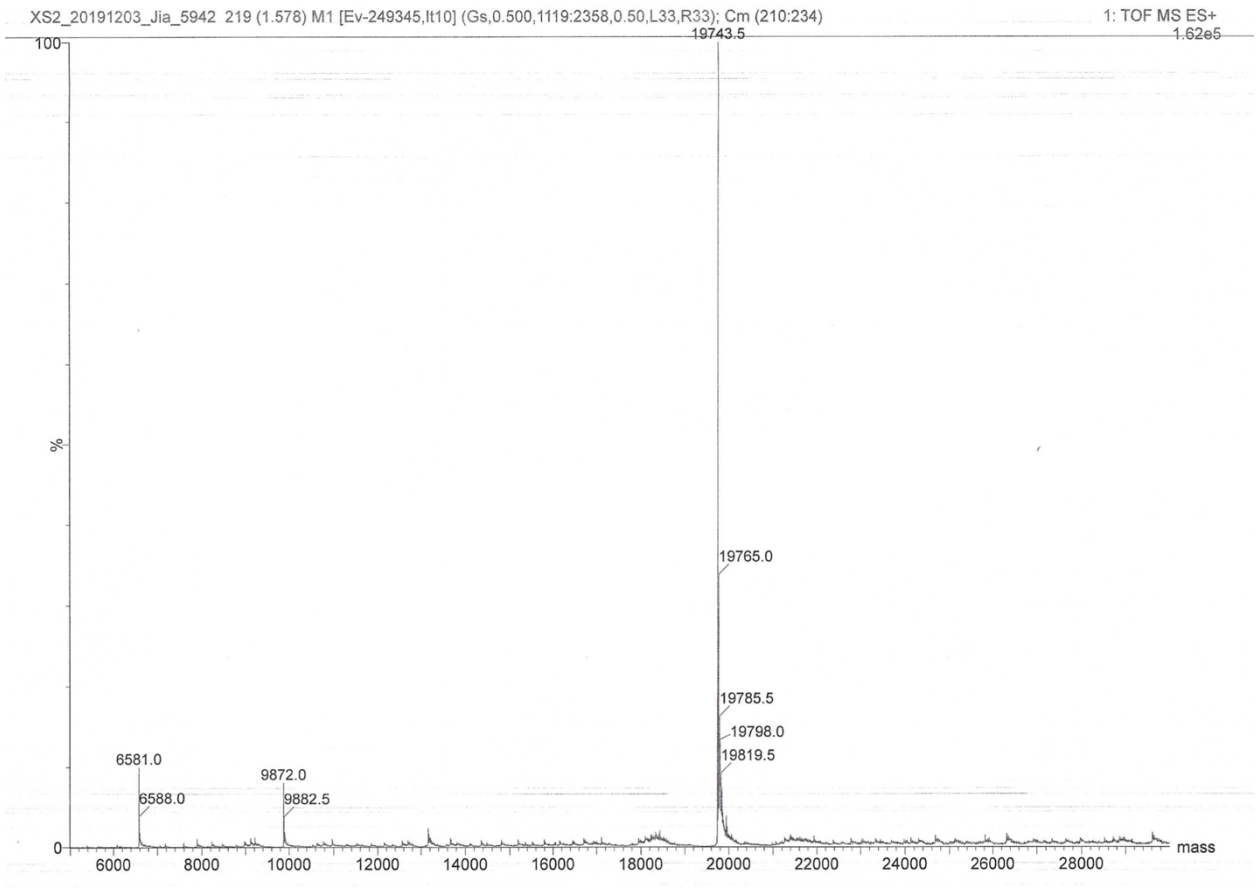


Figure 4.40 ESI-MS of recombinant CD44.

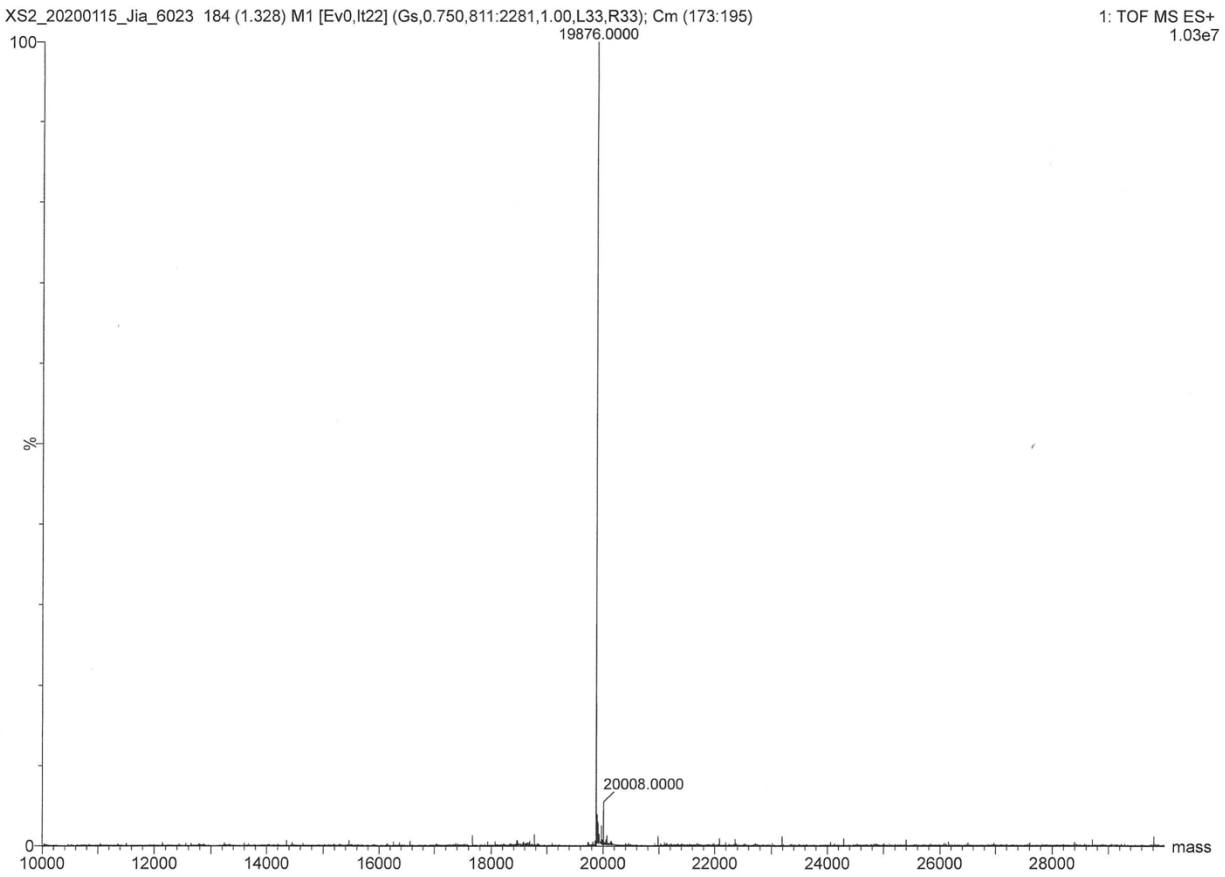


Figure 4.41 ESI-MS of CD44 (*O*-Xyl).

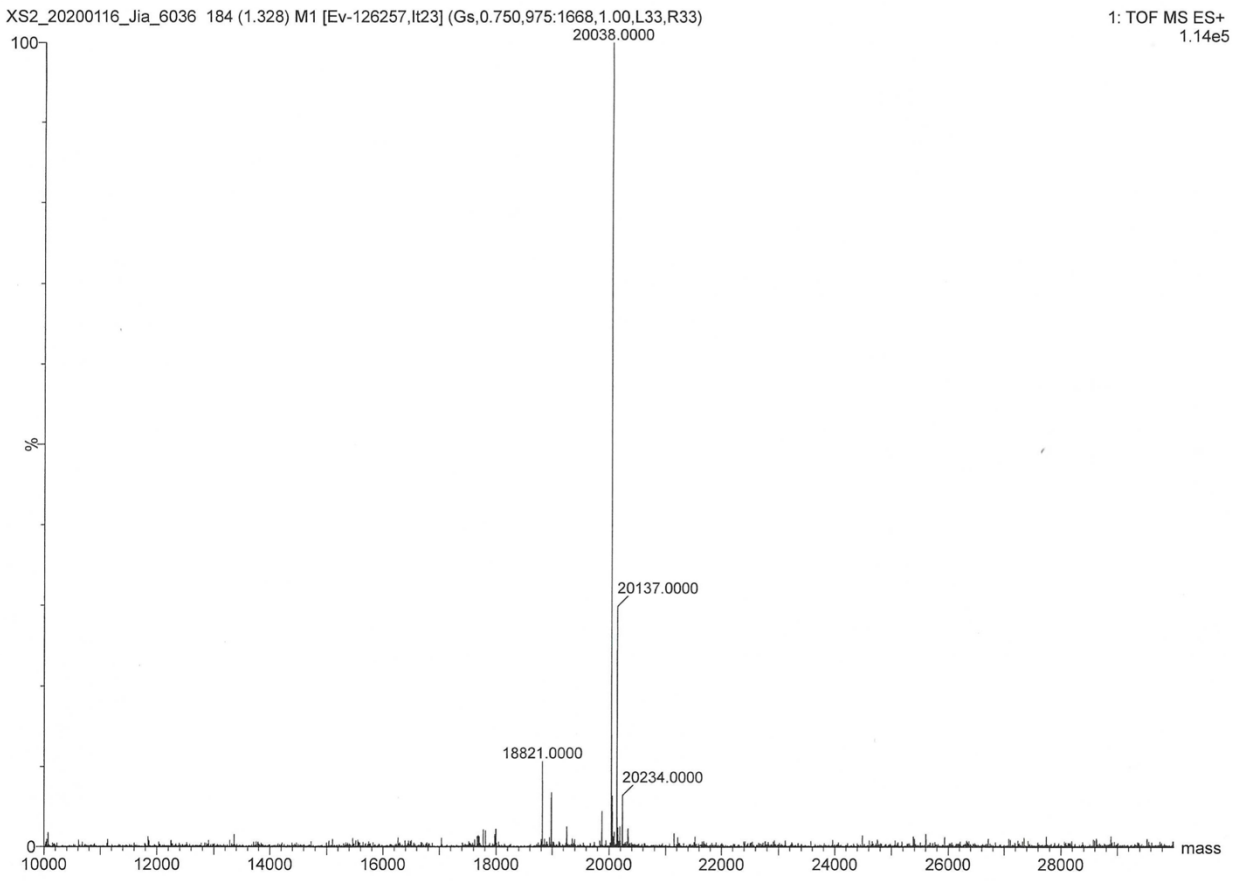


Figure 4.42 ESI-MS of CD44 (*O*-Xyl-Gal).

REFERENCES

REFERENCES

1. Bernfield, M.; Götte, M.; Park, P. W.; Reizes, O.; Fitzgerald, M. L.; Lincecum, J.; Zako, M., Functions of Cell Surface Heparan Sulfate Proteoglycans. *Annu. Rev. Biochem.* **1999,** *68*, 729-777.

2. Couchman, J. R., Syndecans: Proteoglycan Regulators of Cell-Surface Microdomains. *Nat. Rev. Mol. Cell Biol.* **2003,** *4*, 926-937.

3. Schaefer, L.; Schaefer, R. M., Proteoglycans: From Structural Compounds to Signaling Molecules. *Cell Tissue Res.* **2010,** *339*, 237-246.

4. Esko, J. D.; Zhang, L., Influence of Core Protein Sequence on Glycosaminoglycan Assembly. *Curr. Opin. Struc. Biol.* **1996,** *6*, 663-670.

5. Yang, W.; Yoshida, K.; Yang, B.; Huang, X., Obstacles and Solutions for Chemical Synthesis of Syndecan-3 Glycopeptides with Two Heparan Sulfate Chains. *Carbohydr. Res.* **2016,** *435*, 180-194.

6. Yang, W.; Eken, Y.; Zhang, J.; Cole, L. E.; Ramadan, S.; Xu, Y.; Zhang, Z.; Liu, J.; Wilson, A.; 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*, 6393-6404.

7. Li, T.; Yang, W.; Ramadan, S.; Huang, X., Synthesis of *O*-Sulfated Human Syndecan-1-like Glyco-Polypeptides by Incorporating Peptide Ligation and *O*-Sulfated Glycopeptide Cassette Strategies. *Org. Lett.* **2020,** *22*, 6429-6433.

8. Brinkmann, T.; Weilke, C.; Kleesiek, K., Recognition of Acceptor Proteins by UDP-D-Xylose Proteoglycan Core Protein b-D-Xylosyltransferase. *J. Biol. Chem.* **1997,** *272*, 11171-11175.

9. Wilson, I. B. H., Functional Characterization of Drosophila Melanogaster Peptide *O*-Xylosyltransferase, the Key Enzyme for Proteoglycan Chain Initiation and Member of the Core *N*-Acetylglucosaminyltransferase Family. *J. Biol. Chem.* **2002,** *277*, 21207-21212.

10. Kearns, A. E.; Campbell, S. C.; Westley, J.; Schwartz, N. B., Initiation of Chondroitin Sulfate Biosynthesis: A Kinetic Analysis of UDP-D-Xylose:Core Protein B-D-Xylosyltransferase. *Biochemistry* **1991,** *30*, 7477-7483.

11. Weilke, C.; Brinkmann, T.; Kleesiek, K., Determination of Xylosyltransferase Activity in Serum with Recombinant Human Bikunin as Acceptor. *Clin. Chem.* **1997,** *43*, 45-51.

12. Briggs, D. C.; Hohenester, E., Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1. *Structure* **2018,** *26*, 801-809.

13. Almeida, R.; Levery, S. B.; Mandel, U.; Kresse, H.; Schwientek, T.; Bennett, E. P.; Clause, H., Cloning and Expression of a Proteoglycan UDP-Galactose β-Xylose β1,4-Galactosyltransferase I. *J. Biol. Chem.* **1999,** *274*, 26165-26171.

14. Daligault, F.; Rahuel-Clermont, S.; Gulberti, S.; Cung, M. T.; Branlant, G.; Netter, P.; Magdalou, J.; Lattard, V., Thermodynamic Insights into the Structural Basis Governing the Donor Substrate Recognition by Human beta1,4-Galactosyltransferase 7. *Biochem. J.* **2009,** *418*, 605-614.

15. Talhaoui, I.; Bui, C.; Oriol, R.; Mulliert, G.; Gulberti, S.; Netter, P.; Coughtrie, M. W.; Ouzzine, M.; Fournel-Gigleux, S., Identification of Key Functional Residues in the Active Site of Human beta1,4-Galactosyltransferase 7: a Major Enzyme in the Glycosaminoglycan Synthesis Pathway. *J. Biol. Chem.* **2010,** *285*, 37342-37358.

16. Pfeil, U.; Wenzel, K., Purification and Some Properties of UDP-Xylosyltransferase of Rat Ear Cartilage. *Glycobiology* **2000,** *10*, 803-807.

17. Wilson, I. B. H., The Never-Ending Story of Peptide *O*-Xylosyltransferase. *Cell. Mol. Life Sci.* **2004,** *61*, 794-809.

18. Bock, K.; Pedersen, C., A Study of ^{13}C-H Coupling Constants in Hexopyranoses. *J. Chem. Soc., Perkin Trans. 2* **1974**, 293-297.

19. Gao, J.; Lin, P.-H.; Nick, S. T.; Huang, J.; Tykesson, E.; Ellervik, U.; Li, J.; Huang, X., Chemoenzymatic Synthesis of Glycopeptides Bearing Galactose-xylose Disaccharide from the Proteoglycan Linkage Region. *Org. Lett.* **2021,** *23*, 1738–1741.

20. Grebner, E. E.; Hall, C. W.; Neufeld, E. F., Glycosylation of Serine Residues by a Uridine Diphosphate-Xylose Protein Xylosyltransferase from Mouse Mastocytoma. *Arch. Biochem. Biophys.* **1966,** *116*, 391-398.

21. Robinson, H. C.; Telser, A.; Dorfman, A., Studies on Biosynthesis of the Linkage Region of Chondroitin Sulfate-Protein Complex. *Proc. Natl. Acad. Sci. U. S. A.* **1966,** *56*, 1859-1866.

22. Stoolmiller, A. C.; Horwitz, A. L.; Dorfman, A., Biosynthesis of the Chondroitin Sulfate Proteoglycan. *J. Biol. Chem.* **1972,** *247*, 3525-3532.

23. Campbell, P.; Jacobsson, I.; Benzing-Purdie, L.; Roden, L.; Fessler, J. H., Silk — A New Substrate for UDP-D-Xylose:Proteoglycan Core Protein b-D-Xylosyltransferase. *Anal. Biochem.* **1984,** *137*, 505-516.

24. Liu, L.-K.; Finzel, B. C., Fragment-Based Identification of an Inducible Binding Site on Cell Surface Receptor CD44 for the Design of Protein–Carbohydrate Interaction Inhibitors. *J. Med. Chem.* **2014,** *57*, 2714-2725.

25. Wu, Z. L.; Ethen, C. M.; Prather, B.; Machacek, M.; Jiang, W., Universal Phosphatase-Coupled Glycosyltransferase Assay. *Glycobiology* **2011,** *21*, 727-733.

26. Kuhn, J.; Gotting, C.; Beahm, B. J.; Bertozzi, C. R.; Faust, I.; Kuzaj, P.; Knabbe, C.; Hendig, D., Xylosyltransferase II is the Predominant Isoenzyme which is Responsible for the Steady-State Level of Xylosyltransferase Activity in Human Serum. *Biochem. Biophys. Res. Commun.* **2015,** *459*, 469-474.

27. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25*, 1605-1612.

28. Tsutsui, Y.; Ramakrishnan, B.; Qasba, P. K., Crystal Structures of beta-1,4-Galactosyltransferase 7 Enzyme Reveal Conformational Changes and Substrate Binding. *J. Biol. Chem.* **2013,** *288*, 31963-31970.

29. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., Autodock4 And Autodocktools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009,** *30*, 2785-2791.