Automated Solid-Phase Based Methodologies for the Synthesis of Glycosylated Biomolecules

Inaugural dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

Submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

By
Sandra Milena Pinzón Martín

June 2022

This work was performed between January 2018 and December 2021 in the Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces under the supervision of Prof. Dr. Daniel Varón Silva.

1st reviewer: Prof. Dr. Daniel Varón Silva
2nd reviewer: Prof. Dr. Beate Koksch

Date of defense: 18.10.2022

Declaration

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me. This thesis has not been part of another examination.

Berlin, June 2022

Acknowledgments

I am deeply grateful to Prof. Dr. Daniel Varón Silva for his supervision, continuous guidance, constant support and encouragement throughout my Ph.D.

I also want to thank Prof. Dr. Seeberger for giving me the opportunity to perform this research at the Max Planck Institute of Colloids and Interfaces and his support during the last year of my doctoral studies.

I thank Prof. Dr. Koksch for kindly agreeing to review my thesis.

I want to express my gratitude to the current and former members of the GPI group Paula, Antonella and Hyun-Il for the scientific discussions, for making a nice working atmosphere and the motivational music in the lab. I also thank my student Yasmine Meier for her help. Special thanks goes to Antonella Rella for her support, for sharing her knowledge with me and in general her invaluable help. I also want to thank Paula Guerrero for her support and for our talks in Spanish that took me back to Colombia.

I express my gratitude to Dr. Chandradhish Ghosh and Dr. Eric Sletten for proofreading my thesis.

Special thanks are extended to Dorothee Böhme for her help and Eva Settels for her technical support, especially with the analytical devices.

I would like to thank the colleagues I met between 2018-2021 at the department of Biomolecular Systems and the members of vaccine group for adopting me after my group moved to Switzerland, for the scientific discussions and nice working atmosphere. Particularly, I thank the members of SMRD for their company during the last months of my Ph.D., for teaching me English and more notably for playing Latino music on Fridays.

I want to deeply thank all the friends I made at MPIKG: Alexandra, Jasmin, Jhih-Yi, Alonso, Chandradhish (CD), Baris, Eric, Grigori, Majd, Manuel, Marco, Mauro and Nabyl for their support inside and outside the lab, the delicious coffee breaks, lunches and dinners complying my food requirements. Particularly, I express my gratitude to Eric and Manuel for their helpful advices for the automation project. Special thanks goes to Alonso, Jhih-Yi, CD and Majd for their warm friendship and constant support throughout these years.

I want to thank my Colombian friends Alonso Pardo and Laura Agudelo for making me feel at home and for cooking delicious food for me. Also my friend Luisa Martín for our long video calls, her advices and constant support.

Finally and more importantly, quiero agradecer a mis padres Maria del Carmen y Darier y a mis hermanos Cindy y Jaime por su apoyo constante, por su amor incondicional y por todos y cada uno de los esfuerzos que han hecho para que yo sea una mejor persona. Extiendo mi gratitud a mis familiares por su apoyo y el apoyo que han brindado a mi familia mientras yo he estado lejos.

Publications

Pinzón Martín, S.; Seeberger, P. H.; Varón Silva, D., Mucins and Pathogenic Mucin-Like Molecules Are Immunomodulators During Infection and Targets for Diagnostics and Vaccines, *Front. Chem.* **2019**, *7:710,* DOI: 10.3389/fchem.2019.00710.

Scientific Conference and Symposia

"Semi-synthesis of glycosylated Fragments of Human CD59 Protein", 7th RIKEN-MAX Planck Symposium, Ringberg/Germany, September 2019.

"Semi-synthesis of a non-glycosylated variant of the CD59 glycoprotein through sequential Native Chemical Ligation", 8th Peptide Engineering Meeting, Berlin/Germany, November 2018.

Table of Contents

Abbreviations

Ac	Acetyl
Acm	Acetamidomethyl
ACN	Acetonitrile
AGA	Automated glycan assembly
Boc	*Tert*-butoxycarbonyl
BOP	(Benzotriazol-1-yloxy)tris(dimethylamino)phosphonium
Bn	Benzyl
Bz	Benzoyl
Cbz	Benzyloxycarbonyl
Cl-HOBt	6-Chloro-1-hydroxybenzotriazole
DIC	*N,N′*-Diisopropylcarbodiimide
DIPEA	*N,N*-diisopropyl-ethylamine
DMF	Dimethylformamide
DSL	Diselenide selenoester ligation
DTT	Dithiothreitol
EDC	1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide
EDEM	Degradation-enhancing α-mannosidase I-like
ENGases	Endo-β-*N*-acetylglucosaminidase
ER	Endoplasmic reticulum
ESI-MS	Electrospray ionization-mass spectrometry
Fmoc	9-Fluorenylmethoxycarbonyl
Fuc	Fucose
Gal	Galactose
GalN	Galactosamine
GalNAc	*N*-Acetylgalactosamine
Glc	Glucose
GlcN	Glucosamine
GlcNAc	*N*-Acetylglucosamine
Gdn HCl	Guanidinium hydrochloride
GPI	Glycosylphosphatidylinositol
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxidexafluorophosphate
HBTU	*N,N,N′,N′*-Tetramethyl-O-(1H-benzotriazol-1-yl)Uronium hexafluorophosphate
HPLC	High-performance liquid chromatography
Lev	Levulinoyl (4-Oxopentanoyl)
LG	Leaving group
MALDI-MS	Matrix-assisted laser desorption ionization-mass spectrometry

Man	Mannose
NBS	*N*-bromosuccinimide
NCL	Native chemical ligation
NIS	*N*-iodosuccinimide
NMR	Nuclear magnetic resonance
OST	Oligosaccharyltransferase
PEG	Polyetylene glycol
PG	Protective group
PS	Polystyrene
Rha	Rhamnose
RP-HPLC	Reverse-phase HPLC
rt	Room temperature
SAL	Sugar-assisted ligation
STL	Ser/Thr ligation
TFA	Trifluoroacetic acid
TfOH	Trifluoromethanesulfonic acid
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TMSOTf	Trimethylsilyl trifluoromethanesulfonate

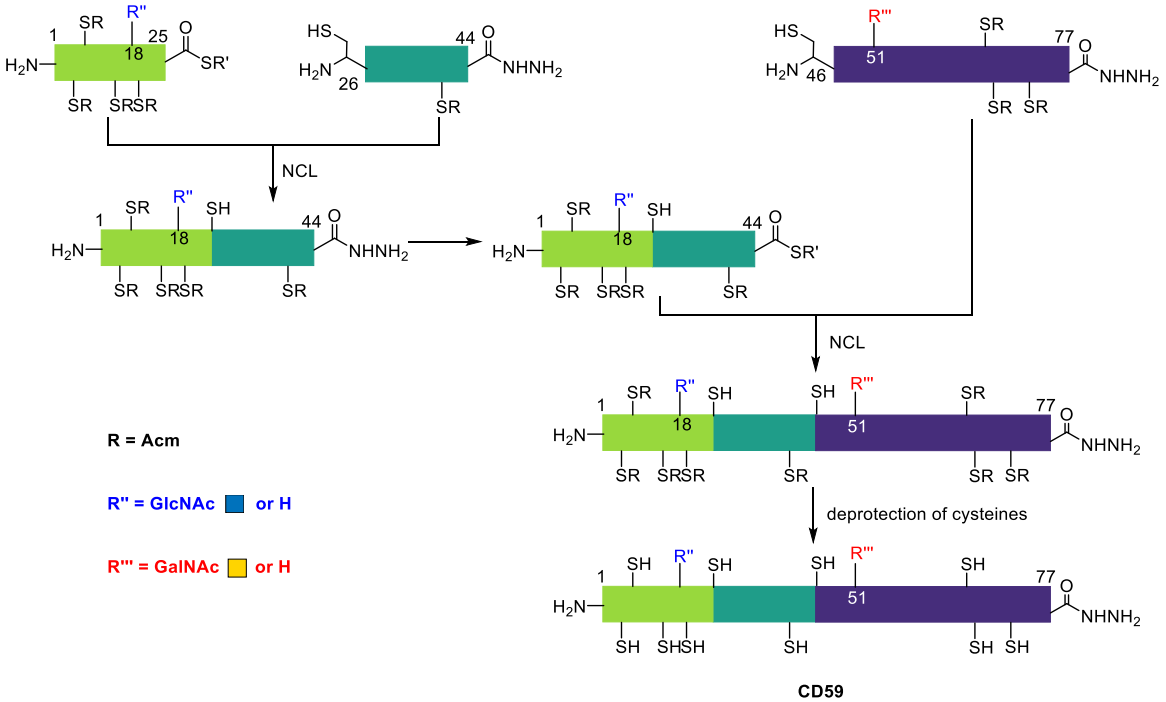
Abbreviations of the 20 canonical amino acids are consistent with the one- and three-letter code recommended by the IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) [Eur. J. Biochem. 1984, 138, 9–37]

Summary

Glycosylation is the most common post-translational modification of proteins in living organisms and participates in several processes such as cell adhesion, signal transduction and immune responses. Glycans present on glycoproteins are associated with health and disease conditions and can be used as markers for diagnosis or as therapeutical agents. However, the main limitation that slows down the research on glycosylated biomolecules is the lack of homogeneous isoforms of glycoproteins. Thus, the need to have homogeneous materials has prompted the development of methods for synthesizing glycosylated proteins.

This work is aimed to develop strategies and methodologies for synthesizing glycosylated molecules. Two projects were executed to synthesize glycosylated amino acids of *Trypanosoma cruzi* and variants of the human glycoprotein CD59. These syntheses relied on a solid-phase strategy and were achieved using automated systems eventually obtaining glycosylated amino acids, peptides, and glycopeptides.

The first project (Chapter 2) focuses on establishing a strategy to synthesize homogeneous variants of human protein CD59. In this chapter three different strategies were investigated to find the suitable synthetic methodology for obtaining glycosylated and non-glycosylated protein variants of CD59. The optimized strategy involved sequential N- to C-terminal assembly of three CD59 fragments 1-25, 26-44, and 45-78 by native chemical ligation of peptides and/or glycopeptides (Scheme I).

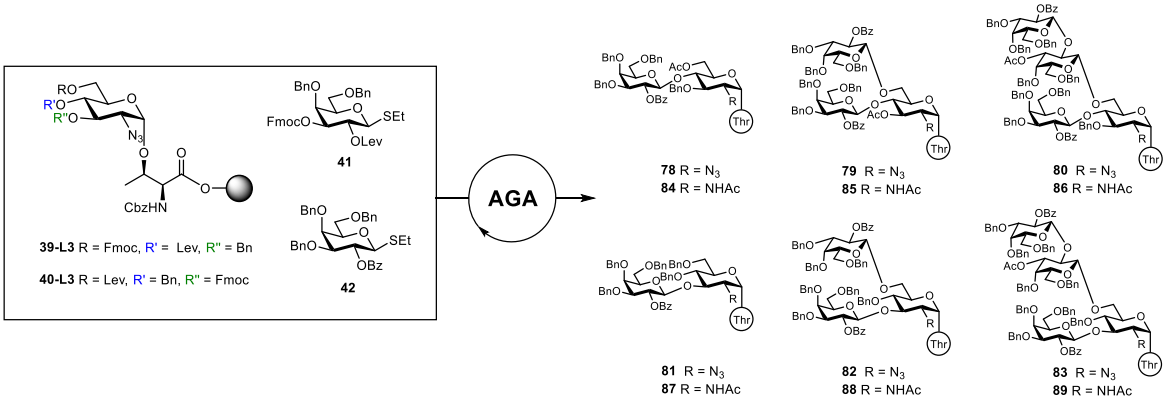


Scheme I - Synthesis of human CD59 protein by native chemical ligation

All fragments were synthesized by solid phase peptide synthesis as peptide hydrazides, using *S*-acetamidomethyl (Acm) protected internal cysteines. Furthermore, two glycopeptides were synthesized by a convergent method that employed glycosylated amino acids to deliver the

fragment 1-25 containing an *N*-acetylglucosamine in position 18, and fragment 45-78 bearing *N*-acetylgalactosamine in position 51. Finally, the conversion of the peptide or glycopeptide hydrazides to thioesters and its further ligation allowed the synthesis of two variants of human CD59 protein.

The second project focuses on establishing a strategy for the solid-phase synthesis of glycosylated molecules using automated glycan assembly, particularly O-glycans derived from *Trypanosoma cruzi* parasite (Y strain) and the results are presented in chapter 3. Four building blocks were designed, prepared, and used to synthesize glycosyl amino acids containing glycans of different sizes (di-, tri-, and tetrasaccharides) under solid-phase conditions (Scheme II). Optimizing different deprotection conditions or post automation steps, a standard methodology to obtain these molecules was developed. This methodology can be applied to synthesize a wide range of molecules from mucin-like molecules or human mucins.



Scheme II - Automated glycan assembly of glycosyl amino acid of T. cruzi Y strain mucins

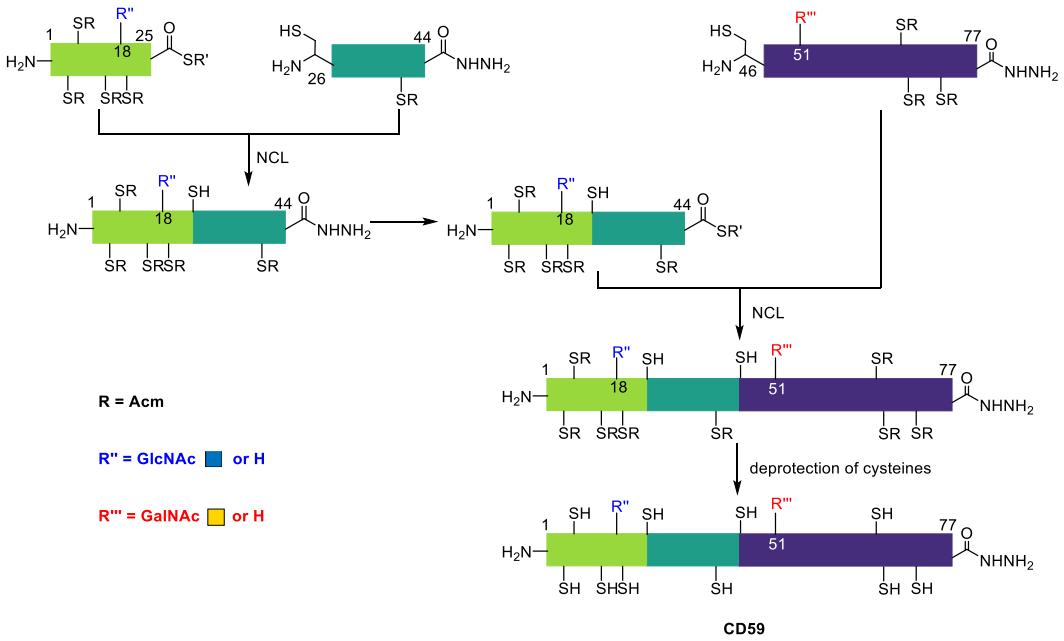
In conclusion, two methodologies were developed employing automated solid-phase based methods for the synthesis of glycosyl amino acids and variants of human CD59 protein. The synthetic methodologies developed in this thesis serve as an efficient strategy to deliver molecules that can act as models for structure-activity relationships studies using pure glycans, glycosyl amino acids and glycoproteins. Moreover, the methodology to synthesize glycosyl amino acids derived from mucins and mucin-like molecules will allow easy access to compounds that can be screened as epitopes for diagnostics, treatments for cancer and infectious diseases, and used as building blocks for glycopeptides or glycoprotein synthesis.

Zusammenfassung

Die Glykosylierung ist die häufigste posttranslationale Modifikation von Proteinen in lebenden Organismen und ist an mehreren Prozessen wie Zelladhäsion, Signaltransduktion und Immunreaktionen beteiligt. Die auf Glykoproteinen vorhandenen Glykane werden mit Gesundheits- und Krankheitszuständen in Verbindung gebracht und können als diagnostische Marker oder als therapeutische Wirkstoffe verwendet werden. Die wesentliche Einschränkung, die die Erforschung glykosylierter Biomoleküle bremst, ist jedoch das Fehlen homogener Isoformen von Glykoproteinen. Der Bedarf an homogenen Materialien hat daher zur Entwicklung von Methoden zur Synthese glykosylierter Proteine geführt.

Ziel dieser Arbeit ist die Entwicklung von Strategien und Methoden für die Synthese von glykosylierten Molekülen. Zwei Projekte wurden durchgeführt: Die Synthese glykosylierter Aminosäuren von Trypanosoma cruzi sowie die Synthese verschiedener Varianten des menschlichen Glykoproteins CD59. Deren Herstellung beruhte auf der Festphasen-Synthese und wurden mit automatisierten Systemen durchgeführt, um schließlich glykosylierte Aminosäuren, Peptide und Glykopeptide zu erhalten.

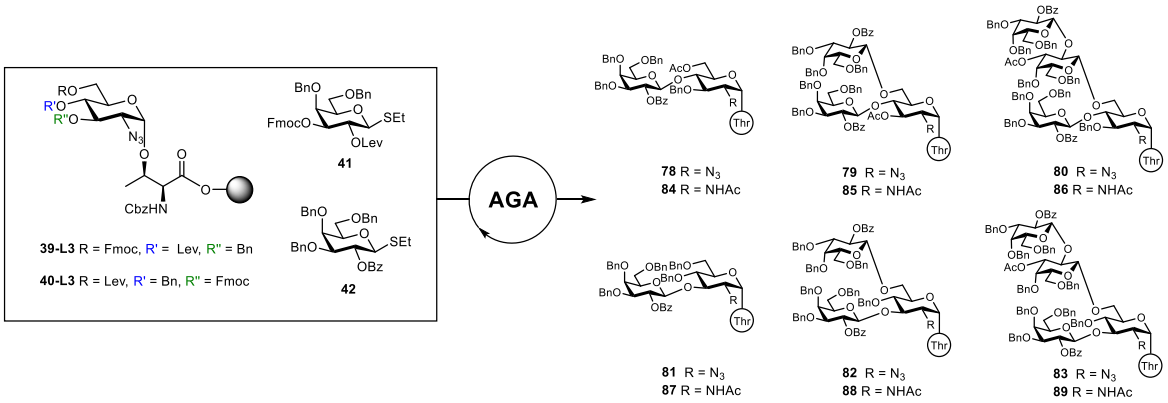
Das erste Projekt (Kapitel 2) konzentriert sich auf die Entwicklung einer Strategie zur Synthese homogener Varianten des menschlichen Proteins CD59. In diesem Kapitel wurden drei verschiedene Strategien untersucht, um die geeignete Synthesemethode zur Gewinnung von glykosylierten und nicht-glykosylierten Proteinvarianten von CD59 zu finden. Die optimierte Strategie umfasste die sequentielle N- nach C-terminale Assemblierung der drei CD59-Fragmente 1-25, 26-44 und 45-78 durch native chemische Ligation von Peptiden und/oder Glykopeptiden (Schema I).



Schema I - Synthese des menschlichen CD59-Proteins durch native chemische Ligation

Alle Fragmente wurden durch Festphasen-Peptidsynthese als Peptidhydrazide synthetisiert, wobei S-Acetamidomethyl (Acm) geschützte interne Cysteine verwendet wurden. Darüber hinaus wurden zwei Glykopeptide durch eine konvergente Methode synthetisiert, bei der glykosylierte Aminosäuren verwendet wurden, um das Fragment 1-25 mit einem N-Acetylglucosamin in Position 18 und das Fragment 45-78 mit einem N-Acetylgalactosamin in Position 51 zu erhalten. Schließlich ermöglichte die Umwandlung der Peptid- oder Glykopeptidhydrazide in Thioester und ihre weitere Ligation die Synthese von zwei Varianten des menschlichen CD59-Proteins.

Das zweite Projekt befasst sich mit der Entwicklung einer Strategie für die Festphasen-Synthese glykosylierter Moleküle unter Verwendung der automatischen Glykan-Assemblierung, insbesondere von O-Glykanen aus dem Parasiten Trypanosoma cruzi (Y-Stamm). Diese Ergebnisse sind in Kapitel 3 dargestellt. Vier Bausteine wurden entworfen, hergestellt und zur Synthese von Glykosylaminosäuren mit Glykanen unterschiedlicher Größe (Di-, Tri- und Tetrasaccharide) unter Festphasenbedingungen verwendet (Schema II). Durch die Optimierung verschiedener Entschützungsbedingungen oder Post-Automatisierungsschritte wurde eine Standardmethodik zur Gewinnung dieser Moleküle entwickelt. Diese Methode kann zur Synthese einer breiten Palette von Molekülen angewandt werden, die von Mucin-ähnlichen Molekülen oder von menschlichen Mucinen abstammen.



Schema II - Automatisierter Glykanaufbau von Glykosylaminosäuren aus T. cruzi

Zusammenfassend wurden zwei Methoden entwickelt, bei denen automatisierte Festphasenverfahren für die Synthese von Glykosylaminosäuren und Varianten des menschlichen CD59-Proteins eingesetzt werden. Die in dieser Arbeit entwickelten Synthesemethoden dienen als effiziente Strategie zur Bereitstellung von Molekülen, die als Modelle für Studien zur Struktur-Aktivitäts-Beziehung mit reinen Glykanen, Glykosylaminosäuren und Glykoproteinen dienen können. Darüber hinaus ermöglicht die Methodik zur Synthese von Glykosylaminosäuren, die von Mucinen und Mucin-ähnlichen Molekülen abgeleitet sind, einen einfachen Zugang zu Verbindungen, die als Epitope für die Diagnostik, die Behandlung von Krebs und Infektionskrankheiten untersucht und als Bausteine für die Synthese von Glykopeptiden oder Glykoproteinen verwendet werden können.

1. Introduction

This chapter includes sections partly modified from the manuscript “Mucins and Pathogenic Mucin-Like Molecules are Immunomodulators During Infection and Targets for Diagnostics and Vaccines, *Front. Chem.* **2019**, *7:710”* .[1]

Proteins are the most ubiquitous biomolecules present in nature and are responsible for several functional and structural processes in living things. The function of a protein is determined by its unique folded structure and the specific sequence of proteinogenic amino acids[2-4]. Once proteins are assembled, a wide range of biochemical modifications can be introduced to the reactive groups at the N- and C-terminus or the side chains of amino acids of the protein. These modifications, called post-translational modifications (PTMs), play essential roles in the structure and function of several proteins[5-6]. The post-translational modifications include glycosylation, lipidation, disulfide bond formation, phosphorylation, methylation, acetylation, and ubiquitination. In addition, PTMs regulate critical cellular events such as gene expression, signal transduction, protein-protein, cell-cell interaction, and communication between intracellular and extracellular environment[7-8].

Protein glycosylation is the most common and complex post-translational modification of proteins[9-10]. It is present in all domains of life[11-12] and plays a significant role in events such as signal transduction[8], host-pathogen interaction[13-16], immune responses[17-18], cell-cell signaling[19-20], cell differentiation[21], cancer progression[22-23], among others. Amino acids such as asparagine (Asn), serine (Ser), threonine (Thr), tryptophan (Trp), and tyrosine (Tyr) are found to be linked to carbohydrates which mainly include *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), glucose (Glc), galactose (Gal) and mannose (Man)[10]. Depending on the position and type of carbohydrate, protein glycosylation can be the covalent attachment of N-linked glycans, O-linked glycans, C-linked mannose, phosphorylated glycans, glycosaminoglycans and glycosylphosphatidylinositol (GPI anchor)[24-25]. Although several types of linkages and moieties are linked to proteins, the most common and most studied types of glycosylation are N- and O-glycosylation. The main aspects of these types will be discussed further later.

The following sections will introduce the function of glycosylated proteins, structural characteristics, and biosynthesis. Then, the different methods for synthesizing glycoproteins will be discussed through subsections covering both the protein and the glycan parts. Subsequently, two prominent examples of glycoproteins will be discussed and outline the aim of the thesis introduced.

1.1. Glycoproteins

Glycosylation of a protein occurs when a glycan is linked to the amino acid side chain of a protein. This process mainly occurs in the endoplasmic reticulum (ER) and Golgi apparatus[25]. Glycoproteins have essential functions in mammals and bacterial cells and the biology of viruses[17, 26-27]. It has been estimated that more than half of all human proteins are expected to be glycosylated[28-29]. The atoms involved in the linkage and the connections between them determine the type of glycoprotein. Depending on the amino acid side chain to which the glycan is linked, proteins can be categorized as N-glycoproteins and O-glycoproteins. Another special modification found in many eukaryotic cells is the anchoring of the protein C-terminus to the membrane by covalent linkage to glycosylphosphatidylinositol (GPI)[30], forming the so-called GPI-anchored glycoproteins (GPI-APs)[31-32].

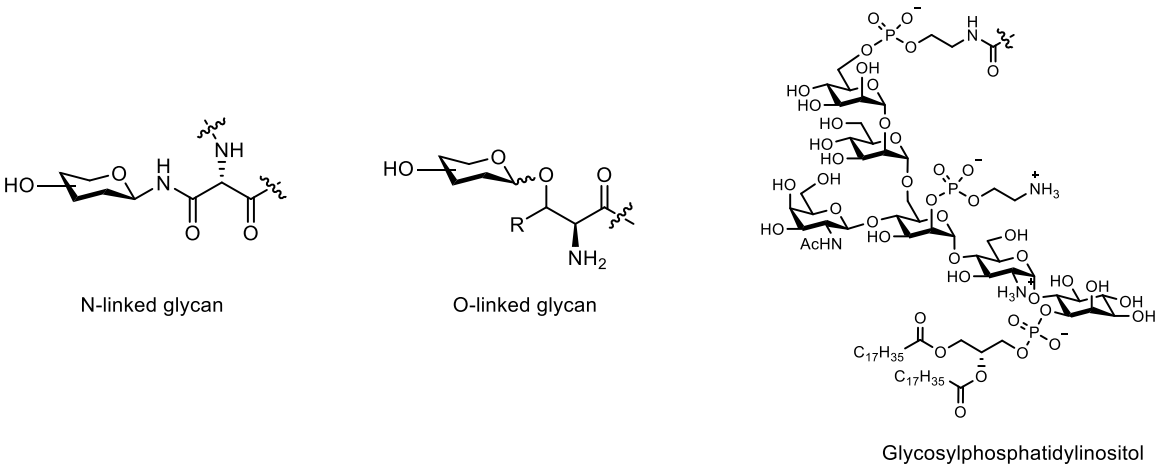


Figure 1 – General structure for N-linked, O-linked, and GPI linked glycans

Protein modification such as N- or O-glycosylation and glypiation (Figure 1) regulate the biophysical properties of proteins; therefore, they are involved in functions related to adhesion of molecules, protease inhibition, complement regulation, and prions[33-34]. Although these glycoproteins seem to play essential roles, and many efforts have been made to isolate and study them, there is a gap in the knowledge of these molecules compared to protein. The main reason is the inherent microheterogeneity observed when they are expressed and isolated[35]. Therefore, considerable efforts must be put into developing approaches that allow the synthesis of homogeneous glycoforms[36] and study these types of molecules.

1.1.1. N-glycosylation

N-glycosylation is the most representative form of glycosylation and has been observed in all three domains of life[37]. N-glycoproteins have been linked to several biological processes, such as the promotion of proper folding of newly synthesized polypeptides and ER-associated degradation pathways in quality control[38-40]. These functions are transcendental since defects in the biosynthesis of glycoconjugates are responsible for congenital disorders[41-42]. Essential proteins

in mammals are N-glycosylated and have clinical and therapeutic significance, for instance, immunoglobulin G antibodies (IgG) and recombinant monoclonal antibodies (mAbs)[43]. In N-glycoproteins, oligosaccharides are covalently attached through a β-linkage between *N*-acetyl-D-glucosamine and the asparagine side chain present in polypeptide chains containing the consensus sequence (Asn-Xaa-Ser/Thr), where Xaa is any amino acid except proline. The N-linked glycans can display a high diversity, and it varies among glycoproteins, cell types, tissues, and species[40].

N-glycans in eukaryotes share a core sequence and are classified as complex, oligomannose, and hybrid types (Figure 2a). They share the main core Manα1,6(Manα1,3)Manβ1,4Glc*N*Acβ1,4Glc*N*Acβ1Asn-X-Ser/Thr and depending on the type of glycan, this core can be further elongated (Figure 2b).

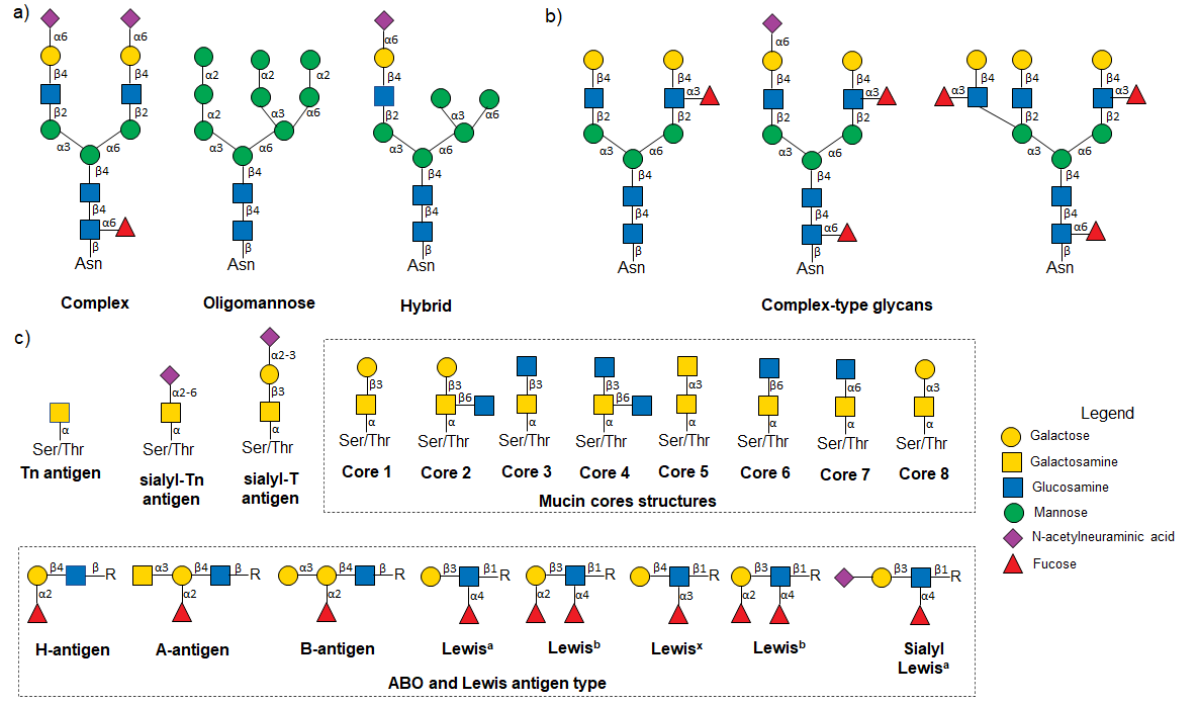


Figure 2 – Selected structures of N- and O-glycans present in human glycoproteins. a) general structure of N-linked glycan types; b) examples of complex-type glycans; c) representatives structures of O-glycans present in human mucins (cores) and peripheral sequences (antigens).

The biosynthesis of N-glycans occurs in two phases and two compartments of eukaryotic cells[44]. The first phase occurs at the ER membrane when an oligosaccharyltransferase (OST) catalyzes the "*en bloc*" transfer of the 14-mer dolichol-phosphate (Dol-P) bound oligosaccharide to the asparagine (Asn) acceptor in selected Asn-X-Ser/Thr site of the nascent protein. The second phase begins with processing reactions that trim the oligosaccharide chains catalyzed by glycosidases and glycosyltransferases in the lumen of the ER, followed by oxidative folding. When the glycoprotein is misfolded, it is recognized by ER degradation-enhancing α-mannosidase I-like (EDEM) proteins and targeted for ER degradation. The further trimming and elongation of the N-glycan continue in the Golgi apparatus (*cis*-Golgi, *medial*-Golgi, *trans*-Golgi). This so-called

terminal glycosylation continues until the N-glycoprotein depending on the species-cell type and or protein has been assembled[25, 37, 40, 44].

1.1.2. O-glycosylation

O-glycosylation is the second most common type of glycosylation known to date[45], and up to 80% of all proteins in the secretory pathway contain this type of glycosylation[46]. O-glycoproteins are involved in various biological processes such as determining protein conformation[47], cellular communication[48], and progression of cancer[49-50]. Structurally, O-glycosylation occurs on amino acids that bear hydroxyl groups on their side chains. For example, in humans, the most common glycans bound to serine and threonine are *N*-acetylglucosamine (through a β-linkage, O-GlcNAc glycosylation), *N*-acetylgalactosamine (through α-linkage, known as mucin-type *O*-glycans)[25]. This class of O-glycans comprises six major basic core structures: terminal GalNAc (Tn), sialyl-Tn antigens, and mucins cores 1-4 (Figure 2c).

The biosynthesis of O-glycoproteins occurs more ambiguously due to the absence of a common glycan core as in N-glycoproteins biosynthesis[51]. O-linked glycosylation occurs in the Golgi apparatus after protein folding and starts with the addition of *N*-acetylgalactosamine on the serine/threonine residue of the protein backbone, catalyzed by the enzyme y *N*-acetylgalactosaminyltransferase (GALNT) [52]. The core structure Ser/Thr-α-GalNAc, called Tn antigen, is further extended by sequential addition of monosaccharides in a stepwise manner catalyzed by respective glycosyltransferases[28, 45]. This extension can be done with either α-(1-3)- or β-(1-3)-linked galactose or by β-(1-6)-, β-(1-3)-, or α-(1-6)-linked *N*-acetylglucosamine; and a high variable peripheral part containing fucose and *N*-acetyl neuraminic acid terminal units[53]. The variability in the monosaccharide that can be linked produces high heterogeneity in linear and branched glycan structures.

Particularities in the biosynthesis of N- and O-glycoproteins make evident the heterogeneity of expressed glycoproteins and their multiple roles in biological processes. They can contain different glycans, multiple glycosylation sites, and slight modifications, as shown in figure 2. Efforts to deepen knowledge of their structure-activity relationship and further applications require the availability of homogeneous isoforms of glycoproteins and glycoconjugates. Therefore the development of approaches to obtain these molecules are of high significance.

1.2. Synthesis of glycoproteins

Glycoproteins play essential roles in nature, and understanding their structure and functions is crucial for developing protein-based therapeutics[54]. Several approaches for synthesizing glycoproteins, such as recombinant expression, enzymatic, chemical, and chemoenzymatic methods, have been established. This section contains information about glycoproteins synthesis using these representative methods. In addition, advancements in solid-phase peptide synthesis to obtain proteins will also be discussed. Further, approaches to synthesize the carbohydrate

components of glycoproteins, including oligosaccharides (specifically O-linked oligosaccharides), will be covered in section 1.2.5.

1.2.1. Recombinant expression

Isolation of homogeneous glycoproteins from living systems is difficult due to the structural diversity of oligosaccharides at a particular glycosylation site (microheterogeneity) and the requirement of multiple purification steps to isolate a homogeneous glycoform. However, recombinant expression allows the production of certain controlled glycoforms of glycoproteins by using genetically-engineered cell lines. An excellent example of this technology is Chinese hamster ovary cell lines (CHO), which express high levels of human α-2,3-sialyltransferase and produce highly sialylated glycoproteins[55].

Different mammalian expression systems have been optimized and used to express glycoproteins. For instance, CHO[56], baby hamster kidney (BHK21)[57], murine myeloma [58] and human cell lines[59-60]. Despite the differences in glycosylation patterns, yeast[61], plants[62], and insect cell[63] systems have also been studied. The biological significance of glycoproteins and the advances in their recombinant expression have contributed to the field of therapeutic glycoproteins. Nowadays, monoclonal antibodies for cancer treatment and clotting factors are available in the pharmaceutical market[64-65].

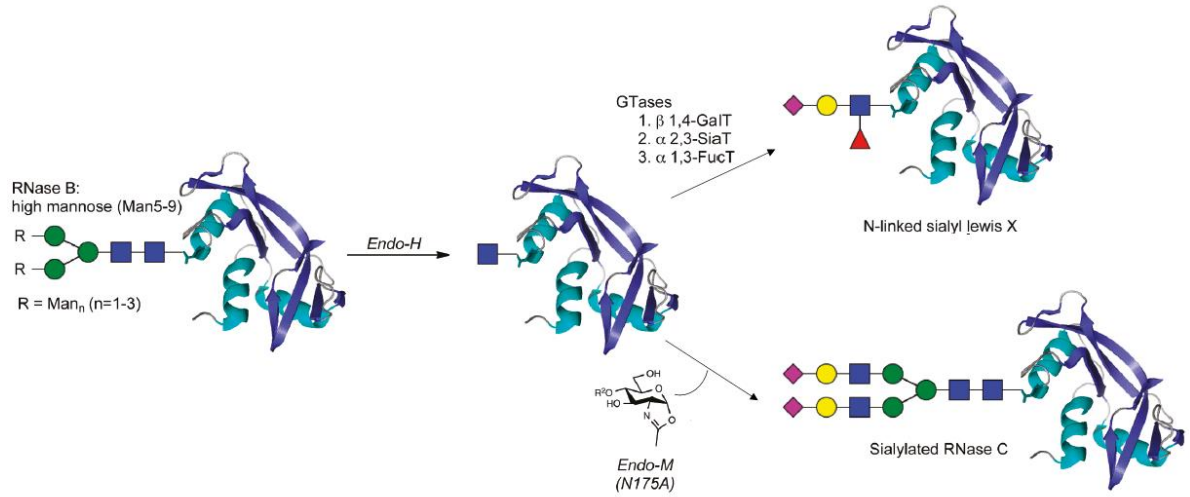
1.2.2. Enzymatic synthesis of glycoproteins

The enzymes employed in oligosaccharide synthesis, glycosyltransferases, and glycosidases, can be used to produce glycan-defined glycoproteins resembling biosynthesis. Glycosyltransferases catalyze the stereo- and regiospecific transfer of monosaccharides from a donor to an acceptor substrate. In addition, glycosidases catalyze the selective hydrolysis of glycosidic linkages either with retention or inversion of the configuration[66-67]. These enzymes have been isolated from natural sources or by heterologous overexpression in foreign hosts and have been proven to accept saccharides, glycolipids, and glycoproteins as substrates[68].

The discovery of a group of endoglycosynthases (glycosidases) derived from the endo-β-*N*-acetylglucosaminidases (ENGases) was important because these enzymes hydrolyze N-glycan between the two *N*-acetylglucosamine residues yielding the protein bearing an *N*-acetylglucosamine residue[69]. More interestingly, it was found that the mutant enzymes were able to transfer an intact N-glycan *en bloc* to a GlcNAc-peptide or protein acceptor to form a new natural N-glycopeptide or N-glycoprotein[70-71]. Other significant contribution in this field includes the oligosaccharyltransferase (PgIB), isolated from bacteria that can transfer N- as well as O-glycans[72], and the *N*-glucosyltransferase (NGT) isolated from *Actinobacillus pleuropneumoniae* that transfers a monosaccharide to the consensus N-glycosylation site (Asn-Xaa-Ser/Thr)[73].

Direct enzymatic glycosylation of proteins is a practical method for preparing homogeneous glycoproteins[66]. For example, the most common glycoform of Ribonuclease B (RNase), which contains high-mannose type oligosaccharides, was used to synthesize a sialyl Lewis X glycoform

of this protein. The high mannose glycan was cleaved and the variant having a single N-acetylglucosamine at Asn 34 was obtained employing endoglycosidase-H. The transfer of UDP-Gal mediated by β-1,4-galactosyltransferase (GALT) allowed the initial elongation to form Galβ1,4-GlcNAc. The disaccharide was further extended to a tetrasaccharide using different enzymes and resembling the stepwise biosynthesis of O-glycans to afford the controlled synthesis of the glycoform NeuAcα2,3-Galβ1,4(Fucβ-1,3)GlcNac-RNase[74]. The optimized mutant of Endo-M (Endo-M N175A) was used in the transglycosylation reaction, transferring a complex-type glycan to produce a sialylated RNase C derivative[75-76] (Scheme 1).



Scheme 1 - Enzymatic synthesis of homogeneous glycoform of Ribonuclease B; adapted from Schmaltz *et al* [77]

Despite the preference of enzymes for industrial-scale synthesis of glycoconjugates, the main limitation of this approach is the reduced number of enzymes available or the high cost of production, and the specificities of glycan substrates. Another problem arises from the heterogeneity of the product, as incomplete transformation in the stepwise enzymatic synthesis could produce mixtures, leading to difficulty in purification.

1.2.3. Chemoenzymatic synthesis

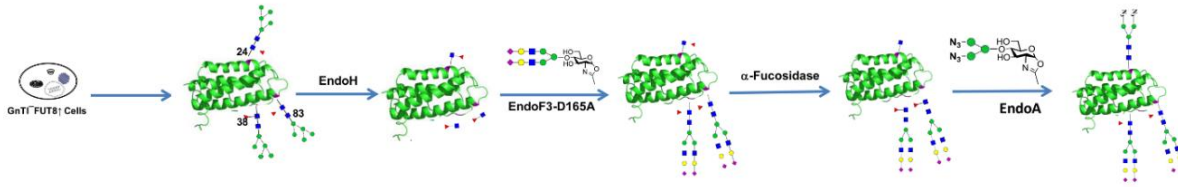
Chemoenzymatic synthesis emerges as a method to overcome some limitations that enzymatic and chemical synthesis (discussed later in sections 1.2.4 and 1.2.5) present. Combining the two methods has allowed the production of complex targets that are difficult to achieve by solely enzymatic or chemical synthesis. Two common examples are represented by the synthesis of saposin C and human erythropoietin (EPO). For the synthesis of saposin C, two molecules were chemically synthesized: a complex-type N-glycan octasaccharide oxazoline and a saposin C variant carrying GlcNAc. Subsequently, the glycan was transferred by transglycosylation reaction using the mutant glycosynthase Endo-M, obtaining the saposin C carrying the N-linked complex type nonasaccharide[78]. The synthesis of sialylated glycoforms of EPO was performed by sequential native chemical ligation of two synthetic peptides and three glycopeptides, the lately obtained by pseudoproline assisted Lansbury aspartylation. The sialylation was achieved by

enzymatic α-2,6-sialyltransferase from *Photobacterium damsela* (ST6). More interestingly, the biological recognition of the synthetic EPO was shown by forming 1:1 complexes with recombinant EPO receptor[79]. This method overcomes the complexity of synthesizing glycans containing sialic acid due to its known lability.

Attractive target, such as the antigen CD52 that bears N- and O-glycans, was synthesized by chemoenzymatic methods. The process involved the solid-phase synthesis of glycopeptides bearing the monosaccharides (Asn-linked GlcNAc and Thr-linked GalNAc), followed by Endo-M N175A transglycosylation with a complex N-glycan and as the last step the extension of the *O*-GalNAc sugar by T-synthase[80]. The synthesis of core-fucosylated triantenary complex type glycoforms of CD52 was also reported after discovering glycosynthase mutants derived from *Elizabethkingia meningoseptica* endoglycosidase F3 (Endo-F3)[81].

A similar chemoenzymatic strategy for the synthesis of O-glycopeptides was also recently presented. It included the chemical synthesis of the peptide core by SPPS and the enzymatic extension of the glycan epitopes using a commercially available peptide synthesizer[82]. This semi-automated methodology can serve as inspiration for designing strategies that allow total chemoenzymatic synthesis of glycoproteins.

Another potential strategy to produce homogeneous glycoproteins is enzymatic glycan remodeling. It requires a glycan trimming by glycosidases and the elongation to a defined structure using glycosyltransferases or mutant glycosidases. The remodeling of glycans can be done in intact isolated or expressed glycoproteins[66]. The site-selective chemoenzymatic glycosylation and remodeling of recombinant EPO show the versatility of this method. After expressing an unusual $Man_5GlcNAc_2Fuc$ glycoform in the engineered HEK293 cell line, the core fucosylated GlcNAc EPO was obtained by employing glycosidase EndoH. Next, this glycoform was used for enzymatic transglycosylation catalyzed by mutant EndoF3-D165A, which led to an unexpected site selectivity for only two locations. Finally, the remaining was defucosylated and elongated with azide-tagged glycan oxazoline employing Endo-A, leading to site-selective glycan remodeling of EPO (Scheme 2)[83].



Scheme 2 - Glycan remodeling of erythropoietin, adapted from Yang *et al* [83]

Another example is the *in vitro* enzymatic galactosylation of heterogeneous IgG glycoform. Recombinant GalT-1 and UDP-galactose allowed the synthesis of a homogeneous IgG glycoform bearing two galactosyl units with more than 98% of conversion on a kilogram scale[84]. Recently, a method for successive glycan remodeling in a solid-phase platform was reported. First, the IgG was immobilized onto protein A resins, and the glycosidases and glycosyltransferases were screened by solid-phase glycan remodeling (SPGR). This strategy allowed fast washing

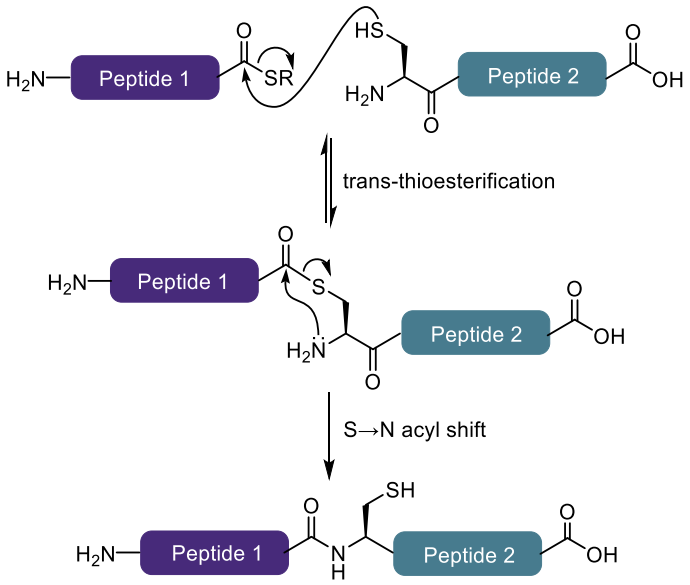
conditioning and product purification filtration, speeding up the multi-step reactions. Then, after trimming and elongating the glycans present in the antibody, the characterization of the IgG products was done by PNGase F treatment, followed by fluorophore labeling of the glycans with further LC-MS analysis[85]. This exciting strategy can expand the platforms for glycan remodeling in glycoproteins.

1.2.4. Chemical synthesis of glycoproteins (protein and peptides)

Chemical synthesis is considered a robust methodology to obtain homogeneous glycoproteins because it allows the modification of the covalent structure of the protein (glycans installation, labeling, among others) as desired[3]. The development of native chemical ligation for protein synthesis (NCL) by Kent in 1994 revolutionized and expedited protein and glycoprotein synthesis. This methodology has been employed and improved during the last three decades, allowing the synthesis of multiple and exciting targets such as acid chemokine monocyte chemotactic glycoprotein-3[86] and glycosylated human interferon γ[87], interleukin 2[88], and several variants of human erythropoietin (EPO)[79, 89].

1.2.4.1. Native Chemical Ligation

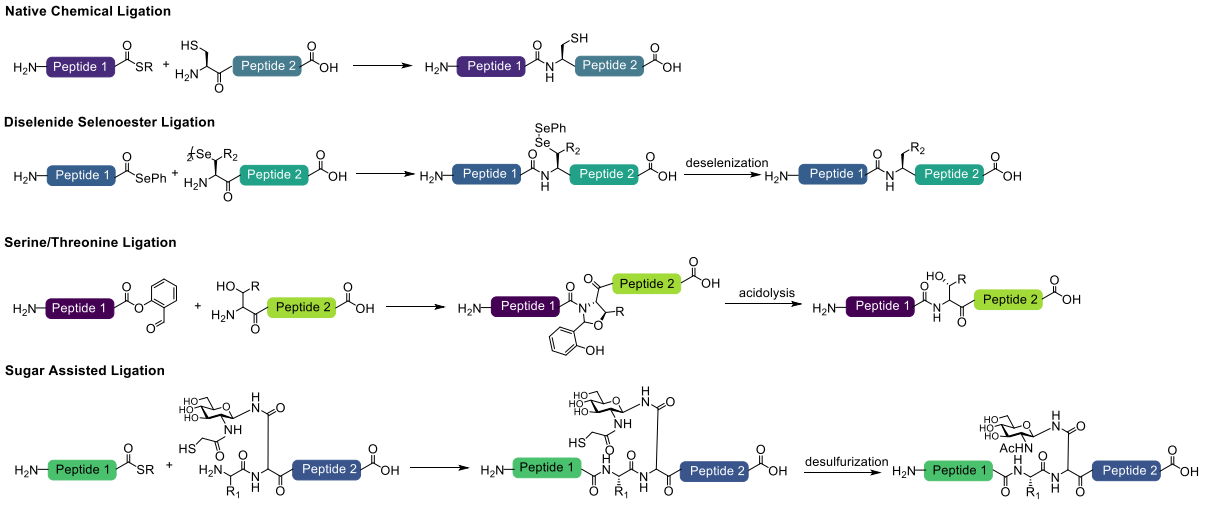
Kent and coworkers introduced the native chemical ligation concept as a chemoselective reaction of two unprotected peptide segments forming an initial thioester-linked species that later rearrange to yield a native peptide bond at the ligation site[90]. The peptides should fulfill specific conditions: one should contain an N-terminal cysteine and the other a C-terminal thioester. The reaction occurs in two steps: 1) reversible thiol/thioester exchange (trans-thioesterification) of the C-terminal thioester with the N-terminal cysteine and 2) spontaneous *S*→*N*-acyl shift that irreversibly forms the native peptide bond[4] (Scheme 3).



Scheme 3 - Mechanism of Native chemical ligation of two unprotected peptides

Sequential ligation strategies have also been developed to achieve one-pot native chemical ligation of proteins. This strategy required the orthogonal protection of the terminal cysteine and an efficient method for its removal, compatible with the NCL conditions. The main groups that have been used in the successful synthesis of proteins and glycoproteins are thiazolidine[91-92], *S*-acetamidomethyl (Acm)[93-94] and disulfide[95-96] among others.

Cysteine is the least abundant proteinogenic amino acid in natural proteins, and it was considered the main limitation of NCL. A significant contribution was the desulfurization of cysteine to yield alanine, which is more abundant in natural proteins. In this method, Ala is selectively replaced by Cys to have suitable disconnection and ligation sites. Therefore, NCL followed by desulfurization has allowed the synthesis of non-cysteine-containing proteins[97-98] and glycoproteins[79]. The extension of this methodology to other targets has been achieved by employing common amino acids such as thiol or selenol derivatives and mercapto and seleno amino acids. For example, in the diselenide selenoester ligation (DSL)[99] a peptide selenophenyl ester reacts with a diselenide peptide giving an amide bond to selenocysteine in a rapid and additive-free manner (Scheme 4). Following the ligation, the products can be chemoselectively deselenized to produce native peptides and proteins, as shown in the synthesis of glycosylated human interferon γ[87].



Scheme 4 - Selected strategies for the synthesis of glycoproteins

Other methods to overcome the limitation caused by the low abundance of cysteine are the Ser/Thr ligation (STL) and sugar-assisted ligation (SAL). STL consists of the chemoselective reaction between unprotected *O*-salicylaldehyde ester at the C-terminus and a peptide bearing a serine or threonine residue at the N-terminus (Scheme 4). This reaction yields an N,O-benzylidene acetal-linked intermediate that can be converted to the natural peptidic bond without further isolation[100]. Sugar-assisted ligation (Scheme 4) allows the synthesis of glycoproteins by employing a modified glycan bearing a sulfhydryl group at the acetamido moiety in C2, which acts as the nucleophile in the trans-thioesterification reaction with a peptide thioester[101]. Initially, this

method allowed the synthesis of cysteine-free β-O-linked and N-linked glycopeptides[102], subsequently, optimization of the protocol allowed the synthesis of α-O-linked glycoprotein Diptericin ε[94].

1.2.4.2. Solid Phase Peptide Synthesis

Peptides are the basic units required for glycoprotein synthesis. As shown in the previous section, peptides and glycopeptides are necessary to assemble glycoproteins employing ligation methods (NCL, DSL, STL, SAL), but also, due to their size and versatility, they are good candidates for peptide-based drugs[103-104]. Solid-phase peptides synthesis (SPPS) was introduced by Merrifield in 1963 and changed the former strategy of solution synthesis for an easier and more efficient methodology[105]. Automation of this process led to versatile and numerous syntheses of peptides using the less hazardous Fmoc-protocol[106].

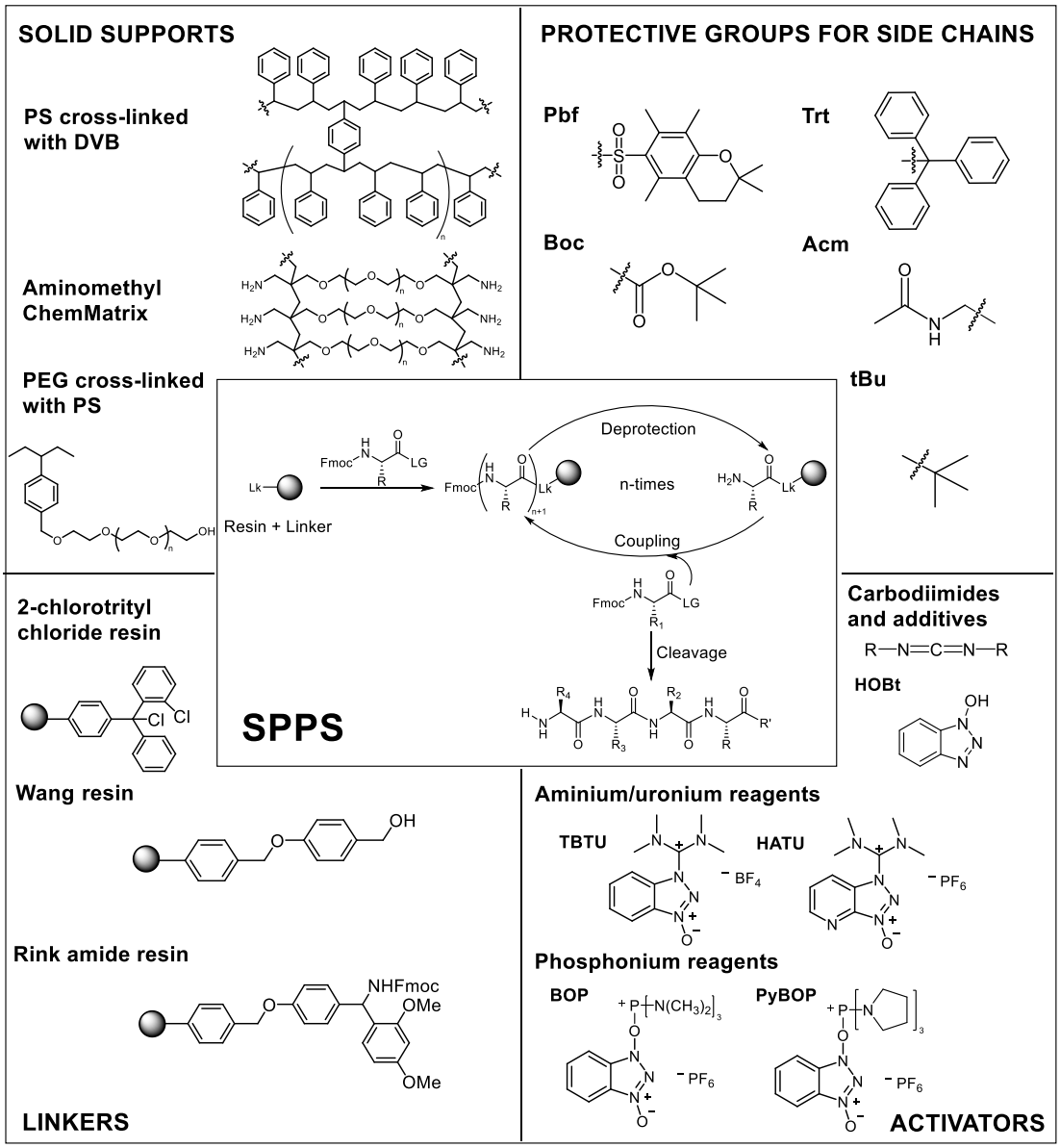


Figure 3 –Solid-phase synthesis scheme and selected examples of solid supports, protective groups, linkers, and activators for SPPS.

Chemical synthesis of peptides occurs from the C-terminal to the N-terminal residue, and the synthesis starts by attaching the first amino acid to the solid support bearing a linker. For this, the amino acid containing orthogonal protective groups in its α-amino and reactive side chain is activated for the coupling to the resin. The process continues with the deprotection of the α-amino group and the subsequent coupling of the following amino acid. Iterative cycles of deprotection and coupling until the desired peptide has been elongated delivers the on-resin fully protected product. After each deprotection and coupling cycle, washes remove the excess reagents and byproducts. Then, removal of the side chain protective groups and the cleavage is completed in the same step to yield the fully deprotected peptide (Figure 3). When a protected peptide is required, mild treatments allow the sole cleavage of the fully protected compound.

In the SPPS using Fmoc-protocol, the deprotection occurs under basic conditions and, therefore, the removal of side protective groups and cleavage in acidic conditions. The selection of solid support, the linker, the protective groups for side chains, and the activator plays a significant role in the outcome of SPPS. Figure 3 represents a summary of the SPPS and the chemical structures of some of these aspects. Their influence on the synthesis will be explained below.

The proper selection of solid support and the linker for SPPS is critical and relies on the resin reactivity, insolubility, and swelling properties[107]. The latter highly influences the diffusion and accessibility of the reagents[108]. Polystyrene (PS) cross-linked divinylbenzene (DVB) offers a better swelling factor than PS itself[109]. Polyethylene glycol-based (PEG) resins offer good swelling in organic and even aqueous solvents[110]. Polyamines and polyethylene glycols PEG-PS are suitable for synthesizing polar substrates[111-112] and fully PEG resins such as ChemMatrix® were reported to improve the synthesis of long and complex peptides[113].

The resins are functionalized with different groups that allow specific chemical reactions to attach the amino acid to the solid support. The most common group attached to PS-resins is chloromethyl to form the Merrifield resin; however, other functionalizations are available[108]. The characteristics of the linker define the used chemistry, the cleavage conditions, and the C-terminal functionalization of the peptide. Typical C-terminal functions are carboxylic acid (Wang resin)[108], amide (Rink amide[114] and MBHA resins), and sulfonamides[115]. When the functionalization is not available in the commercial resin, it can be modified as desired; for instance, a functionalization of Wang[116] and 2-chlorotritylchloride[117-118] resins allowed obtaining peptide hydrazides.

As mentioned above, the protective groups for the side chains must be acid labile to allow the cleavage and deprotection in the same step. Several groups have been studied for each amino acid, and the most common ones are depicted in Figure 3. Permanent protection of cysteine side chain for one-pot ligation purposes is achieved by using stable groups towards acidic and basic conditions such as acetamidomethyl (Acm) or compound forming disulfide bonds (StBu)[108, 119].

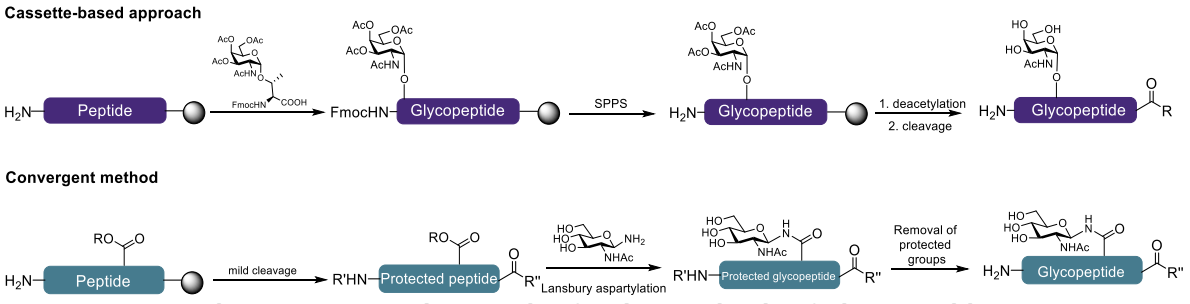
The coupling of the amino acids requires the activation of the α-carboxyl group. Different compounds have been studied for this purpose, and carbodiimide-based reagents have been

traditionally employed. However, they often cause partial racemization due to the high reactivity of the *O*-acylisourea and should be combined with additives such as 1-hydroxybenzotriazole (HOBt)[120] or ethyl-2-cyan-2(hydroxymino)acetate (Oxyma)[121] to form a less reactive and more stable ester, minimizing racemization[108, 122].

Another group of reagents derived from HOBt was developed to overcome the problems caused by carbodiimides during the coupling. The main types of reagents found in this group are aminium/uronium[123] and phosphonium salts[124]. Aminium salts bear a positive charge at a carbon atom, and they tend to be more reactive than phosphonium salts. Therefore, coupling reagents derived from HOAt, such as HATU are the most potent coupling reagents[125] among the existing activators. Although phosphonium salts react slower than uronium salts, their main advantage is that they do not react with the amino function to give the guanidine side product. Therefore, phosphonium salts do not cap the peptide chain. Multiple coupling reagents have been developed; nevertheless, there is not a coupling reagent that can be universally used in all coupling reactions. Particular attention must be given to the type of synthesis (manual, automated) and the functional groups on the molecules involved in the reaction[125].

It is essential to mention that recently a methodology in flow pushed the limits of SPPS, allowing the synthesis of homogeneous proteins. Therefore, the development of automated fast-flow peptide synthesis (AFPS) and its plans to incorporate peptides hydrazides and thioester for ligation reactions[126] could contribute to the synthesis of larger glycosylated proteins.

The synthesis of glycopeptides can be performed using the cassette-based approach and convergent method (Scheme 5). The cassete-based approach consists of the linear elongation of the peptide chain with a glycosylated amino acid[127]. The carbohydrate, usually a protected monosaccharide, is attached to an amino acid residue to generate N-linked and O-linked glycosyl amino acids. This pre-synthesized moiety is coupled to the growing peptide employing the desired activator, and the peptide chain can be further elongated by SPPS. The elongation of the glycan core can be done by chemical or enzymatic synthesis, as described above. This strategy has been widely used to synthesize N- and O-glycopeptides but is preferred for the latter ones.



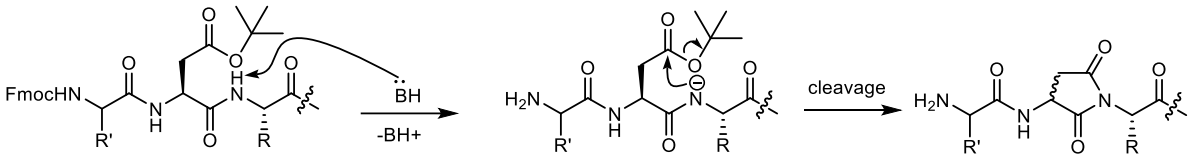
Scheme 5 - General strategies for the synthesis of glycopeptides

The convergent method consists of the direct conjugation of a carbohydrate moiety with the peptide and is mainly used to synthesize N-glycopeptides. The N-glycan having an anomeric

amine is linked to the aspartic acid side chain of a fully assembled peptide through an amide bond employing Lansbury aspartylation conditions[128]. The amino acid side chains must be protected to avoid side reactions. Specifically, the aspartic acid must bear an orthogonal protective group that can be selectively removed prior to activating coupling agents[129].

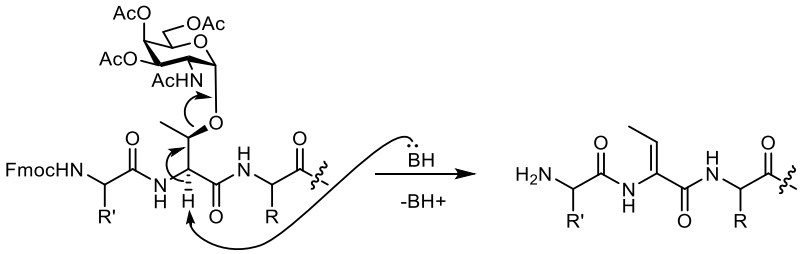
Side reactions during SPPS can occur at various stages and depend on the sequence and the steps involved[130]. Base-catalyzed side reactions represent a problem because the repetitive base treatments lead to the successive formation of the side product, significantly affecting the synthesis of the desired peptide[131].

Aspartimide formation is considered the most documented and studied side reaction in peptide chemistry and is highly sequence and conformation dependent[132-133]. It occurs during the elongation of Asp-containing peptides due to a nucleophilic attack from the amide nitrogen of the preceding residue to the β-carboxyl moiety of Asp (Scheme 6). Several aspects accelerate or delay the extent of the aspartimide formation, such as the type of base, protecting group in the Asp moiety, resin, and solvent used[130-134]. Although several strategies have been developed, these do not achieve complete prevention in demanding sequences; therefore, several strategies might be necessary to prevent aspartimide formation.



Scheme 6 - Base mediated aspartimide formation

Another base-catalyzed side reaction is the glycan β-elimination that occurs during the synthesis of O-glycopeptides. O-glycosylated threonine coupled to the growing peptide can be converted into the β-methyl dehydroalanine during the deprotection cycle eliminating the glycan from the peptide[135] by abstraction of the α-proton (Scheme 7).



Scheme 7 - Base mediated β-elimination

Although the SPPS process has been dramatically improved and optimized during the last 60 years, the SPPS of long peptides (50-60 amino acids) containing difficult sequences still faces challenges[108, 136] due to the possibility of generating byproducts from side reactions, truncation, deletion and aggregation of the growing peptide and glycopeptide[126, 137]. Therefore, the

combination of SPPS to obtain readily available peptides and glycopeptides and its further ligation remains the standard method for the chemical synthesis of glycoproteins.

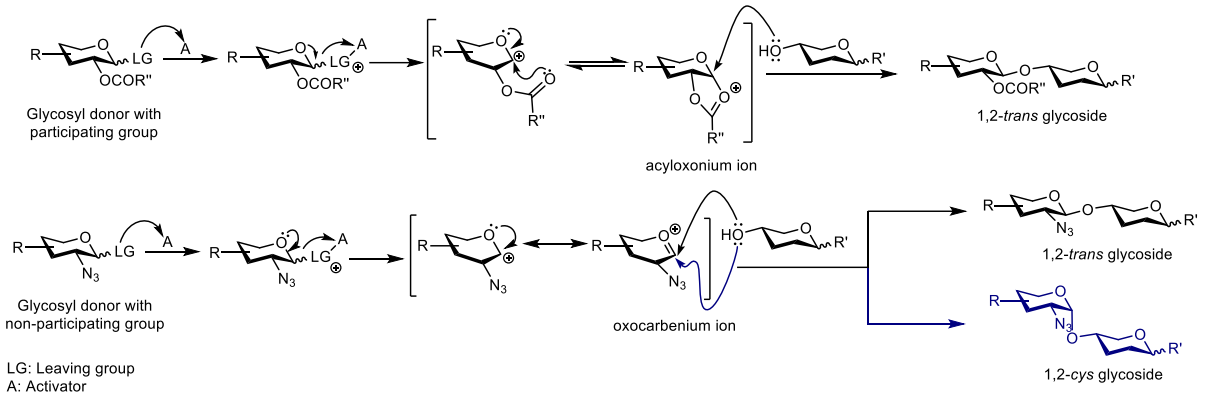
1.2.5. Methods for oligosaccharide synthesis

As described in 1.2, glycoproteins can be accessed through several methods, although the glycan or glycosylated amino acid cores are always required prior to their synthesis. Oligosaccharide synthesis has been achieved using multidisciplinary approaches, including synthetic and biochemical methods. Chemical synthesis can provide access to a broad set of glycan structures; solution and solid-phase synthesis have demonstrated the versatility of the glycosylation reaction to obtain homogeneous polysaccharides. For instance, the synthesis of 150-mer employing Automated Glycan Assembly (AGA)[138] or a 92-mer[139] by solution-phase synthesis. However, these methods face several challenges due to the nature of carbohydrates[140-141] and structural variations. Some possibilities include branching and the configuration of the linkages[142] that demand orthogonal protection and control of stereoselectivity.

1.2.5.1. Chemical glycosylation

The chemical synthesis of oligosaccharides is based on the glycosidic bond formation, and it remains challenging because the glycosyl donor and glycosyl acceptor should be connected with proper stereo and regioselectivity[143]. The reaction consists of the nucleophilic attack of a hydroxyl group from the glycosyl acceptor on the anomeric carbon of the glycosyl donor. Regioselectivity is controlled by orthogonal protection and the reactivity of the free hydroxyl groups present in the glycosyl acceptor[144], while stereoselectivity demands the control of several variables that will be later described.

Several efforts have been made to describe the mechanism of the glycosylation reaction; however, it is not clearly understood yet. It is generally described as an S_N1-like mechanism, but S_N2-like has also been reported in the literature[145]. In its most straightforward way, the mechanism involves the promoter assisted-departure of the leaving group, forming a cationic species (acyloxonium or oxocarbenium), which is then attacked by the nucleophilic hydroxyl group of the glycosyl acceptor[146] (Scheme 8). Nevertheless, the determination of the reaction mechanism is never straightforward and can be influenced by multiple factors[147].

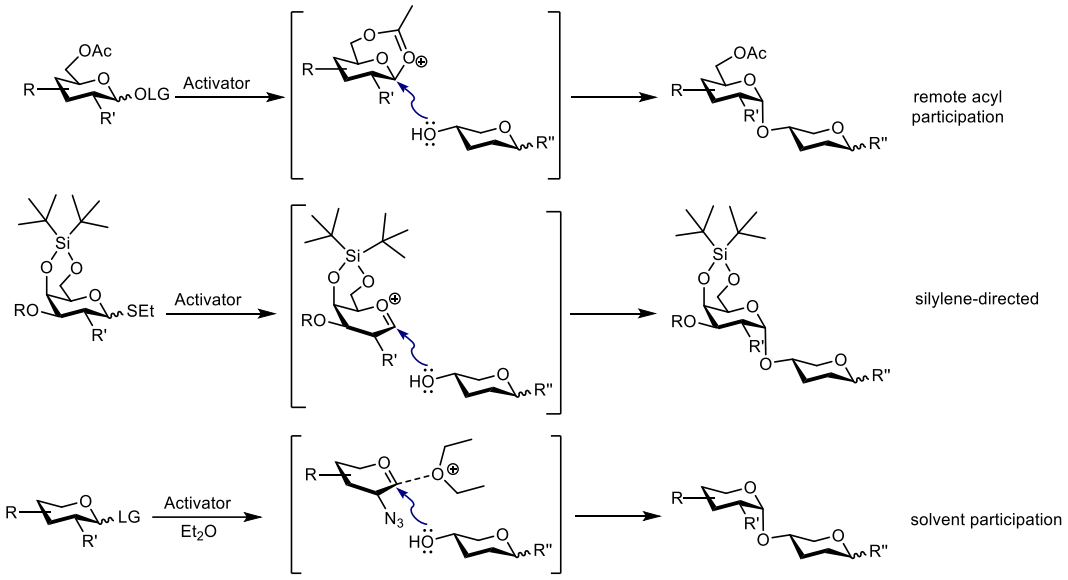


Scheme 8 – General mechanism of the glycosylation reaction with a participating and a non-participating group.

Numerous functional groups have been studied as leaving groups in carbohydrate chemistry[145]. Glycosyl donors are usually prepared as thioglycosides, imidates, phosphates, and halides. After the activation, the nucleophilic attack from the glycosyl acceptor to the cationic specie formed (glycosyl donor) can occur either from the top or the bottom, forming a 1,2-*trans* (mostly β-configuration) or 1,2-*cis* glycosides (mostly α-configuration). Stereoselectivity is usually controlled by different means to achieve these two configurations. For example, neighboring group participation at C-2 allows 1,2-*trans* glycosides with high stereoselectivity.

C2-acyl substituents form an acyloxonium ion complex with C-1, blocking the bottom face and favoring the attack of the nucleophile from the top face. On the other hand, the synthesis of 1,2-*cis* glycosides is complex and requires the control of several factors to favor the formation of 1,2-*cis* linkages. These factors include non-neighboring group participation at C-2, remote anchimeric assistance (remote group participation), conformation-restraining groups, and solvent and temperature effects.

While using a participating group is required to form 1,2-*trans* linkages, the presence of a non-participating group at C-2 is desirable to form 1,2-*cis* linkages. Moreover, remote participation of protective groups in other positions has been used to increase the selectivity for the 1,2-*cis* linkages. For instance, acyl (acetyl, benzoyl) groups at different positions C-6 and C-3 have been reported to increase selectivity for the 1,2-*cis* linkage in $GlcN_3$ sugars[148] and glucosyl donors[149]. Similar studies show the use of di-tert-butylsilylene to afford α-selective galactosylation[150-151]. These effects can be attributed to either steric or electronic shielding of the ring top face favoring the nucleophilic attack from the opposite side[146]. Temperature and solvent also influence stereoselectivity. Glycosylations performed at high temperatures will increase the formation of the more thermodynamically favored α-glycoside due to the anomeric effect[146-147, 152]. The use of ethereal solvents (diethyl ether, 1,4-dioxane) also favors α-glycosylation products[153], while β-glycosylation is predominantly formed when nitriles (acetonitrile) are used[154] (Scheme 9).



Scheme 9 – Examples of strategies to achieve 1,2-*cis* glycosidic linkages.

A plethora of studies to improve the outcome of the glycosylation reactions have been reported during the last century, and these significantly expedited the synthesis of oligosaccharides bearing complex functionalizations[155]. However, it remains challenging due to the high number of manipulations (protection, deprotection) and, therefore, multiple purification steps required to synthesize complex molecules found in nature. Nevertheless, after several months of work, a few milligrams with high purity and homogeneity can be obtained. Solution-phase synthesis is considered time-consuming and limited to a few specialized laboratories[156-157].

1.2.5.2. Automated Glycan Assembly

Solution phase synthesis of carbohydrates has mainly been explored, and with the advances of the last decades, almost any structure can be synthesized using this technique. However, it is considered a long and tedious process[156]. On the other hand, solid-phase synthesis has shown to be efficient for assembling other biopolymers like peptides and oligonucleotides, and this inspired the use of this methodology in the carbohydrates field.

Automated solid-phase synthesis of oligosaccharides was firstly described in 2001[158]. During the last 20 years, this technique has been improved and validated, showing high versatility. Nowadays is known as Automated Glycan Assembly (AGA)[159].

As with other solid-phase synthesis methods, the synthesis starts with functionalized solid support with an appropriate nucleophile required for the glycosylation reaction. Next, a glycosyl donor known as a building block is activated by a promoter or activator and undergoes a glycosylation reaction. The monosaccharide is fully protected with one or more orthogonal groups that can be selectively removed to continue with another glycosylation reaction. Iterative cycles of glycosylation and deprotection allow the on-resin elongation of the carbohydrate as desired. Critical steps such as acetylation of unreacted groups (capping) improve the outcome of the

synthesis[160] and, therefore were included in the iterative cycles for AGA. Once the automated synthesis finishes, the oligosaccharide bound to resin is further treated as required (methanolysis[158], sulfation[161], phosphorylation[162]), then the glycan is cleaved from the solid support releasing a partially protected oligosaccharide that can be purified or used as a crude for hydrogenolysis to yield the fully deprotected moiety (Figure 4).

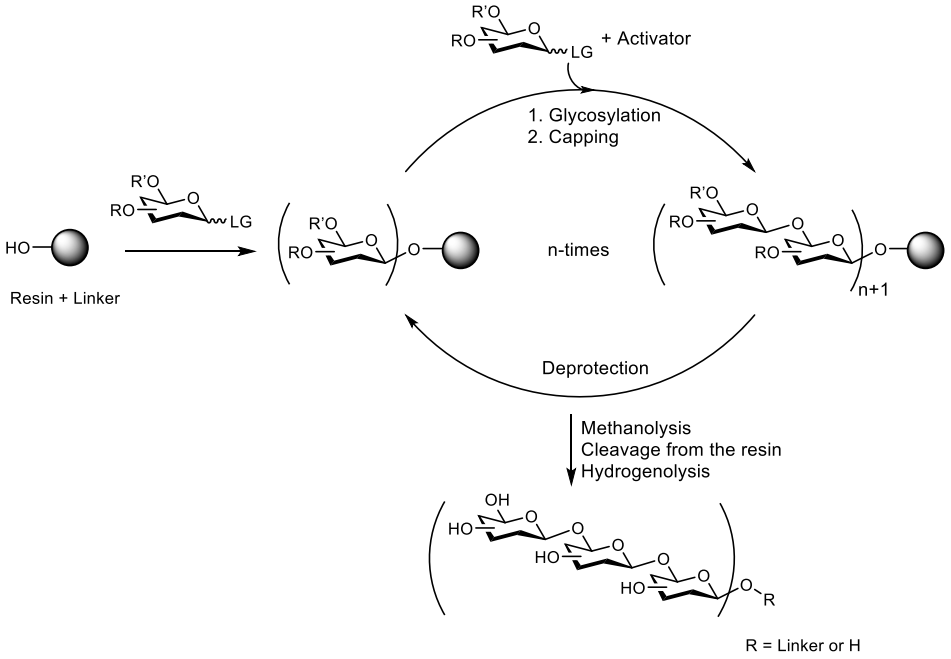


Figure 4 – Automated Glycan Assembly of carbohydrates

The solid support commonly used for AGA is Merrifield resin, a polystyrene cross-linked (divinylbenzene) polymer having good swelling properties. This resin is functionalized with a linker modified with a spacer or a monosaccharide. Some described linkers include metathesis-labile[163-164], base-labile[165], and photocleavable linkers[166]. In addition, recent syntheses employ traceless photolabile linkers that offer the possibility to cleave the glycan with a spacer at the reducing end and are ready for conjugation, as well as free reducing end glycans[167].

The most common building blocks employed in AGA are thioglycosides, which present high reactivity and bench stability[141]. When a more reactive donor is required, thioglycosides are converted to phosphates and imidates. These monosaccharides are equipped with permanent, participating, and non-participating orthogonal protective groups that allow the regioselective formation of oligosaccharides. Stereoselectivity for the formation of 1,2-*trans* linkages is mainly driven by anchimeric remote assistance at C-2. A recent report shows a significant advance in forming 1-2-*cis* linkages that rely on several strategies such as enhanced remote assistance, temperature, and solvent control[168]. However, more optimizations are still required in this field. The most common orthogonal protective groups employed are fluorenyl-9-methoxycarbonyloxy (Fmoc), benzoyl (Bz), and levulinoyl esters (Lev). Moreover, 2-naphthylmethyl ether (Nap) and chloroacetate ester (AcCl) have also been employed[162, 169].

During the last 20 years, AGA has shown a significant advance and substantial versatility for synthesizing linear and branched oligosaccharides derived from mammals[166, 170], bacteria[171-172] and plants[168, 173]. In addition, the rapid access to linear molecules has allowed the structural study of oligosaccharides[174-176]. The most recent improvement to this technique is the integration of microwave to speed up the total time of the synthesis enabling rapid temperature adjustment[169]. Furthermore, automated Glycan Assembly has proven efficient for obtaining several molecules. Nonetheless, the scope of targets should be increased because the synthesis of rare and complex glycans such as bacterial carbohydrates and glycosylphosphatidylinositol is still exclusively performed in solution-phase synthesis.

1.2.5.3. Enzymatic synthesis of oligosaccharides

Enzymatic synthesis represents an alternative to solution and solid-phase synthesis of oligosaccharides. During enzymatic glycosylation, the formation of bonds is always stereo and regioselective, and no protecting group manipulations are required[44, 19]. In stark contrast to chemical synthesis, enzymatic reactions occur in aqueous solutions and only need limited control of the reaction conditions (temperature, pH)[142].

In nature, enzymes make oligosaccharides with high acceptor and donor specificity linearly from the reducing end. The enzymes used in oligosaccharide synthesis are glycosyltransferases and glycosidases/glycosynthases. The glycosyltransferases catalyze the stereo- and regiospecific transfer of a monosaccharide from a donor (nucleotide sugars, lipid phosphor-sugar, and sugar-1-phosphates) to a nucleophilic substrate (oligo and polysaccharides and amino acid residue). Therefore, they have been widely used to prepare essential biomolecules such as carbohydrates and N- and O-glycoproteins[66-67]. In addition, glycosidases catalyze the selective hydrolysis of glycosidic linkages either with retention or inversion of the configuration. Since the hydrolytic reaction is an equilibrium, glycosidases can also be employed in the presence of excess glycosides to favor the product formation. This strategy showed to be efficient for the synthesis of glycosylated amino acids[177].

Although enzymatic synthesis occurs in mild conditions, protection of substrates is not required, and several glycosyltransferases have been characterized and used for oligosaccharide synthesis. Still, the identification and access to more enzymes are required[142]. To date, enzymatic methods are used for industrial-scale synthesis, and chemical methods remain the best practice in academic laboratories to provide new structures for exploratory research[178]. Combining chemical and enzymatic synthesis will bring further advances in glycan synthesis to allow access to complex glycans and glycosylated molecules.

Despite the innovations in producing homogeneous glycoproteins by different methods, challenges should still be faced. Each target represents a specific challenge, and a meticulously choice of strategy for its ligation or expression is required, and novel and more efficient enzymes are still to be discovered. Chemical synthesis allowed access to several targets, but due to the

variability of glycan structures, the optimization and expansion of current methods are necessary. Therefore, it is essential to continue the efforts to develop methodologies that allow the synthesis of complex natural or modified glycoproteins that have biological relevance and can be further used in the pharmaceutical market.

1.3. Selected examples of glycoproteins

As described in the previous sections, glycoproteins are involved in essential functions of living organisms, leading to an interest in using them for diagnostics and treatments. Therefore, understanding of the structure-activity relationship and the finding of suitable epitopes for screening are required. In the following section, two interesting glycoproteins are presented.

1.3.1. CD59

CD59 is a glycosylphosphatidylinositol-anchored protein that inhibits the formation of the membrane attack complex (MAC), therefore protecting cells from undesired complement-mediated cell lysis[179-180]. MAC has opsonizing properties and forms pores in the membrane of pathogens, which leads to an increased influx of water that bursts bacteria[181]. Enveloped viruses such as HIV incorporate CD59 in their lipid bilayer to inactivate the complement system, which prevents them from complement-mediated lysis[182].

The human CD59 glycoprotein (Figure 5), which can be isolated as a complex mixture of glycoforms from erythrocytes, consists of 77 amino acids, one C-terminal GPI anchor (Asn77), a single N-glycosylation site (Asn18) and a limited degree of O-glycosylation (e.g., on Thr51/Thr52)[181]. The structure is stabilized by five disulfide bonds involving cysteines 3-26, 6-13, 19-39, 45-63, and 64-69[183-185]. NMR and crystallographic studies have confirmed that the fold of the protein comprises three β-strands (C, D, E), a single α-helix (α1), and a very small helix represented as (α1'). The N-terminal region is a primarily irregular loop but contains a small β-ribbon (A) packed edge-on with the main β-sheet[183, 186].

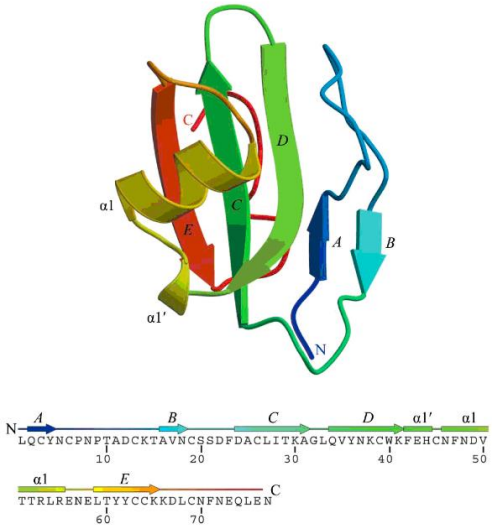


Figure 5 – Crystal structure and amino acid sequence of CD59; taken from Huang *et al* [186]

Human CD59 levels can present a high difference between cell types, but it is expressed in almost all tissues and circulating cells[187-188]. For instance, endothelial cells show higher expression levels than peripheral blood mononuclear cells (PMBCs)[189-190]. The presence of CD59 in several types of tissues and cells suggests that it might play an essential role in health and disease conditions. Neurons and endothelial cells upregulate the expression of CD59 to protect themselves from complement-mediated degeneration[191]. Moreover, studies have shown that on tumor cells/tissue, CD59 levels differ from the normal, which can be used as a biomarker for predicting pathologies[192]. Hyperexpression of CD59 protects tumor cells from immune attack by preventing MAC formation[179]. Inversely, age-related macular degeneration is related to lower expression of CD59[193], and glycoprotein deficiency could cause paroxysmal nocturnal hemoglobinuria[194].

Although CD59 plays a crucial role in protecting normal cells from complement attack, it has been reported to protect cancer cells. In addition, studies have shown that soluble CD59 has potential use as a biomarker for the diagnosis of lung diseases and diabetes[192, 195]. However, there is no advanced understanding of the functions of CD59, especially those related to the immune cells (T lymphocyte, dendritic cells, macrophages, and B lymphocyte cells, among others)[179]. Furthermore, the biological activity of the protein can be related to its structure. Therefore homogeneous glycoforms of this glycoprotein can significantly contribute to the investigation of its functions and the development of efficient therapeutic strategies to apply CD59 as a target in immunotherapy as is desired by the medical science community.

1.3.2. O-glycoproteins: mucins and mucin-like molecules

Mucins and mucin-like molecules (MLMs) are highly O-glycosylated proteins present on the cell surface of mammals and other organisms[13, 196-197]. These glycoproteins are highly diverse in the apoprotein and glycan cores and play a central role in many biological processes and diseases[53, 198]. Mucins are the most abundant macromolecules in mucus and are responsible for their biochemical and biophysical properties[198-199]. Mucin-like molecules cover various protozoan parasites[200-201], fungi[202], and viruses[203]. In humans, modifications in mucin glycosylation are associated with tumors in epithelial tissue[204-205]. These modifications distinguish between normal and abnormal cell[206] conditions and represent important targets for vaccine development against some cancers[22, 207-210]. In addition, mucins and mucin-like molecules derived from pathogens are potential diagnostic markers and targets for therapeutic agents[211-212].

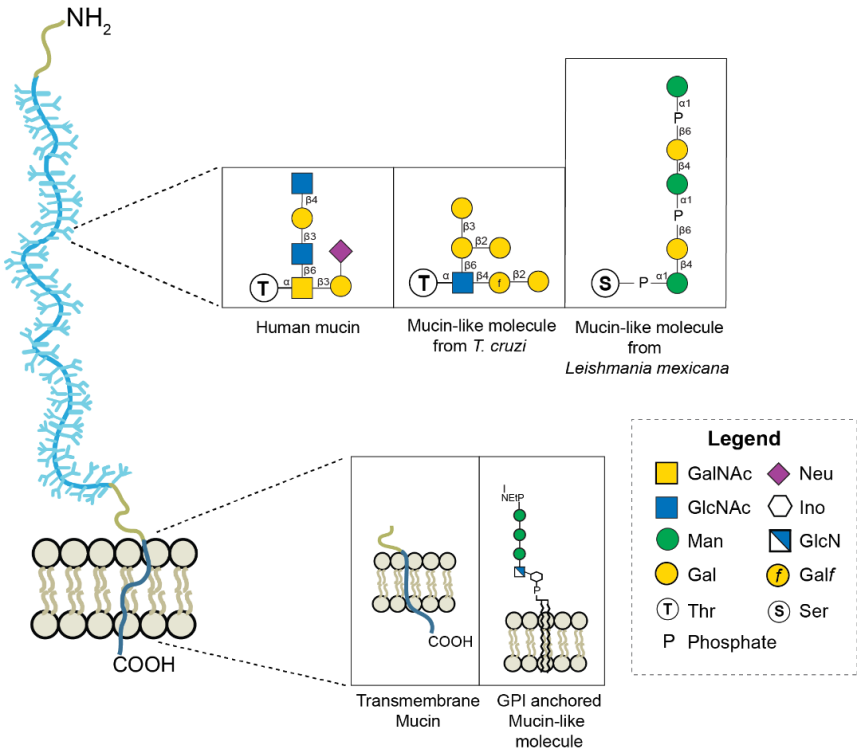


Figure 6 – Variation of the protein-glycan and membrane linkage and glycan structures in human mucins and protozoan MLMs, taken from Pinzón *et al* [1]

Mucins contain variable glycosylated tandem repeat domains rich in proline (Pro), threonine (Thr), and serine (Ser) (PTS domains)[199, 213], and cysteine-rich regions[214] localized at the amino and carboxy terminus and interspersed between the PTS domains[199], membrane-bound mucins have a common structure containing a transmembrane and a cytoplasmatic tail domain[215]. The apomucin, or protein core, and the oligosaccharides are different among mucins[53]. The structure of the O-glycans present in human mucins comprises three main parts: the GalNAc linked to the protein[216-217]; a backbone or extension part corresponding to an elongation of the GalNAc with either α-(1→3)- or β-(1→3)-linked galactose or by β-(1→6)-, β-(1→3)-, or α-(1→6)-linked *N*-acetylglucosamine[218]; and a high variable peripheral part containing fucose and *N*-acetyl neuraminic acid terminal units[53] (Figure 6).

1.3.2.1. O-Glycans in *T. cruzi*

Like human mucins, MLMs have domains rich in Pro, Thr, and Ser containing multiple O-glycosylations. The glycan structure in MLMs from many pathogens is unknown, but some differences have been reported. For example, the surface of the protozoan parasite *T. cruzi* is covered with MLMs and GPI-anchored glycoconjugates, termed mucins and mucin-associated surface proteins[219]. Early reports describing particular features of MLMs glycans derived from *T. cruzi* determined that the glycans are linked to the protein by α-GlcNAc instead of α-GalNAc like in human mucins[220-221] (Figure 7). Galactose constitutes the main residue for elongation, and its configuration varies depending on the strain and the stage of the parasite[222]. For instance, β-

galactofuranose is present in the epimastigote stage[223-226], while α-galactopyranose is more commonly found in trypomastigote mammal state parasite[227]. The other configuration, β-galactopyranose, was identified in all the strains of epimastigotes and trypomastigotes[222-223, 228].

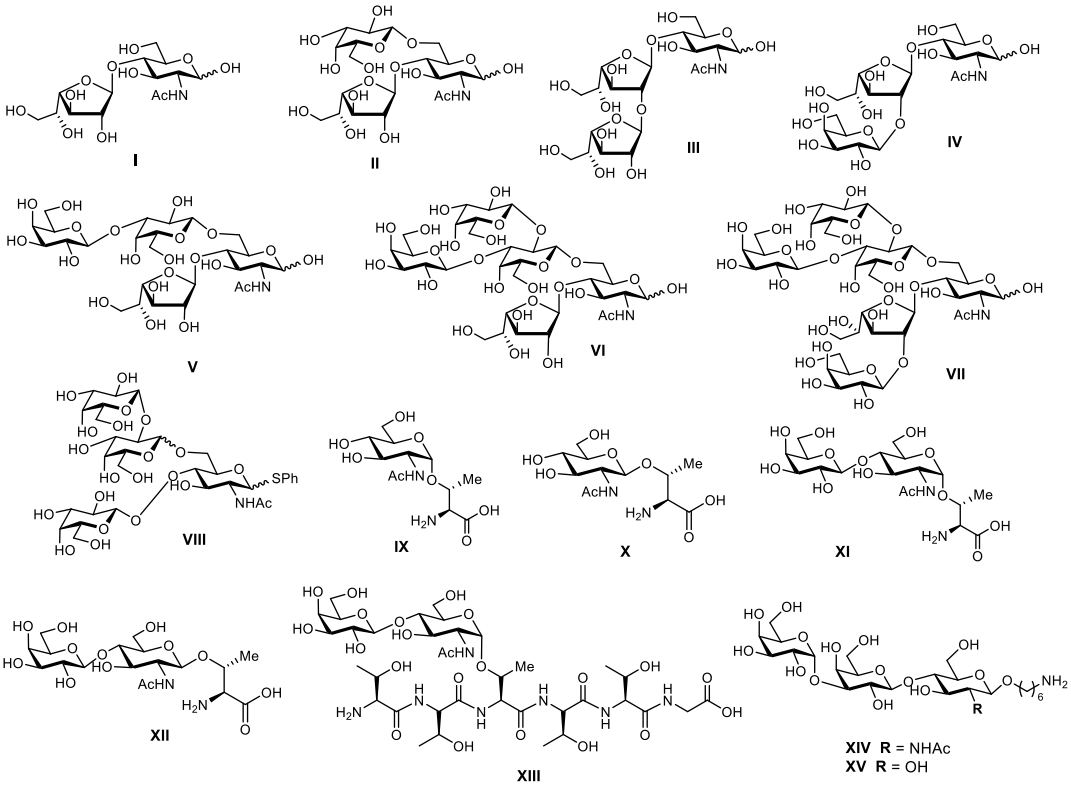


Figure 7 - Chemical structure of synthesized glycans from MLMs of *Trypanosoma cruzi* Colombiana strain (I, II, V-VII), Tulahuen strain (III, IV), and Y strain (VIII-XIII).

The characterization of these glycan molecules and protein core has served as a model to synthesize mucin-like O-glycans, peptides, glycosyl-amino acids, and glycopeptides. Initial synthesis includes the preparation of the O-linked saccharides **I**-**V** (Figure 7) present in *T. cruzi* Colombiana and Tulahuen strains[229]. The first synthetic target was disaccharide **I**[230], which is the basis of synthesizing other molecules, including trisaccharides **II**[231], **III,** and **IV**[232], tetrasaccharide **V**[233], pentasaccharide **VI**[234], and hexasaccharide **VII**[235]. Further reports include the synthesis of glycan **VIII** from the *T. cruz*i Y strain (Figure 7) [236]. Glycosyl amino acids **IX** and **X** and disaccharides glycosides **XI** and **XII** derived from the *T. cruzi* Y strain were synthesized to study the mucins as substrates for *trans*-sialidase activities; i.e. a chemoenzymatic reaction on the glycosyl amino acid **IX** was used to obtain the glycopeptide **XIII**. These studies delivered information about the relaxed acceptor substrate specificity of the *T. cruzi trans*-sialidase, which is essential to understanding the role of this enzyme during *T. cruzi* infections[237].

Mucins and MLMs are becoming important markers for diagnostics and drug and vaccine design targets. For example, human mucin-based structures are used as targets for cancer immunotherapy[212], and antibodies against ML-proteins are employed to discriminate *T. cruzi* lineages and diagnose Chagas disease[211]. However, mucins and MLMs research is still limited by access to pure materials and a poor understanding of the function of these molecules in

diseases. The heterogeneity and difficult characterization of isolated glycoproteins and the need for homogeneous drug and vaccine design material prompted the synthesis of mucin and MLM-related structures.

Human CD59 and glycosyl amino acids of *T. cruzi* are essential targets for diagnostic and treatments; homogeneous glycoforms of these derivatives are required for this purpose. Therefore, the development of strategies that allow the access to several glycoforms of human CD59 and glycosyl amino acids of *T. cruzi* to be screened represents an exciting research field. The contribution made to this field will be summarized in the following section (1.4) and further discussed in chapters 2 and 3.

1.4. Aim of the thesis

The main goal of this work was to develop strategies and methodologies for synthesizing glycosylated molecules. Two projects were conceived to establish the automated solid-phase synthesis of O-glycosyl amino acids and the application of synthetic glycopeptides in glycoprotein synthesis.

The first project focused on establishing a strategy for synthesizing homogeneous variants of human CD59 using native chemical ligation. The aim of the project includes the design of a strategy to allow the synthesis of the protein without glycosylation and bearing and N- and O-linked monosaccharides. In addition, the resulted glycoprotein should be suitable for further transformation, such as transglycosylation reactions, glycan enzymatic elongation, and the attachment of a glycosylphosphatidylinositol glycolipid. The structure of the target molecules, designed synthetic routes to prepare the peptides, and ligation of the fragments by native chemical ligation are described in chapter 2.

The second project should establish a methodology to access glycans and glycosylated amino acids by employing automated glycan assembly targeting glycosyl amino acids of mucin-like molecules present on *Trypanosoma cruzi (Y* strain). A methodology for the synthesis of galactose-containing glycans should be developed and applied to obtain O-glycosyl amino acids suitable for the synthesis of glycopeptides or for screening in biological applications.

The synthetic methodologies described in the thesis should serve as an efficient way to deliver molecules that can act as models for structure-activity relationships studies using pure glycans and glycoproteins. In addition, they would contribute to the knowledge of the therapeutic potential of carbohydrates. The methodology to synthesize glycosyl amino acids derived from mucins and mucin-like molecules should also allow easy access to molecules that can be screened as epitopes for diagnostics, treatments for cancer and infectious diseases, and used as building blocks for glycopeptides or glycoprotein synthesis. Overall, the project results in this thesis should provide model strategies for synthesizing glycoproteins, their protein and carbohydrate cores and provide a platform for studying their biological properties.

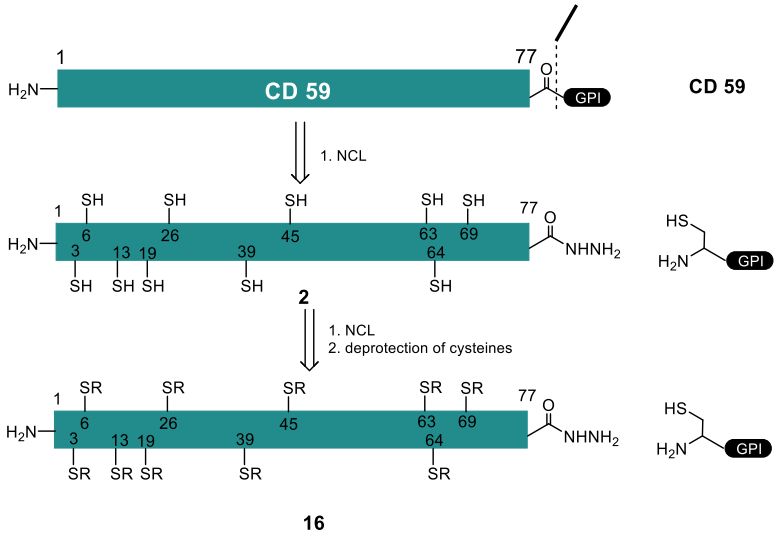
2. Chemical synthesis of human CD59

The development of strategies for the chemical synthesis of proteins and glycoproteins has driven the interest in obtaining several targets of biological interest. However, chemical protein synthesis is still a nascent technology requiring additional procedures and optimization of existing methodologies to overcome difficulties associated with assembling long sequences and complex protein modifications. A fundamental protein synthesis method uses peptides and glycopeptides obtained by the well-established SPPS. Despite the enormous advances, the SPPS of long peptides and peptides containing complex sequences is challenging and suffers from the accumulation of byproducts from side reactions, incomplete couplings, and the aggregation of the growing peptide chains[126, 137]. A significant challenge in glycoprotein synthesis is the synthesis of glycopeptides, where the difficulty inherent to carbohydrate synthesis converges with the limitations of peptide synthesis.

Chemoenzymatic processes emerged as an alternative to solve the difficulties present in chemical glycoprotein synthesis, especially in the protein and glycans elongation steps. In these strategies, the protein part is recombinantly expressed or synthesized by ligation of chemically synthesized fragments. The glycans are incorporated either using synthetic glycopeptides or by enzymatic elongation of monosaccharide or oligosaccharides units. A recombinant protein expression limits the possibilities for functionalization to resemble the post-translational modifications of natural glycoproteins. In contrast, chemical synthesis allows for modifications at specific sites and the synthesis of large, tailored molecules such as glycoproteins.

There is a need for homogeneous glycoforms of glycoproteins to determine the role of glycans in protein activity. This work envisioned establishing a strategy for synthesizing glycoproteins using the glycosylated human CD59 protein as a model. This glycoprotein contains an N- and O-glycosylation at Asn-18 and Thr-51 and a C-terminal GPI anchor[17]. In addition, the project aimed the synthesis of a non-glycosylated protein and glycosylated variant of CD59 bearing monosaccharides to investigate further enzymatic modifications such as transglycosylation, enzymatic O-glycan elongation, and glypiation.

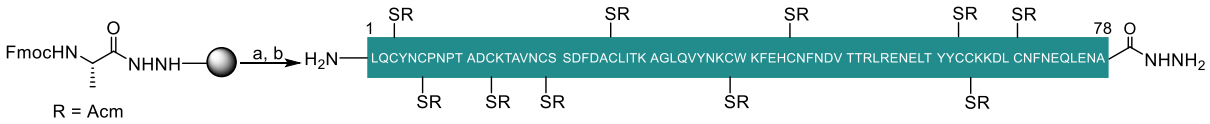
2.1. Strategy I: SPPS of the non-glycosylated CD59



Scheme 10 - Retrosynthesis of a non-glycosylated CD59

The retrosynthetic analysis to obtain CD59 without glycosylation considered one main disconnection between the GPI and the C-terminus of the protein and two fragments, a protein and a glycosylphosphatidylinositol (GPI). The synthesis strategy was envisioned using native chemical ligation between a C-terminal thioester of the non-glycosylated protein (CD59 1-77) and a cysteine-bearing GPI for protein glypiation.

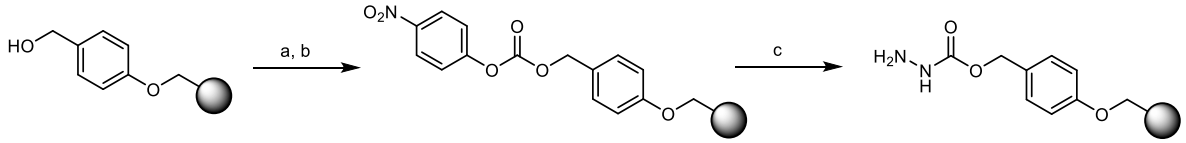
CD59 is a small protein containing 77 amino acids, which is a size commonly found in publications describing the SPPS of long peptides. Therefore, the SPPS of the whole CD59 sequence in one single process was evaluated first (Scheme 10). Considering the different methods to obtain peptide thioester by Fmoc-SPPS, the synthesis of a protein hydrazide as a thioester precursor was selected for this strategy (Scheme 11). Furthermore, peptide hydrazides are stable to the SPPS conditions and can be easily and rapidly converted into a thioester for the NCL reaction[117]. The natural CD59 sequence contains asparagine at the C-terminus (Asn-77), a residue prone to side reactions during the thioester formation and succinimide formation by cyclization of the asparagine side chain with the activated C-terminus[116]. Therefore, an alanine residue was added to the C-terminus to avoid this undesired reaction and facilitate the protein synthesis without alterations to the protein structure.



Scheme 11 – Synthesis of partially protected non-glycosylated CD59. Reagents and conditions: a) MW-SPPS: coupling Fmoc-Amino acid, *N,N'*-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF; b) TFA/phenol/water/thioanisol/EDT (82.5/5/5/5/2.5), rt.

The non-glycosylated protein synthesis was designed involving three main steps: a manual functionalization of the resin to install a linker for peptide hydrazide synthesis, coupling of the C-terminal amino acid, and a MW-assisted SPPS for the elongation of the amino acid chain.

The hydrazide-functionalized resin was prepared from Wang resin as described in the literature by reaction of the hydroxyl group with *p*-nitrophenyl chloroformate in the presence of *N*-methyl morpholine, followed by nucleophilic attack employing hydrazine monohydrate[116] (Scheme 12).



Scheme 12 – Functionalization of NovaPEG Wang resin. Reagents and conditions: a) *p*-nitrophenyl chloroformate, *N*-methyl morpholine, CH_2Cl_2, 0 °C to rt, overnight b) hydrazine monohydrate CH_2Cl_2/DMF (1:1) 0 °C to rt, overnight.

The C-terminal Fmoc-Ala-OH was coupled manually to the hydrazine resin using DIC and 6-Cl-HOBt activation. Considering the number of amino acids in the sequence, this first coupling was used to set the resin substitution to 0.15 mmol/g, which may allow peptide elongation and reduce the peptide aggregation. The resin loading was determined by UV Fmoc removal quantification using reported methods[238]. After obtaining the desired resin loading, unreacted hydrazine groups were capped, employing a solution of Ac_2O/DIPEA/DMF (1:1:10) to deliver the resin ready for peptide elongation.

Initial elongation by MW-assisted SPPS (residues 70-77) was performed at 75 °C using single cycles of 20% piperidine in DMF for Fmoc removal and coupling with DIC/Oxyma amino acid activation. Different conditions were needed only for three amino acids cysteine and histidine coupled at 50 °C and arginine at room temperature following literature protocols to prevent epimerization[239] and δ-lactam formation by arginine[240]. Double deprotection and coupling cycles were performed from the residue 70. This change was necessary to avoid incomplete coupling or deprotections, especially in amino acids appearing double and consecutive in the sequence (Lys -66/Lys-65, Cys-64/Cys-63, and Tyr-62/Tyr-61). Three sites were selected to monitor the peptide elongation and synthesis progress: (Thr-52, Cys-39, and Leu-1). To do this, an aliquot of resin beads was taken from the reaction, a small-scale cleavage employing TFA was performed, and the obtained crudes were analyzed.

LC-MS analysis of the cleaved peptides showed the progression of the synthesis (Figure 8). The chromatogram profile after Thr-52 showed a principal peak corresponding to the expected 26-residues peptide and some deletion sequences (Figure 8a). Since the expected peptide was observed and detected as the main product, the elongation process proceeded until Cys-39. LC-MS analysis at this point revealed the main peak corresponding to the desired product (Figure 8b). However, the peaks corresponding to deletion sequences increased in intensity and number. The presence of the expected 39-residues peptide encouraged to continue the elongation of the sequence. Unfortunately, after completing the synthesis of the CD59 sequence (1-78), an

inseparable mixture of products was obtained, and none of the observed peaks could be assigned by MS to the protein (Figure 8c).

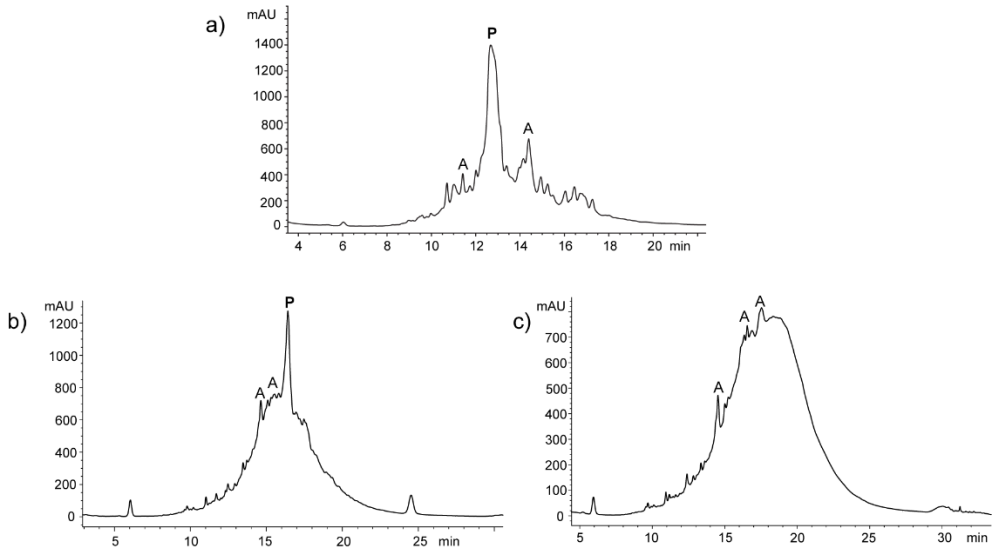


Figure 8 - LC-MS analysis of the crude fragments synthesized with 2-CTC resin. Chromatogram recorded at 214 nm after cleavage a) 52-78; b) 39-78; c) 1-78. (RP-HPLC with C18 column, gradient: 5 to 40% of ACN in water + 0.1% of HCOOH in 25 min).

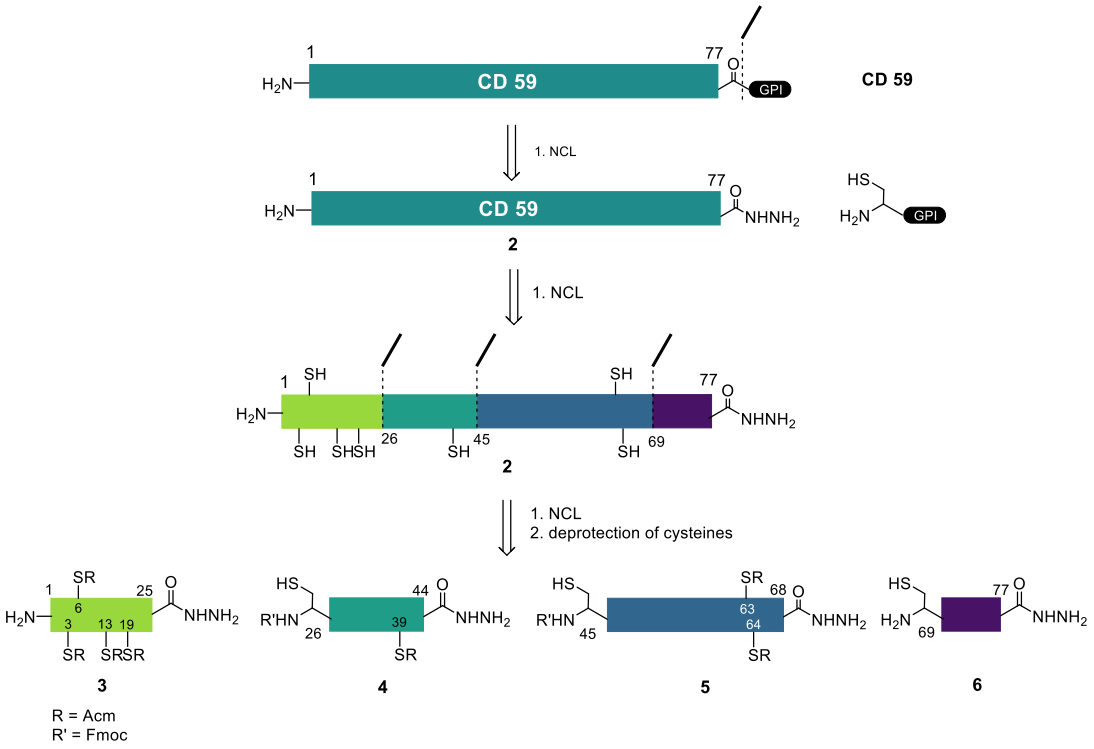
Despite the double cycles employed to assure complete Fmoc removal and amino acid coupling, total protein synthesis by MW-SPPS was unsuccessful, giving a complex mixture of peptides that could not be assigned. Furthermore, in agreement with the literature, the synthesis of this long peptide, containing more than 50 amino acids, presented several difficulties and side reactions without forming the desired peptide[108]. These results prompted a methodology change. Thus, the synthesis was not further optimized, and a new strategy employing native chemical ligation was considered for glycosylated and non-glycosylated CD59 synthesis.

2.2. Strategy II: One-pot NCL C- to N- terminal assembly

MW-SPPS of the full-length CD59 presented difficulties associated with the peptide size and difficulty of the sequence. Native chemical ligation is a methodology allowing the synthesis of long proteins using short peptides and offers advantages in obtaining complex proteins. Thus, this methodology was explored next to synthesize the target protein. Considering the high number of cysteine residues in the CD59 glycoprotein and their potential use as ligation points by NCL, several disconnections are possible for its synthesis by this strategy. Therefore, other different aspects were considered for the peptide selection, such as the sequence length and the position of the glycosylation sites.

2.2.1. Retrosynthetic analysis and strategy design

Retrosynthetic analysis using NCL-based synthesis of non-glycosylated CD59 considered similar size peptide fragments and four ligations as a suitable method to obtain this protein variant. The starting point was the disconnection between the GPI and the C-terminus of the protein that requires a protein hydrazide and GPI to deliver the glypiated CD59 in the last step of the process. Like the previous strategy, this disconnection was selected due to difficulties synthesizing both parts. Then, three further disconnections in the protein part at residues 69, 45, and 26 delivered a set of four peptide fragments, three of them bearing an N-terminal cysteine for NCL (Scheme 13).



Scheme 13 - Retrosynthesis analysis of non-glycosylated CD59 leading to fragments 3-6

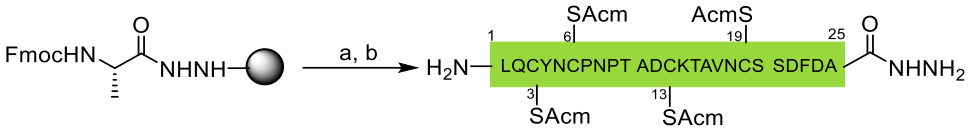
A strategy involving sequential native chemical ligation of the peptide or glycopeptide fragments **3**-**6** having Cys-26, Cys-45, and Cys-69 as junction points required specific modifications at the N- and C-terminus of these peptides. To allow future glypiation via NCL with a modified GPI, the CD59 C-terminus peptide fragment **6** may contain a hydrazide for further conversion into thioester after the protein core synthesis (Scheme 13).

Additional essential factors were considered for the design of the four peptide fragments. First, all peptides may contain a hydrazide for conversion into the corresponding thioester. Second, the internal cysteines needed a protective group due to the high number of cysteines in CD59 and the possible formation of thiolactone side-products during the thioester synthesis. Finally, to perform C- to N-sequential ligations, the N-terminal cysteine of peptide fragments **4** and **5** also required protection. In this case, Fmoc was selected as the orthogonal protective group.

Four peptide hydrazides were designed for this strategy: Leu-1 to Ala-25 (**3**), Cys-26 to His-44 (**4**), Cys-45 to Leu-68 (**5**), and Cys-69 to Ala-78 (**6**). The acetamidomethyl group (Acm) served as a protecting group for the side chains of internal cysteines 3, 6, 13, 19, 39, 63, and 64. The N-terminal of the peptides **4** and **5** design included protection with a carbamate, such as the Fmoc group. The use of peptide hydrazides as thioester surrogates presents several advantages as they can be easily and rapidly converted to a peptide thioester and are unreactive under native chemical ligation unless activated. Therefore, they can be used as masked thioester in the convergent synthesis of proteins[117].

2.2.2. Solid-phase synthesis of CD59 peptide fragments

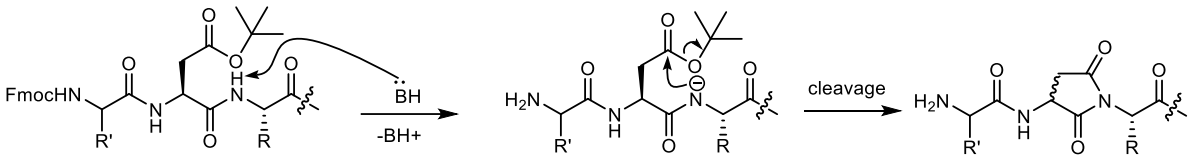
Synthesis of CD59 peptide fragment 3 (1-25)



Scheme 14 – Synthesis of peptide hydrazide 3. Reagents and conditions: a) MW-SPPS: coupling Fmoc-Amino acid, *N,N*'-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF b) TFA/phenol/water/thioanisol/EDT (82.5/5/5/5/2.5), rt.

The process started with the synthesis of the hydrazide-functionalized Wang resin using the protocol early described (Scheme 12)[116]. Next, Fmoc-Ala-OH was activated using DIC and 6-Cl-HOBt and coupled manually to the hydrazine resin. The coupling efficiency was determined by Fmoc quantification, giving a loading of 0.43 mmol/g that corresponded to 84% of the reacted sites. Finally, the remaining unreacted groups were acetylated employing an Ac_2O/DIPEA/DMF (1:1:8) solution.

Initial attempts for synthesizing peptide **3** involved standard couplings with DIC/Oxyma and deprotections at 75 °C. Several products were detected by LC analysis after the peptide synthesis (Figure 9a). MS analysis of these products revealed the formation of various deletion sequences and the corresponding mass of the expected peptide with signals related to $[M-18]^{2+}$, $[M-36]^{2+}$, and $[M-54]^{2+}$. Since the peptide contains three aspartic acid residues (Asp-Ala, Asp-Phe, Asp-Cys(Acm)), the detected masses were associated with aspartimide formation at these residues. These results underlined the need of method optimization for synthesizing peptide **3**. Two strategies focused on improving the synthesis; applying methods to reduce the aspartimide side-reaction and optimizing the coupling conditions to avoid deletion sequences.



Scheme 15 - Base mediated aspartimide formation

Aspartimide formation is a base-catalyzed side reaction commonly occurs in SPPS during the treatments with piperidine for the Fmoc removal[241] (Scheme 15). This side reaction is usually associated with a peptide containing an Asp-Gly motif in the sequence. However, aspartimide can occur in other motifs depending mainly on the neighboring residue of aspartate[133].

Various strategies are reported to overcome efficiently the appearance of this cyclization reaction during SPPS[242]. These strategies were evaluated to improve the synthesis of peptide **3**. The first strategy used piperazine instead of piperidine for Fmoc removal. Different reports described piperazine as a milder base for Fmoc-removal[243] due to its associated pKa. An MW-SPPS of peptide **3** using a piperazine solution (10% w/v) in NMP/EtOH (90:10) showed a slight enhancement in the quality of the obtained products by LC-MS analysis (Figure 9b). The mass of the expected product was present. However, the general synthesis outcome did not improve; deletion sequences and aspartimide were still formed during the sequence elongation.

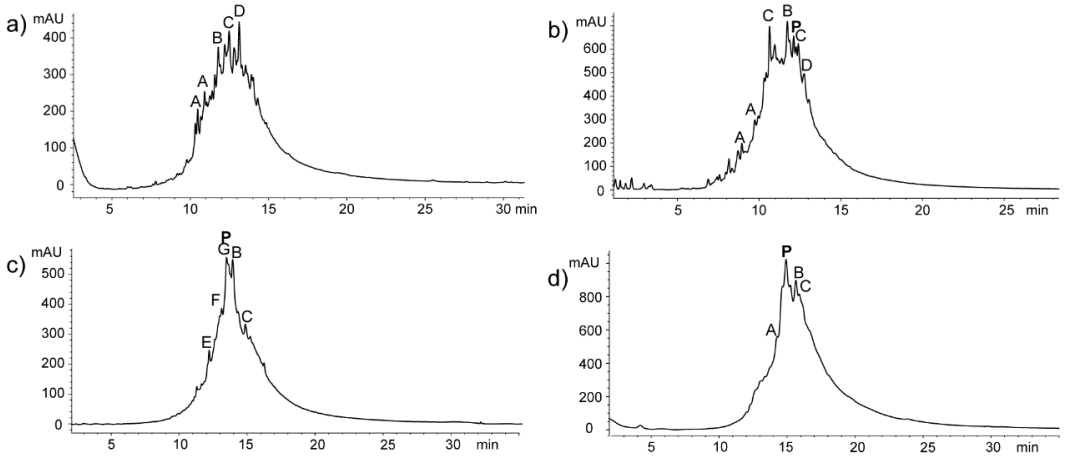
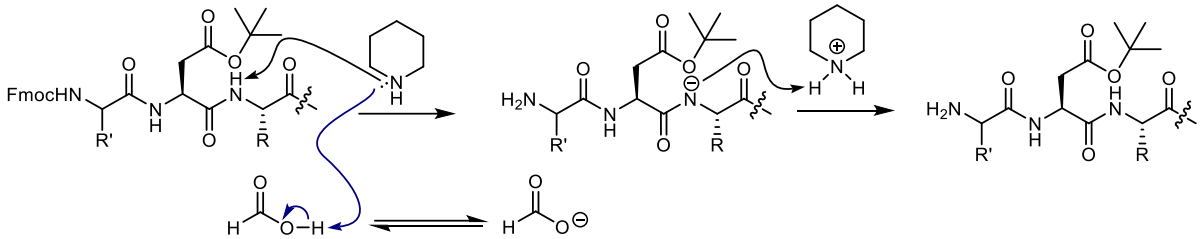


Figure 9 - LC-MS analysis of the crude peptide 3 (CD59 1-25) synthesized under different deprotection conditions. Chromatograms were recorded at 214 nm after small scale cleavage (RP-HPLC with C18 column, gradient: 5 to 40% of ACN in water + 0.1% of HCOOH in 25 min). a) 20% piperidine, 75 °C; b) 10% piperazine, 75 °C; c) 20% piperidine + 0.8% of HCOOH; 90 °C d) 20% piperidine + 0.8% of HCOOH; 75 °C. **P**: Peptide **3**; **A**: deletion by-products; **B**: $[M-18+2H]^{2+}$; **C**: $[M-36+2H]^{2+}$; **D**: $[M-54+2H]^{2+}$; **E**: $[M-Pro+2H]^{2+}$; **F**: $[M-Asn+2H]^{2+}$; **G**: $[M-Ala+2H]^{2+}$; **H**: $[M+tBu+2H]^{2+}$.

Literature reports have also presented aspartimide formation can be reduced using additives and acids in the deprotection solution[241, 244]. However, the two common additives (HOBt, and Oxyma) presented some drawbacks for their application in synthesizing peptide **3**. HOBt is explosive under dry conditions and is not recommended for reactions at high temperatures. Ethyl 2-cyano-2-(hydroxyamino)acetate (Oxyma) is expensive and commonly used in carbodiimide-mediated amino acid activation. The acid-mediated prevention of aspartimide formation studied by Mier and coworkers[241] suggested the addition of formic acid as an efficient and cost-effective method to avoid aspartimide formation in SPPS. Thus, adding 0.8% formic acid to the deprotection solution (20% piperidine in DMF) was tested to synthesize the peptide fragment **3**.

An additional synthesis of **3** performed at 90 °C employing 20% piperazine + 0.8% of formic acid in DMF (v/v) revealed a significant improvement of the process with reduced aspartimide formation observed at only two sites (Figure 9c). Previous reports described the abstraction of the

amide backbone proton as the critical step in the cyclization leading to aspartimide formation. Therefore, the addition of acid increases the concentration of protonated base to prevent the undesired deprotonation of the amide backbone nitrogen, thus decreasing the percentage of aspartimide formation (Scheme 16)[241, 245].



Scheme 16 - Proposed mechanism of the action of the formic acid in the deprotection solution

Changes in the temperature during the Fmoc removal may also contribute to aspartimide formation and its appearance in SPPS processes[130-131, 246]. Therefore, to analyze the temperature effect in the synthesis of peptide **3**, a peptide preparation using deprotection solution with formic acid was also performed at 75 °C. Unfortunately, no further improvement was observed in the synthesis outcome (Figure 9d). Nevertheless, the addition of acid showed to be an effective way to reduce this side reaction in the MW-SPPS, becoming the standard deprotection solution of all following syntheses.

Since aspartimide formation represented a big challenge for synthesizing peptide fragment **3**, several other parameters were considered to minimize this side reaction in further synthesis. An analysis of the amino acids neighboring to aspartate in peptide **3** showed, in agreement with the literature, that the motifs Asp-Ala and Asp-Cys(Acm) had a pronounced tendency to form this side reaction[133]. Due to the position of these amino acids in the sequence, a second approach described in the literature to avoid aspartimide was considered for the synthesis of **3**, which is the introduction of a bulkier protective group for the side chain of aspartic acid. This new protecting group was especially considered for the Asp-Ala motif closer to the C-terminus and should undergo many deprotection cycles. Furthermore, using a bulkier linker such as a trityl linker was also envisaged to reduce aspartimide formation at this position.

Microwave-assisted SPPS increases the deprotection and coupling rates. However, MW heating might also accelerate side reactions such as racemization and/or aspartimide formation[130, 243]. Thus, the suggested temperatures for deprotection and coupling in the system were carefully analyzed in both cycles to diminish the extent of these side reactions. It was observed that reducing the coupling temperature might also reduce the deprotection temperature without a cooling system or long cooling times. By performing couplings at 75 °C, the temperature in the reaction vessel during deprotection presented significant differences between the programmed value (25 °C) and the measured value (50 °C). On the contrary, the difference between these values was more negligible using coupling cycles at 50 °C, which allowed temperature reduction to 38 °C during deprotections (Table 1).

Table 1 – Variation in the temperature under different coupling and deprotection conditions

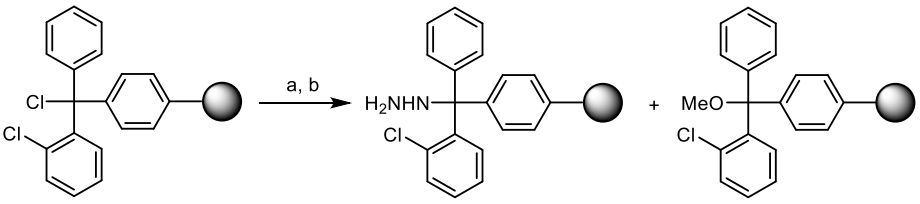
Step	Initial Method	Optimized method
Coupling	75 °C	50 °C
Deprotection (programmed)	25 °C	25 °C
Deprotection (measured)	50 °C	38 °C

The final optimization incorporated all the strategies mentioned above: use of formic acid in the deprotection solution, use of a bulkier protective group for the aspartic acid side chain, reduction of temperatures during deprotection and coupling, and use of a sterically hindered linker on the resin. Previous reports demonstrated the Asp(OMpe) applicability for reducing aspartimide formation[133]. However, the amino acid Asp(OEpe) was preferred because it contains a bulkier group that may protect by substantial steric hindrance the sensitive motifs (Asp-Ala, Asp-Cys(Acm)). A 2-chlorotrityl chloride (CTC) linker on a PEG-PS resin was selected as the best combination of a hindered linker and solid support for synthesizing challenging peptides. In addition, double coupling and capping cycles in selected amino acids were included to avoid deletion sequences and the TFA cocktail cleavage K containing TFA/phenol/water/thioanisol/EDT (82.5/5/5/5/2.5) was incorporated in the strategy. This cocktail solution reduces the reaction between the electrophilic species generated during the global deprotection and the product. The optimized conditions for the synthesis of the peptide **3** are summarized in Table 2.

Table 2 – Optimized conditions for the synthesis of peptide **3**

Step	Optimized method
Deprotection	20% piperidine + 0,8% formic acid in DMF, 25 °C
Coupling	DIC/Oxyma in DMF, 50 °C
Bulkier groups	2-cloro trytyl chloride attached to the resin and Fmoc-Asp(OEpe)-OH
Cocktail cleavage	TFA/phenol/water/thioanisol/EDT (82.5/5/5/5/2.5)

The new strategy to synthesize peptide **3** started with functionalizing the 2-chlorotrityl chloride resin as described in literature [117-118] employing 1 M hydrazine in THF (10 equiv.), followed by capping of unreacted groups employing 5% methanol in DMF (Scheme 17).



Scheme 17 – Functionalization of 2-chloro trityl chloride resin with hydrazine in THF. Reagents and conditions: a) 1 M N_2H_4 in THF in DMF, rt, 2 h; b) 5% MeOH in DMF, rt, 15 min.

The Fmoc-Ala-OH was manually coupled to the resin employing HATU as an activator and DIPEA as a base. Quantification of the Fmoc removal showed substitution of 0.28 mmol/g, corresponding

to 70% conversion of the active sites on the resin. After further capping of remaining reactive groups on the resin, the expensive Fmoc-Asp(OEpe)-OH was manually coupled to the resin using two equivalents and monitoring the coupling efficiency with a qualitative Kaiser test. Finally, the peptide was elongated using the optimized conditions for MW-SPPS.

Analysis of the crude peptide showed an improvement in the synthesis. Although the strategy modifications reduced the reaction rates and fewer side products were observed, the aspartimide formation was not entirely suppressed. Nevertheless, this complex synthesis was significantly optimized, and the peptide hydrazide **3** was obtained with an acceptable 9% yield after purification (Figure 10).

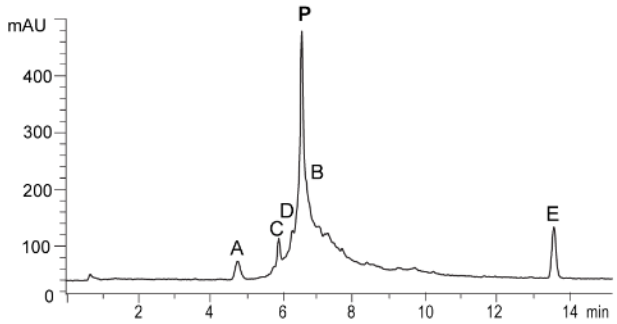
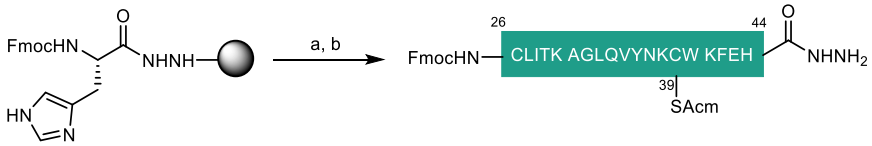


Figure 10 - LC-MS analysis of the crude peptide 3 using the optimized methodology. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 70% of ACN in water + 0.1% of HCOOH in 25 min). **P**: Product **3**; **A**: deletion by-products; **B**: [M-18+2H]$^{2+}$; **C**: [M-Ser+2H]$^{2+}$; **D**: [M-Val+2H]$^{2+}$; **E**: not mass recorded.

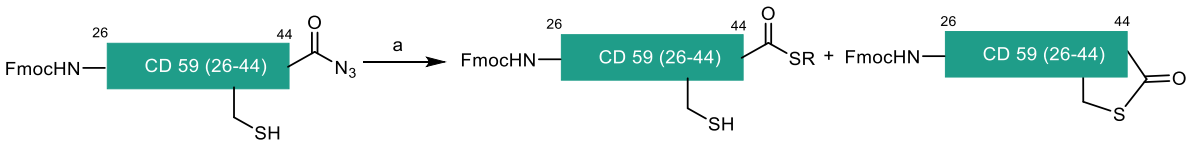
Synthesis of CD59 peptide fragment 4 (26-44)



Scheme 18 - Synthesis of peptide hydrazide 4. a) MW-SPPS: coupling Fmoc-Amino acid, *N,N'*-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF b) TFA/TIPS/water (90:5:5), rt, 3 h.

Sequential native chemical ligation from C- to N-direction required protection of the N-terminal Cys residue of the middle peptide fragments **4** and **5**. This protection is used to avoid polymerization and/or cyclizations caused by intramolecular reactions during the thioester formation and ligations[247]. Reports showed that *N*-allyloxycarbonyl-S-tritylcysteine (Alloc-Cys(Trt)-OH) is an efficiently protected amino acid for one-pot multiple peptide ligation[248]. However, expensive Pd catalyst and other metal complexes are required for the chemoselective deprotection when both Alloc and Acm groups are present in the same peptide[248-249]. In addition, careful chemical manipulation is necessary to prevent side reactions during the deprotection of cysteine with these reagents in aqueous solutions[250]. These considerations and a study by Mandal and coworkers[250] suggested Fmoc-masking as a better group for protecting the N-terminal cysteine peptides.

The synthesis of **4** started with the hydrazide-functionalized resin, which was prepared as described before (Scheme 12). The first amino acid, Fmoc-His(Trt)-OH, was activated using DIC and 6-Cl-HOBt and coupled manually to the resin, and the peptide was elongated using the MW-SPPS optimized conditions (Scheme 18). In this sequence, the internal cysteine (Cys-39) contained an Acm protection to avoid thiolactone formation during the thioester synthesis (Scheme 19). To ensure the N-terminal cysteine protection was intact, Fmoc-Cys(Trt)-OH was manually coupled after the elongation of the peptide by MW-SPPS.



Scheme 19 – Possible side reaction during thioester synthesis having deprotected cysteine. Reagents and conditions: a) thiol, pH 6.8, rt.

An initial attempt to synthesize the peptide using single deprotection and coupling cycles was delivered in four products by LC-MS (Figure 11a). The main peak **P** in the chromatogram corresponded to the expected peptide. However, a second peak showed the main deletion sequence lacking an isoleucine residue (peak A). Two additional peaks derived from a side reaction having a mass of $[M+45]^{2+}$ were attributed to the incomplete removal of the Boc group in tryptophan (peaks B and C). A peak D corresponding to the carboxylate peptide was also detected.

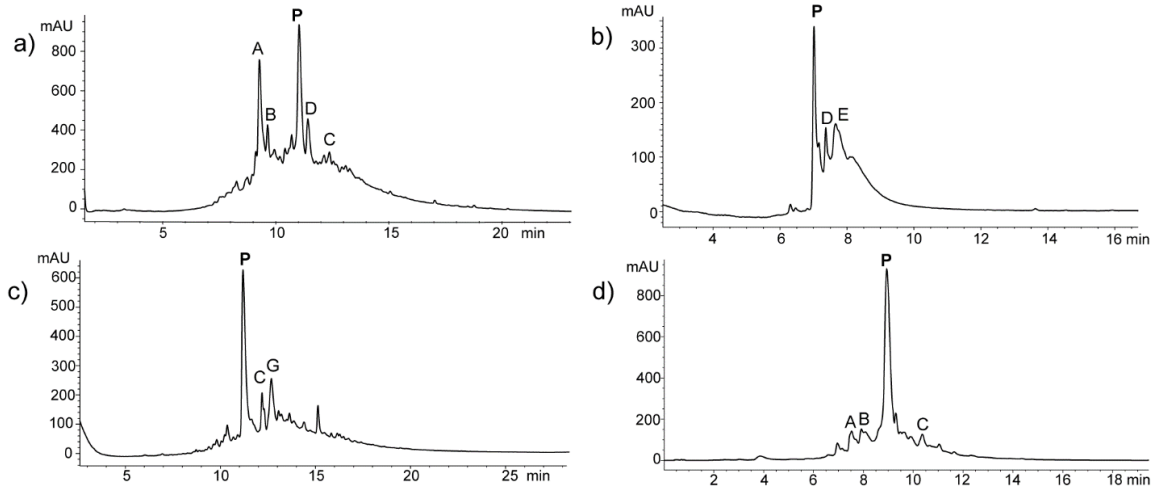
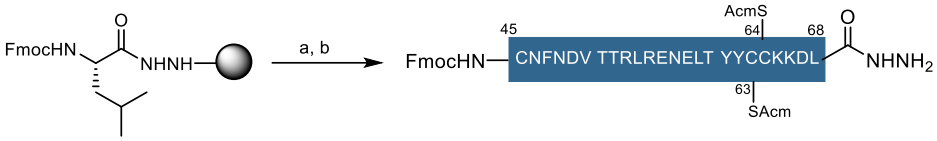


Figure 11 - LC-MS analysis of the crude peptide 4. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 40% of ACN in water + 0.1% of HCOOH in 25 min (a, c, d) or 5 to 70% ACN in water + 0.1% of HCOOH in 17 min (b). **P**: Product **4**; **A**: $[M\text{-Ile}+2H]^{2+}$; **B**: $[M\text{-Ile}+45+2H]^{2+}$; **C**: $[M+45+2H]^{2+}$; **D**: $[M\text{-}14+2H]^{2+}$; **E**: $[M+26+2H]^{2+}$; **G**: $[M\text{-His}+2H]^{2}$.

Synthesis of CD59 peptide fragment 5 (45-68)



Scheme 20 - Synthesis of peptide hydrazide 5. a) MW-SPPS: coupling Fmoc-Amino acid, *N,N*'-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF b) TFA/phenol/water/thioanisol/EDT (82.5/5/5/5/2.5), rt.

The peptide hydrazide **5** also required the protection of the N-terminal cysteine using a Fmoc carbamate and protecting the vicinal Cys-63 and Cys-64 using the Acm group to avoid thiolactone side-products (Scheme 20). Attempts to synthesize this peptide on a hydrazine-functionalized Wang resin are presented in Figure 12. The initial synthesis was performed using standard deprotection with 20% piperidine in DMF, which resulted in product deletion sequences presenting aspartimide formation (Figure 12a). Contrary to the synthesis of peptide **3**, the synthesis of **5** improved using a piperazine solution (10% w/v) in NMP/EtOH (90:10) for deprotection, but the aspartimide was still detected (Figure 12b).

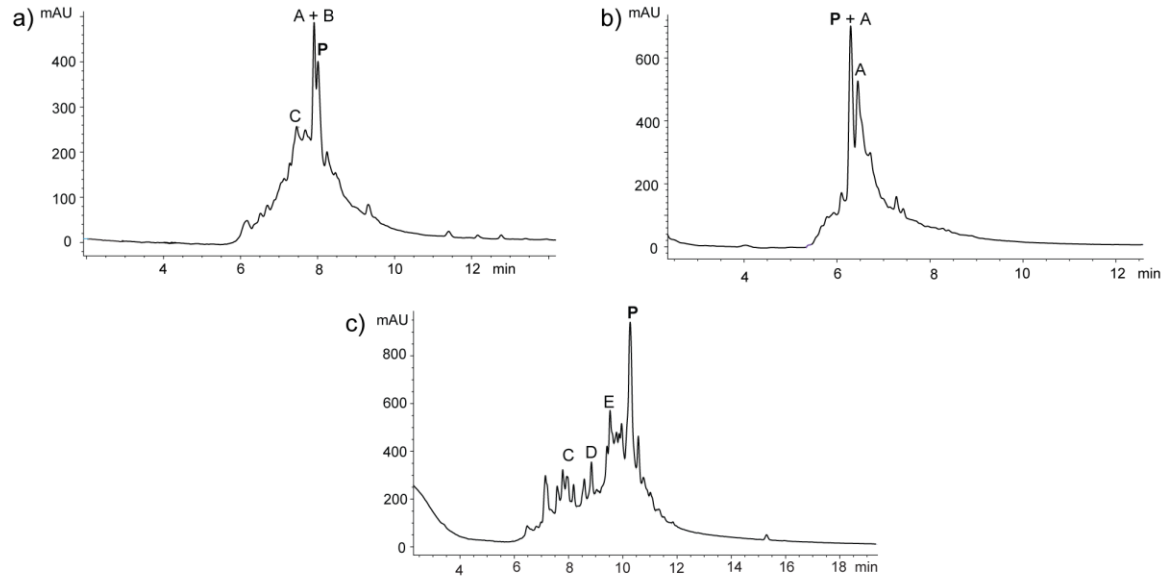
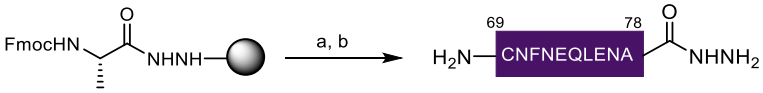


Figure 12 - LC-MS analysis of crude peptide 5 employing different deprotection conditions. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 70% of ACN in water + 0.1% of HCOOH in 30 min (a, b) or 5 to 60% ACN in water + 0.1% of HCOOH in 25 min (c). a) 20% piperidine, 75 °C (Cys-45 to Leu-68); b) 10% piperazine, 75 °C (Thr-52 to Leu-68); c) 20% piperidine + 0.8% of HCOOH 75 °C (Cys-45 to Leu-68); **P**: Product **5**; **A**: $[M-18+2H]^{2+}$; **B**: $[M-36+2H]^{2+}$; **C**: several deletion sequences; **D**: $[M-Tyr+2H]^{2+}$; **E**: $[M-Leu+2H]^{2+}$.

The strategies adopted to optimize the synthesis of peptide **3** were also employed as a suitable method for the synthesis and elongation of peptide **5**. The optimized synthesis was performed using 2-chlorotrityl hydrazine functionalized resin and 20% piperidine with 0.8% formic acid in DMF for the Fmoc removal. The temperature of the coupling and deprotection was reduced to 50 °C and 25 °C, and a capping cycle was included in selected positions. The outcome of the

synthesis significantly changed after applying this modified method (Figure 12c). However, double coupling cycles were necessary for the second amino acid of repeated residues to avoid deletion sequences at these positions. After this optimization, the peptide **5** was obtained in a 9% yield after purification.

Synthesis of CD59 peptide fragment 6 (69-78)



Scheme 21 - Synthesis of peptide hydrazide 6. Reagents and conditions: a) MW-SPPS: coupling Fmoc-Amino acid, *N,N'*-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF b) TFA/TIPS/water (90:5:5), rt, 3 h.

The synthesis of the peptide **6** was completed employing hydrazide-functionalized Wang resin that was also modified with Fmoc-Ala-OH[116]. The first amino acid was coupled using DIC and 6-Cl-HOBt. MW-SPPS peptide elongation was completed using single deprotection and coupling cycles at 75 °C (Scheme 21). LC-MS analysis of the product showed an efficient synthesis of the peptide, and only a tiny amount of a byproduct lacking asparagine was observed (Figure 13a). Since the sequence contains three Asn residues, it was difficult to determine the exact position where the coupling was incomplete. Therefore, a double coupling for Asn-72 and Asn-70 was employed to assure complete conversion. After cleaving the peptide employing TFA/TIPS/water (90:5:5), the product presented major solubility problems hampering the peptide purification. This step was time demanding and represented the main challenge to obtaining the peptide due to the need to use purification only in small batches. The peptide **6** was isolated with a good yield of 40%.

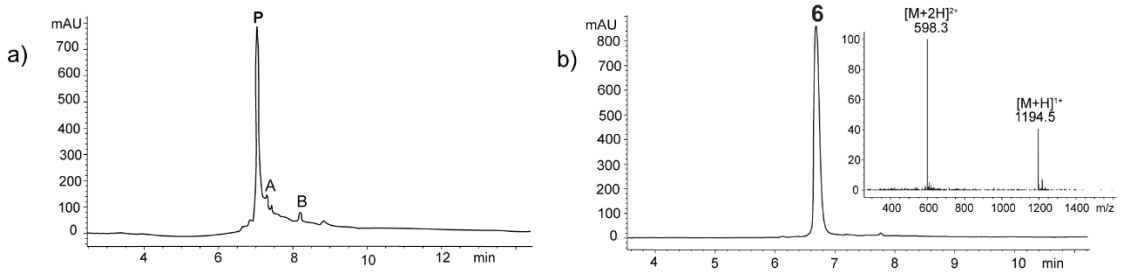
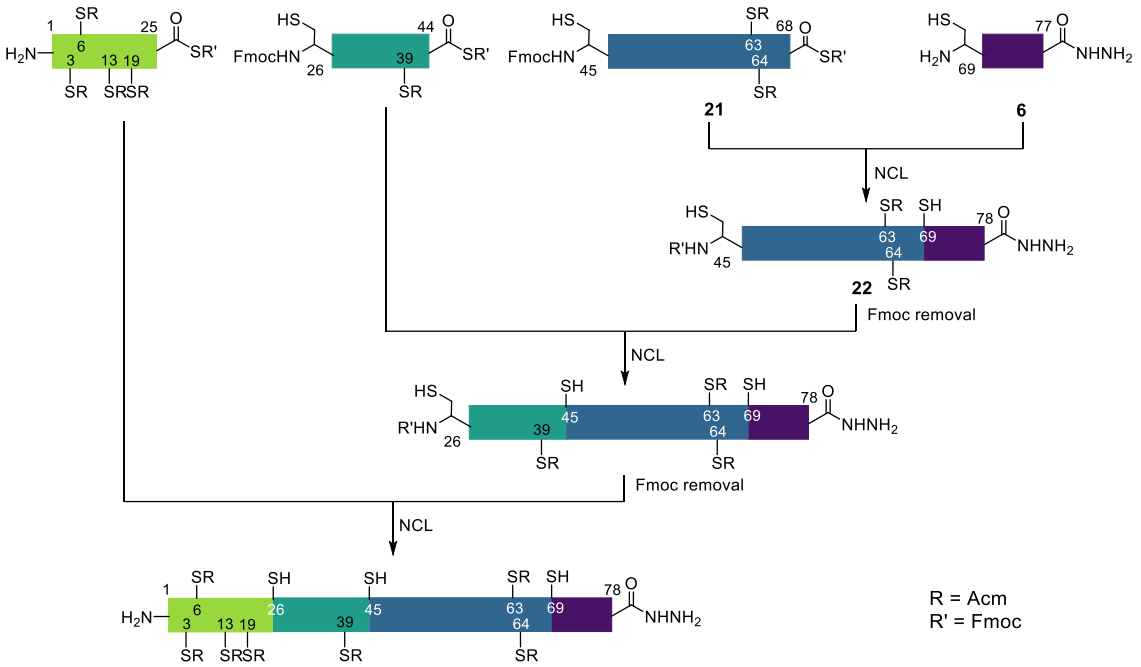


Figure 13 - LC-MS analysis of peptide 6. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 70% of ACN in water + 0.1% of HCOOH in 17 min. a) crude peptide; b) pure peptide. **P**: Product **6**; **A**: $[M\text{-Asn}]^+$.

2.2.3. Ligation of peptides fragments from C- to N-terminus direction

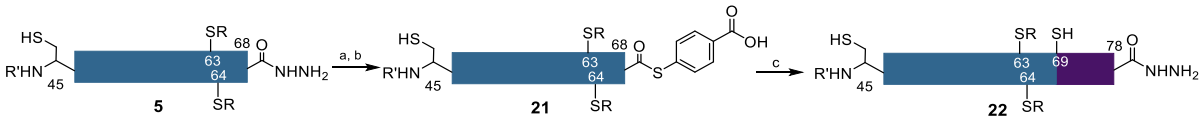
The first NCL approach considered for the synthesis of CD59 was from C- to N-terminus direction using a one-pot process. This strategy allows the sequential ligation of peptides without intermediate purifications[247, 250]. To release the N-terminus after the ligation of the internal fragments **4** and **5**, the methodology proposed by Mandal and coworkers using piperidine was envisioned for the Fmoc removal [250]. The synthesis strategy consisted of three ligations and two deprotection reactions to get the non-glycosylated variant of human CD59 (Scheme 22).



Scheme 22 - Strategy for the one-pot approach for the synthesis of non-glycosylated CD59

Ligation of peptide thioester 21 and peptide 6

The first ligation involved a reaction between fragments **5** (Cys-45 to Leu-68) and **6** (Cys-69 to Ala-78). It required the conversion of the peptide hydrazide **5** into the peptide thioester **21** and the subsequent ligation with the peptide hydrazide **6** (Scheme 23)**.**



Scheme 23 - *in situ* formation of peptide thioester 21 and NCL to obtain peptide 22. Reagents and conditions: a) $NaNO_2$, diazotization buffer pH 3, -10 °C b) MPAA pH 6.8, rt; c) NCL 6 M GdnHCl, 0.1 M NaH_2PO_4, 0.05 M MPAA. 0.015 M TCEP, pH 6.8, not isolated.

The *in-situ* formation of thioester **21** was performed in two steps. First, the C-terminal hydrazide was treated with sodium nitrate at pH 3 to form an acyl azide. Second, a significant excess of 4-mercaptophenylacetic acid (MPAA) is added, and the acyl azide undergoes *in-situ* thiolysis. LC-MS monitoring of the thioester formation revealed the fast conversion of the active azide into the expected product **21** (Figure 14a). The following addition of peptide **6** started the ligation to get peptide **22**. The ligation progress was monitored at reaction times of 0, 3, and 21 h (Figures 14b to 14d). The chromatographic profile showed little formation of product **22**, a large amount of cysteinyl peptide **6,** and disulfide product **6a** resulting from the reaction of **6** and the thiol in the ligation buffer. The oxidation of cysteine seems to play a significant role in avoiding ligation. The oxidized thiol is not nucleophilic and does not undergo the required transthioesterification with **21**, affecting the reaction rate and progress. Efforts to isolate the product **22** using gel filtration resulted in the coelution of **22** and hydrolyzed thioester **21a**.

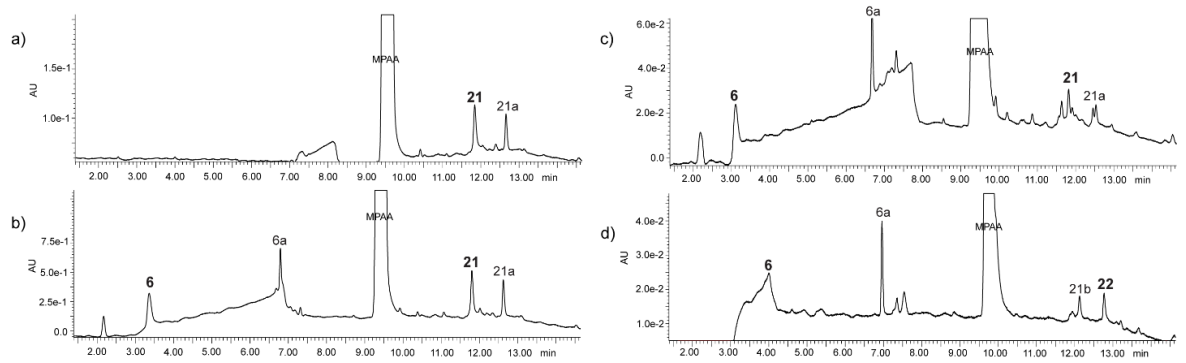


Figure 14 - LC-MS analysis of the reaction progression. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 10 to 50% of ACN in water + 0.1% of HCOOH in 19 min. a) *in situ* formation of thioester **21**, b) chromatogram of the ligation reaction at t = 0 h; c) chromatogram of the ligation reaction at t = 3h; d) chromatogram of the ligation reaction at t = 21 h.

The obtained mixture of **22** and **21a** served for testing the Fmoc removal protocol in an aqueous buffer. Mandal and coworkers proposed the efficient removal of this group using 20% piperidine at pH 11[250]. However, a first experiment using this method delivered a complex mixture of products with incomplete removal of Fmoc from **21**, suggesting potential difficulties in getting a straightforward Fmoc removal after the ligation, primarily due to the high pH that may affect glycosylated peptides having O-glycans. In conclusion, the one-pot C- to N- ligation approach was not appropriate for synthesizing the CD59 glycoprotein. It requires additional product manipulations after ligations that reduce the yields of the process and favor side reactions. Therefore, another strategy was considered for synthesizing CD59 using NCL.

2.3. Strategy III: Synthesis of CD59 N- to C- terminus assembly

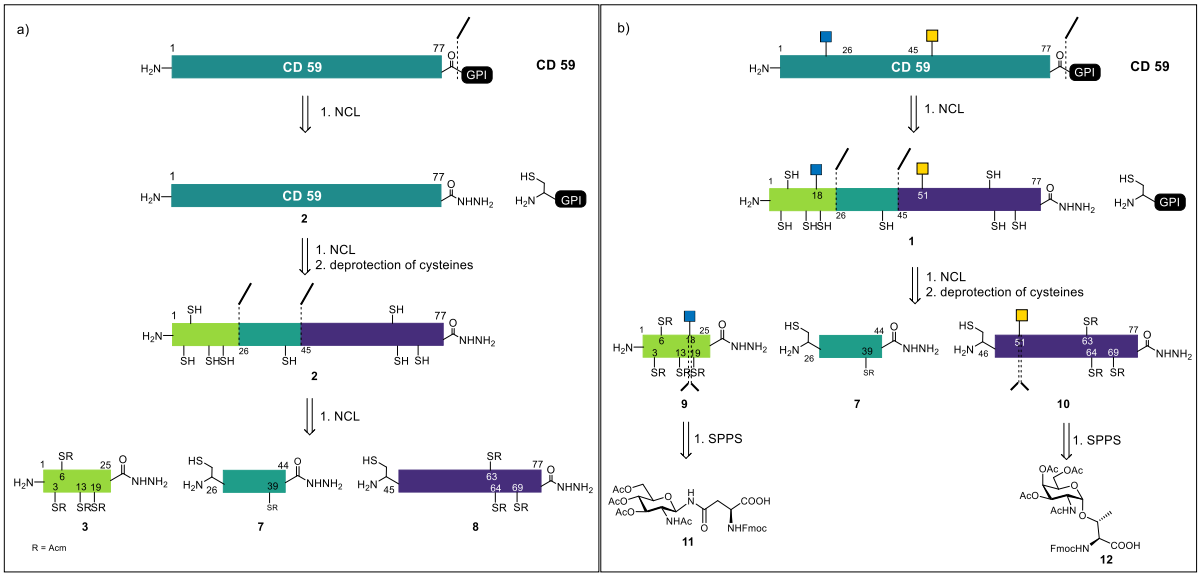
Given the limitations of the previous strategies evaluated for synthesizing CD59, a new approach was considered for a protein synthesis from the N- to C-terminal direction requiring minor manipulations of the ligated fragments. Although this strategy requires purification after each ligation step, pure products enable monitoring the reactions and avoiding side reactions with unreacted fragments.

There was no need for N-terminal protection in terminal fragments in this new strategy, and two new peptide fragments were required. However, considering the difficulties in the ligation of fragments **21** (Cys-45 to Leu-65) and **6** (Cys-69 to Ala-78) and the results obtained in the direct synthesis of non-glycosylated CD59, the synthesis of the extended peptide (Cys-45 to Ala-78) containing both fragments and without affecting the O-glycosylation appeared as the best alternative. Therefore, a new strategy that considered three fragments was evaluated.

2.3.1. Retrosynthetic analysis and strategy design for CD59 synthesis

The retrosynthetic analysis considering the length of the peptides and the glycosylation sites (Asn-18 and Thr-51) showed the first potential disconnection between the GPI and the C-terminus of

the protein again, delivering the glycosylated protein and the modified glycosylphosphatidylinositol bearing a cysteine residue (Scheme 24). Two further disconnections within the protein delivered three peptide fragments for the non-glycosylated variant (Scheme 24a), and one peptide, and two glycopeptides for the glycosylated variant (Scheme 24b)—these glycopeptides required two glycosyl amino acids.



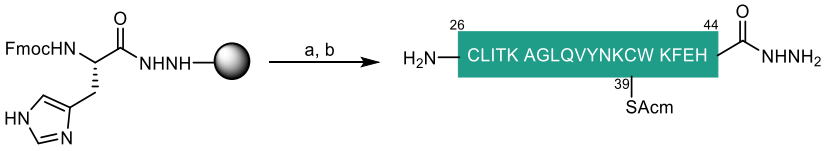
Scheme 24 - Retrosynthesis of non- and glycosylated CD59.

The new strategy involved native chemical ligation of three peptides (or glycopeptide) fragments using Cys-26 and Cys-45 as junction points. In addition, to allow future glypiation via NCL, the C-terminus peptide fragment contains a hydrazide for the further thioester conversion and ligation with a glycosylphosphatidylinositol moiety bearing a cysteine residue.

2.3.2. Solid-phase synthesis of CD59 peptide fragments

The new approach required the synthesis of four new fragments, peptides **7** and **8** and glycopeptides **9** and **10** (Scheme 24).

Synthesis of CD59 peptide fragment 7 (26-44)

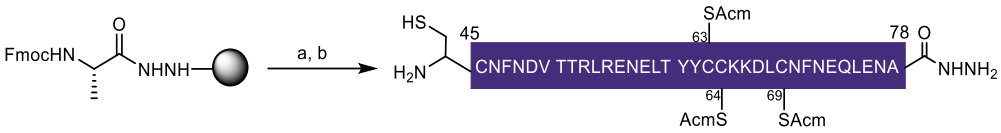


Scheme 25 - Synthesis of peptide hydrazide 7. Reagents and conditions: a) MW-SPPS: coupling Fmoc-Amino acid, *N,N'*-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF b) TFA/TIPS/water (90/5/5)

The synthesis of the fragment Cys-26 to His-44 proceeded straightforwardly, employing the same method and conditions described to obtain **4**. The only additional step was the removal of the Fmoc from the N-terminal residue (Scheme 25). After the synthesis, the peptide was obtained

with a good yield of 27%. The main reason for the yield improvement compared to the yield of peptide **4** was the better solubility of the peptide with the free N-terminus for purification.

Synthesis of CD59 peptide fragment 8 (45-78)



Scheme 26 - Synthesis of peptide hydrazide 8. Reagents and conditions: a) MW-SPPS: coupling Fmoc-Amino acid, *N,N'*-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF b) TFA/phenol/water/thioanisol/EDT (82.5/5/5/5/2.5), rt.

This peptide combined fragments **5** and **6** from the previous strategy**.** Therefore, the elongation of peptide **8** was performed employing the optimized methods for synthesizing peptides **5** and **6**. The synthesis started with a 2-chlorotrityl hydrazine functionalized resin obtained as described before (Scheme 17) and the coupling of the first amino acid Fmoc-Ala-OH employing HATU and DIPEA. MW-SPPS was performed employing 20% piperidine + 0.8% formic acid in DMF for the Fmoc removal (Scheme 26). The first ten amino acids were coupled and deprotected at 75 °C, and the following coupling and deprotection cycles were performed at a lower temperature (50 °C and 25 °C, respectively). The analysis of the crude peptide **8** showed deletion sequences which difficulted the purification. However, the peptide was obtained in an acceptable 10% yield with more than 80% purity (Figure 15).

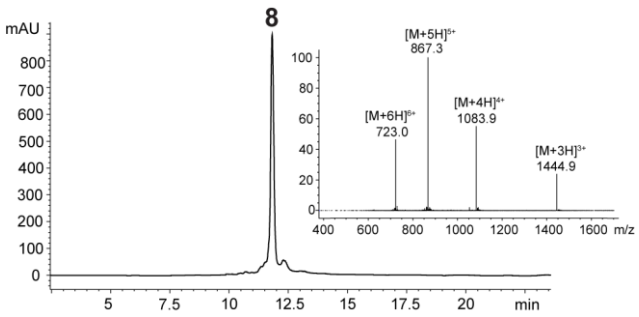
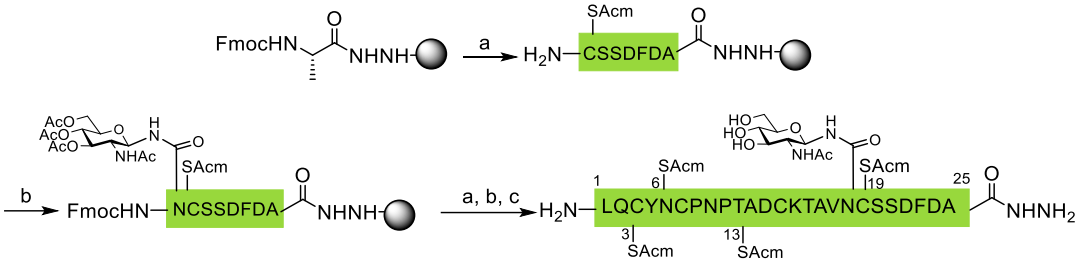


Figure 15 - LC-MS analysis of peptide hydrazide 8. Chromatogram recorded at 214 nm (RP-HPLC with C18 column, gradient: 5 to 60% of ACN in water + 0.1% of HCOOH in 25 min.

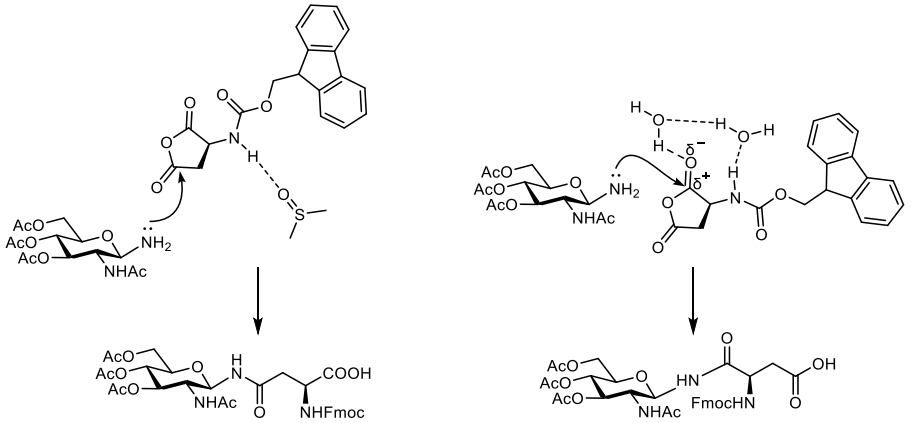
Synthesis of CD59 N-glycopeptide fragment 9 (1-25)



Scheme 27 – Synthesis of glycopeptide fragment 9. Reagents and conditions: a) MW-SPPS: coupling Fmoc-Amino acid, *N,N'*-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF; b) **11**, PyBOP, DIPEA, DMF/DMSO (4:1), rt, overnight; c) 10% $N_2H_4 \cdot H_2O$, THF/MeOH (1:1); d) TFA/phenol/water/thioanisol/EDT (82.5/5/5/5/2.5), rt.

The N-glycopeptide **9** was synthesized by following the optimized protocol for the MW-SPPS of the non-glycosylated fragment **3** (Scheme 27). The synthesis was monitored at several sites (Figure 16). The short peptide fragment (Cys-19 to Ala-25) did not present deletion sequences (Figure 16a).

Building block **11** was synthesized from peracetylated *N*-acetylglucosyl amine, and a succinimide anhydride following previously reported protocols[87, 251]. A difference in regioselectivity was observed in the first attempt to synthesize **11** employing the method reported by Ibatullin *et al.*[251]. The reaction delivered the product of a nucleophilic attack on the α-carbonyl group of Fmoc-aspartic anhydride exclusively, and none of the expected peracetylated glucosamine at the β-carbonyl group[251-252]. Further studies on the reaction concluded that the water content in the DMSO could be responsible for the change in the regioselectivity- The use of a new bottle of anhydrous DMSO delivered the product with the desired regioselectivity. These results showed the possible participation of water in an intermolecular hydrogen bond with the oxygen of the α-carbonyl. This interaction can be strong enough to make the α-carbonyl more electrophilic, therefore, more susceptible to the nucleophilic attack (Scheme 28).



Scheme 28 - A suggested mechanism for the differences in stereoselectivity in the N-Fmoc aspartic anhydride opening in DMSO

Having the proper glycosylated asparagine in hand, the synthesis of **9** continued with the manual coupling of the amino acid employing PyBOP and DIPEA in a solvent mixture DMF/DMSO (4:1).

The phosphonium activator was used in significant excess to fully activate the glycosyl amino acid and did not induce capping at the N-terminus, a common reaction observed when more potent aminium activators like HATU are used. The addition of DMSO was necessary to fully solubilize the glycosylated asparagine. After the overnight reaction, a Kaiser test evidenced incomplete coupling. Thus, a double coupling was performed for additional three hours to obtain complete conversion by Kaiser test and LC-MS analysis after small-scale cleavage (Figure 16b).

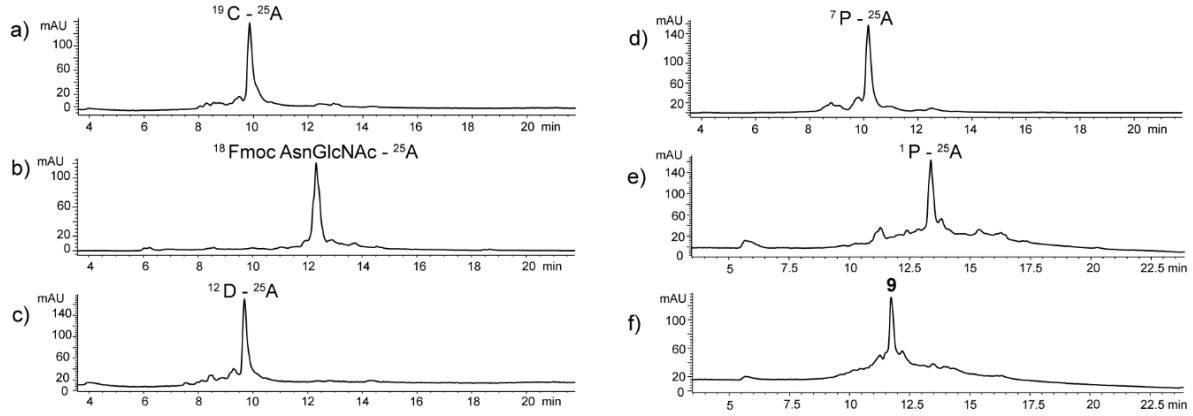
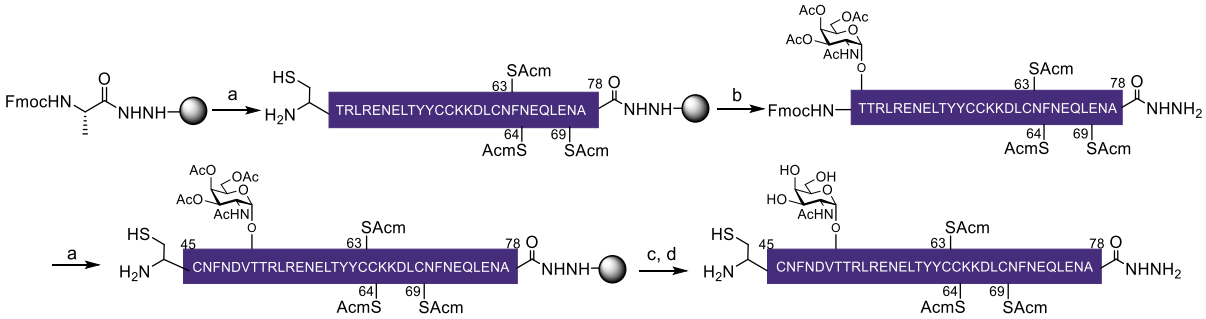


Figure 16 - LC-MS analysis of elongation of the crude N-glycopeptide 9. Chromatogram recorded at 254 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 60% of ACN in water + 0.1% of HCOOH in 25 min (a-d) or 5 to 40% ACN in water + 0.1% of HCOOH in 25 min (e-f); a) Cys-19 to Leu 24; b) FmocAsn((Ac)$_3$-β-D-GlcNAc)-18 to Ala 25 c) Asp-12 to Ala-25; d) Pro-7 to Ala-25; e) 1-25 (Ac)$_3$GlcNAc; f) 1-25 GlcNAc.

Peptide elongation was monitored at Asp-12, Pro-7, and Pro-1 (Figure 16c to 16f). LC-MS analysis showed the formation of the main product, proving that double coupling and capping at selected amino acids diminish the deletion sequences. Furthermore, the aspartimide formation was efficiently minimized by steric hindrance when bulkier moieties were used in the linker and as a side chain protective group for aspartic acid (Fmoc-Asp(OEpe)-OH). Once the elongation was finished, the on-resin deacetylation of glycosylated asparagine was performed employing 10% hydrazine monohydrate[87] in THF/MeOH (1:1) to give a complete conversion after two hours. As a result, the glycopeptide hydrazide **9** was obtained with an acceptable yield of 13% after purification.

Synthesis of CD59 O-glycopeptide fragment 10 (45-78)



Scheme 29 - Synthesis of glycopeptide fragment 10. Reagents and conditions: a) MW-SPPS: coupling Fmoc-Amino acid, *N,N'*-diisopropylcarbodiimide, ethyl cyano(hydroxyimino)acetate, DMF, deprotection: 20% piperidine in DMF; b) **12**, PyBOP, DIPEA, DMF:CH_2Cl_2 (1:1), rt, overnight; c) 10% $N_2H_4 \cdot H_2O$, MeOH; d) TFA/phenol/water/thioanisol/EDT (82.5/5/5/5/2.5), rt.

The O-glycopeptide synthesis required special attention due to the possibility of a β-elimination of the glycan. In this base-catalyzed side reaction, the glycosylated threonine is converted into the β-methyl dehydroalanine, resulting in the glycan removal from the peptide[135]. The β-elimination reaction is commonly used to release O-linked oligosaccharides and allow the structural analysis of glycopeptides such as glycosylation sites and glycan structure determinations[253-254].

The O-glycopeptide was prepared employing the optimized method for synthesizing the peptide fragment **8** (Scheme 29). The MW-SPPS elongation was performed first until Thr-52, and the quality of the peptide was monitored by LC-MS (Figure 17a). This process delivered a product having the desired peptide as the main component. Then, the synthesis continued with the manual coupling of the glycosylated threonine **12** (provided by Dr. Götze). Glycosyl amino acid **12** was coupled by employing the same method as the glycosylated asparagine. The phosphonium reagent PyBOP completely activated the glycosyl amino acid and delivered complete coupling after the overnight reaction (Figure 17b). Next, the Fmoc removal of the glycosylated threonine was performed manually at room temperature. Since the β-elimination is also a base-catalyzed side reaction, 0.8% of formic acid was added to the deprotection solution to suppress it. Monitoring after each deprotection step revealed an excellent effect of the mixture, and the glycan was intact up to completing the synthesis (Figures 17c to 17f).

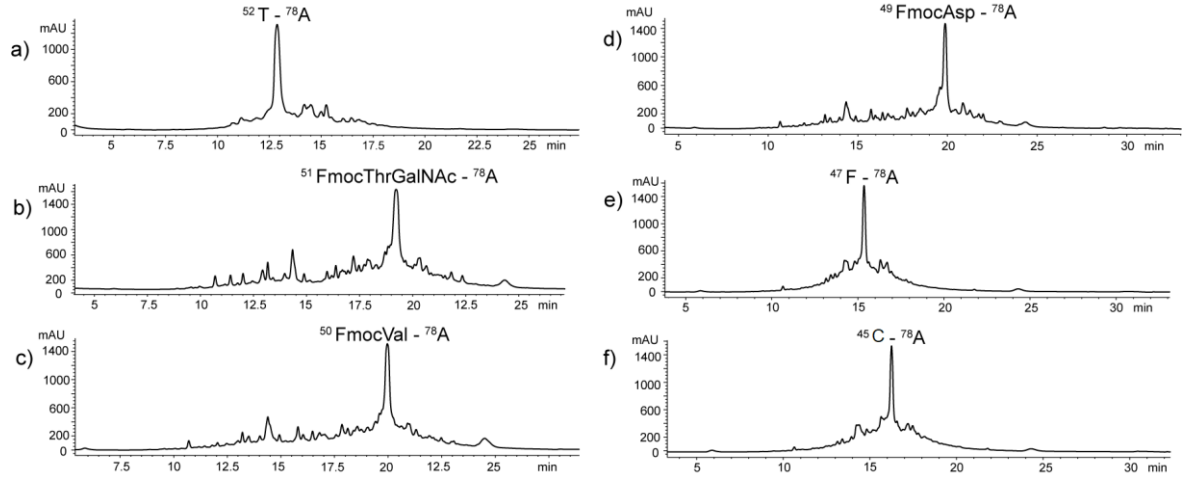


Figure 17 - LC-MS analysis of elongation of the crude glycopeptide 10. Chromatogram recorded at 254 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 40% of ACN in water + 0.1% of HCOOH in 25 min a) Thr-52 to Ala-78 b) FmocThr((Ac)$_3$-β-D-GalNAc)-51 to Ala-78 c) Val-50 to Ala-78; d) FmocAsp-49 to Ala-78; e) Phe-47 to Ala-78; f) 1-25 GlcNAc.

Particular attention was given to the removal of *O*-acetyl groups. On-resin deacetylation of carbohydrates is commonly completed under Zemplen conditions[158]. However, the treatment of O-peptides with this strong base might induce β-elimination. Milder on-resin deacetylation of sensitive sulfated glycans[155] and other methods are reported for on-resin deacetylation of glycopeptides[255]. Thus, several bases (LiOH, NaOH, NH_3, and $N_2H_4 \cdot H_2O$) were screened using methanol as a solvent to determine the most suitable deacetylation conditions for the O-glycopeptide. The inspiration for employing methanol arose from previous reports of Singh et al

where they concluded that the use of protic solvent allows hydrogen bonding, reducing the base reactivity and the acidity of the α-proton[255].

Few milligrams of peptide-resin were treated with each base, and the outcome of the reaction was analyzed by LC-MS after two hours. The chromatogram for the acetylated glycopeptide (peak A) is shown in Figure 18a. The treatment with LiOH (Figures 18b and 18e) showed poor to no conversion and the reaction with a more reactive base NaOH evidenced a faster reaction; however, the β-elimination side product (peak C) was also observed (Figure 18c). Treatment with 7 M ammonia showed faster deacetylation (peak P), some partial deacetylation (peak B), and a high-rate formation of side product (peak C) (Figure 18d). This outcome might be associated with the high ammonia concentration. Complete deprotection with a negligible β-elimination side product was obtained when treating the O-glycopeptide with 10% hydrazine monohydrate. Thus, this methodology was used without further optimization (Figure 18f).

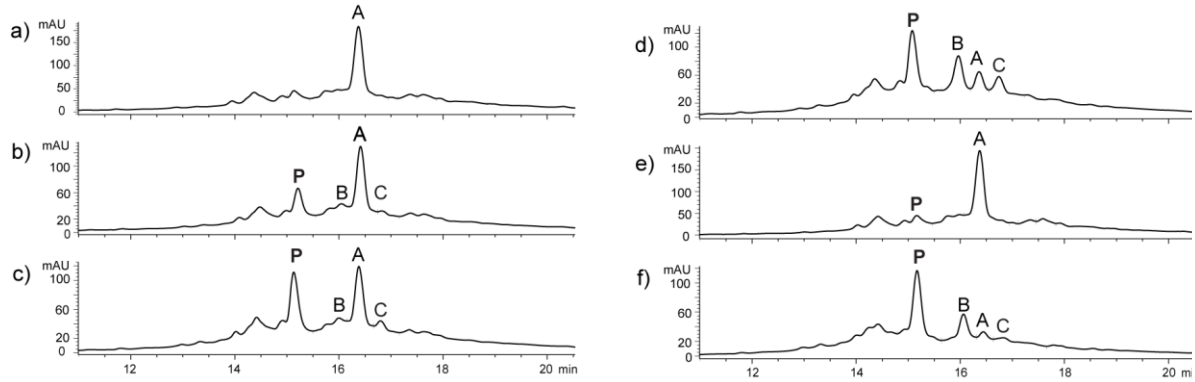
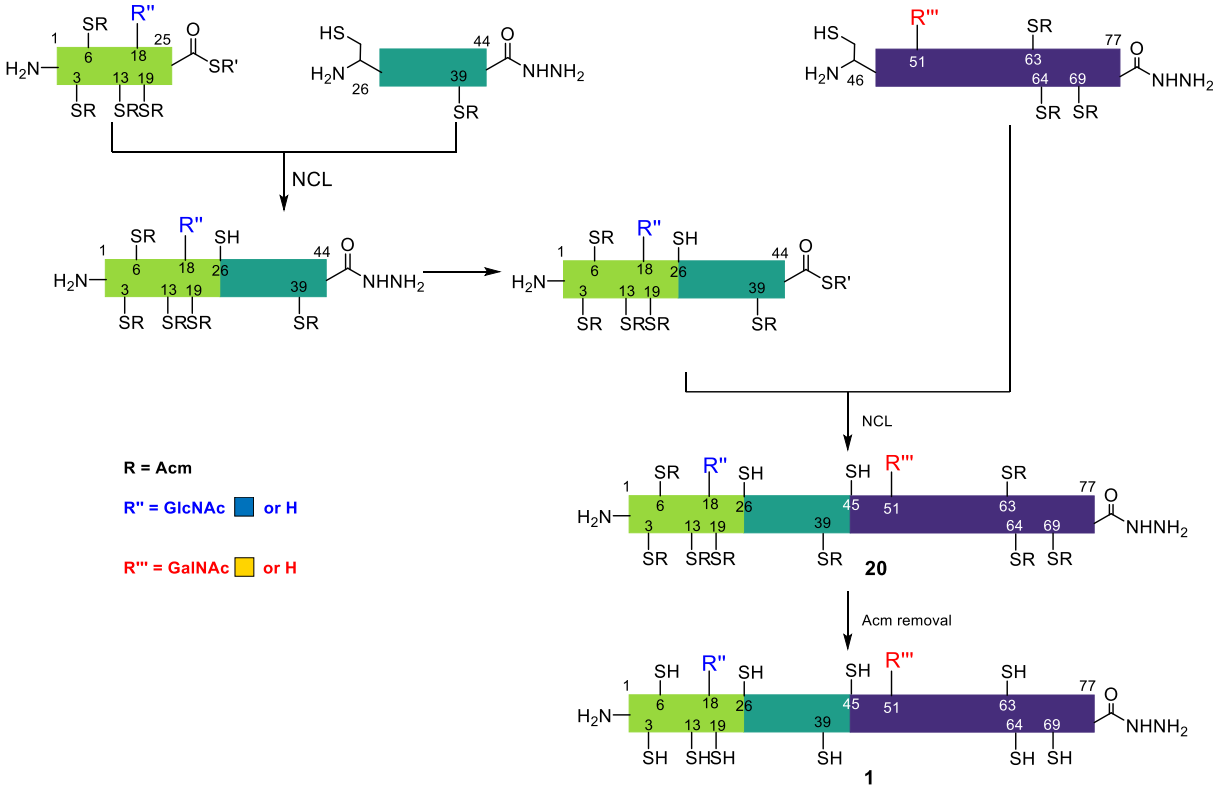


Figure 18 - LC-MS analysis of deacetylation of the O-glycopeptide crude fragment (45-78) employing different bases. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 40% of ACN in water + 0.1% of HCOOH in 25 min. a) starting material; b) LiOH 10 mM in MeOH; c) NaOH 10 mM in MeOH; d) 7 M NH_3 in MeOH; e) LiOH 10 mM in MeOH pH 9; f) 10% $N_2H_4 \cdot H_2O$ in MeOH; **P**: Product **10**; **A**: glycopeptide bearing acetylated galactosamine; **B**: glycopeptide bearing partially acetylated galactosamine; **C**: β-elimination side product.

After optimizing the synthesis and on-resin deacetylation, the deprotected O-glycopeptide was cleaved from the resin using cocktail K to obtain **10** with an acceptable 10% yield. Finally, having all the required peptide fragments in hand, the synthesis of the non-glycosylated and glycosylated variants of the human CD59 was evaluated.

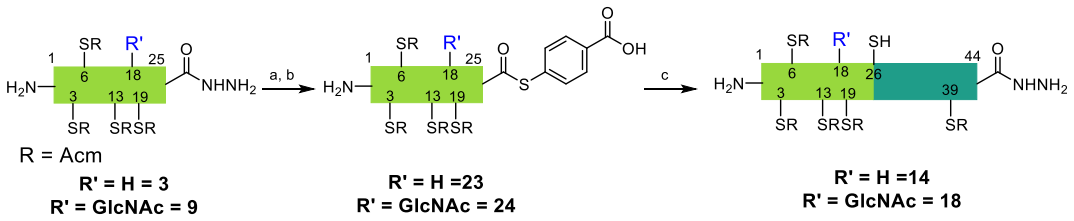
2.3.3. Ligation of peptides fragments from N- to C- terminus direction

The second approach considered for the synthesis of CD59 employing NCL was in the N- to C-direction. This strategy was selected because it requires fewer manipulations (deprotection) when compared with one-pot C- to N-terminus approach. The strategy required two ligation reactions to assemble the non-glycosylated and glycosylated variants of the human CD59 (Scheme 30).



Scheme 30 - Strategy for N- to C- assembly of a glycosylated variant of CD59

Ligation peptide fragments Leu-1 to Ala-25 or N-Leu-1 to Ala-25 and peptide 7



Scheme 31 - *in situ* formation of peptide thioester 13 or 17 and NCL to obtain peptide hydrazide 14 or glycopeptide 18. Reagents and conditions: a) $NaNO_2$, diazotization buffer pH 3, -10 °C b) MPAA pH 6.8, rt, c) **7**, NCL, 6 M GdnHCl, 0.1 M NaH_2PO_4, 0.05 M MPAA. 0.01 M TCEP, pH 6.8, not isolated.

The first ligation involved the reaction of the two fragments closer to the N-terminus. Using separate Eppendorf tubes, this ligation was performed simultaneously for each protein (glycosylated and non-glycosylated variants). The process required the conversion of the peptide hydrazide **3** and glycopeptide hydrazide **9** into the corresponding peptide thioesters **13** and **17**, respectively, and the subsequent ligation of these thioesters with the peptide hydrazide **7** (Scheme 31).

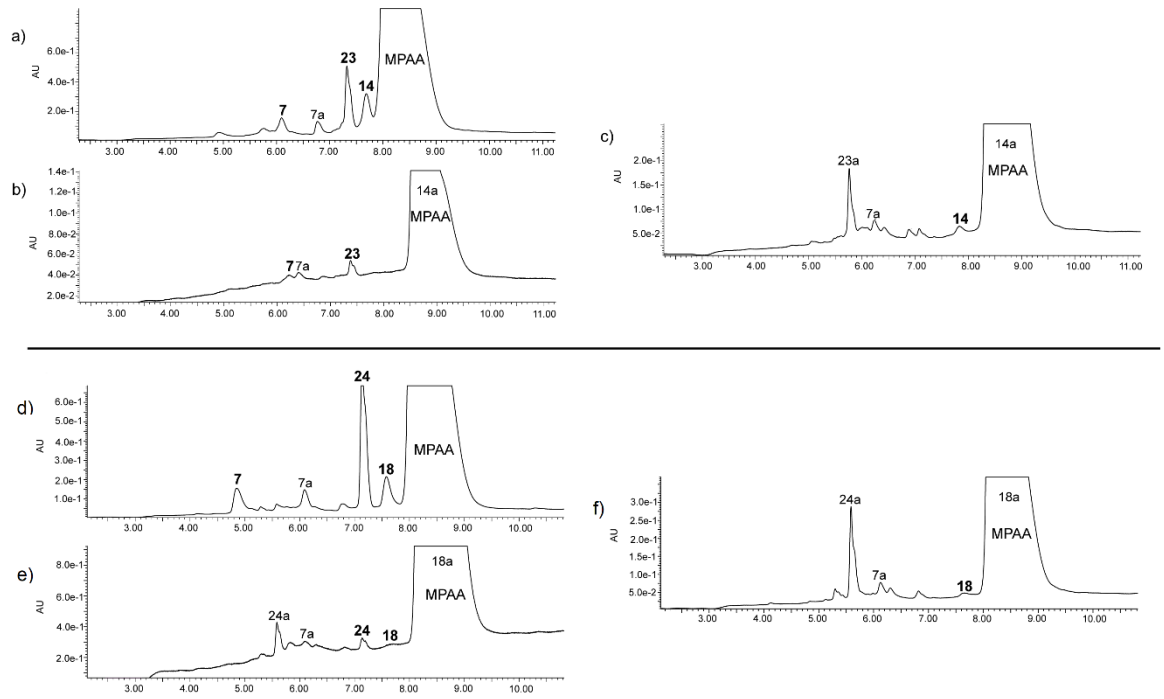
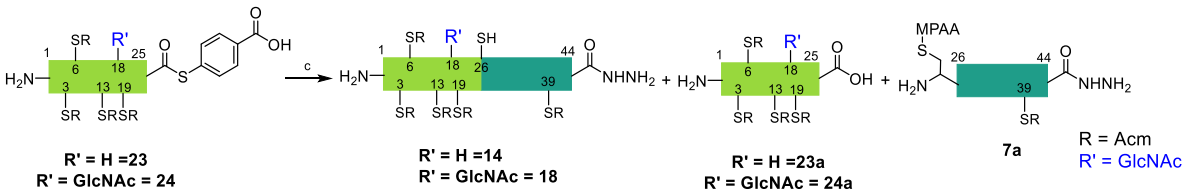


Figure 19 - LC-MS analysis of the reactions progression. Chromatograms recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 10 to 50% of ACN in water + 0.1% of HCOOH in 19 min. Up: a) chromatogram of the ligation reaction at t = 1 h; b) chromatogram of the ligation reaction at t = 3 h; c) chromatogram of the ligation reaction at 21 h; Down: a) chromatogram of the ligation reaction at 1 h; b) chromatogram of the ligation reaction at t = 3 h; c) chromatogram of the ligation reaction at t = 21 h.

The formation of the thioesters was performed in two steps, *in situ* activation with sodium nitrite at pH 3 and thiolysis employing MPAA at pH 7. These reactions were monitored by LC-MS and revealed the rapid formation of the expected thioesters **13** and **17**. The ligation proceeded by adding the peptide hydrazide **7** to each reaction Eppendorf and the progress was monitored at t = 0 h, 3 h, and 17 h (Figure 19). The chromatographic profile after one hour showed the fast formation of the product **14** or **18** and disulfide bond formation between the cysteinyl peptide and the thiol present in the ligation buffer (MPAA), giving **7a**. Analysis after three hours of reaction evidenced a shift in the retention time for the products. A detailed analysis of the mass signals also evidenced a disulfide bond between the products and MPAA (Figure 19c peak **14a** or Figure 19f peak **18a**). Since the reaction was progressing, the mixture was stirred overnight. LC-MS analysis for 21 h reaction revealed that most of the peptide thioesters **23** and **24** hydrolyzed to give **23a** and **24a** (Scheme 32). The reaction was stopped, and the reaction mixtures were separated using gel filtration. However, the products could not be isolated in a significant amount.

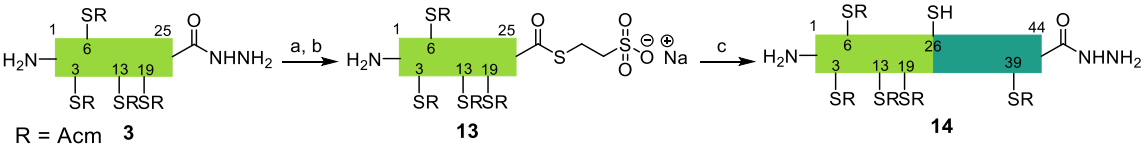


Scheme 32-Ligation products between peptide 23 or glycopeptide 24 and cysteinyl peptide7

The ligation with the peptide **3** and glycosylated peptide **9** to fragment **7** (Scheme 31) revealed several aspects that required modification. First, the high rate of hydrolysis showed that the aromatic thiol employed to obtain the thioesters was too reactive and required a careful pH control when adjusting it from pH 3 to pH 6.8. In addition, the formation of the disulfide bonds between the cysteinyl peptide and MPAA evidenced two main factors. The concentration of TCEP in the ligation buffer was too low, or the samples required a reduction prior to LC-MS analysis. Finally, the purification method required optimization and had to be changed. In conclusion, the optimization of the ligation reaction envisaged the following aspects: the use of an aliphatic thiol (sodium 2-mercaptoethanesulfonate – MESNa) in the thiolysis reaction[256], the increasing of the TCEP concentration in the ligation mixture from 0.010 M to 0.015M, and incubation of the samples with TCEP before injection into the LC-MS and the migration from gel filtration to RP-HPLC for purification.

Assembly of the non-glycosylated variant of human CD59

The synthesis of the non-glycosylated variant of the protein was performed considering optimizing the ligation process. Special attention was given to the thioester formation for fragment Leu-1 to His-44 and the behavior of peptides under the NCL conditions.



Scheme 33 - Formation of peptide thioester 13 and NCL to obtain peptide hydrazide 14. Reagents and conditions: a) $NaNO_2$, diazotization buffer pH 3, -10 °C b) MESNa pH 6.8, rt, 57% over two steps; c) **7**, NCL, 6 M GdnHCl, 0.1 M NaH_2PO_4, 0.05 M MPAA, 0.015 M TCEP, pH 6.8; 37%.

The assembly of CD59 started with the ligation of the peptide thioester **13** and the peptide hydrazide **7** (Scheme 33). The peptide hydrazide **3** was treated with sodium nitrite and MESNa to obtain the peptide thioester **13** in a 57% yield after purification by RP-HPLC. The main reason for adding this purification step was the need to have better control of the thioester amount and the equivalents used in a ligation reaction. At the same time, this step removed side products derived from the poor control of the reaction pH; even if the pH was not correctly adjusted, the thioester was separated from the hydrolysis product. Using the pure thioester **13** the pH for the reaction was controlled by the ligation buffer required to dissolve the peptide fragments, which was carefully adjusted to pH=6.8 prior to addition to the peptides. The progress of the reaction was analyzed by LC-MS (Figure 20). A clean profile was observed at time zero when the reaction started (Figure 20a) and when the reaction finished (Figure 20b). The excellent outcome of the reaction was attributed to the use of a less reactive thioester and the precise control of the pH. The reaction mixture was incubated with 20 equivalents of TCEP for 30 min to avoid problems in the purification. After preparative RP-HPLC, the 44-residue glycopeptide hydrazide **14** was isolated with an acceptable 37% yield.

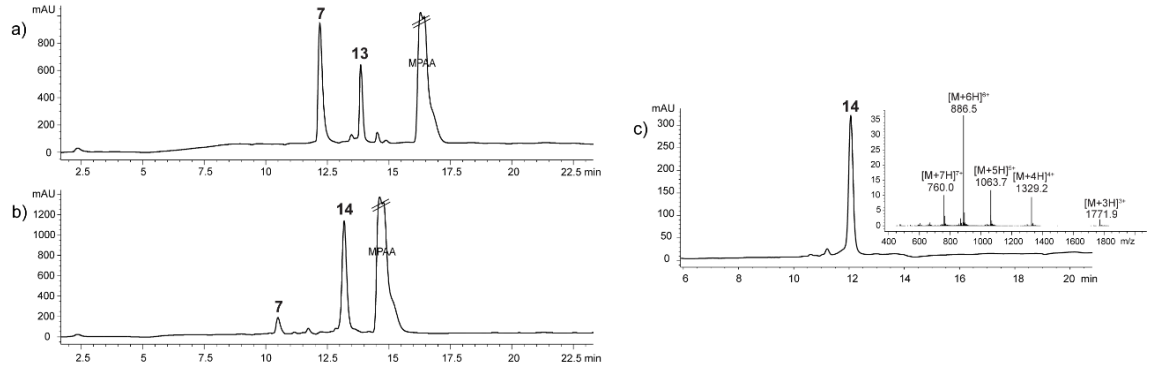
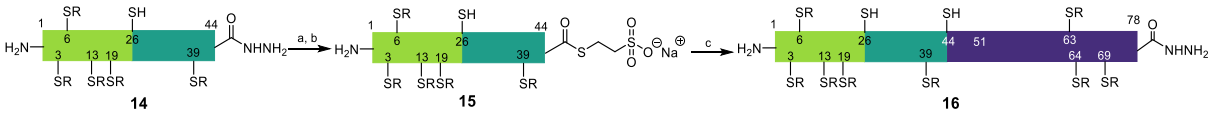


Figure 20 - LC-MS analysis of the reaction progression. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 60% of ACN in water + 0.1% of HCOOH in 25 min. a) chromatogram of the ligation reaction at t = 0 h; b) chromatogram of the ligation reaction at t = 14 h; c) chromatogram and mass spectra of product **14**.

The use of thioester **13** delivered a clean ligation with peptide **7**. However, the yield of isolated peptide hydrazide **14** was lower than expected, mainly due to the loss of material during the RP-HPLC purification. Only one ligation was missing to complete the CD59 synthesis. Therefore, the *in-situ* thioester formation was envisaged for the second ligation reaction (Scheme 34).



Scheme 34 - *in situ* formation of peptide thioester 15 and NCL to obtain a partially protected non-glycosylated variant of human CD59 16. Reagents and conditions: a) $NaNO_2$, diazotization buffer pH 3, -10 °C b) MESNa pH 6.8; c) **8**, NCL, 6 M GdnHCl, 0.1 M NaH_2PO_4, 0.05 M MPAA, 0.015 M TCEP, pH 6.8 rt, not isolated.

The synthesis of the thioester **15** following the described steps delivered a clean reaction according to the LC-MS monitoring (Figure 21a). The following pH adjustment and addition of peptide **8** started the ligation reaction. However, the lack of a suitable pH electrode hampered the determination of the reaction pH for the thiolysis and ligation reactions. Therefore, the pH setting was followed using pH paper. LC-MS analysis after 24 hours showed a complex mixture containing the starting cysteinyl peptide **8** and thioester **15**, hydrolyzed thioester **15a,** and the expected non-glycosylated variant of CD59 **16** (Figure 21c). The hydrolysis of **15** was attributed to the difficulties in adjusting pH 6.8 for the thiolysis.

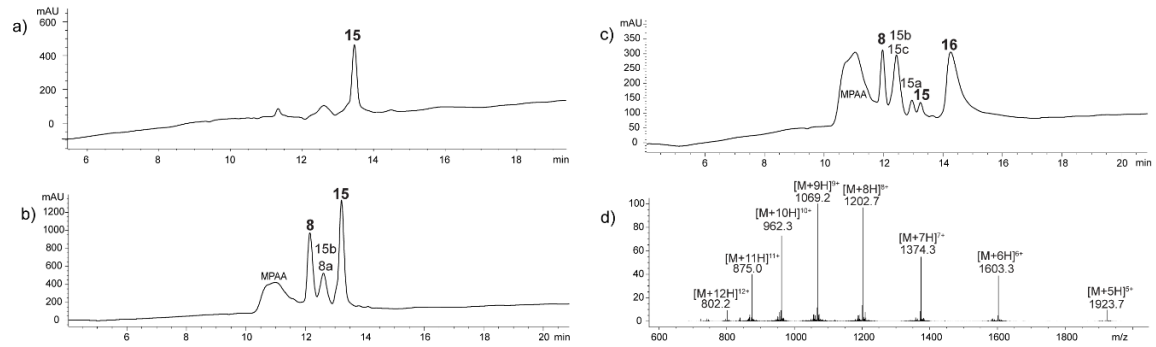
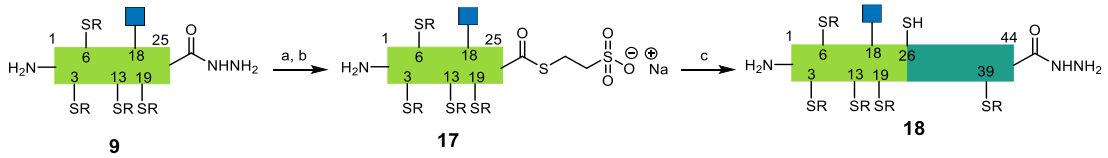


Figure 21 - LC-MS analysis of the reaction progression. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 15 to 55% of ACN in water + 0.1% of HCOOH in 30 min. a) *in situ* peptide thioester formation; b) chromatogram of the ligation reaction at t = 0 h; c) chromatogram of the ligation reaction at 24 h d) mass spectra of the peak corresponding to the product.

Since the thioester was formed *in situ,* a large amount of solid TCEP was required to quench the high thiol concentration present in the reaction mixture (100 equivalents). A complete reduction was observed after incubation for 30 min. The purification of **16** was evaluated using RP-HPLC. However, this was not possible due to the poor product solubility in water acetonitrile mixtures. Further attempts for purification using SEC did not deliver the pure protein and required additional passes until completely pure.

This last strategy was successful in delivering the desired non-glycosylated variant of CD59. The developed methods showed to be efficient for synthesizing the peptide fragments and intermediates. Further studies are still required to improve product purification. However, the methodology optimization and the protocols used to get **16** were applied to assemble the glycoprotein.

Synthesis of the glycosylated variant of human CD59



Scheme 35 - Formation of peptide thioester 17 and NCL to obtain peptide 18. Reagents and conditions: a) $NaNO_2$, diazotization buffer pH 3, -10 °C b) MESNa pH 6.8, rt; c) **7**, NCL, 6 M GdnHCl, 0.1 M NaH_2PO_4, 0.05 M MPAA, 0.015 M TCEP, pH 6.8, 58% over two steps**.**

Following the optimized protocol for the ligation reactions used to get **16**, the synthesis of the glycosylated variant of the protein started with the treatment of glycopeptide hydrazide **9** with sodium nitrite and MESNa to deliver the glycopeptide thioester **17** in 58% yield after RP-HPLC purification (Scheme 35). Next, the thioester **17** was dissolved in the ligation buffer at a pH of 6.8, and the peptide hydrazide **7** was added. The progress of the reaction was followed by LC-MS and is presented in Figure 22.

The ligation proceeded similarly as in the non-glycosylated variant. LC-MS monitoring after one hour of reaction showed already formation of the product **18** (Figure 22b) and complete ligation after 14 h (Figure 22c). Elongation of the reaction time to 17 h did not improve the outcome of the process (Figure 22d).

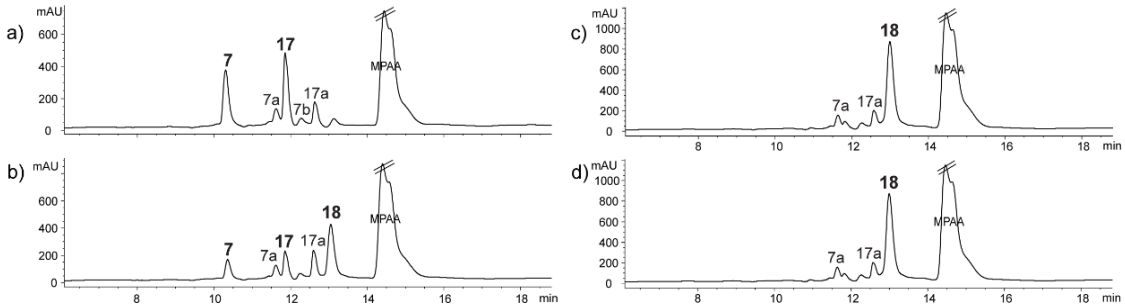
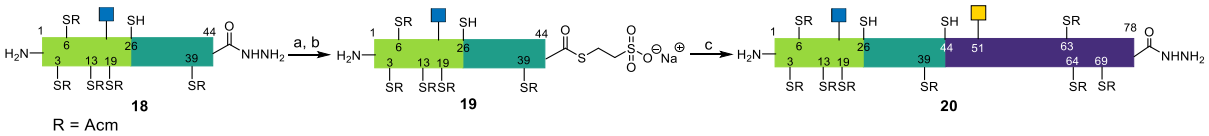


Figure 22 - LC-MS analysis of the reaction progression. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 60% of ACN in water + 0.1% of HCOOH in 25 min. a) chromatogram of the ligation reaction at t = 0 h; b) chromatogram of the ligation reaction at t = 14 h b) chromatogram of the ligation reaction at t = 17 h.

The addition of 20 equivalents of TCEP reduced the oxidation products in the mixture and allowed the RP-HPLC purification of the expected glycopeptide hydrazide **18**, which was isolated in 46% yield with more than 85% purity. With the 44-residues N-glycopeptide, the ligation of **18** with the O-glycopeptide was envisioned using the same protocol as the non-glycosylated variant.



Scheme 36 - *In situ* formation of glycopeptide thioester 19 and NCL to obtain a partially protected glycosylated variant of human CD59 20. Reagents and conditions: a) $NaNO_2$, diazotization buffer pH 3, -10 °C b) **10**, MESNa pH 6.8, rt, 5% over two steps.

The last ligation started with the synthesis of N-glycopeptide thioester **19** using the optimized conditions (Scheme 36). LC-MS analysis showed the formation of the N-glycopeptide thioester **19** without side reactions and any significant difference compared with the synthesis of the non-glycosylated thioester **15** (Figure 23a). Next, O-glycopeptide **10** was added to thioester **19** and started the ligation reaction. Analysis of the reaction progress showed that glycopeptide **10** was coeluting with the thiol (MPAA), mainly because the O-glycan attachment increases the polarity of the fragment (Figure 23b). Monitoring after 24 h indicated the successful formation of the glycoprotein **20** as well as a mixture of subproducts containing hydrolyzed N-glycopeptide thioester **19a,** O-glycopeptide **10**, and a byproduct that could not be identified **19b** (Figure 23c)**.**

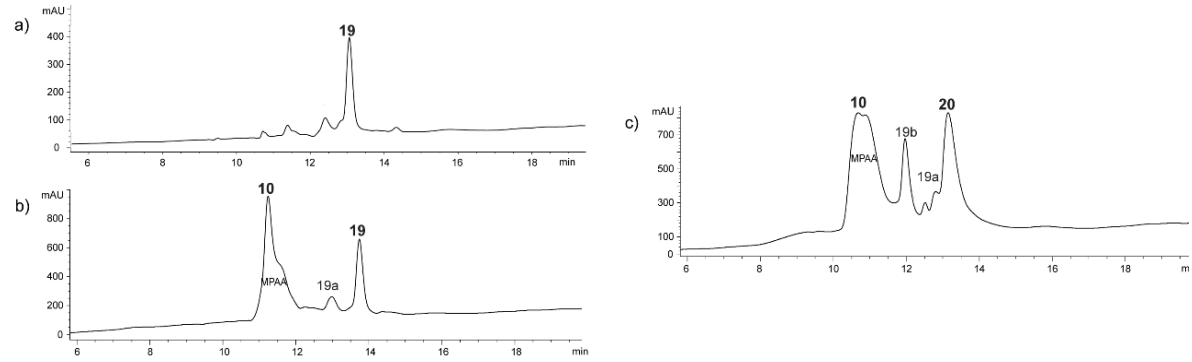


Figure 23 - LC-MS analysis for the synthesis of glycosylated CD59. Chromatograms were recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 15 to 55% of ACN in water + 0.1% of HCOOH in 30 min. a) chromatogram of the *in situ* formation of peptide thioester **19**; b) chromatogram of the ligation reaction at t = 0 h; c) chromatogram of the ligation reaction at 24 h.

Purification of the product was performed by RP-HPLC after TCEP reduction of the mixture. In contrast to the non-glycosylated, purification of the desired glycoprotein was possible by this method and delivered **20** in isolated form in 5% yield together with some mixed fractions. The LC-MS analysis of the pure fraction is presented in Figure 24. The outcome of the process showed that the solubility improvement derived from the protein glycosylation also favored the purification step.

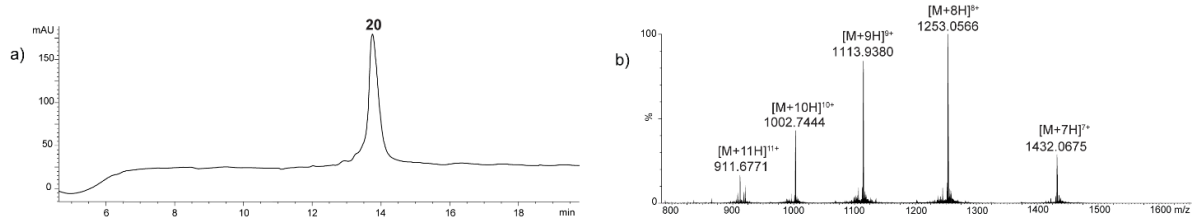


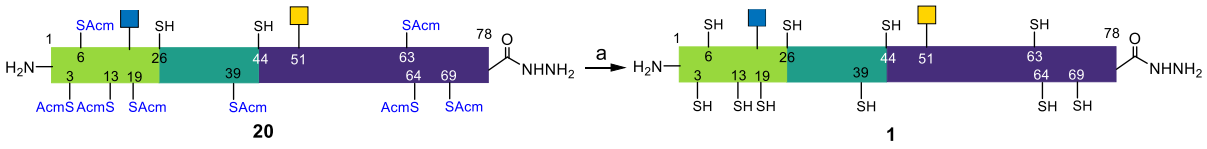
Figure 24 - LC-MS analysis of the partially protected glycoprotein CD59. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradient: 5 to 60% of ACN in water + 0.1% of HCOOH in 25 min. a) chromatogram of pure glycosylated CD59 **20**; b) ESI-MS spectrum of the isolated glycoprotein **20**.

There were no significant differences in the reaction and ligation outcome to the synthesis of non- and glycosylated variants of the human CD59 by NCL. The conversion of the peptide hydrazide into the thioester and the ligation reactions showed similar results and successfully delivered the desired products. This similar behavior was explained by the small glycan size and the similar size and properties of both peptides and glycopeptides. This process displayed glycosylated amino acids having monosaccharides as excellent building blocks for transferring methods in protein and glycoprotein chemical synthesis.

Deprotection of the glycosylated variant of human CD59

The last step to complete the synthesis of the CD59 glycoprotein concerned the removal of the Acm groups. Although this group has been widely used as a permanent protective group for internal cysteines, the removal of Acm remains challenging[247]. Therefore, several deprotection reagents such as silver acetate[257-258], mercury acetate[94], and palladium complexes[79, 93] have been employed in protein synthesis.

Due to the high number of protected residues in glycoprotein **20**, a highly efficient method was required to assure complete Acm removal. Based on previous reports describing problems when using silver or mercury acetate for Acm-removal, a method involving Pd^{II}, specifically $PdCl_2$ [79, 93], was considered the most reliable strategy for deprotection of CD59 glycoprotein (Scheme 37).



Scheme 37 – Deprotection of cysteines for glycoprotein 20. Reagents and conditions: a) 6 M GdmHcl, 0.2 M NaH_2PO_4, $PdCl_2$ (30 equiv), pH 6.9, 4 h, rt.

An initial screening employing the peptide fragment **3** with four Acm groups (Leu-1 to Ala-25) was used to determine the best conditions for the deprotection. Efficient and complete removal of the Acm groups was obtained by treatment with palladium chloride after 2 hours. Subsequently, these conditions were used to remove the Acm group from the glycoprotein in the mixed fractions obtained during the RP-HPLC purification (Figure 25).

The progress of the reaction was made following the peak corresponding to the Acm-protected glycoprotein **20** (Figure 25a). After 2 h of reaction, the palladium in a reaction aliquot was quenched by adding significant excess of DTT to avoid poor ionization in the MS, as described by Unverzagt and coworkers[79]. The LC-MS analysis showed that contrary to the efficient Acm removal observed with peptide **3**, the reaction was not complete and starting material **20** and product **1** was present in the mixture (Figure 25b). Therefore, it was stirred for an additional 2 h (Figure 25c). Complete Acm removal was observed after 4 h of reaction. The need for longer reaction time was associated with a higher number of cysteines in **20**. The initial experiment with peptide **3** was performed to remove four Acm groups, while in the glycoprotein **20**, eight Acm groups had to be removed.

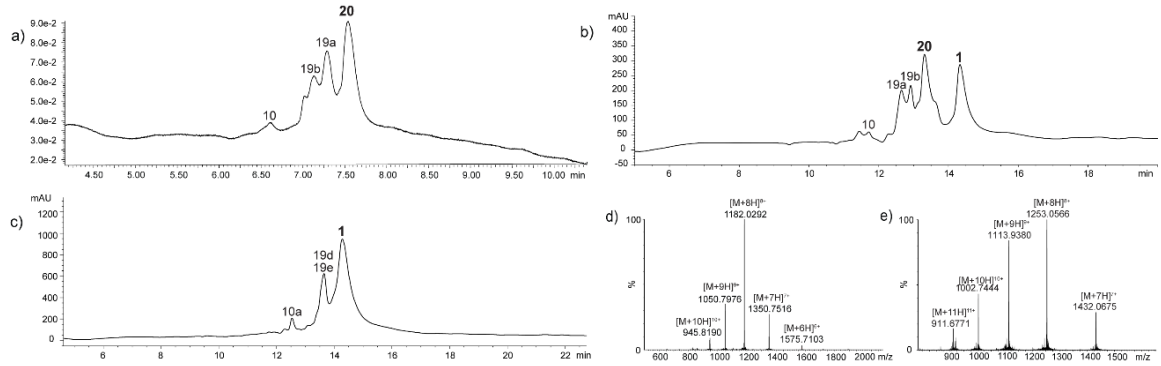


Figure 25 - LC-MS analysis of the deprotection reaction progression. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradients: 5 to 70% of ACN in water + 0.1% of HCOOH in 15 min (a) and 5 to 60% of ACN in water + 0.1% of HCOOH in 25 min (b, c, and d); a) mixed fractions containing glycosylated CD59; b) 2 h reaction; c) 4 h reaction d) MS spectra of deprotected CD59 e) MS spectra of Acm-protected CD59.

Having proved that the Acm-removal employing $PdCl_2$ was a suitable method for the final deprotection of the glycoprotein in the mixed fractions, the reaction was evaluated on a small scale using a pure fraction of **20**. LC-MS monitoring after 4 h showed a mixture of the fully protected and fully deprotected glycoprotein (Figure 26b).

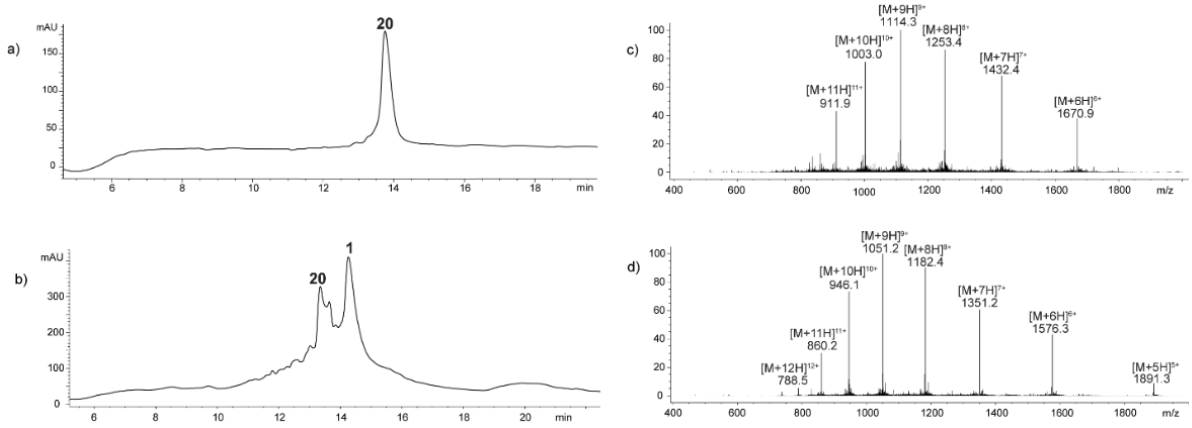


Figure 26 - LC-MS analysis of the deprotection reaction progression. Chromatogram recorded at 214 nm after small-scale cleavage (RP-HPLC with C18 column, gradients: 5 to 60% of ACN in water + 0.1% of HCOOH in 25 min a) pure Acm-protected glycoprotein; b) 4 h reaction; c) MS spectra of protected CD59; d) MS spectra of Acm-target glycoprotein **1**.

The reaction outcome was associated with the solubility problems by dissolving the starting material **20** in the small volume used for the reaction (29 µL), which was not enough to dissolve the protected glycoprotein and bring the protected glycoprotein in contact with the solution containing $PdCl_2$.

These results showed the potency of the developed strategy to obtain a complex glycoprotein, such as CD59. Although the final product was not isolated, several processes and methods were optimized, and a versatile methodology was developed and employed to assemble glycosylated and non-glycosylated variants of the human glycoprotein CD59.

2.4. Conclusions and outlook

A general methodology using NCL was designed and employed for the synthesis of non-glycosylated and glycosylated variants of the human glycoprotein CD59. The primary sequence of the CD59 contains 77 amino acids and is modified with two glycosylation sites on the asparagine at position 18 and threonine at position 51. Different strategies were investigated to prepare the peptides and glycopeptides required to synthesize two protein variants.

Three strategies were designed and tested for the assembly of the non-glycosylated CD59. A direct MW-SPPS synthesis of the whole protein sequence failed and led to a complex mixture of products. This mixture resulted from a combination of factors, including the sequence length (78 residues) and some difficulties in ensuring complete couplings during the elongation, which yielded several deletion sequences. The presence of side reactions also occurred and contributed to the complexity of the synthesis, especially aspartimide formation. These results suggested the need for another strategy, such as NCL, to access the glycoprotein CD59 by chemical synthesis.

NCL strategies were selected to access CD59 chemically. An NCL strategy involving a C- to N ligation approach required the synthesis and the ligation of four peptide fragments. Notably, an optimization was required to synthesize the peptides due to aspartimide formation and deletion sequences. The improvements included preventing aspartimide formation by using bulkier linkers bound to the resin, a bulk protective group for the Asp side chain, and the addition of formic acid in the solutions for Fmoc-removal. In addition, the peptide synthesis was also optimized by introducing double coupling cycles and capping at specific sites to overcome incomplete couplings and deletion sequences.

The assembly of the non-glycosylated variant of the protein using the 4 four fragment strategy was affected by the incomplete removal of the N-terminal protective group, generating a complex product mixture. The need to avoid this deprotection step prompt investigating an NCL strategy with the N- to C- ligation approach.

The N- to C-ligation strategy required the design and syntheses of three peptides and two glycopeptides, which were efficiently completed following the optimized protocols. The methods for the peptide synthesis allowed the synthesis of glycopeptides and the careful optimization of the on-resin deacetylation of GalNAc moiety without base-catalyzed β-elimination. Further optimization of ligation and Acm removal conditions delivered the desired product. This strategy demonstrated versatility and was applied in the synthesis of non- and glycosylated CD59 variants. Although the human CD59 variants were obtained, final products could not be isolated due to difficulties in the purification. Thus, the optimization of the purification methods is required.

The strategy developed in this work delivered two different variants of human CD59 glycoprotein, and these molecules can be used to screen reaction conditions to achieve glypiated variants of the protein and glycosylated variants bearing bigger glycan moieties. Further required steps to obtain a natural functional glycoprotein include the optimization of the folding conditions, the ligation of the glycosylphosphatidylinositol, and the development of a chemoenzymatic approach to elongate the glycans.

2.5. Experimental section

2.5.1. Building blocks synthesis

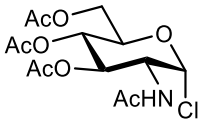
General methods

Commercial grade solvents and reagents were used without further purification unless otherwise noted. Solvents employed for the peptide synthesis were all peptide synthesis grade. L-Fmoc amino acids and activators were purchased from Iris Biotech GmbH or Sigma Aldrich (Germany). 2-CTC Protide resin was purchased from CEM GmbH (Germany). Anhydrous solvents were obtained from a Solvent Drying System (J.C. Meyer) and collected over molecular sieves (4 Å). Analytical thin-layer chromatography (TLC) was performed on 0.25 mm Kieselgel 60 F254 glass-supported plates (Macherey-Nagel). Spots were visualized with UV light (254 nm), sugar stain (0.2 mL of 3-methoxyphenol in 200 mL of EtOH and 6 mL H_2SO_4), ceric ammonium molybdate stain (2 g $Ce(SO4)_2$, 10 g $(NH_4)_6Mo_7O_{24}\cdot 4\ H_2O$, and 20 mL H_2SO_4 in 180 mL H_2O) and vanillin stain (12 g vanilline, 2 mL H_2SO_4 in 200 mL of EtOH). Flash chromatography was performed on Silica gel 60 230-4 00 mesh (Sigma-Aldrich) using air forced flow. NMR spectra were obtained using Bruker Ascend 400 spectrometer at 400 MHz (1H) and 100 MHz (13C) and $CDCl_3$ as solvent. Chemical shifts (δ) are reported in parts per million (ppm) relative to the respective residual solvent peaks ($CDCl_3$: 7.26 ppm ^1H, 77.2 ppm ^{13}C, DMSO-D_6 2.5 ppm ^1H, 39.5 ppm ^{13}C) unless stated otherwise. Bidimensional and non-decoupled experiments were performed to assign identities of peaks showing relevant structural features. NMR spectra were processed using MestreNova 14.2.3 (MestreLab Research).

Analytical high-performance liquid chromatography was performed using an Agilent HP 1100 system equipped with a diode array detector (DAD), an Agilent 1200 series equipped with a Multiple Wavelength Detector (MWD), or a Waters Acquity H-class UPLC equipped with a DAD detector and coupled to a Xevo G2-XS Q-TOF spectrometer (Waters). Analytical HPLC was performed using a YMC Hydrosphere C18 (50 mm X 3.0mm, S-3µm) column [A]. Preparative RP-HPLC purifications were performed on a Knauer Semi-preparative HPLC system using a Synergy 4µm Fusion RP 80 Å (250 x 21.2 mm) column [B] or a YMC Hydrosphere C18 S-5µm, 12nm (150 x 10 mm) column [C]. Separations were completed using a linear gradient of eluent B (0.1% of formic acid in acetonitrile) in solvent A (0.1% of formic acid in water). Products were lyophilize using a Christ Alpha 2-4 LC plus freeze dryer. Centrifugation was carried out using a VWR MicroStar (Kinetic Energy 26 Joules Galaxy Mini Centrifuge) and a VWR Micro Star 17R centrifuge. UV absorbance measurements were recorded on the Shimadzu 1900i UV-Vis spectrophotometer using 1 cm quartz cuvettes. The propylene reactor vessels were shaken on the VWR Heidolph Vibramax 100 system.

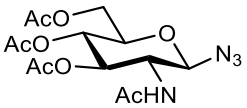
Building Block Syntheses

*2-Acetamido-3,4-6-tri-O-acetyl-2-deoxy-α-D-glycopyranosyl chloride (**25**)*



D-glucosamine hydrochloride (20 g, 90.41 mmol) was suspended in acetyl chloride (70 mL) under argon atmosphere and 27 µL of HCl 37% was added dropwise at 0 °C. The mixture was stirred at rt overnight, diluted with CH_2Cl_2 (200 mL) and extracted with ice water, aqueous saturated $NaHCO_3$ and water. The organic phase was dried over sodium sulfate and concentrated under reduced pressure. The product was purified by flash column chromatography using CH_2Cl_2/methanol (98:2) to give chloride **23** as a white solid (7.0 g, 19.14 mmol, 21%). **R_f =** 0.46 (CH_2Cl_2/MeOH 95:5). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 6.18 (d, *J* = 3.7 Hz, 1H, H-1), 5.87 (d, *J* = 8.8 Hz, 1H, NH), 5.32 (dd, *J* = 10.7, 9.4 Hz, 1H, H-3), 5.21 (t, *J* = 9.8 Hz, 1H, H-4), 4.53 (ddd, *J* = 10.7, 8.7, 3.7 Hz, 1H, H-2), 4.32 – 4.21 (m, 2H, H-6a, H-5), 4.18 – 4.08 (m, 1H, H-6b), 2.10 (s, 3H, CH_3), 2.04 (s, 6H, CH_3), 1.98 (s, 3H, CH_3).**^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 171.6 , 170.7, 170.3, 169.3 (C=O x 4), 93.7 (C-1), 71.0 (C-5), 70.2 (C-3), 67.0 (C-4), 61.2 (C-6), 53.6 (C-2), 23.2 (CH_3-OAc), 20.8 (CH_3-OAc), 20.7 (CH_3-OAc). **HRMS (m/z)**: Calcd for $C_{14}H_{20}ClNO_8Na$ $[M+Na]^+$ 388.0775, observed 388.0782. NMR data agreed with previously reported data [259, 260].

*2-Acetamido-3,4-6-tri-O-acetyl-2-deoxy-β-D-glycopyranoyl azide (**26**)*



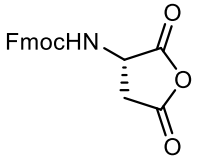
Sodium azide (2.75 g, 42.37 mmol) in 45 mL of saturated $NaHCO_3$ was added to a solution of glycosyl chloride **23** (5.40 g, 14.76 mmol) in CH_2Cl_2 (45 mL). Then, tetrabutylammonium hydrogen sulphate (4.81 g, 14.17 mmol) was added to the reaction and the mixture was stirred vigorously at rt during 2 h. Upon completion by TLC, the reaction mixture was diluted with EtOAc (100 mL) and the aqueous phase was extracted three times with EtOAc. The combined organic layers were washed with saturated aqueous $NaHCO_3$ and brine, dried over Na_2SO_4, filtered and concentrated under reduced pressure. The product was purified by flash column chromatography using EtOAc:hex (6:1) to afford **24** as white solid (5.0 g, 13.43 mmol, 91%). **R_f =** 0.55 (neat EtOAc). **^{1}H NMR** (400 MHz, $CDCl_3$ δ (ppm) 5.82 (d, *J* = 8.9 Hz, 1H, NH), 5.25 (dd, *J* = 10.6, 9.3 Hz, 1H, H-3), 5.09 (t, *J* = 9.7 Hz, 1H, H-4), 4.77 (d, *J* = 9.3 Hz, 1H, H-1), 4.26 (dd, *J* = 12.4, 4.8 Hz, 1H, H-6a), 4.16 (dd, *J* = 12.5, 2.3 Hz, 1H, H-6b), 3.91 (dt, *J* = 10.6, 9.1 Hz, 1H, H-2), 3.80 (ddd, *J* = 10.1, 4.8, 2.3 Hz, 1H, H-5), 2.10 (s, 3H, CH_3), 2.03 (d, *J* = 3.3 Hz, 6H), 1.97 (s, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 171.1, 170.9, 170.6, 169.4 (C=O x 4), 88.5 (C-1), 74.0 (C-5), 72.2 (C-3), 68.1 (C-4), 61.9 (C-6), 54.2 (C-2), 23.4 (CH_3-OAc), 20.9 (CH_3-OAc), 20.8 (CH_3-OAc), 20.7 (CH_3-OAc).

HRMS (m/z): Calcd for $C_{14}H_{21}N_4O_8$ $[M+H]^+$ 373.1359, observed 373.1360. NMR data agreed with previously reported data [259, 260, 87].

*2-Acetamido-3,4-6-tri-O-acetyl-2-deoxy-β-D-glycopyranoyl amine (**27**)*

The glucopyranosyl azide **24** (1.5 g, 4.03 mmol) and 10% Pd/C (85.7 mg, 0.08 mmol) were added to a flask and purged with nitrogen three times. The mixture was dissolved in anhydrous THF (20 mL), triethylamine (0.5 mL, 4.03 mmol) was added, and the flask was purged with hydrogen twice. After stirring the suspension during 3 h at rt, the reaction mixture was diluted with ethyl acetate (100 mL), filtered through Celite, and concentrated under reduced pressure. The crude product was obtained as a grey solid (1.15 g, 3.32 mmol, 83%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 5.79 (d, *J* = 9.2 Hz, 1H, NH), 5.13 – 4.96 (m, 2H, H-3, H-4), 4.20 (dd, *J* = 12.3, 4.9 Hz, 1H, H-6a), 4.15 – 4.05 (m, 2H, H-1, H-6b), 4.04 – 3.96 (m, 1H, H-2), 3.63 (ddd, *J* = 9.6, 4.8, 2.4 Hz, 1H, H-5), 2.08 (s, 3H, CH_3), 2.02 (d, *J* = 5.8 Hz, 6H, CH_3), 1.96 (s, 3H, CH_3). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 171.7, 170.9, 170.7, 169.5 (C=O x 4), 86.5 (C-1), 73.5 (C-5), 72.8 (C-3), 68.5 (C-4), 62.5 (C-6), 54.9 (C-2), 23.4 (CH_3-OAc), 20.9 (CH_3-OAc), 20.9 (CH_3-OAc), 20.8 (CH_3-OAc). **HRMS (m/z):** Calcd for $C_{14}H_{22}N_2O_8Na$ $[M+Na]^+$ 369.1274, observed 369.1276. NMR data agreed with previously reported data [87].

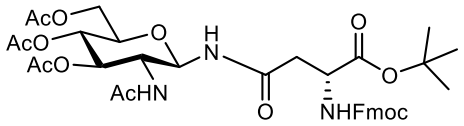
*N-(Fluorenyl-9-methoxycarbonyl)-L-aspartic anhydride (**28**)*



A suspension of *N*-Fmoc-L-aspartic acid 4-tert-butyl ester (4 g, 9.72 mmol) in CH_2Cl_2 (13 mL) was cooled to 0 °C and trifluoroacetic acid (14.9 mL, 194.4 mmol) was added dropwise. The mixture was warmed to rt and stirred for 3 h. After reaction completion, toluene (100 mL) was added to the reaction and the solvent mixture was removed at reduced pressure. Fmoc-L-aspartic acid was obtained as a white powder and used for the next step without further purification (3.45 g, 9.72 mmol, quantitative). Crude *N*-Fmoc-L-aspartic acid (3.45 g, 9.71 mmol) was dissolved in acetic anhydride (28 mL, 297.3 mmol) and the mixture was refluxed for 2 h at 95 °C. The reaction mixture was cooled to rt and a few milliliters of cold ether were added. A formation of a precipitate was observed and more diethyl ether was added. The formed precipitate was filtrated, washed with cold diethyl ether, and dried to give the *N*-Fmoc-aspartic anhydride **26** as a white powder (2.40 g, 7.11 mmol, 73% over two steps). **^{1}H NMR** (400 MHz, DMSO-D_6) δ (ppm)) 8.21 (d, *J* = 7.6 Hz, 1H, NH), 7.93 – 7.86 (m, 2H, H-Ar), 7.67 (d, *J* = 7.5 Hz, 2H, H-Ar), 7.45 – 7.39 (m, 2H, H-Ar), 7.34 (tdd, *J* = 7.4, 2.2, 1.3 Hz, 2H, H-Ar), 4.65 (ddd, *J* = 10.0, 7.7, 6.2 Hz, 1H, H-α), 4.47 – 4.34 (m, 2H,

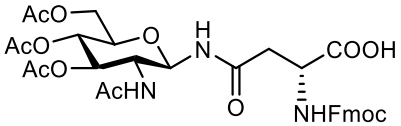
CH_2-Fmoc), 4.25 (t, *J* = 6.5 Hz, 1H, CH-Fmoc), 3.24 (dd, *J* = 18.5, 10.0 Hz, 1H, H-β), 2.86 (dd, *J* = 18.5, 6.2 Hz, 1H, H-β). **^{13}C NMR** (101 MHz, DMSO-D_6) δ (ppm) 172.2 (C=O β), 169.9 (C=O α), 155.9, 143.7, 143.6, 140.8, 127.7, 127.2, 127.1, 125.1, 125.1, 120.2, 120.2 (12 C-Ar), 66.1 (CH_2-Fmoc), 50.4 (C-α), 46.6 (CH-Fmoc), 34.7 (C-β). **HRMS (m/z):** Calcd for $C_{19}H_{15}NO_5Na$ $[M+Na]^+$ 360.0848, observed 360.0854. NMR data agreed with previously reported data.[251].

*N^{α}-(9-fluorenylmethoxycarbonyl)-N-(2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranosyl)-L-asparagine tertbutyl ester (**29**)*



Glycosyl amine **25** (0.7 g, 2.02 mmol), Fmoc-Asp-OtBu (0.83 g, 2.02 mmol) and *N*-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline (0.5 g, 2.02 mmol) were stirred in anhydrous CH_2Cl_2 (20 mL) under argon at room temperature during 16 h. The crude mixture was concentrated in vacuo followed by addition of a mixture of hot EtOAc and petroleum sprit (1:1, v/v, 50 mL). The suspension was filtered, the solid was collected and dried under reduced pressure to yield **27** as a grey solid (1.23 g, 1.66 mmol, 82%).**^{1}H NMR** (400 MHz, DMSO-D_6) δ (ppm) 7.75 (d, *J* = 7.5 Hz, 2H, H-Ar), 7.60 (d, *J* = 7.5 Hz, 2H, H-Ar), 7.39 (t, *J* = 7.5 Hz, 2H, H-Ar), 7.30 (t, *J* = 7.4 Hz, 2H, H-Ar), 7.20 (d, *J* = 8.2 Hz, 1H, NH-Glc), 6.08 (d, *J* = 8.0 Hz, 1H, NH-NHAc), 5.95 (d, *J* = 8.9 Hz, 1H, NH-Asn), 5.19 – 5.00 (m, 3H, H-1, H-3, H-4), 4.52 (dt, *J* = 9.0, 4.5 Hz, 1H, H-α), 4.42 (dd, *J* = 10.4, 7.1 Hz, 1H, H-6a), 4.36 – 4.18 (m, 3H, H-6b, CH-Fmoc, CH_2a-Fmoc), 4.18 – 4.01 (m, 2H, H-2, CH_{2b}-Fmoc), 3.74 (ddd, *J* = 9.8, 4.2, 2.2 Hz, 1H, H-5), 2.85 (dd, *J* = 16.5, 4.7 Hz, 1H, H-βb), 2.70 (dd, *J* = 16.4, 4.4 Hz, 1H, H-βb), 2.13 – 2.00 (m, 9H, CH_3), 1.96 (s, 3H, CH_3), 1.44 (s, 9H, OtBu). **^{13}C NMR** (101 MHz, DMSO-D_6) δ (ppm) 172.5, 172.2, 171.2, 170.8, 170.1, 169.4, 156.2 (C=O x 7), 144.0, 143.9, 141.4, 141.4, 127.8, 127.2, 125.3, 125.3, 120.1 (12 C-Ar), 82.3 (C, tBu)), 80.4 (C-1), 73.6 (C-5), 73.0 (C-3), 67.6 (C-4), 67.2 (C-6), 61.7 (CH_2-Fmoc), 53.6 (H-2), 51.1 (C-α), 47.2 (CH-Fmoc), 38.0 (C-β), 28.0 (CH_3, OtBu), 23.2 (CH_3-OAc), 20.9 (CH_3-OAc), 20.7 (CH_3-OAc). **HRMS (m/z):** Calcd for $C_{37}H_{45}N_3O_{13}Na$ $[M+Na]^+$ 762.2850, observed 762.2861. NMR data agreed with previously reported data [87].

*N^{α}-(9-fluorenylmethoxycarbonyl)-N-(2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranosyl)-L-asparagine (**11**)*



Method A: The glucopyranosyl amine **25** (0.5 g, 1.44 mmol) and the aspartic anhydride **26** (0.49 g, 1.44 mmol) were dissolved in anhydrous DMSO (2 mL) and stirred at rt under argon atmosphere. After 5 h, the mixture was diluted with water (30 mL) and extracted with chloroform

(3 x 100 ml). The organic phase was then washed with brine, dried over Na_2SO_4, and concentrated under reduced pressure. The product glycosylated amino acid **11** was purified by flash column chromatography CH_2Cl_2/MeOH 98:2 to 96:4 and obtained as a white solid (0.537 g, 0.785 mmol, 54%).

Method B: Glycosylated amino acid **27** (0.9 g, 1.22 mmol) was dissolved in a mixture of trifluoroacetic acid/H_2O (10:1 v/v, 16.5 mL) at 0 °C and stirred for 3 h at rt Then, the solvents were removed *in vacuo* and the residue was co-evaporated with toluene (5 x 30 mL) and dried under reduced pressure to yield **11** as a grey solid (0.88, 1.19 mmol, 98%).

Rf = 0.4 (CH_2Cl_2/MeOH 95:5). **^{1}H NMR** (400 MHz, DMSO-D_6) δ (ppm) 8.62 (d, *J* = 9.2 Hz, 1H, NH-Glc), 7.91 (dd, *J* = 13.6, 8.3 Hz, 2H, H-Ar), 7.71 (d, *J* = 7.5 Hz, 2H, H-Ar), 7.55 (d, *J* = 8.5 Hz, 1H, NH-Asn), 7.41 (td, *J* = 7.5, 1.1 Hz, 2H, H-Ar), 7.32 (td, *J* = 7.4, 1.2 Hz, 2H, H-Ar), 5.17 (t, *J* = 9.5 Hz, 1H, H-1), 5.09 (t, *J* = 9.8 Hz, 1H, H-3), 4.81 (t, *J* = 9.8 Hz, 1H, H-4), 4.38 (td, *J* = 7.5, 5.5 Hz, 1H, H-α), 4.32 – 4.14 (m, 4H, H-6a, CH_{2a}-Fmoc, CH-Fmoc), 3.98 – 3.78 (m, 3H, H-2, H-5, CH_{2b}-Fmoc), 2.65 (dd, *J* = 16.3, 5.4 Hz, 1H, H-βa), 2.50 (dd, overlapped by DMSO signal, H- βb) 1.97 (d, *J* = 12.5 Hz, 6H, CH_3), 1.90 (s, 3H, CH_3), 1.71 (s, 3H, CH_3).**^{13}C NMR** (101 MHz, DMSO-D_6) δ (ppm) 173.1, 170.1, 169.9, 169.6, 169.6, 169.4, 155.9, 143.9, 143.8, 140.8, 129.0, 128.3, 127.7, 127.2, 125.4, 125.3, 120.2, 78.1(C-1), 73.4 (C-3), 72.3 (C-5), 68.4 (C-4), 65.8 (C-6), 61.9 (CH_2-Fmoc), 52.1 (C-2), 50.0 (CH-α), 46.6 (CH-Fmoc), 36.9 (C-β), 22.7 (CH_3-OAc), 20.6 (CH_3-OAc), 20.5 (CH_3-OAc), 20.5 (CH_3-OAc). **HRMS (m/z):** Calcd for $C_{33}H_{37}N_3O_{13}Na$ $[M+Na]^+$ 706.2224, observed 706.2239. NMR data agreed with previously reported data[260, 87].

2.5.2. Methods for solid phase synthesis of peptides

General methods for the solid phase peptides synthesis (SPPS)

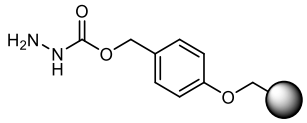
The synthesis of the peptides segments was performed from N- to C-terminus by automated microwave-assisted solid phase synthesis using a Liberty Blue™ peptide synthesizer (CEM corporation, Germany). Syntheses were achieved on a 30 mL Teflon reaction vessel using the CEM suggested conditions for 0.1 mmol scale. All solutions were freshly prepared and kept under nitrogen during the whole synthesis process.

The functionalization of the resins, coupling of the first amino acid and glycosylated amino acids were performed manually in a 10- or 20 mL disposable polypropylene reactor equipped with plunger and 25 µm bottom frit (Multisynthech). Small-scale and full cleavage were performed manually in (2 mL or 10 mL) disposable polypropylene reactor equipped with plunger with 25 µm bottom frit (Multisynthech). All the solutions and solvents were transfer by suction. The reactions were shaken at room temperature on a VWR Heidolph Vibramax 100. The presence of free amino

groups during manual synthesis was detected using qualitative Kaiser test (solution A: 1 mL of a 0.01M aqueous solution of KCN in 49 mL of pyridine, solution B: 1 g of ninhydrin in 20 mL of *n*-butanol, solution C: 40 g of phenol in 20 mL of *n*-butanol). Resin loading was determined by following previously established procedures [238, 261]***.***

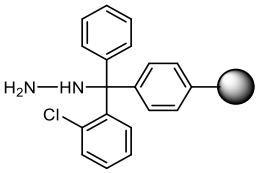
Functionalization of the resins to obtain C-terminal hydrazide peptides

*Functionalization of NovaPEG Wang resin (**R1**)*



NovaPEG Wang resin (500 mg, 0.305 mmol, 0.61 mmol/g commercial loading) was swollen in 15 mL of CH_2Cl_2 at rt during 1 h. A cold solution of 4-nitrophenyl chloroformate (307.3 mg, 1.52 mmol) in CH_2Cl_2 (7 mL) was added, the polypropylene reactor was cooled to -20 °C for 1 h. Then, a solution of *N*-methylmorpholine (169 µl, 1.52 mmol) in CH_2Cl_2 (7 mL) was added to the resin and the mixture was warmed to rt and shaken overnight. The solvent was filtered off, and the resin was washed with CH_2Cl_2, DMF, MeOH, and CH_2Cl_2 (5 x 10 mL each solvent). A cold solution of hydrazine monohydrate (74 µL, 1.52 mmol) in CH_2Cl_2/DMF 1:1 (8 mL) was added to the resin and the mixture was shaken at rt overnight. The solvent was drained, and the resin was washed with DMF, CH_2Cl_2, DMF, MeOH, and CH_2Cl_2 (5 x 10 mL each solvent).

*Functionalization of 2-Cl-Trt-(Cl) (**R2**)*



To Cl-TCP(Cl) ProTide Resin (0.25g, 0.118 mmol, 0.47 mmol/g commercial loading) a solution of 1M hydrazine in THF (1.2 mL, 1.175 mmol) in DMF (1 mL) was added and the suspension was shaken at rt during 1.5 h. The solution was filtered off and the resin was washed once with DMF (5 mL). A freshly prepared solution of hydrazine was added and the mixture was shaken for 30 min. The solution was drained, and the resin was washed with DMF, CH_2Cl_2, and DMF (5 x 5 mL each solvent). Subsequently, a solution of 5% v/v of methanol in DMF (2 mL) was added to the resin and the mixture was shaken for 20 min. The solvent was filtered off and washed with DMF, CH_2Cl_2, MeOH, and DMF (5 x 5 mL each solvent). The functionalized resin was directly used for the coupling of the first amino acid [116].

General method for the coupling of the first amino acid on functionalized solid support

NovaPEG Wang hydrazide

Hydrazine functionalized NovaPEG Wang resin **R1** (0.2 g, 0.122 mmol) was swollen in CH_2Cl_2 (5 mL) for 30 min. The amino acid (0.61 mmol) was activated with DIC (95 µL, 0.5 mmol) and HOBt (93.4 mg, 0.61 mmol) in DMF and after 2 min added to the solid support. The suspension was shaken at rt overnight. The solution was drained, and the resin washed with DMF, CH_2Cl_2, DMF (4 x 4 mL each solvent). Resin loading was determined by following previously established procedures [238, 261]**.**

2CT Hydrazide ProTide

Hydrazine functionalized trityl resin **R2** (0.25 g, 0.118 mmol) was swollen in CH_2Cl_2 (5 mL) at rt for 30 min. The amino acid (0.47 mmol) was activated with HATU (169 mg, 0.44 mmol), DIPEA (164 µL, 0.94 mmol) in DMF (4 mL), and after 30 sec added to the solid support. After shaking the suspension overnight, the solution was drained and the resin was washed with DMF, CH_2Cl_2, and DMF (4 x 4 mL of each solvent). Resin loading was determined by following previously established procedures [238, 261]

General method for the manual coupling of amino acid or glycosylated amino acids

Manual coupling of Fmoc-Cys(Trt)-OH

The peptide-resin (0.1 mmol) was taken out of the synthesizer and placed in a polypropylene reactor. Fmoc-Cys(Trt)-OH (293 mg, 0.5 mmol) was activated with HOBt (77 mg, 0.5 mmol), DIC (77 µL, 0.5 mmol) in DMF (4 mL), and after 2 min added to the solid support. The suspension was shaken at rt overnight. The solution was drained, and the resin was washed with DMF, CH_2Cl_2, and DMF (4 x 4 mL each solvent). Subsequently, a mixture of Ac_2O/DIPEA/NMP (1:1:1) was added to the resin and shaken for 20 min at rt, the solvent was filtered off and the resin was washed with DMF, CH_2Cl_2, MeOH, DMF (4 x 4 mL each solvent).
*Coupling was monitored by Kaiser test and double coupling was performed when needed.

Manual coupling of Fmoc-Asp(OEpe)-OH

The peptide-resin (0.1 mmol) was taken out of the synthesizer and placed in a polypropylene reactor. Fmoc-Asp(OEpe)-OH (91 mg, 0.2 mmol) was activated with HATU (68 mg, 0.18 mmol), DIPEA (70 µL, 0.4 mmol) in DMF (2 mL) and after 30 sec added to the resin. The suspension was shaken at rt during 1h, the solution was drained, and the resin was washed with DMF (2 x 4 mL each solvent). A second coupling was performed with Fmoc-Asp(OEpe)-OH (45 mg, 0.1 mmol), HATU (34 mg, 0.09 mmol), DIPEA (35 µL, 0.2 mmol) in DMF (2 mL) during 1 h. Once draining the

solvent, the resin was washed with DMF, CH_2Cl_2, MeOH, and DMF (4 mL, 4 times each solvent). Subsequently, a mixture of Ac_2O/DIPEA/NMP (1:1:1) was added to the resin and shaken for 30 min at rt, the solvent was filtered off and the resin was washed with DMF, CH_2Cl_2, DMF, MeOH (4 x 4 mL each solvent). The resin was placed back in the synthesizer to continue the elongation.

Manual coupling of glycosylated asparagine

Solid support (0.08 mmol) was taken out of the synthesizer and placed in a polypropylene reactor. Fmoc-L-Asn((Ac)$_3$-α-D-GlcNAc)-OH (219 mg, 0.32 mmol) was activated with PyBOP (208 mg, 0.4 mmol), DIPEA (111 µL, 0.64 mmol) in DMF/DMSO (2.5 mL, 4:1) and after 2 min added to the resin. The suspension was shaken overnight, the solution was drained, and the resin was washed with DMF, CH_2Cl_2, MeOH, and DMF (4 x 4 mL each solvent). Subsequently, a mixture of Ac_2O/DIPEA/NMP (1:1:1) was added to the resin and shaken during 2 h at rt, the solvent was filtered off and the resin washed with DMF, CH_2Cl_2, MeOH, DMF (4 x 4 mL each solvent). The resin was placed back in the synthesizer to continue the elongation.
*Coupling was monitored by Kaiser test and double coupling was performed when needed.

Manual coupling of glycosylated threonine

The peptide-resin (0.08 mmol) was taken out of the synthesizer and placed in a polypropylene reactor. Fmoc-L-Thr((Ac)$_3$-α-D-GalNAc)-OH (215 mg, 0.32 mmol) was activated with PyBOP (208 mg, 0.4 mmol), DIPEA (111 µL, 0.64 mmol) in DMF (3 mL) and after 2 min added to the resin. The suspension was shaken overnight, the solution was drained, and the resin washed with DMF, CH_2Cl_2, MeOH, and DMF (4 x 4 mL each solvent). Subsequently, a mixture of Ac_2O/DIPEA/NMP (1:1:1) was added to the resin and shaken at rt for 30 min, the solvent was removed and the resin was washed with DMF, CH_2Cl_2, MeOH, and DMF (4 x 4 mL each solvent).
*Coupling was monitored by Kaiser test and double coupling was performed when needed.

Manual deprotection after glycosylated amino acid

Deprotection solution 2 mL (20% piperidine, 0.8% formic acid in DMF) was added to the resin, and the suspension was shaken at room temperature for 10 min. The solvent was drained, and the resin was washed with DMF, CH_2Cl_2, MeOH, and DMF (4 x 4 mL of each solvent).
*Deprotection was monitored by Kaiser test and double deprotection was performed when needed.

Manual elongation after glycosylated amino acid

Fmoc amino acid* (0.24 mmol) was activated with HBTU (0.216 mmol) and DIPEA (0.48 mmol) in DMF (3 mL) and after 30 sec added to the resin. The suspension was shaken for 1h at rt, the

solution was drained, and the resin was washed with DMF, CH_2Cl_2, MeOH, DMF (4 x 4 mL each solvent). Subsequently, a mixture of Ac_2O/DIPEA/NMP (1:1:1) was added to the resin and shaken for 30 min at rt, the solvent was filtered off and the resin was washed with DMF, CH_2Cl_2, MeOH, and DMF (4 x 4 mL each solvent).
*Coupling was monitored by Kaiser test and double coupling was performed when needed.

Capping

The peptide-resin (0.08 mmol) was taken out of the synthesizer and placed in a polypropylene reactor and a mixture of Ac_2O/DIPEA/NMP (1:1:1) was added to the resin and shaken for 30 h at rt, the solution was filtered off and the resin washed with DMF, CH_2Cl_2, MeOH, DMF (4 x 4 mL each solvent). The resin was placed back in the synthesizer to continue the elongation.

Microwave assisted solid phase peptide synthesis

Preparation of stock solutions

Amino acid solution: Amino acids were dissolved in DMF to a final concentration of 0.2 M.
Deprotection solution: A solution containing, 20% piperidine, 0.8% of formic acid in DMF was prepared.
Activator solution (Act): A 0.5 M solution of *N,N'*-diisopropylcarbodiimide in DMF was prepared.
Activator base solution (Act base): A 1 M solution of ethyl cyano(hydroxyimino)acetate (Oxyma Pure®) in DMF was prepared.

Cycles used for SPPS

Swelling: Peptide synthesis was performed at 0.1 mmol scale. Solid support bearing the first amino acid was placed in the reaction vessel and swollen in DMF at rt for 10 min. Prior to synthesis, all the main operations were calibrated, and the lines purged. After swelling, the solvent was drained, and the cycle was started.

Fmoc removal (deprotection) and coupling: Deprotection solution (4 mL) was delivered to the reaction vessel and kept under nitrogen bubbling, the solution was drained, and the resin was washed with DMF (3x 4 mL, 17 s each, 10 s draining time). The amino acid solution (2.5 mL), activator (1 mL), and activator base (0.5 mL) were delivered and kept under nitrogen bubbling. After finishing the time given for the coupling, the solvent was filtered off and the resin was washed with DMF (4 mL, 10 s). The temperature of the reaction vessel and the time for deprotection and coupling cycles were adjusted to get various methods. The specifications of each method are shown in the following tables:

Single deprotection, single coupling at 75 °C (11 min)

Step	Operation	Parameters
1	Deprotection	75 °C, 50 W, 3 min
2	Wash	DMF, 4 mL, 17 s, draining time 10 s
3	Wash	DMF, 4 mL, 17 s, draining time 10 s
4	Wash	DMF, 4 mL, 17 s, draining time 10 s
5	Wash	DMF, 4 mL, 17 s, draining time 10 s
6	Coupling	75 °C, 30 W, 5 min
7	Wash	DMF, 4 mL, 17 s, draining time 10 s

Single deprotection, double coupling at 75 °C (16 min)

Step	Operation	Parameters
1	Deprotection	75 °C, 50 W, 3 min
2	Wash	DMF, 4 mL, 17 s, draining time 10 s
3	Wash	DMF, 4 mL, 17 s, draining time 10 s
4	Wash	DMF, 4 mL, 17 s, draining time 10 s
5	Wash	DMF, 4 mL, 17 s, draining time 10 s
6	Coupling	75 °C, 30 W, 5 min
7	Coupling	75 °C, 30 W, 5 min
8	Wash	DMF, 4 mL, 17 s, draining time 10 s

Double deprotection, double coupling at 75 °C (19 min)

Step	Operation	Parameters
1	Deprotection	75 °C, 50 W, 3 min
2	Wash	DMF, 4 mL, 17 s, draining time 10 s
3	Deprotection	75 °C, 50 W, 3 min
4	Wash	DMF, 4 mL, 17 s, draining time 10 s
5	Wash	DMF, 4 mL, 17 s, draining time 10 s
6	Wash	DMF, 4 mL, 17 s, draining time 10 s
7	Coupling	75 °C, 30 W, 5 min
8	Coupling	75 °C, 30 W, 5 min
9	Wash	DMF, 4 mL, 17 s, draining time 10 s

Double deprotection, double coupling at 50 °C (29 min)

Step	Operation	Parameters
1	Deprotection	75 °C, 50 W, 3 min
2	Wash	DMF, 4 mL, 17 s, draining time 10 s
3	Deprotection	75 °C, 50 W, 3 min
4	Wash	DMF, 4 mL, 17 s, draining time 10 s
5	Wash	DMF, 4 mL, 17 s, draining time 10 s
6	Wash	DMF, 4 mL, 17 s, draining time 10 s
7	Coupling	25 °C, 0 W, 2 min / 50 °C, 35 W, 8 min
8	Coupling	25 °C, 0 W, 2 min / 50 °C, 35 W, 8 min
9	Wash	DMF, 4 mL, 17 s, draining time 10 s

Single deprotection at 25 °C, double coupling at 50 °C (52 min)

Step	Operation	Parameters
1	Deprotection	25 °C, 0 W, 10 min
4	Wash	DMF, 4 mL, 17 s, draining time 10 s
5	Wash	DMF, 4 mL, 17 s, draining time 10 s
6	Wash	DMF, 4 mL, 17 s, draining time 10 s
7	Coupling	50 °C, 35 W, 20 min
8	Coupling	50 °C, 35 W, 20 min
9	Wash	DMF, 4 mL, 17 s, draining time 10 s

Double Arg coupling at 75 °C (63 min)

Step	Operation	Parameters
1	Deprotection	75 °C, 50 W, 3 min
2	Wash	DMF, 4 mL, 17 s, draining time 10 s
3	Deprotection	75 °C, 50 W, 3 min
4	Wash	DMF, 4 mL, 17 s, draining time 10 s
5	Wash	DMF, 4 mL, 17 s, draining time 10 s
6	Wash	DMF, 4 mL, 17 s, draining time 10 s
7	Coupling	25 °C, 0 W, 25 min / 75 °C, 35 W, 2 min
8	Coupling	25 °C, 0 W, 25 min / 75 °C, 35 W, 2 min
9	Wash	DMF, 4 mL, 17 s, draining time 10 s

Final deprotection: Once the elongation of the peptide was finished, deprotection solution (4 mL) was delivered to the reaction vessel and kept under nitrogen bubbling. The solution was drained and the resin was washed with DMF (4 x 4 mL, 17 s each, 10 s draining time). The final deprotection of the peptide was performed using the following methods:

Final deprotection at 75 °C (4 min)

Step	Operation	Parameters
1	Deprotection	75 °C, 50 W, 3 min
2	Wash	DMF, 4 mL, 17 s, draining time 10 s
4	Wash	DMF, 4 mL, 17 s, draining time 10 s
5	Wash	DMF, 4 mL, 17 s, draining time 10 s
6	Wash	DMF, 4 mL, 17 s, draining time 10 s

Final deprotection at 25 °C (11 min)

Step	Operation	Parameters
1	Deprotection	25 °C, 0 W, 10 min
2	Wash	DMF, 4 mL, 17 s, draining time 10 s
4	Wash	DMF, 4 mL, 17 s, draining time 10 s
5	Wash	DMF, 4 mL, 17 s, draining time 10 s
6	Wash	DMF, 4 mL, 17 s, draining time 10 s

2.5.3. Post-automated synthesis steps

On-resin deacetylation

The solid support was swollen in DMF/CH_2Cl_2 (1:1) for 30 min. The solvent was removed, a 10% solution of hydrazine monohydrate in MeOH/THF (1:1) was added, and the suspension was shaken at rt for 2 h. The solvent was drained, and the resin was washed with CH_2Cl_2, DMF, MeOH, and CH_2Cl_2 (5 mL, three times each solvent).

Cleavage

Two cocktails were used for the cleavage of peptides and glycopeptides:

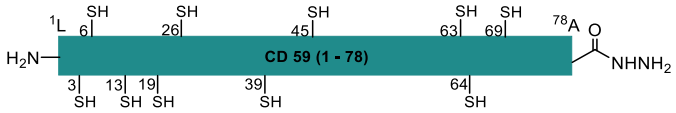
Solution	Reagents	Ratio
K	TFA/phenol/water/thioanisol/EDT	82.5/5/5/5/2.5
B'	TFA/TIPS/Water	90/5/5

Small-scale cleavage: Some beads of dry resin were transfer into a 2 mL polypropylene reactor and 300 µL of the cocktail solution were added. The suspension was shaken at rt, for 2 h. The solution was collected in a 2 mL Eppendorf tube and precipitated with cold Et_2O (1.8 mL), mixed using a vortex and centrifuged. The supernatant was removed and the remaining precipitated was extracted with cold diethyl ether three times. The precipitate was dried using a nitrogen stream and dissolved in 10% of acetonitrile in water. The solution was analyzed by RP-HPLC and LC-MS.

Full cleavage: Dry resin was placed in a 10 mL polypropylene reactor and the cocktail solution (7 mL/200 mg of peptide-resin) was added. The suspension was shaken at rt for 2.5 h. The filtrate was collected in a 50 mL polypropylene tube and precipitated with cold Et_2O (45 mL), mixed using a vortex and centrifuged. The supernatant was removed. The addition of cold diethyl ether, mixing and centrifugation was repeated three times. The precipitate was dried using a nitrogen stream and carried through purification.

2.5.4. Synthesis of peptides and glycopeptides protocols and chromatographic data

*Synthesis of non-glycosylated protein **2***



The synthesis of the non-glycosylated protein **2** was attempted using functionalized resin **R2** (0.3 g, 0.051 mmol, loading 0.17 mmol/g) bearing Fmoc-Ala-OH, as described in the general methods. The peptide was elongated by MW-SPPS using a 10 molar excess of the amino acid. The below table shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue	Cycle
Asn-77, Leu-75, Gln-74, Asn-72, Phe-71, Leu-68, Asp-67, Lys-66	0.1 - Single deprotection single coupling 75 °C
Lys 65	0.1 - Single deprotection double coupling 75 °C
Glu-73, Glu-73, Asn-70	0.1 - Double deprotection double coupling 75 °C
Cys-69, from Cys-64 to Leu-1	0.1 - Double deprotection double coupling 50 °C
Arg-55, Arg-53	0.1 - Double arginine coupling
	Final deprotection 75 °C

To monitor the synthesis, a small-scale cleavage was performed after the coupling of residues Thr-52, Cys-39 and Leu-1. The crude peptide was analyzed by LC-MS using column [A] and a gradient of 5 - 40% MeCN in H_2O (0.1 % HCOOH) over 25 min. The product was not isolated.

*Synthesis of CD59 peptide fragment **3** (1-25)*

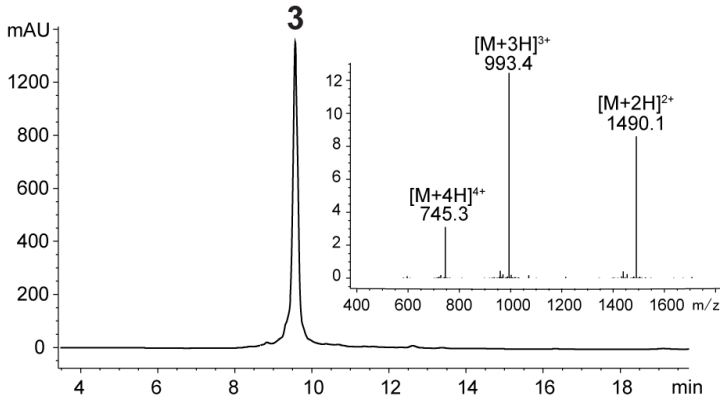
The synthesis of the peptide was performed on functionalized resin **R2** (0.33 g, 0.092 mmol, 0.28 mmol/g). Fmoc-Ala-OH was coupled as described in the general methods. The peptide was elongated by MW-SPPS using a 5.4 molar excess of the amino acid. The table below shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue and position	Cycle
Asp-24*	0.1 - Double deprotection double coupling 75 °C
Phe-23, Ser 21, Ser 20*, Asn-18, Val-17, Ala-16, Thr-15*, Lys-14, Ala-11, Thr-10, Pro-9, Asn-8, Pro-7*, Asn-5*, Tyr-4, Gln-2, Leu-1	0.1 - Single deprotection 25 °C double coupling 50 °C
Cys-19, Cys-13, Cys-6, Cys-3	0.1 - Double deprotection double coupling 50 °C
Asp-22*, Asp-12*,	Manual coupling Final deprotection 25 °C

* Capping cycle
Cys-19, Cys-13, Cys-6, Cys-3: Fmoc-Cys(Acm)-OH
Asp-22 and Asp-12: Fmoc-Asp(OEpe)-OH

The peptide-resin was taken out of the synthesizer and was dried *in vacuo*. The full cleavage was performed using cleavage cocktail K. The crude product (101.6 mg) was dissolved in (20 mL) 10% MeCN/H_2O + 0.1 % TFA, purified by semi-preparative RP-HPLC (column [B], 10-50% MeCN/H_2O + 0.1 % TFA over 35 min at 8 mL/min flow rate) and lyophilized.

Yield of **3**: 8.84 mg (2.97 µmol, 8.7%). LC-MS: column [A], 5 - 60% MeCN/H_2O + 0.1% HCOOH over 25 min. **ESI-TOF-MS** of **3**: *m/z* $C_{121}H_{188}N_{36}O_{44}S_4$ (2977.2 Da) calculated: 1489.6 $[M+2H]^{2+}$, 993.4 $[M+3H]^{3+}$, 745.3 $[M+4H]^{4+}$, observed: 1490.1 $[M+2H]^{2+}$, 993.4 $[M+3H]^{3+}$, 754.3 $[M+4H]^{4+}$.



*Synthesis of CD59 peptide fragment **4** (26-44)*

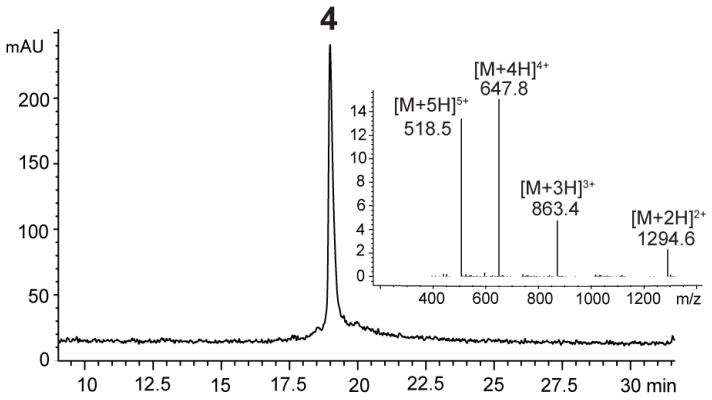
The peptide synthesis was performed on functionalized resin **R1** (0.244 g, 0.04 mmol, loading 0.17 mmol/g). Fmoc-Hys(Trt)-OH was coupled as described in the general methods. The peptide was elongated by MW-SPPS using a 12 molar excess of the amino acid. The table shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue and position	Cycle
Glu-43, Phe-42, Lys-41, Trp-40, Lys-38, Asn-37, Tyr-36, Val-35, Leu-33, Gly-32, Ala-31, Thr-29, Leu-27	0.1 - Single deprotection single coupling 75 °C
Gln-34, Lys-30, Ile-28	0.1 - Double deprotection double coupling 75 °C
Cys-39	0.1 - Double deprotection double coupling 50 °C
Cys-26	Manual coupling

Cys-39: Fmoc-Cys(Acm)-OH
Cys-26 Fmoc-Cys(Trt)-OH

The peptide-resin was dried i*n vacuo and* full cleavage was performed with cleavage cocktail K. The crude product (54.02 mg) was dissolved in (10 mL) 10% MeCN/H_2O + 0.1% TFA, purified by semi-preparative RP-HPLC (column [B], 10 - 50% MeCN/H_2O + 0.1% TFA over 35 min, 8 mL/min) and lyophilized.

Yield of **4**: 11.04 mg (4.26 µmol, 20.4%). LC-MS: column [A], 5 - 40% MeCN/H_2O + 0.1% HCOOH over 25 min. **ESI-TOF-MS** of **4**: *m/z* $C_{123}H_{178}N_{30}O_{28}S_2$ (2587.3 Da) calculated: 1294.6 $[M+2H]^{2+}$, 863.4 $[M+3H]^{3+}$, 647.8 $[M+4H]^{4+}$, 518.5 $[M+5H]^{5+}$, observed: 1294.6 $[M+2H]^{2+}$, 863.4 $[M+3H]^{3+}$, 674.8 $[M+4H]^{4+}$, 518.5 $[M+5H]^{5+}$.



*Synthesis of CD59 peptide fragment **5** (45-68)*

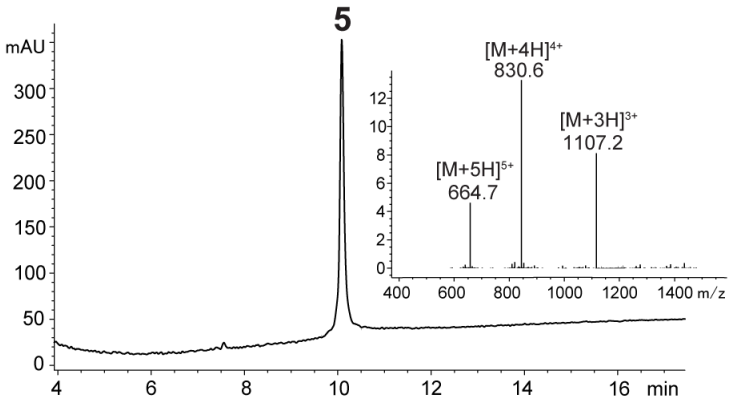
The peptide synthesis was performed on functionalized resin **R2** (0.34 g, 0.09 mmol, loading 0.26 mmol/g). Fmoc-Leu-OH was coupled as described in the general methods. The peptide was elongated by MW-SPPS using a 5.5 molar excess of the amino acid. The table below shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue and position	Cycle
Asp-67	0.1 - Double deprotection double coupling 75 °C
Lys-66, Lys-65, Tyr-62, Tyr-61, Thr-60, Leu-59, Glu-58, Asn-57, Glu-56*, Leu-54, Thr-52, Thr-51, Val-50, Asn-48, Phe-47, Asn-46, Cys-45	0.1 - Single deprotection 25 °C double coupling 50 °C
Cys-64, Cys-63	0.1 - Double deprotection double coupling 50 °C
Arg-55, Arg-53	0.1 - Double arginine coupling
Asp-49*	Manual coupling

* Capping cycle
Cys-64, Cys-63: Fmoc-Cys(Acm)-OH
Asp-49: Fmoc-Asp(OEpe)-OH

The peptide-resin was dried *in vacuo* and full cleavage was performed using cleavage cocktail K. The crude product (58 mg) was dissolved in (10 mL) 10% MeCN/H_2O + 0.1% TFA, purified by semi-preparative RP-HPLC (column [B], 10 - 50% MeCN/H_2O + 0.1% TFA over 35 min, and 8 mL/min flow rate) and lyophilized.

Yield of **5**: 5.43 mg (1.63 µmol, 9.4%). LC-MS: column [A], 5 - 70% MeCN/H_2O + 0.1% HCOOH over 15 min. **ESI-TOF-MS** of **5**: *m/z* $C_{146}H_{219}N_{39}O_{44}S_3$ (3318.5 Da) calculated: 1107.2 $[M+3H]^{3+}$, 830.6 $[M+4H]^{4+}$, 664.7 $[M+5H]^{5+}$, observed: 1107.2 $[M+3H]^{3+}$, 830.6 $[M+4H]^{4+}$, 664.7 $[M+5H]^{5+}$.



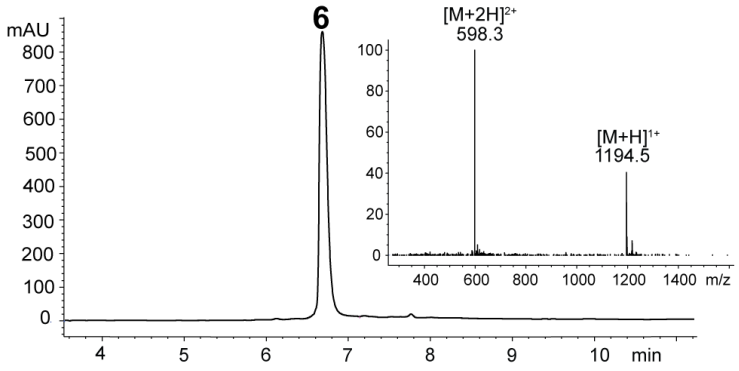
*Synthesis of CD59 peptide fragment **6** (69-78)*

The peptide synthesis was performed on functionalized resin **R1** (0.290 g, 0.092 mmol, loading 0.318 mmol/g). Fmoc-Ala-OH was coupled as described in the general methods. The peptide was elongated by MW-SPPS using a 5.4 molar excess of the amino acid. The table below shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue Number	Cycle
Asn-77, Glu-76, Leu-75, Gln-74, Glu-73, Phe-71	0.1 - Single deprotection single coupling 75 °C
Asn-72, Asn-70	0.1 - Double deprotection double coupling 75 °C
Cys-69	0.1 - Double deprotection double coupling 50 °C Final deprotection 75 °C

The peptide-resin was dried *in vacuo* and full cleavage was performed with cleavage cocktail B'. The crude product (71.7 mg) was dissolved in (15 mL) of 10% MeCN in H_2O (0.1% TFA), purified by semi-preparative RP-HPLC (column [B], 10 - 50% MeCN/H_2O + 0.1% TFA over 35 min, and 8 mL/min flow rate) and lyophilized.

Yield of **6**: 28.8 mg (24.09 µmol, 40.2%). LC-MS: column [A], 5 - 70% MeCN/H_2O + 0.1% HCOOH over 15 min. **ESI-TOF-MS** of **6**: *m/z* $C_{48}H_{74}N_{16}O_{18}S$ (1194.5 Da) calculated: 1195.5 $[M+H]^+$, 598.3 $[M+2H]^2$, observed: 1194.5 $[M+H]^+$, 598.3 $[M+2H]^{2+}$.



Synthesis of CD59 peptide fragment ***7*** *(26-44)*

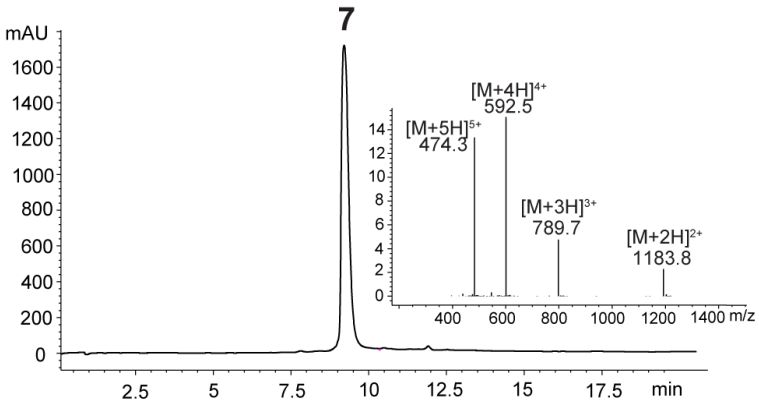
The peptide synthesis was performed on functionalized resin **R1** (0.250 g, 0.105 mmol, loading 0.42 mmol/g). Fmoc-Hys(Trt)-OH was coupled as described in the general methods. The peptide was elongated by MW-SPPS using 4.8 molar excess of the amino acid. The table below shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue Number	Cycle
Glu-43, Phe-42, Lys-41, Trp-40, Lys-38, Asn-37, Tyr-36, Val-35, Leu-33, Gly-32, Ala-31, Thr-29, Leu-27	0.1 - Single deprotection single coupling 75 °C
Gln-34, Lys-30, Ile-28	0.1 - Double deprotection double coupling 75 °C
Cys-39	0.1 - Double deprotection double coupling 50 °C
Cys-26	Manual coupling
	Final deprotection 75 °C

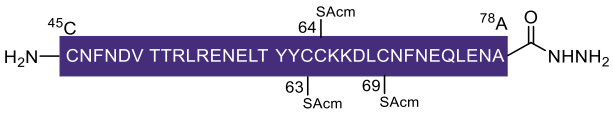
Cys-39: Fmoc-Cys(Acm)-OH
Cys-26 Fmoc-Cys(Trt)-OH

The peptide-resin support was dried *in vacuo* and full cleavage was performed using cleavage cocktail K. The crude product (56.9 mg) was dissolved in (10 mL) 10% MeCN/H_2O + 0.1% TFA, purified by semi-preparative RP-HPLC (column [B], 10 - 50% MeCN in H_2O (0.1% TFA) over 35 min, and 8 mL/min flow rate) and lyophilized.

Yield of **7**: 15.4 mg (6.51 µmol, 27%). LC-MS: column [A], 5 - 60% MeCN/H_2O + 0.1% HCOOH over 25 min. **ESI-TOF-MS** of **7**: *m/z* $C_{108}H_{168}N_{30}O_{26}S_2$ (2365.2 Da) calculated: 1183.6 $[M+2H]^{2+}$, 789.4 $[M+3H]^{3+}$, 592.3 $[M+4H]^{4+}$, 474.0 $[M+5H]^{5+}$, observed: 1183.8 $[M+2H]^{2+}$, 789.7 $[M+3H]^{3+}$, 592.5 $[M+4H]^{4+}$, 474.3 $[M+5H]^{5+}$.



*Synthesis of CD59 peptide fragment **8** (45-78)*



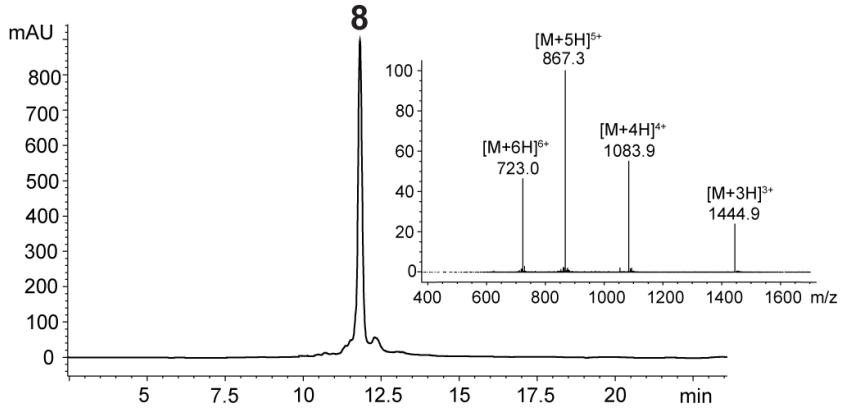
The peptide synthesis was performed on functionalized resin **R2** (0.26 g, 0.08 mmol, loading 0.32 mmol/g). Fmoc-Ala-OH was coupled as described in the general methods. The peptide was elongated by MW-SPPS using a 6.25 molar excess of the amino acid. The table below shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue Number	Cycle
Leu-75, Gln-74, Glu-73*, Phe-71, Ala-68	0.1 - Single deprotection single coupling 75 °C
Asn-77, Glu-76, Asn-72, Asn-70	0.1 - Double deprotection double coupling 75 °C
Cys-69, Cys-64, Cys-63	0.1 - Double deprotection double coupling 50 °C
Lys-66, Lys-65, Tyr-62, Tyr-61, Thr-60*, Leu-59, Glu-58, Asn-57, Glu-56, Leu-54*, Thr-52, Thr-51*, Val-50, Asn-48, Phe-47, Asn-46, Cys-45	0.1 - Single deprotection 25 °C double coupling 50 °C
Arg-55, Arg-53	0.1 - Double Arginine Coupling
Asp-49*	Manual coupling
	Final deprotection 25 °C

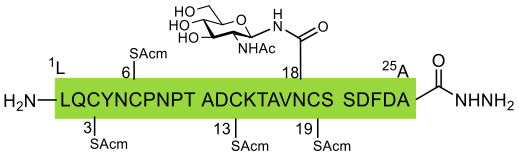
*Capping cycle
Cys-64, Cys-63: Fmoc-Cys(Acm)-OH
Asp-49: Fmoc-Asp(OEpe)-OH

The peptide-resin was dried *in vacuo* and full cleavage was performed with cleavage cocktail K. The crude product (155 mg) was dissolved in (30 mL) 10% MeCN/H_2O + 0.0% TFA, purified by semi-preparative RP-HPLC (column [B], 10 - 50% MeCN/H_2O + 0.1% TFA over 35 min, and 8 mL/min flow rate) and lyophilized.

Yield of **8**: 14.8 mg (3.42 µmol, 9.6%). LC-MS: column [A], gradient 5 - 60 % MeCN in H_2O (0.1 % HCOOH) over 25 min. **ESI-TOF-MS** of **8**: *m/z* $C_{182}H_{284}N_{54}O_{61}S_4$ (4330.0 Da) calculated: 1444.3 $[M+3H]^{3+}$, 1083.5 $[M+4H]^{4+}$, 867.0 $[M+5H]^{5+}$, 722.7 $[M+6H]^{6+}$, observed: 1444.9 $[M+3H]^{3+}$, 1083.9 $[M+4H]^{4+}$, 867.3 $[M+5H]^{5+}$, 723.0 $[M+6H]^{6+}$.



*Synthesis of CD59 glycopeptide fragment **9** (1-25)*



The peptide synthesis was performed on functionalized resin **R2** (0.39 g, 0.08 mmol, loading 0.21 mmol/g). Fmoc-Ala-OH was coupled as described in the general methods. The peptide was elongated by MW-SPPS using a 6.25 molar excess of the amino acid. The table below shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue Number	Cycle
Asp-24*	0.1 - Double deprotection double coupling 75 °C *
Phe-23, Ser 21, Ser 20*, Val-17, Ala-16, Thr-15*, Lys-14, Ala-11, Thr-10, Pro-9, Asn-8, Pro-7*, Asn-5*, Tyr-4, Gln-2, Leu-1	0.1 - Single deprotection 25 °C double coupling 50 °C
Cys-19, Cys-13, Cys-6, Cys-3*	0.1 - Double deprotection double coupling 50 °C
Asn-18*, Asp-22*, Asp-12*	Manual coupling

* Capping cycle

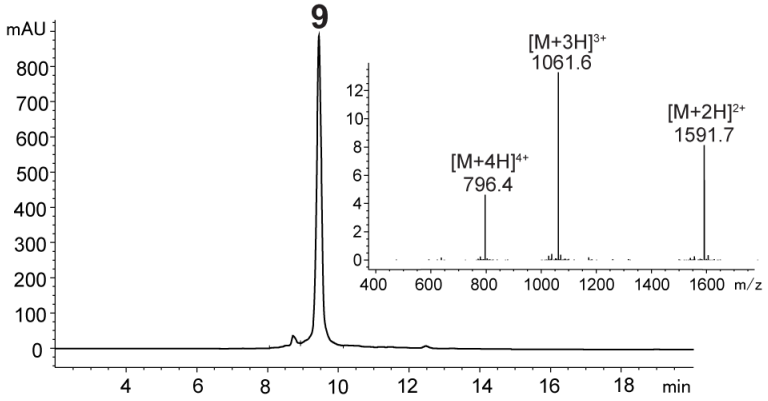
Cys-19, Cys-13, Cys-6, Cys-3: Fmoc-Cys(Acm)-OH

Asn-18: Fmoc-L-Asn((Ac)$_3$-β-D-GlcNAc)-OH

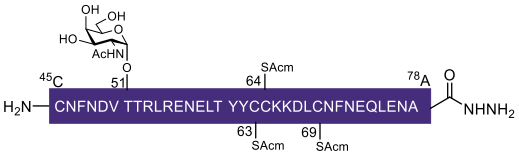
Asp-22 and Asp-12: Fmoc-Asp(OEpe)-OH

The peptide-resin was removed from the synthesizer and treated with a solution 10% $N_2H_4 \cdot H_2O$ in THF/MeOH (1:1) during 2 hours. The resin was washed with CH_2Cl_2, MeOH, CH_2Cl_2 and dried *in vacuo*. The full cleavage was performed with cleavage cocktail K. The crude peptide (132 mg) was dissolved in (30 mL) 10% MeCN/H_2O + 0.1% TFA, purified by semi-preparative RP-HPLC (column [B], 10 - 50% MeCN/H_2O + 0.1% TFA over 35 min, and 8 mL/min flow rate) and lyophilized.

Yield of **9**: 17.13 mg (5.38 µmol, 13%). LC-MS: column [A], 5 - 60% MeCN/H_2O + 0.1% HCOOH over 25 min. **ESI-TOF-MS** of **9**: *m/z* $C_{129}H_{201}N_{37}O_{49}S_4$ (3180.3 Da) calculated: 1591.2 $[M+2H]^{2+}$, 1061.1 $[M+3H]^{3+}$, 796.1 $[M+4H]^{4+}$, observed: 1591.7 $[M+2H]^{2+}$, 1061.6 $[M+3H]^{3+}$, 796.4 $[M+4H]^{4+}$.



*Synthesis of peptide hydrazide **10***



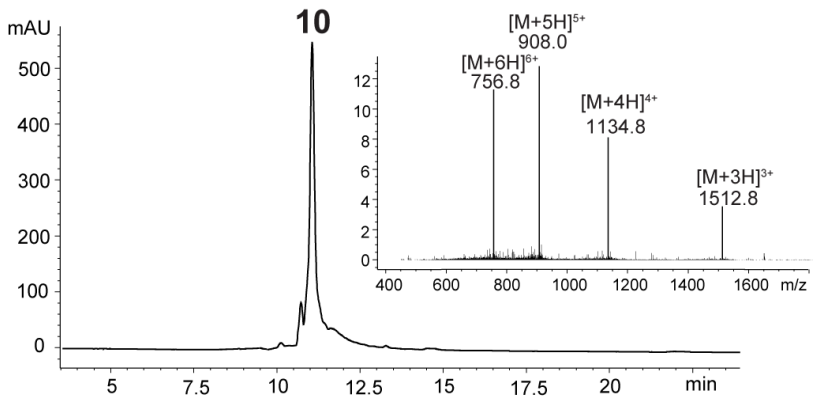
The peptide synthesis was performed on functionalized resin **R2** (0.39 g, 0.08 mmol, loading 0.20 mmol/g). Fmoc-Ala-OH was coupled as described in the general methods. The peptide was elongated by MW-SPPS using a 6.25 molar excess of the amino acid. The table below shows the deprotection and coupling cycles for each amino acid according to the parameters described in 2.5.2.

Residue Number	Cycle
Leu-75, Gln-74, Glu-73*, Phe-71, Ala-68	0.1 - Single Deprotection Single Coupling 75 °C
Asn-77, Glu-76, Asn-72, Asn-70	0.1 - Double Deprotection Double Coupling 75 °C
Cys-69, Cys-64, Cys-63	0.1 - Double Deprotection Double Coupling 50 °C
Lys-66, Lys-65, Tyr-62, Tyr-61, Thr-60*, Leu-59, Glu-58, Asn-57, Glu-56, Leu-54*, Thr-52	0.1 - Single Deprotection 25 °C Double Coupling 50 °C
Arg-55, Arg-53	0.1 - Double Arginine Coupling
Asp-67, Thr-51*, Val-50, Asp-49, Asn-48*, Phe-47, Asn-46, Cys-45	Manual coupling

*Capping cycle
Cys-64, Cys-63: Fmoc-Cys(Acm)-OH
Thr-51: Fmoc-L-Thr((Ac)$_3$-α-D-GalNAc)-OH
Asp-49: Fmoc-Asp(OEpe)-OH

The peptide-resin was removed from the synthesizer and treated with a solution 10% $N_2H_4 \cdot H_2O$ in MeOH during 2 hours. The resin was washed with CH_2Cl_2, MeOH, CH_2Cl_2 and dried *in vacuo*. The full cleavage was performed with cleavage cocktail K. The crude peptide (106 mg) was dissolved in (20 mL) 10% MeCN/H_2O + 0.1 % TFA, purified by semi-preparative RP-HPLC (column [B], 10 - 50 % MeCN/H_2O + 0.1 % TFA over 35 min, and 8 mL/min flow rate) and lyophilized.

Yield of **10**: 10.54 mg (2.32 µmol, 10%). LC-MS: column [A], 5 - 60 % MeCN in H_2O + 0.1 % HCOOH over 25 min. **ESI-TOF-MS** of **10**: *m/z* $C_{190}H_{297}N_{55}O_{66}S_4$ (4533.0 Da) calculated: 1512.0 $[M+3H]^{3+}$, 1134.3 $[M+4H]^{4+}$, 907.6 $[M+5H]^{5+}$, 756.5 $[M+6H]^{6+}$, observed: 1512.8 $[M+3H]^{3+}$, 1083.9 $[M+4H]^{4+}$, 867.3 $[M+5H]^{5+}$, 723.0 $[M+6H]^{6+}$.



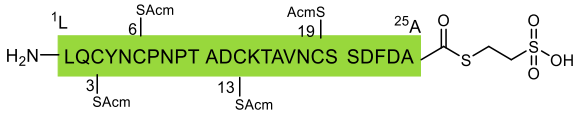
2.5.5. Ligation of peptides and glycopeptides protocols and chromatographic data

General method for the thioester formation and ligation reactions

The buffers were freshly prepared before every ligation reaction. The pH was adjusted using a pH meter VWR phenomenal (VWR international GmbH) previously calibrated at pH 4.0, 7.0, and 10.0. The reactions were performed in 2 mL Eppendorf tubes equipped with a rubber septum and a small magnetic stir bar to ensure the mixing of reactants under inert atmosphere. The employed buffers are listed below:

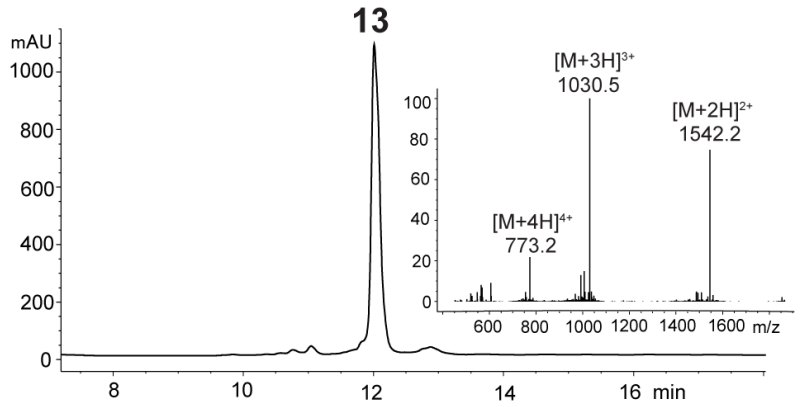
Diazotization Buffer*:* 6 M GdnHCl, 0.2 M NaH_2PO_4, pH 3
Ligation Buffer: 6 M GdnHCl, 0.1 M NaH_2PO_4, 0.05 M MPAA. 0.015 M TCEP, pH 6.8
Acm Cleavage Buffer: 6 M GdnHCl, 0.2 M NaH_2PO_4, pH 6.9. Buffer was degassed with argon for 30 min.

*Synthesis of peptide thioester **13***

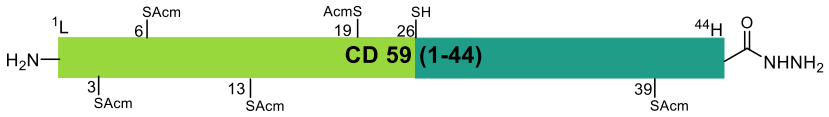


Peptide hydrazide **3** (5.01 mg, 1.68 µmol) was dissolved in 400 µL of diazotization buffer, cooled down to -15 °C, and a 0.5 M aqueous solution of $NaNO_2$ (34 µL, 16.88 µmol) was added under stirring. After 20 min a solution of MESNa (26 mg, 159 µmol) in 50 µL of 1 M NaOH was added and the pH of the reaction was adjusted to 6.8 with 1 M NaOH. The reaction was slowly warmed to room temperature and stirred for 1 h. Subsequently, the reaction was diluted with (1 mL) 10% MeCN in H_2O and the product was purified by semi-preparative RP-HPLC (column [C], 10 - 45% MeCN/H_2O + 0.1% TFA over 30 min, and flow rate 1.8 mL/min) and lyophilized.
Yield of **13**: 2.99 mg (0.96 µmol, 57%). LC-MS: column [A], 5 - 60% MeCN/H_2O + 0.1% HCOOH over 25 min. **ESI-TOF-MS** S of **13**: *m/z* $C_{123}H_{190}N_{34}O_{47}S_6$ (3087.2 Da) calculated: 1544.6 $[M+2H]^{2+}$, 1030.1 $[M+3H]^{3+}$, 772.8 $[M+4H]^{4+}$, observed: 1512.8 $[M+3H]^{3+}$, 1083.9 $[M+4H]^{4+}$, 867.3

$[M+5H]^{5+}$, 723.0 $[M+6H]^{6+}$. **HR-MS** of **13**: *m/z* $C_{123}H_{190}N_{34}O_{47}S_6$ (3087.1847 Da) 1544.5924 $[M+2H]^{2+}$, 1030.0616 $[M+3H]^{3+}$, observed: 1544.5619 $[M+2H]^{2+}$, 1030.0450 $[M+3H]^{3+}$.

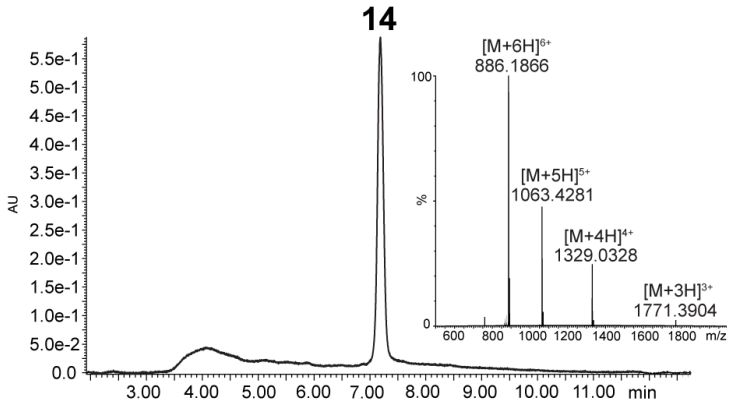


Synthesis of Peptide hydrazide 14

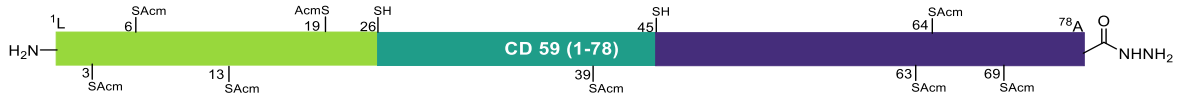


Peptide thioester **13** (2.99 mg, 0.96 µmol) and peptide hydrazide **7** (2.91 mg, 1.23 µmol) were dissolved separately in 240 µL of ligation buffer and bubbled with argon. The solution of peptide **7** was added to the Eppendorf containing **13** (2 mM final concentration) and stirred at rt overnight. Once complete ligation was shown by RP-HPLC, TCEP (5.5 mg, 19.2 µmol) was added to the mixture and incubated during 30 min. The reaction was diluted with (1 mL) 10% MeCN in H_2O and the product was purified in two batches by semi-preparative RP-HPLC (column [C], 5 - 50% MeCN/H_2O + 0.1% TFA over 25 min, 1.8 mL/min) and lyophilized.

Yield of **14**: 1.87 mg (0.35 µmol, 37%). LC-MS: column [A], 5 - 60% MeCN in H_2O (0.1% HCOOH) over 25 min. **HR-MS** of **14**: *m/z* $C_{229}H_{352}N_{64}O_{70}S_6$ (5310.4276 Da) calculated: 1771.1425 $[M+3H]^{3+}$, 1328.6069 $[M+4H]^{4+}$, 1063.0855 $[M+5H]^{5+}$, 886.0713 $[M+6H]^{6+}$, observed: 1771.3904 $[M+3H]^{3+}$,1329.0328 $[M+4H]^{4+}$, 1063.4281 $[M+5H]^{5+}$, 886.1866 $[M+6H]^{6+}$.

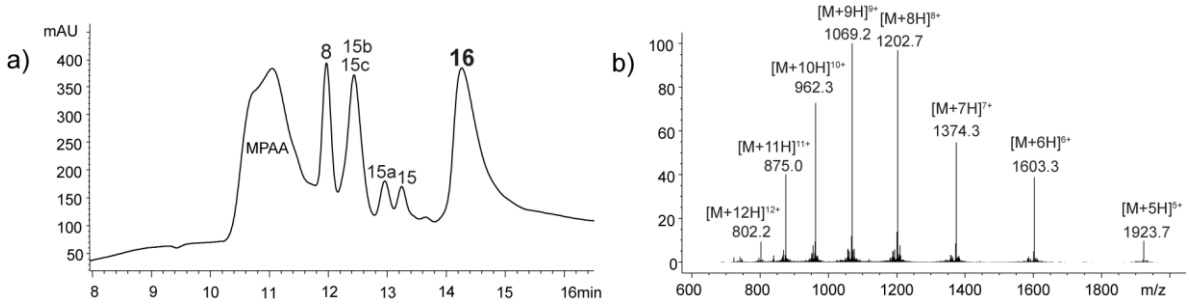


Synthesis of non-glycosylated protein CD59 ***16***

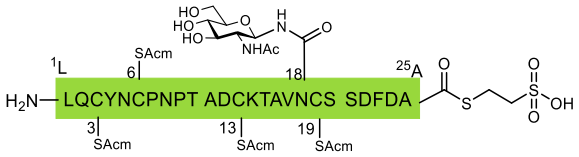


Peptide Hydrazide **14** (1.87 mg, 0.35 µmol) was dissolved in 80 µL of diazotization buffer and cooled down to -15 °C. A 0.5 M aqueous solution of $NaNO_2$ (7 µL, 3.5 µmol) was added and the mixture was stirred for 20 min. Subsequently, a solution of MESNa (5.5 mg, 33.4 µmol) in 20 µL of 0.5 M NaOH was added and the pH of the reaction was adjusted to 6.8 with 1 M NaOH. The reaction was slowly warmed to room temperature and once formation of the thioester **15** was confirmed by RP-HPLC, peptide hydrazide **8** (1.87 mg, 0.43 µmol) dissolved in 80 µL of ligation buffer was added to the reaction Eppendorf and stirred at rt under argon overnight. Once no further conversion was shown by RP-HPLC, TCEP (10.10 mg, 35.2 µmol) and 50 µL of water were added, and the mixture was incubated for 30 min to obtain the Acm-Cys protected non-glycosylated protein CD59 **16.**

Yield of **16**: Not isolated. LC-MS: column [A], 5 - 60 % MeCN/H_2O + 0.1 % HCOOH over 25 min. **ESI-TOF-MS** of **16**: *m/z* $C_{411}H_{632}N_{116}O_{131}S_{10}$ (9608.4 Da) calculated: 1512.0 $[M+3H]^{3+}$, 1134.3 $[M+4H]^{4+}$, 907.6 $[M+5H]^{5+}$, 756.5 $[M+6H]^{6+}$, observed: 1512.8 $[M+3H]^{3+}$, 1083.9 $[M+4H]^{4+}$, 867.3 $[M+5H]^{5+}$, 723.0 $[M+6H]^{6+}$.

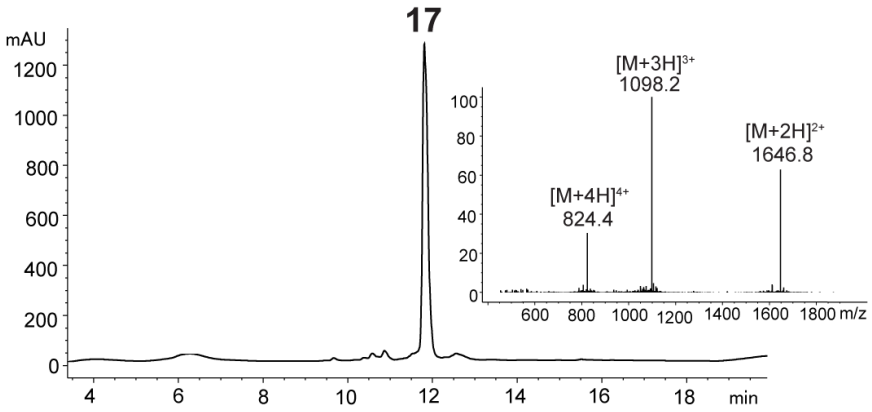


Synthesis of thioester 17

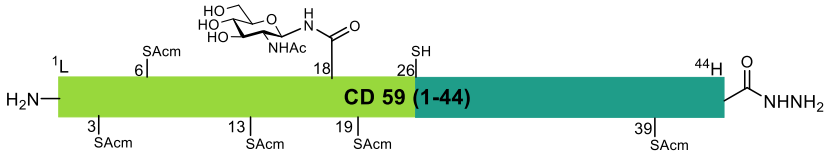


Glycopeptide hydrazide **17** (3.39 mg, 1.06 µmol) was dissolved in 300 µL of diazotization buffer and cooled down to -15 °C. A 0.5 M aqueous solution of $NaNO_2$ (21 µL, 10.65 µmol) was added and the mixture was stirred for 20 min. Subsequently, a pre-cooled to -10 °C solution of MESNa (17 mg, 104 µmol) in 40 µL of 1 M NaOH was added and the pH of the reaction was adjusted to 6.8 with 1 M NaOH. The reaction was slowly warmed to room temperature and stirred for 1 h. After dilution with (1 mL) 10% MeCN/H_2O, the peptide hydrazide **17** was purified by semi-preparative RP-HPLC (column [C], 10 - 45% MeCN in H_2O (0.1 % TFA) over 30 min, and 1.8 mL/min flow rate) and lyophilized.

Yield of **17**: 2.05 mg (0.62 µmol, 58%). LC-MS: column [A], 5 - 60% MeCN/H_2O + 0.1% HCOOH over 25 min. **ESI-TOF-MS** of **17**: *m/z* $C_{131}H_{203}N_{35}O_{52}S_6$ (3290.3 Da) calculated: 1646.1 $[M+2H]^{2+}$, 1097.8 $[M+3H]^{3+}$, 823.6 $[M+4H]^{4+}$, observed: 1646.8 $[M+2H]^{2+}$, 1098.2 $[M+3H]^{3+}$, 824.4 $[M+2H]^{2+}$. **HR-MS** of **17**: *m/z* $C_{123}H_{190}N_{34}O_{47}S_6$ (3290.2641 Da) calculated: 1646.1320 $[M+2H]^{2+}$, 1097.7547 $[M+3H]^{3+}$, observed: 1646.1047 $[M+2H]^{2+}$, 1097.7361 $[M+3H]^{3+}$.

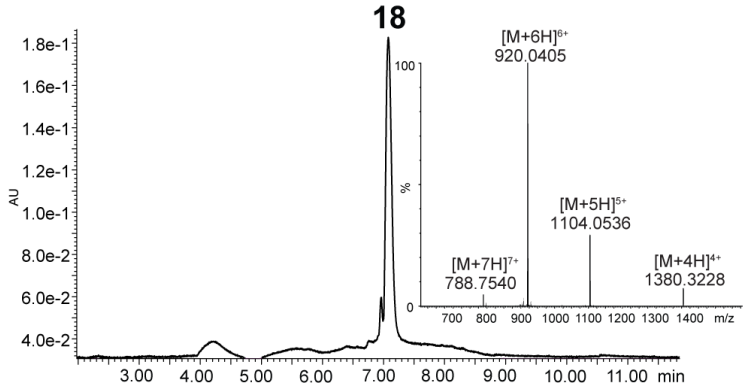


Synthesis of Peptide hydrazide 18

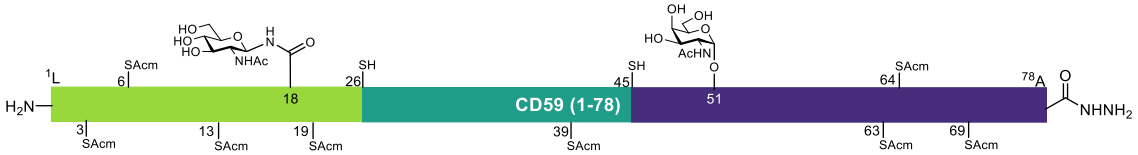


Petide thioester **17** (2.05 mg, 0.62 µmol) and peptide hydrazide **7** (1.62 mg, 0.68 µmol) were dissolved separately in 150 µL of ligation buffer each and bubbled with argon. The solution of peptide **7** was added to the Eppendorf containing **17** (2 mM final concentration) and kept at rt overnight. Once complete ligation was shown by RP-HPLC, TCEP (3.6 mg, 12.5 µmol) was added to the mixture and the reduction was incubated for 30 min. Subsequently, the mixture was diluted with (1 mL) 10% MeCN in H_2O, and the product was purified in two batches by semi-preparative RP-HPLC (column [C], 5 - 50 % MeCN in H_2O (0.1 % TFA) over 25 min, and 1.8 mL/min flow rate) and lyophilized.

Yield of **18**: 1.51 mg (0.27 µmol, 44%). LC-MS: column [A], 5 - 60 % MeCN/H_2O + 0.1 % HCOOH over 25 min. **HR-MS** of **18**: *m/z* $C_{237}H_{365}N_{65}O_{75}S_6$ (5513.5070) calculated: 1379.3767 $[M+4H]^{4+}$, 1103.7014 $[M+5H]^{5+}$, 919.9178 $[M+6H]^{6+}$, 788.6439 $[M+7H]^{7+}$, observed: 1380.3228 $[M+4H]^{4+}$, 1104.0536 $[M+5H]^{5+}$, 920.0405 $[M+6H]^{6+}$, 788.7540 $[M+7H]^{7+}$.

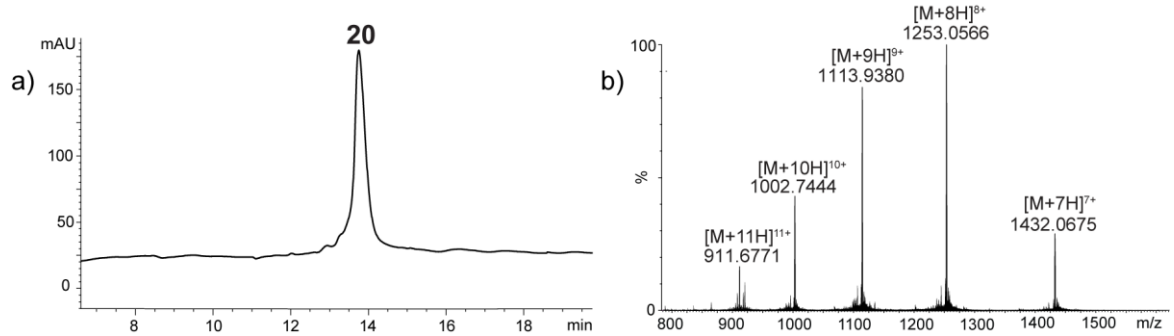


*Synthesis of peptide hydrazide **20***

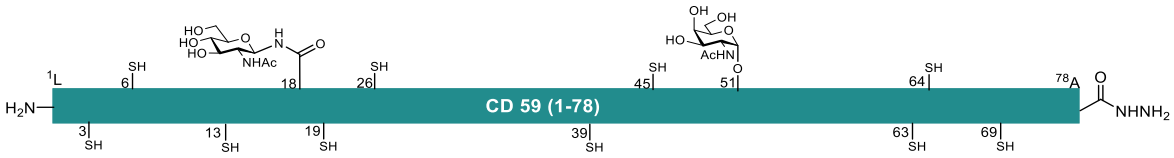


Peptide hydrazide **18** (1.51 mg, 0.27 µmol) was dissolved in 60 µL of diazotization buffer and cooled down to -15 °C. A 0.5 M aqueous solution of $NaNO_2$ (6 µL, 2.97 µmol) was added and the mixture was stirred for 20 min. Subsequently, a cooled to -10 °C solution of MESNa (4 mg, 24.4 µmol) in 20 µL of 0.5 M NaOH was added and the pH of the reaction was adjusted to 6.8 with 1 M NaOH. The reaction was slowly warmed to room temperature. Once RP-HPLC confirmed formation of the thioester **19**, the peptide hydrazide **10** (1.5 mg, 0.33 µmol) dissolved in 60 µL of ligation buffer was added to the reaction and the mixture was stirred at rt under argon overnight. Upon complete ligation by RP-HPLC, TCEP (8 mg, 27.9 µmol) in water (200 µL) was added, and the mixture was incubated for 30 min**.** After final dilution to (1 mL) with 10% MeCN in H_2O, the Acm-protected glycosylated protein CD59 **20** was purified in two batches by semi-preparative RP-HPLC (column [B], 15 - 45 % MeCN in H_2O (0.1% TFA) over 30 min, and 8 mL/min flow rate) and lyophilized.

Yield of **20**: 0.13 mg (0.013 µmol, 5%). LC-MS: column [A], 5 - 60% MeCN/H_2O + 0.1% HCOOH over 25 min. **HR-MS** of **18**: *m/z* $C_{427}H_{658}N_{118}O_{141}S_{10}$ (10014.5153) calculated: 1431.6450 $[M+7H]^{7+}$, 1252.8144 $[M+8H]^{8+}$, 1113.7239 $[M+9H]^{9+}$, 1002.4515 $[M+10H]^{10+}$, 911.4105 $[M+11H]^{11+}$, observed: 1432.0675 $[M+7H]^{7+}$, 1253.0566 $[M+8H]^{8+}$, 1113.9380 $[M+9H]^{9+}$, 1002.7444 $[M+10H]^{10+}$, 911.6771 $[M+11H]^{11+}$.

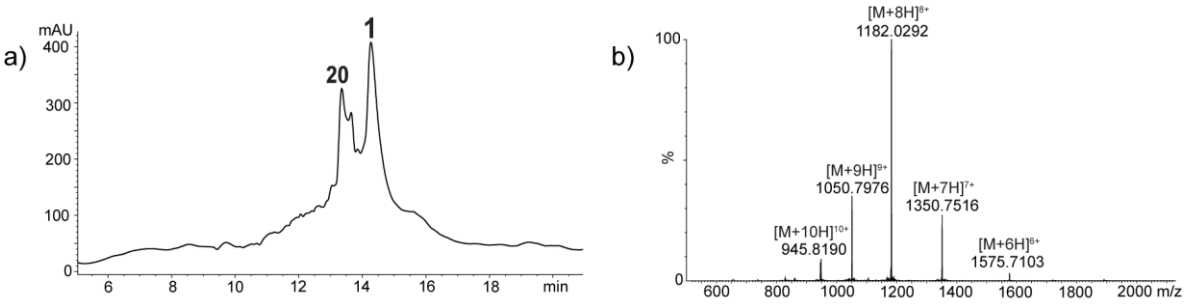


*Removal of Acm group - Synthesis of glycoprotein **1** (Acm removal)*



Acm-protected CD59 glycoprotein **20**

Anhydrous $PdCl_2$ (1.20 mg, 6.77 µmol) was suspended in 500 µL of Acm-cleavage buffer and shaken until complete dissolution to get a 13.5 mM Pd^{II} solution. 29 µL of the Pd^{II} stock solution (68.9 µg $PdCl_2$, 0.389 µmol, 30 eq) were added to the Acm-protected glycoprotein **20**. The mixture was stirred under argon and the exclusion of light. After 2h, A 10 µL aliquot of the reaction mixture was diluted to 100 µL with 10% MeCN in H_2O, quenched by addition of 100 equiv DTT (regarding the Pd content in the aliquot) and analyzed by LC-MS. Since non-representative UV signal was observed in the chromatogram, the rest of the reaction mixture (19 µL) was quenched and analyzed by LC-MS to give a mixture of Acm-protected glycoprotein **20** and fully deprotected CD59 glycoprotein **1**.



Yield of **1**: Not isolated. LC-MS: column [A], 5 - 60 % MeCN/H_2O + 0.1 % HCOOH over 25 min. **HR-MS** of **18**: *m/z* $C_{403}H_{618}N_{110}O_{133}S_{10}$ (9446.2184) calculated: 1575.3697 $[M+6H]^{6+}$, 1350.4598 $[M+7H]^{7+}$, 1181.7773 $[M+8H]^{8+}$, 1050.5798 $[M+9H]^{9+}$, 945.6218 $[M+10H]^{10+}$, observed: 1575.7103 $[M+6H]^{6+}$, 1350.7516 $[M+7H]^{7+}$, 1182.0209 $[M+8H]^{8+}$, 1050.7976 $[M+9H]^{9+}$, 94.8190 $[M+10H]^{10+}$.

3. Establishing a methodology for the automated synthesis of O-glycosyl amino acids

3.1. Glycans in the mucins of *T. cruzi* Y and Colombiana strains

The structure of mucin-like molecules (MLMs) varies in number and distribution of glycans, as explained in section **1.3.2.1**. The O-linked carbohydrates display significant structural variability and changes depending on the strain and the parasite's lifestage[222]. The structure of O-glycans present in the mucins of *T. cruzi* Y strain (**30-36**) and Colombiana strain (**36-38**) was determined in 1995[223] and 2004[224], respectively. Both structures consist of branched glycans that vary in length and have as main constituents galactose β-(1→ *3*), β-(1→ 4), or β-(1→ 6) linked to a core glucosamine, which is α-linked to threonine. Glycans with a 4,6-branched glucosamine are common in both strains. In addition, β-(1→ 3) and β-(1→ 6)-linked galactose is present in Y strain[223]. The main difference between the glycans in Y or Colombiana strain is the occurrence of β-D-Galf(*1→4*) in the MLMs of the Colombiana strain[224] (Figure 27).

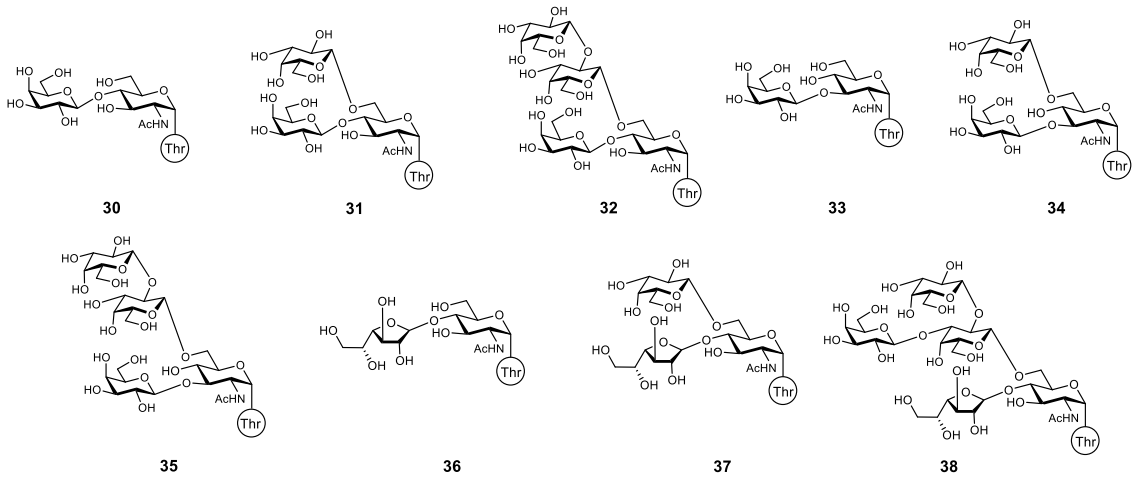


Figure 27 - Chemical structures of O-glycans of *T. cruzi* Y and Colombiana strain

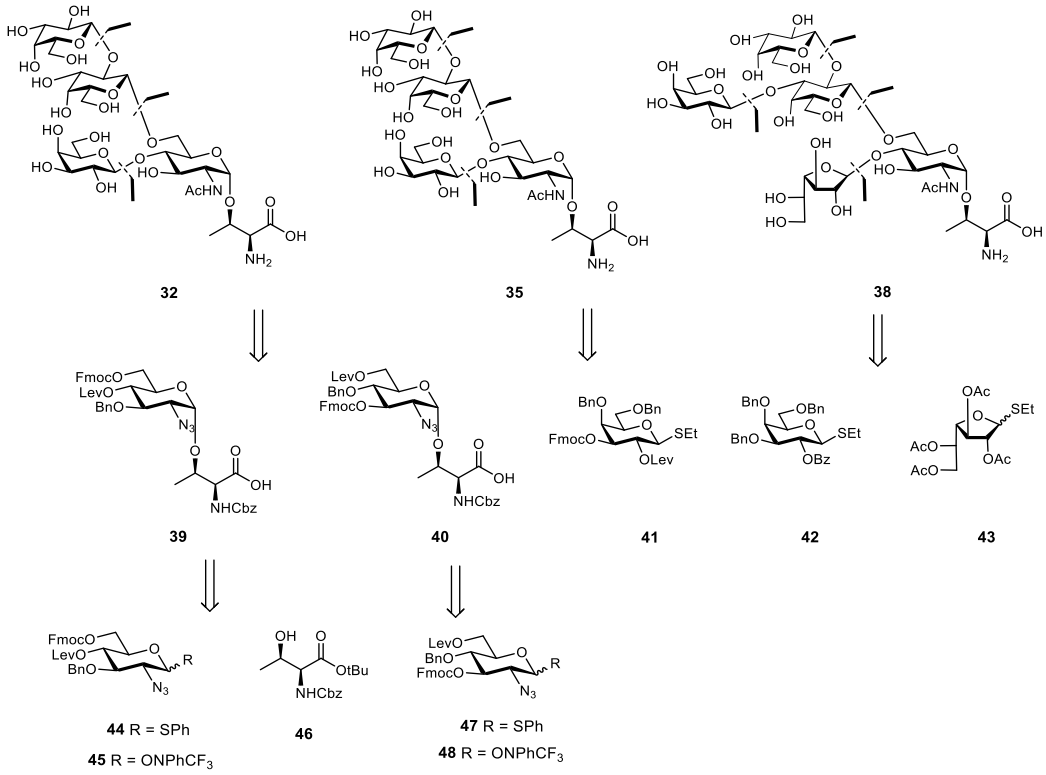
The chemical synthesis of *T. cruzi* glycans has been widely explored over the last 30 years. The reports include synthesis up to hexasaccharides from Colombiana strain [232, 234, 262-266] and tetrasaccharide from *T. cruzi* strain[267]. However, the synthesis of glycosylated amino acids is limited and described only in a few reports[237]. The main limitations for accessing these molecules derive from the intricate formation of the 1,2-cis linkage between the amino acid and the glycan, as well as the molecule lability in basic conditions. The need for homogeneous structures to study the biology of the pathogen motivated the need for their chemical synthesis. In addition, mucin-like molecules are potential markers and targets for diagnostics, prevention, and the prognosis of Chagas disease[211, 268-269]. Furthermore, they can contribute to combating other worldwide impacting, lifelong, life-threatening, or even possible epidemic diseases. Therefore, a significant contribution to the field would be the development of a general methodology for synthesizing glycosylated amino acids derived from *T. cruzi* employing Automated Glycan Assembly (AGA).

3.2. Results and discussion

3.2.1. Retrosynthesis

The O-linked glycans from *T. cruzi* Y and Colombiana strains contain three monosaccharides (D-GlcNAc, D-Gal, D-Gal*f*) and the amino acid (L-Thr). The glycans are connected through β-glycosidic linkages, while an α-linkage is present between the glucosamine and the amino acid. An AGA of these molecules requires at least six orthogonally protected building blocks, five monosaccharides, and one amino acid (Scheme 38).

A retrosynthetic analysis performed for structures **32, 35,** and **38** revealed that these structures are accessible after a series of post-AGA manipulations of fully protected glycans, such as removing permanent protective groups and acetylation of the amine to obtain *N*-acetylglucosamine. The AGA of the structures requires galactose building blocks **41**, **42** and **43** and glycosylated amino acid **39** or **40**, which are constructed from glucosamine **44** or **47** and threonine **46** (Scheme 38).



Scheme 38 - Retrosynthesis of glycans present in *T. cruzi* MLMs of Y and Colombiana strains

Thioglycosides **41**, **42,** and **43** are suitable building blocks for AGA and were selected for stability and reactivity[156, 270]. The hydroxyl groups of these monosaccharides contained Fmoc and levulinoyl ester (Lev) as temporary protecting groups, benzyl ethers as permanent non-participating groups, and a benzoyl ester as a permanent participating group. The amine of glucosamine was masked as an azide to avoid participation and assure α-selectivity. The (Cbz) carbamate was selected for the amino group protection in threonine (Figure 28).

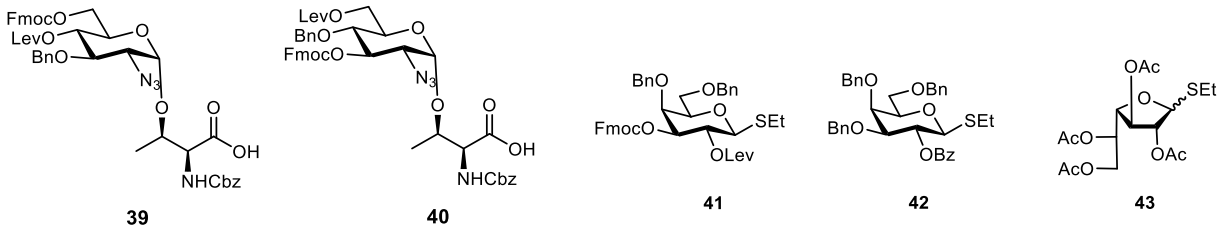


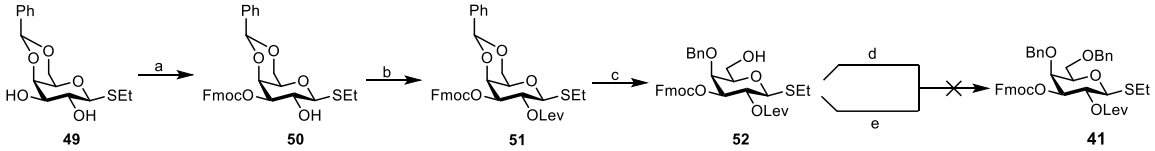
Figure 28 - Set of building blocks for Automated Glycan Assembly

To ensure α-linkage of the amino acid and glycan, the glycosylated amino acids **39** and **40** were synthesized prior to AGA. The process required temporary protection of the α-carboxylate as *tert* butyl ester **46** and glycosyl imidate donors. Thioglycoside building blocks **44** and **47** were converted into imidates **45** and **48** as described in the next section. After glycosylation of Thr **46**, the hydrolysis of the tert butyl ester delivered a carboxylate suitable for esterification of the solid support.

3.2.2. Building block synthesis

Synthesis of galactose building block 41

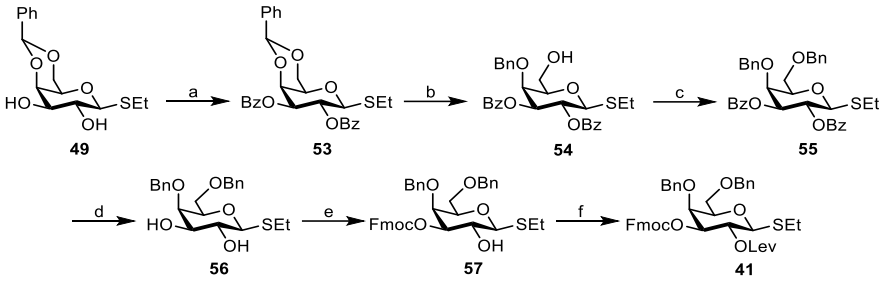
The galactose building block **41** was necessary to synthesize the target structures **38**, **32**, and **35** and requires regioselective protection at the C2 and C3 hydroxyls for branched-chain elongation. The initial strategy to get a synthesis of **41** required five steps from commercially available precursor **49**. The process started with a regioselective Fmoc-protection via a stannylene ketal to enhance the nucleophilicity of the C3-OH[271] position and delivered the alcohol **50** in good yield (Scheme 39).



Scheme 39 - Four-step strategy for the synthesis of 41. Reagent and conditions: a) Bu_2SnO, Fmoc-Cl, toluene, 70%; b) LevOH, DIC, DMAP, CH_2Cl_2, 80%; c) BH_3·THF, TMSOTf, CH_2Cl_2, 65%; d) BnBr, Ag_2O, CH_2Cl_2; e) TriBOT, TfOH, CH_2Cl_2.

The following protection of the C2-OH to obtain intermediate **51** involved the addition of DMAP as a catalyst and showed problems derived from basicity. The base induced the Fmoc group removal and formation of non-desired products reducing the yield to 80%. A regioselective reductive opening of benzylidene with BH_3·THF complex and TMSOTf delivered the C4-O benzylated product **52** in a good yield and was ready for the final benzylation to complete the synthesis of building block **41**. The benzylation of the C6-O under basic conditions presented the main difficulty in the strategy. Due to two base-labile ester groups and their easy removal with NaH, the reactions can result in perbenzylated products and the loss of the desired orthogonality. Benzylation of **52** was attempted under acidic conditions[272] using benzyl trichloroacetimidate and neutral conditions

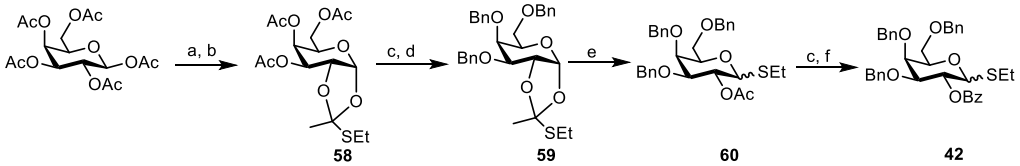
employing silver(I) oxide[273]. However, these reactions showed low to no conversion or ended in the decomposition of the starting material.



Scheme 40 - Six-step strategy for the synthetis of 41. Reagent and conditions: a) Bz_2O, DMAP, Et_3N, CH_2Cl_2, 70%; b) $BH_3 \cdot THF$, TMSOTf, CH_2Cl_2, 75%; c) HMDS, TMSOTf, PhCHO, Et_3SiH, CH_2Cl_2, 92%; d) NaOMe, MeOH, 96%; e) Bu_2SnO, Fmoc-Cl, toluene, 87%; f) LevOH, DIC, DMAP, CH_2Cl_2, 83%.

To overcome the problem with the benzylation, a synthetic route to get **41** was designed using a six-step protocol and alternative temporary protecting groups in the intermediates (Scheme 40). In the first step, commercially available precursor **49** was benzoylated, giving the galactose **53** in 70% yield. A following regioselective reductive opening of the benzylidene, under similar conditions used for **52**, provided compound **54** in good yield. Previous reports described the conversion of **54** into the fully protected **55** under basic conditions[274]. However, this reaction was not reproducible and delivered galactose **55** in a meager yield. Best results delivered a one-pot benzylation of **54** via TMSOTf-catalyzed silylation[275] in benzaldehyde and triethylsilane, giving **55** in a high yield of 92%. Following steps involved saponification of benzoyl esters with sodium methanolate to get diol **56** and the regioselective Fmoc protection of C3-OH via a stannylene acetal. Final esterification of the remaining hydroxyl group in **57** using levulinic acid and DIC/DMAP for activation afforded the target building block **41** in 83% yield. Although this synthetic route included more steps due to alternative protection, the single synthetic steps were high-yielding and delivered the desired building block in 40% overall yield (Scheme 40).

Synthesis of galactose building block 42

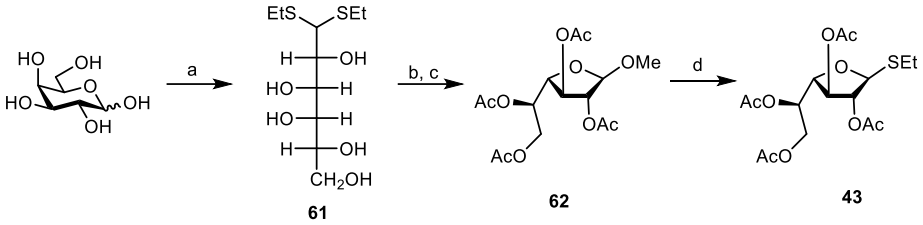


Scheme 41 - Synthetic route for the synthesis of building block 42. Reagents and conditions: a) HBr/AcOH, b) EtSH, Bu_4NBr, 2,6-lutidine, CH_3NO_2, 71% over two steps; c) NaOMe, MeOH; d) BnBr, NaH, DMF, 88% over two steps; e) TMSOTf, EtSH, CH_2Cl_2, 73%; f) BzCl, DMAP, pyridine, CH_2Cl_2, 94% over two steps.

The synthesis of galactose building block **42** (Scheme 41) involved an analogous procedure to the one described by Thijssen[276]. The process started with the bromination of peracetylated galactose and was followed by subsequent bromide conversion into the orthoester **58** by

treatment with ethanethiol and 2,6-lutidine as a sterically-hindered base. The obtained orthoester was deacetylated under Zemplen conditions, and the resulting triol was benzylated via Williamson etherification to give the galactose orthoester **59** in 85% yield. Due to the instability of the orthoester during purification by silica gel column chromatography, these reactions proceeded with modest yields. The process continued with thioglycoside **60** formation by a Lewis acid promoted activation and rearrangement of orthoester **59** in the presence of ethanethiol. The addition of ethanethiol was necessary to displace the equilibrium and avoid the decomposition of the orthoester. Deacetylation under Zemplen conditions, followed by esterification of the resulting alcohol with benzoyl chloride, delivered the target galactose donor **42** in 43% overall yield.

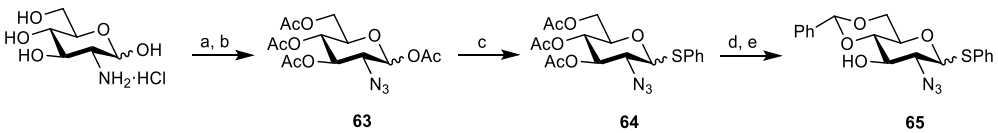
Synthesis of galactofuranose building block 43



Scheme 42 - Synthetic route for the synthesis of building block 43. Reagents and conditions: a) EtSH, HCl, 55%; b) I_2, MeOH; c) Ac_2O, pyridine, 87% over two steps; d) EtSH, $BF_3{\cdot}OEt_2$, CH_2Cl_2, 75%.

The synthesis of building block **43** from D-galactose involved four steps and required the opening of the pyranose ring and the corresponding cyclization to galactofuranose. Starting from D-galactose, a simultaneous acidic opening of the cyclic acetal and reaction of the aldehyde with ethanethiol delivered the dithioacetal **61** in 55 % yield. Next, cyclization of **61** by activating the thiol with iodine in methanol and following acetylation of the remaining hydroxyl groups delivered the methyl galactofuranoside **62**. As described in literature[277], the reaction delivered a 9:1 (β:α) anomeric mixture that was separated to give the β-linked methyl galactoside in good yield. Finally, hydrolysis of the methyl acetal in **62** with $BF_3{\cdot}Et_2O$ and following reaction with ethanethiol delivered the target thioglycoside **43** in 36% overall yield (Scheme 42).

Synthesis of glucosamine building blocks 44 and 47

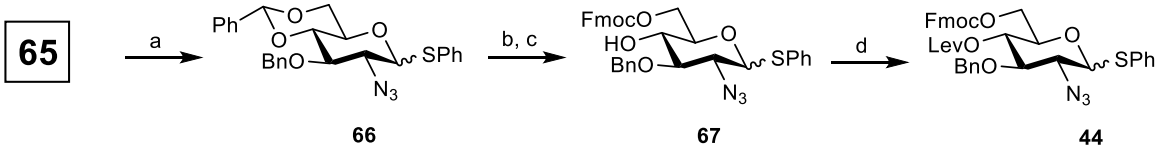


Scheme 43 - Synthetic route to glucosamine intermediate 65. Reagents and conditions: a) TfN_3, $CuSO_4$, Et_3N, ACN; b) Ac_2O, pyridine, DMAP, 77% over two steps; c) PhSH, $BF_3{\cdot}OEt_2$, CH_2Cl_2, 64%; d) NaOMe, MeOH; e) benzaldehyde dimethyl acetal, CSA, ACN, 89%.

The synthesis of glucosamines **44** and **47** required the preparation of the benzylidene acetal **65** as a common intermediate (Scheme 43). This precursor allows selective functionalization on 3-OH and mild acidic cleavage of the benzylidene permit further protection. Following reported

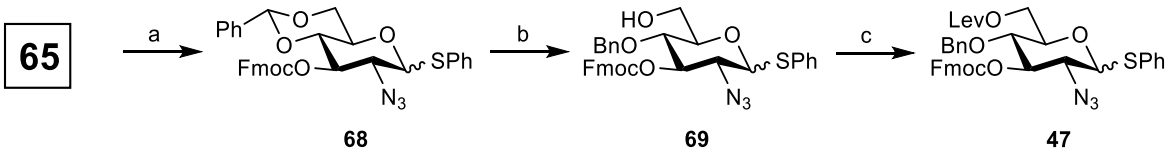
methods[278], the process initiated with the conversion of D-glucosamine hydrochloride into the corresponding glycosyl azide **63** by a diazo transfer reaction using triflic azide and the following peracetylation with acetic anhydride. Treatment of **63** with thiophenol and $BF_3{\cdot}OEt_2$ yielded thioglycoside **64**. This step showed low yield due to the difficulty of getting complete conversion of the starting material, even by adding an excess of Lewis acid. Alternate reactions were screened, changing the Lewis acid to $SnCl_4$ and TfOH; however, the outcome was the overreaction of the thiol and subsequent substitution of the azide. Thus, the reaction was scaled up under $BF_3{\cdot}Et_2O$ activation. Finally, compound **64** was deacetylated under Zemplen conditions, and the obtained crude triol was reacted with benzaldehyde dimethyl acetal and CSA to get the benzylidene acetal **65** in 89% yield. The obtained derivative **65** was split into two portions and used to synthesize **44** and **47** as described below.

The synthesis of the glucosamine **44** with orthogonal groups for 4,6 branching started with the reaction of intermediate **65** with benzyl bromide and sodium hydride to afford **66** in a good yield of 81%. Next, the benzylidene acetal was hydrolyzed under acidic conditions, and the resulted diol was protected at the primary alcohol using stochiometric Fmoc-Cl in pyridine to produce **67**. Finally, Steglich esterification of the remaining hydroxyl group with levulinic acid, DIC, and DMAP afforded the target building block **44** in 77% yield (Scheme 44).



Scheme 44 - Synthesis of glucosamine building block 44. Reagents and conditions: a) BnBr, NaH, DMF, 81%; b) p-TSA, MeOH:CH_2Cl_2 (4:1); c) Fmoc-Cl, pyridine, CH_2Cl_2, 80% over two steps d) LevOH, DIC, DMAP, CH_2Cl_2, 77%

The glucosamine functionalized with orthogonal protecting groups at positions 3 and 6 was synthesized from the intermediate **65** (Scheme 45)**.** The first step involved the C3-OH protection using Fmoc-Cl and pyridine to obtain **68**. Next, this product was subjected to the regioselective reductive opening of benzylidene acetal using $BH_3{\cdot}THF$ and catalytic amounts of TMSOTf to afford **69** in 62 %. In the final step, the remaining primary alcohol of **69** was reacted with levulinic acid, DIC, and DMAP to yield the target glucosamine building block **47**.

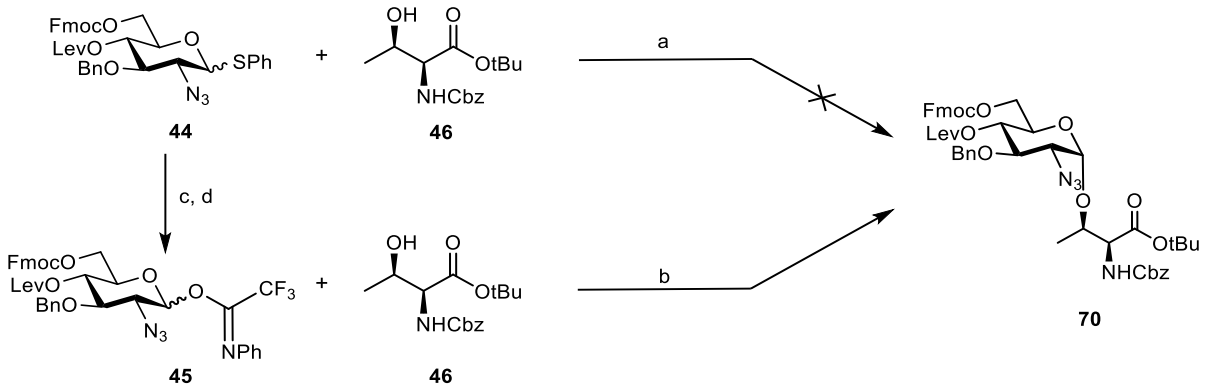


Scheme 45 - Synthetic route to building block 47. Reagents and conditions: a) Fmoc-Cl, pyridine, CH_2Cl_2, 78%; b) $BH_3{\cdot}THF$, TMSOTf, CH_2Cl_2, 62%; c) LevOH, DIC, DMAP, CH_2Cl_2, 68%.

Synthesis of glycosylated amino acids 39 and 40

The control of the stereochemical outcome of a glycosylation reaction depends on various factors, including the neighboring group participation for PG at C2, remote anchimeric assistance (remote group participation), conformation-restraining groups, solvent, and temperature effects. However, the introduction of 1,2-*cis* glycosidic bonds still represents a challenge in carbohydrate chemistry[146]. Products having α-configuration are accessible by using several strategies and can profit from various effects and the thermodynamic stabilization given by an anomeric effect. Thus, the synthesis of glycosylated amino acids **70** and **71** was envisioned considering the absence of neighboring group participation at C2, remote anchimeric assistance, and solvent and temperature effects.

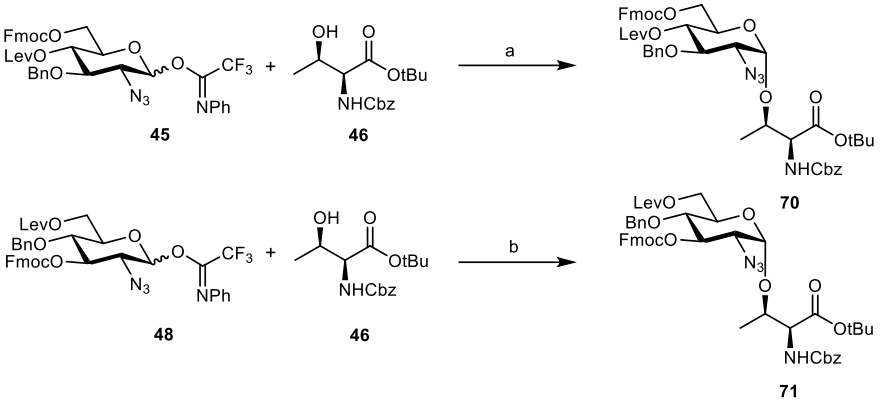
The amine group from glucosamine was masked as azide, a C2 non-participating moiety. The glycosylations of the amino acid were performed in a solvent mixture containing ether and at room temperature. Ethers can interact with the oxocarbenium ion formed during the glycan donor activation blocking the β-face for an attack of the nucleophile and favoring the selective α-product formation[279]. In addition, the α-glycoside is thermodynamically favored, and it is predominantly formed at high temperatures[146].



Scheme 46 - Glycosylation conditions for the synthesis of 70. Reagents and conditions: a) NIS/TfOH, Et_2O:CH_2Cl_2 (6:1) 0 °C to rt; b) TMSOTf, Et_2O:CH_2Cl_2 (6:1), (44%, only α), -15 °C to r.t; c) NBS, acetone:H_2O (10:1) d) ClC=N(Ph)CF_3, Cs_2CO_3, DMF, 83% over two steps.

Glycosylation of donor **44** and threonine acceptor **46** was initially attempted using NIS and triflic acid (TfOH) as promoters in Et_2O:CH_2Cl_2 (6:1) as a solvent mixture. Under these conditions, no product was detected after 45 min and further additions of NIS and TfOH (Scheme 46). Identical results were found during the attempts to glycosylate the amino acid **46** with the phenyl thioglycoside **47**. These results showed a lack of reactivity from the thioglycosides to form the glycosylated amino acid. TLC and mass spectrometry monitoring of the reaction progress indicated that the unmodified donor was present during the whole reaction, suggesting that activation was not occurring. Likely, the lack of reactivity of the thioglycoside donor was derived from the electron-withdrawing effect of the ester protecting groups. Since the building block was designed for AGA, the levulinoyl ester could not be changed. Therefore more reactive donors, imidates **45** and **48,** were selected for the synthesis.

Glucosamine donors **44** and **47** were converted into the corresponding imidates by hydrolysis of the thioglycosides employing NBS for activation (Scheme 46). The reaction of the obtained hemiacetals with 2,2,-trifluoro-*N*-phenylacetimidoyl chloride in the presence of cesium carbonate delivered the *N*-phenyl trifluoroacetimidates **45** and **48** in good yield after two steps.



Scheme 47 - Synthesis of protected glycosylated amino acids 70 and 71. Reagents and conditions: a) TMSOTf, Et_2O:CH_2Cl_2 (6:1), -15 °C to rt

Glycosylation of the imidate donor **45** or **48** with threonine acceptor **46** in Et_2O:CH_2Cl_2 delivered the desired products **70** and **71** in excellent α-selectivity (Scheme 47). It is not clear which modification of the donor contributed to the high selectivity; it seems to result from a combination of solvent and temperature effects, a non-participating group at C2, and remote anchimeric assistance from the levulinoyl ester and the Fmoc group at the C6-position. Large amounts of hydrolyzed donors accompanied the formation of the products, which were isolated in modest yields (~42%). Further optimization of the reaction conditions using a lower temperature (-25 °C) did not affect the reaction outcome but reduced the reaction rate.

Table 3 – Evaluation of no acid **46** glycosylation using donors **45** and **48** at different scales

Entry	Donor	Donor µmol	Acceptor µmol	Promoter µmol	Product µmol	Yield %
1	45	63.55	76.26	11.44	28	44
2	48	63.55	76.26	11.44	27.12	43
3	45	259.3	259.3	25.3	99.23	38*
4	48	317.8	317.8	31.78	134.51	42
5	45	1270	1270	127.1	153.3	12
6	45	1450	1450	145	549.1	38*

* Accumulated yield for five different batches

The glycosylation outcome with of amino acid **46** with donors **45** and **48** was evaluated at different scales and the results are compiled in Table 3. In the initial experiment, the acceptor was added in a 1.2 molar excess (Entry 1). Albeit that threonine was always recovered after chromatographic purification, this reaction delivered the product in only 44% yield and represented the best conditions for glycosylation. A reaction of **46** with imidate **48** delivered the product in a similar yield

to the glycosylation with donor **46** (Entry 2), suggesting a similar reactivity of both building blocks. Using an equimolar ratio of acceptor and donor did not induce differences in the reaction result, but it slightly reduced the yield to 42% when the reaction was scaled up five fold (Entry 4).

The next step was upscaling the reaction to obtain enough material for AGA. Unexpectedly, glycosylation using ~1.3 mmol of both the donor and acceptor (upscaling of 20 times) resulted in a drastic drop in the product yield to 12% (Entry 5). Further optimizations and experiments did not show any improvement in the reaction results. Due to the reaction with imidate **48** delivered better results using ~300 µmol of material (medium-size scale), the reaction of **46** with **45** was performed in multiple batches at this scale. After combining the crude product of all batches in a single purification, the product **70** was isolated with a modest yield of 38% (Entry 6). This strategy was used to obtain enough material to continue with the deprotection of the carboxylate.

Having the fully protected glycosylated amino acids **70** and **71**, the hydrolysis of the *t*-butyl ester was performed as reported in the literature for analogous molecules[280-281] employing trifluoroacetic acid. Treatment of the fully protected compounds **70** and **71** afforded **39** and **40** in yields of 75% and 70%, respectively. Despite the reported stability for the Cbz protective group, the concentrated acid affected the carbamate and partially removed the protection from the α-amino group delivering a mixture of products that required purification.

3.2.3. Development of a strategy for the assembly of glycosylated amino acids of *T. cruzi* Y strain

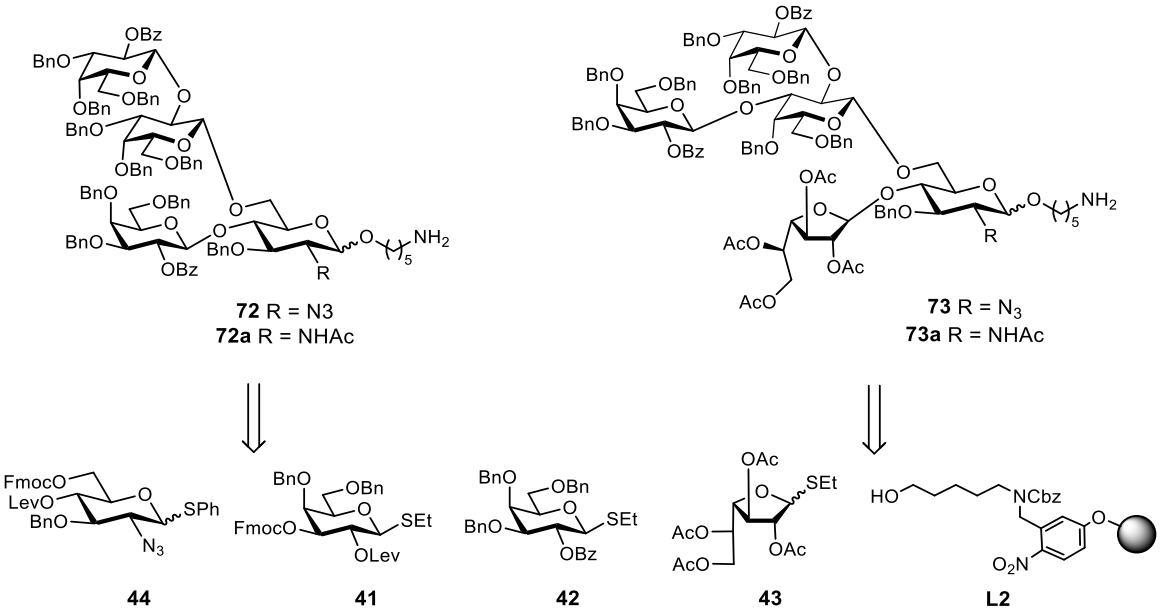
During the last 20 years, AGA has shown an excellent versatility for synthesizing linear and branched oligosaccharides derived from mammals, bacteria, and plants[141]. The methodology efficiently delivered molecules with a broad scope of monosaccharides such as Glc, Gal, Man, GalNAc, GlcNAc, GlcA, Rha, and Fuc. However, an expansion of AGA to obtain structures having glycans modifications is still underexplored. In a pioneer study, the synthesis of glycosylated threonine and a short protected glycopeptide was demonstrated using this methodology[282]. However, the reported method presents regular selectivity and any strategy for deprotection of the obtained molecules was presented. Despite the enormous progress in the glycan assembly and the importance of glycosylated amino acids, no further synthesis of this type of molecule using AGA has been reported. Considering the need to expand the limits of AGA and explore its applicability to obtain glycosylated amino acids that can be used as building blocks for the synthesis of glycopeptides or be screened as epitopes in biological applications, the development of a strategy for the AGA of glycosylated threonine derivatives from MLMs of *T. cruzi* was explored in this work.

Automated glycan assembly methodology requires a Merrifield resin equipped with a linker and involves iterative cycles of glycosylation, capping, and selective deprotection steps. Some requirements and challenges in AGA were considered for synthesizing glycosylated amino acid building blocks for glycopeptide synthesis, i. e. the need for a free carboxylate in the final product

for further activation, which required modifications of the AGA standard conditions and linkers. Further considerations and possible side reactions to be studied during the synthesis included amino acid epimerization and β-elimination of the attached O-glycans in basic conditions.

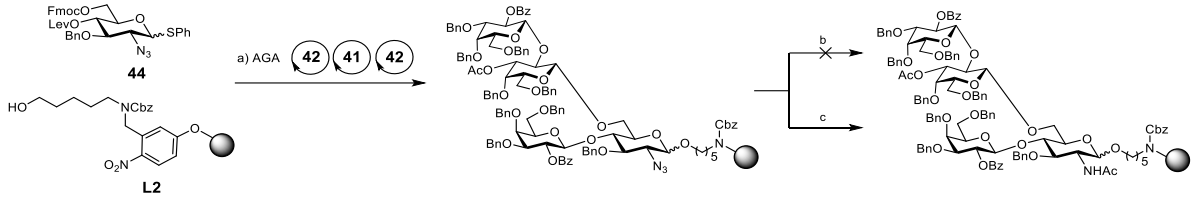
Initial screening of viability of automated glycan assembly

To establish the glycosylation conditions for the assembly of glycosylated amino acids from *T. cruzi* MLM, two representative structures (**72** and **73**) of *T. cruzi* Y and Colombiana strains were initially synthesized (Scheme 48). The synthesis of these molecules was evaluated to verify the suitability of building blocks **41-44** and the AGA conditions for the synthesis of the tetrasaccharides **72** and **73** using previously reported conditions for analogous building blocks[170]. The process progress and product characterization involved small-scale cleavage of the compounds from the solid support and MS monitoring by MALDI-TOF MS.



Scheme 48 – Retrosynthesis of target molecules for initial synthesis

The automated glycosylations were completed on a home-built synthesizer using six equivalents of thioglycoside donor per cycle (based on resin loading) and NIS/TfOH activation. The general method comprised delivering building blocks to the reaction vessel at -20 °C and an incubation time of 5 min followed by activator delivery. Subsequently, the temperature of the reaction mixture was increased to 0 °C and maintained for 20 min. The capping of unreacted hydroxyl groups was performed with acetic anhydride and methanesulfonic acid. The selective removal of the temporary protecting groups was completed prior to glycosylations. Fmoc removal was accomplished by treatment with piperidine and the levulinoyl ester was removed with hydrazine acetate. After assembly of the oligosaccharide, the resin was transferred to a polypropylene reactor to continue with the Post-AGA steps.



Scheme 49 – Synthesis of tetrasaccharide 72 of *T. cruzi* Y strain MLM. Reagents and conditions: a) AGA: Protective group removal (Fmoc: piperidine in DMF, Lev: hydrazine acetate in pyridine/AcOH/H_2O; Acidic wash: TMSOTf in CH_2Cl_2; Glycosylation: building block (4 eq), NIS/TfOH, CH_2Cl_2, -20 °C (5 min) to 0 °C (20 min); Capping: Ac_2O, MsOH, CH_2Cl_2; b) PBu_3, THF:H_2O, o.n.; c) AcSH, o.n.

The assembly of the carbohydrates was initially performed using the amino pentanol linker **L2** to standardize the conditions for the glycan synthesis, then this knowledge was transferred to the preparation of the further glycosyl amino acids. The synthesis of tetrasaccharide **72** started with the glycosylation of Merrifield resin bearing a photo linker **L2** with glucosamine **44** (Scheme 49). Then, tetrasaccharide **72** was assembled using iterative cycles of protective group removal (Fmoc or Lev), glycosylation with galactose building blocks **42** and **41**, and capping. Due to the low reactivity observed with glucosamine **44,** glycosylations with this building block required 40 min of reaction time for completion. MALDI-TOF MS analysis of the crude product showed an efficient synthesis of the tetrasaccharide without deletion sequences (Figure 29).

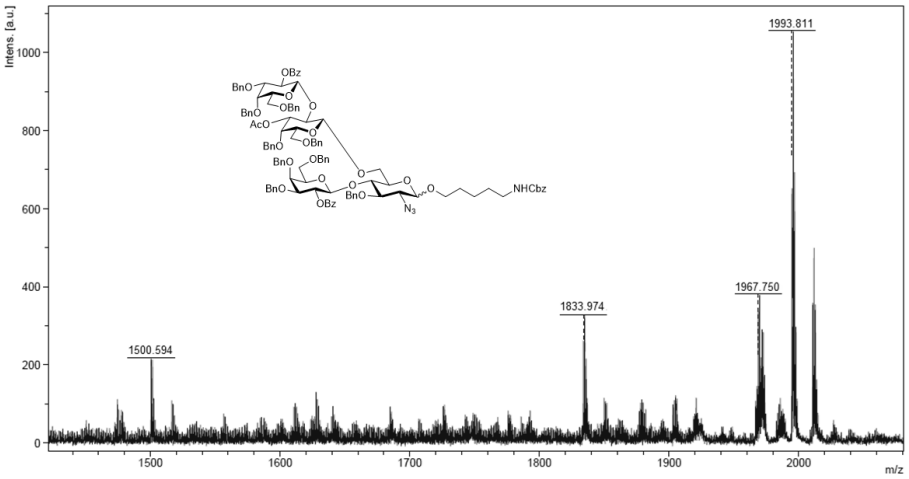
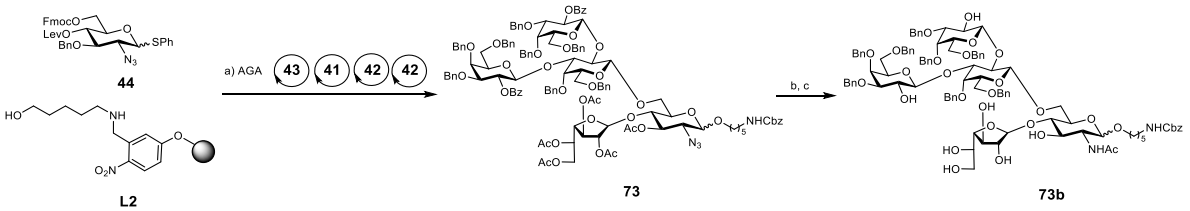


Figure 29 – MS spectra after AGA of tetrasaccharide 72. MALDI-TOF MS of **72**: m/z Calc for $C_{116}H_{122}N_4O_{25}Na$ [M+Na]+ 1993.830, observed 1993.811.

Oligosaccharide **72** linked to resin was used to standardize the reduction of the azide and the corresponding acetylation of the resulting amine to *N*-acetyl glucosamine (Scheme 49). Different conditions reported in literature for analogous structures were evaluated[282-284]. Staudinger reduction with tributylphosphine in THF did not deliver the expected on-resin azide reduction. MS analysis showed conversion of the starting material and formation of a product assigned to the iminophosphorane intermediate, which did not hydrolyze to the corresponding amine. Further addition of fresh reagents did not modify the outcome of the reaction. A change of the phosphine to trimethylphosphine resulted in no reaction; MALDI-TOF MS neither showed the mass of the product nor from the iminophosphorane.

Additional attempts to perform the hydrolysis of the product obtained with tributylphosphine during the amine acetylation with acetic anhydride and pyridine did not show the further reaction of this material; thus, this methodology was abandoned. On the contrary, the treatment of the azide **72** with an excess of thioacetic acid for two days reduced the azide and delivered on-resin the acetylated amine in a single step. However, further efforts to reduce the reaction time were not successful and resulted in incomplete acetylation of glucosamine. Therefore, despite the reaction duration, this method was preferred as an effective methodology to perform on-resin conversion of azide into an acetamide in a one-pot process.

Next, the synthesis of pentasaccharide **73** was evaluated as a proof of concept for the suitability of the strategy when synthesizing complex branched structures (Scheme 50). This structure displays more complexity than **72**. Oligosacharide **73** contains a branched glycan with a double substitution on the glucosamine and galactose residues, requiring orthogonal removal of temporary protecting groups during the AGA. In addition, the assembly requires glycosylation with a galactofuranose donor. Considering the possible difficulties in following steps and the results described in a previous report[285], the peracetylated thioglycoside **43** was selected as the most convenient donor.



Scheme 50 – Synthesis of pentasaccharide 73 of T. cruzi colombiana strain. Reagents and conditions: a) AGA: Protective group removal (Fmoc: piperidine in DMF, Lev: hydrazine acetate in pyridine/AcOH/H_2O; Acidic wash: TMSOTf in CH_2Cl_2; Glycosylation: building block (4 eq), NIS/TfOH, CH_2Cl_2, -20 °C (5 min) to 0 °C (20 min); Capping: Ac_2O, MsOH, CH_2Cl_2; b) AcSH, o.n. c) NaOMe in THF/MeOH (4:1), 2 h.

Solid support **L2** was initially glycosylated with glucosamine **44,** employing the conditions evaluated in the synthesis of **72**. Unreacted groups on the linker were capped under the standard conditions, the Lev group was removed with hydrazine acetate, and the resin was washed with DMF and CH_2Cl_2. Elongation at C4-position of glucosamine was performed first due to the know lower nuclephilicity of C4-OH. Glycosylation with galactofuranose **43** was followed with elongation of the glycan at the C6-position of glucosamine by iterative cycles with galactose building blocks **41**, **42** up to complete the structure.

The analysis of the product obtained after the assembly by MALDI-TOF MS showed the mass corresponding to the desired product **73** and the mass of a product containing a deletion of galactose **42**. Two glycosylations were performed with this building block; thus, it was not possible to determine by MS which glycosylation step was incomplete. Considering the different reactivity of galactose hydroxyl groups, C3-OH > C2-OH[286], this product was hypothesized to an incomplete

formation of the β-Gal-(1→2)-β-Gal linkage. Nevertheless, the product formation in this preliminary synthesis demonstrated the viability of the strategy. Thus, the remaining material was used to evaluate the post-AGA steps to complete the synthesis of **73**. Treatment of the glycan-resin with thioacetic acid showed complete conversion of the resulting amine. Lastly, previously described methods for on-resin ester saponification with sodium methoxide were used to hydrolyze the benzoyl and acetyl esters. The reaction proceeded smoothly, with sodium methoxide delivering the desired pentasaccharide.

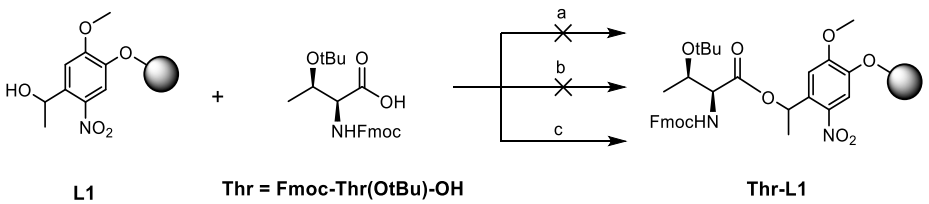
The syntheses of **72** and **73** were performed as a preliminary study to evaluate the AGA viability of these structures employing building blocks **41** – **44**. The glycans bound to the resins did not receive further treatment. However, the established protocols were the base for synthesizing the glycosylated threonine using AGA.

Coupling of the glycosylated amino acid to the solid support

Screening of the coupling to the solid support (traceless linker)

Several reports have demonstrated the versatility of the traceless linker in glycan synthesis obtaining molecules with a free reducing end for direct use[162, 167, 284] or for further transformation into glycosyl donors to synthesize polysaccharides[138]. Nevertheless, building blocks with a free carboxylate are required for using glycosylated amino acids in glycopeptide synthesis. Thus, it was necessary establishing the application of an ester for the amino acid-resin linkage and its late-stage hydrolysis to release the free carboxylic acid. Some available resins and linkers for SPPS fulfill this requirement. However, the Merrifield resin bearing the traceless photolabile linker[167] was envisioned to avoid further AGA optimization.

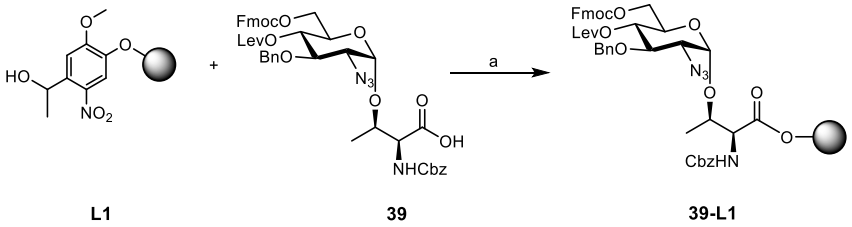
The coupling of threonine to the secondary alcohol of a nitrophenyl linker was considered a good alternative for attaching the amino acid to the resin. This linkage is stable to the AGA conditions and enables access to the carboxylate by a photocleavage. Initial coupling conditions involved an 4 molar excess of activated Fmoc-Thr-(OtBu)-OH and overnight reaction with the hydroxyl group of the linker **L1** (Scheme 51). Fmoc cleavage and quantification of the dibenzofulvene-piperidine adduct were used to determine the coupling efficiency. Three general methods were evaluated for activating the amino acid and the coupling: a) by the formation of asymmetric anhydride of the amino acid, b) activation using the aminium reagent HATU to produce the active ester, and c) via formation of an acyl *N*-methylimidazolium cation.



Scheme 51 - Threonine coupling to the solid support bearing a traceless linker. Reagents and conditions: a) Thr, DIC, pyridine, CH_2Cl_2:DMF (1:1), o.n.; b) Thr, HATU, DIPEA, CH_2Cl_2:DMF (1:1), o.n.; c) Thr, DIC, *N*-methylimidazole, CH_2Cl_2:DMF (1:1), o.n.

Although these three methodologies are commonly efficient in amide bond formation, two of them failed to give the desired ester. The poor reactivity was attributed to the low nucleophilicity of alcohols compared to amines, especially that of the secondary alcohol in the photolabile linker to attack the carbonyl group. In contrast, acyl activation using *N*-methylimidazolium cation formation afforded the formation of the ester **Thr-L1**. In this case, the high reactivity of the electrophile balanced the poor nucleophilicity of the alcohol in the reaction.

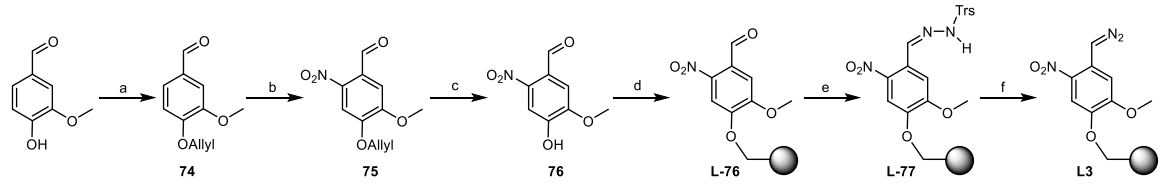
The next step was to establish the conditions for coupling a glycosylated amino acid. Building block **39** activated with DIC was treated with N-methylimidazole and added to the resin-linker **L1** (Scheme 52). Previously used glycosylation protocols with a mannose building block resulted in loading for the traceless linker of 0.34 mmol/g[167]. Quantification of the Fmoc removal showed an efficient glycosylated amino acid coupling to the resin and delivered a resin loading of 0.24 mmol/g, which was in agreement with the expected values. MALDI-TOF MS monitoring of the small-scale photocleavage products showed the effective glycosylated amino acid removal from the resin and the carboxylate function integrity.



L1 **39** **39-L1**

Scheme 52 - Coupling of glycosylated amino acid to the traceless linker. Reagents and conditions: a) DIC*, N*-methylimidazole, CH_2Cl_2:DMF (1:1), o.n.

Despite the excellent result of the first coupling of the glycosylated amino acids **39** and **40** to the resin-linker **L1**, the reaction was not reproducible even by using a significant excess of expensive building blocks **39** or **40**. The variations in the loading and the low nucleophilicity observed for resin-linker **L1** affected the assembly of more complex structures. Thus, the standard form of this traceless linker was discharged for the synthesis of more significant glycosyl amino acids using AGA. A new linker modification was required to get an easy ester bond formation while fulfilling the stability requirements for AGA. To accomplish this a linker was designed as a photolabile nitroveratryl (NV) type-linker bearing a diazo function as a leaving group. The efficient synthetic approach to get this linker attached to the solid support was envisioned considering previous literature reports[287-288], and it involved the linker synthesis and its conjugation of the resin.



74 **75** **76** **L-76** **L-77** **L3**

Scheme 53 - Synthesis of photolabile traceless diazo linker. Reagents and conditions: a) Allyl bromide, K_2CO_3, ethanol, 96%; b) TFA, KNO_3, 60 °C 88%, c) $Pd(PPh_3)_4$, K_2CO_3, MeOH, 79%; d) Merrifield resin, **76**, Cs_2CO_3, TBAI, CH_2Cl_2; e) 2,4,6-triisopropylbenzenesulfonyl hydrazide, THF; f) DBU, DMF.

The linker synthesis began with the allyl protection of vanillin using allyl bromide and potassium carbonate to obtain **74**, which was nitrated with potassium nitrate and TFA giving **75** in 88% yield. Next, compound **75** was deallylated by treatment with $Pd(PPh_3)_4$ in methanol, and the resulting alcohol **76** was directly attached to Merrifield resin using cesium carbonate and TBAI giving the resin-linked aldehyde **L-76**. The coupling progress was monitored and confirmed by FT-IR analysis of the resin and the appearance of the IR signals of the aldehyde (1689 cm^{-1}), nitro (1515 cm^{-1}), C-N stretching, and C-O stretching (1280 cm^{-1}, 1060 cm^{-1}), respectively (Figure 30). Next, the resin equipped with **L-76** was treated with 2,4,6-triisopropylbenzenesulfonyl hydrazide to give the corresponding hydrazone **L-77**, exhibiting a disappearance of the IR carbonyl signal at 1689 cm^{-1}. A final treatment of L-77 with DBU delivered the dark orange photolabile diazo functionalized resin **L3** (Scheme 53).

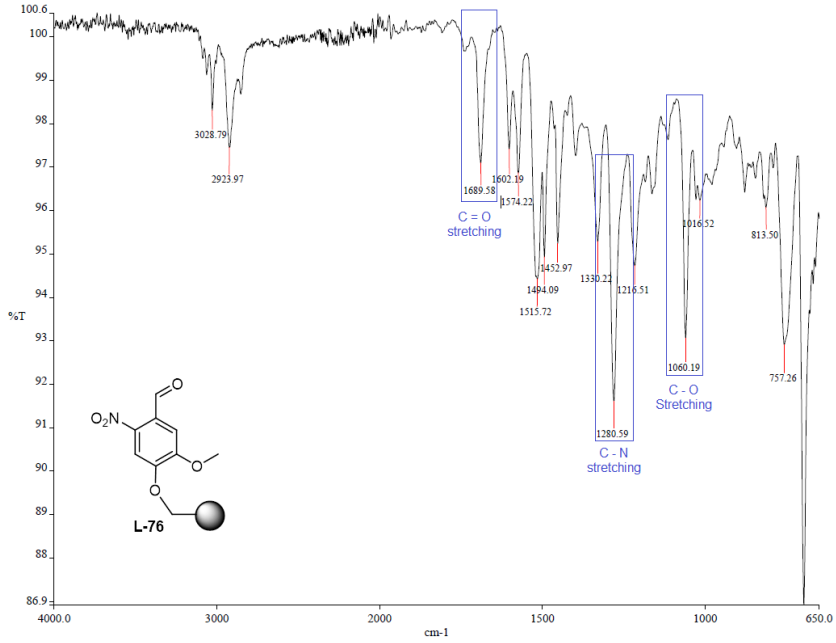


Figure 30 - IR spectra of functionalized Merrifield resin with photolabile linker

Screening of the coupling to the solid support (photolabile traceless diazo linker)

To determine the applicability of the synthesized photolabile diazo functionalized solid support for amino acid coupling using an ester bond formation, a solution of Fmoc-Phe-OH was added to the resin. The reaction of the amino acid and the resin showed the formation of gas and a change in the resin color (dark to light orange). Quantification of resin loading using Fmoc removal showed a value of 0.5 mmol/g. In contrast, the synthesis of a methyl-6-nitroveratryl (MeNV) was efficient from apocynin (acetovanillon), but Fmoc-Phe-OH could not be coupled to this linker, and therefore **L-76** was the resin of choice.

To confirm the compatibility of the linker with solid-phase synthesis and the feasibility of the photocleavage, the Fmoc group of the Phe was removed, and the peptide was elongated to a hexapeptide using standard SPPS. MALDI-TOF MS after small-scale cleavage showed the

expected mass for the peptide (Figure 31). However, the time required to remove the compound had to be increased to 2 h to assure full cleavage of the molecule.

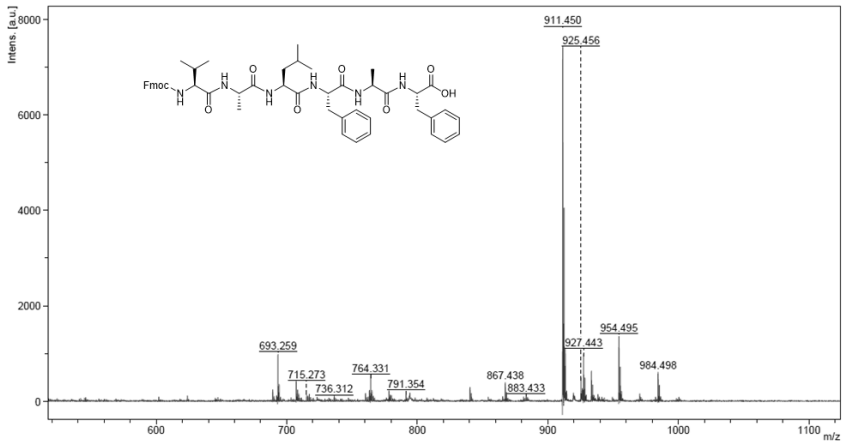
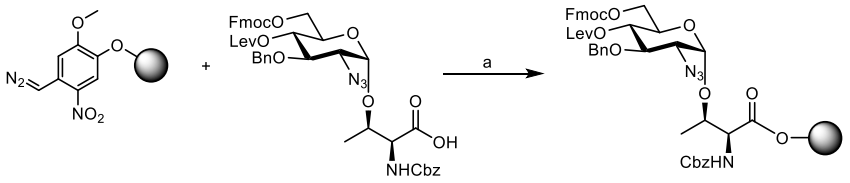


Figure 31 – MS spectra of the synthesized hexapeptide. MALDI-TOF MS of **72**: m/z Calc for $C_{50}H_{60}N_6O_9Na$ $[M+Na]^+$ 911.432, observed 911.450.

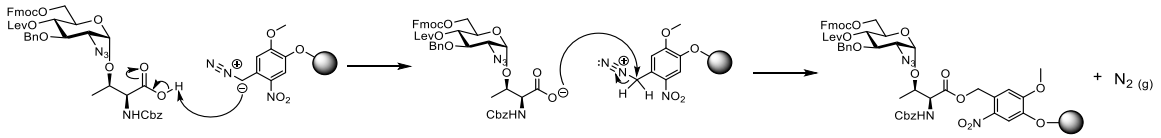
After the successful coupling of Fmoc-Phe-OH and the peptide elongation and photocleavage, the coupling of glycosylated amino acid was envisioned using similar conditions (Scheme 54).



L3 **39** **39-L3**

Scheme 54 - Coupling of the glycosylated amino acid to photolabile diazo solid support L3. Reagents and conditions: CH_2Cl_2, o.n.

The coupling of **39** to the solid support **L3** was tested by adding a solution of **39** (1.3 equiv) in CH_2Cl_2 to the resin. Within seconds, bubbling (evolution of gas) and a change in the resin color were observed. The reaction proceeds by protonation of the diazo α-position by the carboxylate proton and a following nucleophilic attack from the carboxylate to the same position, resulting in the liberation of nitrogen and the formation of the ester bond (Scheme 55). Fmoc removal quantification revealed a successful and efficient glycosylated amino acid coupling to the resin in a single process. The use of a slight excess of building block without activation afforded an excellent conversion of **39-L3** and the observed loading of 0.35 mmol/g was suitable for the assembly of the carbohydrates.



Scheme 55 - Mechanism for coupling of glycosylated amino acid to the photolabile diazo linker

The use of the new linker **L3** showed better performance regarding previous results obtained for **L1**. Table 4 compares the two linkers employed to couple the glycosylated amino acid to the solid support. Coupling of a building block to **L1** (Entry 1) required a significant excess of the activated amino acid (4-5 eq), making the strategy unsuitable for recovering the expensive glycosylated amino acid. On the other hand, the ester bond formation with the newly developed resin bearing the photolabile diazo linker **L3** (Entry 2) was reproducible and performed better than the traceless linker. A significant reduction in the number of equivalents employed and the possibility of recovering the excess glycosylated amino acid after the coupling were the most significant advantages of this methodology.

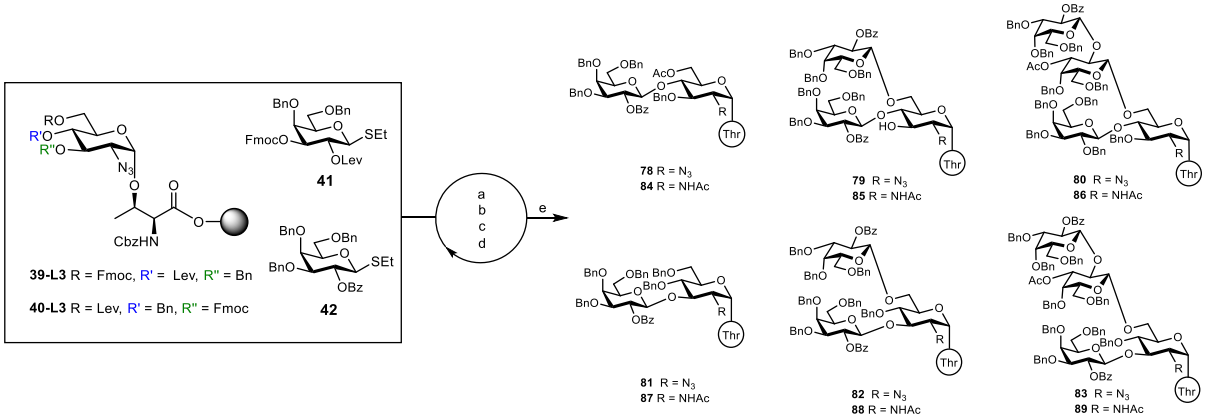
Table 4 - Comparison of conditions for coupling of **39** to functionalized solid supports

Entry	Linker	BB Eq	Reagent	Reagent Eq	Solvent
1	Traceless photolabile (**L1**)	4	DIC, *N*-methyl imidazol	4, 4	CH_2Cl_2:DMF 1:1
2	Photolabile diazo (**L3)**	1.3	-	-	CH_2Cl_2

Assembly of the oligosaccharides and glycosylated amino acids of *T. cruzi* MLM

The results and the evaluated conditions for the AGA of *T. cruzi* glycans employing amino pentanol linker were considered for designing the solid-phase synthesis of glycosylated amino acids using the functionalized solid supports **39-L3** and **40-L3**. First, the initial step of the process was eliminated with glycosyl amino acid coupling. Then, the subsequent elongation of the oligosaccharide was performed using two glycosylation cycles and four molar equivalents of building blocks per cycle.

For the AGA run on the home-built synthesizer, the functionalized solid support **39-L3** or **40-L3** was placed in the reaction vessel, and the resin was swollen for 20 min. Subsequently, selective removal of the temporary protecting group (Fmoc or Lev) was performed to reveal the appropriate hydroxyl for elongation, which was then followed by an acidic wash. Next, the thioglycoside donor (four equivalents based on resin loading) was delivered to the reaction vessel at an initial temperature of -20 °C, followed by the addition of the required activator solution. After 5 min, the temperature of the reaction mixture was increased to 0 °C and maintained for 20 min and subsequent capping of unreacted groups was performed with acetic anhydride and methanesulfonic acid. Iterative cycles of protective group removal, glycosylation, and capping allowed the assembly of the fully protected oligosaccharide. Finally, the resin was transferred to a polypropylene reactor to continue with the post-AGA steps on resin.

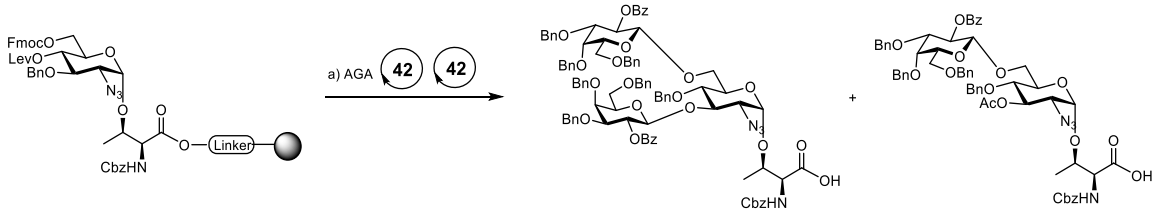


Scheme 56 - Automated assembly of glycosylated amino acids of *T. cruzi* MLMs. Reagents and conditions: a) Protective group removal (Fmoc: piperidine in DMF, Lev: hydrazine acetate in pyridine/AcOH/H_2O; b) TMSOTf in CH_2Cl_2; c) building block (4 eq), NIS/TfOH, CH_2Cl_2, -20 °C (5 min) to 0 °C (20 min); d) Ac_2O, MsOH, CH_2Cl_2; e) *hv* (356 nm).

This AGA development study was designed and completed to determine the best conditions for glycan synthesis. The protocols were completed on a small scale and served mainly for analytical purposes. MALDI-TOF MS Analysis after small-scale cleavage delivered information about products and intermediates formed in each reaction step. In addition, one of the products after the global deprotection was analyzed by LC-MS.

Automated glycan assembly employing the optimized conditions on functionalized resin **39-L3** and **40-L3** and building blocks **41** and **42** allowed the assembly of a library of fully protected oligosaccharides from disaccharides to tetrasaccharides **78**-**83** (Scheme 56). A double glycosylation cycle with four equivalents of building blocks per cycle was enough to deliver the structures; however, some deletion sequences were observed in the trisaccharide and tetrasaccharide synthesis. Since all the structures are branched and contain the same building block **42** in both chains, it was difficult to determine by MALDI-TOF MS which glycosylation was incomplete.

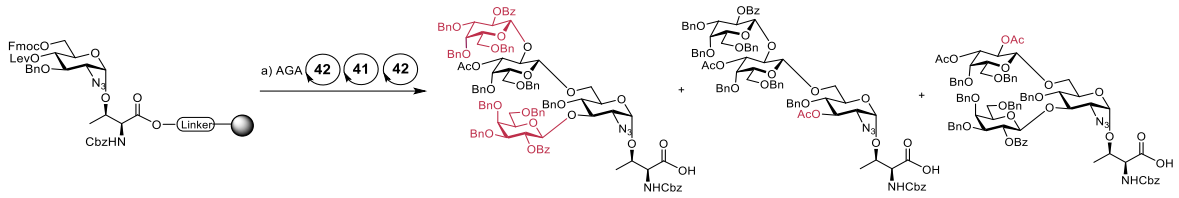
In the synthesis of trisaccharide **79** and **82**, it can be hypothesized that due to the higher reactivity and accessibility of the primary C6-OH this reaction was complete and therefore the deletion sequence corresponds to incomplete glycosylation in position C4-OH for **79** and C3-OH for **82** (Scheme 57).



Scheme 57 – Possible products in the synthesis of trisaccharide 82

The assembly of tetrasaccharide **83** occurred with a deletion sequence resulting in the mixture of the product and a trisaccharide. Nonetheless, the assignment of the deletion sequence becomes difficult because the building block **42** is attached to two different carbohydrates (GlcNAc and Gal)

(Scheme 58). Therefore, specific analysis such as stepwise monitoring by MALDI-TOF MS or NMR is required to certainly assigned which glycosylation was not complete.

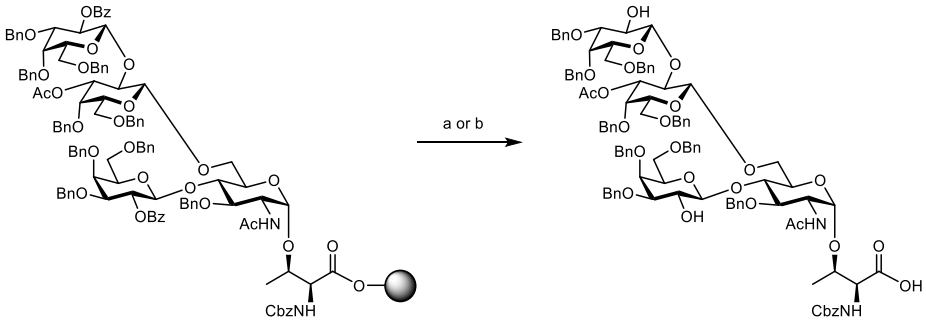


Scheme 58 – Possible products in the synthesis of tetrasaccharide 83

The established methodology for reducing the azide and acetylation of amines employing thioacetic acid was used with the library of glycosylated amino acids. One-pot reaction on resin allowed complete conversion of the azides (**78**-**83**) into acetamides (**84**-**89**), proving that the length of the structures and branching do not affect the reaction outcome.

Screening to standardize deprotection conditions

The use of benzoyl esters as participating groups to assure β-linkage was necessary during Automated Glycan Assembly. However, ester hydrolysis under basic conditions is challenging and can lead to the β-elimination of the glycan from the amino acid. Therefore, the hydrolysis of benzoates and acetates on the resin was tested with lithium hydroxide and hydrazine monohydrate. It is essential to highlight that these treatments hydrolyzed the ester linkage and released the glycosylated amino acid from the resin, reducing the process by one step.



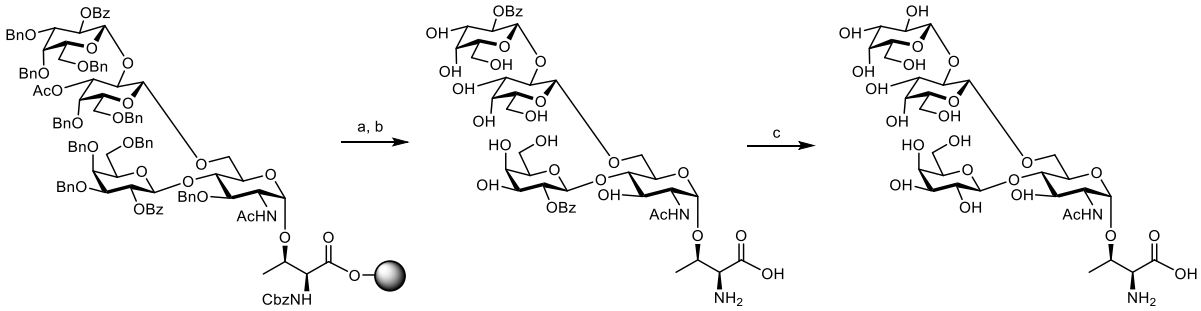
Scheme 59 – On-resin ester hydrolysis of 86 for removal of benzoates and cleavage from solid support. Reagents and conditions a) LiOH in THF/MeOH (4:1), 2 h, rt; b) 10% $N_2H_4 \cdot H_2O$ in THF:MeOH (1:1), 2 days, rt.

On-resin ester hydrolysis of tetrasaccharide **86** (Scheme 59) using mild conditions (LiOH, 0.08 M) was performed in an optimal mixture of THF/MeOH (4:1). THF was necessary for resin swelling, and methanol improves lithium hydroxide solubility[161] and reduces the nucleophilicity of the base and the acidity of the C-α-proton by hydrogen bonding[255]. After two hours, the glycosylated amino acid was cleaved from the resin, and the acetates were removed. Unfortunately, the β-elimination product and incomplete removal of benzoyls were observed. This trial indicated that the methodology was not suitable to perform the intended two-step reaction.

A second methodology to remove the benzoyl group involved the treatment of resin-bound trisaccharide **88** with 10% hydrazine monohydrate in THF/MeOH (1:1). Reaction for five hours resulted in the removal of products from the resin under the formation of a hydrazide. This product was attributed to the nucleophilicity of the hydrazine. Despite this result, the successful hydrolysis of benzoyl without the formation of other side reactions suggested this methodology as an alternative to complete the process in solution. The hydrazide product was dissolved in methanol (18 mg/mL), and hydrazine monohydrate (20%) was added. The reaction progress was monitored for five days until MALDI-TOF MS showed the complete hydrolysis of benzoyl esters. Despite this good outcome, β-elimination of the glycan was also detected. However, this strategy appeared appropriate for the ester hydrolysis and suggested two possible experiments.

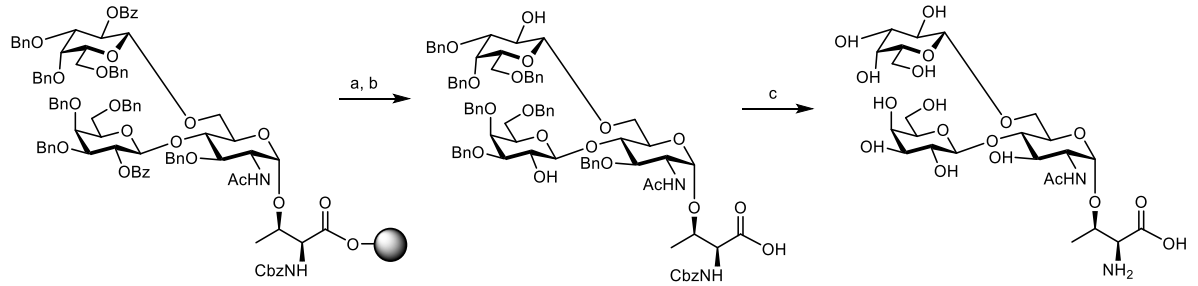
Screening to standardize global deprotection of the molecule

A new approach was envisioned to find an efficient protocol for the global deprotection of the molecules. The strategy involved the photocleavage of the moieties **85** and **89** from the resin in delivering the glycosylated amino acid as carboxylate. Then, further treatment of the cleaved material in solution follows two different routes: 1) hydrogenolysis, ester hydrolysis (Scheme 60) and 2) ester hydrolysis, hydrogenolysis (Scheme 61). These strategies were evaluated to determine if a change in the order of the steps could affect the outcome of the reaction.



Scheme 60 – Global deprotection strategy I. Reagents and conditions: a) *hv* (356 nm); b) Pd/C, *t*-BuOH:EtOAc:THF:H_2O (1:1:1:0.25), rt, 2 days; c) $N_2H_4 \cdot H_2O$, MeOH, rt, 5 days.

Strategy I started with the *N*-benzyloxycarbonyl (Cbz) and benzyl (Bn) removal from **89** by heterogeneous catalytic hydrogenolysis (Scheme 60). The initial reaction resulted in no conversion or reaction progress, even with the addition of a significant excess of Pd/C. Presumably, some thioacetic acid used to convert azide into acetamide remained after the reaction workup, and this reagent poisoned the palladium catalyst used in the experiment. Purification of protected **89** by gel filtration chromatography employing Sephadex LH 20 and CH_2Cl_2/MeOH (2:1) and the following hydrogenolysis of the isolated material showed efficient removal of the Cbz and Bn groups. Finally, treatment of the debenzylated material with hydrazine hydrate in methanol removed the acetate and benzoates, giving the desired glycosyl amino acid.



Scheme 61 - Global deprotection strategy II. Reagents and conditions: a) *hv* (356 nm); b) $N_2H_4 \cdot H_2O$, MeOH, rt, 5 days; c) Pd/C, *t*-BuOH:EtOAc:THF:H_2O (1:1:1:0.25), rt, 2 days.

An analogous protocol was evaluated by treating trisaccharide **85** with 20% hydrazine monohydrate in methanol at 1 mg/mL glycan concentration (Scheme 61). These changes were considered to assure slow conversion and reduce the possibility of getting the β-elimination product. Monitoring the reaction progression by MALDI-TOF MS showed that benzoyl groups were slowly removed, and no side product was observed (Figure 32). After five days, the partially protected glycosylated amino acid was delivered. Final removal of the *N*-benzyloxycarbonyl (Cbz) and benzyl (Bn) groups by heterogeneous catalytic hydrogenolysis, removal of the catalyst by filtration, and RP-HPLC purification delivered the desired fully deprotected product **32** (Figure 33).

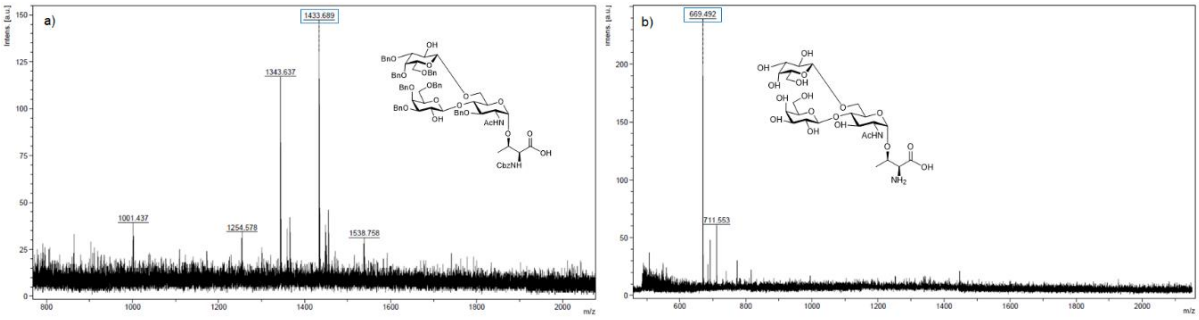


Figure 32 - MS spectra of reaction progression strategy 2. a) MALDI-TOF MS of **85c**: m/z Calc for $C_{81}H_{90}N_{20}O_{20}Na$ $[M+Na]^+$ 1433.598, observed 1433.689; b) MALDI-TOF MS of **32**: m/z Calc for $C_{24}H_{42}N_2O_{18}Na$ $[M+Na]^+$ 669.233, observed 669.492.

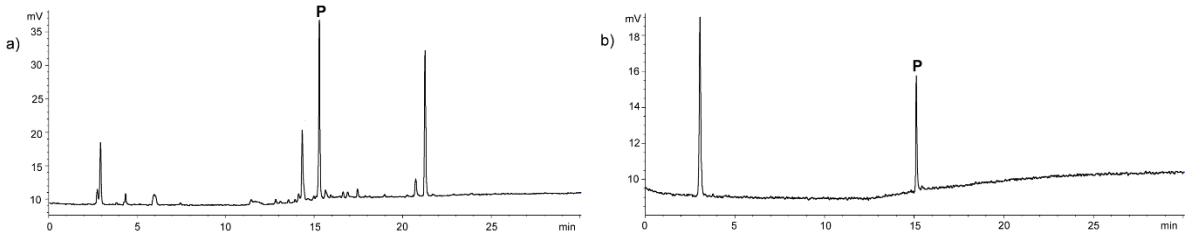


Figure 33 - Analytical RP-HPLC chromatograms of 32 a) crude reaction mixture P = $[M+H]^+$ = 647 expected product; b) pure product after RP-HPLC purification.

Although the two strategies were efficient for the global deprotection, strategy I required the gel filtration of the compound to remove the excess of thioacetic acid, thus strategy II was chosen as the most suitable methodology for the global deprotection. Under optimized conditions, the post-AGA process consisted of four steps: 1) on-resin azide to acetamide conversion using thioacetic acid, 2) removal of the compounds from the resin by photocleavage, 3) ester hydrolysis with 20% hydrazine monohydrate to delivered the partially protected molecule without β-elimination, and 4) catalytic hydrogenolysis for Cbz and Bn removal.

3.3. Conclusions and Outlook

A general methodology for synthesizing glycosylated amino acids derived from *T. cruzi* was developed, representing the first report on the solid-phase synthesis of these molecules. A collection of six structures (**78**-**83**) was synthesized from two monosaccharide building blocks and an α-glycosylated amino acid bound to the resin. Some key modifications were essential for the optimization of the process: replacing thioglycosides **44** and **47** for more reactive imidates **45** and **48** and replacing the traceless photolabile linker bound to Merrifield resin **L1** to the photolabile diazo linker **L3**. This change was necessary to assure a good resin loading for glycan assembly, which was not afforded during conjugation of **39** to **L1**, presumably due to the low nucleophilicity of the secondary alcohol of the linker.

Employing Automated Glycan Assembly, two linear and four branched structures were assembled, proving the generality of the protocol and the feasibility for the assembly of complex and straightforward structures under these conditions. However, the reproducibility of the glycosylations under the used conditions still require optimization to avoid deletion sequences and variations within different synthesis. The use of a more reactive donor (phosphate) for **42** should be tested to drive the reactions to complete conversion. A global deprotection strategy of ester hydrolysis and hydrogenolysis delivered the fully deprotected glycosylated amino acid **32**. Additional work must be performed to complete the global deprotection of the remaining molecules.

A new photolinker was designed, synthesized and evaluated to reduce the number of equivalents employed for the loading of amino acids. This linker proven to be efficient for automated glycan assembly and releasing carboxylate by photocleavage. Further developments include optimization of the acetamide synthesis from azide, application of gel filtration to avoid catalyst poisoning, and optimization of ester hydrolysis to avoid the β-elimination using a slow reaction of the diluted glycosylated amino acid with hydrazine monohydrate in methanol.

The method developed in this work can be used to obtain other targets, such as mucin-like molecules of the *T. cruzi* Colombiana strain or O-glycosyl amino acids from human mucins. In future work, the synthetic glycosylated amino acids from *T. cruzi* MLMs can be used to evaluate the diagnostic potential of these molecules to detect Chagas infections in sera samples. This analysis will require printing the molecules onto activated glass slides and incubating them with sera from Chagas' patients. The presence of anti-glycan antibodies will be detected using fluorescence tagged secondary antibodies. The fluorescence intensities of the different spots will be quantified to determine the best glycan binder.

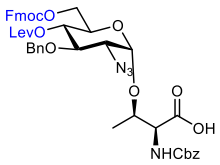
3.6. Experimental Section

3.6.1. Building blocks for AGA

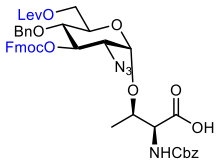
General Methods

Solvents and reagents were of commercial grade and were used without further purification unless otherwise noted. Anhydrous solvents were obtained from a Solvent Drying System (J.C. Meyer) and further dried with molecular sieves (4 Å). Analytical thin layer chromatography was performed on 0.25 mm silica gel 60 F254 glass-supported plates (Macherey-Nagel). Compounds were visualized with UV light (254 nm), 3-methoxyphenol (0.2 mL in 200 mL of EtOH and 6 mL H_2SO_4, sugar stain), ceric ammonium molybdate stain (2 g $Ce(SO4)_2$ g, 10 g$(NH_4)_6Mo_7O_{24}$·4 H_2O and 20 mL H_2SO_4 in 180 mL H_2O) and vanillin stain (12 g vanilline, 2 mL H_2SO_4 in 200 mL of EtOH). Flash chromatography was performed using Silica gel 60 230-400 mesh (Sigma- Aldrich). NMR spectra were obtained using a Bruker Ascend 400 spectrometer at 400 MHz (^{1}H) and 100 MHz (^{13}C). $CDCl_3$ and DMSO-D_6 were used as solvents and chemical shifts (δ) are reported in parts per million (ppm) relative to the respective residual solvent peaks ($CDCl_3$: 7.26 ppm ^{1}H, 77.16 ppm ^{13}C, DMSO-d_6 2.50 ppm ^{1}H, 39.52 ppm ^{13}C) unless stated otherwise. Bidimensional and non-decoupled experiments were performed to assign identities of peaks showing relevant structural features. NMR spectra were processed using MestreNova 14.2.3 (MestreLab Research). Specific rotations were measured with an UniPol L1000 polarimeter (Schmidt & Haensch) at 25 °C and λ = 589 nm. Concentration (*c*) is expressed in g/100 mL in the solvent noted in parentheses. IR spectra were measured with a Spectrum 100 ATR-FTIR spectrometer (Perkin Elmer) and are reported in terms of frequency of absorption (v, cm^{-1}). MALDI-TOF MS spectra were obtained on a Bruker Daltonics Autoflex Speed spectrometer. High-resolution mass spectra (ESI-HRMS) were recorded with a Xevo G2-XS Q-Tof mass spectrometer (Waters) coupled to an Acquity H-class UPLC.

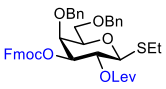
Building Blocks and resins for AGA



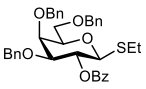
39



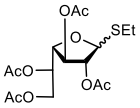
40



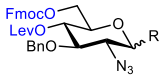
41



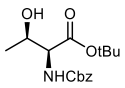
42



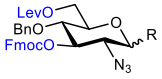
43



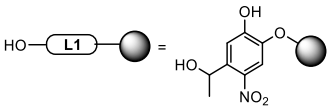
44 R = SPh
45 R = $ONPhCF_3$



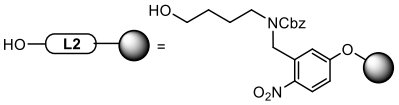
46



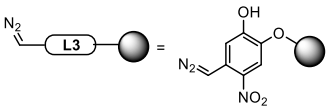
47 R = SPh
48 R = $ONPhCF_3$



L1



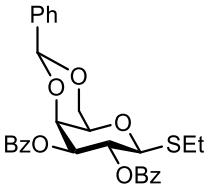
L2



L3

Galactose Building Blocks

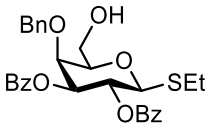
*Ethyl 2,3-di-O-benzoyl-4,6-O-benzylidene-1-thio-β-D-galactopyranose (***53***)*



To a solution of ethyl 4,6-O-benzylidene1-thio-β-D-galactopyranoside (5 g, 16.0 mmol) in 150 mL anhydrous CH_2Cl_2, benzoic anhydride (14.5 g, 64.1 mmol) and DMAP (0.97 g, 8 mmol) were added under an Argon atmosphere. The solution was cooled to 0 °C and triethylamine (17.85 mL, 0.128 mmol) was slowly added. The reaction mixture was stirred overnight at rt, diluted with CH_2Cl_2 and quenched by addition of MeOH at 0 °C. The solvent mixture was evaporated under reduced pressure. The residue was then dissolved in ethyl acetate and subsequently extracted with 1 M HCl, saturated $NaHCO_3$, and brine. The organic phase was dried over Na_2SO_4, filtered and concentrated under reduced pressure. Purification of the product by flash chromatography hexanes/EtOAc (5:1) as eluent, afforded **53** as white solid (8.2 g, 15.7 mmol, 98%). **R_f =** 0.4 (hexanes/EtOAc 5:1), **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.98 (m, *J* = 8.5, 4.1, 1.4 Hz, 4H, H-Ar benzoyl), 7.57 – 7.45 (m, 4H, H-Ar benzoyl), 7.43 – 7.31 (m, 7H, H-Ar benzoyl and benzylidene), 5.96 (t, *J* = 9.9 Hz, 1H, H-2), 5.54 (s, 1H, CH benzylidene), 5.40 (dd, *J* = 10.0, 3.5 Hz, 1H, H-3), 4.74 (d, *J* = 9.9 Hz, 1H, H-1), 4.63 (dd, *J* = 3.5, 1.0 Hz, 1H, H-4), 4.42 (dd, *J* = 12.5, 1.6 Hz, 1H, H-6a), 4.10 (dd, *J* = 12.5, 1.8 Hz, 1H, H-6b), 3.73 (q, *J* = 1.5 Hz, 1H, H-5), 2.96 (m, *J* = 12.3, 7.5 Hz, 1H, SCH_2CH_3), 2.81 (m, *J* = 12.2, 7.4 Hz, 1H, SCH_2CH_3),), 1.31 (t, *J* = 7.5 Hz, 3H, SCH_2CH_3).

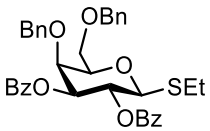
^{13}C NMR (101 MHz, $CDCl_3$) δ (ppm) 166.28 (C=O Bz), 165.47 (C=O Bz), 137.7, 133.5, 133.3, 130.0, 129.9, 129.6, 129.2, 129.1, 128.5, 128.5, 128.3, 126.4, 101.1 (CH benzylidene), 83.0 (C-1), 73.9 (C-3), 73.9 (C-4), 70.0 (C-5), 69.3 (C-6), 67.2 (C-2), 23.0 ($S\underline{C}H_2CH_3$), 14.9 ($SCH_2\underline{C}H_3$). **ESI-HRMS** (m/z): Calcd for $C_{29}H_{28}O_7SNa$ $[M+Na]^+$ 543.1453, observed 543.1455. NMR data agreed with previously reported data[289].

*Ethyl 2,3-di-O-benzoyl-4-O-benzyl-1-thio-β-D-galactopyranose (***54***)*



To a solution of **53** (8.1 g, 15.6 mmol) in anhydrous dichloromethane (60 mL) at 0 °C, borane in THF (1 M, 46 mL, 46.2 mmol) and TMSOTf (560 µL, 3.1 mmol) were sequentially added under an Ar atmosphere. The mixture was stirred for 1.5 h and the reaction was quenched by portion-wise addition of a Et_3N/MeOH (1:10) solution until no further gas evolved. The resulting mixture was coevaporated with MeOH to dryness. The product was purified by flash chromatography using 2:1 hexanes/EtOAc to yield **54** as a white solid (6.02 g, 11.5 mmol, 75%) **R_f =** 0.39 (hexanes/EtOAc 1:1), $[\alpha]_D^{25}$ = +87.09 (c = 1, $CHCl_3$). **IR (ATR)** (neat) ν_{max} cm^{-1} 3067 (C-H Ar), 3035, 2930 (C-H Ar), 2872 (C-H Ar), 1720 (C=O), 1602, 1585, 1495, 1452, 1354, 1315, 1273 (C-O ester), 1178, 1132, 1088 (C-O ether), 1069, 1027, 1001, 987, 937, 870, 802, 756, 735, 707, 687, 675. **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 8.00 – 7.93 (m, 4H, H-Ar benzoyl), 7.51 (q, *J* = 7.6 Hz, 2H, H-Ar benzoyl), 7.37 (td, *J* = 7.7, 3.1 Hz, 4H, H-Ar benzoyl and benzylidene), 7.26 (d, *J* = 2.7 Hz, 5H, H-Ar benzylidene), 5.90 (t, *J* = 10.0 Hz, 1H, H-2), 5.38 (dd, *J* = 10.0, 3.0 Hz, 1H, H-3), 4.78 (d, *J* = 11.7 Hz, 1H, C$\underline{H}$HPh), 4.69 (d, *J* = 9.9 Hz, 1H, H-1), 4.50 (d, *J* = 11.7 Hz, 1H, CH$\underline{H}$Ph), 4.17 (d, *J* = 3.0 Hz, 1H, H-4), 3.88 (dd, *J* = 11.2, 7.0 Hz, 1H, H-6a), 3.77 (t, *J* = 6.1 Hz, 1H, H-5), 3.57 (dd, *J* = 11.2, 5.1 Hz, 1H, H-6b), 2.77 (qdd, *J* = 14.9, 9.9, 6.3 Hz, 2H, $SC\underline{H_2}CH_3$), 1.25 (t, *J* = 7.4 Hz, 5H, $SCH_2C\underline{H_3}$). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 166.1 (C=O), 165.6 (C=O), 137.4, 133.6, 133.3, 130.0, 129.9, 129.6, 129.0, 128.7, 128.6, 128.5, 128.5, 128.2 (15 C-Ar), 84.0 (C-1), 79.1 (C-5), 76.0 (C-3), 74.8 (CH_2-Ph), 73.6 (C-4), 68.5 (C-2), 62.0 (C-6), 24.1($S\underline{C}H_2CH_3$), 15.0 ($SCH_2\underline{C}H_3$). **ESI-HRMS** (m/z): Calcd for $C_{29}H_{30}O_7SNa$ $[M+Na]^+$ 545.1610, observed 545.1606.

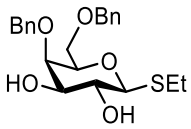
*Ethyl 2,3-di-O-benzoyl-4,6-di-O-benzyl-1-thio-β-D-galactopyranose (***55***)*



To a solution of **54** (6.0 g, 11.48 mmol) in CH_2Cl_2 (100 mL), TMSOTf (95 µL, 0.526 mmol) and HMDS (2.87 mL, 13.78 mmol) were added. The reaction mixture was stirred for 45 min at rt. The reaction was then flushed with nitrogen to remove the ammonia byproduct and the solvent was

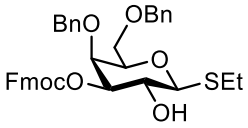
then removed under reduced pressure. The residue was dissolved in anhydrous CH_2Cl_2 (100 mL), followed by addition of freshly activated 4Å molecular sieves (2 g) and benzaldehyde (2.40 mL, 23.54 mmol). The mixture was stirred for 1 h and then cooled to -15 °C, at this temperature triethylsilane (9.2 mL, 57.75 mmol) and TMSOTf (477 µL, 2.64 mmol) were added, and stirred for 1 h. The crude mixture was neutralized with Et_3N and concentrated. Purification by flash column chromatography (hexanes/ethyl acetate 5:1) gave **55** as a white solid (6.48 g, 10.6 mmol, 92%). **R_f =** 0.65 (hexanes/EtOAc 5:1), $[\alpha]_D^{25}$ = +43.21 (c = 1, $CHCl_3$). **IR (ATR)** (neat) ν_{max} cm^{-1} 3065 (C-H Ar), 3033 (C-H Ar), 2930, 2870, 1722 (C=O), 1602, 1585, 1496, 1452, 1353, 1315, 1273 (C-O ester), 1212, 1178, 1152, 1094 (C-O ether), 1070, 1027, 1001, 940, 889, 857, 802, 752, 734, 707, 675.**^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 8.04 – 7.93 (m, 4H, H-Ar benzoyl), 7.52 (td, *J* = 7.5, 3.5 Hz, 2H, H-Ar benzoyl), 7.43 – 7.16 (m, 14H. H-Ar benzyl), 5.91 (t, *J* = 9.9 Hz, 1H, H-2), 5.42 (dd, *J* = 10.0, 3.0 Hz, 1H, H-3), 4.73 (dd, *J* = 14.3, 10.7 Hz, 2H, 1-H, C<u>H</u>HPh), 4.57 – 4.45 (m, 3H, H-1, CH<u>H</u>Ph, C<u>H</u><u>H</u>Ph), 4.28 (d, *J* = 3.0 Hz, 1H, H-4), 3.95 (t, *J* = 6.6 Hz, 1H, H-5), 3.76 – 3.64 (m, 2H, H-6a/b), 2.89 – 2.69 (m, 2H, SC<u>H</u>$_2$CH$_3$), 1.28 (t, *J* = 7.4 Hz, 4H, SCH$_2$C<u>H</u>$_3$). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 166.0 (C=O), 165.6 (C=O), 137.9, 137.8, 133.5, 133.2, 130.0, 129.9, 129.6, 129.2, 128.6, 128.4, 128.4, 128.1, 128.0, 128.0, 127.8, 83.8 (C-1), 77.5 (C-5), 75.8 (C-3), 75.1 (CH_2-Bn), 74.3 (C-4), 73.6 (CH_2-Ph), 68.6 (C-2), 68.2 (C-6), 24.1 (S<u>C</u>H$_2$CH$_3$), 15.0 (SCH$_2$<u>C</u>H$_3$). **ESI-HRMS** (m/z): Calcd for $C_{36}H_{36}O_7SNa$ $[M+Na]^+$ 635.2079, observed 635.2068.

*Ethyl 4,6-di-O-benzyl-1-thio-β-D-galactopyranose (***56***)*



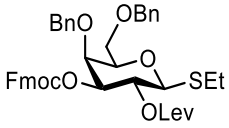
Compound **55** (6.40 g, 5.83 mmol) was dissolved in methanol (75 mL) at rt and 10 mL of NaOMe solution (1 M in methanol, 10.04 mmol) was added dropwise. The mixture was stirred for 2 h, then neutralized by addition of Amberlite IR 120 H^+ form, filtered and concentrated under reduced pressure. The product was purified by flash column chromatography (hexanes/EtOAc 1:2) to afford **56** (4.06 g, 10.04 mmol, 96%) as a white solid. **R_f =** 0.25 (hexanes/EtOAc 2:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.39 – 7.25 (m, 10H, H-Ar benzyl), 4.70 (d, *J* = 1.9 Hz, 2H, CH_2-Bn), 4.50 (q, *J* = 11.7 Hz, 2H, CH_2-Bn), 4.29 (d, *J* = 9.4 Hz, 1H, H-1), 3.92 (dd, *J* = 3.3, 1.0 Hz, 1H, H-4), 3.73 – 3.63 (m, 4H, H-2, H-5, H-6a/b), 3.60 (dd, *J* = 9.2, 3.3 Hz, 1H, H-3), 2.73 (qd, *J* = 7.5, 4.2 Hz, 2H, SC<u>H</u>$_2$CH$_3$), 1.30 (t, *J* = 7.4 Hz, 3H, SCH$_2$C<u>H</u>$_3$). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 138.4, 137.8, 128.6, 128.6, 128.0, 128.0, 128.0, 128.0 (C-Ar), 86.3 (C-1), 77.6 (C-5), 76.2 (C-4), 75.4 (C-3), 75.3 (CH_2-Bn), 73.7 (CH_2-Ph), 71.0 (C-2), 68.5 (C-6), 24.6 (S<u>C</u>H$_2$CH$_3$), 15.4 (SCH$_2$<u>C</u>H$_3$). **ESI-HRMS** (m/z): Calcd for $C_{22}H_{28}O_5SNa$ $[M+Na]^+$ 427.1555, observed 427.1576. NMR data agreed with previously reported data[290].

*Ethyl 3-O-(9-fluorenylmethoxycarbonyl)-4,6-di-O-benzyl-1-thio-β-D-galactopyranose (***57***)*



Compound **56** (4.0 g, 9.89 mmol) and dibutyltin oxide (4.92g, 19.78 mmol) were dissolved in 80 mL of anhydrous toluene and refluxed using a Dean-Stark trap for 6 h. The mixture was cooled down to rt, followed by addition of Fmoc-Cl (3.07 g, 11.87 mmol). The reaction was stirred overnight and the solvent was removed under reduce pressure. The product was purified by flash column chromatography (hexanes/EtOAc 3:1) to yield **57** (5.40 g, 8.62 mmol, 87%) as a white solid. **R_f =** 0.51 (hexanes/EtOAc 2:1), $[\alpha]_D^{25}$ = + 35.51; (c = 1, $CHCl_3$). **IR (ATR)** (neat) v_{max} cm^{-1} = 3069, 3033, 2930, 2871, 1720, 1602, 1585, 1495, 1452, 1354, 1315, 1274, 1178, 1132, 1090, 1070, 1027, 1001, 960, 889, 870, 802, 754, 736, 707, 675. **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.73 (tt, *J* = 7.5, 0.9 Hz, 2H, H-Ar, Fmoc), 7.62 (ddd, *J* = 14.1, 7.5, 1.0 Hz, 2H, H-Ar Fmoc), 7.47 – 7.16 (m, 14H, H-Ar Fmoc and benzyl), 4.77 (dd, *J* = 9.7, 3.1 Hz, 1H, H-3), 4.65 (d, *J* = 11.4 Hz, 1H, C<u>H</u>HPh), 4.54 – 4.47 (m, 2H, C<u>H</u>H-Fmoc, CH_2-Ph), 4.48 – 4.40 (m, 3H, CH<u>H</u>Ph, CH<u>H</u>-Fmoc, CH_2-Ph), 4.38 (d, *J* = 9.7 Hz, 1H, 1-H), 4.27 (t, *J* = 7.1 Hz, 1H, CH-Fmoc), 4.10 – 3.98 (m, 2H, H-2, H-4), 3.75 (ddd, *J* = 7.1, 5.8, 1.1 Hz, 1H, H-5), 3.66 – 3.57 (m, 2H, H6a/b), 2.83 – 2.65 (m, 2H, SC<u>H$_2$</u>CH_3), 1.30 (t, *J* = 7.5 Hz, 3H, SCH_2C<u>H$_3$</u>) **^{13}C NMR** (101 MHz, $CDCl_3$) δ 154.8 (C=O), 143.6, 143.0, 141.5, 141.4, 138.1, 137.8, 128.6, 128.4, 128.2, 128.0, 128.0, 128.0, 127.8, 127.3, 127.3, 125.3, 125.2, 120.2, 86.7 (C-1), 80.8 (C-3), 77.2 (C-5), 75.2 (CH_2-Bn), 74.2 (C-4), 73.6 (CH_2-Bn), 70.1 (CH_2-Fmoc), 68.3 (C-6), 68.1 (C-2), 46.8 (CH-Fmoc), 24.6 (S<u>C</u>H_2CH_3), 15.4 (SCH_2<u>C</u>H_3). **ESI-HRMS** (m/z): Calcd for $C_{37}H_{38}O_7SNa$ $[M+Na]^+$ 649.2236, observed 649.2221. NMR data agreed with previously reported data[165].

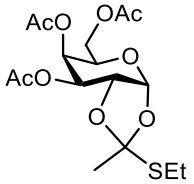
*Ethyl 2-O-levulinoyl-3-O-(9-fluorenylmethoxycarbonyl)-4,6-di-O-benzyl-1-thio-β-D-galactopyranose (***41***)*



Compound **57** (5.40 g, 8.62 mmol) and levulinic acid (1.60 g, 13.79 mmol) were dissolved in CH_2Cl_2 (150 mL) and a solution of DIC (2.14 mL, 13.80 mmol) and DMAP (316 mg, 2.58 mmol) in CH_2Cl_2 (10 mL) was added at 0 °C. The mixture was warmed to rt and stirred for 2 h, then diluted with CH_2Cl_2 (100 mL) and filtered through a celite pad. The filtrate was extracted with 1 M HCl, saturated $NaHCO_3$, and brine. The organic phase was dried over Na_2SO_4, filtered and concentrated under reduced pressure. Purification by flash column chromatography using hexanes/EtOAc (3:1) gave the final building block **41** (5.16 g, 7.12 mmol, 83%) as a colorless syrup. **R_f =** 0.47 (hexanes/EtOAc 2:1), $[\alpha]_D^{25}$ = +29.48 (c = 1, $CHCl_3$**). IR (ATR)** (neat) v_{max} cm^{-1} =3031 (C-H Ar), 2927 (C-H Ar), 2871, 1748 (C=O), 1608, 1497, 1478, 1452, 1362, 1259 (C-O

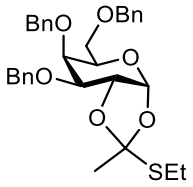
ester), 1152, 1074 (C-O ether), 1028, 984, 909, 877, 784, 738, 697, 666.**^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.74 (ddt, *J* = 7.6, 4.2, 1.0 Hz, 2H, H-Ar, Fmoc), 7.66 – 7.58 (m, 2H, H-Ar, Fmoc), 7.44 – 7.19 (m, 14H, H-Ar Fmoc and Benzyl), 5.47 (t, *J* = 9.9 Hz, 1H, H-2), 4.91 (dd, *J* = 10.0, 3.0 Hz, 1H, H-3), 4.76 (d, *J* = 11.4 Hz, 1H, C<u>H</u>H-Ph), 4.55 – 4.34 (m, 6H, C<u>H</u>H-Ph, CH_2-Bn, H-1, CH_2-Fmoc), 4.29 (t, *J* = 7.3 Hz, 1H, CH-Fmoc), 4.06 (dd, *J* = 3.1, 1.1 Hz, 1H, H-4), 3.74 (ddd, *J* = 7.6, 5.5, 1.1 Hz, 1H, H-5), 3.67 – 3.56 (m, 2H, H6a/b), 2.84 – 2.50 (m, 6H, CH_2-Lev, S<u>C</u>$H_2$$CH_3$), 2.11 (s, 3H, CH_3-Lev), 1.25 (t, *J* = 7.4 Hz, 3H, SCH_2<u>C</u>H_3). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 206.2 (C=O, ketone Lev), 171.6 (C=O Lev), 154.6 (C=O Fmoc), 143.6, 143.1, 141.4, 141.3, 138.0, 137.8, 128.5, 128.4, 128.2, 128.0, 128.0, 128.0, 128.0, 127.8, 127.3, 125.4, 125.3, 120.2, 83.6 (C-1), 78.8 (C-3), 77.2 (C-5), 75.1 (<u>C</u>H_2-Ph), 74.0 (C-4), 73.6 (<u>C</u>H_2-Ph), 70.3 (<u>C</u>H_2-Fmoc), 68.3 (C-2), 68.1 (C-6), 46.7 (<u>C</u>H-Fmoc), 38.0 (CH_2-Lev), 28.1 (CH_2-Lev), 29.9 (CH_3-Lev) 23.9 (S<u>C</u>$H_2$$CH_3$), 14.9 ($SCH_2$<u>C</u>$H_3$). **ESI-HRMS** (m/z): Calcd for $C_{42}H_{44}O_9SNa$ $[M+Na]^+$ 747.2604, observed 747.2661.

*3,4,6-tri-O-acetyl-α-D-galactopyranose-1,2-O-(1-(ethylthio)-orthoacetate (***58***)*



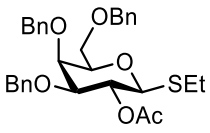
Peracetylated galactopyranose (5.01 g, 12.84 mmol) was dissolved in CH_2Cl_2 (100 mL) and 33% HBr in AcOH (16 mL) was added dropwise at 0 °C. The reaction mixture was stirred at rt overnight, diluted with CH_2Cl_2, and washed with water (3 x 70 mL), saturated $NaHCO_3$ (2 x 70 mL), and brine. The organic phase was dried under Na_2SO_4 and filtered. The solvent was removed under reduced pressure and the residue was taken up in nitromethane (30 mL), followed by addition of 2,6-lutidine (2.24 mL, 19.26 mmol), ethanethiol (3.8 mL, 51.36 mmol) and tetrabutylammonium bromide (412.6 mg, 1.28 mmol). The mixture was stirred at rt for 72 h and the solution was then partitioned between EtOAc (50 mL) and aqueous saturated $NaHCO_3$ (50 mL). The aqueous layer was extracted with EtOAc, dried under Na_2SO_4, filtered and evaporated under reduced pressure. The product was purified by flash column chromatography (hexanes/EtOAc 5:1) to give **58** (3.59 g, 9.15 mmol, 71% over two steps) as a colorless oil. **R_f = 0.74** (hexanes/EtOAc 1:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 5.83 (d, *J* = 5.0 Hz, 1H, H-1), 5.33 (dd, *J* = 3.7, 2.4 Hz, 1H, H-3), 5.00 (dd, *J* = 6.6, 3.7 Hz, 1H, H-4), 4.29 – 4.21 (m, 2H, H-2, H-5), 4.13 – 4.00 (m, 2H, H-6), 2.58 (qd, *J* = 7.5, 2.1 Hz, 2H, S<u>C</u>$H_2$$CH_3$), 2.04 (s, 3H, CH_3-acetyl), 2.00 (d, *J* = 1.2 Hz, 6H, CH_3-acetyl), 1.84 (s, 3H, CH_3 orthoacetate), 1.19 (t, *J* = 7.4 Hz, 3H, SCH_2<u>C</u>H_3). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 170.3, 169.9, 169.6 (3C, C=O acetyl) 115.1 (CH-orthoacetate), 98.1 (C-1), 72.2 (C-2), 70.9 (C-4), 68.9 (C-6), 65.4 (C-3), 61.4 (C-6), 28.7 (CH_3-orthoacetate), 24.6 (S<u>C</u>$H_2$$CH_3$), 20.6 ($CH_3$-acetyl), 20.4 ($CH_3$-acetyl), 14.8 ($SCH_2$<u>C</u>$H_3$). **ESI-HRMS** (m/z): Calcd for $C_{16}H_{24}O_9SNa$ $[M+Na]^+$ 415.1039; observed 415.1050. NMR data agreed with previously reported data[291].

*3,4,6-tri-O-benzyl-α-D-galactopyranose-1,2-O-(1-(ethylthio)-orthoacetate (***59***)*



Acetylated orthoester **58** (3.50 g, 8.92 mmol) was dissolved in MeOH (100 mL) and freshly prepared solution of NaOMe (1 M in methanol) was added dropwise (2.5 mL) and the mixture was stirred at rt for 2 h. The reaction was stopped by addition of Amberlite resin H^+ form until neutral. The resin was removed by filtration through cotton and the solvent was removed under reduced pressure. The residue was dissolved in DMF (70 mL), cooled to 0 °C and sodium hydride (1.29 g, 53.62 mmol) was added to the reaction mixture. The resulting suspension was stirred for 15 min, followed by addition of benzyl bromide (4.78 mL, 40.22 mmol). The reaction mixture was allowed to warm up to rt and stirred overnight. The reaction was quenched by adding methanol, and volatiles were removed under reduced pressure. The residue was dissolved in CH_2Cl_2 and water, extracted three times with CH_2Cl_2 and the combined organic phases were dried over Na_2SO_4, filtered and concentrated under reduced pressure. The product was purified by flash column chromatography using hexanes/EtOAc (5:1) to obtain **59** (4.22 g, 7.86 mmol, 88% over two steps) as a colorless oil. **R_f = 0.78** (hexanes/EtOAc 1:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.44 – 7.24 (m, 15H, H-Ar), 5.82 (d, *J* = 4.7 Hz, 1H, H-1), 4.93 (d, *J* = 11.5 Hz, 1H, C<u>H</u>H-Ph), 4.80 (d, *J* = 12.2 Hz, 1H, CH<u>H</u>-Ph), 4.64 (dd, *J* = 21.8, 11.9 Hz, 2H, C<u>H</u>H-Ph), 4.53 – 4.40 (m, 3H, CH<u>H</u>-Ph, H-2), 4.04 (s, 1H, H-4), 3.99 (t, *J* = 2.4 Hz, 1H, H-5), 3.67 – 3.58 (m, 3H, H-3, H-6), 2.65 (q, *J* = 7.4 Hz, 2H, SC<u>H$_2$</u>CH$_3$), 1.84 (s, 3H, CH_3 orthoacetate), 1.27 (t, *J* = 7.5 Hz, 3H, SCH$_2$C<u>H$_3$</u>).**^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 138.4, 138.1, 137.9, 128.5, 128.5, 128.4, 128.1, 128.0, 127.9, 127.8, 127.8, 127.7 (18C-Ar), 115.6 (C-orthoacetate), 98.6 (C-1), 80.3 (C-3), 78.0 (C-2), 74.6 (CH_2-Ph), 73.6 (CH_2-Ph), 73.4 (C-4), 72.6 (C-5), 71.5 (CH_2-Ph), 68.2 (C-6), 29.4 (S<u>C</u>H$_2$CH$_3$), 24.8 (CH_3 orthoacetate), 15.2 (SCH$_2$<u>C</u>H$_3$). **ESI-HRMS** (m/z): Calcd for $C_{31}H_{36}O_6SNa$ $[M+Na]^+$ 559.2130; observed 559.2142. NMR data agreed with previously reported data[291].

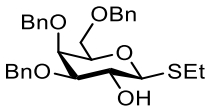
*Ethyl 2-O-acetyl-3,4,6-tri-O-benzyl-1-thio-β-D-galactopyranose (***60***)*



To a solution of orthoester **59** (4.20 g, 7.83 mmol) in CH_2Cl_2 (100 mL), 4Å molecular sieves (1.5 g) and ethanethiol (0.58 mL, 7.83 mmol) were added. The solution was stirred for 1 h, cooled to 0 °C and TMSOTf (89 µL, 0.49 mmol) was added. The reaction mixture was stirred for 4 h and quenched by the addition of Et_3N, filtered through celite and evaporated to dryness under reduced pressure. The product was purified by flash column chromatography using hexanes/EtOAc (3:1)

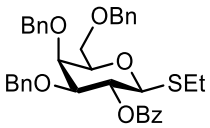
to yield **60** (3.03 g, 5.65 mmol, 73%) as a white solid. **R_f =** 0.68 (hexanes/EtOAc 1:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.23 (ttd, *J* = 11.8, 6.1, 5.5, 2.5 Hz, 16H, H-Ar), 5.34 (t, *J* = 9.8 Hz, 1H, H-2), 4.87 (d, *J* = 11.7 Hz, 1H, CHH-Ph), 4.60 (d, *J* = 12.2 Hz, 1H, CHH-Ph), 4.48 (dd, *J* = 19.9, 12.0 Hz, 2H, CHH-Ph), 4.41 – 4.30 (m, 2H, CHH-Ph), 4.26 (d, *J* = 9.9 Hz, 1H, H-1), 3.92 (d, *J* = 2.7 Hz, 1H, H-4), 3.53 (s, 3H, H-5, H-6), 3.46 (dd, *J* = 9.7, 2.8 Hz, 1H, H-3), 2.72 – 2.52 (m, 2H, SCH_2CH_3), 1.96 (s, 3H, CH_3-acetyl), 1.15 (t, *J* = 7.4 Hz, 3H, SCH_2CH_3).**^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 169.5 (C=O acetyl), 138.4, 137.8, 137.6, 128.2, 128.2, 127.9, 127.8, 127.8, 127.7, 127.5, 127.3, 127.2 (18H, C-Ar), 83.4 (C-1), 81.2 (C-3), 77.2 (C-5), 74.2 (CH_2-Ph), 73.4 (CH_2-Ph), 72.6 (C-4), 71.7 (CH_2-Ph), 69.4 (C-2), 68.3 (C-6), 23.4 (SCH_2CH_3), 20. 9 (CH_3-acetyl), 14.6 (SCH_2CH_3). **ESI-HRMS** (m/z): Calcd for $C_{31}H_{36}O_6SNa$ $[M+Na]^+$ 559.2130, observed 559.2139. NMR data agreed with previously reported data[292].

*Ethyl 3,4,6-tri-O-benzyl-1-thio-β-D-galactopyranose (**60a**)*



Galactose **60** (3.03 g, 5.65 mmol) was dissolved in methanol (50 mL) and a freshly prepared solution of NaOMe (1M in methanol) was added dropwise (2 mL) until pH 10. The mixture was stirred at rt for 2 h. The reaction was stopped by addition of IR120 Amberlite resin (H^+ form) until neutral pH. The resin was removed by filtration through cotton and the solvent was removed under reduced pressure. Purification by flash column chromatography (hexanes/EtOAc 4:1) afforded **60a** (2.79 g, 5.64 mmol, quant.) as a colorless oil. **R_f =** 0.60 (hexanes/EtOAc 1:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.40 – 7.06 (m, 16H, H-Ar), 4.82 (d, *J* = 11.6 Hz, 1H, CHH-Ph), 4.72 – 4.56 (m, 2H, CHH-Ph), 4.53 (d, *J* = 11.6 Hz, 1H, CHH-Ph), 4.45 – 4.30 (m, 2H, CHH-Ph), 4.25 (d, *J* = 9.6 Hz, 1H, 1-H), 3.95 – 3.86 (m, 2H, H-2, H-4), 3.60 – 3.48 (m, 3H, H-5, H-6), 3.38 (dd, *J* = 9.3, 2.8 Hz, 1H, H-3), 2.74 – 2.53 (m, 2H, SCH_2CH_3), 1.22 (t, *J* = 7.4 Hz, 3H, SCH_2CH_3). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 138.4, 137.9, 137.6, 128.3, 128.3, 128.0, 127.8, 127.6, 127.6, 127.5, 127.3 (18H, C-Ar), 86.1 (C-1), 83.0 (C-3), 77.4 (C-5), 74.3 (CH_2-Ph), 73.4 (CH_2-Ph), 72.9 (C-4), 72.1 (CH_2-Ph), 69.5 (C-2), 68.4 (C-6), 24.0 (SCH_2CH_3), 15.1 (SCH_2CH_3). **ESI-HRMS** (m/z): Calcd for $C_{29}H_{34}O_6SNa$ $[M+Na]^+$ 517.2025; observed 517.2042. NMR data agreed with previously reported data[292].

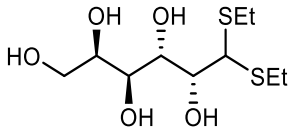
*Ethyl 2-O-benzoyl-3,4,6-tri-O-benzyl-1-thio-β-D-galactopyranose (**42**)*



Deacetylated derivative **60a** (2.79 g, 5.64 mmol) was dissolved in pyridine (35 mL) and benzoyl chloride (0.85 mL, 7.33 mmol) and DMAP (896 mg, 7.33 mmol) were added. The mixture was stirred at rt overnight. The reaction mixture was diluted with CH_2Cl_2 and quenched by addition of

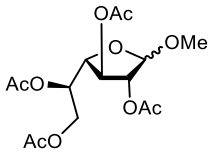
MeOH at 0 °C. The solvent was removed under reduced pressure, the residue was dissolved in ethyl acetate and subsequently extracted with 1 M HCl, saturated $NaHCO_3$, and brine. The product was purified by flash column chromatography using hexanes and ethyl acetate as eluent to yield building block **42** (3.17g, 5.29 mmol, 94%) as a white solid. **R_f =** 0.66 (hexanes/EtOAc 1:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 8.00 – 7.89 (m, 2H, H-Ar Bz), 7.54 – 7.46 (m, 1H, H-Ar Bz), 7.37 (t, *J* = 7.7 Hz, 2H, H-Ar Bz), 7.33 – 7.15 (m, 10H, H-Ar Bn), 7.09 (q, *J* = 7.4, 6.4 Hz, 5H, H-Ar Bn), 5.62 (t, *J* = 9.7 Hz, 1H, H-2), 4.92 (d, *J* = 11.6 Hz, 1H, C<u>H</u>H-Ph), 4.56 (dd, *J* = 12.0, 7.7 Hz, 2H, C<u>HH</u>-Ph), 4.47 – 4.33 (m, 4H, C<u>HH</u>-Ph, H-1), 3.98 (d, *J* = 2.7 Hz, 1H, H-4), 3.65 – 3.50 (m, 4H, H-3, H-5, H-6), 2.72 – 2.53 (m, 2H, SC<u>H$_2$</u>CH$_3$), 1.13 (t, *J* = 7.4 Hz, 3H, SCH$_2$C<u>H$_3$</u>). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 165.2 (C=O, Bz), 138.4, 137.6, 137.4, 132.8, 129.9, 129.7, 128.3, 128.1, 128.0, 127.9, 127.8, 127.7, 127.5, 127.3 (24C, C-Ar), 83.5 (C-1), 80.8 (C-3), 77.3 (C-5), 74.2 (CH_2-Ph), 73.4 (CH_2-Ph), 72.5 (C-4), 71.5 (CH_2-Ph), 69.9 (C-2), 68.3 (C-6), 23.5 (S<u>C</u>H$_2$CH$_3$), 14.6 (SCH$_2$<u>C</u>H$_3$). **ESI-HRMS** (m/z): Calcd for $C_{36}H_{38}O_6SNa$ $[M+Na]^+$ 621.2287; observed 621.2296. NMR data agreed with previously reported data[293].

*D-galactose diethyl dithioacetal (***61***)*



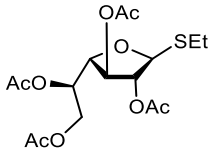
D-galactose (15 g, 0.083 mmol) was dissolved in concentrated HCl (12 mL) at rt and ethanethiol (12.3 mL, 0.166 mmol) was slowly added. The reaction mixture was stirred vigorously, releasing the pressure occasionally. After 5 min when temperature increased, a little amount ice water was added and the mixture stirred for 5 min more. Then, more ice water was added and the solid product was filtered and washed with small amount of cold water. Recrystallization from pure ethanol afforded **61** (13.20 g, 46.09 mmol, 55%) as a white crystalline solid. **R_f =** 0.42 (CH_2Cl_2/MeOH 6:1). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) 4.52 (d, *J* = 7.8 Hz, 1H, OH), 4.45 (s, 1H, OH), 4.28 (d, *J* = 8.0 Hz, 1H, OH), 4.17 (d, *J* = 6.7 Hz, 1H, OH), 4.08 (d, *J* = 7.9 Hz, 1H, OH), 4.02 (d, *J* = 9.2 Hz, 1H, H-1), 3.87 (m, *J* = 9.4, 7.9, 1.5 Hz, 1H, H-4), 3.71 (m, *J* = 9.8, 7.1, 1.4 Hz, 2H, H-2, H-5), 3.48 – 3.34 (m, 3H, H-3, H-6), 2.61 (m, *J* = 7.3, 5.5 Hz, 4H, SC<u>H$_2$</u>CH$_3$), 1.18 (t, *J* = 7.4 Hz, 6H, SCH$_2$C<u>H$_3$</u>). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) 71.6 (C-2), 69.9 (C-5), 69.5 (C-3), 69.3 (C-4), 63.2 (C-6), 54.7 (C-1), 24.2 (S<u>C</u>H$_2$CH$_3$), 23.7 (S<u>C</u>H$_2$CH$_3$), 14.5 (SCH$_2$<u>C</u>H$_3$). **ESI-TOF MS** (m/z): Calcd for $C_{10}H_{22}O_5S_2Na$ $[M+Na]^+$ 309.1; observed 309.2. NMR data agreed with previously reported data[294,277].

*Methyl 2,3,5,6-tetra-O-acetyl-β-D-galactofuranoside(***61***)*



Dithioacetal **61** (3.0 g, 10.45 mmol) was dissolved in a solution of 2% I_2 in MeOH (w/v, 250 mL) and stirred at rt for 7 h. Excess I_2 was quenched by addition of solid sodium thiosulfate ($Na_2S_2O_3$) until the color diminished, then solid sodium bicarbonate ($NaHCO_3$) was added to neutralize the reaction mixture. Solvents were removed under reduced pressure to give a yellow solid, which was taken to the next step without purification. The crude was dissolved in equal volumes of acetic anhydride/pyridine (50 mL), the mixture was cooled to 0 °C and DMAP (128 mg, 1.05 mmol) was added. The reaction mixture was stirred overnight at rt and diluted with CH_2Cl_2 (100 mL), extracted with water, 1 M HCl, saturated $NaHCO_3$, and brine. The organic phase was dried over Na_2SO_4, filtered, and the solvent was removed under reduced pressure. The crude was purified by flash chromatography using hexanes/EtOAc (3:2) to give **62** (β:α = 9:1), β isomer was isolated after column chromatography (3.28 g, 9.05 mmol, 87% over two steps). **R_f =** 0.47 (hexanes/EtOAc 1:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) δ 5.37 (dt, *J* = 7.0, 4.1 Hz, 1H, H-5), 5.07 – 4.95 (m, 2H, H-2, H-3), 4.90 (s, 1H, H-1), 4.32 (ddd, *J* = 11.9, 9.6, 4.4 Hz, 1H, H-6a), 4.27 – 3.98 (m, 2H, H-4, H-6b), 3.36 (d, *J* = 2.7 Hz, 3H, OCH_3), 2.12 (s, 3H, CH_3-acetyl), 2.09 (s, 3H, CH_3-acetyl), 2.07 (s, 3H, CH_3-acetyl), 2.04 (s, 3H, CH_3-acetyl). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 170.6 (C=O, acetyl), 170.1 (C=O, acetyl), 170.1 (C=O, acetyl), 169.7 (C=O, acetyl), 106.6 (C-1), 81.3 (C-2), 79.9 (C-4), 76.5 (C-3), 69.3 (C-5), 62.6 (C-6), 55.0 (OCH_3), 20.9 (CH_3-acetyl), 20.8 (CH_3-acetyl), 20.8 (CH_3-acetyl), 20.7 (CH_3-acetyl). **ESI-TOF MS** (m/z): Calcd for $C_{15}H_{22}O_{10}Na$ $[M+Na]^+$ 385.1; observed 385.3. NMR data agreed with previously reported data[295].

*Ethyl 2,3,5,6-tetra-O-acetyl-1-thio-β-D-galactofuranoside (***43***)*



A solution of **62** (3.18 g, 8.78 mmol) and ethanethiol (800 µL, 10.80 mmol) in anhydrous CH_2Cl_2 (100 mL) was cooled to 0 °C, followed by dropwise addition of $BF_3{\cdot}OEt_2$ (3.25 mL, 26.33 mmol) over 10min. The reaction mixture was stirred overnight allowing it to warm up to rt gradually, diluted with CH_2Cl_2 (200 mL), washed with aq. saturated $NaHCO_3$, brine and dried with Na_2SO_4. The solvent was removed under reduced pressure and the product purified by flash column chromatography using hexanes/EtOAc (3:1 to 2:1) to yield building block **43** (2.6 g, 6.63 mmol, 75%) as a colorless solid. **R_f =** 0.5 (hexanes/EtOAc 1:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 5.37 (ddd, *J* = 7.2, 4.5, 3.7 Hz, 1H, H-5), 5.30 (d, *J* = 2.2 Hz, 1H, H-1), 5.03 (t, *J* = 2.3 Hz, 1H, H-2), 5.00 (ddd, *J* = 6.1, 2.4, 0.7 Hz, 1H, H-3), 4.36 (ddd, *J* = 6.1, 3.7, 0.7 Hz, 1H, H-4), 4.30 (dd, *J* =

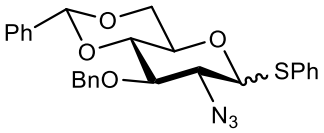
11.8, 4.6 Hz, 1H, H-6a), 4.16 (dd, *J* = 11.7, 7.1 Hz, 1H, H-6a), 2.73 – 2.52 (m, 2H, CH_2-SEt), 2.11 (s, 3H, CH_3-acetyl), 2.07 (d, *J* = 5.5 Hz, 6H, CH_3-acetyl), 2.02 (s, 3H, CH_3-acetyl), 1.26 (t, *J* = 7.4 Hz, 3H, CH_3-SEt). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 170.6 (C=O, acetyl), 170.1 (C=O, acetyl), 170.0 (C=O, acetyl), 169.7 (C=O, acetyl), 87.7 (C-1), 81.8 (C-2), 79.3 (C-4), 76.7 (C-3), 69.1 (C-5), 62.6 (C-6), 25.3 ($S\underline{C}H_2CH_3$), 20.9 (CH_3-acetyl), 20.9 (CH_3-acetyl), 20.8 (CH_3-acetyl), 20.7 (CH_3-acetyl), 14.9 ($SCH_2\underline{C}H_3$). **ESI-TOF MS** (m/z): Calcd for $C_{16}H_{24}O_9SNa$ $[M+Na]^+$ 415.1; observed 415.2. NMR data agreed with previously reported data[296].

Synthesis of glucosamine building blocks

*Phenyl 2-azido-4,6-O-benzylidene-2-deoxy-1-thio-α-D-glucopyranose (**65**)*

Benzylidene acetal was synthesized from glucosamine hydrochloride (20 g, 92.75 mmol) in three steps following reported protocols[278]. The identity of the product was determined by NMR and comparison with the data from the literature[278]. Compound **65** was obtained as a white solid (11.13 g, 28.88 mmol, 44% over three steps). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.43 (m, 4H, H-Ar), 7.37 – 7.19 (m, 6H, H-Ar), 5.48 (d, *J* = 6.2 Hz, 2H, H-1, CH-benzylidene), 4.31 (m, 1H, H-5), 4.15 (m, 1H, H-6a), 3.97 (td, *J* = 9.5, 2.8 Hz, 1H, H-3), 3.83 (dd, *J* = 10.0, 5.6 Hz, 1H, H-2), 3.67 (t, *J* = 10.3 Hz, 1H, H-6b), 3.49 (t, *J* = 9.3 Hz, 1H, H-4).**^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 132.5 (CH-Ar), 129.5 (CH-Ar), 129.2 (CH-Ar), 128.5 (CH-Ar), 128.1 (CH-Ar), 126.7 (CH-Ar), 102.2 (CH-benzylidene), 87.8 (C-1), 81.7 (C-4), 70.8 (C-3), 68.5 (C-6), 63.9 (C-2), 63.4 (C-5). **ESI-HRMS** (m/z): Calcd for $C_{19}H_{19}N_3O_4SNa$ $[M+Na]^+$ 408.0993; observed 408.0978.

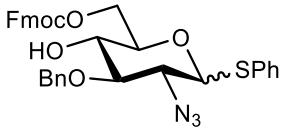
*Phenyl 2-azido-3-O-benzyl-4,6-O-benzylidene-2-deoxy-1-thio-α,β-D-glucopyranose (**66**)*



Benzylidene acetal 65 (5.56 g, 14.43 mmol) was dissolved in anhydrous DMF (80 mL). The solution was cooled to 0 °C and sodium hydride (692.3 mg, 28.85 mmol) was added to the reaction mixture. After stirring the suspension for 15 min, benzyl bromide (2.57 mL, 21.64 mmol) was added dropwise. The reaction mixture was allowed to warm up to rt, stirred for 5 h and quenched by addition of methanol at 0 °C. Volatiles were removed under reduced pressure. The residue was dissolved in CH2Cl2 (100 mL) and extracted twice with saturated NaHCO3 and brine, dried over Na2SO4, filtered and concentrated under reduced pressure. Product 66 was purified by flash column chromatography using 0 to 20% EtOAc in hexanes to yield a white solid (5.55 g, 11.67

mmol, 81% yield). R_f = 0.54 (hexanes/EtOAc, 4:1). Anomeric mixture α:β 1:0.7. $[\alpha]_D^{25}$= - 84.9 (c 1, $CHCl_3$). **IR (ATR)** (neat) v_{max} cm^{-1} 2961 (C-H Ar), 2927 (C-H Ar), 2110 (N=N=N), 1455, 1376, 1276, 1095 (C-O ether), 970, 743, 697. **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.59 – 7.45 (m, 7H, H-Ar), 7.45 – 7.27 (m, 20H, H-Ar), 5.61 (s, 1H, CH-benzylidene), 5.58 (d, J = 4.7 Hz, 1H, H-1 α:β), 5.05 – 4.89 (m, 2H, CHHPh), 4.88 – 4.75 (m, 2H, CHH-Ph), 4.57 (s, 1H, H-5), 4.53 – 4.36 (m, 2H, H-4), 4.24 (dd, J = 10.4, 4.9 Hz, 1H, H-6a), 4.03 – 3.93 (m, 2H, H-2), 3.85 – 3.72 (m, 3H, H-3 and H-6b) **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 132.6 (CH-Ar), 129.8 (CH-Ar), 128.6 (CH-Ar), 128.5 (CH-Ar), 128.4 (CH-Ar), 128.1 (CH-Ar), 126.1 (CH-Ar), 101.6 (CH benzylidene), 87.9 (C-1), 82.8 (C-3), 77.9 (C-2), 75.3 (CH_2-Ph), 72.2 (C-5), 68.7 (CH_2-C6), 63.9 (C-4). **ESI-HRMS** (m/z) Calc. for $C_{26}H_{25}N_3O_4SNa$, $[M+Na]^+$ 498.1463; observed 498.1472.

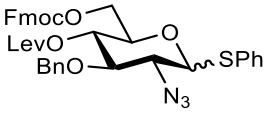
*Phenyl 2-azido-3-O-benzyl-6-O-(9-fluorenylmethoxycarbonyl)-2-deoxy-1-thio-α,β-D-glucopyranose (***67***)*



To a solution of **66** (5.55 g, 11.67 mmol) in 75 mL of MeOH/CH_2Cl_2 (4:1), p-TSA (1 g, 5.84 mmol) was added and the mixture was stirred at 35 °C for 5 h. The reaction mixture was diluted with CH_2Cl_2, extracted with saturated $NaHCO_3$ and water. The organic phase was dried over Na_2SO_4 and the solvent was removed under reduce pressure to get a glucosamine diol, which was used in the next step without further purification. The crude compound (4.52 g, 11.67 mmol) was dissolved in anhydrous CH_2Cl_2 (150 mL) and pyridine (2.14 mL, 23.3 mmol) was added. The mixture was stirred at rt for 15 min, subsequently cooled to 0 °C and Fmoc chloride (3.17 g, 12.25 mmol) was slowly added. The reaction was stirred at rt for 2 h, diluted with CH_2Cl_2 (200 mL) and extracted with aqueous citric acid 0.5 M. The aqueous layer was extracted with CH_2Cl_2 and the organic layers were combined, dried over Na_2SO_4 and filtered. The solvent was removed under reduced pressure and the residue was purified by flash column chromatography (hexanes/EtOAc 3:1) to yield **67** as a white solid (5.7 g, 9.35 mmol, 80% over two steps). **R_f =** 0.46 (hexanes/EtOAc, 2:1), $[\alpha]_D^{25}$= - 43.2 (c 1, $CHCl_3$), **IR (ATR)** (neat) v_{max} cm^{-1} 3503 (C-H Ar), 3067 (C-H Ar), 2957 (C-H Ar), 2110 (N=N=N), 1748 (C=O), 1584, 1479, 1452, 1262 (C-O ester), 1077 (C-O ether), 970, 787, 759, 741, 698. Anomeric mixture α:β 1:2 **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.82 – 7.74 (m, 3H, H Ar Fmoc), 7.61 (t, *J* = 6.5 Hz, 4H, H Ar Fmoc), 7.52 (dd, *J* = 7.9, 1.8 Hz, 2H, H-Ar), 7.47 – 7.22 (m, 18H, H-Ar), 5.59 (d, *J* = 5.5 Hz, 1H, H-1), 4.97 (dd, *J* = 18.3, 11.1 Hz, 2H, CHHPh), 4.81 (dd, *J* = 24.0, 11.1 Hz, 2H, CHH-Ph), 4.55 (dd, *J* = 12.0, 4.4 Hz, 1H, H-6a), 4.49 – 4.36 (m, 5H, H-5, H-6b, CH_2-Fmoc), 4.33 – 4.21 (m, 2H, CH-Fmoc), 3.92 (dd, *J* = 10.1, 5.5 Hz, 1H, H-2), 3.69 (dd, *J* = 10.1, 8.7 Hz, 1H, H-3), 3.65 – 3.55 (m, 1H, H-4) **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 155.8 (C=O Fmoc), 143.3, 143.3, 141.4, 137.8, 133.8, 132.3, 129.3, 128.9, 128.4, 128.1, 127.3, 125.3, 125.2, 120.2 (24C-Ar), 87.3 (C-1), 81.2 (C-3), 75.8 (CH_2-Ph), 70.5 (C-4), 70.2 (CH_2-Fmoc),

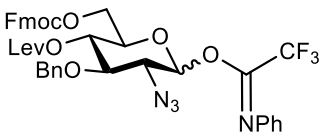
69.6 (C-5), 66.4 (C-6), 64.7, 63.7 (C-2), 46.8 (CH-Fmoc). **ESI-HRMS** (m/z) Calc. for $C_{34}H_{31}N_3O_6SNa$ $[M+Na]^+$ 632.1831, observed 632.1843.

*Phenyl 2-azido-3-O-benzyl-4-O-levulinoyl-6-O-(9-fluorenylmethoxycarbonyl)-2-deoxy-1-thio-α,β-D-glucopyranose (***44***)*



Compound **67** (5.70 g, 9.35 mmol) was dissolved in anhydrous 170 mL of CH_2Cl_2 and cooled down to 0 °C. Subsequently, levulinic acid (1.53 mL, 14.96 mmol), DMAP (685 mg, 5.61 mmol) and DIC (2.32 mL, 14.96 mmol) were added and stirred for 3 h at rt The reaction was washed with 1 M HCl, saturated $NaHCO_3$, and brine. The solvent was removed under reduced pressure and the product was purified by flash column chromatography (hexanes/EtOAc 3:2) to yield **44** as a colorless foam (5.1 g, 7.21 mmol, 77%). **R_f =** 0.44 (hexanes/EtOAc, 1:1), $[\alpha]_D^{25}$= -41.3 (c 1, $CHCl_3$). **IR (ATR)** (neat) ν_{max} cm^{-1} 3023 (C-H Ar), 2956 (C-H Ar), 2903 (C-H Ar), 2109 (N=N=N), 1748 (C=O), 1718 (C=O), 1584, 1498, 1479, 1451, 1441, 1401, 1363, 1255 (C-O ester), 1208, 1177, 1151, 1070 (C-O ether), 1043, 1026, 1000, 970, 875, 786, 739, 691, 667. Anomeric mixture α:β 1:2 **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.67 (d, *J* = 7.5 Hz, 2H, Ar, Fmoc), 7.52 (dd, *J* = 7.5, 1.3 Hz, 2H, Ar, Fmoc), 7.43 (dd, *J* = 7.6, 1.9 Hz, 2H, Ar), 7.37 – 7.12 (m, 14H, Ar), 5.50 (d, *J* = 5.5 Hz, 1H, H-1), 5.03 (dd, *J* = 10.3, 9.0 Hz, 1H, H-4), 4.77 (d, *J* = 11.2 Hz, 1H, CH_2-Bn), 4.63 (d, *J* = 11.1 Hz, 1H, C<u>H</u>H-Ph), 4.49 (m, 1H, H-5), 4.38 – 4.08 (m, 5H, H-6a, H-6b, CH Fmoc, CH_2-Fmoc), 3.92 (dd, *J* = 10.2, 5.5 Hz, 1H, H-2), 3.75 (dd, *J* = 10.2, 9.0 Hz, 1H, H-3), 2.70 – 2-36 (m, 3H, $CH_{2a/b}$-Lev), 2.27 (ddd, *J* = 17.2, 6.7, 5.7 Hz, 1H, CH_{2b}-Lev), 2.01 (s, 3H, CH_3-Lev). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 206.3 (C=O ketone), 171.8 (C=O ester), 155.0 (C=O Fmoc), 143.6, 143.4, 141.5, 141.2, 137.5, 132.6, 129.4, 128.6, 128.2, 128.0, 127.3, 125.4, 125.3, 120.2 (24C, Ar), 87.1 (C-1), 79.2 (C-3), 75.5 (CH_2-Ph), 70.8 (C-4), 70.1 (CH_2-Fmoc), 69.1 (C-5), 66.0 (C-6), 63.7 (C-2), 46.8 (CH-Fmoc), 37.9 (CH_2-Lev), 29.8 (CH_3-Lev), 27.9 (CH_2-Lev). **ESI-HRMS** (m/z) Calc. for $C_{39}H_{37}N_3O_8SNa$ $[M+Na]^+$ 730.2199, observed 730.2218.

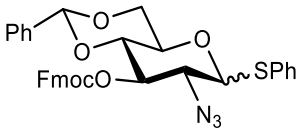
*Phenyl 2-azido-3-O-benzyl-4-O-levulinoyl-6-O-(9-fluorenylmethoxycarbonyl)-2-deoxy-D-glucopyranosyl N-phenyltrifluoroacetimidate (***45***)*



To a solution of **44** (2 g, 2.83 mmol) in acetone/water (9:1, 60 mL), *N*-bromo succinimide (1.76 g, 9.89 mmol) was added and stirred overnight at rt. Upon completion, the reaction was diluted with CH_2Cl_2 (100 mL), washed with 10% aqueous solution of $Na_2S_2O_3$, water and brine. The organic phase was dried over Na_2SO_4, filtered and concentrated *in vacuo*. The glucosamine lactol was

used for the next step without further purification. The crude compound (1.74 g, 2.83 mmol) was dissolved in anhydrous CH_2Cl_2 (12 mL) under an Argon atmosphere, subsequently Cs_2CO_3 (1.84 g, 5.65 mmol) and 2,2,2-trifluoro-*N*-phenylacetimidoyl chloride (1.83 mL, 11.31 mmol) were added to the solution. The reaction was vigorously stirred at rt for 40 min, diluted with 100 mL of CH_2Cl_2, filtered through a plug of Celite and concentrated under reduced pressure. The product was purified by flash column chromatography hexanes/EtOAc (2:1) to yield **45** as a white foam (1.85 g, 2.36 mmol, 83% over two steps). **R_f =** 0.44 (hexanes/EtOAc, 2:1). $[\alpha]_D^{25}$ = 3.47 (c 1, $CHCl_3$). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.67 (dd, *J* = 8.4, 0.9 Hz, 2H, Ar, Fmoc), 7.52 (dd, *J* = 9.9, 8.6 Hz, 2H, Ar, Fmoc), 7.37 – 7.11 (m, 11H, Ar), 7.04 – 6.95 (m, 1H, Ar, NHPh), 6.79 (d, *J* = 7.6 Hz, 2H, Ar, NHPh), 5.50 (s, 1H, 1-H), 5.03 (t, *J* = 9.6 Hz, 1H, H-4), 4.73 (s, 1H, C<u>H</u>H-Ph), 4.63 (d, *J* = 11.3 Hz, 1H CH<u>H</u>-Ph), 4.33 (dd, *J* = 10.1, 7.2 Hz, 1H, CH_{2a}-Fmoc), 4.26 – 4.10 (m, 4H, H-5, H-6, CH-Fmoc, CH_{2b}-Fmoc), 3.65 (d, *J* = 9.0 Hz, 1H, H-2), 3.47 (d, *J* = 9.9 Hz, 1H, H-3), 2.70 – 2.46 (m, 2H, CH_2a-Lev), 2.39 (ddd, *J* = 17.3, 7.8, 5.3 Hz, 1H, CH_2b-Lev), 2.26 (dt, *J* = 17.3, 6.1 Hz, 1H, CH_{2b}-Lev), 1.98 (s, 3H, CH_3-Lev).**^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 206.3 (C=O ketone), 171.7 (C=O ester), 154.9 (C=O Fmoc), 143.5, 143.4, 141.4, 141.3, 137.5, 129.0, 128.7, 128.6, 128.2, 128.2, 128.0, 127.3, 125.5, 125.4, 124.7, 120.2, 120.1 (24C, Ar), 119.3 (CF_3), 95.4 (C-1), 80.1 (C-3), 75.3 (CH_2-Ph), 72.9 (C-5), 70.3 (CH_2-Fmoc), 69.8 (C-4), 65.7 (C-6), 64.9 (C-2), 46.7 (CH-Fmoc), 37.9 (CH_{2a}-Lev), 29.8 (CH_3-Lev), 27.9 (CH_{2b}-Lev). **ESI-HRMS** (m/z) Calc. for $C_{41}H_{37}F_3N_3O_9Na$ $[M+Na]^+$ 809.2410, observed 809.2361.

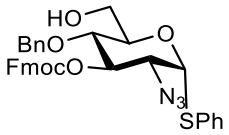
*Phenyl 2-azido-3-O-(9-fluorenylmethoxycarbonyl)-4,6-O-benzylidene-2-deoxy-1-thio-α-D-glucopyranose (***68***)*



To a solution of **65** (5.56 g, 12.97 mmol) in 100 mL of anhydrous CH_2Cl_2, pyridine (2.37 mL, 29.32 mmol) was added and stirred at rt for 15 min. The reaction mixture was subsequently cooled down to 0 °C and Fmoc chloride (7.58 g, 29.32 mmol) was added. The mixture was stirred at rt for 5 h, diluted with CH_2Cl_2 (250 mL), and extracted with aqueous citric acid 0.5M. The aqueous layer was extracted with with CH_2Cl_2 and the organic layers were combined, dried over Na_2SO_4 and filtered. The solvent was removed under reduced pressure and the residue was purified by flash column chromatography with hexanes/EtOAc (3:1) to yield **68** (6.90 g, 11.35 mmol, 78%) as a white solid. **R_f =** 0.5 (hexanes/EtOAc, 3:1), $[\alpha]_D^{20}$= -84.9 (c 1, $CHCl_3$). **IR (ATR)** (neat) ν_{max} cm^{-1} 3067 (C-H Ar), 2957 (C-H Ar). 2109 (N=N=N), 1756 (C=O), 1451, 1378, 1256 (C-O ester), 1097, 977, 758, 740, 698. **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.69 (d, *J* = 7.6 Hz, 2H, H-Ar Fmoc), 7.53 (t, *J* = 7.9 Hz, 2H, H-Ar Fmoc), 7.48 – 7.42 (m, 2H, H-Ar SPh), 7.41 – 7.11 (m, 12H. H-Ar), 5.59 (d, *J* = 5.7 Hz, 1H, H-1), 5.47 (s, 1H, CH benzylidene), 5.23 (t, *J* = 9.9 Hz, 1H, H-3), 4.49 – 4.34 (m, 3H, H-5, CH_2-Fmoc), 4.25 – 4.13 (m, 2H, H-6a, CH Fmoc), 4.05 (dd, *J* = 10.3, 5.7 Hz, 1H, H-2), 3.71 (td, *J*

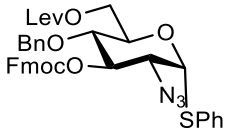
= 10.0, 6.5 Hz, 2H, H-4, H-6b). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 154.3 (C=O, Fmoc), 143.3, 143.3, 141.4, 136.8, 132.8, 132.6, 129.4, 129.3, 128.4, 128.4, 128.0, 127.3, 126.3, 125.3, 125.3, 120.2 (18C, Ar), 101.9 (CH benzylidene), 87.8 (C-1), 79.4 (C-4), 74.9 (C-3), 70.6 (CH_2-Fmoc), 68.6 (C-6), 63.9 (C-5), 62.6 (C-2), 46.7 (CH-Fmoc). **ESI-HRMS** (m/z): Calcd for $C_{34}H_{29}N_3O_6SNa$ $[M+Na]^+$ 630.1674, observed 630.1678.

*Phenyl 2-azido-3-O-(9-fluorenylmethoxycarbonyl)-4-O-benzyl-2-deoxy-1-thio-α-D-glucopyranose (***69***)*



A solution of **68** (6.90 g, 11.35 mmol) in CH_2Cl_2 (250 mL) was cooled to 0 °C, subsequently a borane solution (1 M in THF, 45.42 mL, 45.2 mmol) and TMSOTf (450 µL, 2.50 mmol) were added. The mixture was stirred at rt for 3 h and the reaction quenched by portion-wise addition of a Et_3N/MeOH (1:10) solution until no further gas evolved. The resulting mixture was coevaporated with MeOH to dryness. The product was purified by flash column chromatography (hexanes/EtOAc 3:1) to afford **69** (4.30 g, 7.05 mmol, 62%). **R_f =** 0.33 (hexanes/EtOAc, 2:1), $[\alpha]_D^{25}$ = -84.9 (c 1, $CHCl_3$), **IR (ATR)** (neat) ν_{max} cm^{-1} 3066 (C-H Ar), 2956 (C-H Ar), 2925 (C-H Ar), 2109 (N=N=N), 1755 (C=O), 1584, 1479, 1452, 1387, 1258 (C-O ester), 1082 (C-O ether), 1027, 998, 859, 783, 759, 739, 698. **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.80 – 7.72 (m, 2H, H-Ar Fmoc), 7.67 – 7.57 (m, 2H, H-Ar Fmoc), 7.54 – 7.44 (m, 2H, H-Ar), 7.43 – 7.23 (m, 12H, H-Ar), 5.63 (d, *J* = 5.6 Hz, 1H, H-1), 5.24 (dd, *J* = 10.6, 9.1 Hz, 1H, H-3), 4.71 – 4.52 (m, 3H, CHH-Ph, CH_2-Fmoc), 4.40 (dd, *J* = 10.5, 7.4 Hz, 1H, CH_2-Fmoc), 4.33 – 4.20 (m, 2H, H-6a, CH-Fmoc), 4.01 (dd, *J* = 10.6, 5.5 Hz, 1H, H-2), 3.80 – 3.70 (m, 3H, H-4. H-5, H-6b). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 154.5 (C=O, Fmoc), 143.4, 143.2, 141.4, 137.3, 132.8, 132.7, 129.4, 128.6, 128.2, 128.2, 128.1, 128.1, 127.4, 127.3, 125.3, 125.2, 120.2 (18C, Ar), 87.0 (C-1), 78.4 (C-4), 75.6 (C-3), 75.0 (CH_2-Ph), 72.2 (C-5), 70.5 (CH_2-Fmoc), 62.4 (C-6), 61.4 (C-2), 46.8 (CH-Fmoc). **ESI-HRMS** (m/z): Calcd for $C_{34}H_{31}N_3O_6SNa$ $[M+Na]^+$ 632.1831; observed 632.1822.

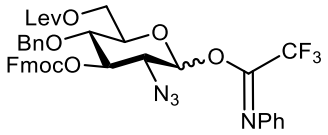
*Phenyl 2-azido-3-O-(9-fluorenylmethoxycarbonyl)-4-O-benzyl-6-O-levulinoyl-2-deoxy-1-thio-α,β-D-glucopyranose (***47***)*



Compound **69** (4.30 g, 5.0 mmol) was dissolved in anhydrous CH_2Cl_2 (150 mL), cooled down to 0 °C and subsequently treated with DMAP (258 mg, 2.12 mmol), DIC (1.75 mL, 11.28 mmol), and levulinic acid (1.16 mL, 11.28 mmol). The solution was stirred and after completion as monitored

by TLC (3 h), the reaction was diluted with CH_2Cl_2 (150 mL), extracted with HCl 1M, saturated $NaHCO_3$, and brine. The organic layer was dried under Na_2SO_4, filtered through a celite pad and concentrated under reduced pressure. The product was purified by flash column chromatography using hexanes/EtOAc (3:2) to yield glucosamine building block **47** as a colorless foam (3.40 g, 4.80 mmol, 68%) $\mathbf{R_f}$ **=** 0.46 (hexanes/EtOAc, 1:1). $[\alpha]_D^{25}$ = -41.3 (c 1, $CHCl_3$). **IR (ATR)** (neat) ν_{max} cm^{-1} 3066 (C-H Ar), 2957 (C-H Ar), 2108 (N=N=N), 1752 (C=O), 1718, 1584, 1498, 1479, 1452, 1441, 1387, 1360, 1253 (C-O ester), 1231, 1182, 1157, 1083 (C-O ether), 1027, 998, 967, 857, 739, 699, 667. **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.78 (ddt, *J* = 7.5, 5.4, 0.9 Hz, 2H, H-Ar Fmoc), 7.65 (ddd, *J* = 11.5, 7.5, 1.0 Hz, 2H, H-Ar Fmoc), 7.56 – 7.46 (m, 2H, H-Ar), 7.46 – 7.21 (m, 12H, H-Ar), 5.66 (d, *J* = 5.5 Hz, 1H, H-1, α anomer), 5.26 (dd, *J* = 10.7, 9.0 Hz, 1H, H-3), 4.70 – 4.48 (m, 4H, H-5, CHH-Ph, CH_{2a}-Fmoc), 4.46 – 4.33 (m, 2H, H-6a, CH_{2b}-Fmoc), 4.31 – 4.19 (m, 2H, H-6b, CH-Fmoc), 4.06 (dd, *J* = 10.6, 5.5 Hz, 1H, H-2), 3.74 – 3.64 (m, 1H, H-4), 2.79 – 2.64 (m, 1H, CH_2-Lev), 2.55 (t, *J* = 6.3 Hz, 2H, CH_2-Lev), 2.20 (s, 3H, CH_3-Lev).**^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 206.5 (C=O ketone), 172.5 (C=O ester), 154.4 (C=O Fmoc), 143.4, 143.2, 141.4, 137.1, 132.8, 132.4, 129.3, 128.6, 128.2, 128.2, 128.1, 128.1, 128.0, 127.4, 127.3, 125.2, 125.2, 120.2 (18C, Ar), 86.7 (C-1), 78.4 (C-3), 75.9(C-4), 75.0 (CH_2-Ph), 70.5 (CH_2-Fmoc), 69.9 (C-5), 62.8 (C-6), 62.2 (C-2), 46.8 (CH-Fmoc), 37.9 (CH_2-Lev), 30.0 (CH_3-Lev), 27.8 (CH_2-Lev). **ESI-HRMS** (m/z): Calcd for $C_{39}H_{37}N_3O_8SNa$ $[M+Na]^+$ 730.2199, observed 730.2170,

*Phenyl 2-azido-3-O-(9-fluorenylmethoxycarbonyl)-4-O-benzyl-6-O-levulinoyl-2-deoxy-D-glucopyranosyl N-phenyltrifluoroacetimidate (***48***)*

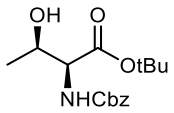


To a solution of **47** (2 g, 2.83 mmol) in acetone/water (9:1, 60 mL), *N*-bromo succinimide (1.76 g, 9.89 mmol) was added and stirred overnight at rt Upon completion, the reaction was diluted with CH_2Cl_2 (100 mL), washed with 10% aqueous solution of $Na_2S_2O_3$, water, and brine. The organic phase was dried over Na_2SO_4, filtered and concentrated *in vacuo*. The glucosamine lactol was used for the next step without further purification. The crude compound (1.74 g, 2.83 mmol) was dissolved in anhydrous CH_2Cl_2 (12 mL) under Argon atmosphere, subsequently Cs_2CO_3 (1.84 g, 5.65 mmol) and 2,2,2-trifluoro-*N*-phenylacetimidoyl chloride (1.83 mL, 11.31 mmol) were added to the solution. The reaction was vigorously stirred at rt for 40 min, diluted with 100 mL of CH_2Cl_2, filtered through a plug of Celite and concentrated under reduced pressure. The product was purified by flash column chromatography hexanes/EtOAc (2:1) to yield **48** as a white solid (1,60 g, 2.03 mmol, 72% over two steps). $\mathbf{R_f}$ **= 0.38** (hexanes/EtOAc, 2:1). $[\alpha]_D^{25}$ = 7.88 (c 1, $CHCl_3$). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.77 (ddt, *J* = 6.6, 5.6, 0.9 Hz, 2H, H-Fmoc), 7.68 – 7.55 (m, 2H, H-Fmoc), 7.46 – 7.16 (m, 12H, H-Ar), 7.16 – 7.10 (m, 1H, H-Ar NPh), 6.86 (d, *J* = 7.8 Hz, 2H, H-

Ar NPh), 5.64 (s, 1H, H-1), 4.91 (s, 1H, H-3), 4.64 – 4.55 (m, 2H, CHH-Ph, CH_2-Fmoc), 4.51 (d, *J* = 11.1 Hz, 1H, CHH-Ph), 4.43 (dd, *J* = 10.5, 7.1 Hz, 1H, CH_2-Fmoc), 4.26 (dd, *J* = 9.2, 5.1 Hz, 3H, H-6, CH-Fmoc), 3.82 – 3.64 (m, 2H, H-2, H-4), 2.73 (dt, *J* = 11.9, 6.4 Hz, 2H, CH_{2a}-Lev), 2.65 – 2.52 (m, 2H, CH_{2b}-Lev), 2.18 (s, 3H, CH_3-Lev).**^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 206.5 (C=O ketone), 172.4 (C=O ester), 154.5 (C=O Fmoc), 143.2, 143.1, 141.5, 141.4, 136.9, 129.0, 128.6, 128.3, 128.1, 127.4, 125.2, 125.1, 124.8, 120.3 (24C-Ar), 119.3 (CF_3), 95.5 (C-1), 78.8 (C-3), 75.0 (CH_2-Ph), 74.8 (C-4), 73.8 (C-5), 70.5 (CH_2-Fmoc), 63.3 (C-2), 62.2 (C-6), 46.8 (CH-Fmoc), 37.9 (CH_{2a}-Lev), 30.0 (CH_3-Lev), 27.9 (CH_{2b}-Lev). **ESI-HRMS** (m/z) Calc. for $C_{41}H_{37}F_3N_3O_9Na$ $[M+Na]^+$ 809.2410, observed 809.2361.

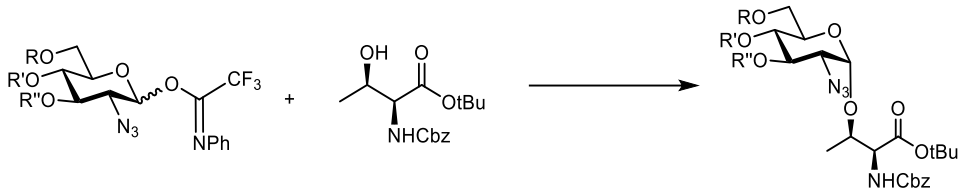
Synthesis of the Threonine Amino Acid Acceptor

*N^{α}-(carbobenzyloxy)-L-threonine tert-butyl ester (***46***)*



A mixture of t-BuOH (5.0 mL, 52.30 mmol), DIC (6.13mL, 39.62 mmol), and CuCl (227mg, 2.30 mmol) was stirred at rt under the exclusion of light for 5 d. The mixture was diluted with CH_2Cl_2 (25 mL), and a solution of Cbz-Thr-OH in CH_2Cl_2 (15 mL) and DIEA (300 µL) were added dropwise. The reaction mixture was stirred at rt for two days, diluted with CH_2Cl_2 (100 mL) and filtered through a celite pad. The organic layer was washed with aq. saturated $NaHCO_3$, brine and dried over anhydrous Na_2SO_4. The solvent was evaporated under reduced pressure and the product was purified by flash column chromatography using hexanes/EtOAc (3:1) to yield the threonine acceptor **46** as a white solid (2.68 g, 8.66 mmol, 73%). **R_f =** 0.43 (hexanes/EtOAc, 3:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.35 (s, 5H, H-Ar), 5.59 (s, 1H, NH), 5.12 (s, 2H, CH_2-Cbz), 4.31 – 4.17 (m, 2H, H-α Thr, H-β Thr), 1.47 (s, 9H, CH_3, OtBu), 1.23 (d, *J* = 6.4 Hz, 3H, CH_3-Thr).**^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 170.3 (C=O tBu ester), 156.8 (C=O Cbz), 136.3, 128.6, 128.3, 128.2 (5C-Ar), 82.7 (*q*C), 68.5 (C-β Thr), 67.2 (CH_2-Cbz), 59.7 (C-α Thr), 28.1 (CH_3, OtBu), 20.1 (CH_3-Thr). **ESI-HRMS** (m/z) Calc. for $C_{16}H_{23}NO_5Na$ $[M+Na]^+$ 332.1474, observed 332.1469. NMR data agreed with previously reported data[297].

Glycosylation of Threonine with glucosamine donors



The glycosylation of threonine **46** with glucosamine building blocks **45** and **48** was evaluated under the following conditions:

Table 5 – Conditions screened for glycosylation reaction

Entry	Donor	Donor g	Acceptor g	TMSOTf uL	Solvent mL	Concentration mM	Product g	Yield %
1	4,6	0.05	0.023	2	5	13	0.025	44
2	3,6	0.05	0.023	2	5	13	0.024	43
3	4,6	0.204	0.08	5	9	29	0.090	38*
4	3,6	0.250	0.098	6	10	29	0.122	42
5	4,6	1.00	0.39	23	45	28	0.139	12
6	4,6	1.14	0.448	26	50	29	0.498	38*

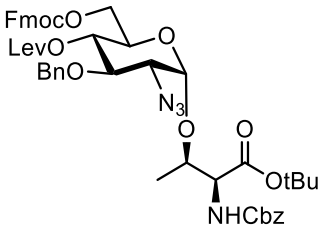
* Accumulated yield for 5 different batches

After optimization of the process, large scale glycosylation was performed in five batches of 200 mg each (total 1 g) .

For medium scale: A mixture of glucosamine donor (250 mg, 0.318 mmol) and threonine acceptor **46** (98.3 mg, 0.318 mmol) was dissolved, co-evaporated twice with toluene and dried under high vacuum overnight. The remaining was dissolved in an anhydrous mixture of Et_2O/CH_2Cl_2 (6:1, 11 mL). Freshly activated 4Å molecular sieves were added and the suspension was stirred for 30 min at rt under Argon. Subsequently, the mixture was cooled down to -15 °C, TMSOTf (7 µL, 0.04 mmol) was added and stirred for 30 min at this temperature, then the ice bath was removed and reaction was stirred at rt for additional 30 min.

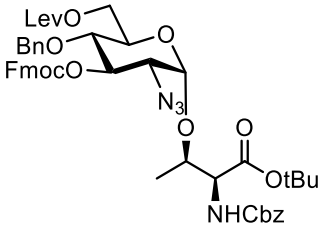
The purification of the glycosylated threonine products was performed following two different protocols:

*N^{α}-(carbobenzyloxy)-O-(2-azido-3-O-benzyl-4-O-levulinoyl-6-O-(9-fluorenylmethoxycarbonyl)-2-deoxy-α-D-glucopyranosyl)-L-threonine tert-butyl ester (***70***)*



Compound **70** was separated by flash column chromatography using a stepwise gradient from 40 to 50% of EtOAc in hexanes (for large scale the raw products were pooled and purified at once). The fractions containing the product were collected and the solvent was removed under reduced pressure. The remaining was treated with 7 mL of Ac_2O/pyridine (1:1) and stirred at rt for 1h. The mixture was diluted with toluene and co-evaporated twice. The product was isolated by flash column chromatography using (2:1 hexanes/EtOAc) to give pure α-glycosylated amino acid **70** (25.5 mg, 0.028 mmol 44%, 498 mg, 0.549 mmol, 38%). **R_f =** 0.51 (hexanes/EtOAc 2:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.77 (d, *J* = 7.5 Hz, 2H, H Ar Fmoc), 7.61 (dd, *J* = 7.5, 4.2 Hz, 2H, H Ar Fmoc), 7.48 – 7.23 (m, 14H, H-Ar), 5.57 (d, *J* = 9.5 Hz, 1H, NH), 5.17 (s, 2H, CH_2-Cbz), 5.11 – 5.03 (m, 1H, H-4), 4.99 (d, *J* = 3.8 Hz, 1H, H-1, (α-linkage)), 4.81 (d, *J* = 11.1 Hz, 1H, CH_{2a}-Bn), 4.69 (d, *J* = 11.1 Hz, 1H, CH_{2b}-Bn), 4.48 – 4.22 (m, 7H, H-α, H-β, H6, CH-Fmoc, CH_2-Fmoc), 4.01 (ddd, *J* = 9.0, 5.9, 2.6 Hz, 1H, H-5), 3.93 (d, *J* = 9.5 Hz, 1H, H-3), 3.40 (dd, *J* = 10.3, 3.8 Hz, 1H, H-2), 2.84 – 2.59 (m, 2H, CH_{2a}-Lev), 2.57 – 2.47 (m, 1H, CH_{2b}-Lev), 2.41 – 2.29 (m, 1H, CH_{2b}-Lev), 2.09 (s, 3H, CH_3-Lev), 1.50 (s, 9H, OtBu), 1.32 (d, *J* = 6.4 Hz, 3H, CH_3-Thr). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 206.3 (C=O ketone), 171.8(C=O ester), 169.2 (C=O carboxylate), 157.0 (C=O Cbz), 155.0 (C=O Fmoc), 143.5, 143.4, 141.4, 141.4, 137.7, 136.4, 128.7, 128.7, 128.6, 128.4, 128.3, 128.2, 128.0, 128.0, 127.9, 127.3, 125.4, 125.3, 124.9, 120.2 (24C-Ar), 99.1 (C-1), 83.0 (C-tBu), 77.4 (C-3), 77.0 (C-β Thr), 75.1 (CH_2-Bn), 70.8 (C-4), 70.2 (CH_2-Fmoc), 68.6 (C-5), 67.3 (CH_2-Cbz), 66.1 (C-6), 63.3 (C-2), 59.5 (C-α Thr), 46.8 (CH-Fmoc), 37.9 (CH_{2a}-Lev), 29.8 (CH_3-Lev), 28.1 (CH_3, OtBu), 27.9 (CH_{2b}-Lev), 19.1 (CH_3-Thr). **ESI-HRMS** (m/z) Calc. for $C_{49}H_{54}N_4O_{13}Na$ $[M+Na]^+$ 929.3585, observed 929.3599.

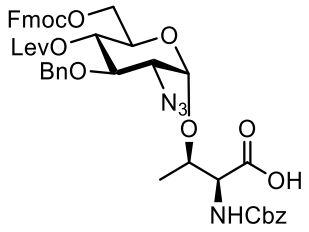
*N^{α}-(carbobenzyloxy)-O-(2-azido-3-O-(9-fluorenylmethoxycarbonyl)-4-O-benzyl-6-O-levulinoyl-2-deoxy-α-D-glucopyranosyl)-L-threonine tert-butyl ester (***71***)*



The product was purified by flash column chromatography using a stepwise gradient from 30 to 40% of EtOAc in hexanes. For the large scale synthesis of glycosylated threonine **71**, the crude

of the five batches were pooled and purified at once. The pure α-glycosylated amino acid **71** (0.122 mg, 0.105 mmol) was obtained as an amorphous white solid in 42% yield. **R_f =** 0.38 (hexanes/EtOAc 3:2). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.65 (t, *J* = 6.6 Hz, 2H, H-Ar Fmoc), 7.50 (dd, *J* = 12.6, 7.5 Hz, 2H, H-Ar Fmoc), 7.34 – 7.08 (m, 14H. H-Ar), 5.45 (d, *J* = 9.8 Hz, 1H, NH), 5.21 (dd, *J* = 10.7, 8.9 Hz, 1H, H-3), 5.03 (s, 2H, CH_2-Cbz), 4.96 (d, *J* = 3.8 Hz, 1H, H-1 α anomer), 4.54 (d, *J* = 11.1 Hz, 1H, CH_{2a}-Bn), 4.48 – 4.39 (m, 2H, CH_{2b}-Bn, CH_{2a}-Fmoc), 4.39 – 4.23 (m, 2H, H-β Thr, CH_{2b}-Fmoc), 4.25 – 4.09 (m, 4H, CH-Fmoc, H-α Thr, H-6), 3.88 (dt, *J* = 10.5, 3.4 Hz, 1H, H-5), 3.51 (t, *J* = 9.5 Hz, 1H, H-4), 3.12 (dd, *J* = 10.7, 3.8 Hz, 1H, H-2), 2.65 (dt, *J* = 9.6, 6.3 Hz, 2H, CH_{2a}-Lev), 2.47 (td, *J* = 6.7, 3.2 Hz, 2H, CH_{2b}-Lev), 2.09 (s, 3H, CH_3-Lev), 1.38 (s, 9H, OtBu), 1.20 (d, *J* = 6.5 Hz, 3H, CH_3-Thr). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 206.5 (C=O ketone), 172.5 (C=O ester), 169.2 (C=O carboxylate), 157.0 (C=O Cbz), 154.6 (C=O Fmoc), 143.3, 143.2, 141.4, 137.1, 136.4, 128.9, 128.6, 128.6, 128.3, 128.2, 128.2, 128.1, 128.1, 128.0, 127.3, 127.3, 125.2, 125.1, 120.2 (18C-Ar), 99.4 (C-1), 83.0 (C tBu), 77.4 (C-3), 76.7 (C-β Thr), 75.8 (C-4), 75.1 (CH_2-Bn), 70.4 (CH_2-Fmoc), 69.2 (C-2), 67.2 (CH_2-Cbz), 62.7 (C-6), 61.5 (C-2), 59.4 (C-α Thr), 46.8 (CH-Fmoc), 37.9 (CH_2-Lev), 30.0 (CH_3-Lev), 28.0 (CH_3, OtBu), 27.8 (CH_2-Lev), 19.1 (CH_3 (Thr)). **ESI-HRMS** (m/z) Calc. for $C_{49}H_{54}N_4O_{13}Na$ $[M+Na]^+$ 929.3585, observed 929.3561.

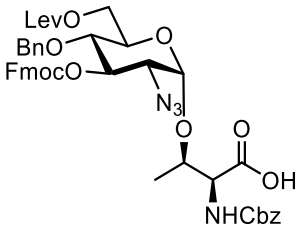
*N^{α}-(carbobenzyloxy)-O-(2-azido-3-O-benzyl-4-O-levulinoyl-6-O-(9-fluorenylmethoxycarbonyl)-2-deoxy-α-D-glucopyranosyl)-L-threonine (***39***)*



A solution of **70** (251 mg, 0.277 mmol) was dissolved in a mixture trifluoroacetic acid/H_2O (10:1 v/v, 16.5 mL) at 0 °C and stirred for 2 h at rt. Then the solvents were removed *in vacuo* and the residue was co-evaporated with toluene (5 x 30 mL). The products were purified by flash column chromatography (CH_2Cl_2/MeOH/AcOH 94.5:5:0.5) to give the glycosylated threonine **39** (177 mg, 0.208 mmol, 75%) as a white solid. R_f = 0.4 (CH_2Cl_2/MeOH/AcOH 94.5:5:0.5). **R_f =** 0.4 (CH_2Cl_2/MeOH/AcOH 94.5:5:0.5). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.76 (d, *J* = 7.5 Hz, 2H, H Ar Fmoc), 7.66 – 7.57 (m, 2H, H Ar Fmoc), 7.46 – 7.24 (m, 14H, H-Ar), 5.66 (d, *J* = 8.5 Hz, 1H, NH), 5.14 (d, *J* = 1.7 Hz, 2H, CH_2-Cbz), 5.11 – 5.02 (m, 2H, H-1 (α-linkage), H-4), 4.79 (d, *J* = 11.1 Hz, 1H, CH_2-Bn), 4.68 (d, *J* = 11.1 Hz, 1H, CH_2-Bn), 4.51 – 4.17 (m, 7H, H-α, H-β, H-6, CH-Fmoc, CH_2-Fmoc), 4.01 (dq, *J* = 7.6, 2.9 Hz, 1H, H-5), 3.90 (t, *J* = 9.6 Hz, 1H, H-3), 3.51 (dd, *J* = 10.1, 3.8 Hz, 1H, H-2), 2.79 – 2.57 (m, 2H, CH_{2a}-Lev), 2.51 (ddd, *J* = 17.2, 8.0, 5.3 Hz, 1H, CH_{2b}-Lev), 2.40 – 2.26 (m, 1H, CH_{2b}-Lev), 2.09 (s, 3H, CH_3-Lev), 1.29 (d, *J* = 6.4 Hz, 4H, CH_3-Thr). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 206.5 (C=O ketone), 171.8 (C=O ester), 156.6 (C=O Cbz), 155.0

(C=O Fmoc), 143.5, 143.3, 141.4, 141.4, 137.5, 136.0, 128.8, 128.6, 128.5, 128.4, 128.1, 128.1, 128.0, 127.3, 125.4, 125.3, 120.2 (24C-Ar), 98.5 (C-1), 77.9 (C-3), 76.5 (C-β Thr), 75.2 (CH_2-Bn), 70.7 (C-4), 70.2 (CH_2-Fmoc), 68.8 (C-5), 67.6 (CH_2-Cbz), 66.0 (C-6), 63.4 (C-2), 58.1(C-α Thr), 46.7 (CH-Fmoc), 37.9 (CH_{2a}-Lev), 29.8 (CH_3-Lev), 27.9 (CH_{2b}-Lev), 18.0 (CH_3-Thr). **ESI-HRMS** (m/z) Calc. for $C_{45}H_{46}N_4O_{13}Na$ $[M+Na]^+$ 873.2959, observed 873.2984.

*N^{α}-(carbobenzyloxy)-O-(2-azido-3-O-(9-fluorenylmethoxycarbonyl)-4-O-benzyl-6-O-levulinoyl-2-deoxy-α-D-glucopyranosyl)-L-threonine (***40***)*



A solution of glycosylated amino acid **71** (166 mg, 0.183 mmol) was dissolved in a mixture trifluoroacetic acid/H_2O (10:1 v/v, 11 mL) at 0 °C and stirred for 2 h at rt. Then, the solvents were removed *in vacuo* and the residue was co-evaporated with toluene (5 x 30 mL). The product was purified by flash column chromatography (CH_2Cl_2/MeOH/AcOH 94.5:5:0.5) to give glycosylated amino acid **40** (109 mg, 0.128 mmol, 70%) as a white solid. $\mathbf{R_f}$ = 0.4 (CH_2Cl_2/MeOH/AcOH 94.5:5:0.5). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 7.73 (t, *J* = 6.9 Hz, 2H, H-Ar Fmoc), 7.59 (dd, *J* = 16.6, 7.5 Hz, 2H, H-Ar Fmoc), 7.43 – 7.10 (m, 14H, H-Ar), 5.71 (d, *J* = 9.3 Hz, 1H, NH), 5.28 (d, *J* = 1.7 Hz, 1H, H-3), 5.10 (dd, *J* = 17.4, 5.1 Hz, 3H, H-1, CH_2-Cbz), 4.64 – 4.42 (m, 5H, H-α Thr, H-β Thr,CH_2-Bn, CH_2-Fmoc-), 4.38 (dd, *J* = 10.5, 7.3 Hz, 1H, CH_2-Fmoc), 4.31 – 4.17 (m, 3H, H6, CH-Fmoc), 3.96 (dt, *J* = 10.3, 3.4 Hz, 1H, H-5), 3.59 (t, *J* = 9.4 Hz, 1H, H-4), 3.35 (dd, *J* = 10.7, 3.8 Hz, 1H, H-2), 2.82 – 2.63 (m, 2H, CH_{2a}-Lev), 2.54 (td, *J* = 6.3, 3.2 Hz, 2H, CH_{2b}-Lev), 2.18 (s, 3H, CH_3-Lev), 1.30 (d, *J* = 6.5 Hz, 3H, CH_3-Thr). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 206.9, (C=O ketone), 173.6 (C=O ester), 172.5 (C=O carboxylate), 157.0 (C=O Cbz), 154.6 (C=O Fmoc), 143.4, 143.1, 141.4, 141.4, 137.0, 136.1, 129.0, 128.9, 128.7, 128.6, 128.6, 128.6, 128.3, 128.2, 128.2, 128.2, 128.1, 128.0, 128.0, 127.3, 127.3, 126.3, 125.2, 125.1, 120.2, (18C-Ar) 98.8 (C-1), 77.4 (C-3), 76.4 (C-β Thr), 75.8 (C-4), 75.1 (CH_2-Bn), 70.4 (CH_2-Fmoc), 69.4 (C-5), 67.5 (CH_2-Cbz), 62.6 (C-6), 61.7 (C-5), 58.3 (C-α Thr), 46.8 (CH-Fmoc), 37.9 (CH_{2a}-Lev), 30.0 (CH_3-Lev), 27.8 (CH_{2b}-Lev), 18.6 (CH_3-Thr). **ESI-HRMS** (m/z) Calc. for $C_{45}H_{46}N_4O_{13}Na$ $[M+Na]^+$ 873.2959, observed 873.2933.

3.6.2. Development of a strategy for the solid phase peptide synthesis of glycosylated amino acids of *T. cruzi* Y strain MLMs

General materials and methods

The functionalization of the resins, coupling of the first amino acid and glycosylated amino acids were performed manually in a 2 mL disposable polypropylene reactor equipped with 25 µm bottom frit (Multisynthech), and all solutions and solvents were transfer by suction. All solvents employed were HPLC-grade. The solvents used to dissolve the building blocks and to prepare the solutions for activation, acidic wash and capping were taken from an anhydrous solvent system (J.C Meyer) and further dried over molecular sieves (4 Å). The automated syntheses were performed in a home-built synthesizer developed at the Max Plank Institute of Colloids and Interfaces. The building blocks were co-evaporated three times with toluene and dried overnight on high vacuum before use. Flask for preparation of moisture sensitive solutions and building blocks were oven dried and argon flushed prior to use. All solutions were freshly prepared and kept under argon for the automation run. The photocleavable linkers traceless photolinker (L1), aminopentyl (L2) and photolabile diazo linker (L3) were synthesized according to established protocols[167,287]. Resin loading was determined by following previously established procedures[238,261]. Analytical and preparative high-performance liquid chromatography was performed using Agilent 1200 series equipped with a Multiple Wavelength Detector (MWD) and a Evaporative Light Scattering Detector (ELSD) using the following columns [A] Hypercarb column, 150 x 4.60 mm and [B] Hypercarb column 5 µm 150 x 10. The solvent system used was eluent A: 0.1% of formic acid in water, eluent B: acetonitrile. Products were lyophilized using a Chris Alpha 2-4 LC plus freeze dryer.

Coupling of the glycosylated amino acid to the solid support

Screening of the coupling to solid support (traceless linker)

Different conditions were screened to standardize the method for coupling the glycosylated amino acid to the solid support bearing the traceless photo labile linker (**L1**)[167].

Entry	Resin	BB Eq	Reagent	Eq	Solvent	Outcome
1	**L1**	4	DIC	2	CH_2Cl_2:DMF 1:1	Not coupled
2	**L1**	4	HATU, DIPEA	3,8:4	CH_2Cl_2:DMF 1:1	Not coupled
3	**L1**	4	DIC, *N*-methyl imidazole	4:4	CH_2Cl_2:DMF 1:1	Coupled

Symmetric anhydride method

The solid support **L1** (40 mg, 0.014 mmol, loading = 0.34 mmol/g) was swollen in CH_2Cl_2 (1.5 mL) for 30 min and solvent was drained. Fmoc-Thr-(OtBu)-OH (43.2 mg, 0.109 mmol) was dissolved in a mixture CH_2Cl_2/DMF 1:1 (1 mL) and subsequently DIC (8.4 µL, 0.054 mmol) and pyridine (9µL, 0.109) were added and stirred for 10 min. The activated amino acid was transferred to the polypropylene reactor and shaken overnight at rt. Coupling was monitored by determining the loading through Fmoc removal quantification at 290 nm using previously reported methods[238,261].

Aminium/Uronium method

Solid support **L1** (40 mg, 0.014 mmol, loading = 0.34 mmol/g) was swollen in CH_2Cl_2 (1.5 mL) for 30 min and drained. Fmoc-Thr(OtBu) (21.62 mg, 0.054 mmol) was dissolved in a mixture CH_2Cl_2/DMF 1:1 (1 mL). HATU (20.68 mg, 0.052 mmol), and DIEA (14 µL, 0.051 mmol) were subsequently added and the mixture was stirred for 30 seconds. The activated amino acid was transferred to the polypropylene reactor containing the resin and shaken overnight at rt Solid support was washed with CH_2Cl_2 (3 x 1.5 mL), DMF (3 x 1.5 mL), and CH_2Cl_2 (3 x 1.5 mL) and dried. Coupling was monitored by determining the loading through Fmoc removal quantification using previously reported methods[238].

N-methylimidazolium method

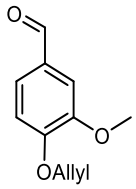
Solid support **L1** (40 mg, 0.014 mmol, loading = 0.34 mmol/g) was swollen in CH_2Cl_2 (1.5 mL) for 30 min and drained. Fmoc-Thr(OtBu)-OH (21.62 mg, 0.054 mmol) was dissolved in a mixture CH_2Cl_2/DMF 1:1 (1 mL), subsequently DIC (8 µL, 0.054 mmol), and *N*-methylimidazole (6 µL, 0.075 mmol) were added and stirred for 1 min. The activated amino acid was transferred to the polypropylene reactor containing the resin and shaken overnight at rt Solid support was washed with CH_2Cl_2 (3 x 1.5 mL), DMF (3 x 1.5 mL), and CH_2Cl_2 (3 x 1.5 mL) and dried. Coupling was monitored by determining the loading through Fmoc removal quantification using previously reported methods[238].

Coupling of the glycosylated amino acid to the solid support (traceless photolabile linker)

Solid support **L1** (40 mg, 0.014 mmol, loading = 0.34 mmol/g) was swollen in CH_2Cl_2 (1.5 mL) for 30 min and then drained. **39** (46.3 mg, 0.054 mmol) was dissolved in a mixture CH_2Cl_2/DMF 1:1 (1 mL), subsequently DIC (8.5 µL, 0.054 mmol), and *N*-methylimidazole (6 µL, 0.075 mmol) were added and stirred for 1 min. The activated BB was transferred to the reactor containing the resin and shaken overnight at rt. Solid support was washed with CH_2Cl_2 (3 x 1.5 mL), DMF (3 x 1.5 mL), and CH_2Cl_2 (3 x 1.5 mL) and dried. Coupling efficiency was determined by MALDI-TOF MS after small-scale cleavage and by Fmoc removal quantification[261].The new loading of **L1** bearing **39** was 0.24 mmol/g.

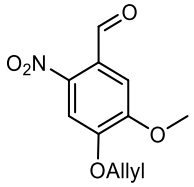
Coupling of the glycolsylated amino acid to the solid support (diazo traceless photolabile linker)

*4-(allyloxy)-3-methoxybenzaldehyde (**74**)*



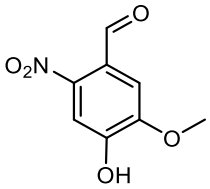
To a suspension of vanillin (2 g, 13.15 mmol) in ethanol (50 mL) was added K_2CO_3 (2 g, 14.46 mmol) and allyl bromide (1.25 mL, 14.46 mmol). The solution was refluxed at 70 °C until complete consumption of starting material was indicated by TLC. The solution was filtered to remove solid material and the filtrate was evaporated. The mixture was extracted with CH_2Cl_2 and the organic layer was dried over Na_2SO_4, filtered, concentrated under reduced pressure and purified by flash column chromatography (hexanes/EtOAc 2:1) to obtain compound **74** (2.42 g, 12.59 mmol, 96%). **R_f =** 0.43 (hexanes/EtOAc 2:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 9.83 (s, 1H), 7.41 (d, *J* = 7.6 Hz, 2H), 7.00 – 6.93 (m, 1H), 6.07 (m, *J* = 17.2, 10.6, 5.4 Hz, 1H), 5.43 (m, *J* = 17.2, 1.5 Hz, 1H), 5.33 (m, *J* = 10.5, 1.3 Hz, 1H), 4.69 (m, *J* = 5.4, 1.5 Hz, 2H), 3.93 (s, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 191.1, 153.5, 149.9, 132.3, 130.2, 126.8, 118.9, 111.9, 109.2, 69.9, 56.1. **HRMS** (m/z) Calc. for $C_{11}H_{13}O_3$ $[M+H]^+$ 193.0864 , observed 193.0867. NMR data agreed with previously reported data[298].

*4-(allyloxy)-5-methoxy-2-nitrobenzaldehyde (**75**)*



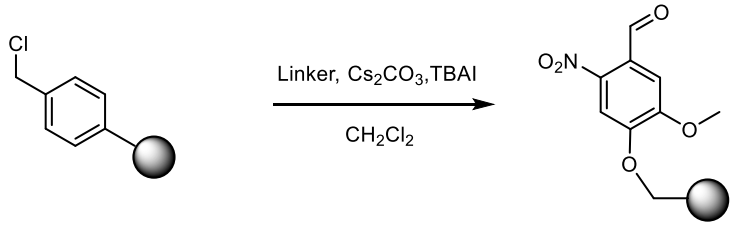
To a solution of compound **74** (2.42 g, 12.59 mmol) in trifluoroacetic acid (9 mL) was added KNO_3 (1.40 g, 13.83 mmol) portion-wise at 10 °C. The resulting black solution was heated at 60 °C for 4 h under argon. Upon complete consumption of starting material, the mixture was diluted with 50 mL of CH_2Cl_2 and extracted with 50 mL of cold water. The organic phase was separated, and the aqueous layer was extracted once with CH_2Cl_2. The organic layers were combined and neutralized with saturated $NaHCO_3$ until gas evolution ceased. The organic layer was dried over Na_2SO_4, filtered, concentrated under reduced pressure and purified by flash column chromatography (hexanes/EA 2:1) to provide compound **75** (2.61 g, 11.00 mmol, 88%)**.** $\mathbf{R_f}$ **=** 0.53 (hexanes/EtOAc 2:1). **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) 10.43 (s, 1H), 7.61 (s, 1H), 7.41 (s, 1H), 6.07 (m, *J* = 17.2, 10.7, 5.4 Hz, 1H), 5.48 (m, *J* = 17.2, 1.4 Hz, 1H), 5.40 (m, *J* = 10.5, 1.2 Hz, 1H), 4.75 (m, *J* = 5.6, 1.5 Hz, 2H), 4.01 (s, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) 187.9, 153.6, 151.4, 143.8, 131.3, 125.7, 119.9, 110.0, 108.6, 70.5, 56.9. **HRMS** (m/z) Calc. for $C_{11}H_{12}NO_5$ $[M+H]^+$ 283.0715, observed 283.0721. NMR data agreed with previously reported data[299].

*4-hydroxy-5-methoxy-2-nitrobenzaldehyde (**76**)*



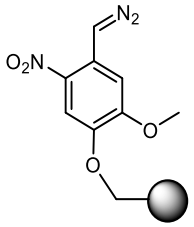
To a stirred solution of compound **75** (2.60 g, 10.96 mmol) in anhydrous MeOH (30 mL) a catalytic amount of $Pd(PPh_3)_4$ (380 mg, 0.33 mmol) was added under a nitrogen atmosphere. The solution was stirred for five minutes, and K_2CO_3 (4.54 g, 32.88 mmol) was added. The reaction mixture was stirred overnight at rt, the solvent was removed under reduced pressure and the residue was treated with aqueous citric acid 1M. The aqueous phase was extracted three times with CH_2Cl_2 (50 mL). The combined organic layer was dried over Na_2SO_4, filtered, concentrated and purified by flash column chromatography (CH_2Cl_2/EtOAc = 15:1) to give **76** (1.72 g, 8.72 mmol, 79%)**.** $\mathbf{R_f}$ **=** 0.43 (CH_2Cl_2/EtOAc 2:1). **^{1}H NMR** (400 MHz, $(CD_3)_2CO$) δ (ppm) 10.27 (s, 1H), 7.57 (s, 1H), 7.43 (s, 1H), 4.05 (s, 3H). **^{13}C NMR** (101 MHz, $(CD_3)_2CO$) δ (ppm) 188.2, 152.6, 151.7, 145.5, 125.3, 112.1, 111.1, 57.0. **HRMS** (m/z) Calc. for $C_8H_8NO_5$ $[M+Na]^+$ 198.0402, observed 198.0410. NMR data agreed with previously reported data[300].

Conjugation of photolinker to resin (**L-76**)



Merrifield resin HL (100-200 mesh, initial loading 1.0 mmol/g, 1g) was washed three times with CH_2Cl_2 swollen in CH_2Cl_2 for 30 min and drained. Subsequently, a mixture of 4-hydroxy-5-methoxy-2-nitrobenzaldehyde (0.79 g, 4 mmol), cesium carbonate (1.63 g, 5 mmol) and TBAI (0.185 g, 0.5 mmol) in CH_2Cl_2 (15 mL) was added to the resin. The suspension was shaken overnight at rt, the solution was drained and the resin was washed successively with CH_2Cl_2, MeOH, CH_2Cl_2, DMF, MeOH, CH_2Cl_2 (3 x 15 mL, each solvent).

On resin synthesis of the photolabile diazo linker **(L3)**



Photolabile aldehyde carbamate modified Merrifield resin (1 g, 1.0 mmol) was treated with 2,4,6-triisopropylbenzenesulfonyl hydrazide[287] (0.90 g, 3 mmol) in anhydrous THF (15 mL). The suspension was shaken at rt for two days, drained, and washed THF, CH_2Cl_2, MeOH, THF, and DMF (15 mL, 3 x each solvent), to give a resin bound hydrazone L-77. A solution of 1,8-diazabicyclo[5.4.0]undec-7-en (465 µL, 3 mmol) in DMF was added and the suspension was softly stirred for 7 min at rt, during this time, the resin turned red wine. Subsequently, the solid support was washed with THF, CH_2Cl_2, THF, DMF, and CH_2Cl_2 (15 mL, 3 x each), dried under vacuum and stored under the absence of light.

Optimization of the coupling to solid support

Solid support **L3** (50 mg, 0.05 mmol) was swollen in CH_2Cl_2 for 30 min and then drained. A solution of Fmoc-Phe-OH (25 mg, 0.06 mmol) in 1 mL of CH_2Cl_2:DMF (1:1) was added to the resin and shaken overnight at rt After the addition of the amino acid bubbling (evolution of gas) and the change of color of the resin (deep red to orange) was observed. Coupling was monitored by determining the loading through Fmoc removal quantification using previously reported methods[238]. A loading of 0.511 mmol/g was obtained. To confirm that the resin was compatible for the synthesis, a peptide was elongated using standard SPPS by Fmoc and characterized by MALDI-TOF MS after small-scale cleavage.

Coupling of the glycosylated amino acid to the solid support

Small-scale: Solid support L3 (50 mg, 0.02 mmol) was swollen in CH_2Cl_2 for 30 min and drained. A solution of glycosylated amino acid 39 or 40 (22.2 mg, 0.026 mmol) in 1 mL of CH_2Cl_2 was added to the resin, after some seconds bubbling (evolution of gas) was observed, the polypropylene reactor was closed and the suspension was shaken at rt overnight. Coupling efficiency was determined by MALDI-TOF MS after small-scale cleavage and by Fmoc removal quantification[261] to obtain a loading of (0.35 mmol/g), corresponding to 71% efficiency .

Large-scale: Solid support L3 (100 mg, 0.04 mmol) was swollen in CH_2Cl_2 for 30 min and drained. A solution of glycosylated amino acid 39 or 40 (46 mg, 0.054 mmol) in 2 mL of CH_2Cl_2 was added to the resin, after some seconds the evolution of gas was observed, the mixture was shaken at rt overnight. Coupling determination by MALDI-TOF MS after small-scale cleavage and through Fmoc removal quantification showed a loading of 0.33 mmol/g, corresponding to 67% efficiency.

Procedure for automated solid-phase synthesis

Preparation of stock solutions

Building block solution: A 60 mM solution of each thioglycoside building block was prepared in anhydrous CH_2Cl_2.
Acidic wash solution: A 41.3 mM solution of TMSOTf in anhydrous CH_2Cl_2 was prepared.
Activator solution: A solution 0.15 M of recrystallized NIS and 0.015 M of triflic acid in a mixture containing anhydrous CH_2Cl_2 and anhydrous dioxane (2:1 v/v) was prepared. The solution was kept under ice-bath for the duration of the automation run.
Pre-capping solution: A 10% (v/v) solution of piperidine in DMF was prepared.
Capping solution: A solution containing (v/v) 10% acetic anhydride, 2% methanesulfonic acid in CH_2Cl_2 was prepared.
Fmoc removal solution: A solution containing 20% piperidine in DMF was prepared.
Lev removal solution: A 0.15 M solution of hydrazine acetate was prepared in a mixture containing pyridine/AcOH/H_2O (v/v 32:8:2) and sonicated for 10 min.

Modules

Resin swelling (20 min)

All automated synthesis were performed on 0.013 mmol scale. Solid support with different linkers, 0.35 mmol/g traceless photolinker (**L1**), 0.30 mmol/g aminopentyl linker (**L2**) or 0.33 mmol/g

photolabile diazo linker (**L3**), bearing the glycosylated amino acid was placed in the synthesizer and swollen in CH_2Cl_2 for 20 min at rt prior to synthesis. During this time, all reagents lines were washed and primed. After swelling, the resin was washed with DMF, THF, and CH_2Cl_2 (2 mL, 3 x each solvent, 25 s).

Acidic wash (20 min)

The resin was swollen in 2 mL and the temperature of the reaction vessel was adjusted to -20 °C. Upon reaching the temperature, the acidic wash solution (1 mL) was added dropwise to the reaction vessel and mixed by Argon bubbling for 3 min. The acidic solution was drained, the resin was washed with 2 mL CH_2Cl_2 for 25 s and drained.

Action	Cycles	Solution	Amount	T (°C)	Incubation time
Cooling	-	-	-	-20	15 min*
Deliver	1	CH_2Cl_2	2 mL	-20	-
Deliver	1	Acidic wash	1 mL	-20	3 min
Wash	1	CH_2Cl_2	2 mL	-20	25 s

*time required to reach the desired temperature

Glycosylation (35 – 55 min)

A solution containing the appropriate building block (1 mL) was delivered to the reaction vessel. After reaching the set temperature, the activator solution (1 mL) was added dropwise and the reaction mixture is kept at the same temperature for 25 minutes under argon bubbling. Once the time is completed, the solution was drained and the resin was washed with CH_2Cl_2, dioxane, CH_2Cl_2/dioxane (1:2, 3 mL, 20 s), CH_2Cl_2 (2 x 2 mL, 25 s each). The temperature of the reaction vessel was increased to 25 °C for the next cycle.

Action	Cycles	Solution	Amount	T (°C)	Incubation time
Cooling	-	-	-	-20	-
Deliver	1	Building block	1 mL	-20	-
Deliver	1	Activator	1 mL	-20	-
Reaction time	1			-20 to 0	5 min 20 min
Wash	1	CH_2Cl_2	2 mL	0	5 s
Wash	1	CH_2Cl_2/Dioxane	2 mL	0	20 s
Heating	-	-	-	25	-
Wash	2	CH_2Cl_2	2 mL	> 0	25 s

Capping (30 min)

The resin was washed with DMF (2 x 2 mL, 25 s each) and the pre-capping solution (2 mL) was delivered to the reaction vessel. After 1 min, the solution was drained and the resin was washed with CH_2Cl_2 (3 mL, three times, 25 s each) and drained. Then, capping solution (4 mL) was delivered and kept at 25 °C for 20 min, then the solution was drained and the resin washed with CH_2Cl_2 (3 mL, three times, 25 s each).

Action	Cycles	Solution	Amount	T (°C)	Incubation time
Heating	-	-	-	25	5 min*
Wash	2	DMF	2 mL	25	25 s
Deliver	1	Pre-capping	2 mL	25	1 min
Wash	3	CH_2Cl_2	2 mL	25	25 s
Deliver	1	Capping solution	4 mL	25	20 min
Wash	3	CH_2Cl_2	2 mL	25	25 s

Fmoc removal (9 min)

The resin was washed with DMF (2 mL, three times, 25 s each) and the Fmoc removal solution (2 mL) was delivered to the reaction vessel and kept at 25 °C under argon bubbling for 5 min. The reaction vessel was drained and the resin was washed with DMF (3 x 3 mL, 25 s each) and CH_2Cl_2 (5 x 2 mL, 25 s each). The temperature of the reaction vessel was decreased to 20 °C for the acidic wash.

Action	Cycles	Solution	Amount	T (°C)	incubation time
Wash	3	DMF	2 mL	25	25 s
Deliver	1	Fmoc removal	2 mL	25	5 min
Wash	1	DMF	2 mL	25	25 s
Cooling	-	-	-	-20	-
Wash	3	DMF	2 mL	< 25	25 s
Wash	5	CH_2Cl_2	2mL	< 25	25 s

Lev removal (65 min)

The resin was washed with CH_2Cl_2 (3 x 2 mL, 25 s each), after draining the reaction vessel, CH_2Cl_2 (1.3 mL) was delivered to the reaction vessel and the temperature was adjusted to 25 °C. Lev removal solution (2 mL) was added to the reaction vessel and kept under Ar bubbling for 30 min. After repeating the procedure twice, the solvent was drained and the resin washed with DMF (3 mL, three times, 25 s) and CH_2Cl_2 (2 mL, five times, 25 s each).

Action	Cycles	Solution	Amount	T (°C)	incubation time
Wash	3	DMF	2 mL	25	25 sec
Deliver	1	Lev removal	2 mL	25	30 min
Wash	1	DMF	2 mL	25	25 sec
Cooling	-	-	-	-20	-
Wash	3	DMF	2 mL	< 25	25 sec
Wash	5	CH_2Cl_2	2mL	< 25	25 sec

Post synthesizer manipulations

On-resin azide reduction and acetylation

Staudinger reduction: The solid support (42 mg, 0.013 mmol) was swollen in THF (3 mL) for 15 min and the solvent was drained. A solution of tributylphosphine 1 M in THF (0.5 mL, 0.50 mmol) in 5% of water in THF (1.5 mL) was added to the propylene reactor and shaken overnight. The mixture was drained, the resin washed with THF (3 x 3 mL), CH_2Cl_2 (3 x 3 mL), THF (3 x 3 mL), CH_2Cl_2 (3 x 3 mL) and a small-scale cleavage was performed to monitor the reaction by MALDI-TOF MS.

Thioacetic acid (one-pot azide to amide conversion): The resin (42 mg, 0.013 mmol) was swollen in THF (3 mL) for 15 min and the solvent was drained. Thioacetic acid (1 mL, 14 mmol) was added to the reactor and the suspension was shaken at rt for two days. The solvent was drained and the resin washed with THF (3 x 3 mL), CH_2Cl_2 (3 x 3 mL), THF (3 x 3 mL), CH_2Cl_2 (3 x 3 mL). A small-scale cleavage was performed to monitor the reaction by MALDI-TOF MS.

On resin hydrolysis

Method A-Methanolysis: The glycan bound resin was transferred to a propylene reactor and swollen in THF for 30 min and the solvent was drained. Subsequently, a 0.5 M solution of NaOMe in MeOH (0.4 mL, 0.045 M) in THF/MeOH (4:1, 4 mL) was added and the resin was gently shaken for 2 h at rt After collection of the solution, the resin was washed with MeOH, THF and CH_2Cl_2 (4 mL, three times each) and the washes were also collected. The solvent was removed under reduce pressure and the residue was analyzed by MALDI-TOF MS. Possible remaining compounds on resin were analyzed by small-scale photocleavage and MALDI-TOF MS.

Method B-LiOH: The resin was transferred to a propylene reactor and swollen in THF (3 mL) for 30 min and the solvent was drained. Then, a 1 M aqueous solution of LiOH (0.35 mL, 0.08 M) in THF/MeOH (4:1, 4 mL) was added and the suspension was gently shaken for 2 h at rt After collection of the solution, the resin was washed with MeOH, THF and CH_2Cl_2 (4 mL, three times each) and the washes were also collected. The solvent was removed under reduce pressure and the residue was analyzed by MALDI-TOF MS. Remaining resin was submitted to small-scale photocleavage and MALDI-TOF MS.

Method C-Hydrazine monohydrate: The resin was transferred to a propylene reactor and swollen in THF (3 mL) for 30 min and the solvent was drained. Subsequently, hydrazine monohydrate (300 µL) in THF:MeOH (1:1, 3 mL) was added and the suspension was gently shaken for 2 h at

rt. After removal of the reaction solution by filtration, the resin was washed with MeOH, THF, and CH_2Cl_2 (3 x 4 mL each solvent). Reaction was monitored by MALDI-TOF MS.

Cleavage from solid support

Small-scale cleavage: To a small amount of resin (~0.5 mg) DMF (0.5 mL) was added and irradiated with a UV lamp (6 watt, 356 nm) for 1 h. DMF was removed using nitrogen stream and acetonitrile was then added to the resin and the resulting solution was analyzed by MALDI-TOF MS.

Full cleavage: Resin was suspended in DMF (4 mL) and gently stirred while irradiated overnight using a LED lamp (43 watt, 370 nm, distance 3 cm) (Kessil PR160L). The solvent was drained and the resin washed with DMF (3 x 4 mL), after collecting the washes the solvent was removed under reduced pressure.

Hydrolysis in solution

Method A: Hydrazine monohydrate 10%: After photocleavage of the glycosylated amino acid, it was dissolved in a solution of hydrazine monohydrate (300 µL) in MeOH (3 mL) and was stirred at rt. To follow the progress of the reaction, a small aliquot of the reaction was monitored by MALDI-TOF MS. Once the reaction finished, it was cooled to 0 °C and acetone was slowly added and stirred for 30 min. Solvent was then removed under reduced pressure.

Method B: Hydrazine monohydrate 20%: After photocleavage of the glycosylated amino acid, it was dissolved in a solution of hydrazine monohydrate (100 µL) in MeOH (0.5 mL, final concentration 18 mg/mL) and was stirred at rt. To follow the progress of the reaction, a small aliquot of the reaction was monitored by MALDI-TOF MS. Once the reaction finished, it was cooled to 0 °C and acetone was slowly added and stirred for 30 min. Solvent was then removed under reduced pressure.

Method C: Hydrazine monohydrate 20%: After photocleavage of the glycosylated amino acid, it was dissolved in a solution of hydrazine monohydrate (600 µL) in MeOH (3 mL, final concentration 1 mg/mL) and was stirred at rt. To follow the progress of the reaction, a small aliquot of the reaction was monitored by MALDI-TOF MS. Once the reaction finished, it was cooled to 0 °C and acetone was slowly added and stirred for 30 min. Solvent was then removed under reduced pressure.

Hydrogenolysis

The crude compound was dissolved in 3.25 mL of *t*-BuOH:EtOAc:THF:H_2O (1:1:1:0.25). To the vial, 10% Pd/C catalyst (four times the weight of the starting material) was added and the reaction sealed with a rubber septum and stirred while connected to a H_2 balloon. The reaction progress was monitored by MALDI-TOF MS. Upon completion, the reaction was filtered and washed with *t*-BuOH and H_2O. The filtrate was then concentrated under reduced pressure.

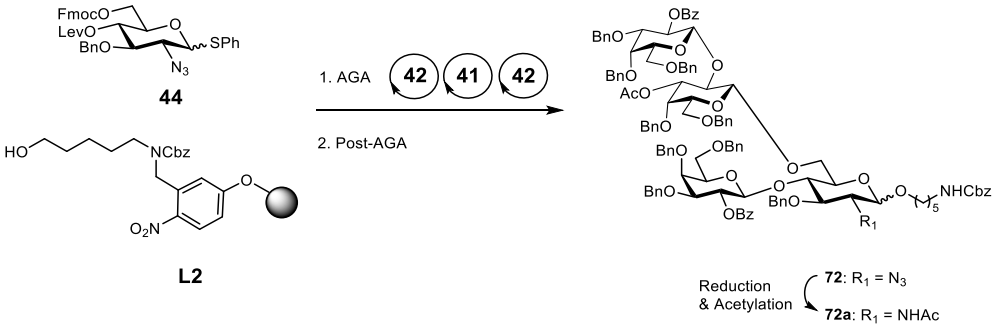
Purification

After global deprotection, the final compound was analyzed and purified using analytical RP-HPLC Agilent 1200 Series using method A and purified using method B.

Method A: RP-HPLC-MS: column [A], 0 - 70 % MeCN/H_2O + 0.1 % HCOOH over 30 min, 0.7 mL/min. ELSD detector: 45 °C.
Method B: RP-HPLC: column [B], 0 - 30 % MeCN/H_2O + 0.1 % HCOOH over 30 min, 3 mL/min.

3.6.3. Assembly of the oligosaccharides and glycosyl amino acids of *T. cruzi* MLMs

*N-Benzyloxycarbonyl-5-aminopentyl 2-O-benzoyl-3,4,6-tri-O-benzyl-β-D-galactopyranosyl-(1→2)-3-O-acetyl-4,6-di-O-benzyl-β-D-galactopyranosyl-(1→6)-[2-O-benzoyl-3,4,6-tri-O-benzyl-β-D-galactopyranosyl-(1→4)]-2-N-acetyl-2-deoxy-α/β-D-glucopyranoside (***72a)**



Step	Modules	Parameters
AGA	Resin swelling Acidic wash Initiation Glycosylation Capping Lev removal Acidic wash Glycosylation Capping Fmoc Removal Acidic wash Glycosylation Capping Lev Removal Acidic wash Glycosylation Capping Fmoc Removal Capping	BB **44** (6 equiv, 1 cycle, -20 °C for 5 min, 0 °C for 40 min) BB **42** (6 equiv, 1 cycles, -20 °C for 5 min, 0 °C for 20 min) BB **41** (6 equiv, 1 cycles, -20 °C for 5 min, 0 °C for 20 min) BB **42** (6 equiv, 1 cycles, -20 °C for 5 min, 0 °C for 20 min)
Post-AGA	On-resin Staudinger Reduction and acetylation Full cleavage	

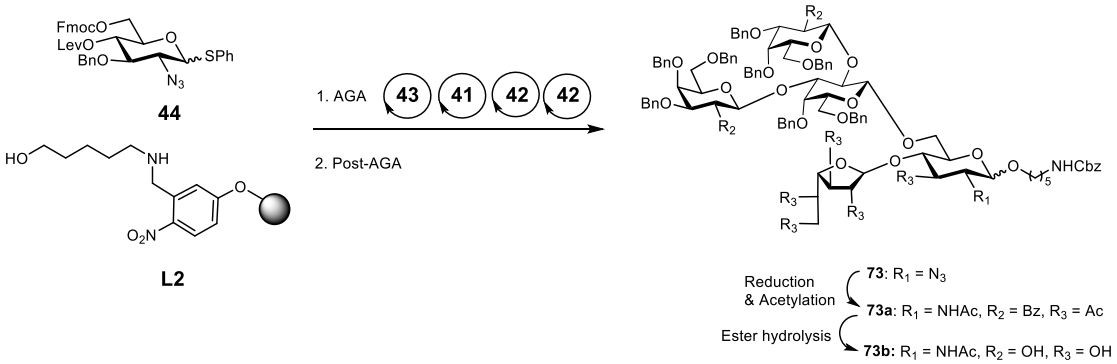
Resin was split and two different methodologies for azide reduction and acetylation were screened (On-resin Staudinger and one-pot azide to amide conversion).

Mass assignment tetrasaccharide 72a

Step	Type of cleavage	Expected mass	Observed mass
AGA	Small-scale cleavage	1970.839	$[M+Na]^+$ 1993.811 [**72**]
On-resin Staudinger	Small-scale cleavage	1986.859	$[M+H]^+$ 2146.300 Iminophosphorane
Reduction and acetylation	Small-scale cleavage	1986.859	$[M+Na]^+$ 2009.849 [**72a**]

[SM] = Product after AGA process, [P] = Expected product after the corresponding step

*N-Benzyloxycarbonyl-5-aminopentyl 3,4,6-tri-O-benzyl-β-D-galactopyranosyl-(1→2)-[3,4,6-tri-O-benzyl-β-D-galactopyranosyl-(1→3)]-(4,6-di-O-benzyl-β-D-galactopyranosyl-(1→6)-4-O-(β-D-galactofuranosyl)-2-N-acetyl-2-deoxy-α/β-D-glucopyranoside (**73b**)*



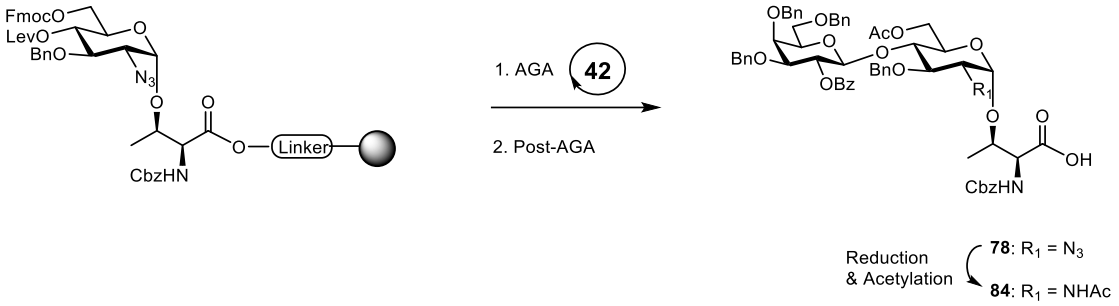
Step	Modules	Parameters
AGA	Resin swelling Acid wash Glycosylation Capping Lev Removal	BB **44**(4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
	Acid wash Glycosylation Capping Fmoc Removal	BB **43** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
	Acidic wash Glycosylation Capping Lev Removal	BB **41** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
	Acid wash Glycosylation Capping Fmoc Removal	BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
	Acid wash Glycosylation Capping	BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
Post-AGA	Reduction and acetylation Large-scale cleavage Methanolysis	

Mass assignment pentasaccharide 73b

Step	Type of cleavage	Expected mass	Observed mass
AGA	Small-scale cleavage	2258.924	$[M+Na]^+$ 2281.417 [**73**] [M+Na]+ 1787.417 [**73** - **42**]
Reduction and acetylation	Small-scale cleavage	2274.944	$[M+Na]^+$ 2297.509 [**73a**] [M+Na]+ 1803.730 [**73a** - **42**]
Methanolysis	Small-scale cleavage	1898.849	$[M+Na]^+$ 1921.606 [**73b**] [M+Na]+ 1489.610 [**73b** - **42**]

[SM] = Product after AGA process, [P] = Expected product after the corresponding step

*Nᵅ-(carbobenzyloxy)-O-(2-N-acetyl-3-O-benzyl-4-O-(3,4,6-tri-O-tribenzyl-2-O-benzoyl-β-D-galactopyranosyl)-6-O-acetyl-2-deoxy-α-D-glucopyranosyl)-L-threonine (**84**)*



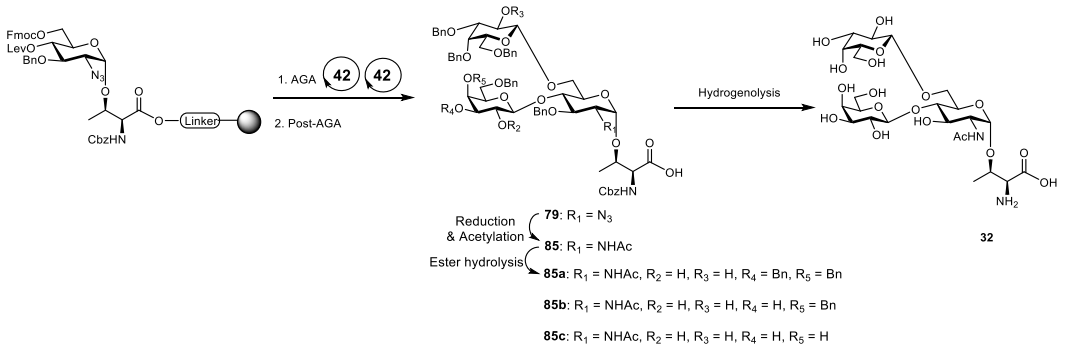
Step	Modules	Parameters
AGA	Resin swelling Lev Removal Acidic wash Glycosylation Capping Fmoc Removal Capping	BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20min)
Post-AGA	Reduction and acetylation	

Mass assignment disaccharide 84

Step	Type of cleavage	Expected mass	Observed mass
AGA	Small-scale cleavage	1108.432	$[M+Na]^+$ 1131.682 [**78**]
Reduction and acetylation	Small-scale cleavage	1124.452	$[M+Na]^+$ 1147.543 [**84**]

[SM] = Product after AGA process, [P] = Expected product after the corresponding step

*O-(2-N-acetyl-4-O-(β-D-galactopyranosyl)-6-O-(β-D-galactopyranosyl)-2-deoxy-α-D-glucopyranosyl)-L-threonine (***32***)*



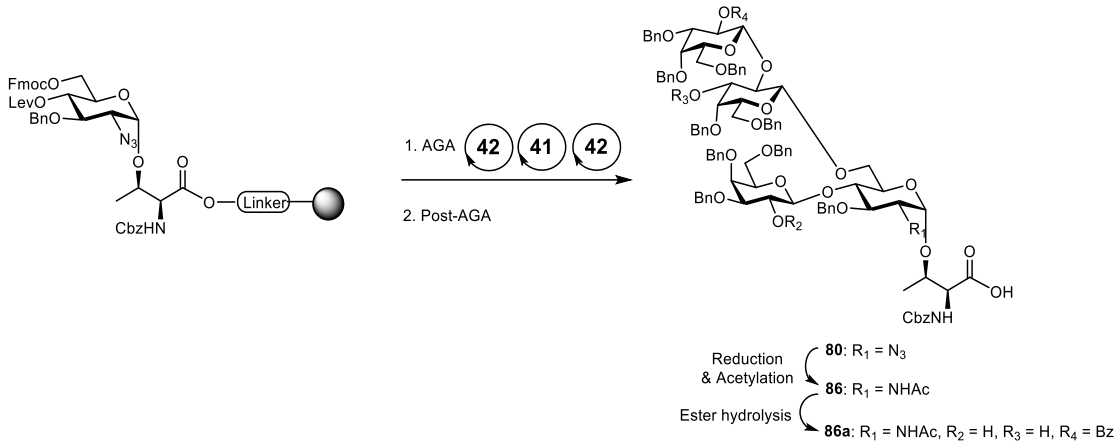
Step	Modules	Parameters
AGA	Resin swelling Lev Removal Acidic wash Glycosylation Capping Fmoc Removal Acidic wash Glycosylation Capping	BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min) BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
Post-AGA	Reduction and acetylation Photocleavage Ester hydrolysis on solution Hydrogenolysis	

Mass assignment trisaccharide 32

Step	Type of cleavage	Expected mass	Observed mass
AGA	Small-scale cleavage	1602.641	$[M+Na]^+$ 1626.150 [**79**]
Reduction and acetylation	Small-scale cleavage	1618.661	$[M+Na]^+$ 1642.793 [**85**]
Photocleavage	Large-scale cleavage	1618.661	$[M+Na]^+$ 1642.793 [**85**]
Ester hydrolysis in solution Method B (5 days)	n/a	1410.601	$[M+Na]^+$ 1433.689 [**85a**] [M+Na]+ 1001.740 [**85a** - **42**] $[M+Na]^+$ 1343.637 [**85b**] [M+Na]+ 911.860 [**85b** - **42**] $[M+Na]^+$ 1254.578 [**85c**]
Hydrogenolysis (2 days)	n/a	646,243	$[M+Na]^+$ 669,492 [**32**]

[SM] = Product after AGA process, [P] = Expected product after the corresponding step

*N^{α}-(carbobenzyloxy)-O-(2-N-acetyl-3-O-benzyl-4-(3,4,6-tri-O-benzyl-β-D-galactopyranosyl-(1→2)-4,6-tri-O-benzyl-β-D-galactopyranosyl)-2-deoxy-α-D-glucopyranosyl)-L-threonine (***86a***)*



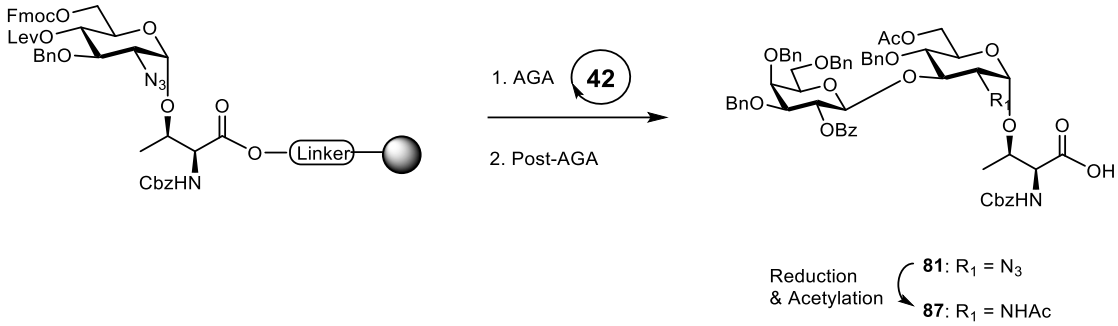
Step	Modules	Parameters
AGA	Resin swelling Lev Removal Acidic wash Glycosylation Capping Fmoc Removal	BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
	Acidic wash Glycosylation Capping Lev Removal	BB **41** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
	Acidic wash Glycosylation Capping Fmoc Removal Capping	BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
Post-AGA	Reduction and acetylation Ester hydrolysis on resin LiOH	

Mass assignment tetrasaccharide 86a

Step	Type of cleavage	Expected mass	Observed mass
AGA	Small-scale cleavage	1988.798	$[M+Na]^+$ 2009.795 [**80**]
Reduction and acetylation	Small-scale cleavage	2002.818	$[M+Na]^+$ 2025.841 [**86**]
On-resin hydrolysis Method B	Ester hydrolysis	1736.735	$[M+Na]^+$ 1879.823 [**86a**] $[M+Na]^+$ 1643.212 [β-elimination side product]

[SM] = Product after AGA process, [P] = Expected product after the corresponding step

*N^{α}-(carbobenzyloxy)-O-(2-N-acetyl-3-O-(3,4,6-tri-O-tribenzyl-2-O-benzoyl-β-D-galactopyranosyl)-4-O-benzyl-6-O-acetyl-2-deoxy-α-D-glucopyranosyl)-L-threonine (***87**)*



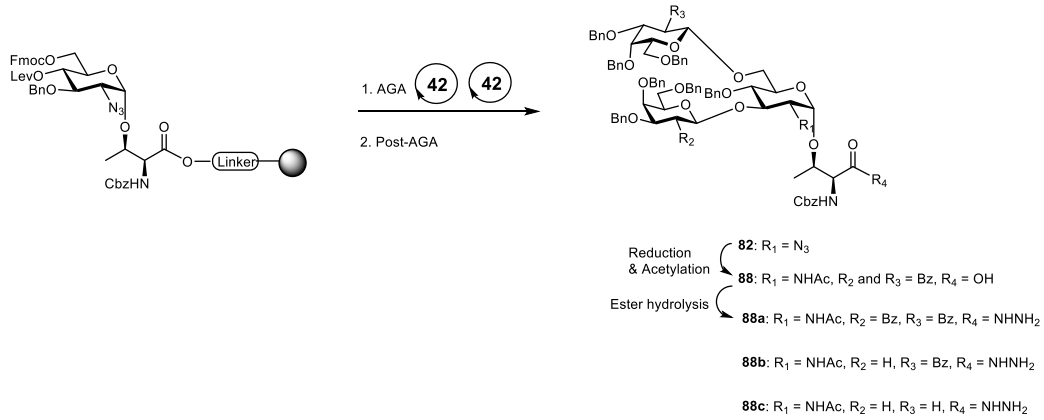
Step	Modules	Parameters
AGA	Resin swelling Fmoc Removal Acid wash Glycosylation Capping Lev Removal Capping	BB **40** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
Post-AGA	Reduction and acetylation	

Mass assignment disaccharide 75

Step	Type of cleavage	Expected mass	Observed mass
AGA	Small-scale cleavage	1108.432	$[M+Na]^+$ 1132.695 [**81**]
Reduction and acetylation	Small-scale cleavage	1124.452	$[M+Na]^+$ 1147.543 [**87**]

[SM] = Product after AGA process, [P] = Expected product after the corresponding step

*N^{α}-(carbobenzyloxy)-O-(2-N-acetyl-3-O-(3,4,6-tri-O-benzyl-β-D-galactopyranosyl)-4-O-benzyl-6-O-(3,4,6-tri-O-benzyl-β-D-galactopyranosyl)-2-deoxy-α-D-glucopyranosyl)-L-threonine hydrazide (***88c***)*



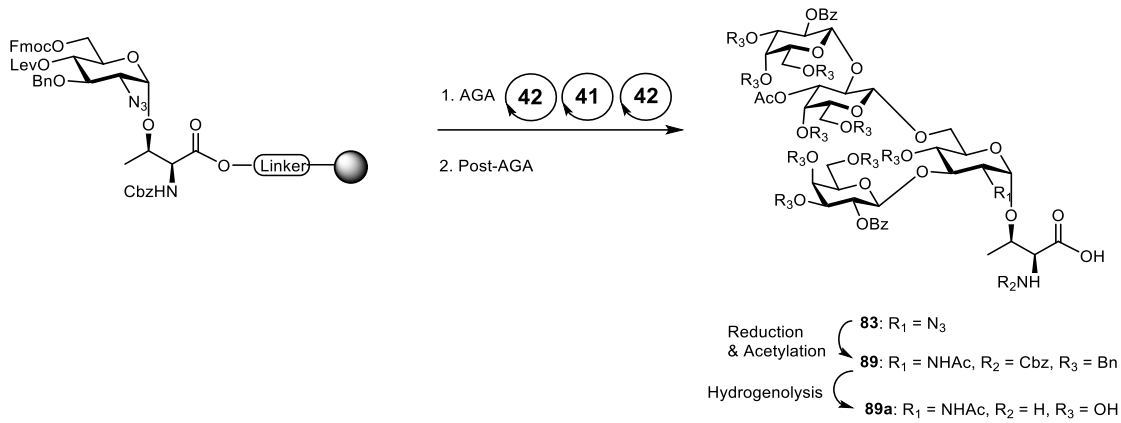
Step	Modules	Parameters
AGA	Resin swelling Fmoc Removal Acid wash Glycosylation Capping Lev Removal Acid wash Glycosylation Capping	BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min) BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
Post-AGA	Reduction and acetylation Ester hydrolysis on Resin Ester hydrolysis on solution	

Mass assignment trisaccharide 88c

Step	Type of cleavage	Expected mass	Observed mass
AGA	Small-scale cleavage	1602.641	$[M+Na]^+$ 1625.407 [**82**]
Reduction and acetylation	Small-scale cleavage	1618.661	$[M+Na]^+$ 1643.303 [**88**]
Photocleavage	Large-scale cleavage	1618.661	$[M+Na]^+$ 1643.303 [**88**]
Ester hydrolysis on resin Method C	Ester hydrolysis	1410.609	$[M+Na]^+$ 1726.792 [**88a** + THF adduct] $[M+Na]^+$ 1656.722 [**88a**] [M+Na]+ 1621.743 [**88b** + THF adduct] $[M+Na]^+$ 1552.720 [**88b**] [M+Na]+ 1189.500 [**88c** + THF adduct - **42**] $[M+Na]^+$ 1448.688 [**88c**]
Ester hydrolysis in solution Method B (5 days)	n/a	1410.609	$[M+Na]^+$ 1212.549 [β-elimination side product] $[M+Na]^+$ 1015.439 [**88c** - **42**] $[M+Na]^+$ 780.280 [[β-elimination side product - **42**]

[SM] = Product after AGA process, [P] = Expected product after the corresponding step

*O-(2-N-acetyl-3-O-(2-O-benzoyl-β-D-galactopyranosyl)-6-(2-O-benzoyl-β-D-galactopyranosyl-(1→2)-3-O-acetyl-β-D-galactopyranosyl)-2-deoxy-α-D-glucopyranosyl)-L-threonine (***89a***)*



Step	Modules	Parameters
AGA	Resin swelling Fmoc Removal Acid wash Glycosylation Capping Lev Removal Acid wash Glycosylation Capping Lev Removal Acidic wash Glycosylation Capping Fmoc Removal Capping	BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min) BB **41** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min) BB **42** (4 equiv, 2 cycles, -20 °C for 5 min, 0 °C for 20 min)
Post-AGA	Reduction and acetylation Large-scale photocleavage Hydrogenolysis	

Mass assignment tetrasaccharide 89a

Step	Type of cleavage	Expected mass	Observed mass
AGA	Small-scale cleavage	1988.798	$[M+Na]^+$ 2009.698 [**83**] $[M+Na]^+$ 1515.692 [**83** - **42**]
Reduction and acetylation	Small-scale cleavage	2002.818	$[M+Na]^+$ 2025.872 [**89**] $[M+Na]^+$ 1531.780 [**89** - **42**]
Photocleavage	Large-scale cleavage	2002.818	$[M+Na]^+$ 2025.872 [**89**] $[M+Na]^+$ 1531.780 [**89** - **42**]
Hydrogenolysis	n/a	1058.319	$[M+Na]^+$ 1081.518 [**89a**] $[M+Na]^+$ 857.369 [**89a** - **42**]

[SM] = Product after AGA process, [P] = Expected product after the corresponding step

References

1. Pinzón Martín, S.; Seeberger, P. H.; Varón Silva, D., *Front. Chem.* **2019,** *7* (710).

2. Hanna, C. C.; Kriegesmann, J.; Dowman, L. J.; Becker, C. F. W.; Payne, R. J., *Angew. Chem. Int. Ed.* **2022,** *61* (15), e202111266.

3. Kent, S. B. H., *Chem. Soc. Rev.* **2009,** *38* (2), 338-351.

4. Kulkarni, S. S.; Sayers, J.; Premdjee, B.; Payne, R. J., *Nat. Rev. Chem.* **2018,** *2* (4), 0122.

5. Müller, M. M., *Biochemistry.* **2018,** *57* (2), 177-185.

6. Walsh, C., *Posttranslational modification of proteins: expanding nature's inventory*. Roberts and Company Publishers: 2006.

7. Wang, Y.-C.; Peterson, S. E.; Loring, J. F., *Cell Res.* **2014,** *24* (2), 143-160.

8. Deribe, Y. L.; Pawson, T.; Dikic, I., *Nat. Struct. Mol. Bio.* **2010,** *17* (6), 666-672.

9. Ho, W.-L.; Hsu, W.-M.; Huang, M.-C.; Kadomatsu, K.; Nakagawara, A., *J. Hematol. Oncol.* **2016,** *9* (1), 100.

10. Spiro, R. G., *Glycobiology.* **2002,** *12* (4), 43R-56R.

11. Abu-Qarn, M.; Eichler, J.; Sharon, N., *Curr. Opin. Struct. Biol.* **2008,** *18* (5), 544-550.

12. Eichler, J., *Nat. Rev. Microbiol.* **2013,** *11* (3), 151-156.

13. Buscaglia, C. A.; Campo, V. A.; Frasch, A. C.; Di Noia, J. M., *Nat. Rev. Microbiol.* **2006,** *4* (3), 229-36.

14. Wanyiri, J.; Ward, H., *Future Microbiol.* **2006,** *1* (2), 201-208.

15. Bhalchandra, S.; Ludington, J.; Coppens, I.; Ward, H. D., *Infect. Immun.* **2013,** *81* (9), 3356.

16. Cancela, M.; Santos, G. B.; Carmona, C.; Ferreira, H. B.; Tort, J. F.; Zaha, A., *Parasitology.* **2015,** *142* (14), 1673-1681.

17. Rudd Pauline, M.; Elliott, T.; Cresswell, P.; Wilson Ian, A.; Dwek Raymond, A., *Science.* **2001,** *291* (5512), 2370-2376.

18. Marth, J. D.; Grewal, P. K., *Nat. Rev. Immunol.* **2008,** *8* (11), 874-887.

19. Rudd, P. M.; Wormald, M. R.; Stanfield, R. L.; Huang, M.; Mattsson, N.; Speir, J. A.; DiGennaro, J. A., et al., *J. Mol. Biol.* **1999,** *293* (2), 351-366.

20. Ohtsubo, K.; Marth, J. D., *Cell.* **2006,** *126* (5), 855-867.

21. Berger, R. P.; Dookwah, M.; Steet, R.; Dalton, S., *BioEssays.* **2016,** *38* (12), 1255-1265.

22. Nath, S.; Mukherjee, P., *T. Mol. Med.* **2014,** *20* (6), 332-342.

23. Kufe, D. W., *Nat Rev Cancer* **2009,** *9* (12), 874-85.

24. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., *Nat. Rev. Mol. Cell Biol.* **2012,** *13* (7), 448-462.

25. Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J., *Nat. Rev. Nephrol.* **2019,** *15* (6), 346-366.

26. Varki, A., *Glycobiology.* **1993,** *3* (2), 97-130.

27. Bertozzi Carolyn, R.; Kiessling; Laura, L., *Science.* **2001,** *291* (5512), 2357-2364.

28. Lee, H. S.; Qi, Y.; Im, W., *Sci. Rep.* **2015,** *5* (1), 8926.

29. Apweiler, R.; Hermjakob, H.; Sharon, N., *Biochim. Biophys. Acta. Gen. Subj.* **1999,** *1473* (1), 4-8.

30. Brown, D.; Waneck, G. L., *J. Am. Soc. Nephrol.* **1992,** *3* (4), 895.

31. Vogt Marian, S.; Schmitz Gesa, F.; Varón Silva, D.; Mösch, H.-U.; Essen, L.-O., *Proc. Natl. Acad. Sci.* **2020,** *117* (36), 22061-22067.

32. Imjeti, N. S.; Lebreton, S.; Paladino, S.; de la Fuente, E.; Gonzalez, A.; Zurzolo, C., *Mol. Biol. Cell.* **2011,** *22* (23), 4621-4634.

33. Kinoshita, T., *Open. Biol. 10* (3), 190290.

34. Dwek, R. A., *Chem. Rev.* **1996,** *96* (2), 683-720.

35. Gottschalk, A., *Nature.* **1969,** *222* (5192), 452-454.

36. Davis, B. G., *Chem. Rev.* **2002,** *102* (2), 579-602.

37. Schwarz, F.; Aebi, M., *Curr. Opin. Struct. Biol.* **2011,** *21* (5), 576-582.

38. Roth, J.; Zuber, C., *Histochem. Cell Biol.* **2017,** *147* (2), 269-284.

39. Imperiali, B.; O'Connor, S. E., *Curr. Opin. Chem. Biol.* **1999,** *3* (6), 643-649.

40. Helenius, A.; Aebi, M., *Science.* **2001,** *291* (5512), 2364-2369.

41. Marquardt, T.; Denecke, J., *Eur. J. Pediatr.* **2003,** *162* (6), 359-379.

42. Jaeken, J.; Matthijs, G., *Annu. Rev. Genomics. Hum. Genet.* **2001,** *2* (1), 129-151.

43. Fairbanks, A. J., *Chem. Soc. Rev.* **2017,** *46* (16), 5128-5146.

44. Ajit Varki; Richard D Cummings; Jeffrey D Esko; Pamela Stanley; Gerald W Hart; Markus Aebi; Alan G Darvill, et al., *Essentials of Glycobiology, 3rd edition*. Cold Spring Harbor Laboratory Press: 2017.

45. Corfield, A., *Histochem. Cell Biol.* **2017,** *147* (2), 119-147.

46. Steentoft, C.; Vakhrushev, S. Y.; Joshi, H. J.; Kong, Y.; Vester-Christensen, M. B.; Schjoldager, K. T. B. G.; Lavrsen, K., et al., *EMBO J.* **2013,** *32* (10), 1478-1488.

47. Hounsell, E. F.; Davies, M. J.; Renouf, D. V., *Glycoconj. J.* **1996,** *13* (1), 19-26.

48. Brandley, B. K.; Schnaar, R. L., *J. Leukoc. Biol,* **1986,** *40* (1), 97-111.

49. Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C. A., *Cancer Cell* **2019,** *36* (1), 6-16.

50. Pinho, S. S.; Reis, C. A., *Nat. Rev. Cancer.* **2015,** *15* (9), 540-555.

51. Wilkinson, H.; Saldova, R., *J. Proteome Res.* **2020,** *19* (10), 3890-3905.

52. Tuccillo, F. M.; de Laurentiis, A.; Palmieri, C.; Fiume, G.; Bonelli, P.; Borrelli, A.; Tassone, P., et al., *Biomed. Res. Int.* **2014,** *2014*, 742831.

53. Corfield, A. P., *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015,** *1850* (1), 236-252.

54. Dalziel, M.; Crispin, M.; Scanlan Christopher, N.; Zitzmann, N.; Dwek Raymond, A., *Science.* **2014,** *343* (6166), 1235681.

55. Weikert, S.; Papac, D.; Briggs, J.; Cowfer, D.; Tom, S.; Gawlitzek, M.; Lofgren, J., et al., *Nat. Biotech.* **1999,** *17* (11), 1116-1121.

56. Yang, Z.; Wang, S.; Halim, A.; Schulz, M. A.; Frodin, M.; Rahman, S. H.; Vester-Christensen, M. B., et al., *Nat. Biotech.* **2015,** *33* (8), 842-844.

57. Fernández-Núñez, E. G.; de Rezende, A. G.; Puglia, A. L. P.; Leme, J.; Boldorini, V. L. L.; Caricati, C. P.; Tonso, A., *Biotechnol. Lett.* **2015,** *37* (6), 1153-1163.

58. De Bruyne, E.; Andersen, T. L.; De Raeve, H.; Van Valckenborgh, E.; Caers, J.; Van Camp, B.; Delaissé, J. M.; Van Riet, I.; Vanderkerken, K., *Leukemia* **2006,** *20* (10), 1870-1879.

59. Lalonde, M.-E.; Durocher, Y., *J. Biotechnol.* **2017,** *251*, 128-140.

60. Dumont, J.; Euwart, D.; Mei, B.; Estes, S.; Kshirsagar, R., *Crit. Rev. Biotechnol.* **2016,** *36* (6), 1110-1122.

61. Amano, K.; Chiba, Y.; Kasahara, Y.; Kato, Y.; Kaneko Mika, K.; Kuno, A.; Ito, H., et al., *Proc. Natl. Acad. Sci.* **2008,** *105* (9), 3232-3237.

62. Margolin, E.; Allen, J. D.; Verbeek, M.; van Diepen, M.; Ximba, P.; Chapman, R.; Meyers, A., et al., *Front. Plant. Sci.* **2021,** *12*.

63. Harrison, R. L.; Jarvis, D. L., *Protein N-Glycosylation in the Baculovirus–Insect Cell Expression System and Engineering of Insect Cells to Produce "Mammalianized" Recombinant Glycoproteins*. Academic Press: 2006; Vol. 68, p 159-191.

64. Ghaderi, D.; Zhang, M.; Hurtado-Ziola, N.; Varki, A., *Biotechnol. Genet. Eng. Rev.* **2012,** *28* (1), 147-176.

65. Rosenlöcher, J.; Sandig, G.; Kannicht, C.; Blanchard, V.; Reinke, S. O.; Hinderlich, S., *J. Proteomics.* **2016,** *134*, 85-92.

66. Li, C.; Wang, L.-X., *Chem. Rev.* **2018,** *118* (17), 8359-8413.

67. Liang, D.-M.; Liu, J.-H.; Wu, H.; Wang, B.-B.; Zhu, H.-J.; Qiao, J.-J., *Chem. Soc. Rev.* **2015,** *44* (22), 8350-8374.

68. Bill, R. M.; Flitsch, S. L., *Chem. Biol.* **1996,** *3* (3), 145-149.

69. Wang, L.-X.; Huang, W., *Curr. Opin. Chem. Biol.* **2009,** *13* (5), 592-600.

70. Wang, L.-X., *Carbohydr. Res.* **2008,** *343* (10), 1509-1522.

71. Yamamoto, K.; Fujimori, K.; Haneda, K.; Mizuno, M.; Inazu, T.; Kumagai, H., *Carbohydr. Res.* **1997,** *305 3-4*, 415-22.

72. Glover, K. J.; Weerapana, E.; Numao, S.; Imperiali, B., *Chem. Biol.* **2005,** *12* (12), 1311-1316.

73. Schwarz, F.; Fan, Y.-Y.; Schubert, M.; Aebi, M., *J. Biol. Chem.* **2011,** *286* (40), 35267-35274.

74. Witte, K.; Sears, P.; Martin, R.; Wong, C.-H., *J. Am. Chem. Soc.* **1997,** *119* (9), 2114-2118.

75. Huang, W.; Yang, Q.; Umekawa, M.; Yamamoto, K.; Wang, L.-X., *ChemBioChem* **2010,** *11* (10), 1350-1355.

76. Huang, W.; Li, C.; Li, B.; Umekawa, M.; Yamamoto, K.; Zhang, X.; Wang, L.-X., *J. Am. Chem. Soc.* **2009,** *131* (6), 2214-2223.

77. Schmaltz, R. M.; Hanson, S. R.; Wong, C.-H., *Chem. Rev.* **2011,** *111* (7), 4259-4307.

78. Hojo, H.; Tanaka, H.; Hagiwara, M.; Asahina, Y.; Ueki, A.; Katayama, H.; Nakahara, Y., et al., *J. Org. Chem.* **2012,** *77* (21), 9437-9446.

79. Hessefort, H.; Gross, A.; Seeleithner, S.; Hessefort, M.; Kirsch, T.; Perkams, L.; Bundgaard, K. O., et al., *Angew. Chem. Int. Ed.* **2021,** *60* (49), 25922-25932.

80. Huang, W.; Zhang, X.; Ju, T.; Cummings, R. D.; Wang, L.-X., *Org. Biomol. Chem.* **2010,** *8* (22), 5224-5233.

81. Giddens, J. P.; Lomino, J. V.; Amin, M. N.; Wang, L.-X., *J. Biol. Chem.* **2016,** *291* (17), 9356-9370.

82. Zhang, J.; Liu, D.; Saikam, V.; Gadi, M. R.; Gibbons, C.; Fu, X.; Song, H., et al., *Angew. Chem. Int. Ed.* **2020,** *59* (45), 19825-19829.

83. Yang, Q.; An, Y.; Zhu, S.; Zhang, R.; Loke, C. M.; Cipollo, J. F.; Wang, L.-X., *ACS Chem. Biol.* **2017,** *12* (6), 1665-1673.

84. Warnock, D.; Bai, X.; Autote, K.; Gonzales, J.; Kinealy, K.; Yan, B.; Qian, J., et al., *Biotechnol. Bioeng.* **2005,** *92* (7), 831-842.

85. Hsu, Y.-P.; Verma, D.; Sun, S.; McGregor, C.; Mangion, I.; Mann, B. F., *Commun. Biol.* **2022,** *5* (1), 328.

86. Yamamoto, N.; Tanabe, Y.; Okamoto, R.; Dawson, P. E.; Kajihara, Y., *J. Am. Chem. Soc.* **2008,** *130* (2), 501-510.

87. Wang, X.; Ashhurst, A. S.; Dowman, L. J.; Watson, E. E.; Li, H. Y.; Fairbanks, A. J.; Larance, M.; Kwan, A.; Payne, R. J., *Org. Lett.* **2020,** *22* (17), 6863-6867.

88. Murar, C. E.; Ninomiya, M.; Shimura, S.; Karakus, U.; Boyman, O.; Bode, J. W., *Angew. Chem. Int. Ed.* **2020,** *59* (22), 8425-8429.

89. Wang, P.; Dong, S.; Shieh, J.-H.; Peguero, E.; Hendrickson, R.; Moore Malcolm, A. S.; Danishefsky Samuel, J., *Science.* **2013,** *342* (6164), 1357-1360.

90. Dawson Philip, E.; Muir Tom, W.; Clark-Lewis, I.; Kent Stephen, B. H., *Science.* **1994,** *266* (5186), 776-779.

91. Jbara, M.; Maity, S. K.; Seenaiah, M.; Brik, A., *J. Am. Chem. Soc.* **2016,** *138* (15), 5069-5075.

92. Sato, K.; Tanaka, S.; Yamamoto, K.; Tashiro, Y.; Narumi, T.; Mase, N., *Chem. Commun.* **2018,** *54* (66), 9127-9130.

93. Maity, S. K.; Jbara, M.; Laps, S.; Brik, A., *Angew. Chem. Int. Ed.* **2016,** *55* (28), 8108-8112.

94. Yang, Y.-Y.; Ficht, S.; Brik, A.; Wong, C.-H., *J. Am. Chem. Soc.* **2007,** *129* (24), 7690-7701.

95. Diezmann, F.; Eberhard, H.; Seitz, O., *Pept. Sci.* **2010,** *94* (4), 397-404.

96. Reif, A.; Lam, K.; Weidler, S.; Lott, M.; Boos, I.; Lokau, J.; Bretscher, C., et al., *Angew. Chem. Int. Ed.* **2021,** *60* (24), 13380-13387.

97. Yan, L. Z.; Dawson, P. E., *J. Am. Chem. Soc.* **2001,** *123* (4), 526-533.

98. Chiang, K. P.; Jensen, M. S.; McGinty, R. K.; Muir, T. W., *ChemBioChem* **2009,** *10* (13), 2182-2187.

99. Kulkarni, S. S.; Watson, E. E.; Premdjee, B.; Conde-Frieboes, K. W.; Payne, R. J., *Nat. Prot.* **2019,** *14* (7), 2229-2257.

100. Zhang, Y.; Xu, C.; Lam Hiu, Y.; Lee Chi, L.; Li, X., *Proc. Natl. Acad. Sci.* **2013,** *110* (17), 6657-6662.

101. Brik, A.; Yang, Y.-Y.; Ficht, S.; Wong, C.-H., *J. Am. Chem. Soc.* **2006,** *128* (17), 5626-5627.

102. Brik, A.; Ficht, S.; Yang, Y.-Y.; Wong, C.-H., *J. Am. Chem. Soc.* **2006,** *128* (46), 15026-15033.

103. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D., *Chem. Biol. Drug Des.* **2013,** *81* (1), 136-147.

104. Usmani, S. S.; Bedi, G.; Samuel, J. S.; Singh, S.; Kalra, S.; Kumar, P.; Ahuja, A. A., et al., *PLOS ONE* **2017,** *12* (7), e0181748.

105. Merrifield, R. B., *J. Am. Chem. Soc.* **1963,** *85* (14), 2149-2154.

106. Carpino, L. A.; Han, G. Y., *J. Am. Chem. Soc.* **1970,** *92* (19), 5748-5749.

107. Merrifield, B., *Br. Polym. J.* **1984,** *16* (4), 173-178.

108. Palomo, J. M., *RSC Adv.* **2014,** *4* (62), 32658-32672.

109. Santini, R.; Griffith, M. C.; Qi, M., *Tetrahedron Lett.* **1998,** *39* (49), 8951-8954.

110. Wang, Z.; Yang, R.; Zhu, J.; Zhu, X., *Sci. China Chem.* **2010,** *53* (9), 1844-1852.

111. De Marco, R.; Tolomelli, A.; Greco, A.; Gentilucci, L., *ACS Sustain. Chem. Eng.* **2013,** *1* (6), 566-569.

112. Kim, M.; Park, Y.-S.; Shin, D.-S.; Lee, S.; Lee, Y.-S., *Tetrahedron Lett.* **2012,** *53* (34), 4576-4579.

113. García-Ramos, Y.; Paradís-Bas, M.; Tulla-Puche, J.; Albericio, F., *J. Pept. Sci.* **2010,** *16* (12), 675-678.

114. Rink, H., *Tetrahedron Lett.* **1987,** *28* (33), 3787-3790.

115. Palomo, J. M.; Lumbierres, M.; Waldmann, H., *Angew. Chem. Int. Ed.* **2006,** *45* (3), 477-481.

116. Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L., *Angew. Chem. Int. Ed.* **2011,** *50* (33), 7645-7649.

117. Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L., *Nat. Prot.* **2013,** *8* (12), 2483-2495.

118. Huang, Y.-C.; Chen, C.-C.; Li, S.-J.; Gao, S.; Shi, J.; Li, Y.-M., *Tetrahedron.* **2014,** *70* (18), 2951-2955.

119. Behrendt, R.; White, P.; Offer, J., *J. Pept. Sci.* **2016,** *22* (1), 4-27.

120. König, W.; Geiger, R., *Chem. Ber.* **1970,** *103* (3), 788-798.

121. Subirós-Funosas, R.; Prohens, R.; Barbas, R.; El-Faham, A.; Albericio, F., *Chem. Eur. J.* **2009,** *15* (37), 9394-9403.

122. Han, S.-Y.; Kim, Y.-A., *Tetrahedron.* **2004,** *60*, 2447-2467.

123. Albericio, F.; Bofill, J. M.; El-Faham, A.; Kates, S. A., *J. Org. Chem.* **1998,** *63* (26), 9678-9683.

124. Castro, B.; Dormoy, J. R.; Evin, G.; Selve, C., *Tetrahedron Lett.* **1975,** *16* (14), 1219-1222.

125. El-Faham, A.; Albericio, F., *Chem. Rev.* **2011,** *111* (11), 6557-6602.

126. Hartrampf, N.; Saebi, A.; Poskus, M.; Gates, Z. P.; Callahan, A. J.; Cowfer, A. E.; Hanna, S., et al., *Science.* **2020,** *368* (6494), 980-987.

127. Brik, A., 6.03 - New Strategies for Glycopeptide, Neoglycopeptide, and Glycoprotein Synthesis. In *Comprehensive Natural Products II*, Liu, H.-W.; Mander, L., Eds. Elsevier: Oxford, 2010; pp 55-89.

128. Cohen-Anisfeld, S., .; Lansbury, P., *J. Am. Chem. Soc.* **1993,** *115*, 10531-10537.

129. Fernández-Tejada, A.; Brailsford, J.; Zhang, Q.; Shieh, J.-H.; Moore, M. A. S.; Danishefsky, S. J., Total Synthesis of Glycosylated Proteins. In *Protein Ligation and Total Synthesis I*, Liu, L., Ed. Springer International Publishing: Cham, 2015; pp 1-26.

130. Subirós-Funosas, R.; El-Faham, A.; Albericio, F., *Tetrahedron.* **2011,** *67* (45), 8595-8606.

131. Mergler, M.; Dick, F.; Sax, B.; Weiler, P.; Vorherr, T., *J. Pept. Sci.* **2003,** *9* (1), 36-46.

132. Wade, J. D.; Mathieu, M. N.; Macris, M.; Tregear, G. W., *Lett. Pept. Sci.* **2000,** *7* (2), 107-112.

133. Mergler, M.; Dick, F.; Sax, B.; Stähelin, C.; Vorherr, T., *J. Pept. Sci.* **2003,** *9* (8), 518-526.

134. Mergler, M.; Dick, F., *J. Pept. Sci.* **2005,** *11* (10), 650-657.

135. Li, W.; Backlund, P. S.; Boykins, R. A.; Wang, G.; Chen, H.-C., *Anal. Biochem.* **2003,** *323* (1), 94-102.

136. Paradís-Bas, M.; Tulla-Puche, J.; Albericio, F., *Chem. Soc. Rev.* **2016,** *45* (3), 631-654.

137. Zompra, A. A.; Galanis, A. S.; Werbitzky, O.; Albericio, F., *Future Med. Chem.* **2009,** *1* (2), 361-377.

138. Abragam, J.; Pardo-Vargas, A.; Seeberger, P., *J. Am. Chem. Soc.* **2020,** *142* (19), 8561-8564.

139. Wu, Y.; Xiong, D.-C.; Chen, S.-C.; Wang, Y.-S.; Ye, X.-S., *Nat. Commun.* **2017,** *8* (1), 14851.

140. Krasnova, L.; Wong, C.-H., *J. Am. Chem. Soc.* **2019,** *141* (9), 3735-3754.

141. Guberman, M.; Seeberger, P. H., *J. Am. Chem. Soc.* **2019,** *141* (14), 5581-5592.

142. Wen, L.; Edmunds, G.; Gibbons, C.; Zhang, J.; Gadi, M. R.; Zhu, H.; Fang, J., et al., *Chem. Rev.* **2018,** *118* (17), 8151-8187.

143. Das, R.; Mukhopadhyay, B., *ChemistryOpen.* **2016,** *5* (5), 401-433.

144. Stick, R. V.; Williams, S. J., *Carbohydrates: The Essential Molecules of Life (Second Edition)*. Elsevier: 2009.

145. Ranade, S. C.; Demchenko, A. V., *J. Carbohydr. Chem.* **2013,** *32* (1), 1-43.

146. Nigudkar, S. S.; Demchenko, A. V., *Chem. Sci.* **2015,** *6* (5), 2687-2704.

147. Chatterjee, S.; Moon, S.; Hentschel, F.; Gilmore, K.; Seeberger, P. H., *J. Am. Chem. Soc.* **2018,** *140* (38), 11942-11953.

148. Kalikanda, J.; Li, Z., *J. Org. Chem.* **2011,** *76* (13), 5207-5218.

149. Baek, J. Y.; Lee, B.-Y.; Jo, M. G.; Kim, K. S., *J. Am. Chem. Soc.* **2010,** *132* (20), 7229-7229.

150. Imamura, A.; Ando, H.; Korogi, S.; Tanabe, G.; Muraoka, O.; Ishida, H.; Kiso, M., *Tetrahedron Lett.* **2003,** *44* (35), 6725-6728.

151. Imamura, A.; Matsuzawa, N.; Sakai, S.; Udagawa, T.; Nakashima, S.; Ando, H.; Ishida, H.; Kiso, M., *J. Org. Chem.* **2016,** *81* (19), 9086-9104.

152. Tvaroŝka, I.; Bleha, T., Anomeric and Exo-Anomeric Effects in Carbohydrate Chemistry. In *Advances in Carbohydrate Chemistry and Biochemistry*, Tipson, R. S.; Horton, D., Eds. Academic Press: 1989; Vol. 47, pp 45-123.

153. Demchenko, A. V.; Rousson, E.; Boons, G.-J., *Tetrahedron Lett.* **1999,** *40* (36), 6523-6526.

154. Satoh, H.; Hansen, H. S.; Manabe, S.; van Gunsteren, W. F.; Hünenberger, P. H., *J. Chem. Theory Comput.* **2010,** *6* (6), 1783-1797.

155. Pawar, N. J.; Wang, L.; Higo, T.; Bhattacharya, C.; Kancharla, P. K.; Zhang, F.; Baryal, K., et al., *Angew. Chem. Int. Ed.* **2019,** *58* (51), 18577-18583.

156. Seeberger, P. H.; Haase, W.-C., *Chem. Rev.* **2000,** *100* (12), 4349-4394.

157. Seeberger, P. H.; Werz, D. B., *Nat. Rev. Drug Discov.* **2005,** *4* (9), 751-763.

158. Plante Obadiah, J.; Palmacci Emma, R.; Seeberger Peter, H., *Science.* **2001,** *291* (5508), 1523-1527.

159. Seeberger, P. H., *Acc. Chem. Res.* **2015,** *48* (5), 1450-1463.

160. Yu, Y.; Kononov, A.; Delbianco, M.; Seeberger, P. H., *Chem. Eur. J.* **2018,** *24* (23), 6075-6078.

161. Tyrikos-Ergas, T.; Sletten, E. T.; Huang, J.-Y.; Seeberger, P. H.; Delbianco, M., *Chem. Sci.* **2022,** *13* (7), 2115-2120.

162. Sletten, E.; Danglad Flores, J.; Leichnitz, S.; Joseph, A.; Seeberger, P., *Carbohydr. Res.* **2022,** *511*, 108489.

163. Routenberg Love, K.; Seeberger, P. H., *Angew. Chem. Int. Ed.* **2004,** *43* (5), 602-605.

164. de Jong, A. R.; Volbeda, A. G.; Hagen, B.; van den Elst, H.; Overkleeft, H. S.; van der Marel, G. A.; Codée, J. D. C., *Eur. J. Org. Chem.* **2013,** *2013* (29), 6644-6655.

165. Kröck, L.; Esposito, D.; Castagner, B.; Wang, C.-C.; Bindschädler, P.; Seeberger, P., *Chem. Sci.* **2012,** *3*, 1617-1622.

166. Eller, S.; Collot, M.; Yin, J.; Hahm, H. S.; Seeberger, P. H., *Angew. Chem. Int. Ed.* **2013,** *52* (22), 5858-5861.

167. Le Mai Hoang, K.; Pardo-Vargas, A.; Zhu, Y.; Yu, Y.; Loria, M.; Delbianco, M.; Seeberger, P. H., *J. Am. Chem. Soc.* **2019,** *141* (22), 9079-9086.

168. Zhu, Y.; Delbianco, M.; Seeberger, P. H., *J. Am. Chem. Soc.* **2021,** *143* (26), 9758-9768.

169. Danglad-Flores, J.; Leichnitz, S.; Sletten, E. T.; Abragam Joseph, A.; Bienert, K.; Le Mai Hoang, K.; Seeberger, P. H., *J. Am. Chem. Soc.* **2021,** *143* (23), 8893-8901.

170. Guberman, M.; Bräutigam, M.; Seeberger, P. H., *Chem. Sci.* **2019,** *10* (21), 5634-5640.

171. Pardo-Vargas, A.; Bharate, P.; Delbianco, M.; Seeberger, P. H., *Beilstein J. Org. Chem.* **2019,** *15*, 2936-2940.

172. Lisboa Marilda, P.; Khan, N.; Martin, C.; Xu, F.-F.; Reppe, K.; Geissner, A.; Govindan, S., et al., *Proc. Natl. Acad. Sci.* **2017,** *114* (42), 11063-11068.

173. Bartetzko, M. P.; Pfrengle, F., *ChemBioChem* **2019,** *20* (7), 877-885.

174. Yu, Y.; Tyrikos-Ergas, T.; Zhu, Y.; Fittolani, G.; Bordoni, V.; Singhal, A.; Fair, R. J., et al., *Angew. Chem. Int. Ed.* **2019,** *58* (37), 13127-13132.

175. Anggara, K.; Zhu, Y.; Fittolani, G.; Yu, Y.; Tyrikos-Ergas, T.; Delbianco, M.; Rauschenbach, S., et al., *Proc. Natl. Acad. Sci.* **2021,** *118* (23), e2102168118.

176. Yu, Y.; Gim, S.; Kim, D.; Arnon, Z. A.; Gazit, E.; Seeberger, P. H.; Delbianco, M., *J. Am. Chem. Soc.* **2019,** *141* (12), 4833-4838.

177. Singh, S.; Scigelova, M.; Vic, G.; Crout, D. H. G., *J. Chem. Soc., Perkin trans. 1.* **1996,** (16), 1921-1926.

178. Krasnova, L.; Wong, C.-H., *Annu. Rev. Biochem.* **2016,** *85* (1), 599-630.

179. Zhang, R.; Liu, Q.; Liao, Q.; Zhao, Y., *Future Oncol.* **2018,** *14* (8), 781-791.

180. Davies, A.; Simmons, D. L.; Hale, G.; Harrison, R. A.; Tighe, H.; Lachmann, P. J.; Waldmann, H., *J. Exp. Med.* **1989,** *170* (3), 637-654.

181. Rudd, P. M.; Morgan, B. P.; Wormald, M. R.; Harvey, D. J.; van den Berg, C. W.; Davis, S. J.; Ferguson, M. A.; Dwek, R. A., *J. Biol. Chem.* **1997,** *272* (11), 7229-44.

182. Yu, Q.; Yu, R.; Qin, X., *Cell. Mol. Immunol.* **2010,** *7* (5), 334-40.

183. Fletcher, C. M.; Harrison, R. A.; Lachmann, P. J.; Neuhaus, D., *Structure* **1994,** *2* (3), 185-199.

184. Sugita, Y.; Nakano, Y.; Oda, E.; Noda, K.; Tobe, T.; Miura, N.-H.; Tomita, M., *J. Biochem.* **1993,** *114* (4), 473-477.

185. Kieffer, B.; Driscoll, P. C.; Campbell, I. D.; Willis, A. C.; van der Merwe, P. A.; Davis, S. J., *Biochemistry.* **1994,** *33* (15), 4471-4482.

186. Huang, Y.; Fedarovich, A.; Tomlinson, S.; Davies, C., *Acta Cryst.* **2007,** *63* (Pt 6), 714-21.

187. Kimberley, F. C.; Sivasankar, B.; Paul Morgan, B., *Mol. Immunol.* **2007,** *44* (1), 73-81.

188. Michielsen, L. A.; Budding, K.; Drop, D.; van de Graaf, E. A.; Kardol-Hoefnagel, T.; Verhaar, M. C.; van Zuilen, A. D.; Otten, H. G., *Front. immunol.* **2018,** *8*.

189. Budding, K.; van de Graaf, E. A.; Otten, H. G., *Transpl. Immunol.* **2014,** *31* (4), 260-265.

190. Rao, S. P.; Sancho, J.; Campos-Rivera, J.; Boutin, P. M.; Severy, P. B.; Weeden, T.; Shankara, S.; Roberts, B. L.; Kaplan, J. M., *PLOS ONE* **2012,** *7* (6), e39416.

191. Kolev, M. V.; Tediose, T.; Sivasankar, B.; Harris, C. L.; Thome, J.; Morgan, B. P.; Donev, R. M., *Pharmacogenomics J.* **2010,** *10* (1), 12-19.

192. Budding, K.; van de Graaf, E. A.; Kardol-Hoefnagel, T.; Kwakkel-van Erp, J. M.; Luijk, B. D.; Oudijk, E.-J. D.; van Kessel, D. A., et al., *Sci. Rep.* **2016,** *6* (1), 26274.

193. Bora, N. S.; Kaliappan, S.; Jha, P.; Xu, Q.; Sivasankar, B.; Harris, C. L.; Morgan, B. P.; Bora, P. S., *J Immunol* **2007,** *178* (3), 1783.

194. Ruiz-Delgado, G. J.; Vázquez-Garza, E.; Méndez-Ramírez, N.; Gómez-Almaguer, D., *Hematology* **2009,** *14* (1), 33-37.

195. Ghosh, P.; Sahoo, R.; Vaidya, A.; Chorev, M.; Halperin, J. A., *Endocr. Rev.* **2015,** *36* (3), 272-288.

196. Hansson, G. C., *J. Int. Med.* **2019,** *285* (5), 479-490.

197. Wagner, C. E.; Wheeler, K. M.; Ribbeck, K., *Annu. Rev. Cell Dev. Biol.* **2018,** *34* (1), 189-215.

198. Petrou, G.; Crouzier, T., *Biomater. Sci.* **2018,** *6* (9), 2282-2297.

199. Bansil, R.; Turner, B. S., *Adv. Drug. Deliv. Rev.* **2018,** *124*, 3-15.

200. Di Noia, J. M.; Buscaglia, C. A.; De Marchi, C. R.; Almeida, I. C.; Frasch, A. C., *J. Exp. Med.* **2002,** *195* (4), 401-13.

201. Tomita, T.; Ma, Y.; Weiss, L., *Parasitol. Res.* **2018,** *117* (8), 2457-2466.

202. Altgärde, N.; Eriksson, C.; Peerboom, N.; Phan-Xuan, T.; Moeller, S.; Schnabelrauch, M.; Svedhem, S., et al., *J. Biol. Chem.* **2015,** *290* (35), 21473-21485.

203. Lee, J. E.; Fusco, M. L.; Hessell, A. J.; Oswald, W. B.; Burton, D. R.; Saphire, E. O., *Nature.* **2008,** *454*, 177.

204. van Putten, J. P. M.; Strijbis, K., *J. Innate. Immun.* **2017,** *9* (3), 281-299.

205. Linden, S. K.; Sutton, P.; Karlsson, N. G.; Korolik, V.; McGuckin, M. A., *Mucosal Immunol.* **2008,** *1* (3), 183-97.

206. Dhanisha, S. S.; Guruvayoorappan, C.; Drishya, S.; Abeesh, P., *Crit. Rev. Oncol. Hematol.* **2018,** *122*, 98-122.

207. Brockhausen, I., *Glycodynamics of Mucin Biosynthesis in Gastrointestinal Tumor Cells*. Springer: Boston, MA, 2003; Vol. 535, p 163-168.

208. Kasprzak, A.; Adamek, A., *Int. J. Mol. Sci.* **2019,** *20* (6).

209. Sheng, Y. H.; Hasnain, S. Z.; Florin, T. H. J.; McGuckin, M. A., *J. Gastroen. Hepatol.* **2012,** *27* (1), 28-38.

210. Behera, S. K.; Praharaj, A. B.; Dehury, B.; Negi, S., *Glycoconj. J.* **2015,** *32* (8), 575-613.

211. Bhattacharyya, T.; Falconar, A. K.; Luquetti, A. O.; Costales, J. A.; Grijalva, M. J.; Lewis, M. D.; Messenger, L. A., et al., *PLOS Negl. Trop. Dis.* **2014,** *8* (5), e2892.

212. Martínez-Sáez, N.; Peregrina, J. M.; Corzana, F., *Chem. Soc. Rev.* **2017,** *46* (23), 7154-7175.

213. Dekker, J.; Rossen, J. W. A.; Büller, H. A.; Einerhand, A. W. C., *Trends Biochem. Sci.* **2002,** *27* (3), 126-131.

214. Desseyn, J.-L., *Mol. Phylogenet. Evol.* **2009,** *52* (2), 284-292.

215. Xu, D.; Pavlidis, P.; Thamadilok, S.; Redwood, E.; Fox, S.; Blekhman, R.; Ruhl, S.; Gokcumen, O., *Sci. Rep.* **2016,** *6* (1), 31791.

216. Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A., *Glycobiology.* **2011,** *22* (6), 736-756.

217. Brockhausen, I., *EMBO reports.* **2006,** *7* (6), 599-604.

218. Jin, C.; Kenny, D. T.; Skoog, E. C.; Padra, M.; Adamczyk, B.; Vitizeva, V.; Thorell, A., et al., *Mol. Cell. Proteomics.* **2017,** *16* (5), 743-758.

219. El-Sayed, N. M.; Myler, P. J.; Bartholomeu, D. C.; Nilsson, D.; Aggarwal, G.; Tran, A.-N.; Ghedin, E., et al., *Science.* **2005,** *309* (5733), 409-415.

220. Mendonça-Previato, L.; Penha, L.; Garcez, T. C.; Jones, C.; Previato, J. O., *Glycoconj. J.* **2013,** *30* (7), 659-666.

221. Previato, J. O.; Sola-Penna, M.; Agrellos, O. A.; Jones, C.; Oeltmann, T.; Travassos, L. R.; Mendonça-Previato, L., *J. Biol. Chem.* **1998,** *273* (24), 14982-14988.

222. Giorgi, M. E.; de Lederkremer, R. M., *Molecules.* **2020,** *25* (17), 3913.

223. Previato, J. O.; Jones, C.; Xavier, M. T.; Wait, R.; Travassos, L. R.; Parodi, A. J.; Mendonca-Previato, L., *J. Biol. Chem.* **1995,** *270* (13), 7241-50.

224. Jones, C.; Todeschini, A. R.; Agrellos, O. A.; Previato, J. O.; Mendonca-Previato, L., *Biochemistry.* **2004,** *43* (37), 11889-97.

225. Agrellos, O. A.; Jones, C.; Todeschini, A. R.; Previato, J. O.; Mendonca-Previato, L., *Mol. Biochem. Parasitol.* **2003,** *126* (1), 93-6.

226. Todeschini, A. R.; de Almeida, E. G.; Agrellos, O. A.; Jones, C.; Previato, J. O.; Mendonca-Previato, L., *Mem. Inst. Oswaldo Cruz.* **2009,** *104 Suppl 1*, 270-4.

227. Almeida, I. C.; Ferguson, M. A.; Schenkman, S.; Travassos, L. R., *Biochem. J.* **1994,** *304 (Pt 3)*, 793-802.

228. Previato, J. O.; Jones, C.; Goncalves, L. P.; Wait, R.; Travassos, L. R.; Mendonca-Previato, L., *Biochem. J.* **1994,** *301 (Pt 1)*, 151-9.

229. de Lederkremer, R. M.; Agusti, R., *Adv. Carbohydr. Chem. Biochem.* **2009,** *62*, 311-66.

230. Gallo-Rodriguez, C.; Varela, O.; Lederkremer, R. M., *Journal of Organic Chemistry* **1996,** *61* (5), 1886-1889.

231. Gallo-Rodriguez, C.; Varela, O.; Lederkremer, R., *Carbohydrate Research* **1998,** *305*, 163-170.

232. Mendoza, V. M.; Kashiwagi, G. A.; de Lederkremer, R. M.; Gallo-Rodriguez, C., *Carbohydr. Res.* **2010,** *345* (3), 385-96.

233. Gallo-Rodriguez, C.; Gil Libarona, M.; M. Mendoza, V.; Lederkremer, R., *ChemInform* **2003,** *34*.

234. Mendoza, V. M.; Agusti, R.; Gallo-Rodriguez, C.; de Lederkremer, R. M., *Carbohydr. Res.* **2006,** *341* (10), 1488-97.

235. Agusti, R.; Giorgi, M. E.; Mendoza, V. M.; Kashiwagi, G. A.; de Lederkremer, R. M.; Gallo-Rodriguez, C., *Bioorganic and Medicinal Chemistry* **2015,** *23* (6), 1213-22.

236. van Well, R. M.; Collet, B.; Field, R., *Synlett* **2008,** *14*, 2175-2177.

237. Campo, V. L.; Carvalho, I.; Allman, S.; Davis, B. G.; Field, R. A., *Organic & Biomolecular Chemistry* **2007,** *5* (16), 2645-2657.

238. Gude, M.; Ryf, J.; White, P. D., *Lett. Pept. Sci.* **2002,** *9* (4), 203-206.

239. Collins, J. M.; Porter, K. A.; Singh, S. K.; Vanier, G. S., *Org. Lett.* **2014,** *16* (3), 940-943.

240. Collins, J. M.; Palasek, S. A.; Cox, Z. J., Expanding Applications of Microwave Enhanced SPPS. In *29th European Peptide Symposium*, Gdansk, Poland, 2006.

241. Michels, T.; Dölling, R.; Haberkorn, U.; Mier, W., *Org. Lett.* **2012,** *14* (20), 5218-5221.

242. Offer, J.; Quibell, M.; Johnson, T., *J. Chem. Soc., Perkin trans. 1.* **1996,** (2), 175-182.

243. Palasek, S. A.; Cox, Z. J.; Collins, J. M., *J. Pept. Sci.* **2007,** *13* (3), 143-148.

244. Subirós-Funosas, R.; El-Faham, A.; Albericio, F., *Pept. Sci.* **2012,** *98* (2), 89-97.

245. Martinez, J.; Bodanszky, M., *Int. J. Pept. Protein Res.* **1978,** *12* (5), 277-283.

246. Desfougères, Y.; Jardin, J.; Lechevalier, V.; Pezennec, S.; Nau, F., *Biomacromolecules.* **2011,** *12* (1), 156-166.

247. Bang, D.; Kent, S. B. H., *Angew. Chem. Int. Ed.* **2004,** *43* (19), 2534-2538.

248. Kamo, N.; Hayashi, G.; Okamoto, A., *Angew. Chem. Int. Ed.* **2018,** *57* (50), 16533-16537.

249. Kamo, N.; Hayashi, G.; Okamoto, A., *Org. Lett.* **2019,** *21* (20), 8378-8382.

250. Kar, A.; Mannuthodikayil, J.; Singh, S.; Biswas, A.; Dubey, P.; Das, A.; Mandal, K., *Angew. Chem. Int. Ed.* **2020,** *59* (35), 14796-14801.

251. Ibatullin, F. M.; Selivanov, S. I., *Tetrahedron Lett.* **2009,** *50* (46), 6351-6354.

252. Huang, X.; Luo, X.; Roupioz, Y.; Keillor, J. W., *J. Org. Chem.* **1997,** *62* (25), 8821-8825.

253. Furukawa, J.-i.; Fujitani, N.; Araki, K.; Takegawa, Y.; Kodama, K.; Shinohara, Y., *Analytical Chemistry* **2011,** *83* (23), 9060-9067.

254. An, H. J.; Froehlich, J. W.; Lebrilla, C. B., *Curr. Opin. Chem. Biol.* **2009,** *13* (4), 421-426.

255. Singh, Y.; Rodriguez Benavente, M. C.; Al-Huniti, M. H.; Beckwith, D.; Ayyalasomayajula, R.; Patino, E.; Miranda, W. S.; Wade, A.; Cudic, M., *J. Org. Chem.* **2020,** *85* (3), 1434-1445.

256. Thompson, R. E.; Liu, X.; Alonso-García, N.; Pereira, P. J. B.; Jolliffe, K. A.; Payne, R. J., *J. Am. Chem. Soc.* **2014,** *136* (23), 8161-8164.

257. Lechner, C. C.; Agashe, N. D.; Fierz, B., *Angew. Chem. Int. Ed.* **2016,** *55* (8), 2903-2906.

258. Seenaiah, M.; Jbara, M.; Mali, S. M.; Brik, A., *Angew. Chem. Int. Ed.* **2015,** *54* (42), 12374-12378.

259. Paiotta, A.; D'Orazio, G.; Palorini, R.; Ricciardiello, F.; Zoia, L.; Votta, G.; De Gioia, L.; Chiaradonna, F.; La Ferla, B., *Eur. J. Org. Chem.* **2018,** *2018* (17), 1946-1952.

260. Premdjee, B.; Adams, A. L.; Macmillan, D., *Bioorg. Med. Chem. Lett.* **2011,** *21* (17), 4973-4975.

261. Chan, W. C.; White, P. D., *Fmoc Solid Phase Peptide Synthesis: A practical Approach*. Oxford University Press: 2000.

262. Gallo-Rodriguez, C.; Varela, O.; Lederkremer, R. M., *J. Org. Chem.* **1996,** *61* (5), 1886-1889.

263. Gallo-Rodriguez, C.; Varela, O.; Lederkremer, R., *Carbohydr. Res.* **1997,** *305*, 163-170.

264. Gallo-Rodriguez, C.; Gil Libarona, M.; M. Mendoza, V.; Lederkremer, R., *Tetrahedron.* **2002,** *58* (46), 9373-9380.

265. Agusti, R.; Giorgi, M. E.; Mendoza, V. M.; Kashiwagi, G. A.; de Lederkremer, R. M.; Gallo-Rodriguez, C., *Bioorg. Med. Chem.* **2015,** *23* (6), 1213-22.

266. Kashiwagi, G. A.; Cori, C. R.; de Lederkremer, R. M.; Gallo-Rodriguez, C., *Carbohydr. Res.* **2019,** *482*, 107734.

267. Van Well, R. M.; Collet, B.; Field, R., *Synlett.* **2008,** *14*, 2175-2177.

268. Lopez, R.; Giorgi, M. E.; Melgarejo, L. T.; Ducrey, I.; Balouz, V.; Gonzalez-Salas, D.; Camara, M. L. M., et al., *Carbohydr. Res.* **2019,** *478*, 58-67.

269. Portillo, S.; Zepeda, B. G.; Iniguez, E.; Olivas, J. J.; Karimi, N. H.; Moreira, O. C.; Marques, A. F., et al., *NPJ Vaccines.* **2019,** *4*, 13.

270. Smith, R.; Müller-Bunz, H.; Zhu, X., *Org. Lett.* **2016,** *18* (15), 3578-3581.

271. Xu, H.; Lu, Y.; Zhou, Y.; Ren, B.; Pei, Y.; Dong, H.; Pei, Z., *Adv. Synth. Catal.* **2014,** *356* (8), 1735-1740.

272. Yamada, K.; Fujita, H.; Kunishima, M., *Org. Lett.* **2012,** *14* (19), 5026-5029.

273. Wang, L.; Hashidoko, Y.; Hashimoto, M., *J. Org. Chem.* **2016,** *81* (11), 4464-4474.

274. Bartetzko, M. P.; Schuhmacher, F.; Hahm, H. S.; Seeberger, P. H.; Pfrengle, F., *Org. Lett.* **2015,** *17* (17), 4344-4347.

275. Wang, C.-C.; Lee, J.-C.; Luo, S.-Y.; Kulkarni, S. S.; Huang, Y.-W.; Lee, C.-C.; Chang, K.-L.; Hung, S.-C., *Nature.* **2007,** *446* (7138), 896-899.

276. Thijssen, M.-J. L.; Halkes, K. M.; Kamerling, J. P.; Vliegenthart, J. F. G., *Bioorg. Med. Chem.* **1994,** *2* (11), 1309-1317.

277. Completo, G. C.; Lowary, T. L., *J. Org. Chem.* **2008,** *73* (12), 4513-4525.

278. Glibstrup, E.; Pedersen, C. M., *Org. Lett.* **2016,** *18* (17), 4424-4427.

279. Ishiwata, A.; Munemura, Y.; Ito, Y., *Tetrahedron.* **2008,** *64* (1), 92-102.

280. Wu, X.; McKay, C.; Pett, C.; Yu, J.; Schorlemer, M.; Ramadan, S.; Lang, S., et al., *ACS Chem. Biol.* **2019,** *14* (10), 2176-2184.

281. Orii, R.; Sakamoto, N.; Fukami, D.; Tsuda, S.; Izumi, M.; Kajihara, Y.; Okamoto, R., *Chem. Eur. J.* **2017,** *23* (39), 9253-9257.

282. Hurevich, M.; Seeberger, P. H., *Chem. Commun.* **2014,** *50* (15), 1851-1853.

283. Senf, D.; Ruprecht, C.; de Kruijff, G. H. M.; Simonetti, S. O.; Schuhmacher, F.; Seeberger, P. H.; Pfrengle, F., *Chem. Eur. J.* **2017,** *23* (13), 3197-3205.

284. Zhu, Y.; Tyrikos-Ergas, T.; Schiefelbein, K.; Grafmüller, A.; Seeberger, P. H.; Delbianco, M., *Org. Biomol. Chem.* **2020,** *18* (7), 1349-1353.

285. Randell, K. D.; Johnston, B. D.; Brown, P. N.; Pinto, B. M., *Carbohydr. Res.* **2000,** *325* (4), 253-264.

286. van der Vorm, S.; van Hengst, J. M. A.; Bakker, M.; Overkleeft, H. S.; van der Marel, G. A.; Codée, J. D. C., *Angew. Chem. Int. Ed.* **2018,** *57* (27), 8240-8244.

287. Bhalay, G.; Dunstan, A. R., *Tetrahedron Lett.* **1998,** *39* (42), 7803-7806.

288. Frawley Cass, S. M.; Reid, G. E.; Tepe, J. J., *Org. Biomol. Chem.* **2009,** *7* (16), 3291-3299.

289. Lindberg, J.; Svensson, S.; Påhlsson, P.; Konradsson, P., *Tetrahedron.* **2002,** *58*, 5109-5117.

290. Buskas, T.; Li, Y.; Boons, G. J., *Chem. Eur. J.* **2005,** *11* (18), 5457-67.

291. Plé, K.; Chwalek, M.; Voutquenne-Nazabadioko, L., *Tetrahedron.* **2005,** *61* (18), 4347-4362.

292. Thijssen, M.-J. L.; Halkes, K. M.; Kamerling, J. P.; Vliegenthart, J. F. G., *Bioorg. Med. Chem. Lett.* **1994,** *2* (11), 1309-1317.

293. Grube, M.; Lee, B.-Y.; Garg, M.; Michel, D.; Vilotijević, I.; Malik, A.; Seeberger, P. H.; Varón Silva, D., *Chem. Eur. J.* **2018,** *24* (13), 3271-3282.

294. Completo, G.; Sangalang, R.; Pique, B.; Nacario, R., *KIMIKA.* **2017,** *27*, 38-49.

295. Masui, S.; Manabe, Y.; Hirao, K.; Shimoyama, A.; Fukuyama, T.; Ryu, I.; Fukase, K., *Synlett.* **2019,** *30* (04), 397-400.

296. Gelin, M.; Ferrières, V.; Plusquellec, D., *Eur. J. Org. Chem.* **2000,** *2000* (8), 1423-1431.

297. Wiegmann, D.; Spork, A. P.; Niro, G.; Ducho, C., *Synlett.* **2018,** *29* (04), 440-446.

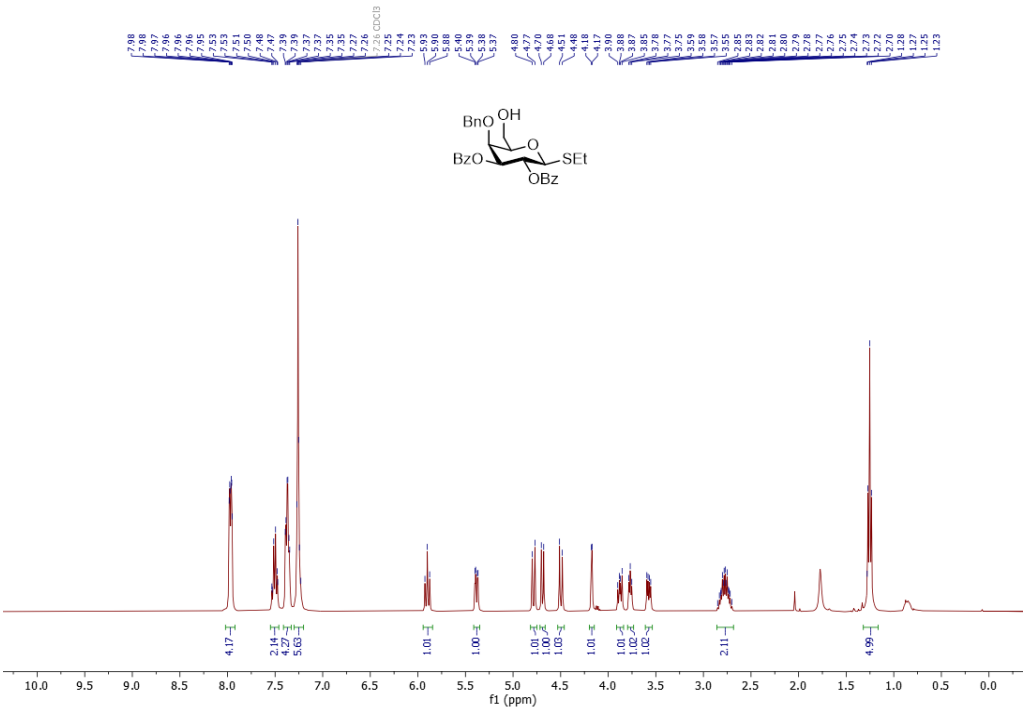
298. Gross, P. J.; Bräse, S., *Chem. Eur. J.* **2010,** *16 42*, 12660-7.

299. Hoffmann, J.; Kazmaier, U., *Synthesis.* **2015,** *47* (3), 411-420.

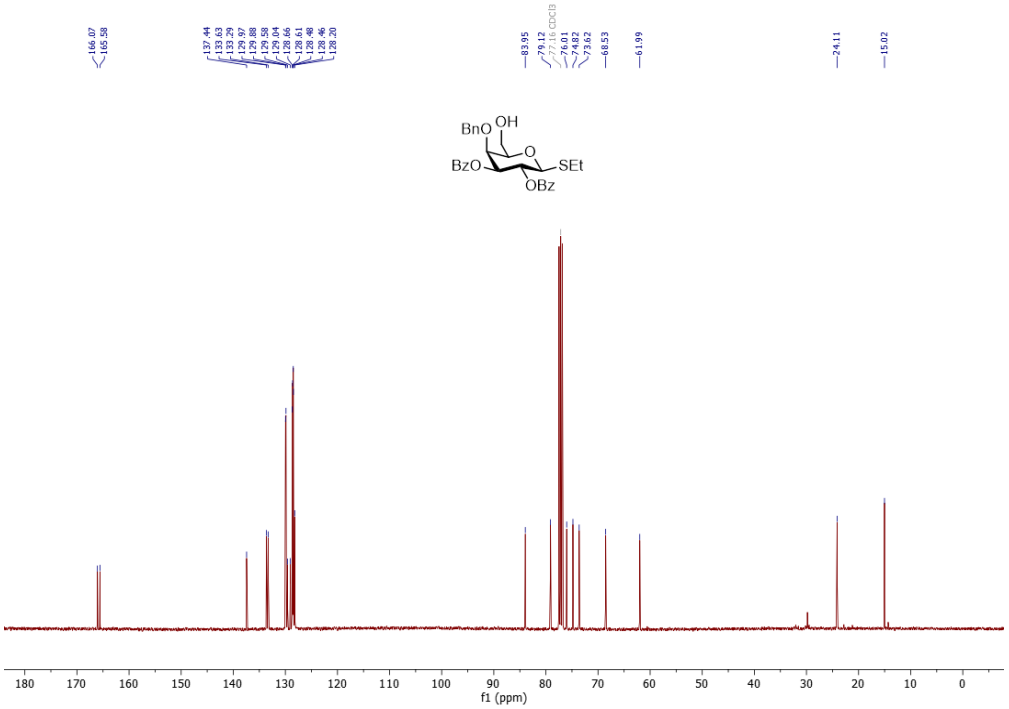
300. Azagarsamy, M. A.; Alge, D. L.; Radhakrishnan, S. J.; Tibbitt, M. W.; Anseth, K. S., *Biomacromolecules.* **2012,** *13* (8), 2219-2224.

Appendix: NMR Spectra of New Compounds

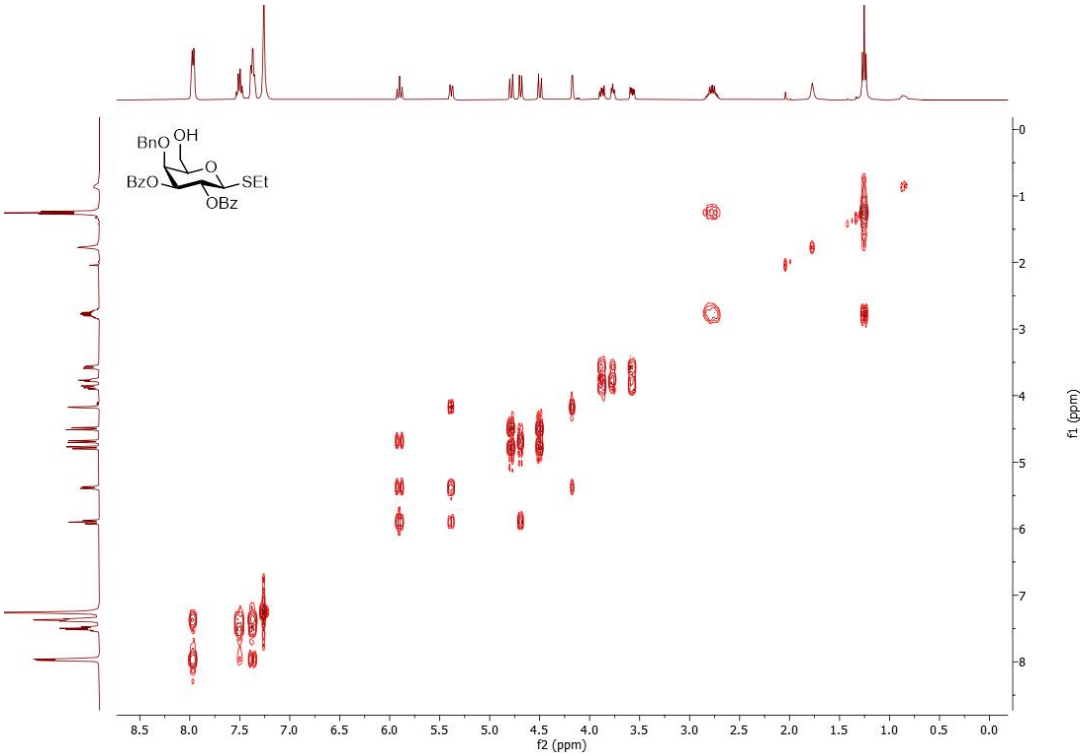
^{1}H NMR (400 MHz, $CDCl_3$) of **54**



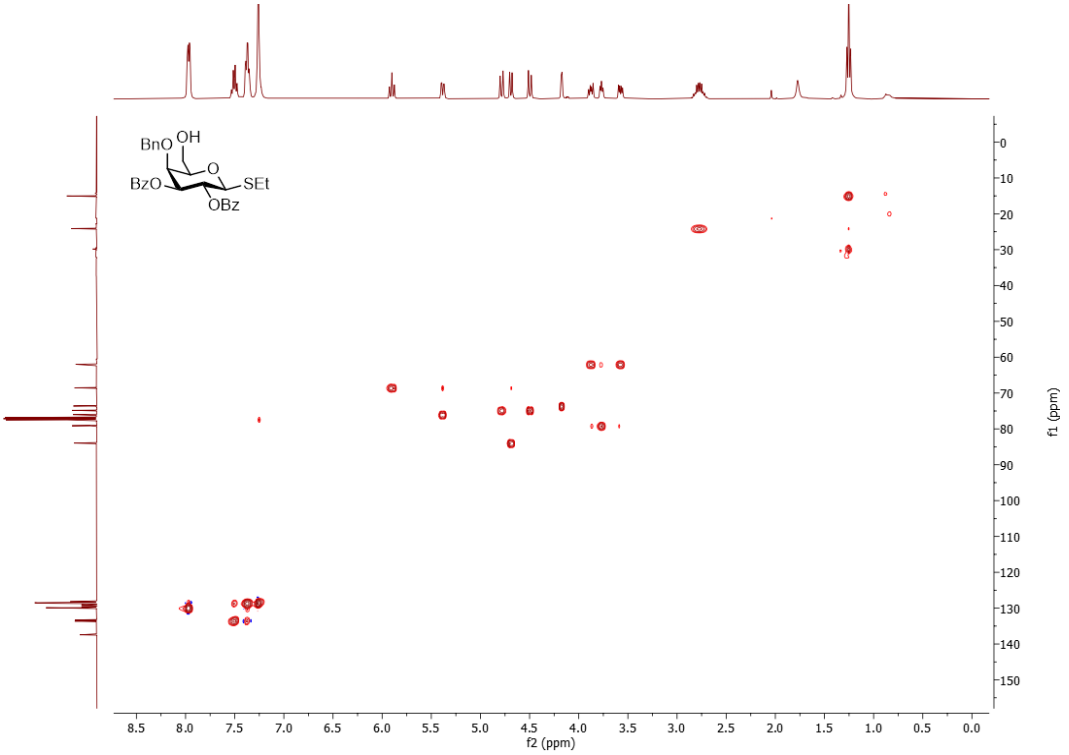
^{13}C NMR (101 MHz, $CDCl_3$) of **54**



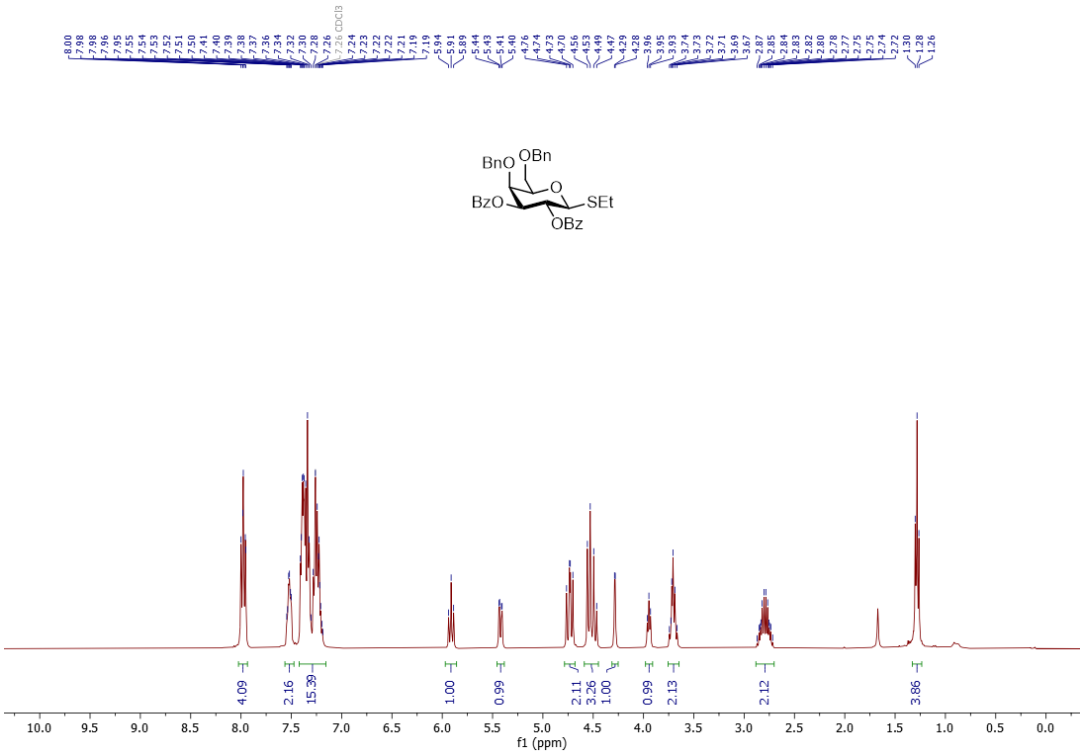
^{1}H, ^{1}H COSY of **54**



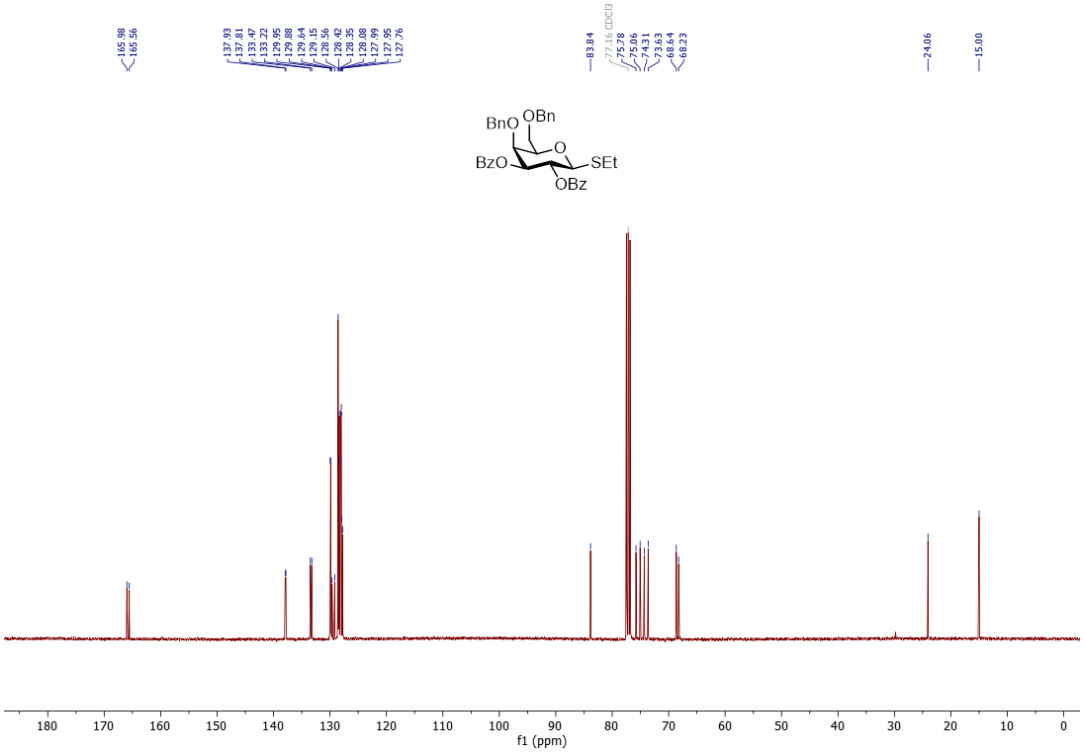
^{13}C, ^{1}H HSQC of **54**



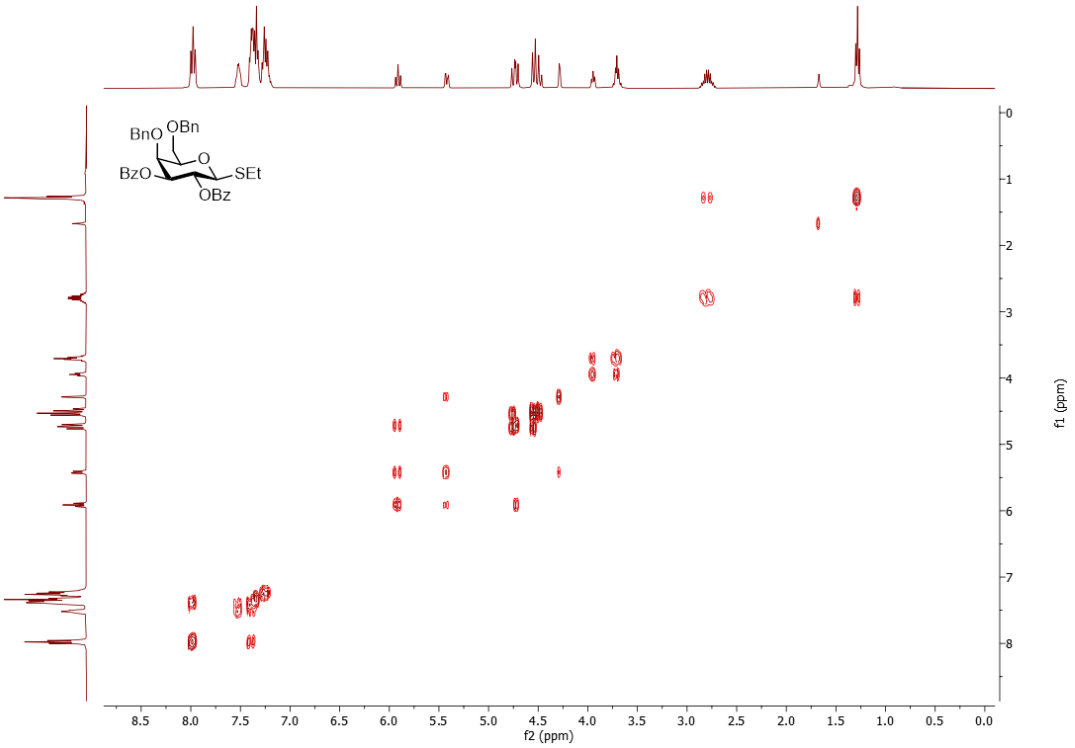
^{1}H NMR (400 MHz, $CDCl_3$) of **55**



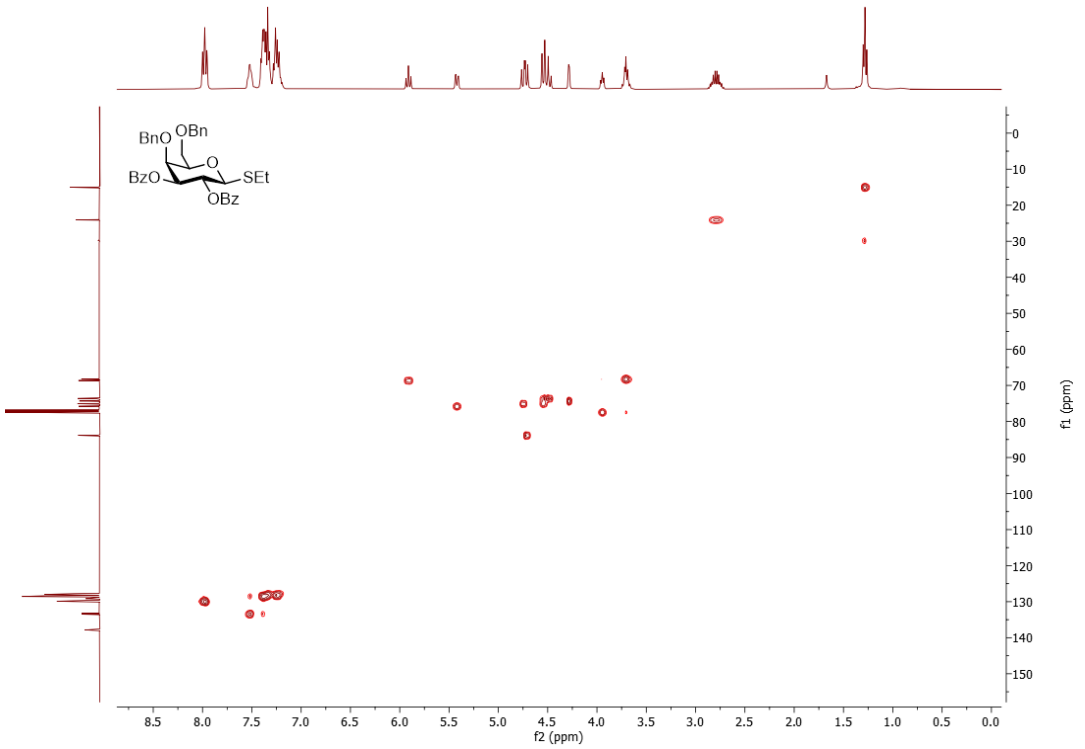
^{13}C NMR (101 MHz, $CDCl_3$) of **55**



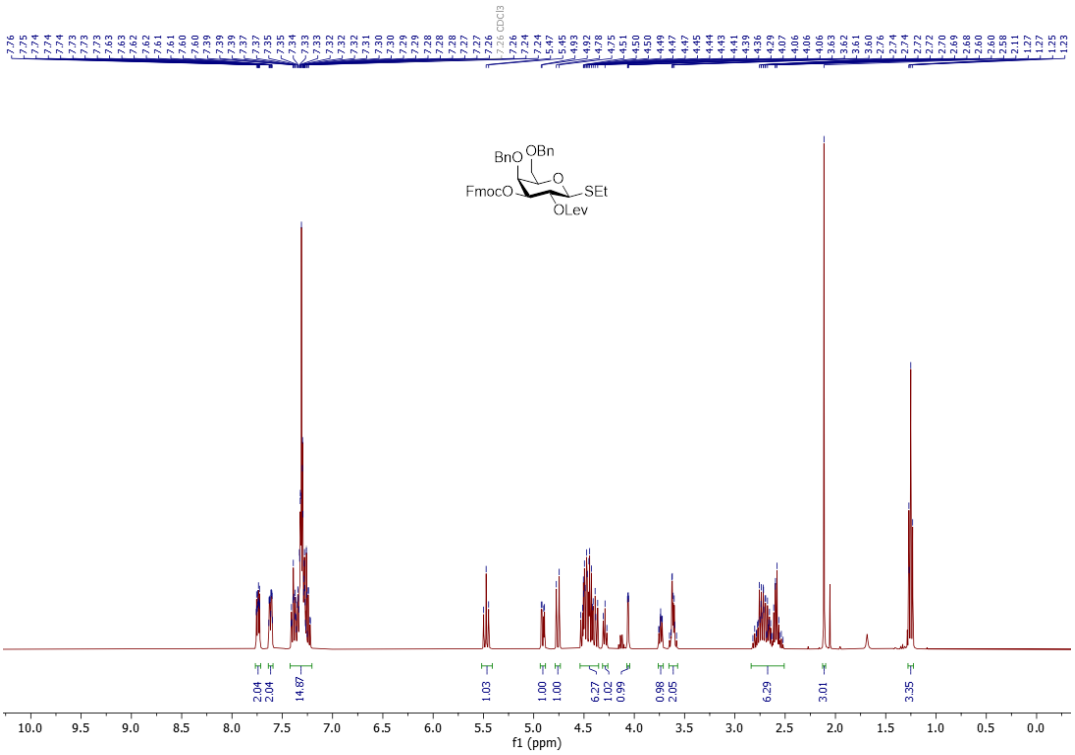
^{1}H, ^{1}H COSY of **55**



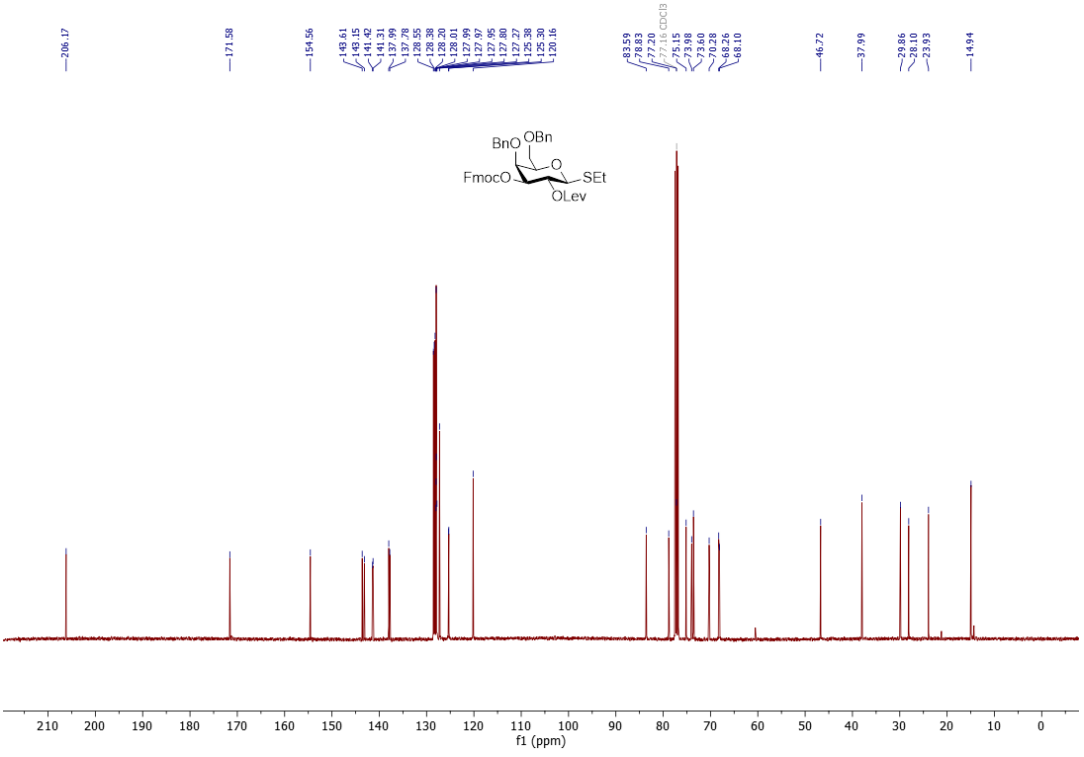
^{13}C, ^{1}H HSQC of **55**



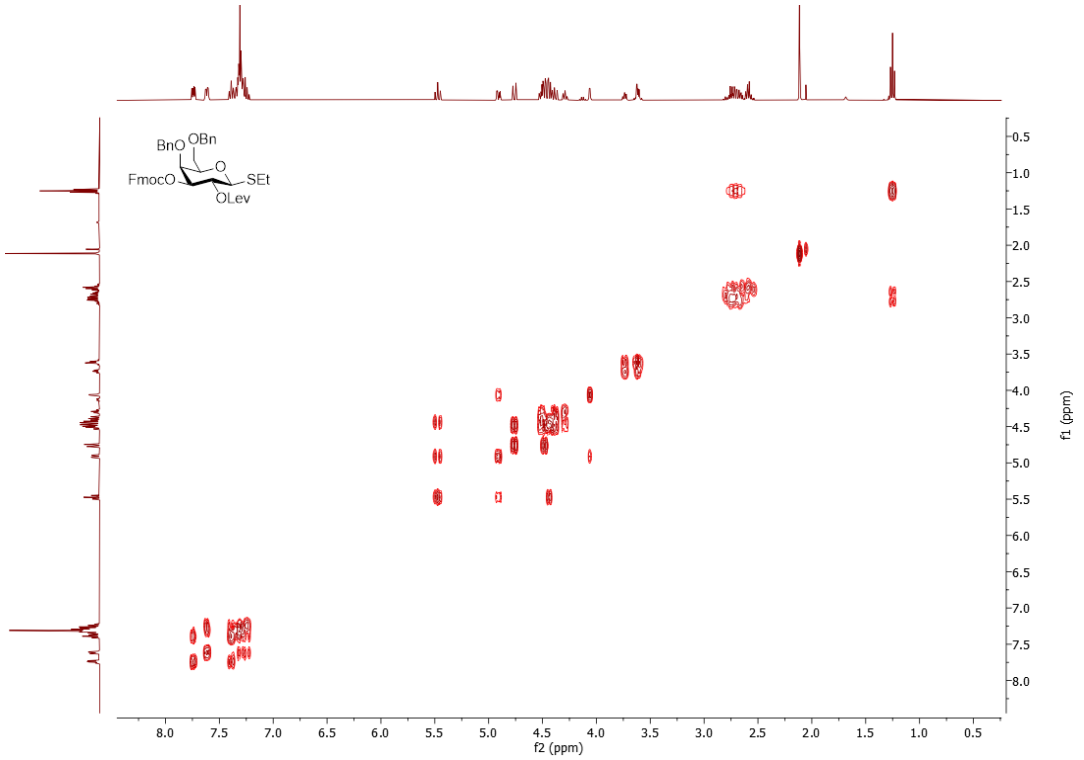
^{1}H NMR (400 MHz, $CDCl_3$) of **41**



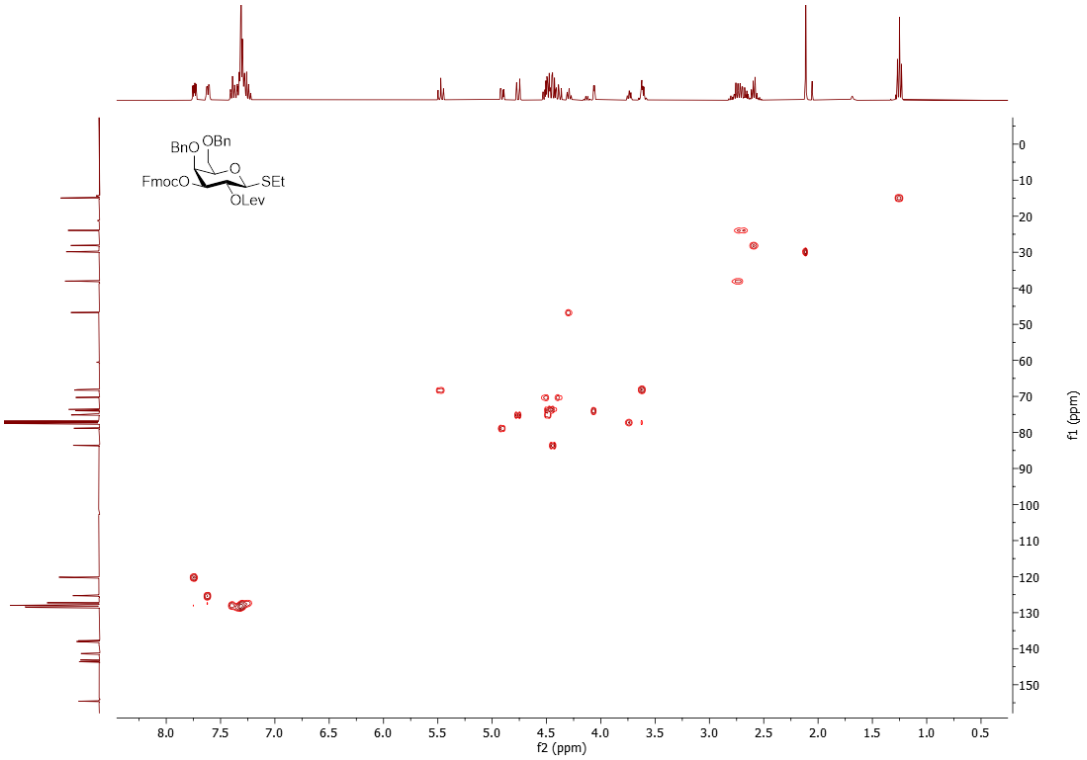
^{13}C NMR (101 MHz, $CDCl_3$) of **41**



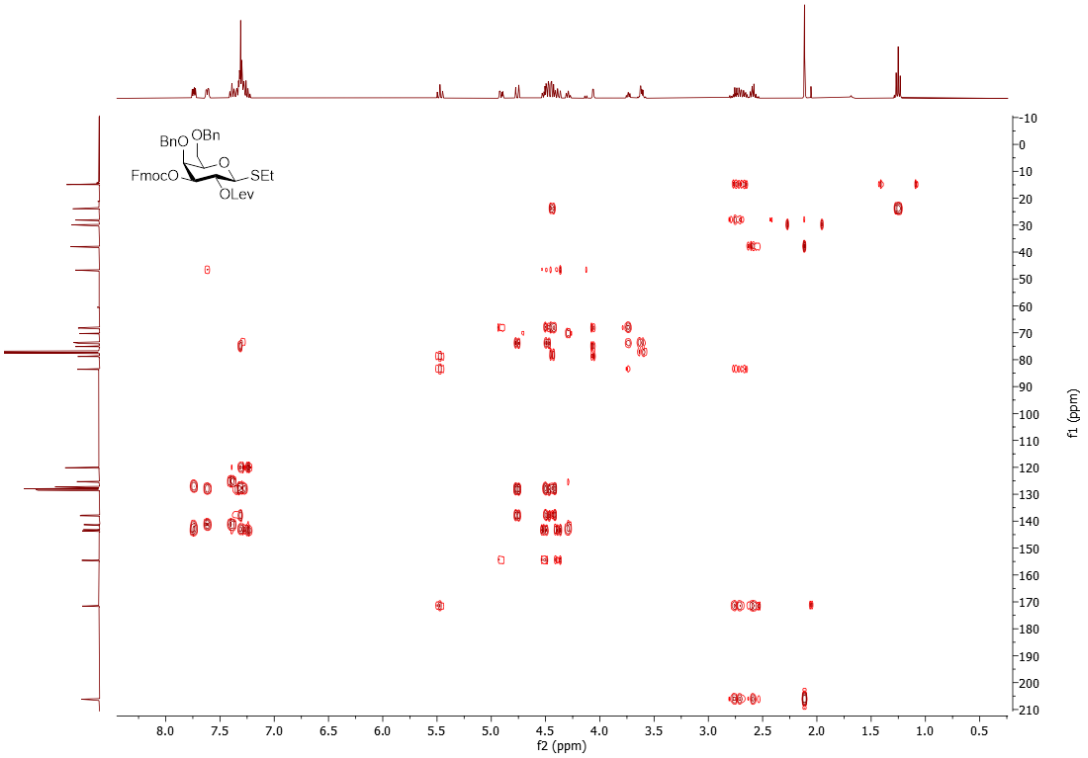
^{1}H, ^{1}H COSY of **41**



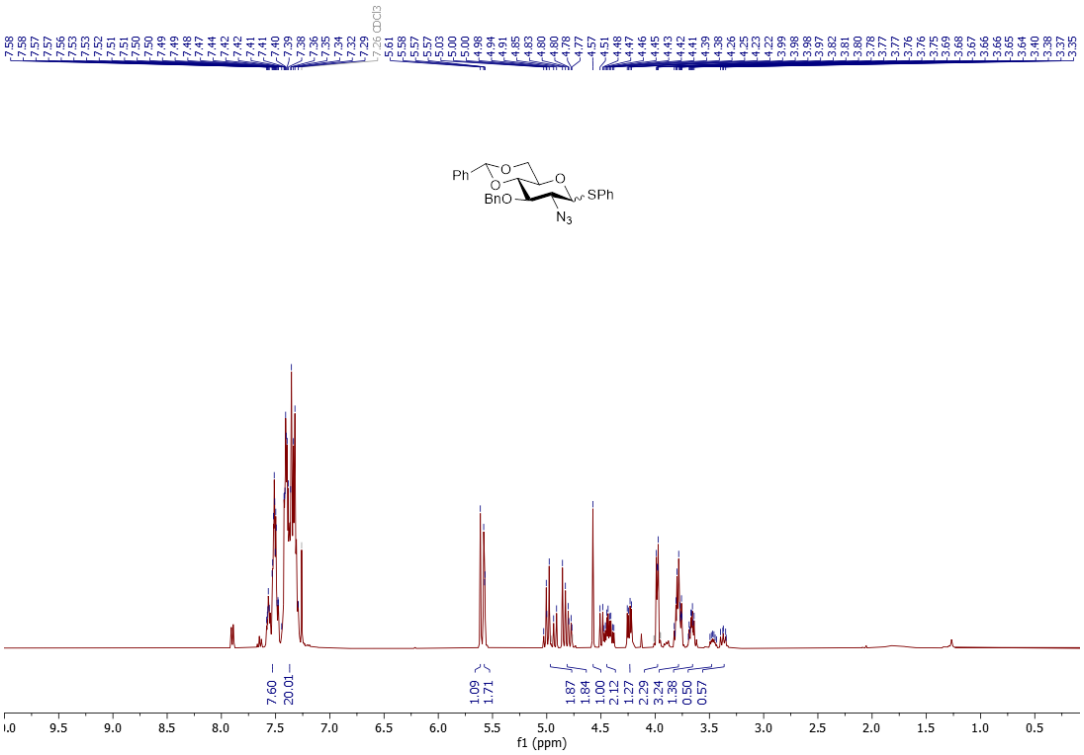
^{13}C, ^{1}H HSQC of **41**



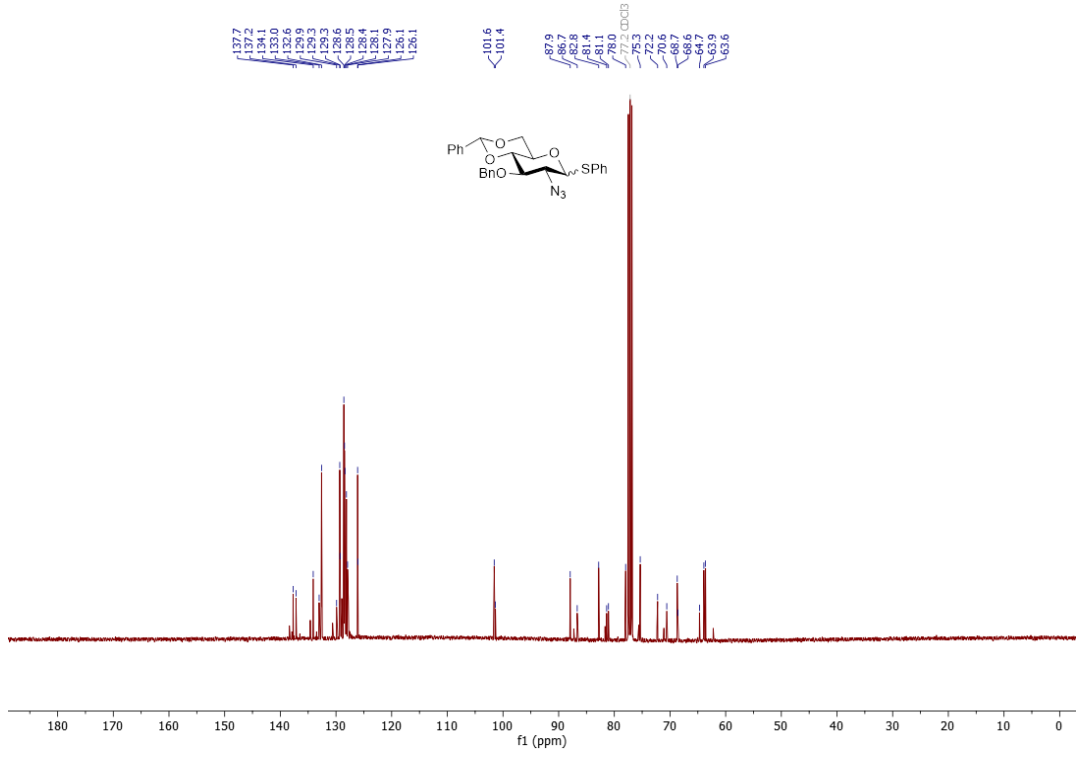
^{13}C, ^{1}H HMBC of **41**



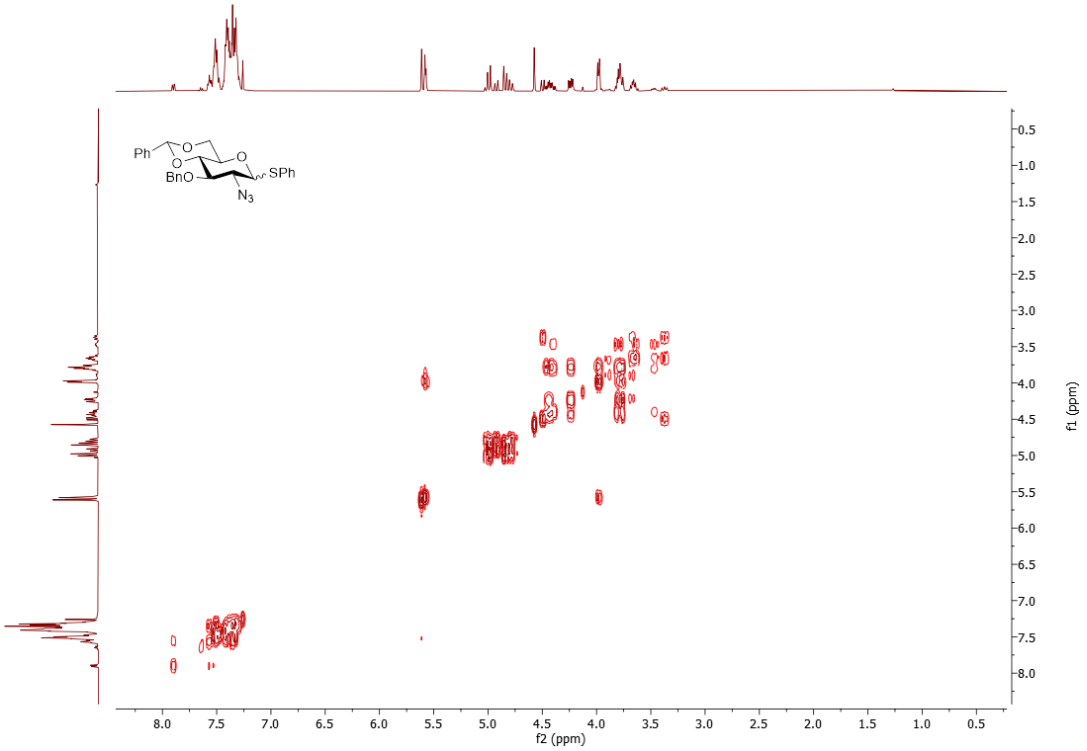
^{1}H NMR (400 MHz, $CDCl_3$) of **66**



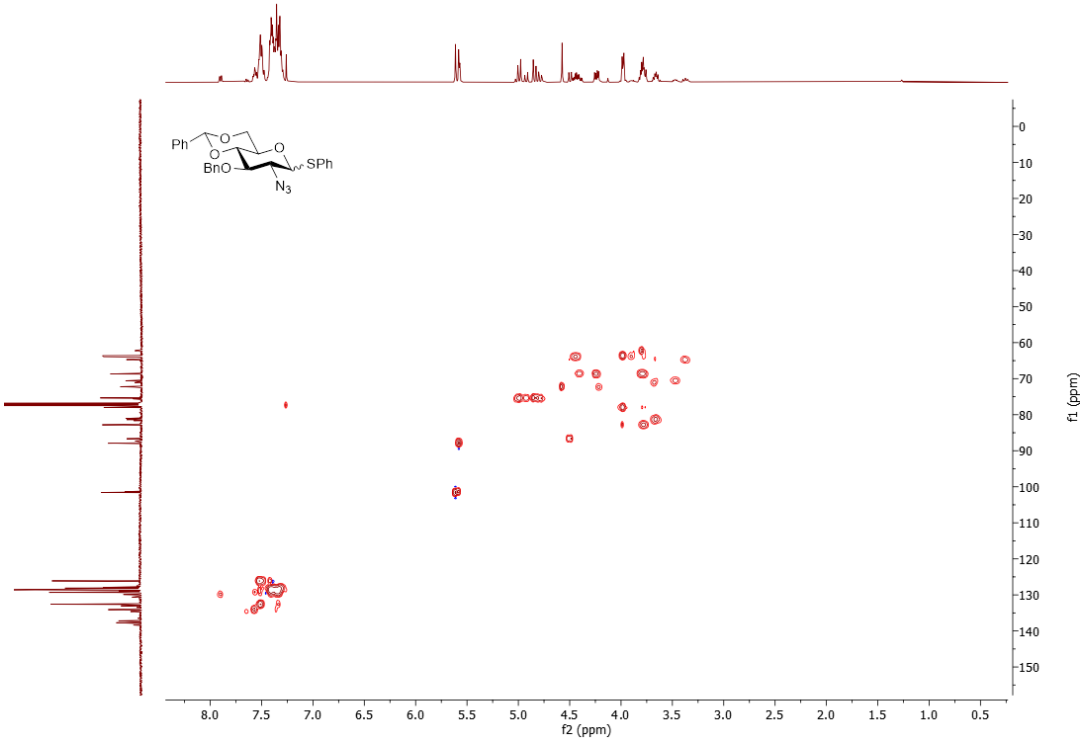
^{13}C NMR (101 MHz, $CDCl_3$) of **66**



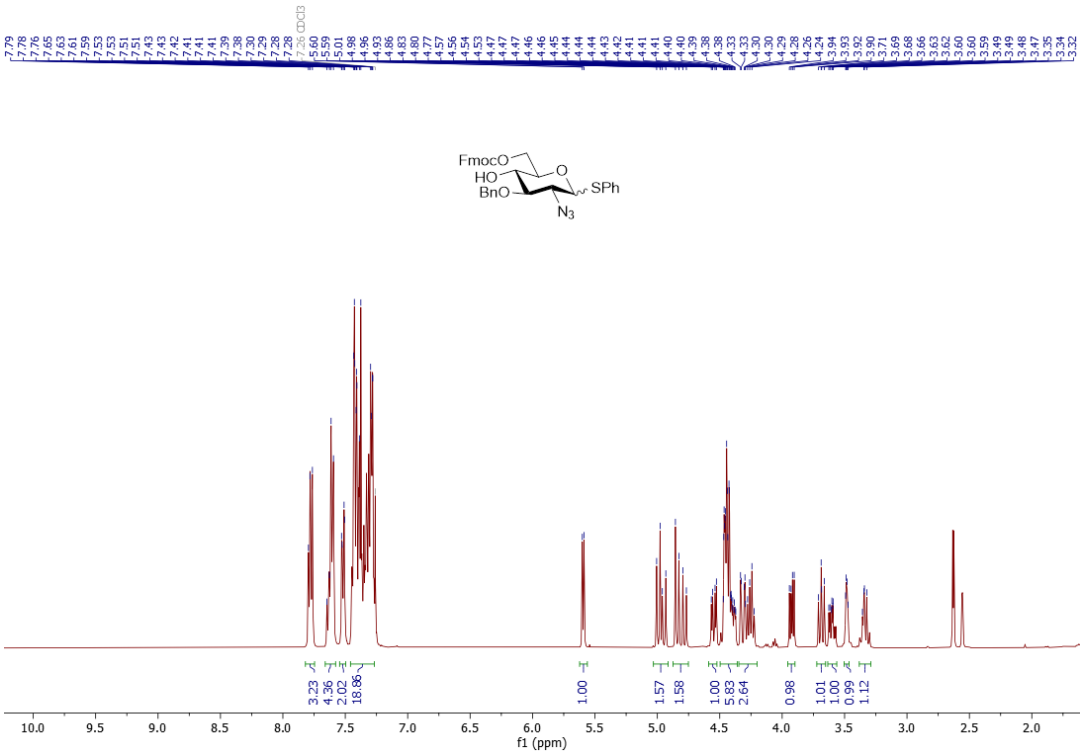
^{1}H, ^{1}H COSY of **66**



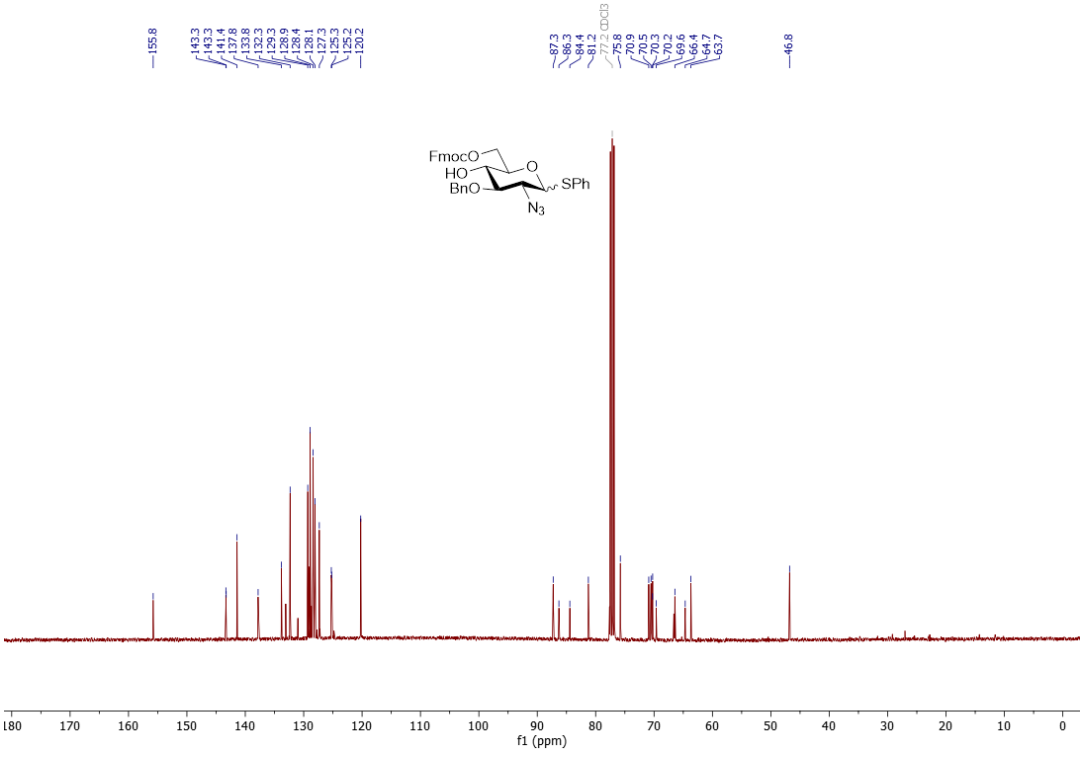
^{13}C, ^{1}H HSQC of **66**



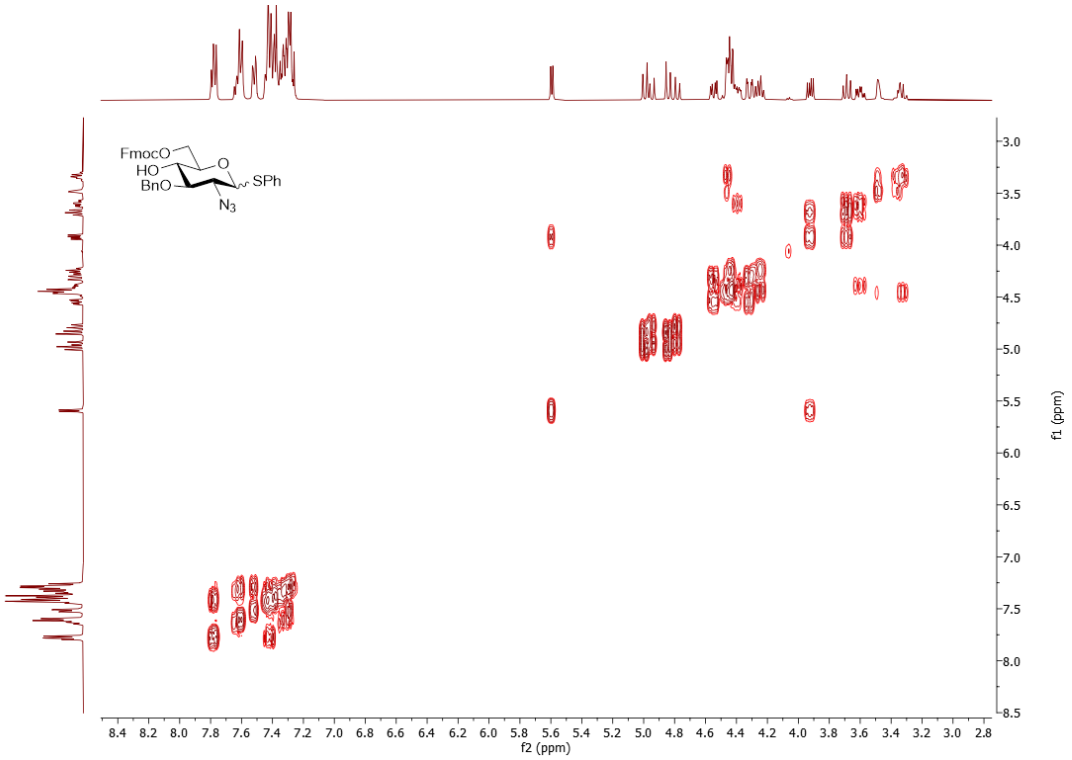
^{1}H NMR (400 MHz, $CDCl_3$) of **67**



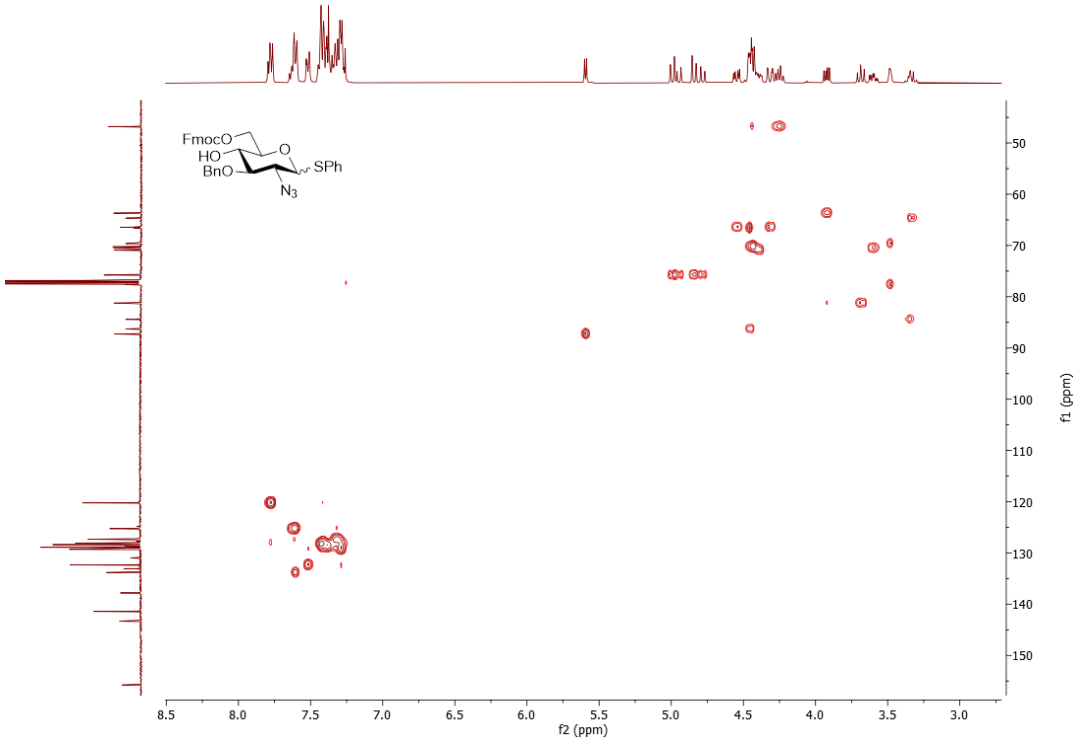
^{13}C NMR (101 MHz, $CDCl_3$) of **67**



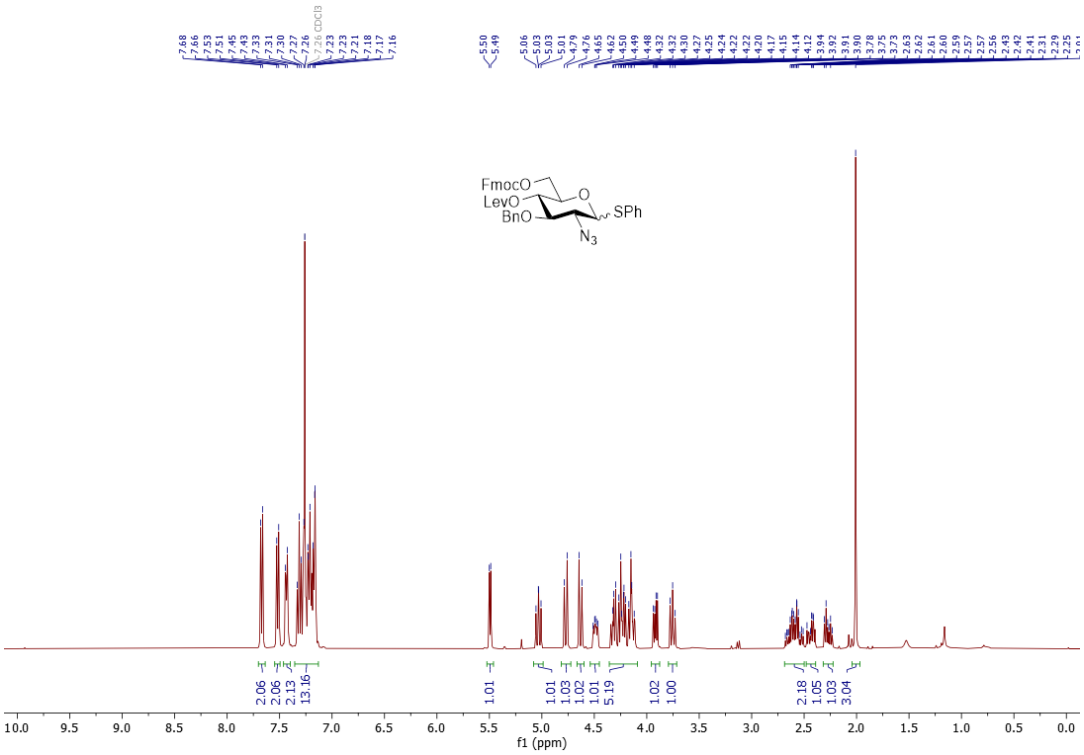
^{1}H, ^{1}H COSY of **67**



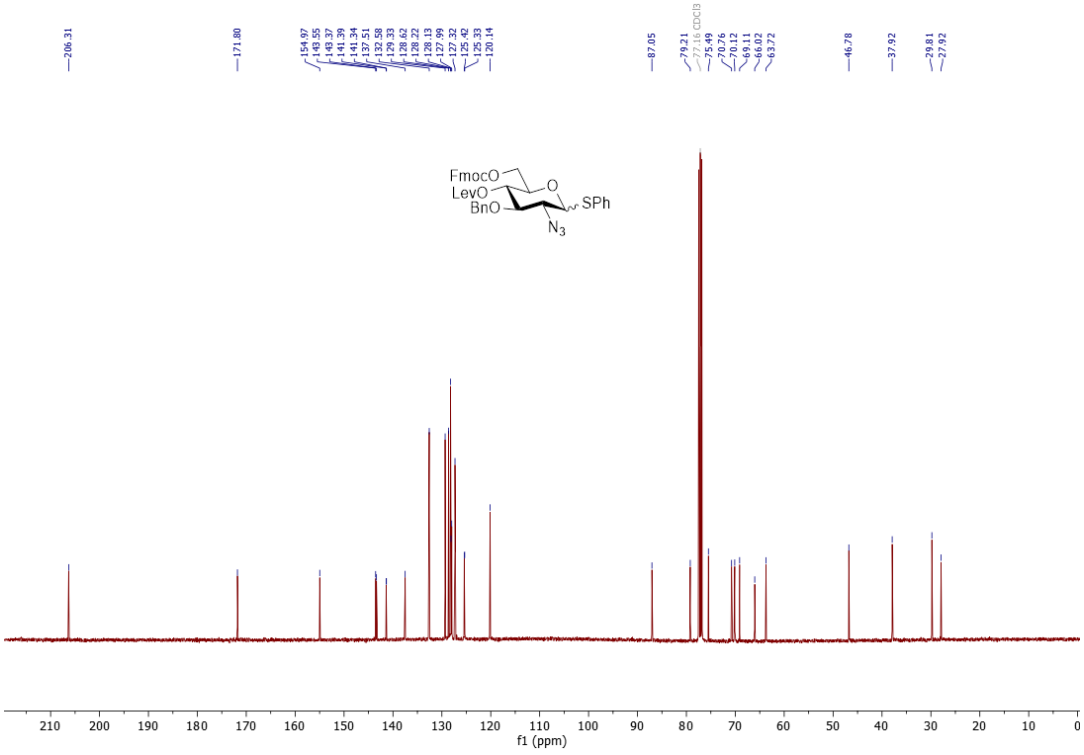
^{13}C, ^{1}H HSQC of **67**



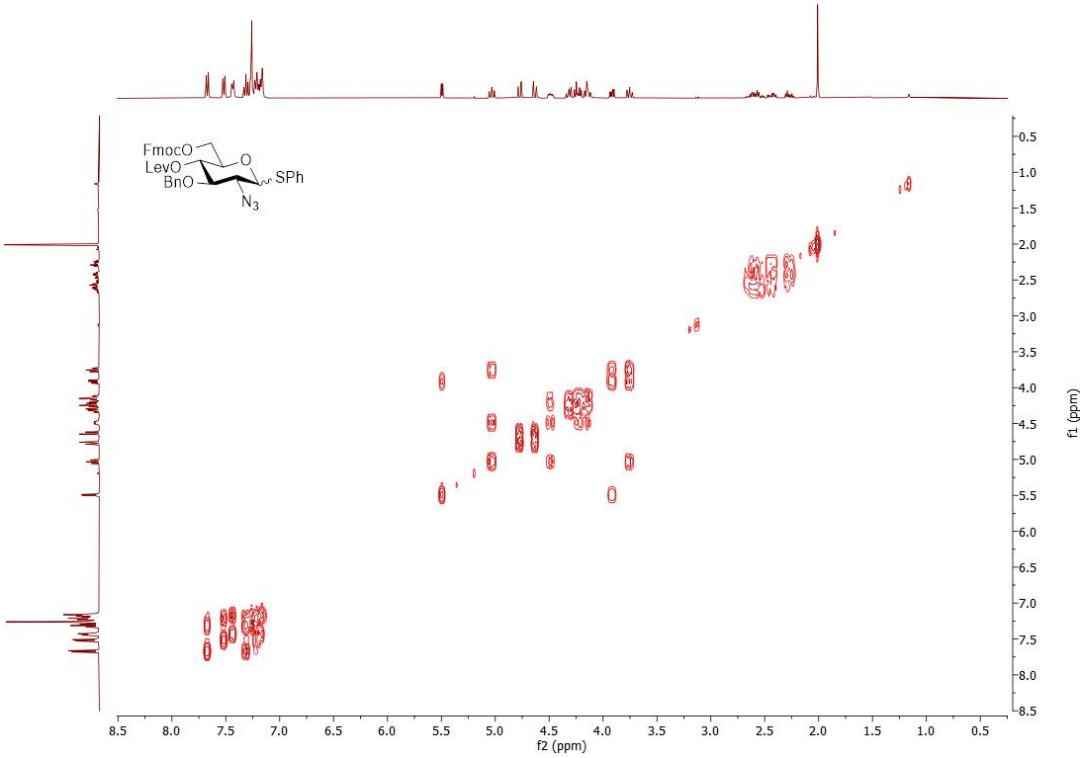
^{1}H NMR (400 MHz, $CDCl_3$) of **44**



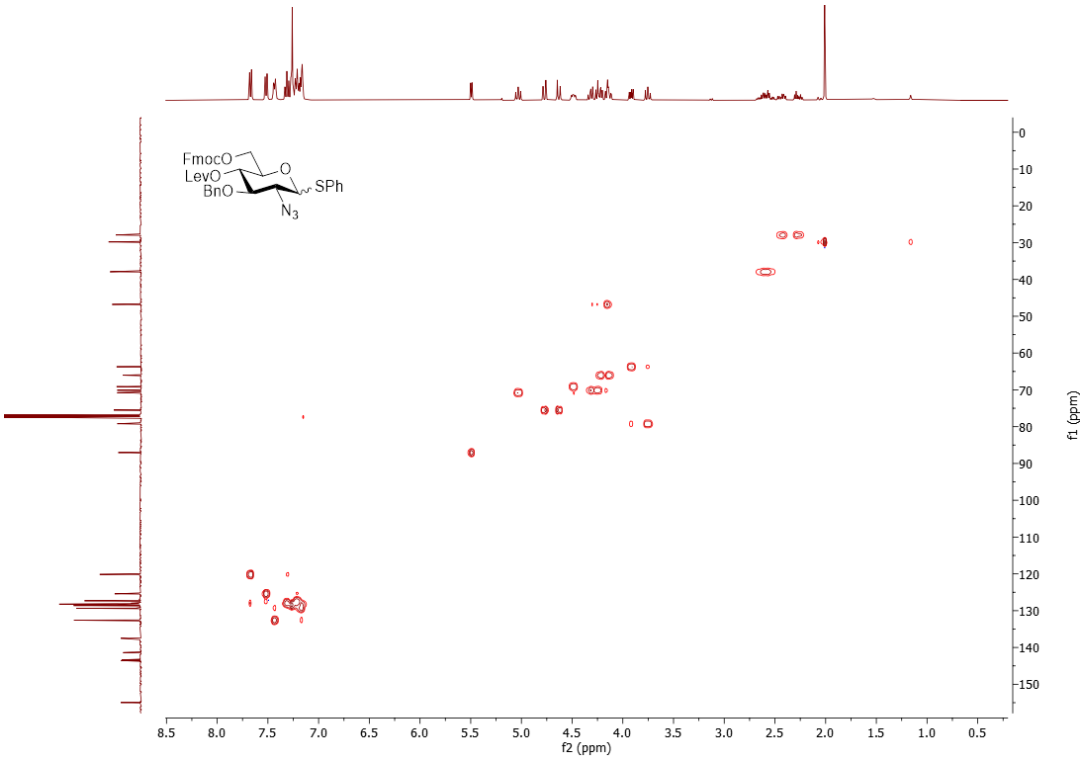
^{13}C NMR (101 MHz, $CDCl_3$) of **44**



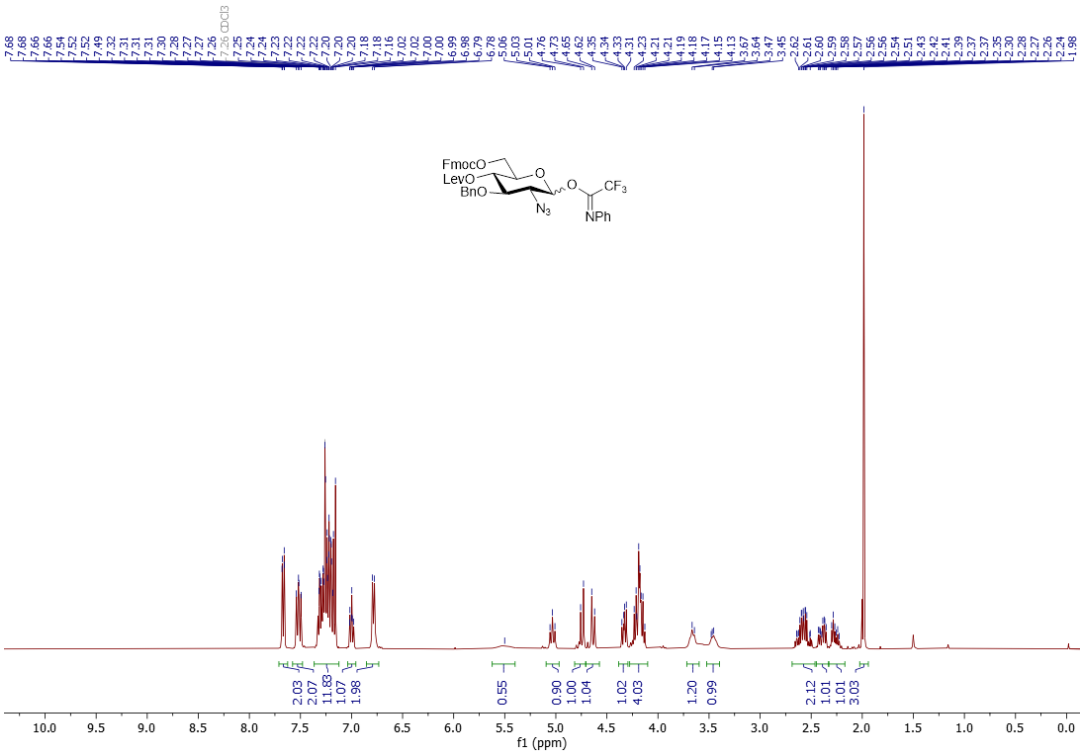
^{1}H, ^{1}H COSY of **44**



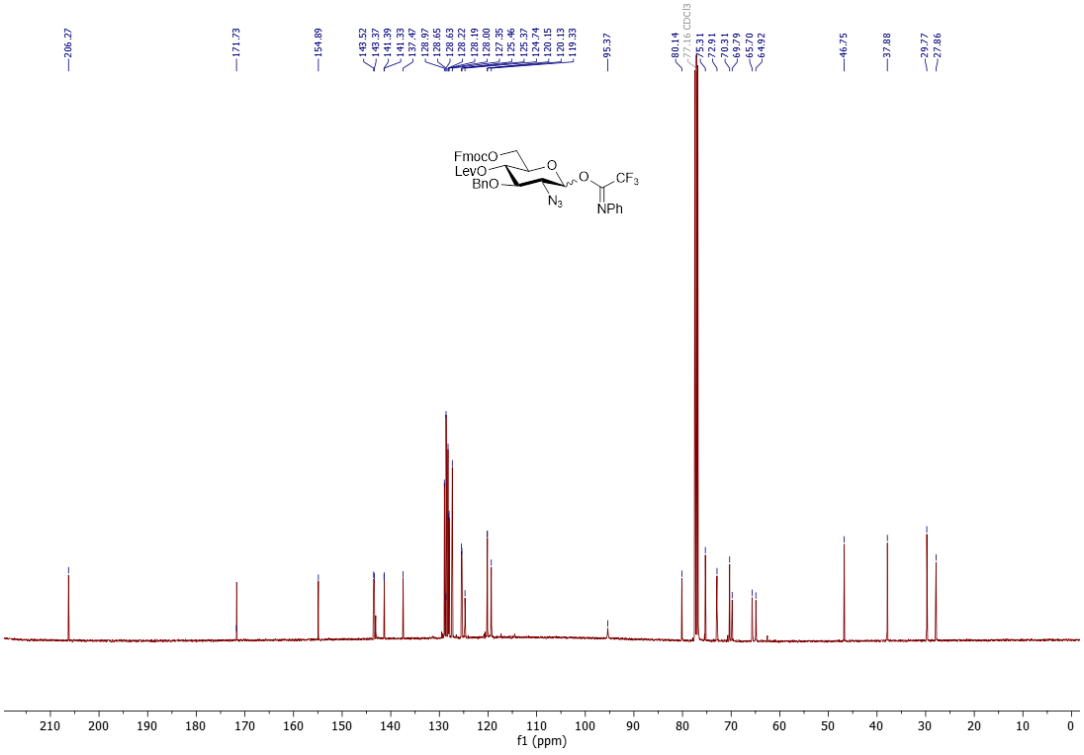
^{13}C, ^{1}H HSQC of **44**



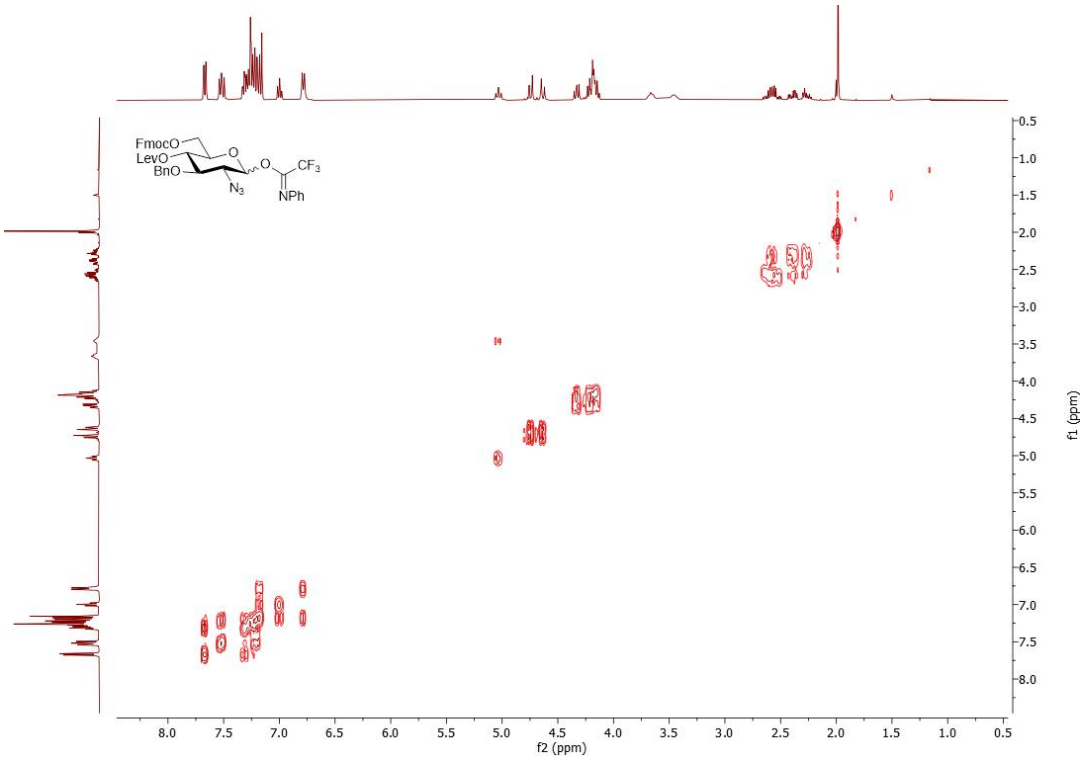
^{1}H NMR (400 MHz, $CDCl_3$) of **45**



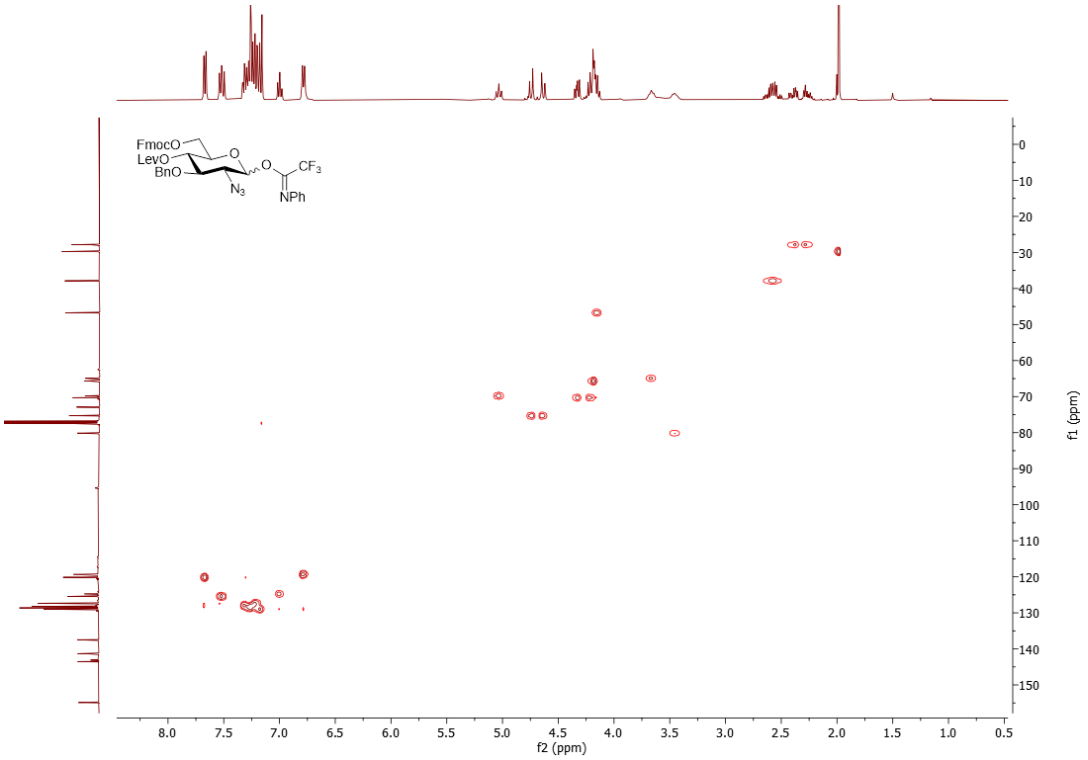
^{13}C NMR (101 MHz, $CDCl_3$) of **45**



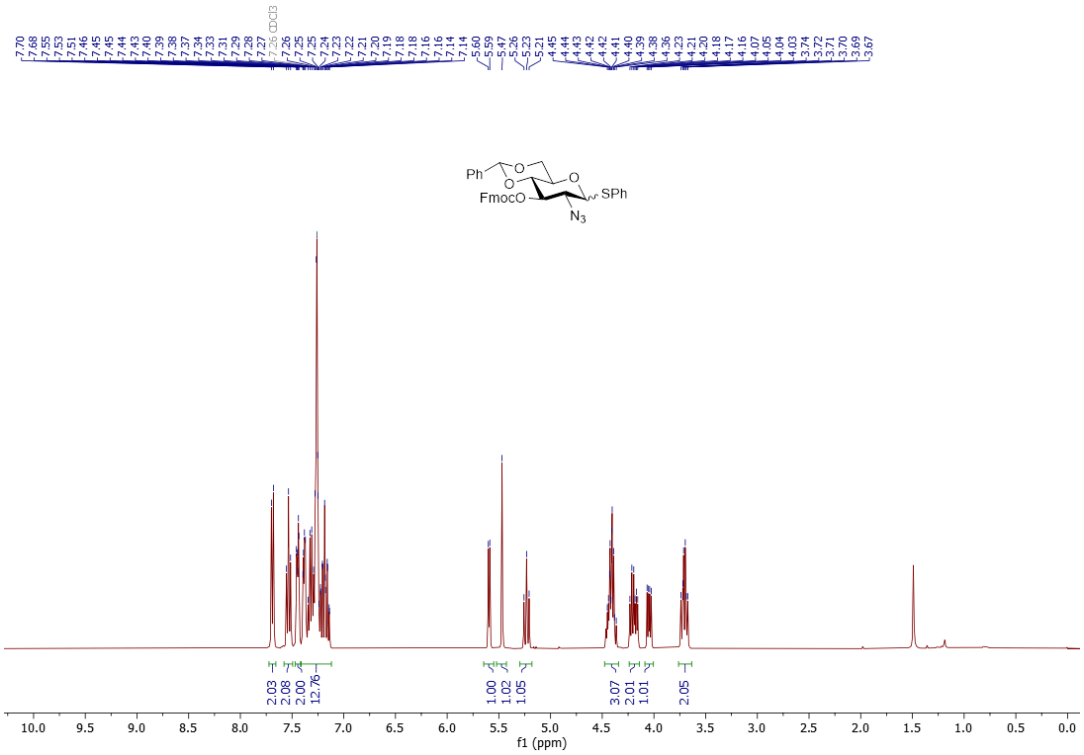
^{1}H, ^{1}H COSY of **45**



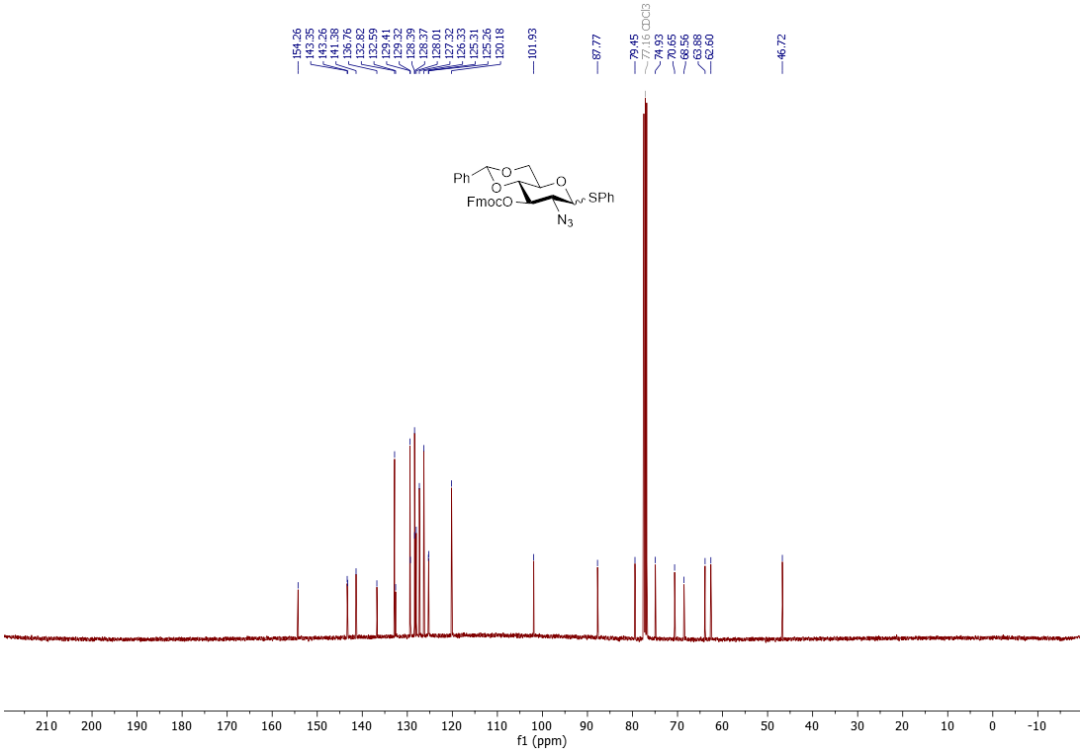
^{13}C, ^{1}H HSQC of **45**



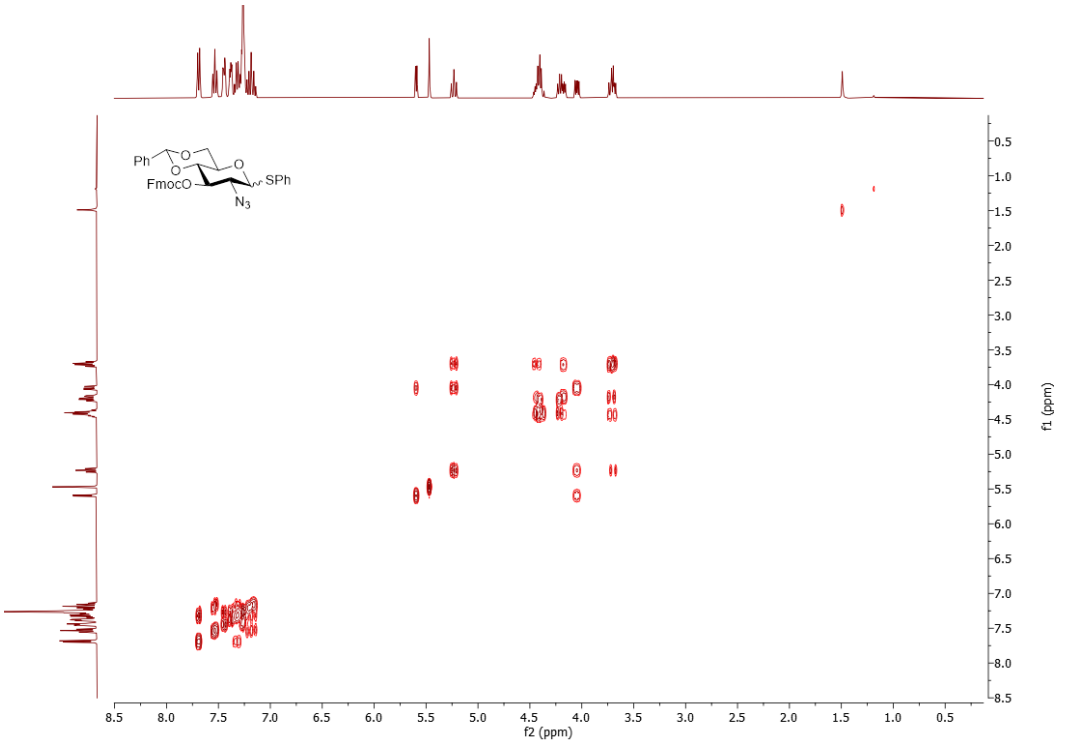
^{1}H NMR (400 MHz, $CDCl_3$) of **68**



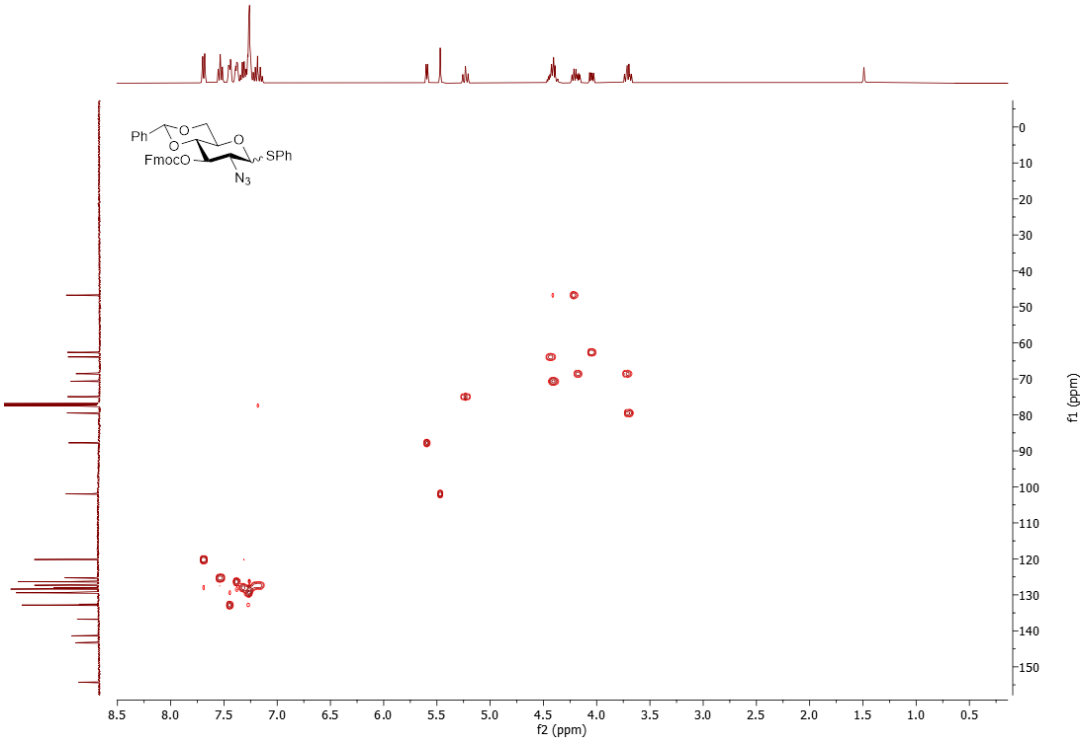
^{13}C NMR (101 MHz, $CDCl_3$) of **68**



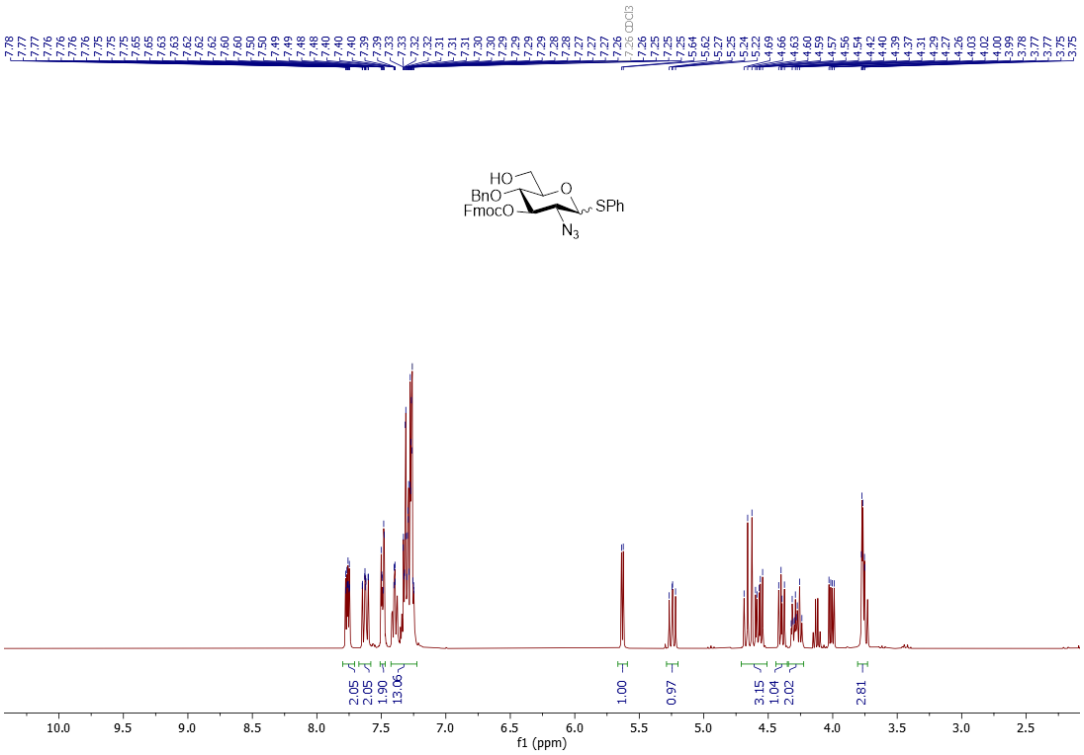
^{1}H, ^{1}H COSY of **68**



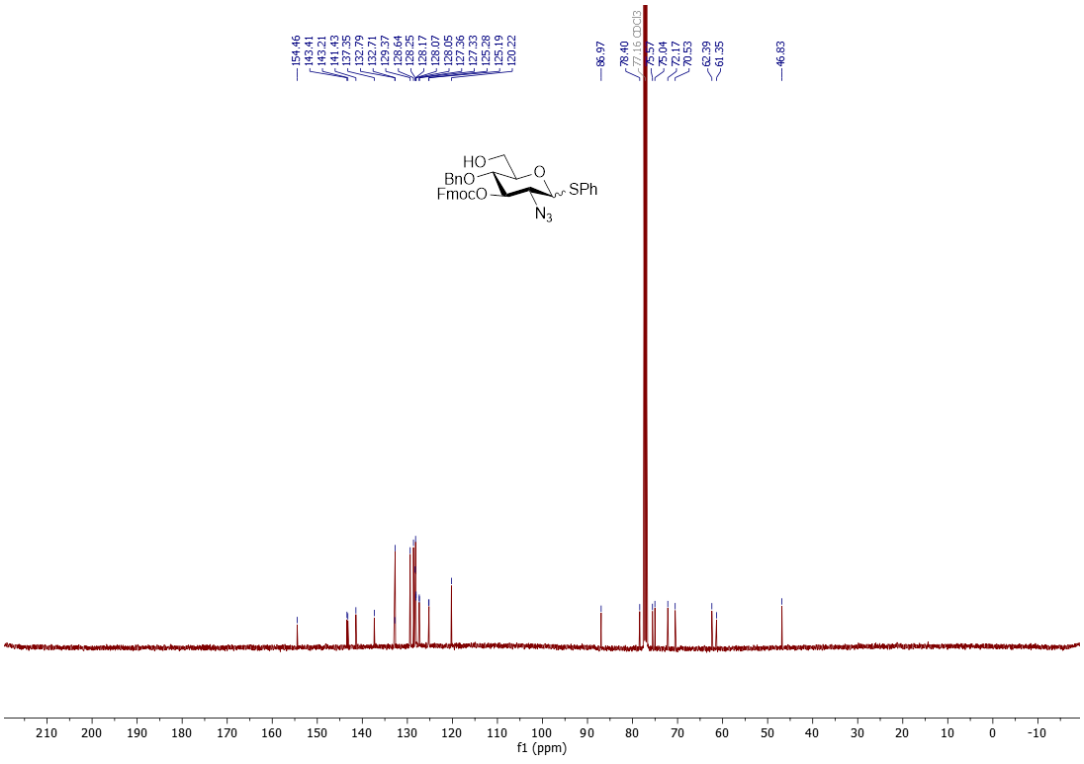
^{13}C, ^{1}H HSQC of **66**



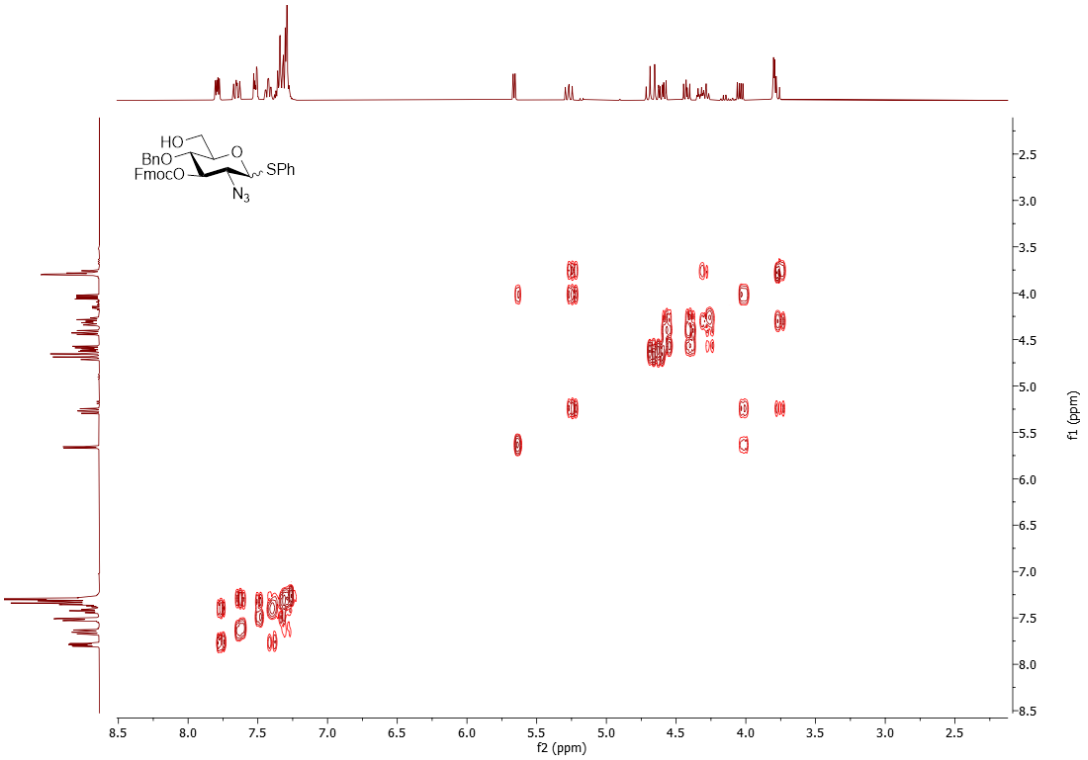
^{1}H NMR (400 MHz, $CDCl_3$) of **69**



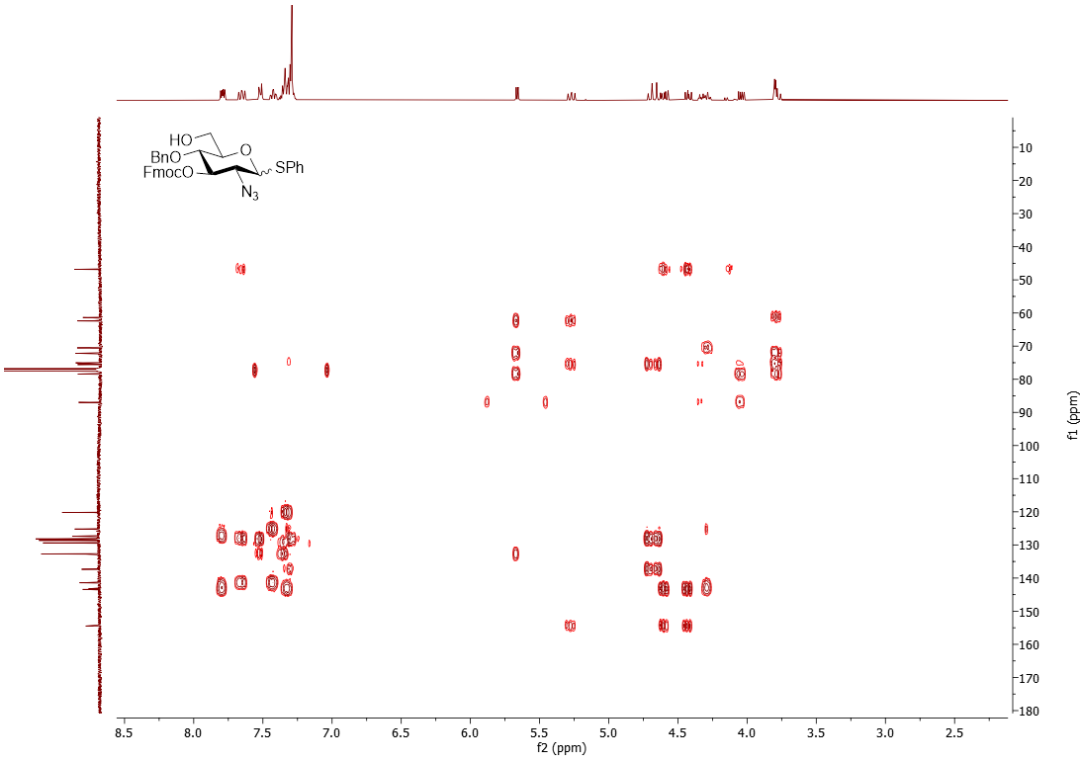
^{13}C NMR (101 MHz, $CDCl_3$) of **69**



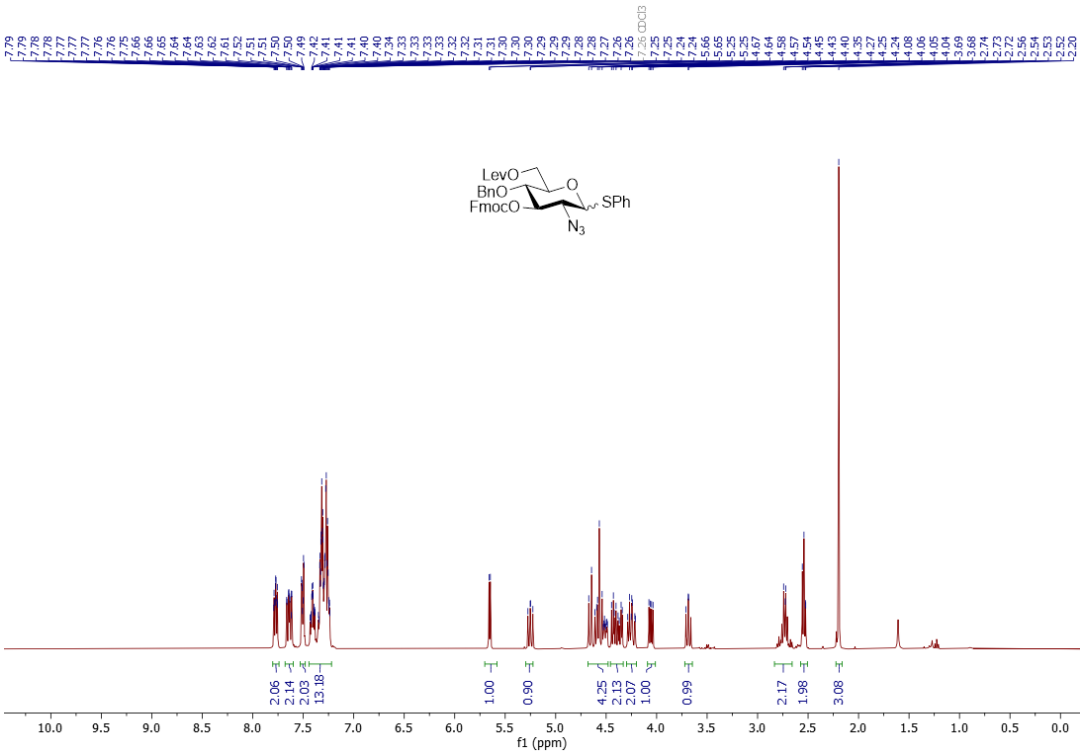
^{1}H, ^{1}H COSY of **69**



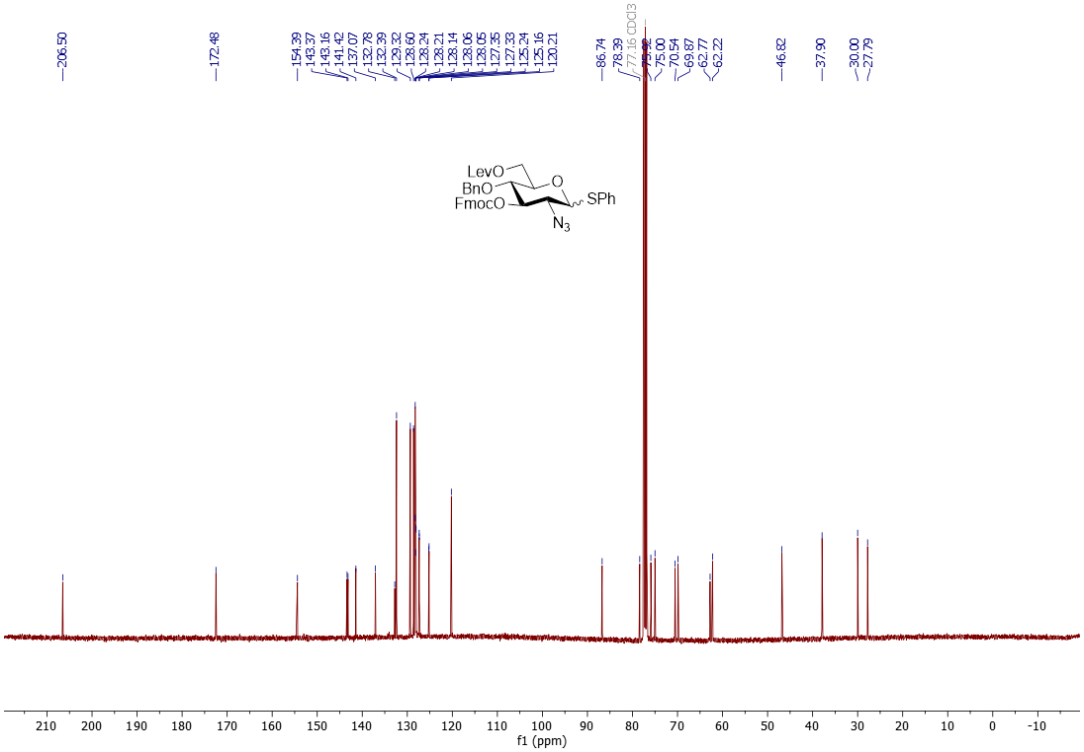
^{13}C, ^{1}H HSQC of **69**



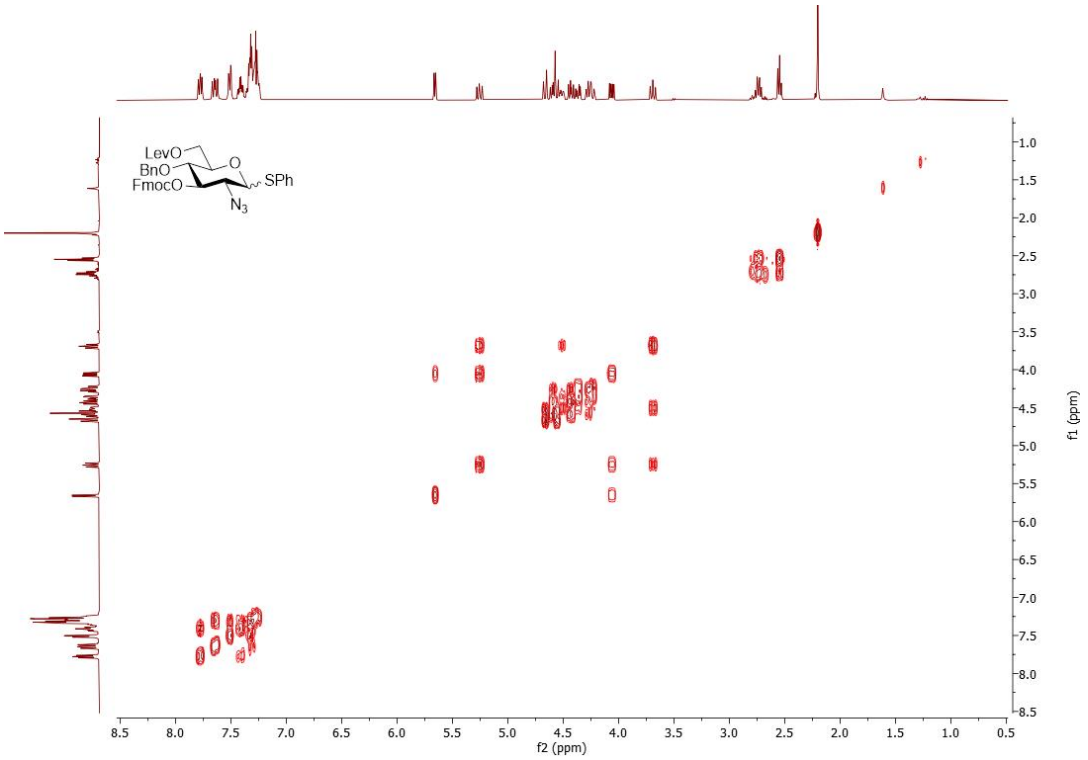
^{1}H NMR (400 MHz, $CDCl_3$) of **47**



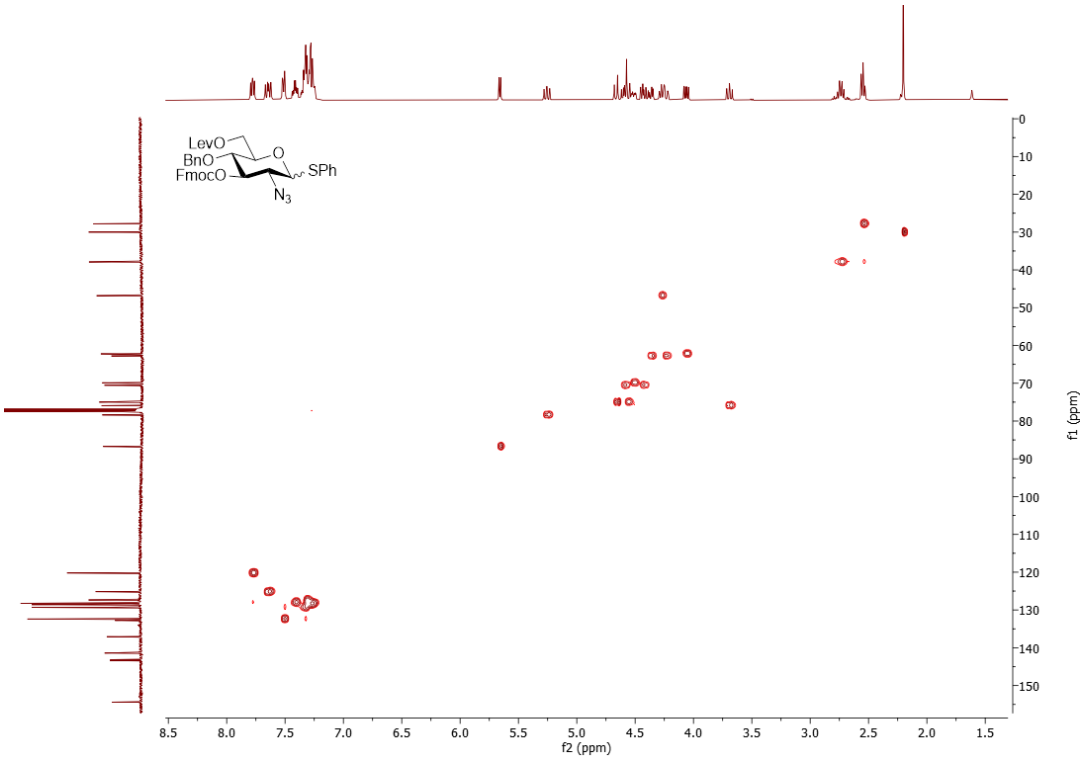
^{13}C NMR (101 MHz, $CDCl_3$) of **47**



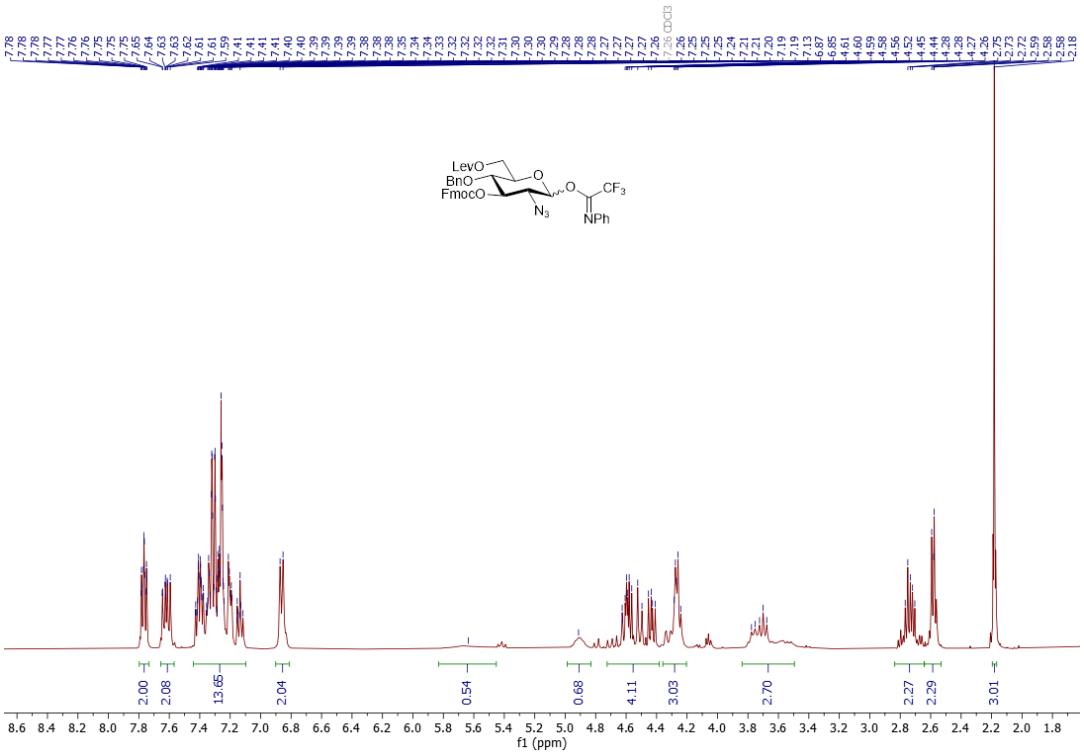
^{1}H, ^{1}H COSY of **47**



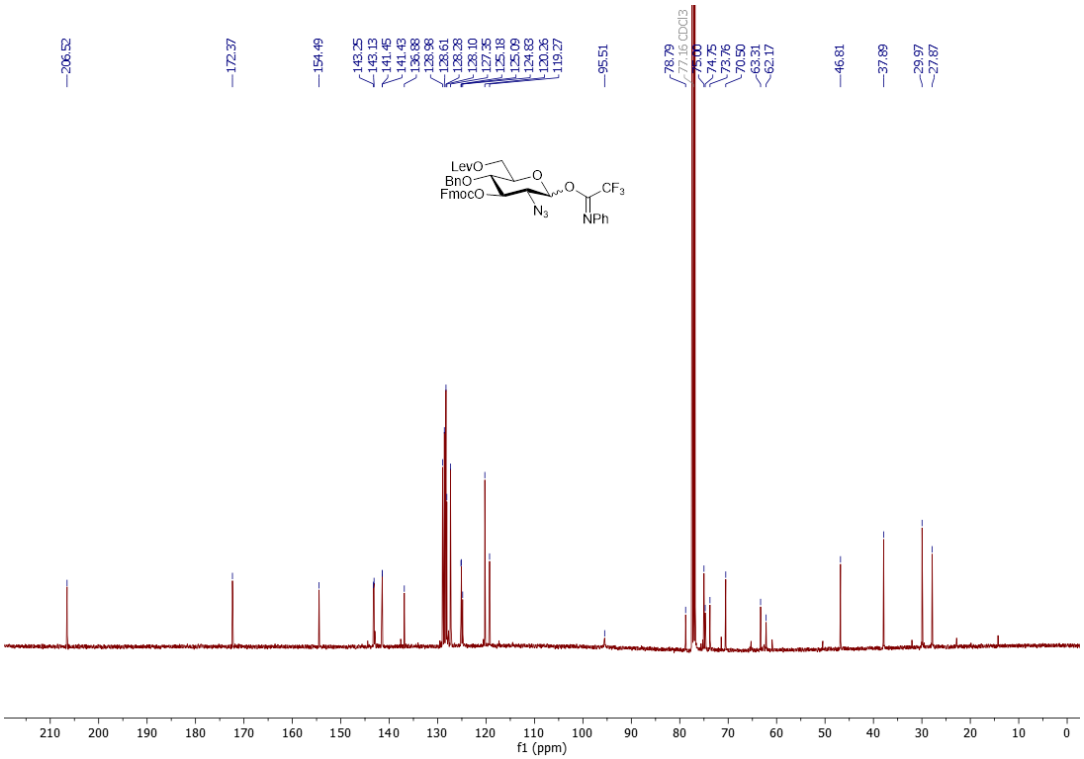
^{13}C, ^{1}H HSQC of **47**



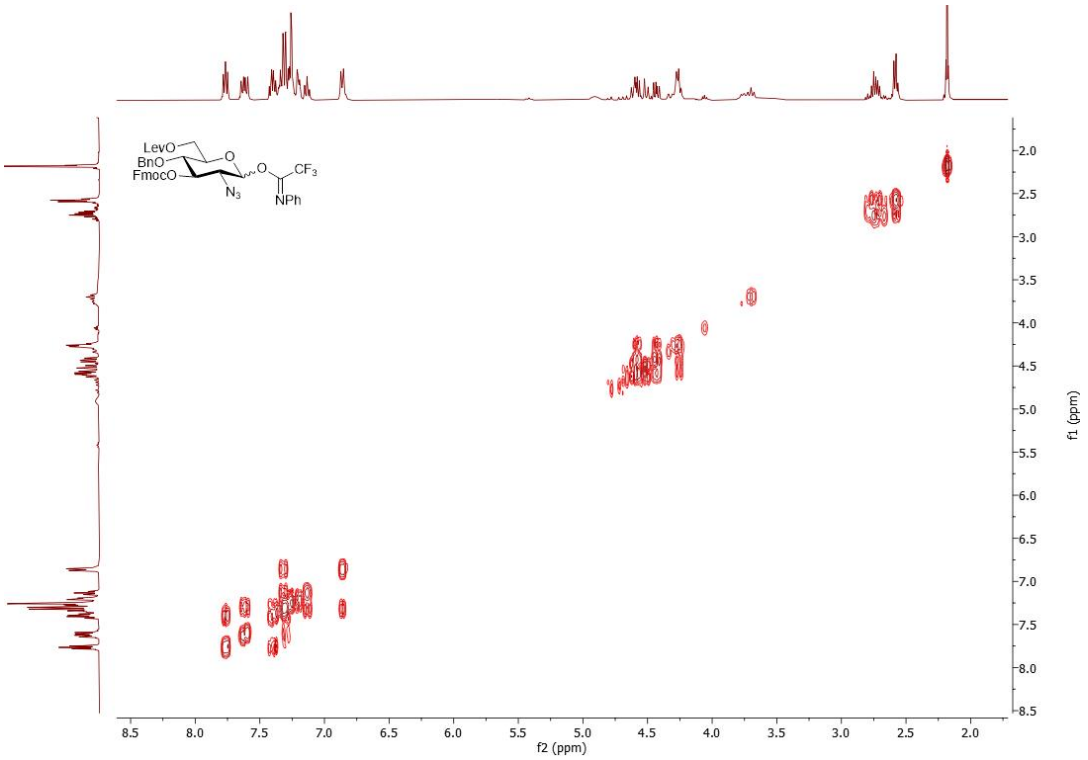
^{1}H NMR (400 MHz, $CDCl_3$) of **48**



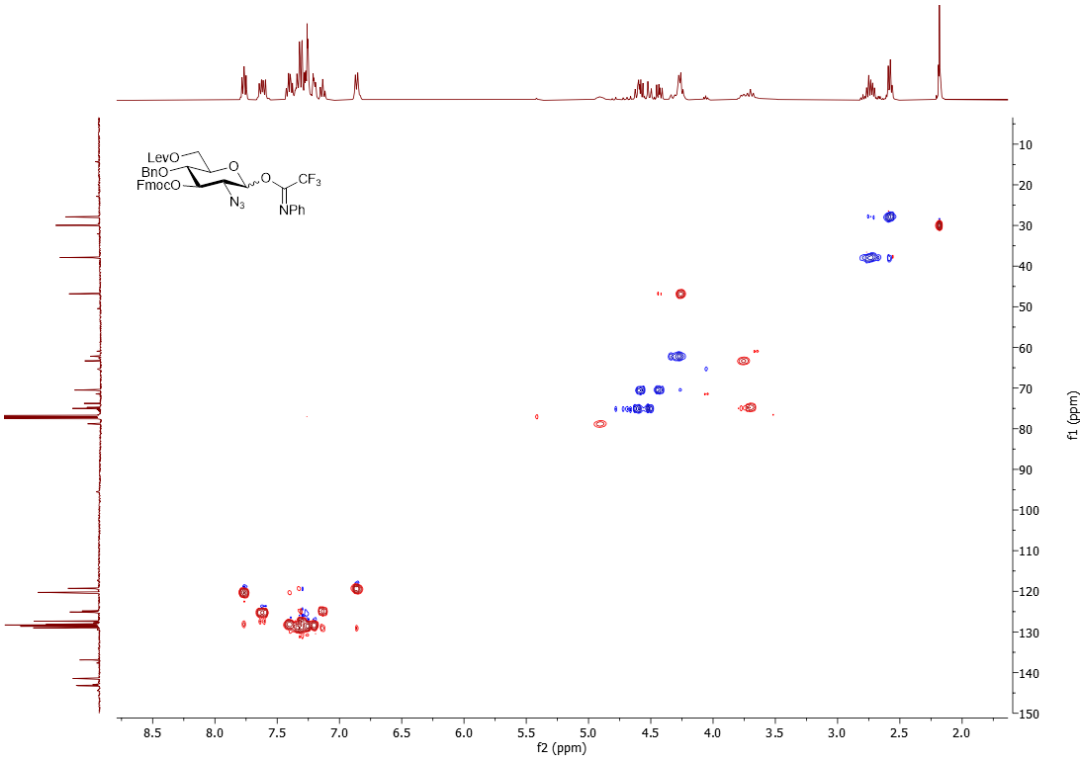
^{13}C NMR (101 MHz, $CDCl_3$) of **48**



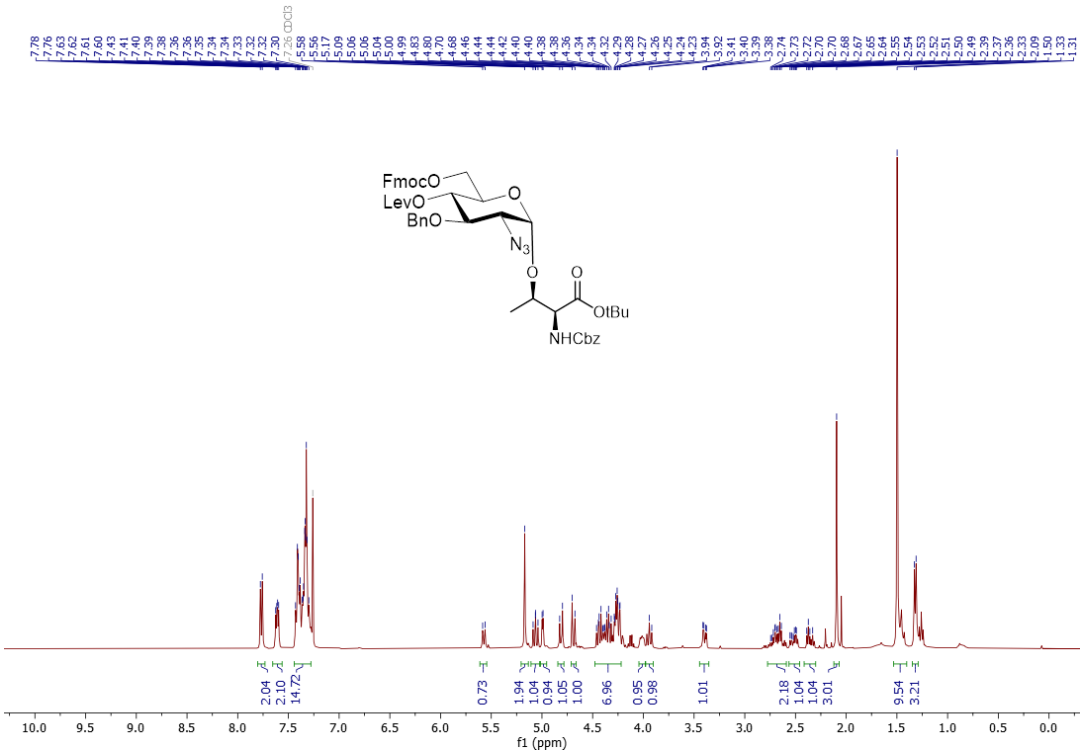
^{1}H, ^{1}H COSY of **48**



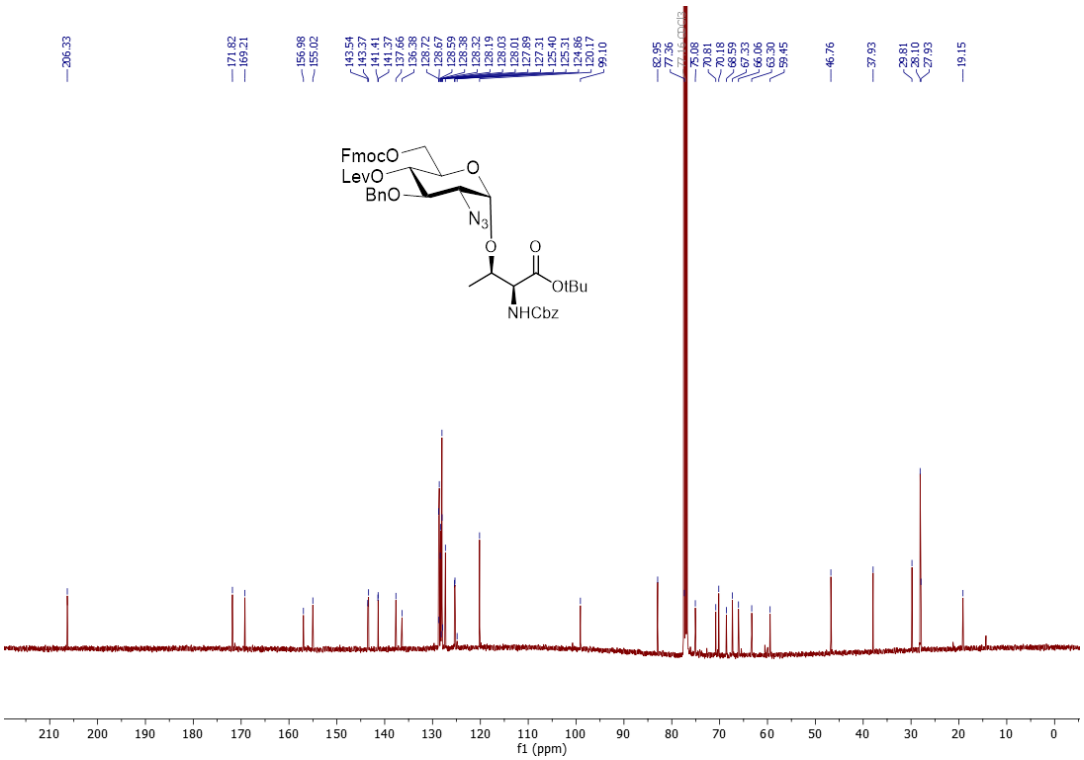
^{13}C, ^{1}H HSQC of **48**



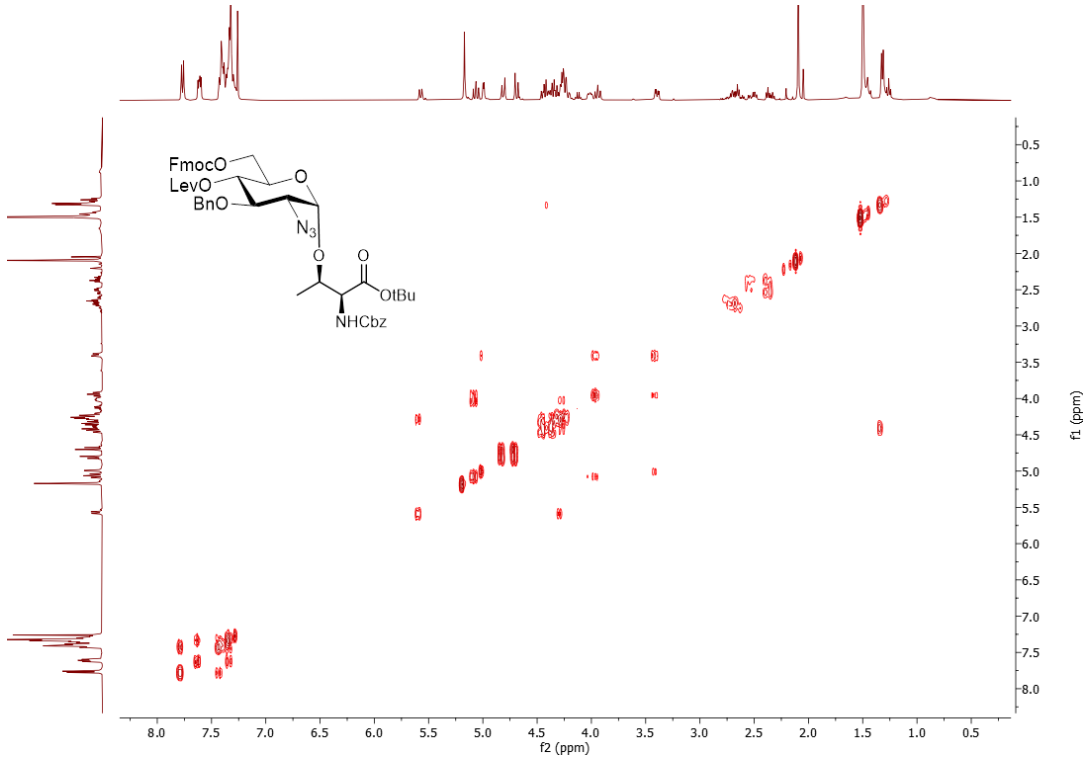
^{1}H NMR (400 MHz, $CDCl_3$) of **70**



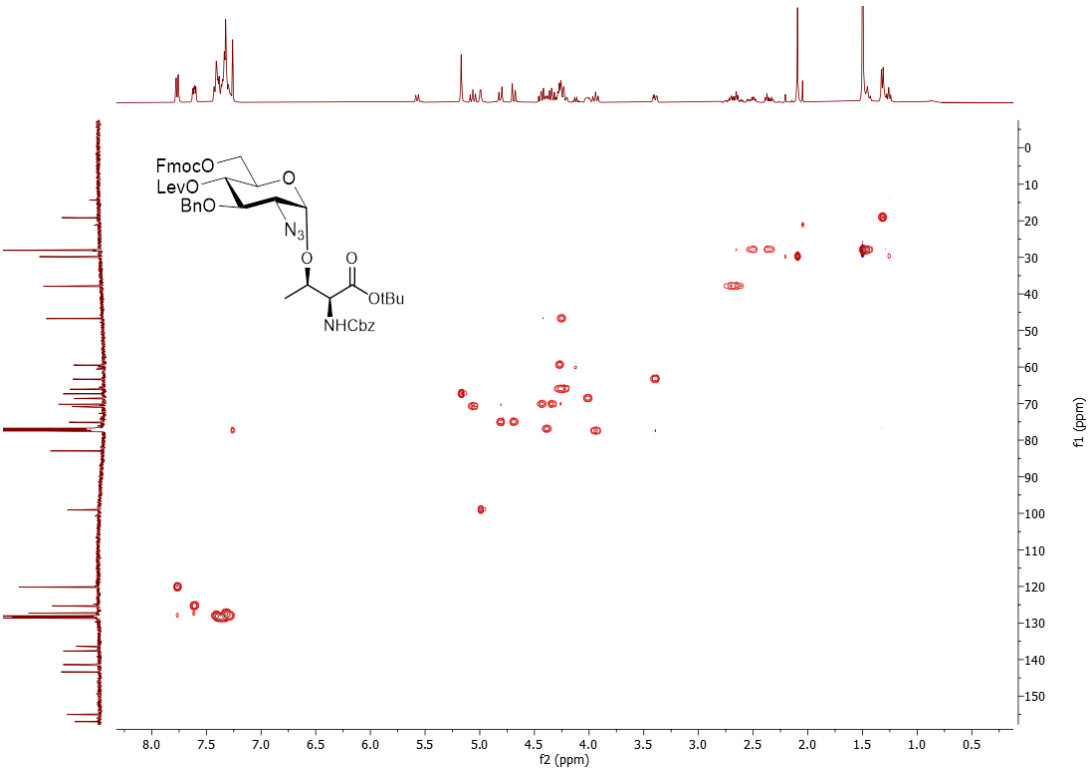
^{13}C NMR (101 MHz, $CDCl_3$) of **70**



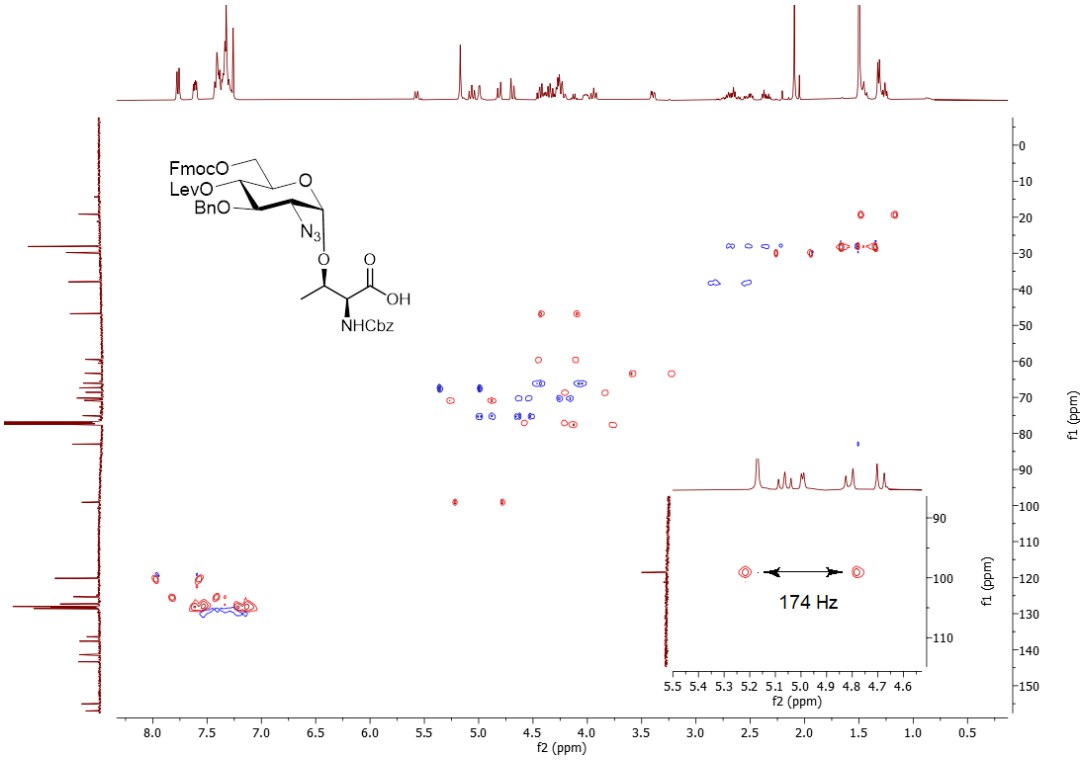
^{1}H, ^{1}H COSY of **70**



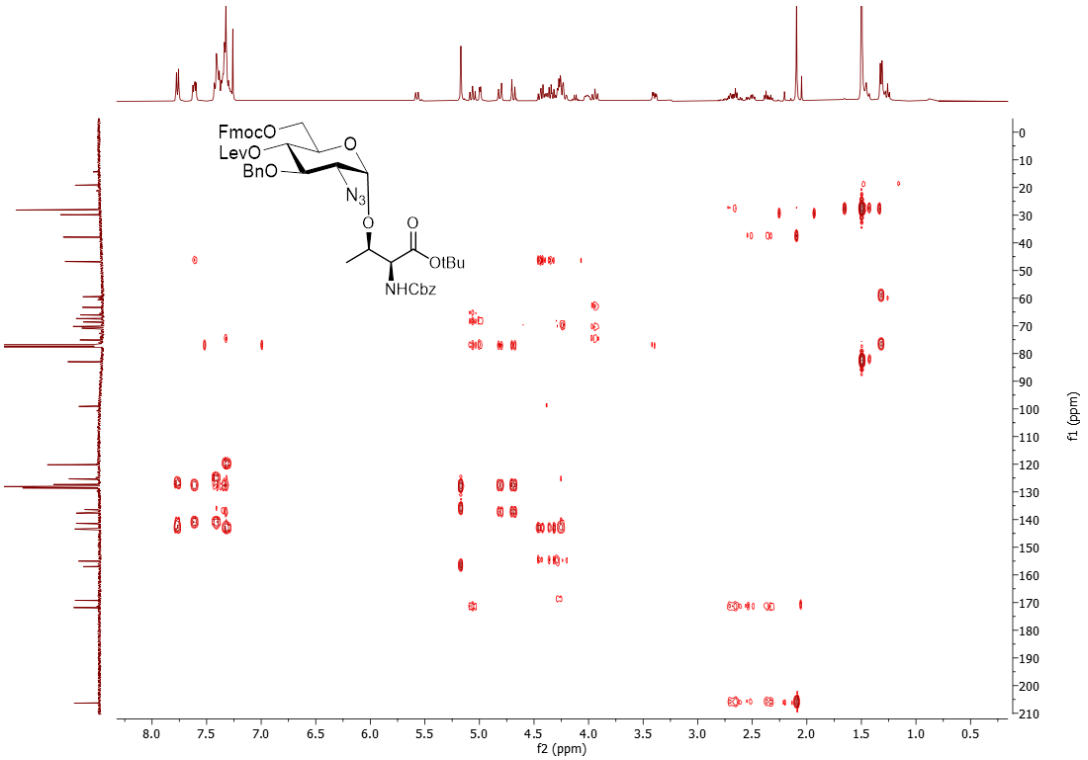
^{13}C, ^{1}H HSQC of **70**



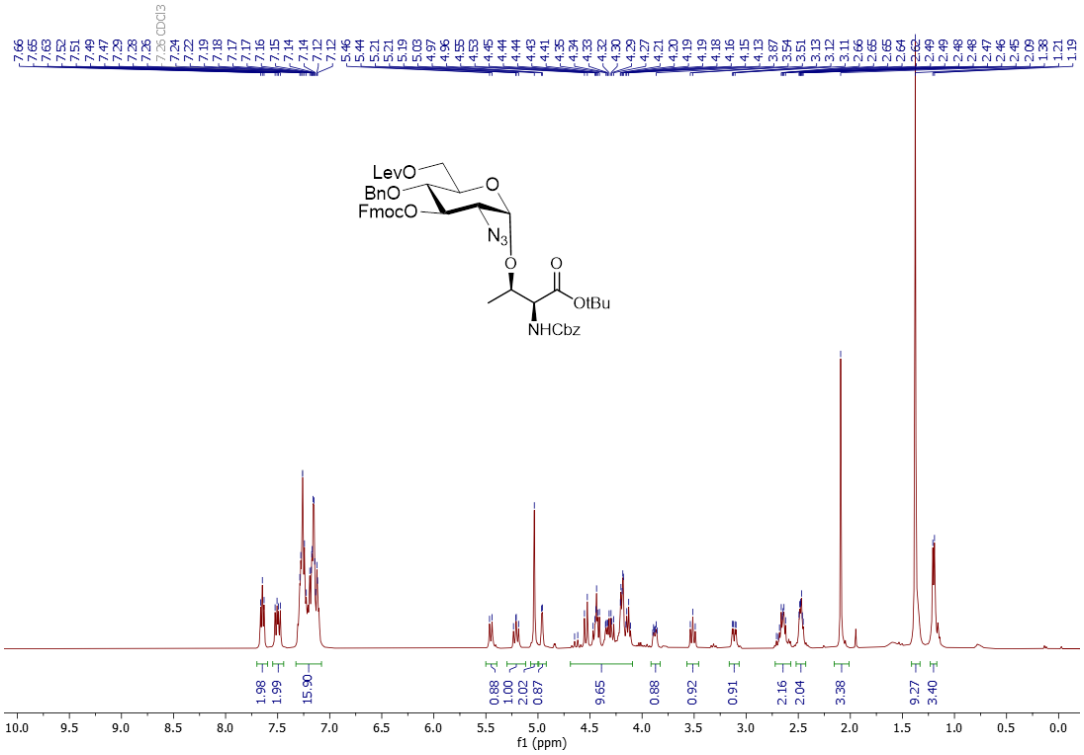
Coupled ^{13}C, ^{1}H HSQC of **70**



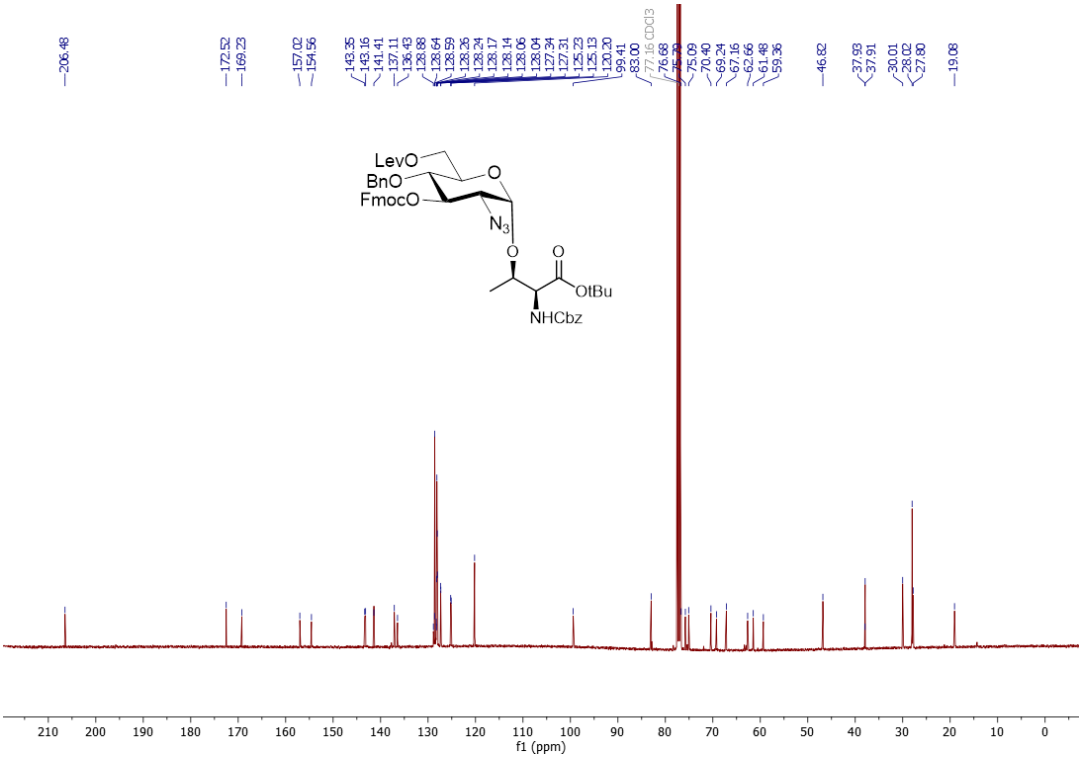
^{13}C, ^{1}H HMBC of **70**



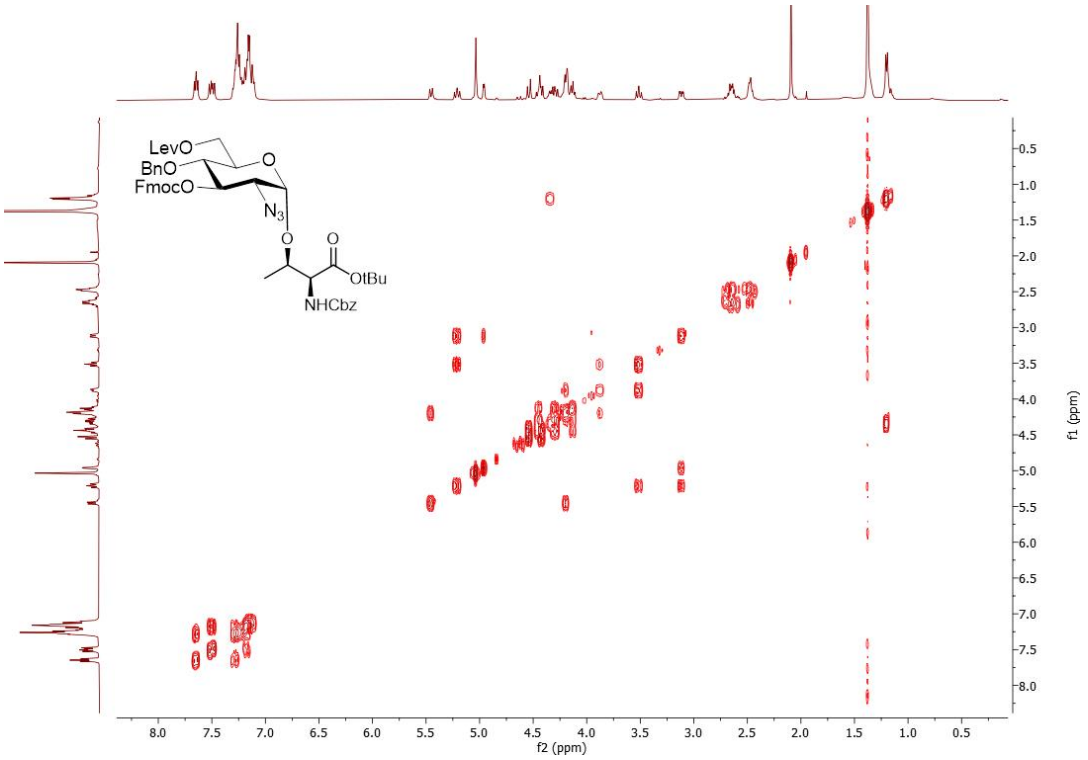
^{1}H NMR (400 MHz, $CDCl_3$) of **71**



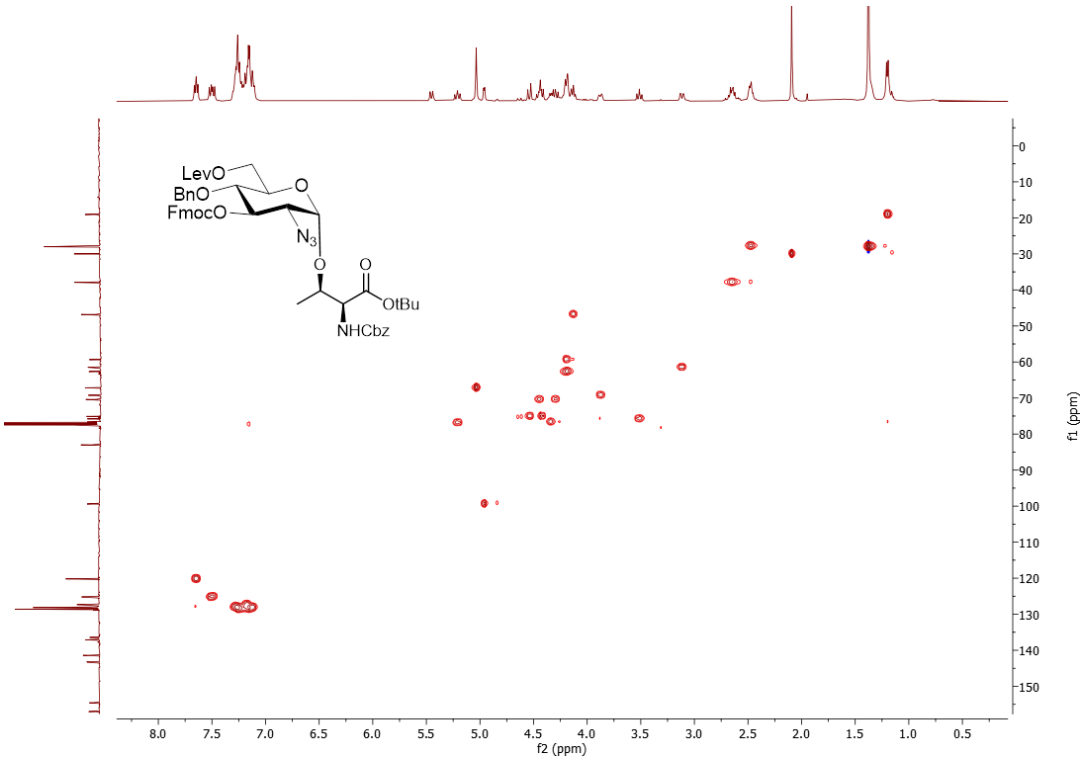
^{13}C NMR (101 MHz, $CDCl_3$) of **71**



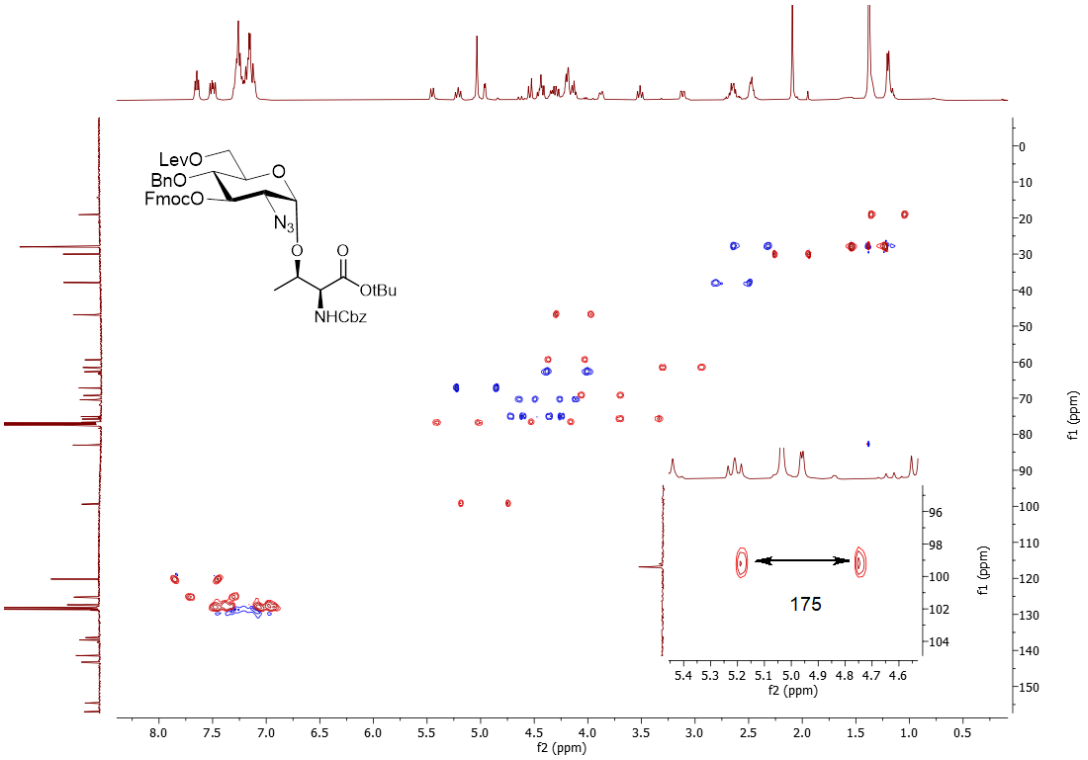
^{1}H, ^{1}H COSY of **71**



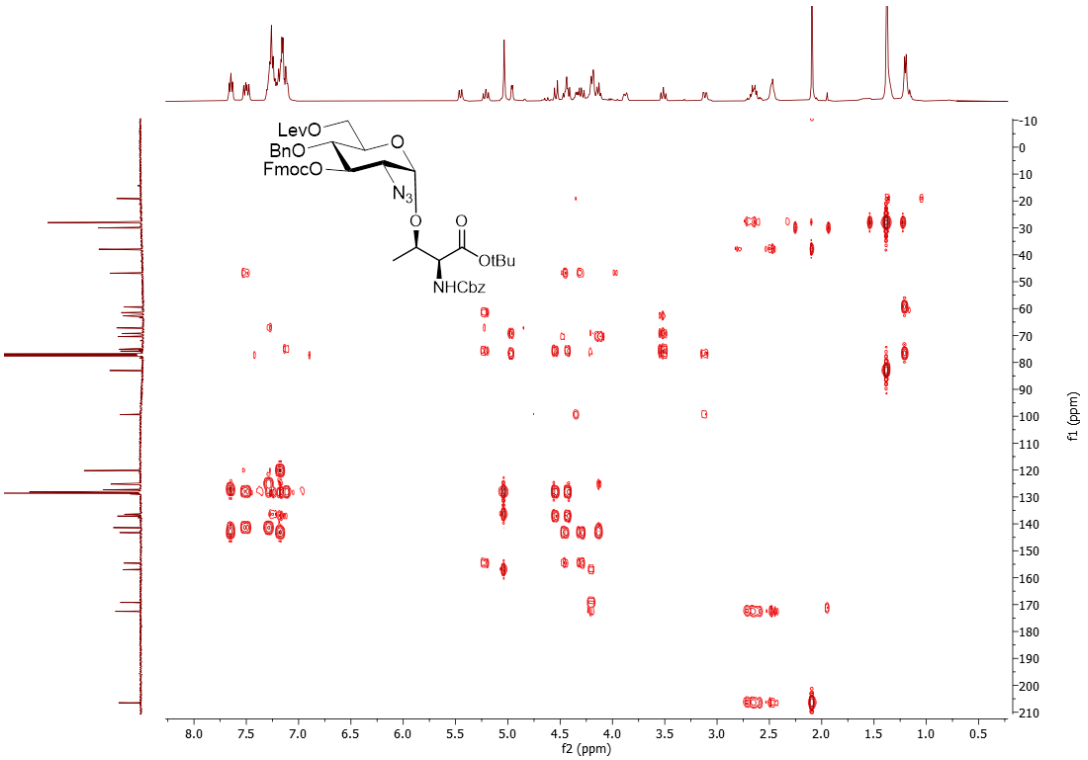
^{13}C, ^{1}H HSQC of **71**



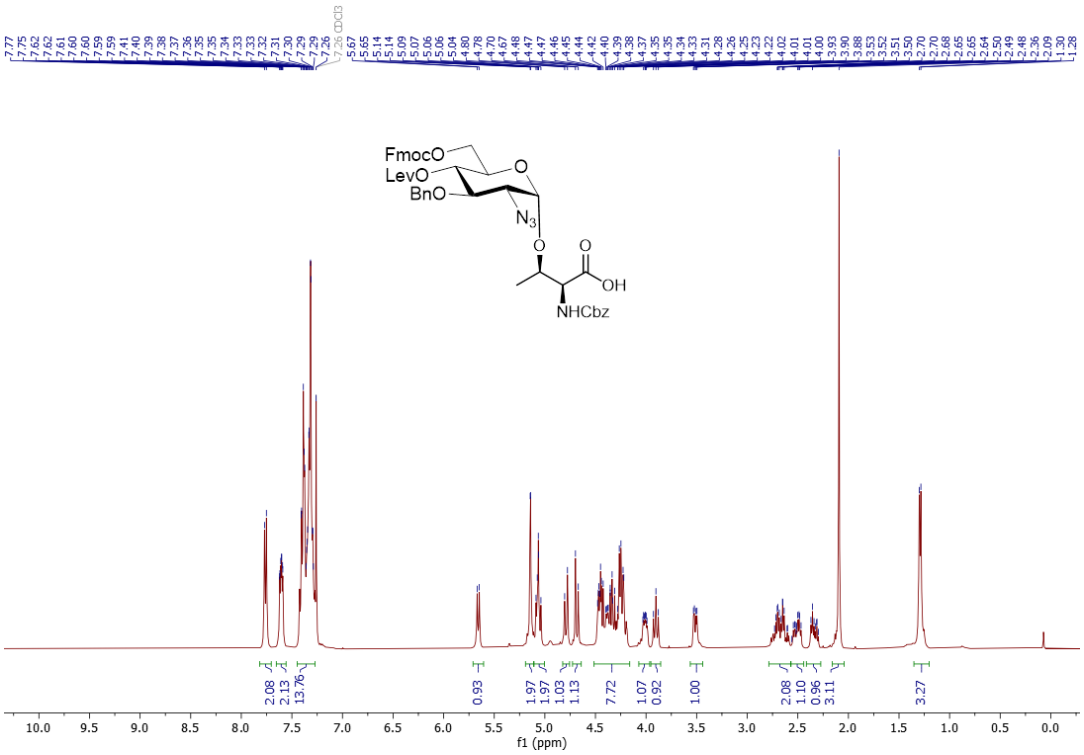
Coupled ^{13}C, ^{1}H HSQC of **71**



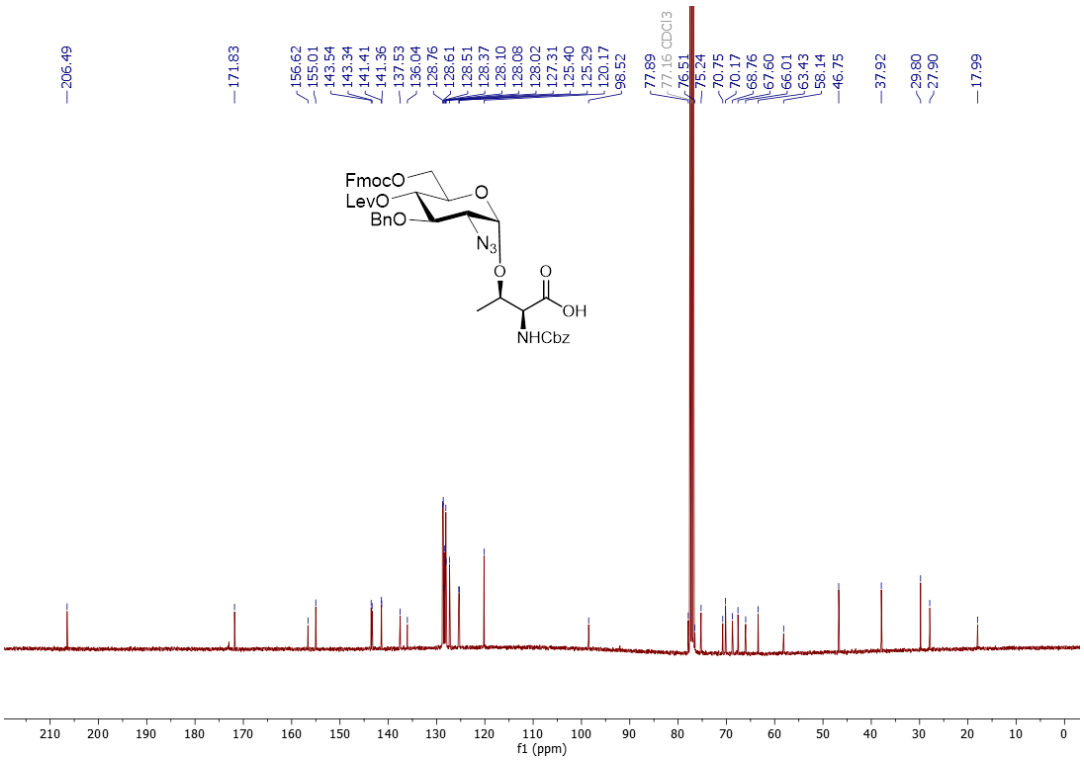
^{13}C, ^{1}H HMBC of **71**



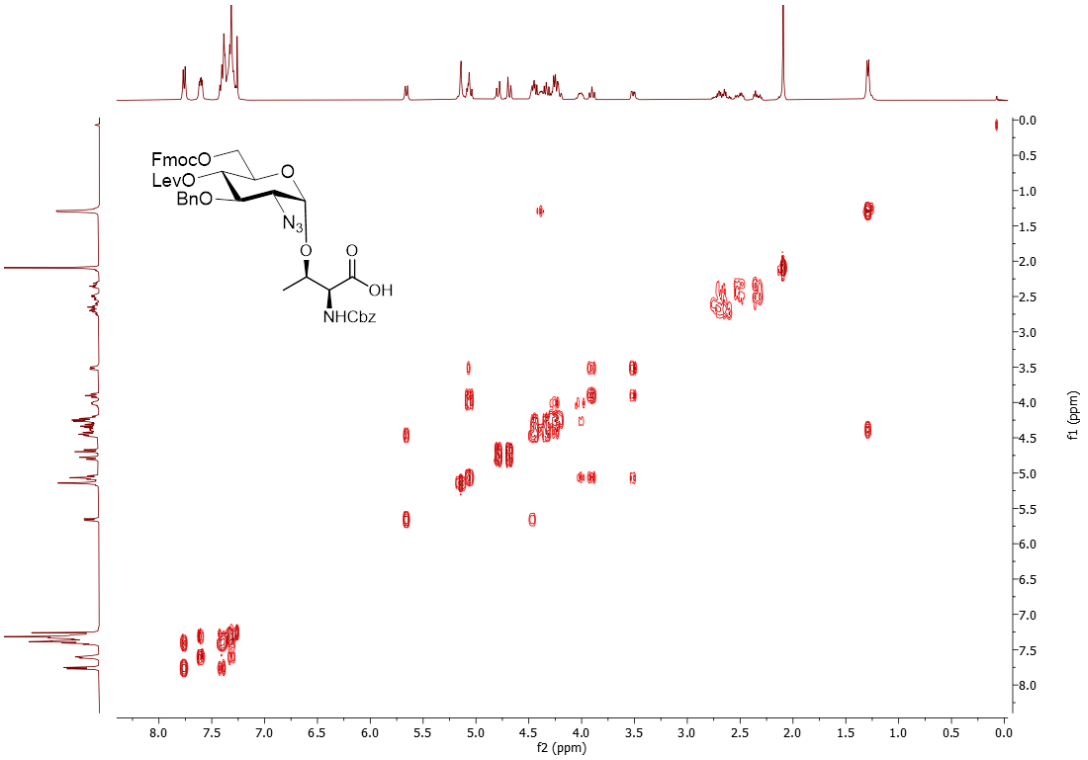
^{1}H NMR (400 MHz, $CDCl_3$) of **39**



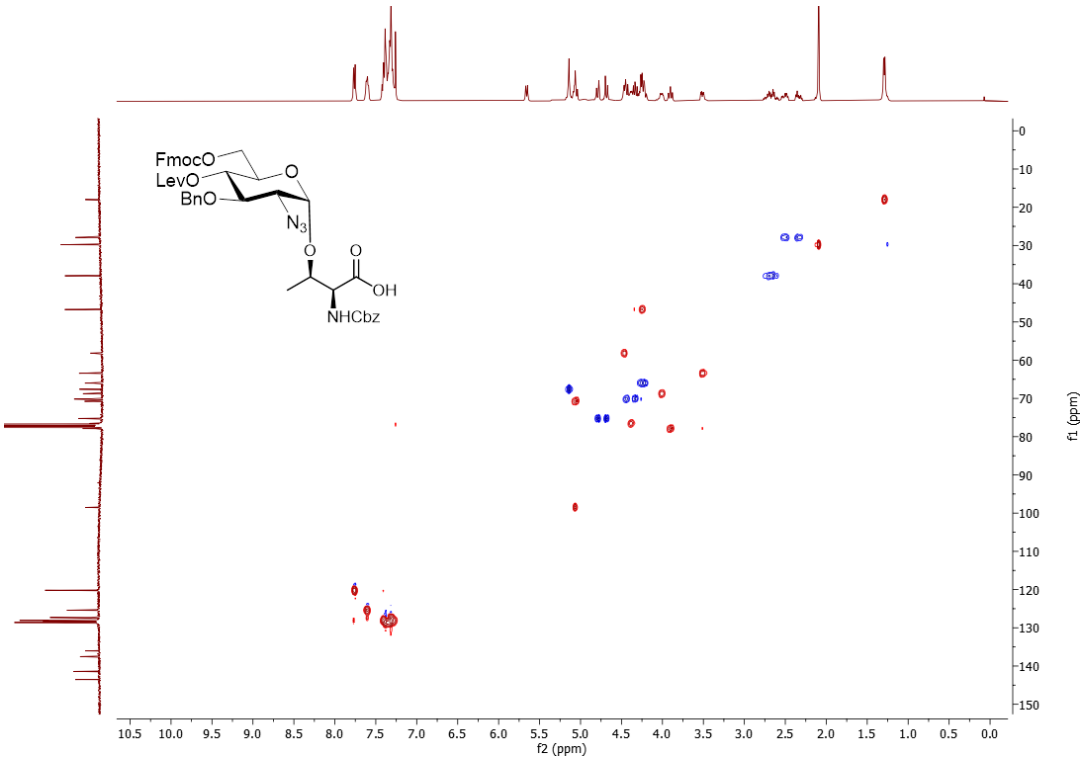
^{13}C NMR (101 MHz, $CDCl_3$) of **39**



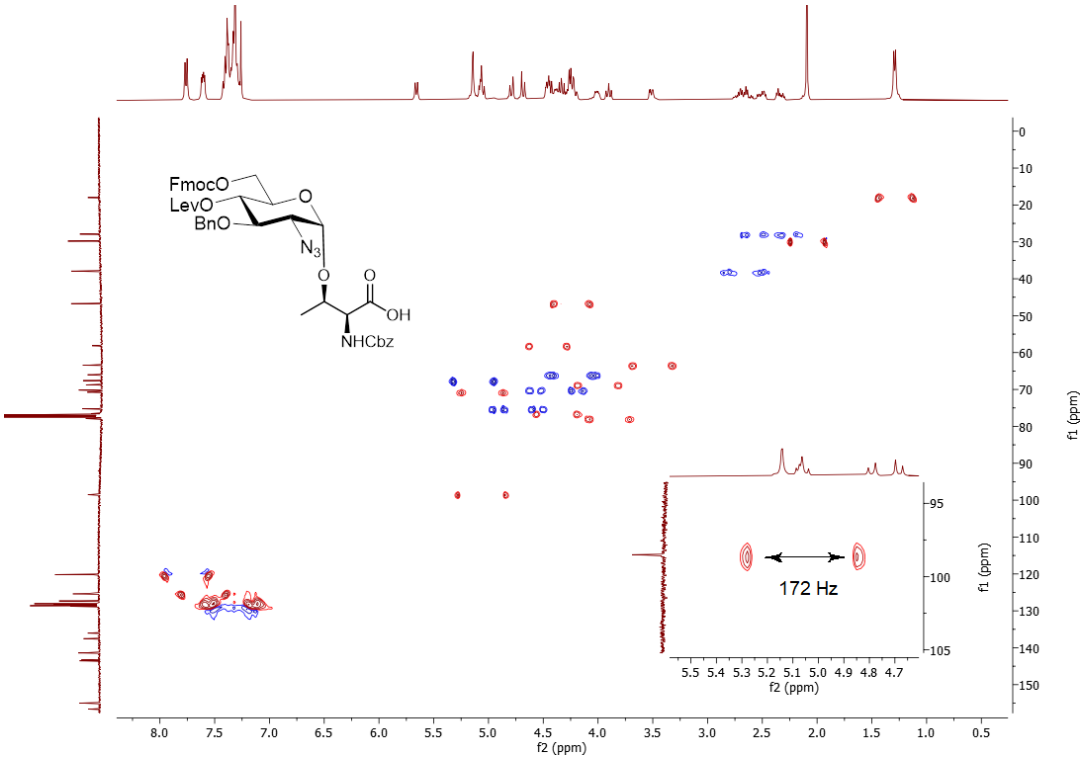
^{1}H, ^{1}H COSY of **39**



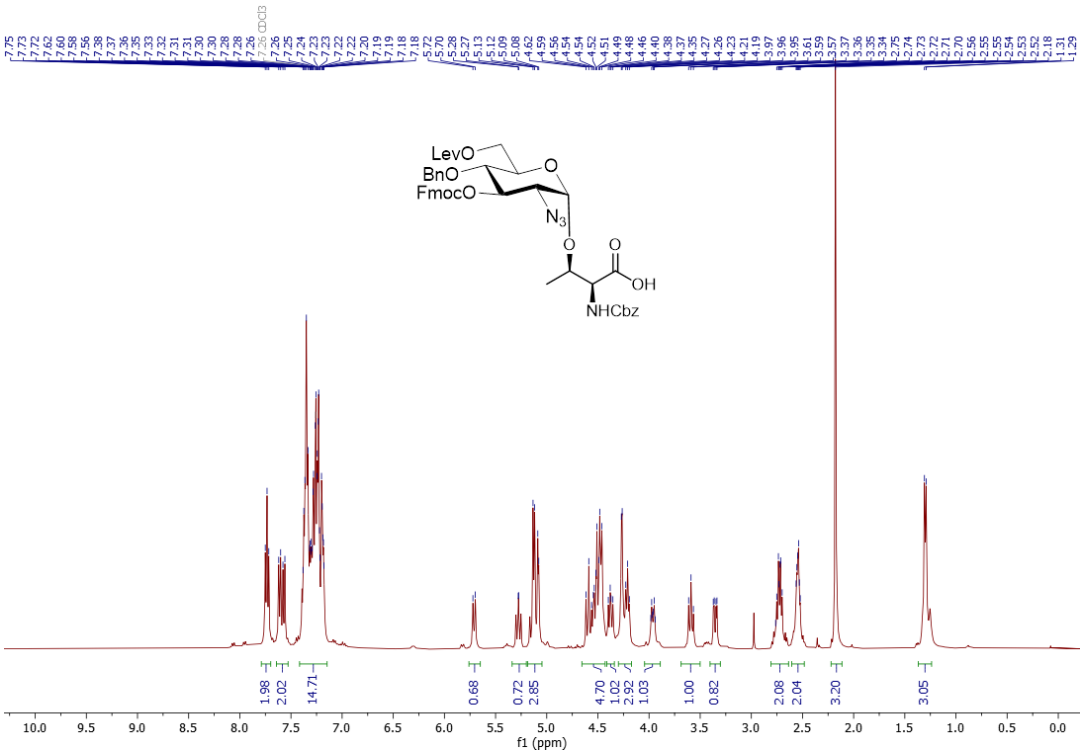
^{13}C, ^{1}H HSQC of **39**



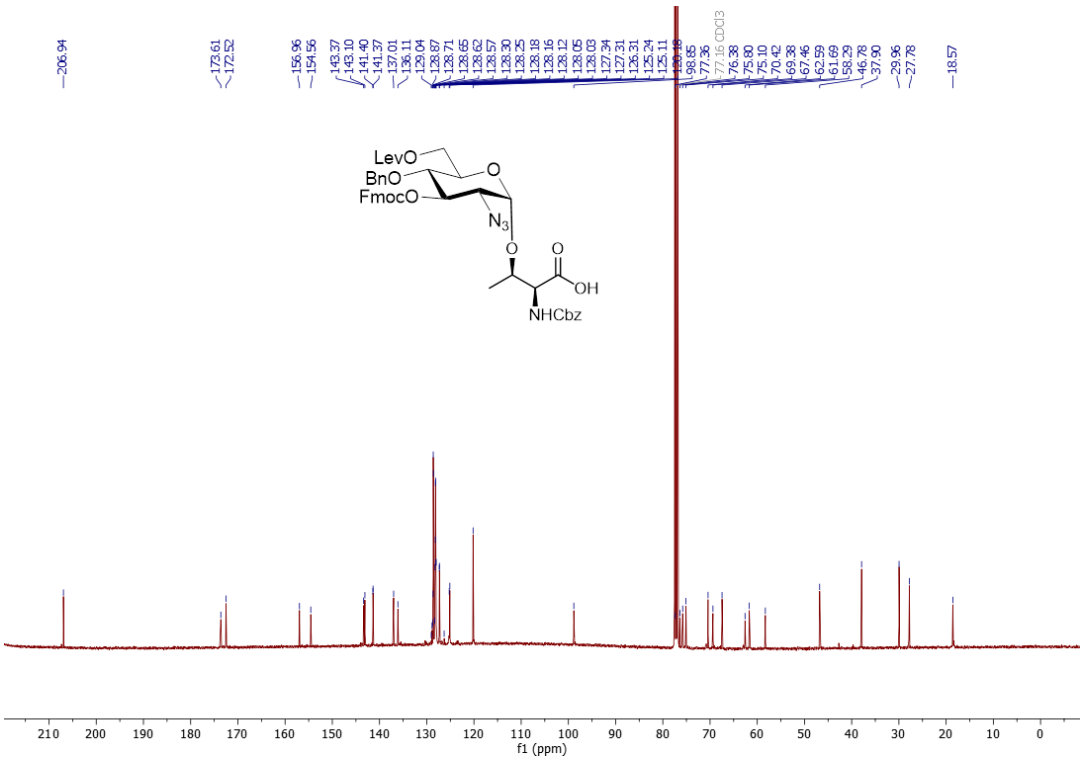
Coupled ^{13}C, ^{1}H HSQC of **39**



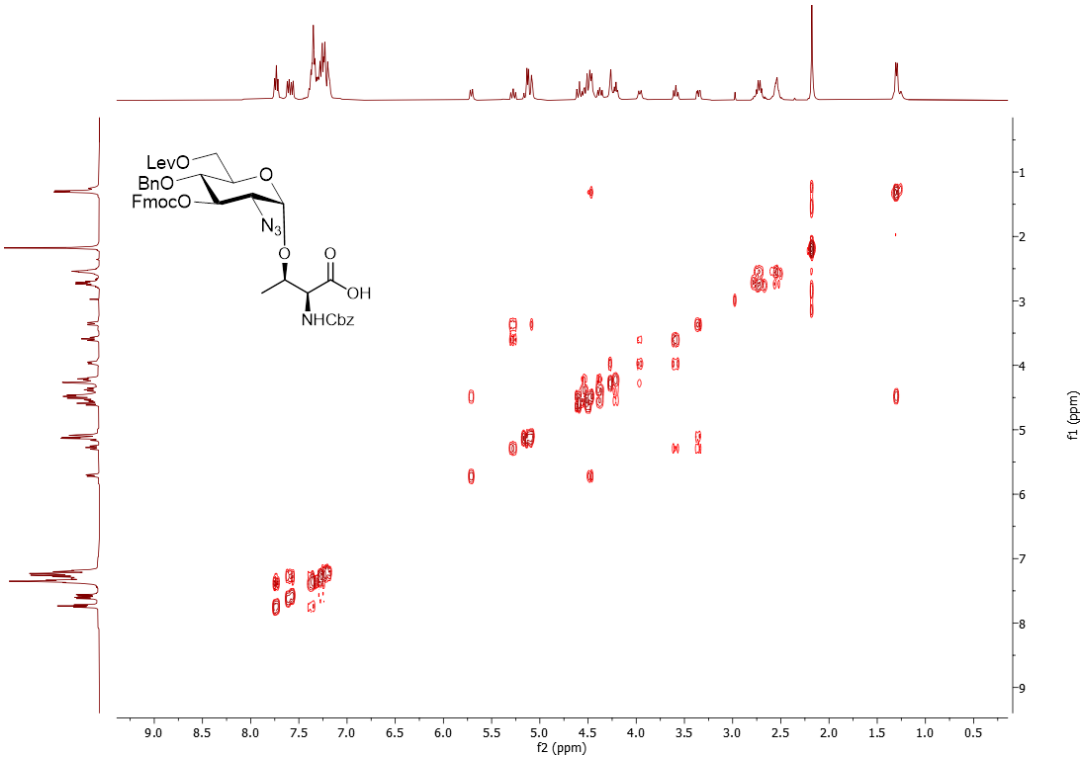
^{1}H NMR (400 MHz, $CDCl_3$) of **40**



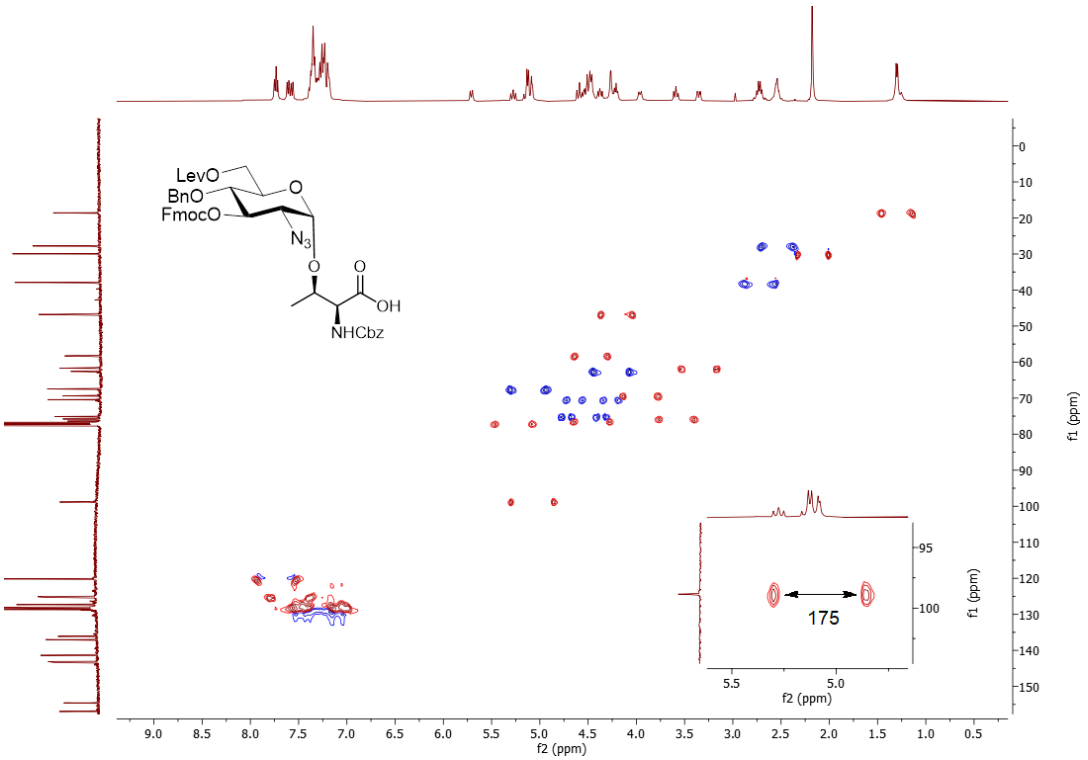
^{13}C NMR (101 MHz, $CDCl_3$) of **40**



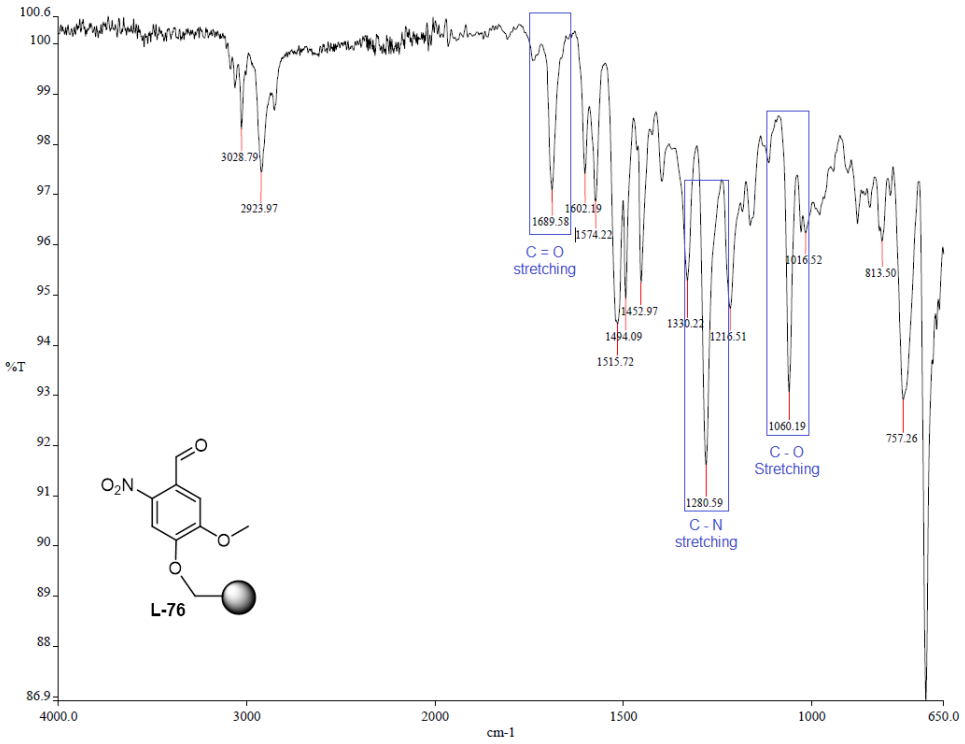
^{1}H, ^{1}H COSY of **40**

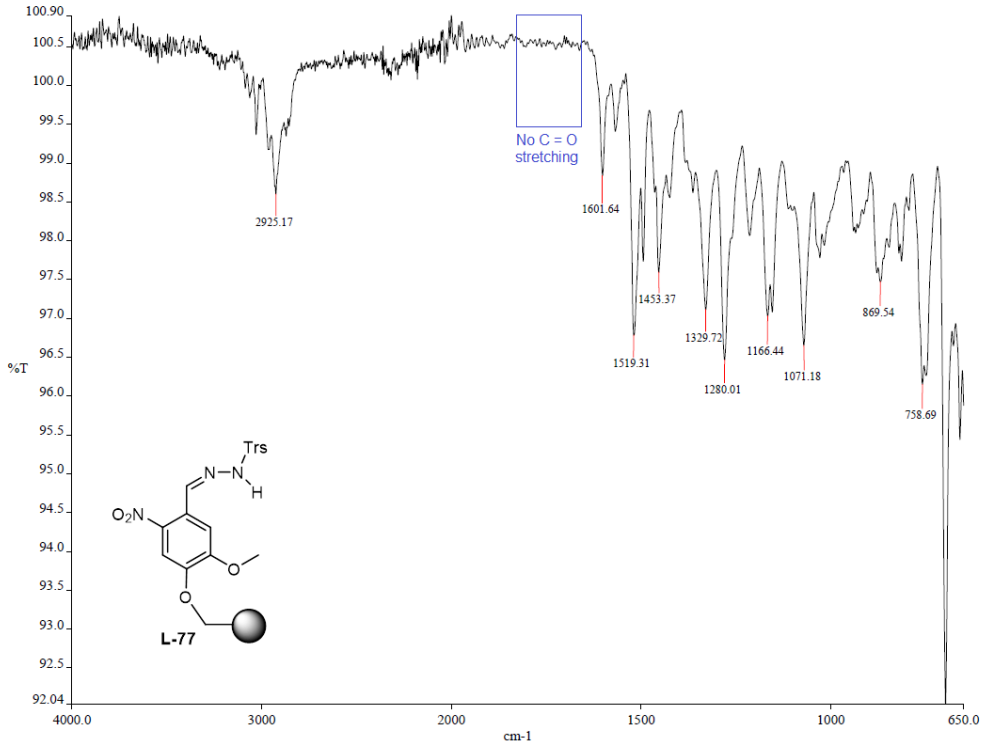


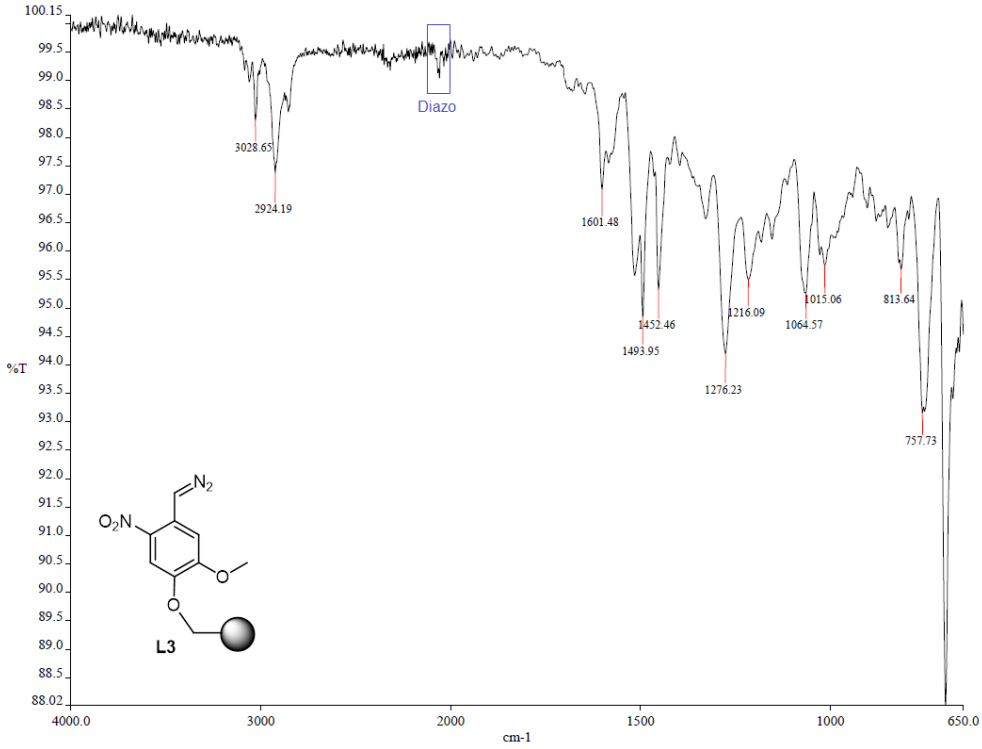
^{13}C, ^{1}H HSQC of **40**



FT-IR spectrum for the functionalization of the Merrifield resin with photolabile diazo liner






100.15
99.5
99.0
98.5
98.0
97.5
97.0
96.5
96.0
95.5
95.0
94.5
94.0
93.5
93.0
92.5
92.0
91.5
91.0
90.5
90.0
89.5
89.0
88.5
88.02
%T
Diazo
3028.65
2924.19
1601.48
1452.46
1493.95
1276.23
1216.09
1064.57
1015.06
813.64
757.73
N_2
O_2N
O
O
L3
4000.0
3000
2000
1500
1000
650.0
cm-1