# Synthetic glycans for structural studies: the importance of the modification pattern

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
Theodore Tyrikos-Ergas
from Kalymnos/Greece

2021

This work was performed between June 2018 and November 2021 under the supervision of Dr. Martina Delbianco and Prof. Dr. Peter H. Seeberger in the Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces Potsdam, and the Institute of Chemistry and Biochemistry, Freie Universität Berlin.

1st reviewer: Prof. Peter H. Seeberger

2nd reviewer: Prof. Rainer Haag

Date of oral defense: 29/03/2022

Declaration of independence

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree.

# Acknowledgement

Foremost, I would like to thank my supervisor Dr. Martina Delbianco for her guidance and for her continuous support on this thesis. Thank you for always being patient with me and constantly encouraging me throughout the hard moments in my PhD.

I would like to thank Prof. Dr. Peter H. Seeberger for offering the excellent environment and equipment in Department of Biomolecular Systems at the Max-Planck-Institute of Colloids and Interfaces. Besides, thank you for reviewing this thesis and for the constant support.

I thank Prof. Dr. Rainer Haag for kindly agreeing to review this thesis.

I would like to thank all the current and previous members of *carbohydrate materials* group, Dr. Yang Yu, Dr. Yuntao Zhu, Dr. Vittorio Bordoni, Dr. Denisa Vargova, Zhouxiang Zhao, Jhih-Yi Huang, Marlene Dal Colle, Soeun Gim, Giulio Fittolani and Surusch Djalali for the insightful and beneficial discussions and the amazing atmosphere in the lab that they create. I owe a big chunk of this achievement to each one of you. I would also like to thank all members of *automated glycan assembly* group. I am particularly grateful to Sandra Pinzon and Dr. Eric Sletten for the nice collaboration we had and to Surusch Djalali for the German translation of the abstract in this thesis.

My sincere thanks also goes to Dorothee Böhme for the administrative support, Eva Settels for her help with the HPLC, to Olaf Niemeyer for the maintenance of NMRs. Special thanks to my mentors in computational chemistry, Dr. Andrea Grafmueller, Dr. Ankush Singal and to the IT staff especially Marko Ehlert and Rene Genz for warranting GPU resources, managing all the data, and always taking the trouble to promptly address my trouble tickets.

My deep gratitude also goes to my family for their support and love, and for their constant presence in my life. Above all I would like to thank my wife Katerina for her love and constant support, for all the late nights and early mornings, and for keeping me sane over the past few months. This accomplishment would not have been possible without your backing.

# Table of Contents

# Summary

Carbohydrates are the most abundant organic compounds in nature. They serve as energy sources, regulate a plethora of biological processes, and are essential structural components in animals, plants and microorganisms. The structural diversity of carbohydrates results in materials with extremely different properties. Still, structure-property correlations are hardly established for carbohydrates due to the difficulty in obtaining pure, well-defined molecules and the lack of suitable analytical methods. A comprehensive understanding of carbohydrate function requires a detailed understanding and thorough elucidation of the carbohydrate's structure. The ultimate goal of this thesis is to establish correlations between the structure and the properties of carbohydrates and shine light on how small modifications affect the shape of carbohydrates. To achieve this goal, automated glycan assembly (AGA) is used as a platform to produce well-defined oligosaccharide probes. Molecular dynamic (MD) simulations are performed to address conformational aspects of oligosaccharides at the atomic level and to support the structural analysis.

In Chapter **2**, well-defined unnatural oligosaccharides including methylated, deoxygenated, deoxyfluorinated, as well as carboxymethylated cellulose analogues were prepared with full control over degree and pattern of substitution. My contributions to this project was to guide the synthesis using MD by investigating how single site substitutions affect the geometry and the properties of tailor-made cellulose analogues. A detailed dihedral analysis depicted how each glycosidic bond is affected by the modifications with consequences on the overall structure. Compounds with the same degree of modification, but different substitution patterns, behave drastically different.

In Chapter **3**, I focused on how the pattern of ionic substituents affect oligosaccharide structures. MD simulations demonstrated how the positive charges on the chitosan backbone promote local interactions that play a major role in the flexibility and overall conformation of such oligosaccharides. Intramolecular hydrogen bonds that stabilize new geometries were predicted by MD and detected by NMR. These observations is particularly interesting to clarify why Nature is producing sequence-specific chitosans to trigger different biological processes.

The realization that ionic charges play an important role in determining oligosaccharide conformation inspired me to develop efficient synthetic protocols to access more complex ionic oligosaccharides. In Chapter **4,** I developed a general protocol for the on resin synthesis of sulfated glycans. I addressed several issues related to the synthesis of sulfated glycans on solid support including sulfation, deprotection, and purification. With this method, a diverse collection of complex sulfated glycans was synthesized in good to excellent yield requiring only one final purification step.

Recently, it was discovered that some *E. coli* strains produce cellulose modified with phosphoethanolamine (pEtN) groups as part of their protective biofilms, providing increased adhesion. In Chapter **5**, I developed a model system to study how this modification affects the formation of bacterial biofilms, nanocomposites of cellulose and proteins. I prepared synthetic oligomers representative of each component with full control over their chemical composition. Co-assembly experiments revealed that different degree and pattern of

pEtN substitution in the oligosaccharide modulated the length and aggregation tendency of the peptide fibers. The mechanical properties of the protein-carbohydrate network were affected by the chemical nature of the carbohydrate component. Moreover, I identified synthetic oligosaccharides capable of interrupting fibrillary assembly that could serve as promising drug candidates for the treatment of neurological diseases or as antibacterial agents.

Overall, the synthetic oligosaccharide prepared and studied in this thesis offered the basis for the fundamental understanding of carbohydrate structure-function relationship and will serve as starting point for the rational design of carbohydrate-based materials.

# Zusammenfassung

Kohlenhydrate sind die am häufigsten vorkommenden organischen Verbindungen in der Natur. Sie dienen als Energiequelle, regulieren eine Vielzahl biologischer Prozesse und zählen zu den wesentlichen Strukturbausteinen in Tieren, Pflanzen und Mikroorganismen. Die strukturelle Vielfalt der Kohlenhydrate führt zu Materialien mit extrem unterschiedlichen Eigenschaften. Dennoch sind Struktur-Eigenschafts-Korrelationen für Kohlenhydrate kaum etabliert, da nicht nur die Gewinnung von reinen und definierten Strukturen ein Problem darstellt, sondern auch der Mangel an geeigneten Analysemethoden. Ein umfassendes Verständnis der Funktion von Kohlenhydraten erfordert eine detaillierte Untersuchung und Aufklärung der Molekularstruktur. Ziel dieser Arbeit ist es, Parallelen zwischen der dreidimensionalen Struktur und den Eigenschaften von Kohlenhydraten herzustellen und aufzuzeigen, welchen Einfluss kleine Veränderungen der Primärstruktur haben können. Um dieses Ziel zu erreichen, wird die automated glycan assembly (AGA) Plattform zur Herstellung wohldefinierter Oligosaccharide verwendet. Molekulardynamische (MD) Simulationen dienen als unterstützende Methode, um das Konformationsverhalten der Oligosaccharide auf atomarer Ebene zu untersuchen und die Strukturanalyse zu begleiten.

Im zweiten Kapitel wurden wohldefinierte, nicht natürlich vorkommende Oligosaccharide, einschließlich methylierter, desoxygenierter, desoxyfluorierter sowie carboxymethylierter Celluloseanaloga synthetisiert, wobei der Grad und das Muster der Substitution vollständig kontrolliert werden konnten. Mein Beitrag bei diesem Projekt bestand darin, die Synthese mit Hilfe von MD zu führen. Dabei wurde untersucht wie die einzelne Substitutionen die Geometrie und die Eigenschaften der synthetischen Cellulose-Analoga beeinflussen. Mit Hilfe von detaillierten Diederwinkel-Analysen konnte aufgezeigt werden, wie die einzelnen glyosidischen Bindungen durch die Modifikationen beeinflusst werden und wie sich diese Veränderungen auf die Gesamtstruktur auswirken. Dabei stellte sich heraus, dass Verbindungen mit gleichem Modifikationsgrad und unterschiedlichen Substitutionsmustern sich drastisch in ihrem Verhalten unterschieden.

Im dritten Kapitel habe ich die Frage behandelt, wie das ionische Substitutionsmuster die Oligosaccharidstruktur beeinflusst. MD-Simulationen konnten im Vorfeld zeigen, dass die positiven Ladungen auf dem Chitosan-Grundgerüst lokale Wechselwirkungen fördern und damit einen wichtigen Beitrag für die Flexibilität und die Gesamtkonformation der Oligosaccharide leisten. Das Auftreten von intramolekularen Wasserstoffbrückenbindungen, die neue Geometrien stabilisieren, konnte mittels MD vorhergesagt und im Anschluss durch NMR nachgewiesen werden. Diese Beobachtungen sind von besonderem Interesse, um bei der Beantwortung der Frage, warum die Natur sequenzspezifische Chitosane produziert, um verschiedene biologische Prozesse auszulösen, zu helfen.

Die Erkenntnis, dass Ladungen eine essentielle Rolle bei der Bestimmung der Oligosaccharidkonformation spielen, inspirierte mich dazu ein neues, effizientes Syntheseprotokoll für den Zugang von komplexeren ionischen Oligosacchariden zu entwickeln. Im vierten Kapitel habe ich ein allgemeines Protokoll für die Festphasen-Synthese von sulfonierten Glykanen erarbeitet. Dabei habe ich verschiedene Aspekte der Synthese wie die Sulfonierung, Entschützung und Reinigung beleuchtet. Mit dieser Methode wurde anschließend eine Reihe komplexer sulfonierter Glykane in guter bis ausgezeichneter Ausbeute synthetisiert.

Kürzlich wurde entdeckt, dass einige E. Coli-Stämme mit Phosphoethanolamin (pEtN)-Gruppen modifizierte Cellulose als Teil ihrer schützenden Biofilme produzieren und so eine erhöhte Adhäsion aufbauen können. Im fünften Kapitel habe ich ein Modellsystem entwickelt, um zu untersuchen, wie sich diese Modifikation auf die Bildung von bakteriellen Biofilmen, Nanokompositen aus Cellulose und Proteinen auswirkt. Ich habe synthetische Oligomere hergestellt, die die jeweilige Komponente repräsentieren. Co-assembly-Experimente dieser konnten zeigen, dass der unterschiedlich Grad und das Muster der pEtN-Substitution sowohl Einfluss auf die Länge des Oligosaccharids, als auch auf die Aggregation der Peptidfasern haben. Die mechanischen Eigenschaften des Protein-Kohlenhydrat-Netzwerks wird vor allem durch die chemische Natur der Kohlenhydratkomponente bestimmt. Darüber hinaus konnte ich synthetische Oligosaccharide identifizieren, die in der Lage sind, die Bildung von Fibrillen bei der Aggregation zu unterdrücken. Diese könnten in Zukunft als vielversprechende Arzneimittelkandidaten bei der Behandlung von neurologischen Erkrankungen oder als antibakterielle Wirkstoffe eingesetzt werden.

Insgesamt leisten die in dieser Arbeit hergestellten und untersuchten synthetischen Oligosaccharide einen bedeutenden Beitrag für das grundlegende Verständnis der Struktur-Funktions-Beziehung von Kohlenhydraten. Sie können des Weiteren als Ausgangspunkt für das rationale Design von Materialien auf Kohlenhydratbasis herangezogen werden.

# List of Publications

## Scientific Publications and Reviews

- *Systematic Hydrogen-Bond Manipulations to Establish Polysaccharide Structure-Property Correlations*
  Y. Yu*, **T. Tyrikos-Ergas***, Y. Zhu, G. Fittolani, V. Bordoni, A. Singhal, R.J. Fair, A. Grafmüller, P.H. Seeberger, M. Delbianco, *Angew. Chem., Int. Ed.*, **2019**, 58 (37), 13127-13132

- *Systematic Structural Characterization of Chitooligosaccharides Enabled by Automated Glycan Assembly*
  **T. Tyrikos-Ergas**, V. Bordoni, G. Fittolani, M. A. Chaube, A. Grafmüller, P.H. Seeberger, M. Delbianco *Chem. Eur. J.*, **2021**, 27, 2321-2325

- *Structural Studies Using Unnatural Oligosaccharides: Toward Sugar Foldamers*
  **T. Tyrikos-Ergas*,** G. Fittolani*, P.H. Seeberger, M. Delbianco, *Biomacromolecules*, **2020**, *21* (1), 18-29

- *Automated access to well-defined ionic oligosaccharides*
  Y. Zhu, **T. Tyrikos-Ergas**, K. Schiefelbein, A. Grafmüller, P.H. Seeberger, M. Delbianco, *Org. Biomol. Chem.*, **2020**, 18, 1349-1353

- *Identifying the origin of local flexibility in a carbohydrate polymer*
  K. Anggara, Y. Zhu, G. Fittolani, Y. Yu, **T. Tyrikos-Ergas**, M. Delbianco, S. Rauschenback, S. Abb, P.H. Seeberger, K. Kern, *Proc. Natl. Acad. Sci.*,**2021**, 118, e2102168118

- *Progress and challenges in the synthesis of sequence controlled polysaccharides*
  G. Fittolani, **T. Tyrikos-Ergas**, D. Vargová, M.A. Chaube, M. Delbianco, *Beilstein J. Org. Chem.*, **2021**, 17, 1981-2025

  ***equal contribution**

## Scientific Conferences and Symposia Presentations:

August 2021, ACS Fall 2021 webinar, (USA) (oral)

June 2021, NIH & FDA Glycoscience Research Day 2021 webinar, (USA) (poster)

July 2020, 17th RSC Carbohydrate ECR Webinar, Manchester (UK) (oral)

October 2019, 3rd Potsdamer Chemistry Symposium, Potsdam (Germany) (oral)

June 2019, XLIV "A. Corbella" International Summer School on Organic Synthesis (ISOS 2019), Gargnano (Italy) (poster)

# List of Abbreviations

| | |
|---|---|
| AFM | atomic force microscopy |
| AGA | automated glycan assembly |
| BBs | building blocks |
| Bn | benzyl |
| Bz | benzoyl |
| CD | circular dichroism |
| COS | chitooligosaccharide |
| Cryo-EM | cryo-electron microscopy |
| DA | degree of acetylation |
| DP | degree of polymerization |
| E.Coli | Escherichia coli |
| ECM | extracellular matrix |
| EPR | electron paramagnetic resonance |
| ESI | electrospray ionization |
| ES-IBD | electrospray ion-beam deposition |
| FA | fraction of acetylation |
| Fmoc | fluorenylmethyloxycarbonyl group |
| FRET | fluorescence resonance energy transfer |
| GAGs | glycosaminoglycans |
| H | heparin |
| HPC | high performance computing |
| HPLC | high-performance liquid chromatography |
| HS | heparan sulfate |
| IR | Infrared |
| Lev | levulinic acid |
| LG | leaving group |
| LJ | Lennard-Jones |
| MD | molecular dynamics |
| NMR | nuclear magnetic resonance |
| pEtN | phosphoethanolamine cellulose |
| REMD | replica exchange molecular dynamics |

| | |
|---|---|
| RMSD | root-mean-square deviation |
| ROG | radius of gyration |
| SAXS | small angle X-ray scattering |
| SEM | scanning electron microscopy |
| ssNMR | solid state nuclear magnetic resonance |
| STM | scanning tunneling microscopy |
| TEM | transmission electron microscopy |
| TLR | toll-like receptors |
| TOCSY | total correlation spectroscopy |
| tPG | temporary protecting groups |
| XRD | X-ray diffraction |

# 1. Introduction

This chapter has been modified in part from the following articles:

Y. Yu,* **T. Tyrikos-Ergas**,* Y. Zhu, G. Fittolani, V. Bordoni, A. Singhal, R. J. Fair, A. Grafmüller, P. H. Seeberger, M. Delbianco, Systematic Hydrogen Bond Manipulations to Establish Polysaccharide Structure-Property Correlations. *Angew. Chem. Int. Ed.,* **2019**, 131, 13261.

**T. Tyrikos-Ergas**, V. Bordoni, G. Fittolani, M. A. Chaube, A. Grafmüller, P.H. Seeberger, M. Delbianco, Systematic Structural Characterization of Chitooligosaccharides Enabled by Automated Glycan Assembly. *Chem. Eur. J.,* **2021**, 27, 2321-2325.

**T. Tyrikos-Ergas*,** G. Fittolani*, P.H. Seeberger, M. Delbianco, Structural Studies Using Unnatural Oligosaccharides: Toward Sugar Foldamers. *Biomacromolecules*, **2020**, 21 (1), 18-29

G. Fittolani, **T. Tyrikos-Ergas**, D. Vargová, M.A. Chaube, M. Delbianco, Progress and challenges in the synthesis of sequence controlled polysaccharides. *Beilstein J. Org. Chem.,* **2021**, 17, 1981-2025.

## 1.1. Carbohydrates

Proteins, nucleic acids, carbohydrates and lipids are the main building blocks of life[1]. Carbohydrates are the most abundant organic compounds in Nature and perform numerous roles in living organisms. Polysaccharides serve as essential source of energy and are involved in a myriad of biological processes, promoting cell recognition and signaling[2]. Additionally, carbohydrates are able to assemble in high ordered architectures, providing materials with exceptional properties and structurally supporting plant cells, fungi, and all of arthropods and crustaceans.

Glycans are generally composed of simple monomers called monosaccharides. The monosaccharides are assembled into complex structures, ranging from short oligomers to long polysaccharides. The enormous diversity arises not only from the many monosaccharide units, but also from the fact that the anomeric carbon of a monosaccharide can be linked to any of the hydroxyl groups of another sugar unit (i.e., different connectivity). Compared to nucleic acids and peptides that are linear polymers, polysaccharides can form branched structures. In addition, the formation of a glycosidic bond generates a stereogenic center that can adopt the alpha or beta configuration (**Figure 1**).

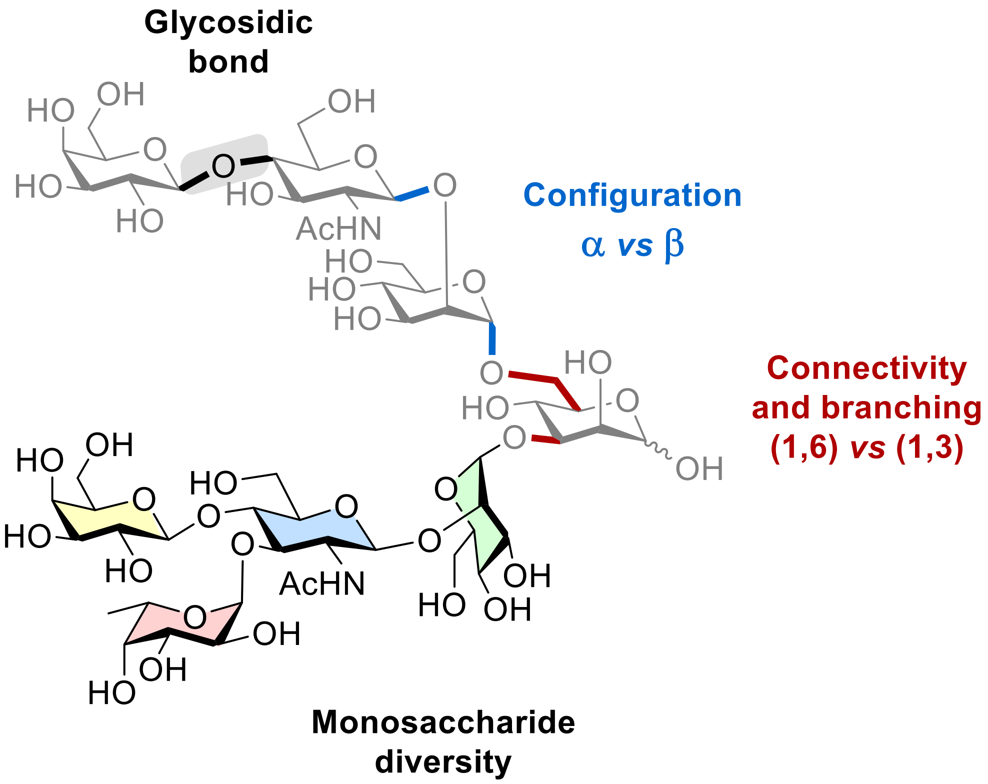


**Figure 1**. Example of an oligosaccharide to depict the structural complexity of glycans.

### 1.1.1. Polysaccharides as biomaterials

An immense interest is devoted to the study of biomaterials, because they are an attractive resource of biodegradable materials for multiple applications[3]. Natural polysaccharides, such as cellulose and chitin, are considered as unlimited sources of raw materials[4]. These polysaccharides have a strong tendency to aggregate in well-defined architectures and provide material with different properties. The formation of high ordered structures is due to inter- and intramolecular interactions originated from the many hydroxyl and other functional groups. Chemical modifications have been widely used to tune polysaccharide properties[5] and broaden their applications[6]. However, non-regioselective derivatization often used to modify polysaccharides result in ill-defined patterns that do not allow for proper structure-function correlations[7]. In the context of this thesis, I designed and synthesized well-defined glycans to investigate how single-site modifications affect their structure and material properties.

### 1.1.2. Cellulose and chitin

Cellulose is the most abundant biopolymer on Earth and the principal constituent of all plant cell walls. The worldwide natural production rate of cellulose has been estimated to be approximately $10^{11}$ tons annually. It consists of linear chains of β-(1,4)-linked glucose repeating units[8]. Cellulose is a material featured with extraordinary physicochemical properties such as high rigidity, compressive strength, and high thermal stability. Cellulose is neither meltable nor soluble in conventional solvents which limits its application[9]. A dense network of cooperative hydrogen bonds between the numerous hydroxyl groups stabilizes the structure of cellulose (**Figure 2**). In addition, recent quantum chemical calculations[10] have shown that London dispersion interactions, which arise from temporary polarization, significantly contribute to the stabilization of cellulose. These interactions explain the dissolution resistance of cellulose[11] that cannot be exclusively rationalized with the mere contribution of hydrogen bonds[12].

The crystal structure of cellulose was determined using synchrotron and neutron diffraction from oriented fibrous samples[13]. Cellulose is polymorphic[14], but naturally exists in two different crystal forms, cellulose Iα (triclinic unit cell) and cellulose Iβ (monoclinic unit cell) with the chains packed in a parallel manner. The ratio of those crystalline forms varies depending on the source of origin, with Iα being more prevalent in higher plants such as flax and cotton. Alkali treatment of cellulose I results in a new allomorph, Cellulose II, which is the most stable form of cellulose, having antiparallel oriented chains (**Figure 2**). There are several other cellulose allomorphs, such as Cellulose III$_{\text{I}}$ which arises from the chemical treatment of algal cellulose with ethyldiamine.

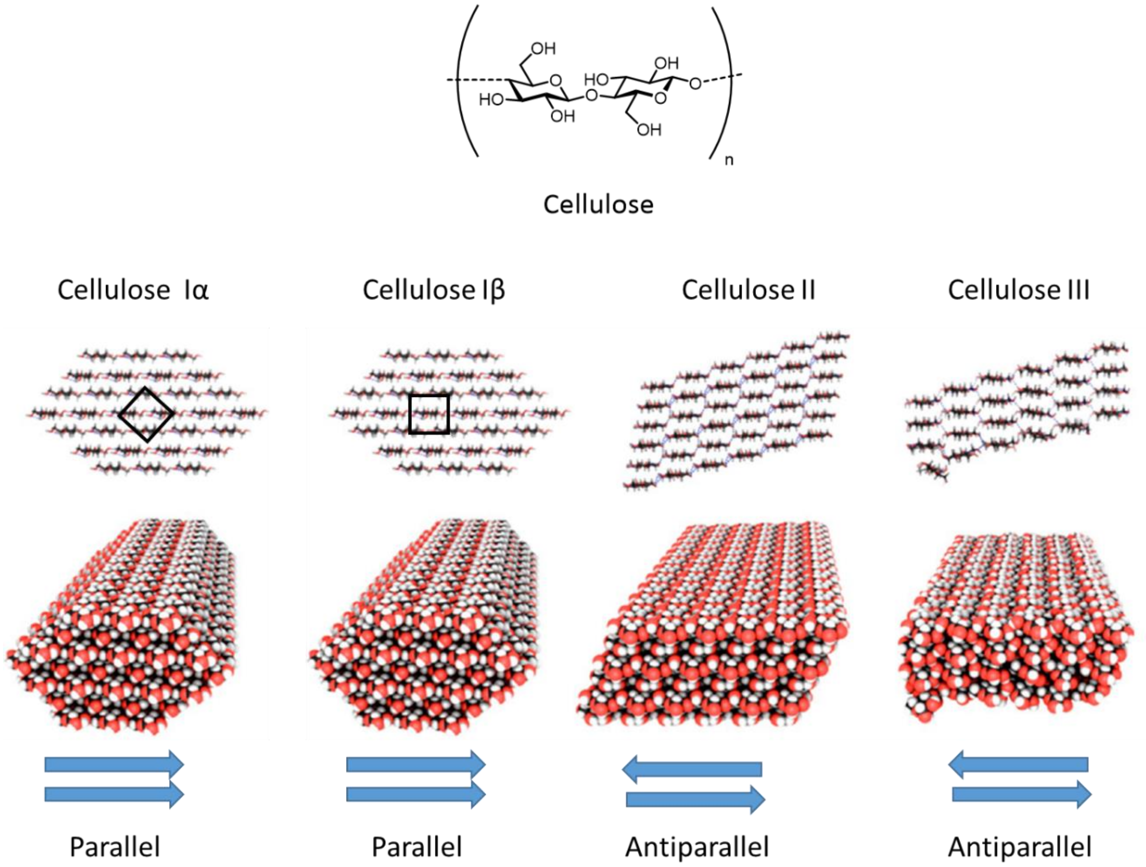


**Figure 2.** Chemical structure of cellulose and molecular models of cellulose allomorphs. Cellulose Iα (triclinic unit cell) and cellulose Iβ (monoclinic unit cell) are the native form of cellulose in plant cell walls. Cellulose II and III are generated by chemical treatment of Cellulose I and the chains are oriented antiparallel.

The many hydroxyl groups on the cellulose surface can be chemically modified to afford novel cellulosic materials with exceptional physicochemical properties. Many of these cellulose derivatives have found applications in various fields and can be mostly categorized into two types[7]: cellulose ethers and cellulose esters. Cellulose ethers are widely used for pharmaceutical applications such as drug delivery, coating drugs, gelatin, and thickening agents[7]. Organic and inorganic esters of cellulose like cellulose acetate, cellulose acetate propionate, and cellulose acetate butyrate have long played an important role in coatings applications[5].

Despite the simplicity of the synthetic event (i.e., polymerization of glucose units), unraveling the molecular mechanism that underline cellulose formation in higher plants is extremely challenging[15,16]. In contrast, the bacterial biosynthesis of cellulose is less complex and studied thoroughly[17,18]. Bacteria such as *Gluconacetobacter* and *Aerobacter* are major producers of high purity cellulose[19]. Recently, it was reported that *Escherichia coli* (*E. coli)* and *Salmonella* produce modified cellulose[20] in which every second glucosyl residue carries a phosphoethanolamine (pEtN) substituent at the 6-OH [21].This modification play a crucial role in adjusting the macroscopic morphology of bacteria biofilms (**Figure 3**). Still, the pattern of modification, the length of the pEtN cellulose, and the mode of interaction with other extracellular biomolecules remain unknown[22].

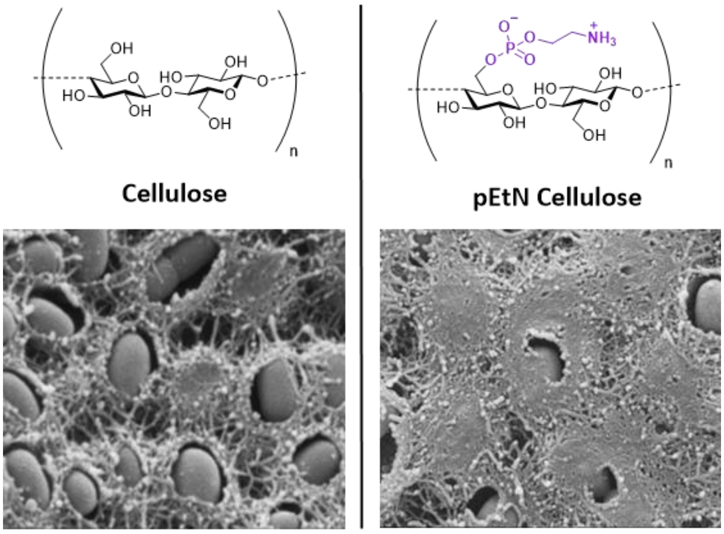


**Figure 3**. High resolution imaging of biofilms by scanning electron microscopy. Biofilms containing cellulose (left) and pEtN cellulose (right). Adapted with the permission from Thongsomboon[20] et al. Copyright (2018) AAAS.

After cellulose, the second most abundant natural biopolymer is chitin[23], a nitrogenous polysaccharide forming the primary component of fungal cell walls, crystaceans, arthropods, and certain algae. Its structure is similar to cellulose and consists of β-(1,4)-linked *N*-acetyl-D-glucosamine units (**Figure 4**). Chitin mainly occurs in three different polymeric forms α-, *β*-, and γ. α-Chitin is arranged in anti-parallel strands and it is the most abundant and stable allomorph found in nature. Chitin is highly hydrophobic; thus sharing the same insolubility issues and limitations for large scale applications as cellulose. Chemical treatment of chitin with concentrated sodium hydroxide, or its biotransformation by chitin deacetylase, results in chitosan, a pseudo-natural cationic polymer (**Figure 4**). When the degree of acetylation of chitin reaches about 40-60%, the amino groups drastically increase chitosan solubility[24]. Chitin and chitosan have been widely employed as biomedical materials and as decontaminant of residual waters containing heavy metals[25,26]. Chitooligosaccharide (COS) are short oligomers produced by the degradation of chitin and exhibit biological activities such as antitumor, antioxidant, antimicrobial and elicitor of plant immunity[27].

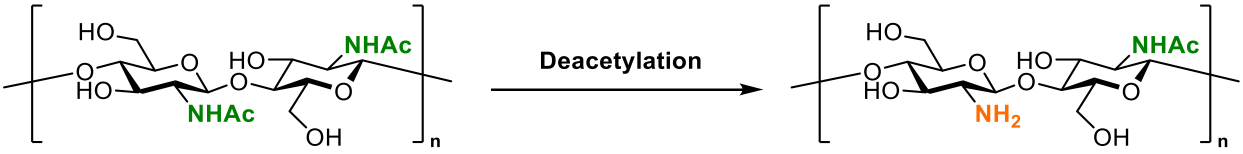


**Figure 4.** Structure of chitin and its partially deacetylated analogue chitosan.

The biological activity of chitosans is highly depended on their degree of polymerization (DP) and degree of acetylation (DA)[28]. The cationic nature of chitosan allows for electrostatic interaction with negatively charged species, triggering different biological and chemical processes. Understanding chitin and chitosan molecular structure and their conformational behavior in solution is essential to understand their mode of action and establish structure-function relationships. To date, structural and biological studies were hampered by the lack of synthetic methods that can provide standards with full control over length and pattern. Chemical or enzymatic *N*-(de)acetylation are common manipulations, but in most cases yield ill-defined products with varying DP and DA[29]. Biotransformations of chitin to well-defined chitosans through enzymatic deacetylation can be achieved with chitin deacetylases[30], but only few of the required enzymes are available. Chemical synthesis using orthogonal protecting groups[31] and glycosylation conditions[32] offers an alterative to access COS, albeit is laborious and limited to simple motifs.

### 1.1.3. Ionic oligosaccharides

Ionic polysaccharides are ubiquitous in living organisms where they regulate a multitude of cell functions. While only two monosaccharides in eukaryotes are ionic (hexuronic and sialic acids), there is a plethora of biotransformations to install an ionic group on the sugar backbone such as sulphate, pyruvate, phosphate, phosphorylcholine, phosphoethanolamine, and aminoethylphosphonates[33]. Bacteria are generally covered by a layer comprised of tightly packed repeating ionic polysaccharides[34]. These polysaccharides are often zwitterionic at physiological pH, featured by positively charged amino groups and negatively charged carboxylates or phosphates. The ionic moieties control the shape of such compounds by promoting repulsive and attractive interactions[35]. Sulfated glycans are a structurally complex and widely diverse class of carbohydrates. In mammals, glycosaminoglycans (GAGs) are the major structural components of the extracellular matrix (ECM). GAGs are linear, negatively charged polysaccharides composed mainly of repeating disaccharide units[36] (**Figure 5**). The anionic backbone, bearing sulfate moieties (except for hyaluronans), is responsible for their interactions with proteins and the regulation of a large number of biological processes[37–39]. GAGs are water retaining polysaccharides, thus modulating the hydration and the water homeostasis in tissues[40]. Among GAGs, heparin (H), heparan sulfate (HS), and heparin-like structures have been studied extensively due to their biological importance[38]. HS binds to over hundred proteins depending on sulfation pattern and chain length. Generally, a minimal length (in most cases an octasaccharide) is required to trigger specific biological events, including interaction with cytokines and chemokines growth factors[41].

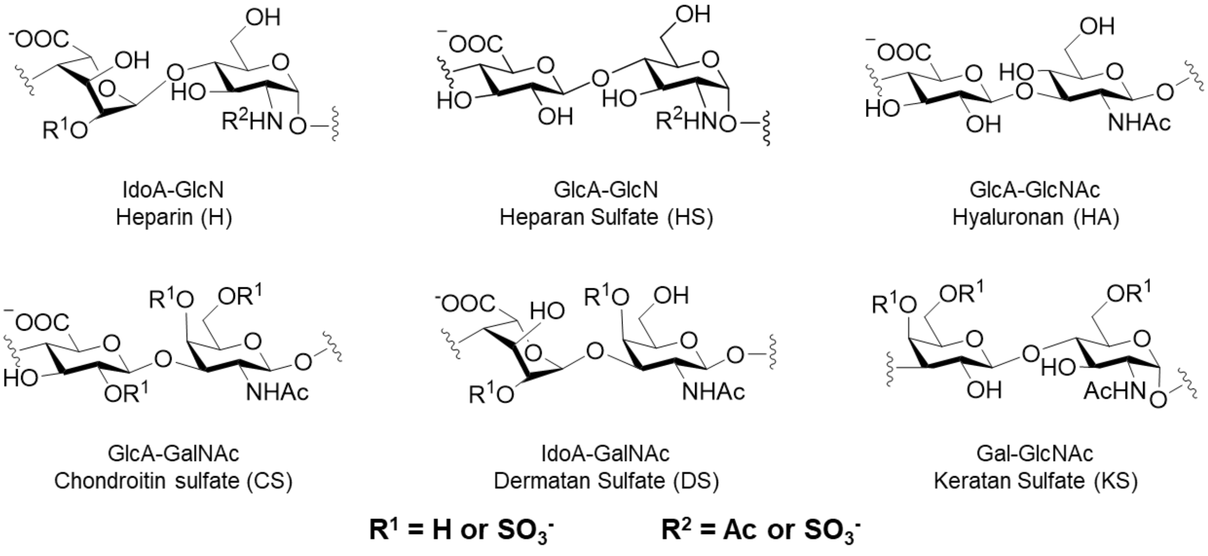


**Figure 5**. Chemical structure of the representative disaccharide units forming the GAG backbones.

GAGs can be obtained from animal sources (e.g., chondroitin sulfate is extracted from shark cartilage), albeit in heterogeneous mixtures that hampered quality control and reproducibility[42]. The heterogeneous nature of these samples is particularly problematic because the spatial orientation of the sulfate groups (sulfation code)[43,44] and the length of the polysaccharide chain affects the structural and biological features of GAGs. This became obvious during the "heparin crisis" in 2008, when batches of heparin, contaminated with oversulfated chondroitin sulfate, entered the marketplace costing the life of hundreds of people[45]. This episode underscored the urgency to develop reliable methods to access well-defined GAGs, with full control over substitution pattern and chain length. Well-defined GAGs are also valuable standards to uncover the mechanism of action of heparin-like drugs (e.g., fondaparinux).

Sulfated fucans and sulfated galactans offer a non-GAG alternative to sulfated glycans with anticoagulant heparin-like properties. Such sulfated glycans are extracted from invertebrate animals and red algal[46,47] and possess a more regular backbone composition than GAGs, often with more defined distributions of the sulfate groups.

To unlock the “sulfation code” and establish structure–function correlations, the synthesis of compounds, with well-defined composition, patterns, and lengths is needed. However, the synthesis of sulfated glycans remains extremely challenging[48]. Selective sulfation mandates the rational placement of orthogonal protecting groups, thus adding extra complexity to the already challenging synthesis of the oligosaccharide backbone. A major bottleneck arises from the lability of the sulfate groups, limiting PG manipulations or functional group transformations. Lengthy and laborious purification steps of the amphiphilic intermediates are often unavoidable[49].

### 1.1.4. Polysaccharide synthesis

Polysaccharides are mostly extracted from natural sources[50]. This process is often cumbersome and low yielding, requiring extensive purifications. In most cases, carbohydrates are obtained as heterogeneous mixtures that hamper characterization, reproducibility, and quality control. The heterogeneity of naturally sourced samples poses a severe bottleneck to the molecular characterization of polysaccharides that dwarfs in comparison to other biomolecules like peptides and nucleic acids. Small contaminations can heavily affect the polysaccharides’ material and biological properties as exemplified by the “heparin crisis”. Pure compounds with full control over pattern and length are needed to study polysaccharides at the molecular level and identify structure–activity relationships. In most cases, synthesis remains the only way access to pure samples.

Enzymatic and chemoenzymatic approaches led to the heroic synthesis of complex polysaccharides such as heparin and heparan sulfate structures[51]. Enzymes offer the possibility to use unprotected carbohydrates as substrates and guarantee remarkable control over the regio- and stereochemistry during glycosylation. Despite the numerous advantages of this approach, the limited enzyme availability, the high cost of the activated sugars as well as their high specificity narrowed the substrate scope (**Figure 6a**). In particular, the high specificity has been a significant limitation to the enzymatic synthesis of unnatural oligosaccharides. Recently, the portfolio of enzymes for complex oligosaccharides synthesis has expanded, including engineered enzymes[52] capable of catalyzing reactions with unnatural building blocks.

Chemical synthesis of polysaccharides provides compounds with well-defined length and substitution pattern, but requires a substantial synthetic effort (numerous synthetic steps) and expertise (**Figure 6b**). Properly designed building blocks (BBs) with judiciously chosen orthogonal protecting groups are needed to ensure the stereo- and regioselectivity of the glycosylation reactions. Linear approaches where the BBs are added sequentially to the growing polysaccharide chain or convergent approaches that couple pre-assembled blocks have been reported. A milestone in carbohydrate chemistry was the synthesis in solution phase of a 92mer[53] mycobacterial arabinogalactan using a preactivation-based glycosylation and block coupling strategies. Still, especially for complex polysaccharide targets, difficulties associated with side products and tedious purifications steps have to be considered. For this reason, only few examples of polysaccharides prepared by chemical synthesis have been reported.

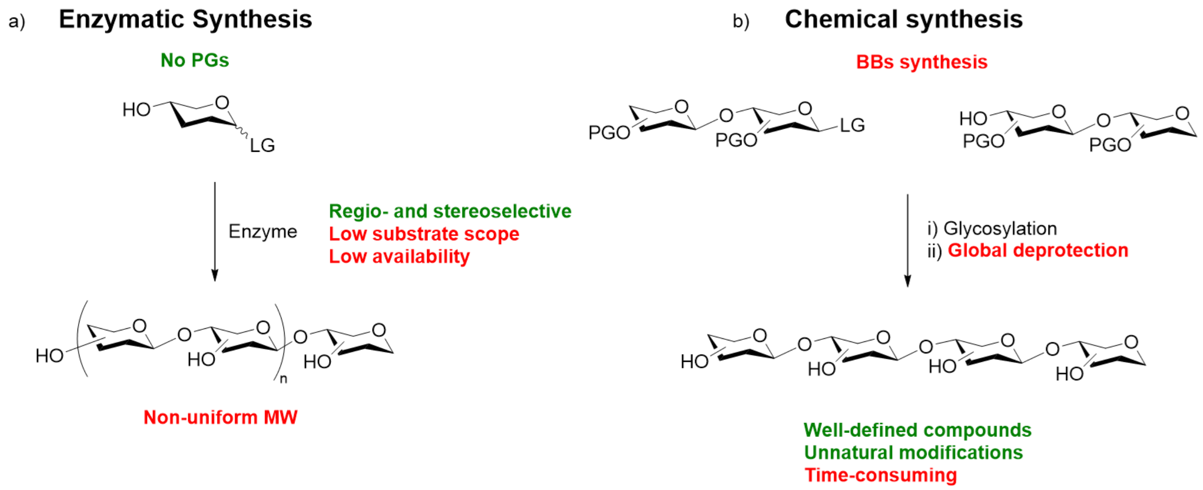


**Figure 6**. a) Enzymatic and b) chemical approach for the synthesis of polysaccharides. For each method, advantages (green) and disadvantages (red) are highlighted.

Automated synthetic platforms have been developed to reduce the tedious manual work and eliminate the potential for human error. Automated platforms based on electrochemical assembly[54], fluorous-assisted solution-phase[55], and HPLC-assisted[56] synthesis have been employed for the synthesis of oligomers up to hexasaccharides. To date, intrinsic instrumental and chemistry-related limitations impeded the exploitation of these strategies to longer and more complex structures. Among those automated platforms, Automated Glycan Assembly[57] (AGA) stood out for the fast and reliable solid-phase synthesis of long carbohydrates.

### 1.1.4.1. Automated glycan assembly (AGA)

Solid-phase synthesis has considerably facilitated access to biomolecules such as peptides[58] and DNAs[59]. Intermediates are not isolated and excess of reagents can be utilized to drive reactions to completion. By-products and excess reagents are removed by filtration and no purification is required after each step. AGA is a powerful synthetic platform for the solid phase synthesis of glycans that enabled rapid access to synthetic polysaccharides[60]. In AGA, a solid support (polystyrene-based Merrifield resin) is equipped with a cleavable linker which serve as an anchor to couple building blocks (BBs) and assemble the oligosaccharide chain (**Figure 7**). The BBs are equipped with a reactive leaving group (LG) that is chemically activated with Lewis acids and coupled to the hydroxyl group of the glycosyl acceptor. After glycosylation, the excess of BB and reactants are removed by filtration. The BBs carry temporary protecting groups (tPG), such as levulinyl group (Lev) or fluorenylmethyloxycarbonyl group (Fmoc), strategically placed for the regioselective liberation of a hydroxyl group, upon coupling. Selective cleavage of the tPG generates the new gycosyl acceptor for the subsequent glycosylation. Permanent protecting groups, such as benzyl (Bn) ethers and benzoyl (Bz) esters, remain until the end of the assembly and are removed during the final global deprotection. The AGA workflow is amenable to automation with a computer-controlled platform. Upon completion of the assembly, cleavage from the solid support is performed and methanolysis and hydrogenolysis of the remaining PGs yield the desired product. Throughout the years, different linkers have been developed such as the base-labile[61], the metathesis-labile[62], and many photocleavable linkers[63]. The photosensitive linkers are currently the most widely used. To improve the efficiency of photocleavage, a flow device is employed.

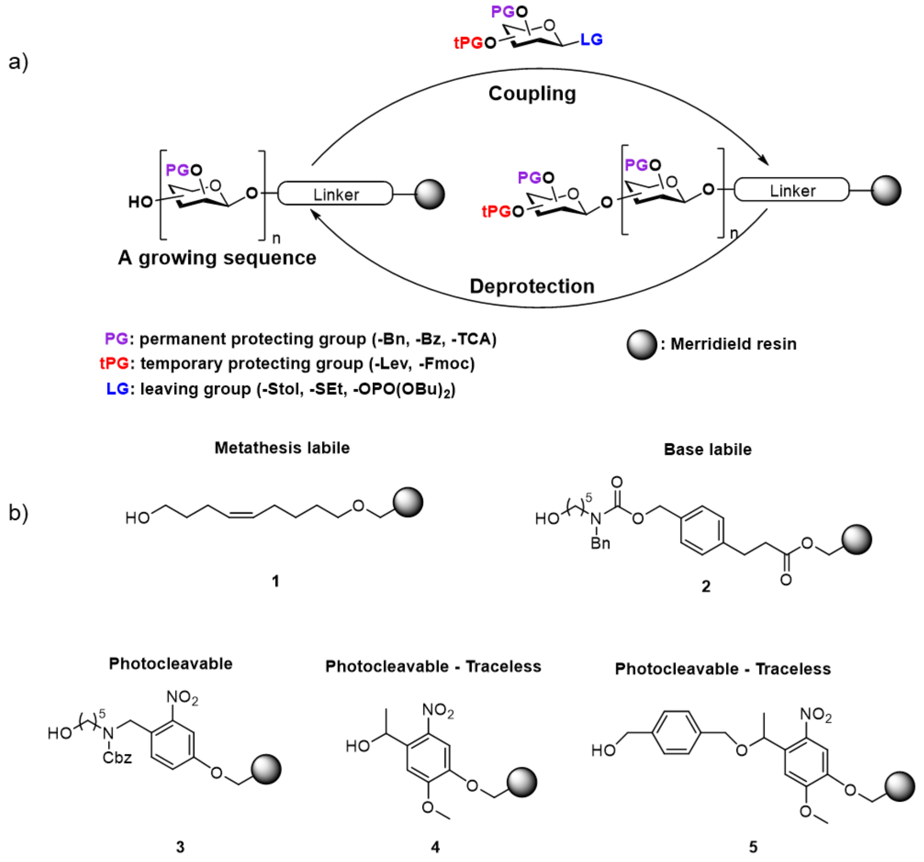
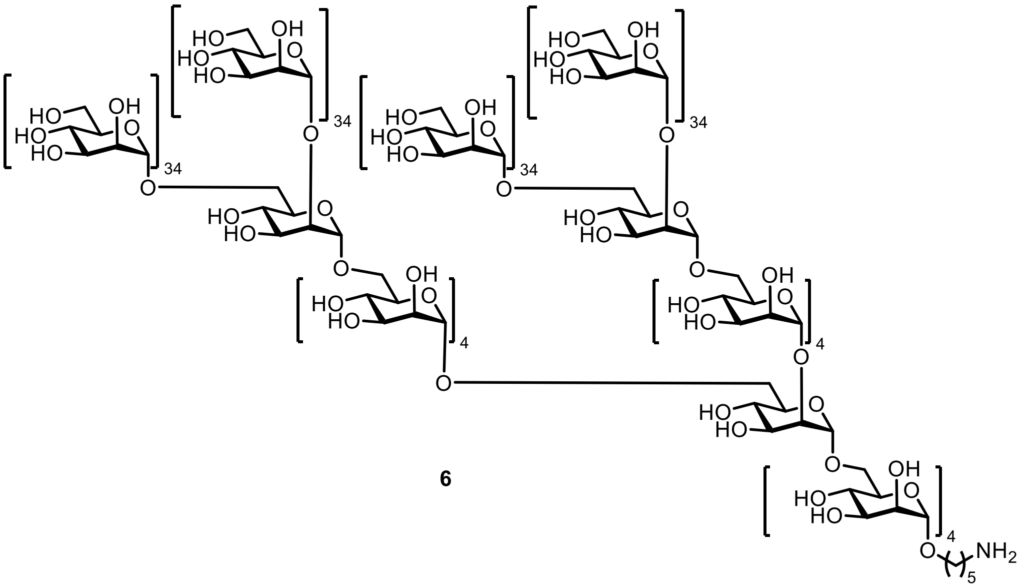


**Figure 7**. a) Synthetic cycle of AGA and b) different linkers developed for AGA.

AGA has been utilized for the synthesis of complex carbohydrate structures, including biological relevant compounds[64]. The continuous optimization and development of AGA led to the incorporation of a capping a step into the coupling cycle[65], affording the construction of linear and branched polysaccharides, up to 151-mers (**Figure 8**)[66], in high yield.



**6**

**Figure 8**. Branched 151-mer polymannoside obtained by AGA.

The AGA technology offers the opportunity to prepare large collections of related, well-defined oligosaccharides in a relative short period of time for detailed structural characterization. Preliminary studies revealed that different oligosaccharides adopt distinct secondary structures and present different geometry and flexibility depending on their carbohydrate composition[67].

## 1.2. Carbohydrate structural analysis

The function of a molecule is strongly connected to its three dimensional (3-D) structure. Thus, the elucidation of the shape and dynamics of a biomolecules is key to understand its function[68]. Many carbohydrate sequences are carriers of biological information that can only be deciphered by understanding their structure and mode of interactions with other biomolecules, such as membrane proteins. For example, lectins are able to discriminate between sugar conformers[69] stressing the importance of elucidating glycan conformation. Structural studies on synthetic zwitterionic *Streptococcuspneumoniae* serotype 1 oligosaccharides revealed that antibodies recognize particular conformations for binding[35]. A correlation between the length of glycan chain and the antibody binding affinity was demonstrated, with the highest affinity for the nonasaccharide **8**, capable of adopting a full helical turn (**Figure 9**).

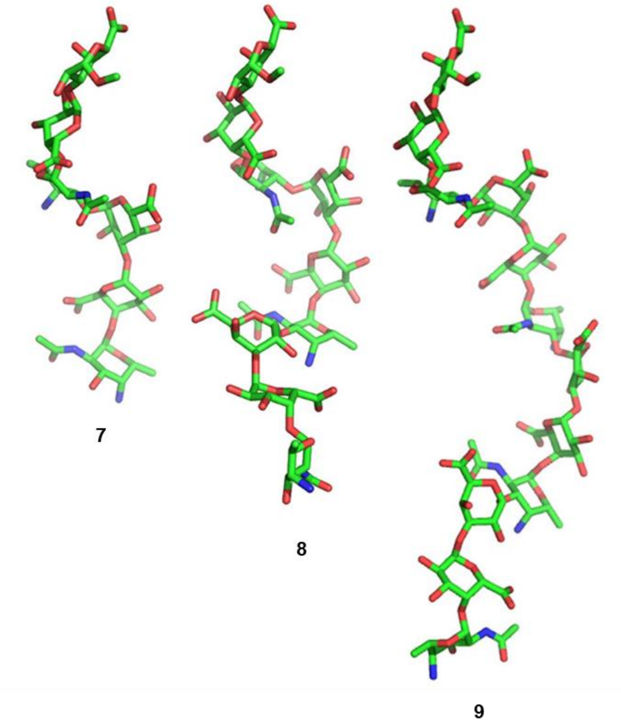


**Figure 9**. Representative conformations of zwitterionic *Streptococcus pneumoniae* serotype 1 hexasaccharide **7**, nonasaccharide **8**, and dodecasaccharide **9** as obtained by MD simulations. Reprinted with permission from Ref. [35]. Copyright, 2019, American Chemical Society

The shape of a polysaccharides tremendously affects also the aggregation process that generates different architectures. A comparison of the structure of cellulose, amylose and chitin reveals how small differences in the sugar backbone drastically affect their material properties (**Figure 10**). The purest natural form of cellulose is presented in cotton fibers whereas amylose, composed of α-linked glucose units, is a component of starch. The chemical structure of chitin shares many similarities with that of cellulose; however, the amides on chitin's backbone affect the aggregation mode and generate new materials. Studying the molecular interactions and how modifications are able to perturb the overall architecture of such biomaterials will fuel our understanding of polysaccharides and will lay the foundation for the rational design of novel carbohydrate materials.

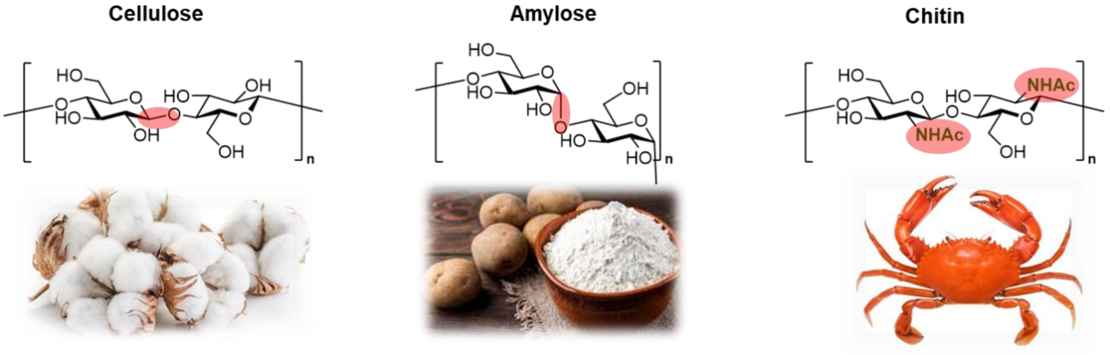


**Figure 10**. Chemical structure of the most abundant natural carbohydrate materials.

#### 1.2.1.1. Challenges in glycan structural analysis

The main variable that determines polysaccharide conformation is the geometry of the glycosidic linkage, with the monosaccharide units generally considered rigid. The torsion angles Φ (H1−C1−Ox−Cx) and Ψ (C1−Ox−Cx−Hx) define the relative orientation of two monosaccharides involved in a glycosidic bond (**Figure 11**). For (1,6)-linkages, the ω torsion angle (O6−C6−C5−O5) provides additional flexibility. The most populated conformation is generally the exo-syn(Φ)[70], due to hyperconjugation between the exocyclic oxygen lone electron pair and the antibonding orbital (σ*) of the endocyclic C−O bond (exo-anomeric effect) (**Figure 11**)[71]. The Ψ dihedral is more sensitive to sterics[72], favoring the anti (Ψ) conformer.

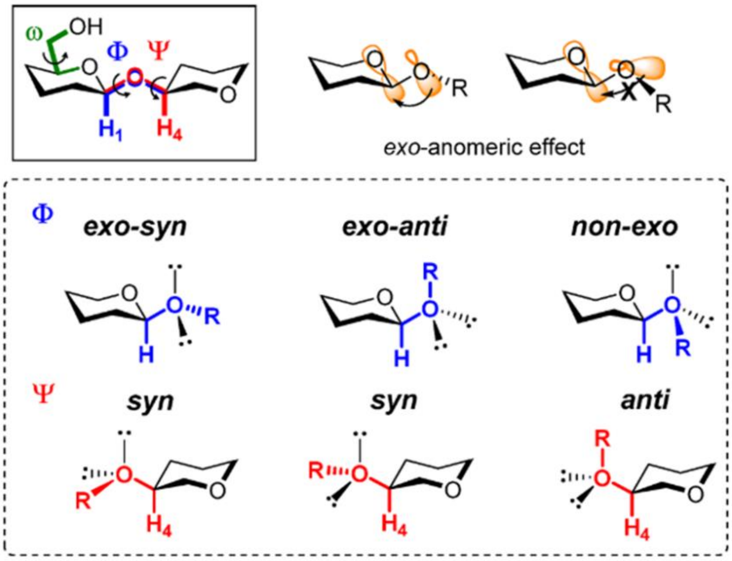


**Figure 11**. Standard definition of dihedral angles used for the description of a glycosidic bond, exemplified for a β-glycosidic linkage.

Intramolecular hydrogen bonds between hydroxyl groups may play a role in stabilizing particular conformational states of oligosaccharides, but they are not generally responsible for conformational changes of the glycosidic linkages[73]. Moreover, water contributes to the flexibility of sugars by disrupting internal hydrogen bonds[74]. Thus, glycans in solution exhibit as an ensemble of multiple conformations separated by low energy barriers, making the description of glycan structures highly challenging.

The lack of reliable characterizing methods limits the understanding of carbohydrate structures. The intrinsic flexibility of oligosaccharides impedes a folding behavior similar to proteins, which form rigid tertiary structures. Hence, sugars are difficult to crystallize and structural data from X-ray studies are rare[75]. NMR spectroscopy can only suggest an average 3-D conformation, which arise from the contribution from more than one conformational state. A major breakthrough in the structural characterization of oligosaccharides came with the implementation of single molecule imaging. Using a combination of electrospray ionization (ESI) and scanning tunneling microscopy (STM) at cold temperature, it was possible to visualize recurrent structural features of natural oligosaccharides in high resolution (**Figure 12**). The atomic level resolution obtained with ESI-STM single molecule imaging is tremendously helpful and generates important information about the shape of the oligosaccharides.

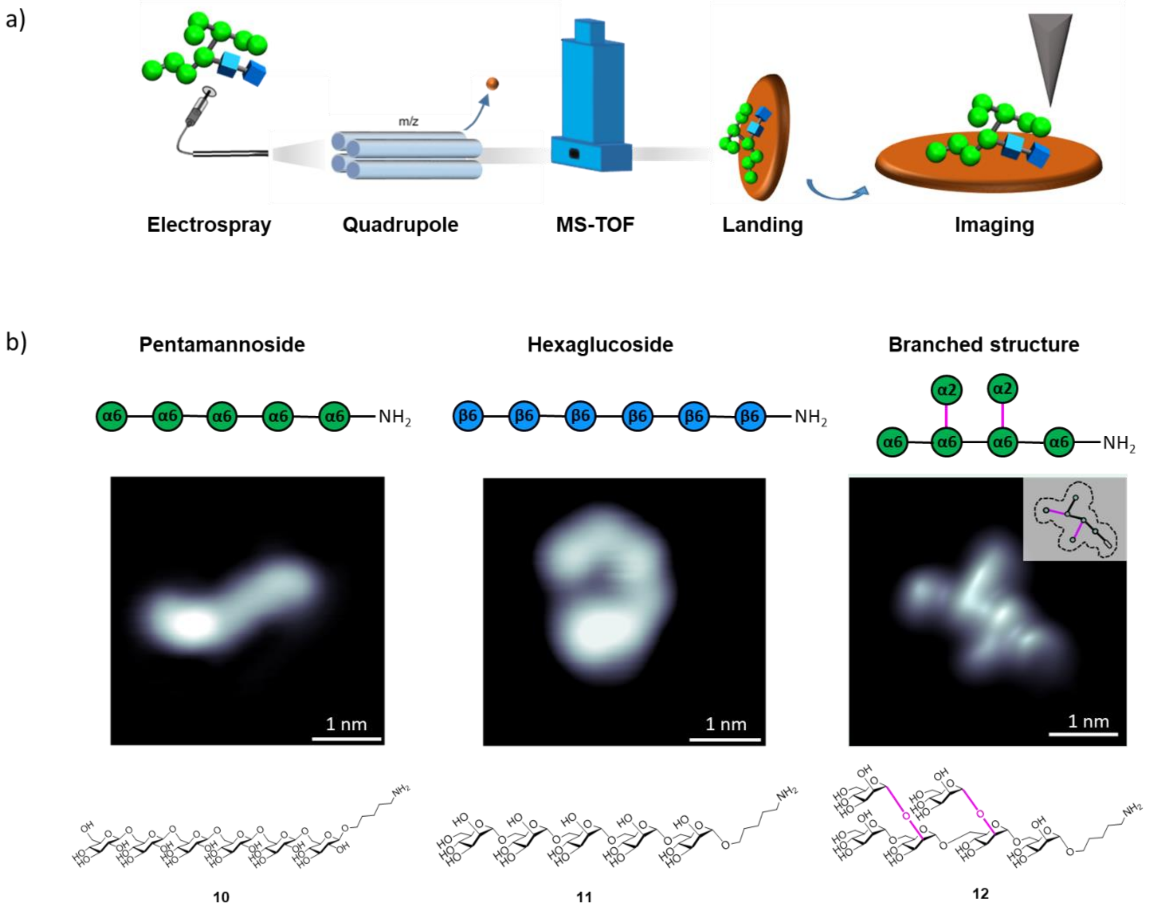


**Figure 12**. a) Schematic workflow of glycan electrospray and deposition combined with single-glycan STM imaging, and b) examples of STM topography images of glycans.

To date, most experimental methods generate either averaged ensemble structure or static results that neglect the highly dynamic nature of glycans. An attractive alternative is to model atomic-level motions computationally, based on first principles of physics, and provide a dynamic description of the system. To this end, molecular modeling in tandem with NMR has played a critical role in determining and analyzing 3-D structures of oligosaccharides[76].

## 1.3. Computational modeling of carbohydrates

The increasing attention to biomolecular simulations is propelled mainly by the need of understanding how the molecules are interacting at the atomic level, how they respond to structural perturbations, and how the environment (i.e., solvent) determines internal motions. The development of carbohydrate modeling methods was motivated by the desire to interpret solution state data from NMR spectroscopy. The importance of water to oligosaccharides conformation and dynamics push the replacement of the old Monte Carlo modeling approaches with solvated MD simulations, which incorporate advanced and refined carbohydrate force fields. Modern force fields, optimized for carbohydrates, reproduce substantially more accurate torsional potentials for α- and β-glycosdic linkages. Different water models[77] (e.g., tip3p and tip5p) have been developed to simulate aqueous systems with explicit solvent, replacing the previous implicit solvation models. The increased number of studies using molecular dynamics to simulate oligosaccharides has been fueled by the general availability of user-friendly software and the advancement in high performance computing (HPC). Although atomistic MD simulations do not model the underlying quantum physics exactly, they can provide a sufficiently close approximation and observe atomic details which are difficult or impossible to be captured by any other biophysical technique and wet-lab experiments (**Figure 13**).

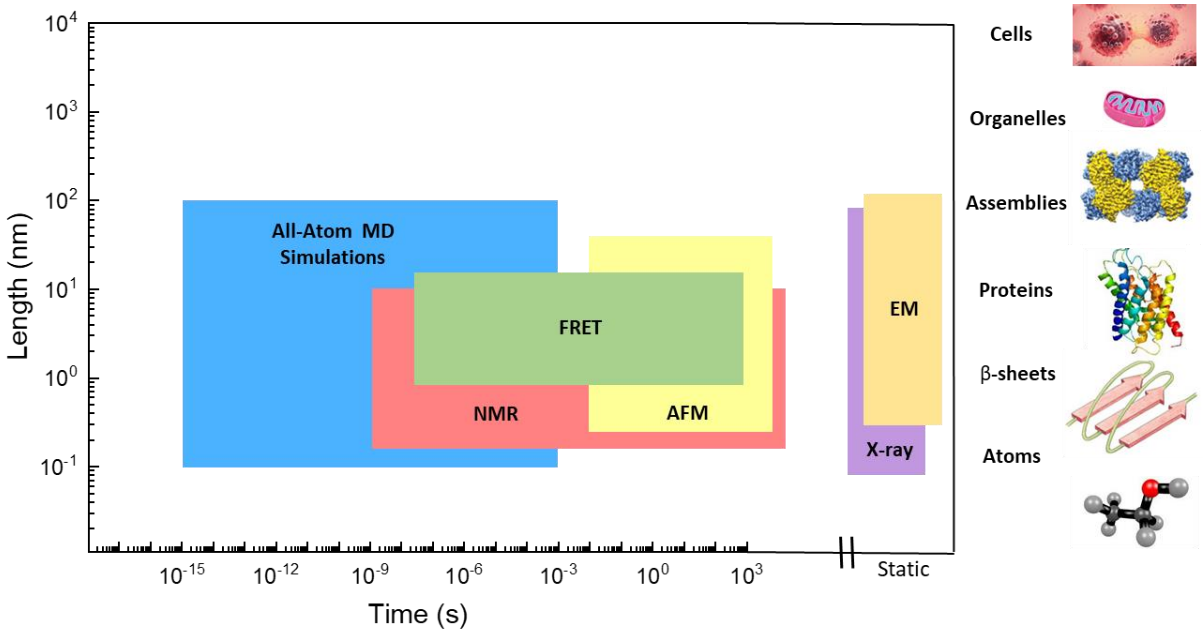


**Figure 13**. Temporal and spatial resolution of various techniques indicated by color boxes. MD simulations can probe a wide range of timescales and capture events at femtoseconds (fs) scale. For comparison, the timescale is increased for analytical techniques such as NMR, fluorescence resonance energy transfer (FRET) and atomic force microscopy (AFM).

### 1.3.1. Principles of molecular dynamic simulations

MD simulation is an all-atom deterministic approach where the atoms are presented as van der Waals spheres with electronic point charges, allowed to interact following the low of classical mechanics. The concept behind MD simulations relies on the repeating calculation of the displacement of the atoms as a function of the simulation time (**Figure 14a**). The positions and the velocities of the atoms are calculated by integrating Newton’s equation of motion. The forces acting on the atoms derive from the overall potential energy, which is described by a set of interactions and corresponding parameters referred to as the force field. The total potential energy includes bonded and non-bonded terms to describe covalent bonds and electrostatic interactions, respectively. The non-bonded interactions are most commonly given by the Lennard-Jones (LJ) and the Coulomb potential (**Figure 14b**).The parameters in the force field are obtained from experimental values and high-level quantum mechanical calculations on simple molecules and transferred to more complex systems.

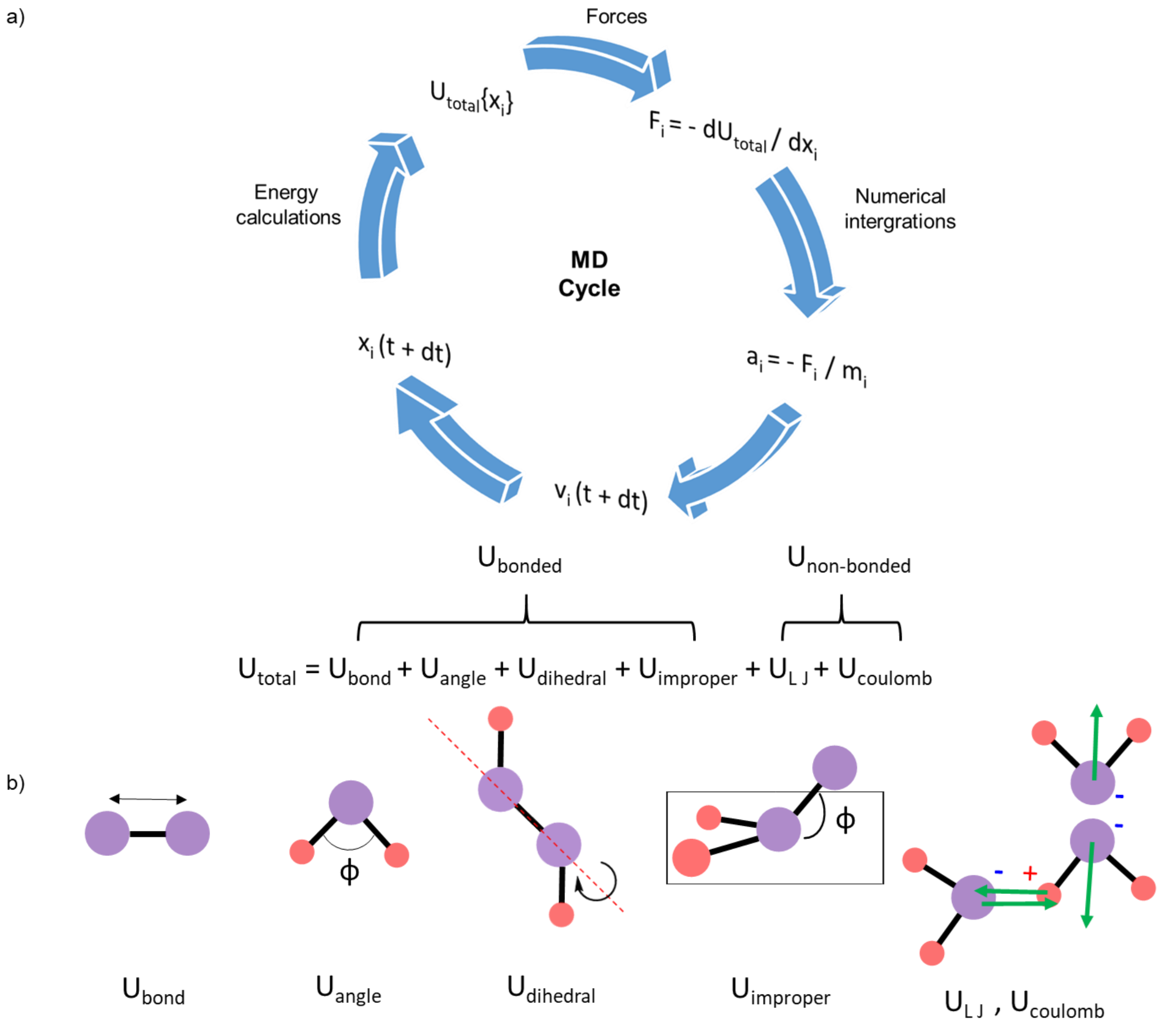


**Figure 14**. a) Basic molecular dynamics algorithm, b) the total potential energy, required for the calculation of the forces, is a sum of the energy of the corresponding bonded and non-bonded interactions.

MD simulations give a detailed picture of how a system changes from one conformation to another over a period of time and are often used in combination with a wide variety of experimental structural biology techniques, including small-angle X-ray scattering[78] (SAXS), cryo-electron microscopy[79] (cryo-EM), nuclear magnetic resonance[80] (NMR), and electron paramagnetic resonance[81] (EPR).

Software that parallelize MD force calculations across multiple computer processors enable simulating systems containing thousands of atoms for simulation times ranging from a few picoseconds to microseconds[82]. While these numbers are certainly respectable, biomolecules may undergo dynamic events on longer time scales. The current timestep for MD relies in femtoseconds and guarantees the conservation of energy and a smooth process of calculations. The increase of the timestep is one of the main challenges in computational chemistry. Recently developed super computers[83] overcome this problem by increasing the computational power. Nowadays, all-atom MD simulations can be performed by super machines, at up to 25 microseconds per day for $5x10^5$ atoms using more than 512 processing cores.

### 1.3.1.1. MD analysis of carbohydrates

Within the framework of this thesis, atomistic MD simulations are performed employing a modified version of the GLYCAM06 carbohydrate force field[70] to assess the conformational space of oligosaccharides. The Groningen Machine for Chemical Simulations (GROMACS) package permitted the analysis of the results. The root-mean-square deviation (RMSD), radius of gyration (ROG) and the end-to end distance are calculated to evaluate the dynamic nature and the conformational distribution around the mean value (**Figure 15**). Additional determinants of the glycan molecular shape are the dihedrals of the glyosidic linkages (Φ and Ψ) and therefore are investigated thoroughly. The side-chain hydroxyl group of carbohydrates populates generally three conformations called *gauche,gauche* (*gg*), *gauche,trans* (*gt*), and *trans,gauche* (*tg*). Monitoring these dihedral rationalizes structural preferences and captures unique inter and intramolecular interactions. Simulations with multiple solutes are prepared to resemble a concentrated environment. Such simulation will fuel our understanding on how carbohydrates are interacting with one another and, most importantly, if the structural changes observed in single solute are transferable to crowded systems.

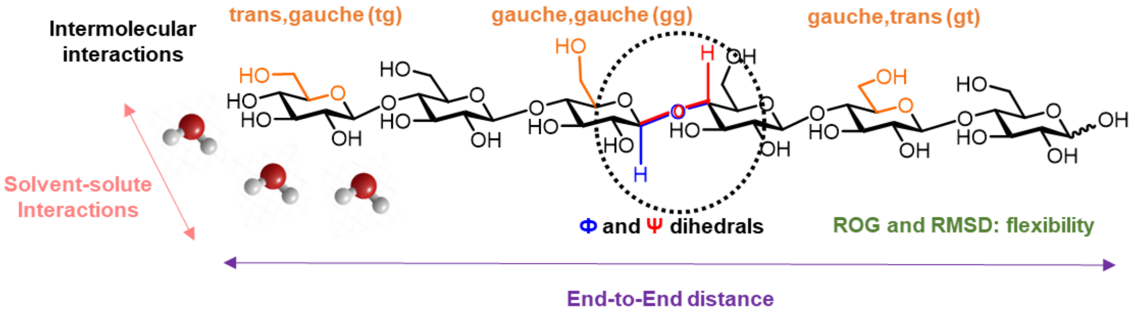


**Figure 15.** Conformational determinants of the glycan molecular shape.

## 1.4. Aims of this thesis

The general aim of this thesis was to establish correlations between structure and properties of carbohydrates, necessary for the future rational design of novel carbohydrate-based materials. Particular focus was given to the elucidation of how specific modifications, common in Nature or artificially designed, affect the structure and properties of polysaccharides. To fulfill this aim: 1) I designed MD simulations to address conformational aspects of oligosaccharides and guide their synthesis; 2) I optimized synthetic procedures to produce well-defined compounds; 3) I generated collections of complex oligomers to study their interaction with other biomolecules.

The first aim of this thesis was to evaluate how single site modifications perturb the 3-D structure and consequently the aggregation properties of cellulose and chitin analogues. My main contribution to this project was to use Molecular Dynamics to guide the synthesis and simulate a collection of tailor-made cellulose and chitin/chitosan analogues to analyze their conformational space and aggregation tendency. To explore the interaction occurring between oligosaccharides, I developed a simulation in crowded environment of solutes.

This project required the implementation of the following steps:

i) BBs bearing unnatural modifications need to be synthesized and a collection of cellulose/chitin oligosaccharide analogues with specific modification patterns has to be designed. The synthesis was a collaborative effort with several colleagues.
ii) MD simulations need to be performed to guide the synthesis and analyze the results. The oligomers need to be analyzed in terms of global conformation using end-to-end distance, the radius of gyration (ROG) and the root-mean-square deviation (RMSD). The local conformation needs to be evaluated monitoring the changes of the dihedral values.
iii) MD simulations of a crowded environment need to be designed to study the aggregation tendency of the oligosaccharides.
iv) NMR spectroscopy needs to be implemented to assess structural changes and corroborate the theoretical models.
v) The collections of oligosaccharides need to be analyzed in terms of solubility and crystallinity (XRD) and the results compared to the simulations.

The second aim of this thesis was the development of a general protocol for the on resin synthesis of sulfated glycans that bypasses many of the previous challenges associated with sulfated oligosaccharide synthesis.

To this end:

i) Monosaccharide BBs need to be designed and synthesized.
ii) The syntheses of many oligosaccharide backbones need to be performed by AGA.
iii) Post-AGA on resin chemical manipulations need to be implemented to generate a new flexible protocol for the synthesis of sulfated oligosaccharides. In particular, the on resin sulfation reaction and the further hydrolysis of ester PGs need to be optimized.

iv) The synthesis of a diverse collection of sulfated glycans need to be performed to demonstrate the scope of the new procedure.

The third goal of this thesis was the synthesis of zwitterionic oligosaccharides, naturally produce by some bacteria as part of their protective biofilms, to understand their role in extracellular matrix assembly.

To this end:

i) A collection of well-defined oligosaccharides with full control over length and modification pattern needs to be designed.
ii) Synthetic procedures to obtain the zwitterionic oligosaccharides need to be developed.
iii) A synthetic amylogenic peptide present in bacterial biofilms need to be prepared.
iv) An assay to generate and analyze artificial biofilm models had to be developed (by Soeun Gim).
v) NMR spectroscopy needs to be used to shed light on the interactions between the major components (i.e., oligosaccharides and peptides) of the artificial biofilm.
vi) The mechanical properties of the artificial biofilms had to be measured and correlated to the chemical structure of the oligosaccharide components (by Soeun Gim).

# 2. Systematic hydrogen bond manipulations to establish polysaccharide structure-property correlations

This chapter has been modified in part from the following article:

Y. Yu,* **T. Tyrikos-Ergas**,* Y. Zhu, G. Fittolani, V. Bordoni, A. Singhal, R. J. Fair, A. Grafmüller, P. H. Seeberger, M. Delbianco, Systematic Hydrogen Bond Manipulations to Establish Polysaccharide Structure-Property Correlations. *Angew. Chem. Int. Ed.,* **2019**, 131, 13261.

Y. Zhu, **T. Tyrikos-Ergas**, K. Schiefelbein, A. Grafmüller, P.H. Seeberger, M. Delbianco, Automated access to well-defined ionic oligosaccharides. *Org. Biomol. Chem.*, **2020**, 18, 1349-1353.

K. Anggara, Y. Zhu, G. Fittolani, Y. Yu, **T. Tyrikos-Ergas**, M. Delbianco, S. Rauschenback, S. Abb, P.H. Seeberger, K. Kern, *Proc. Natl. Acad. Sci.,* **2021**, 118, e2102168118.

## 2.1. Introduction

The structure of cellulose has been a subject of intensive research[13]. The stability, crystallinity, and poor water solubility of cellulose are the result of a dense network of inter- and intramolecular hydrogen bonds, dispersion forces, and hydrophobic effects that create allomorphs with different properties[10]. In particular, the hydrogen bond between the OH(3) and the O(5) of the ring stabilizes the cellobiose repeating unit, with additional stabilization gained from interactions involving OH(6) and OH(2). The apolar faces of the glucose units consisting of CH groups additionally contribute to the chain stacking through hydrophobic interactions. Cellulose has been subjected to a wide range of chemical modifications[84] to alter the properties and improve its processing with nonpolar matrices[85]. Non-regioselective derivatization results in diverse materials with respect to degree and modification patterns that do not allow for proper structure-function correlations. Efforts towards understanding the effect of the modifications at the atomic level using computational techniques were hampered by the lack of well-defined standards to experimentally validate the predicted conformational changes.

Well-defined natural and unnatural glycans are useful probes for systematic structural investigations. A collection of oligosaccharide prepared by AGA revealed that hexasaccharides adopt distinct secondary structures depending on their primary sequence[67]. Here, cellulose was selected as model system, to investigate how single site modifications are able to perturb the single chain geometry and consequently affect the materials properties. Tailor-made cellulose derivatives were designed to selectively disrupt H-bond networks and/or alter the electronic properties and establish structure-property correlations. Charges were introduced to the system prompting the formation of new interactions that could potentially affect the aggregation properties (**Figure 16**).

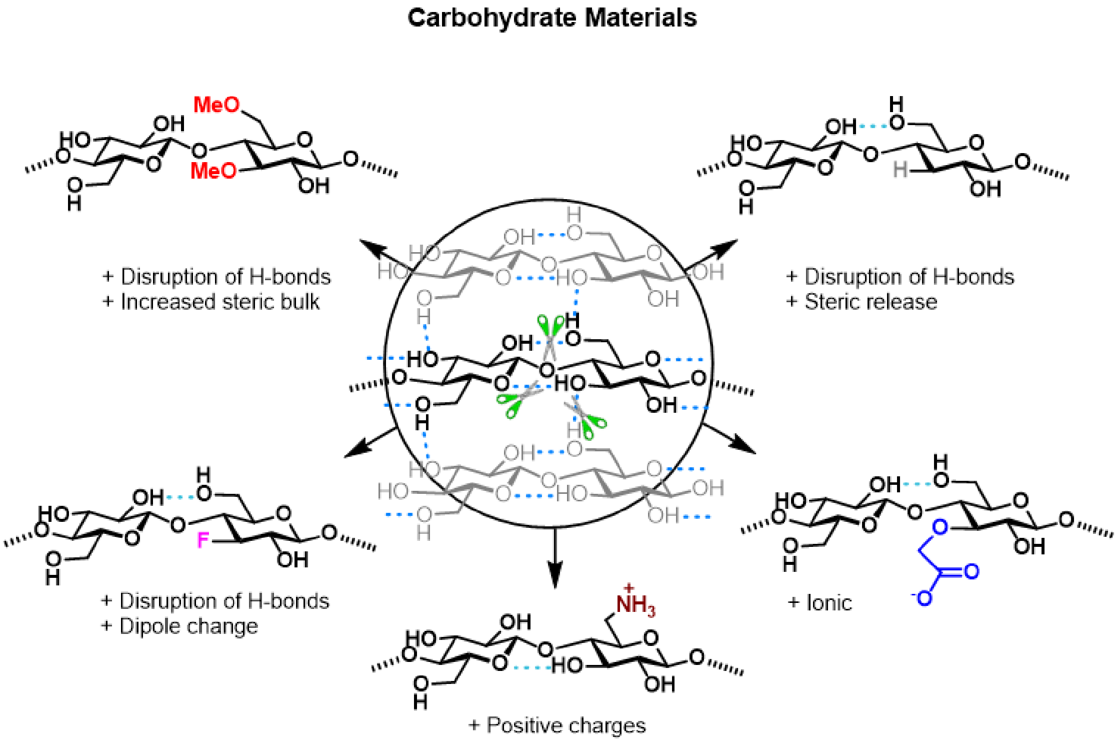


**Figure 16.** Chemical modifications (symbolized by green scissors) can be used to manipulate the hydrogen bond network and/or alter the electronic properties of cellulose.

This project was a collaborative effort of several scientists (**Table 1**); in this chapter, the major focus will be given to my contribution. Methylated, carboxymethylated, deoxygenated, deoxyfluorinated, as well as deoxyamino analogues, are prepared with full control over the length, pattern, and degree of substitution. MD simulations guided the synthesis by correlating the disruption of the H-bond network with the increased flexibility of the glycan chain and predicting the 3-D shape of the modified oligosaccharides. Structural analysis revealed the important role of the single-site modifications, responsible for conformational changes.

**Table 1**. Contribution table

| Contribution | Name |
|---|---|
| MD simulations | Theodore Tyrikos-Ergas |
| Synthesis of methylated cellulose, solubility test, XRD measurements | Dr. Yang Yu |
| Synthesis of deoxygenated, carboxymethyl and deoxyamino cellulose | Dr. Yuntao Zhu |
| Synthesis of chitin analogues | Dr. Vittorio Bordoni |
| Synthesis of deoxyfluorinated cellulose | Giulio Fittolani |

## 2.2. Results

### 2.2.1. Synthesis of modified cellulose

A collection of well-defined cellulose derivatives was prepared using AGA (**Figure 17, Table 1**). Two natural cellulose oligomers (hexamer $\mathbf{A_6}$ and dodecamer $\mathbf{A_{12}}$) served as standards for the structural analysis. Unnatural analogues with defined substitution patterns were prepared to tune the conformation and properties of the material. Regioselective functionalization was achieved with nine “unnatural” monosaccharide building blocks **14**-**22** (**Figure 17**). Each BB is equipped with a reactive thioglycoside leaving group and a temporary Fmoc protecting group that is easily removed after glycosylation to release a free hydroxyl group that serves as the new glycosyl acceptor in the next coupling cycle. Iterative glycosylation and deprotection cycles allow for the step-wise elongation of polysaccharides and the insertion of specific modifications in defined positions of the chain. The fully protected glycan target with a free reducing end is released from the solid support upon cleavage of the UV-labile linkers **4** or **5**. Global deprotection afforded oligosaccharide derivatives with complete control over the length, pattern, and degree of functionalization.

Six hexa- and four dodecamers, with different methylation patterns, were synthesized using **15** and **16**, that contain the 3-methyl and 3,6-dimethyl motifs. The position of the substituents was chosen to selectively disrupt H-bonds that play a fundamental role in cellulose rigidity. Methylation of OH(3) impedes the H-bond between O(5) and OH(3), while 6-methylation hinders the inter- and intra-chain stabilization offered by OH(6). Structures with a regular methylation pattern (e.g., $\mathbf{(AB)_3}$), di-block analogues (e.g., $\mathbf{A_3B_3}$), as well as irregularly functionalized structures (e.g., $\mathbf{(ABA)_2}$) were assembled to assess the effect of methylation patterns on the overall cellulose conformation. Similar considerations were followed for the synthesis of 3-deoxyfluorinated and 3-deoxygenated cellulose analogues. Such modifications, in addition to selectively disrupting H-bonds, are expected to modulate the steric hindrance and dipole orientation within the sugar unit. Additionally, carboxymethylation and deoxyamination were introduced to assess the effect of charges on the overall structure (**Figure 17,Table 1**).

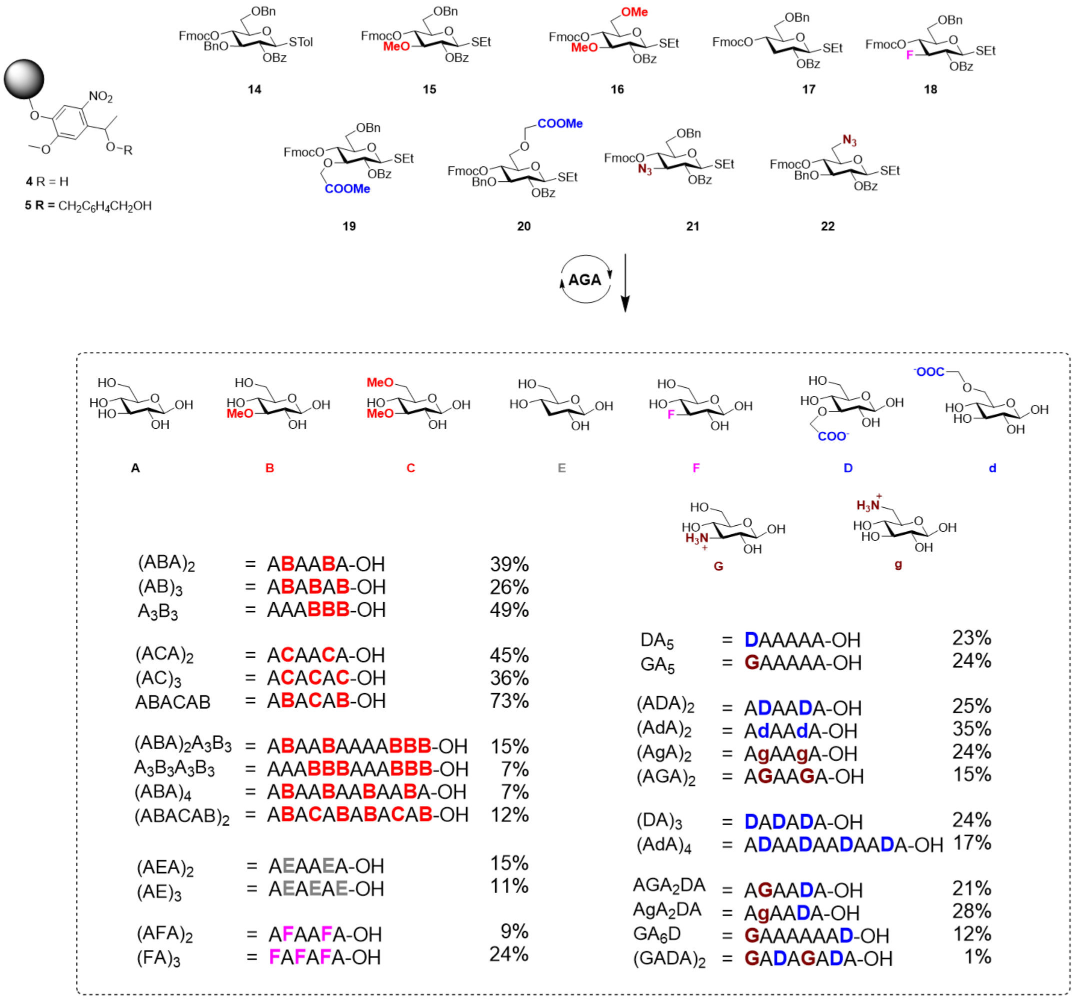


**Figure 17**. AGA synthesis and nomenclature of tailor-made cellulose oligosaccharides

## 2.2.2. Structural evaluation of tailor-made cellulose analogues

The perturbation of the 3-D shape of the oligosaccharides as a result of single-site substitutions was modelled using MD simulations, employing a modified version of the GLYCAM06 carbohydrate force field[70,86]. The effect of the neighboring monomer's substitution on the torsion angles (ω, Ψ, Φ) was monitored and compared with the unsubstituted analogue $\mathbf{A_6}$. Particular attention was paid to the changes in the population of Ψ, directly related to the presence of a hydrogen bond. To monitor the overall conformation of the hexamers, the end-to-end distance and the ROG were calculated as a function of time.

A preliminary analysis of Cremer-Pople parameters[87] showed surprisingly frequent $^4C_1$-to-$^1C_4$ interconversions for all methylated analogues (**B** and **C**), during the simulation time (even at the monomer level). Different simulation conditions using a 3-site water model (tip3p) or modifying the definition of dihedral angles did not lead to major changes. To support this observation with experimental data, NMR analysis was performed to study the chair conformation of a glucose monomer and its methylated analogue (**Figure 18a**). The values of $^3J_{H1H2}$ and $^1J_{C1H1}$ couplings (**Figure 18b-d**) disproved the presence of $^1C_4$ chair conformation and dihedral restraints were applied to these monomers (**B** and **C**), to prevent the "flipping- chair" artifact.

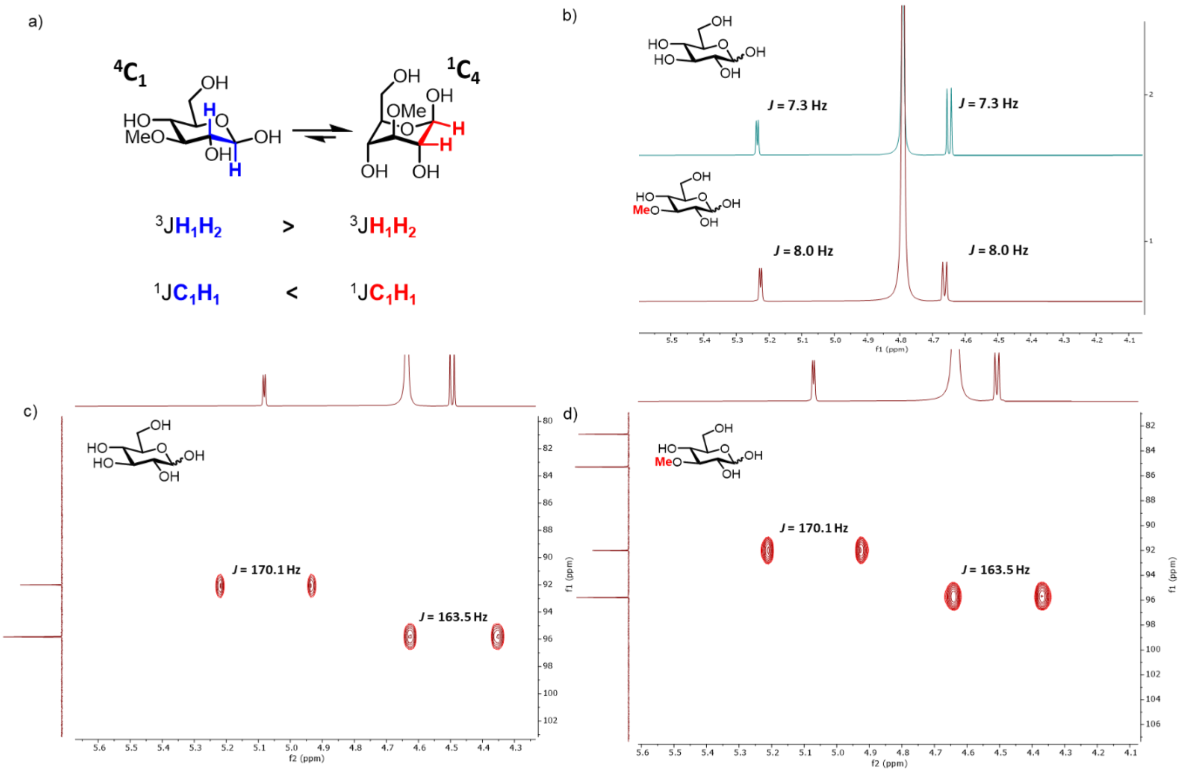


**Figure 18**. a) $^4C_1$ and $^1C_4$ chair conformations and comparison of $^3J_{H1H2}$ and $^1J_{C1H1}$ couplings between the different conformers, b) stacked $^1H$ NMR spectra of glucose ($J$ = 7.3 Hz) and methylated glucose at OH(3) ($J$ = 8.0 Hz), c) C-H coupling for glucose ($J$ = 170.1 Hz, $J$ = 163.5 Hz) and d) for methylated glucose at OH(3) ($J$ = 170.1 Hz, $J$ = 163.5 Hz), respectively.

Cellulose **$A_6$** revealed a fairly rigid backbone core with low conformational variability (average end-to-end distance 2.76 ± 0.17 nm) (**Figure 19b**) and tends to adopt an extended helical conformation. To examine how specific modifications affect such organized structures, the series of methylated analogues was studied (**Figure 19**). A regular alternated substitution pattern, as in the case of **$(AB)_3$**, revealed a moderate, yet important, decrease of the population of Ψ at negative degrees (-27°), as a result of the increased distance between OMe(3) and O(5) due to the decreased tendency to form hydrogen bonds and the increased steric bulk (**Figure 19**).

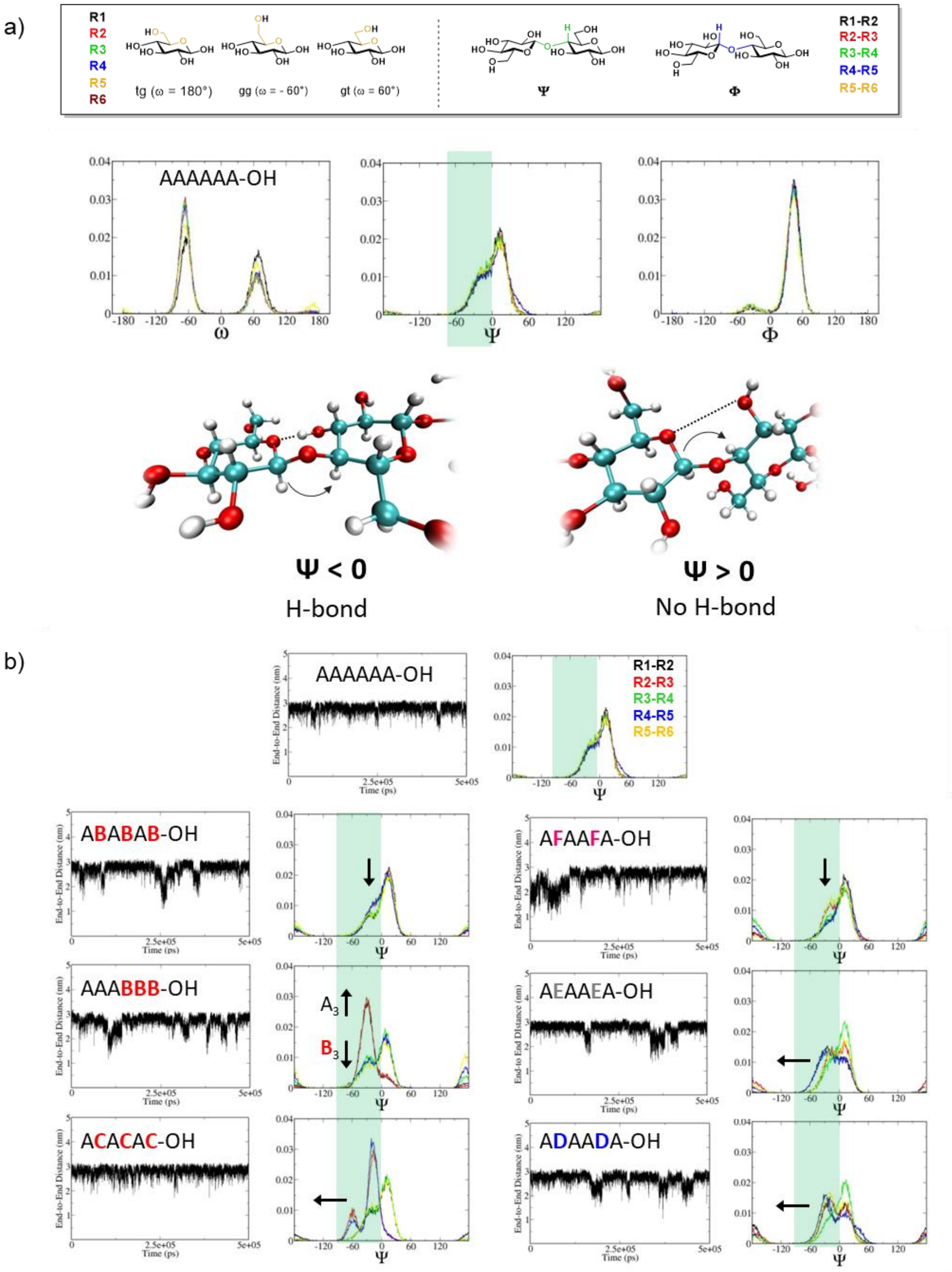


**Figure 19**. a) Analysis of the dihedral distributions obtained by MD simulations for **$A_6$**. Negative degrees of Ψ (green box) are stabilised by the hydrogen bond between OH(3) and O(5), whereas the increased distance between these two residues is reflected by positive Ψ. The residues are numbered from the non-reducing end (R1) to the reducing end (R6), b) Analysis of end-to-end distances as a function of MD time and Ψ distribution obtained by MD simulations. The end-to-end distance was monitored over 500 ns. Large fluctuations are observed for all modified analogues, indicating that these molecules are more flexible. Changes in the population of Ψ at negative degrees (green box) are correlated to changes in hydrogen bonding between OH(3) and O(5). Changes in intensity are correlated with a decreased (↓) or increased (↑) rigidity, as compared to **$A_6$**; shifts (⟵) indicate that new geometries become accessible.

The same degree of methylation with a block distribution **$A_3B_3$** resulted in dramatic changes. A significantly more flexible bent shape (**Figure 20c**) with an end-to-end distance of 2.65 ± 0.26 nm was observed for most of the simulation time. Surprisingly, the OH(3)···O(5) hydrogen bond between the first two glucose monomers was detected for most of the simulation time, suggesting the coexistence of a rigid rod block (**$A_3$**) and a very flexible counterpart (**$B_3$**).

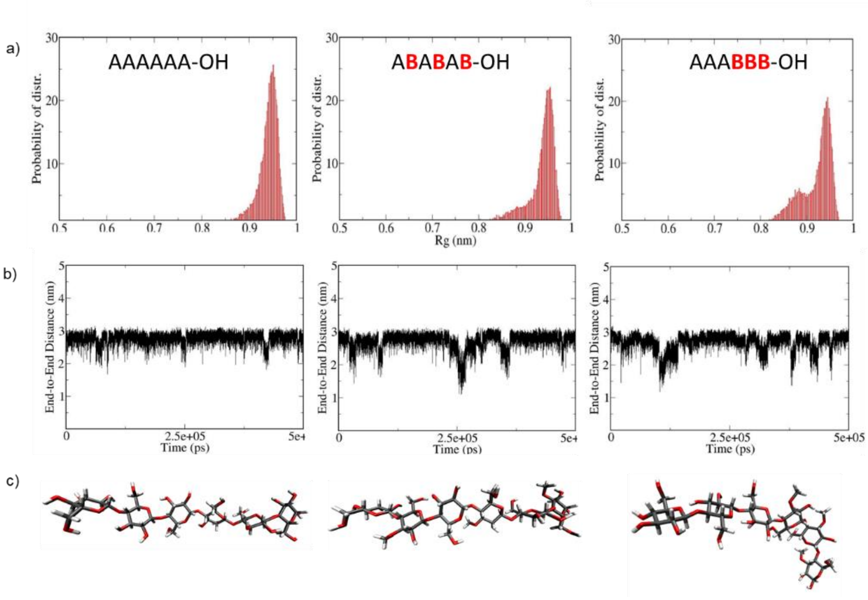


**Figure 20**. a) Radius of gyration and b) end-to-end distances as a functions of MD time, c) representative conformations as obtain by MD simulations.

Methylation at the 3 and 6 positions (**C**), aimed to reduce inter- and intramolecular hydrogen bonds, disrupts the "standard" dihedral values, resulting in a completely new geometry. Ramachandran plots of the dodecamers confirmed that increased length enhances resistance to deformation, since the cooperativity of intramolecular H-bonding interactions stabilises the overall structure. Nevertheless, a noticeable deviation from the main population of **$A_{12}$** was observed for all substituted analogues. An irregular substitution pattern appears to be important to drastically change the cellulose conformation (e.g., **$(ABA)_2A_3B_3$**). A regular substitution pattern such as **$(ABA)_4$** maintains a higher cellulose character while improving water solubility (**Figure 21**).

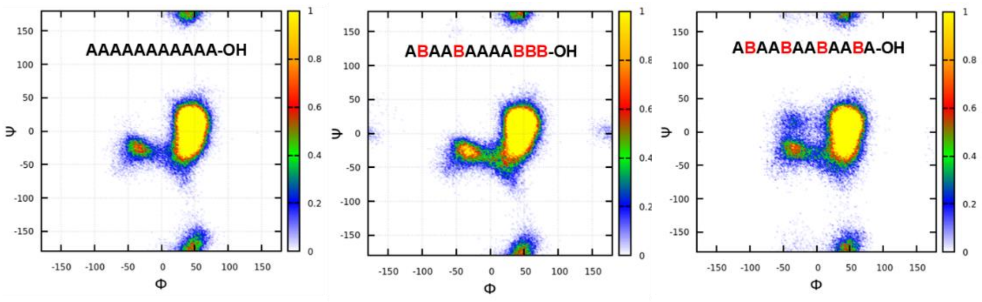


**Figure 21**. Ramachandran plots of **$A_{12}$, $(ABA)_2A_3B_3)$** and **$(ABA)_4$** dodecamers. The modified compounds reveal an increase conformational freedom compared to **$A_{12}$.**

Like methylation, deoxyfluorination and deoxygenation prevent the formation of H-bonds between O(5) and OH(3). In addition, these substitutions influence the electron density along the chain (electronegative F) and the steric hindrance (deoxygenation). Since dipoles are key to cellulose stability, the replacement of OH(3) with the electron withdrawing F is expected to greatly influence the resulting material conformation. The calculated mean RoG for **$(AFA)_2$** shows a large dispersion and the average end-to-end distance is among the lowest (2.61 ± 0.34 nm), indicative of a very flexible system (**Figure 19**) with a lower population at negative degrees for $\Psi_1$, $\Psi_4$. This effect extends beyond the single AF glycosidic bond, with significant variation of the $\Psi_3$. 3-Deoxygenation had an even bigger effect on the $\Psi$ distribution for **$(AEA)_2$**, as reduced steric hindrance allows for more conformational freedom. The insertion of a carboxylic groups (e.g., **$(ADA)_2$**) resulted in a highly flexible, mostly linear conformation (**Figure 19**). Moreover, the carboxylate can engage in additional H-bonds, as observed between $COO^-$ and OH(2) of the same residue, as well as between $COO^-$ and OH(6) of the adjacent previous sugar (OH(6)···$COO^-$···OH(2)).

All functionalized cellulose analogues increased the conformational variation and result in more flexible structures. Imaging of these oligomers (performed by Dr.Anggara) at subnanometer resolution by combining scanning tunnelling microscopy (STM) and electrospray ion-beam deposition (ES-IBD)[88] confirmed the increase conformational freedom of the modified compounds predicted by MD simulations. The hexamer of cellulose **$A_6$** mainly adopts a straight geometry, while the substituted analogues adopt both straight and bent shape geometries (**Figure 22**)[89].

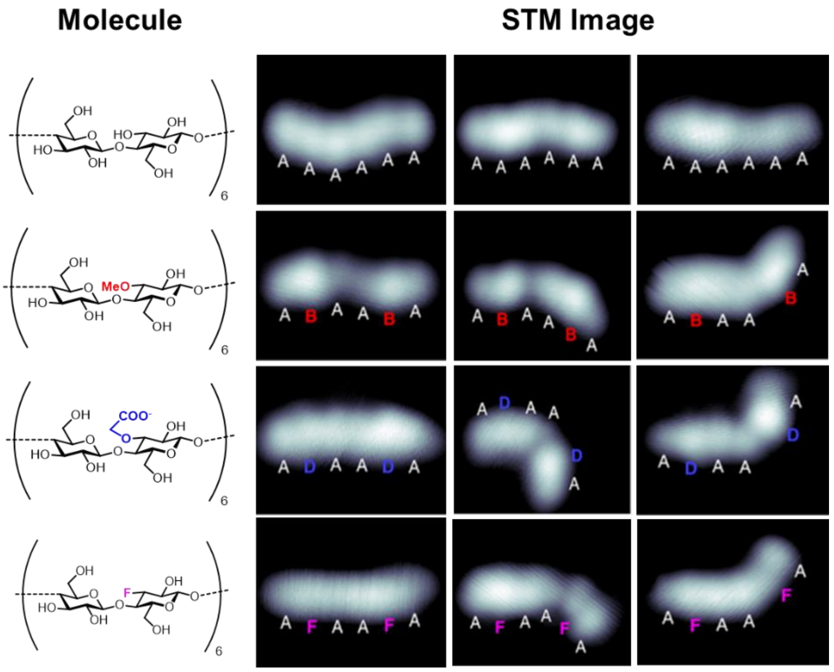


**Figure 22.** STM imaging of single cellulose chains. The modified cellulose analogues revealed an increased flexibility, adopting multiple conformations compared to **$A_6$**.

To further examine the roles played by the ionic moieties in the oligosaccharide conformation, a systematic structural study of ionic cellulose analogues was performed. MD simulations suggested that specific intramolecular hydrogen bonds stabilise particular conformations that can trigger the formation of different materials. In particular, the substitution at position 6 (**g** and **d**) produces a drastic effect on the ω torsion angles, much less affected by the modification at position 3 (**G** and **D**). The interaction between the $NH_3^+$(6) and the OH(3) of the adjacent sugar (R + 1) stabilises the *gt* rotamers (red and yellow plots, **Figure 23**). This also affects the glycosidic bond geometry with a high population at negative degrees (−172°) for $\Psi_1$, $\Psi_4$ (**(AgA)$_2$** red boxes). The opposite trend is observed when position 6 is substituted with $COO^-$ (**d**). Due to sterics, the *gg* conformation is preferred, with a small percentage of *tg* stabilised by a $COO^-$···OH(2) interaction (**Figure 23**). This interaction also affects the dihedrals of the adjacent glycosidic bond (**(AdA)$_2$** blue boxes). The substitutions in position 3 influence predominantly the glycosidic bond geometry. The interaction between the $NH_3^+$(3) and the O(5) of the previous residue (R-1) preserved a cellulose-like character (**(AGA)$_2$** red boxes). In contrast, the carboxylate at position 3 can engage in additional H-bonds, as observed between $COO^-$ and OH(2) of the same residue as well as the OH(6) of the previous sugar (R-1), resulting in remarkable changes of the Φ and Ψ dihedrals (**(ADA)$_2$** blue boxes).

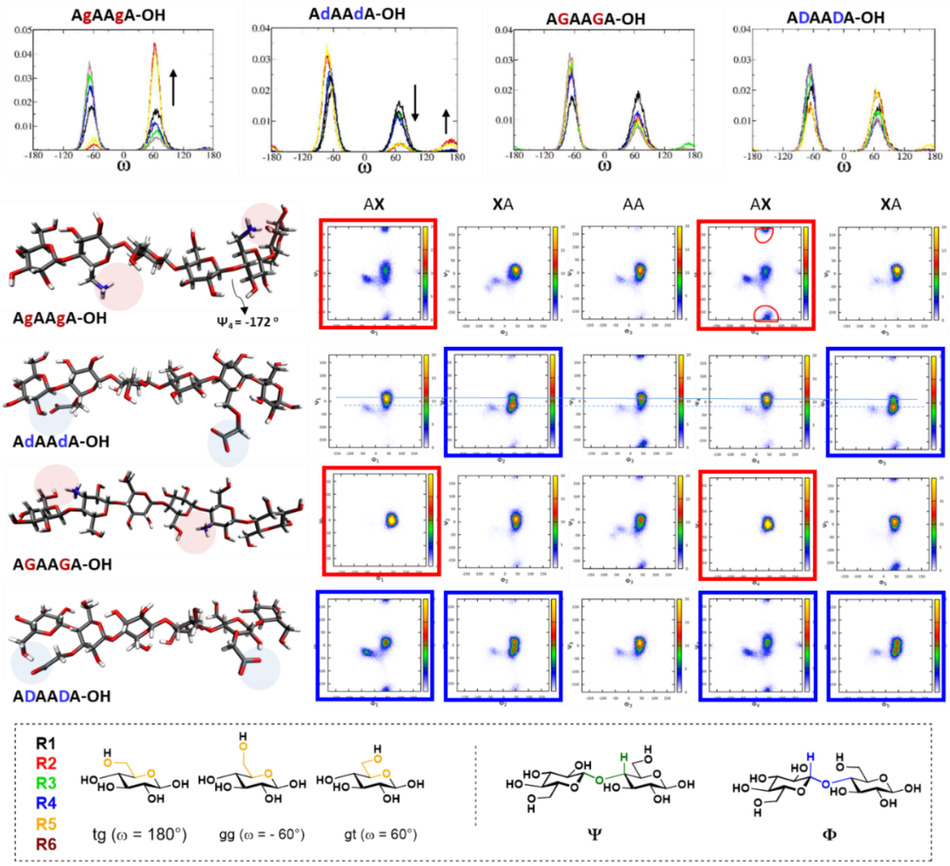


**Figure 23.** Dihedral analysis (ω, Ψ and Φ) obtained by MD simulations and representative snapshots of the oligomers showing specific intramolecular interactions (highlighted with circles) due to the modifications. The effect of the modification on the ω torsion angles is highlighted with arrows (top graphs). The most affected Ψ and Φ dihedrals are marked by red or blue boxes. For **(AdA)$_2$** the two blue lines serve as guide to the eye for comparing different minima. The residues are numbered from the non-reducing end (R1) to the reducing end (R6).

### 2.2.2.1. MD simulations of multiple solutes

Oligosaccharide behavior in a crowded environment was studied and correlated to the material crystallinity and solubility. Long MD simulations (1 µs production run) of a crowded environment of solutes (**Figure 24a**) aimed to elucidate molecular interactions. Radial distribution functions (RDFs) were used to characterize the spatial correlations in the systems (**Figure 24b**). The RDF for **$A_6$** shows three sharp signals at small distances and remains large for distances up to 1.5 nm, indicating high aggregation tendencies of such oligosaccharides. The more soluble methylated analogue **$(AB)_3$** shows some tendency to aggregate at high concentrations. However, a significantly decreased signal at 0.5 nm indicates the lower probability to find two chains in very close proximity, as compared to cellulose oligomers. RDF peaks are only found at shorter distances, revealing a lower tendency for cluster formation and a less organized structure, with a homogeneous distribution of molecules beyond the nearest neighbors. No aggregation was detected for **$A_3B_3$**, as expected from the high flexibility of such compound that should prevent chain-stacking (**Figure 24b**).

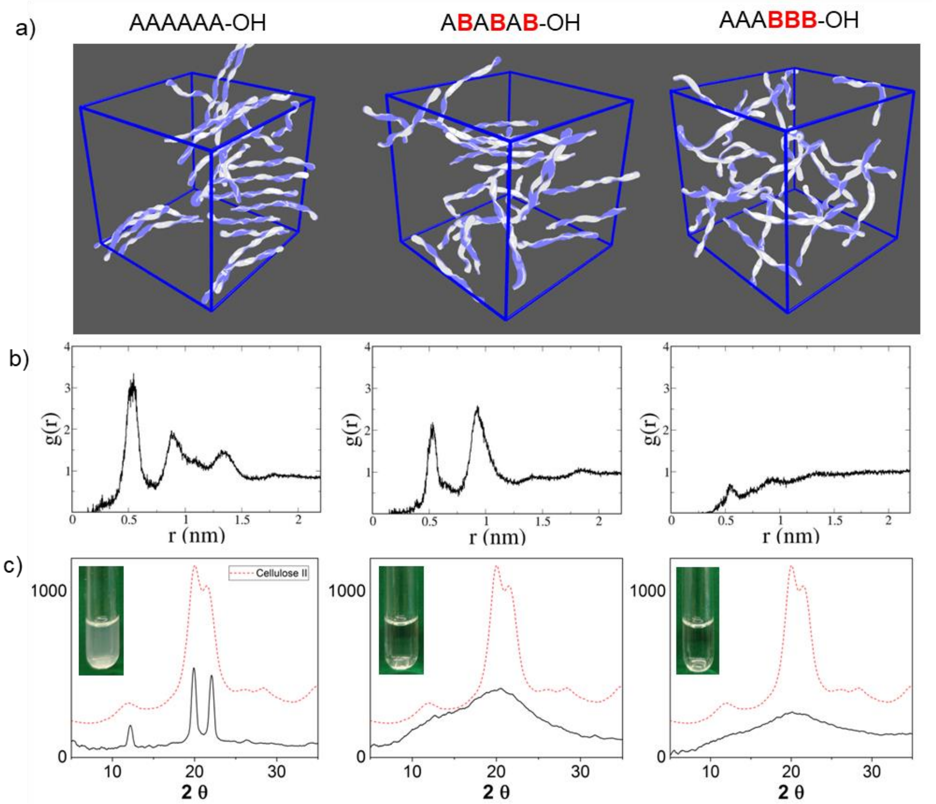


**Figure 24**. Representative snapshots of MD simulations of concentrated solutions (a), RDFs (b), XRD (c), and solubility test (inset).

X-ray diffraction and solubility data (**Figure 24c**, performed by Dr. Yang Yu) supported the calculations. As anticipated, **$A_6$** is very poorly soluble in water (less than 1 mg/mL), due to the formation of cellulose-like aggregates. Powder XRD measurements of **$A_6$** gave sharp peaks, that are distinctive for cellulose II, indicating that short oligomers adopt the same aggregation pattern and the same H-bonding arrangement of cellulose. The flat XRD profile of the di-block analogue **$A_3B_3$** indicates the absence of any ordered structural organization, as predicted by the theoretical model (**Figure 24a-b**). The alternating methylation pattern of **(AB)$_3$**, renders the material more sensitive to the X-ray beam angle and, while the XRD peaks are still broad, they resemble the cellulose II structure, as predicted by MD simulations.

The dihedral distributions in the concentrated experiments were then evaluated. The torsion angles of all the 25 oligomers were calculated to clarify whether aggregations affects the chain rigidity (**Figure 25**). A remarkable change in the population of Ψ was observed for both **A$_6$** and **(AB)$_3$**, directly related to the enhancement of the intramolecular hydrogen bonds. A considerable population at negative degrees of the Φ dihedrals was measured for **(AB)$_3$** , revealing the existence of a non-exo conformation, probably due to sterics. All the torsion angles for **A$_3$B$_3$** remained unchanged, as expected from such a flexible compound which predominantly prefers to interact with water.

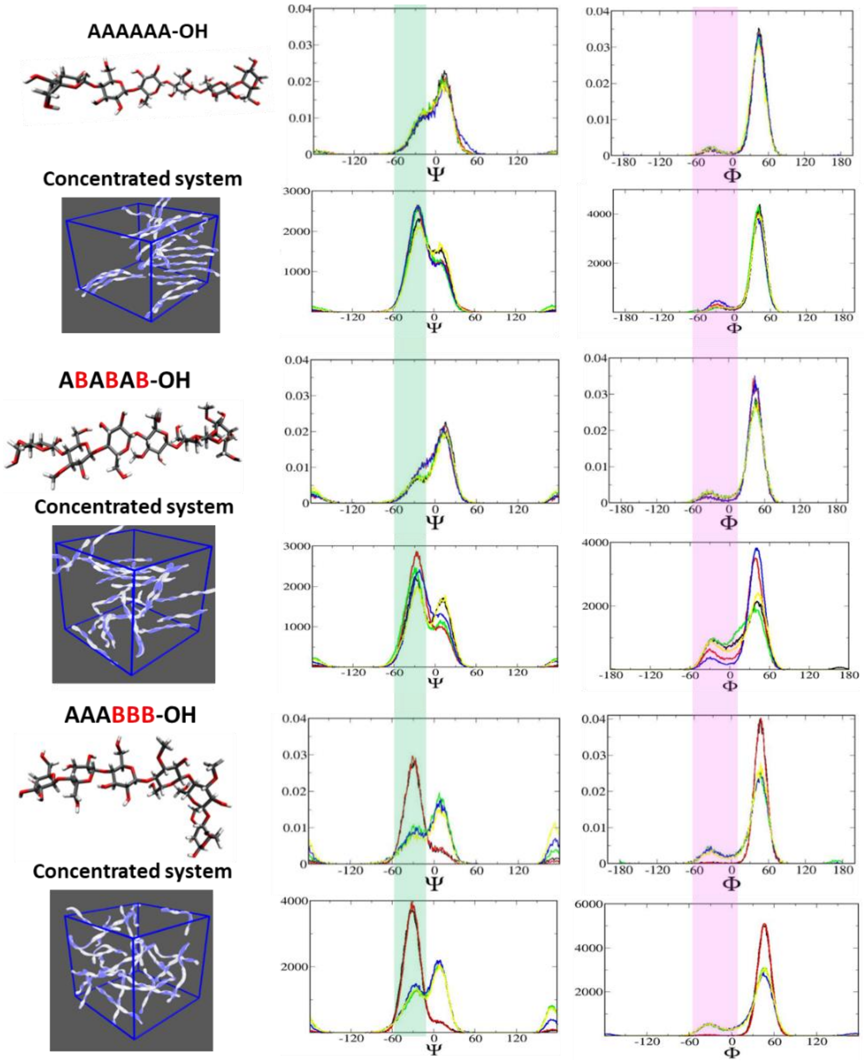


**Figure 25**. Overall dihedral analysis for the single chains and concentrated experiments of **A$_6$**, **(AB)$_3$** and **A$_3$B$_3$**. The increase of the population at negative Ψ values (green line) for the concentrated systems is related to the enhancement of the intramolecular hydrogen. The population at negative degrees of the Φ dihedrals (red line) observed for **(AB)$_3$** reveals the presence of an non-exo conformation.

## 2.3. Conclusion

In summary, I successfully designed MD simulations to address conformational aspects of oligosaccharides and guide their synthesis. A collection of cellulose analogues, prepared by sequential addition of monomeric BBs using AGA, allows for control over the length and substitution patterns. A detailed dihedral analysis depicted how each glyosidic bond is affected by the modifications, with consequences on the overall structure such as fluctuation of the end-to-end distance during the simulation time. Compounds with the same degree of methylation, but different substitution patterns, behave drastically different. Regular substitution patterns result in quasi-linear structures, whereas more bent geometries are observed with a block arrangement. These structural features control the aggregation process that is expressed in high crystallinity for the natural compound and amorphous organization for irregular or block substituted analogues. A more significant disruption of the "standard" dihedral values was observed with methylation at the OH(3) and OH(6) positions. MD simulations demonstrate how the nature (positive vs negative) and the position of the charges (3 vs 6) plays a major role in determining the flexibility and conformational preference of such oligosaccharides. These differences could play a major role in the formation of supramolecular assembly based on charged polysaccharides. All unnatural analogues are drastically more soluble, due to the more flexible backbone. Novel biomaterials with tuned properties that could be engineered depending on nature and pattern of the substituents can be envisioned.

# 3. Systematic structural characterization of chitooligosaccharides enabled by Automated Glycan Assembly

This chapter has been modified in part from the following article:

**T. Tyrikos-Ergas**, V. Bordoni, G. Fittolani, M. A. Chaube, A. Grafmüller, P.H. Seeberger, M. Delbianco, Systematic Structural Characterization of Chitooligosaccharides Enabled by Automated Glycan Assembly, *Chem. Eur. J.,* **2021**, 27, 2321-2325.

## 3.1. Introduction

In the previous chapter, I demonstrated how the nature, the position, and the pattern of a chemical modification affect the 3-D shape of cellulose. Next to cellulose, chitin is the second most abundant polysaccharide in Nature and it is composed of *N*-acetylglucosamine repeating units. The partial deacetylation of chitin results in chitosan, a pseudo-natural polycationic biopolymer. The cationic nature of chitosan confers higher water solubility, easy processability and is important for industrial applications such as coating material, ingredient in cosmetics, and pharmaceutical excipient[23]. Chitin and chitosan are commonly used to produce fibers, particles, and composites with exceptional biological and mechanical properties[90]. Degree of polymerization (DP) and fraction of acetylation (FA) offer the opportunity to tune the stiffness, solubility and the transparency of the resulting materials[91].

Chitin degradation produces chitooligosaccharides (COS). These short oligomers are known to trigger an innate immune response in humans and antifungal defense mechanisms in plants[91–93]. The DP of COS is crucial for the biological response, as size-dependent recognition was observed in plant chitin receptors as well as in toll-like receptors (TLR2)[94]. It has been suggested that the acetylation pattern (AP) of COS modulates the biological activity[95] and may explain the existence of sequence-specific chitosan hydrolases in most organisms[52].

A detailed molecular description of chitin, chitosan, and COS structure-function relation is missing, as most studies are performed with ill-defined samples[96–99]. Computationally, several all-atom models have been applied to study the conformational space of COS, showing that DP, FA, AP as well as pH strongly affect the conformation and control the aggregation. Coarse grained (CG) computational methods provide further insights on the COS interactions in solution, aiming for a description of chitin and chitosan polymers[100–102]. However, due to the intrinsic CG approximation, chemical details are lost. The lack of standards to validate the theoretical models remains the major bottleneck, leading, in some cases, to contradictory theories[101].

Well-defined samples with controlled DP, FA, and AP are important targets to shine light on molecular conformations and interaction mechanisms. COS commonly obtained by partial degradation of polymeric chitin and chitosan require extensive purification and typically exist as mixtures.[92] Chemical or enzymatic *N*-(de)acetylation are common manipulations, but in most cases yield ill-defined products with varying DP, FA, and AP. Sequence-specificity or regio-selectivity may be achieved enzymatically[31,103], but few of the required enzymes are available. To date, no general method to produce all possible patterns exists.[95] Alternatively, well-defined COS can be prepared by chemical synthesis using orthogonal protecting groups[31], but are too laborious to prepare a large collection of well-defined COS required for systematic structural studies. Solid phase based automated techniques offer the ideal solution for the quick production of large series of related compounds.[60,104] Still, their scope is limited by problems with sequences that form rigid tertiary structures, such as cellulose and chitin oligomers.[104]

In this work, a collection of hexasaccharides, including well-defined COS, as well as hybrid chitin-cellulose oligomers was synthesized by Automated Glycan Assembly (AGA) (see **Table 2**). These unnatural analogues were designed to explore the importance of the amino groups in COS. My contribution to this project was to explore how the pattern of acetylation and the introduction of positive charges from the protonated amines affect the conformation of the synthetic oligomers. Molecular dynamics (MD) simulations permitted a systematic structural analysis and NMR experiments were performed to validate the predicted structural changes.

**Table 2**. Contribution table

| Contribution | Name |
|---|---|
| MD simulations and NMR analysis | Theodore Tyrikos-Ergas |
| Synthesis of chitin/chitosan analogues | Dr. Vittorio Bordoni, Dr. Manishkumar Chaube |
| XRD analysis | Giulio Fittolani |

## 3.2. Results

### 3.2.1. Synthesis of chitin/chitosan oligosaccharides

Nine hexasaccharides are assembled by AGA employing differentially protected thioglycoside or glycosyl phosphate monosaccharide BBs (**Figure 26**). Two glucosamine BBs are designed for the introduction of either *N*-acetyl (**N**) or free (**K**) glucosamine units. The amino group in **23** is equipped with the trichloroacetyl (TCA) group, whereas **24** bears a carboxybenzyl (Cbz) group. The glucose unit (**A**), required for the chitin-cellulose oligomers, is installed using **25**. The desired sequence is assembled on solid support, following iterative cycles of glycosylation and deprotection. The protected oligomer is released from the solid support upon cleavage of the UV-labile linker **4**.[63] Hydrogenolysis removes all the benzyl (Bn) ether protecting groups and allows for the concomitant TCA reduction and Cbz cleavage, affording the desired COS with defined AP. The hybrid structures require basic hydrolysis of the benzoate (Bz) esters, prior to hydrogenolysis. Glycosyl phosphate **23b** performed significantly better as indicated by an increase in yield of $\mathbf{N_6}$ from 8% to 34%. Therefore, all COS were synthesized using glycosyl phosphate building blocks.

Nine oligomers including the chitin oligomer $\mathbf{N_6}$, four COS with different acetylation degree and patterns, and four hybrid chitin-cellulose analogues were assembled and were found to be highly water soluble (**Figure 26**).

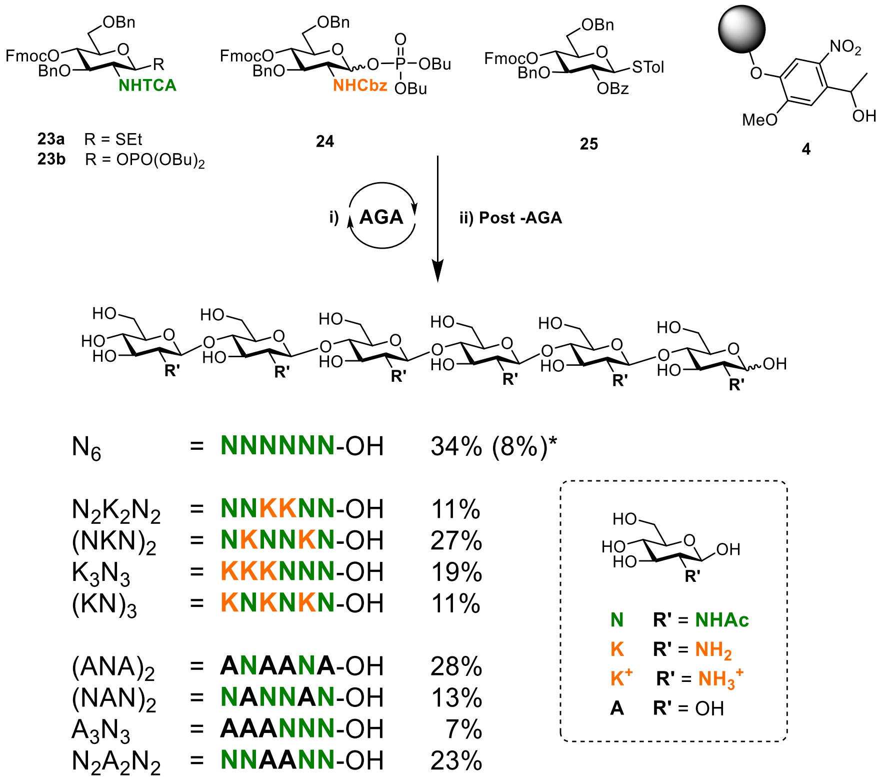


**Figure 26.** AGA of COS and cellulose-chitin hybrids. Isolated yields after AGA and post-AGA (deprotection and purifications) are reported. *Yield obtained when AGA was performed using **23a**.

### 3.2.2. Structural analysis of chitin and chitosan analogues

The synthetic oligomers were modelled using MD simulations, employing a modified version of the GLYCAM06 carbohydrate force field.[70] The partially deacetylated COS and the hybrid cellulose-chitin oligomers were compared to the reference chitin oligomer **N$_6$**. Amino substituted structures were simulated with neutral NH$_2$ (**K**) as well as with protonated NH$_3^+$ (**K$^+$**), as representative models of COS at different pH. All the modified analogues result in more flexible structures when compared to **N$_6$**. Ramachandran plots were used to compare changes on the glycosidic bond torsion angles (Ψ, Φ) (**Figure 27**). No significant differences are observed for Φ, stabilised by the exo-anomeric effect. The high population of Ψ at negative degrees (**Figure 27c**, red circle) is related to the presence of the conventional O(5)···OH(3) hydrogen bond (**Figure 27b**), which rigidifies the chitin structure (**N$_6$**). All modified oligomers show an increased population at positive Ψ (**Figure 27c**, blue circles), albeit with different intensity. Major disruption of the conventional H-bond is observed for the charged COS (e.g., **(NK$^+$N)$_2$**).

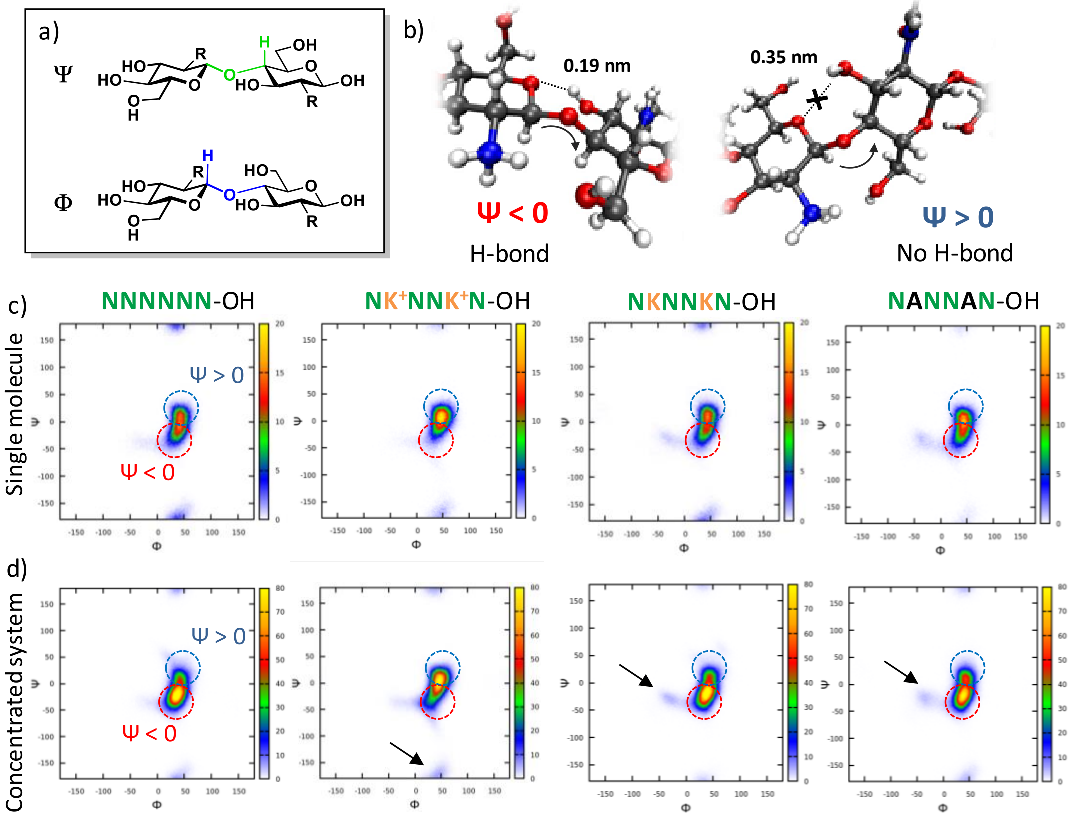


**Figure 27**. Definition of the glycosidic bond torsion angles (Ψ and Φ) (a) and representative snapshots indicating the presence (Ψ < 0) or absence (Ψ > 0) of the conventional O(5)···OH(3) H-bond (b). Analysis of the Ramachandran plots obtained by MD simulations for a single molecule (c) and for a concentrated system with 25 molecules (d). Negative degrees of Ψ (red circles) indicate the presence of the conventional O(5)···OH(3) H-bond, whereas the increased distance between these two residues is reflected by positive Ψ (blue circles). Deviations from the main conformations are highlighted with arrows.

To further assess the intermolecular interactions, long MD atomistic simulations (1 µs production run) of concentrated experiments (25 molecules) were performed, to resemble a crowded environment. Deviations from the main population are observed for the modified compounds (**Figure 27d**, arrows), suggesting a higher conformational freedom and a less regular packing than **$N_6$**. The low aggregation tendency of the ionic COS **$(NK^+N)_2$** is confirmed by the similarity of the Ramachandran plots obtained for the single molecule and the concentrated experiments. A remarkable increase in the population at negative Ψ is detected for **$N_6$** and for all the modified uncharged structures (**Figure 27d**, red circles). Stacking reduces the conformational freedom, favoring the formation of the O(5)···OH(3) hydrogen bond as well as inter-molecular H-bonds. Radial distribution functions (RDFs) were used to characterize the spatial correlations in the systems (**Figure 28**). The presence of intense RDF signals at the interval of 0.5 nm to 1.5 nm indicates the high probability of cluster formation (**Figure 28a)**. In stark contrast, the RDF for **$(NK^+N)_2$** suggests a low probability of chain-stacking (**Figure 28a**) and high tendency to interact with the solvent molecules (**Figure 28b)**. The decreased number of intermolecular hydrogen bonds between the solutes (**Figure 28c**) confirms the higher hydrophilicity of **$(NK^+N)_2$.**

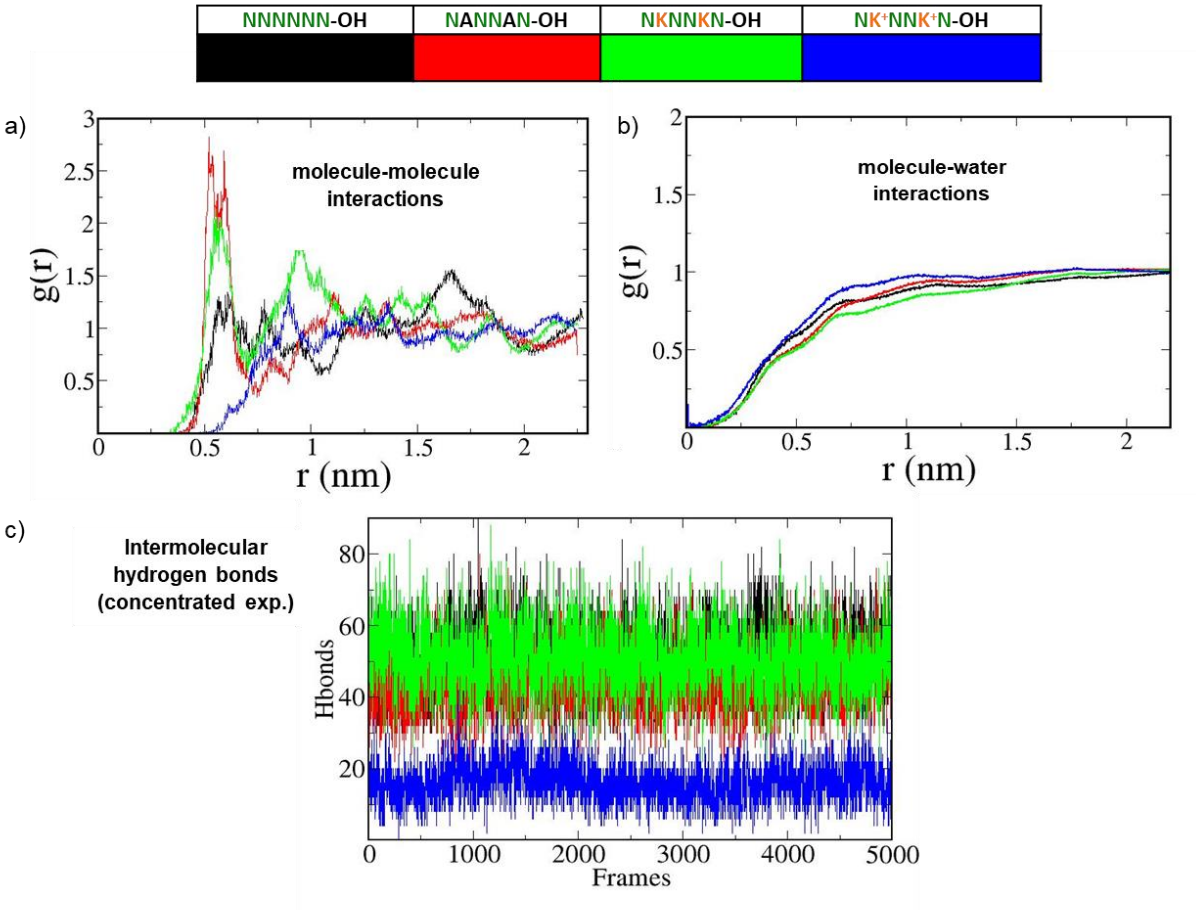


**Figure 28**. a-b) Radial Distribution Functions of **N6** (black), **(NAN)$_2$** (red), **(NKN)$_2$** (green), and **(NK$^+$N)$_2$** (blue) and c) number of intermolecular hydrogen bonds of the concentrated systems.

In agreement with the computational model, powder XRD analysis (done by Giulio Fittolani, **Table 2**) shows sharp peaks for **N$_6$**, indicating the tendency to pack with a regular architecture (**Figure 29**). While amorphous XRD profiles are measured for all the modified compounds, a peak at 21.3° was observed for the block pattern cellulose-chitin hybrid **A$_3$N$_3$** suggesting a different type of crystal packing than natural cellulose and chitin (**Figure 29**).

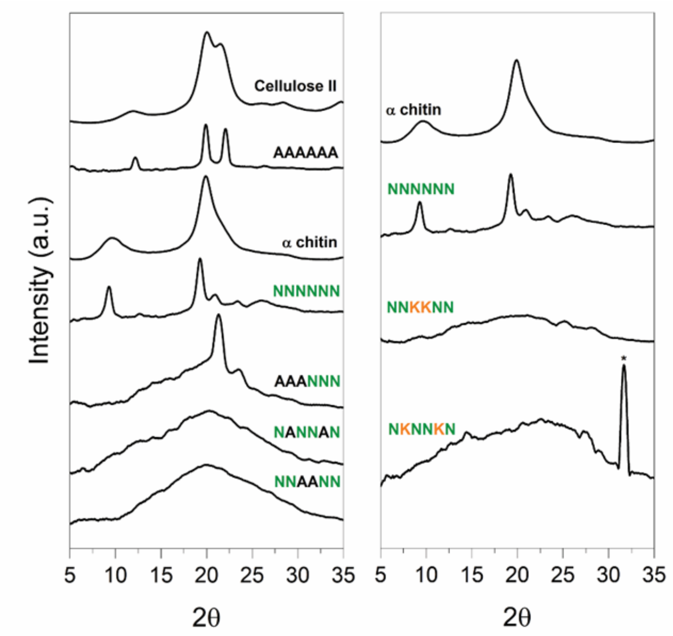


**Figure 29.** Powder XRD profiles of synthetic COS and hybrid chitin-cellulose oligomers. The XRD profiles obtained for the polysaccharide α-chitin and cellulose II are reported for comparison.*Crystalline carbonate peak resulting from the prolonged exposure of the polyamino compound to air[105].

A closer look at the atomistic model shows a significant percentage of *tg* rotamers (orientation of the C6 side chain) for the charged COS (**Figure 30a-b**). MD suggests the formation of an intramolecular hydrogen bond between the $NH_3^+$(2) and the OH(6) of the adjacent residue (R+1) (**Figure 30c**).

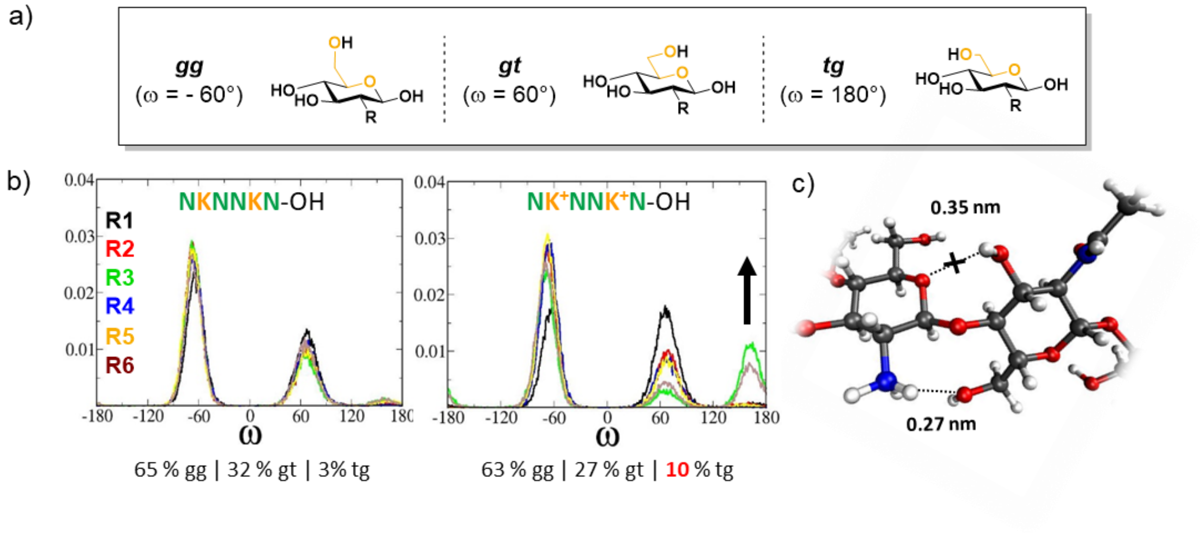


**Figure 30.** a) Definition of the ω torsion angle and b) ω distribution obtained by MD simulations for the **(NKN)₂** analogue with different protonation. c) Representative snapshots indicating the presence of an intramolecular hydrogen bond between the $NH_3^+$(2) and the OH(6) of the adjacent residue (R+1).

NMR measurements were performed to confirm this interaction. Selective 1D HOHAHA-NMR experiments on the model dimer **26** were performed to reduce the complexity of the system that suffers from chemical shift degeneracy (**Figure 31**).

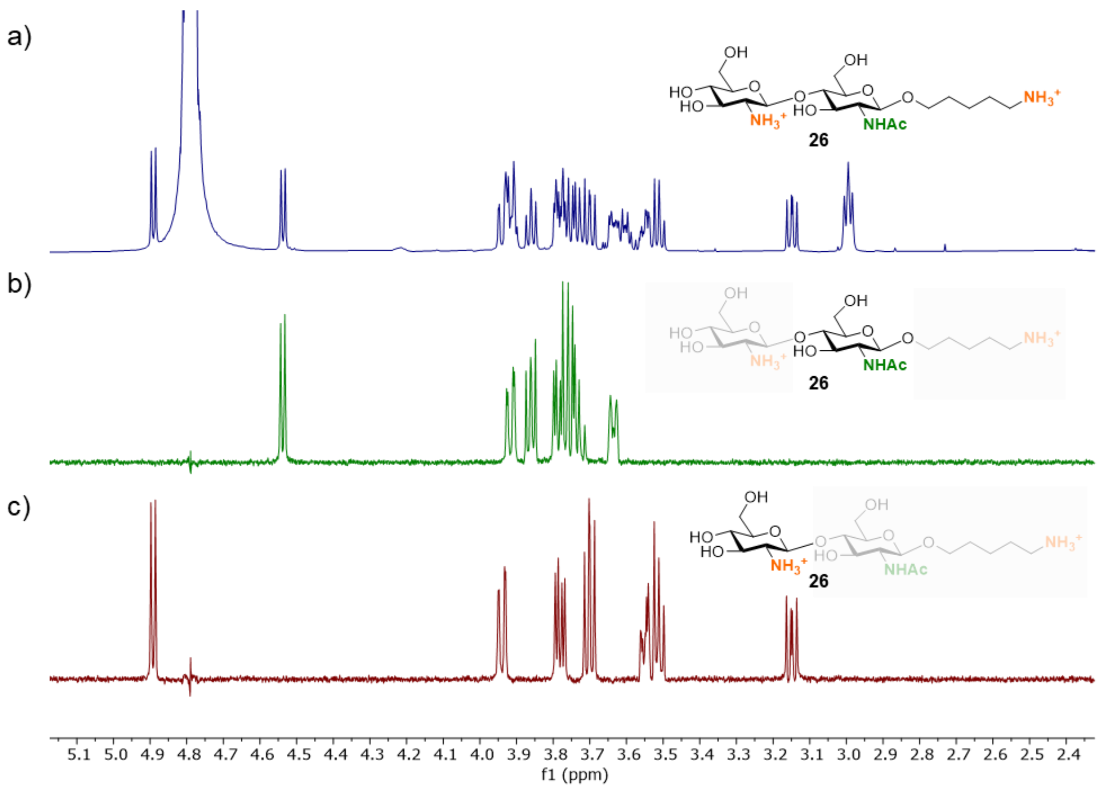
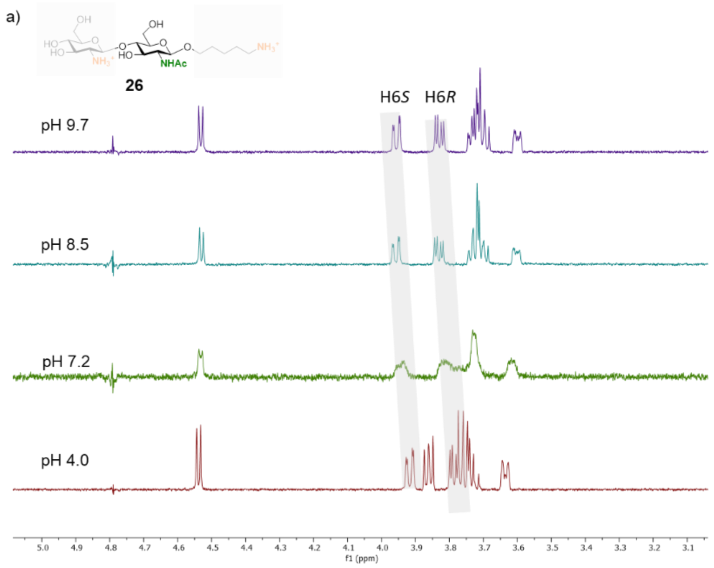


**Figure 31**. a) Representative $^1$H-NMR spectrum of KN-NH$_2$ recorded at pH 4.0. b) Selective 1D HOHAHA-NMR spectrum with selective excitation of the GlcNAc (N) anomeric proton at 4.5 ppm. c) Selective 1D HOHAHA-NMR spectrum with selective excitation of GlcN (K) anomeric proton at 4.9 ppm..

$^3J_{H5H6}$ coupling constants were measured at different pH. The experimental calculated coupling constants $^3J_{H5H6R}$ and $^3J_{H5H6S}$ can be converted to rotamers percentage using empirical equations[106] (**Figure 32b**).



b)

| $^3J_{H5H6R}$ | | | $^3J_{H5H6S}$ | | |
|---|---|---|---|---|---|
| *gg* | *gt* | *tg* | *gg* | *gt* | *tg* |
| 1.0 | 11.0 | 4.8 | 2.2 | 2.5 | 10.2 |

$$^3J_{H5H6R} = {}^3J_{R,gg}f_{gg} + {}^3J_{R,gt}f_{gt} + {}^3J_{R,tg}f_{tg} \qquad [1]$$

$$^3J_{H5H6s} = {}^3J_{S,gg}f_{gg} + {}^3J_{S,gt}f_{gt} + {}^3J_{S,tg}f_{tg} \qquad [2]$$

$$1 = f_{gg} + f_{gt} + f_{tg} \qquad [3]$$

| | **pH 8.5** | **pH 4.0** |
|---|---|---|
| $^3J_{H5H6R}$ | 5.1 Hz | 4.9 Hz |
| $^3J_{H5H6S}$ | 2.3 Hz | 2.6 Hz |
| ***gg*** | 59 % | 59 % |
| ***gt*** | 41 % | 38 % |
| ***tg*** | 0 % | **4 %** |

**Figure 32.** a) Selective 1D HOHAHA-NMR analysis of the $^3J_{H5H6}$ coupling constants measured for the model dimer **26** at different pH and b) conversion of the coupling constants to rotamers (f) percentage using empirical equations [1-3][106].

A small percentage of *tg* rotamers (4%) was detected at pH 4, when the amine is fully protonated. No *tg* rotamers were detected at basic pH (amine not protonated), in agreement with the predictions. Although NMR data indicated the existence of a small *tg* population at acidic pH, the calculated fraction of *tg* is significantly lower than the predicted value. Different simulation conditions using i) a different water model (tip3p), ii) N3 angle parameters derived in the context of GAG molecules, iii) increased ionic strength, and iv) reduced Lennard-Jones interactions of the nitrogen atoms (consistent with the changes introduced in the GLYCAM$_{OSMO,r14}$ force field) did not lead to major changes in the simulation results (**Figure 33**). This overrepresentation of the observed H-bonds trend demonstrates the need of further optimization of the dihedral potentials, especially in the presence of ionic moieties (e.g., amines).

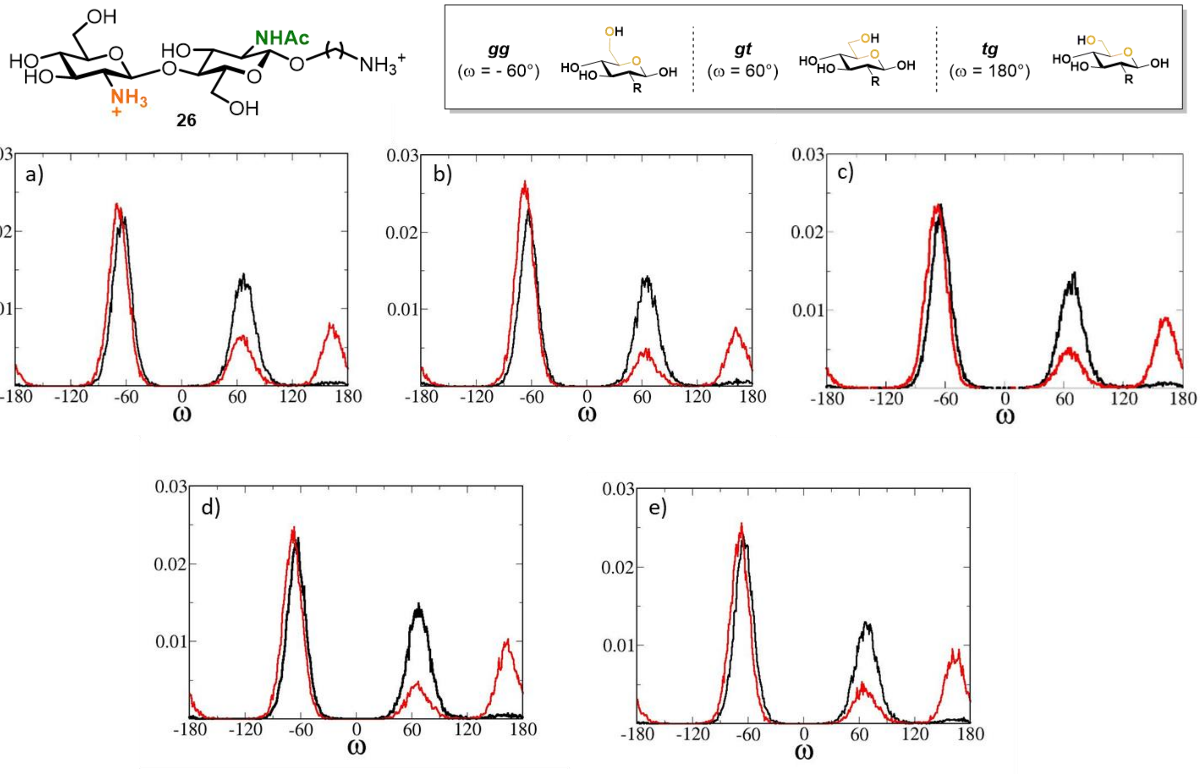


**Figure 33**. Omega torsional angles ($\omega_1$ black and $\omega_2$ red) of K$^+$N-NH$_3$ measured with: (a) tip5p water model, (b) tip3p water model, (c) N$_3$ angle parameters derived in the context of GAG, (d) increased ionic strength (50 solutes of Na$^+$ and Cl$^-$), and (e) reduced Lennard-Jones interactions.

## 3.3. Conclusion

Five COS were assembled by AGA with full-control over length as well as acetylation degree and pattern. Four unnatural hybrid chitin-cellulose oligomers were prepared to study the importance of the amino group in chitosan. I successfully designed the MD simulations to evaluate the conformational space of well-defined chitin and chitosan analogues. Single molecule as well as concentrated MD simulations showed that all analogues have more conformational freedom than the fully *N*-acetylated hexamer **N$_6$**, resulting in amorphous aggregation upon drying. The hybrid compounds showed a similar conformational behavior as the neutral partially acetylated COS. Amine protonation is crucial for the formation of novel intramolecular interactions, detected by NMR, that stabilise new geometries. This observation is particularly relevant to describe molecular mechanisms of chitosan-protein interactions, as glycoside hydrolases binding is affected by the orientation of the C6 side chain[107].

# 4. On resin synthesis of sulfated oligosaccharides

## 4.1. Introduction

The study of chitosan oligomers, discussed in Chapter 3, revealed that positive charges displayed on the chitin backbone are capable of triggering new interactions leading to structural changes. The addition of ionic moieties to neutral polysaccharide backbones is a common modification in nature. The insertion of charges increases the solubility and induces structural changes. Sulfation of glycans is a significant post-glycosylation modification[108] that introduces negative charges to neutral oligosaccharides. The polyanionic nature of sulfated glycans results in strong affinity to positively charged surfaces on lectins, facilitating a multitude of interactions that mediate essential biological processes[109]. The specific function of these sulfated compounds strongly depends on the degree of sulfation as well as on the sulfation pattern along the glycan chain[43,44]. Access to well-defined sulfated glycans is valuable to clarify how the spatial orientation of the sulfate groups, responsible for conformational changes, regulates the function (sulfation code).

Protocols for the selective introduction of sulfate groups on complex oligosaccharide structures have been developed[110]. Nevertheless, the synthesis of sulfated glycans remains exceedingly challenging. Selective sulfation mandates the rational placement of orthogonal protecting groups (PGs), thus adding extra complexity to the already challenging synthesis of the oligosaccharide backbone. Upon sulfation, the removal of the residual PGs can be troublesome and further synthetic manipulations must avoid migration and/or cleavage of the labile sulfate moieties[48].

The identification of biologically active sulfation patterns requires versatile, adaptable, and rapid synthetic methodologies. Over the past decades, a variety of solution-phase methodologies has been developed, permitting the assembly of highly complex structures[111–113]. Still, the production of these compounds is far from trivial, with each target often requiring further optimization[114]. Tedious chromatographic purifications after each step (normal phase, reverse phase and/or size exclusion) are unavoidable, since the accumulation of unreacted reagents, side-products, and the presence of salts can affect the next reaction steps (**Figure 34a**). Moreover, sulfation of multiple positions becomes progressively more difficult, due to steric hindrance and anion crowding commonly leading to incomplete reactions[115]. The increase in product polarity impedes purifications due to the amphiphilic nature of the partially protected intermediates.

Solid-phase approaches (**Figure 34b**) offer an alternative to traditional solution-phase synthesis[66], as exemplified by the preparation of a small collection of keratan sulfate (KS) tetrasaccharides[116]. In this example, the backbone, prepared by Automated Glycan Assembly (AGA), was sulfated on resin, before solution-phase removal of all the PGs yielded the target compounds. Still, to date, solid-phase approaches were powerful to access neutral backbones[117,118], but have shown a limited substrate scope, poor reproducibility and, in most cases, proved unsuccessful in obtaining the final deprotected sulfated targets[64,119]. Solid-phase approaches are currently less scalable than solution phase syntheses, making purification and deprotection of the “off resin” compound very challenging, often resulting in loss of compounds.

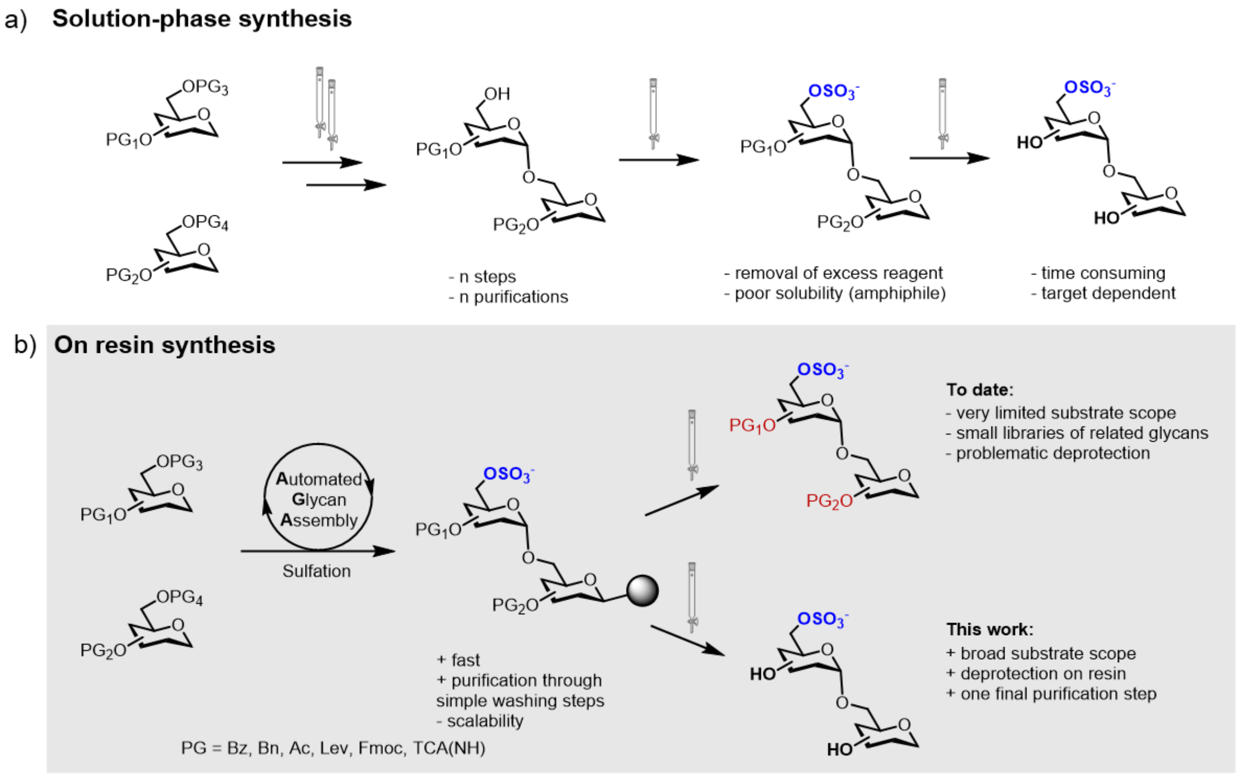


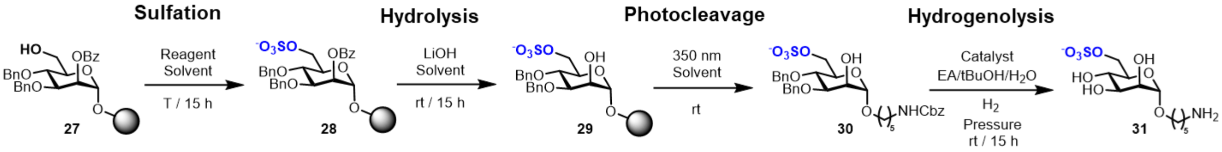
**Figure 34**. Synthetic approaches to access sulfaed glycans: classical solution-phase synthesis (a), on resin synthesis (this work, b).

In this chapter, I report a general strategy for the solid-phase synthesis of various sulfated oligosaccharides (**Figure 34b**). The glycan backbone is constructed by AGA. Optimized sulfation and ester hydrolysis procedures yield the partially protected intermediate on solid support. Only simple washing steps are required for the removal of excess reagents. Cleavage from the solid support followed by hydrogenolysis of the remaining PGs allows for access to the desired sulfated glycan. Throughout the protocol, only a final purification step is required. Synthetic bottlenecks associated with particular structural features are identified and overcome. The general approach is showcased in the synthesis of seven diverse, biologically relevant targets.

## 4.2. Results

### 4.2.1. Synthesis of sulfated mannose 31

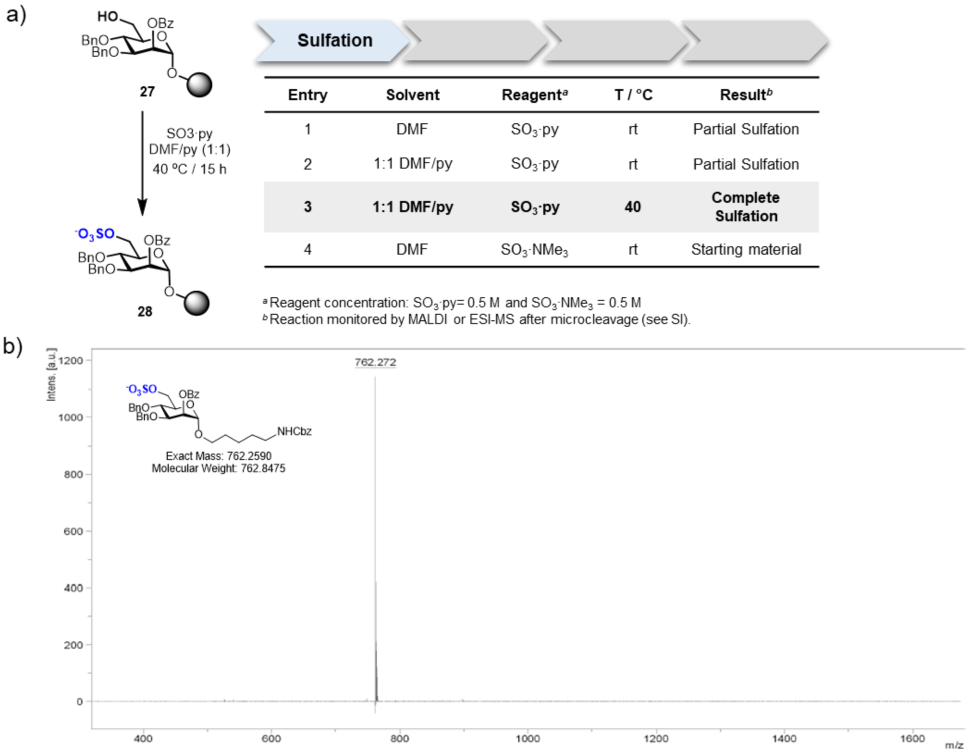
Solid-supported mannoside **27** bearing a free hydroxyl group at C6 was prepared by AGA and selected as model substrate for the development of the sulfation protocol (**Scheme 1**). To obtain the sulfated target **31**, four key steps (i.e., sulfation, hydrolysis, photocleavage, hydrogenolysis) had to be optimized to reach full conversion and avoid the accumulation of side products.



**Scheme 1**. Synthetic steps towards the synthesis of **31** for the development of the sulfation protocol.

#### 4.2.1.1. On resin sulfation

In search for an efficient sulfating procedure compatible with the solid support, I screened different conditions based on sulfur trioxide amine complexes. Incomplete sulfation was observed using sulfur trioxide pyridine complex ($SO_3$·py) at ambient temperature (**Figure 35a, Entries 1-2**). Compound **27** was readily sulfated when the reaction was conducted at 40 °C for 15 h in a DMF/Py mixture (**Figure 35a, Entries 3**). No sulfation occurred when $SO_3$·$NMe_3$ was employed (**Figure 35a, Entries 4**).



| Entry | Solvent | Reagent[a] | T / °C | Result[b] |
|---|---|---|---|---|
| 1 | DMF | $SO_3$·py | rt | Partial Sulfation |
| 2 | 1:1 DMF/py | $SO_3$·py | rt | Partial Sulfation |
| **3** | **1:1 DMF/py** | **$SO_3$·py** | **40** | **Complete Sulfation** |
| 4 | DMF | $SO_3$·$NMe_3$ | rt | Starting material |

[a] Reagent concentration: $SO_3$·py= 0.5 M and $SO_3$·$NMe_3$ = 0.5 M

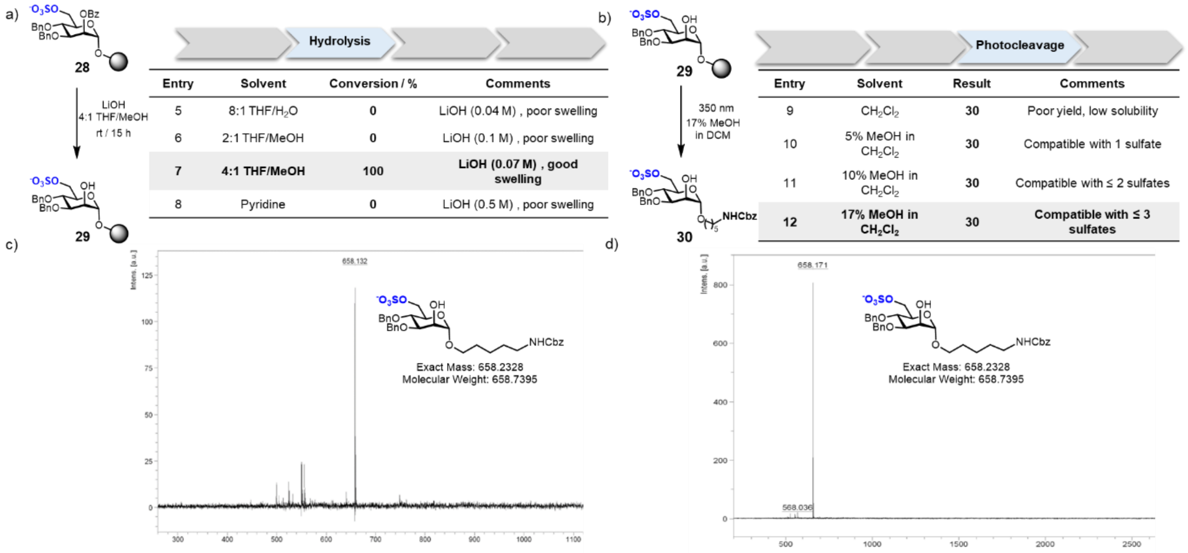
[b] Reaction monitored by MALDI or ESI-MS after microcleavage (see SI).

**Figure 35**. a) Screening of conditions for sulfation of **27** on resin and b) MALDI-TOF of compound **28** after microcleavage (negative mode).

### 4.2.1.2. On resin hydrolysis and photocleavage

Lithium hydroxide (LiOH) is frequently utilized for ester hydrolysis of sulfated oligosaccharides[120]. To translate this procedure to the solid-phase paradigm, I set to identify a solvent system that dissolves LiOH and swells the polystyrene resin.

The compromise proved troublesome since LiOH is poorly soluble in apolar solvents, often employed to guarantee proper resin swelling. The low LiOH concentration accessible in an 8:1 THF/water mixture (0.04 M) together with poor swelling of the resin resulted in no conversion (**Figure 36a, Entry 5**). Replacing water with MeOH permitted to increase the LiOH concentration (0.07 M), while maintaining suitable resin swelling. Full conversion was obtained when hydrolysis was performed in a 4:1 THF/MeOH mixture (**Figure 36a, Entry 7**). In an effort to increase the concentration of LiOH further, pyridine was added, but despite higher concentrations (0.5 M), no product was observed (**Figure 36a, Entry 8**). The hydrophilic nature of compound **29** required the addition of MeOH to the photocleavage cocktail, generally performed[63] in pure $CH_2Cl_2$ (**Figure 36b, Entry 9-12**).



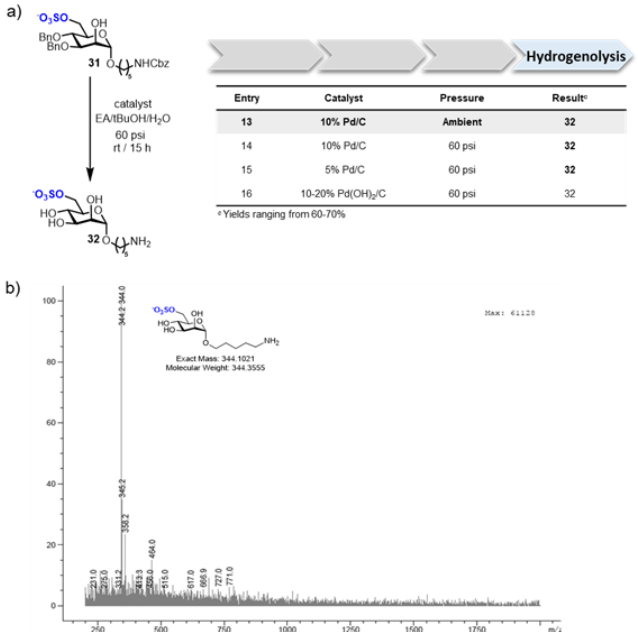
| Entry | Solvent | Conversion / % | Comments |
|---|---|---|---|
| 5 | 8:1 THF/$H_2O$ | 0 | LiOH (0.04 M) , poor swelling |
| 6 | 2:1 THF/MeOH | 0 | LiOH (0.1 M) , poor swelling |
| **7** | **4:1 THF/MeOH** | **100** | **LiOH (0.07 M) , good swelling** |
| 8 | Pyridine | 0 | LiOH (0.5 M) , poor swelling |

| Entry | Solvent | Result | Comments |
|---|---|---|---|
| 9 | $CH_2Cl_2$ | 30 | Poor yield, low solubility |
| 10 | 5% MeOH in $CH_2Cl_2$ | 30 | Compatible with 1 sulfate |
| 11 | 10% MeOH in $CH_2Cl_2$ | 30 | Compatible with ≤ 2 sulfates |
| **12** | **17% MeOH in $CH_2Cl_2$** | **30** | **Compatible with ≤ 3 sulfates** |

**Figure 36.** a) Screening of conditions for on resin hydrolysis of the Bz esters of **28** and b) screening conditions for photocleavage of **29**. c,d) MALDI-TOF of compound **29** and **30** after micro and full photocleavage respectively (negative mode).

### 4.2.1.3. Hydrogenolysis

The removal of the benzyl (Bn) groups was performed by catalytic heterogeneous hydrogenolysis. Different catalysts were tested, showing no significant differences in the reaction outcome (**Figure 37a, Entry 13-16**). The catalyst was removed by filtration and the product purified by a reverse phase C18 and LH-20 size exclusion chromatography. The target compound was obtained as sodium salt upon treatment with $Na^+$ exchange resin for a 72% overall yield (over five steps including AGA, sulfation and deprotections).



| Entry | Catalyst | Pressure | Result[c] |
|---|---|---|---|
| **13** | **10% Pd/C** | **Ambient** | **32** |
| 14 | 10% Pd/C | 60 psi | **32** |
| 15 | 5% Pd/C | 60 psi | **32** |
| 16 | 10-20% $Pd(OH)_2$/C | 60 psi | 32 |

[c] Yields ranging from 60-70%

**Figure 37**. a) Screening of conditions for the hydrogenolysis of **31** and b) MALDI-TOF of compound **32** (negative mode).

### 4.2.2. Expanding the substrate scope

To assess the generality of the protocol, six monosaccharides and five oligosaccharides based on different monosaccharide units, PGs, and glycosidic linkages were prepared. All glycan backbones were assembled following standard AGA protocols (**Figure 38**).

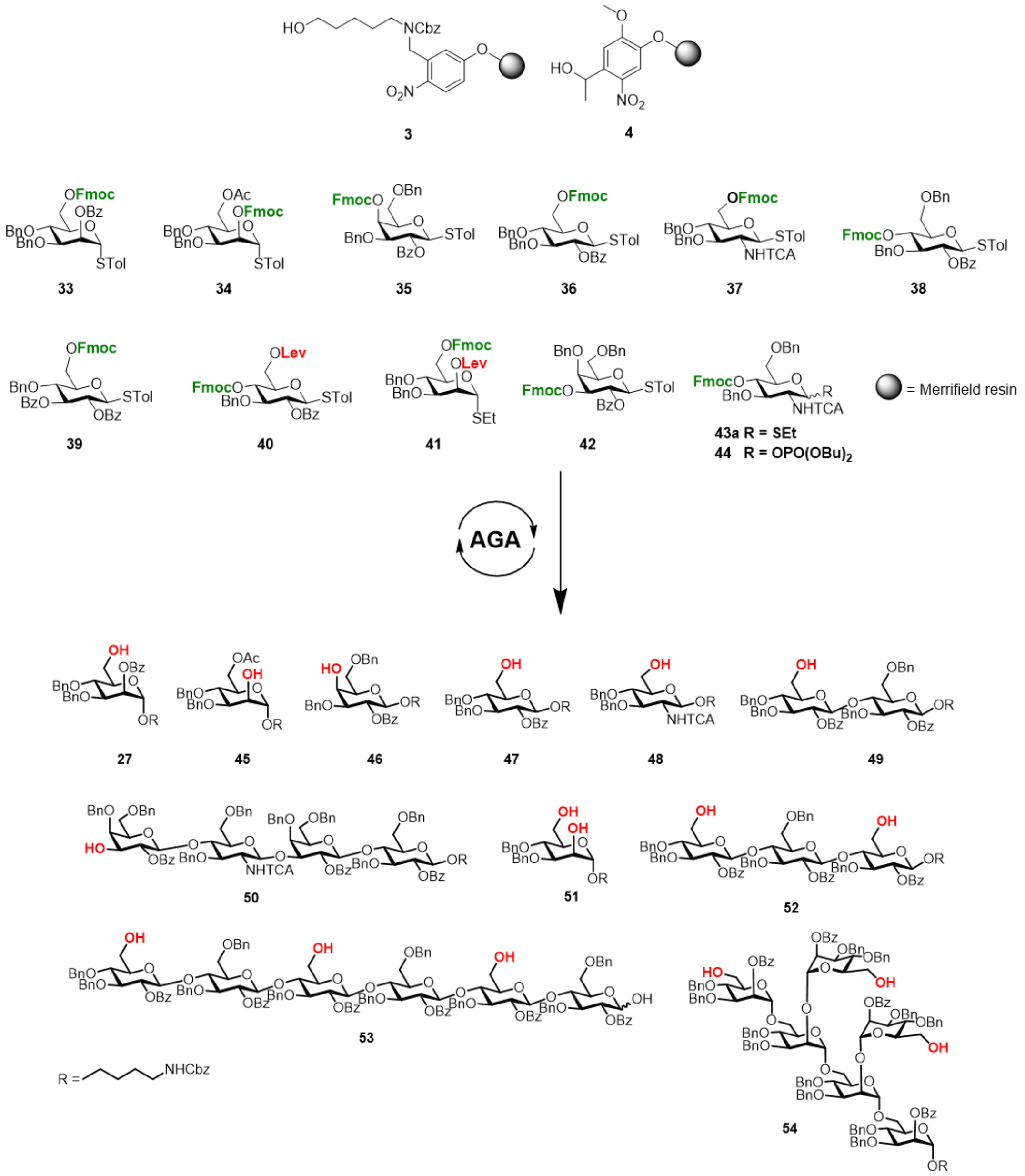


**Figure 38.** Glycan backbones assembled by AGA. The liberated hydroxyl groups (highlighted in red) were subjected to sulfation.

The optimized sulfation procedure worked smoothly for all primary alcohols, regardless of the glycan length. Full conversion was also observed for secondary hydroxyl groups, including the poorly reactive hydroxyl groups at the C-2 position of mannose **56** and the C-4 position of galactose **57** (**Figure 39a**). Multiple sulfate groups were introduced on a glycan (**63-65**) as well as on the same monosaccharide (**62**); proving ion crowding was not a challenge. Complex structures, including the trisulfated branched hexamannoside (**64**) and cellohexasaccharide (**65**), were successfully sulfated without any adjustments (**Figure 39b**).

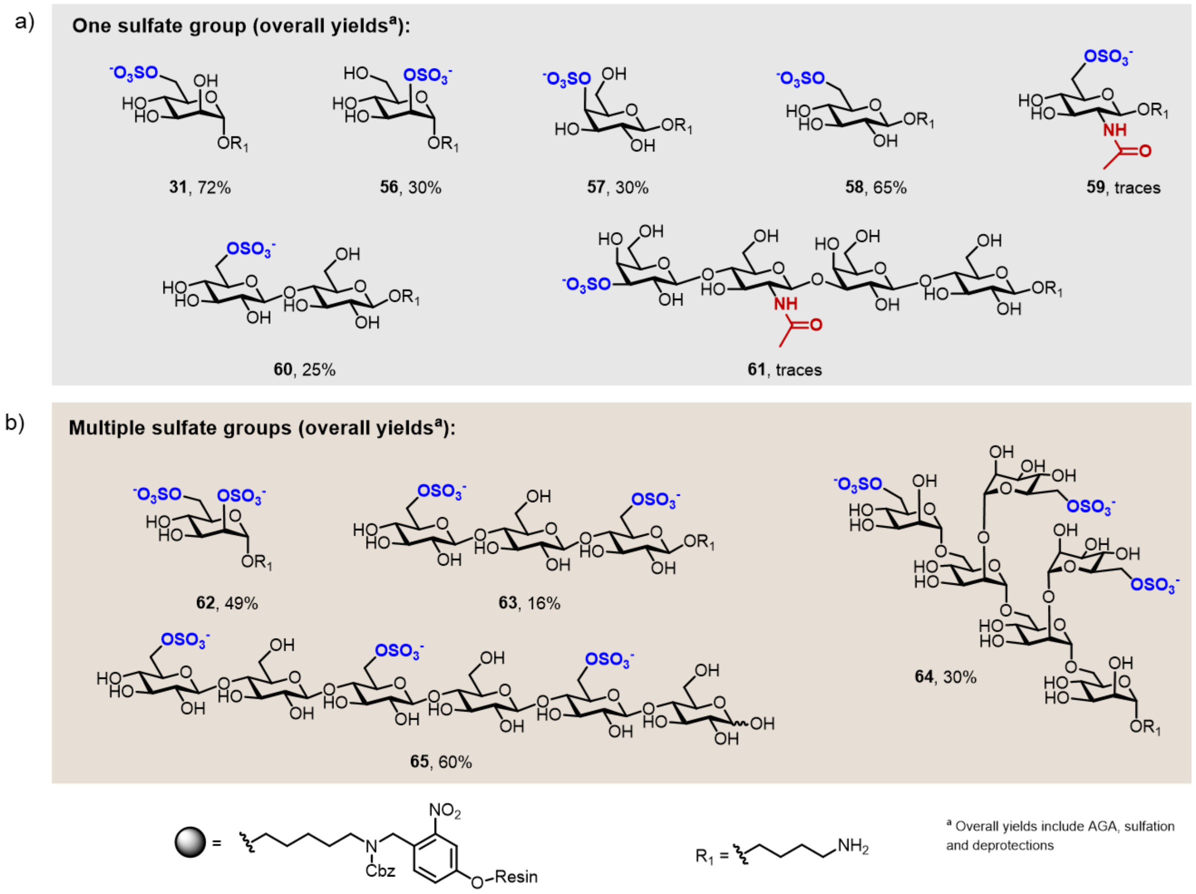


**Figure 39**. a) Collection of mono- and b) multi-sulfated compounds obtained with the optimized conditions reported in section 4.2.1.2.

Hydrolysis of the ester PGs was completed within 15 h for the majority of the sulfated targets. Longer reaction times were required for the sulfated cellulose oligomers **63** (24 h) and **65** (48 h) and lacto-*N*-tetraose (LNT) **61** (120 h), as previously reported for non-sulfated analogues[121]. The mild conditions were compatible with long debenzoylation reactions as no side-products were observed. Reaction progress was easily monitored by ESI-mass spectrometry analysis on a minute reaction sample after microcleavage. Removal of Lev or Fmoc PGs can be achieved during the hydrolysis step or selectively using hydrazine or triethylamine, respectively. Adjustments to the photocleavage solvent system were required for structures containing multiple sulfate groups to guarantee good solubility of the cleaved compound while maintaining ample resin swelling (**Figure 36b, Entry 10-12**). The final hydrogenolysis of the benzyl ethers proceeded smoothly under standard conditions (**Figure 37a**). With this protocol, linear and branched oligosaccharides were obtained in good to excellent overall yields. The entire protocol, including AGA, sulfation, and deprotections required less than 72 h, even for complex structures like **64** and **65**. Simple access to collections of complex sulfated mannans (**64**), highly abundant in seaweeds, will fuel the characterization of marine glycans and their processing enzymes[122]. Moreover, pattern-controlled sulfated cellulose analogues like **65** are interesting targets to mimic the bioactivity of heparin[123,124], in which the anticoagulant activity is influenced by the position of sulfates. As well, sulfated cellulose structures are important compounds in materials science[125].

#### 4.2.2.1. Overcoming hydrogenolysis issues

A major limitation of the approach was revealed when attempting to synthesize compounds containing an acetylamino-2-deoxy-glucoside (GlcNAc) in their backbone (**59** and **61**). In these instances, the desired products were only obtained in trace amounts. The problem seemed associated with the final hydrogenolysis step, as mass spectrometry indicated full conversions for all prior transformations and a significant amount of crude product was obtained after photocleavage. Only trace amounts of product were recovered upon hydrogenolysis of **59a** using 10% Pd/C (**Figure 40, Entry 1-2**) even though MS-QTOF indicated complete removal of the Bn PGs (**Figure 40a**).

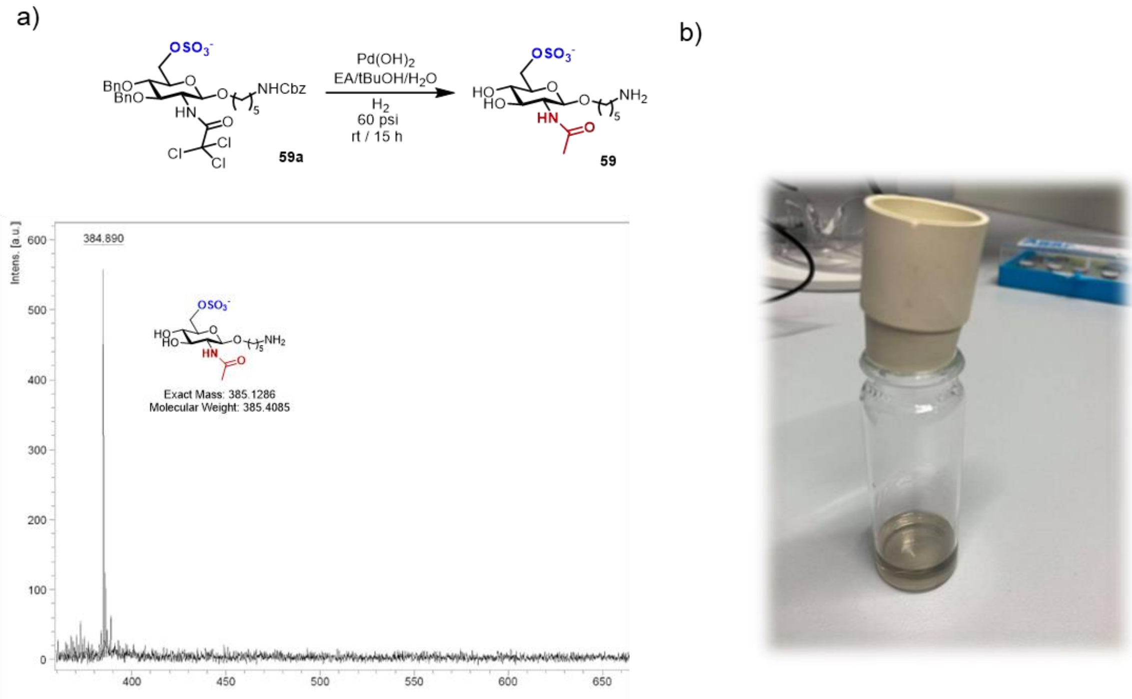
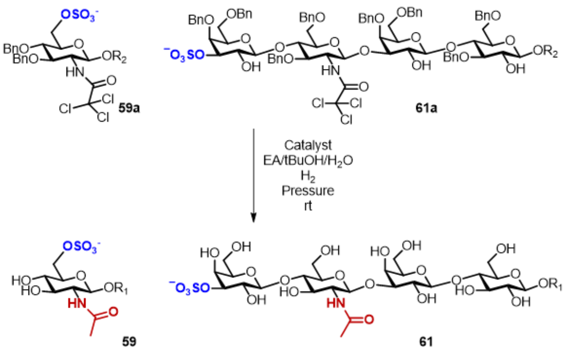


**Figure 40.** a) MALDI-TOF analysis of the crude reaction indicating completion of the hydrogenolysis step (negative mode). b) Upon completion of hydrogenolysis, filtration resulted in a black solution resulting in the loss of the compound **59**.

Since sulfated *N*-acetyl glucosamine residues are an important part of biologically active glycans,[126] I set out to overcome this limitation by screening different catalysts and conditions using compound **59a** as model system (**Figure 41**). The use of a more disperse Pd/C catalyst[127] can improve hydrogenolysis yields; however, in our case, the use of 5% Pd/C was not beneficial (**Figure 41, Entry 3**). A remarkable increase in yield was observed using PdPt/C or $Pd(OH)_2/C$ as catalysts in a pressurized reactor (**Figure 41, Entry 4-5**). However, these catalysts failed when the reaction was performed with **61a** as starting material (**Figure 41, Entry 7**). Loss of sulfated compounds during hydrogenolysis is a known issue; similar observations have been reported during the synthesis of hyaluronans[128] and in the synthesis of a 24-mer of chondroitin sulfate, where the yield drastically dropped during hydrogenolysis[111]. The significant drop in yields are often tolerated because solution phase synthesis are commonly performed on a larger scale, but becomes much more dramatic when dealing with the smaller quantities compatible with solid-phase approaches, often resulting in complete loss of compound[119].



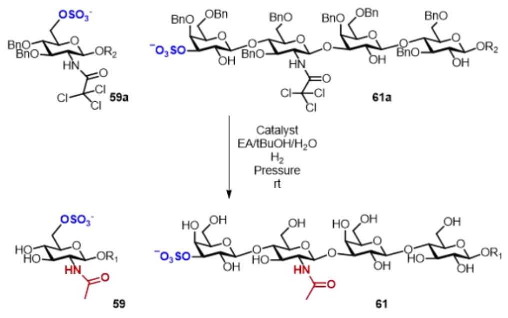
| Entry | SM | Catalyst | Pressure | Overall Yield[b] |
|---|---|---|---|---|
| 1 | 59a | 10% Pd/C | Ambient | 2% |
| 2 | 59a | 10% Pd/C | 60 psi | 3% |
| 3 | 59a | 5% Pd/C | 60 psi | 5% |
| 4 | 59a | 4% 1% PdPt/C | 60 psi | 19% |
| 5 | 59a | 10-20% Pd(OH)$_2$/C | 60 psi | 24% |
| 6 | 61a | 10% Pd/C | Ambient | 2% |
| 7 | 61a | 10-20% Pd(OH)$_2$/C | 60 psi | 5% |

[b] Overall yields including AGA, sulfation, and deprotections.

**Figure 41**. Screening of conditions for the hydrogenolysis of **59a** and **61a**.

I hypothesized that the low yields were the result of the strong coordination of the target compounds to the palladium catalyst. Thus, I envisioned the treatment of the reaction mixture, upon completion of hydrogenolysis, with a palladium scavenger to promote ligand exchange and release the desired product. Much higher yields were obtained when the crude mixture was treated with thiourea, a molecule with high affinity to palladium (**Figure 42, Entry 6 and 9**). The overall yield of sulfated **61** significantly improved from 5 to 38% (**Figure 42, Entry 9**).

a)



| Entry | SM | Catalyst | Additive[a] | Pressure | Overall Yield[b] |
|---|---|---|---|---|---|
| 5 | 59a | 10-20% Pd(OH)$_2$/C | - | 60 psi | 24% |
| **6** | **59a** | **10-20% Pd(OH)$_2$/C** | **Thiourea** | **60 psi** | **34%** |
| 7 | 61a | 10% Pd/C | - | Ambient | 2% |
| 8 | 61a | 10-20% Pd(OH)$_2$/C | - | 60 psi | 5% |
| **9** | **61a** | **10-20% Pd(OH)$_2$/C** | **Thiourea** | **60 psi** | **38%** |

[a] Upon completion of hydrogenolysis, prior to filtration, the crude mixture was treated with thiourea (10 equiv.) [b] Overall yields including AGA, sulfation, and deprotections.

b)

**Figure 42**. a) Screening of conditions for the hydrogenolysis of **59a** and **61a**. Increased yields were observed when the crude mixture was treated with thiourea, b) Picture of the sample solutions obtained upon filtration after hydogenolysis without (top) and with (bottom) thiourea treatment.

Next, I explored the scope of the new protocol for the synthesis of biologically relevant compounds bearing the acetylamino moiety (**Figure 44**). The glycan backbones were assembled with AGA (**Figure 43**).

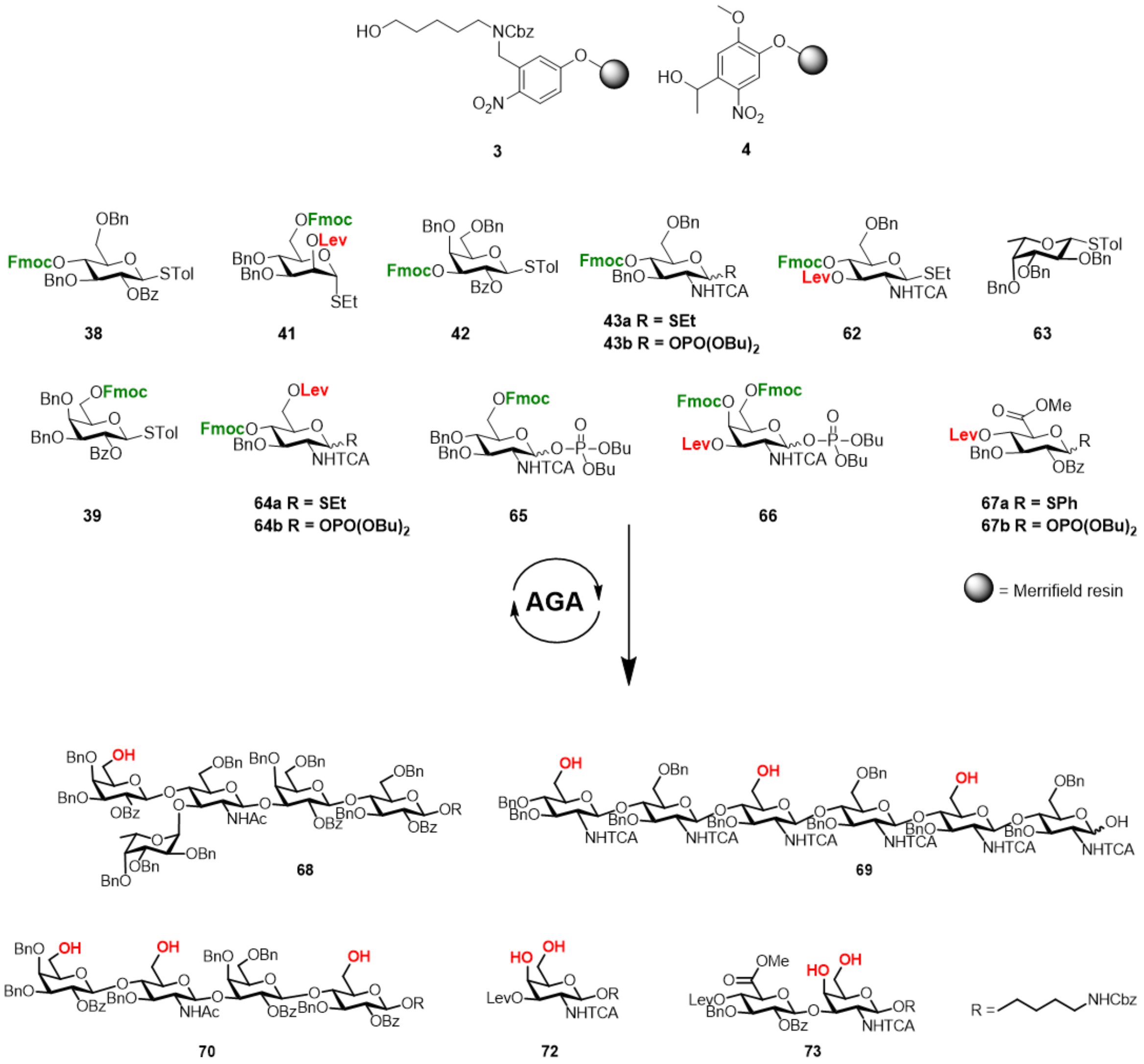
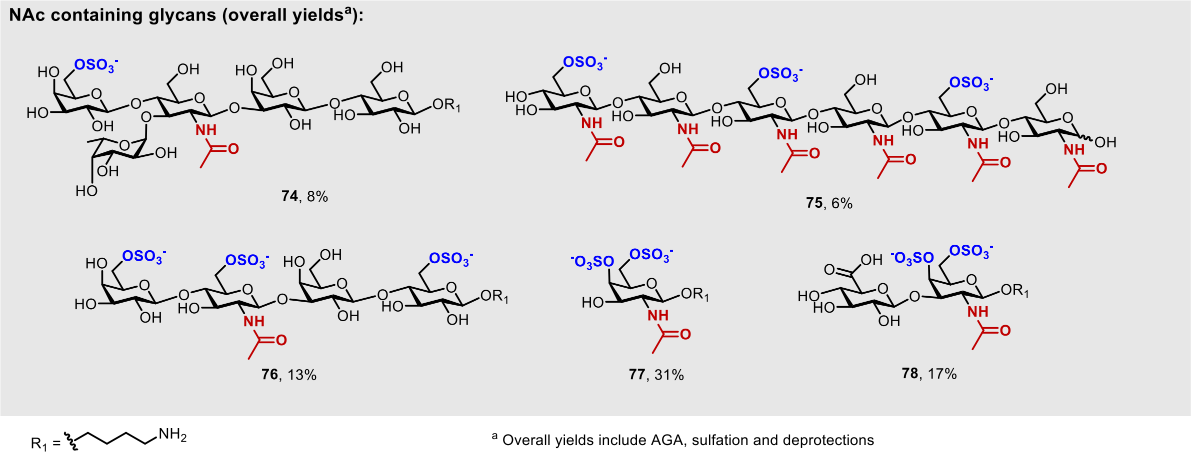


**Figure 43**. Glycan backbones assembled by AGA. The liberated hydroxyl groups (highlighted in red) were subjected to sulfation.

Two sulfated LNT structures (**61** and **76**) as well as the 6-*O*-$SO_3^-$ Lewis$^x$ antigen **74** were prepared to study their interaction with the human galectin-4[129] and L-selectins[130], respectively. The synthesis of the trisulfated chitin oligomer **75** yielded a heparin-like anticoagulant structure with a defined degree and pattern of substitution, crucial to avoid toxicity[131]. During the synthesis of tetrasaccharide **76**, treatment with lithium hydroxide unexpectedly caused the cleavage of the TCA PGs. The modular methodology allowed for fast, selective acetylation of the free amino group directly on resin, prior to photocleavage, using 15% acetic anhydride in DMF. The loss of the TCA group was also observed for compounds **77** and **78**, where the subsequent re-acetylation was not fully selective such that some hydroxyl groups were acetylated. Advantageously, the unwanted *O*-acetyls were chemoselectively removed by repeating the on resin LiOH hydrolysis step, facilitating the purification and ultimately yielding the final compounds in good overall yields. The successful synthesis of compounds **77** and **78** indicate that the protocol offers a promising approach for the assembly of well-defined glycosaminoglycan targets.



[a] Overall yields include AGA, sulfation and deprotections

**Figure 44**. Biologically relevant GlcNAc and GalNAc containing glycans obtained with the optimized protocol highlighted in gray panel.

#### 4.2.2.2. Towards the synthesis of chondroitin sulfate (CS) tetramer

After the successful synthesis of CS precursors such as **77** and **78**, I aimed to showcase my method for the synthesis a complex CS tetrasaccharide, a stimulator of the neuronal growth[132]. A previously reported convergent strategy gave access to the CS tetramer[132]. Still, the synthesis is far from trivial with many low yielding and challenging purifications steps. The assembly of the CS tetramer was attempted using AGA and BB **66** and **67b** containing Lev and Fmoc temporary protecting groups (**Figure 45**). The selective deprotection of Lev allowed for chain elongation whereas the Fmoc cleavage at the end of the synthesis unmasks the hydroxyl groups for further post-AGA transformations. To overcome the issue of the general low reactivity of the glucuronic acid (GlcA) monosaccharide, the thioglycosides were converted into phosphate BBs and a double cycle of glycosylation was performed for each glycosylation step. Glycosyl phosphate donors performed significantly better and optimization of the glycosylation temperature minimized the presence of deletion sequences. The standard conditions (20% piperidine in DMF) used for Fmoc removal resulted in *β*-elimination side-product **81** commonly observed in CS synthesis[132,133]. Milder basic conditions (5% TEA in DMF) afforded the target oligosaccharide backbone. The optimization of the sulfation protocol (to introduce 4 sulfate groups) and the analysis of the products is currently ongoing.

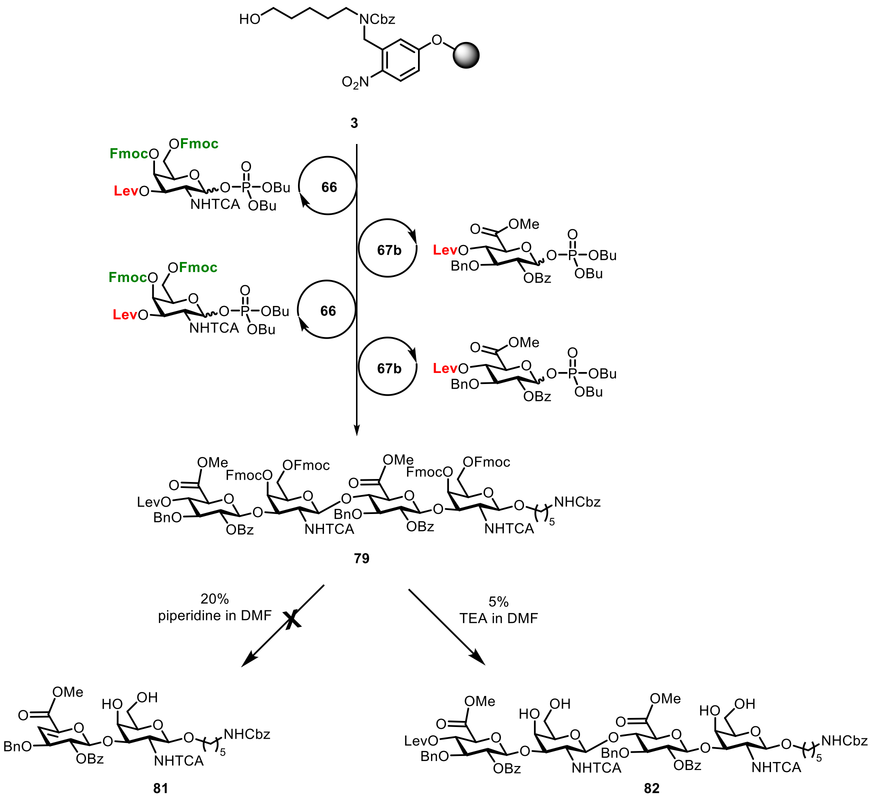


**Figure 45**. AGA of CS tetramer using BBs **66** and **67b**. Deprotection of the Fmoc protecting groups using TEA resulted in the desired tetramer **82.**

## 4.3. Conclusion

I have addressed several issues related to the synthesis of sulfated glycans on solid support. I identified $SO_3$py as a generally effective reagent for on resin sulfation. The resin swelling related issues in the solid phase ester hydrolysis were overcame by careful selection of reagents and solvent mixtures. Solubility issues from the partially protected sulfated glycans during the photocleavage step were encountered. Moreover, I identified metal coordination as the most likely detriment to the high yielding debenzylation of acetylamino-containing sulfated glycans. The resolution of these problems resulted in a general on resin approach for the synthesis of sulfated glycans with a broad reaction scope. A diverse collection of sulfated glycans, including multi-sulfated linear and branched hexasaccharides was prepared. The precise position of sulfate groups along the carbohydrate backbone permits sulfated glycans to encode information in a sequence-specific manner. The robust approach presented in this Chapter, will fuel the production of collections of sulfated glycans with defined patterns, necessary to unlock the “sulfation code” and establish structure–function correlations.

# 5. Dissecting bacterial biofilms with synthetic phosphoethanolamine-modified oligosaccharides

## 5.1. Introduction

In previous chapters the syntheses and the structural evaluation of oligosaccharides bearing positive or negative charges were described. The study of the modified cellulose analogues revealed how single site modifications can perturb the structure and alter the material properties. Recently, it was discovered that some bacteria (e.g., *E. coli* and *Salmonella enterica*) produce chemically modified cellulose bearing phosphoethanolamine (pEtN) substituents[20] to form highly ordered extracellular architectures, upon interactions with other biomolecules. Bacterial biofilms[134] (**Figure 46a**) are a typical example of natural assemblies; bacteria secrete various biomolecules to create extensive networks of extracellular matrix (ECM). These biofilms, often associated with pathogenic infections[135], have gained popularities for their remarkable mechanical properties, transforming bacteria into elegant biofactories of smart materials[136–139]. The major components of the ECM of *Escherichia coli (E. coli)* biofilms are curli fibrils – bacterial functional amyloids[140,141] - and cellulose[142]. Genetic engineering approaches[143] permitted the programing of bacterial amyloid production[144,145] to generate tunable bioplastics[146]. Similar strategies to tune the production of bacterial polysaccharides[138,147,148], the other major components of bacteria biofilms, are limited by complex biosynthetic pathways.

The composite of curli and pEtN cellulose generates biofilm (**Figure 46b**) with enhanced elasticity and adhesion to bladder epithelial cells[149]. This exciting discovery suggests that the carbohydrate component tunes the biofilm properties and may be the basis for tailoring cellulosic materials for applications in tissue engineering, biotechnology, and the food industry[22,150]. With the genes responsible for the pEtN modification identified,[20] genetically engineered bacteria could be imagined for the production of specifically modified cellulose[147,151–153].

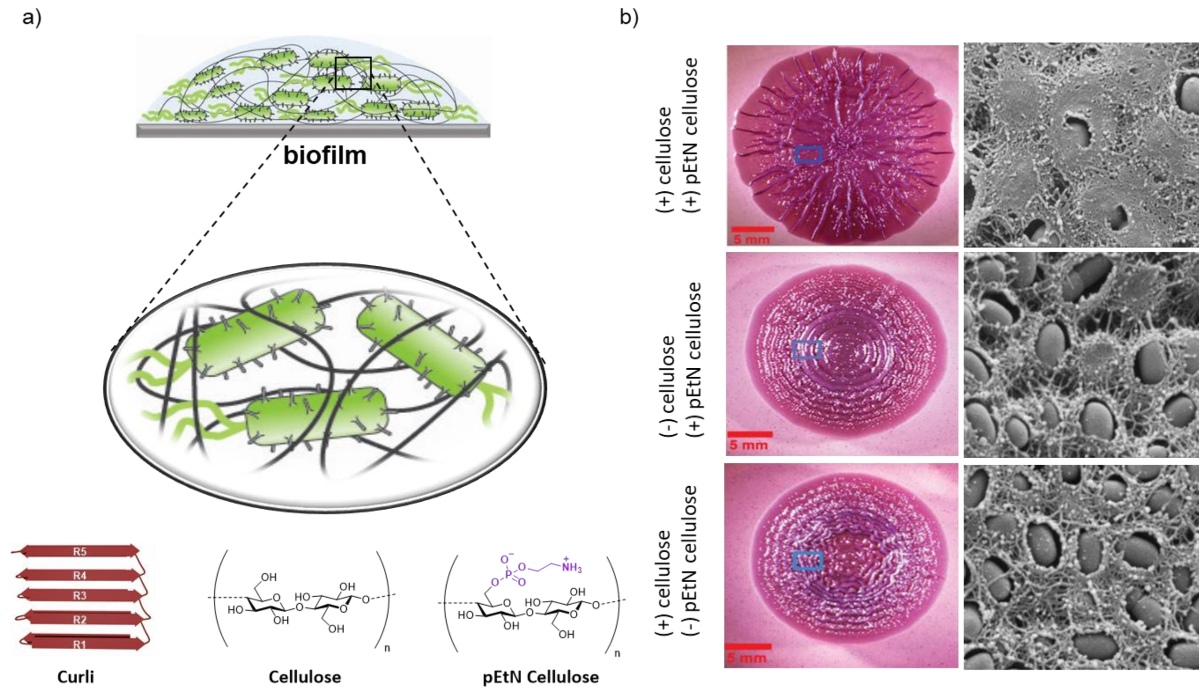


**Figure 46**. a) Schematic representation of the bacterial biofilms and its major components, amyloid curli and cellulose. b) The phenotype of the surface of the microcolonies (left) and high resolution imaging of biofilms by scanning electron microscopy (right). Adapted with the permission from Thongsomboon et al. Copyright (2018) AAAS.

Several fundamental aspects remain to be elucidated before pEtN cellulose can be exploit to its full potential. Approximately half of the glucose units of cellulose are substituted at the C6 hydroxyl group with a pEtN moiety and the ratio of curli-to-pEtN cellulose varies among different *E. coli* strains[21]. The pattern of modification, the length of the pEtN cellulose, and the mode of interaction with curli remain unknown[22]. Pure, well-defined oligosaccharide standards are essential to better understand pEtN cellulose and its role in the ECM in anticipation of applications. While isolation of pEtN cellulose from natural sources generates ill-defined mixtures and may alter its chemical structure[149], chemical synthesis can provide standards with precise control over sequence, length, and substitution pattern[48,154]. However, to date, the inherent complexity of carbohydrate synthesis has prevented the production of pEtN cellulose oligomers beyond a disaccharide[155].

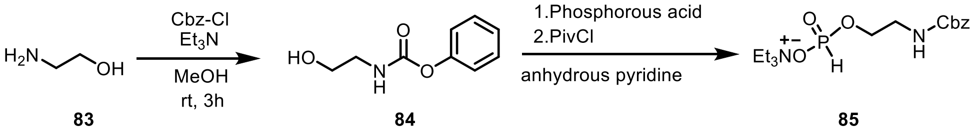
Here, I report a model system to create artificial biofilm matrices and study the complex interactions occurred in the ECM, based on synthetic epitomes. I synthesised six pEtN cellulose oligosaccharides (**86-88**) with different degree and pattern of pEtN substitution with AGA[60]. These compounds are important standards to characterize the natural polysaccharide, whose pattern is still unknown. As a synthetic epitome for the protein part, I have synthesized a subunit (**R5**) of the main protein (CsgA) exist in the ECM. The interactions of the glycans with a representative amylogenic peptide of curli (**R5**)[156] were studied. Co-assembly experiments generated artificial fibers and matrices with morphologies and mechanical properties depending on the oligosaccharide structure (carried out by Soeun Gim).

Unnatural, synthetic oligosaccharides such as chitin and sulfated cellulose (discussed in previous chapters), triggered the disruption or modulation of the artificial fibers. These results demonstrate that synthetic oligosaccharides are powerful tools to characterize complex ECMs and suggest that polysaccharide modification is a valuable approach to generate tunable biofilm-inspired materials.

## 5.2. Results

### 5.2.1. Synthesis of pEtN cellulose analogues

The pEtN-substituted oligosaccharides were prepared by a combination of AGA and post-AGA steps. The cellulose backbone was constructed by AGA, following cycles of glycosylation and Fmoc deprotection on solid support **4** (**Figure 47b**). Building block **38** allowed for linear chain elongation and **40** was designed with a levulinoyl (Lev) ester at C6 that can be selectively hydrolyzed to unmask the hydroxyl group for the subsequent introduction of pEtN. Strategically assembly of **38** and **40** generated oligomers with the desired pattern of hydroxyl groups. Building block **39** was employed in the last cycle of the assembly. After Lev removal, the oligosaccharide backbone was cleaved from the solid support and subjected to post-AGA transformations. The available hydroxyl groups were coupled to the H-phosphonate **85** to give the protected phosphorylated compounds (**Figure 47b**), upon oxidation with aqueous iodine[157]. For the synthesis of phosphorylating agent **85**, ethanolamine **83** was protected with benzyl chloroformate (Cbz-Cl) to afford compound **84**. Pivaloyl chloride used to promote the coupling between phosphorous acid ($H_3PO_3$) and **84** to yield H phosphonate **85** (**Scheme 2**).



**Scheme 2**. Synthesis of H-phosphonate **85**.

The synthesis of simple phosphorylated a mono-, di- and trisaccharide was tackled to optimize the post-AGA conditions and identify issues that could be encountered during the synthesis of more complex structures. Steric hindrance made multi-phosphorylation progressively more difficult, requiring five equiv. of **85** and pivaloyl chloride (PivCl) per hydroxyl group to reach full conversion. Six zwitterionic hexasaccharides bearing one, two or three pEtN units were prepared (**Figure 47b**). The neutral cellulose analogue $\mathbf{A_6}$ was synthesized as a control.

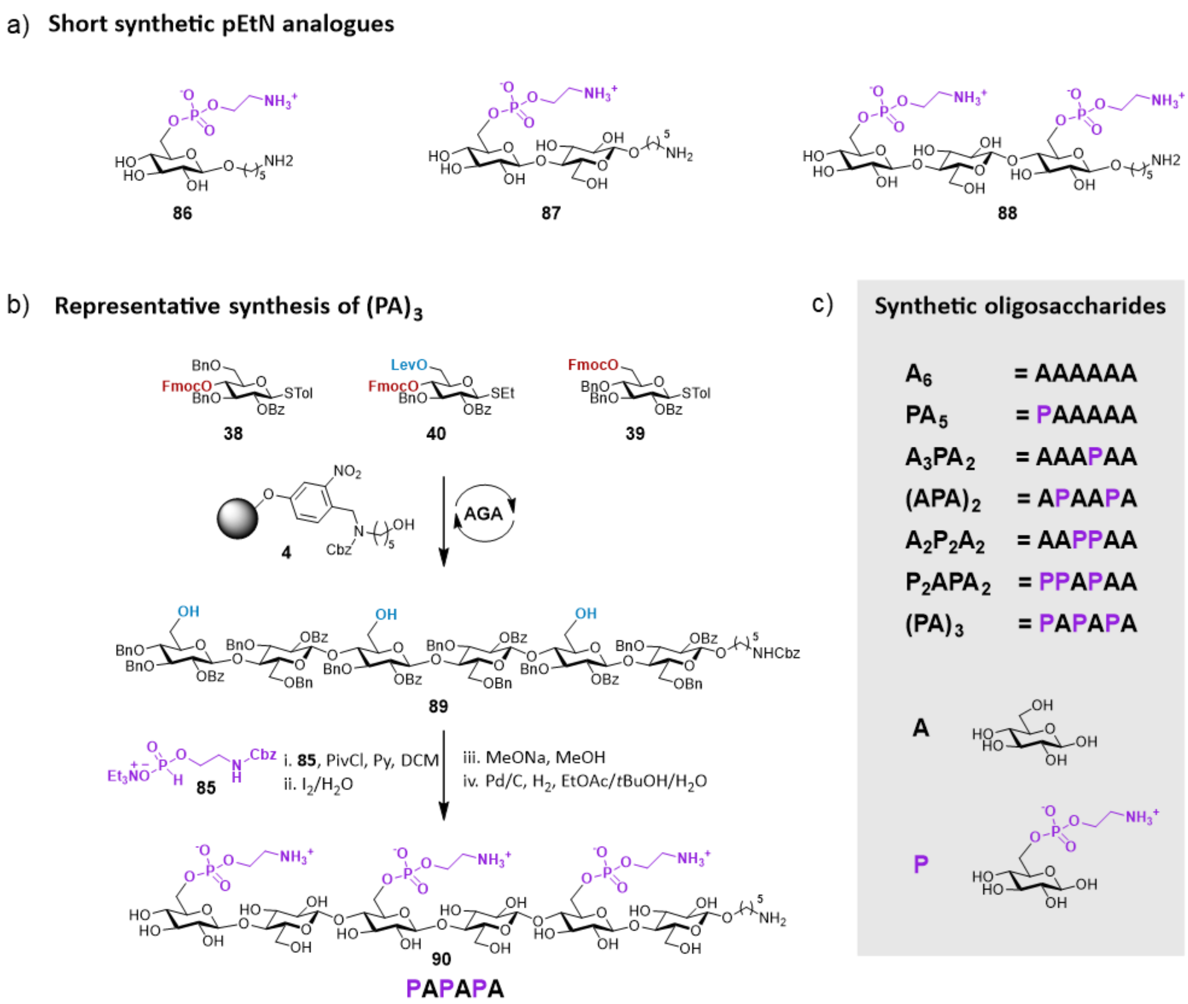


**Figure 47**. a) Model structures synthesized to develop the post-AGA protocol. b) Representative synthesis of a pEtN hexasaccharide. AGA includes cycles of glycosylation, capping, and Fmoc deprotection. A final Lev deprotection liberates the hydroxyl groups that are functionalized in post-AGA steps. PivCl = pivaloyl chloride, py = pyridine, MeONa = sodium methoxide. c) Collection of oligosaccharides synthesized in this work.

Removal of all the remaining protecting groups (PGs) via methanolysis and hydrogenolysis required a careful optimization of the reaction conditions to avoid aggregation/precipitation of the amphiphilic intermediates[158]. The methanolysis of the esters proved to be very slow. Heating to 40 ºC, increasing the concentration of MeONa, or addition of fresh reagents (after work-up of the reaction mixture) did not show any significant improvement. The reaction proceeded with a rough rate of one ester cleaved per day (**Figure 48**). Further difficulties arose as the reaction progressed, with solubility issues further slowing down the methanolysis. Addition of fresh reagents and THF finally afforded the desired semi-deprotected compounds. Notably, the addition of THF reduced the reaction time from 11 to 3 days.

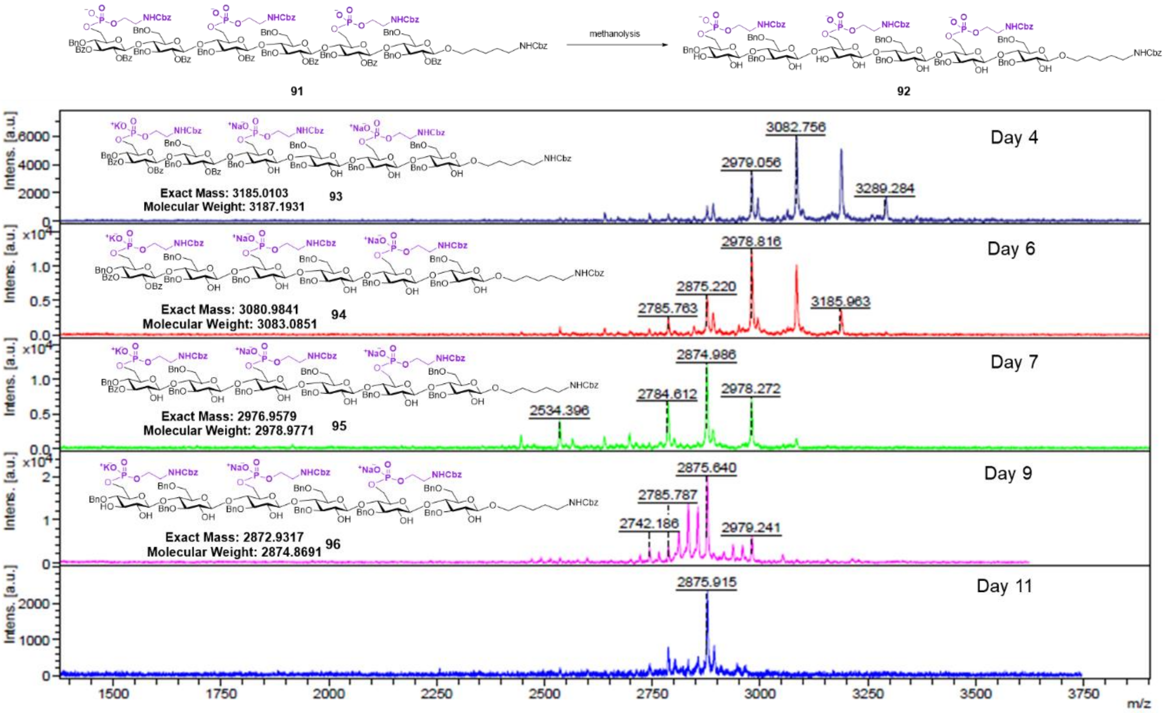


**Figure 48**. a) Stacked MALDI-TOF spectra, showing the slow progression of the methanolysis of **91.**

More difficulties were encountered during the hydrogenolysis of the benzyl ethers. The crude mixture of **97**, as obtained from methanolysis reaction step, was subjected to hydrogenation in the presence of Pd/C**.** No benzyl groups was cleaved at ambient pressure, while an over-night hydrogenation in a pressurized reactor (3 bar) cleaved only one ether group. The addition of THF to enhance the solubility of **97** during the hydrogenolysis, led to the unexpected side-product **98** (**Figure 49**).

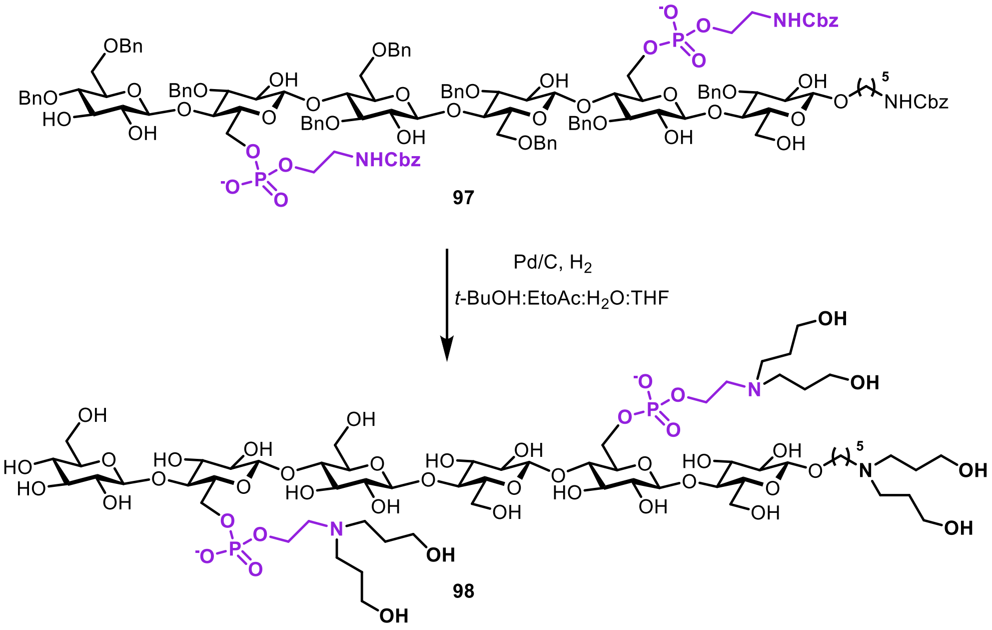


**Figure 49**. Unexpected side-product observed during the synthesis of the di-substituted pEtN cellulose.

A different solvent system (EA:*t*-Bu:AcOH—4:2:1) resulted in a mixture of 3 components: the desired product, the monosubstituted hexamer and the hexamer of glucose. When hydrogenolysis was performed (under pressure) before methanolysis, unexpected degradation of the starting material was observed. A purification of the semi-deprotected products by size exclusion solved the problem of the slow debenzylation without further reaction adjustments. A general optimized purification protocol was followed to obtain all the other pEtN oligomers (**Figure 50**).

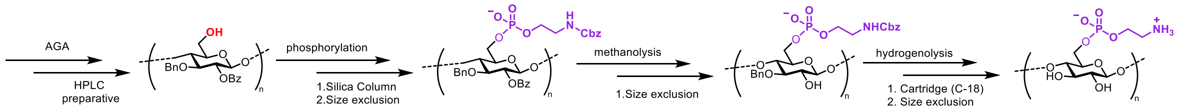


**Figure 50**. General purification protocol followed to obtain the pEtN oligosaccharides.

### 5.2.2. Synthesis of peptide R5

To enable the co-assembly experiments, the synthesis of a peptide representative of the natural curli protein (CsgA) was tackled. Short hexapeptide repeats that are included within the CsgA protein and contain conserved residues, have been reported to be amylogenic[159]. An automated microwave peptide synthesizer (CEM Liberty Blue) was used for the solid phase peptide synthesis (SPPS) of the two proposed oligopeptides **99** and **100**. In Fmoc-SPPS the α-amino group is temporary masked with an Fmoc group, which is removed by piperidine (20% in DMF) permitting chain elongation. The side chains are protected with acid-labile, typically *tert*-butyl- and trityl-based protecting groups. Both the linker (**101,** trityl linker coupled on Merrifield resin) and side chain protecting groups are cleaved in the final step by using 90-95% TFA (**Figure 51**).

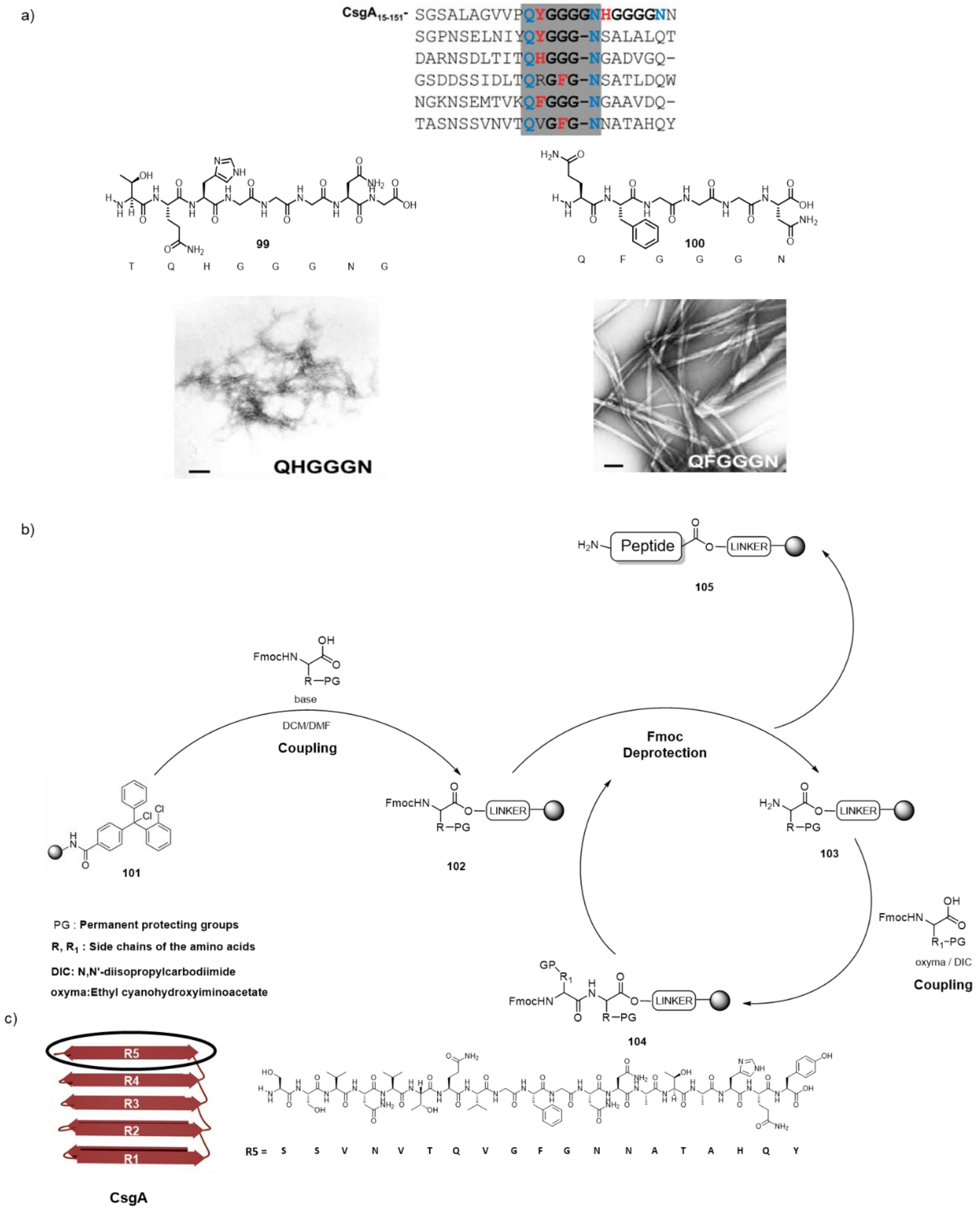


**Figure 51.** a) Identification of CsgA oligopeptide repeats. Highly conserved residues within the repeated domains are indicated with bold letters and coloured according to the residue type. TEM analysis of the self-assembly properties of oligopeptide repeats of hexapeptides **99** and **100**. b) The principle of Fmoc-SPPS. c) CsgA curli protein and chemical structure of **R5**.

Surprisingly, the cleavage from the solid support, after completion of the synthesis, led to a cyclized side-products and to cumbersome purification process. A head-to-tail cyclization was promoted by the acidic conditions of the cleavage. A stronger evidence from the literature[160] for a more amylogenic sequence led me to target the synthesis of **R5**, a CsgA whole repeating unit. The synthesis was successful and, after cleavage from the solid support, the crude mixture was purified by HPLC. The purification proved to be extremely challenging since the crude mixture was poorly soluble in water. Efforts to enhance the solubility by sonication and increased temperature failed to produce a clear solution. After filtration, the resulted transparent solution progressively turns into a suspension, indicating the formation of new aggregates. A significant part of **R5** was lost during either filtration or HPLC purification, resulting in a low isolated yield (2%). This issue was overcame by using a manual C-18 column and by decreasing the time of the purification process, to avoid aggregation of **R5** inside the column. The isolated yield increased to 11%.

### 5.2.3. Assembly of artificial amyloid fibers

Well-defined pEtN oligosaccharides provided the bases for exploring the role of the carbohydrate component in *E. coli* biofilm formation. I envisioned an artificial model system consisting of synthetic molecules representatives of the major components of the *E. coli* ECM. **R5** was selected as epitome for the protein part (**Figure 52**). To generate artificial curli fibers (done by Soeun Gim), **R5** and the different oligosaccharide were dissolved in hexafluoroisopropanol (HFIP)[161,162]. HFIP was then removed under nitrogen purging followed by evaporation under high vacuum. Addition of water triggered a structural transition from an alpha helix to a beta-sheet conformation, as confirmed by circular dichroism (CD) spectroscopy.

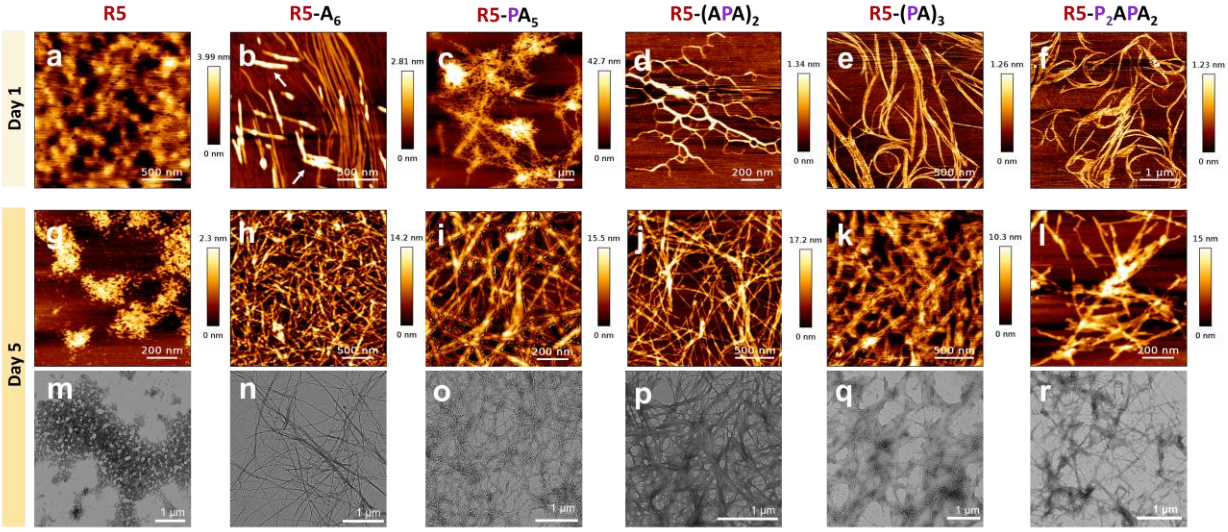


**Figure 52**. Assembly of **R5** in the presence of selected oligosaccharides. AFM (Day 1: **a**-**f** and Day 5: **g**-**l**) and TEM (Day 5: **m**-**r**) images of the samples containing **R5** alone or in the presence of different hexasaccharides. Fibers are generated in all the samples containing an oligosaccharide, whereas the **R5** alone sample results in random aggregation. In **Figure 52b**, the aggregation of **$A_6$** is indicated with white arrows.

Microscopic analysis (done by Soeun Gim) showed that the **R5** sample containing the cellulose oligomer **$A_6$** assembled into thin fibrils (**Figure 52b**, Day 1) that developed into a fibrous network within five days (**Figure 52h**, Day 5). Next, we examined the effect of the six different pEtN cellulose hexasaccharides on the assembly of **R5**. While the samples containing the three-substituted oligomers **$(PA)_3$** and **$P_2APA_2$**

showed long and defined fibrils already on Day 1 (**Figure 52e-f**), the less substituted analogues formed shorter aggregates (**Figure 52c-d**). Interestingly, the fibers observed for **R5** and **(PA)$_3$** adopt the classical curled shape responsible for the name of the natural analogue (**Figure 52e**)[163]. On Day 5, all samples formed fibrous networks.

## 5.2.4. NMR analysis

To gain insights into the molecular interaction between **R5** and the oligosaccharides, we employed solution-state NMR spectroscopy following an approach that revealed key interactions between synthetic heparin oligosaccharides and amyloid fibers[164]. 2D $^1$H-$^1$H total correlated spectroscopy (TOCSY) helped the assignment of the nineteen amidic protons of **R5** (**Figure 53**).

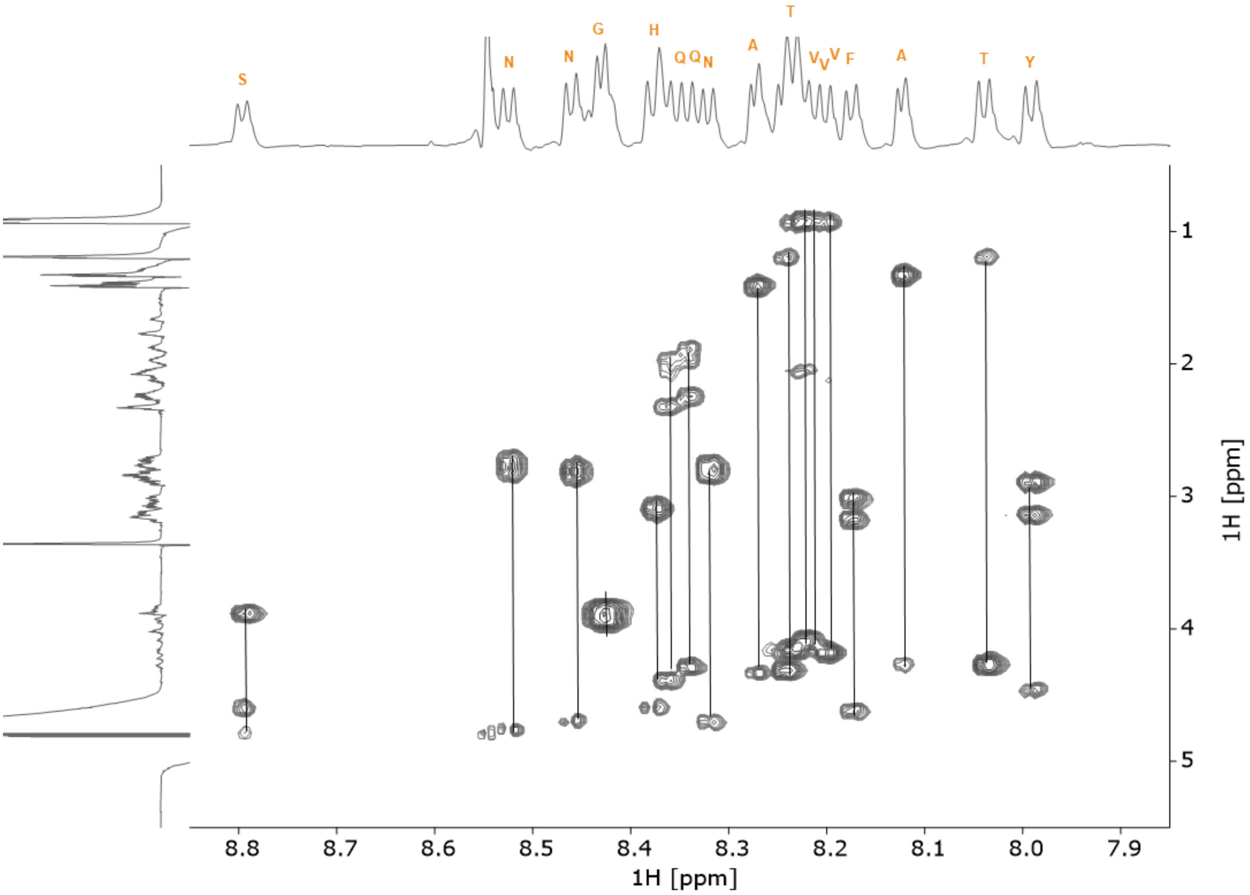


**Figure 53**. **R5** assignment of the –NH signals. The 2D $^1$H-$^1$H TOCSY spectrum was recorded with a **R5** concentration of 200 µM in $H_2O/D_2O$ (9:1) at 25 °C.

This sample suffered from poor solubility due to aggregation, as shown by the broadening and decreased intensity of the NMR signal with time. The three samples containing both **R5** and **A$_6$**, **(PA)$_3$** and **P$_2$APA$_2$** respectively showed chemical shift perturbations for selected amino acids, suggesting their role in the interaction with the oligosaccharides (**Figure 54a**). Tyrosine, glutamine, histidine and serine were the most affected amino acids in all three samples, albeit to a different extent (**Figure 54a**, top panels). The $^{31}$P-NMR signals of the pEtN groups did not show any significant perturbation (**Figure 54b**), suggesting that the pEtN groups are not directly involved in the interaction with **R5**. The interaction between **R5** and the oligosaccharides might be the reason for the slower **R5** transition into the beta-sheet conformation, favoring the formation of long amyloid fibers over ill-defined aggregates[165,166].

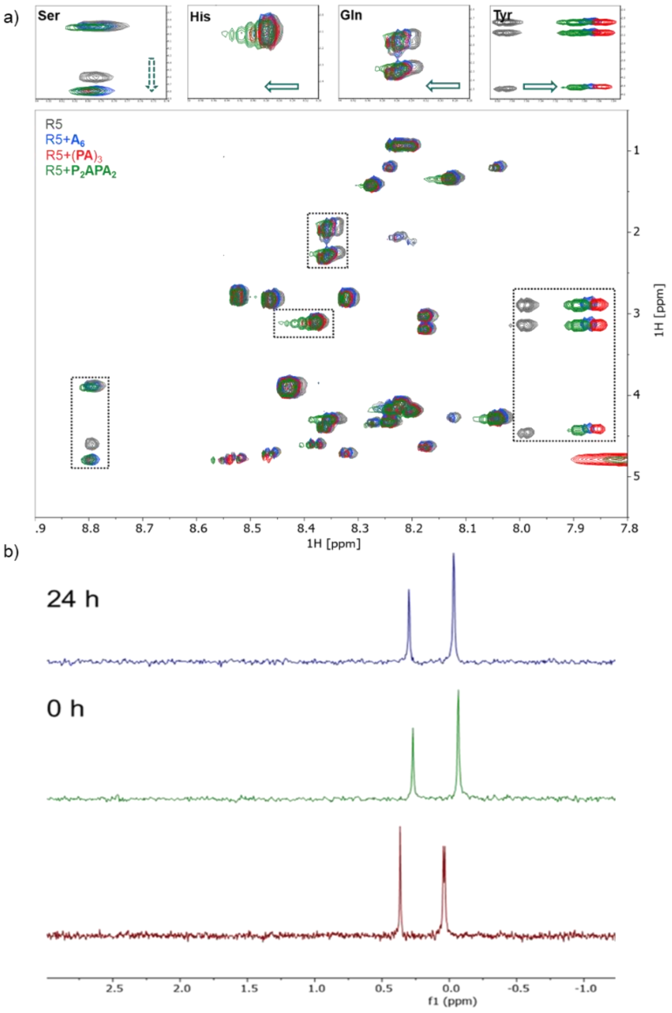


**Figure 54**. a) Structural analysis of the fibers generated from the assembly of R5 in the presence of three different oligosaccharides. Overlay of a selected region of the $^1$H-$^1$H TOCSY spectra for the samples containing **R5** alone (grey) and in the presence of **A$_6$** (blue), **(PA)$_3$** (red) and **P$_2$APA$_2$** (green). Each spectrum was recorded with a **R5** concentration of 200 µM in $H_2O/D_2O$ (9:1) at 25 °C. The four amide protons mostly affected by the interaction are highlighted (top panels), showing a change in chemical shift (His, Gln, Tyr) or signal broadening (Ser). b) $^{31}$P NMR of **(PA)$_3$** (brown) and **(PA)$_3$** in the presence of **R5** (green and blue) recorded at different time intervals in $H_2O/D_2O$ (9:1) at 25 °C.

## 5.2.5. Mechanical properties of artificial biofilm matrices

The mechanical properties of artificial biofilms were probed using AFM force-distance curve analysis (performed by Soeun Gim) (**Figure 55**). A stiffness of ca. 12 MPa for all the matrices was measured with AFM nanoindentation experiments, indicating that the peptide fibres are the major structural component of the artificial biofilm matrix. The presence of the pEtN-modified oligosaccharides dramatically enhanced adhesion. The adhesion force for the sample containing **R5** and **(PA)$_3$** was around 130 nN, six times higher than the value obtained for the sample containing **R5** and **A$_6$**. No direct correlation between the number of pEtN groups and the adhesion was found. The highest values were measured for compounds with the pEtN moiety coupled to the non-reducing end glucose (i.e., **$(PA)_3$** and **$PA_5$**). Multiple pEtN substituents in close vicinity (e.g., **$A_2P_2A_2$**) resulted in much lower adhesion forces, underscoring the importance of the substitution pattern in determining the mechanical properties of the biofilm.

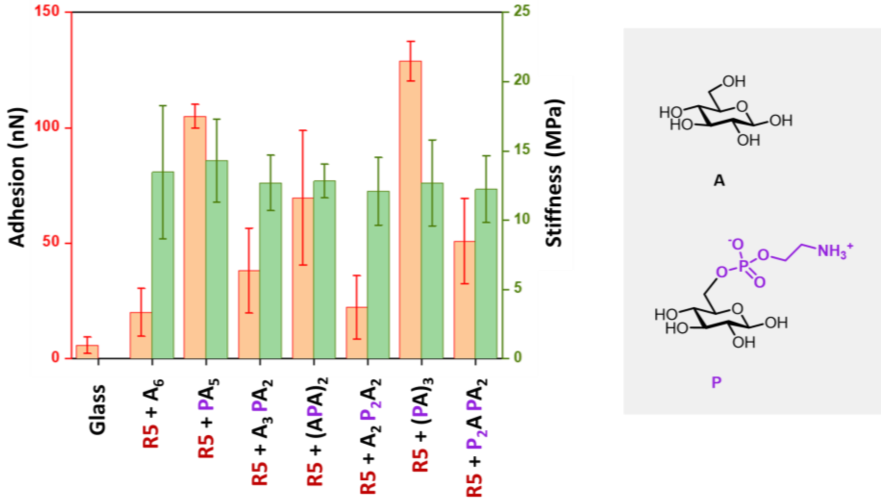


**Figure 55**. Mechanical properties of artificial biofilm matrices generated from the assembly of **R5** in the presence of different oligosaccharides.

### 5.2.6. Tuning artificial biofilms with synthetic oligosaccharides

The discovery of the naturally modified pEtN-cellulose opened up opportunities to generate tuneable materials upon engineering of the carbohydrate components[22]. It has been shown that carbohydrates can modulate the formation of neurotoxic amylogenic fibrils, with chitosan oligosaccharides inhibiting aggregation[167] and heparan sulfates promoting fiber formation[168]. In an effort to tune the morphology and properties of bacterial biofilm matrices, two oligosaccharides not present in natural bacterial biofilms were prepared following established protocols[169] (**Figure 56**). In the presence of the *N*-acetyl glucosamine hexasaccharide **$N_6$**, fibrils shorter than 1 µm that further aggregated into supramolecular bundles were formed (**Figure 56a-c**). In contrast, the negatively charged sulfated hexasaccharide **$(SA)_3$** interrupted the formation of fibrils (**Figure 56e-g**). The ability of the sulfated hexasaccharide, **$(SA)_3$**, to inhibit amyloid formation renders this compound an interesting starting point for novel approaches towards the treatment of neurological diseases or as antibacterial agent[170,171].

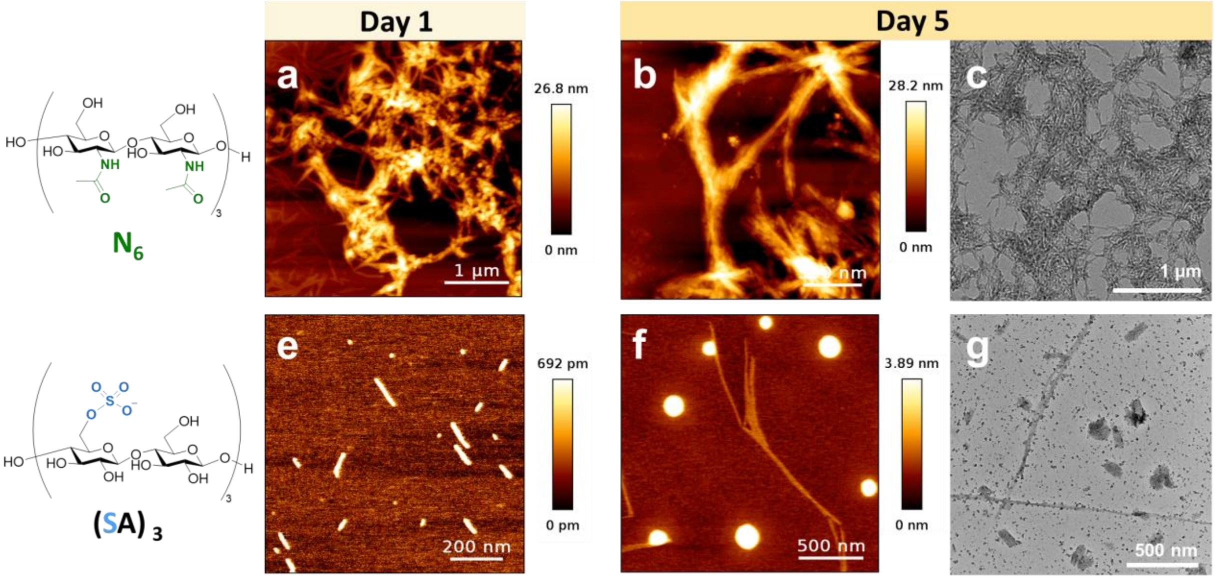


**Figure 56.** Exploring the effect of unnatural oligosaccharides on the assembly of **R5**. AFM (Day 1: a,e and Day 5 : b,f) and TEM (Day 5: c,g) images, of amyloid fibrils prepared with **R5** in the presence of $\mathbf{N_6}$ (top) or $(SA)_3$ (bottom).

## 5.3. Conclusions

I have synthesized a collection of pEtN-modified oligosaccharides with full control over degree and pattern of substitutions. In addition, I have synthesized short peptides as representatives of CsgA, choosing **R5** for our model system. These compounds are essential standards to elucidate the role of pEtN cellulose in the formation of *E. coli* biofilm. The oligosaccharides were incubated with a synthetic peptide, **R5**, representative of curli, to generate a modular model of the *E. coli* biofilm ECM and break down its complexity. Full control over the chemical structure of the individual components permitted to explore the peptide-carbohydrate interactions involved in biofilm formation, offering a complementary approach to genetic engineering. The oligosaccharides slowed down the secondary structure transition into beta-sheet of the amylogenic peptide, **R5**, inducing the growth of extended fibrous structures. The oligosaccharide fine structure dramatically affected the fiber growth rate and the mechanical properties of the composite. We demonstrated that not only the degree, but also the pEtN pattern influences film adhesion. In contrast, stiffness remains unchanged for all samples indicating its strong connection to the peptide component. These results demonstrate that much more attention should be devoted to the design of bacterial polysaccharides to develop new classes of tunable matrices. Unnatural oligosaccharides were screened, delivering interesting targets for the future production of antibacterial agents or engineered biofilms. Metabolic engineering[172] and/or directed evolution approaches[173,174] may introduce such modifications *in vivo* and produce novel cellulosic materials with non-natural modifications.

# 6. Conclusion and outlook

The ultimate goal of my thesis was to improve the fundamental understanding of carbohydrate structure and to establish structure-property correlations. Particular focus was given to how the conformational space accessible by an oligosaccharide defines its aggregation properties, providing important information for the rational design and application of carbohydrate materials. This ambitious goal required the implementation of state-of-the-art modelling techniques, synthetic approaches, and assays to characterize polysaccharide interactions. Future developments can be imagined for all these topics and are listed below.

I employed **Molecular Dynamics** to gain insight into the conformation of cellulose-like oligomers at the molecular level and guide their synthesis. An atomistic analysis of the 3-D conformation revealed key interactions and structural changes than can affect and rationalize their molecular packing. To depict the interactions occurring between oligosaccharides, I developed a simulation protocol that involves a crowded environment of solutes. These simulations demonstrated how the different geometries observed in the single solute analysis affect the aggregation process, but also how aggregation has an effect on the molecular conformations of the single chains. Still, even though atomistic MD of concentrated environments gave a tremendous insight on the stacking process, its application to predict crystalline structures is limited by its algorithmic efficiency and the available computing power[175]. Coarse-grained[176] models are computationally more effective and enable simulations of much longer time-scales and/or larger sizes of the systems, despite losing some chemical information[101]. Using the information generated by atomistic experiments, multiscale modelling will be implemented to explore the formation of carbohydrate materials (**Figure 57**). Another development of this project involves the implementation of Replica-Exchange Molecular Dynamics[177] (REMD). This is a recent technique used to enhance sampling of standard molecular dynamic simulations by allowing systems of similar potential energies to sample conformations at different temperatures. By doing so, energy barriers on the potential energy surface might be overcome, allowing for the exploration of new conformational space. Applying REMD to oligosaccharides will be considered to deliver a more reliable descriptions of the studied sample.

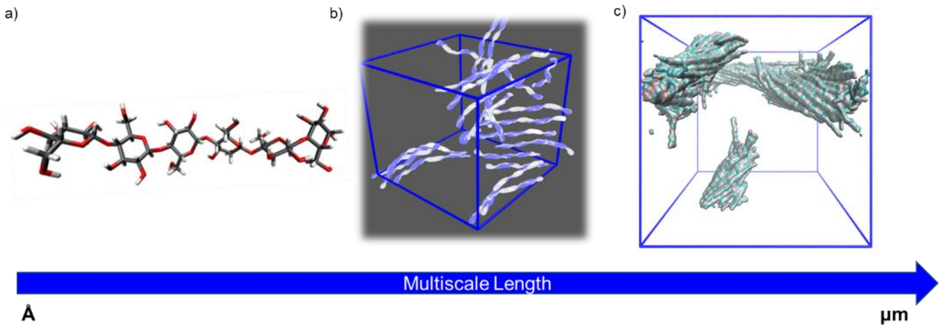


**Figure 57.** The multiscale length of modeling approaches. a) Classical all-atom MD, b) Classical all-atom MD of concentrated environment. c) Coarse-grained molecular simulations.

Experiments play an essential role in validating the simulation methodology: comparisons of simulation and experimental data serve to test the accuracy of the calculated results and to provide criteria for improving the modelling. MD predictions were supported by XRD analysis, solubility measurements (Chapter 2) and NMR spectroscopy (Chapter 3). In some cases, ionic interactions were overestimated, demanding further optimization of the dihedral potentials, especially in the presence of ionic moieties (e.g amines).

A prerequisite for the molecular level understanding of oligosaccharide structures is the availability of efficient **synthetic methods** to obtain pure and well-defined structures. In the context of this dissertation, I have developed new protocols and optimized reactions to access complex glycans as probes to study their interactions with other biomolecules. A general protocol for the on resin synthesis of sulfated glycans that bypasses many of the previously challenges associated with sulfated oligosaccharide synthesis has been described (Chapter 4). This project can be further extended to allow for *N*-sulfonation, another common modification found in GAGs. More in general, new modifications could be implemented and optimized to be performed on resin, avoiding issues associated with purification steps often encountered in solution phase synthesis (see Chapter 5). I envision new on resin protocols to access various charged oligosaccharides, including phosphorylated compounds. Such targets are important standards to enable the structural characterization of ionic oligosaccharides, implicated in several biological events (**Figure 58**).

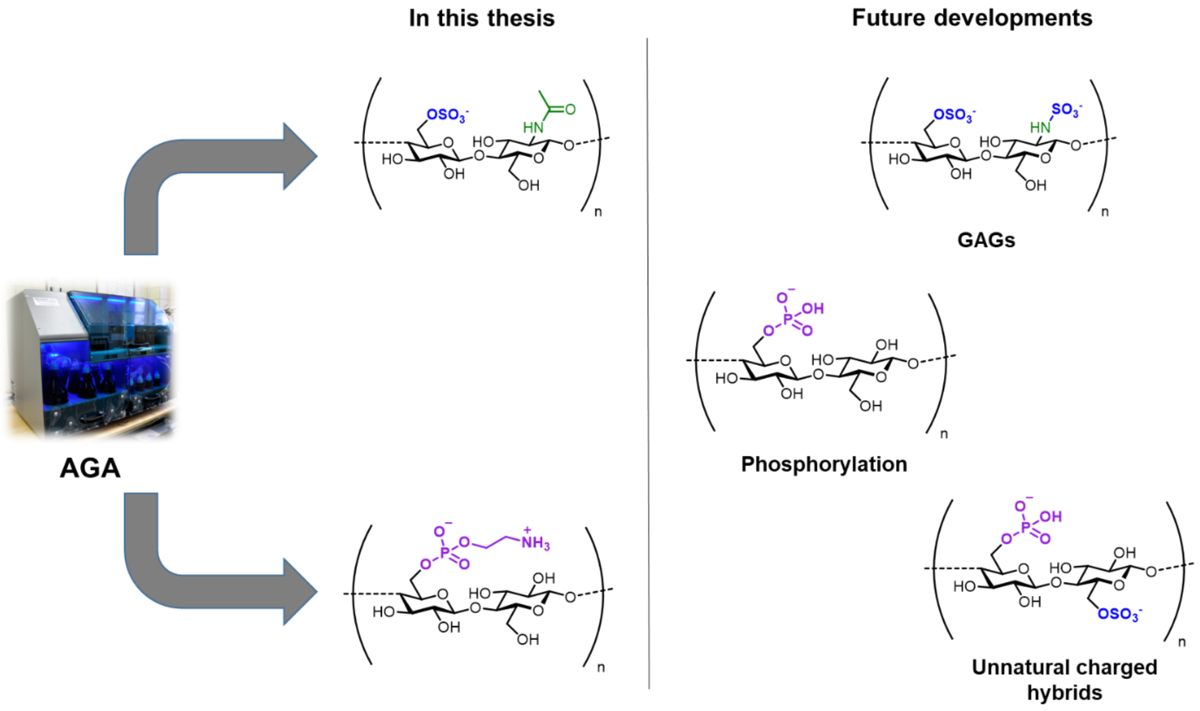


**Figure 58.** Synthetic ionic compounds obtained in the context of this thesis (right) and suggested follow up work to implement new on resin reactions (left).

In Nature, polysaccharide modifications might offer a strategy to alter the properties of a material, as exemplified by the introduction of pEtN groups onto the cellulose backbone (Chapter 5). For this project, I have prepared synthetic oligomers representative of extracellular components with full control over their chemical composition. These synthetic epitomes were used to create artificial biofilm matrices and assess carbohydrate-protein **interactions** in *E.coli* ECM. I demonstrated that the polysaccharide component plays a central role in tuning the mechanical properties of biofilms. In the future, synthetic oligosaccharides could be applied to living bacterial cells, to explore their ability to direct the production of tunable matrices. This work could offer a complementary approach to genetic engineering approaches[138] to develop new engineered living materials (**Figure 59**). Moreover, our model allowed me to screen the effect of oligosaccharides bearing unnatural modifications on biofilm formation. The ability of the sulfated hexasaccharide to inhibit amyloid formation renders this compound an interesting starting point for novel approaches towards the treatment of neurological diseases or as antibacterial agent[170,171].

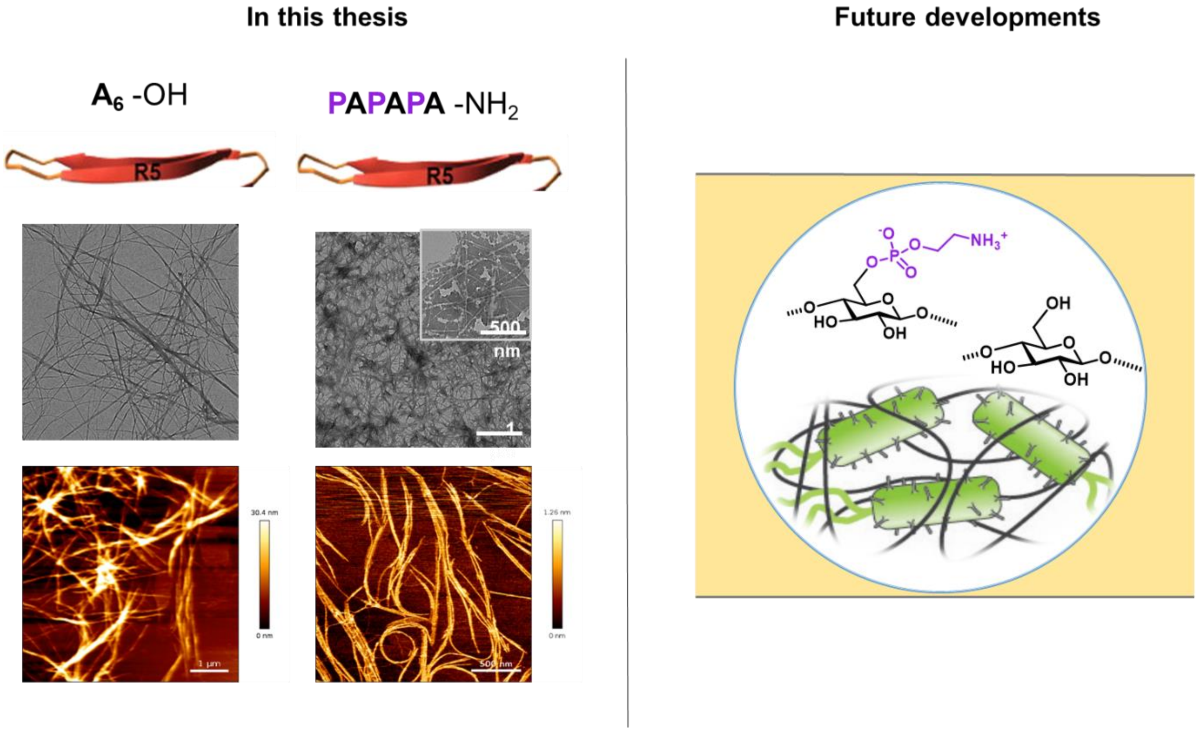


**Figure 59.** Synthetic pEtN oligomers demonstrated that the polysaccharide component plays a central role in orchestrating the formation of ECM (left). Synthetic oligosaccharides could be incubated with bacterial living cells to produce new matrices (right).

# 7. Experimental section

In this section, only my contributions to each project will be included unless otherwise specified. Additional experimental data can be found in the publication mentioned at the beginning of each chapter.

## 7.1. Molecular Dynamic simulations

Initial conformations for single hexamer and dodecamer simulations were constructed with the Glycam Carbohydrate builder and tleap. The topology was subsequently converted using the glycam2gmx.plscript and solvated with 2100 TIP5P[178] water molecules using gromacs tools. Partial charges for non-standard monomers were derived using the R.E.D. tools scripts following the GLYCAM06 protocol. Structure optimization for the charge derivation was performed using Gaussian at the HF/6-31G* level of theory for neutral and cationic groups, and the HF/6-31++G** level of theory for anions Concentrated systems were prepared by randomly inserting, 25 hexamer molecules in a simulation box (6 nm x 6 nm x 6 nm) with gmx_insert-molecules. The system was then solvated with 84 TIP5P water molecules per hexamer.

Molecular Dynamics simulations for each system were performed using gromacs 5.1.2[179]. The systems were kept at a constant temperature of 303 K using a Nosé- Hoover thermostat and at constant pressure of 1 bar with the Parrinello-Rahman barostat. Non-bonded interactions were cut-off at 1.4 nm, long range electrostatics were calculated using the particle mesh ewald method. Bonds involving hydrogens were constrained using the LINCS[180] to allow a 2 fs time step algorithm; water molecules were kept rigid with SETTLE[181].

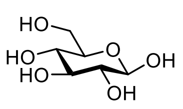
After energy minimization (steepest descent algorithm) and before the production run, the systems were equilibrated at 300 K for 20 ns in a canonical (NVT) ensemble (constant number of particles, volume and temperature) and subsequently at 300 K and 1 atm for 20 ns in an isothermal-isobaric (NPT) ensemble. A longer equilibration procedure (50ns of npt and 400ns of NVT) was performed for the concentrated experiments.

All hexamers were simulated for 500 ns, all dodecamers for 400ns and concentrated solutions for 1 µs. For all methylated (A and B) and carboxymethylated (D) monosaccharides dihedral restraints were applied to the dihedral angle $C_1$-$C_2$-$C_3$-$C_4$, using an equilibrium angle of -47° and a force constant of 20 20 20 (x, y, z ; kJ/(mol rad^2) ), in order to prevent the ”ring flip” artifact.

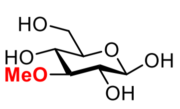
Radial distribution functions were calculated between the centers of mass of the oligosaccharides by invoking the gmx rdf command with the option (whole_mole_com).

# SF1

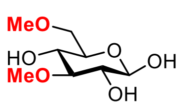
Root-mean-square deviation of atomic positions (**RMSD**)



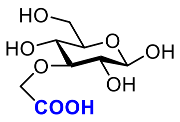
**A**



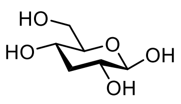
**B**



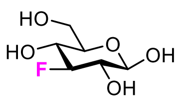
**C**



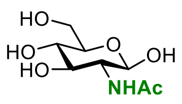
**D**



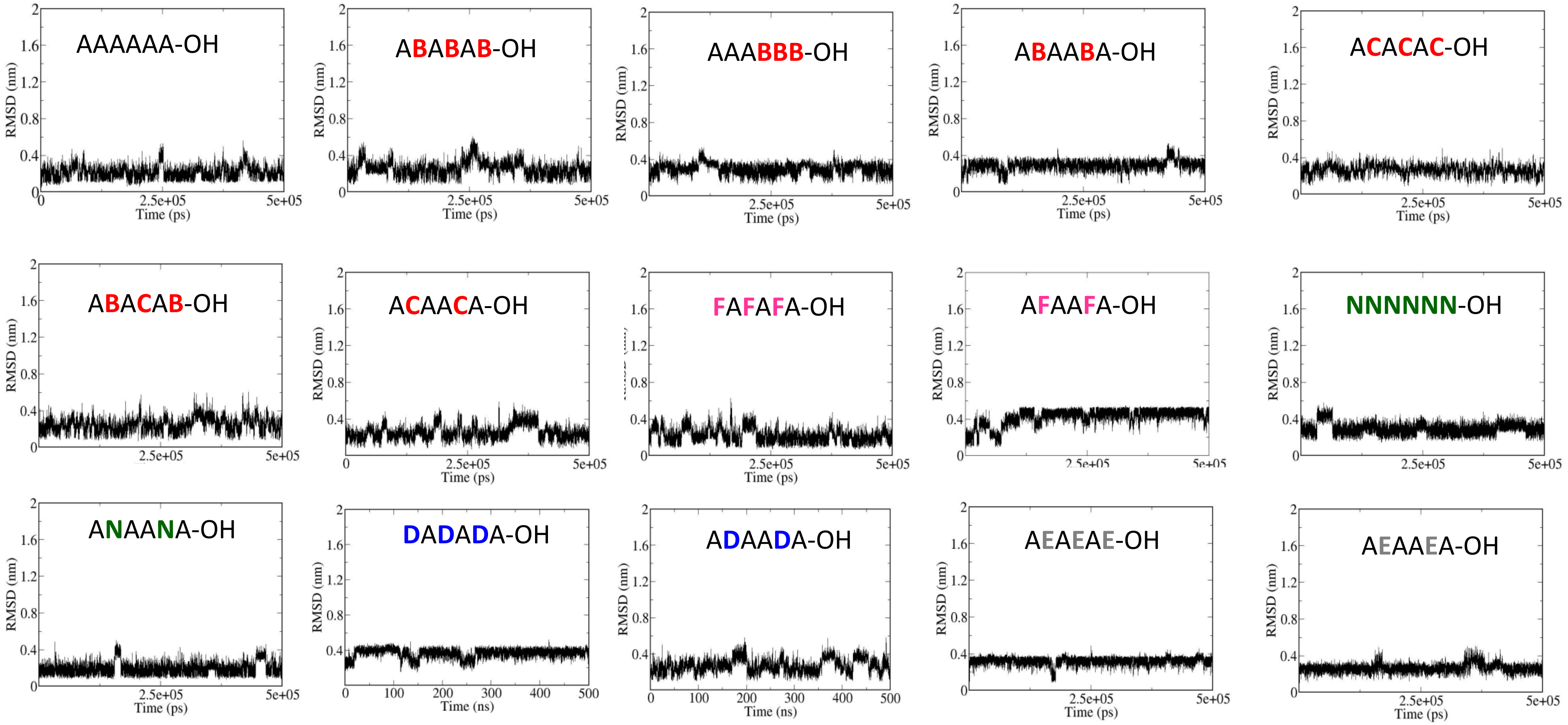
**E**



**F**

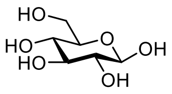


**N**

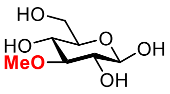


# SF2

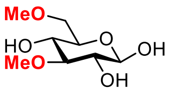
Histograms of radius of gyration (**RoG**)



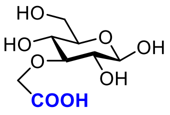
**A**



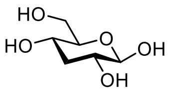
**B**



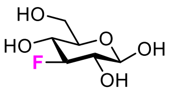
**C**



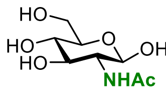
**D**



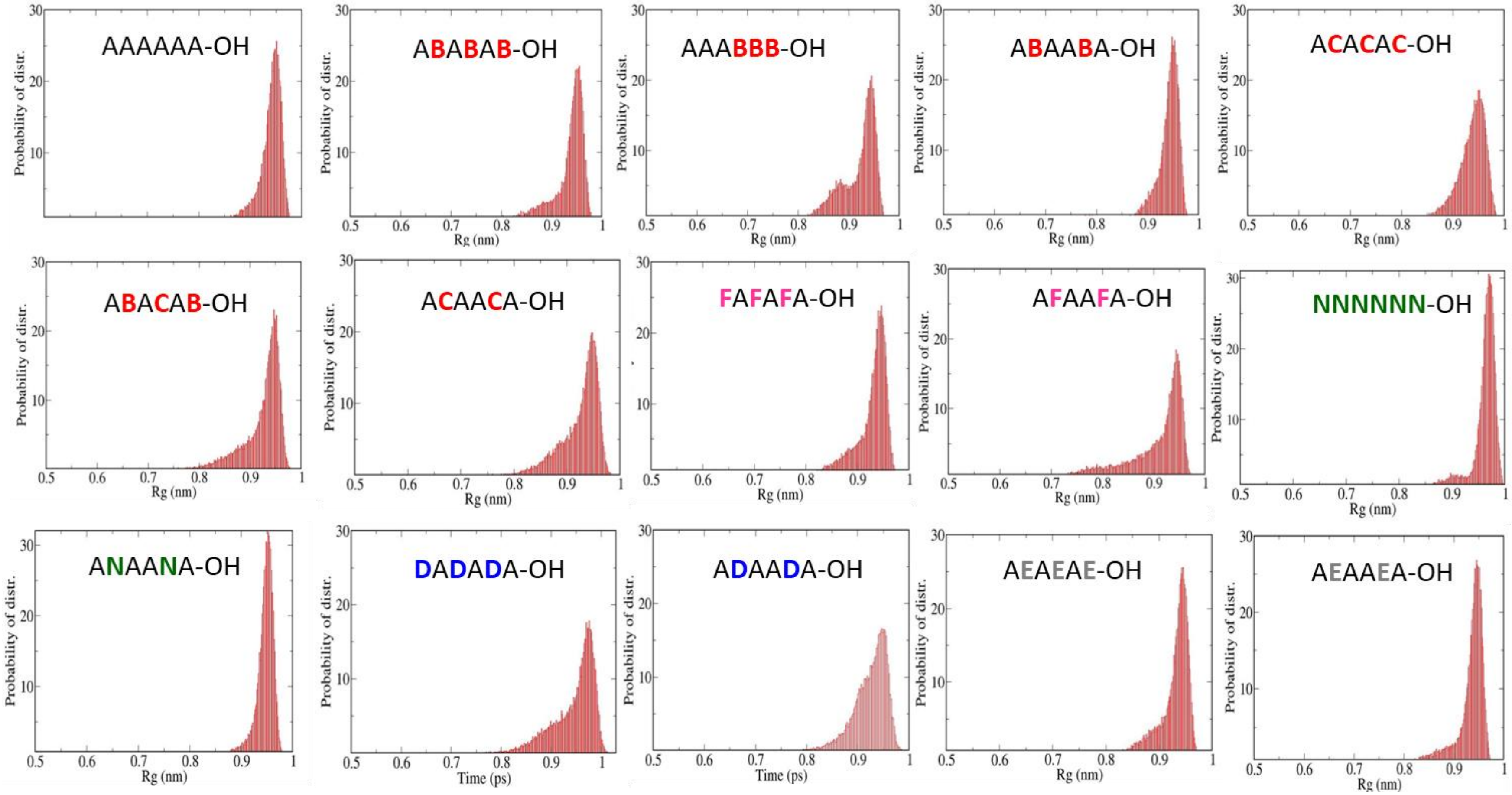
**E**



**F**

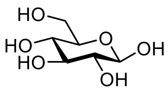


**N**

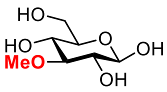


# SF3

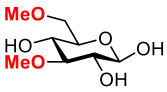
## End-to-end distance as a function of time



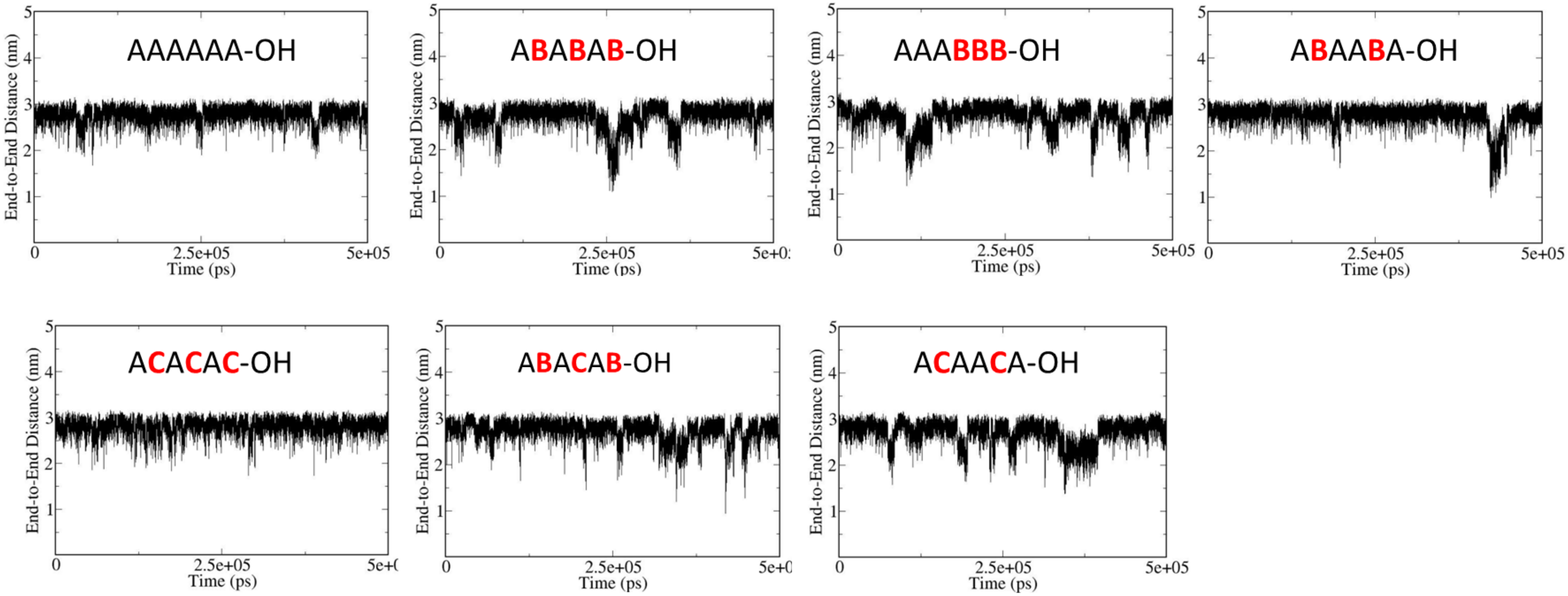
**A**



**B**

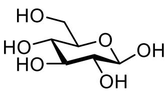


**C**

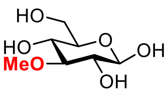


| Hexamer | AAAAAA-OH | ABABAB-OH | AAABBB-OH | ABAABA-OH | ACACAC-OH | ABACAB-OH | ACAACA-OH |
|---|---|---|---|---|---|---|---|
| Average Distance (nm) | 2.76 | 2.71 | **2.65** | 2.75 | 2.80 | 2.73 | 2.72 |
| Standard Deviation | 0.17 | 0.25 | **0.26** | 0.24 | 0.17 | 0.23 | 0.25 |

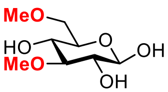
# End-to-end distance as a function of time



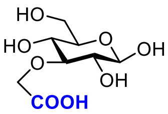
A



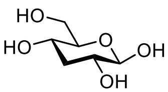
B



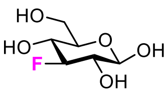
C



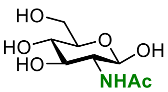
D



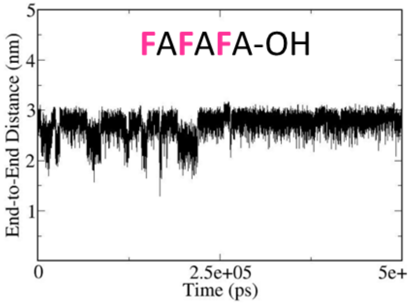
E

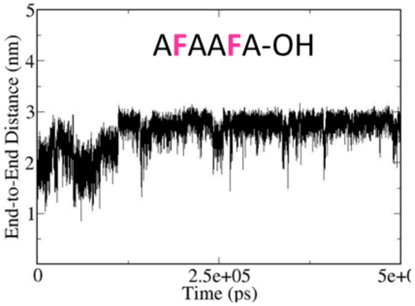


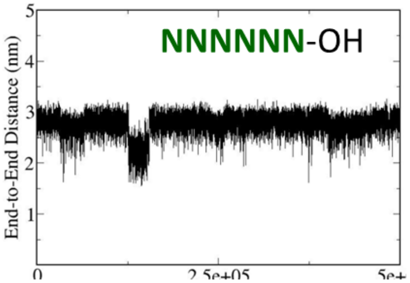
F

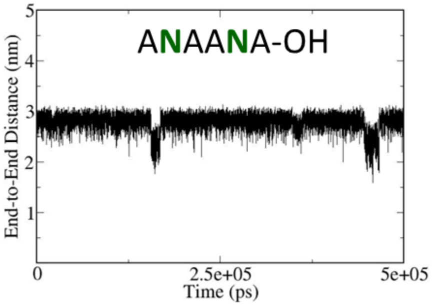


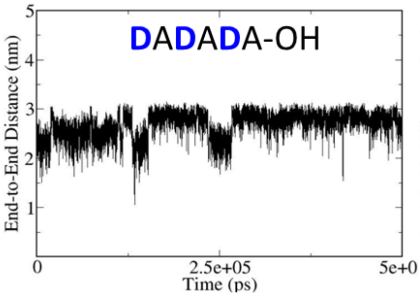
N

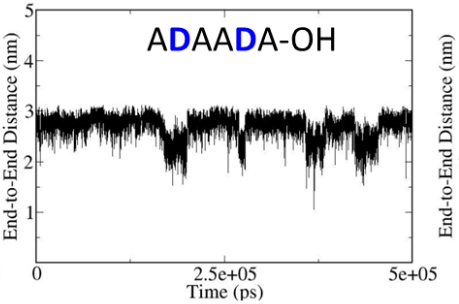


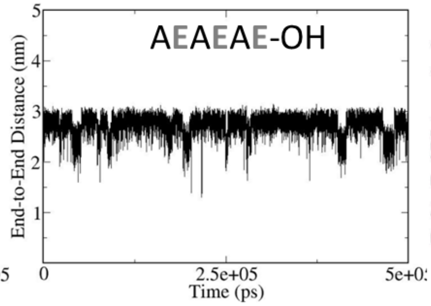


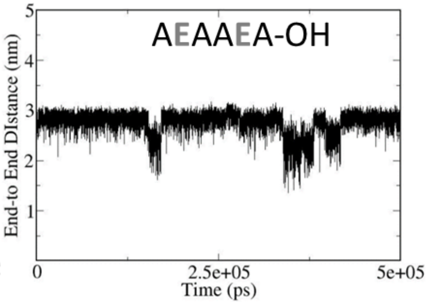












| Hexamer | FAFAFA-OH | AFAAFA-OH | NNNNNN-OH | ANAANA-OH | ADAADA-OH | DADADA-OH | AEAEAE-OH | AEAAEA-OH |
|---|---|---|---|---|---|---|---|---|
| Average Distance (nm) | 2.72 | **2.61** | 2.76 | 2.79 | 2.70 | 2.69 | 2.73 | 2.76 |
| Standard Deviation | 0.26 | **0.34** | 0.22 | 0.17 | 0.24 | 0.26 | 0.21 | 0.22 |

**SF4**

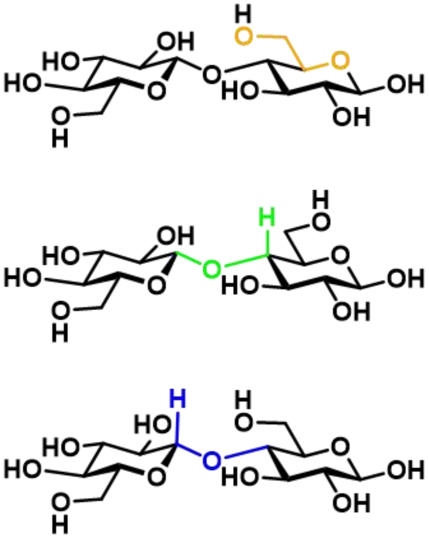
# Definition of dihedrals in a glucose disaccharide

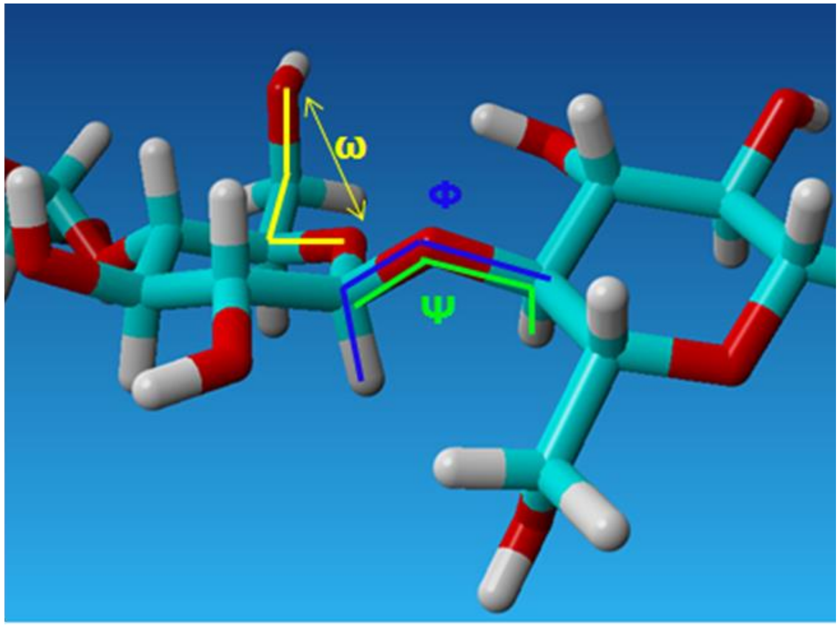
ω = O5C5C6O6

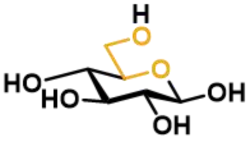
Ψ = C1O4C4H4

Φ = H1C1O4C4

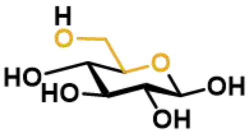
(The atoms in red belong to the following residue)



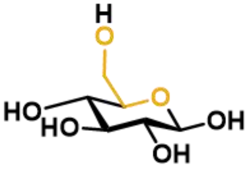




gt (ω = 60°)

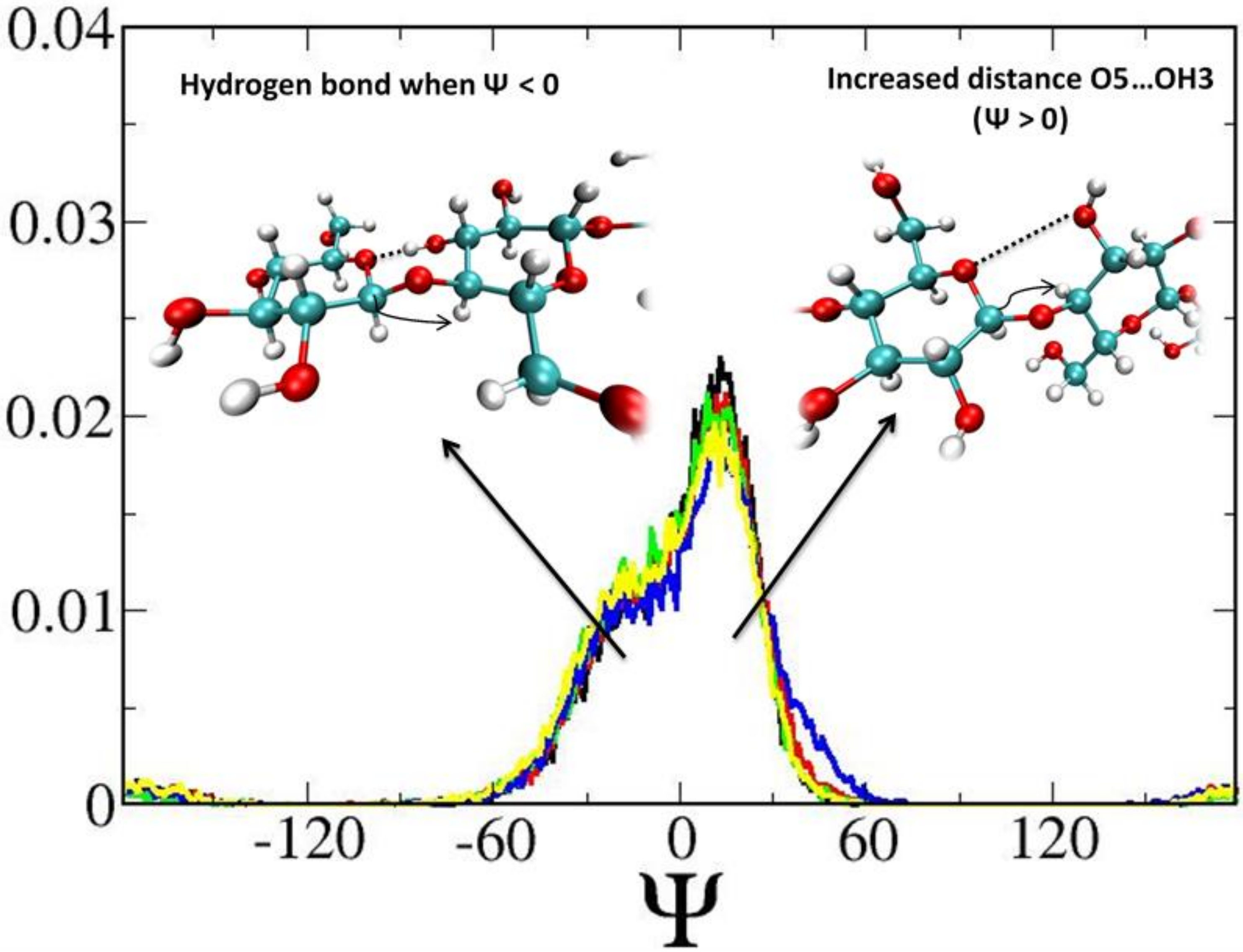


tg (ω = 180°)



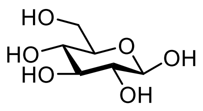
gg (ω = - 60°)

**SF5**

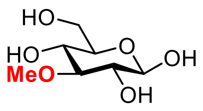

Hydrogen bond when Ψ < 0
Increased distance O5…OH3
(Ψ > 0)
0.04
0.03
0.02
0.01
0
-120
-60
0
60
120
Ψ

# SF6

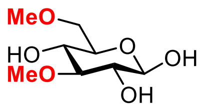
## Omega (ω) torsion angles.



**A**



**B**

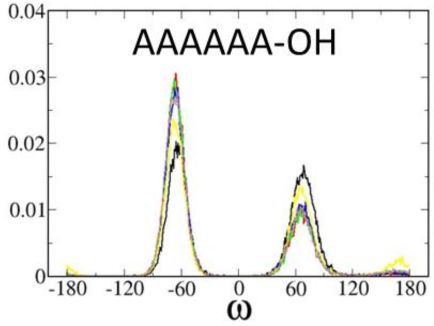


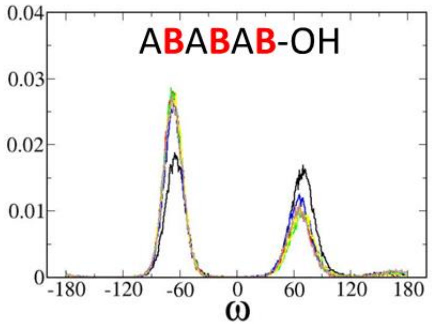
**C**

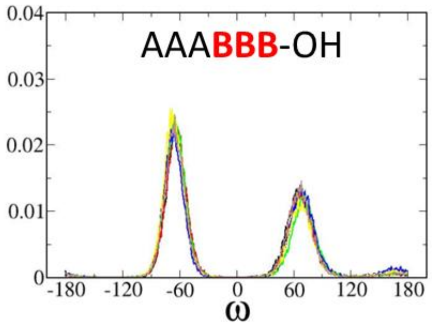
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Omega (ω) 1 | ⬛ | Omega (ω) 2 | 🟥 | Omega (ω) 3 | 🟩 | Omega (ω) 4 | 🟦 | Omega (ω) 5 | 🟨 | Omega (ω) 6 | 🟩 |

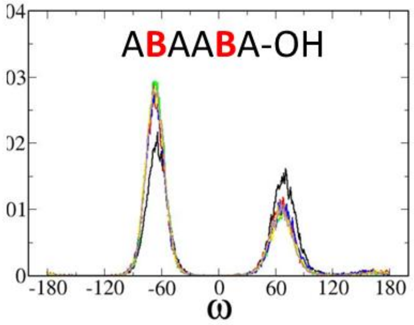
Non-reducing end

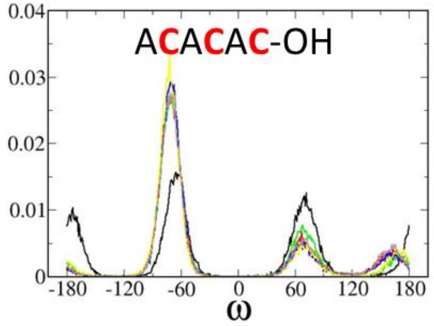
Reducing end

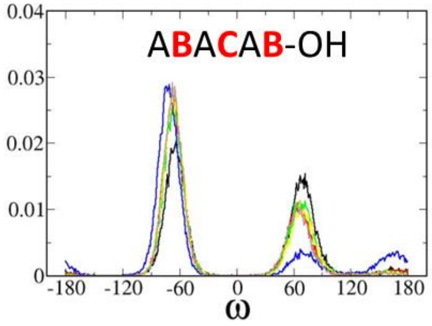


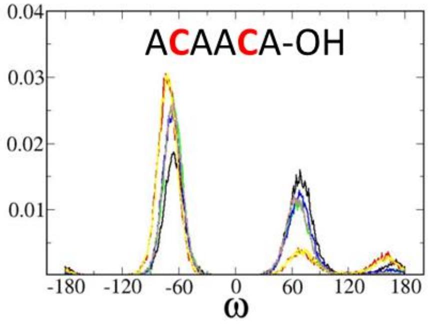




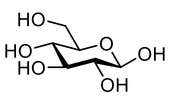




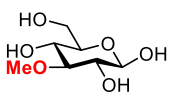




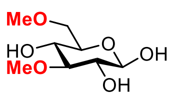
# Omega (ω) torsion angles.



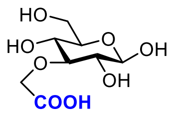
A



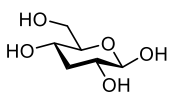
B



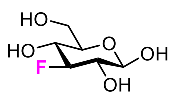
C



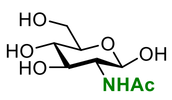
D



E



F

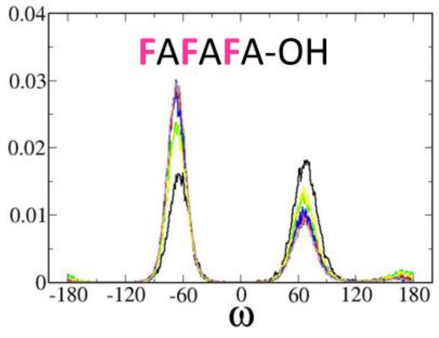


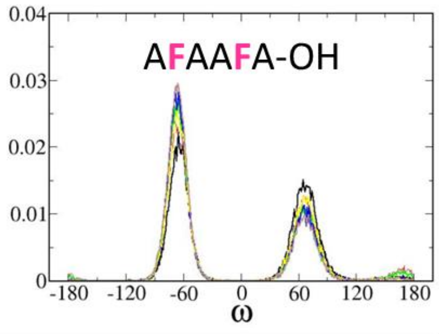
N

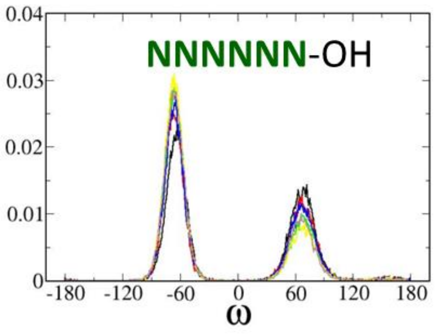
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Omega (ω) 1 | ⬛ | Omega (ω) 2 | 🟥 | Omega (ω) 3 | 🟩 | Omega (ω) 4 | 🟦 | Omega (ω) 5 | 🟨 | Omega (ω) 6 | 🟩 |

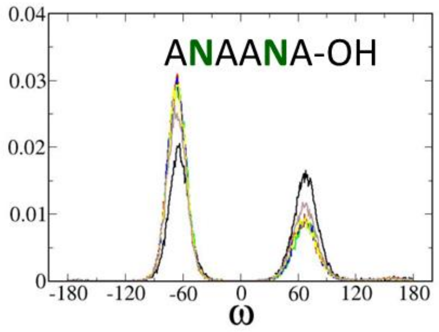
Non-reducing end

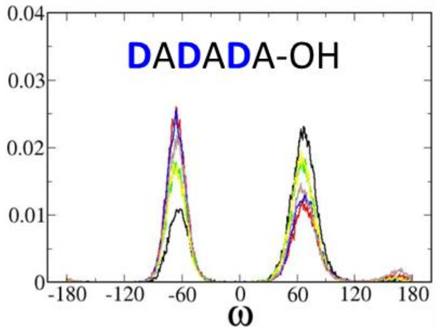
Reducing end

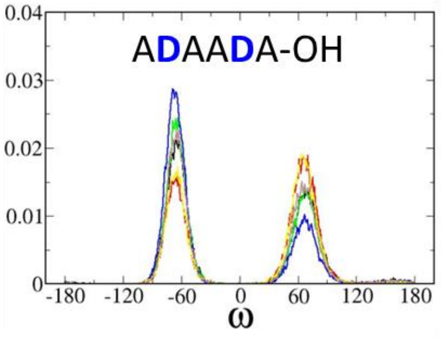


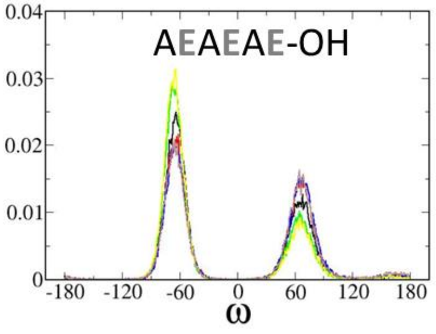


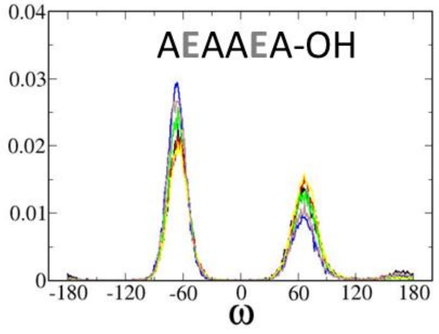






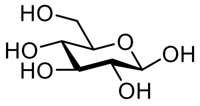




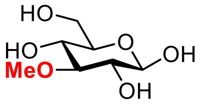


# SF7

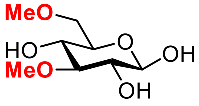
Phi (φ) torsion angles.



**A**



**B**

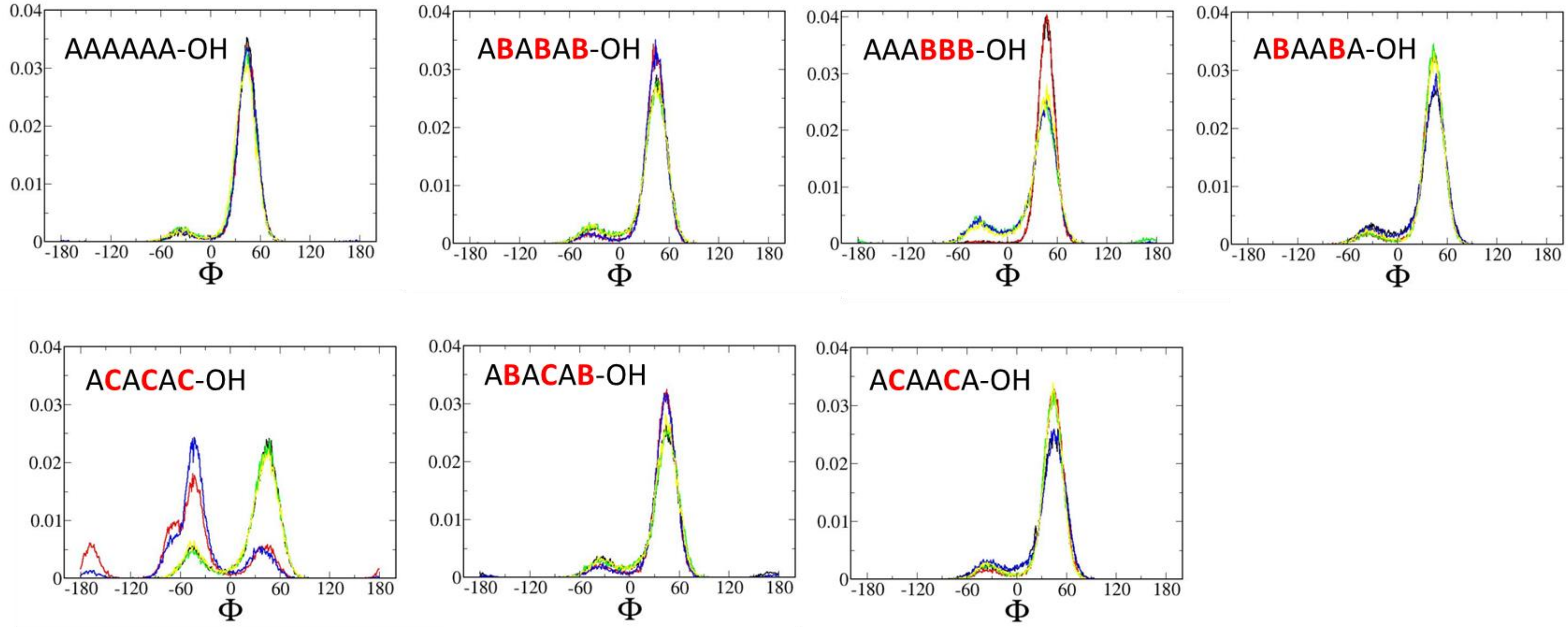


**C**

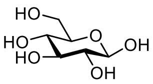
| Phi (φ) 1 | ⬛ | Phi (φ) 2 | 🟥 | Phi (φ) 3 | 🟩 | Phi (φ) 4 | 🟦 | Phi (φ) 5 | 🟨 |
|---|---|---|---|---|---|---|---|---|---|

Non-reducing end

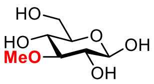
Reducing end



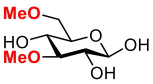
# Phi (φ) torsion angles.



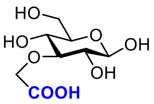
A



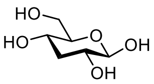
B



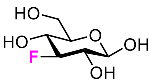
C



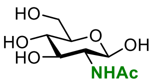
D



E



F

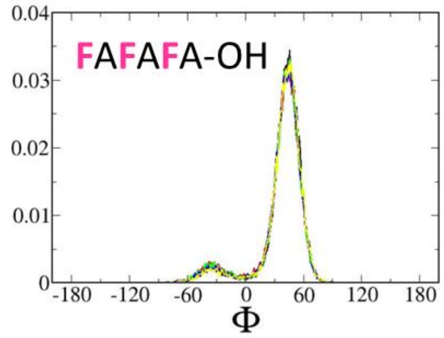


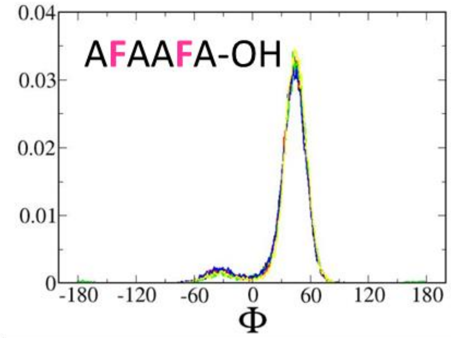
N

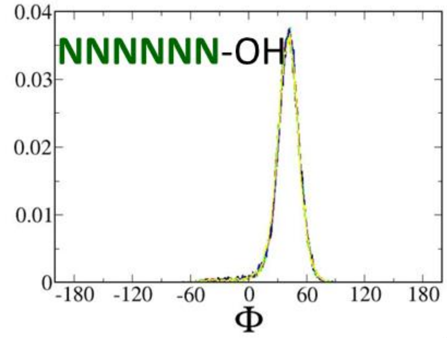
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Phi (φ) 1 | ⬛ | Phi (φ) 2 | 🟥 | Phi (φ) 3 | 🟩 | Phi (φ) 4 | 🟦 | Phi (φ) 5 | 🟨 |

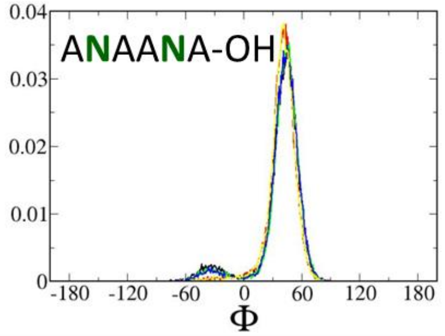
Non-reducing end

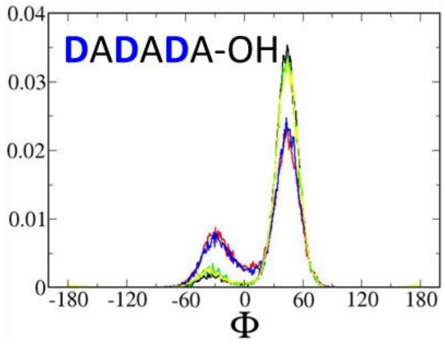
Reducing end

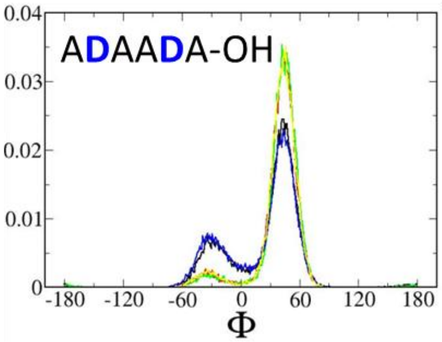


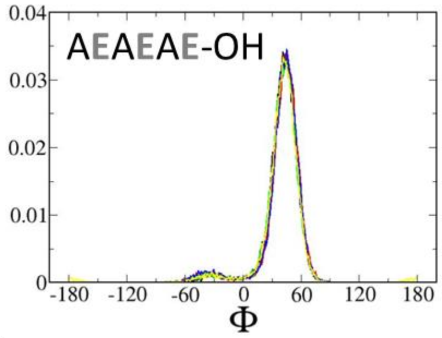


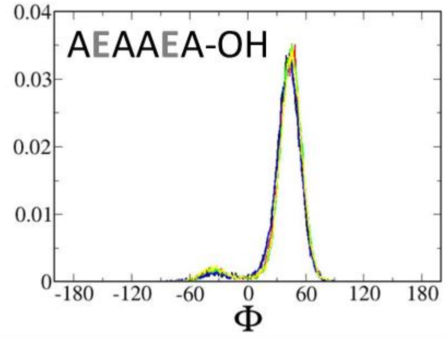






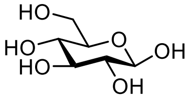




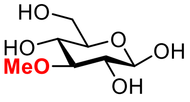


# SF8

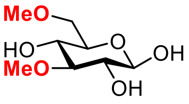
Psi (ψ) torsion angles.



**A**



**B**

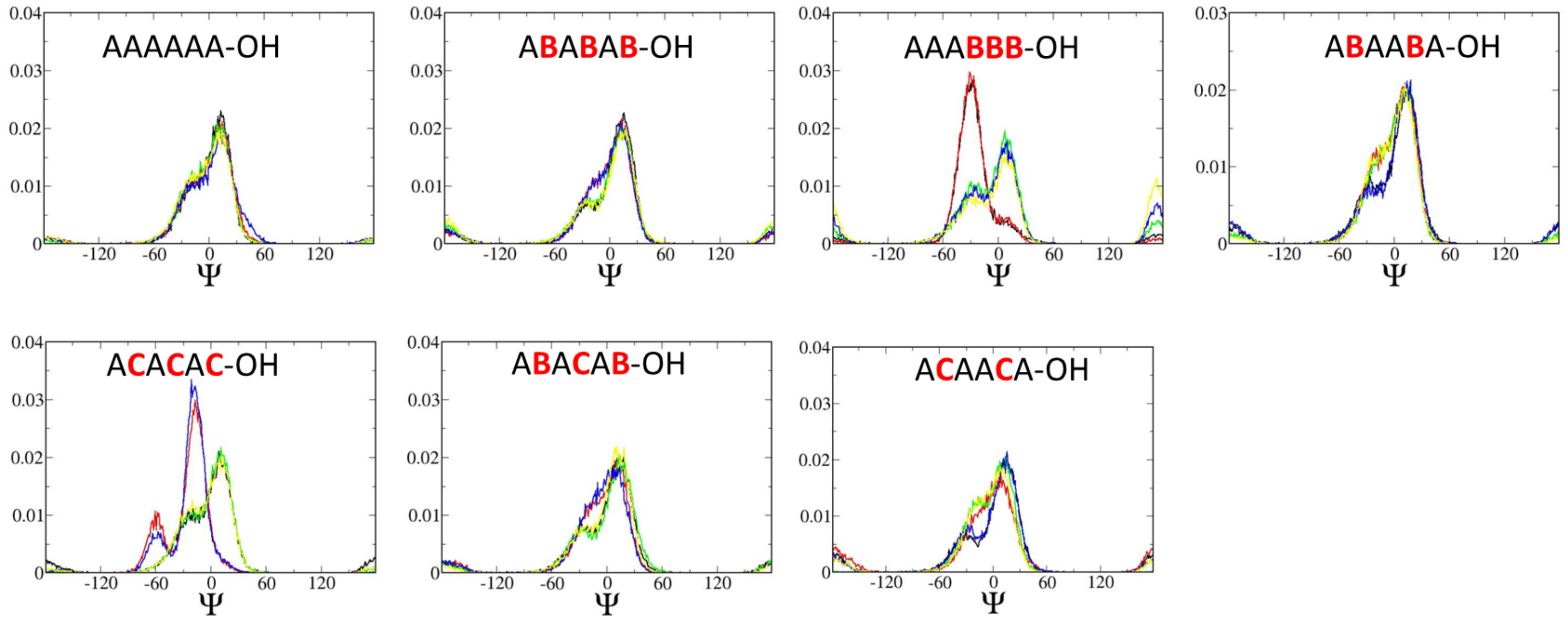


**C**

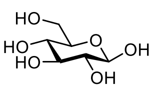
| Psi (ψ) 1 | ⬛ | Psi (ψ) 2 | 🟥 | Psi (ψ) 3 | 🟩 | Psi (ψ) 4 | 🟦 | Psi (ψ) 5 | 🟨 |
|---|---|---|---|---|---|---|---|---|---|

Non-reducing end

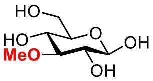
Reducing end



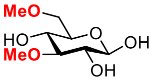
# Psi (ψ) torsion angles.



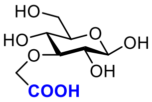
A



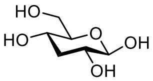
B



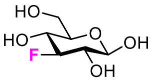
C



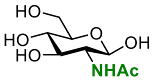
D



E



F

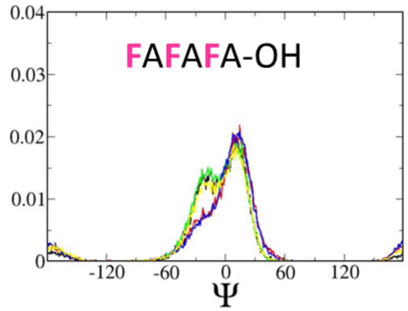


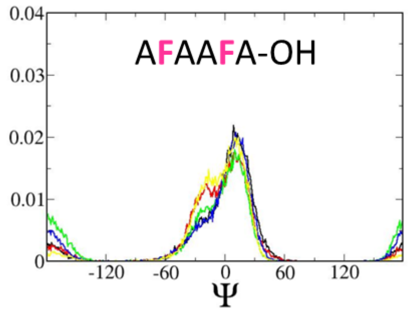
N

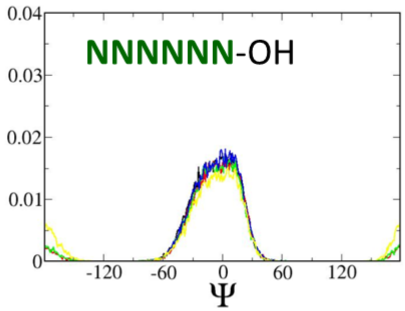
| Psi (ψ) 1 | ⬛ | Psi (ψ) 2 | 🟥 | Psi (ψ) 3 | 🟩 | Psi (ψ) 4 | 🟦 | Psi (ψ) 5 | 🟨 |
|---|---|---|---|---|---|---|---|---|---|

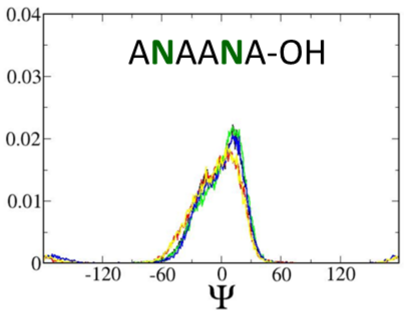
Non-reducing end

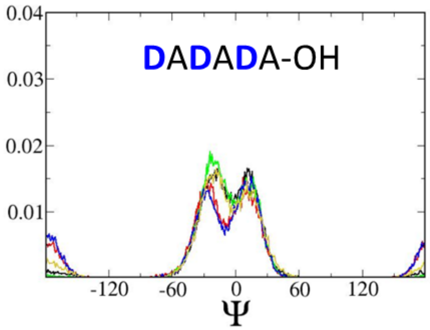
Reducing end

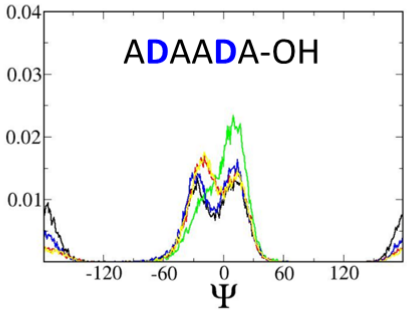


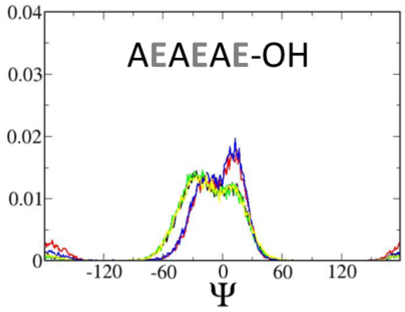


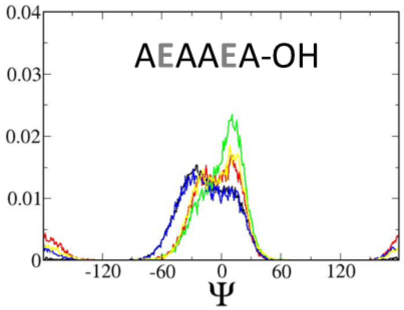






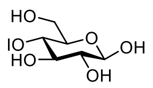




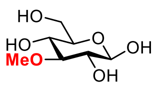


# SF9

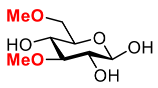
Overimposed puckering theta (**θ**) values



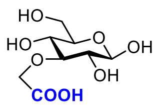
**A**



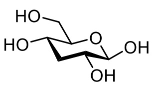
**B**



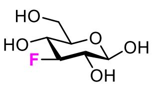
**C**



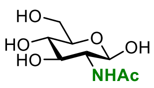
**D**



**E**



**F**

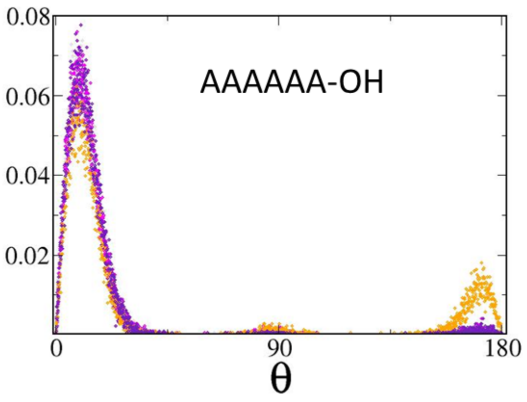


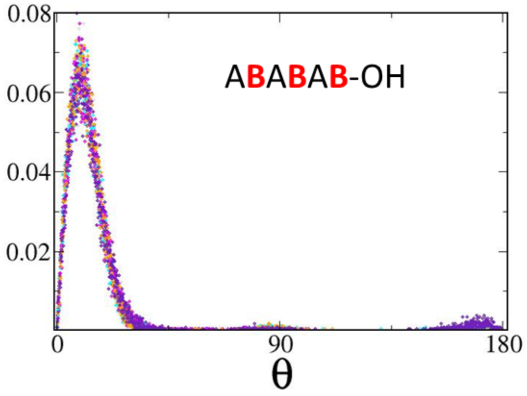
**N**

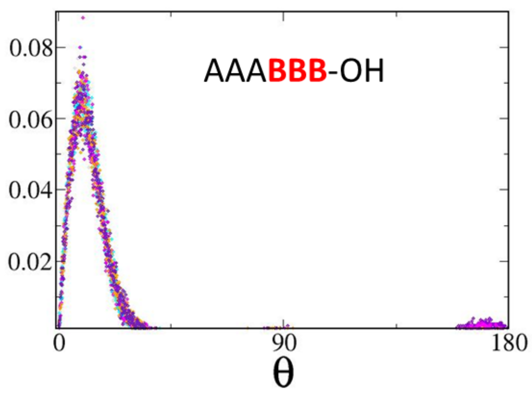
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Theta (**θ**) 1 | ■ | Theta (**θ**) 2 | ■ | Theta (**θ**) 3 | ■ | Theta (**θ**) 4 | ■ | Theta (**θ**) 5 | ■ | Theta (**θ**) 6 | ■ |

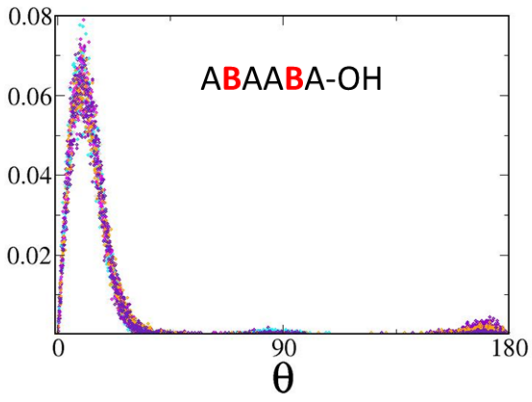
Non-reducing end

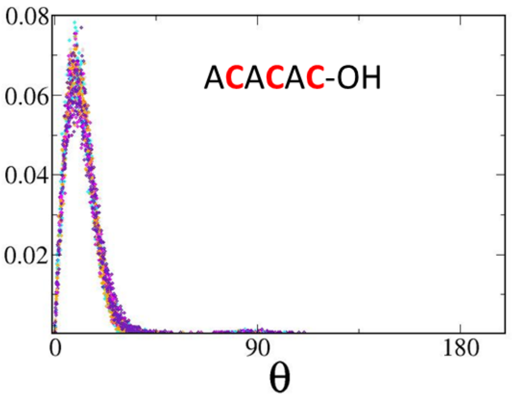
Reducing end

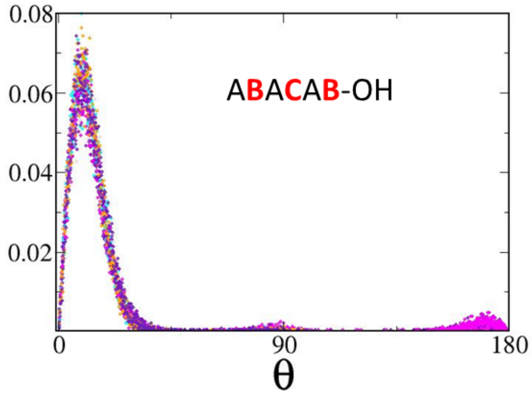




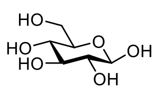




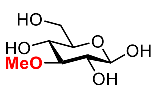




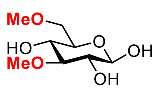
# Overimposed puckering theta (**θ**) values



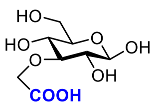
**A**



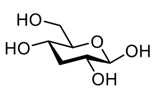
**B**



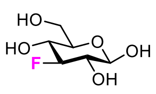
**C**



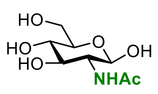
**D**



**E**



**F**

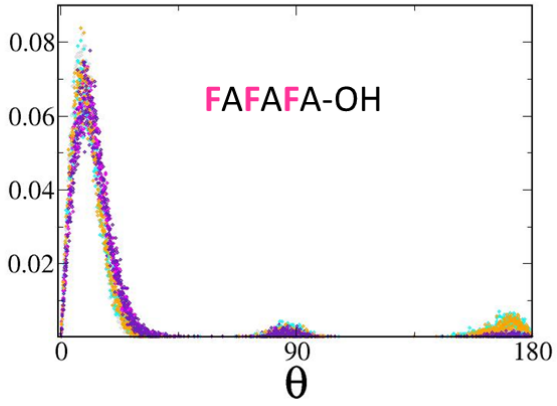


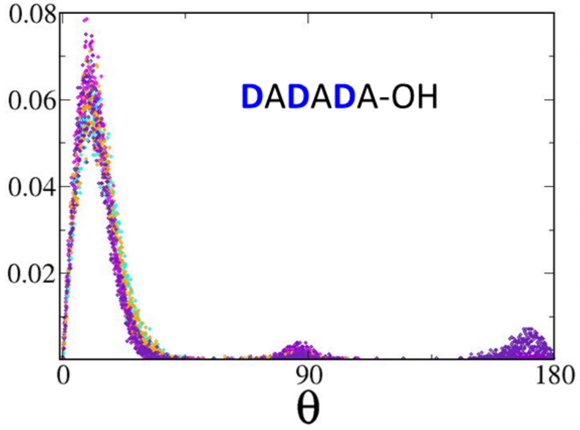
**N**

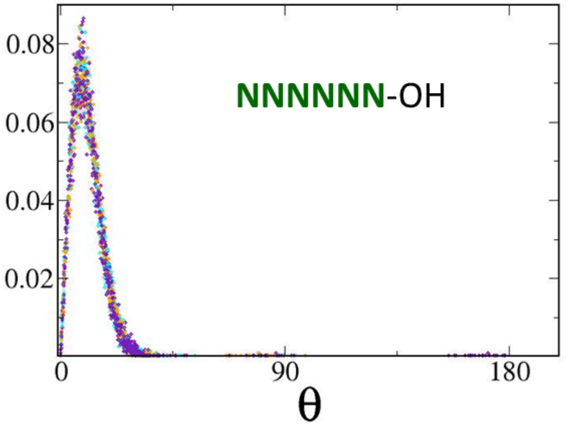
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Theta (**θ**) 1 | ■ | Theta (**θ**) 2 | ■ | Theta (**θ**) 3 | ■ | Theta (**θ**) 4 | ■ | Theta (**θ**) 5 | ■ | Theta (**θ**) 6 | ■ |

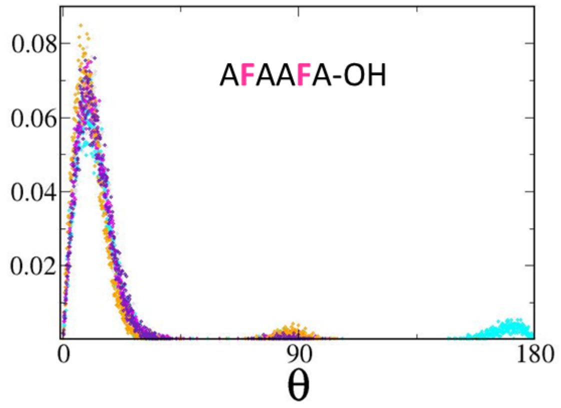
Non-reducing end

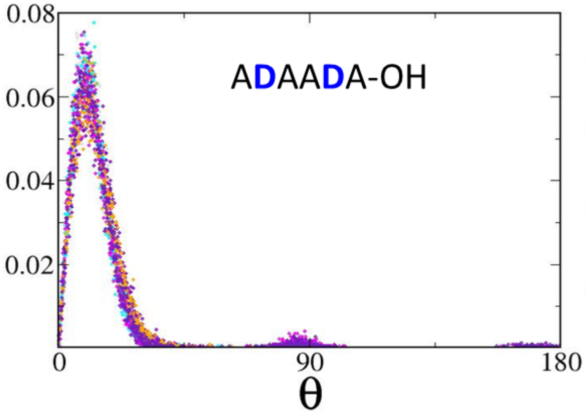
Reducing end

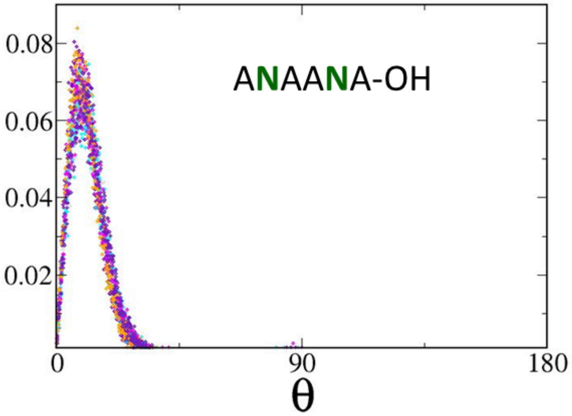








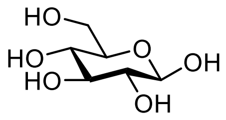




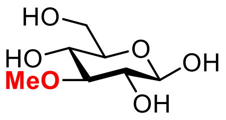
# SF10

Ramachandran plots

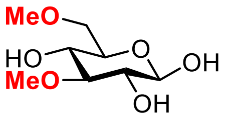
Overimposed phi and psi torsion angles



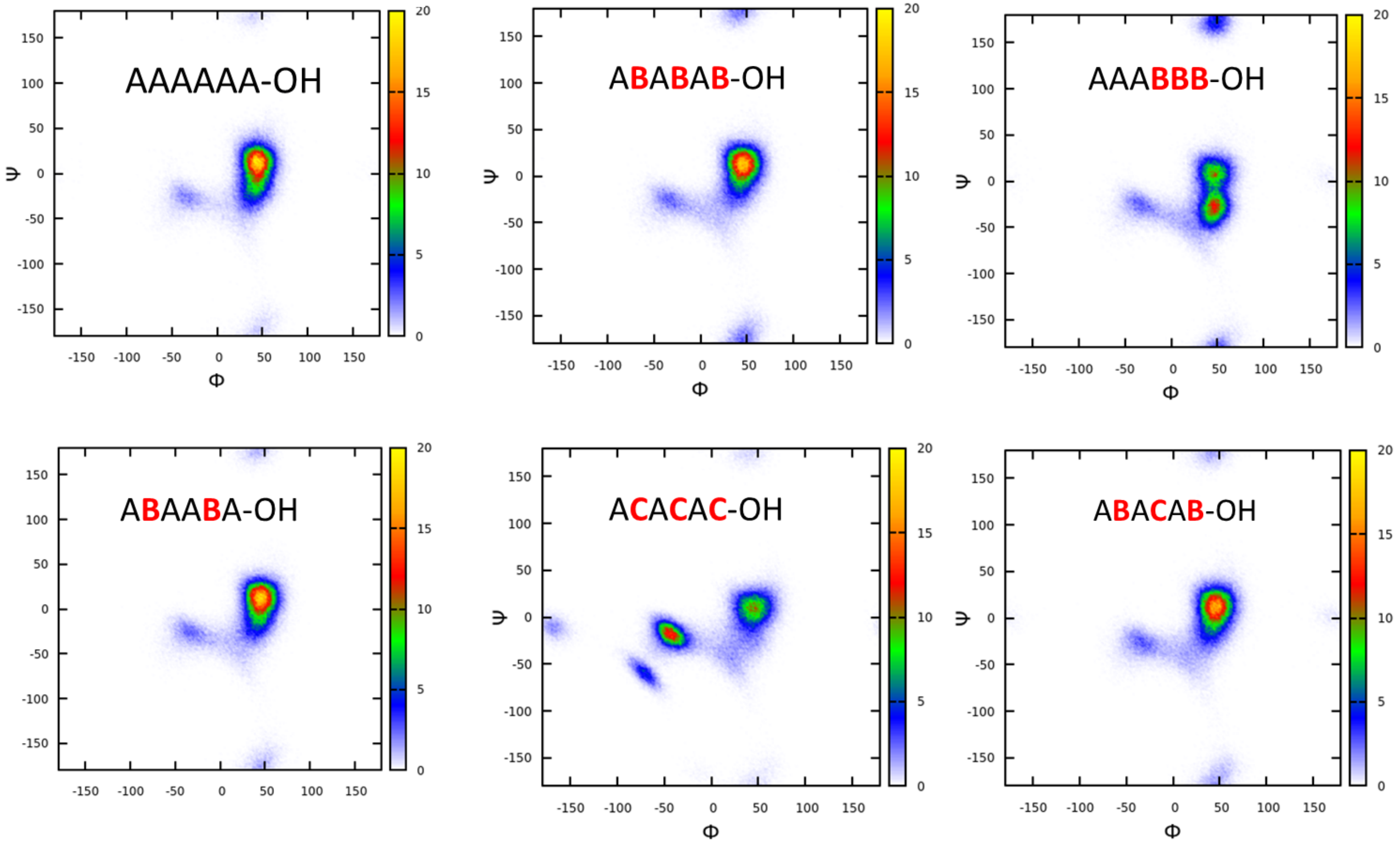
A



B

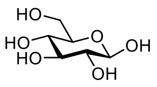


C

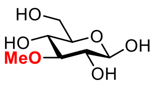


# Ramachandran plots

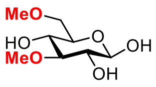
## Overimposed phi and psi torsion angles



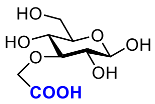
**A**



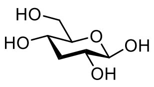
**B**



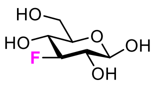
**C**



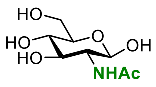
**D**



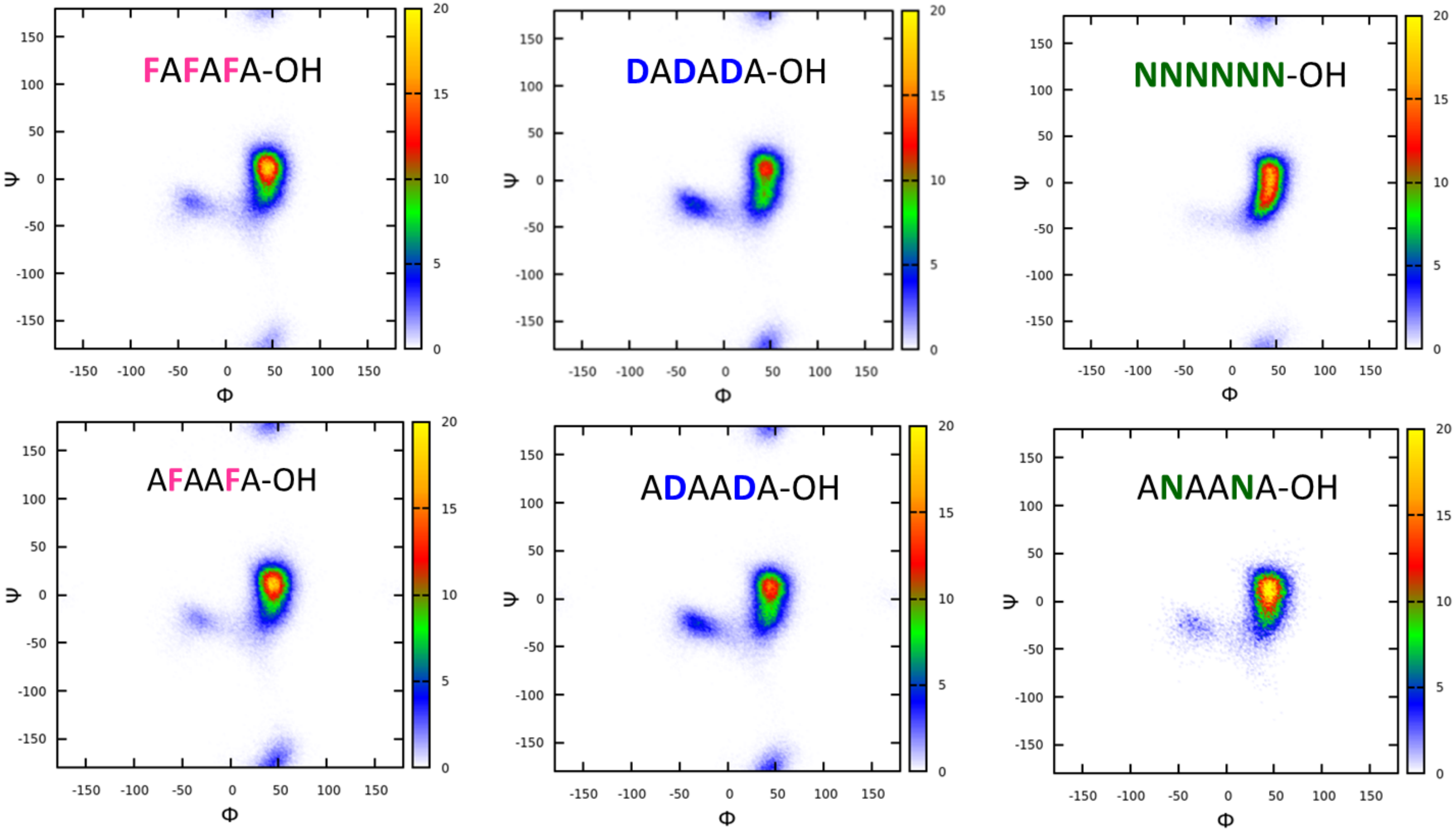
**E**



**F**

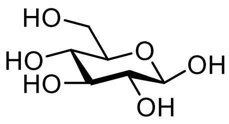


**N**

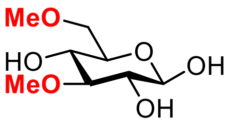


## Ramachandran plots

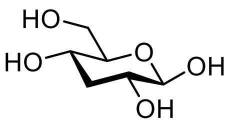
Overimposed phi and psi torsion angles



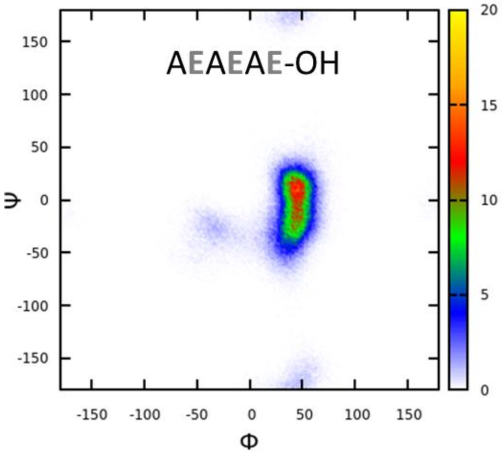
**A**

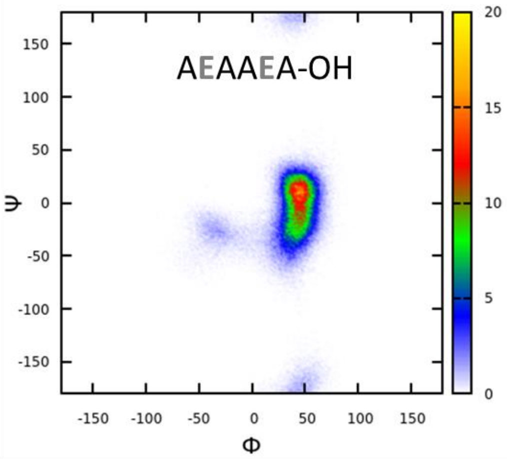


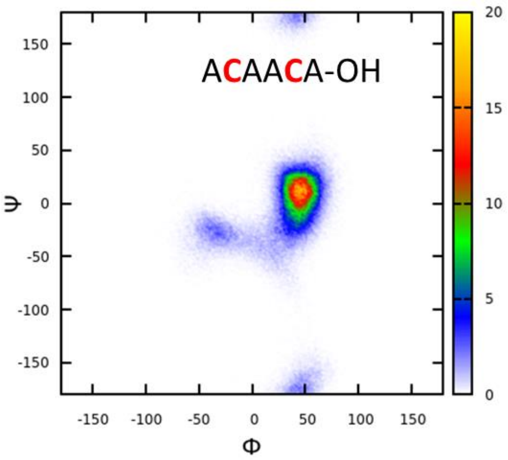
**C**



**E**

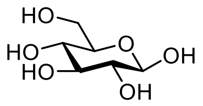




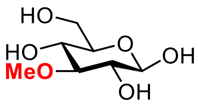


# Ramachandran plots

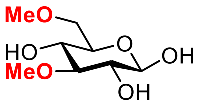
Individual phi and psi torsion angles



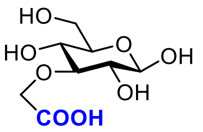
**A**



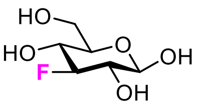
**B**



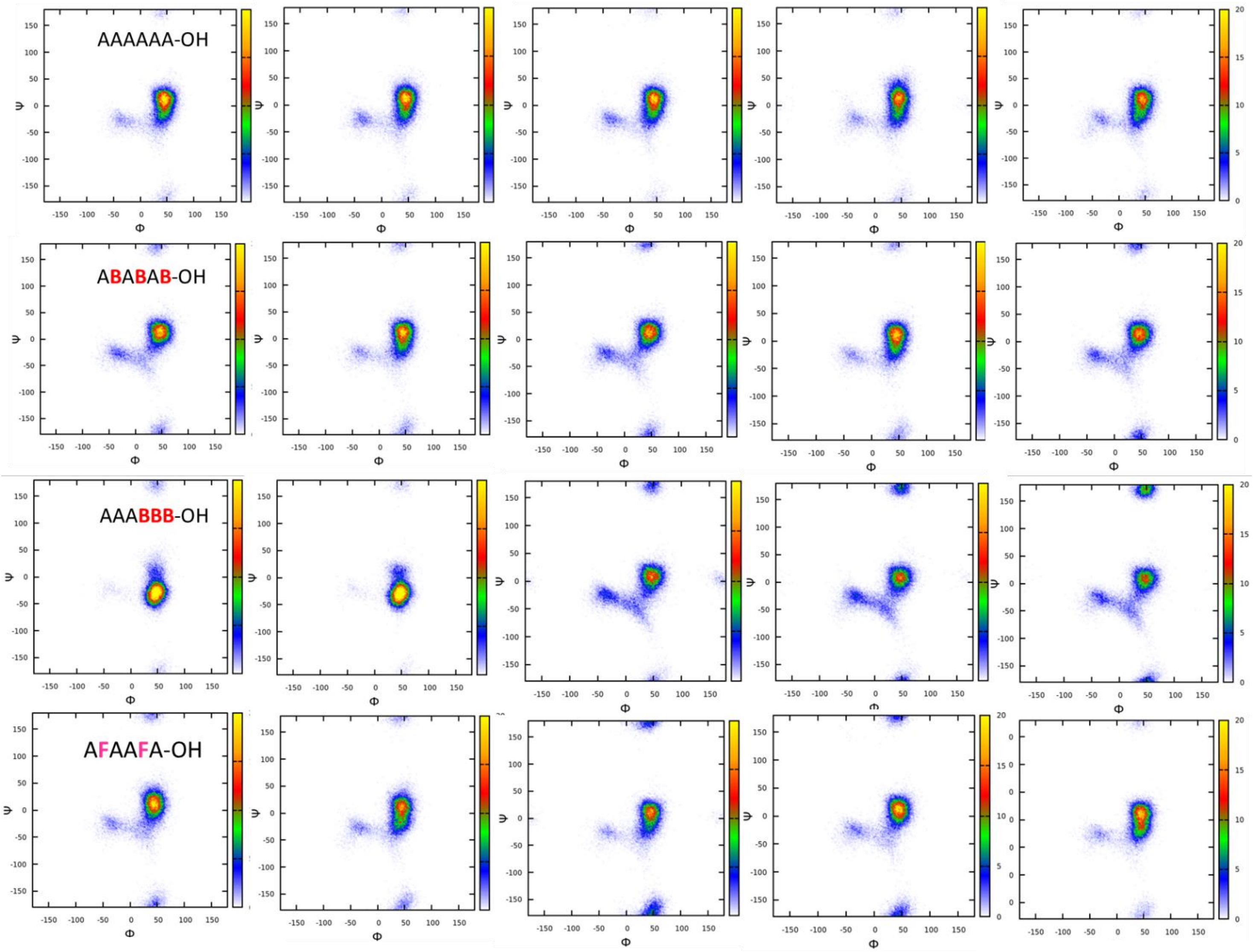
**C**



**D**

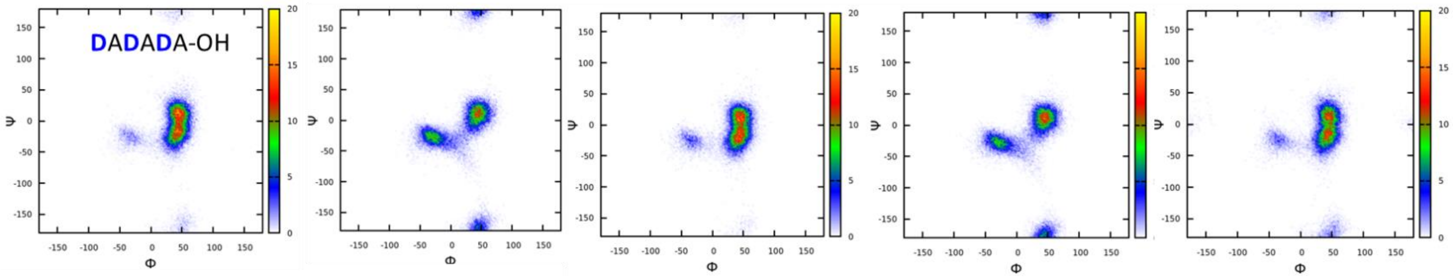


**F**



Non-reducing end

Reducing end



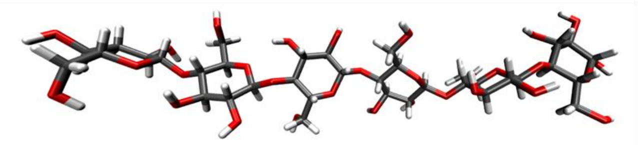
Non-reducing end

Reducing end

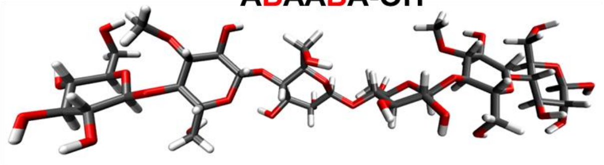
**SF11**

# Selected representative structures as obtained from MD simulations

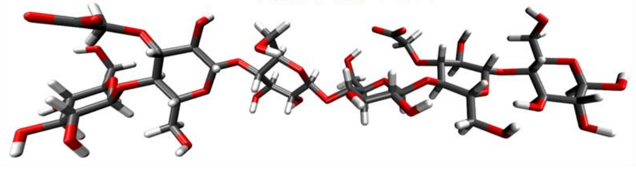
**AAAAAA-OH**



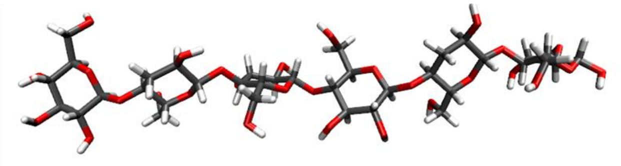
**ABAABA-OH**



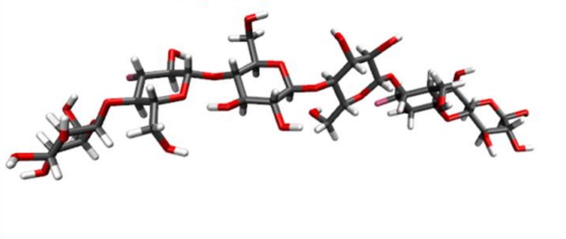
A**D**AA**D**A-OH



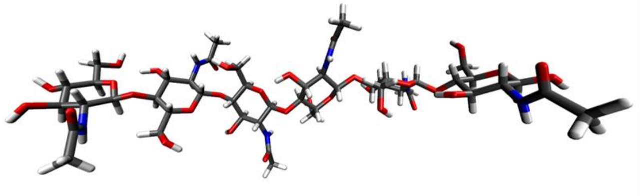
AEAAEA-OH



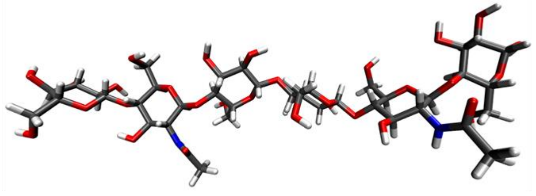
**AFAAFA-OH**



**NNNNNN-OH**

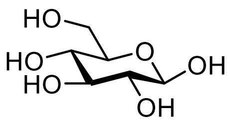


A**N**AA**N**A-OH

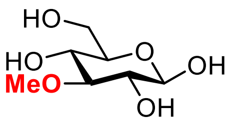


**SF12**

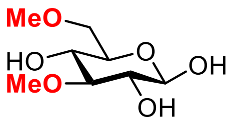
End-to-end distance as a function of time



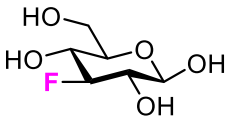
A



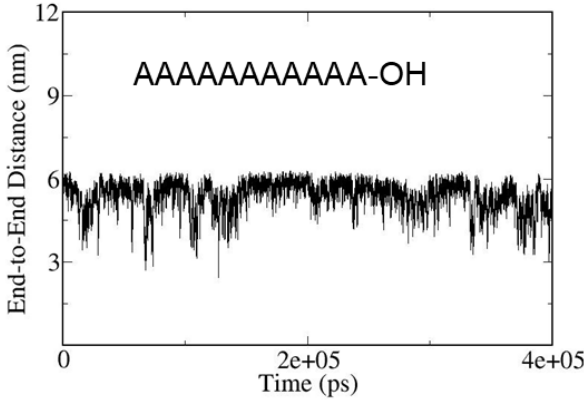
B

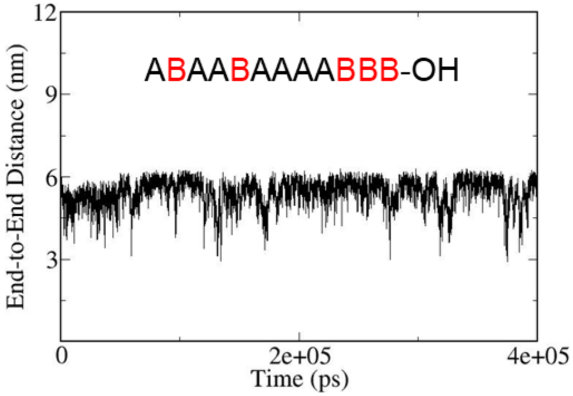


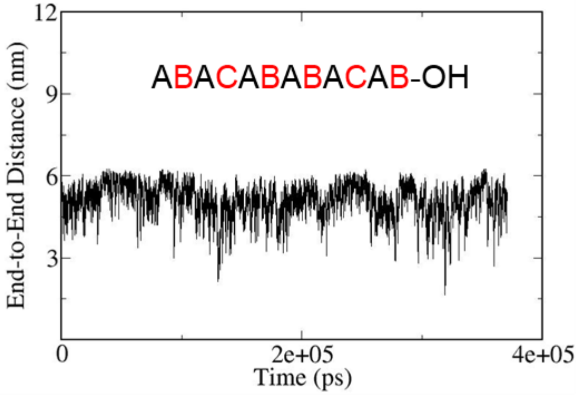
C

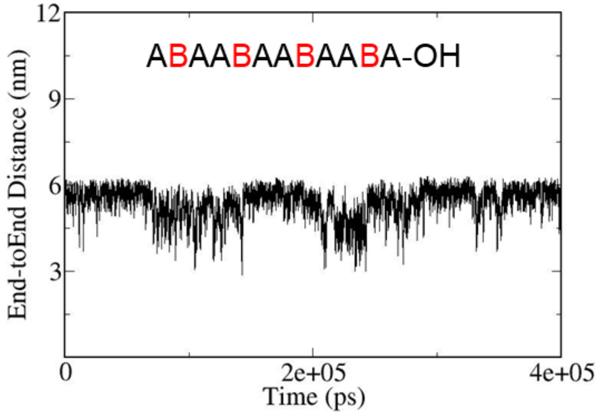


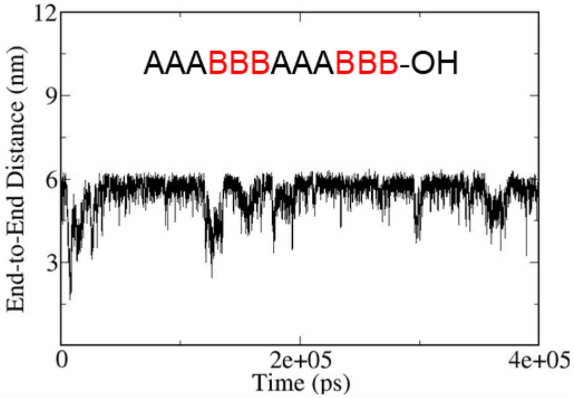
F

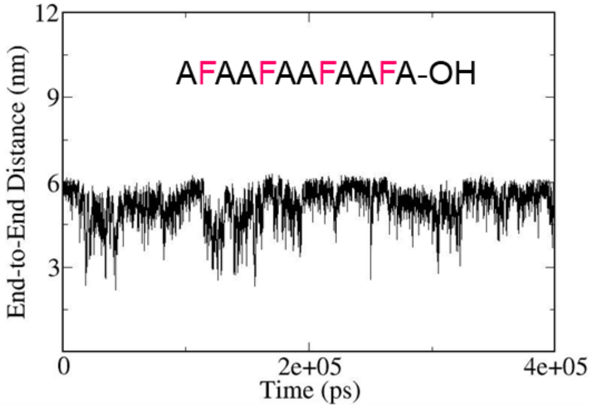








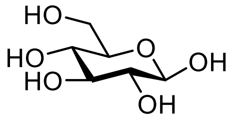




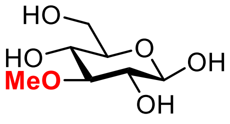
# SF13

Ramachandran plots

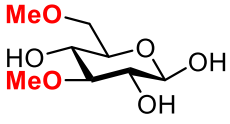
Overimposed phi and psi torsion angles



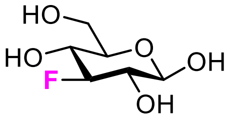
**A**



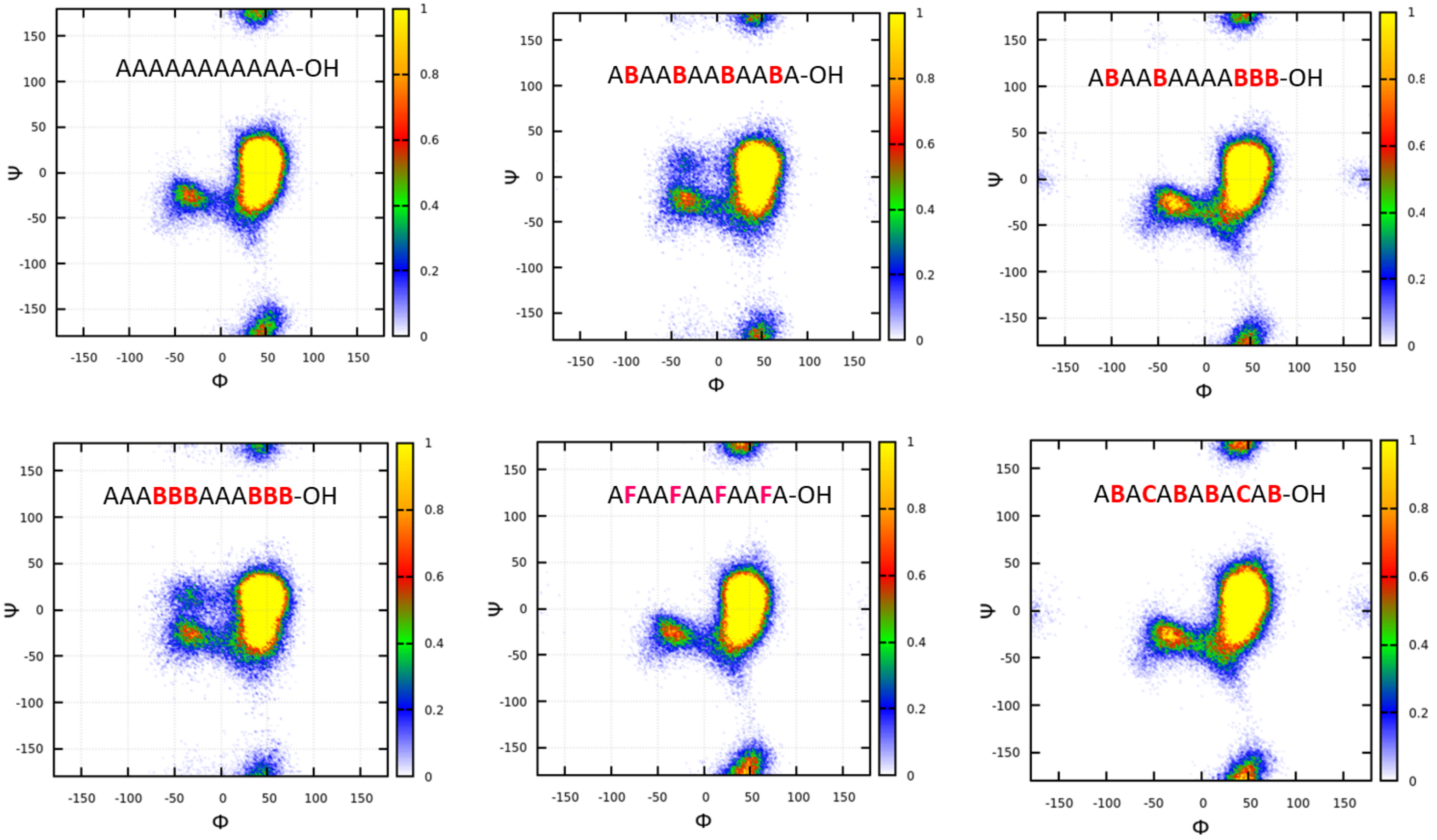
**B**



**C**

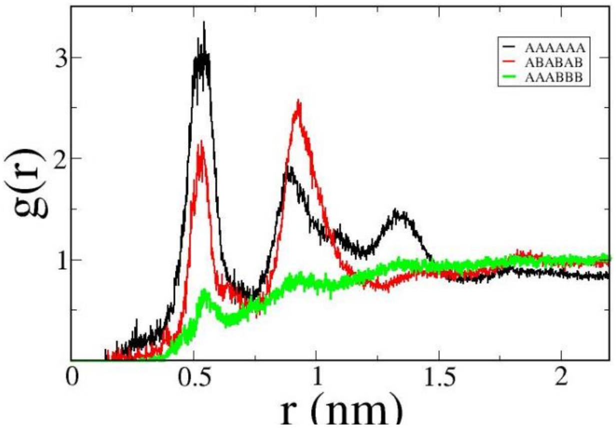


**F**

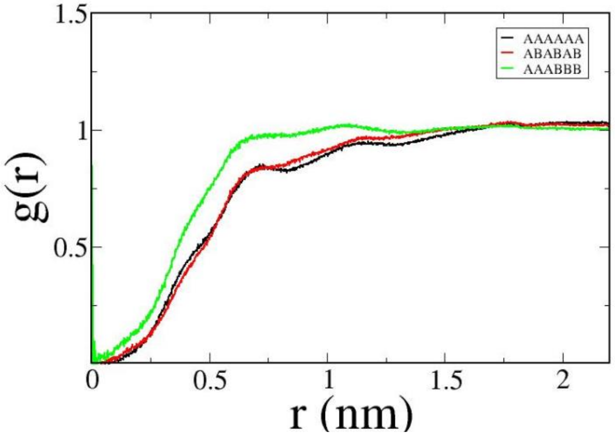


# SF14

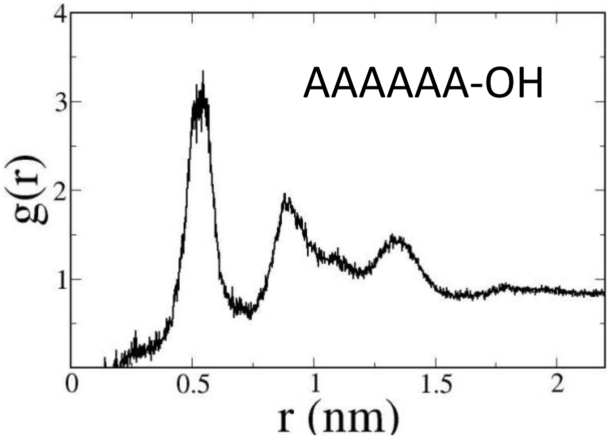
## Radial Distribution Functions

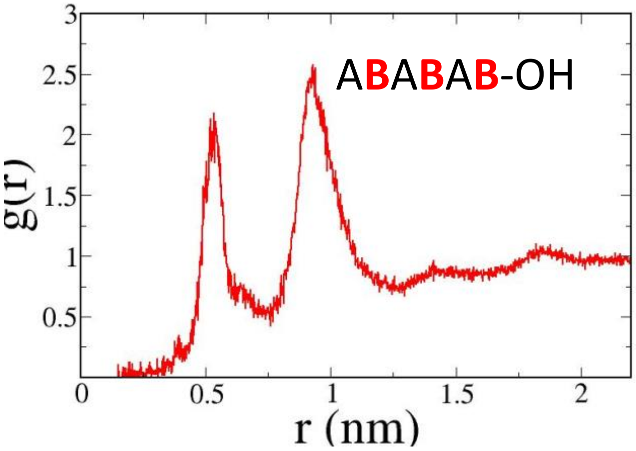


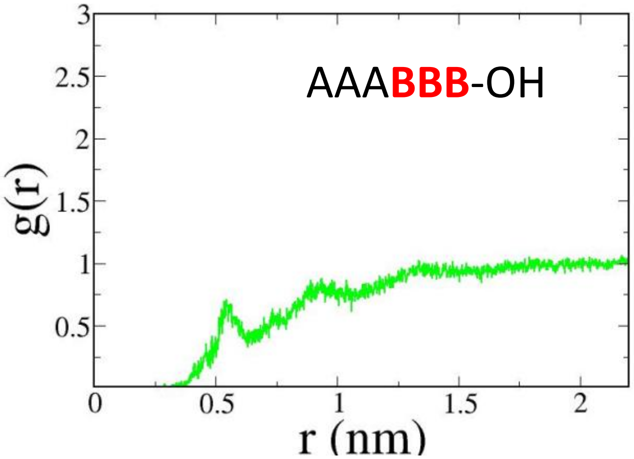
molecule - molecule interactions



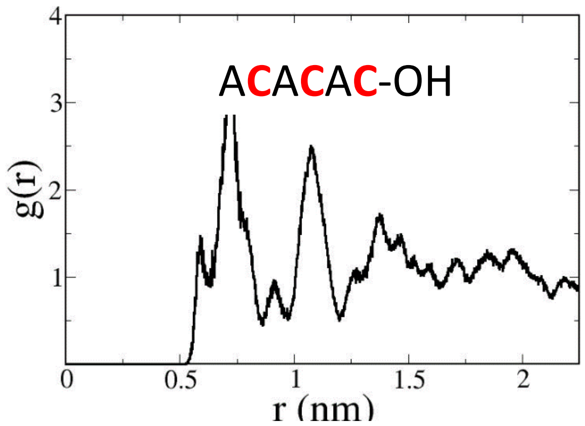
molecule - water interactions

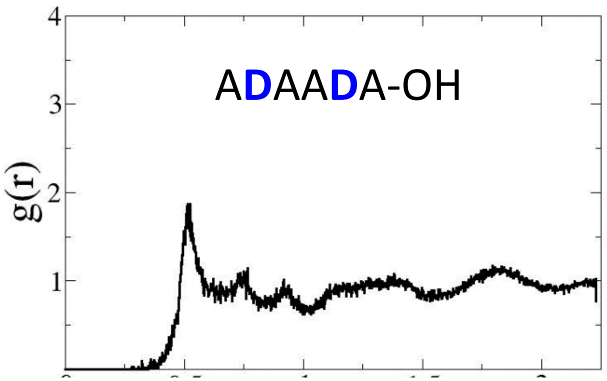


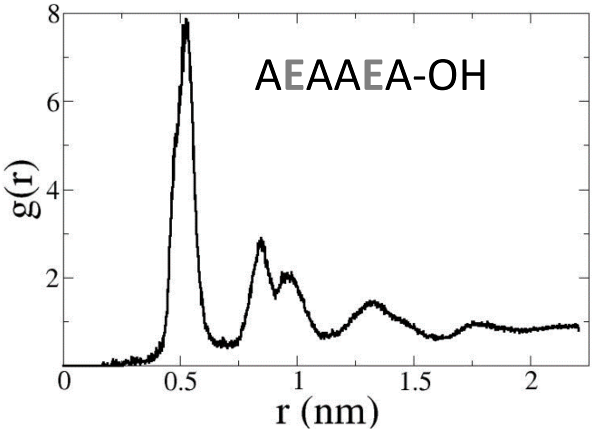


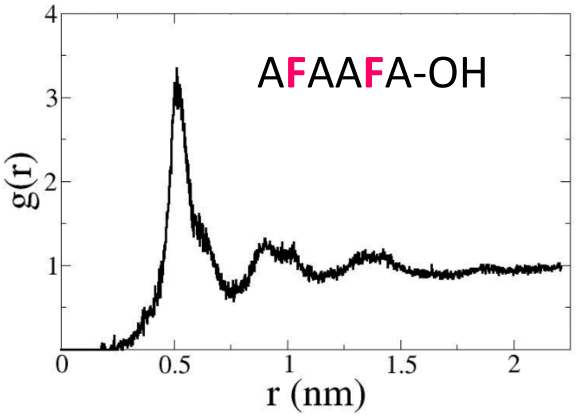


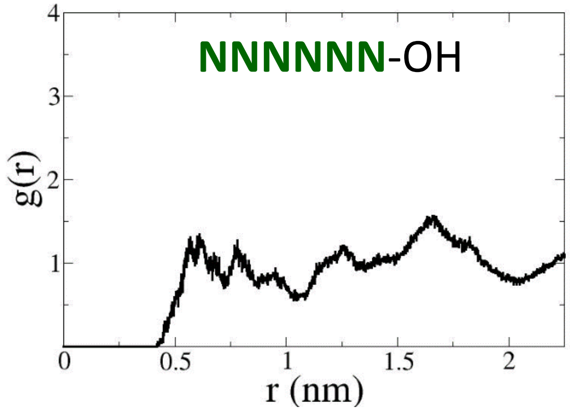
# Radial Distribution Functions





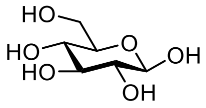




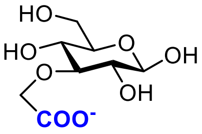


# SF15

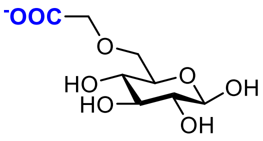
End-to-end distance as a function of time



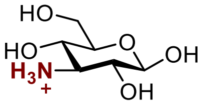
**A**



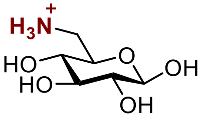
**D**



**d**

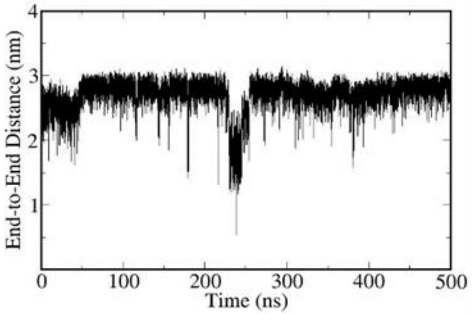


**G**

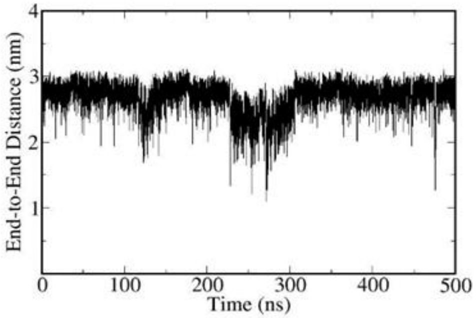


**g**

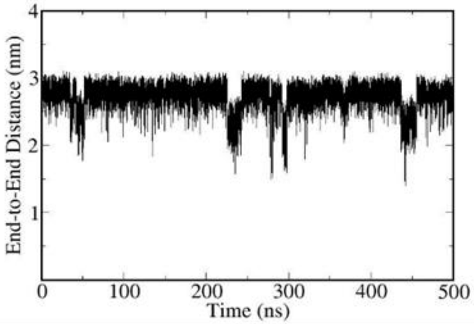
**AGAAGA-OH**



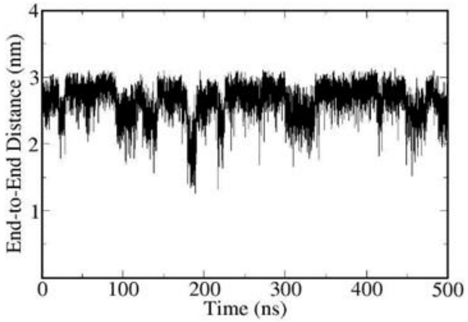
**ADAADA-OH**



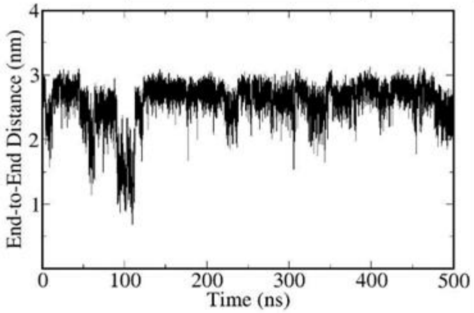
**AGAADA-OH**



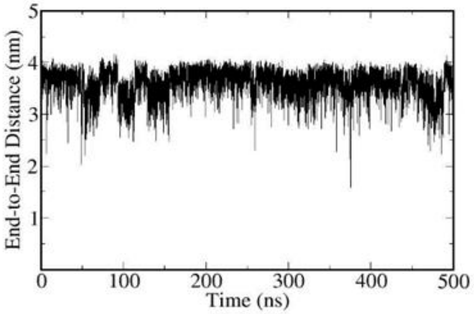
**AgAAgA-OH**



**AdAAdA-OH**

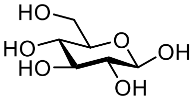


**GADAGADA -OH**

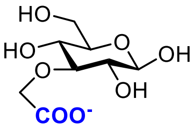


# SF16

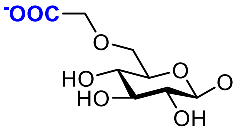
Histograms of radius of gyration (**RoG**)



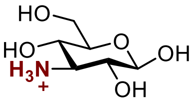
**A**



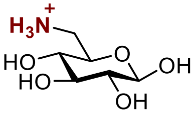
**D**



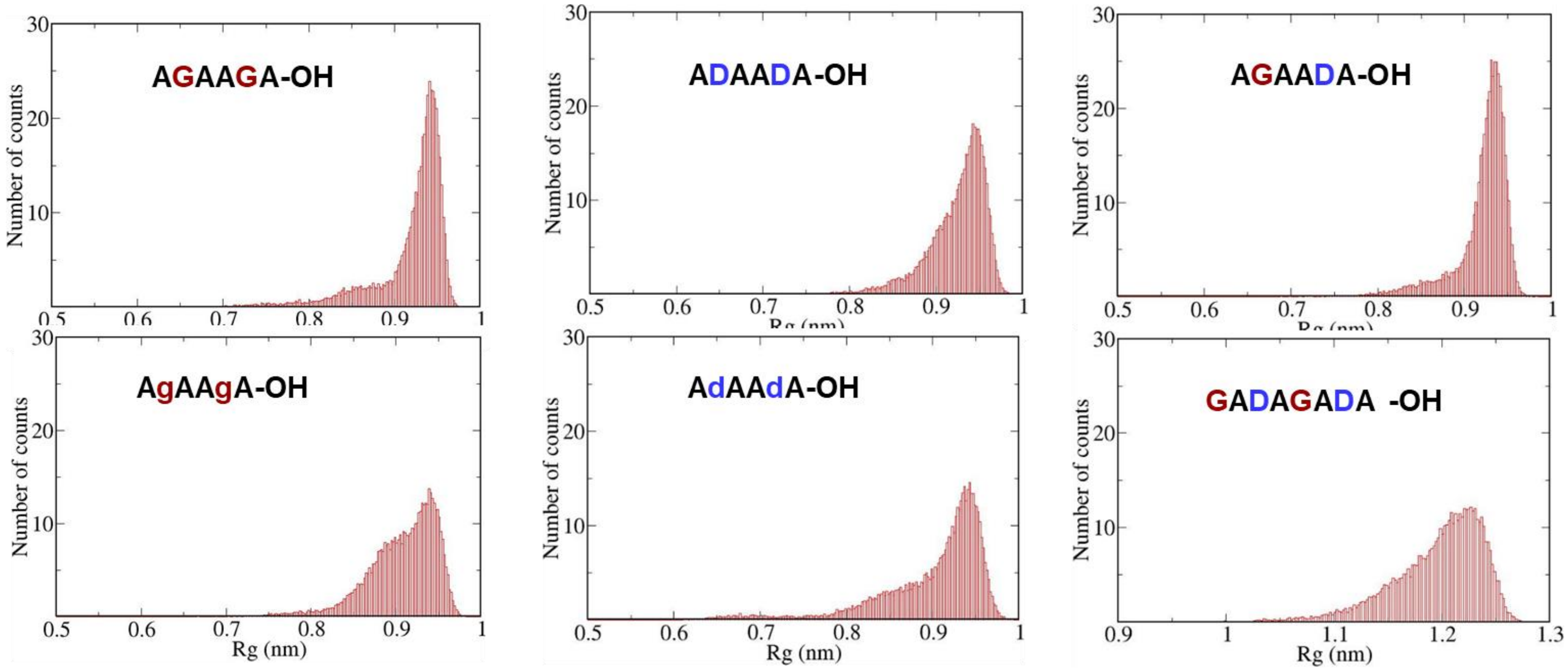
**d**



**G**

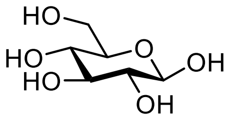


**g**

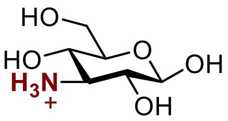


# SF17

Ramachandran plots

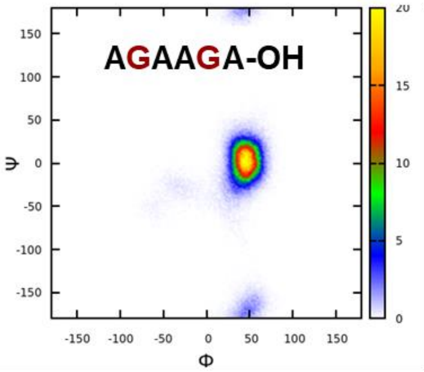


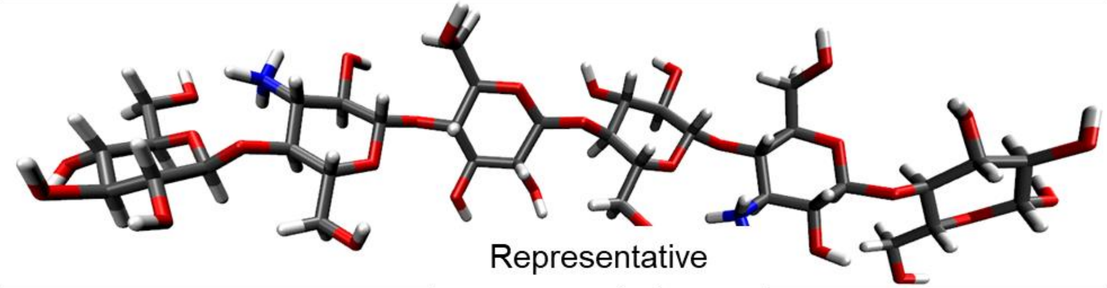
**A**



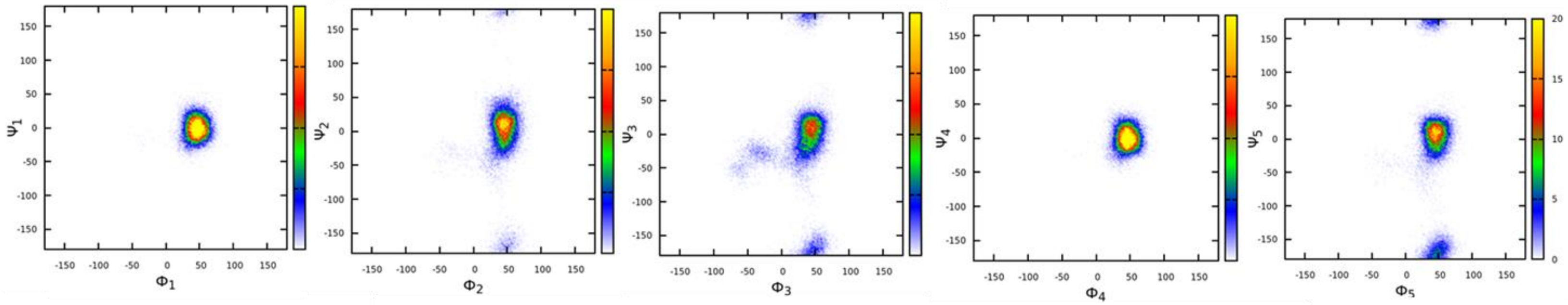
**G**

Overimposed phi and psi torsion angles





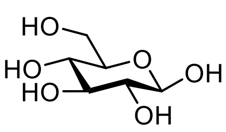
Representative snapshot



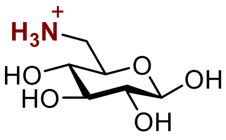
Non-reducing end

Reducing end

# Ramachandran plots



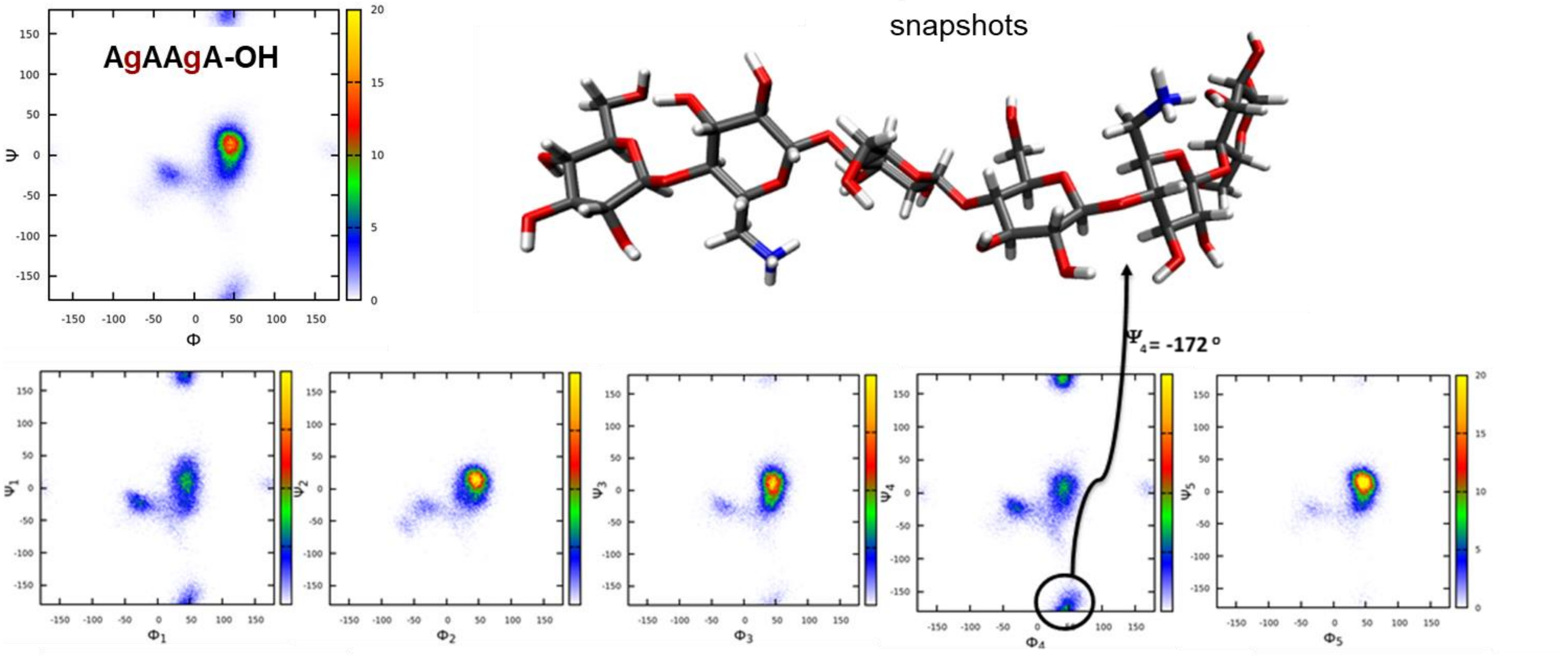
**A**



**g**

Overimposed phi and psi torsion angles

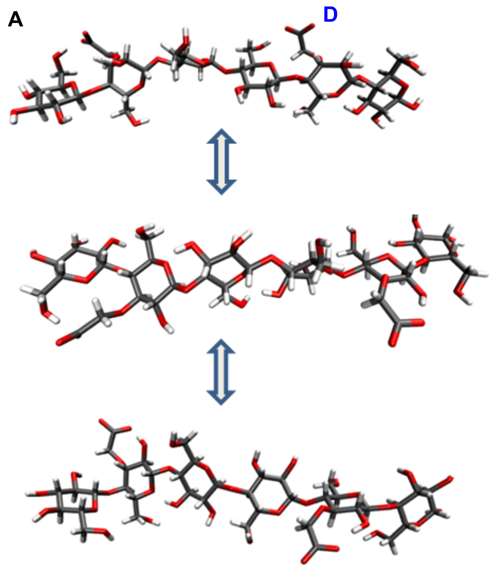
Representative snapshots



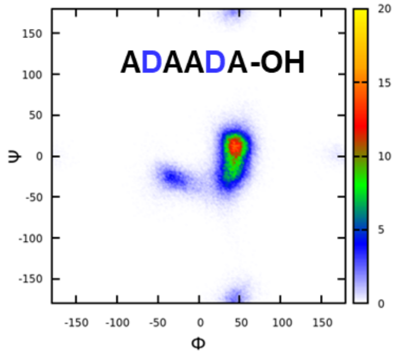
Non-reducing end

Reducing end

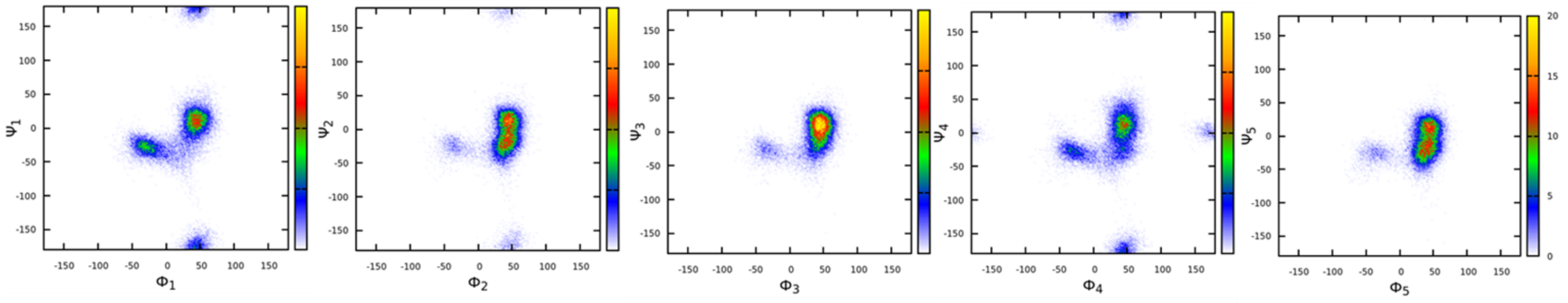
# Ramachandran plots



Overimposed phi and psi torsion angles



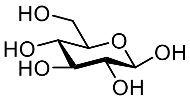
Representative snapshots



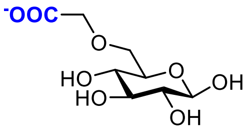
Non-reducing end

Reducing end

# Ramachandran plots

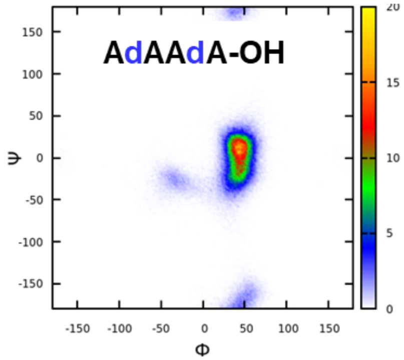


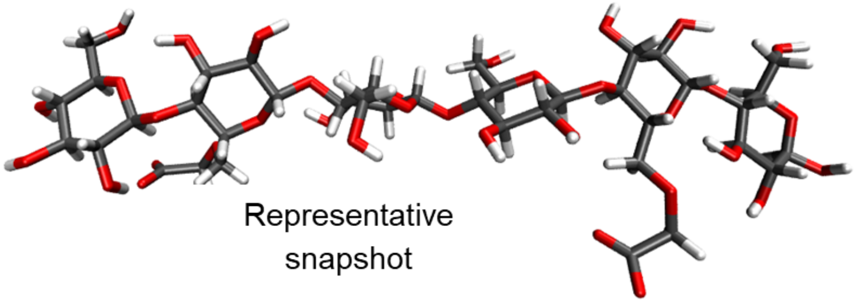
**A**



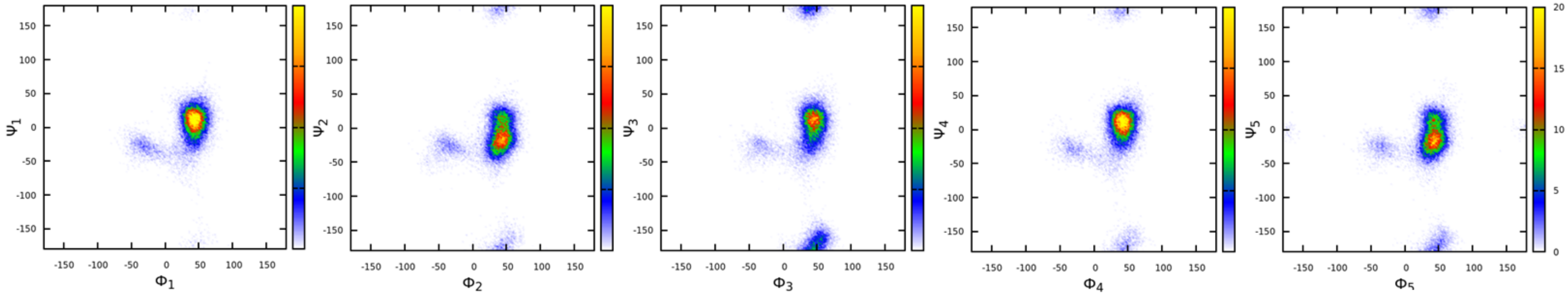
**d**

Overimposed phi and psi torsion angles





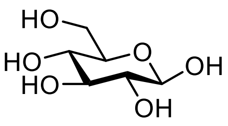
Representative snapshot



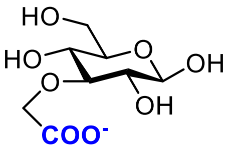
Non-reducing end

Reducing end

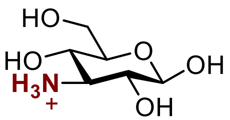
# Ramachandran plots



**A**

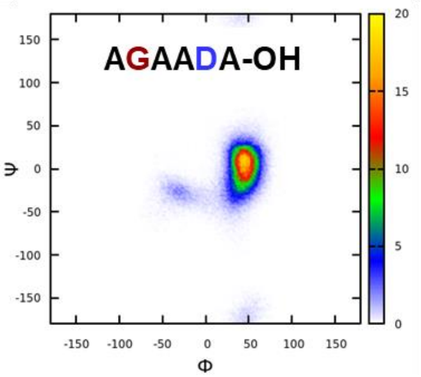


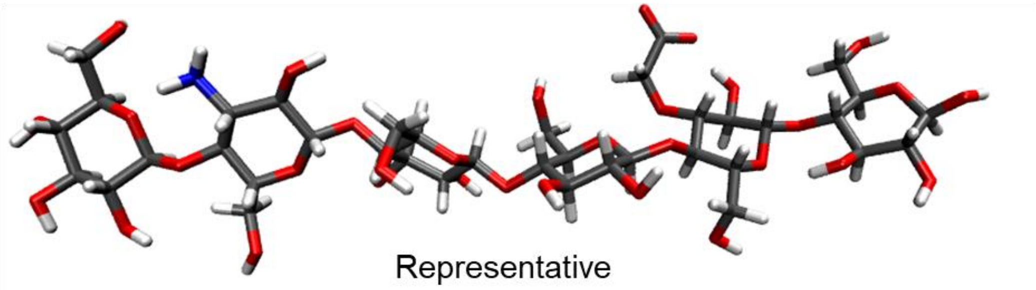
**D**



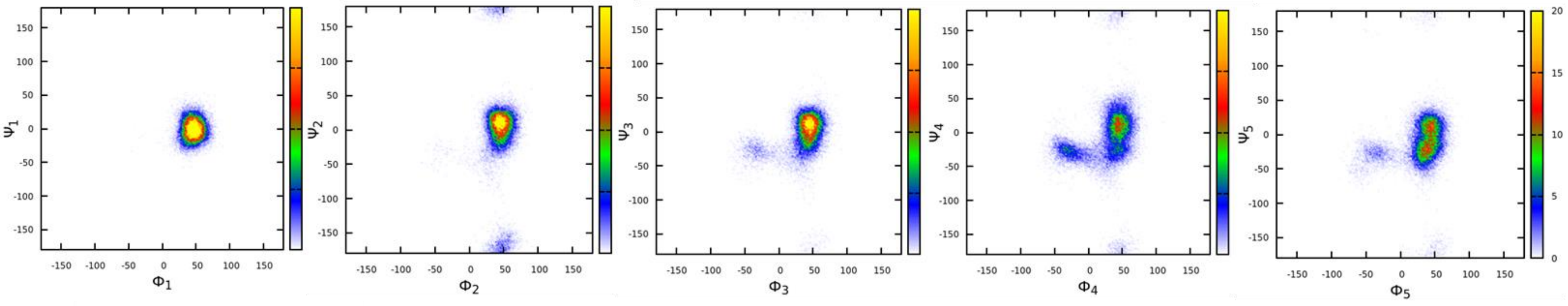
**G**

Overimposed phi and psi torsion angles



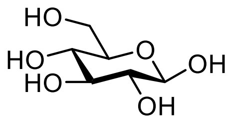


Representative snapshot

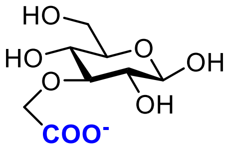


Non-reducing end

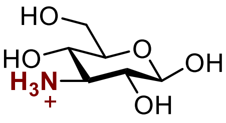
Reducing end



**A**



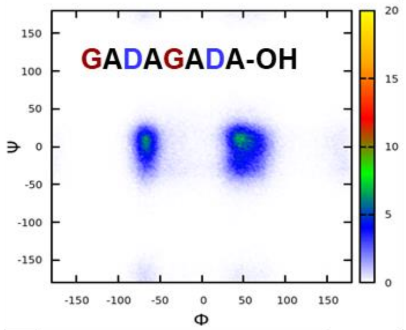
**D**



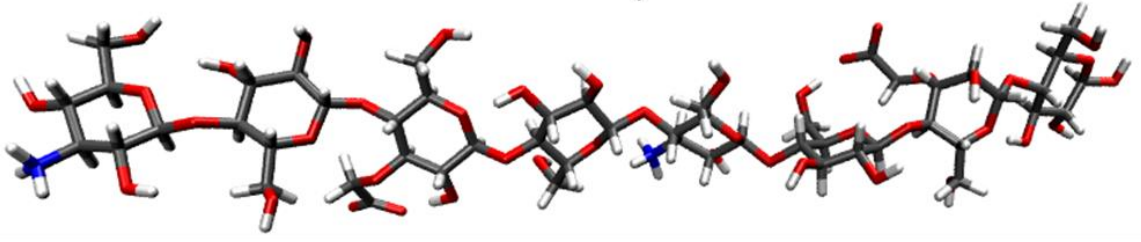
**G**

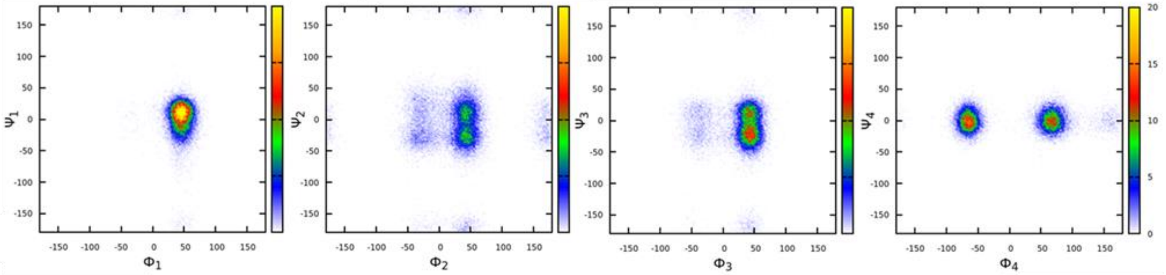
## Ramachandran plots

Overimposed phi and psi torsion angles

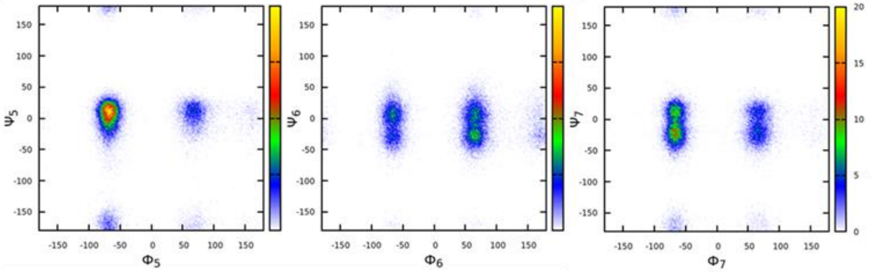


Representative snapshot





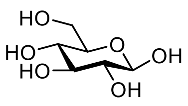
Non-reducing end



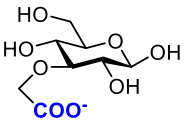
Reducing end

# SF18

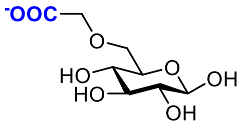
Superimposed Omega (**ω**) torsion angles.



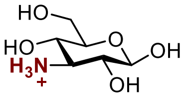
**A**



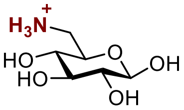
**D**



**d**



**G**

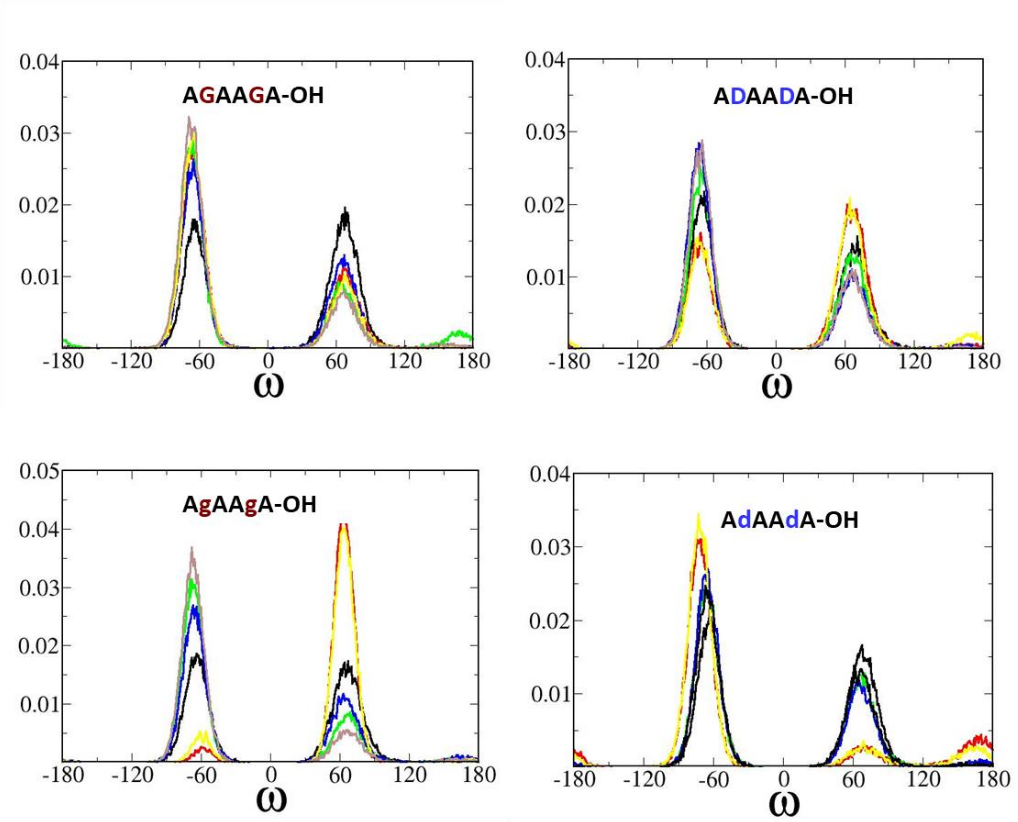


**g**

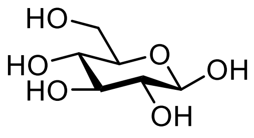
| Omega (**ω**) 1 | ⬛ | Omega (**ω**) 2 | 🟥 | Omega (**ω**) 3 | 🟩 | Omega (**ω**) 4 | 🟦 | Omega (**ω**) 5 | 🟨 | Omega (**ω**) 6 | 🟫 |
|---|---|---|---|---|---|---|---|---|---|---|---|

Non-reducing end

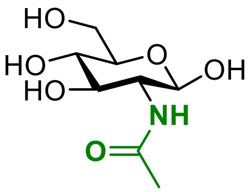
Reducing-end



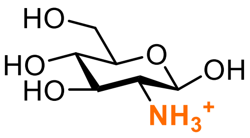
# SF19 End-to-end distance as a function of time



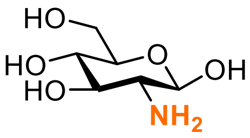
**A**



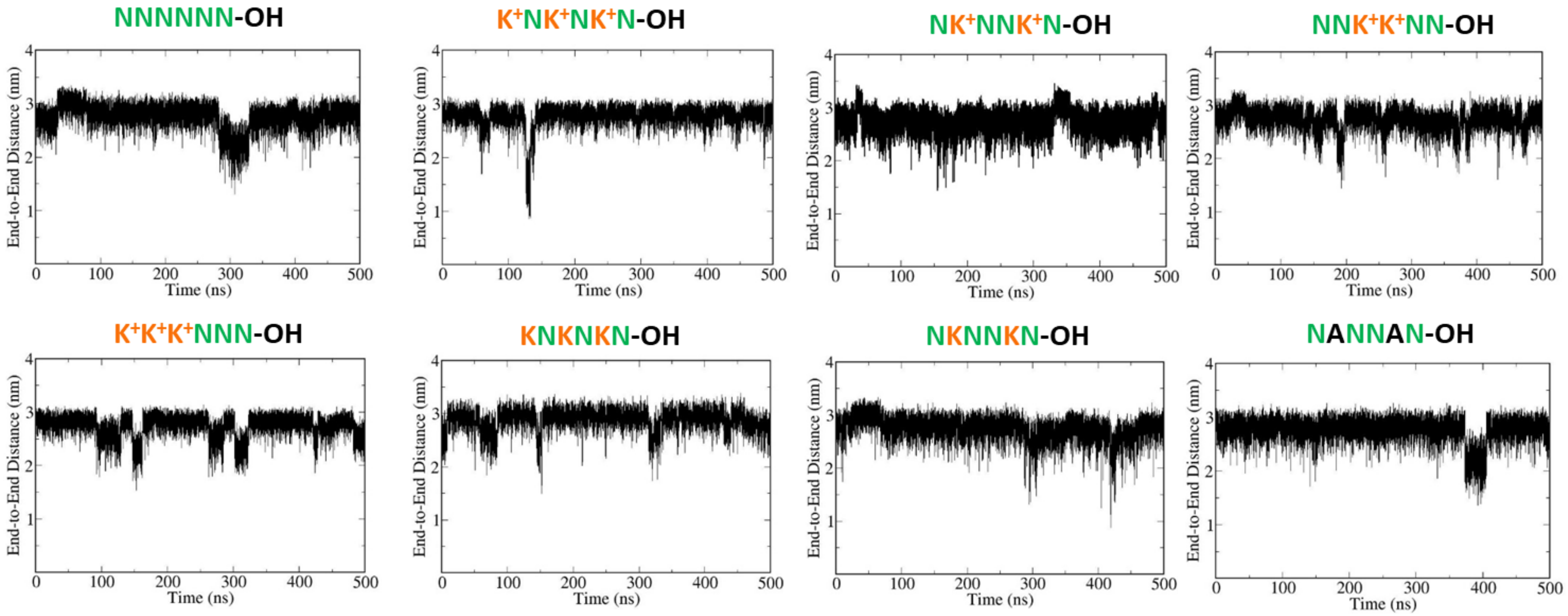
**N**



**$K^+$**



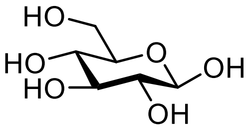
**K**



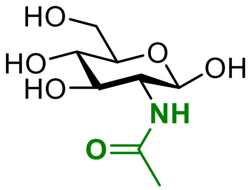
| Hexamer | NNNNNN-OH | $K^+NK^+NK^+N$-OH | KNKNKN-OH | $NNK^+K^+NN$-OH | NNKKNN-OH | $K^+K^+K^+NNN$-OH | $NK^+NNK^+N$-OH | NKNNKN-OH | NANNAN-OH |
|---|---|---|---|---|---|---|---|---|---|
| Average Distance (nm) | 2.76 | 2.76 | 2.88 | 2.73 | 2.73 | 2.73 | 2.73 | 2.73 | 2.75 |
| Standard Deviation | 0.22 | 0.22 | 0.20 | 0.21 | 0.21 | 0.21 | 0.21 | 0.23 | 0.23 |

# SF20

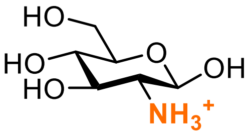
Histograms of radius of gyration (**RoG**)



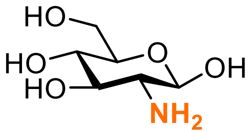
**A**



**N**

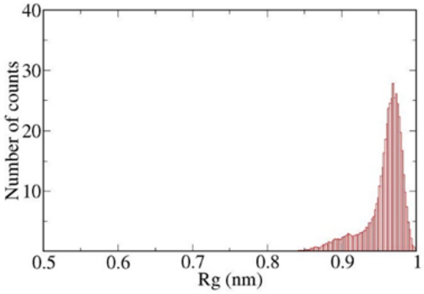


**K$^+$**

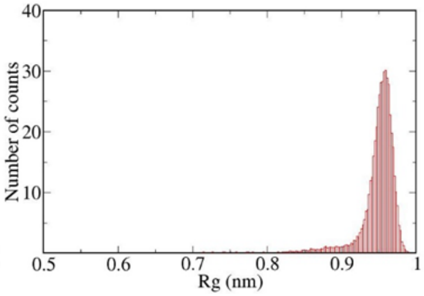


**K**

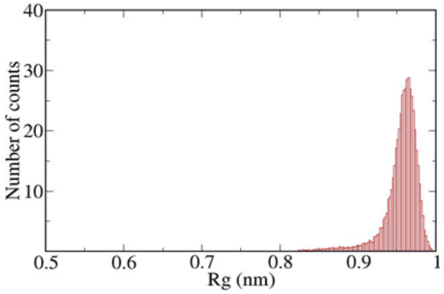
**NNNNNN-OH**



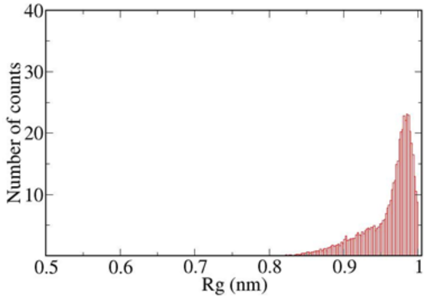
**K$^+$NK$^+$NK$^+$N-OH**



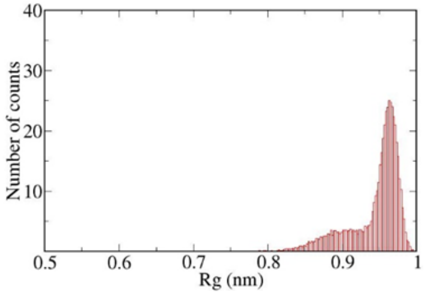
**NK$^+$NNK$^+$N-OH**



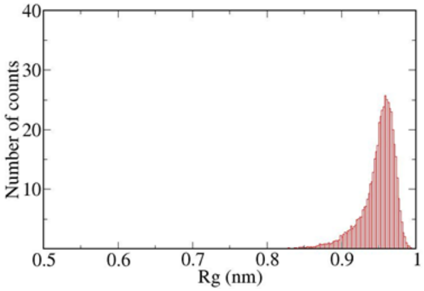
**NNK$^+$K$^+$NN-OH**



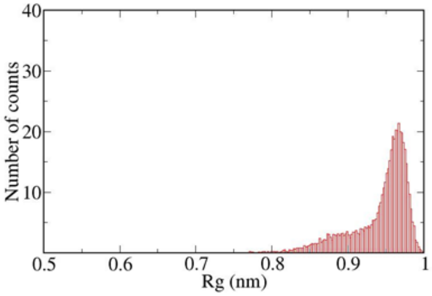
**K$^+$K$^+$K$^+$NNN-OH**



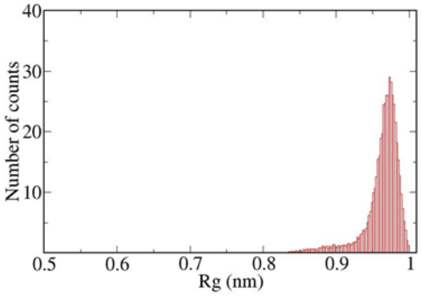
**KNKNKN-OH**



**NKNNKN-OH**

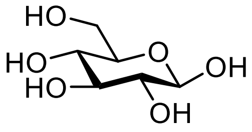


**NANNAN-OH**

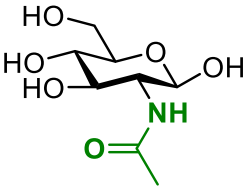


# SF21

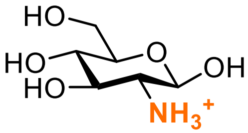
Dihedral analysis of
**N6**



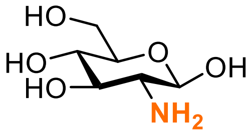
**A**



**N**

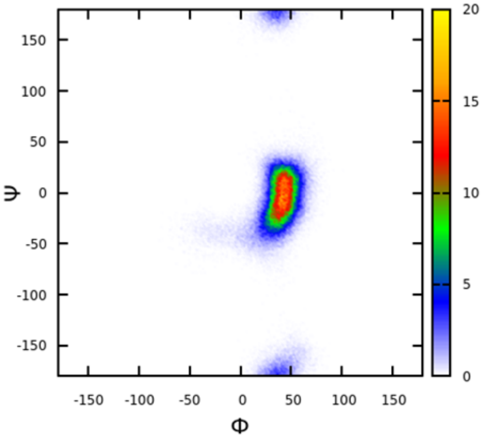


**K+**

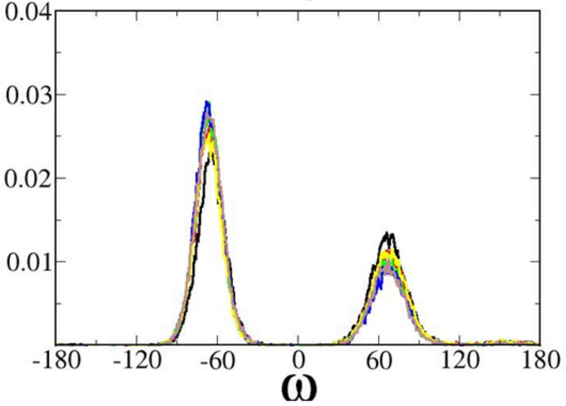


**K**

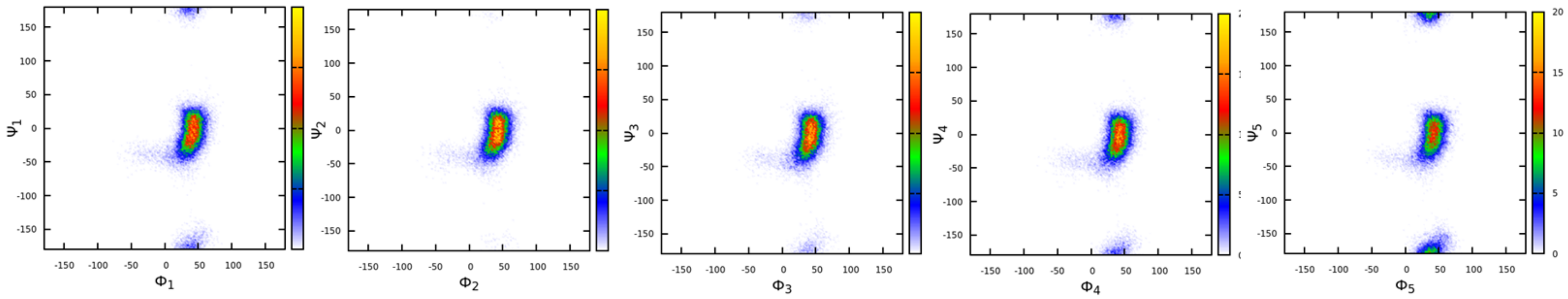
Overimposed phi and psi torsion angles



Overimposed omega torsion angles



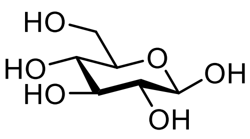
| | gg(%) | gt(%) | tg(%) |
|---|---|---|---|
| **ω1** | 56 | 43 | 1 |
| **ω2** | 61 | 38 | 1 |
| **ω3** | 68 | 31 | 1 |
| **ω4** | 63 | 36 | 1 |
| **ω5** | 73 | 26 | 1 |
| **ω6** | 70 | 29 | 1 |



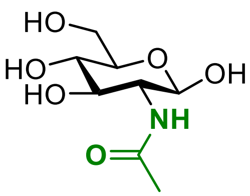
Non-reducing end

Reducing end

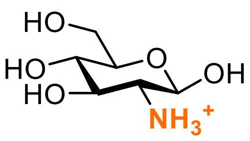
# Dihedral analysis of (K$^+$N)$_3$



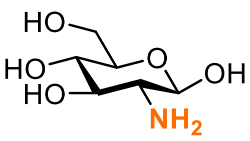
**A**



**N**

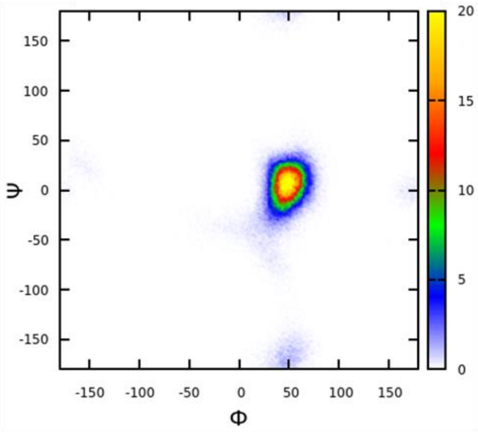


**K$^+$**

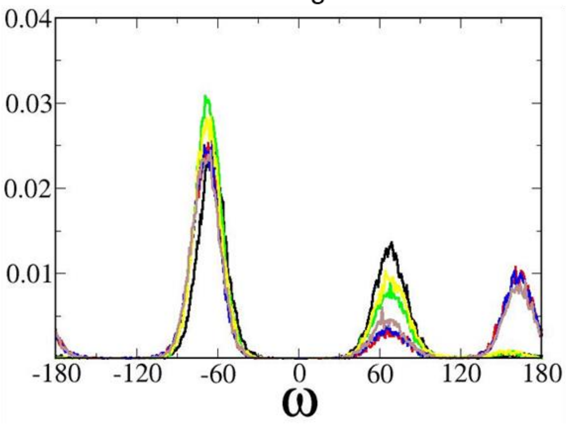


**K**

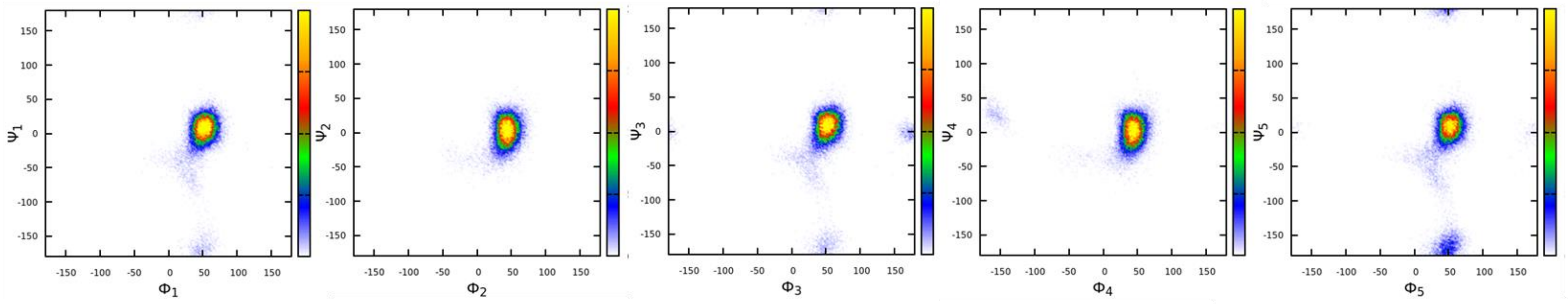
Overimposed phi and psi torsion angles



Overimposed omega torsion angles



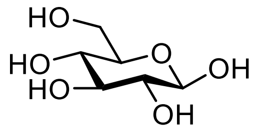
| | gg(%) | gt(%) | tg(%) |
|---|---|---|---|
| **ω1** | 60 | 38 | 2 |
| **ω2** | 62 | 10 | **28** |
| **ω3** | 73 | 25 | 2 |
| **ω4** | 62 | 11 | **27** |
| **ω5** | 68 | 30 | 2 |
| **ω6** | 62 | 14 | **24** |



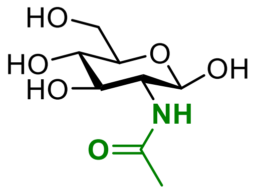
Non-reducing end

Reducing end

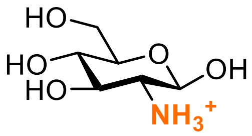
# Dihedral analysis of (KN)$_3$



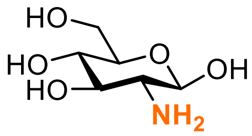
**A**



**N**

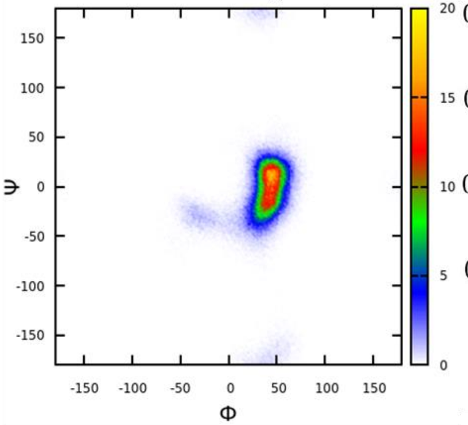


**$K^+$**

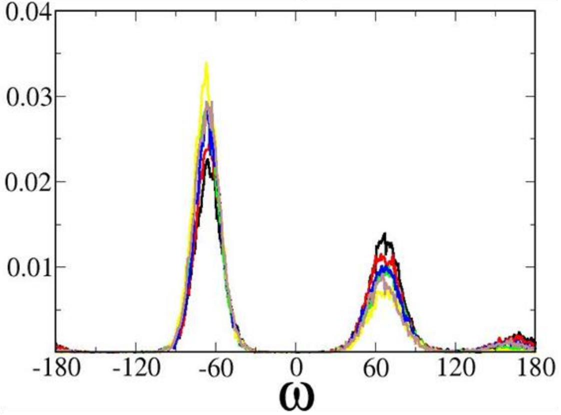


**K**

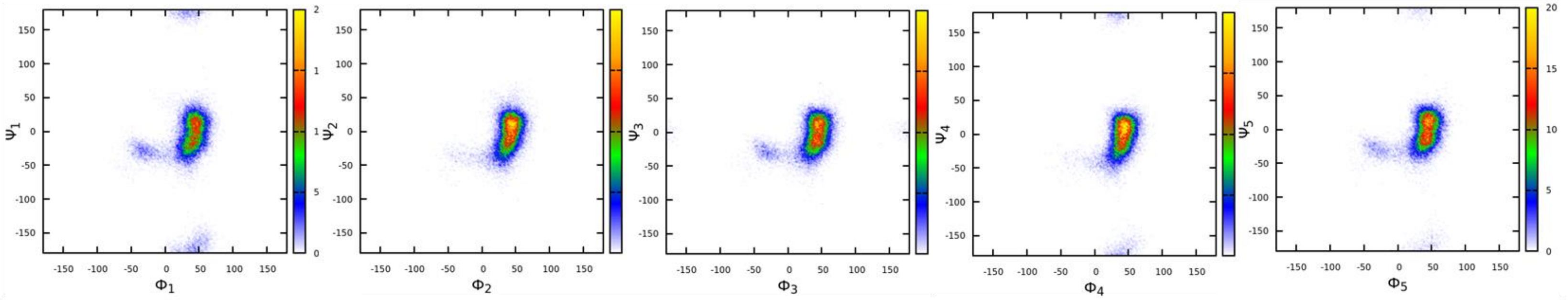
Overimposed phi and psi torsion angles



Overimposed omega torsion angles



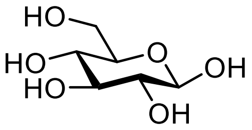
| | gg(%) | gt(%) | tg(%) |
|---|---|---|---|
| **ω1** | 54 | 40 | 6 |
| **ω2** | 59 | 36 | 5 |
| **ω3** | 68 | 30 | 2 |
| **ω4** | 67 | 31 | 4 |
| **ω5** | 76 | 22 | 1 |
| **ω6** | 70 | 25 | 5 |



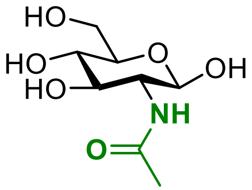
Non-reducing end

Reducing end

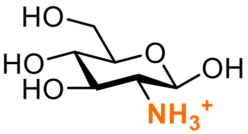
## Dihedral analysis of NN$K^+K^+$NN



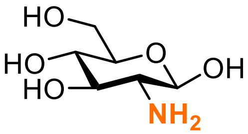
**A**



**N**

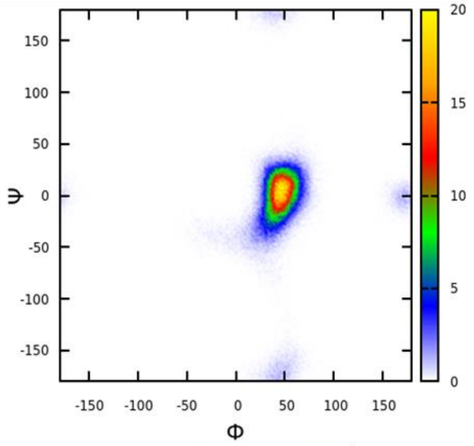


**$K^+$**

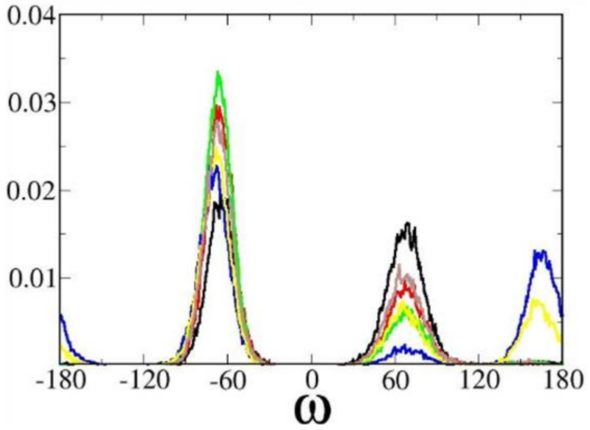


**K**

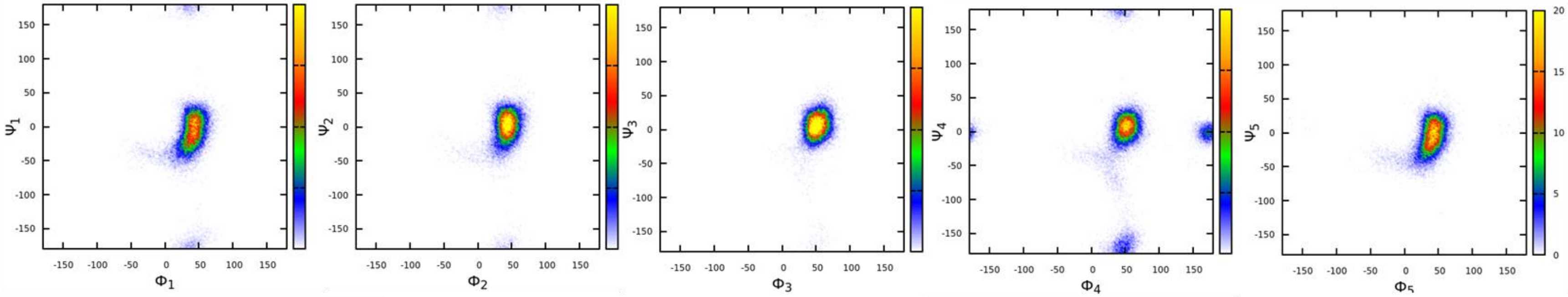
Overimposed phi and psi torsion angles



Overimposed omega torsion angles



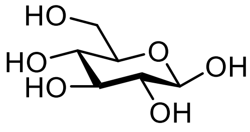
| | gg(%) | gt(%) | tg(%) |
|---|---|---|---|
| ω1 | 50 | 49 | 1 |
| ω2 | 70 | 29 | 1 |
| ω3 | 79 | 19 | 2 |
| ω4 | 57 | 6 | 37 |
| ω5 | 60 | 20 | 20 |
| ω6 | 66 | 33 | 1 |



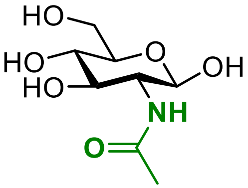
Non-reducing end

Reducing end

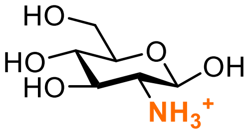
# Dihedral analysis of $K^+$ $K^+K^+$NNN



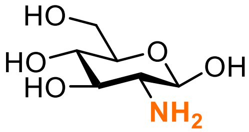
**A**



**N**

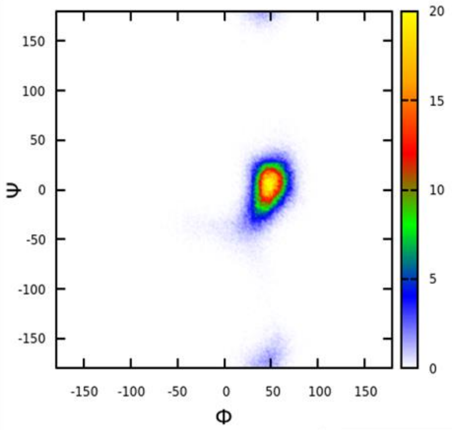


**$K^+$**

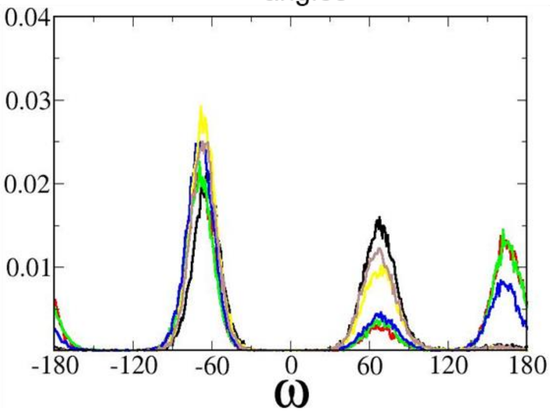


**K**

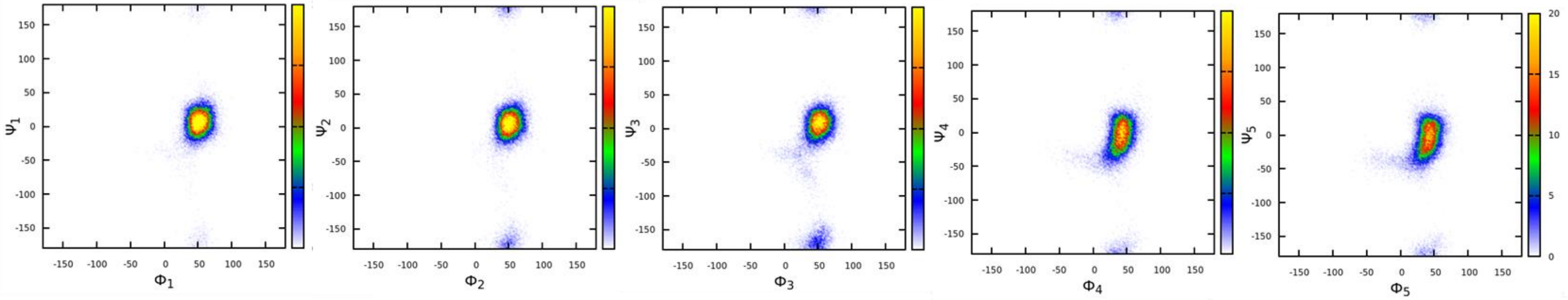
Overimposed phi and psi torsion angles



Overimposed omega torsion angles



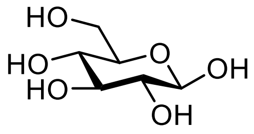
| | gg(%) | gt(%) | tg(%) |
|---|---|---|---|
| **ω1** | 52 | 46 | 2 |
| **ω2** | 54 | 9 | **38** |
| **ω3** | 53 | 10 | **37** |
| **ω4** | 63 | 13 | **23** |
| **ω5** | 67 | 32 | 1 |
| **ω6** | 61 | 37 | 2 |



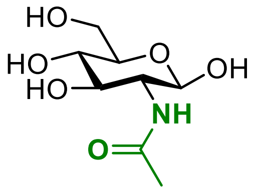
Non-reducing end

Reducing end

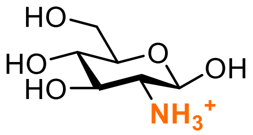
## Dihedral analysis of NK$^+$NNK$^+$N



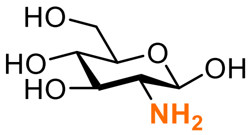
**A**



**N**

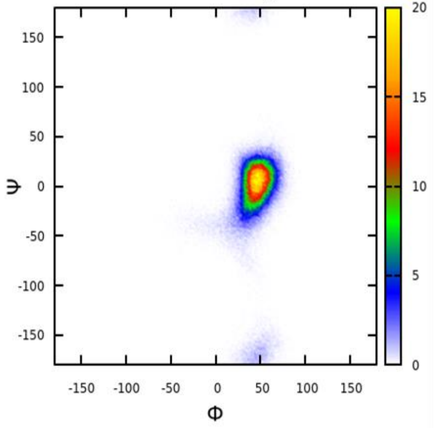


**K$^+$**

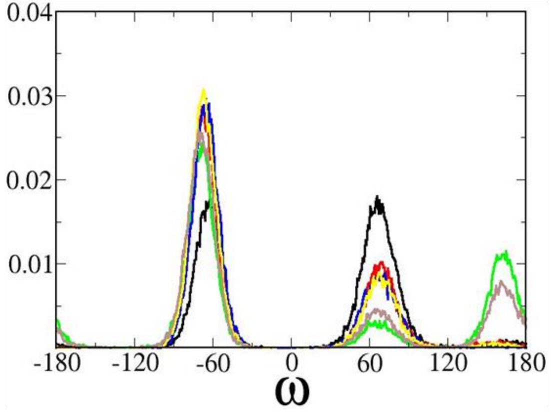


**K**

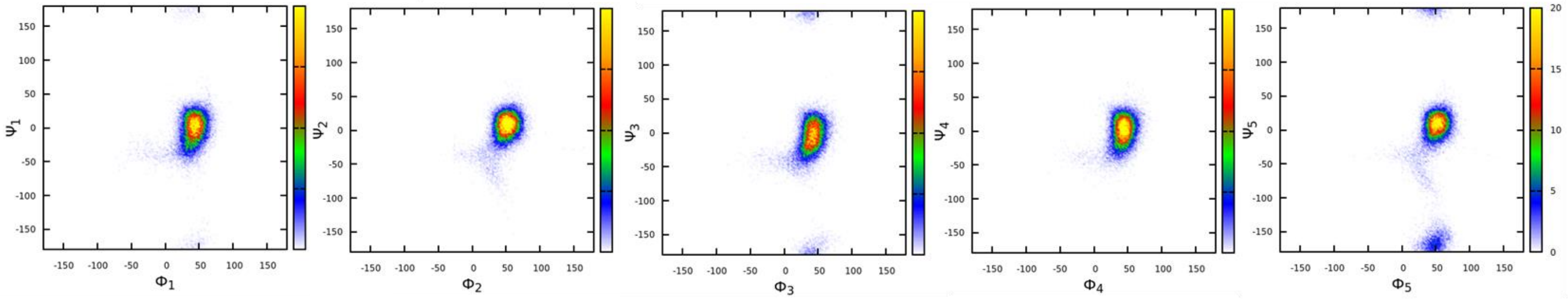
Overimposed phi and psi torsion angles



Overimposed omega torsion angles



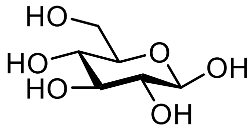
| | gg(%) | gt(%) | tg(%) |
|---|---|---|---|
| **ω1** | 44 | 54 | 2 |
| **ω2** | 68 | 30 | 2 |
| **ω3** | 59 | 9 | **32** |
| **ω4** | 71 | 28 | 1 |
| **ω5** | 72 | 26 | 1 |
| **ω6** | 65 | 13 | **22** |



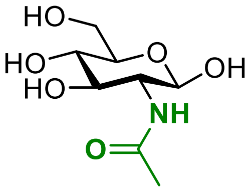
Non-reducing end

Reducing end

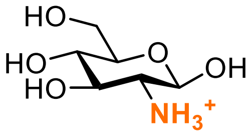
# Dihedral analysis of **NKNNKN**



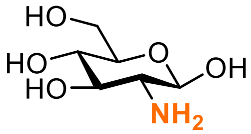
**A**



**N**

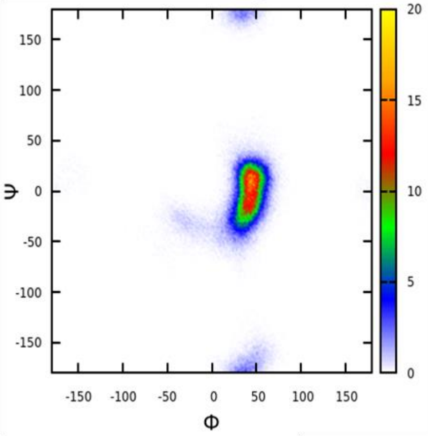


**$K^+$**

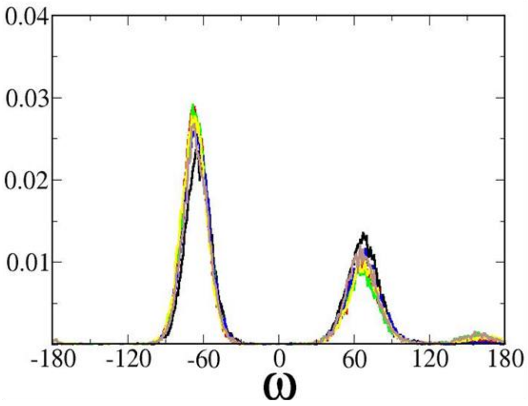


**K**

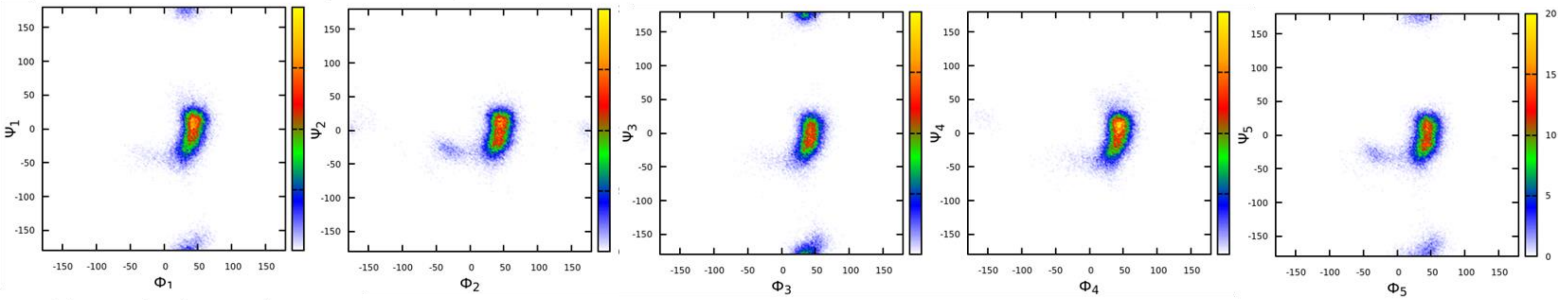
Overimposed phi and psi torsion angles



Overimposed omega torsion angles

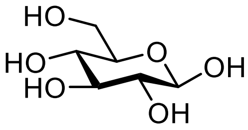


| | gg(%) | gt(%) | tg(%) |
|---|---|---|---|
| **ω1** | 59 | 40 | 1 |
| **ω2** | 68 | 30 | 2 |
| **ω3** | 68 | 28 | 4 |
| **ω4** | 65 | 34 | 1 |
| **ω5** | 68 | 29 | 3 |
| **ω6** | 62 | 34 | 4 |

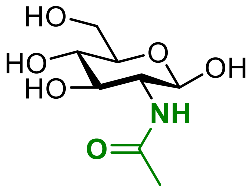


Non-reducing end

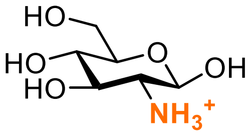
## Dihedral analysis of **NANNAN**



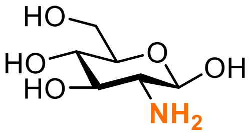
**A**



**N**

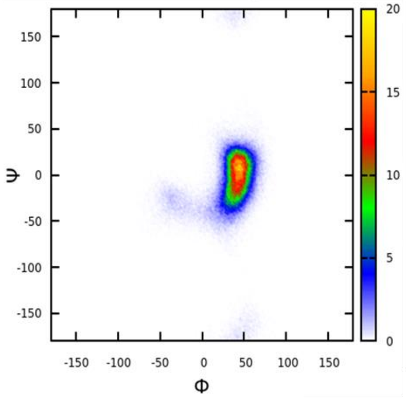


**$K^+$**

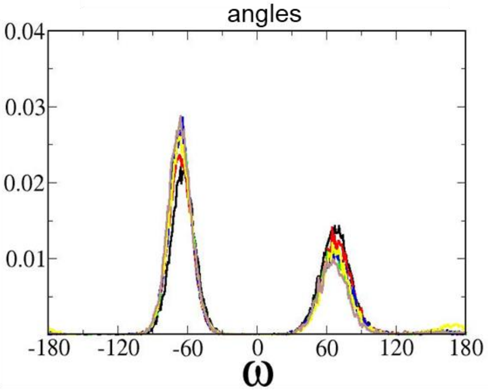


**K**

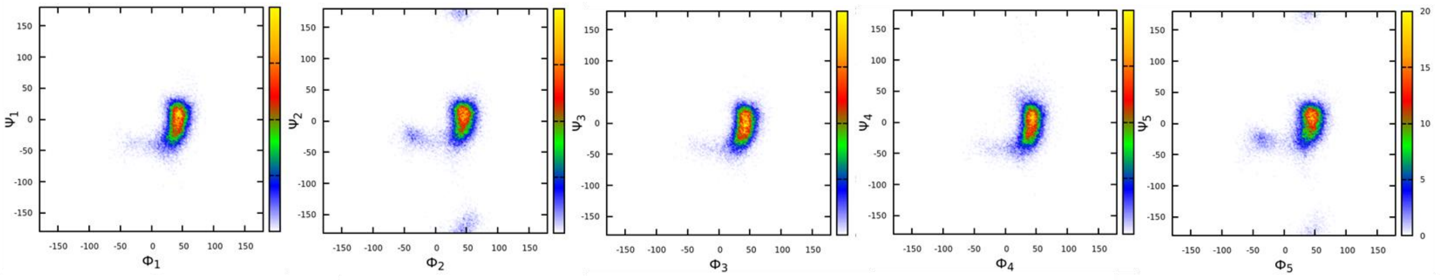
Overimposed phi and psi torsion angles



Overimposed omega torsion angles



| | gg(%) | gt(%) | tg(%) |
|---|---|---|---|
| **ω1** | 65 | 34 | 1 |
| **ω2** | 65 | 32 | 3 |
| **ω3** | 69 | 30 | 1 |
| **ω4** | 65 | 34 | 1 |
| **ω5** | 64 | 35 | 1 |
| **ω6** | 75 | 24 | 1 |

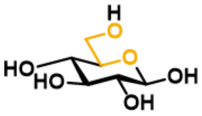


Non-reducing end

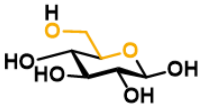
Reducing end

# SF22

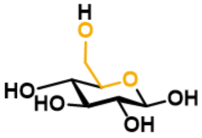
Superimposed Omega (**ω**) torsion angles.



gt (ω = 60°)

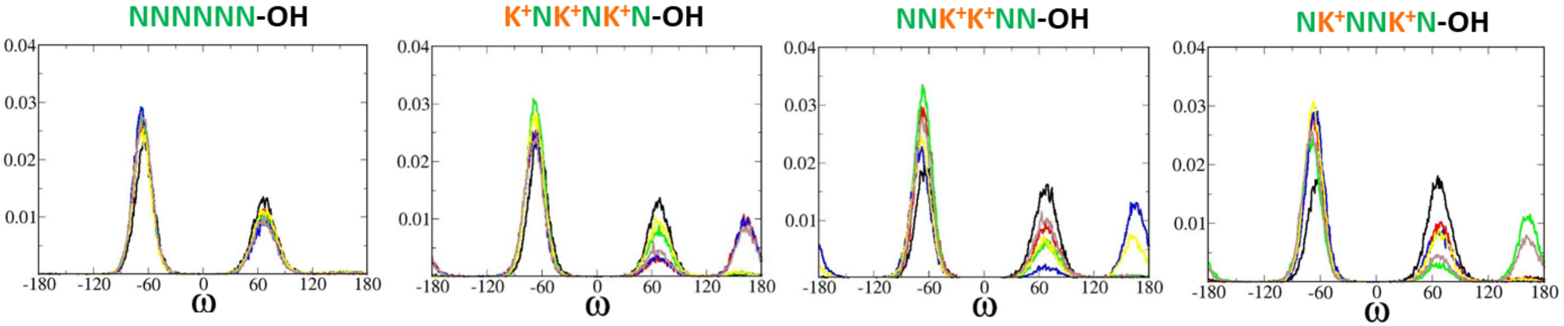


tg (ω = 180°)



gg (ω = - 60°)

| Non-reducing end | ω1 | ⬛ | ω2 | 🟥 | ω3 | 🟩 | ω4 | 🟦 | ω5 | 🟨 | ω6 | ▩ | Reducing end |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

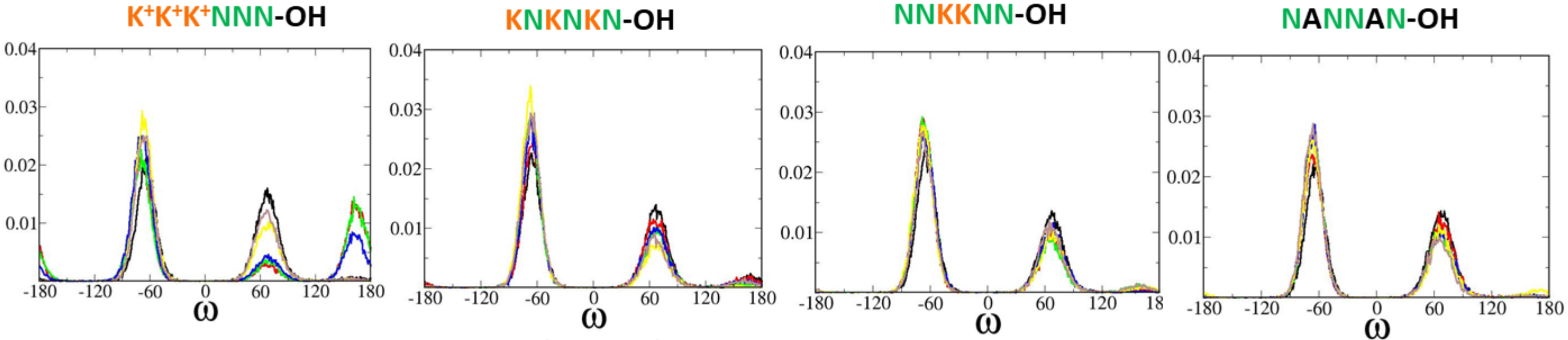


65 % gg | 34 % gt | 1% tg

61 % gg | 24 % gt | 14 % tg

63 % gg | 26 % gt | 11 % tg

63 % gg | 27 % gt | 10 % tg



57 % gg | 26 % gt | 11 % tg

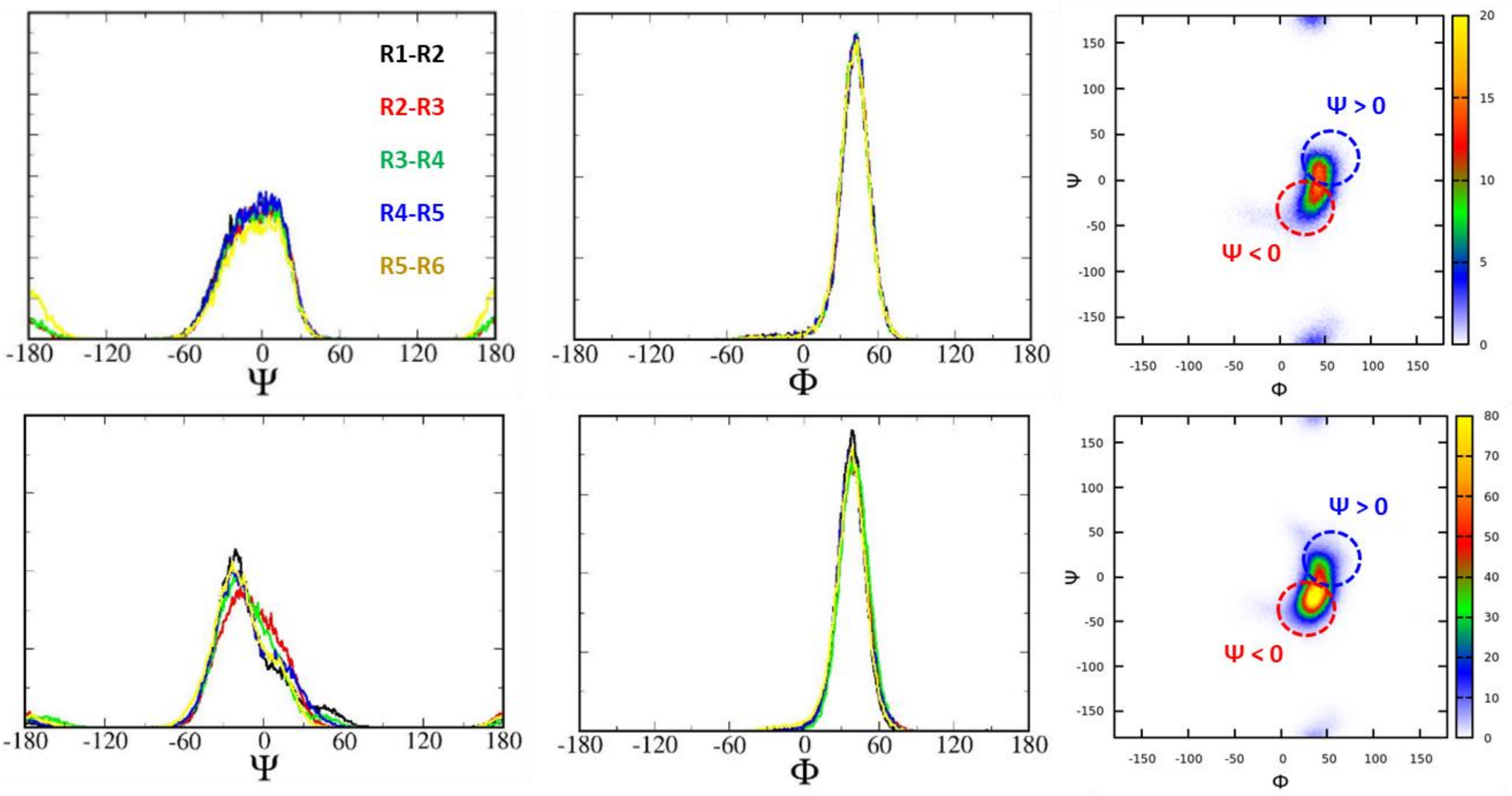
65 % gg | 31 % gt | 4% tg

65 % gg | 32 % gt | 3% tg

66 % gg | 33 % gt | 1% tg

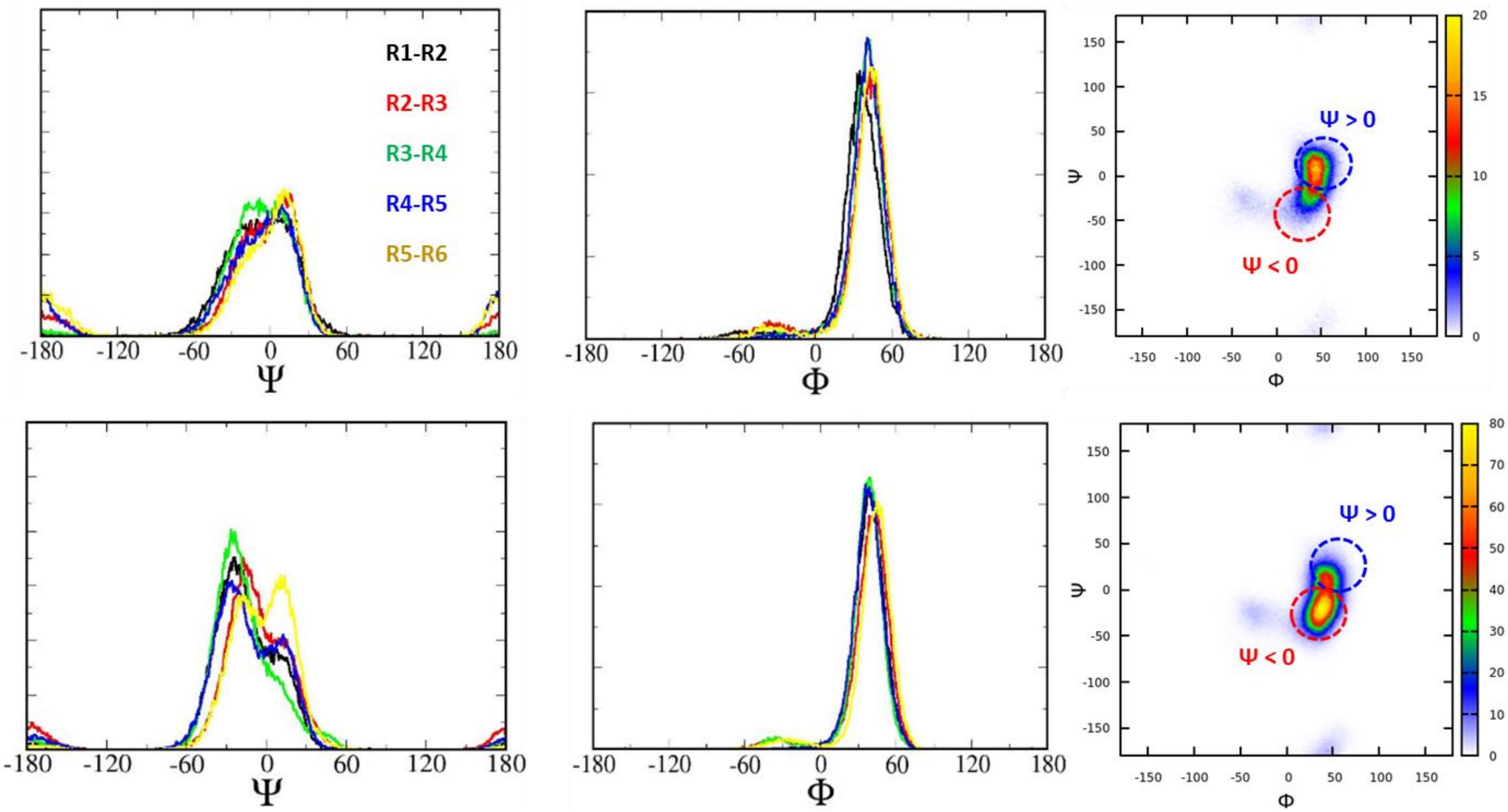
**SF23** Superimposed **ψ** and **φ** torsion angles for a single chain (*top*) *vs* the concentrated system (*bottom*).

**NNNNNN-OH**



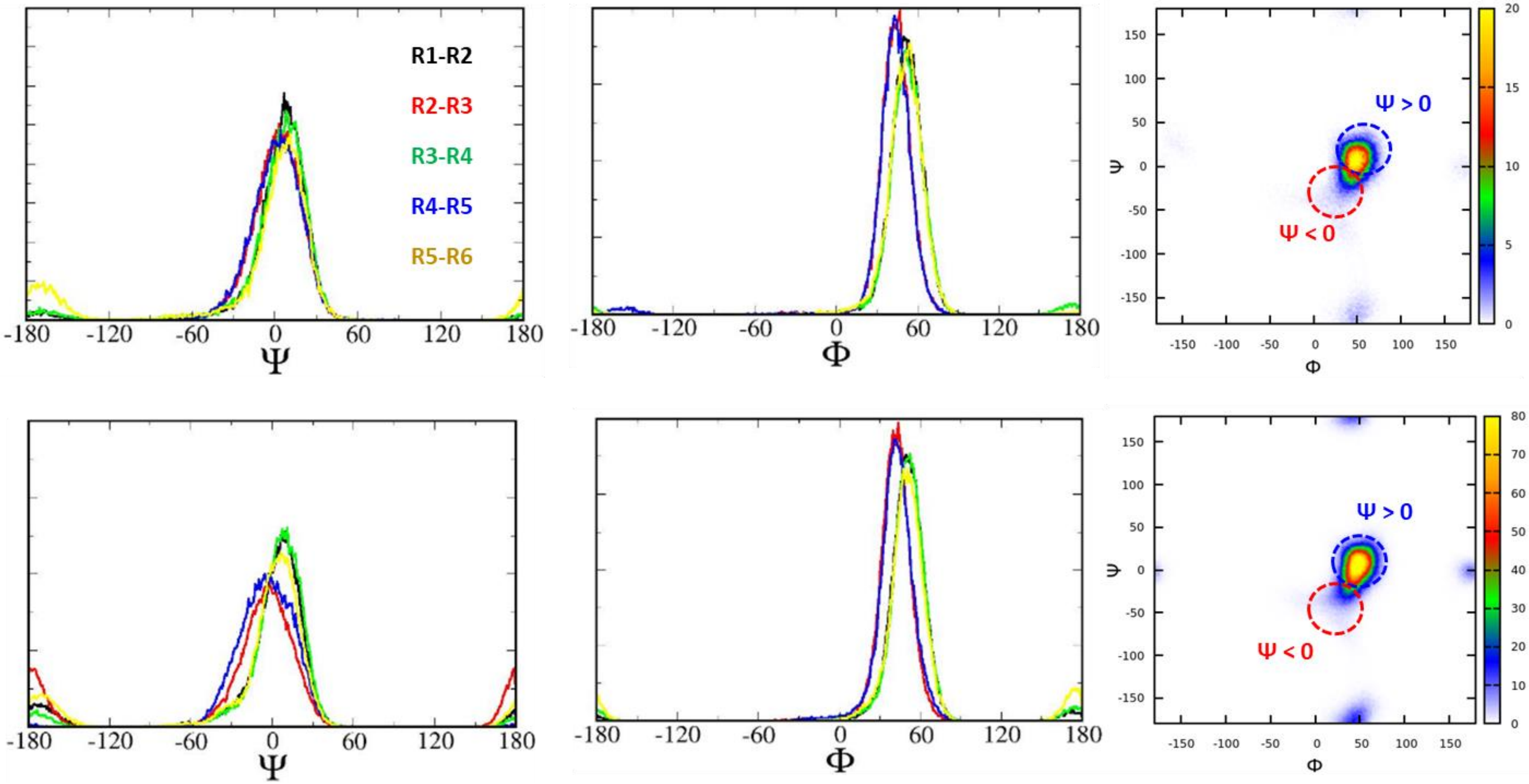
Superimposed **ψ** and **φ** torsion angles for a single chain (*top*) *vs* the concentrated system (*bottom*).

## **NANNAN-OH**



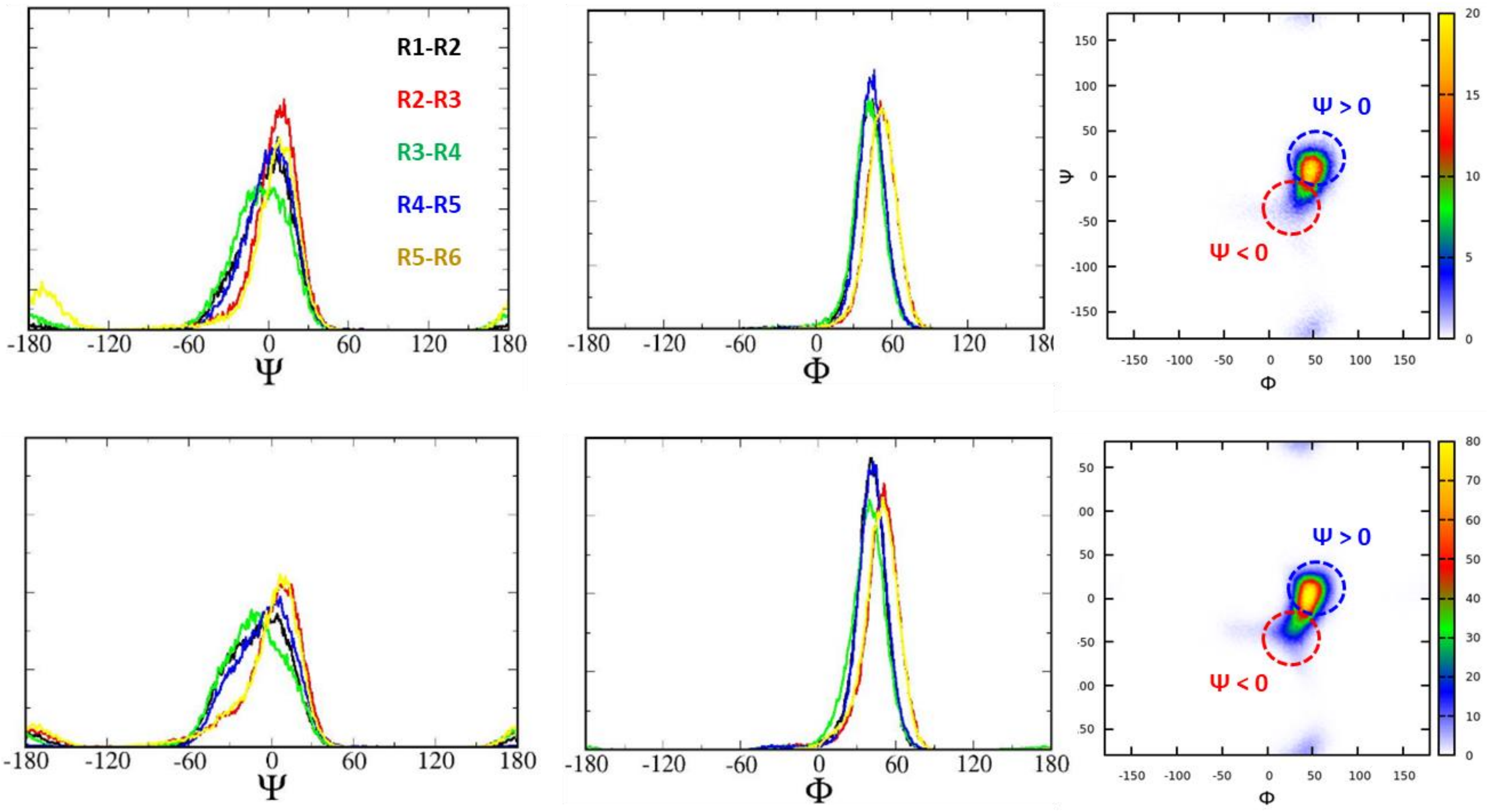
Superimposed **ψ** and **φ** torsion angles for a single chain (*top*) *vs* the concentrated system (*bottom*).

**K$^+$NK$^+$NK$^+$N-OH**



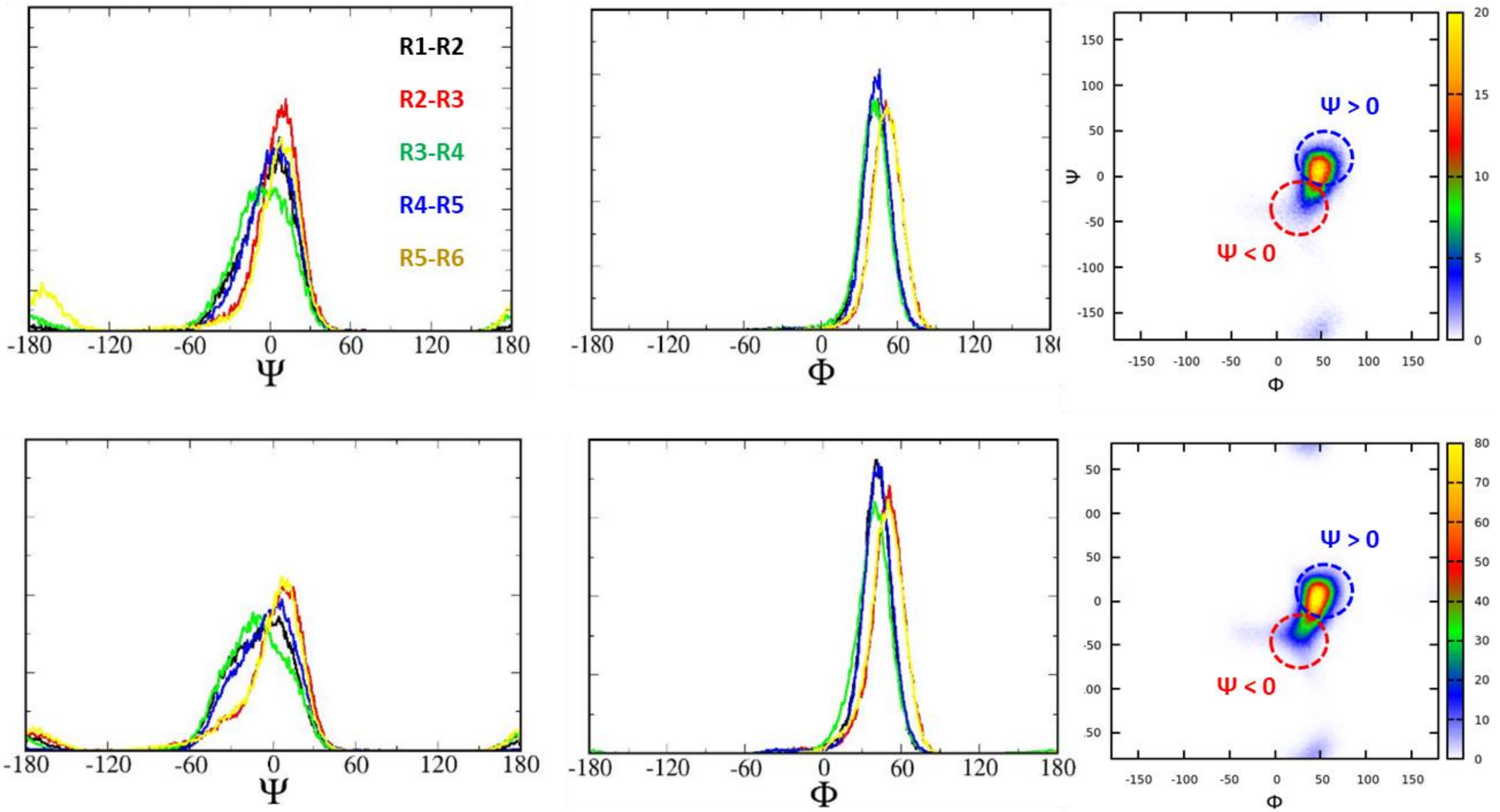
Superimposed **ψ** and **φ** torsion angles for a single chain (*top*) *vs* the concentrated system (*bottom*).

**NK$^+$NNK$^+$N-OH**



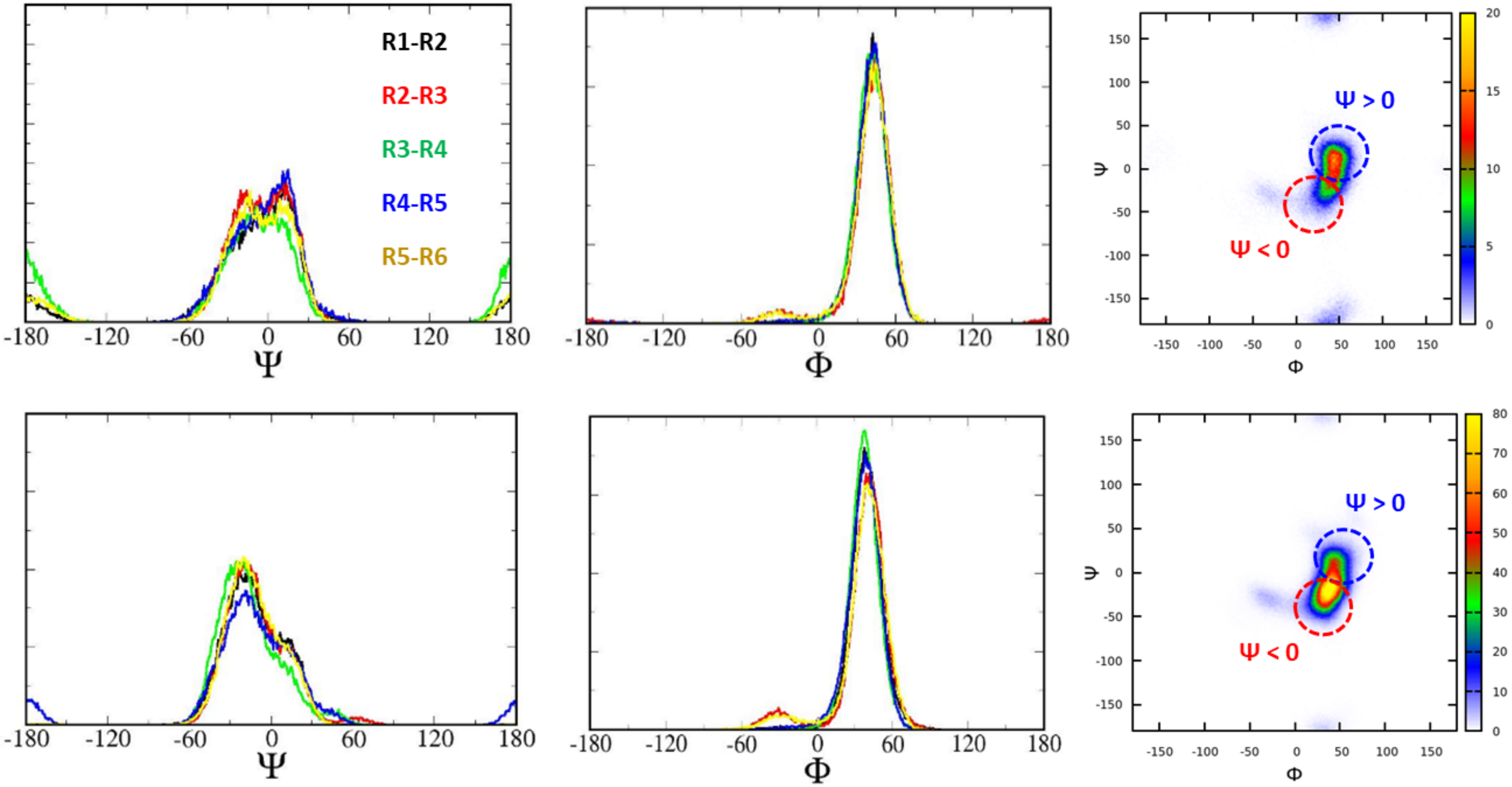
## Superimposed **ψ** and **φ** torsion angles for a single chain (*top*) *vs* the concentrated system (*bottom*).

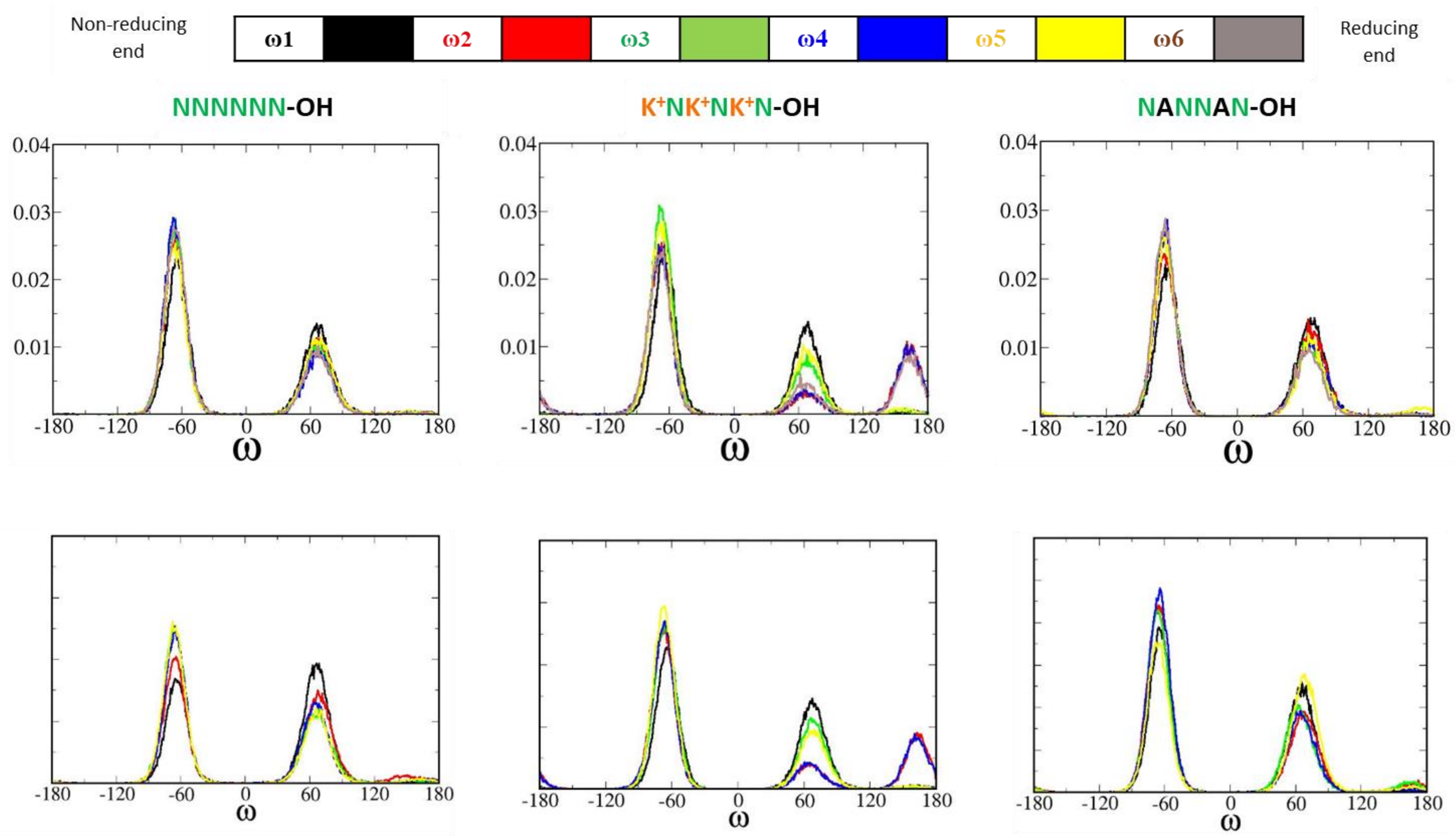
**$\mathrm{NK^{+}NNK^{+}N}$-OH**



Superimposed **ψ** and **φ** torsion angles for a single chain (*top*) *vs* the concentrated system (*bottom*).

**NKNNKN-OH**

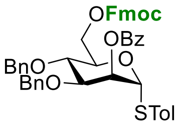



Non-reducing end
ω1
ω2
ω3
ω4
ω5
ω6
Reducing end
NNNNNN-OH
K+NK+NK+N-OH
NANNAN-OH
0.04
0.03
0.02
0.01
-180
-120
-60
0
60
120
180
ω

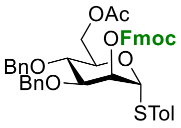
## 7.2. General materials and methods

All chemicals used were reagent grade and used as supplied unless otherwise noted. The automated syntheses were performed on a home-built synthesizer developed at the Max Planck Institute of Colloids and Interfaces. Analytical thin-layer chromatography (TLC) was performed on Merck silica gel 60 F254 plates (0.25 mm). Compounds were visualized by UV irradiation or dipping the plate in a staining solution (sugar stain: 10% $H_2SO_4$ in EtOH; CAM: 48 g/L ammonium molybdate, 60 g/L ceric ammonium molybdate in 6% $H_2SO_4$ aqueous solution). Flash column chromatography was carried out by using forced flow of the indicated solvent on Fluka Kieselgel 60 M (0.04 – 0.063 mm). Analysis and purification by normal and reverse phase HPLC was performed by using an Agilent 1200 series. Products were lyophilized using a Christ Alpha 2-4 LD plus freeze dryer. $^1H$, $^{13}C$ and HSQC NMR spectra were recorded on a Varian 400-MR (400 MHz), a Varian 600-MR (600 MHz) or a Varian 700-MR (700 MHz) spectrometer. Spectra were recorded in $CDCl_3$ by using the solvent residual peak chemical shift as the internal standard ($CDCl_3$: 7.26 ppm $^1H$, 77.0 ppm $^{13}C$) or in $D_2O$ using the solvent as the internal standard in $^1H$ NMR ($D_2O$: 4.79 ppm $^1H$). High resolution mass spectra were obtained using a 6210 ESI-TOF mass spectrometer (Agilent) and a MALDI-TOF autoflex$^{TM}$ (Bruker). MALDI and ESI mass spectra were run on IonSpec Ultima instruments. All solution-state NMR experiments regarding **R5** and its interaction with oligosaccharides were performed using a Bruker Ascend$^{TM}$ (AvanceIII HD) 700 MHz NMR spectrometer with water suppression. All spectra were recorded at 297 K. In order to monitor the aggregation of R5, $^1H$ spectra were acquired on Day0, Day1 and Day5. Data were processed with MestReNova.

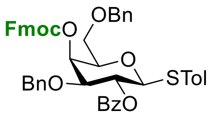
## 7.3. Building blocks for AGA



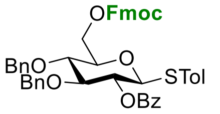
**33**



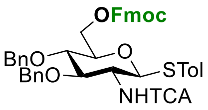
**34**



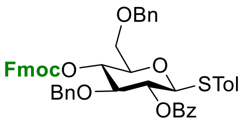
**35**



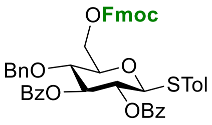
**36**



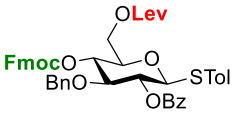
**37**



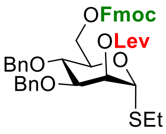
**38**



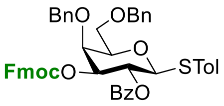
**39**



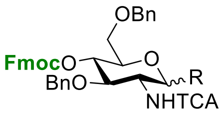
**40**



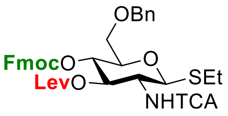
**41**



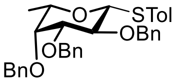
**42**



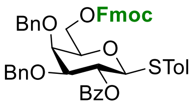
**43a R = SEt**
**44 R = OPO(OBu)$_2$**



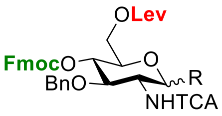
**62**



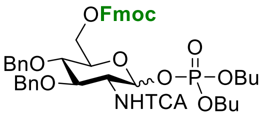
**63**



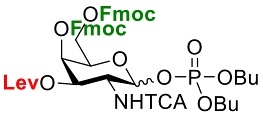
**64**



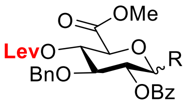
**65a R = SEt**
**65b R = OPO(OBu)$_2$**



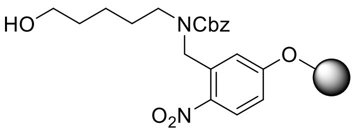
**66**



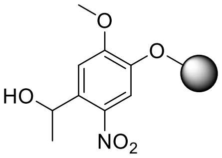
**67**



**68a R = SPh**
**68b R = OPO(OBu)$_2$**



**3**



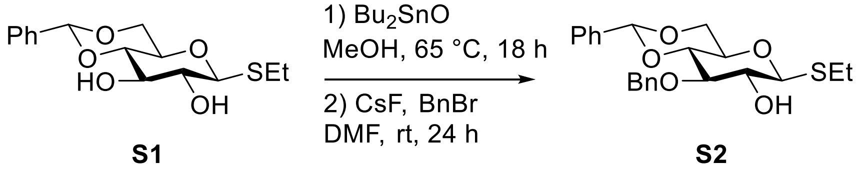
**4**

● = Merrifield resin

All BBs were purchased from GlycoUniverse, apart from **40, 43b, 65b, 66, 67, 68b** and their synthesis is reported below. Merrifield resin equipped photocleavable linkers (**3**, loading 0.34 mmol/g and **4**, loading 0.30 mmol/g) were prepared according to previous literature.[63]

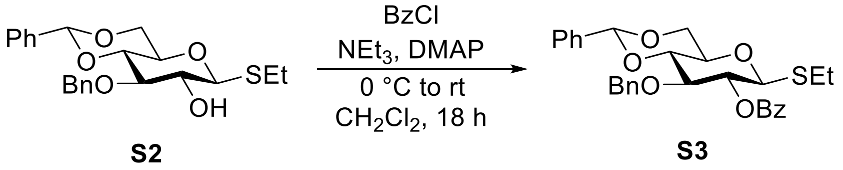
## 7.3.1. Synthesis of 40

**Ethyl 4,6-*O*-benzylidene-3-*O*-benzyl-1-thio-β-glucopyranoside, S2**



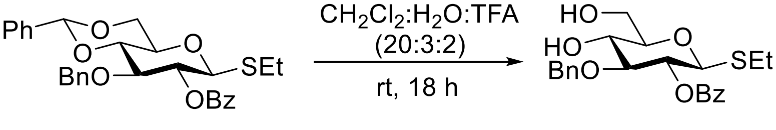
Compound **S1** (10.0 g, 32.0 mmol) was dissolved in MeOH (250 mL), di-*n*-butyltin oxide (9.6 g, 38.4 mmol) was added and the reaction mixture heated to 65 °C for 18 h. The reaction mixture was then cooled to rt, concentrated *in vacuo* and the residue dissolved in DMF (200 mL). Benzyl bromide (6.6 g, 38.4 mmol) and cesium(I) fluoride (6.32 g, 41.6 mmol) were added and the mixture stirred at rt for 24 h. The reaction mixture was concentrated *in vacuo* and the residue dissolved in $CH_2Cl_2$ (250 mL). The organic layer was washed with 1M potassium fluoride (100 mL, aq.), dried over ($MgSO_4$), filtered and concentrated *in vacuo*. The resulting crude was purified by column chromatography (Hexanes : EtOAc = 2:1) to give **S2** as a white solid (9.0 g, 22.3 mmol, 67%). $^1$H NMR (400 MHz, $CDCl_3$) δ 7.51 – 7.47 (m, 2H), 7.41 – 7.30 (m, 8H), 5.58 (s, 1H), 4.98 (d, *J* = 11.6 Hz, 1H), 4.82 (d, *J* = 11.6 Hz, 1H), 4.47 (d, *J* = 9.7 Hz, 1H), 4.36 (dd, *J* = 10.5, 5.0 Hz, 1H), 3.83 – 3.64 (m, 3H), 3.62 – 3.44 (m, 2H), 2.82 – 2.68 (m, 2H), 1.32 (t, *J* = 7.4 Hz, 3H). NMR data were in agreement with previously reported.[182]

**Ethyl 4,6-*O*-benzylidene-3-*O*-benzyl-2-*O*-benzoyl-1-thio-β-glucopyranoside, S3**



Compound **S2** (9.0 g, 22.3 mmol) was dissolved in anhydrous $CH_2Cl_2$ (200 mL). Triethylamine (8.7 mL, 67 mmol) and DMAP (825 mg, 6.7 mmol) were added slowly to the solution while stirring. Benzoyl chloride (3.9 mL, 33.5 mmol) was slowly added at 0 °C and the reaction allowed to rt. Upon completion (18 h) the reaction was quenched with sat. aq. solution of $NaHCO_3$. The mixture was washed three times with sat. aq. solution of $NaHCO_3$ and one time with brine. The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure. The crude product was purified with flash chromatography (Hexanes : EtOAc = 6:1) to obtain **S3** as a white solid (10.5 g, 20.3 mmol, 92%). $^1$H NMR (400 MHz, $CDCl_3$) δ 8.04 (dt, *J* = 8.4, 1.2 Hz, 2H), 7.64 – 7.38 (m, 9H), 7.20 – 7.08 (m, 5H), 5.64 (s, 1H), 5.37 (m, 1H), 4.90 – 4.70 (m, 2H), 4.65 (d, *J* = 10.1 Hz, 1H), 4.43 (dd, *J* = 10.5, 4.9 Hz, 1H), 3.98 – 3.80 (m, 3H), 3.58 (td, *J* = 9.5, 5.0 Hz, 1H), 2.74 (m, 2H), 1.24 (t, *J* = 7.4 Hz, 3H). NMR data were in agreement with previously reported.[183]

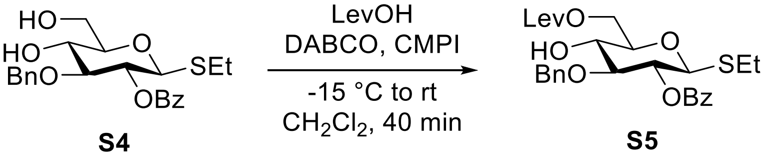
**Ethyl 3-*O*-benzyl-2-*O*-benzoyl-1-thio-β-glucopyranoside, S4**



**S3** **S4**

TFA (17 mL) and water (18 mL) were added to a solution of compound **S3** (10.5 g, 20.7 mmol) in $CH_2Cl_2$ (180 mL) and the mixture was stirred at rt for 18 h. The reaction was quenched with sat. aq. $NaHCO_3$ and extracted three times with $CH_2Cl_2$. The organic layer was dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by flash chromatography (Hexanes : EtOAc = 1:1) to give compound **S4** as a white solid (7.0 g, 16.7 mmol, 80%). $^1H$ NMR (400 MHz, $CDCl_3$) δ 8.06 (dd, *J* = 8.0, 1.5 Hz, 2H), 7.60 (td, *J* = 7.3, 1.5 Hz, 1H), 7.47 (t, *J* = 7.8 Hz, 2H), 7.22 (m, 5H), 5.34 – 5.23 (m, 1H), 4.75 (d, *J* = 11.4 Hz, 1H), 4.66 – 4.54 (m, 2H), 3.98 – 3.67 (m, 4H), 3.48 (m, 1H), 2.71 (m, 2H), 2.20 (s, 2H), 1.23 (t, *J* = 7.4 Hz, 3H); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ 165.4, 137.8, 133.5, 130.0, 129.8, 128.7, 128.7, 128.7, 128.2, 128.2, 84.0, 84.0, 79.6, 74.9, 72.4, 70.6, 62.8, 24.3, 15.0; $[\alpha]_D^{20}$ = +87.09; IR (neat) $\nu_{max}$= 2988, 1739, 1373, 1236, 1044; $R_f$ = 0.15 ($SiO_2$, Hexanes : EtOAc = 1:1); HRMS (QToF): Calcd for $C_{22}H_{26}O_6SNa$ $[M+Na]^+$ 441.1362; found 441.1391.

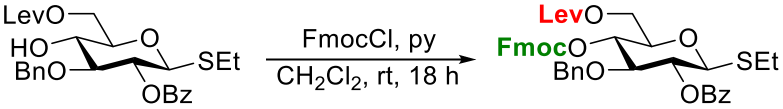
**Ethyl 3-*O*-benzyl-2-*O*-benzoyl-6-O-levulinoyl-1-thio-β-glucopyranoside, S5**



**S4** **S5**

Compound **S4** (7.0 g, 16.75 mmol) was dissolved in $CH_2Cl_2$ (200 mL). Levulinic acid (3.45 mL, 33.5 mmol) and 2-chloro-1-methylpyridinium iodide (8.5 gr, 33.5 mmol) were added. The reaction was stirred for 15 min, then cooled to -15 °C and DABCO (7.5 g, 67 mmol) was added. The reaction mixture was stirred for 40 min and then filtered over a plug of celite and concentrated *in vacuo*. The reaction mixture was quenched with sat. aq. $NaHCO_3$ (200 mL) and extracted with $CH_2Cl_2$. The organic layer was washed with brine and dried over $Na_2SO_4$. Solvent removed by reduced pressure and purification by flash chromatography (Hexanes : EtOAc = 1:1) afforded compound **S5** as a white solid (7.3 g, 14.1 mmol, 84%). $^1H$ NMR (400 MHz, $CDCl_3$) δ 8.11 (dt, *J* = 8.1, 1.1 Hz, 2H), 7.70 – 7.63 (m, 1H), 7.57 – 7.49 (m, 2H), 7.26 (s, 5H), 5.39 – 5.29 (m, 1H), 4.78 (q, *J* = 11.4 Hz, 2H), 4.68 – 4.52 (m, 2H), 4.39 (dd, *J* = 12.2, 2.1 Hz, 1H), 3.82 – 3.70 (m, 2H), 2.98 – 2.66 (m, 7H), 2.27 (s, 3H), 1.41 – 1.25 (m, 3H); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ 207.0, 173.5, 165.4, 137.9, 133.4, 130.0, 129.9, 128.6, 128.2, 128.0, 84.0, 83.4, 78.1, 75.0, 72.2, 70.2, 63.5, 38.14, 30.0, 28.0, 24.3, 15.0; $[\alpha]_D^{20}$ = + 35.51; IR (neat) $\nu_{max}$= 2930, 1721, 1361, 1273.8, 1070,749, 713; $R_f$ = 0.5 ($SiO_2$, Hexanes : EtOAc = 1:1); HRMS (QToF): Calcd for $C_{27}H_{32}O_6SNa$ $[M+Na]^+$ 539.1716; found 539.1716.

**Ethyl 3-*O*-benzyl-2-*O*-benzoyl-6-*O*-levulinoyl-4-*O*-(9-fluorenylmethoxycarbonyl)-1-thio-β-glucopyranoside, 40**

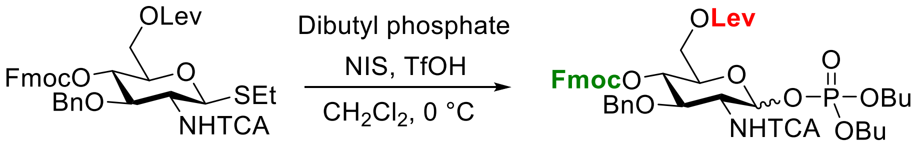


**S5** **40**

Compound **S5** (7.3 g, 14.1 mmol) was dissolved in $CH_2Cl_2$ (100 mL) and pyridine was added (3.5 mL, 42.4 mmol). FmocCl (7.3 g, 28.3 mmol) was dissolved in $CH_2Cl_2$ (100 mL) and added to the reaction mixture. The yellow solution was stirred for 3 h and then quenched with 1 M solution of HCl. The organic layer was washed one time with 1 M HCl, one time with sat. aq. solution of $NaHCO_3$ and one time with brine. The crude compound was purified with flash column chromatography (Hexanes : EtOAc = 2:1) to give compound **40** as white solid (6.7 g, 10.7 mmol, 68%). $^1H$ NMR (400 MHz, $CDCl_3$) δ 8.29 – 8.19 (m, 2H), 7.97 (ddt, *J* = 7.2, 6.3, 0.9 Hz, 2H), 7.87 – 7.78 (m, 3H), 7.68 (t, *J* = 7.8 Hz, 2H), 7.65 – 7.57 (m, 2H), 7.55 – 7.47 (m, 2H), 7.33 – 7.22 (m, 5H), 5.57 (dd, *J* = 10.0, 9.1 Hz, 1H), 5.22 (dd, *J* = 10.1, 9.3 Hz, 1H), 4.85 – 4.66 (m, 4H), 4.61 (dd, *J* = 10.5, 7.2 Hz, 1H), 4.53 – 4.40 (m, 3H), 4.12 (t, *J* = 9.2 Hz, 1H), 3.97 (m, 1H), 3.06 – 2.89 (m, 4H), 2.89 – 2.76 (m, 2H), 2.40 (s, 3H), 1.45 (d, *J* = 7.4 Hz, 3H); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ 206.5, 172.4, 165.0, 154.2, 143.3, 143.1, 141.3, 141.3, 137.2, 133.4, 129.9, 129.6, 128.5, 128.2, 128.0, 128.0, 127.9, 127.7, 127.3, 127.3, 125.1, 125.0, 120.1, 120.1, 83.8, 80.9, 75.8, 74.5, 74.4, 71.8, 70.2, 62.7, 46.7, 37.9, 29.9, 27.9, 24.2, 14.9; $[\alpha]_D^{20}$ = +29.48; IR (neat) $\nu_{max}$= 3661, 2982, 1463, 1383, 1252, 1153, 1073, 955, 816; $R_f$ = 0.5 ($SiO_2$, Hexanes : EtOAc = 2:1); HRMS (QToF): Calcd for $C_{42}H_{42}O_{10}SNa$ $[M+Na]^+$ 761.2396; found 761.2405.

## 7.3.2. Synthesis of 64b

**Dibutoxyphosphoryloxy 3-*O*-benzyl-4-*O*-(9-fluorenylmethoxycarbonyl)-2-deoxy-2-*N*-trichloroacetyl-6-*O*-levulinoyl-α/β-galactopyranoside, 64b**

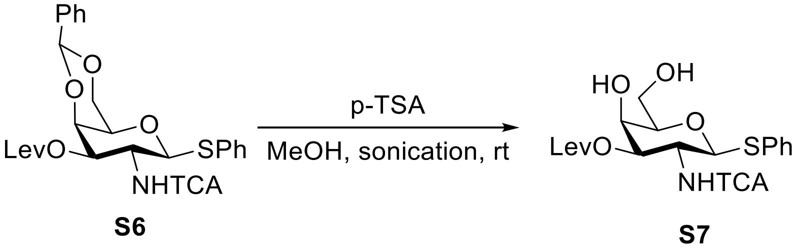


**64a** **64b**

An oven dried round bottom flask containing a solution of thioglycoside **64a** (0.5 g, 0.64 mmol) and dibutyl hydrogen phosphate (0.25 mL, 1.28 mmol, 2.0 equiv) in anhydrous $CH_2Cl_2$ (7 mL) was cooled to 0 °C under Ar atmosphere. After 15 min, *N*-iodosuccinimide (215 mg, 0.96 mmol, 1.5 equiv.) was added followed by the dropwise addition of TfOH (6 µL, 0.06 mmol, 0.1 equiv.) at 0 °C. The reaction progress was checked every 30 min until the starting material was fully consumed. The reaction mixture was diluted with $CH_2Cl_2$ (50 mL) and a sodium thiosulfate solution (10% w/w in water, 20 mL) was added. The organic layer was then separated, washed with a $NaHCO_3$ saturated solution (20 mL), dried over $Na_2SO_4$, filtered, and concentrated. Compound **BB15b** was obtained after purification by column chromatography ($SiO_2$, Hexanes : EtOAc = 2:1) as a white solid (0.25 g, 0.27 mmol, 42%). $^1$H NMR (400 MHz, $D_2O$) δ 7.75 (t, *J* = 7.7 Hz, 3H), 7.59 (dd, *J* = 13.6, 7.5 Hz, 3H), 7.39 (q, *J* = 7.4 Hz, 3H), 7.31 – 7.28 (m, 2H), 7.23 – 7.15 (m, 6H), 6.84 (d, *J* = 8.7 Hz, 1H), 5.72 (dd, *J* = 5.9, 3.2 Hz, 1H), 5.06 (t, *J* = 9.6 Hz, 1H), 4.62 – 4.45 (m, 4H), 4.41 – 4.26 (m, 4H), 4.23 – 4.17 (m, 3H), 4.13 – 4.03 (m, 4H), 3.94 (t, *J* = 9.9 Hz, 1H), 2.86 – 2.52 (m, 6H), 2.17 (s, 3H), 1.71 – 1.62 (m, 7H), 1.45 – 1.33 (m, 5H), 0.94 (dd, *J* = 12.2, 7.3 Hz, 6H).; $^{13}$C NMR (101 MHz, $D_2O$) δ 206.5, 172.4, 162.1, 154.1, 143.4, 143.1, 141.5, 141.4, 137.0, 128.7, 128.6, 128.5, 128.2, 128.1, 128.1, 128.1, 128.0, 128.0, 127.7, 127.4, 125.3, 125.2, 125.0, 120.3, 120.2, 95.5, 95.5, 92.2, 91.2, 76.9, 76.1, 74.4, 73.9, 70.5, 69.9, 68.6, 68.5, 68.5, 62.6, 61.7, 54.4, 54.3, 46.8, 38.0, 32.4, 32.4, 32.4, 32.3, 30.3, 30.1, 30.0, 29.8, 28.3, 28.95, 18.8, 13.7.; $[\alpha]_D^{20}$ = + 0.19; IR (neat) $\nu_{max}$= 3255, 2961, 2928, 2876, 1750, 1717, 1521, 1452, 1358, 1256, 1154, 1107, 1058, 1027, 955, 838, 823, 784, 760, 742; $R_f$ = 0.37 ($SiO_2$, Hexanes : EtOAc = 1:1); HRMS (QToF): Calcd for $C_{43}H_{51}Cl_3NO_{13}PNa$ $[M+Na]^+$ 948.2061; found 948.2087.

## 7.3.3. Synthesis of 66

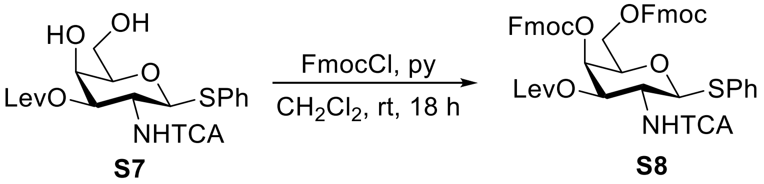
**Phenyl 2-deoxy-2-*N*-trichloroacetylamino-3-O-levulinoyl-1-thio-β-galactopyranoside, S7**



**S6** was obtained following previously establish procedures.[184]

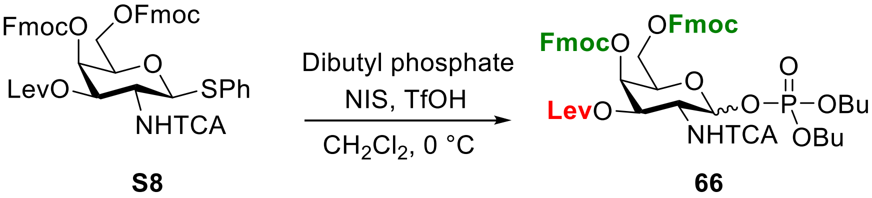
*p*-Toluenesulfonic acid (0.21 g, 1.1 mmol, 0.2 equiv) was added to a suspension of **S6** (3.4 g, 5.65 mmol) in MeOH (28 mL, 0.2 M) The reaction mixture was sonicated for 1.5 h at rt and monitored for completion by TLC (30% EtOAc in Hexanes). The reaction was then quenched with triethylamine until neutral. The solvent was removed and compound **S7** as obtained after purification by column chromatography ($SiO_2$, Hexanes : EtOAc = 1:1 to 0:1) as a white solid (2.1 g, 4.1 mmol, 72%). $^1H$ NMR (400 MHz, $CDCl_3$) δ 7.54 – 7.47 (m, 2H), 7.34 – 7.27 (m, 3H), 6.91 (d, *J* = 9.0 Hz, 1H), 5.16 (dd, *J* = 10.7, 3.0 Hz, 1H), 5.00 (d, *J* = 10.4 Hz, 1H), 4.29 (dd, *J* = 10.9, 9.0 Hz, 1H), 4.22 (t, *J* = 3.4 Hz, 1H), 4.01 – 3.92 (m, 1H), 3.92 – 3.83 (m, 1H), 3.68 (t, *J* = 5.3 Hz, 1H), 3.29 (d, *J* = 3.6 Hz, 1H), 3.14 (s, 1H), 2.80 – 2.73 (m, 2H), 2.66 – 2.43 (m, 3H), 2.17 (s, 3H); $^{13}C$ NMR (101 MHz, $CDCl_3$) δ 208.2, 172.4, 161.9, 132.7, 132.4, 129.3, 128.4, 92.5, 86.5, 78.2, 73.8, 67.5, 62.9, 51.2, 38.3, 30.0, 28.3. $[\alpha]_D^{20}$ = +46.85; IR (neat) $\nu_{max}$= 3335, 2941, 1697, 1526, 1480, 1366, 1275, 1148, 1067, 819, 741, 690; $R_f$ = 0.1 ($SiO_2$, Hexanes : EtOAc = 1:1); HRMS (QToF): Calcd for $C_{19}H_{22}Cl_3NO_7SNa$ $[M+Na]^+$ 536.0075; found 536.0075.

### Phenyl 4,6-di-*O*-(9-fluorenylmethoxycarbonyl)-2-deoxy-2-*N*-trichloroacetylamino-3-*O*-levulinoyl-1-thio-β-galactopyranoside, S8



Pyridine (5 mL, 61.2 mmol, 15 equiv) and FmocCl (3.1 g, 12.2 mmol, 3 equiv) were added to a solution of **S7** (2.1 g, 4.1 mmol) in $CH_2Cl_2$ (40 mL, 0.1 M). The reaction mixture was stirred overnight at rt and monitored for completion by TLC (100% EtOAc). The solvent was removed and compound **S8** was obtained after purification by column chromatography ($SiO_2$, Hexanes : EtOAc = 1:1 to 0:1, containing 10% $CH_2Cl_2$) as a yellow solid (3.3 g, 84%). $^1$H NMR (400 MHz, $CDCl_3$) δ 7.77 (t, *J* = 7.0 Hz, 4H), 7.66 – 7.55 (m, 6H), 7.45 – 7.26 (m, 12H), 6.86 (d, *J* = 8.7 Hz, 1H), 5.41 (dd, *J* = 10.8, 3.2 Hz, 1H), 5.30 (d, *J* = 3.4 Hz, 1H), 5.08 (d, *J* = 10.4 Hz, 1H), 4.48 – 4.35 (m, 5H), 4.34 – 4.17 (m, 4H), 4.09 – 4.04 (m, 1H), 2.69 – 2.29 (m, 4H), 2.03 (s, 3H)*;* $^{13}$C NMR (101 MHz, $CDCl_3$) δ 206.3, 172.1, 161.9, 154.8, 154.8, 143.4, 143.3, 143.3, 143.1, 141.4, 141.4, 133.0, 132.1, 129.2, 128.6, 128.2, 128.1, 128.1, 127.5, 127.5, 127.4, 127.3, 125.5, 125.3, 125.3, 120.3, 120.2, 120.2, 92.3, 86.4, 74.6, 71.2, 70.7, 70.4, 65.4, 51.5, 46.8, 46.7, 37.7, 29.8, 29.7, 27.9; $[\alpha]_D^{20}$ = + 0.63; IR (neat) $\nu_{max}$= 2385, 2359, 2344, 1748, 1525, 1450, 1241, 1147, 819, 783, 759, 738; $R_f$ = 0.57 ($SiO_2$, Hexanes : EtOAc = 1:1); HRMS (QToF): Calcd for $C_{49}H_{42}Cl_3NO_{11}SNa$ $[M+Na]^+$ 980.1436; found 980.1507.

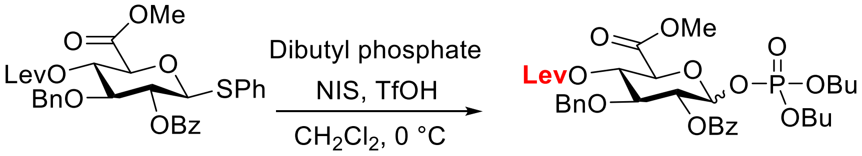
### Dibutoxyphosphoryloxy 4,6-di-*O*-(9-fluorenylmethoxycarbonyl)-2-deoxy-2-*N*-trichloroacetylamino-3-*O*-levulinoyl-α/β-galactopyranoside, 66



An oven dried round bottom flask containing a solution of thioglycoside **S8** (1.5 g, 1.56 mmol) and dibutyl hydrogen phosphate (0.62 mL, 3.13 mmol, 2.0 equiv) in anhydrous $CH_2Cl_2$ (15 mL) was cooled to 0 °C under Ar atmosphere. After 15 min, *N*-iodosuccinimide (525 mg, 2.35 mmol, 1.5 equiv.) was added followed by the dropwise addition of TfOH (15 µL, 0.16 mmol, 0.1 equiv.) at 0 °C. The reaction progress was checked every 30 min until the starting material was fully consumed after 1 h. The reaction mixture was diluted with $CH_2Cl_2$ (50 mL) and a sodium thiosulfate solution (10% w/w in water, 20 mL) was added. The organic layer was then separated, washed with sat. aq. solution of $NaHCO_3$ (20 mL), dried over $Na_2SO_4$, filtered, and concentrated. **66** was obtained after purification by column chromatography ($SiO_2$, Hexanes : EtOAc = 2:1) as a white solid (1.1 g, 2.06 mmol, 66%). $^1$H NMR (400 MHz, CDCl$_3$) δ 7.80 – 7.73 (m, 6H), 7.67 (d, *J* = 6.8 Hz, 1.4H), 7.63 – 7.54 (m, 4.6H), 7.45 – 7.29 (m, 13H), 6.98 (d, *J* = 9.9 Hz, 0.4H), 5.94 – 5.84 (m, 0.4H), 5.50 (t, *J* = 7.9 Hz, 1H), 5.42 – 5.34 (m, 0.8H), 5.34 – 5.28 (m, 2H), 4.76 (t, *J* = 9.8 Hz, 0.4H), 4.62 – 4.47 (m, 1.4H), 4.48 – 4.31 (m, 10.4H), 4.24 (m, 1.4H), 4.19 – 3.98 (m, 7H), 2.69 – 2.27 (m, 6H), 2.06 (s, 1.2H), 2.04 (s, 3H), 1.74 – 1.60 (m, 9H), 1.40 (p, *J* = 8.1 Hz, 6H), 0.92 (t, *J* = 7.3 Hz, 9H); $^{13}$C NMR (151 MHz, CDCl$_3$) δ 205.9, 205.9, 172.5, 172.0, 162.6, 162.3, 154.7, 154.7, 154.7, 154.6, 143.3, 143.2, 143.2, 143.2, 143.1, 143.0, 142.9, 141.3, 141.3, 141.3, 141.2, 128.1, 128.0, 128.0, 128.0, 127.9, 127.5, 127.5, 127.4, 127.4, 127.2, 127.2, 125.4, 125.4, 125.2, 125.2, 125.1, 125.1, 120.1, 120.1, 120.1, 96.6, 96.6, 95.6, 95.6, 92.5, 92.0, 71.6, 71.0, 70.8, 70.8, 70.5, 70.4, 70.3, 70.3, 68.7, 68.6, 68.6, 68.5, 68.4, 68.4, 67.9, 65.0, 64.9, 52.4, 52.3, 49.9, 49.8, 46.6, 46.6, 46.5, 46.5, 37.6, 37.5, 32.3, 32.2, 32.2, 32.2, 32.1, 32.1, 32.0, 32.0, 29.6, 29.6, 27.8, 27.8, 22.7, 18.7, 18.6, 18.6, 18.6, 14.2, 13.6, 13.6, 13.6; $[\alpha]_D^{20}$ = + 1.35; IR (neat) $\nu_{max}$= 2960, 1751, 1719, 1529, 1451, 1385, 1247, 1026, 964, 822, 758, 739; $R_f$ = 0.5 and 0.3 (SiO$_2$, Hexanes : EtOAc = 1:1); HRMS (QToF): Calcd for $C_{51}H_{55}Cl_3NO_{15}PNa$ [M+Na]$^+$ 1080.2267; found 1080.2310.

## 7.3.4. Synthesis of 67b

**Methyl (Dibutoxyphosphoryloxy 4-O-levulinoyl-3-O-benzyl-2-O-benzoyl-α/β-glucopyranosyluronate), 67b**



**67a** **67b**

An oven dried round bottom flask containing a solution of thioglycoside **S9**[*] (1.0 g, 1.69 mmol) and dibutyl hydrogen phosphate (0.67 mL, 3.38 mmol, 2.0 equiv) in anhydrous $CH_2Cl_2$ (16 mL) was cooled to 0 °C under Ar atmosphere. After 15 min, *N*-iodosuccinimide (567 mg, 2.53 mmol, 1.5 equiv.) was added followed by the dropwise addition of TfOH (15 µL, 0.16 mmol, 0.1 equiv.) at 0 °C. The reaction progress was checked every 30 min until the starting material was fully consumed. The reaction mixture was diluted with $CH_2Cl_2$ (50 mL) and a sodium thiosulfate solution (10% w/w in water, 20 mL) was added. The organic layer was then separated, washed with a $NaHCO_3$ saturated solution (20 mL), dried over $Na_2SO_4$, filtered, and concentrated. Compound **67b** was obtained after purification by column chromatography ($SiO_2$, Hexanes : EtOAc = 2:1) as a white solid (1.0 g, 1.4 mmol, 85%). $^1$H NMR (400 MHz, $CDCl_3$) δ 8.12 (d, *J* = 7.1 Hz, 2H), 7.71 (t, *J* = 7.5 Hz, 1H), 7.57 (t, *J* = 7.8 Hz, 2H), 5.59 – 5.50 (m, 2H), 5.41 (t, *J* = 9.5 Hz, 1H), 4.79 (d, *J* = 11.7 Hz, 1H), 4.72 (d, *J* = 11.8 Hz, 1H), 4.23 (d, *J* = 9.9 Hz, 1H), 4.21 – 4.13 (m, 2H), 4.03 (t, *J* = 9.0 Hz, 1H), 3.85 (s, 3H), 2.82 (t, *J* = 6.1 Hz, 2H), 2.74 – 2.57 (m, 2H), 2.30 (s, 3H), 1.78 – 1.69 (m, 2H), 1.54 – 1.43 (m, 2H), 1.42 – 1.34 (m, 2H), 1.15 – 1.08 (m, 2H), 1.02 (t, *J* = 7.3 Hz, 3H), 0.78 (t, *J* = 7.4 Hz, 3H); $^{13}$C NMR (151 MHz, $CDCl_3$) δ 206.1, 171.5, 167.0, 164.9, 137.3, 133.6, 130.0, 130.0, 130.0, 129.2, 128.6, 128.6, 128.5, 128.4, 128.2, 128.0, 127.9, 127.9, 96.4, 96.3, 90.4, 78.7, 76.0, 74.9, 74.4, 73.1, 72.8, 72.6, 72.5, 71.2, 69.2, 68.4, 68.4, 68.2, 68.2, 53.0, 53.0, 37.8, 37.7, 37.7, 32.3, 32.3, 32.1, 32.0, 31.8, 31.8, 30.0, 29.7, 27.8, 27.8, 18.8, 18.6, 18.3, 13.7, 13.7, 13.5; $[\alpha]_D^{20}$ = + 46.43; IR (neat) $\nu_{max}$= 2962, 1722, 1454, 1365, 1267, 1151, 1028, 909, 713; $R_f$ = 0.37 and 0.25 ($SiO_2$, Hexanes : EtOAc = 1:1); HRMS (QToF): Calcd for $C_{34}H_{45}O_{13}PNa$ $[M+Na]^+$ 715.2490; found 715.2509.

[*] (purchased from GlycoUniverse)

## 7.4. Automated Glycan Assembly

### 7.4.1. General materials and method

The automated syntheses were performed on a home built synthesizer developed at the Max Planck Institute of Colloids and Interfaces. All solvents used were HPLC-grade. The solvents used for the building block, activator, TMSOTf and capping solutions were taken from an anhydrous solvent system (J.C. Meyer) and further dried with molecular sieves (4 Å) for moisture sensitive solutions. The building blocks were co-evaporated three times with toluene and dried for 1 h on high vacuum before use. Oven dried, argon flushed flasks were used to prepare all moisture sensitive solutions. Activator, capping, deprotection, acidic wash, and building block solutions were freshly prepared and kept under argon during the automation run. All yields of products obtained by AGA were calculated on the basis of resin loading. Resin loading was determined following previously established procedures.

### 7.4.2. Preparation of stock solutions

- **Building block solution**: Between 0.06 and 0.10 mmol of building block (depending on the BB, see Module C1 and C2) was dissolved in $CH_2Cl_2$ (1 mL).
- **NIS/TfOH activator solution**: 1.35 g (6.0 mmol) of recrystallized NIS was dissolved in 40 mL of a 2:1 v/v mixture of anhydrous $CH_2Cl_2$ and anhydrous dioxane. Then triflic acid (55 µL, 0.6 mmol) was added. The solution was kept at 0 °C for the duration of the automation run.
- **Fmoc deprotection solution**: A solution of 20% piperidine in DMF (v/v) was prepared.
- **Lev deprotection solution**: Hydrazine acetate (550 mg, 5.97 mmol) was dissolved in pyridine/AcOH/$H_2O$ (40 mL, v/v, 32:8:2) and sonicated for 10 min.
- **TMSOTf solution**: TMSOTf (0.45 mL, 2.49 mmol) was added to $CH_2Cl_2$ (40 mL) or for glycosyl phosphate activation; TMSOTf (0.9 mL, 5.0 mmol) was added to $CH_2Cl_2$ (40 mL).
- **Capping solution**: A solution of 10% acetic anhydride and 2% methanesulfonic acid in $CH_2Cl_2$ (v/v) was prepared.

## 7.4.3. Modules for automated synthesis

**Module A: Resin preparation for synthesis (20 min)**

All automated syntheses were performed on 0.0135 mmol scale. Resin (**L1**, 45 mg or **L2**, 35 mg) was placed in the reaction vessel and swollen in $CH_2Cl_2$ for 20 min at rt prior to synthesis. During this time, all reagent lines needed for the synthesis were washed and primed. After the swelling, the resin was washed with DMF, THF, and $CH_2Cl_2$ (three times each with 2 mL for 25 s).

**Module B: Acidic wash with TMSOTf solution (20 min)**

The resin was swollen in 2 mL $CH_2Cl_2$ and the temperature of the reaction vessel was adjusted to -20 °C. Upon reaching the low temperature, TMSOTf solution (1 mL) was added drop wise to the reaction vessel. After bubbling for 3 min, the acidic solution was drained and the resin was washed with 2 mL $CH_2Cl_2$ for 25 s.

| Action | Cycles | Solution | Amount | T (°C) | Incubation time |
|---|---|---|---|---|---|
| Cooling | - | - | - | -20 | (15 min)* |
| Deliver | 1 | $CH_2Cl_2$ | 2 mL | -20 | - |
| Deliver | 1 | TMSOTf solution | 1 mL | -20 | 3 min |
| Wash | 1 | $CH_2Cl_2$ | 2 mL | -20 | 25 sec |

*Time required to reach the desired temperature.

**Module C1: Thioglycoside glycosylation (35 - 55 min)**

The building block solution was delivered to the reaction vessel. After the set temperature was reached, the reaction was started by dropwise addition of the NIS/TfOH activator solution (1.0 mL). After completion of the reaction, the solution was drained and the resin was washed with $CH_2Cl_2$ , $CH_2Cl_2$:dioxane (1:2, 3 mL for 20 s) and $CH_2Cl_2$ (two times, each with 2 mL for 25 s). The temperature of the reaction vessel was increased to 25 °C for the next module.

| Action | Cycles | Solution | Amount | T (°C) | Incubation time |
|---|---|---|---|---|---|
| Cooling | - | - | - | -20 | - |
| Deliver | 1 | BB solution | 1 mL | -20 | - |
| Deliver | 1 | NIS/TfOH activator solution | 1 mL | -20 | - |
| Reaction time | 1 | | | -20 | 5 min |
| | | | | to 0 | 20 min |
| Wash | 1 | $CH_2Cl_2$ | 2 mL | 0 | 5 sec |
| Wash | 1 | $CH_2Cl_2$: Dioxane (1:2) | 2 mL | 0 | 20 sec |
| Heating | - | - | - | 25 | - |
| Wash | 2 | $CH_2Cl_2$ | 2 mL | > 0 | 25 sec |

**Module C2: Glycosyl phosphate glycosylation (45 min)**

The building block solution (0.06 mmol of BB in 1 mL of $CH_2Cl_2$ per glycosylation) was delivered to the reaction vessel. After the set temperature was reached, the reaction was started by drop wise addition of the TMSOTf solution (1.0 mL, same equiv). After completion of the reaction, the solution was drained and the resin washed with $CH_2Cl_2$ (six times, each with 2 mL for 25 s). The temperature of the reaction vessel was increased to 25 °C for the next module.

| Action | Cycles | Solution | Amount | T (°C) | Incubation time |
|---|---|---|---|---|---|
| Cooling | - | - | - | -30 | - |
| Deliver | 1 | BB solution | 1 mL | -30 | - |
| Deliver | 1 | TMSOTf solution | 1 mL | -30 | - |
| Reaction time | 1 | | | -30 | 5 min |
| | | | | to -10 | 40 min |
| Wash | 1 | $CH_2Cl_2$ | 2 mL | -10 | 5 sec |
| Heating | - | - | - | 25 | - |
| Wash | 6 | $CH_2Cl_2$ | 2 mL | > 0 | 25 sec |

**Module D: Capping (30 min)**

The resin was washed with DMF (two times with 2 mL for 25 s) and the temperature of the reaction vessel was adjusted to 25 °C. 2 mL of Pyridine solution (10% in DMF) was delivered into the reaction vessel. After 1 min, the reaction solution was drained and the resin washed with $CH_2Cl_2$ (three times with 3 mL for 25 s). 4 mL of capping solution was delivered into the reaction vessel. After 20 min, the reaction solution was drained and the resin washed with $CH_2Cl_2$ (three times with 3 mL for 25 s).

| Action | Cycles | Solution | Amount | T (°C) | Incubation time |
|---|---|---|---|---|---|
| Heating | - | - | - | 25 | (5 min)* |
| Wash | 2 | DMF | 2 mL | 25 | 25 sec |
| Deliver | 1 | 10% Pyridine in DMF | 2 mL | 25 | 1 min |
| Wash | 3 | $CH_2Cl_2$ | 2 mL | 25 | 25 sec |
| Deliver | 1 | Capping Solution | 4 mL | 25 | 20 min |
| Wash | 3 | $CH_2Cl_2$ | 2 mL | 25 | 25 sec |

*Time required to reach the desired temperature.

**Module E1: Fmoc deprotection (9 min)**

The resin was washed with DMF (three times with 2 mL for 25 s) and the temperature of the reaction vessel was adjusted to 25 °C. 2 mL of Fmoc deprotection solution was delivered to the reaction vessel and kept under Ar bubbling. After 5 min, the reaction solution was drained and the resin washed with DMF (three times with 3 mL for 25 s) and $CH_2Cl_2$ (five times each with 2 mL for 25 s). The temperature of the reaction vessel was decreased to -20 °C for the next module.

| Action | Cycles | Solution | Amount | T (°C) | Incubation time |
|---|---|---|---|---|---|
| Wash | 3 | DMF | 2 mL | 25 | 25 sec |
| Deliver | 1 | Fmoc depr. solution | 2 mL | 25 | 5 min |
| Wash | 1 | DMF | 2 mL | | |
| Cooling | - | - | - | -20 | - |
| Wash | 3 | DMF | 2 mL | < 25 | 25 sec |
| Wash | 5 | $CH_2Cl_2$ | 2 mL | < 25 | 25 sec |

**Module E2: Lev deprotection (65 min)**

The resin was washed with $CH_2Cl_2$ (three times with 2 mL for 25 s). $CH_2Cl_2$ (1.3 mL) was delivered to the reaction vessel and the temperature of the reaction vessel was adjusted to 25 °C. 2 mL of Lev deprotection solution was delivered to the reaction vessel that was kept under pulsed Ar bubbling for 30 min. This procedure was repeated twice. The reaction solution was drained and the resin washed with DMF (three times with 3 mL for 25 s) and $CH_2Cl_2$ (five times each with 2 mL for 25 s).

| Action | Cycles | Solution | Amount | T (°C) | Incubation time |
|---|---|---|---|---|---|
| Wash | 3 | DMF | 2 mL | 25 | 25 sec |
| Deliver | 2 | Lev depr. solution | 2 mL | 25 | 30 min |
| Wash | 1 | DMF | 2 mL | | |
| Cooling | - | - | - | -20 | - |
| Wash | 3 | DMF | 2 mL | < 25 | 25 sec |
| Wash | 5 | $CH_2Cl_2$ | 2 mL | < 25 | 25 sec |

### 7.4.4. Post-synthesizer manipulations (Post-AGA)

**Module F: On-resin sulfation**

The resin was suspended in 4 mL of a 0.5 M $SO_3$·py solution (DMF/pyridine, 1:1). The reaction was rotated for 12 h at 40 °C, after which time the resin was repeatedly washed with DMF (5 x 4 mL), MeOH (5 x 4 mL) and $CH_2Cl_2$ (5 x 4 mL).

**Module G: On-resin hydrolysis**

The resin was suspended in THF:MeOH (4:1, 4 mL) and a solution of LiOH in water (150 µL, 1 M) was added. The mixture was gently shaken at rt. After microcleavage (see Module G1) indicated the complete hydrolysis of all ester groups, the resin was repeatedly washed with MeOH (5 x 4 mL) and $CH_2Cl_2$ (5 x 4 mL). The reaction time is variable and it is indicated for each synthesis.

**Module G1: On-resin acetylation**

The resin was suspended in a 4 mL solution of acetic anhydride in DMF (15% v/v) and the mixture gently shaken at rt for 3 h, after which time the resin was repeatedly washed with DMF (5 x 4 mL), MeOH (5 x 4 mL) and CH2Cl2 (5 x 4 mL).

**Module H: Cleavage from solid support**

The oligosaccharides were cleaved from the solid support using a continuous-flow photoreactor as described previously. A 20% MeOH in $CH_2Cl_2$ solvent system was used due to the presence of sulfate groups in the glycan.

**Module H1: Micro-cleavage from solid support**

Trace amount of resin (around 20 beads) was dispersed in $CH_2Cl_2$ (0.1 mL) and irradiated with a UV lamp (6 watt, 356 nm) for 20 min. ACN was then added to the resin and the resulting solution analyzed by MS-Q-TOF or MALDI.

**Module I: Hydrogenolysis at ambient pressure**[a]

The crude compound obtained from Module H was dissolved in 2 mL of *t*-BuOH:$H_2O$ (1:1). The Pd catalyst (2.5 times the weight of the starting material) was added and the reaction was stirred in a flask equipped with a $H_2$ balloon. The reaction progress was monitored to avoid undesired side products formation. Upon completion, the reaction was filtered and washed with *t*-BuOH and $H_2O$. The filtrates were concentrated *in vacuo.*

**Module I1: Hydrogenolysis**[a]

The crude compound obtained from Module H was dissolved in 2 mL of *t*-BuOH:$H_2O$ (1:1). Pd catalyst (2.5 times the weight of starting material) was added and the reaction was stirred in a high pressure reactor (60 psi $H_2$). The reaction progress was monitored to avoid undesired side products formation. Upon completion, the reaction was filtered and washed with *t*-BuOH and $H_2O$. The filtrates were concentrated *in vacuo.* [b]

[a]Reaction times and type of catalyst are indicated for each synthesis.

**[b]**Upon completion of hydrogenolysis, prior to filtration, the crude mixtures of compounds containing GlcNAc or GalNAc were treated with thiourea (10 equiv.).

### Module J: Purification

The final compounds after global deprotection were purified by **Method $B_1$ or Method $C_2$** followed by **Method $A_1$** and analyzed using analytical HPLC (Agilent 1200 Series spectrometer, **Method $C_1$**).

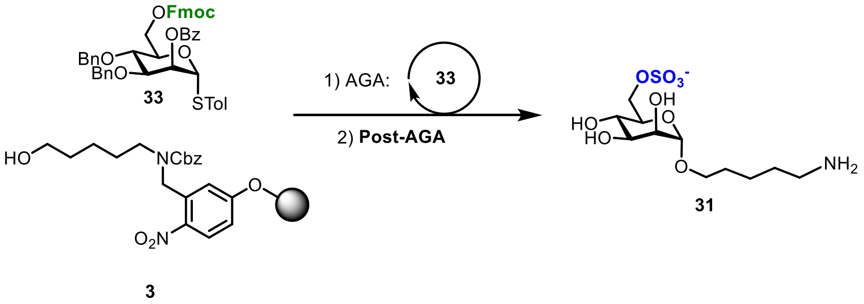
- **Method $A_1$:** Sephadex® LH-20 column with $H_20$:MeOH (1:1) as eluent, isocratic.
- **Method $B_1$:** (Manual reverse phase $C_{18}$ silica gel column chromatography): $H_2O$ (10 mL), 5% MeOH (10 mL), 7.5% MeOH (10 mL), 10% MeOH (10 mL), 15% MeOH (10 mL), 20% MeOH (10 mL).
- **Method $C_1$:** (Hypercarb column, 150 x 4.60 mm) flow rate of 1.0 mL / min. 0 to 70% of B in 30 min (A = 0.01 M $NH_4HCO_3$, B = ACN); ELSD Detector: 45 °C
- **Method $C_2$:** (Hypercarb column, 150 x 10 mm) flow rate of 3.5 mL / min. 0 to 70% of B in 30 min (A = 0.01 M $NH_4HCO_3$, B = ACN); ELSD Detector: 45 °C
- **Method $C_3$:** (Hypercarb column, 150 x 10 mm) flow rate of 3.5 mL / min. 0 to 60% of B in 40 min (A = 0.01 M $NH_4HCO_3$, B = ACN); ELSD Detector: 45 °C

### Module K: Ion exchange

The final purified compounds were passed through an Amberlite resin-$Na^+$ bed (2 cm diameter x 10 cm length, pre-swollen in water, eluent system: water).

## 7.5. Oligosaccharides synthesis

### 7.5.1. Synthesis of 31

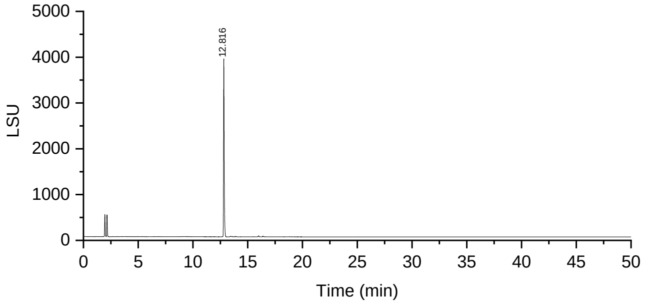


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **33** | **B, C1, D, E1** | **C1:** (**33**, -20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

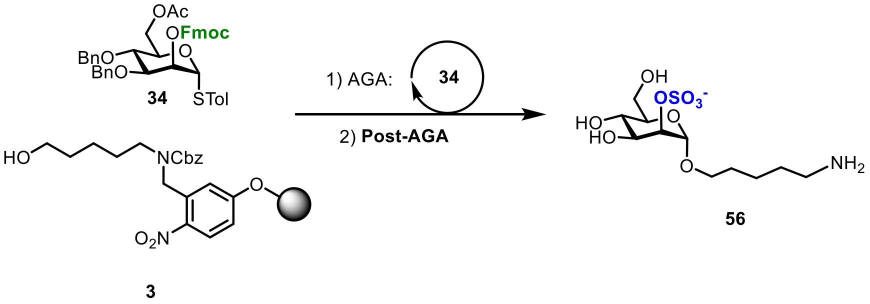
Compound 31 was obtained as a white solid (3.2 mg, 70% overall yield).

Analytical data for 31: $^1$H NMR (400 MHz, D$_2$O) δ 4.81 (d, *J* = 1.8 Hz, 1H), 4.34 (dd, *J* = 11.3, 2.1 Hz, 1H), 4.12 (dd, *J* = 11.3, 7.0 Hz, 1H), 3.89 (dd, *J* = 3.4, 1.7 Hz, 1H), 3.82 (m, 1H), 3.78 – 3.69 (m, 2H), 3.66 – 3.52 (m, 2H), 2.95 (t, *J* = 7.6 Hz, 2H), 1.64 (m, 4H), 1.49 – 1.35 (m, 2H); $^{13}$C NMR (101 MHz, D$_2$O) δ 99.7, 70.8, 70.4, 69.8, 67.5, 66.4, 39.3, 27.8, 26.3, 22.2; HRMS (QToF): Calcd for $C_{11}H_{22}NO_9S$ [M]$^-$ 344.1021; found 344.1028.

**RP-HPLC of 31 (ELSD trace, Method C$_1$, t$_R$= 12.82 min)**



## 7.5.2. Synthesis of 56

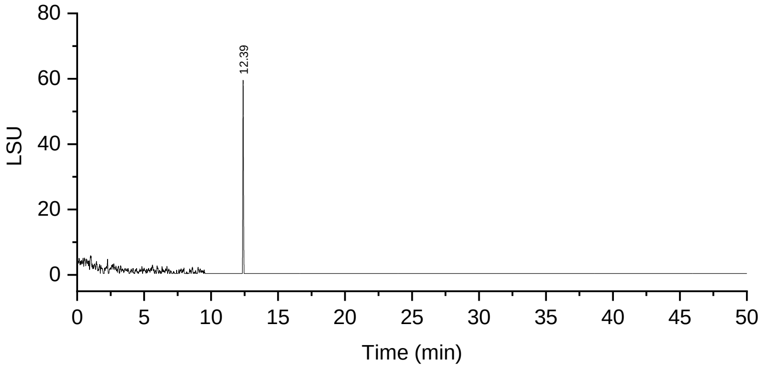


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **34** | **B, C1, D, E1** | **C1:** (**34, -**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

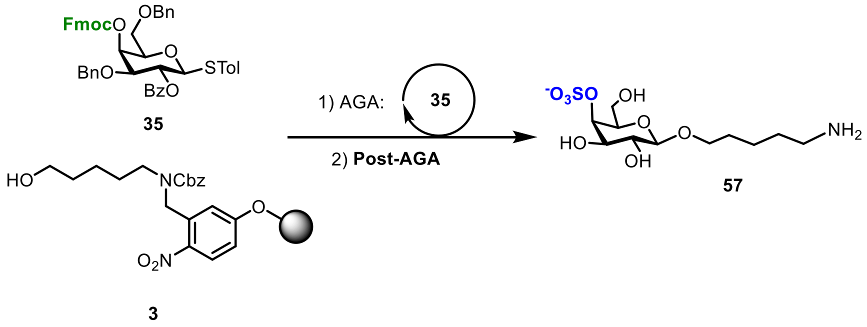
Compound **56** was obtained as a yellowish solid (1.4 mg, 30% overall yield).

Analytical data for **56**: $^{1}$H NMR (400 MHz, $D_2O$) δ 5.10 (d, *J* = 1.8 Hz, 1H), 4.47 (dd, *J* = 3.6, 1.7 Hz, 1H), 3.92 – 3.79 (m, 2H), 3.76 – 3.67 (m, 2H), 3.65 – 3.51 (m, 3H), 3.01 – 2.92 (m, 2H), 1.64 (m, 4H), 1.50 – 1.36 (m, 2H); $^{13}$C NMR (101 MHz, $D_2O$) δ 97.1, 76.9, 72.6, 69.0, 67.5, 66.5, 60.6, 39.2, 27.7, 26.4, 22.2; HRMS (QToF): Calcd for $C_{11}H_{22}NO_9S$ $[M]^-$ 344.1021; found 344.1013.

**RP-HPLC of 56 (ELSD trace, Method $C_1$, $t_R$= 12.39 min)**



## 7.5.3. Synthesis of 57

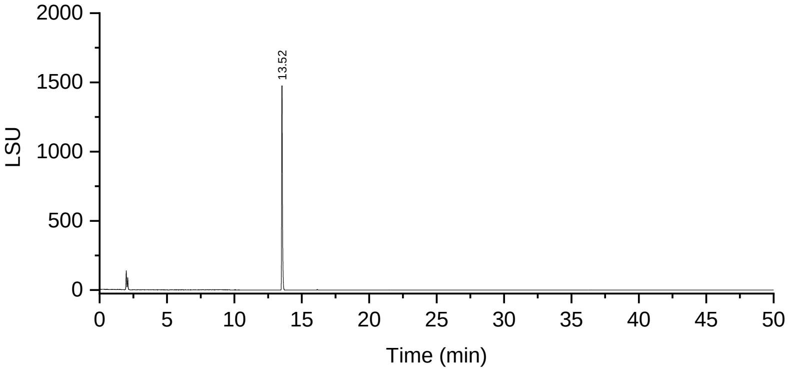


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **35** | **B, C1, D, E1** | **C1:** (**35**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

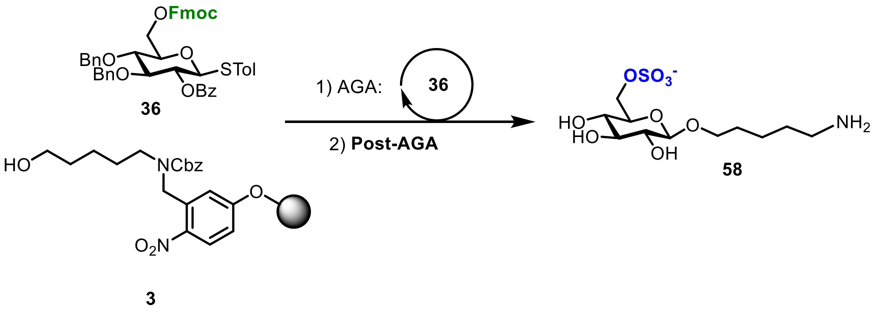
Compound **57** was obtained as a white solid (1.4 mg, 30% overall yield).

Analytical data for **57**: $^1$H NMR (400 MHz, $D_2O$) δ 4.61 (d, *J* = 3.3 Hz, 1H), 4.39 (dd, *J* = 7.9, 3.3 Hz, 1H), 3.90 (dt, *J* = 10.0, 6.5 Hz, 1H), 3.82 – 3.60 (m, 7H), 3.51 – 3.44 (m, 1H), 2.96 (dd, *J* = 9.0, 6.0 Hz, 2H), 1.68 – 1.58 (m, 4H), 1.45 – 1.39 (m, 2H); $^{13}$C NMR (101 MHz, $D_2O$) δ 102.5, 76.4, 74.2, 71.7, 70.7, 69.9, 60.8, 39.2, 28.0, 26.3, 22.0; HRMS (QToF): Calcd for $C_{11}H_{22}NO_9S$ $[M]^-$ 344.1021; found 344.1020.

**P-HPLC of 57 (ELSD trace, Method $C_1$, $t_R$= 13.52 min)**



## 7.5.4. Synthesis of 58

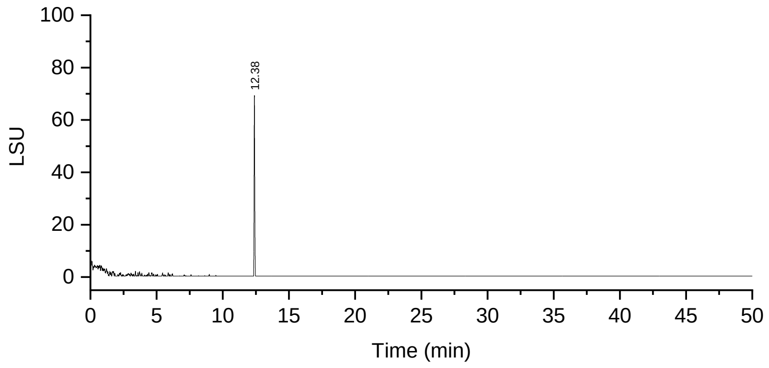


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **36** | **B, C1, D, E1** | **C1:** (**36**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

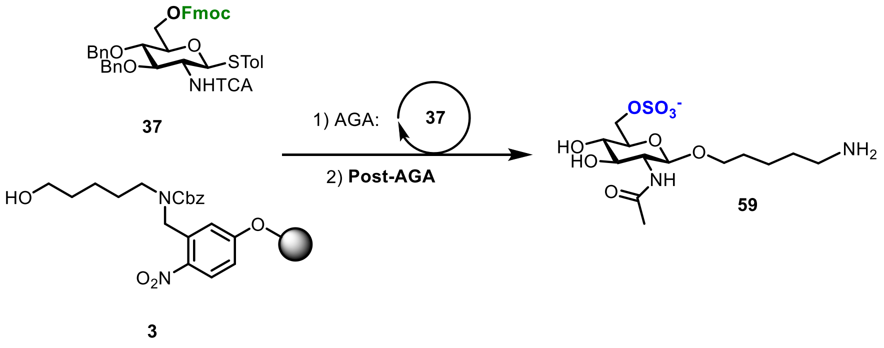
Compound **58** was obtained as a white solid (3 mg, 65% overall yield).

Analytical data for **58**: $^1$H NMR (400 MHz, $D_2O$) δ 4.42 (d, *J* = 8.0 Hz, 1H), 4.29 (dd, *J* = 11.2, 2.1 Hz, 1H), 4.14 (dd, *J* = 11.2, 5.6 Hz, 1H), 3.86 (dt, *J* = 10.2, 6.5 Hz, 1H), 3.75 – 3.58 (m, 2H), 3.51 – 3.35 (m, 2H), 3.23 (dd, *J* = 9.0, 8.0 Hz, 1H), 2.95 (t, *J* = 7.5 Hz, 2H), 1.75 – 1.54 (m, 4H), 1.51 – 1.33 (m, 2H).*;* $^{13}$C NMR (101 MHz, $D_2O$) δ 102.9, 76.2, 74.2, 73.6, 70.9, 69.8, 67.6, 39.9, 28.8, 26.9, 22.5*;* HRMS (QToF): Calcd for $C_{11}H_{22}NO_9S$ $[M]^-$ 344.1021; found 344.1016.

**RP-HPLC of 58 (ELSD trace, Method C1, $t_R$= 12.38 min)**



### 7.5.5. Synthesis of 59



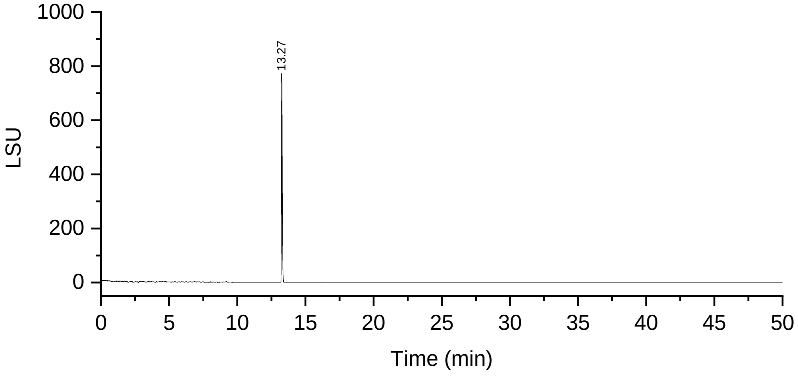
| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **37** | **B, C1, D, E1** | **C1:** (**37**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrogenolysis** | **I1**$^{a}$ | **I1:** 10-20% $Pd(OH)_2/C$ (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

$^{a}$ treated with thiourea upon completion of hydrogenolysis

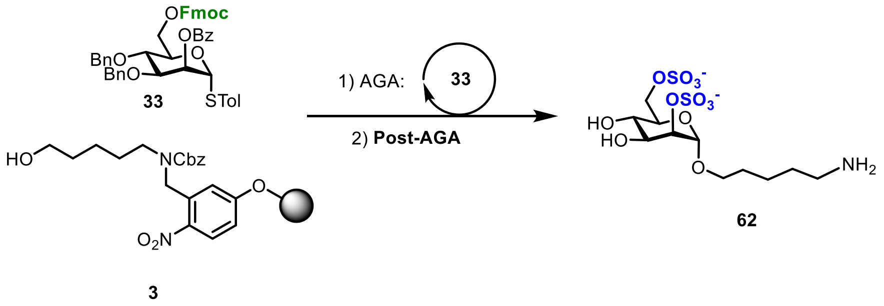
Compound **59** was obtained as a white solid (2 mg, 38% overall yield).

Analytical data for **59**: $^{1}$H NMR (600 MHz, $D_2O$) δ 4.54 (d, *J* = 8.5 Hz, 1H), 4.37 (dd, *J* = 11.3, 2.1 Hz, 1H), 4.22 (dd, *J* = 11.2, 5.8 Hz, 1H), 3.89 (dt, *J* = 10.3, 6.2 Hz, 1H), 3.74 – 3.60 (m, 4H), 3.60 – 3.47 (m, 3H), 3.01 (t, *J* = 7.6 Hz, 2H), 2.05 (s, 3H), 1.69 (m, 2H), 1.63 (m, 2H), 1.43 (m, 2H); $^{13}$C NMR (151 MHz, $D_2O$) δ 174.5, 101.2, 73.7, 73.6, 70.2, 69.6, 67.1, 55.5, 39.3, 28.1, 26.2, 22.1, 22.0; HRMS (QToF): Calcd for $C_{13}H_{25}N_2O_9S$ $[M]^-$ 385.1286; found 385.1288.

**RP-HPLC of 59 (ELSD trace, Method C1, $t_R$= 13.3 min)**



### 7.5.6. Synthesis of 62

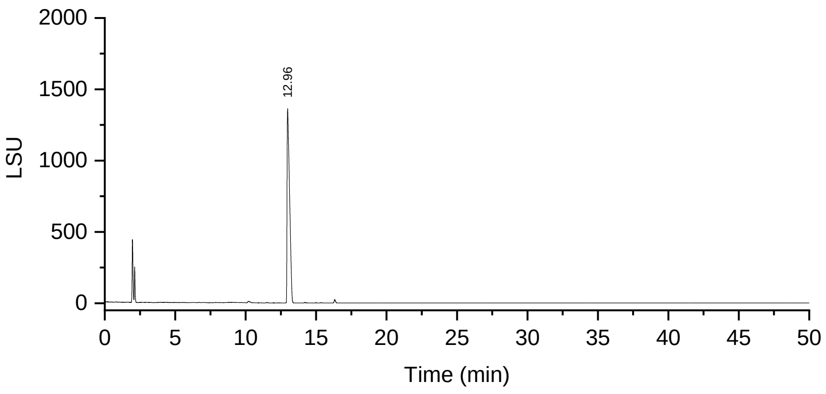


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **33** | **B, C1, D, E1** | **C1:** (**33**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Sulfation** | **F** | |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

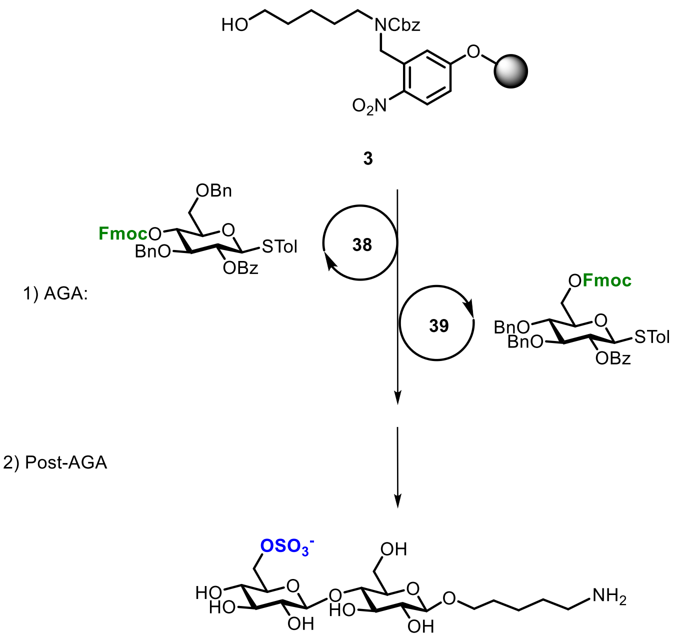
Compound **62** was obtained as yellowish solid (2.7 mg, 49% overall yield).

Analytical data for **62**: $^1$H NMR (400 MHz, $D_2O$) δ 5.09 (d, *J* = 1.7 Hz, 1H), 4.47 (dd, *J* = 3.5, 1.7 Hz, 1H), 4.34 (dd, *J* = 11.3, 2.1 Hz, 1H), 4.13 (dd, *J* = 11.4, 7.0 Hz, 1H), 3.92 – 3.82 (m, 2H), 3.73 (m, 1H), 3.65 – 3.53 (m, 2H), 2.96 (t, *J* = 7.6 Hz, 2H), 1.64 (m, 4H), 1.42 (m, 2H); $^{13}$C NMR (101 MHz, $D_2O$) δ 97.3, 76.7, 70.8, 69.0, 67.7, 67.4, 66.4, 39.3, 27.7, 26.2, 22.2; HRMS (QToF): Calcd for $C_{11}H_{22}NO_{12}S_2$ $[M+H]^-$ 424.0589; found 424.0598.

**RP-HPLC of 62 (ELSD trace, Method $C_1$, $t_R$= 13.0 min)**



## 7.5.7. Synthesis of 60



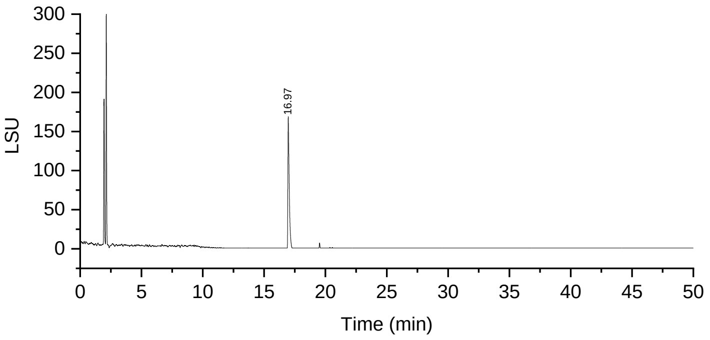
**60**

| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **38** | **B, C1, D, E1** | **C1:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C1, D, E1** | **C:1** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

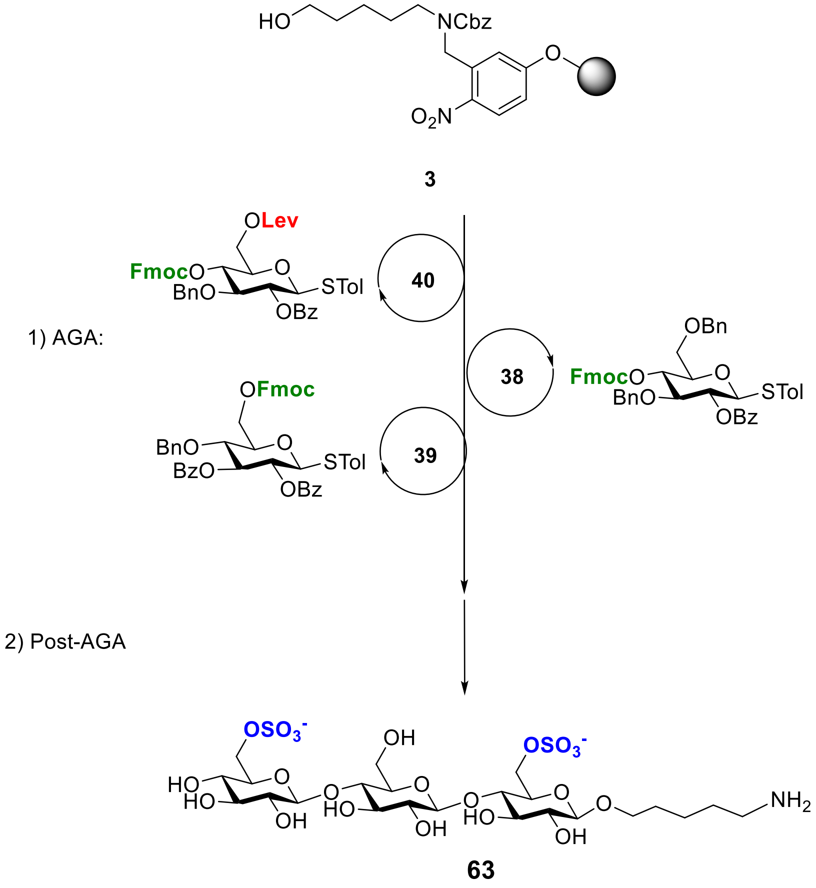
Compound **60** was obtained as a white solid (1.7 mg, 25% overall yield).

Analytical data for **60**: $^1$H NMR (400 MHz, $D_2O$) δ 4.48 (d, *J* = 7.9 Hz, 1H), 4.44 (d, *J* = 8.1 Hz, 1H), 4.30 (dd, *J* = 11.2, 2.2 Hz, 1H), 4.16 (dd, *J* = 11.2, 5.7 Hz, 1H), 3.90 (dt, *J* = 13.0, 9.4 Hz, 2H), 3.78 – 3.53 (m, 6H), 3.50 – 3.39 (m, 2H), 3.27 (td, *J* = 9.1, 7.9 Hz, 2H), 2.99 – 2.91 (m, 2H), 1.70 – 1.58 (m, 4H), 1.42 (q, *J* = 8.2 Hz, 2H); $^{13}$C NMR (101 MHz, $D_2O$) δ 102.5, 101.8, 79.3, 75.2, 74.6, 74.2, 73.6, 72.9, 72.7, 69.9, 69.0, 66.9, 60.0, 39.2, 28.0, 26.3, 22.0; HRMS (QToF): Calcd for $C_{17}H_{32}N_1O_{14}S$ $[M]^-$ 506.1549; found 506.1537.

**RP-HPLC of 60 (ELSD trace, Method C1, $t_R$= 17.0 min)**



## 7.5.8. Synthesis of 63

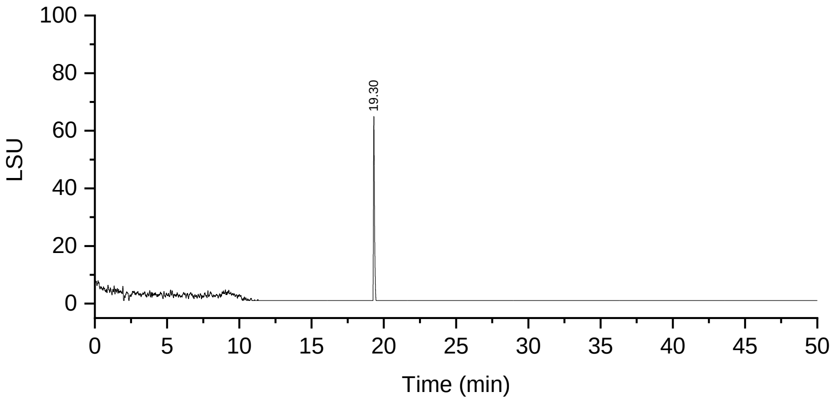


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **40** | **B, C1, D, E1** | **C1:** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **38** | **B, C1, D, E1** | **C1:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C1, D, E1, E2** | **C1:** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (24 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (20 h) |
| | **Purification** | **J(B1), J(A1), K** | |

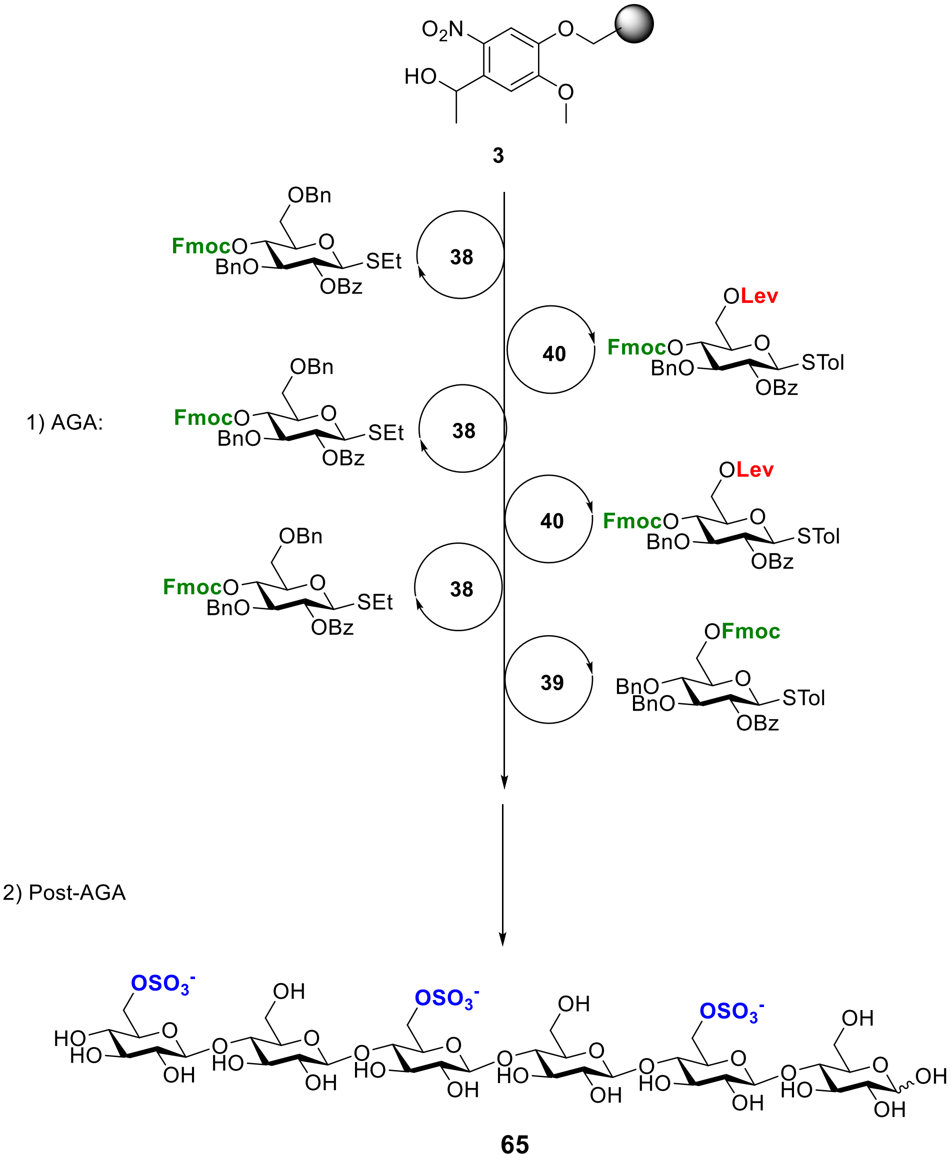
Compound **63** was obtained as a white solid (1.2 mg, 16% overall yield).

Analytical data for **63**: $^{1}$H NMR (700 MHz, $D_2O$) δ 4.59 (d, *J* = 7.9 Hz, 1H), 4.51 (dd, *J* = 11.1, 8.0 Hz, 2H), 4.40 (d, *J* = 10.9 Hz, 1H), 4.36 (d, *J* = 11.1 Hz, 1H), 4.30 (dd, *J* = 11.2, 4.5 Hz, 1H), 4.22 (dd, *J* = 11.2, 5.4 Hz, 1H), 3.99 (d, *J* = 12.2 Hz, 1H), 3.92 (dd, *J* = 10.7, 6.2 Hz, 1H), 3.86 – 3.60 (m, 9H), 3.55 – 3.45 (m, 2H), 3.34 (p, *J* = 9.2 Hz, 3H), 3.01 (t, *J* = 7.5 Hz, 2H), 1.69 (dp, *J* = 14.3, 7.0 Hz, 4H), 1.47 (dq, *J* = 15.6, 7.0 Hz, 2H); $^{13}$C NMR (176 MHz, $D_2O$) δ 102.7, 102.1, 101.8, 79.0, 77.4, 75.2, 74.7, 74.2, 74.0, 73.7, 73.0, 72.9, 72.8, 72.5, 70.3, 69.1, 67.0, 66.3, 59.9, 39.4, 28.2, 26.3, 22.0; HRMS (QToF): Calcd for $C_{23}H_{41}N_{1}O_{22}S_{2}$ $[M]^{2-}$ 373.5786; found 373.5791.

**RP-HPLC of 63 (ELSD trace, Method C1, $t_R$= 19.3 min)**



## 7.5.9. Synthesis of 65

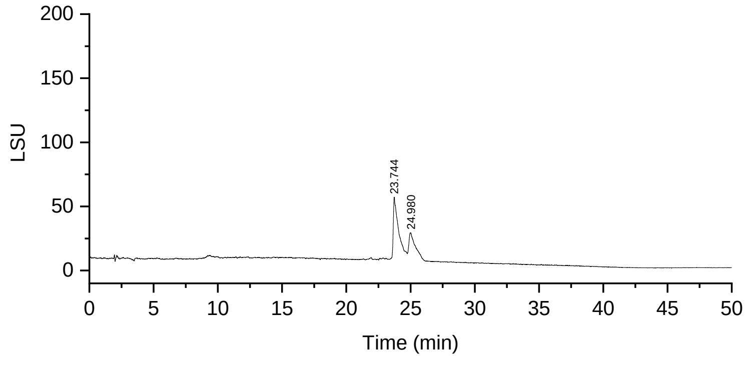


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **38** | **B, C1, D, E1** | **C1:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **40** | **B, C1, D, E1** | **C:1** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **38** | **B, C1, D, E1** | **C1:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **40** | **B, C1, D, E1** | **C1:** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **38** | **B, C1, D, E1** | **C1:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C1, D, E1, E2** | **C1:** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (72 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (48 h) |
| | **Purification** | **J(B1), J(A1), K** | |

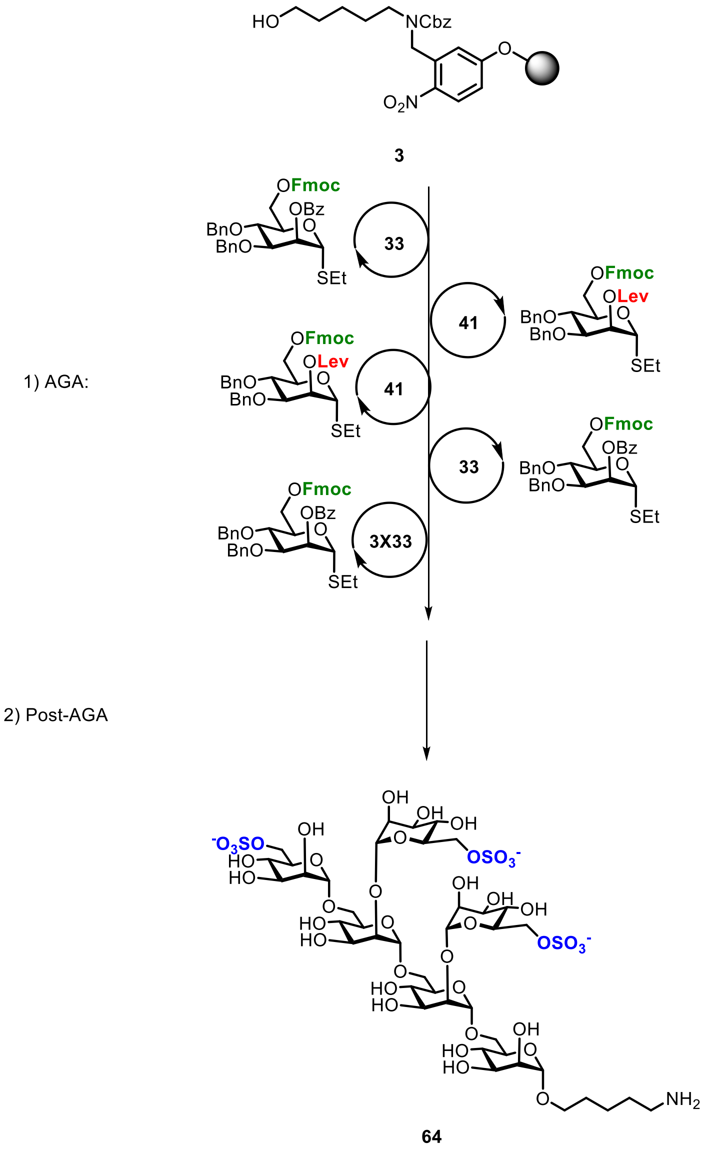
Compound **65** was obtained as a white solid (10 mg, 60% overall yield).

Analytical data for **65**: $^1$H NMR (700 MHz, $D_2O$) δ 5.17 (d, *J* = 3.8 Hz, 1H), 4.61 (d, *J* = 8.0 Hz, 1H), 4.56 – 4.45 (m, 5H), 4.39 – 4.24 (m, 5H), 4.17 (dd, *J* = 11.2, 5.3 Hz, 1H), 3.92 (dd, *J* = 16.0, 12.8 Hz, 4H), 3.83 – 3.73 (m, 6H), 3.73 – 3.67 (m, 2H), 3.67 – 3.52 (m, 10H), 3.50 – 3.40 (m, 3H), 3.37 – 3.21 (m, 4H); $^{13}$C NMR (151 MHz, $D_2O$) δ 102.7, 102.4, 101.8, 101.7, 95.6, 79.2, 79.0, 78.9, 78.6, 77.1, 77.0, 75.1, 74.6, 74.1, 73.9, 73.9, 73.7, 73.6, 72.8, 72.8, 72.7, 72.7, 72.4, 71.2, 69.9, 69.0, 66.9, 66.1, 60.0, 59.8; HRMS (QToF): Calcd for $C_{36}H_{59}O_{40}S_3$ $[M]^{3-}$ 409.0587; found 409.0597.

**RP-HPLC of 65 (ELSD trace, Method C1, $t_R$= 23.70 min and 25.0 min)**



## 7.5.10. Synthesis of 64

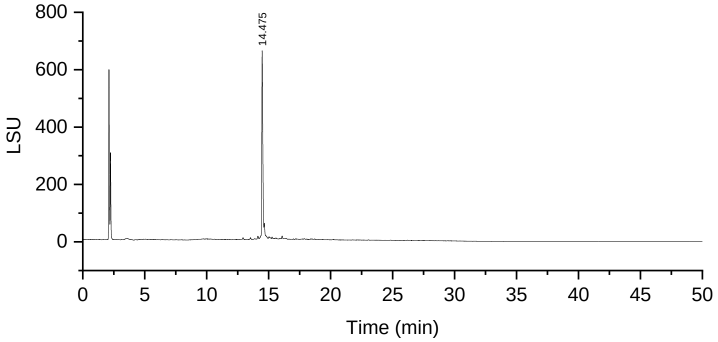


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **33** | **B, C1, D, E1** | **C1:** (**33**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **41** | **B, C1, D, E1** | **C1:** (**41**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **41** | **B, C1, D, E1** | **C1:** (**41**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **33** | **B, C1, D, E2** | **C1:** (**33**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **33** | **B, C1(x3), D, E1** | **C1:** (**33**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

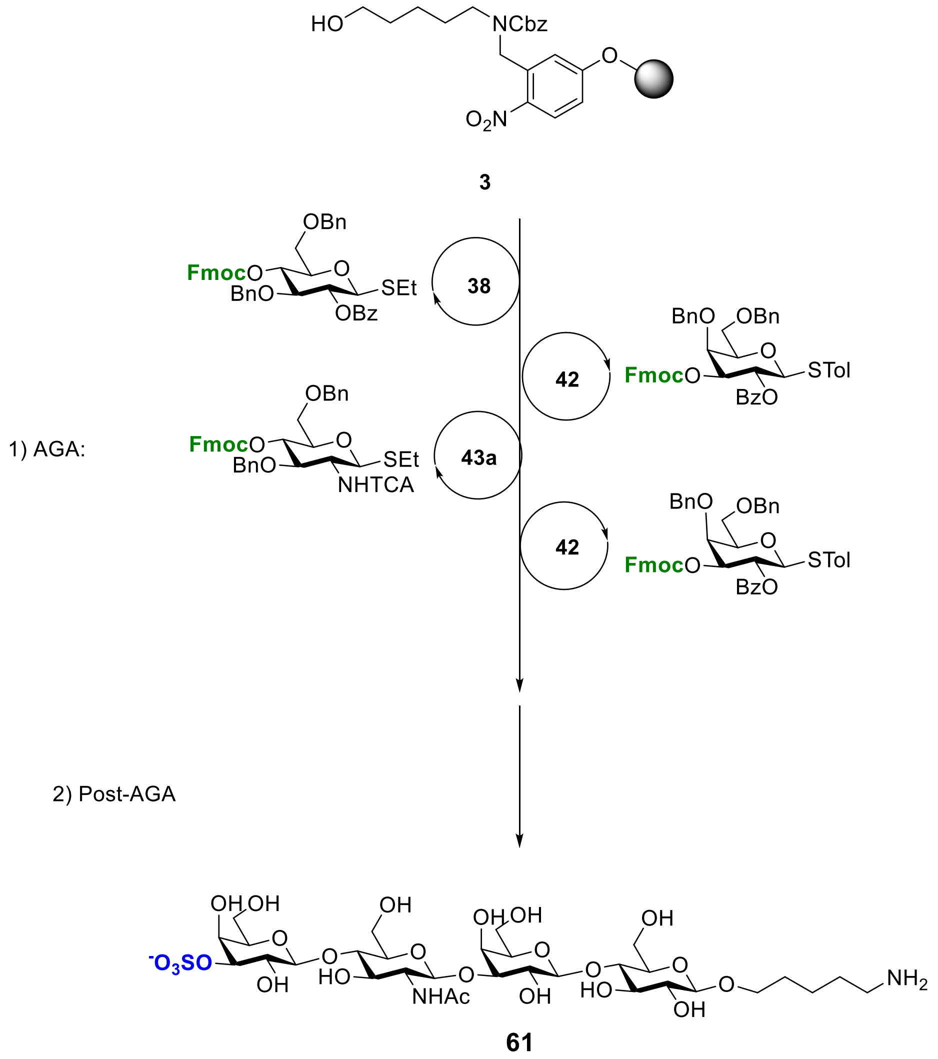
Compound **64** was obtained as a white solid (4.8 mg, 30% overall yield).

Analytical data for **64**: $^1$H NMR (600 MHz, $D_2O$) δ 5.08 (dd, *J* = 12.2, 1.8 Hz, 2H), 5.05 – 4.98 (m, 2H), 4.98 – 4.86 (m, 1H), 4.83 (d, *J* = 1.8 Hz, 1H), 4.36 – 4.17 (m, 6H), 4.11 – 3.54 (m, 32H), 2.98 (q, *J* = 7.2 Hz, 2H), 1.67 (td, *J* = 15.1, 7.4 Hz, 4H), 1.49 – 1.37 (m, 2H); $^{13}$C NMR (151 MHz, $D_2O$) δ 92.7, 92.7, 90.2, 89.9, 88.3, 88.2, 69.5, 69.0, 61.5, 61.4, 61.2, 61.1, 61.0, 60.8, 60.7, 60.6, 60.6, 60.4, 60.2, 60.2, 57.9, 57.70, 57.57, 57.1, 56.9, 56.7, 56.6, 56.5, 55.9, 55.5, 29.7, 18.3, 16.8, 12.8; HRMS (QToF): Calcd for $C_{41}H_{71}NO_{40}S_3$ $[M+H]^{2-}$ 656.6362; found 656.6385.

**RP-HPLC of 64 (ELSD trace, Method C1, $t_R$= 14.48 min)**



### 7.5.11. Synthesis of 61



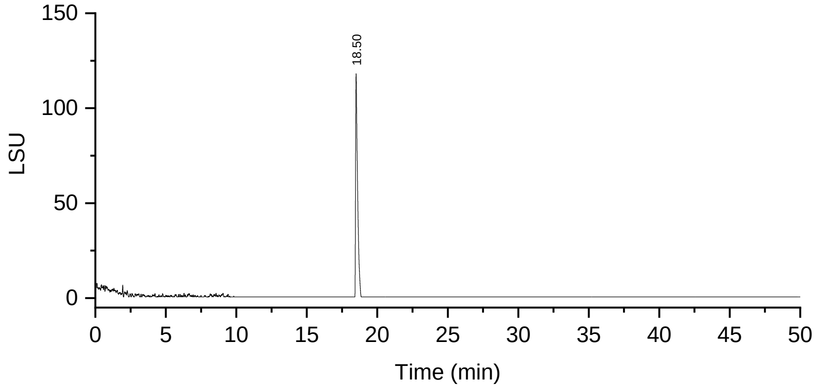
| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **38** | **B, C1(x2), D, E1** | **C1:** (**83**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **42** | **B, C1(x2), D, E1** | **C1:** (**42**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **43a** | **B, C1(x2), D, E1** | **C1:** (**43a**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **42** | **B, C1(x2), D, E1,** | **C1:** (**42**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (120 h) |
| | **Hydrogenolysis** | **I1**[a] | **I1:** 10-20% $Pd(OH)_2/C$ (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

[a] treated with thiourea upon completion of hydrogenolysis

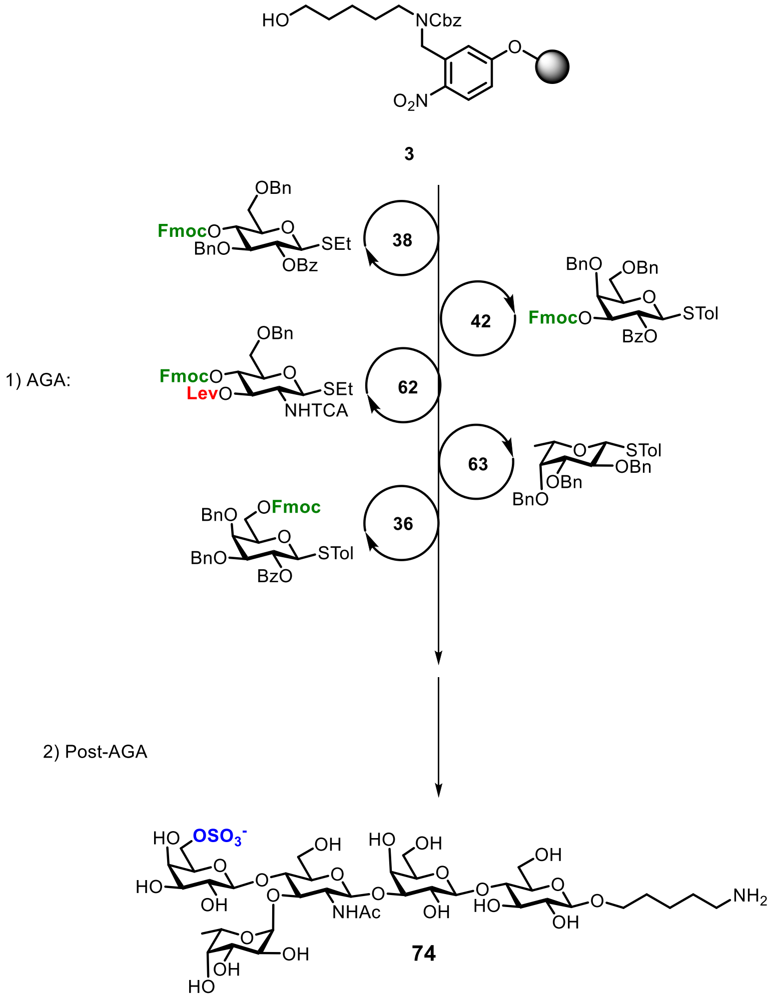
Compound **61** was obtained as a white solid (2 mg, 38% overall yield).

Analytical data for **61**: $^1$H NMR (700 MHz, $D_2O$) δ 4.71 (d, *J* = 8.4 Hz, 1H), 4.60 (d, *J* = 7.8 Hz, 1H), 4.49 (d, *J* = 8.0 Hz, 1H), 4.44 (d, *J* = 7.9 Hz, 1H), 4.35 (dd, *J* = 9.9, 3.3 Hz, 1H), 4.30 (d, *J* = 3.3 Hz, 1H), 4.17 (d, *J* = 3.4 Hz, 1H), 4.01 – 3.91 (m, 3H), 3.91 – 3.56 (m, 16H), 3.01 (t, *J* = 7.6 Hz, 3H), 2.04 (s, 3H), 1.69 (dp, *J* = 14.7, 7.2 Hz, 4H), 1.50 – 1.44 (m, 2H)*;* $^{13}$C NMR (176 MHz, $D_2O$) $^{13}$C NMR (176 MHz, $D_2O$) δ 174.9, 102.9, 102.8, 102.4, 101.9, 82.0, 80.0, 78.4, 78.01, 75.0, 74.9, 74.8, 74.5, 74.4, 72.8, 72.2, 70.0, 70.0, 69.1, 68.3, 66.8, 61.0, 60.9, 60.1, 59.8, 55.2, 39.4, 28.1, 26.4, 23.2, 22.2, 22.0*;* HRMS (QToF)**:** Calcd for $C_{31}H_{55}N_2O_{24}S$ $[M]^-$ 871.2870; found 871.2866.

**RP-HPLC of 61 (ELSD trace, Method C1, $t_R$= 18.50 min)**



### 7.5.12. Synthesis of 74



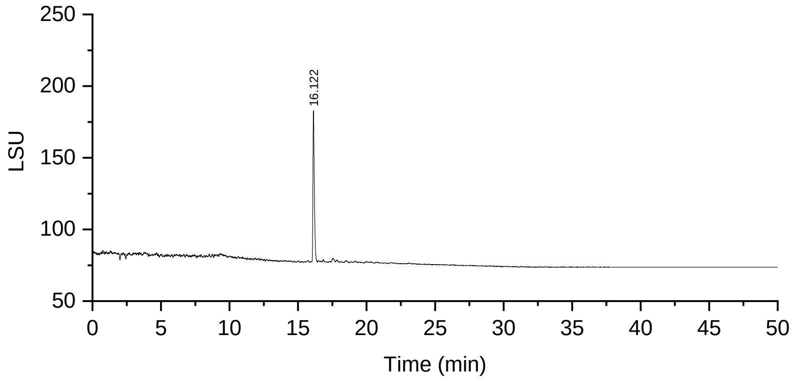
| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **38** | **B, C1, D, E1** | **C1:** (**38**, -20 °C for 5 min, 0 °C for 20 min) |
| | **42** | **B, C1, D, E1** | **C1:** (**42**, -20 °C for 5 min, 0 °C for 20 min) |
| | **62** | **B, C1(x2), D, E2** | **C1:** (**62**, -20 °C for 5 min, 0 °C for 20 min) |
| | **63** | **B, C1(x2), D, E1** | **C1:** (**63**, -20 °C for 5 min, 0 °C for 20 min) |
| | **36** | **B, C1(x2), D, E1,** | **C1:** (**36**, -20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (120 h) |
| | **Hydrogenolysis** | **I1**[a] | **I1:** 10-20% $Pd(OH)_2/C$ (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

[a] treated with thiourea upon completion of hydrogenolysis

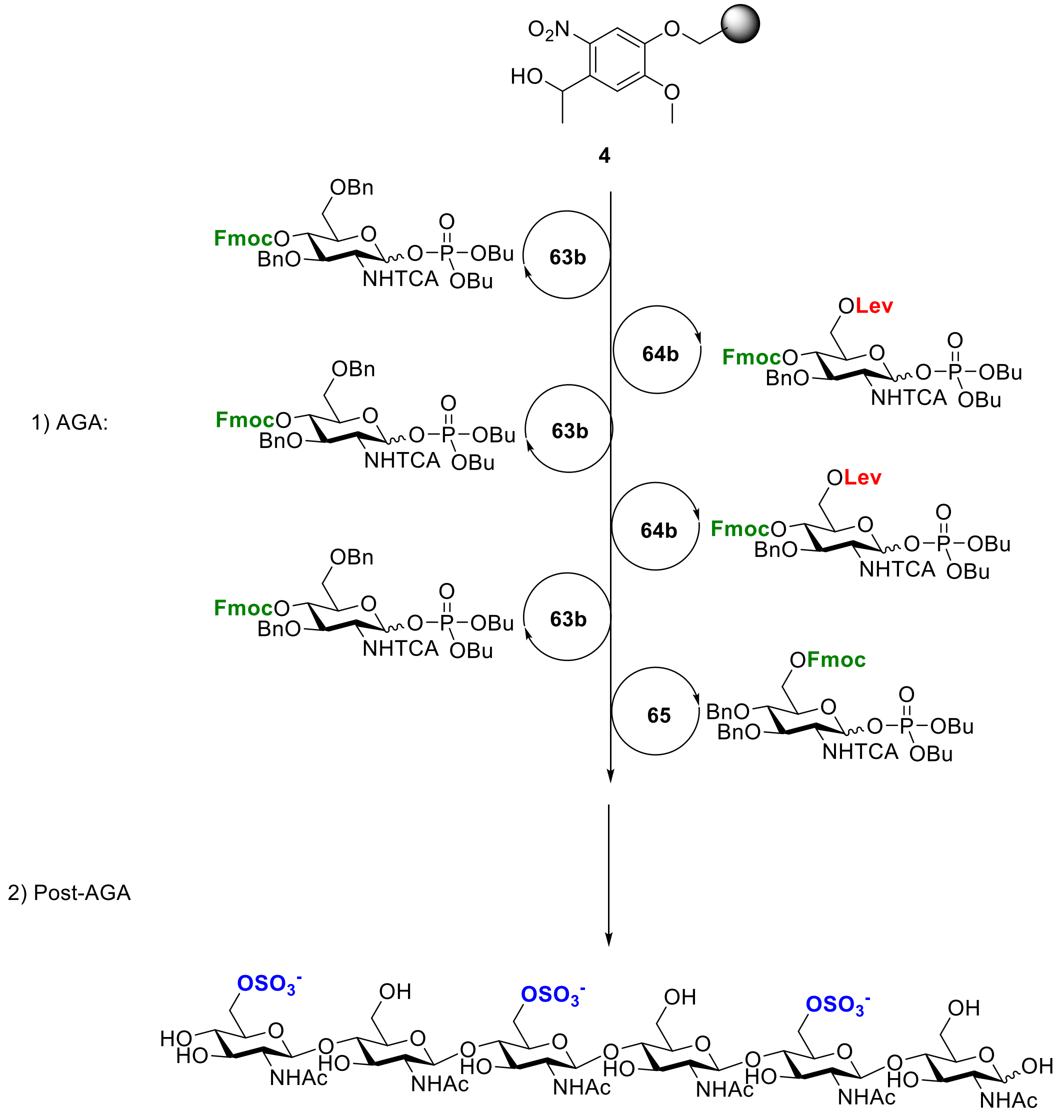
Compound **74** was obtained as a white solid (1 mg, 7% overall yield).

Analytical data for **74**: $^1H$ NMR (700 MHz, $D_2O$) δ 5.12 (d, *J* = 4.0 Hz, 1H), 4.72 (d, *J* = 8.3 Hz, 1H), 4.49 (d, *J* = 8.0 Hz, 2H), 4.44 (d, *J* = 7.8 Hz, 1H), 4.21 – 4.15 (m, 3H), 4.00 – 3.51 (m, 26H), 3.01 (t, *J* = 7.5 Hz, 2H), 2.03 (s, 3H), 1.69 (dt, *J* = 14.8, 7.6 Hz, 4H), 1.47 (p, *J* = 7.9 Hz, 2H), 1.18 (d, *J* = 6.6 Hz, 3H); $^{13}C$ NMR (176 MHz, $D_2O$) δ 174.6, 102.9, 102.5, 102.0, 101.6, 98.5, 82.1, 78.45, 75.2, 74.0, 74.9, 74.8, 74.5, 73.4, 72.8, 72.3, 72.2, 72.0, 70.9, 70.0, 69.9, 69.1, 68.3, 67.9, 67.8, 67.0, 66.7, 61.0, 60.1, 59.8, 56.0, 39.4, 28.1, 26.4, 22.3, 22.0, 20.0, 15.3; HRMS (QToF): Calcd for $C_{37}H_{65}N_2O_{28}S$ $[M]^-$ 1017.3450; found 1017.3487.

**RP-HPLC of 74 (ELSD trace, Method C1, $t_R$= 16.12 min)**



## 7.5.13. Synthesis of 75



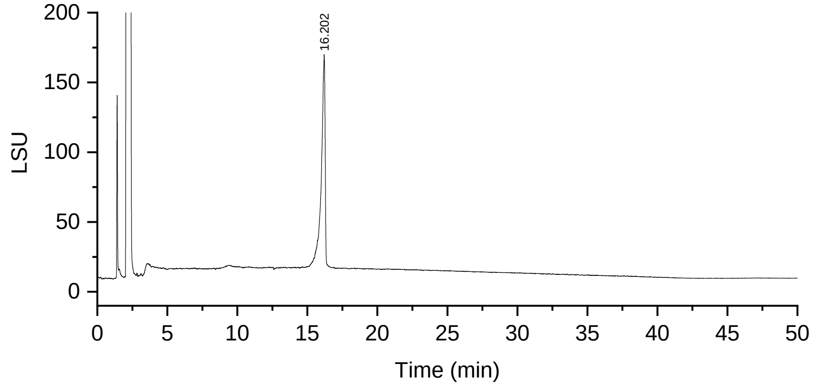
**75**

| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **63b** | **B, C2, D, E1** | **C2:** (**63b**, -30 °C for 5 min, -10 °C for 40 min) |
| | **64b** | **B, C2, D, E1** | **C2:** (**64b**, -30 °C for 5 min, -10 °C for 40 min) |
| | **63b** | **B, C2, D, E1** | **C2:** (**63b**, -30 °C for 5 min, -10 °C for 40 min) |
| | **64b** | **B, C2, D, E1** | **C2:** (**64b**, -30 °C for 5 min, -10 °C for 40 min) |
| | **63b** | **B, C2, D, E1** | **C2:** (**63b**, -30 °C for 5 min, -10 °C for 40 min) |
| | **65** | **B, C2, D, E1, E2** | **C2:** (**65**, -30 °C for 5 min, -10 °C for 40 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrogenolysis** | **I1** | **I1:** 10-20% $Pd(OH)_2/C$ (24 h) |
| | **Purification** | **J(B1), J(A1), K** | |

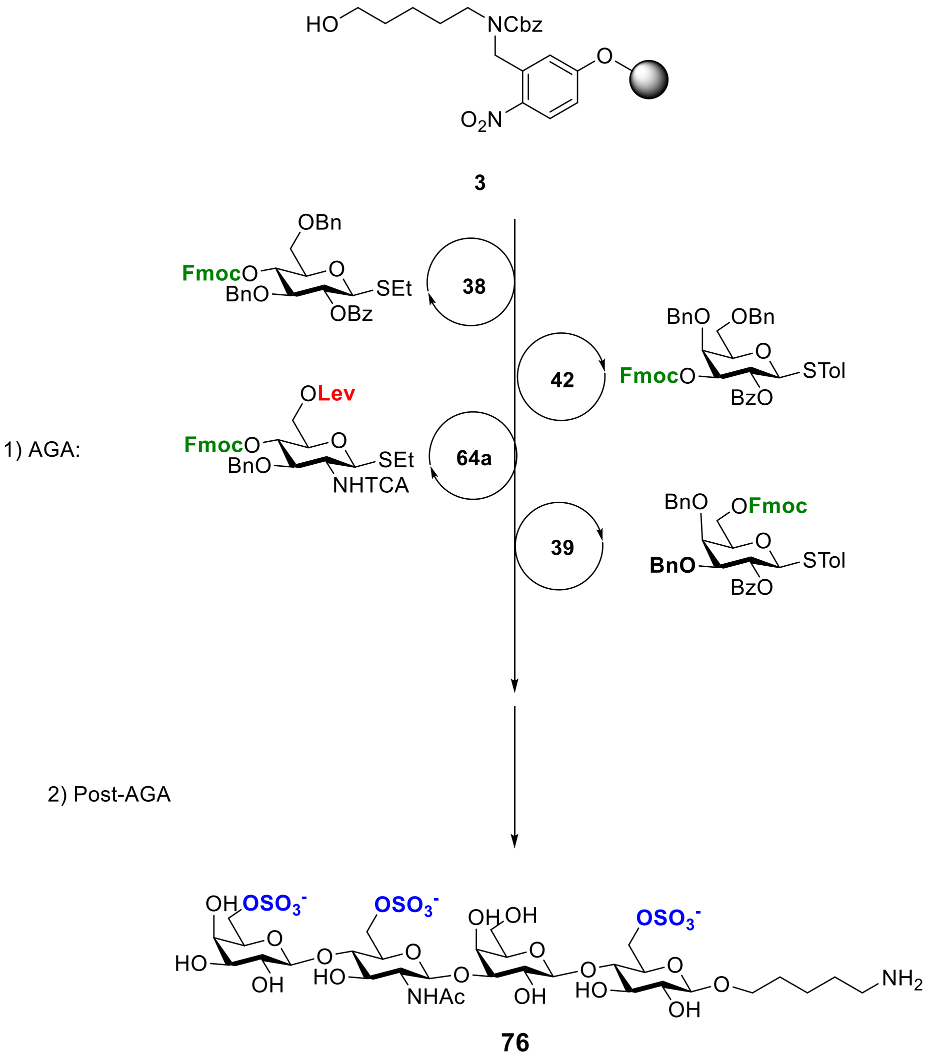
Compound **75** was obtained as a white solid (0.7 mg, 6% overall yield).

Analytical data for **75**: $^1$H NMR (700 MHz, $D_2O$) δ 5.20 (d, *J* = 3.7 Hz, 0.60H, α-H1), 4.69 (d, *J* = 6.1 Hz, 0.40H, β-H1), 4.61 (dt, *J* = 14.0, 6.5 Hz, 5H), 4.36 (d, *J* = 10.8 Hz, 1H), 4.29 (d, *J* = 10.8 Hz, 2H), 4.23 (dd, *J* = 11.3, 5.5 Hz, 1H), 4.12 (dd, *J* = 12.7, 5.9 Hz, 2H), 3.95 – 3.46 (m, 31H), 2.11 – 2.03 (m, 18H); $^{13}$C NMR (176 MHz, $D_2O$) δ 101.5, 101.4, 101.3, 101.2, 102.1, 94.6, 90.2.$^*$Only the anomeric carbons are reported due to low amount; HRMS (QToF): Calcd for $C_{48}H_{77}N_6O_{40}S_3$ $[M]^{3-}$ 491.1118; found 491.1127.

**RP-HPLC of 75 (ELSD trace, Method C1, $t_R$= 16.2 min)**



### 7.5.14. Synthesis of 76



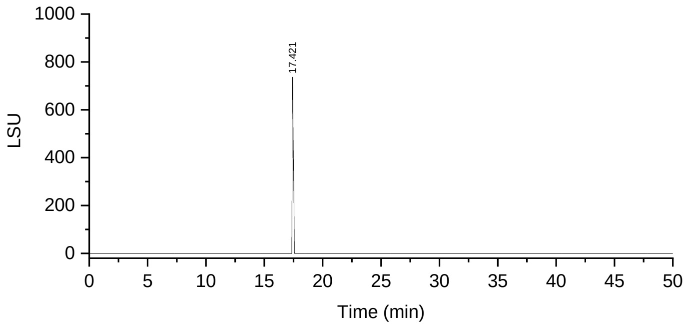
| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **38** | **B, C1(x2), D, E1** | **C1:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **42** | **B, C1(x2), D, E1** | **C1:** (**42**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **64a** | **B, C1(x2), D, E1** | **C1:** (**64a**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C1(x2), D, E1, E2** | **C1:** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (120 h) |
| | **Acetylation** | | **G1** |
| | **Hydrogenolysis** | **I1**[a] | **I1:** 10-20% $Pd(OH)_2/C$ (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

[a] treated with thiourea upon completion of hydrogenolysis

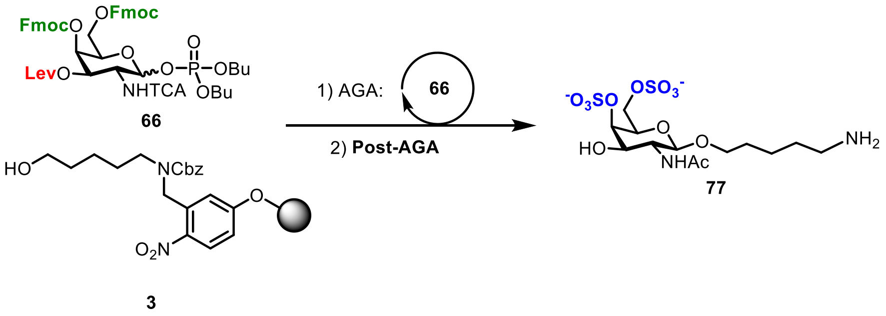
Compound **76** was obtained as a white solid (0.7 mg, 6% overall yield).

Analytical data for **76**: $^1$H NMR (700 MHz, $D_2O$) δ 4.74 – 4.71 (m, 1H), 4.57 – 4.48 (m, 3H), 4.44 – 4.39 (m, 1H), 4.34 – 4.18 (m, 5H), 4.02 – 3.52 (m, 20H), 3.37 – 3.30 (m, 1H), 3.02 (t, *J* = 7.6 Hz, 2H), 2.05 (s, 2H), 1.69 (dp, *J* = 18.0, 6.6 Hz, 4H), 1.52 – 1.42 (m, 2H); $^{13}$C NMR (176 MHz, $D_2O$) δ 174.9, 102.9, 102.8, 102.7, 102.1, 82.4, 78.7, 77.7, 75.0, 74.4, 74.4, 72.7, 72.5, 72.3, 72.3, 72.2, 72.1, 70.8, 70.3, 69.9, 68.3, 68.2, 67.0, 66.6, 66.4, 61.1, 55.1, 39.4, 39.3, 28.2, 26.2, 22.2, 22.2, 22.0; HRMS (QToF): Calcd for $C_{31}H_{54}N_2O_{30}S_3^{3-}$ $[M]^{2-}$ 515.0967; found 515.0961.

**RP-HPLC of 76 (ELSD trace, Method C1, $t_R$= 17.42 min)**



## 7.5.15. Synthesis of 77

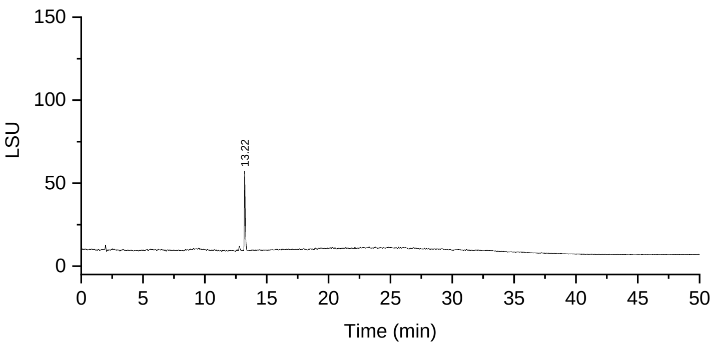


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **66** | **B, C2, D, E1** | **C2:** (**66**,-30 °C for 5 min, -10 °C for 40 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Acetylation** | | **G1** |
| | **Hydrolysis** | **G** | **G:** (6 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

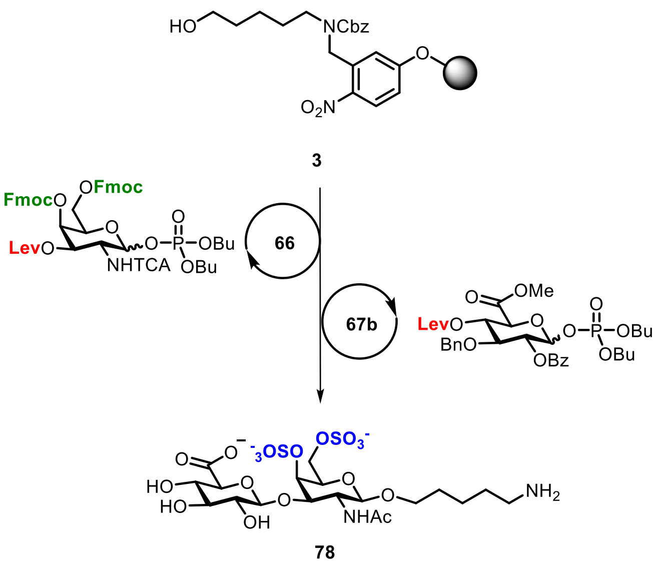
Compound **77** was obtained as a white solid (2 mg, 31% overall yield).

Analytical data for **77**: $^1$H NMR (400 MHz, $D_2O$) δ 4.71 – 4.67 (m, 1H), 4.53 – 4.48 (m, 1H), 4.30 (dd, *J* = 11.4, 2.7 Hz, 1H), 4.20 (t, *J* = 10.3 Hz, 1H), 4.08 – 4.01 (m, 1H), 3.92 – 3.83 (m, 3H), 3.70 – 3.61 (m, 1H), 2.97 (t, *J* = 7.8 Hz, 2H), 2.02 (s, 3H), 1.71 – 1.56 (m, 4H), 1.49 – 1.34 (m, 2H); $^{13}$C NMR (101 MHz, $D_2O$) δ 174.7, 101.4, 75.5, 72.1, 70.3, 69.9, 67.9, 52.6, 39.3, 28.1, 26.3, 22.1, 21.9.; HRMS (QToF): Calcd for $C_{13}H_{25}N_2O_{12}S_2$ $[M+H]^-$ 465.0865; found 465.0875.

**RP-HPLC of 77 (ELSD trace, Method $C_1$, $t_R$= 13.22 min)**



## 7.5.16. Synthesis of 78



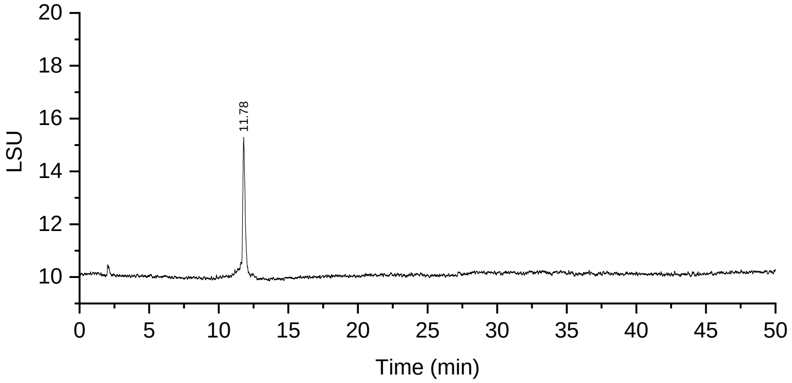
| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **66** | **B, C2, D, E2** | **C2:** (**66,**-30 °C for 5 min, -10 °C for 40 min) |
| | **67b** | **B, C2(x2), D, E2, E1*** | **C2:** (**67b,**-30 °C for 5 min, -10 °C for 40 min) |
| Post-AGA | **Sulfation** | **F** | |
| | **Hydrolysis** | **G** | **G:** (12 h) |
| | **Acetylation** | | **G1** |
| | **Hydrolysis** | **G** | **G:** (6 h) |
| | **Hydrogenolysis** | **I** | **I:** 10% Pd/C (12 h) |
| | **Purification** | **J(B1), J(A1), K** | |

*To avoid possible elimination side-reactions, the Fmoc deprotection E1 was carried out with 5% of TEA in DMF .

Compound **78** was obtained as a white solid (1.5 mg, 17% overall yield).

Analytical data for **78**: $^1$H NMR (600 MHz, $D_2O$) δ 4.51 (d, *J* = 8.3 Hz, 1H), 4.47 (d, *J* = 7.8 Hz, 1H), 4.29 (dd, *J* = 11.4, 2.8 Hz, 1H), 4.19 (dd, *J* = 11.4, 9.0 Hz, 1H), 4.10 – 4.02 (m, 3H), 3.88 (dt, *J* = 10.3, 6.2 Hz, 1H), 3.70 – 3.67 (m, 1H), 3.65 (d, *J* = 9.8 Hz, 1H), 3.52 (t, *J* = 9.4 Hz, 1H), 3.45 (t, *J* = 9.2 Hz, 1H), 3.36 (dd, *J* = 9.4, 7.8 Hz, 1H), 2.98 (t, *J* = 7.6 Hz, 2H), 2.01 (s, 3H), 1.70 – 1.59 (m, 4H), 1.46 – 1.37 (m, 2H); $^{13}$C NMR (151 MHz, $D_2O$) δ 175.9, 174.6, 103.0, 101.2, 76.4, 76.1, 75.1, 74.3, 72.4, 72.2, 71.7, 70.3, 68.0, 51.7, 39.3, 28.0, 26.2, 22.2, 21.9; HRMS (QToF): Calcd for $C_{19}H_{33}N_2O_{18}S_2$ [M+H+H]$^-$ 641.1175; found 641.1179.

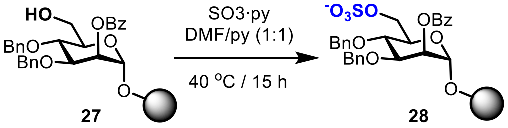
## RP-HPLC of 78 (ELSD trace, Method C1, $t_R$= 11.76 min)



## 7.6. Mass spectrometry and additional information

Representative examples of reaction monitoring are reported. MALDI and/or ESI-MS is performed after microcleavage from the solid support at each step of the synthetic process.

### 7.6.1. Mass spectrometry analysis of the intermediate steps for the synthesis of 31



**27** **28**

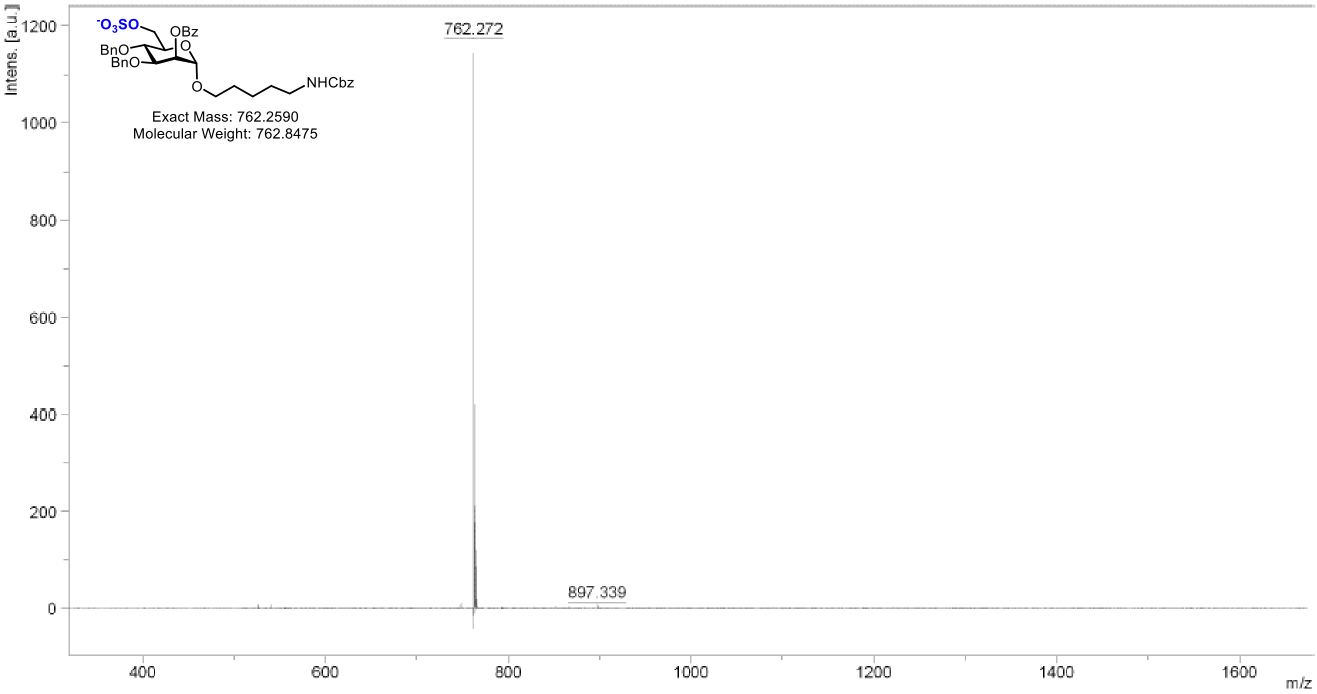
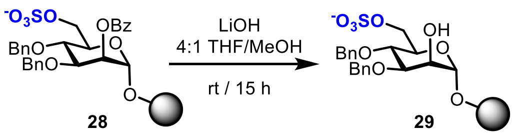


**Figure S2.** MALDI-TOF of compound **28** after microcleavage (negative mode).



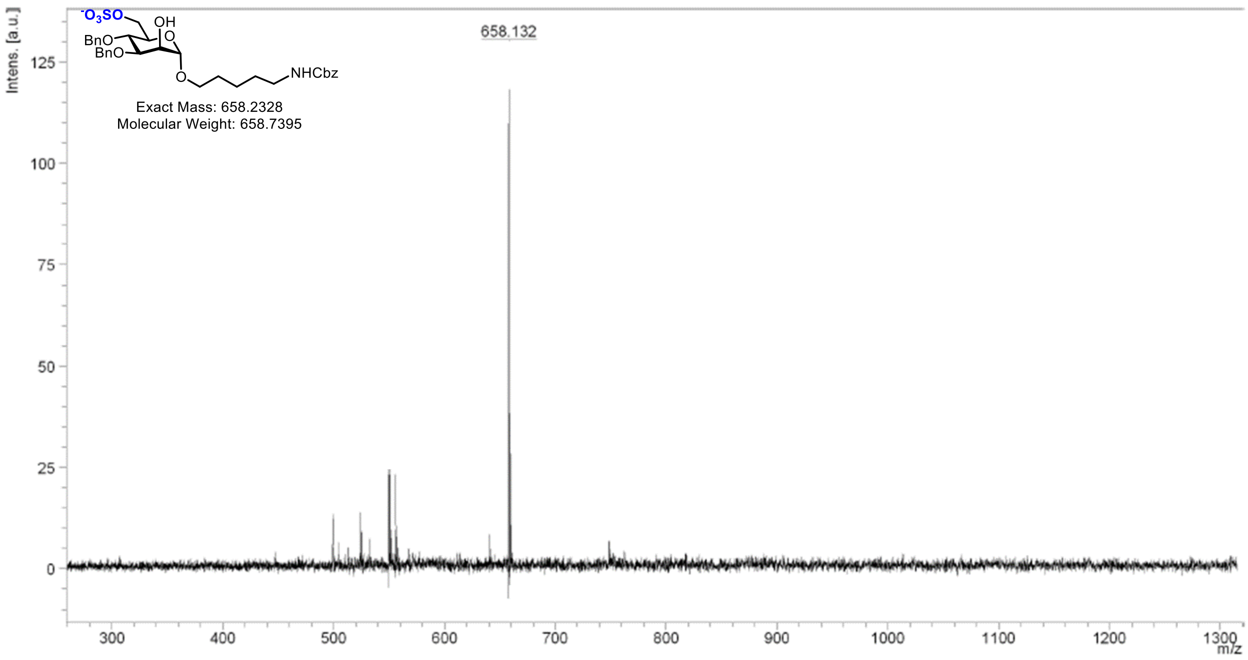
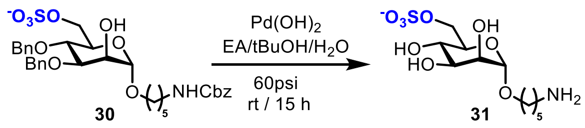


**Figure S3**. MALDI-TOF of compound **29** after microcleavage (negative mode).



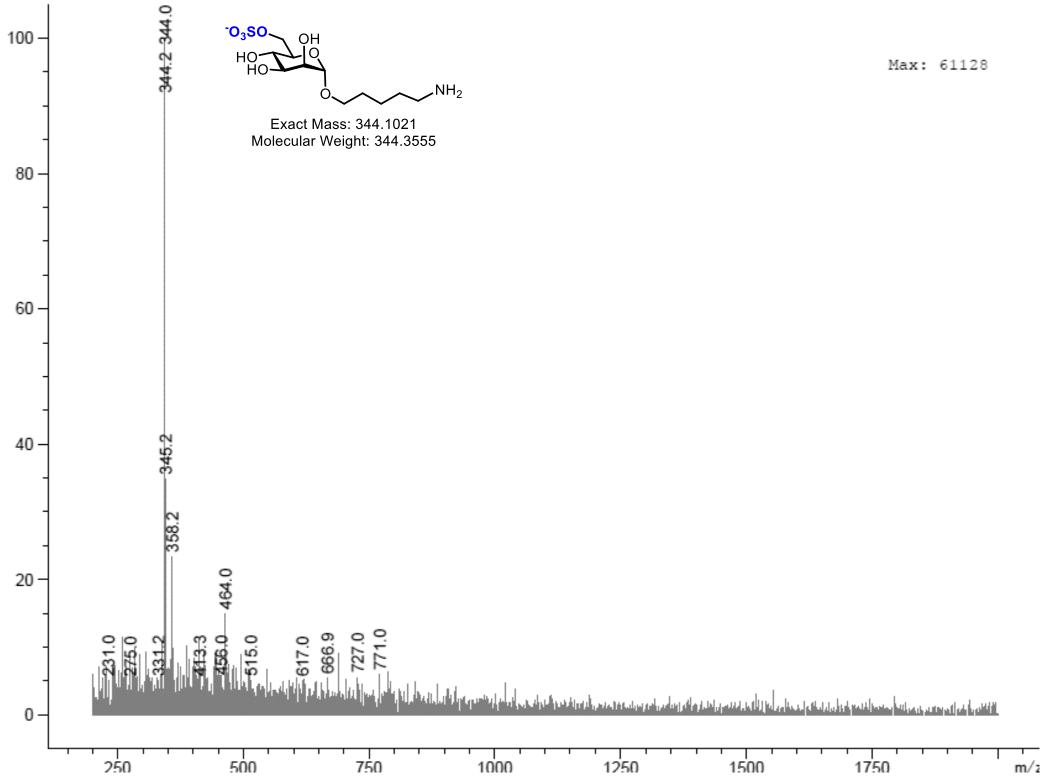
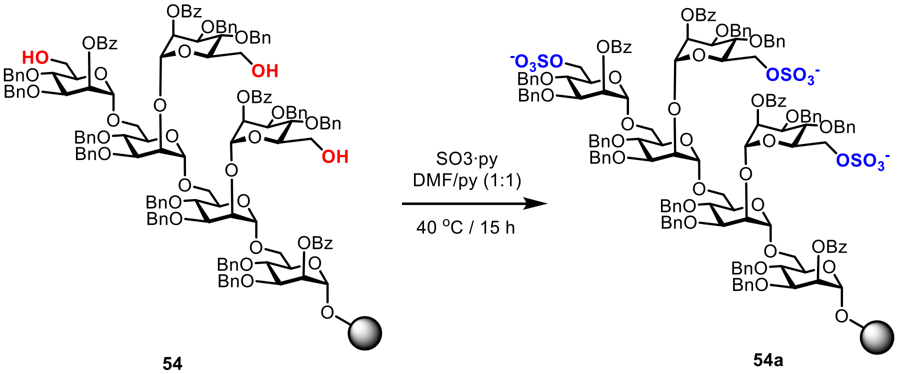


**Figure S4**. ESI-MS of crude compound **31** (negative mode).

### 7.6.2. Mass spectrometry analysis of the intermediate steps for the synthesis of 64



**54**

**54a**

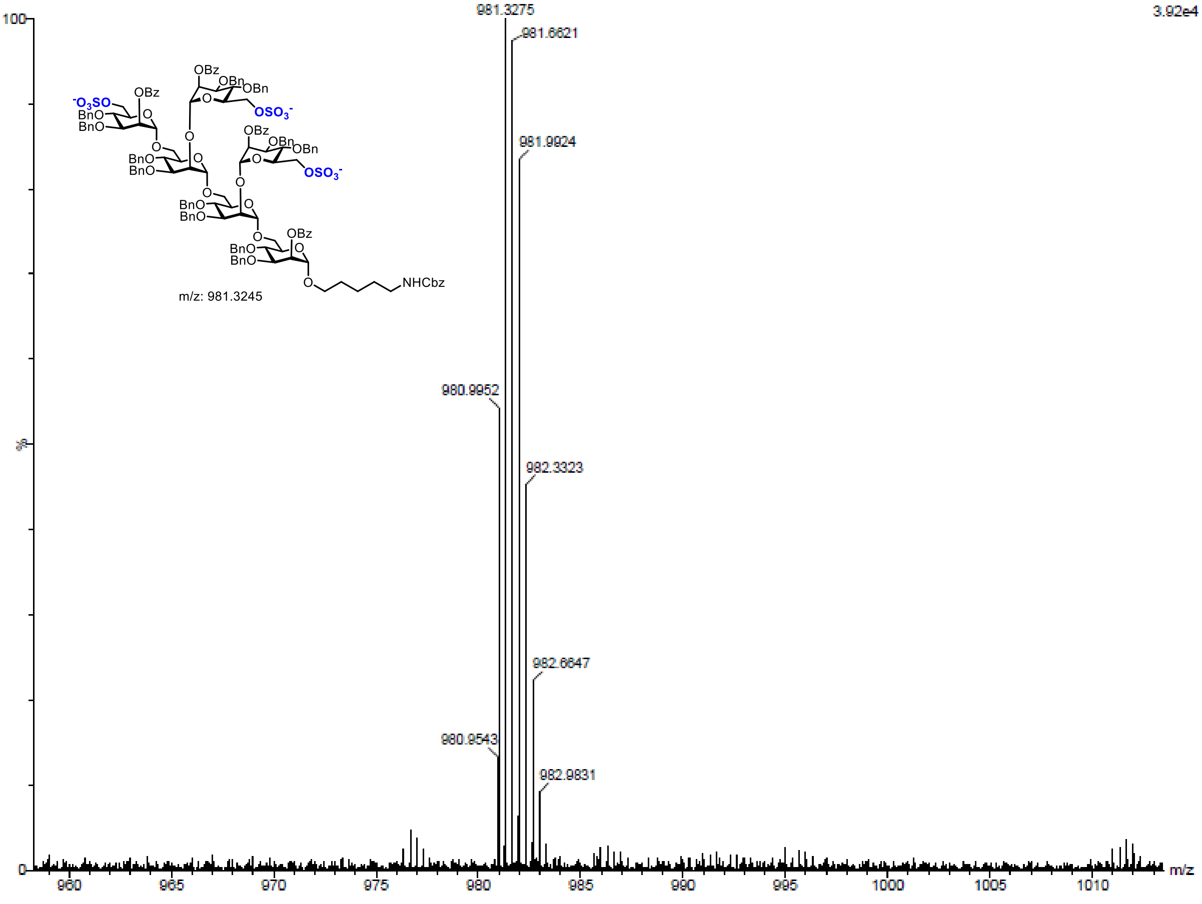
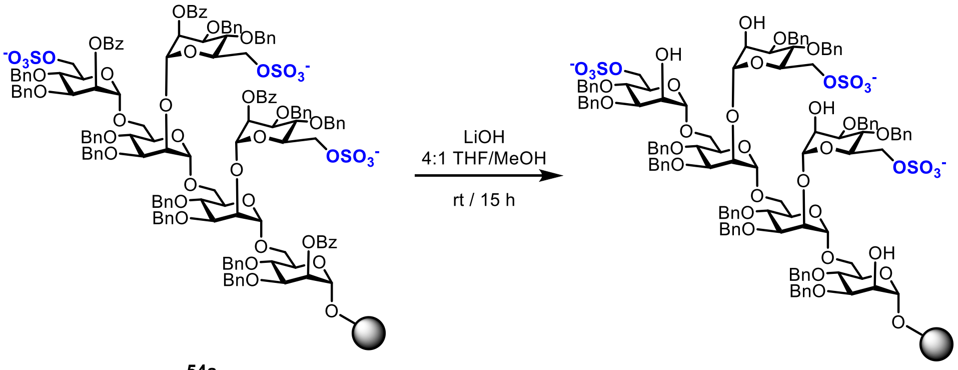


**Figure S5.** QTOF-MS of compound **54a** after microcleavage (negative mode).



**54a**

**54b**

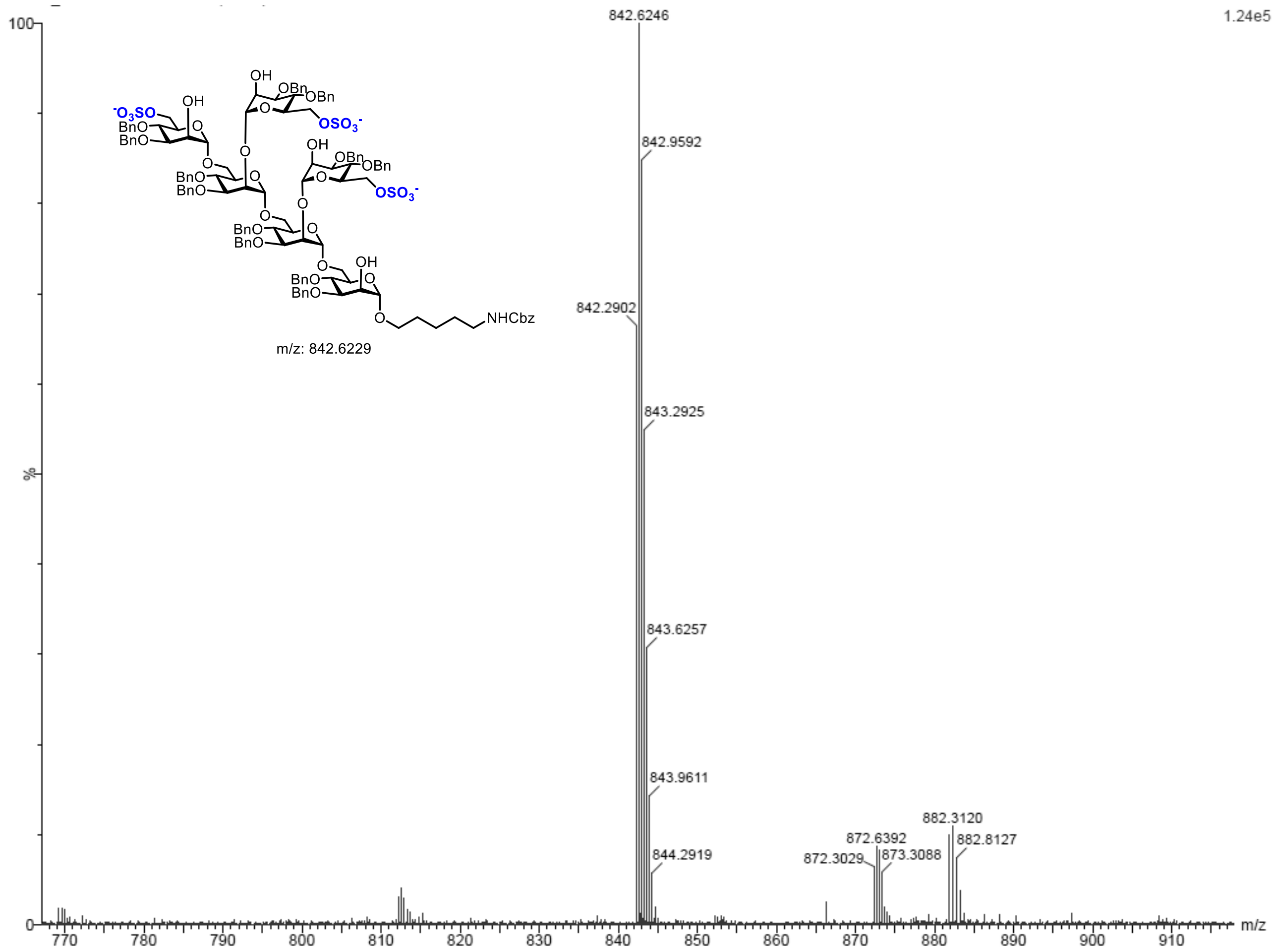
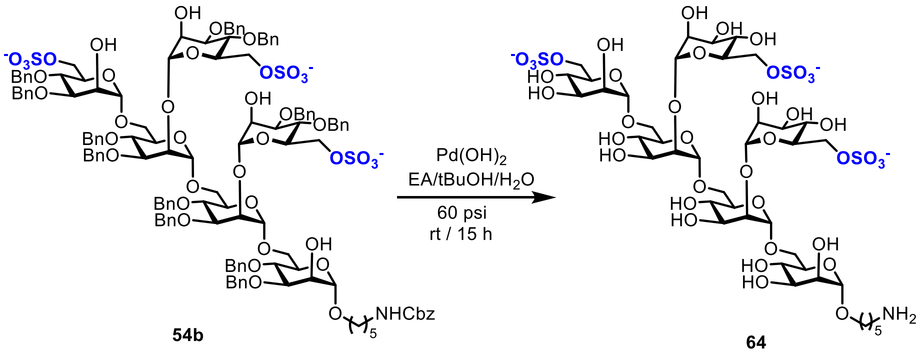


**Figure S6.** QTOF-MS of compound **54b** after microcleavage (negative mode).



**54b** **64**

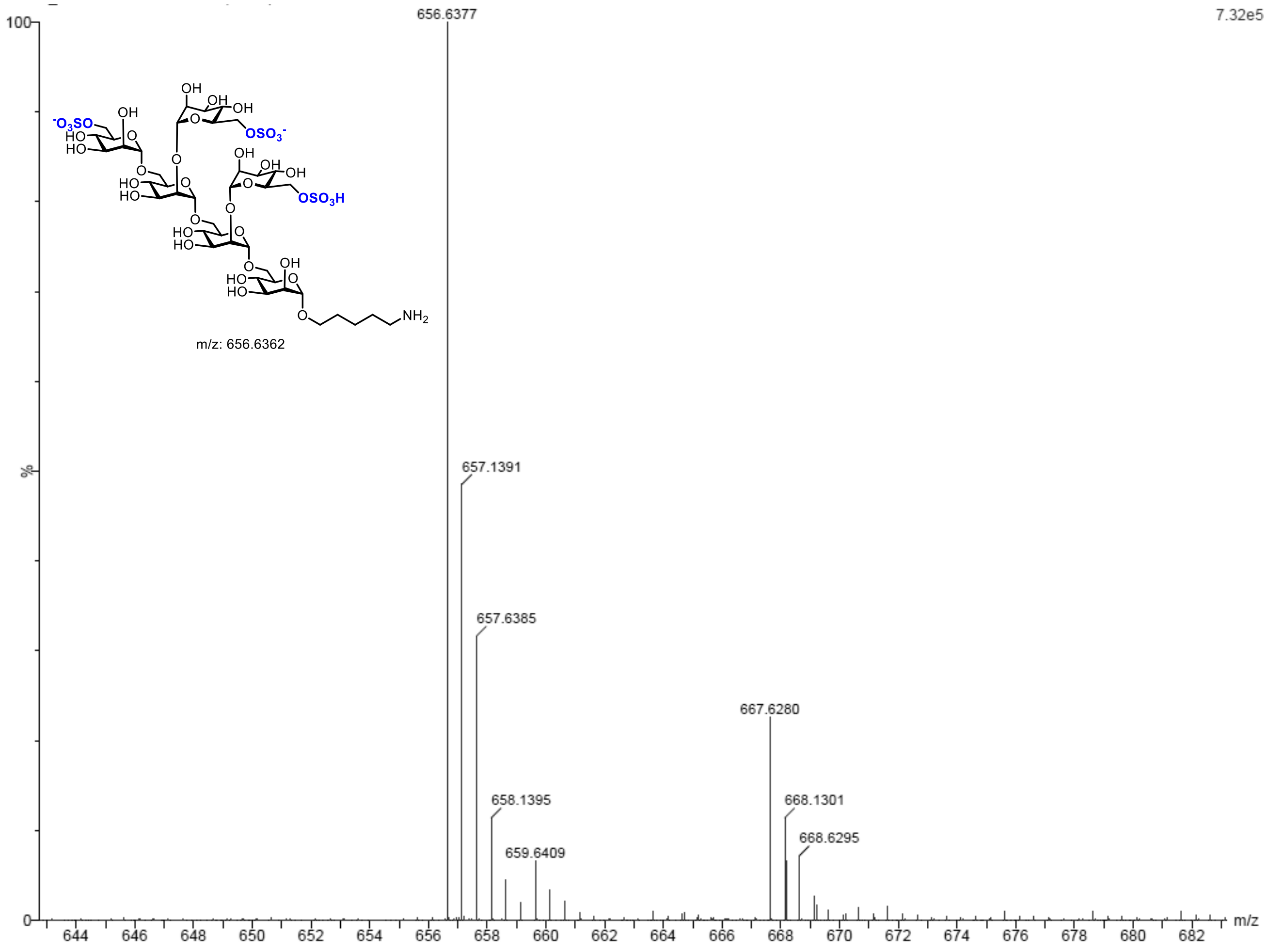
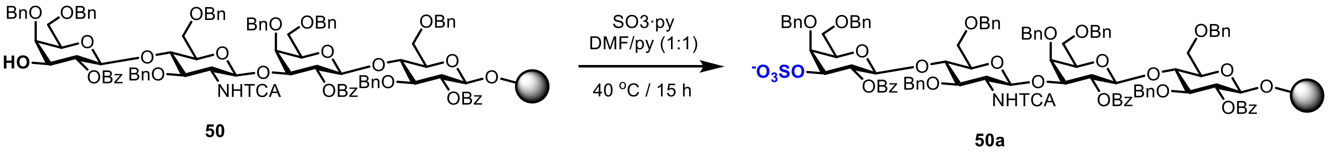


**Figure S7.** QTOF-MS of compound **64** after hydrogenolysis (negative mode).

### 7.6.3. Mass spectrometry analysis of the intermediate steps for the synthesis of 61



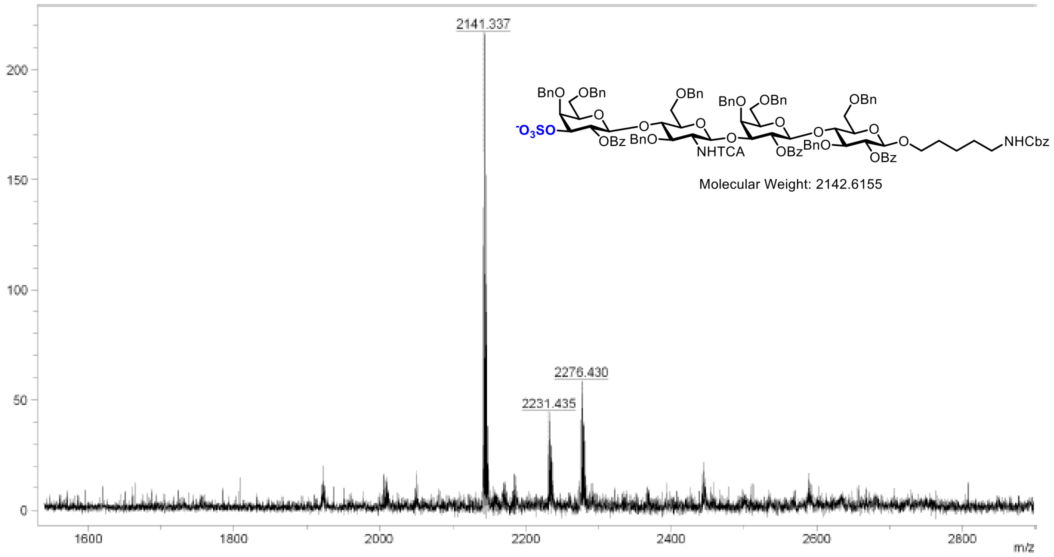
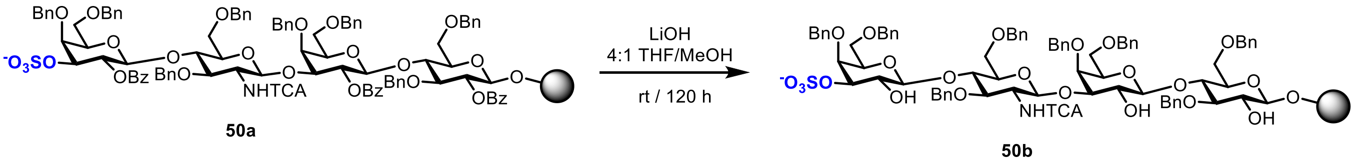


**Figure S8**. MALDI-TOF of compound **50a** after microcleavage (negative mode).



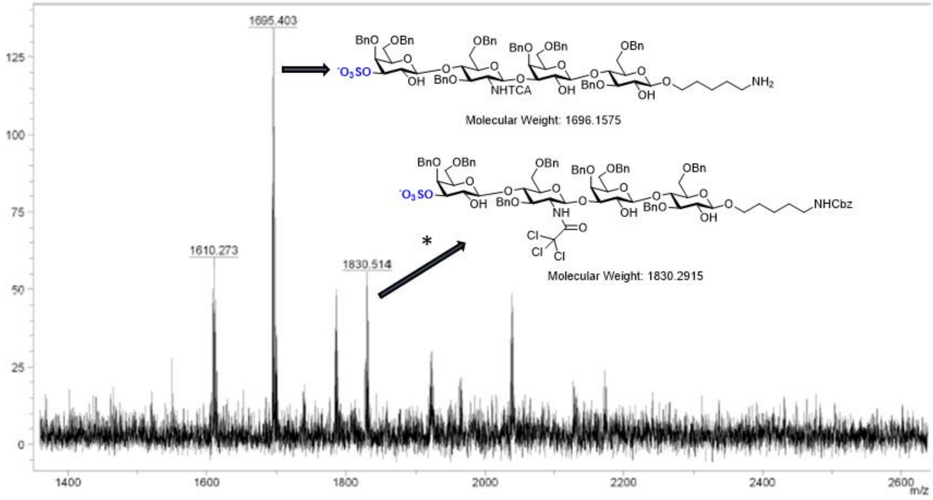
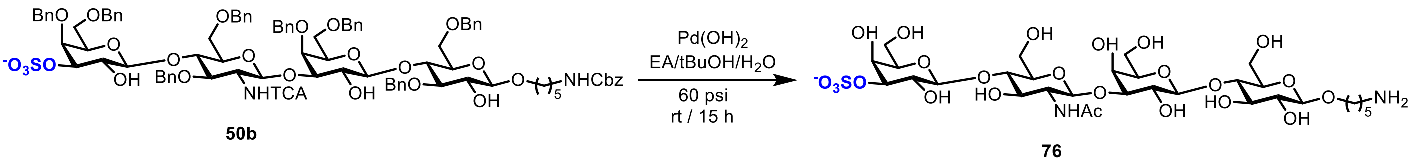


**Figure S9**. MALDI-TOF of compound **28** after microcleavage (negative mode).*Note: The loss of Cbz protecting group is commonly detected by MALDI and it is due to the longtime exposed to light during microcleavage.



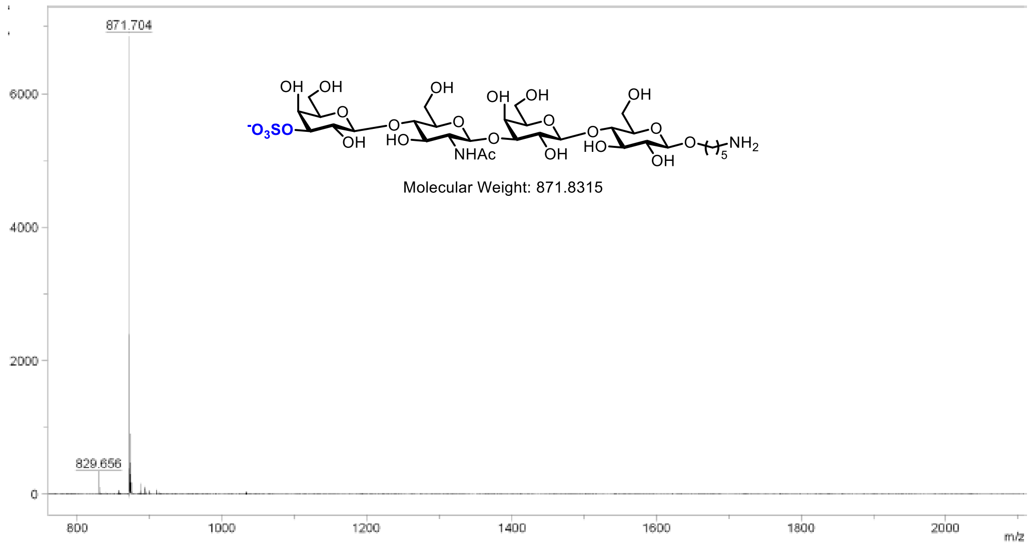


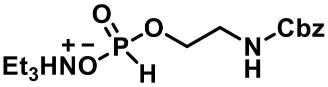
**Figure S10**. MALDI-TOF of compound **76** after hydrogenolysis (negative mode).

## 7.7. Post-synthesizer manipulations (Post-AGA) for pEtN oligomers synthesis

**Module L: Cleavage from solid Support**

The oligosaccharides were cleaved from the solid support using a continuous-flow photoreactor as described previously.[185]

**Module M: Phosphorylation**



**85**

**85** was prepared according to a previously established procedure.[186]

The partially protected oligosaccharide obtained from *Module F* was mixed with **85** (4 equiv.), co-evaporated with pyridine for three times and dried under high vacuum for 2 h. The ratio between the oligosaccharide and **85** changed depending on the oligosaccharide structure (here we report the most common set of conditions, variations are reported in the specific procedures). The mixture was dissolved in anhydrous pyridine (2 mL) and a solution of pivaloyl chloride (equimolar to **85**) in pyridine (1 mL) was added. The solution was stirred for 12 h at RT, after which time iodine (10 equiv.) and water (0.5 mL) were added and the reaction was stirred for additional 2 h. The reaction mixture was quenched with $Na_2S_2O_3$ and extracted with $CH_2Cl_2$. The organic layers were combined and evaporated.

**Module N: Solution-phase methanolysis**

The oligosaccharide was dissolved in MeOH : DCM (1.5 mL, 1:1). NaOMe in MeOH (0.5 M, 3 equiv. per benzoyl ester) was added and the solution was stirred at RT for 12 h, neutralized with Amberlite IR-120 ($H^+$ form), filtered and concentrated *in vacuo.*

**Module O: Hydrogenolysis at ambient pressure**

The compound obtained from *Module H* was dissolved in 2 mL of EA:*t*-BuOH:$H_2O$ (1:0.5:0.5). 100 % by weight Pd/C (10 %) was added and the reaction was stirred under $H_2$–atmosphere for 12 h. The reaction was filtered through celite and washed with *t*-BuOH and $H_2O$. The filtrates were concentrated *in vacuo.*

**Module P: Purification**

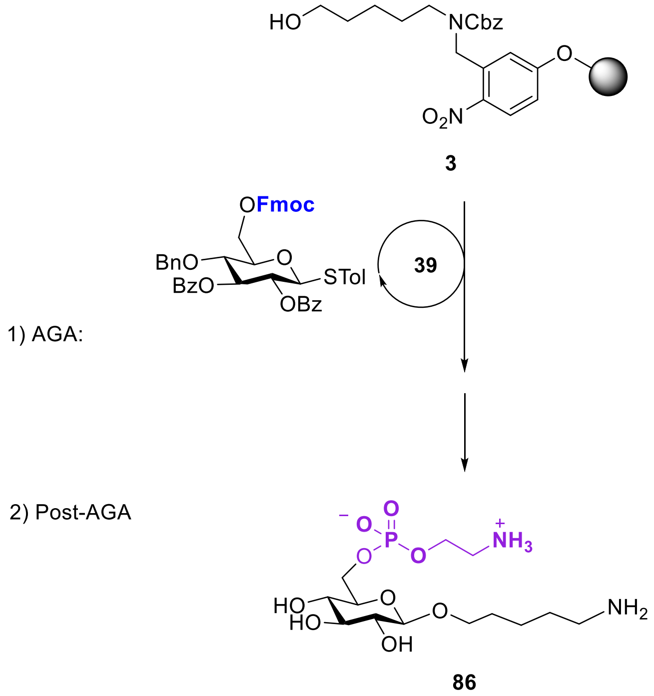
After photovleavage, crudes were analyzed and purified using analytical and preparative HPLC (Agilent 1200 Series spectrometer, **Method $Q_1$** and **Method $Q_2$,** respectively). The protected phosphorylated crudes were purified with **Method R** and **Method S**. After methanolysis, the semi-deprotected compounds were purified with **Method S**. The final compounds were purified with **Method U** and analyzed using analytical HPLC (Agilent 1200 Series spectrometer, **Method V**).

- **Method $Q_1$:** (YMC-Diol-300 column, 150 x 4.6 mm) flow rate of 1.0 mL / min with Hex – 20% EtOAc as eluent [isocratic 20% EtOAc (5 min), linear gradient to 55% EtOAc (35 min), linear gradient to 100% EtOAc (5 min)].
- **Method $Q_2$:** (YMC-Diol-300 column, 150 x 20 mm) flow rate of 15 mL / min with Hex – 20% EtOAc as eluent [isocratic 20% EtOAc (5 min), linear gradient to 55% EtOAc (35 min), linear gradient to 100% EtOAc (5 min)].
- **Method R**: Manual silica gel column chromatography quenched with 0.1% $Et_3N$ in DCM. Solvent system DCM:MeOH, gradient from 0 to 10% MeOH.
- **Method S**: Sephadex® LH-20 column with DCM:MeOH (1:1) as eluent, isocratic.
- **Method T:** Sephadex® LH-20 column with $H_20$:MeOH (1:1) as eluent, isocratic.
- **Method U:** Manual reverse phase $C_{18}$ silica gel column chromatography. Solvent system $H_2O$:MeOH, gradient from 0 to 50% MeOH).
- **Method V:** (Hypercarb column, 150 x 10 mm) flow rate of 0.7 mL / min with $H_2O$ (0.1% formic acid) as eluents [isocratic (5 min), linear gradient to 50% ACN (30 min), linear gradient to 100% ACN (5 min)].

Following final purification, all deprotected products were lyophilized on a Christ Alpha 2-4 LD plus freeze dryer prior to characterization.

## 7.8. Synthesis of pEtN oligosaccharides

### 7.8.1. Synthesis of 86

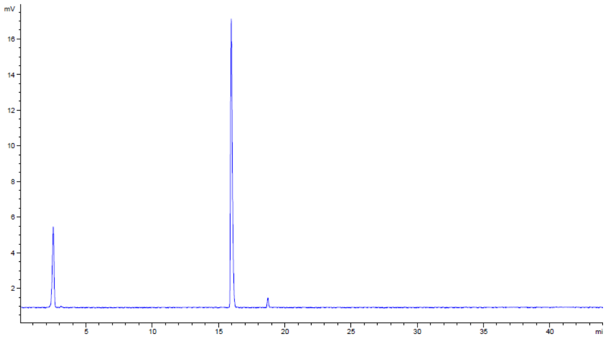


| Step | | Modules | Notes |
|---|---|---|---|
| **AGA** | | **A** | |
| | **39** | **B**, **C**, **D**, **E** | **C**:( **39**, -20° for 5 min, 0° for 20 min) |
| **Post-AGA** | | **L, P** | **P:** (Method **Q**$_2$, $t_R$ = 18.9 min) |
| | **Phosphorylation** | **M, P1, P2,** | **P1:** (Method **R**)<br>**P2:** (Method **S**) |
| | **Deprotection** | **N, P1, O, P2** | **P1:** (Method **S**) **P2:** (Method **U** and **T**) |

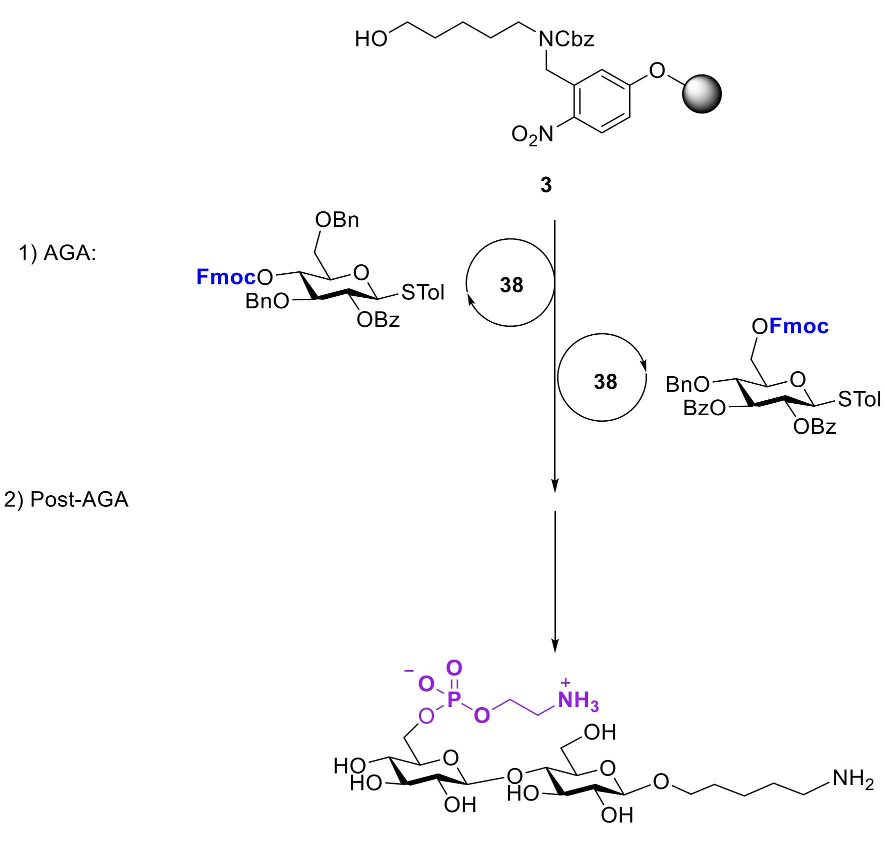
Compound **86** was obtained as a white solid (1.1 mg, 25 % overall yield).

Analytical data for **86**: $^1$H NMR (400 MHz, $D_2O$) δ 4.40 (t, *J* = 7.7 Hz, 1H), 4.17 – 3.95 (m, 4H), 3.85 (dt, *J* = 10.0, 6.5 Hz, 1H), 3.75 – 3.35 (m, 5H), 3.27 (s, 1H), 3.25 – 3.12 (m, 3H), 2.93 (t, *J* = 7.5 Hz, 2H), 1.60 (p, *J* = 7.5 Hz, 4H), 1.50 – 1.29 (m, 2H).; $^{13}$C NMR (101 MHz, $D_2O$) δ 102.19, 75.54, 74.2, 72.97, 70.07, 68.95, 64.47, 40.04, 39.22, 28.06, 26.27, 21.93.; $^{31}$P NMR (162 MHz, $D_2O$) δ 0.24. m/z (HRMS+) 389.1671 [M+H]$^+$ ($C_{13}H_{30}N_2O_9P$ requires 389.1689).

**RP-HPLC of 86 (ELSD trace, Method F$_1$, t$_R$= 15.9 min)**



## 7.8.2. Synthesis of 87



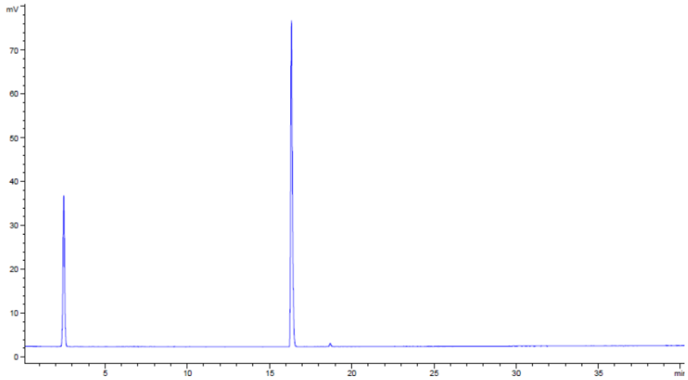
**87**

| Step | | Modules | Notes |
|---|---|---|---|
| **AGA** | | **A** | |
| | **38** | **B**, **C**, **D**, **E** | **C**:( **38**, -20° for 5 min, 0° for 20 min) |
| | **39** | **B**, **C**, **D**, **E** | **C**:( **39**, -20° for 5 min, 0° for 20 min) |
| **Post-AGA** | | **L, P** | **P:** (Method $\mathbf{Q_2}$, $t_R$ = 23.6 min) |
| | **Phosphorylation** | **M, P1, P2,** | **P1:** (Method **R**)<br>**P2:** (Method **S**) |
| | **Deprotection** | **N, P1, O, P2** | **J1:** (Method **S**) **J2:** (Method **U** and **T**) |

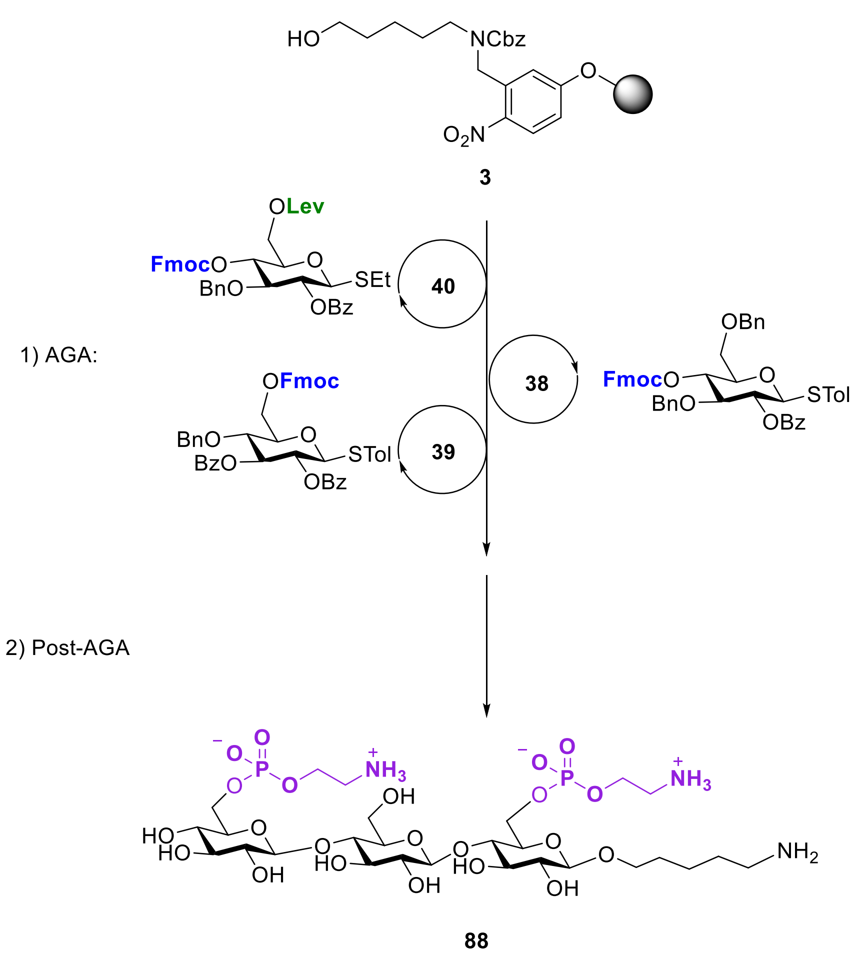
Compound **87** was obtained as a white solid (2.1 mg, 49 % overall yield).

Analytical data for **87** : $^1$H NMR (400 MHz, $D_2O$) δ 4.46 (d, *J* = 8.0 Hz, 1H), 4.42 (d, *J* = 8.0 Hz, 1H), 4.22 – 3.78 (m, 6H), 3.74 – 3.42 (m, 8H), 3.29 – 3.16 (m, 4H), 2.97 – 2.89 (m, 2H), 1.61 (h, *J* = 7.0, 6.3 Hz, 4H), 1.44 – 1.34 (m, 2H).; $^{13}$C NMR (101 MHz, $D_2O$) δ 104.82, 103.93, 81.37, 77.25, 76.66, 76.40, 75.03, 74.84, 72.09, 71.00, 64.05, 62.02, 41.32, 28.42, 24.09.$^{31}$P NMR (162 MHz, $D_2O$) δ 0.26.; m/z (HRMS+) 551.2217 $[M+H]^+$ ($C_{19}H_{39}N_2O_{14}P$ requires 551.2217).

**RP-HPLC of 87 (ELSD trace, Method $F_1$, $t_R$= 16.3 min)**



### 7.8.3. Synthesis of 88

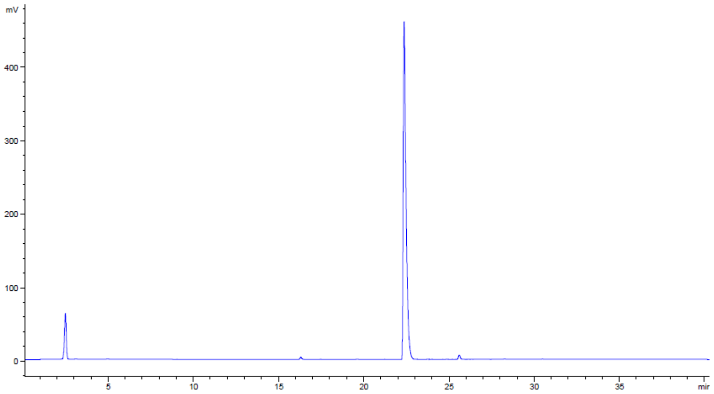


| Step | | Modules | Notes |
|---|---|---|---|
| **AGA** | | **A** | |
| | **40** | **B, C, D, E** | **C**:( **40**, -20° for 5 min, 0° for 20 min) |
| | **38** | **B, C, D, E** | **C**:( **38**, -20° for 5 min, 0° for 20 min) |
| | **39** | **B, C, D, E, $E_2$** | **C**:( **39**, -20° for 5 min, 0° for 20 min) |
| **Post-AGA** | | **L, P** | **P:** (Method **$Q_2$**, $t_R$ = 35.2 min) |
| | **Phosphorylation** | **M, P1, P2** | **M:** ( 5eq. of compound **85** per -OH)<br>**P1:** (Method **R**)<br>**P2:** (Method **S**) |
| | **Deprotection** | **N, P1, O, P2** | **J1:** (Method **S**) **J2:** (Method **U** and **T**) |

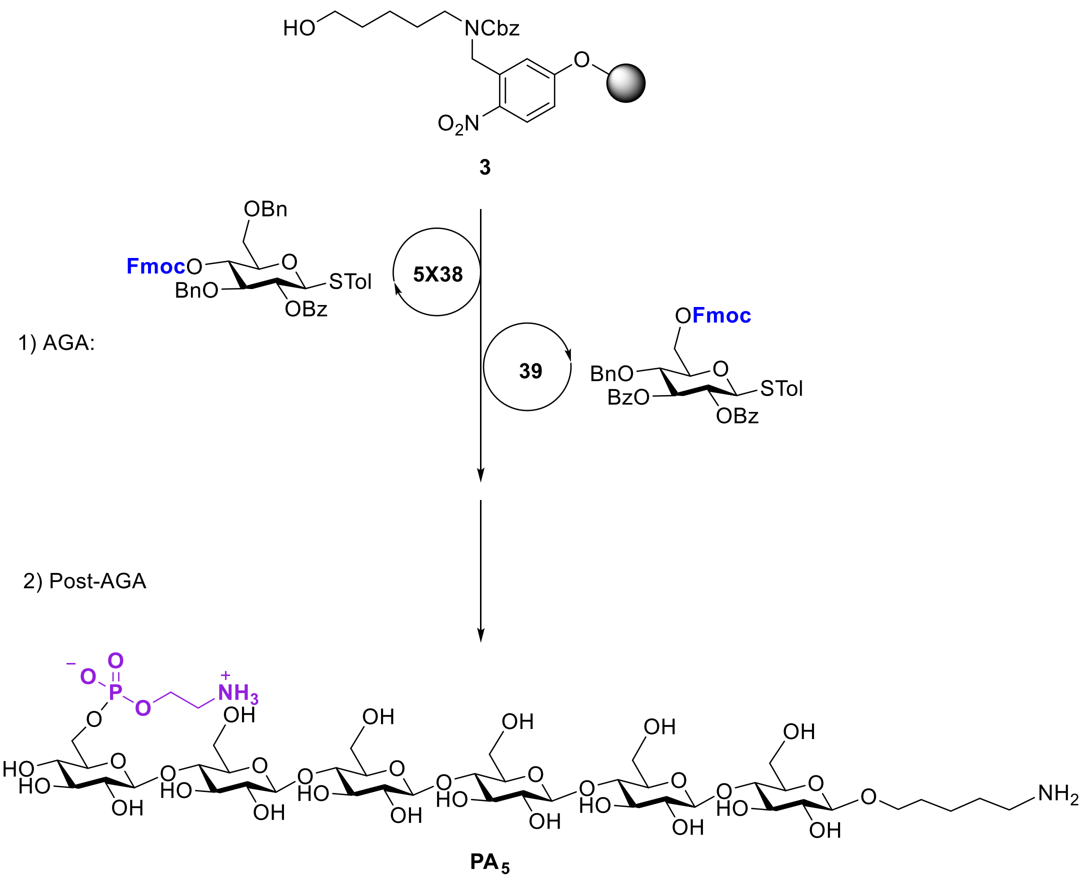
Compound **88** was obtained as a white solid (2.7 mg, 25 % overall yield).

Analytical data for **88:** $^1$H NMR (600 MHz, $D_2O$) δ 4.61 (d, *J* = 8.0 Hz, 1H), 4.55 (d, *J* = 8.0 Hz, 1H), 4.52 (d, *J* = 8.0 Hz, 1H), 4.29 – 3.49 (m, 22H), 3.40 – 3.14 (m, 8H), 3.06 – 3.01 (m, 2H), 1.76 – 1.65 (m, 4H), 1.48 (ddd, *J* = 10.8, 8.7, 4.9 Hz, 2H).; $^{13}$C NMR (151 MHz, $D_2O$) δ 102.78, 102.12, 102.04, 79.11, 78.00, 75.19, 74.65, 74.54, 74.48, 74.17, 74.13, 73.33, 72.97, 72.88, 72.81, 70.15, 68.98, 61.89, 61.86, 59.85, 52.23, 46.62, 39.99, 39.93, 39.30, 28.09, 26.30, 21.96.; $^{31}$P NMR (243 MHz, $D_2O$) δ 0.30, -0.04.; m/z (HRMS+) 836.2854 $[M+H]^+$ ($C_{27}H_{56}N_3O_{22}P_2$ requires 836.2831).

**RP-HPLC of 88 (ELSD trace, Method $F_1$, $t_R$= 22.2 min)**



### 7.8.4. Synthesis of PA$_5$

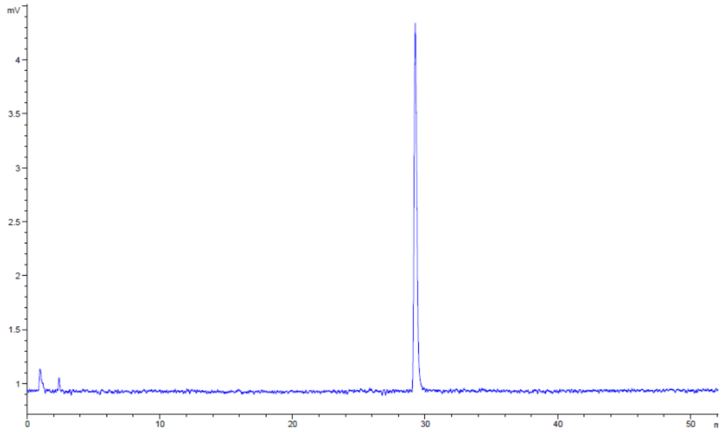


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **5x38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C, D, E** | **C:** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | | **L, P** | **P:** (Method **$Q_2$**, $t_R$ = 37.2 min) |
| | | **M, P1, P2,** | **M:** (4 equiv. of **85** per -OH)<br>**J1:** (Method **R**)<br>**J2:** (Method **S**) |
| | | **N,P1, O, P2** | **J1:** (Method **S**) **J2:** (Method **U** and **T**) |

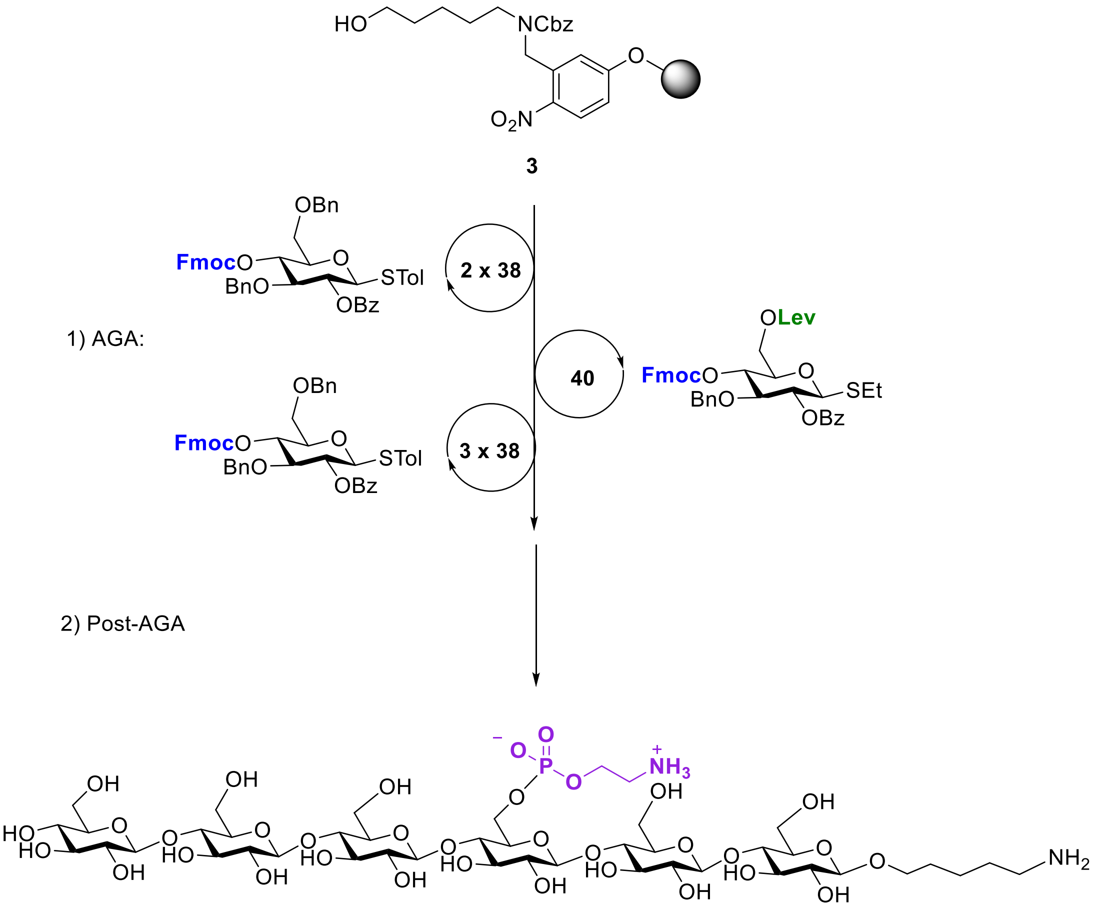
Compound **$PA_5$** was obtained as a white solid (2.4 mg, 16% overall yield).

Analytical data for **$PA_5$:** $^1H$ NMR (600 MHz, $D_2O$) δ 4.55 (dd, *J* = 7.9, 2.3 Hz, 5H), 4.50 (d, *J* = 8.0 Hz, 1H), 4.20 (ddd, *J* = 11.5, 5.3, 2.1 Hz, 1H), 4.16 – 4.11 (m, 2H), 4.09 (dt, *J* = 11.1, 5.4 Hz, 1H), 4.03 – 3.93 (m, 6H), 3.83 (m, 5H), 3.73 – 3.62 (m, 16H), 3.53 (dd, *J* = 8.6, 6.9 Hz, 2H), 3.41 – 3.28 (m, 8H), 3.05 – 3.00 (m, 2H), 1.76 – 1.66 (m, 4H), 1.51 – 1.43 (m, 2H); $^{13}C$ NMR (151 MHz, $D_2O$) δ 102.73, 102.28, 102.20, 101.94, 79.02, 78.51, 78.20, 75.20, 74.75, 74.71, 74.65, 74.49, 74.28, 74.01, 73.93, 72.98, 72.86, 72.81, 70.02, 68.97, 61.89, 59.78, 39.99, 39.29, 28.09, 26.34, 22.01; $^{31}P$ NMR (243 MHz, $D_2O$) δ 0.30; m/z (HRMS+) 600.2223 $[M+2H]^{2+}$ ($C_{43}H_{81}N_2O_{34}P$ requires 600.2198).

**RP-HPLC of $PA_5$ (ELSD trace, Method F, $t_R$= 29.2 min)**



### 7.8.5. Synthesis of $A_3PA_2$



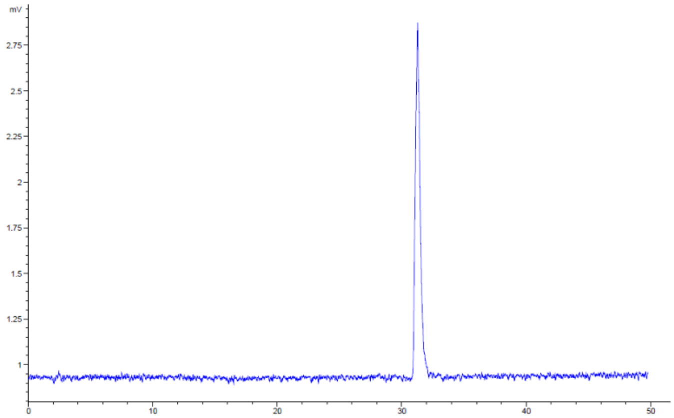
$A_3PA_2$

| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | A | |
| | 2x38 | B, C, D, E | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | 40 | B, C, D, E | **C:** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | 3x38 | B, C, D, E, D, E2 | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | | **L, P** | **P:** (Method **$Q_2$**, $t_R$ = 36.9 min) |
| | | **M, P1, P2,** | **M:** (4 equiv. of **85** per -OH)<br>**P1:** (Method **R**)<br>**P2:** (Method **S**) |
| | | **N,P1, O, P2** | **P1:** (Method **S**) **P2:** (Method **U** and **T**) |

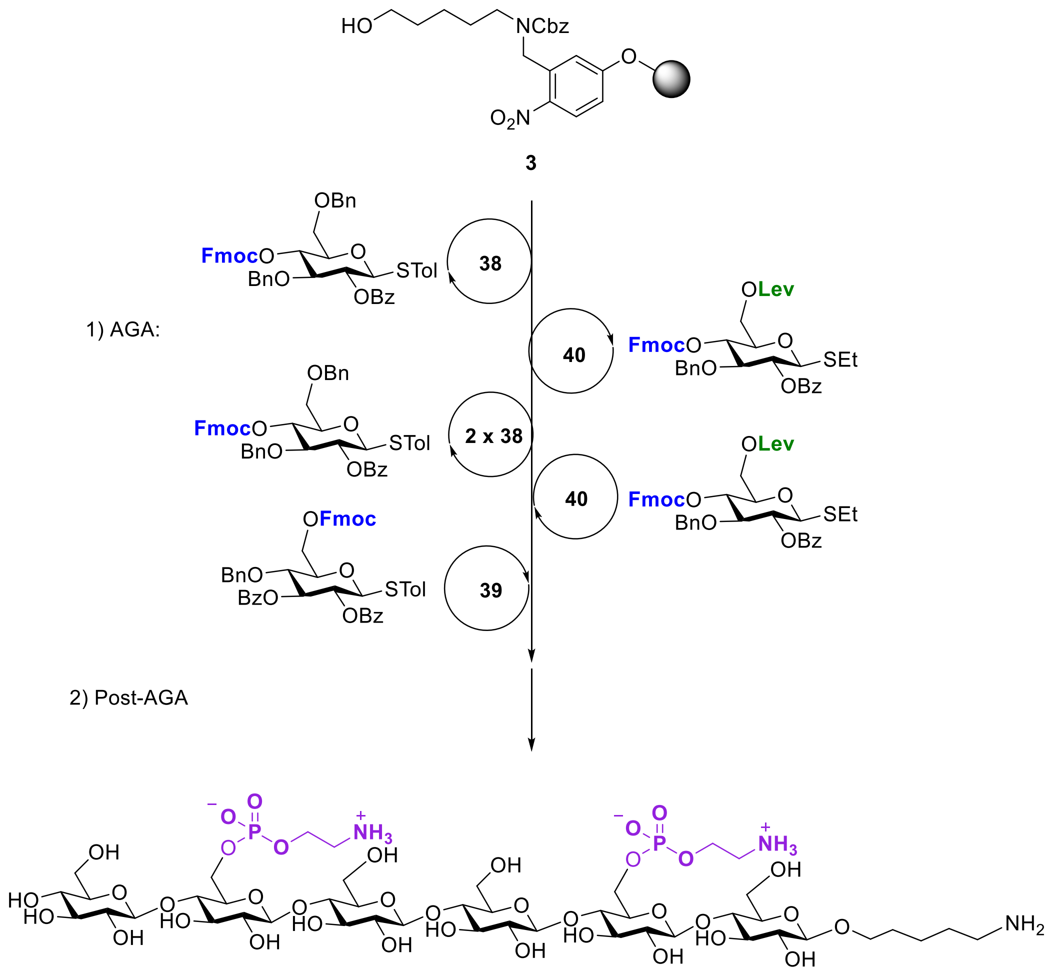
Compound **$A_3PA_2$** was obtained as a white solid (1.3 mg, 11% overall yield).

Analytical data for **$A_3PA_2$:** $^{1}$H NMR (700 MHz, $D_2O$) δ 4.63 – 4.47 (m, 6H), 4.27 (dd, *J* = 10.8, 5.0 Hz, 1H), 4.21 – 4.08 (m, 3H), 4.03 – 3.89 (m, 6H), 3.87 – 3.58 (m, 21H), 3.56 – 3.26 (m, 11H), 3.06 – 2.98 (m, 2H), 1.93 (s, 2H), 1.72 – 1.64 (m, 4H), 1.53 – 1.43 (m, 2H)*;* $^{13}$C NMR (176 MHz, $D_2O$) δ 102.49, 102.28, 101.95, 75.91, 74.75, 74.27, 73.97, 73.80, 72.85, 70.02, 69.37, 59.82, 39.28, 28.08, 26.34, 22.00*;* $^{31}$P NMR (243 MHz, $D_2O$) δ -0.01; m/z (HRMS+) 600.2222 $[M+2H]^{2+}$ ($C_{43}H_{81}N_2O_{34}P$ requires 600.2198).

**RP-HPLC of $A_3PA_2$ (ELSD trace, Method F, $t_R$= 31.2 min)**



## 7.8.6. Synthesis of (APA)$_2$



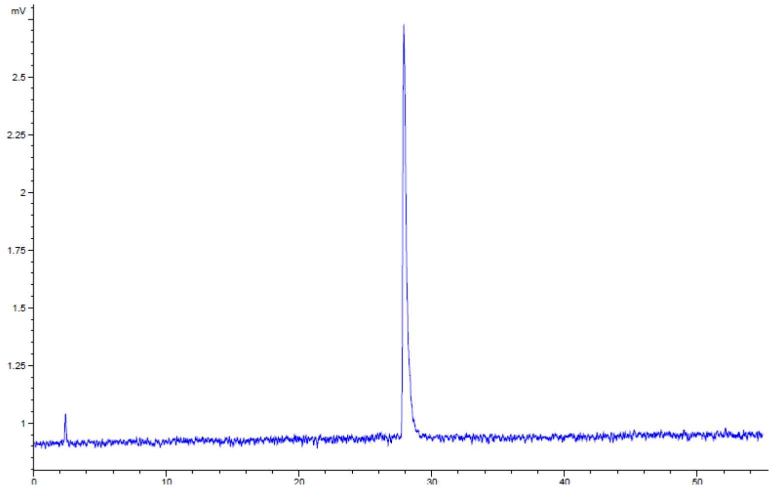
**(APA)$_2$**

| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **40** | **B, C, D, E** | **C:** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **2x38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **40** | **B, C, D, E** | **C:** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C, D, E, D, E2** | **C:** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | | **L, P** | **P:** (Method **Q$_2$**, $t_R$ = 46.1 min) |
| | | **M, P1, P2,** | **M:** (5 equiv. of **4** per -OH)<br>**P1:** (Method **B**)<br>**P2:** (Method **C**) |
| | | **N,P1, O, P2** | **P1:** (Method **S**) **P2:** (Method **U** and **T**) |

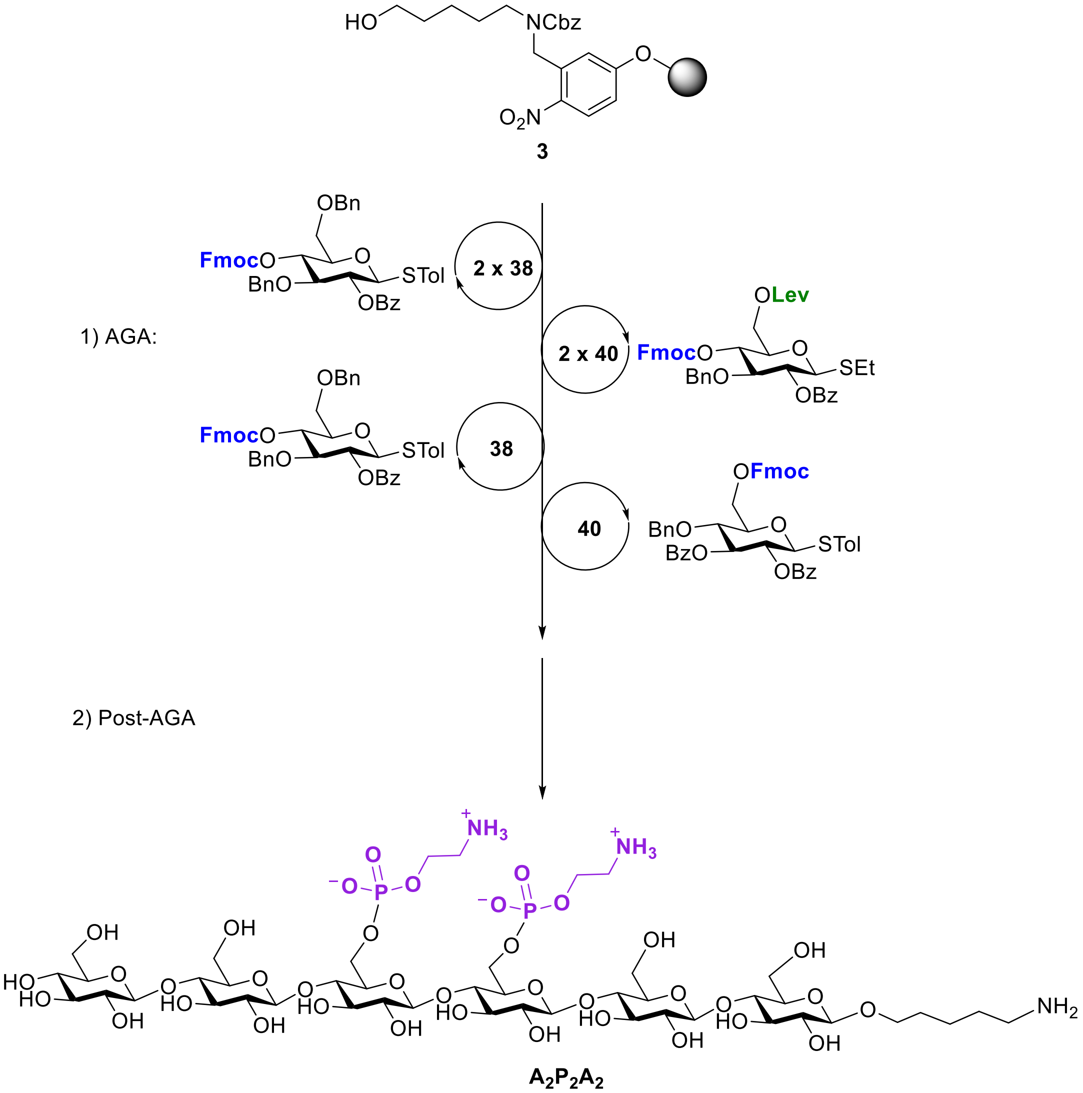
Compound **(APA)$_2$** was obtained as a white solid (4 mg, 11% overall yield).

Analytical data for **(APA)$_2$:** $^1$H NMR (600 MHz, $D_2O$) δ 4.64 – 4.46 (m, 6H), 4.31 – 3.27 (m, 46H), 3.04 – 3.00 (m, 2H), 1.70 (dp, *J* = 13.9, 7.2 Hz, 4H), 1.48 (q, *J* = 8.0 Hz, 2H)*;* $^{13}$C NMR (151 MHz, $D_2O$) δ 102.51, 102.49, 102.37, 102.23, 102.16, 101.91, 79.16, 78.84, 78.17, 77.80, 75.91, 75.46, 74.75, 74.67, 74.62, 74.32, 73.96, 73.83, 73.34, 73.17, 72.94, 72.83, 72.79, 70.03, 69.43, 61.84, 60.50, 59.97, 59.81, 40.01, 39.96, 39.28, 28.09, 26.33, 22.01*;* $^{31}$P NMR (243 MHz, $D_2O$) δ -0.02, -0.05; m/z (HRMS+) 661.7258 $[M+2H]^{+2}$ ($C_{45}H_{87}N3O_{37}P_2$ requires 661.7241).

**RP-HPLC of (APA)$_2$ (ELSD trace, Method F, $t_R$= 27.8 min)**



### 7.8.7. Synthesis of $A_2P_2A_2$

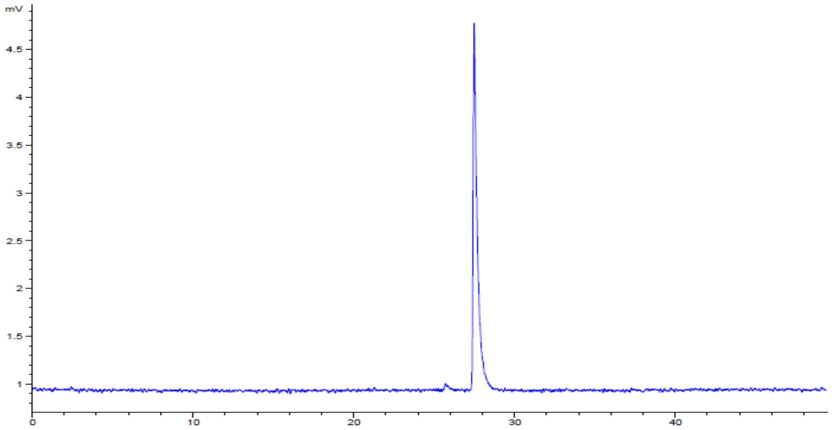


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **2x38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **2X40** | **B, C, D, E** | **C:** 40, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C, D, E, D, E2** | **C:** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | | **L, P** | **P:** (Method **$Q_2$**, $t_R$ = 41.2 min) |
| | | **M, P1, P2,** | **M:** (5 equiv. of **4** per -OH)<br>**P1:** (Method **R**)<br>**P2:** (Method **S**) |
| | | **N,P1, O, P2** | **P1:** (Method **S**) **P2:** (Method **U** and **T**) |

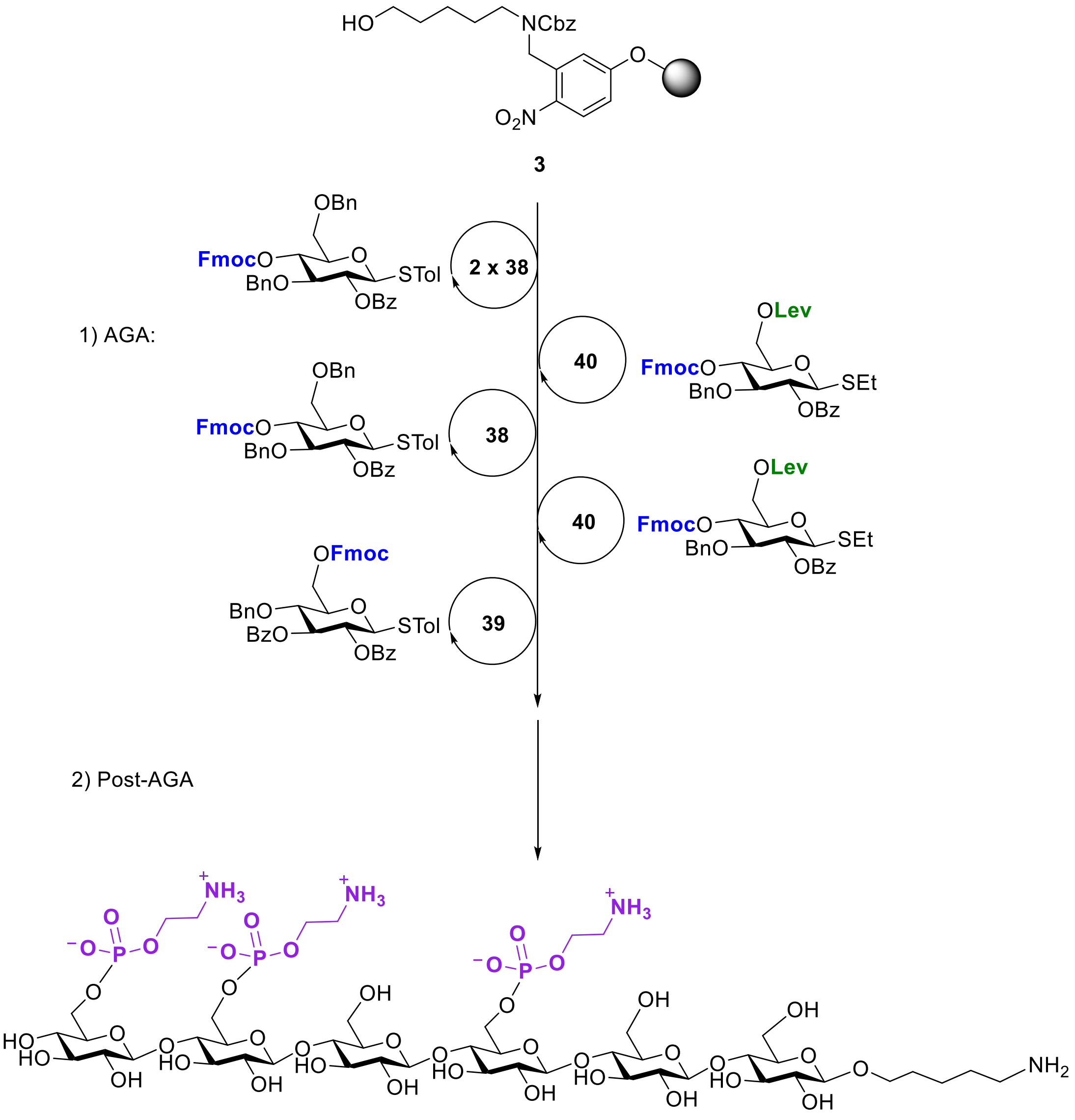
Compound **$A_2P_2A_2$** was obtained as a white solid (3.3 mg, 9 % overall yield).

Analytical data for **$A_2P_2A_2$:** $^1$H NMR (600 MHz, $D_2O$) δ 4.61 (t, *J* = 8.4 Hz, 2H), 4.57 (d, *J* = 7.9 Hz, 1H), 4.55 (d, *J* = 7.9 Hz, 1H), 4.52 (d, *J* = 7.9 Hz, 1H), 4.50 (d, *J* = 8.0 Hz, 1H), 4.27 (dd, *J* = 11.0, 4.9 Hz, 2H), 4.22 – 4.07 (m, 6H), 4.03 – 3.90 (m, 5H), 3.86 – 3.59 (m, 19H), 3.55 – 3.47 (m, 2H), 3.45 – 3.27 (m, 11H), 3.05 – 3.00 (m, 2H), 1.74 – 1.66 (m, 4H), 1.51 – 1.44 (m, 2H)); $^{13}$C NMR (151 MHz, $D_2O$) δ 102.49, 102.45, 102.24, 102.13, 101.94, 78.79, 78.47, 78.41, 78.33, 77.66, 75.91, 75.39, 74.75, 74.70, 74.68, 74.27, 74.00, 73.98, 73.85, 73.83, 73.33, 73.07, 72.93, 72.87, 72.85, 72.75, 70.03, 69.38, 63.81, 61.90, 61.87, 60.50, 59.93, 59.82, 58.38, 46.62, 40.02, 39.97, 39.29, 28.09, 26.34, 23.18, 22.01, 8.15; $^{31}$P NMR (243 MHz, $D_2O$) δ -0.01, -0.05; m/z (HRMS+) 661.7283 [M+2H]$^{+2}$ ($C_{45}H_{87}N3O_{37}P_2$ requires 661.7241).

**RP-HPLC of $A_2P_2A_2$ (ELSD trace, Method F, $t_R$= 27.4 min)**



## 7.8.8. Synthesis of P₂APA₂



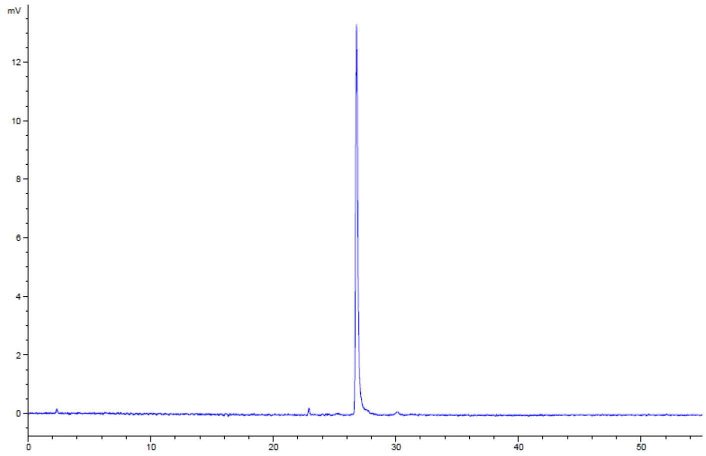
**$P_2APA_2$**

| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **2x38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **40** | **B, C, D, E** | **C:** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **40** | **B, C, D, E** | **C:** (**40 -**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C, D, E, E2** | **C:** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | | **L, P** | **P:** (Method **Q$_2$**, t$_R$ = 48.7 min) |
| | | **M, P1, P2,** | **M:** (5 equiv. of **4** per -OH)<br>**P1:** (Method **R**)<br>**P2:** (Method **S**) |
| | | **N,P1, O, P2** | **P1:** (Method **S**) **P2:** (Method **U** and **T**) |

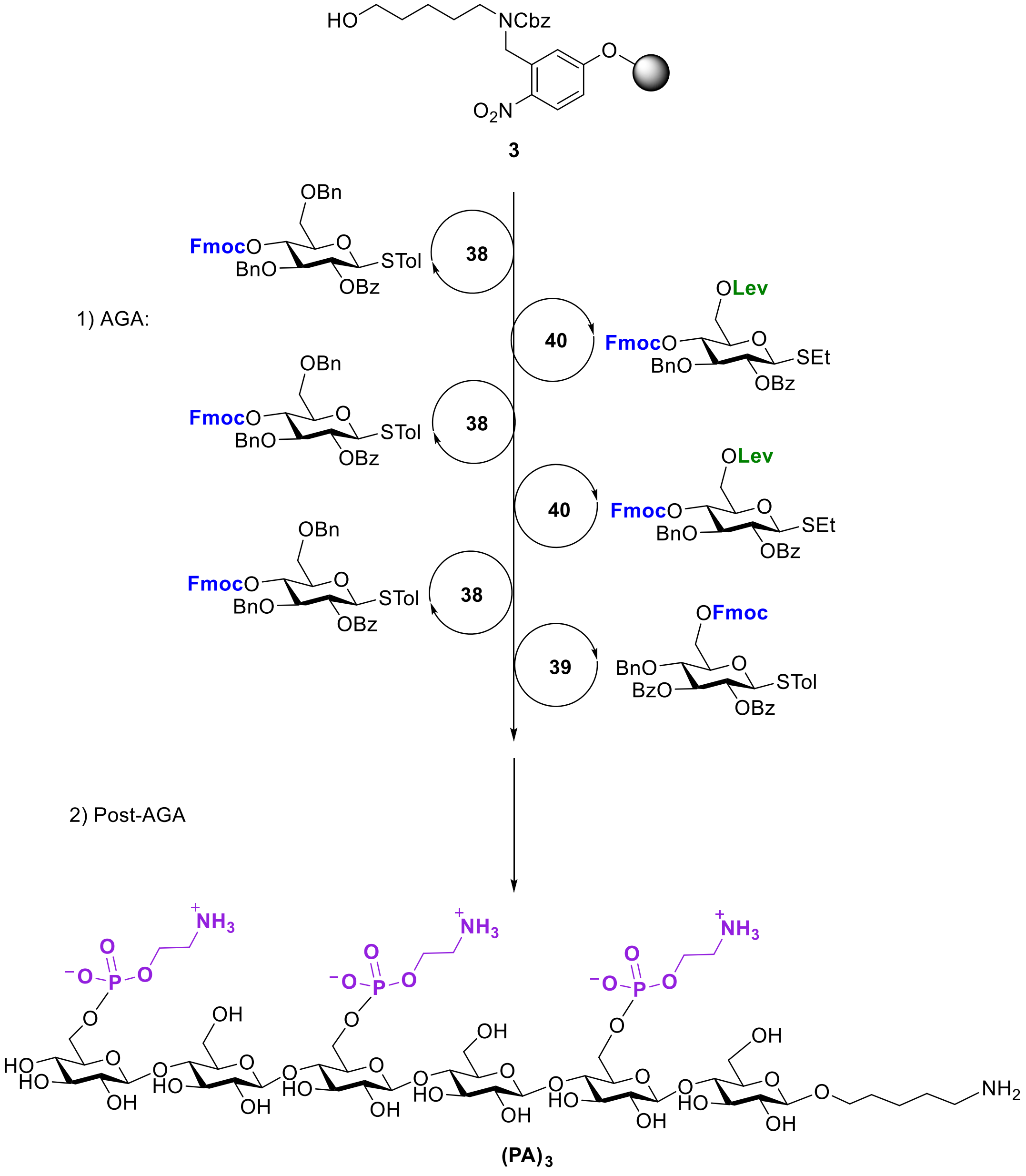
Compound **P$_2$APA$_2$** was obtained as a white solid (3.1 mg, 8 % overall yield).

Analytical data for **P$_2$APA$_2$:** $^{1}$H NMR (600 MHz, d$_2$o) δ 4.58 (ddd, *J* = 20.7, 10.4, 6.8 Hz, 5H), 4.50 (d, *J* = 7.9 Hz, 1H), 4.32 – 4.05 (m, 11H), 4.03 – 3.90 (m, 4H), 3.89 – 3.48 (m, 22H), 3.47 – 3.23 (m, 11H), 3.02 (t, *J* = 7.6 Hz, 2H), 1.70 (dp, *J* = 14.1, 7.2 Hz, 4H), 1.47 (p, *J* = 7.8 Hz, 2H)*;* $^{13}$C NMR (151 MHz, D$_2$O) δ 102.49, 102.45, 102.24, 102.13, 101.94, 78.79, 78.47, 78.41, 78.33, 77.66, 75.91, 75.39, 74.75, 74.70, 74.68, 74.27, 74.00, 73.98, 73.85, 73.83, 73.33, 73.07, 72.93, 72.87, 72.85, 72.75, 70.03, 69.38, 63.81, 61.90, 61.87, 60.50, 59.93, 59.82, 58.38, 46.62, 40.02, 39.97, 39.29, 28.09, 26.34, 23.18, 22.01, 8.15*;* $^{31}$P NMR (243 MHz, D$_2$O) δ 0.31, 0.04, -0.02*;* m/z (HRMS+) 482.4858 [M+3H]$^{+3}$ ($C_{47}H_{94}N_4O_{40}P_3$ requires 482.4880).

**RP-HPLC of P$_2$APA$_2$ (ELSD trace, Method F, t$_R$= 26.7 min)**



### 7.8.9. Synthesis of (PA)$_3$

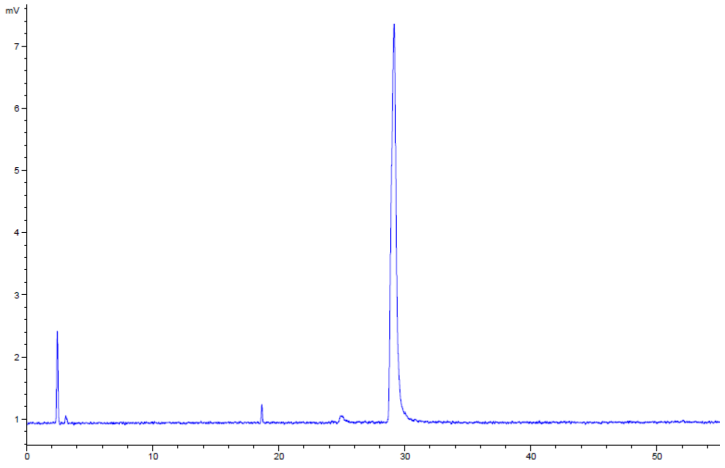


| Step | | Modules | Notes |
|---|---|---|---|
| AGA | | **A** | |
| | **38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **40** | **B, C, D, E** | **C:** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **40** | **B, C, D, E** | **C:** (**40**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **38** | **B, C, D, E** | **C:** (**38**, **-**20 °C for 5 min, 0 °C for 20 min) |
| | **39** | **B, C, D, E, E2** | **C:** (**39**, **-**20 °C for 5 min, 0 °C for 20 min) |
| Post-AGA | | **L, P** | **P** (Method **$Q_2$**, $t_R$ = 46.1 min) |
| | | **M, P1, P2,** | **M:** (5 equiv. of **4** per -OH)<br>**P1:** (Method **R**)<br>**P2:** (Method **S**) |
| | | **N,P1, O, P2** | **P1:** (Method **S**) **P2:** (Method **U** and **T**) |

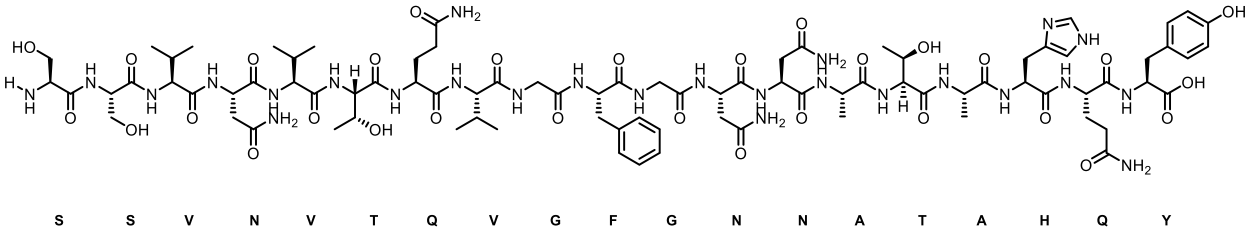
Compound **(PA)$_3$** was obtained as a white solid (0.5 mg, 3 % overall yield).

Analytical data for **(PA)$_3$:** $^1$H NMR (600 MHz, $D_2O$) δ 4.63 – 4.52 (m, 5H), 4.50 (d, *J* = 8.0 Hz, 1H), 4.26 (t, *J* = 6.9 Hz, 2H), 4.18 (td, *J* = 13.8, 11.4, 8.5 Hz, 3H), 4.15 – 4.08 (m, 6H), 4.03 – 3.93 (m, 4H), 3.87 – 3.61 (m, 19H), 3.57 – 3.50 (m, 2H), 3.40 – 3.31 (m, 4H), 3.25 (m, 5H), 3.02 (t, *J* = 7.6 Hz, 2H), 1.76 – 1.64 (m, 4H), 1.47 (m, 2H)*;* $^{13}$C NMR (151 MHz, $D_2O$) δ 102.6, 102.5, 102.4, 102.1, 101.8, 78.8, 77.7, 75.17, 74.62, 72.79, 70.1, 70.01, 68.96, 69.7, 63.25, 59.7, 40.02, 39.7, 39.29, 28.3, 26.35, 22.01, 21.8*;* $^{31}$P NMR (243 MHz, $D_2O$) δ 0.37, 0.05, 0.03*;* m/z (HRMS+) 482.4851 $[M+3H]^{+3}$ ($C_{47}H_{94}N_4O_{40}P_3$ requires 482.4880).

**RP-HPLC of (PA)$_3$ (ELSD trace, Method F, $t_R$= 29.1 min)**



## 7.9. Synthesis of R5



**R5**

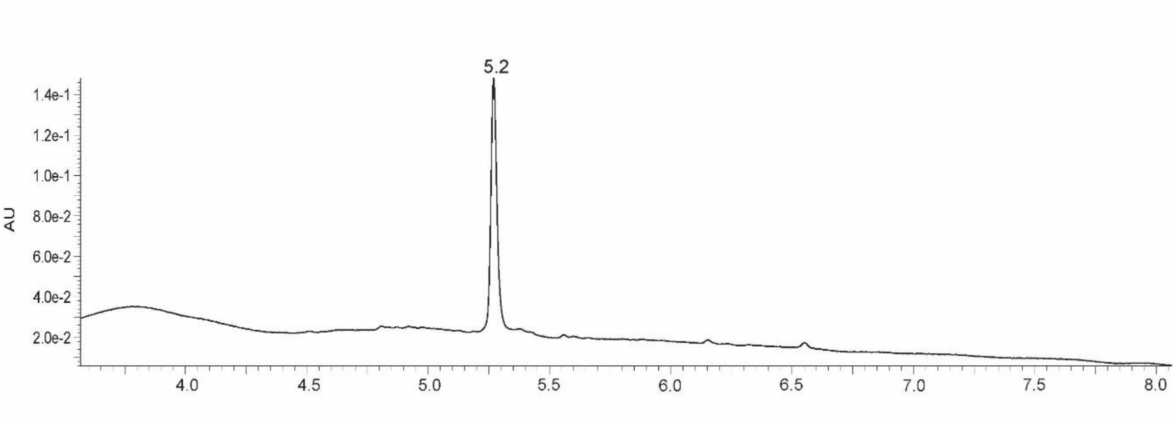
The solid-phase peptides synthesis was carried out with a microwave-assisted peptide synthesizer (Liberty Blue, CEM, USA). 2-Cl-Trt-Cl Protide resin was swollen in dichloromethane for 30 min. The first amino acid was coupled manually using 4 equiv. of Fmoc-Tyr(OtBu)-OH (with respect to the resin loading) and 8 equiv. of diisopropylethylamine (DIEA) in 3 mL of DCM, shaking at RT overnight. The resin was washed with DMF and DCM. Unreacted groups were capped with a solution of DCM/$CH_3OH$/DIEA (17:2:1 v/v/v) (1 h at RT). Then, the resin was washed with DMF, DCM and DMF and transferred to the reaction vessel of the synthesizer. A resin loading of 0.24 mmol/g was determined using Fmoc quantification at 290 nm. Peptide coupling was performed with 0.25 M solutions of Fmoc-Xaa-COOH in DMF, 1 M ethyl cyano(hydroxyimino)acetate (Oxyma Pure®) in DMF and 0.5 M N,N′-diisopropylcarbodiimide (DIC) for activation (5 min at 75°C). The fluorenylmehtyloxycarbonyl (Fmoc) was removed with 20% piperidine in DMF solution (3 min at 75°C). Histidine coupling required a double cycle performed at RT (2 x 1 h ). After the fifth residue, all coupling and deprotection cycles were performed twice. After synthesis, the resin was dried under vacuum and the peptide was cleaved by treatment with trifluoroacetic acid containing 2.5% of $H_2O$ and 2.5% triisopropylsilane as scavengers (2 h at RT). The cleaved peptide was precipitated and washed with ice-cold diethyl ether three times and analyzed by RP-HPLC as described below. Peptide **R5** was obtained as a fluffy white powder (7 mg, 11%).

### Purification

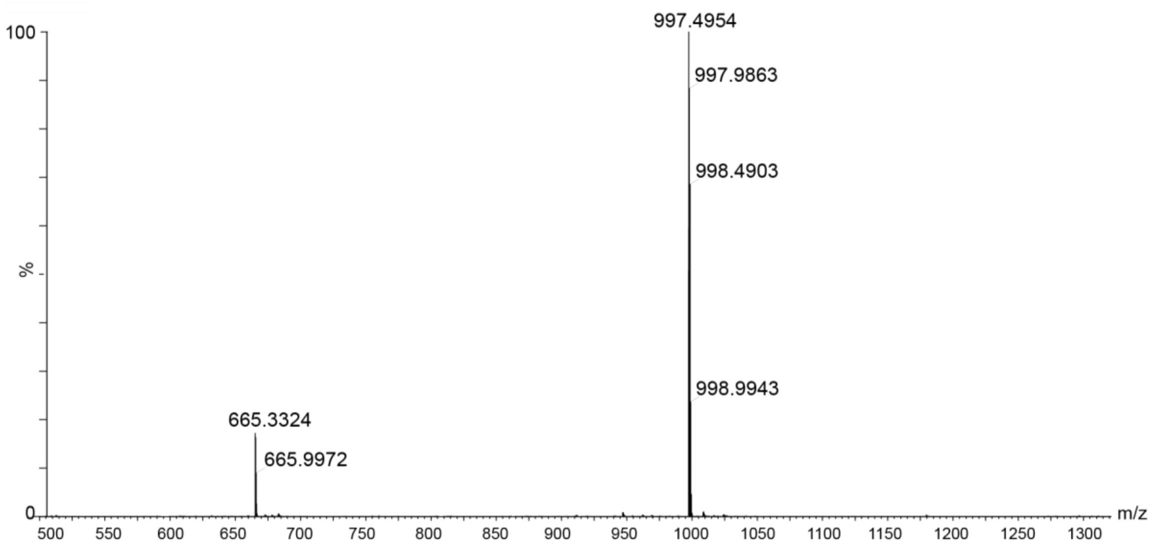
After cleavage, the crudes were analyzed with **Method W** using analytical HPLC Agilent HP 1100 and purified with **Method X**. The final pure compound was analyzed by LC-MS using a Waters Acquity UPLC coupled to a Xevo Q2-XS Qtof (**Method Y**).

- **Method W:** (YMC Hydrosphere C18 column 50 mm X 3.0mm, S-3 μm) flow rate of 0.5 mL/min with 5% ACN in $H_2O$ (0.1% formic acid) as eluent [isocratic 5% (2min)], linear gradient to 70% of ACN (15 min), linear gradient to 100% ACN (3 min)]
- **Method X:** (Manual reverse phase $C_{18}$ silica gel column chromatography) solvent system 0.1% formic acid in $H_2O$:MeOH, gradient from 0 to 100% MeOH.
- **Method Y**: (YMC Hydrosphere C18 column 50 mm X 3.0mm, S-3 μm) flow rate of 0.5 mL/min with 5%ACN (0.1% formic acid) in H2O (0.1% formic acid) as eluent [isocratic 5% (2min)], linear gradient to 70% of CAN (0.1% formic acid) (13 min), linear gradient to 100% ACN (3 min)]

**RP-HPLC of R5 (UV 214nm trace, Method Y, $t_R$= 5.2 min)**



**ESI-HRMS of R5**



# 8. References

[1] H.-J. Gabius, J. Roth, *Histochem. Cell Biol.* **2017**, *147*, 111–117.

[2] I. Bucior, M. M. Burger, *Curr. Opin. Struct. Biol.* **2004**, *14*, 631–637.

[3] M. Rinaudo, *Polym. Int.* **2008**, *57*, 397–430.

[4] D. de Wit, L. Maat, A. P. G. Kieboom, *Ind. Crops Prod.* **1993**, *2*, 1–12.

[5] K. J. Edgar, C. M. Buchanan, J. S. Debenham, P. A. Rundquist, B. D. Seiler, M. C. Shelton, D. Tindall, *Prog. Polym. Sci.* **2001**, *26*, 1605–1688.

[6] L. Fu, J. Zhang, G. Yang, *Carbohydr. Polym.* **2013**, *92*, 1432–1442.

[7] H. C. Arca, L. I. Mosquera-Giraldo, V. Bi, D. Xu, L. S. Taylor, K. J. Edgar, *Biomacromolecules* **2018**, *19*, 2351–2376.

[8] D. Klemm, B. Heublein, H.-P. Fink, A. Bohn, *Angew. Chem. Int. Ed.* **2005**, *44*, 3358–3393.

[9] J. Zhang, J. Wu, J. Yu, X. Zhang, J. He, J. Zhang, *Mater. Chem. Front.* **2017**, *1*, 1273–1290.

[10] Y. Nishiyama, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *376*, 20170047.

[11] P. Kumar Gupta, S. Sai Raghunath, D. Venkatesh Prasanna, P. Venkat, V. Shree, C. Chithananthan, S. Choudhary, K. Surender, K. Geetha, in *Cellulose*, IntechOpen, **2019**, 1-21.

[12] B. Medronho, A. Romano, M. G. Miguel, L. Stigsson, B. Lindman, *Cellulose* **2012**, *19*, 581–587.

[13] Y. Nishiyama, P. Langan, H. Chanzy, *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.

[14] D. Ciolacu, V. I. Popa, *Cellulose Allomorphs - Overview and Perspectives*, **2010,** 1-38.

[15] J. K. Polko, J. J. Kieber, *Plant Cell* **2019**, *31*, 282–296.

[16] D. P. Delmer, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 245–276.

[17] M. Liu, L. Liu, S. Jia, S. Li, Y. Zou, C. Zhong, *Sci. Rep.* **2018**, *8*, 6266.

[18] U. Römling, *Res. Microbiol.* **2002**, *153*, 205–212.

[19] Gorgieva, Trček, *Nanomaterials* **2019**, *9*, 1352.

[20] W. Thongsomboon, D. O. Serra, A. Possling, C. Hadjineophytou, R. Hengge, L. Cegelski, *Science* **2018**, *359*, 334–338.

[21] J. Jeffries, W. Thongsomboon, J. A. Visser, K. Enriquez, D. Yager, L. Cegelski, *Biopolymers* **2021**, *112*, 1–11.

[22] J. Jeffries, G. G. Fuller, L. Cegelski, *Microbiol. Insights* **2019**, *12*, 117863611986523.

[23] M. Rinaudo, *Prog. Polym. Sci.* **2006**, *31*, 603–632.

[24] E. F. Franca, R. D. Lins, L. C. G. Freitas, T. P. Straatsma, *J. Chem. Theory Comput.* **2008**, *4*, 2141–2149.

[25] I. Aranaz, M. Mengibar, R. Harris, B. Miralles, N. Acosta, L. Calderon, A. Sanchez, A. Heras, *Curr. Chem. Biol.* **2014**, *8*, 27–42.

[26] M. N. . Ravi Kumar, *React. Funct. Polym.* **2000**, *46*, 1–27.

[27] F. Liaqat, R. Eltem, *Carbohydr. Polym.* **2018**, *184*, 243–259.

[28] S. Basa, M. Nampally, T. Honorato, S. N. Das, A. R. Podile, N. E. El Gueddari, B. M. Moerschbacher, *J. Am. Chem. Soc.* **2020**, *142*, 1975–1986.

[29] M. X. Weinhold, J. C. M. Sauvageau, J. Kumirska, J. Thöming, *Carbohydr. Polym.* **2009**, *78*, 678–684.

[30] M. B. Kaczmarek, K. Struszczyk-Swita, X. Li, M. Szczęsna-Antczak, M. Daroch, *Front. Bioeng. Biotechnol.* **2019**, *7*, DOI 10.3389/fbioe.2019.00243.

[31] R. Enugala, L. C. R. Carvalho, M. J. Dias Pires, M. M. B. Marques, *Chem. – An Asian J.* **2012**, *7*, 2482–2501.

[32] Y. Yang, B. Yu, *Tetrahedron* **2014**, *70*, 1023–1046.

[33] K. Paschinger, I. B. H. Wilson, *Glycoconj. J.* **2020**, *37*, 27–40.

[34] B. A. Cobb, D. L. Kasper, *Cell. Microbiol.* **2005**, *7*, 1398–1403.

[35] Q. Zhang, A. Gimeno, D. Santana, Z. Wang, Y. Valdés-Balbin, L. M. Rodríguez-Noda, T. Hansen, L. Kong, M. Shen, H. S. Overkleeft, V. Vérez-Bencomo, G. A. van der Marel, J. Jiménez-Barbero, F. Chiodo, J. D. C. Codée, *ACS Cent. Sci.* **2019**, *5*, 1407–1416.

[36] M. Mende, C. Bednarek, M. Wawryszyn, P. Sauter, M. B. Biskup, U. Schepers, S. Bräse, *Chem. Rev.* **2016**, *116*, 8193–8255.

[37] M. Yu, T. Zhang, W. Zhang, Q. Sun, H. Li, J. Li, *Front. Mol. Biosci.* **2021**, *7*, 1–13.

[38] S. Ricard-Blum, *Perspect. Sci.* **2017**, *11*, 62–69.

[39] Q. Zhang, C. Z. Chen, M. Swaroop, M. Xu, L. Wang, J. Lee, A. Q. Wang, M. Pradhan, N. Hagen, L. Chen, M. Shen, Z. Luo, X. Xu, Y. Xu, W. Huang, W. Zheng, Y. Ye, *Cell Discov.* **2020**, *6*, 80.

[40] J. M. Mattson, R. Turcotte, Y. Zhang, *Biomech. Model. Mechanobiol.* **2017**, *16*, 213–225.

[41] K. Arnold, Y.-E. Liao, J. Liu, *Biomedicines* **2020**, *8*, 503.

[42] A. Badri, A. Williams, R. J. Linhardt, M. A. Koffas, *Curr. Opin. Biotechnol.* **2018**, *53*, 85–92.

[43] C. I. Gama, S. E. Tully, N. Sotogaku, P. M. Clark, M. Rawat, N. Vaidehi, W. A. Goddard, A. Nishi, L. C. Hsieh-Wilson, *Nat. Chem. Biol.* **2006**, *2*, 467–473.

[44] H. Habuchi, O. Habuchi, K. Kimata, *Glycoconj. J.* **2004**, *21*, 47–52.

[45] M. Guerrini, D. Beccati, Z. Shriver, A. Naggi, K. Viswanathan, A. Bisio, I. Capila, J. C. Lansing, S. Guglieri, B. Fraser, A. Al-Hakim, N. S. Gunay, Z. Zhang, L. Robinson, L. Buhse, M. Nasr, J. Woodcock, R. Langer, G. Venkataraman, R. J. Linhardt, B. Casu, G. Torri, R. Sasisekharan, *Nat. Biotechnol.* **2008**, *26*, 669–675.

[46] A. Vasconcelos, V. Pomin, *Microorganisms* **2017**, *5*, 51.

[47] V. Pomin, *Pharmaceuticals* **2015**, *8*, 848–864.

[48] G. Fittolani, T. Tyrikos-Ergas, D. Vargová, M. A. Chaube, M. Delbianco, *Beilstein J. Org. Chem.* **2021**, *17*, 1981–2025.

[49] J. Liu, R. J. Linhardt, *Nat. Prod. Rep.* **2014**, *31*, 1676–1685.

[50] Y. Yu, M. Shen, Q. Song, J. Xie, *Carbohydr. Polym.* **2018**, *183*, 91–101.

[51] X. Zhang, V. Pagadala, H. M. Jester, A. M. Lim, T. Q. Pham, A. M. P. Goulas, J. Liu, R. J. Linhardt, *Chem. Sci.* **2017**, *8*, 7932–7940.

[52] D. Dube, *Curr. Opin. Chem. Biol.* **2003**, *7*, 616–625.

[53] Y. Wu, D.-C. Xiong, S.-C. Chen, Y.-S. Wang, X.-S. Ye, *Nat. Commun.* **2017**, *8*, 14851.

[54] T. Nokami, R. Hayashi, Y. Saigusa, A. Shimizu, C. Y. Liu, K. K. T. Mong, J. I. Yoshida, *Org. Lett.* **2013**, *15*, 4520–4523.

[55] S. L. Tang, L. B. Linz, B. C. Bonning, N. L. B. Pohl, *J. Org. Chem.* **2015**, *80*, 10482–10489.

[56] N. V. Ganesh, K. Fujikawa, Y. H. Tan, K. J. Stine, A. V. Demchenko, *Org. Lett.* **2012**, *14*, 3036–3039.

[57] P. H. Seeberger, *Acc. Chem. Res.* **2015**, *48*, 1450–1463.

[58] R. A. Houghten, *Proc. Natl Acad. Sci. USA* **1985**, *82*, 5131–5135.

[59] A. Grajkowski, C. Ausín, J. S. Kauffman, J. Snyder, S. Hess, J. R. Lloyd, S. L. Beaucage, *J. Org. Chem.* **2007**, *72*, 805–815.

[60] M. Guberman, P. H. Seeberger, *J. Am. Chem. Soc.* **2019**, *141*, 5581–5592.

[61] X. Wu, M. Grathwohl, R. R. Schmidt, *Org. Lett.* **2001**, *3*, 747–750.

[62] M. T. C. Walvoort, A. G. Volbeda, N. R. M. Reintjens, H. Van Den Elst, O. J. Plante, H. S. Overkleeft, G. A. Van Der Marel, J. D. C. Codée, *Org. Lett.* **2012**, *14*, 3776–3779.

[63] K. Le Mai Hoang, A. Pardo-Vargas, Y. Zhu, Y. Yu, M. Loria, M. Delbianco, P. H. Seeberger, *J. Am. Chem. Soc.* **2019**, *141*, 9079–9086.

[64] S. Eller, M. Collot, J. Yin, H. S. Hahm, P. H. Seeberger, *Angew. Chem. Int. Ed.* **2013**, *52*, 5858–5861.

[65] Y. Yu, A. Kononov, M. Delbianco, P. H. Seeberger, *Chem. Eur. J.* **2018**, *24*, 6075–6078.

[66] A. A. Joseph, A. Pardo-Vargas, P. H. Seeberger, *J. Am. Chem. Soc.* **2020**, *142*, 8561–8564.

[67] M. Delbianco, A. Kononov, A. Poveda, Y. Yu, T. Diercks, J. Jiménez-Barbero, P. H. Seeberger, *J. Am. Chem. Soc.* **2018**, *140*, 5421–5426.

[68] I. Moraes, A. Quigley, *Biology (Basel).* **2021**, *10*, 245.

[69] M. Nagae, M. Kanagawa, K. Morita-Matsumoto, S. Hanashima, Y. Kizuka, N. Taniguchi, Y. Yamaguchi, *Sci. Rep.* **2016**, *6*, 22973.

[70] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.

[71] K. B. Wiberg, H. Castejon, W. F. Bailey, J. Ochterski, *J. Org. Chem.* **2000**, *65*, 1181–1187.

[72] R. J. Woods, *Chem. Rev.* **2018**, *118*, 8005–8024.

[73] S. Jo, Y. Qi, W. Im, *Glycobiology* **2015**, *26*, cwv083.

[74] A. Almond, *Carbohydr. Res.* **2005**, *340*, 907–920.

[75] A. Imberty, S. Pérez, *Chem. Rev.* **2000**, *100*, 4567–4588.

[76] G. Widmalm, *Carbohydr. Res.* **2013**, *378*, 123–132.

[77] B. M. Sattelle, B. Bose-Basu, M. Tessier, R. J. Woods, A. S. Serianni, A. Almond, *J. Phys. Chem. B* **2012**, *116*, 6380–6386.

[78] S. Jo, D. Myatt, Y. Qi, J. Doutch, L. A. Clifton, W. Im, G. Widmalm, *J. Phys. Chem. B* **2018**, *122*, 1169–1175.

[79] Ł. Nierzwicki, G. Palermo, *Front. Mol. Biosci.* **2021**, *8*, 1–6.

[80] O. Fisette, P. Lagüe, S. Gagné, S. Morin, *J. Biomed. Biotechnol.* **2012**, *2012*, 1–12.

[81] H.-J. Steinhoff, M. Müller, C. Beier, M. Pfeiffer, *J. Mol. Liq.* **2000**, *84*, 17–27.

[82] M. Karplus, J. A. McCammon, *Nat. Struct. Biol.* **2002**, *9*, 646–652.

[83] R. K. Karmani, G. Agha, M. S. Squillante, J. Seiferas, M. Brezina, J. Hu, R. Tuminaro, P. Sanders, J. L. Träffe, R. A. Geijn, J. L. Träff, R. A. Geijn, M. B. Sander, J. L. Gustafson, R. O. Dror, C. Young, D. E. Shaw, C. Lin, J.-K. Lee, R.-G. Chang, C.-B. Kuan, G. Kollias, A. Y. Grama, Z. Li, R. C. Whaley, R. W. Vuduc, in *Encycl. Parallel Comput.*, Springer US, Boston, MA, **2011**, 60–71.

[84] H. Yuan, Y. Nishiyama, M. Wada, S. Kuga, *Biomacromolecules* **2006**, *7*, 696–700.

[85] H. Yuan, Y. Nishiyama, M. Wada, S. Kuga, *Biomacromolecules* **2006**, *7*, 696–700.

[86] J. Sauter, A. Grafmüller, *J. Chem. Theory Comput.* **2016**, *12*, 4375–4384.

[87] D. Cremer, J. A. Pople, *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

[88] S. Rauschenbach, M. Ternes, L. Harnau, K. Kern, *Annu. Rev. Anal. Chem.* **2016**, *9*, 473–498.

[89] K. Anggara, Y. Zhu, G. Fittolani, Y. Yu, T. Tyrikos-Ergas, M. Delbianco, S. Rauschenbach, S. Abb, P. H. Seeberger, K. Kern, *Proc. Natl Acad. Sci. USA* **2021**, *118*, e2102168118.

[90] S. Gim, Y. Zhu, P. H. Seeberger, M. Delbianco, *WIREs Nanomedicine and Nanobiotechnology* **2019**, *11*, 1–29.

[91] L. P. Erwig, N. A. R. Gow, *Nat. Rev. Microbiol.* **2016**, *14*, 163–176.

[92] S. Naqvi, B. M. Moerschbacher, *Crit. Rev. Biotechnol.* **2017**, *37*, 11–25.

[93] K. Li, R. Xing, S. Liu, P. Li, *J. Agric. Food Chem.* **2020**, *68*, 12203–12211.

[94] K. Fuchs, Y. Cardona Gloria, O. Wolz, F. Herster, L. Sharma, C. A. Dillen, C. Täumer, S. Dickhöfer, Z. Bittner, T. Dang, A. Singh, D. Haischer, M. A. Schlöffel, K. J. Koymans, T. Sanmuganantham, M. Krach, T. Roger, D. Le Roy, N. A. Schilling, F. Frauhammer, L. S. Miller, T. Nürnberger, S. LeibundGut-Landmann, A. A. Gust, B. Macek, M. Frank, C. Gouttefangeas, C. S. Dela Cruz, D. Hartl, A. N. Weber, *EMBO Rep.* **2018**, *19*, 1–14.

[95] S. Basa, M. Nampally, T. Honorato, S. N. Das, A. R. Podile, N. E. El Gueddari, B. M. Moerschbacher, *J. Am. Chem. Soc.* **2020**, *142*, 1975–1986.

[96] S. Skovstrup, S. G. Hansen, T. Skrydstrup, B. Schiøtt, *Biomacromolecules* **2010**, *11*, 3196–3207.

[97] E. F. Franca, R. D. Lins, L. C. G. Freitas, T. P. Straatsma, *J. Chem. Theory Comput.* **2008**, *4*, 2141–2149.

[98] A. Almond, *Glycobiology* **2003**, *13*, 255–264.

[99] M. T. McDonnell, D. A. Greeley, K. M. Kit, D. J. Keffer, *J. Phys. Chem. B* **2016**, *120*, 8997–9010.

[100] Z. Yu, D. Lau, *J. Mol. Model.* **2015**, *21*, DOI 10.1007/s00894-015-2670-9.

[101] L. Tsereteli, A. Grafmüller, *An Accurate Coarse-Grained Model for Chitosan Polysaccharides in Aqueous Solution*, **2017**.

[102] J. D. Schneible, A. Singhal, R. L. Lilova, C. K. Hall, A. Grafmüller, S. Menegatti, *Biomacromolecules* **2019**, *20*, 3126–3141.

[103] S. N. Hamer, S. Cord-Landwehr, X. Biarnés, A. Planas, H. Waegeman, B. M. Moerschbacher, S. Kolkenbrock, *Sci. Rep.* **2015**, *5*, 8716.

[104] M. Delbianco, A. Kononov, A. Poveda, Y. Yu, T. Diercks, J. Jiménez-Barbero, P. H. Seeberger, *J.*

*Am. Chem. Soc.* **2018**, *140*, 5421–5426.

[105] M. Eddya, B. Tbib, K. EL-Hami, *Heliyon* **2020**, *6*, e03486.

[106] R. Stenutz, I. Carmichael, G. Widmalm, A. S. Serianni, *J. Org. Chem.* **2002**, *67*, 949–958.

[107] J. C. K. Quirke, D. Crich, *J. Am. Chem. Soc.* **2020**, *142*, 16965–16973.

[108] N. Taniguchi, T. Endo, G. W. Hart, P. H. Seeberger, C. H. Wong, *Glycoscience: Biology and Medicine*, Springer Japan, Tokyo, **2015**.

[109] S. M. Muthana, C. T. Campbell, J. C. Gildersleeve, *ACS Chem. Biol.* **2012**, *7*, 31–43.

[110] H. E. Caputo, J. E. Straub, M. W. Grinstaff, *Chem. Soc. Rev.* **2019**, *48*, 2338–2365.

[111] S. Ramadan, T. Li, W. Yang, J. Zhang, Z. Rashidijahanabad, Z. Tan, N. Parameswaran, X. Huang, *ACS Cent. Sci.* **2020**, *6*, 913–920.

[112] S. U. Hansen, G. J. Miller, M. J. Cliff, G. C. Jayson, J. M. Gardiner, *Chem. Sci.* **2015**, *6*, 6158–6164.

[113] Y. Xu, E. H. Pempe, J. Liu, *J. Biol. Chem.* **2012**, *287*, 29054–29061.

[114] X. Lu, M. N. Kamat, L. Huang, X. Huang, *J. Org. Chem.* **2009**, *74*, 7608–7617.

[115] R. A. Al-Horani, U. R. Desai, *Tetrahedron* **2010**, *66*, 2907–2918.

[116] H. S. Hahm, F. Broecker, F. Kawasaki, M. Mietzsch, R. Heilbronn, M. Fukuda, P. H. Seeberger, *Chem* **2017**, *2*, 114–124.

[117] M. Guberman, M. Bräutigam, P. H. Seeberger, *Chem. Sci.* **2019**, *10*, 5634–5640.

[118] P. Czechura, N. Guedes, S. Kopitzki, N. Vazquez, M. Martin-Lomas, N. C. Reichardt, *Chem. Commun.* **2011**, *47*, 2390–2392.

[119] J. Kandasamy, F. Schuhmacher, H. S. Hahm, J. C. Klein, P. H. Seeberger, *Chem. Commun.* **2014**, *50*, 1875–1877.

[120] N. J. Pawar, L. Wang, T. Higo, C. Bhattacharya, P. K. Kancharla, F. Zhang, K. Baryal, C. Huo, J. Liu, R. J. Linhardt, X. Huang, L. C. Hsieh-Wilson, *Angew. Chem. Int. Ed.* **2019**, *58*, 18577–18583.

[121] G. Fittolani, E. Shanina, M. Guberman, P. H. Seeberger, C. Rademacher, M. Delbianco, *Angew. Chem. Int. Ed.* **2021**, *60*, 13302–13309.

[122] L. R. S. Moreira, E. X. F. Filho, *Appl. Microbiol. Biotechnol.* **2008**, *79*, 165–178.

[123] T. Groth, W. Wagenknecht, *Biomaterials* **2001**, *22*, 2719–2729.

[124] N. Aggarwal, N. Altgärde, S. Svedhem, K. Zhang, S. Fischer, T. Groth, *Langmuir* **2013**, *29*, 13853–13864.

[125] D.-C. Wang, H.-Y. Yu, D. Qi, Y. Wu, L. Chen, Z. Li, *J. Am. Chem. Soc.* **2021**, *143*, 11620–11630.

[126] P. D. Leiphrakpam, P. P. Patil, N. Remmers, B. Swanson, P. M. Grandgenett, F. Qiu, F. Yu, P. Radhakrishnan, *Sci. Rep.* **2019**, *9*, 9665.

[127] C. Crawford, S. Oscarson, *Eur. J. Org. Chem.* **2020**, *2020*, 3332–3337.

[128] C.-J. Yeh, M. M. L. Zulueta, Y.-K. Li, S.-C. Hung, *Org. Biomol. Chem.* **2020**, *18*, 5370–5387.

[129] K. Bum-Erdene, H. Leffler, U. J. Nilsson, H. Blanchard, *FEBS J.* **2015**, *282*, 3348–3367.

[130] S. D. Rosen, *Annu. Rev. Immunol.* **2004**, *22*, 129–156.

[131] R. Jayakumar, N. Nwe, S. Tokura, H. Tamura, *Int. J. Biol. Macromol.* **2007**, *40*, 175–181.

[132] S. E. Tully, R. Mabon, C. I. Gama, S. M. Tsai, X. Liu, L. C. Hsieh-Wilson, *J. Am. Chem. Soc.* **2004**, *126*, 7736–7737.

[133] M. Torres-Rico, S. Maza, J. L. de Paz, P. M. Nieto, *Org. Biomol. Chem.* **2021**, *19*, 5312–5326.

[134] K. JEFFERSON, *FEMS Microbiol. Lett.* **2004**, *236*, 163–173.

[135] H.-C. Flemming, J. Wingender, *Nat. Rev. Microbiol.* **2010**, *8*, 623–633.

[136] C. Gilbert, T.-C. Tang, W. Ott, B. A. Dorr, W. M. Shaw, G. L. Sun, T. K. Lu, T. Ellis, *Nat. Mater.* **2021**, *20*, 691–700.

[137] M. F. Moradali, B. H. A. Rehm, *Nat. Rev. Microbiol.* **2020**, *18*, 195–210.

[138] P. Q. Nguyen, N.-M. D. Courchesne, A. Duraj-Thatte, P. Praveschotinunt, N. S. Joshi, *Adv. Mater.* **2018**, *30*, 1704847.

[139] E. N. Hayta, M. J. Ertelt, M. Kretschmer, O. Lieleg, *Adv. Mater. Interfaces* **2021**, *8*, 2101024.

[140] M. L. Evans, M. R. Chapman, *Biochim. Biophys. Acta* **2014**, *1843*, 1551–1558.

[141] P. C. Ke, R. Zhou, L. C. Serpell, R. Riek, T. P. J. Knowles, H. A. Lashuel, E. Gazit, I. W. Hamley, T. P. Davis, M. Fändrich, D. E. Otzen, M. R. Chapman, C. M. Dobson, D. S. Eisenberg, R. Mezzenga, *Chem. Soc. Rev.* **2020**, *49*, 5473–5509.

[142] D. O. Serra, A. M. Richter, R. Hengge, *J. Bacteriol.* **2013**, *195*, 5540–5554.

[143] D. E. Cameron, C. J. Bashor, J. J. Collins, *Nat. Rev. Microbiol.* **2014**, *12*, 381–390.

[144] A. Y. Chen, Z. Deng, A. N. Billings, U. O. S. Seker, M. Y. Lu, R. J. Citorik, B. Zakeri, T. K. Lu, *Nat. Mater.* **2014**, *13*, 515–523.

[145] P. Q. Nguyen, Z. Botyanszki, P. K. R. Tay, N. S. Joshi, *Nat. Commun.* **2014**, *5*, 4945.

[146] A. M. Duraj-Thatte, A. Manjula-Basavanna, N.-M. D. Courchesne, G. I. Cannici, A. Sánchez-Ferrer, B. P. Frank, L. van't Hag, S. K. Cotts, D. H. Fairbrother, R. Mezzenga, N. S. Joshi, *Nat. Chem. Biol.* **2021**, *17*, 732–738.

[147] M. Florea, H. Hagemann, G. Santosa, J. Abbott, C. N. Micklem, X. Spencer-Milnes, L. de Arroyo Garcia, D. Paschou, C. Lazenbatt, D. Kong, H. Chughtai, K. Jensen, P. S. Freemont, R. Kitney, B. Reeve, T. Ellis, *Proc. Natl Acad. Sci. USA* **2016**, *113*, E3431–E3440.

[148] J. Zhang, C. L. Poh, *Sci. Rep.* **2018**, *8*, 13127.

[149] E. C. Hollenbeck, A. Antonoplis, C. Chai, W. Thongsomboon, G. G. Fuller, L. Cegelski, *Proc. Natl Acad. Sci. USA* **2018**, *115*, 10106–10111.

[150] A. Singh, K. T. Walker, R. Ledesma-Amaro, T. Ellis, *Int. J. Mol. Sci.* **2020**, *21*, 9185.

[151] W. Yi, X. Liu, Y. Li, J. Li, C. Xia, G. Zhou, W. Zhang, W. Zhao, X. Chen, P. G. Wang, *Proc. Natl Acad. Sci. USA* **2009**, *106*, 4207–4212.

[152] S. Goon, B. Schilling, M. V. Tullius, B. W. Gibson, C. R. Bertozzi, *Proc. Natl Acad. Sci. USA* **2003**, *100*, 3089–3094.

[153] B. H. A. Rehm, *Nat. Rev. Microbiol.* **2010**, *8*, 578–592.

[154] Y. Yu, T. Tyrikos-Ergas, Y. Zhu, G. Fittolani, V. Bordoni, A. Singhal, R. J. Fair, A. Grafmüller, P. H. Seeberger, M. Delbianco, *Angew. Chem. Int. Ed.* **2019**, *58*, 13127–13132.

[155] J. M. Nguyen, R. E. Moore, S. K. Spicer, J. A. Gaddy, S. D. Townsend, *ChemBioChem* **2021**, *22*, 2540–2545.

[156] X. Wang, D. R. Smith, J. W. Jones, M. R. Chapman, *J. Biol. Chem.* **2007**, *282*, 3713–3719.

[157] Y.-H. Tsai, S. Götze, I. Vilotijevic, M. Grube, D. V. Silva, P. H. Seeberger, *Chem. Sci.* **2013**, *4*, 468–481.

[158] Y. Yu, S. Gim, D. Kim, Z. A. Arnon, E. Gazit, P. H. Seeberger, M. Delbianco, *J. Am. Chem. Soc.* **2019**, *141*, 4833–4838.

[159] I. Cherny, L. Rockah, O. Levy-Nissenbaum, U. Gophna, E. Z. Ron, E. Gazit, *J. Mol. Biol.* **2005**, *352*, 245–252.

[160] P. Lembre, C. Vendrely, P. Martino, *Protein Pept. Lett.* **2013**, *20*, 942–946.

[161] S. Tomaselli, V. Esposito, P. Vangone, N. A. J. van Nuland, A. M. J. J. Bonvin, R. Guerrini, T. Tancredi, P. A. Temussi, D. Picone, *ChemBioChem* **2006**, *7*, 257–267.

[162] N. Hirota, Y. Goto, K. Mizuno, *Protein Sci.* **1997**, *6*, 416–421.

[163] M. Sleutel, I. Van den Broeck, N. Van Gerven, C. Feuillie, W. Jonckheere, C. Valotteau, Y. F. Dufrêne, H. Remaut, *Nat. Chem. Biol.* **2017**, *13*, 902–908.

[164] P. Wang, J. Zhao, S. Hossaini Nasr, S. A. Otieno, F. Zhang, W. Qiang, R. J. Linhardt, X. Huang, *ACS Chem. Biol.* **2021**, *16*, 1894–1899.

[165] G. Wei, Z. Su, N. P. Reynolds, P. Arosio, I. W. Hamley, E. Gazit, R. Mezzenga, *Chem. Soc. Rev.* **2017**, *46*, 4661–4708.

[166] I. M. Ilie, A. Caflisch, *Chem. Rev.* **2019**, *119*, 6956–6993.

[167] X. Dai, W. Hou, Y. Sun, Z. Gao, S. Zhu, Z. Jiang, *Int. J. Mol. Sci.* **2015**, *16*, 10526–10536.

[168] C.-C. Liu, N. Zhao, Y. Yamaguchi, J. R. Cirrito, T. Kanekiyo, D. M. Holtzman, G. Bu, *Sci. Transl. Med.* **2016**, *8*, 332ra44-332ra44.

[169] T. Tyrikos-Ergas, V. Bordoni, G. Fittolani, M. A. Chaube, A. Grafmüller, P. H. Seeberger, M. Delbianco, *Chem. Eur. J.* **2021**, *27*, 2321–2325.

[170] S. A. Sievers, J. Karanicolas, H. W. Chang, A. Zhao, L. Jiang, O. Zirafi, J. T. Stevens, J. Münch, D. Baker, D. Eisenberg, *Nature* **2011**, *475*, 96–100.

[171] E. Chuang, A. M. Hori, C. D. Hesketh, J. Shorter, *J. Cell Sci.* **2018**, *131*, 1–18.

[172] S. T. Laughlin, C. R. Bertozzi, *Nat. Protoc.* **2007**, *2*, 2930–2944.

[173] P. A. Romero, F. H. Arnold, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866–876.

[174] A. Kan, N. S. Joshi, *MRS Commun.* **2019**, *9*, 441–455.

[175] M. Vendruscolo, C. M. Dobson, *Curr. Biol.* **2011**, *21*, R68–R70.

[176] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, A. Kolinski, *Chem. Rev.* **2016**, *116*, 7898–7936.

[177] T.-Q. Yu, J. Lu, C. F. Abrams, E. Vanden-Eijnden, *Proc. Natl Acad. Sci. USA* **2016**, *113*, 11744–11749.

[178] M. W. Mahoney, W. L. Jorgensen, *J. Chem. Phys.* **2001**, *114*, 363–366.

[179] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701–1718.

[180] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463–1472.

[181] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952–962.

[182] K. Daragics, P. Fügedi, *Org. Lett.* **2010**, *12*, 2076–2079.

[183] K. Daragics, P. Fügedi, *Tetrahedron* **2010**, *66*, 8036–8046.

[184] P. Chassagne, L. Raibaut, C. Guerreiro, L. A. Mulard, *Tetrahedron* **2013**, *69*, 10337–10350.

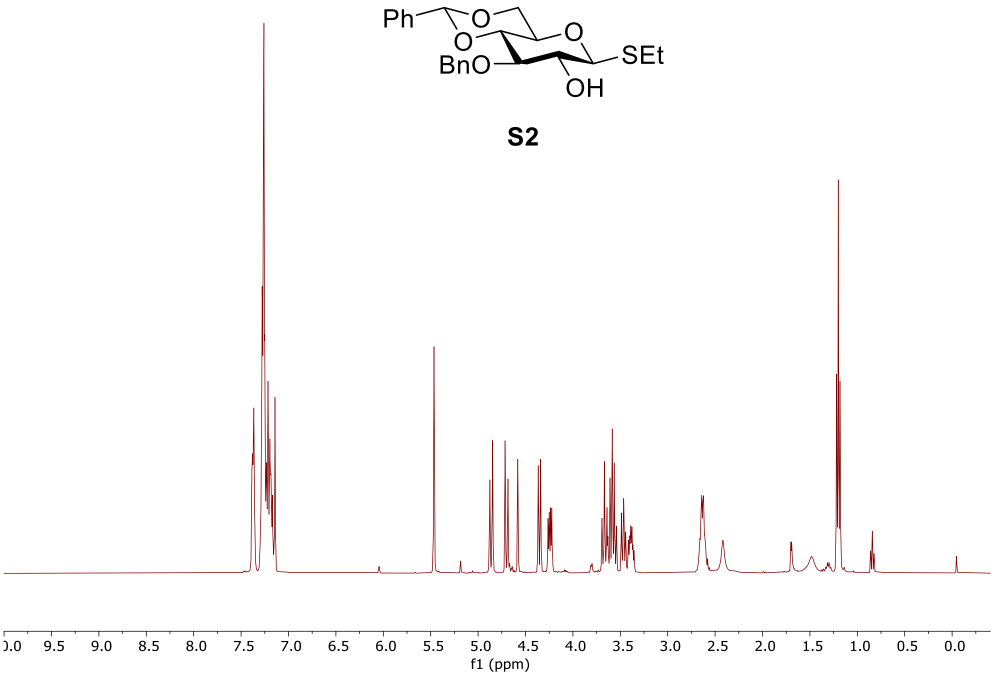
[185] L. Kröck, D. Esposito, B. Castagner, C.-C. Wang, P. Bindschädler, P. H. Seeberger, *Chem. Sci.* **2012**, *3*, 1617.

[186] C. Murakata, T. Ogawa, *Tetrahedron Lett.* **1991**, *32*, 671–674.

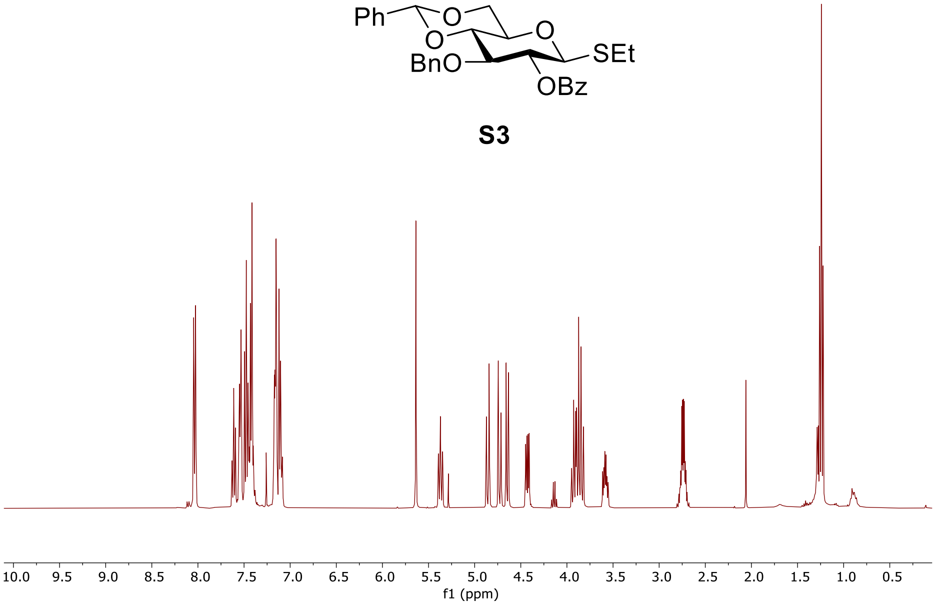
# 9. Appendix

## 9.1. NMR spectra

**$^{1}$H NMR of S2 (400 MHz, $CDCl_3$)**

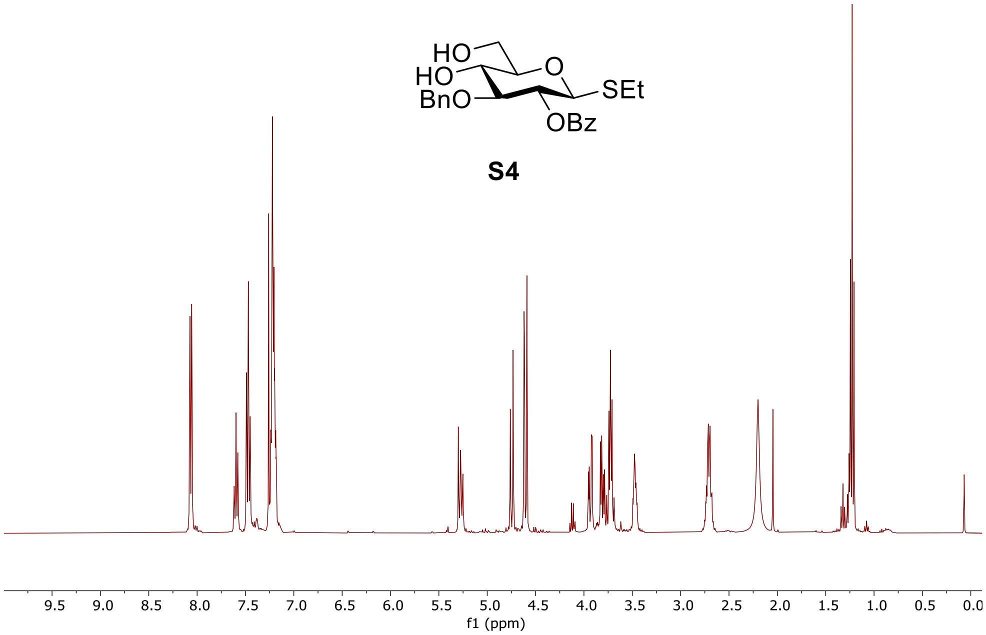


## $^1$H NMR of S3 (400 MHz, CDCl$_3$)



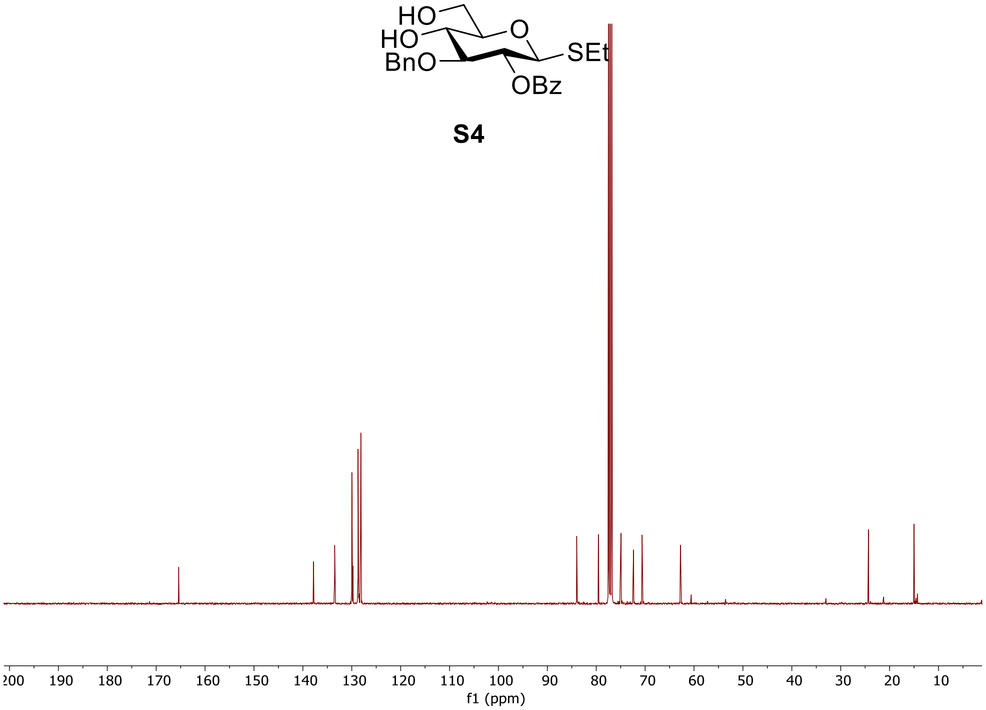
**S3**

## $^1$H NMR of S4 (400 MHz, CDCl$_3$)



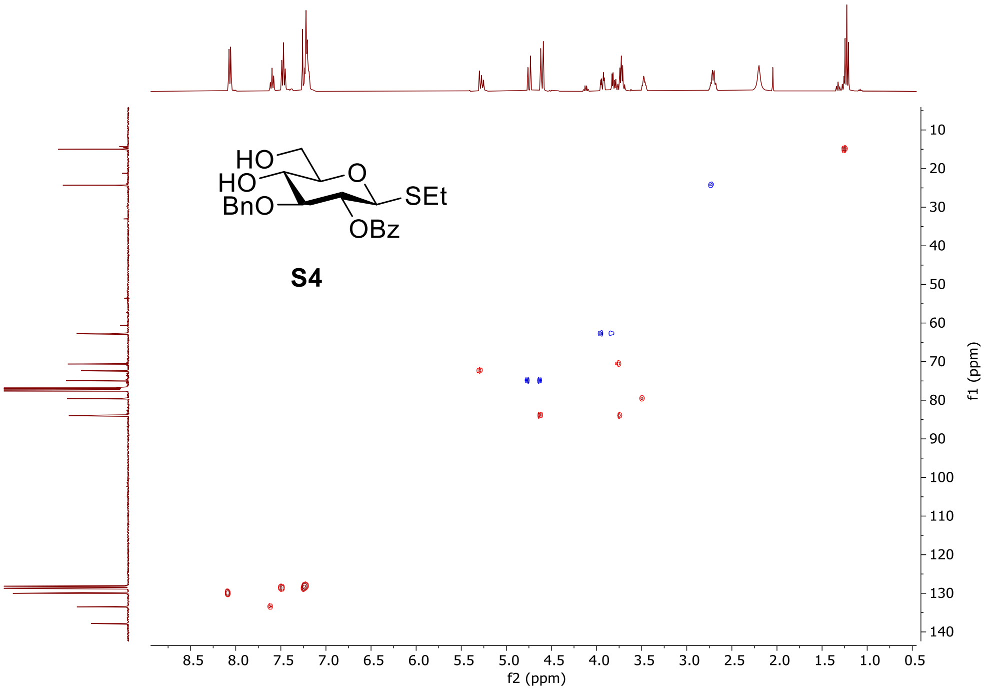
**S4**

## $^{13}$C NMR of S4 (101 MHz, CDCl$_3$)

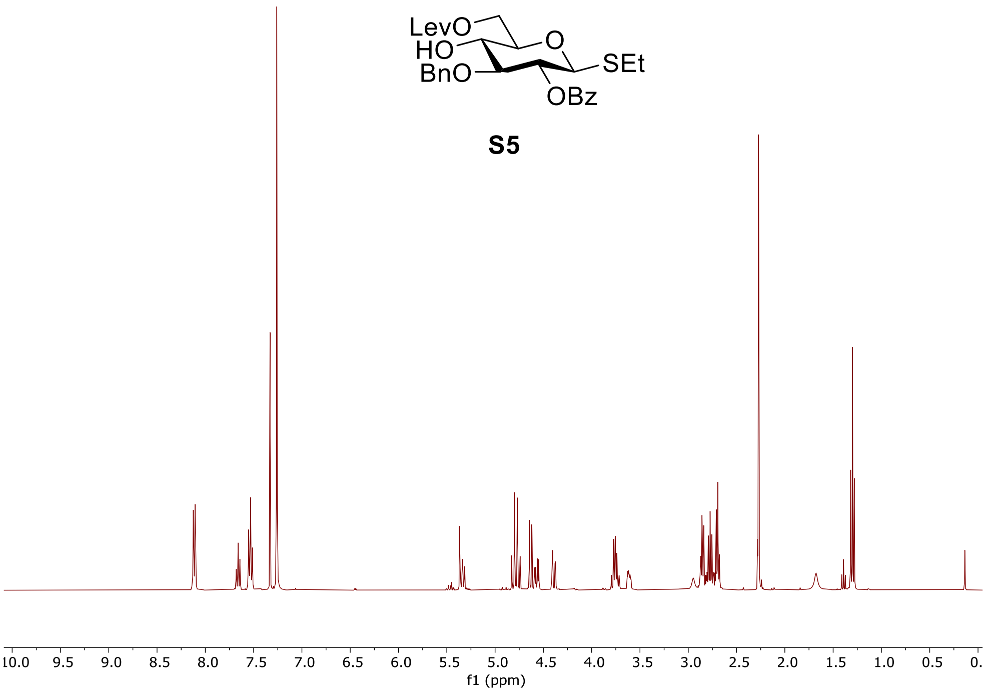


**S4**

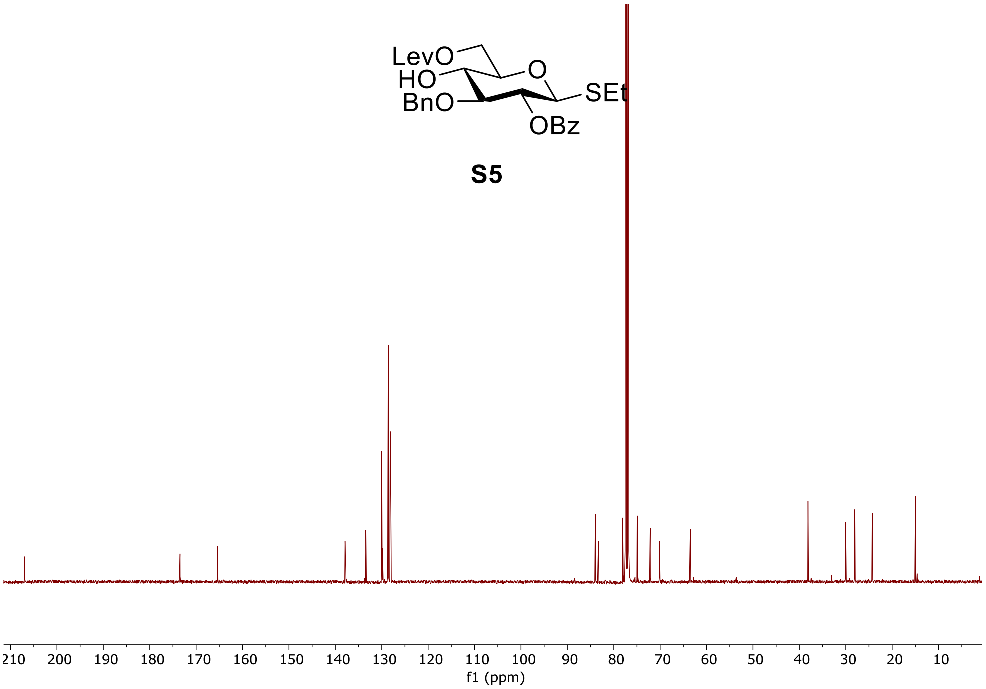
## HSQC NMR of S4 ($CDCl_3$)



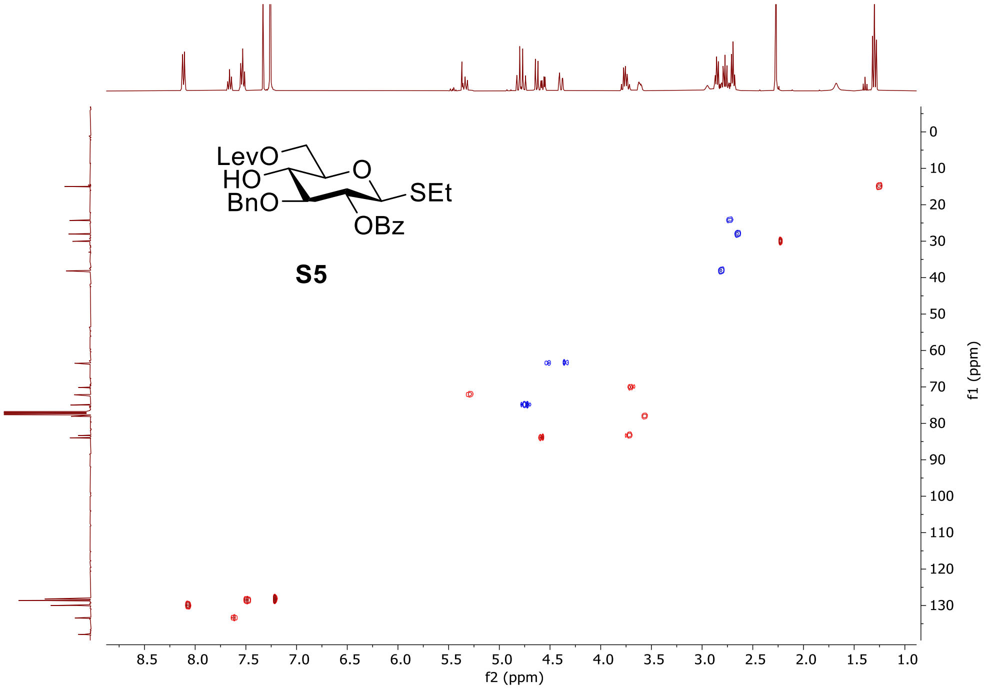
## $^1$H NMR of S5 (400 MHz, $CDCl_3$)



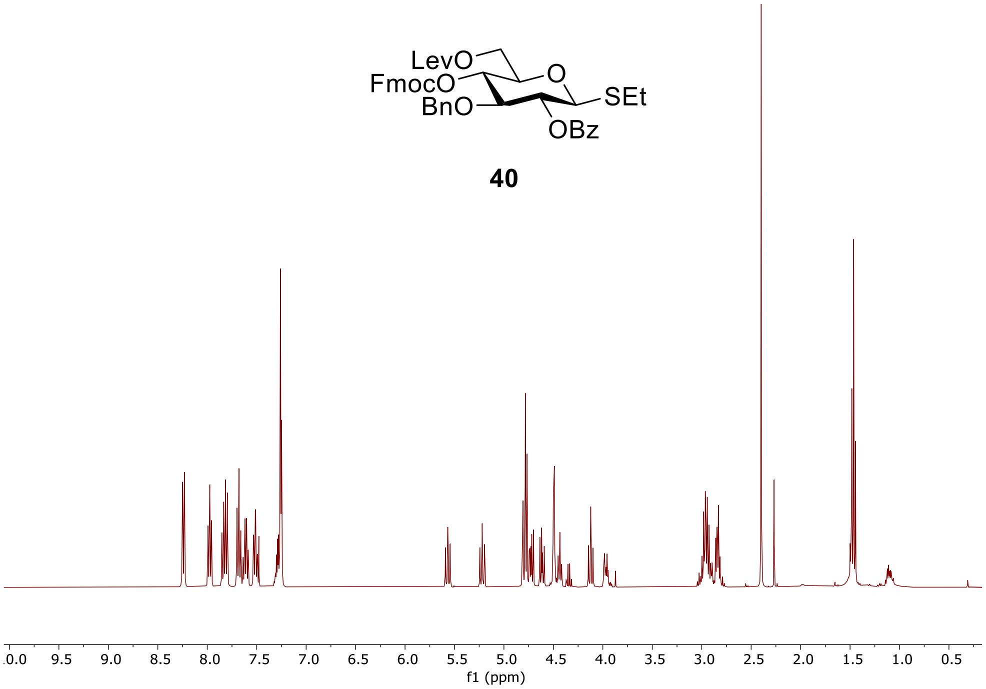
## $^{13}$C NMR of S5 (101 MHz, CDCl$_3$)



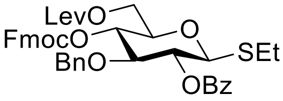
## HSQC NMR of S5 (CDCl$_3$)



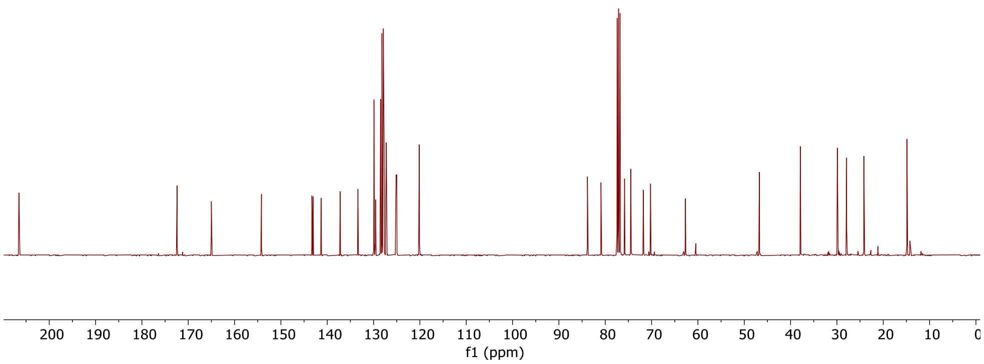
## $^1$H NMR of 49 (400 MHz, CDCl$_3$)



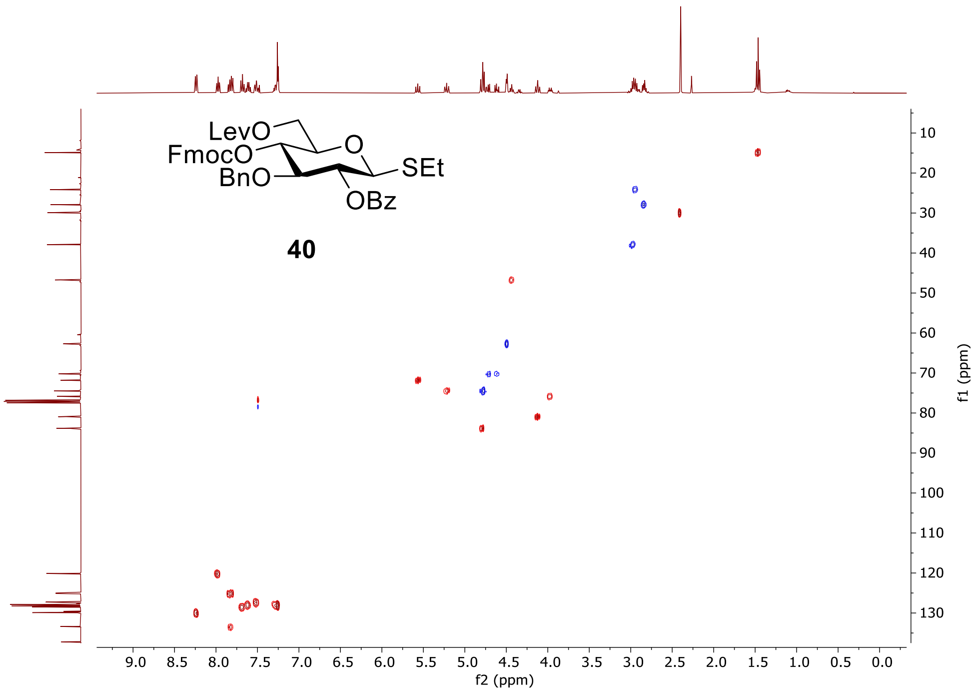
## $^{13}$C NMR of 40 (101 MHz, CDCl$_3$)



**40**

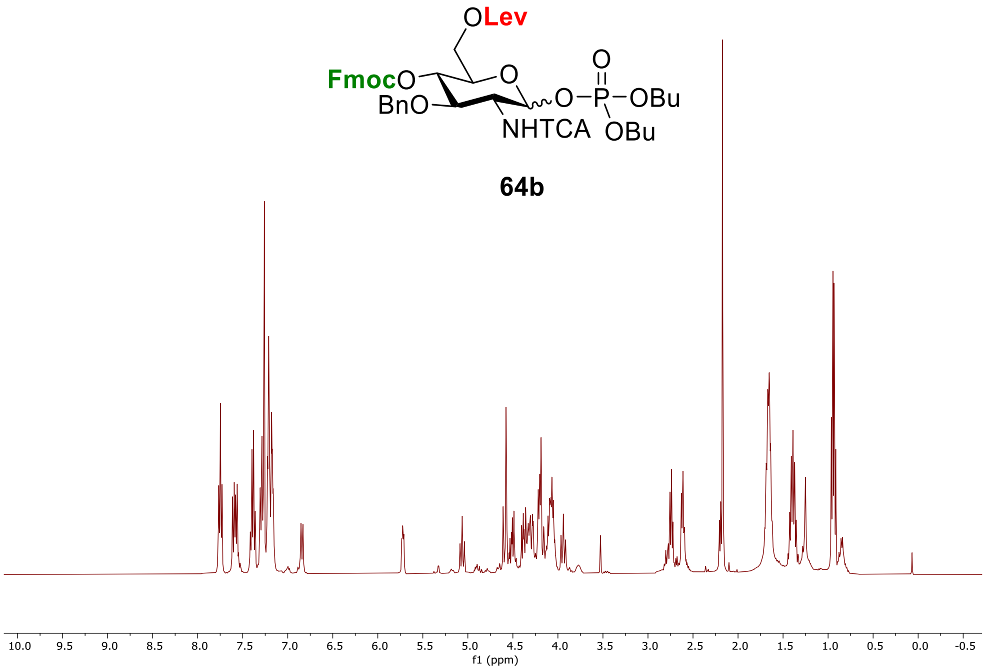


## HSQC NMR of 40 ($CDCl_3$)



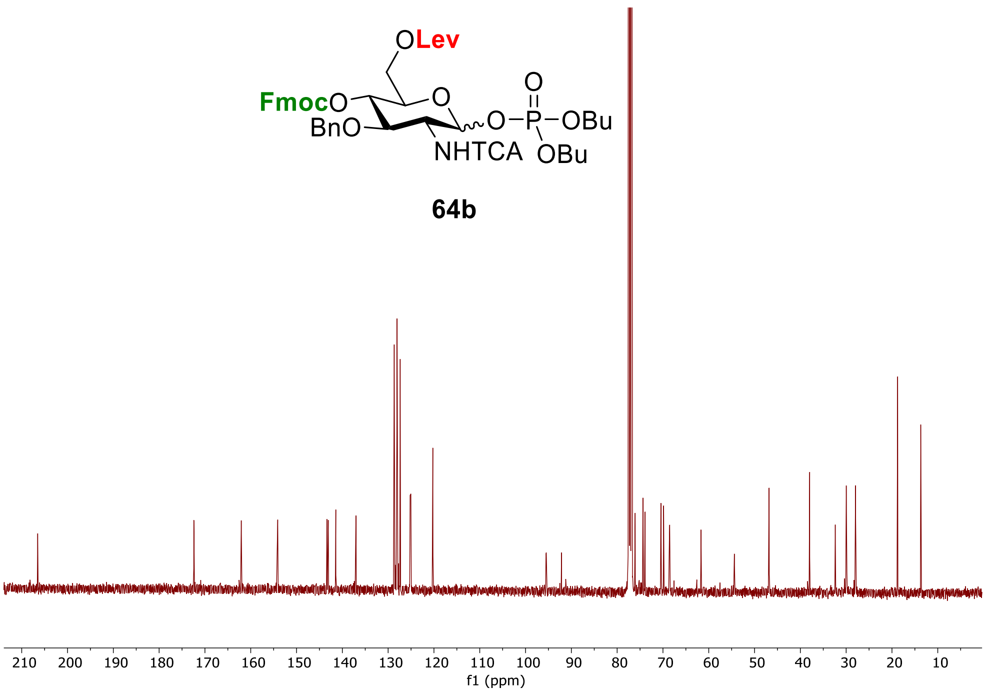
**40**

## $^1$H NMR of 64b (400 MHz, $CDCl_3$)

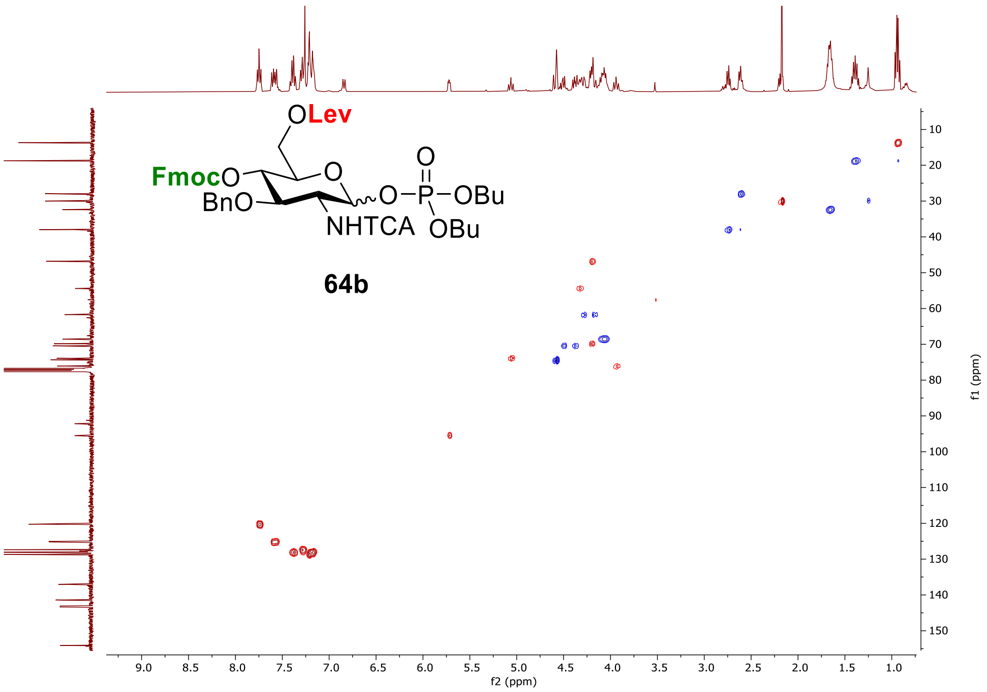


**64b**

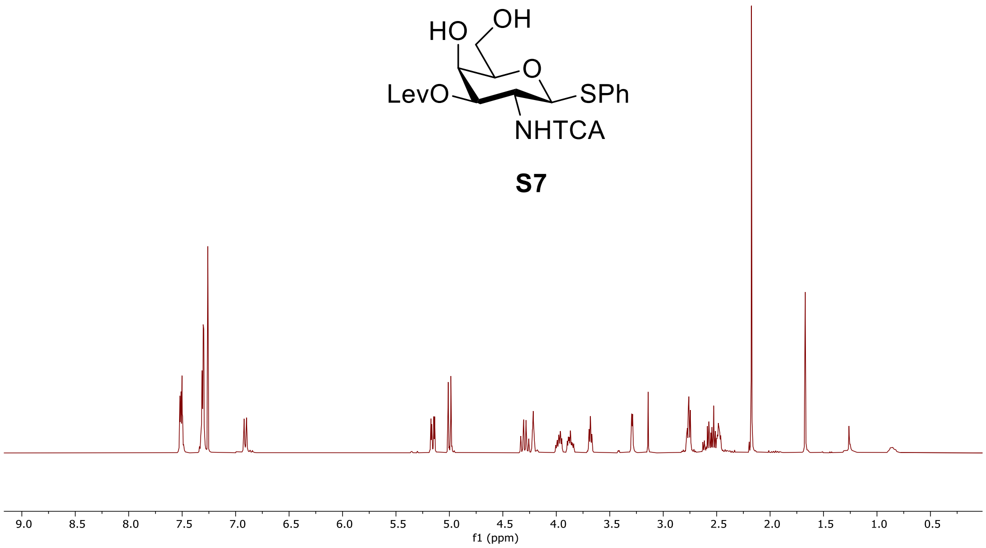
## $^{13}$C NMR of 64b (101 MHz, CDCl$_3$)



## HSQC NMR of 64b (CDCl$_3$)

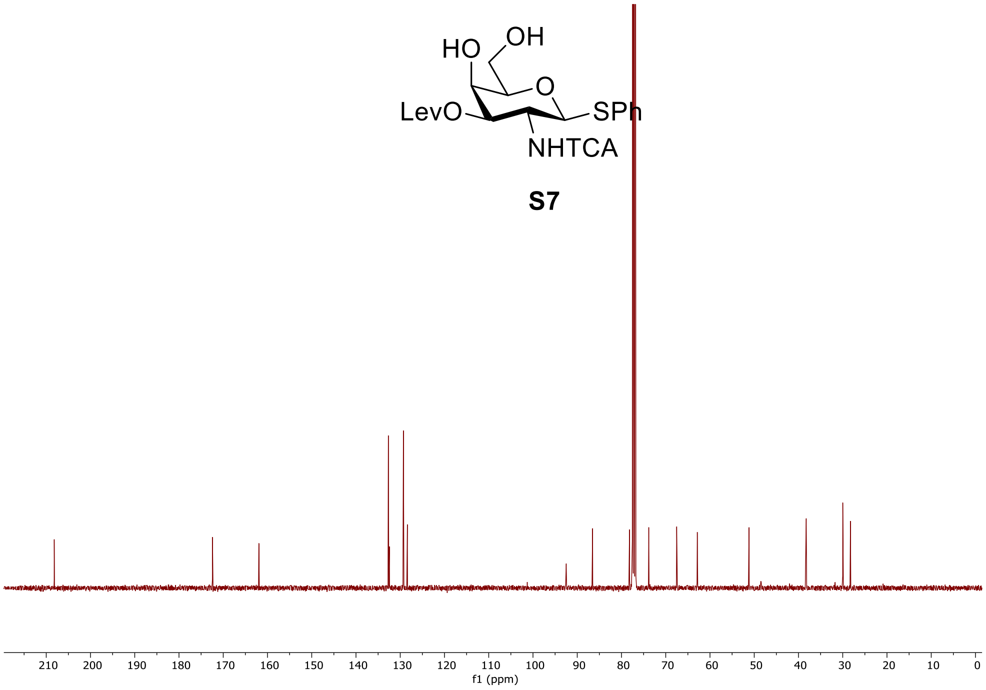


## $^{1}$H NMR of S7 (400 MHz, $CDCl_3$)



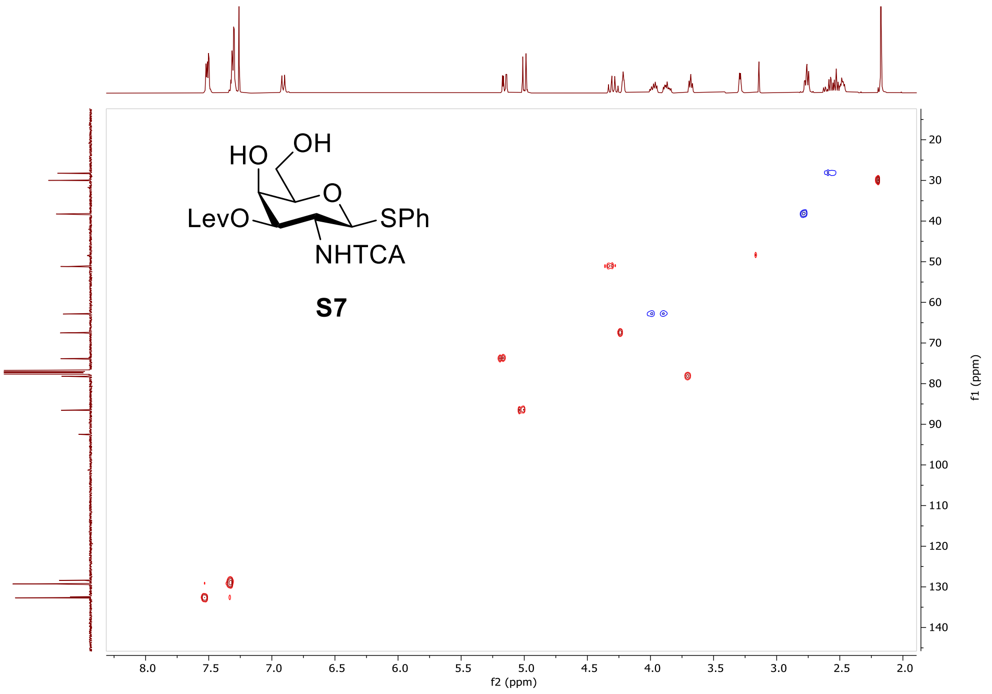
**S7**

## $^{13}$C NMR of S7 (101 MHz, $CDCl_3$)



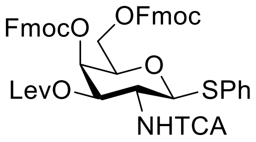
**S7**

## HSQC NMR of S7 ($CDCl_3$)

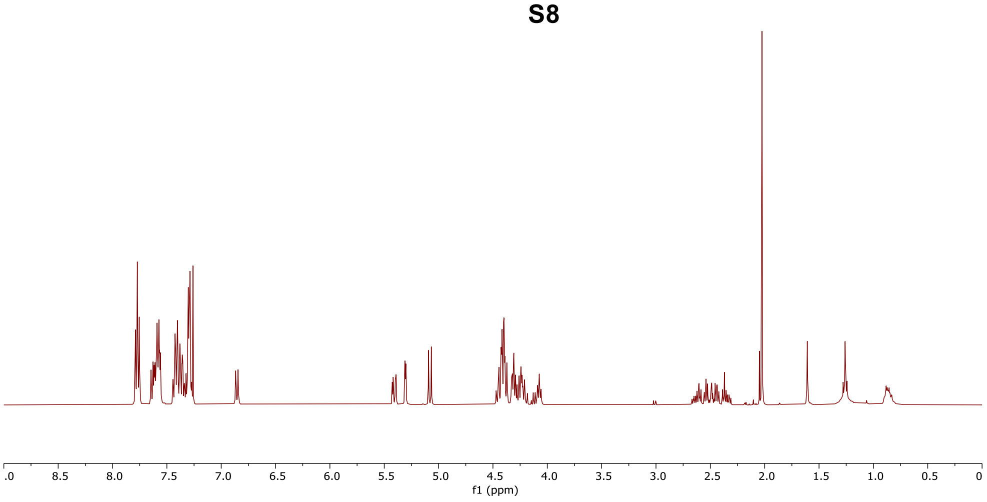


**S7**

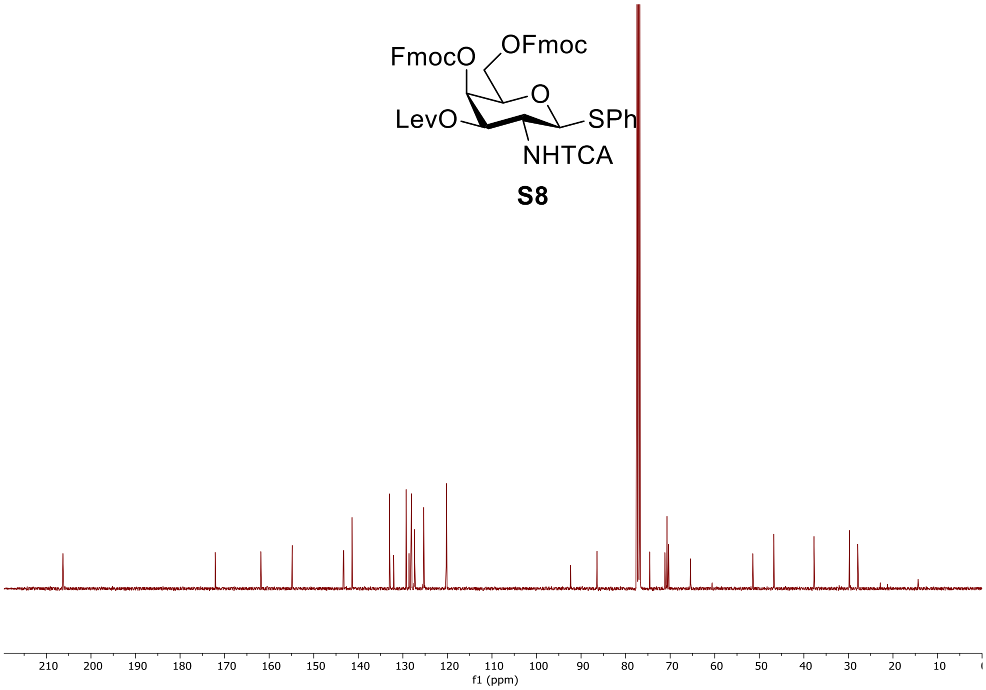
## $^1$H NMR of S8 (400 MHz, $CDCl_3$)



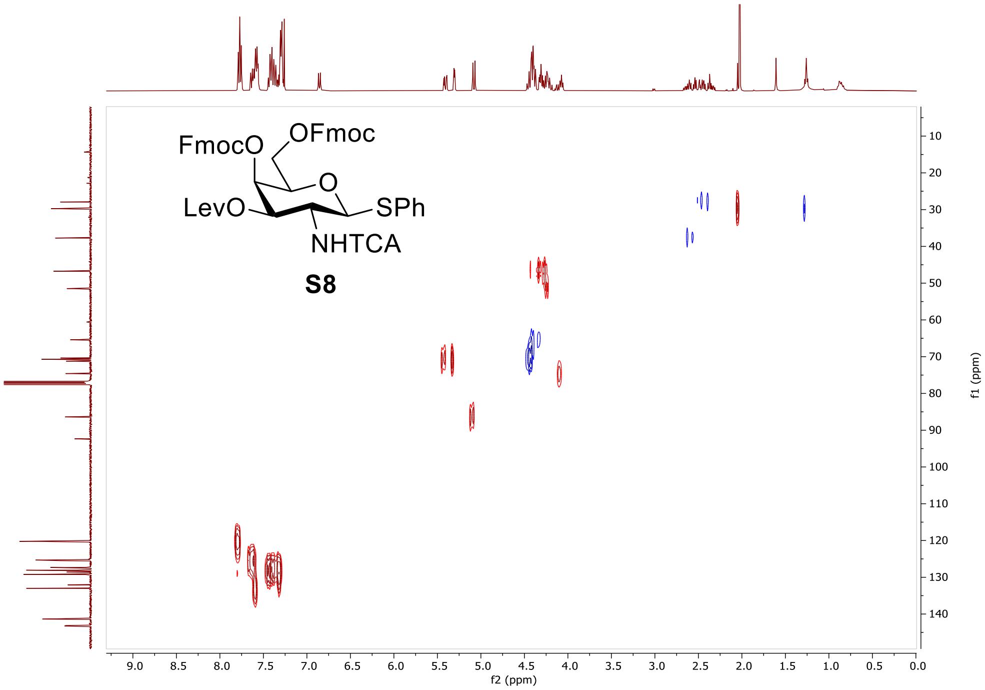
**S8**



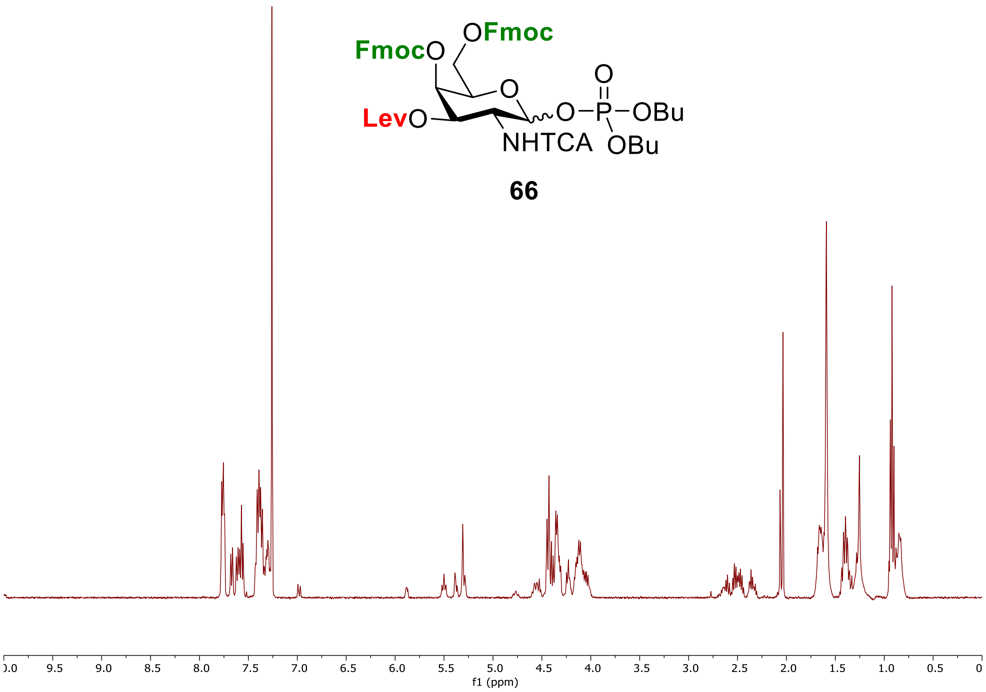
## $^{13}$C NMR of S8 (101 MHz, CDCl$_3$)



## HSQC NMR of S8 (CDCl$_3$)

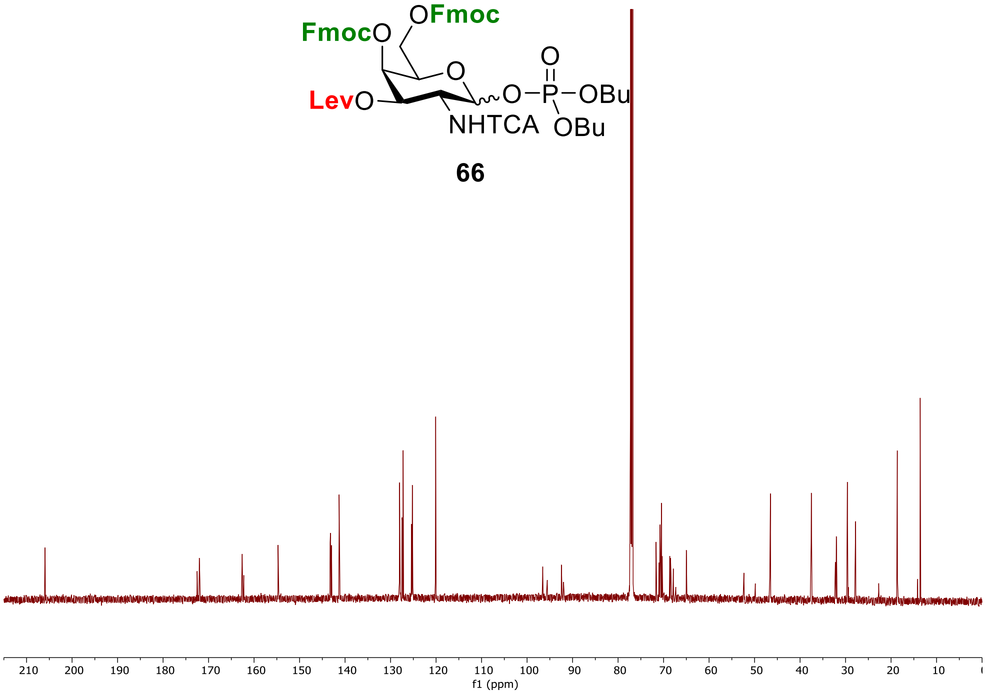


## $^1$H NMR of 66 (400 MHz, CDCl$_3$)



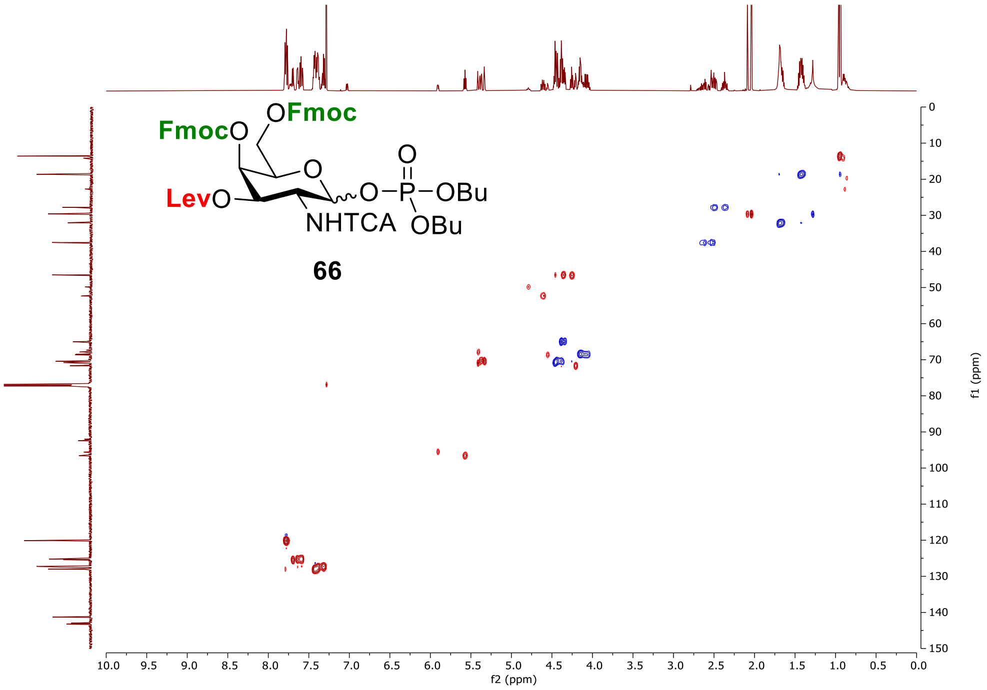
**66**

## $^{13}$C NMR of 66 (151 MHz, CDCl$_3$)



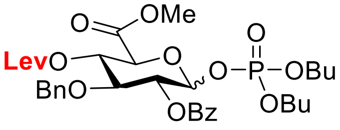
**66**

## HSQC NMR of 66 ($CDCl_3$)

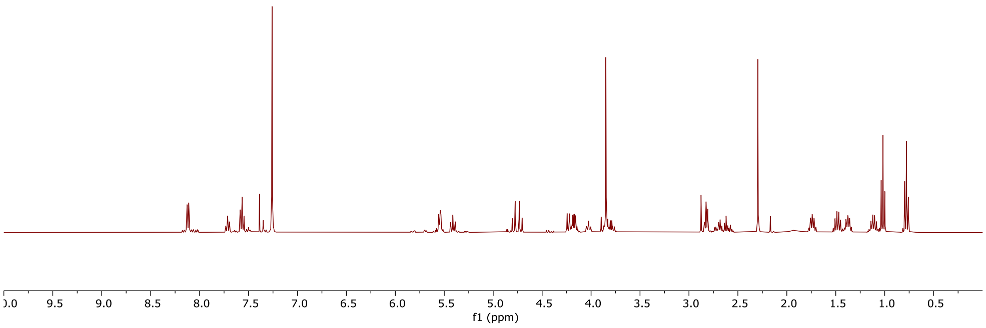


**66**

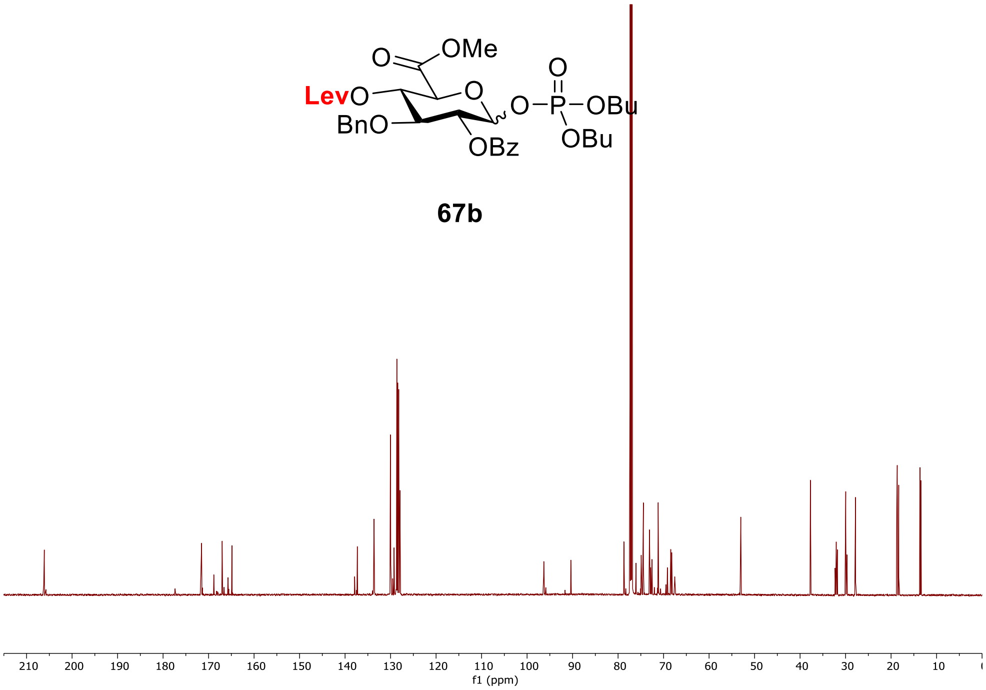
## $^1H$ NMR of 67b (400 MHz, $CDCl_3$)



**67b**

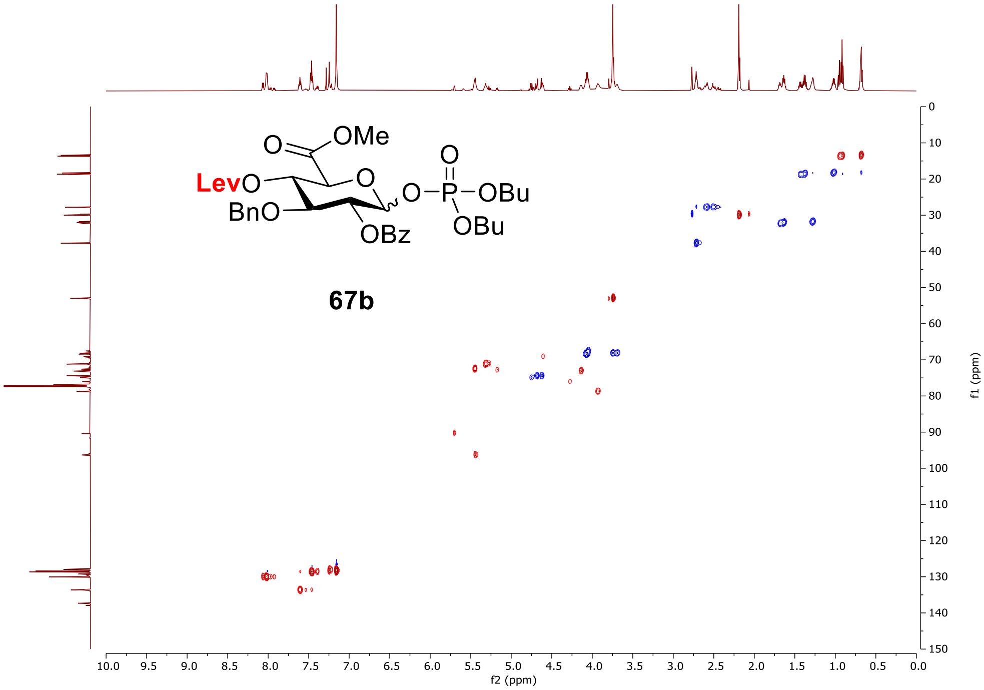


## $^{13}$C NMR of 67b (151 MHz, CDCl$_3$)



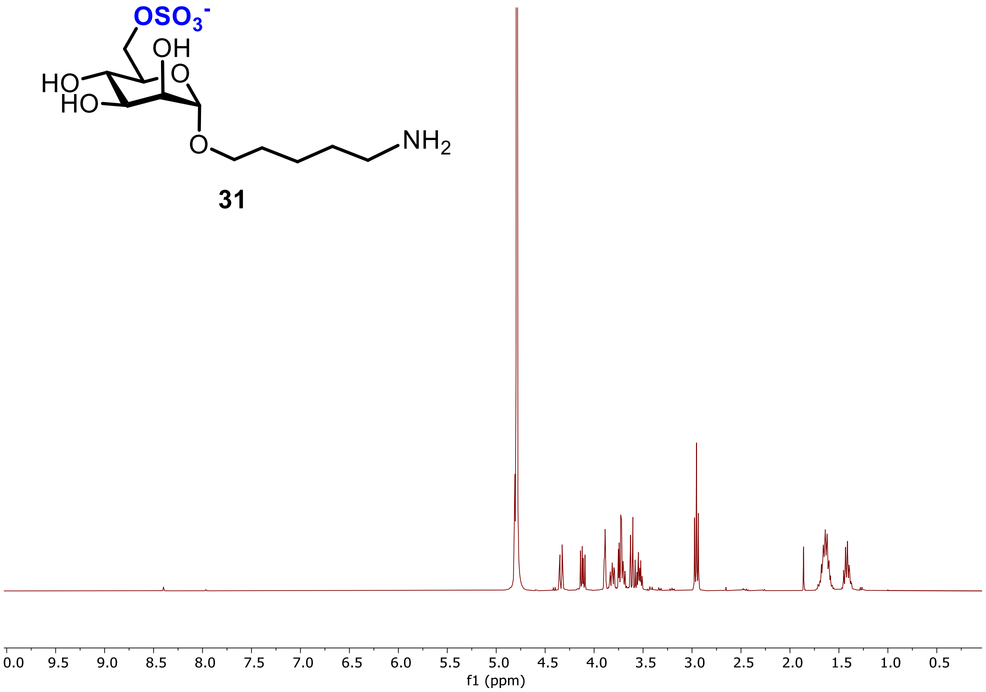
**67b**

## HSQC NMR of 67b (CDCl$_3$)

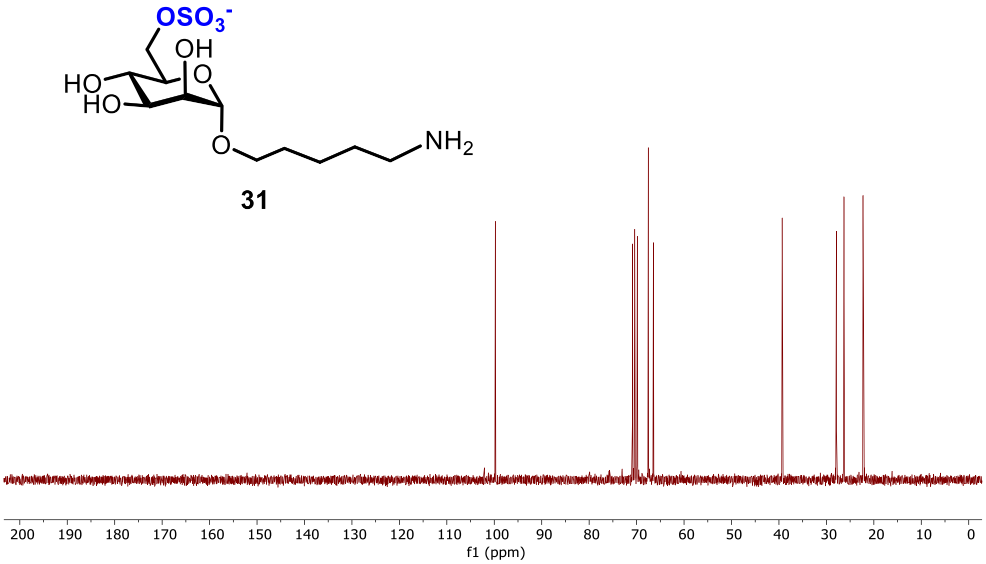


**67b**

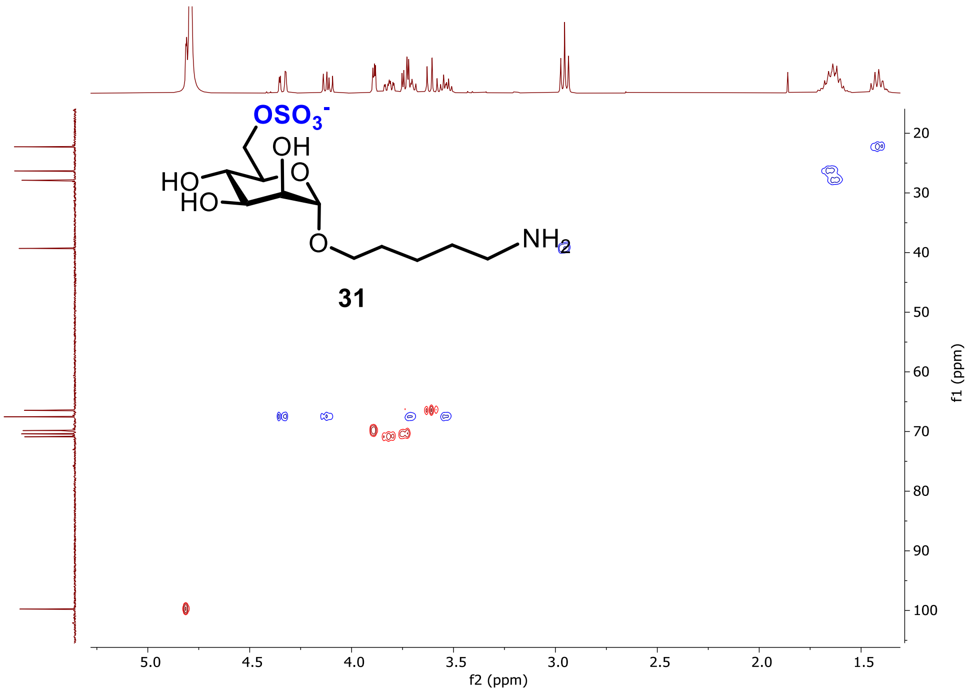
## $^{1}$H NMR of 31 (400 MHz, $D_2O$)



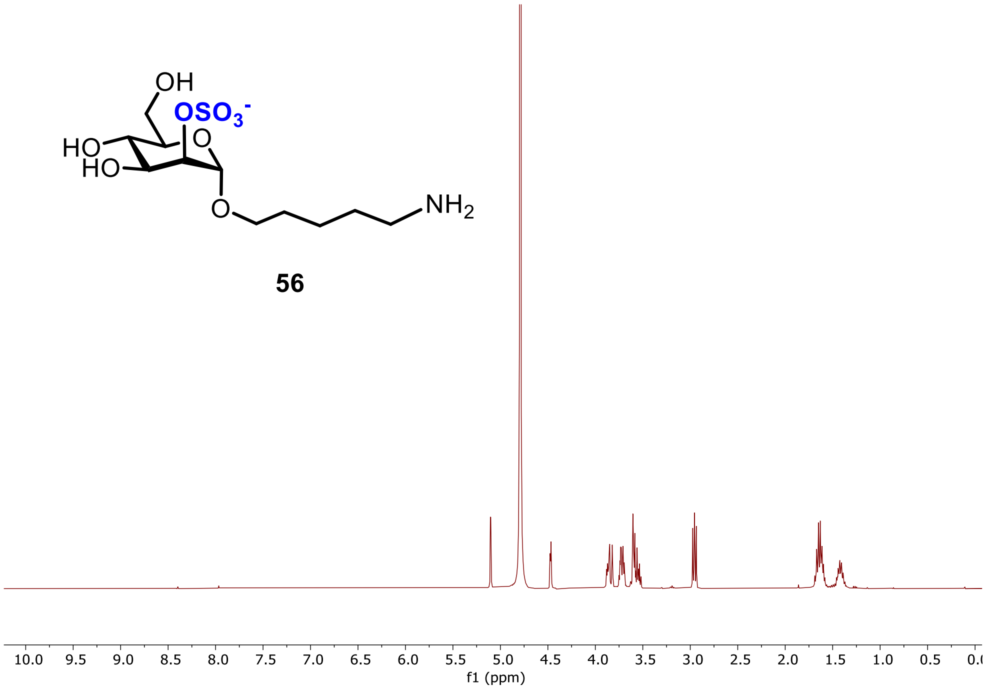
## $^{13}$C NMR of 31 (101 MHz, $D_2O$)



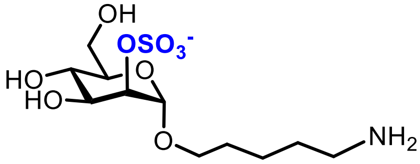
## HSQC NMR of 31 ($D_2O$)



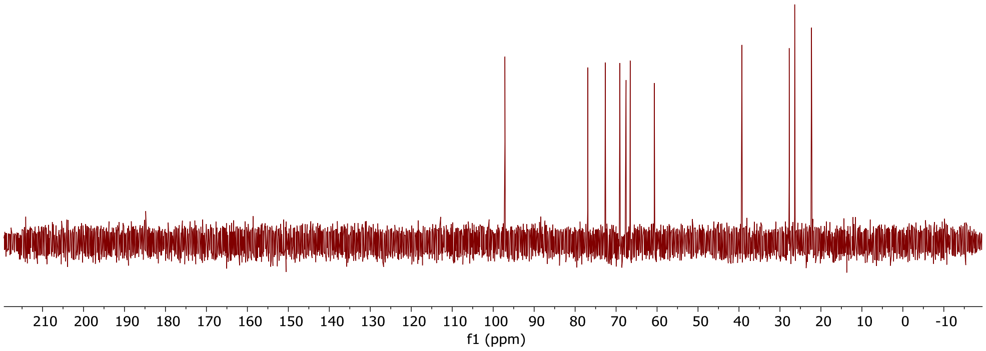
## $^1$H NMR of 56 (400 MHz, $D_2O$)



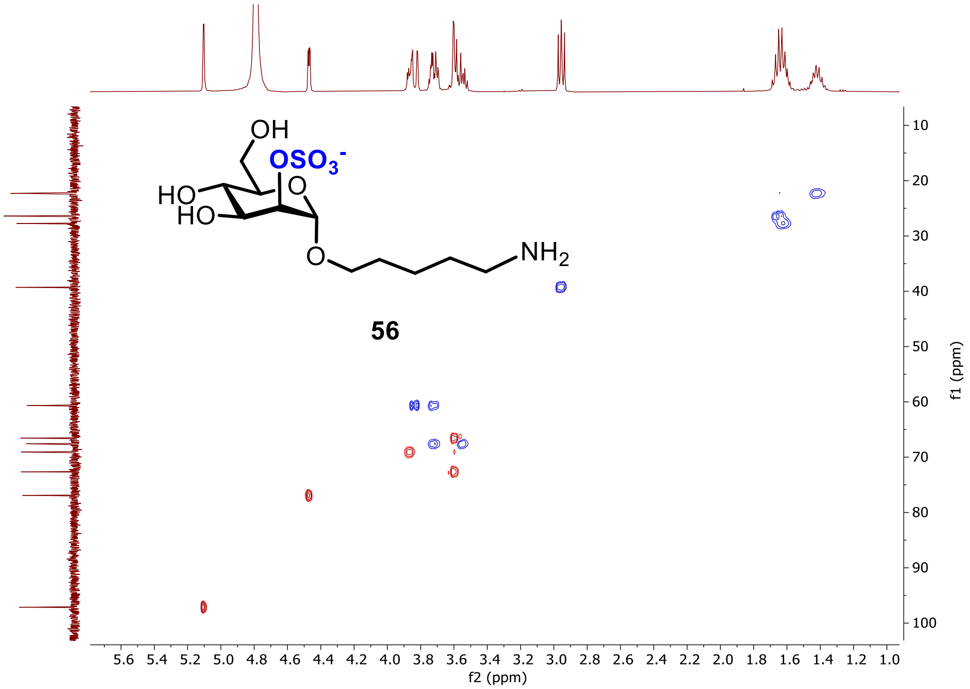
## $^{13}$C NMR of 56 (101 MHz, $D_2O$)



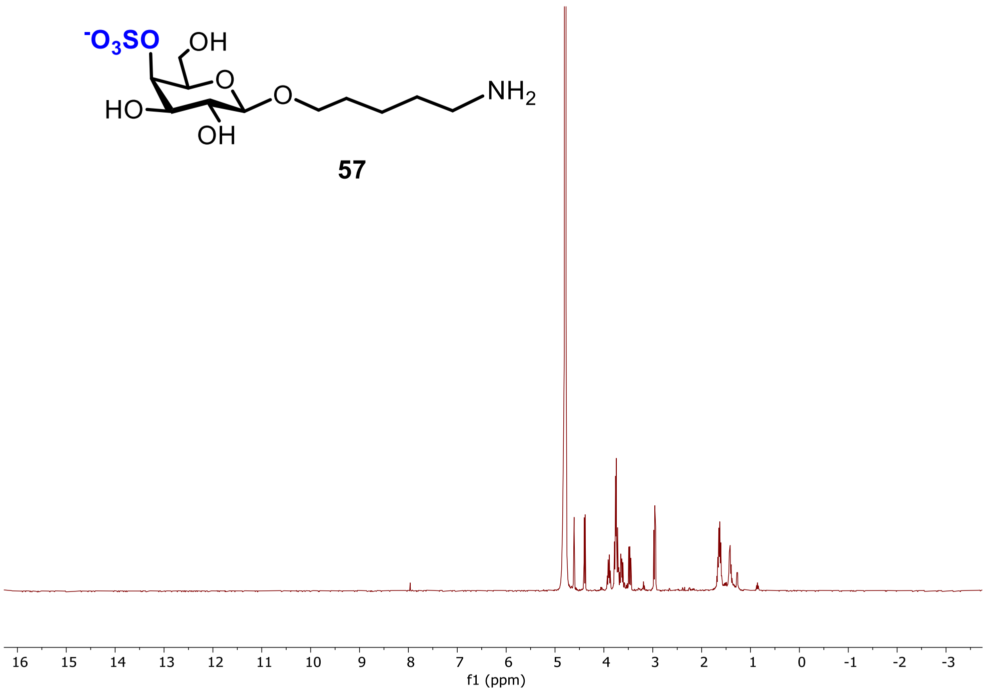
**56**



## HSQC NMR of 56 ($D_2O$)

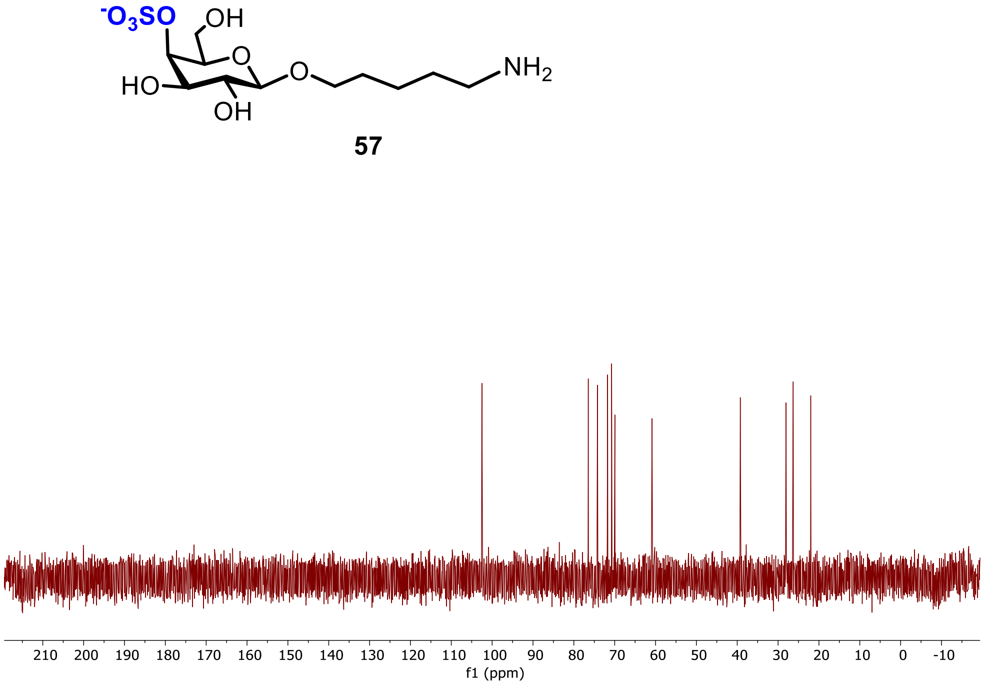


## $^1$H NMR of 7 (400 MHz, D$_2$O)



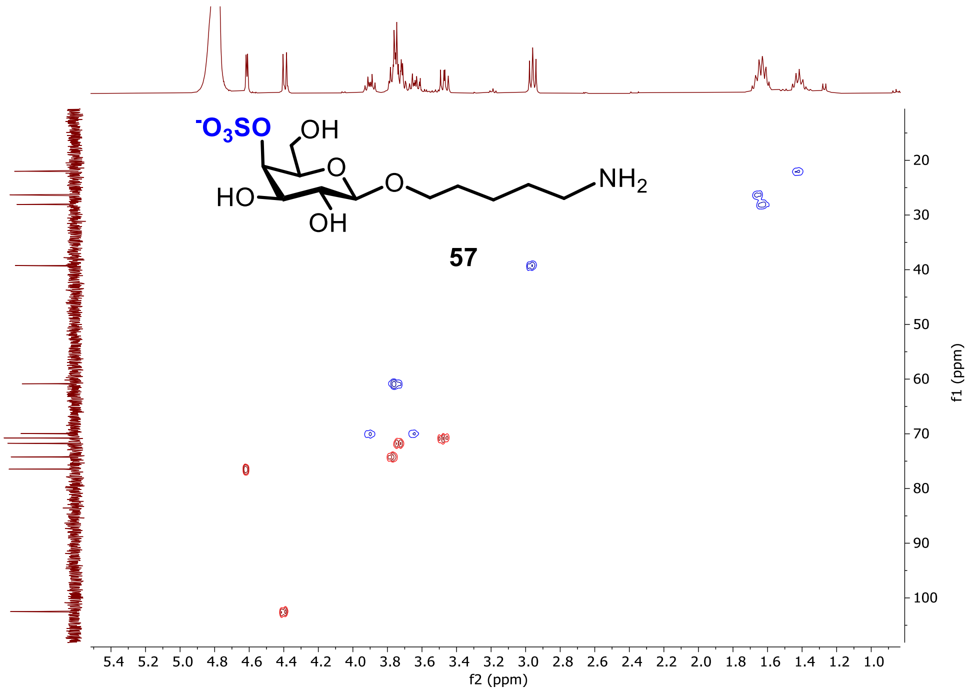
**57**

## $^{13}$C NMR of 57 (101 MHz, D$_2$O)



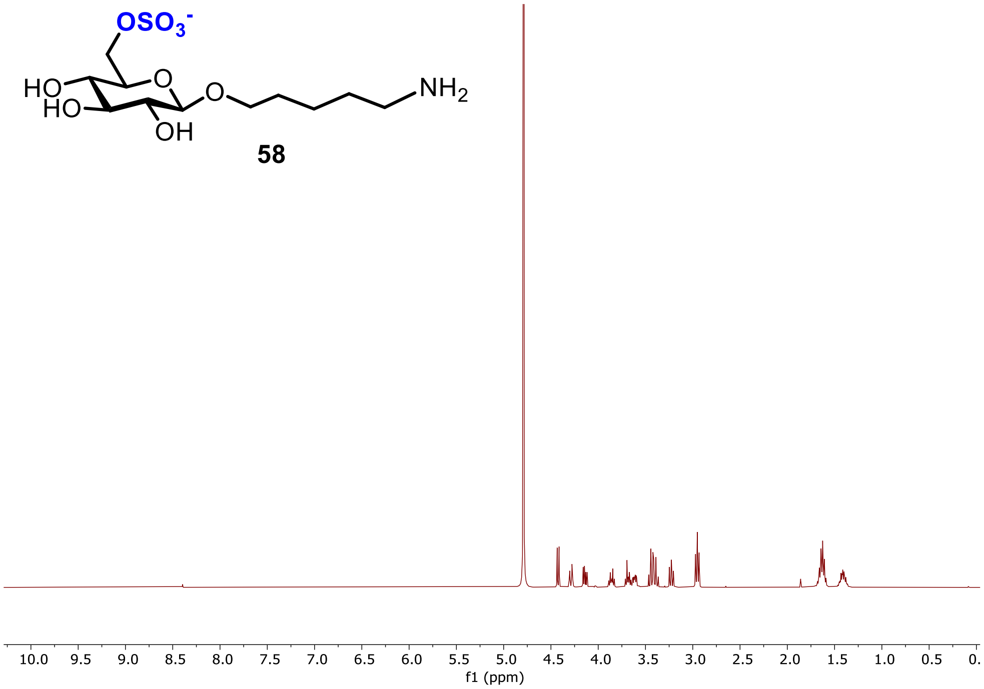
**57**

## HSQC NMR of 57 ($D_2O$)



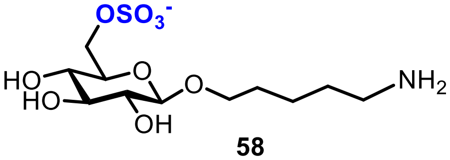
**57**

## $^1$H NMR of 58 (400 MHz, $D_2O$)

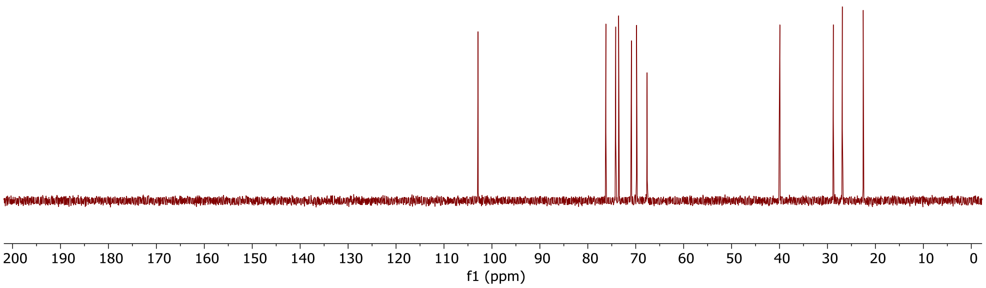


**58**

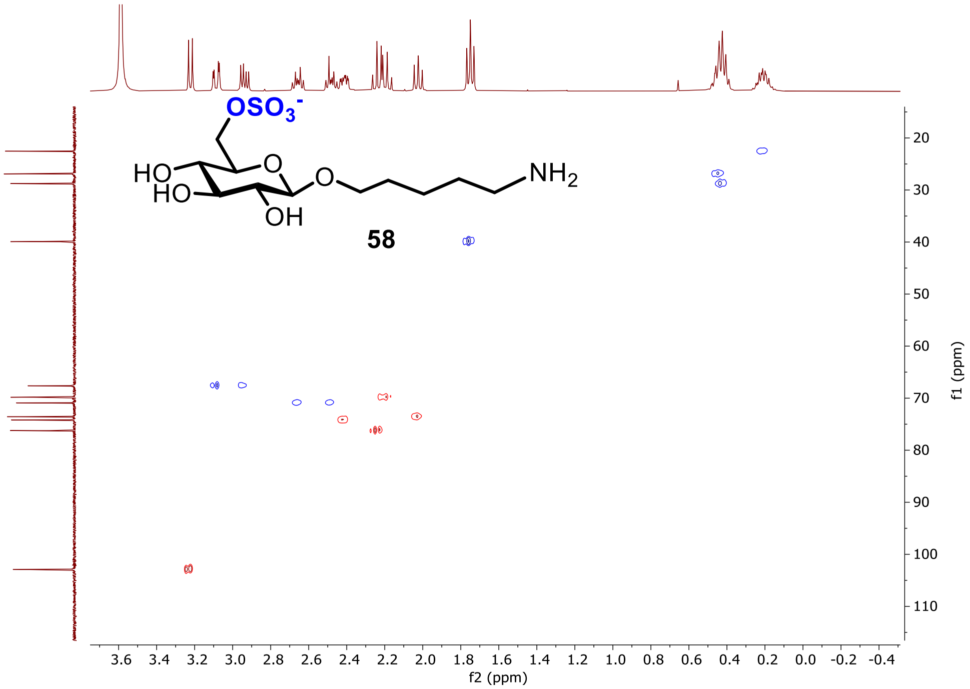
## $^{13}$C NMR of 58 (101 MHz, D$_2$O)



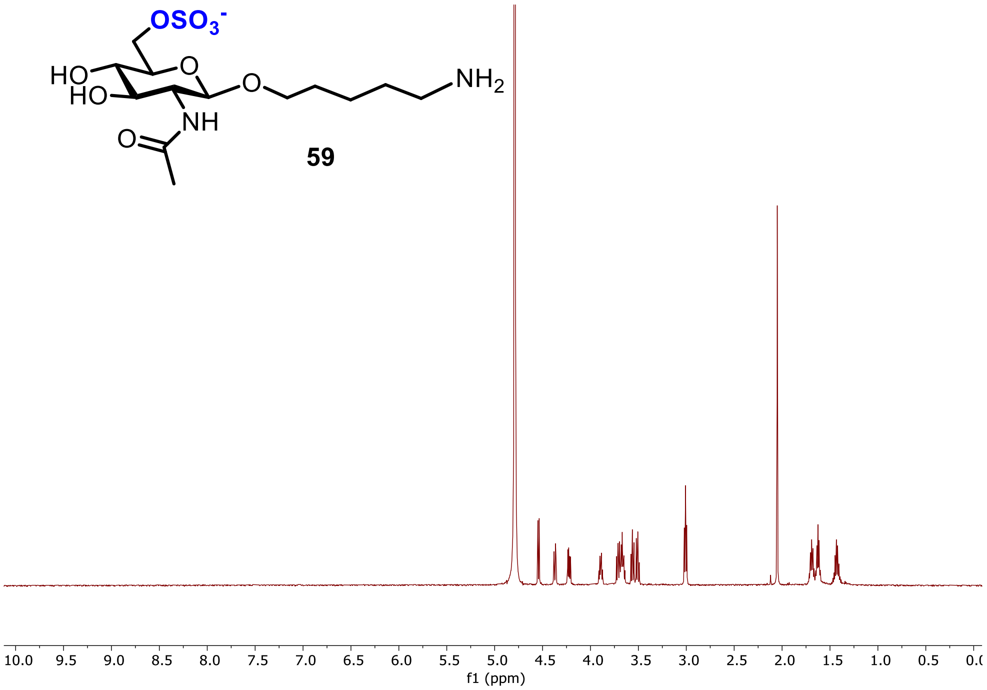
**58**



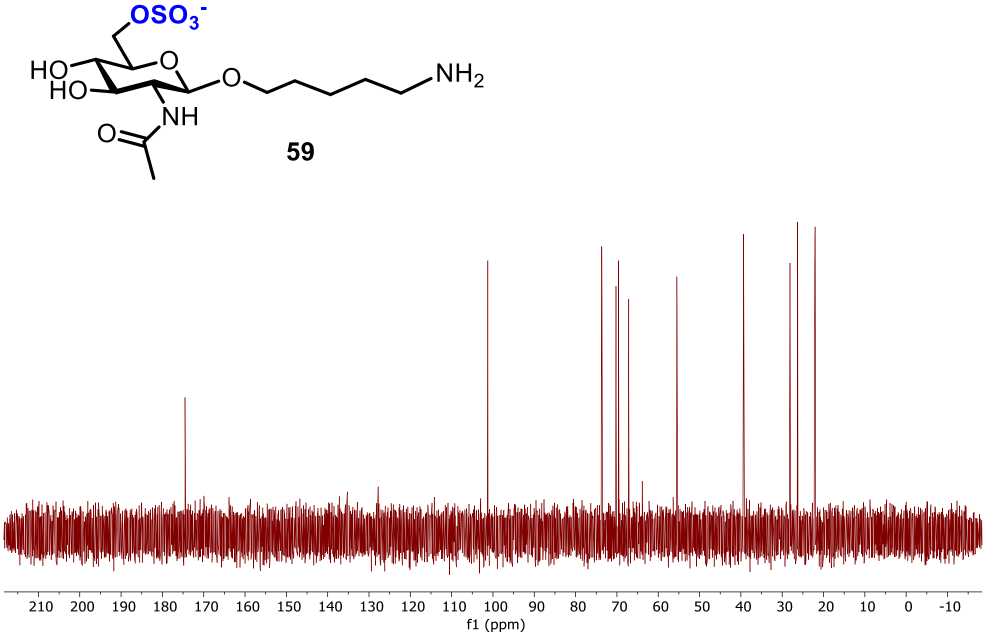
## HSQC NMR of 58 (D$_2$O)



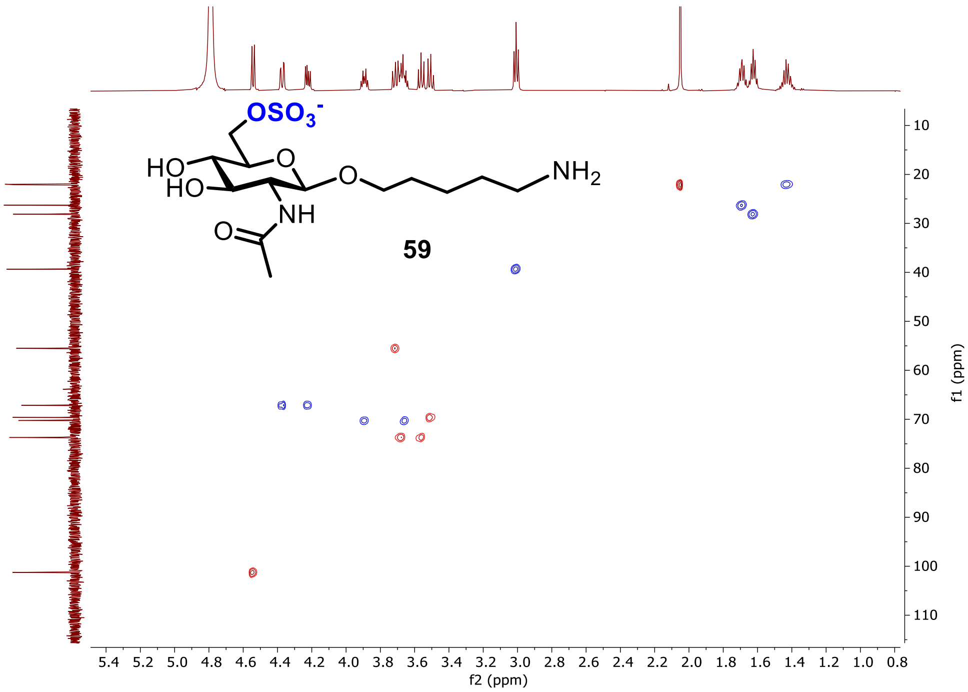
## $^{1}$H NMR of 59 (600 MHz, $D_2O$)



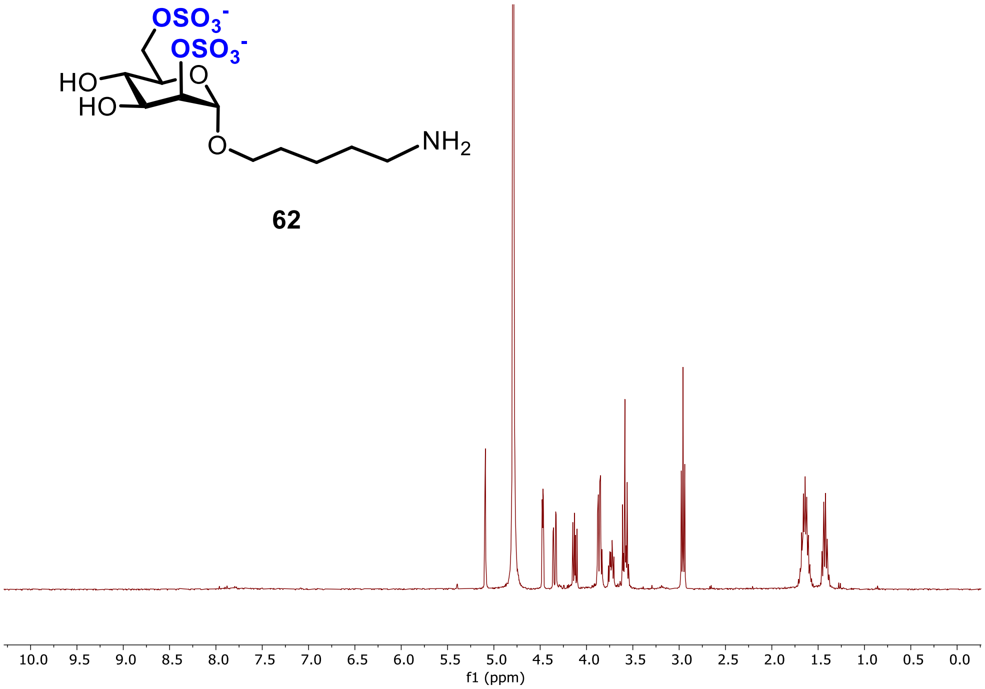
## $^{13}$C NMR of 59 (151 MHz, $D_2O$)



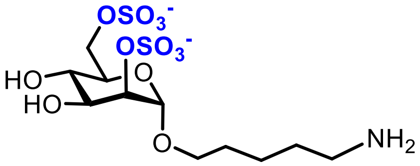
## HSQC NMR of 59 ($D_2O$)



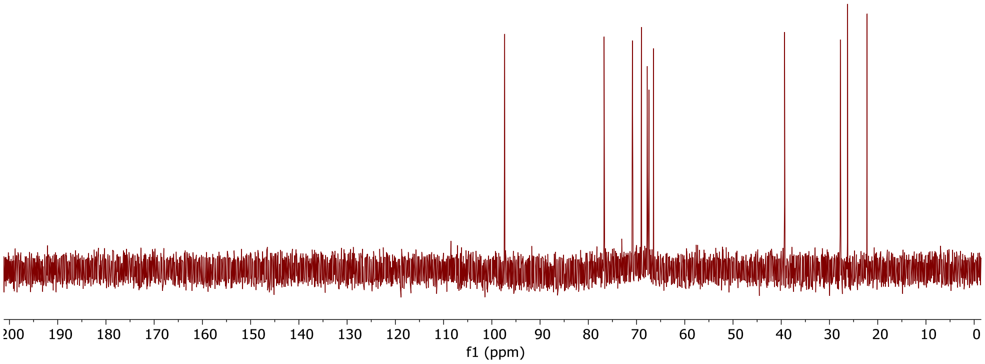
## $^1$H NMR of 62 (400 MHz, $D_2O$)



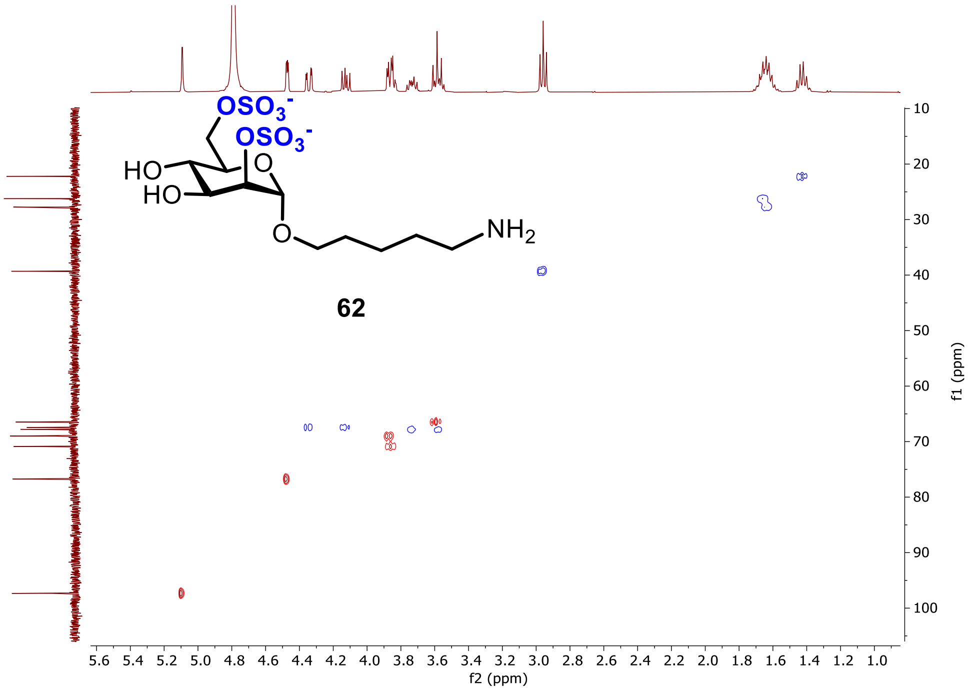
## $^{13}C$ NMR of 10 (101 MHz, $D_2O$)



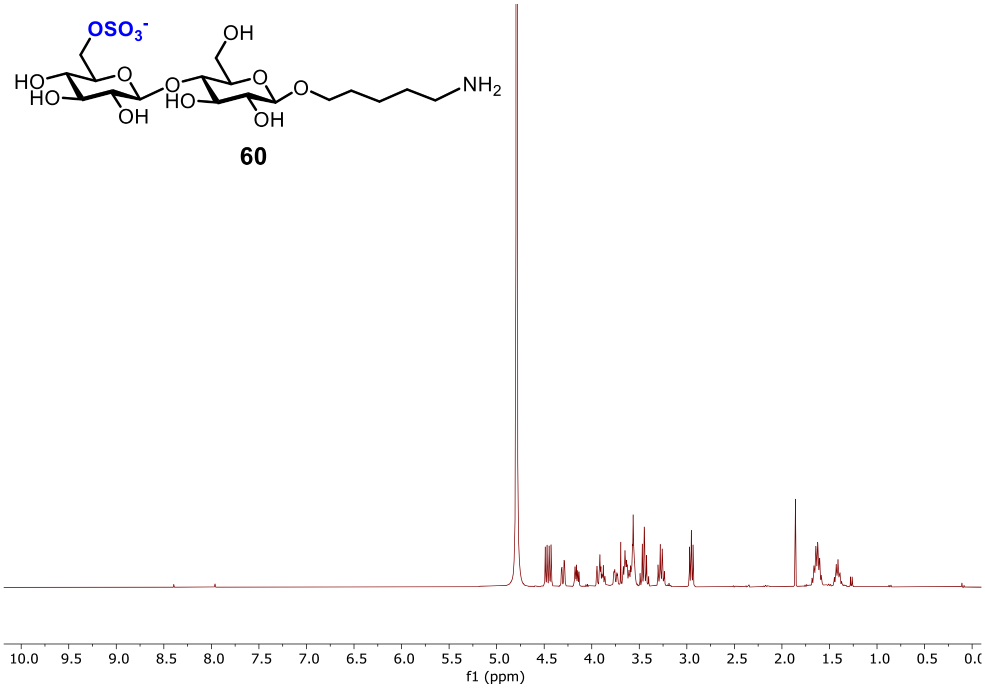
**62**



## HSQC NMR of 62 ($D_2O$)

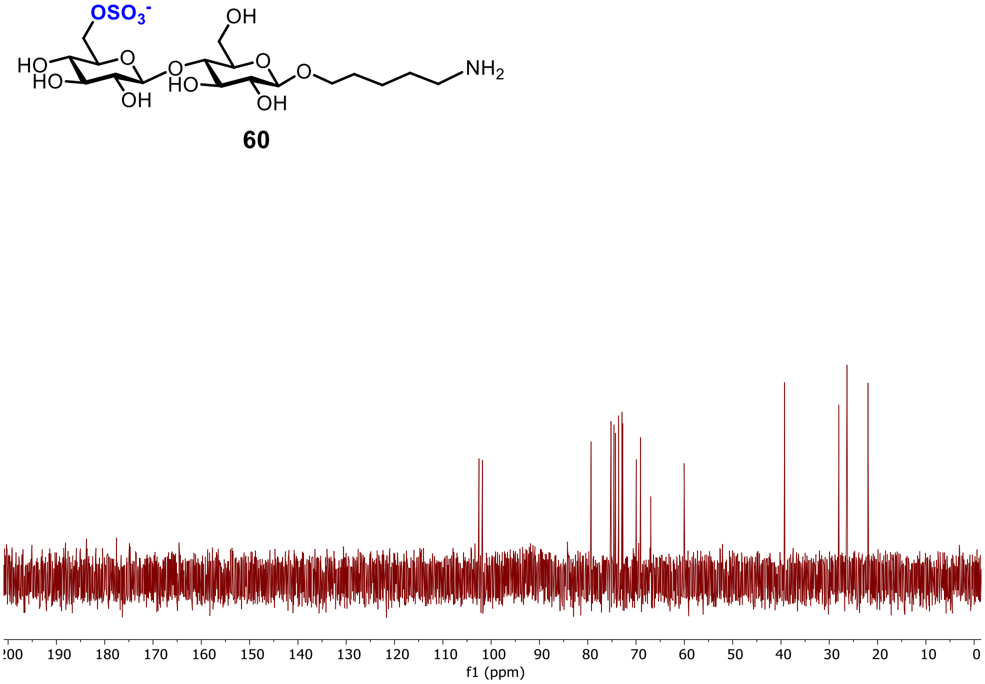


## $^1$H NMR of 60 (400 MHz, $D_2O$)



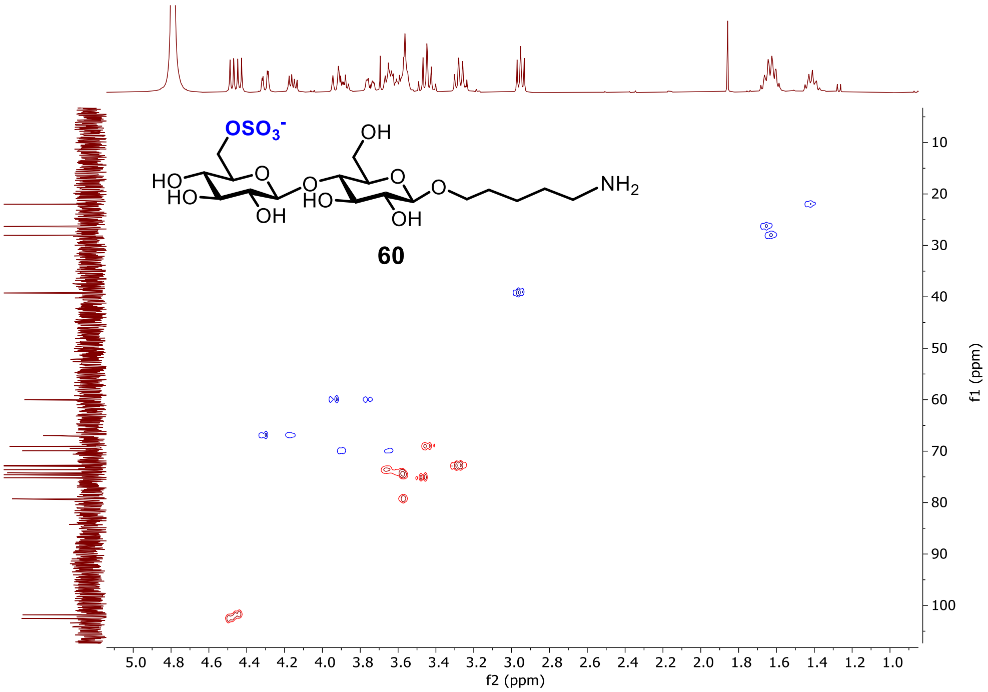
**60**

## $^{13}$C NMR of 60 (101 MHz, $D_2O$)



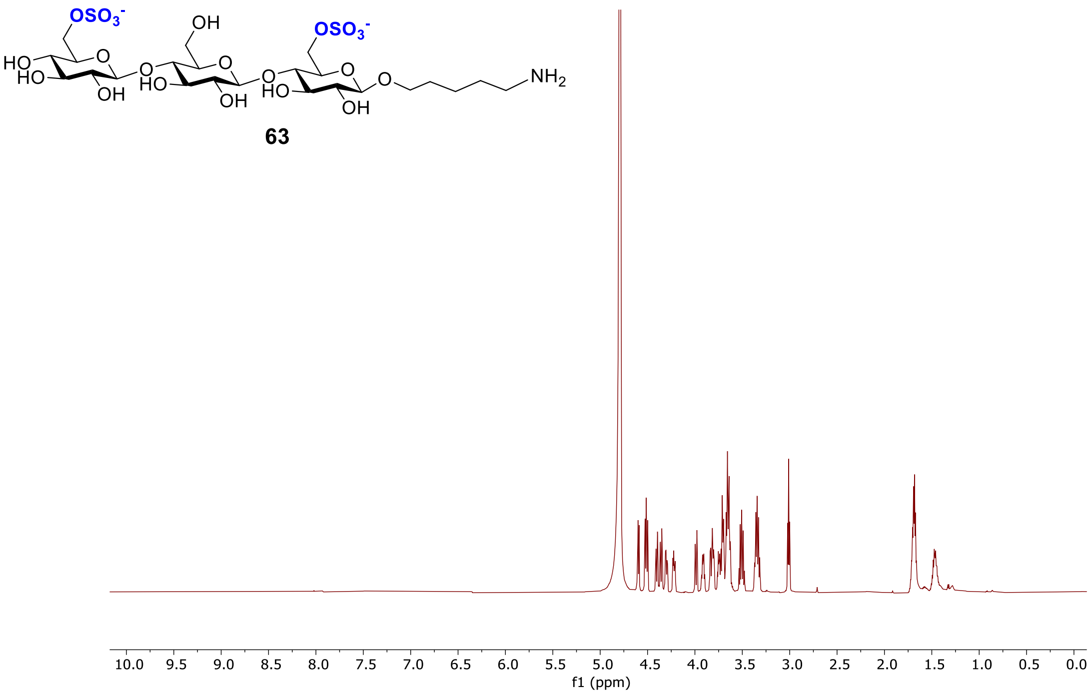
**60**

## HSQC NMR of 60 ($D_2O$)



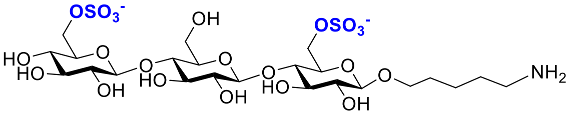
**60**

## 1H NMR of 63 (700 MHz, $D_2O$)

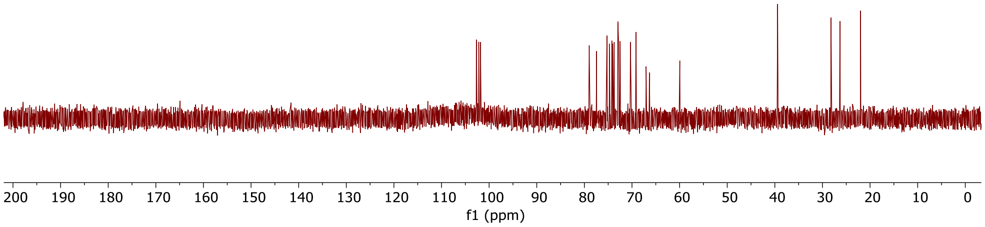


**63**

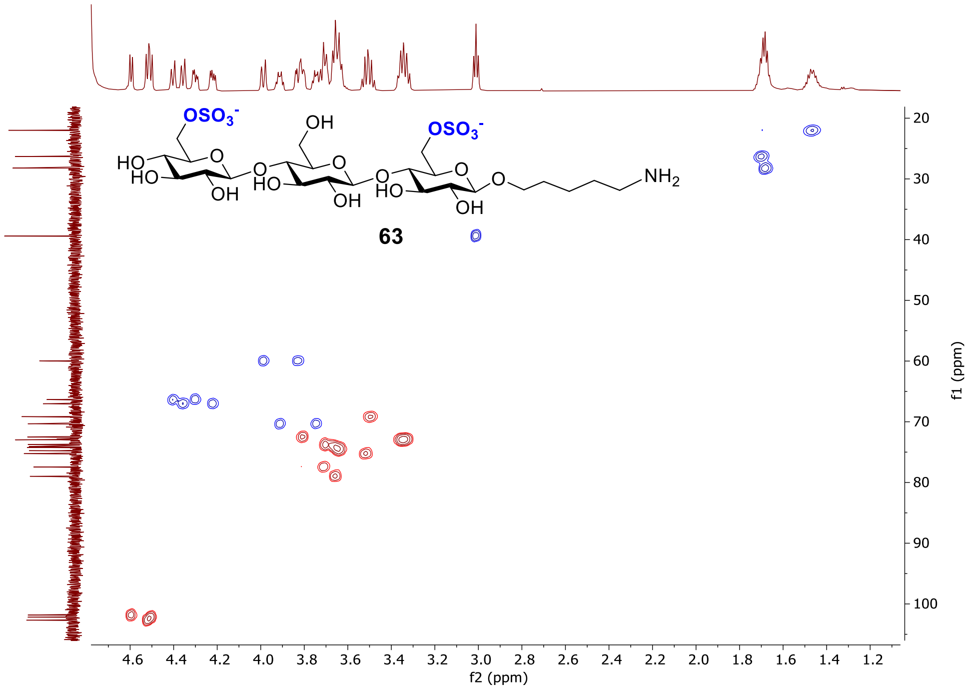
**$^{13}C$ NMR of 63 (176 MHz, $D_2O$)**



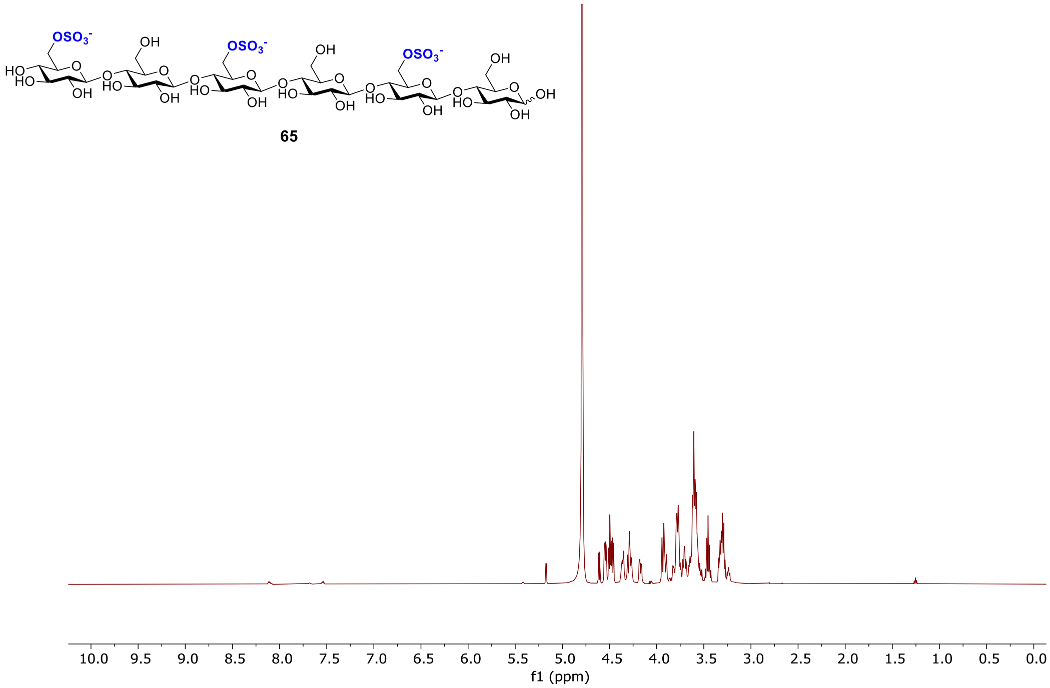
**63**



**HSQC NMR of 63 ($D_2O$)**

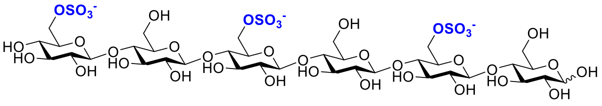


## 1H NMR of 65 (700 MHz, D$_2$O)

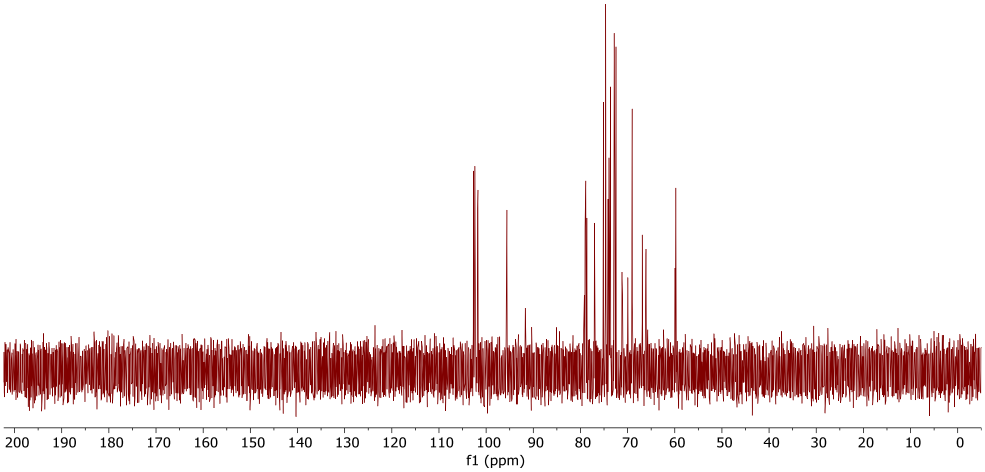


**65**

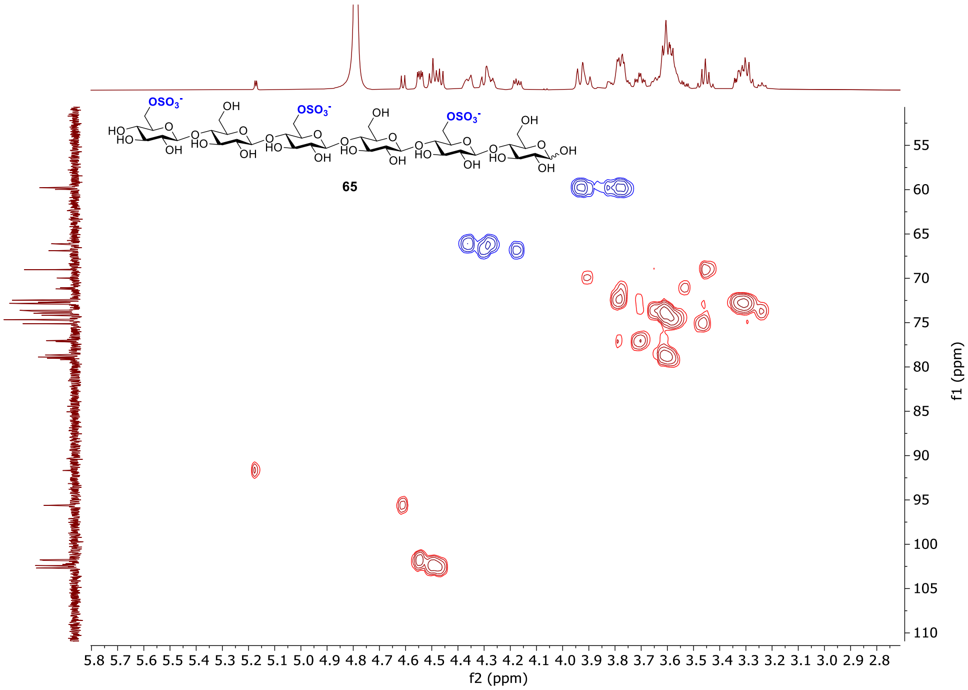
## $^{13}$C NMR of 65 (176 MHz, D$_2$O)



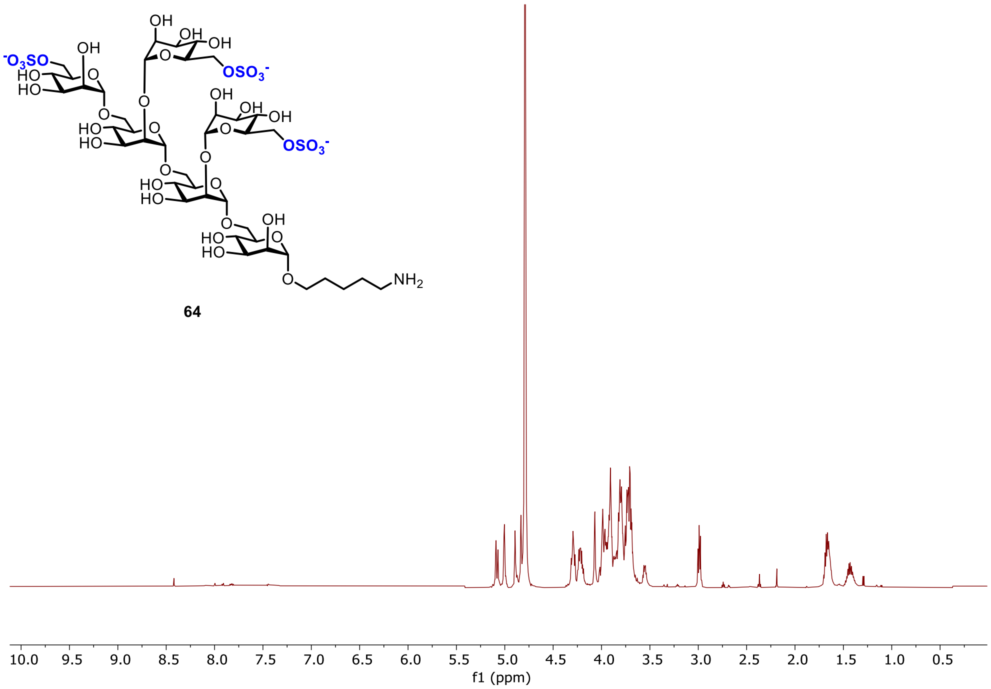
**65**



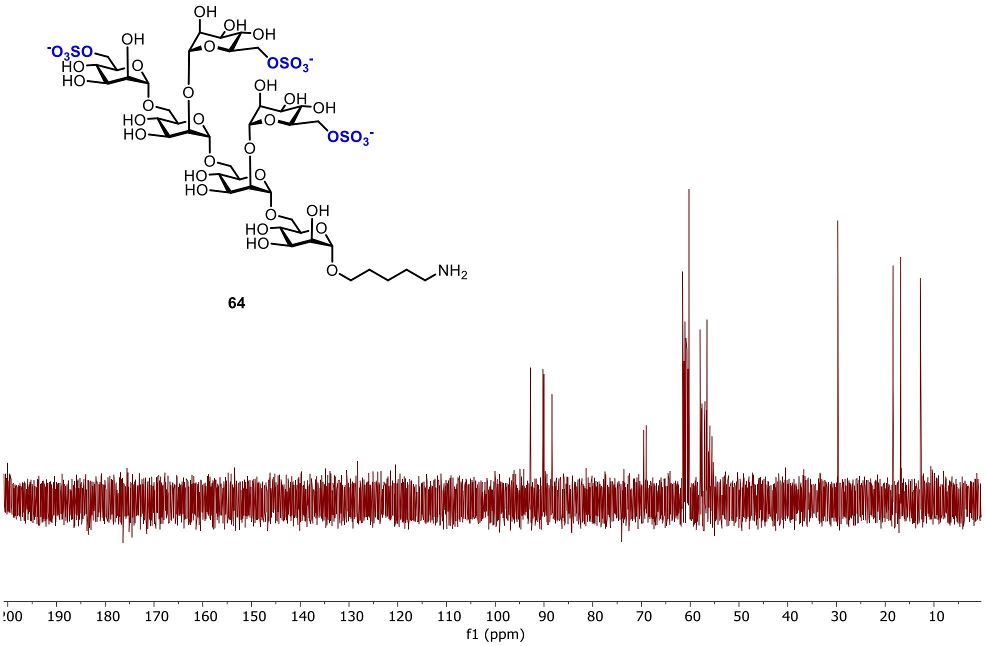
## HSQC NMR of 65 ($D_2$0)



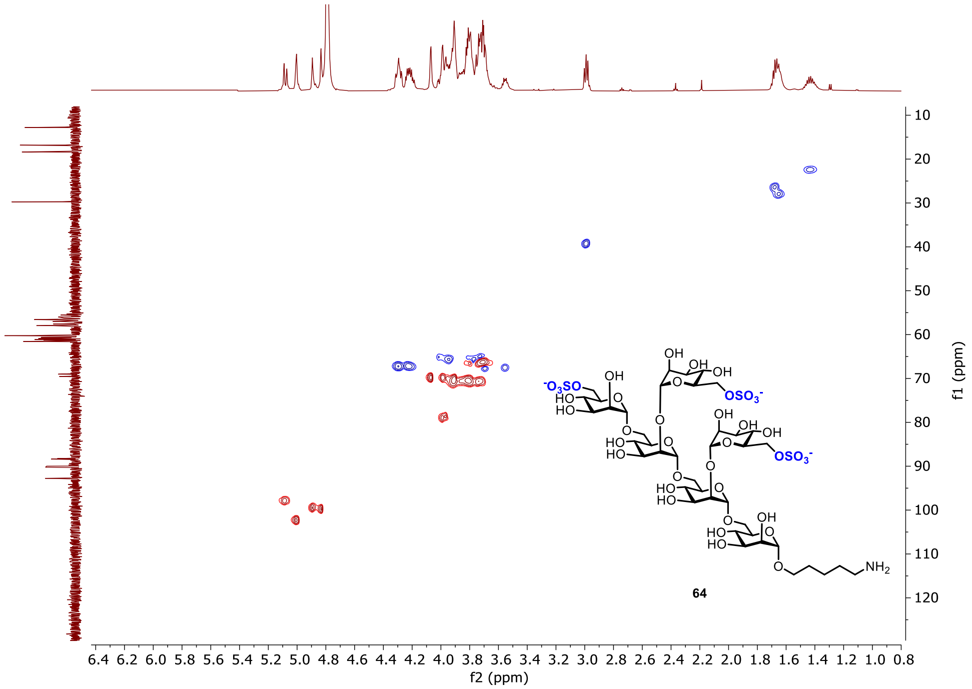
## 1H NMR of 64 (600 MHz, $D_2O$)



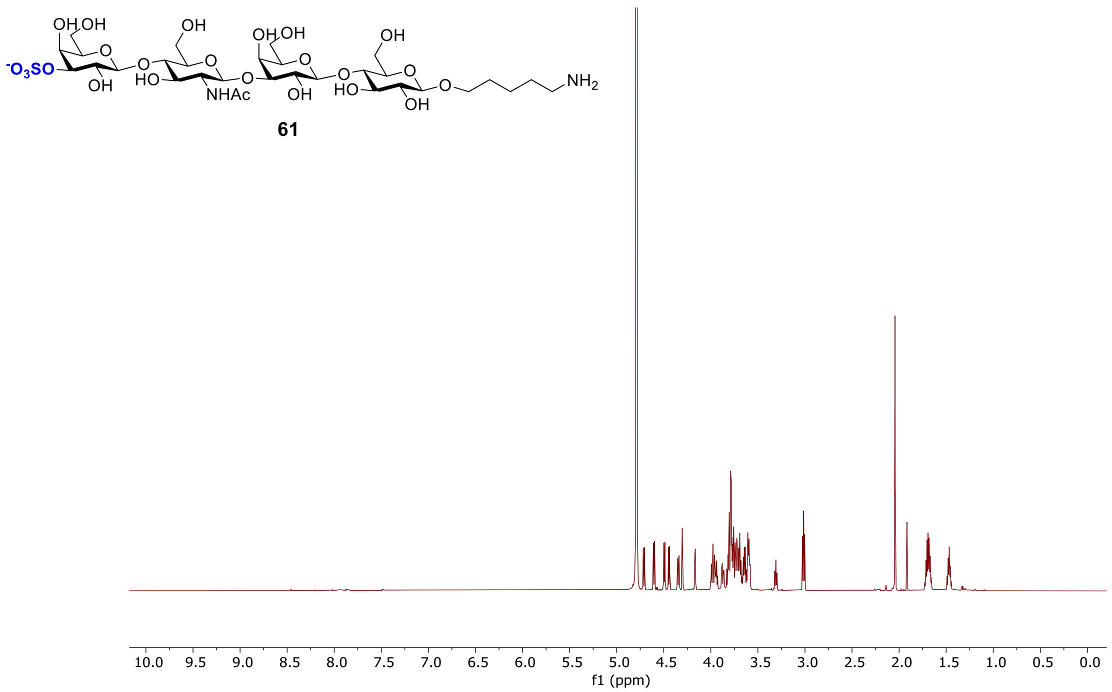
## $^{13}$C NMR of 64 (151 MHz, D$_2$O)



## HSQC NMR of 64 (D$_2$O)

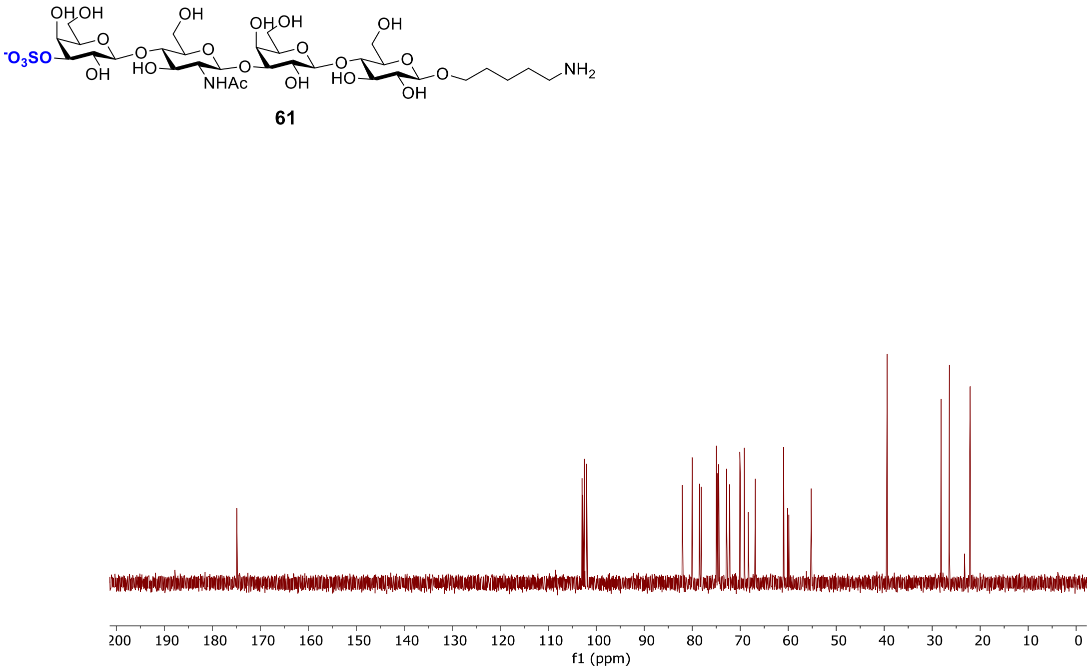


## 1H NMR of 61 (700 MHz, $D_2O$)



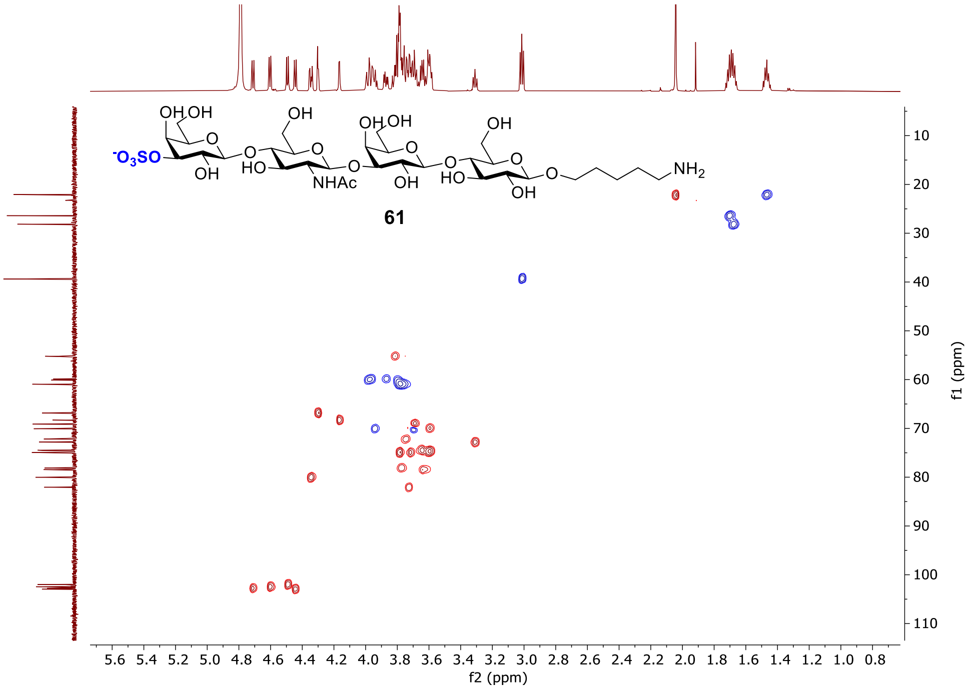
**61**

## $^{13}$C NMR of 61 (176 MHz, $D_2O$)

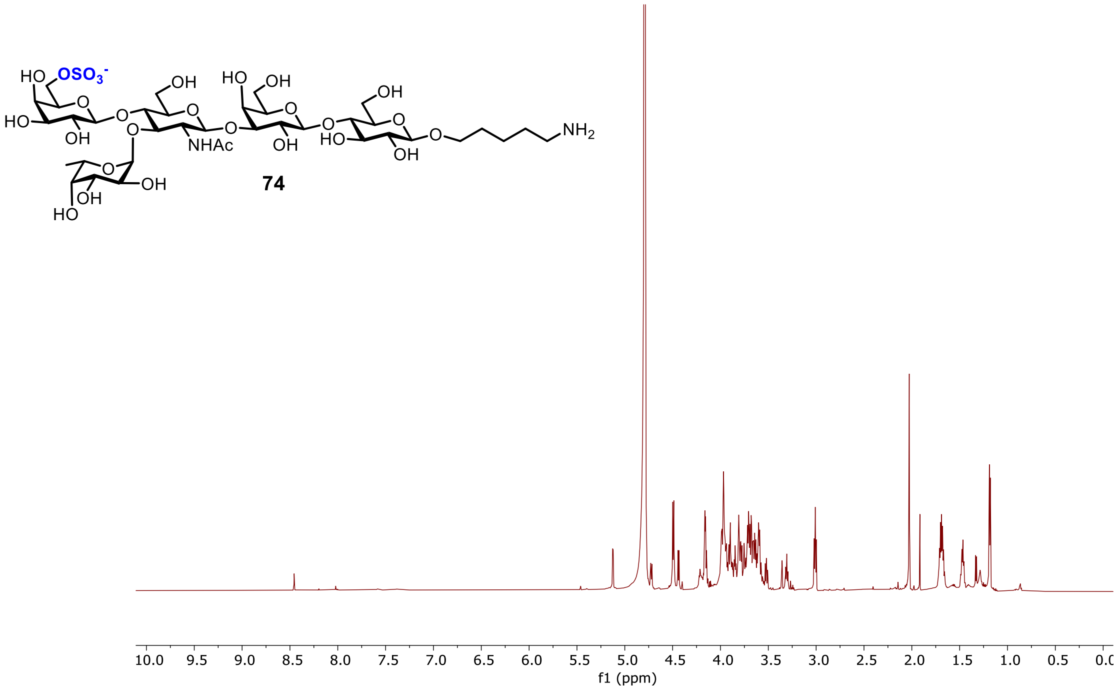


**61**

# HSQC NMR of 61 ($D_2O$)

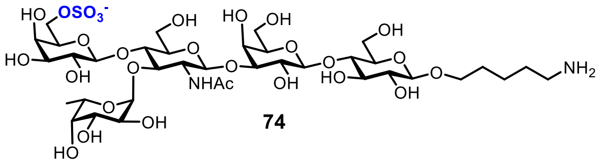


## 1H NMR of 74 (700 MHz, $D_2O$)

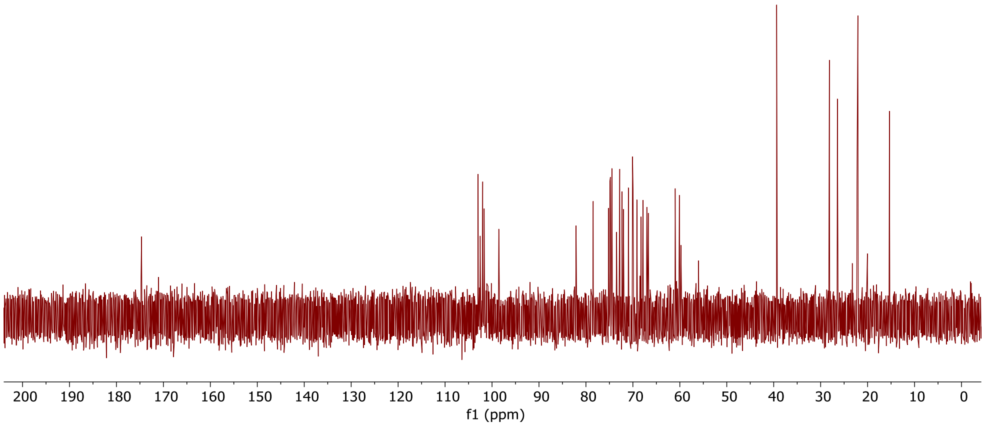


74

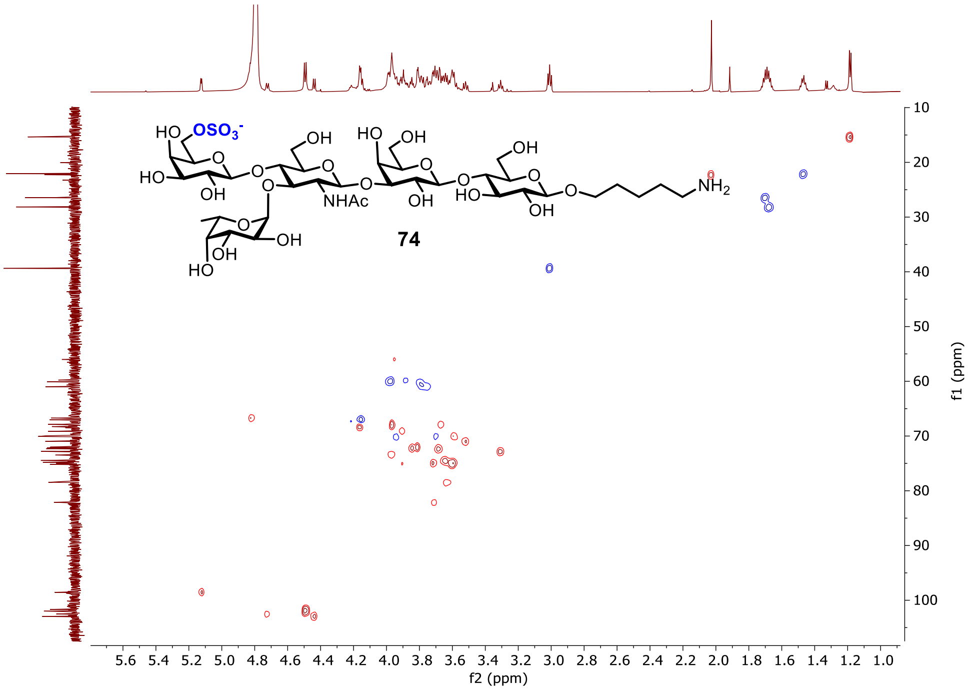
## $^{13}C$ NMR of 74 (176 MHz, $D_2O$)



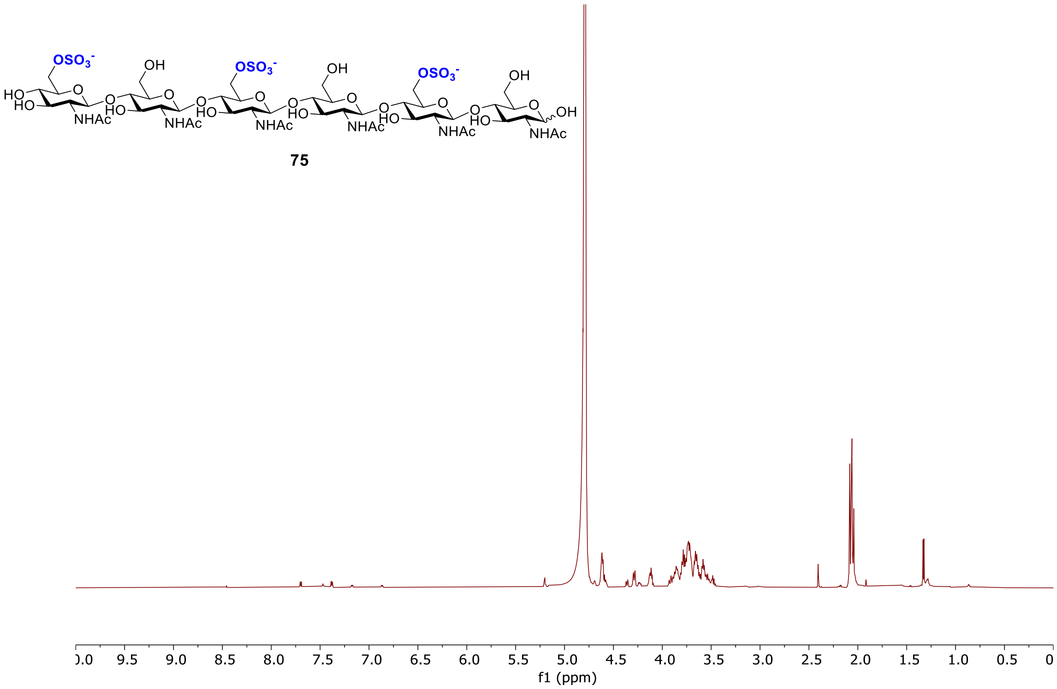
74



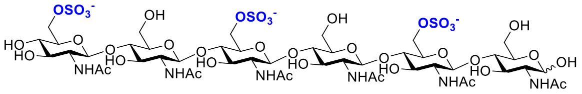
## HSQC NMR of 74 ($D_2O$)



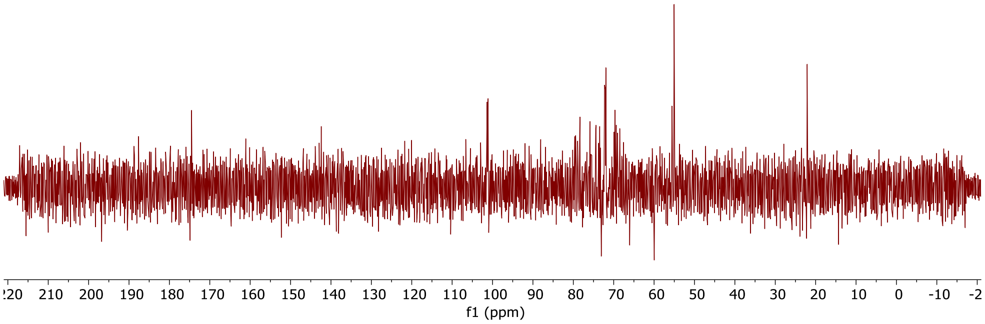
## $^1$H NMR of 75 (700 MHz, $D_2O$)



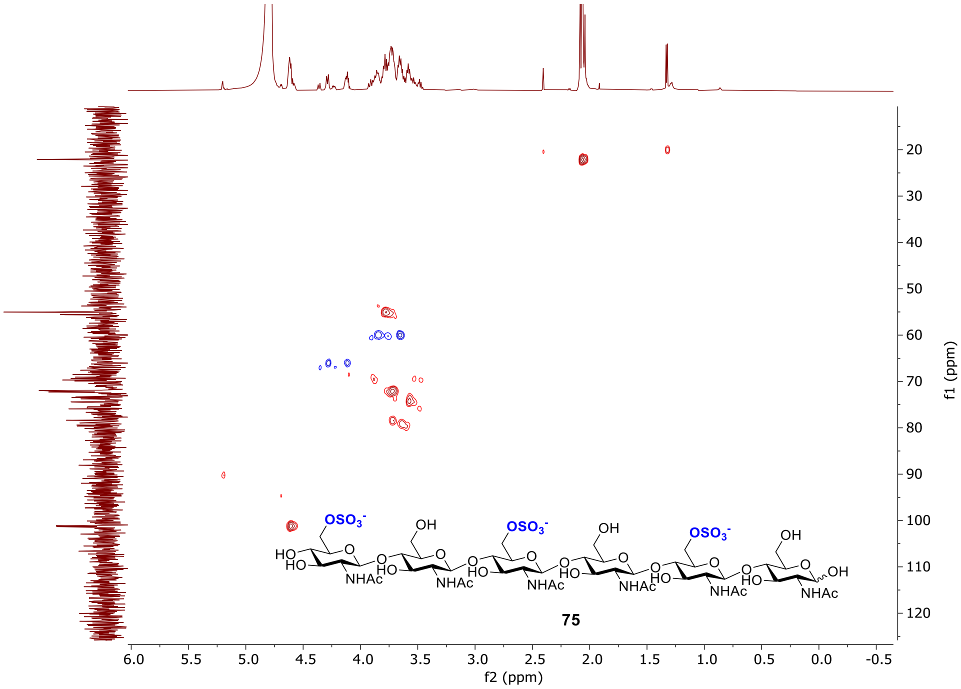
## $^{13}$C NMR of 75 (176 MHz, D$_2$O)



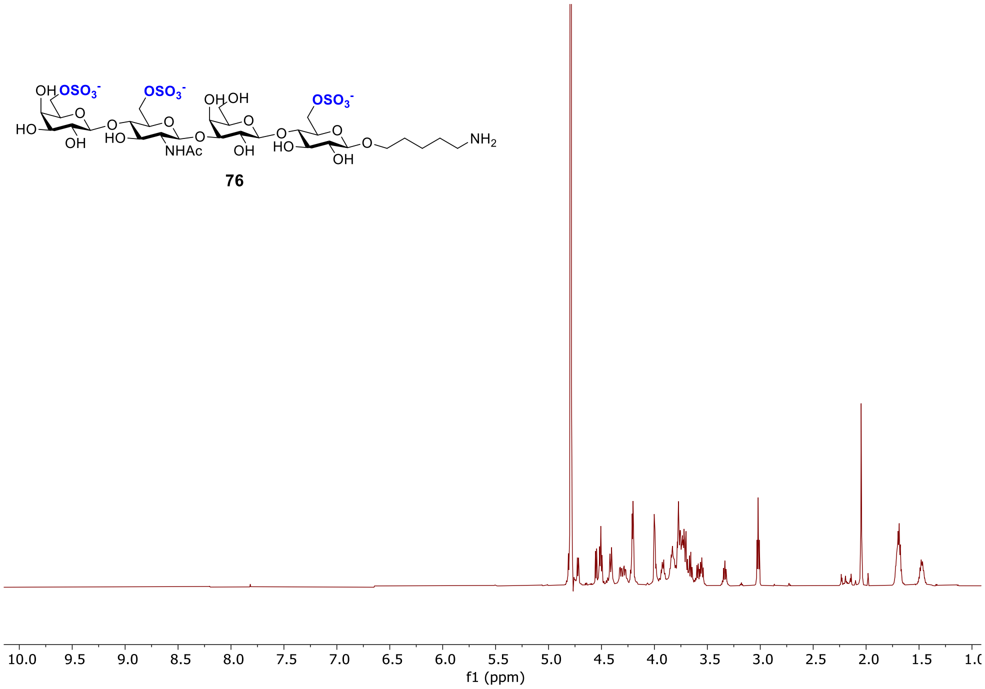
**75**



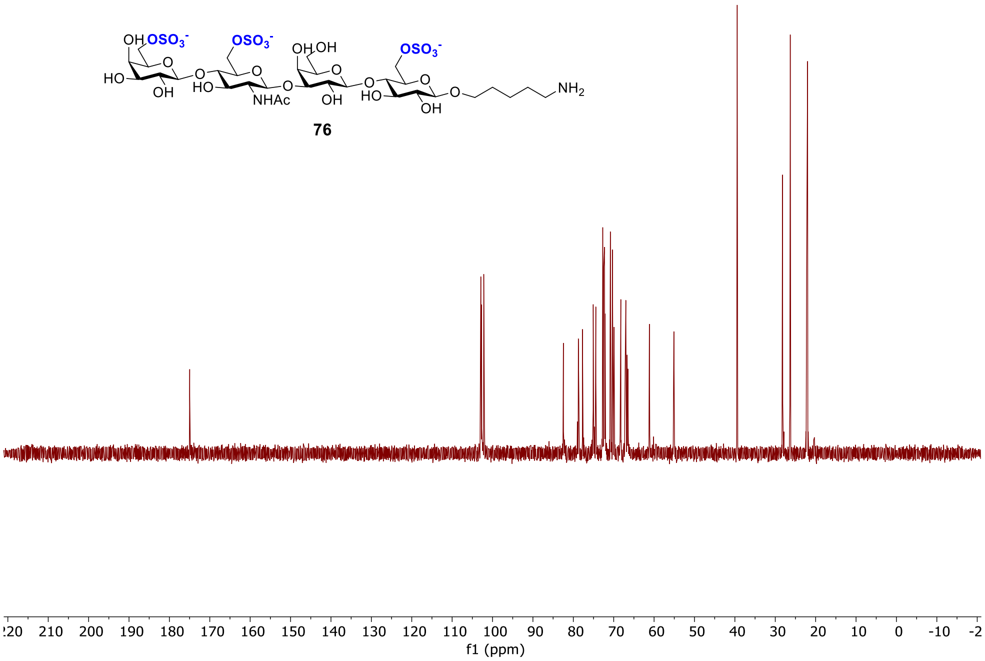
## HSQC NMR of 75 (D$_2$O)



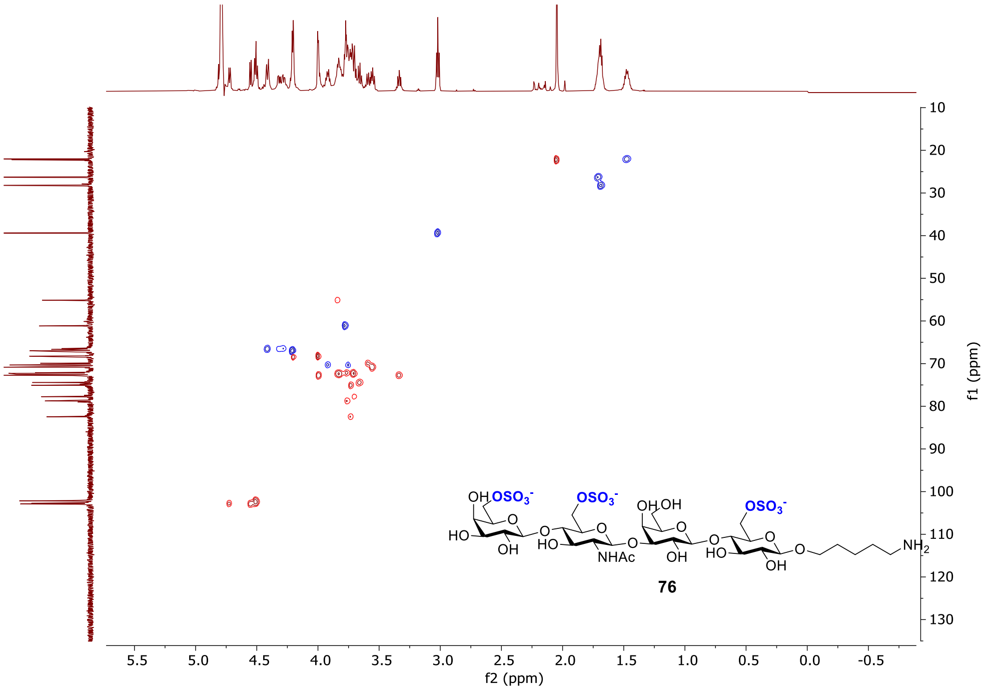
## $^{1}$H NMR of 76 (700 MHz, $D_2O$)



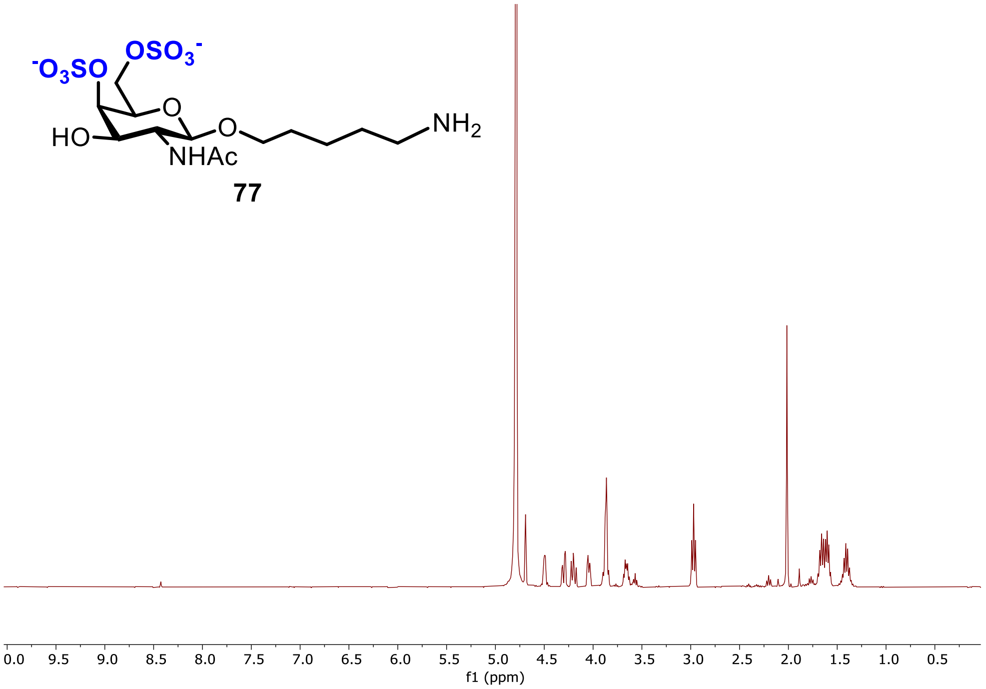
## $^{13}$C NMR of 76 (176 MHz, $D_2O$)



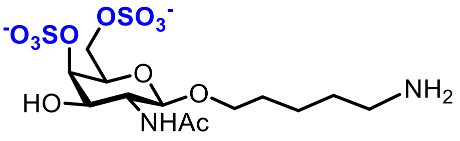
## HSQC NMR of 76 ($D_2O$)



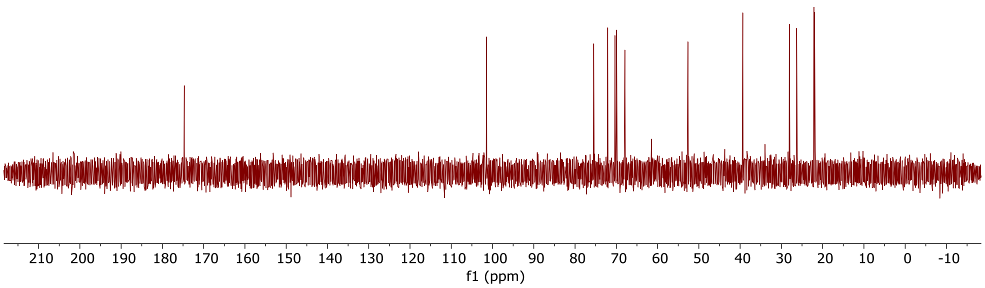
## $^1$H NMR of 77 (400 MHz, $D_2O$)



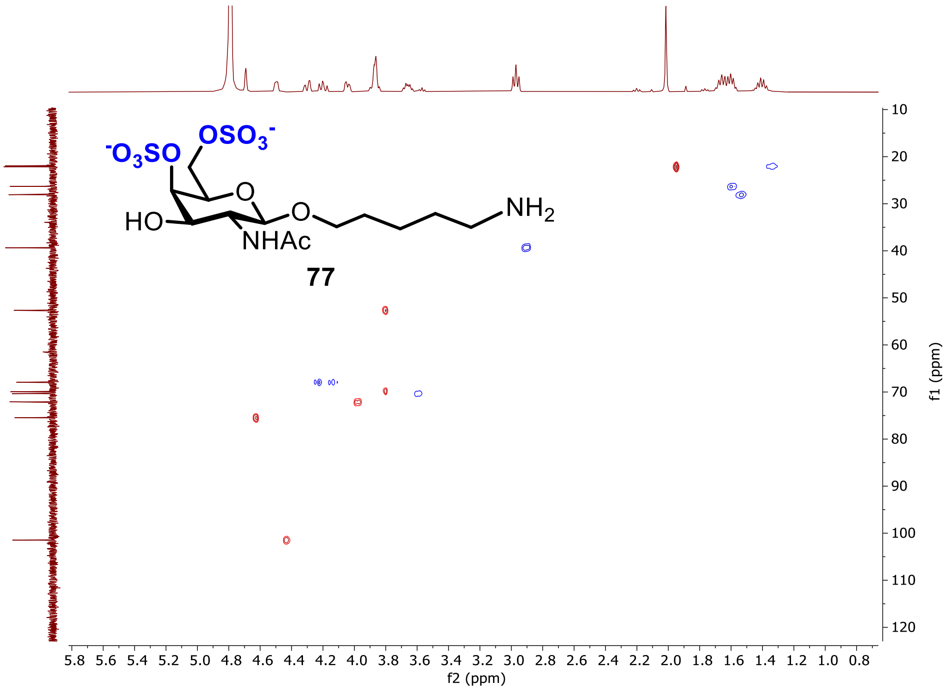
## $^{13}$C NMR of 77 (101 MHz, $D_2O$)



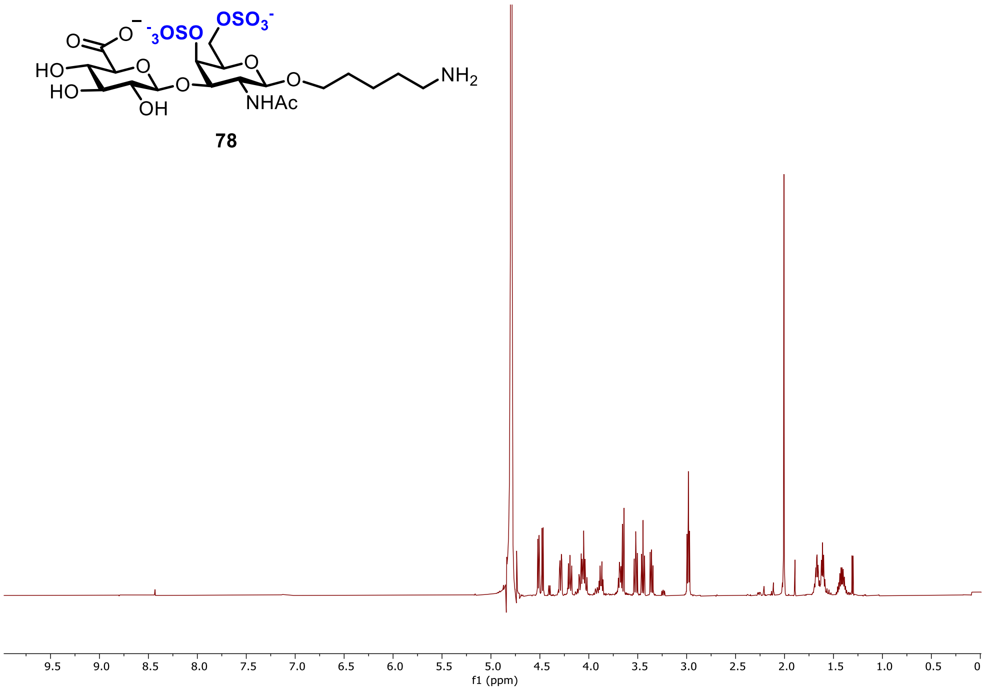
**77**



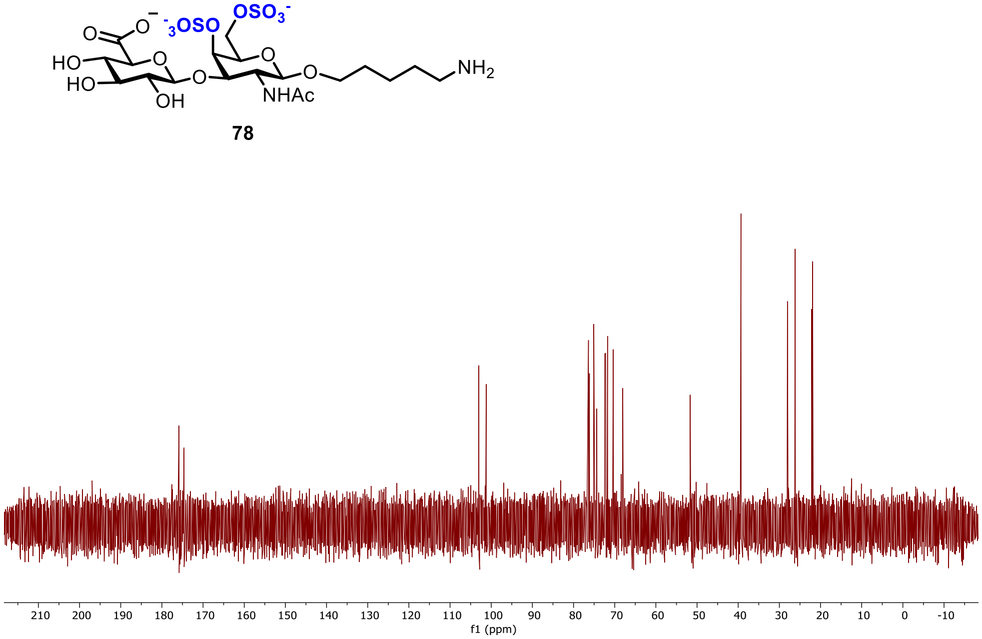
## HSQC NMR of 77 ($D_2O$)



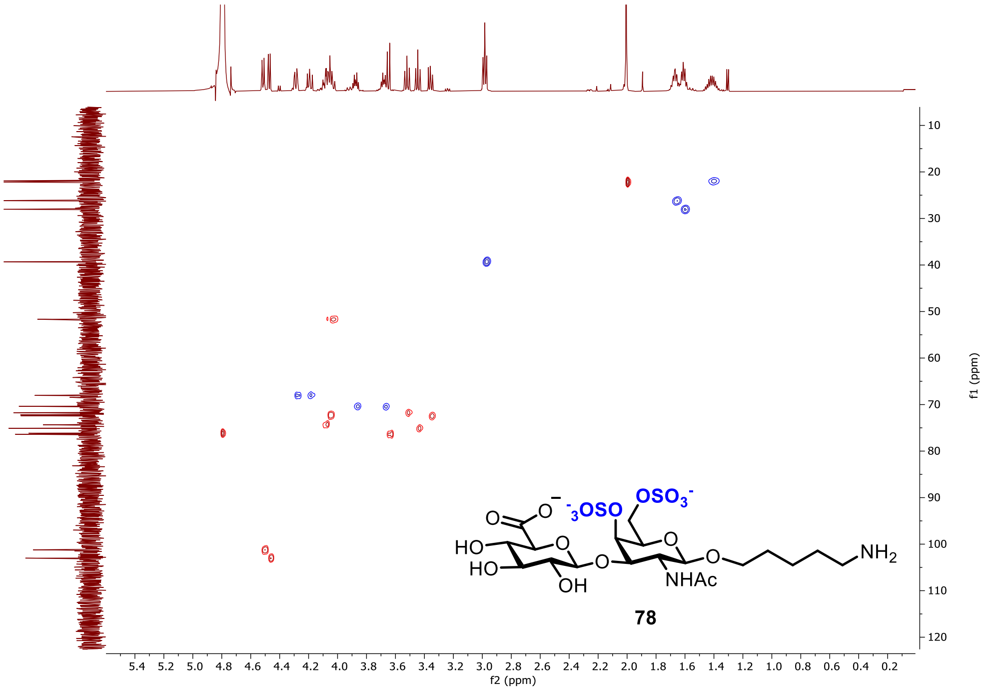
## $^{1}H$ NMR of 78 (600 MHz, D2O)



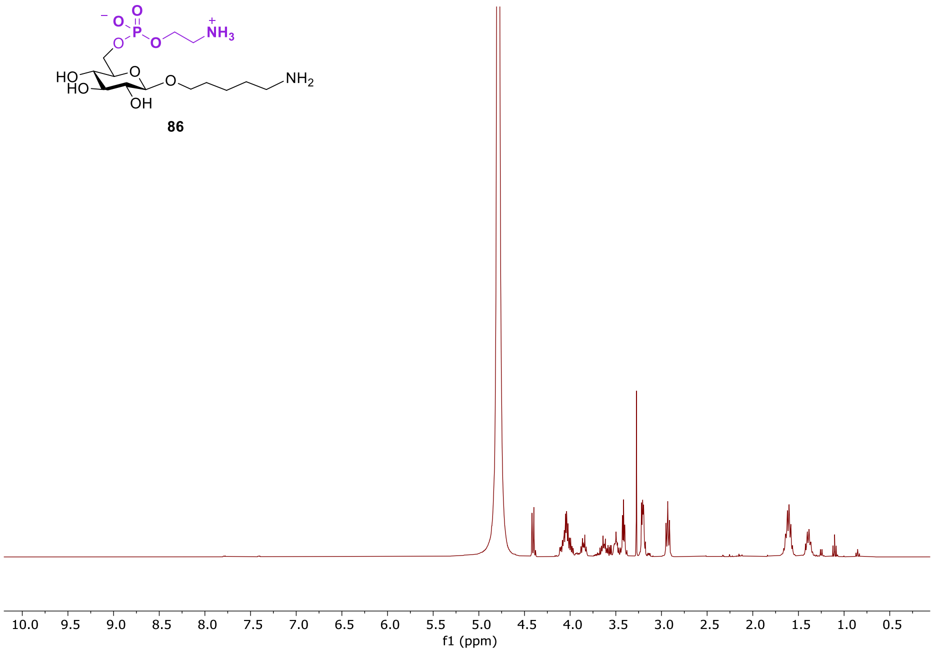
## $^{13}C$ NMR of 78 (151 MHz, $D_2O$)



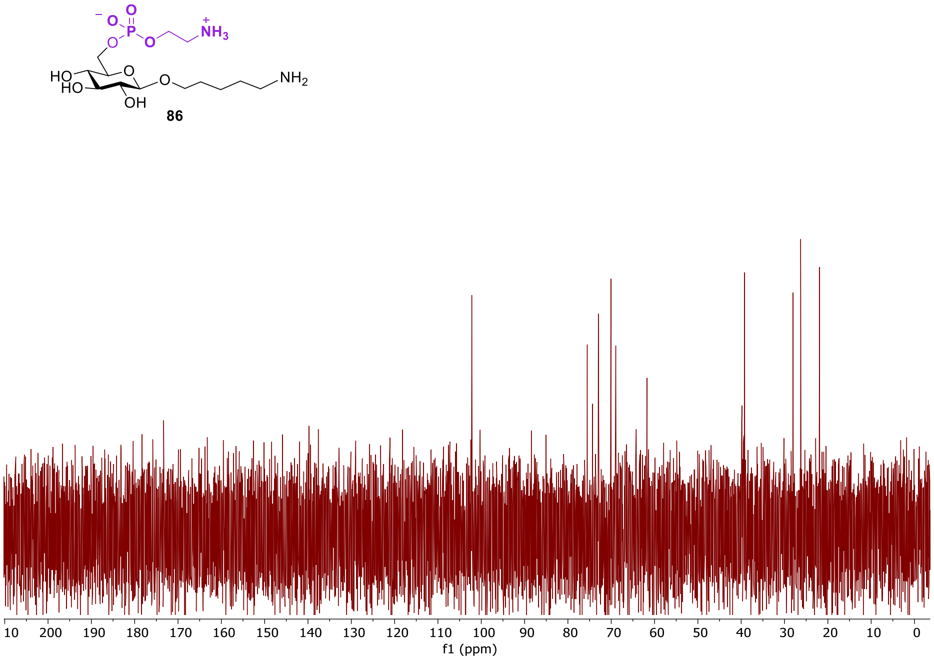
# HSQC NMR of 78 ($D_2O$)



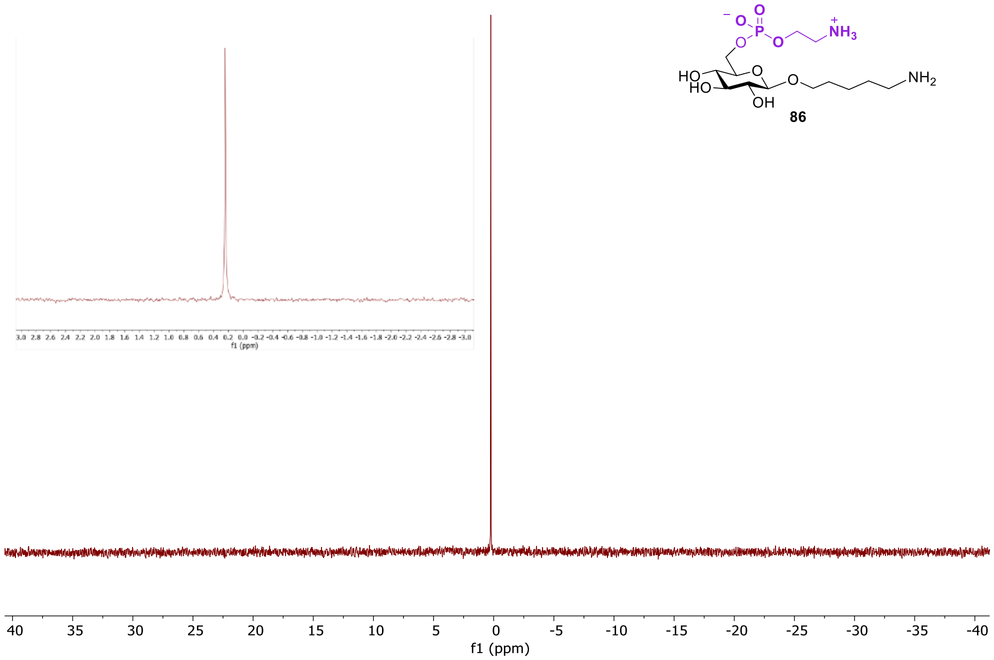
## $^1$H NMR of 86 (400 MHz, $D_2O$)



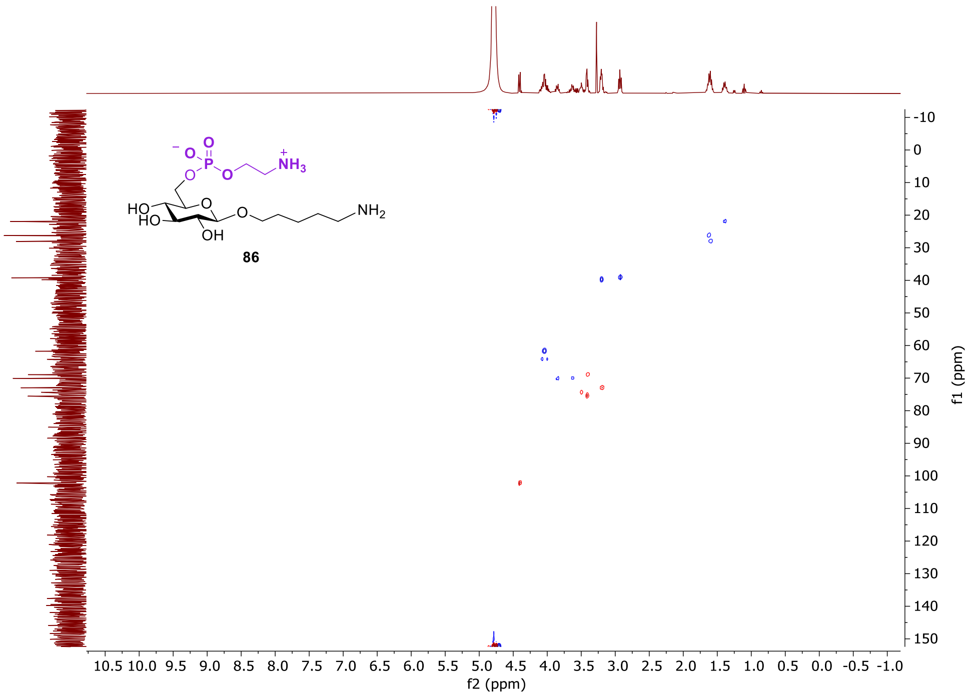
## $^{13}$C NMR of 86 (101 MHz, $D_2O$)



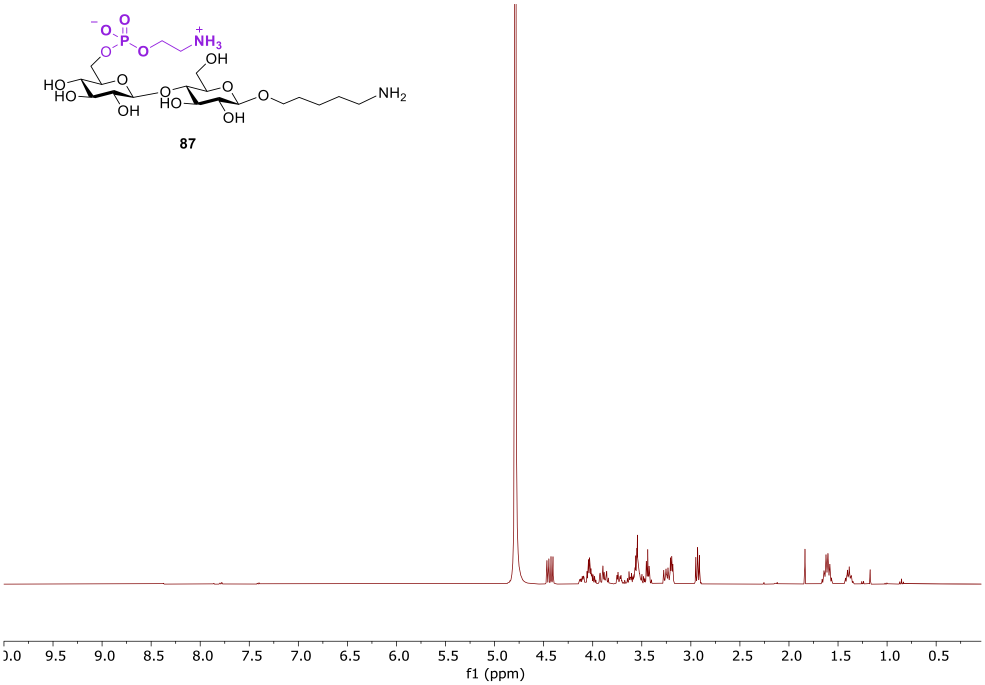
## $^{31}$P NMR of 86 (162 MHz, $D_2O$)



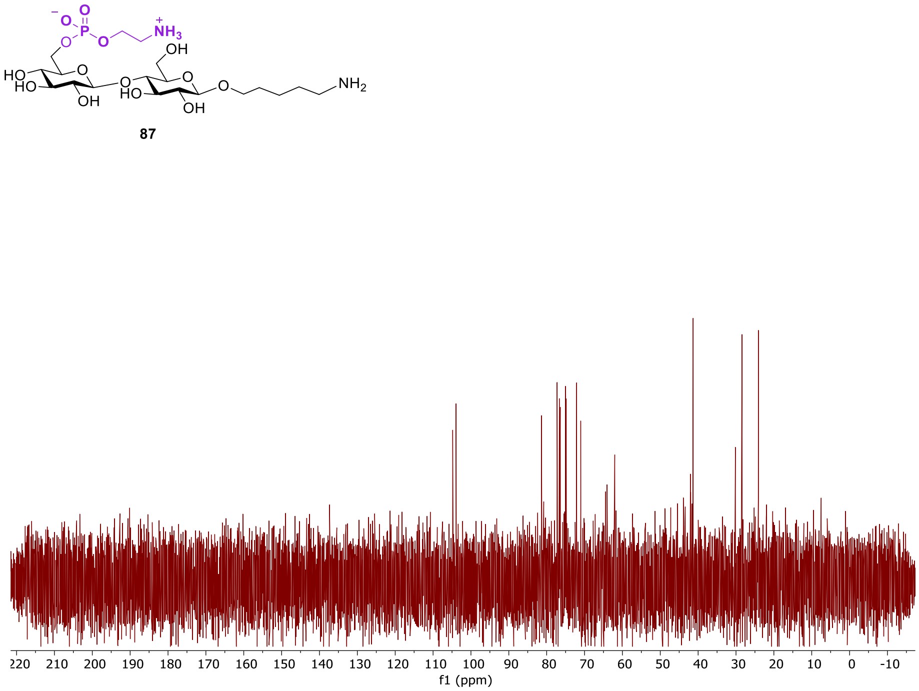
## HSQC NMR of 86 ($D_2O$)



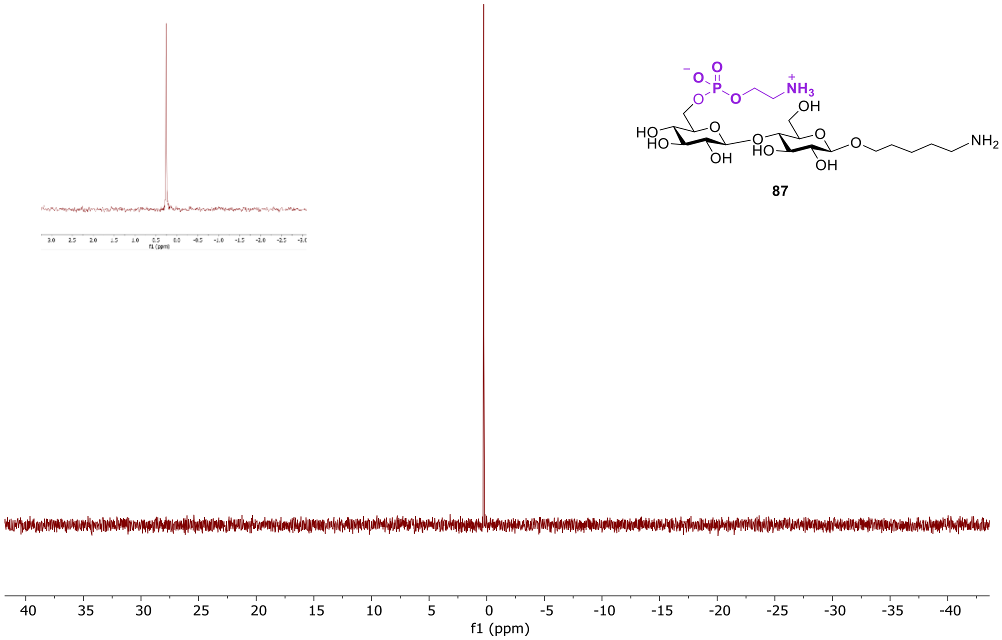
# $^1$H NMR of 87 (400 MHz, D$_2$O)



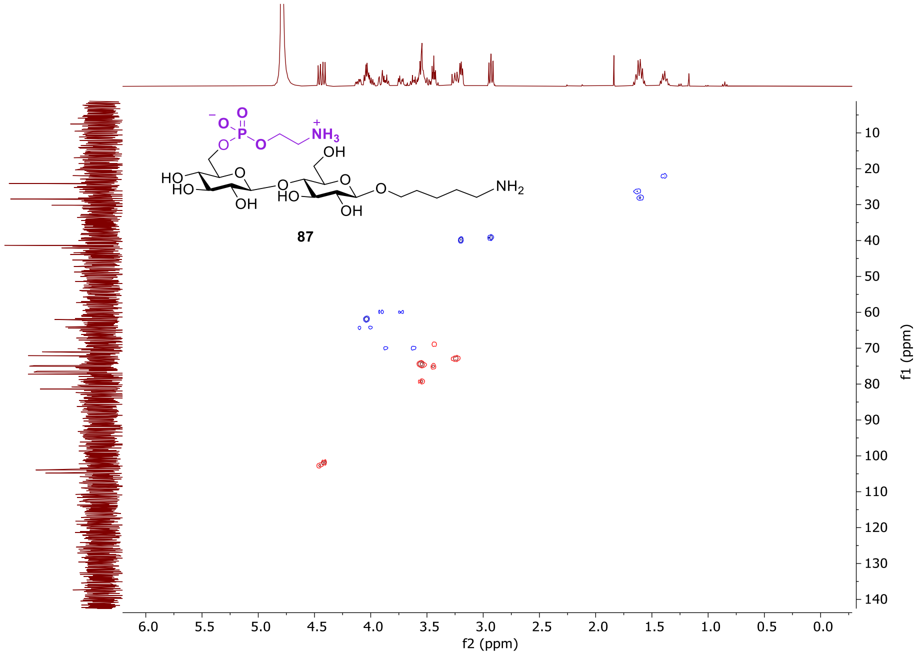
# $^{13}$C NMR of 87 (101 MHz, D$_2$O)



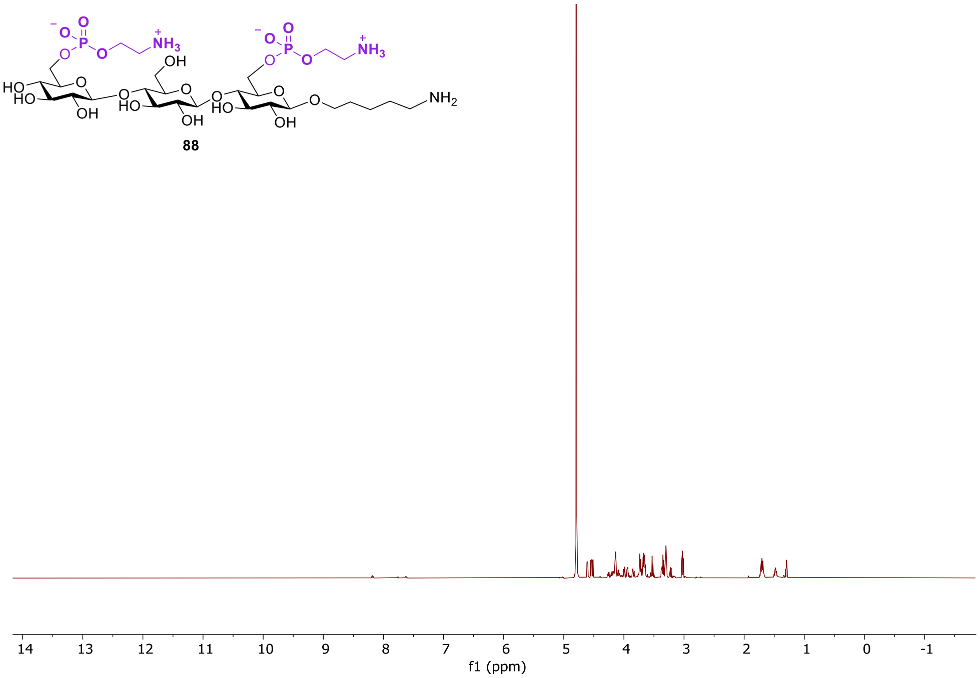
## $^{31}$P NMR of 87 (162 MHz, $D_2O$)



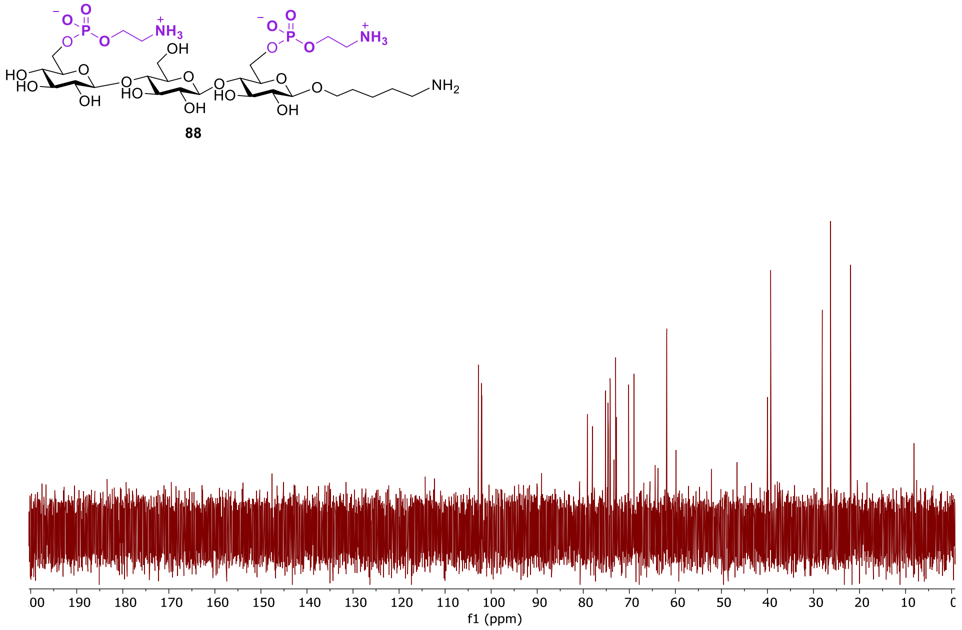
## HSQC NMR of 87 ($D_2O$)



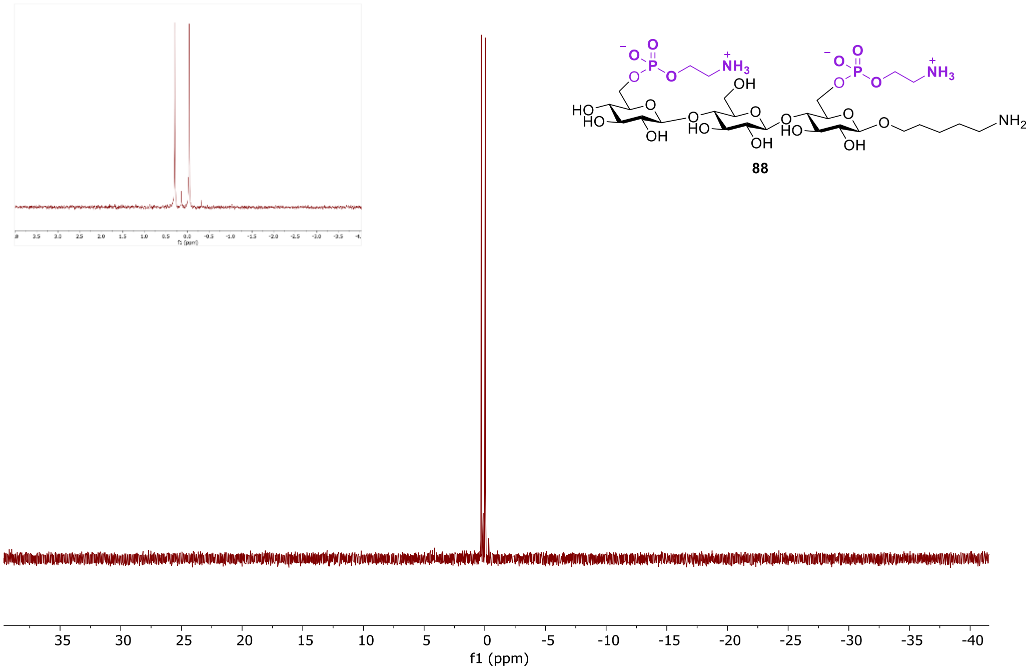
## $^1$H NMR of 88 ( 600 MHz, D$_2$O)


88

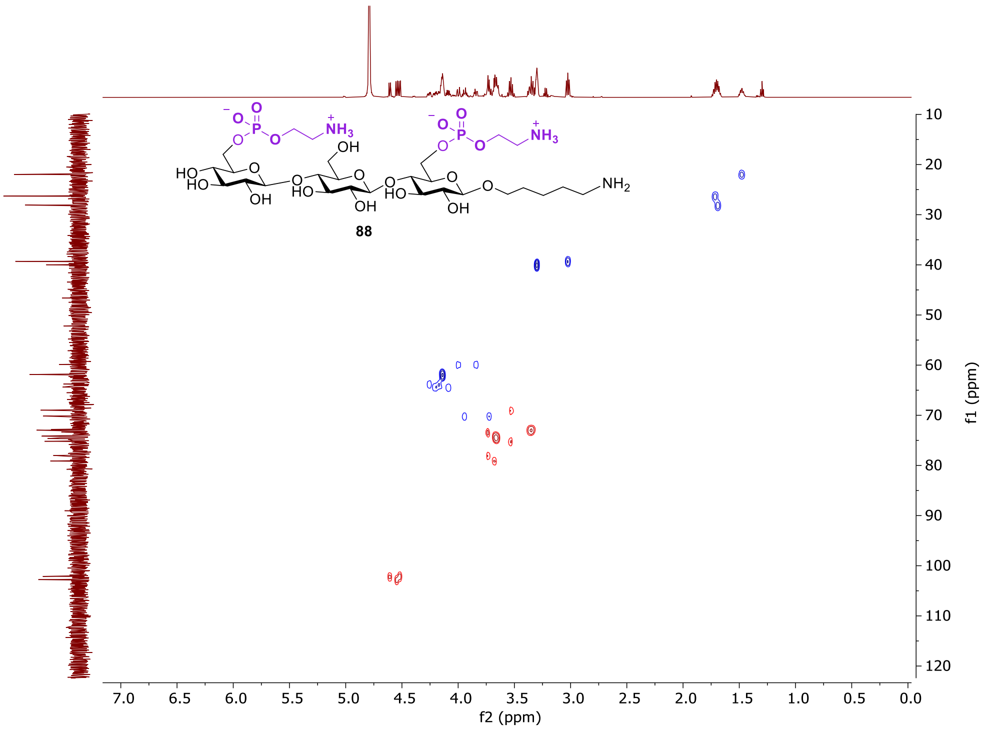
## $^{13}$C NMR of 88 (151 MHz, D$_2$O)


88

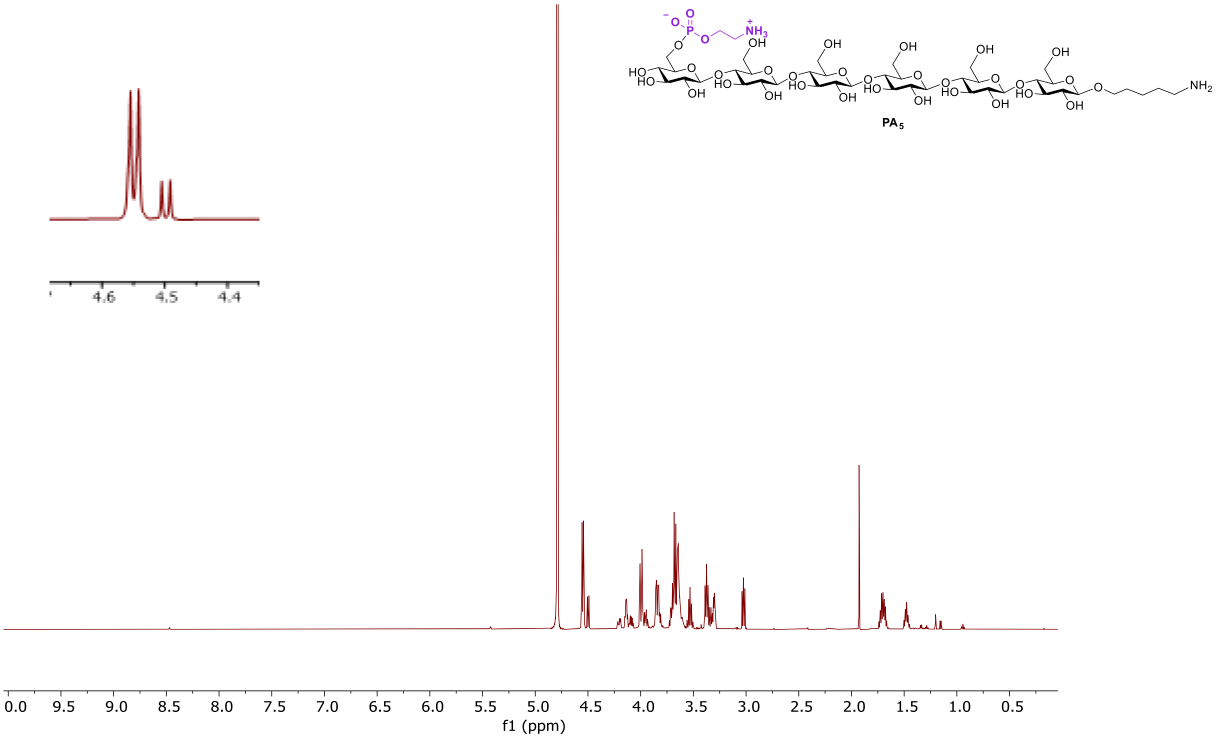
# $^{31}$P NMR of 88 (243 MHz, D$_2$O)



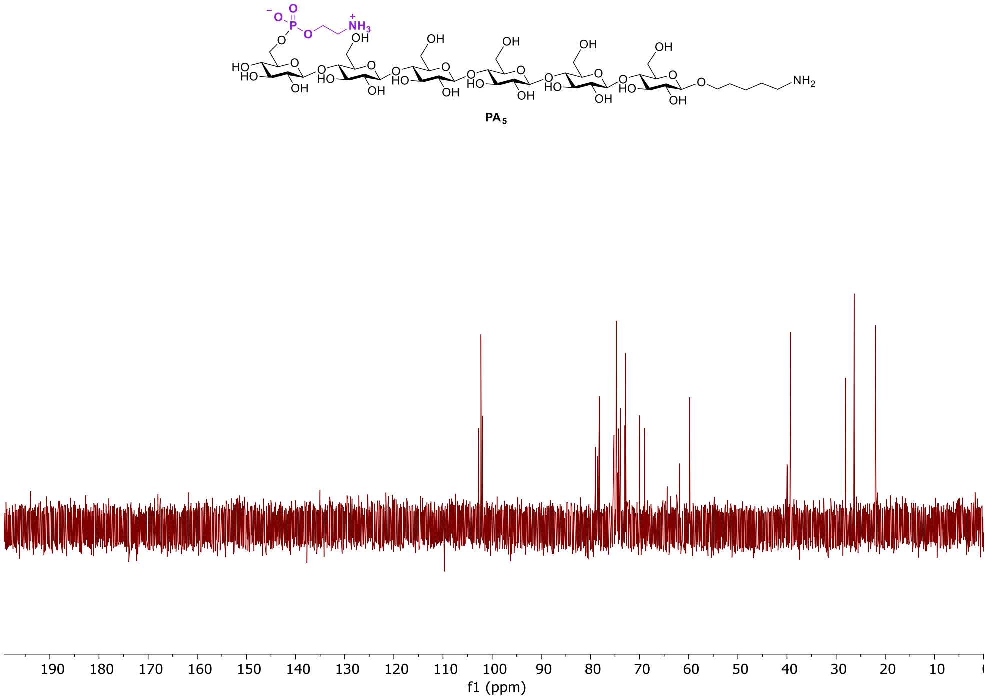
# HSQC NMR of 88 (D$_2$0)



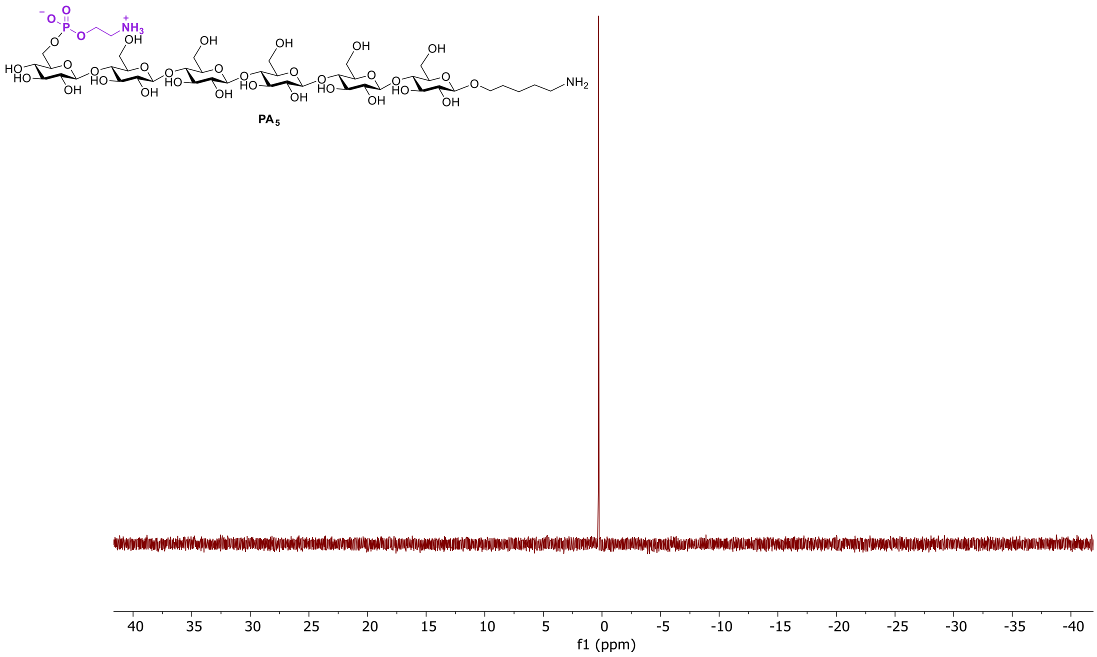
## $^{1}$H NMR of PA$_5$ (600 MHz, D$_2$O)



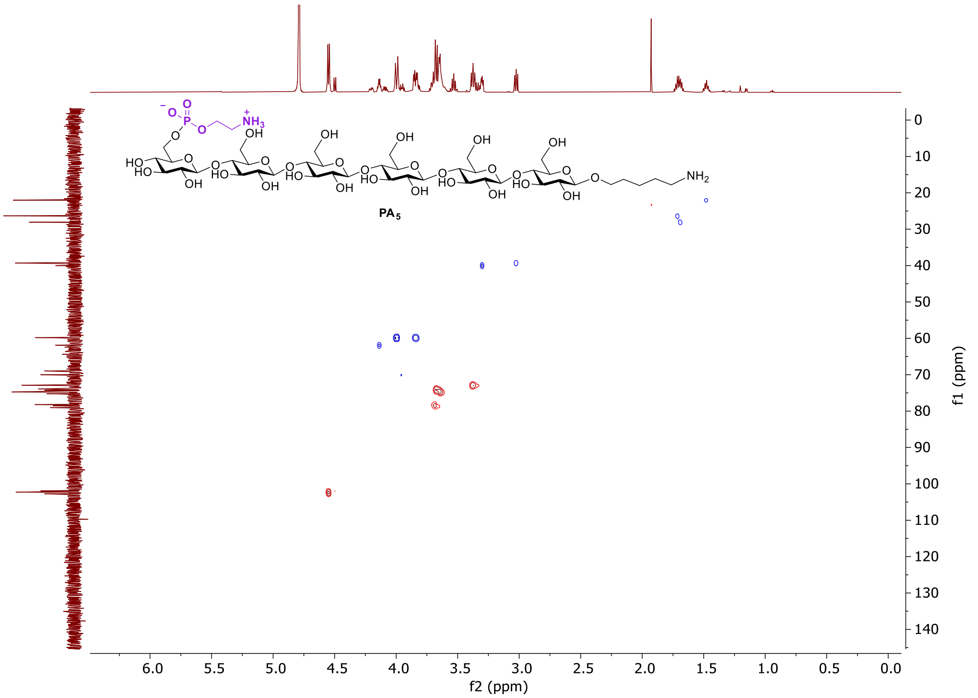
## $^{13}$C NMR of PA$_5$ (151 MHz, D$_2$O)



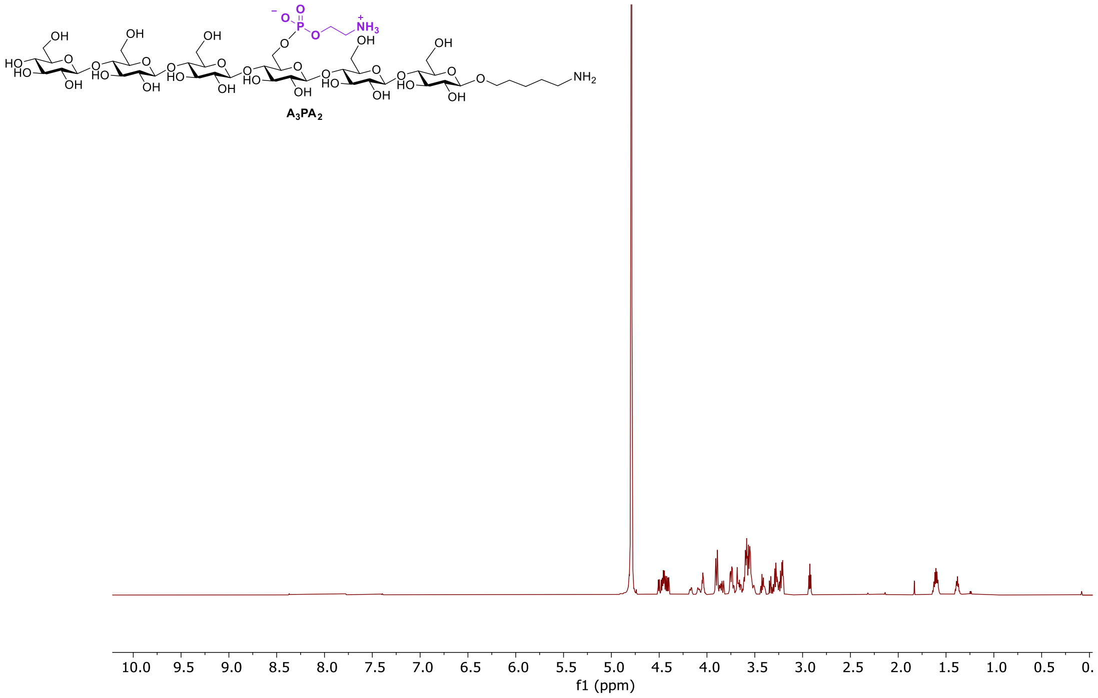
## $^{31}P$ NMR of $PA_5$ (243 MHz, $D_2O$)



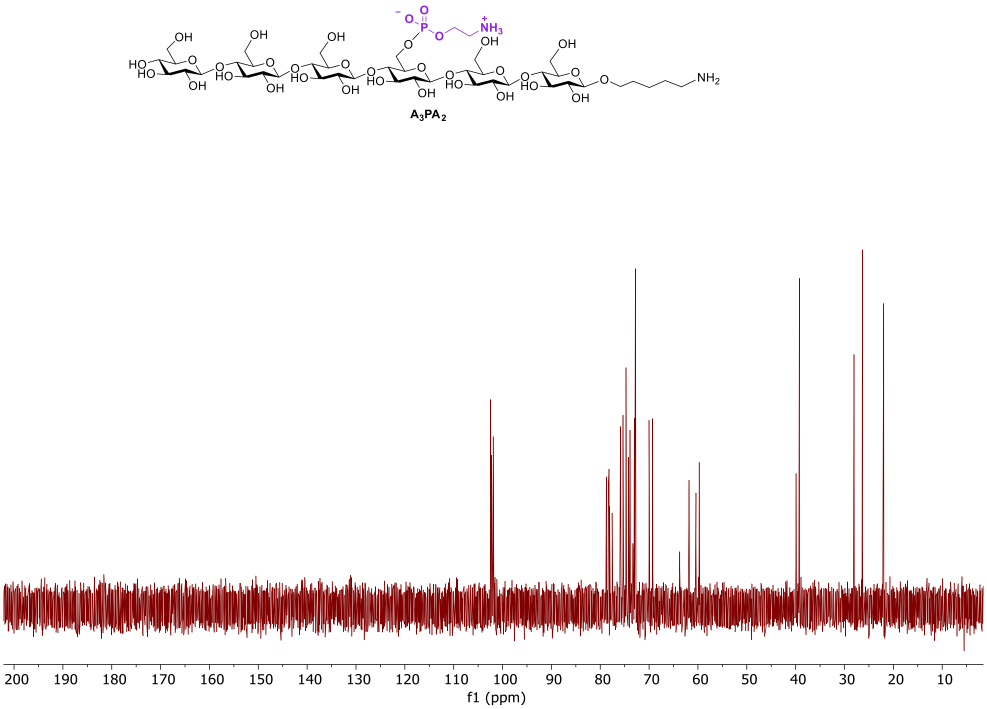
## HSQC NMR of $PA_5$ ($D_2O$)



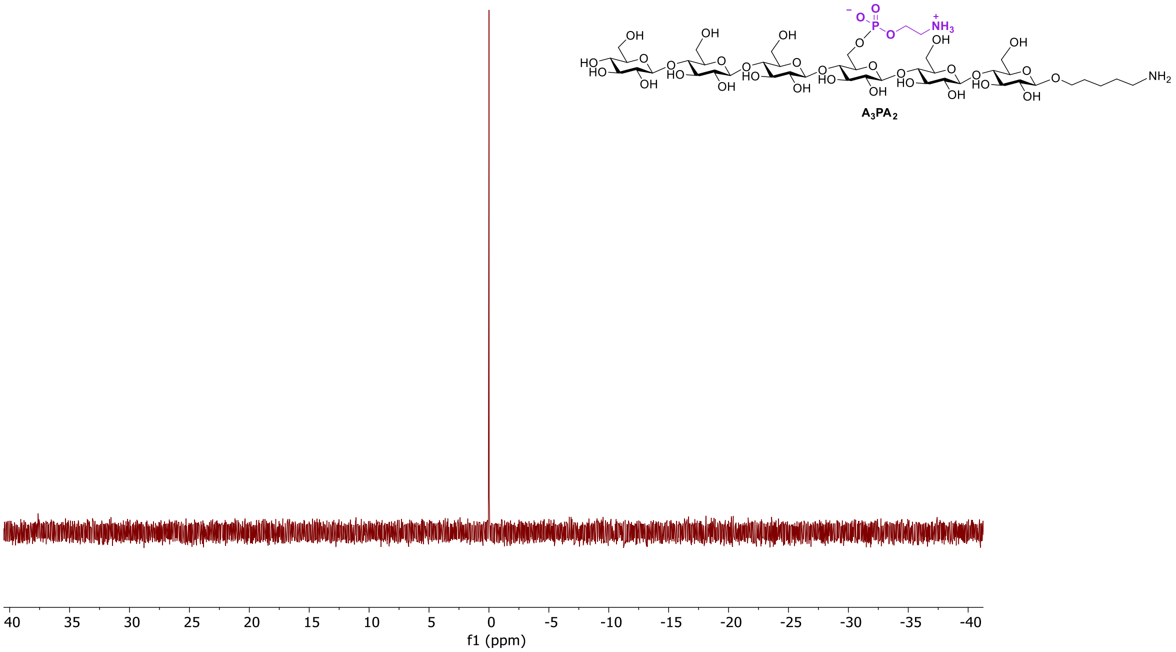
## $^1H$ NMR of $A_3PA_2$ (700 MHz, $D_2O$)



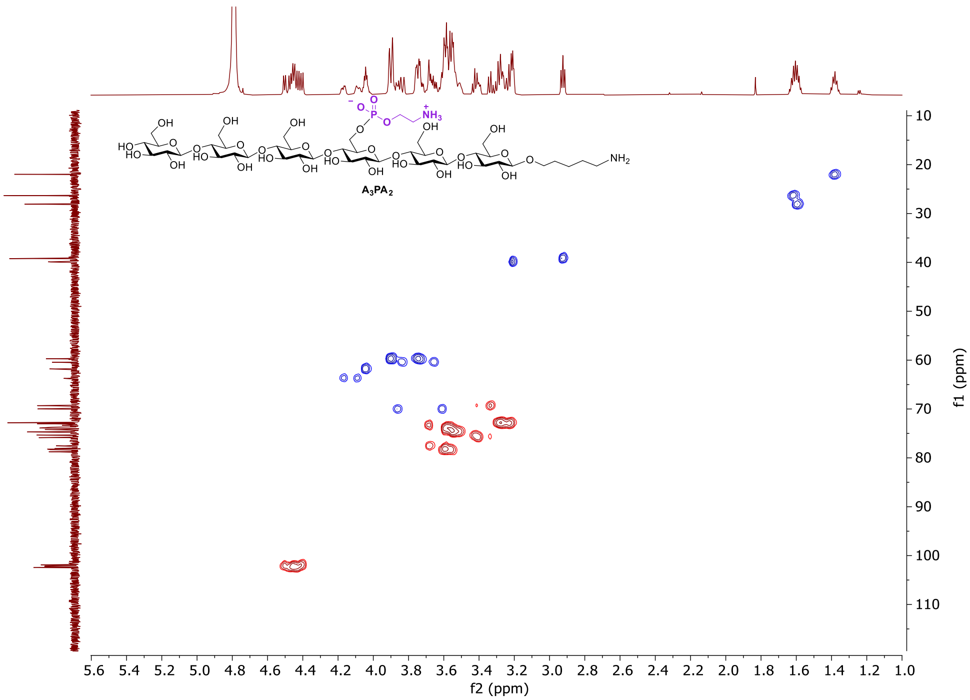
## $^{13}C$ NMR of $A_3PA_2$ (176 MHz, $D_2O$)



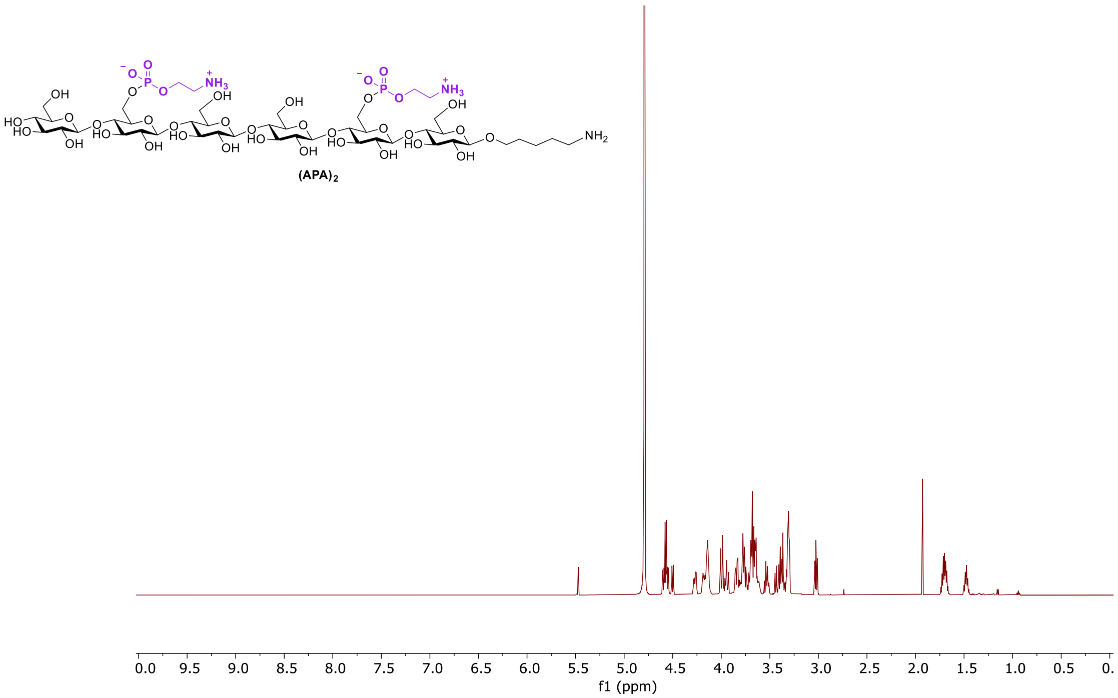
## $^{31}P$ NMR of $A_3PA_2$ (243 MHz, $D_2O$)



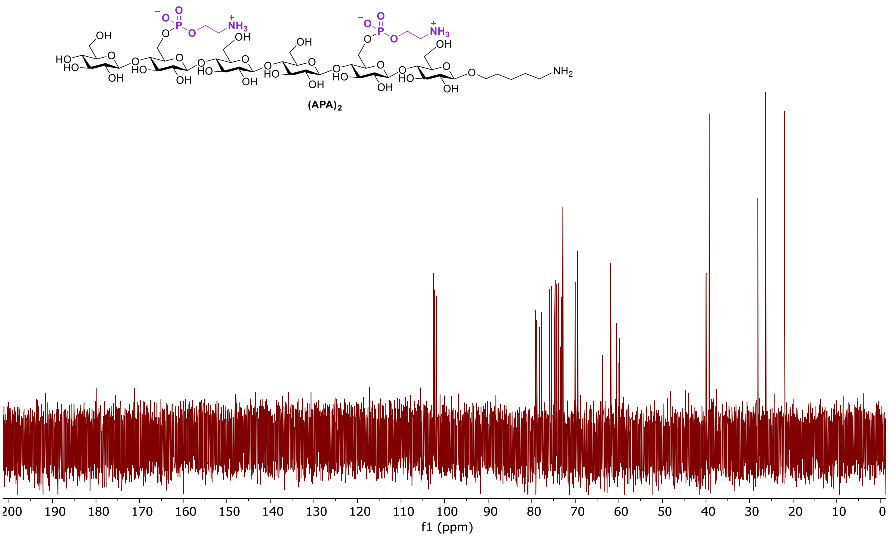
## HSQC NMR of $A_3PA_2$ ($D_2O$)



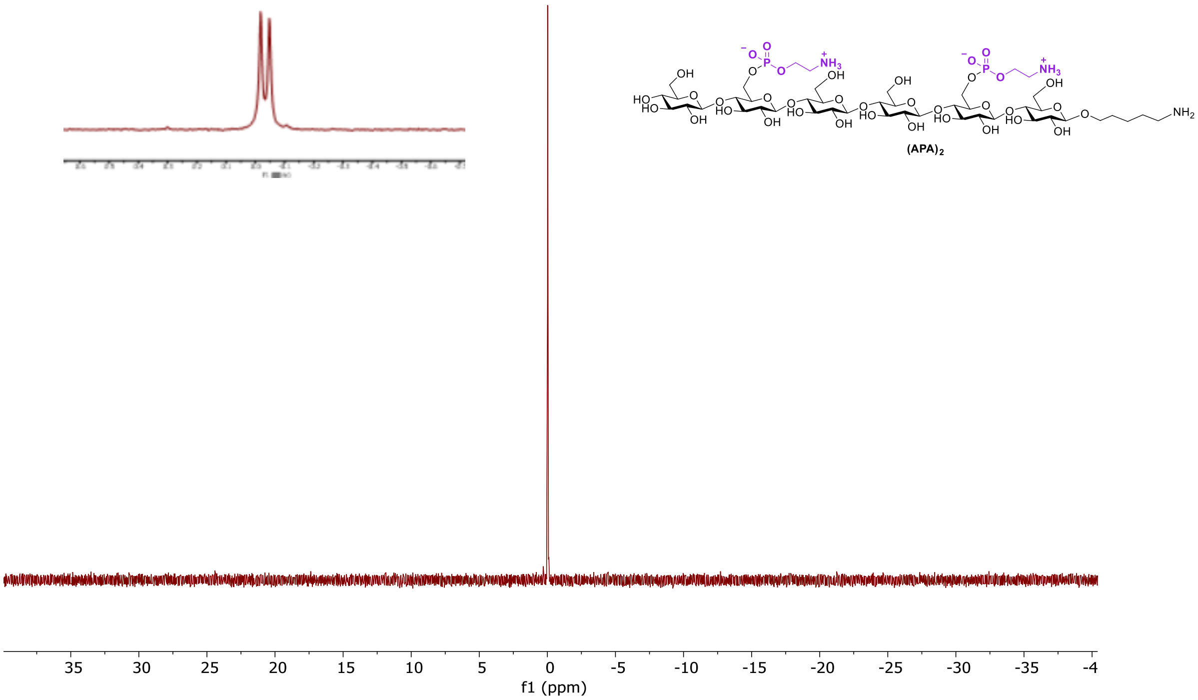
## $^1H$ NMR of $(APA)_2$ (600 MHz, $D_2O$)



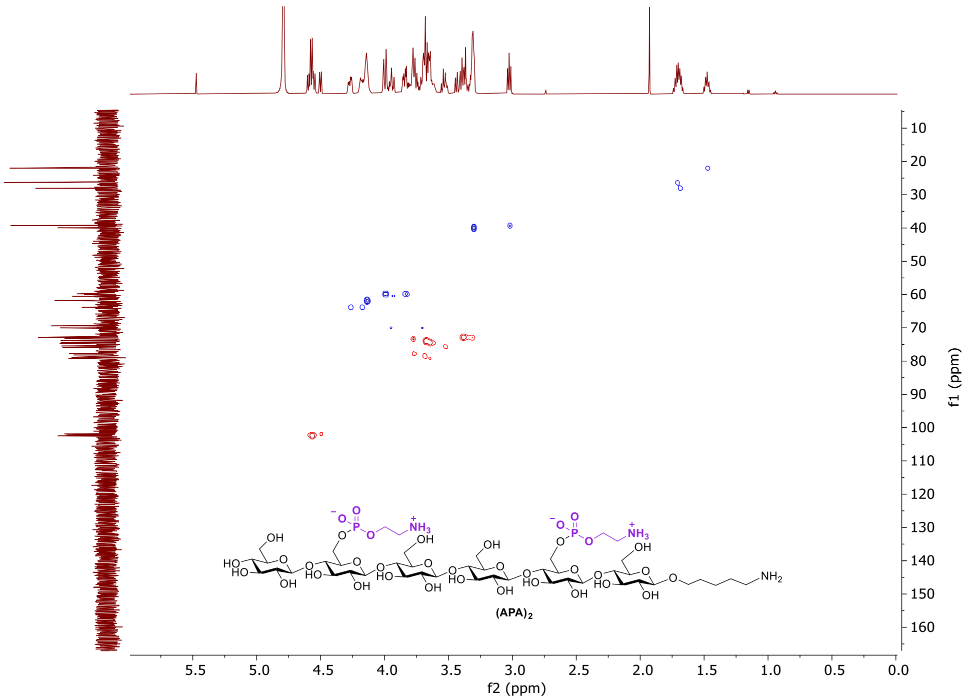
## $^{13}$C NMR of $(APA)_2$ (151 MHz, $D_2O$)



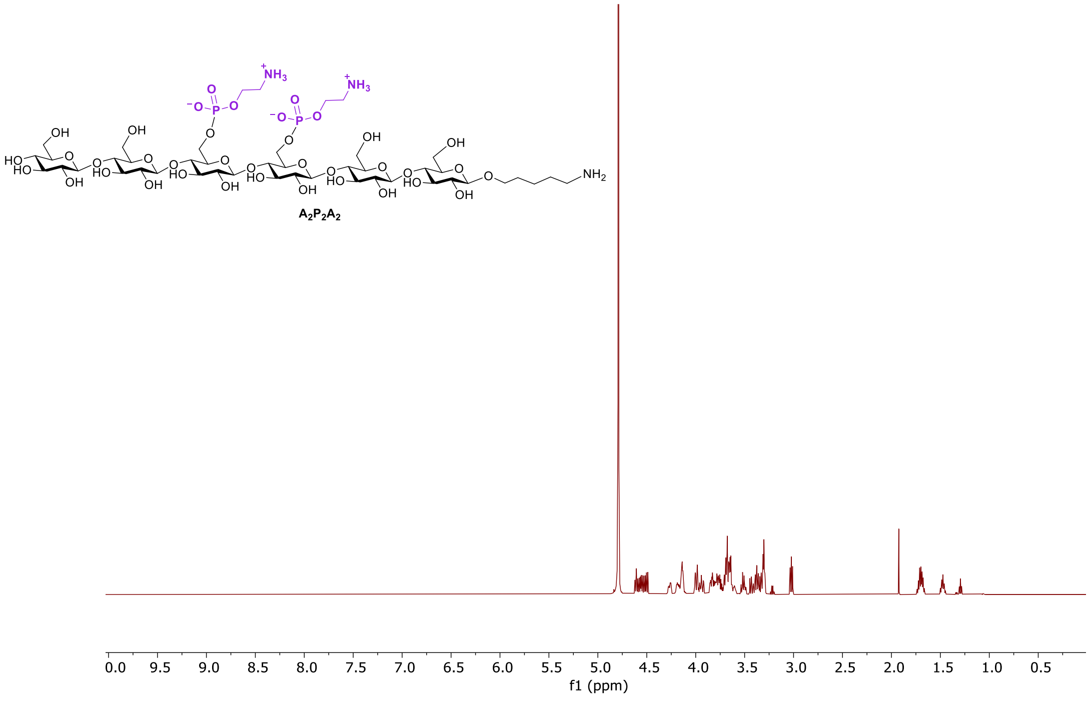
## $^{31}$P NMR of $(APA)_2$ (243 MHz, $D_2O$)



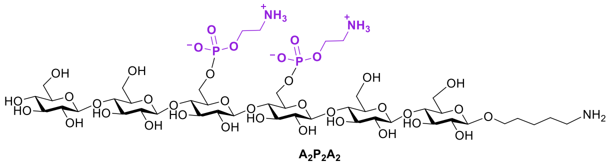
## HSQC NMR of $(APA)_2$ ($D_2O$)



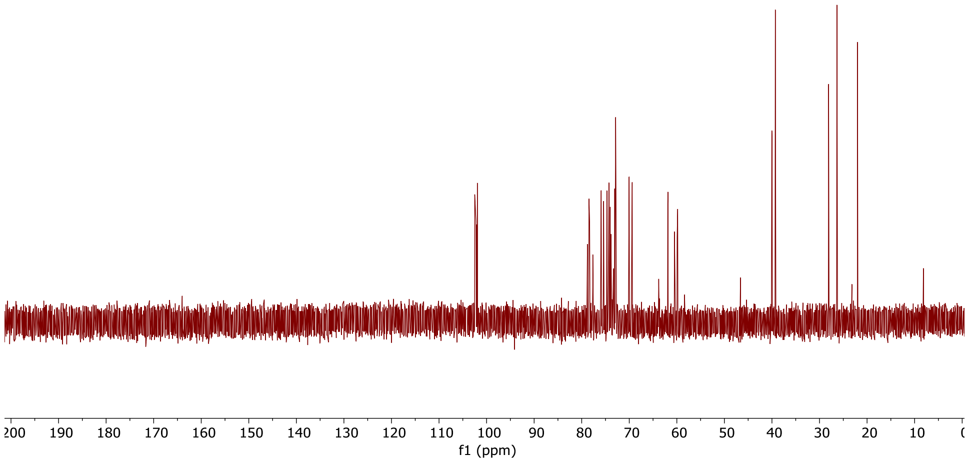
## $^1H$ NMR of $A_2P_2A_2$ (600 MHz, $D_2O$)



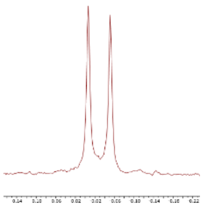
## $^{13}$C NMR of $A_2P_2A_2$ (151 MHz, $D_2O$)

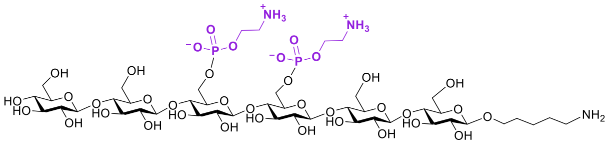


$A_2P_2A_2$

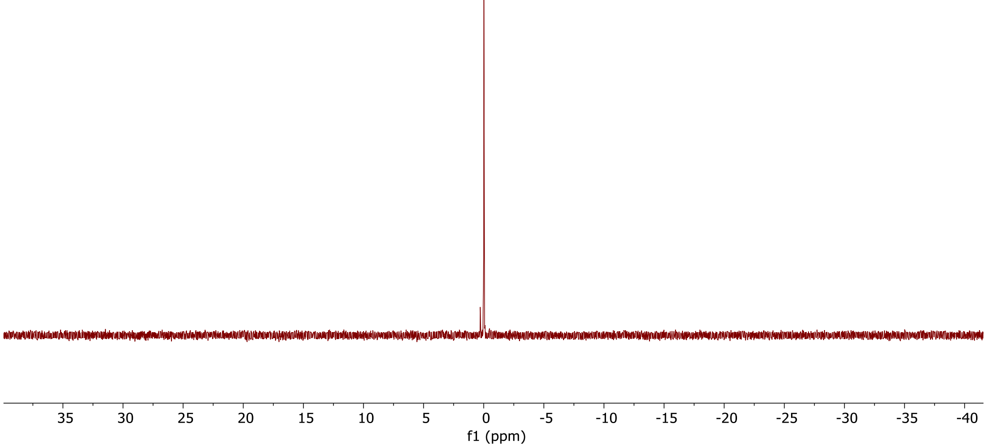


## $^{31}$P NMR of $A_2P_2A_2$ (243 MHz, $D_2O$)

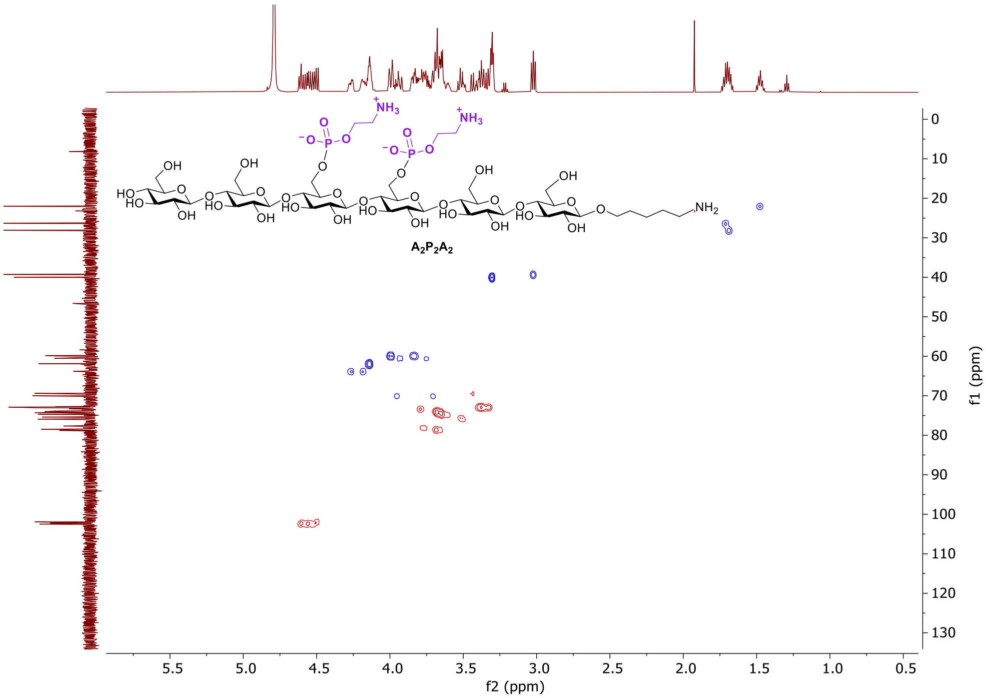




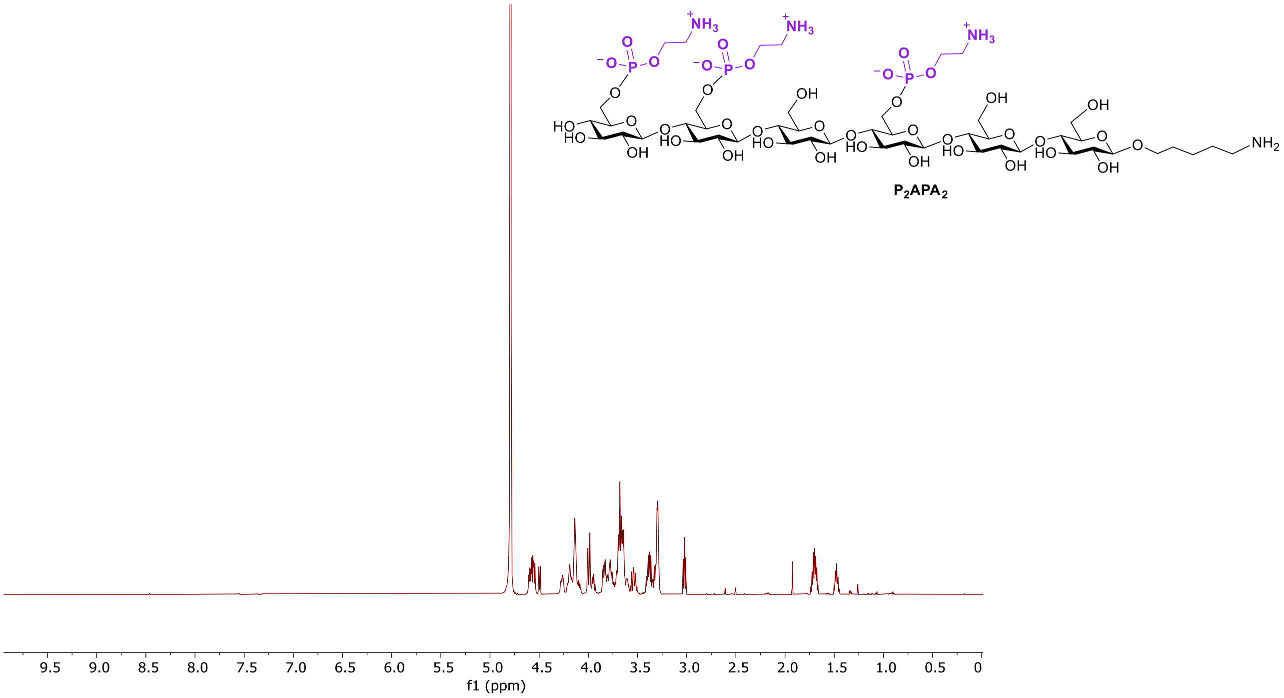
$A_2P_2A_2$



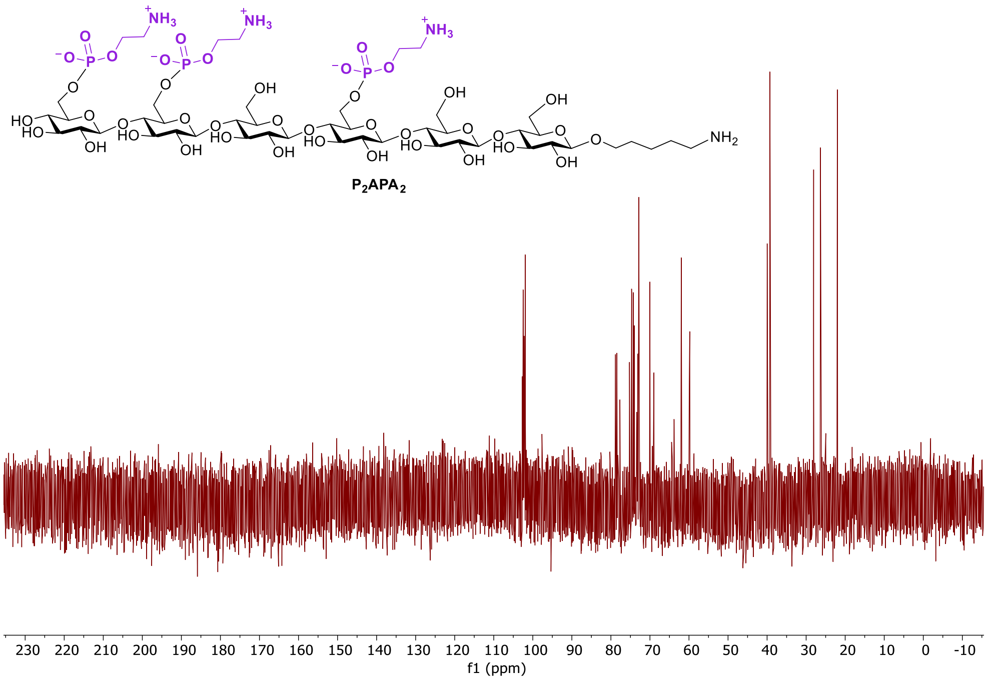
## HSQC NMR of $A_2P_2A_2$ ($D_2O$)



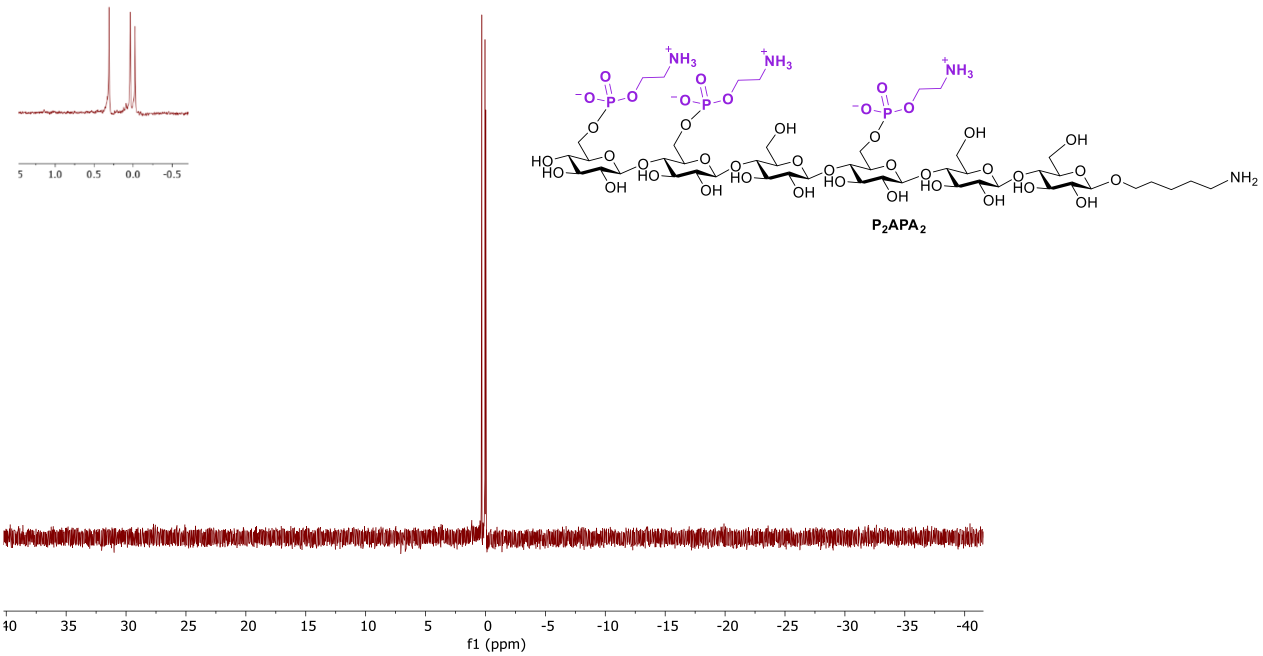
## $^1$H NMR of $P_2APA_2$ (600 MHz, $D_2O$)



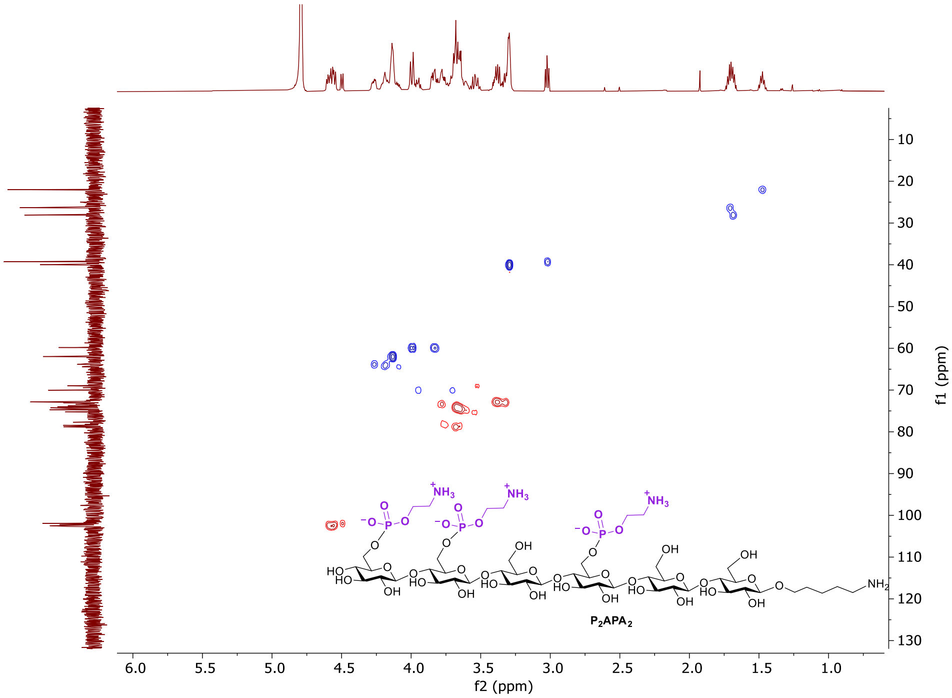
## $^{13}$C NMR of $P_2APA_2$ (151 MHz, $D_2O$)



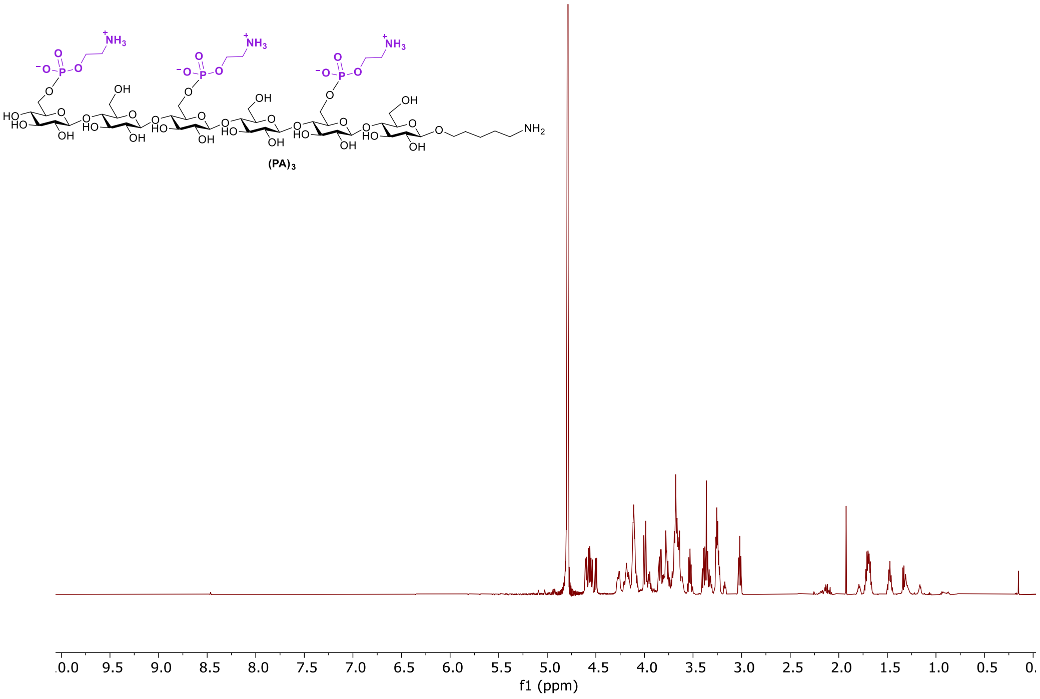
## $^{31}$P NMR of $P_2APA_2$ (243 MHz, $D_2O$)



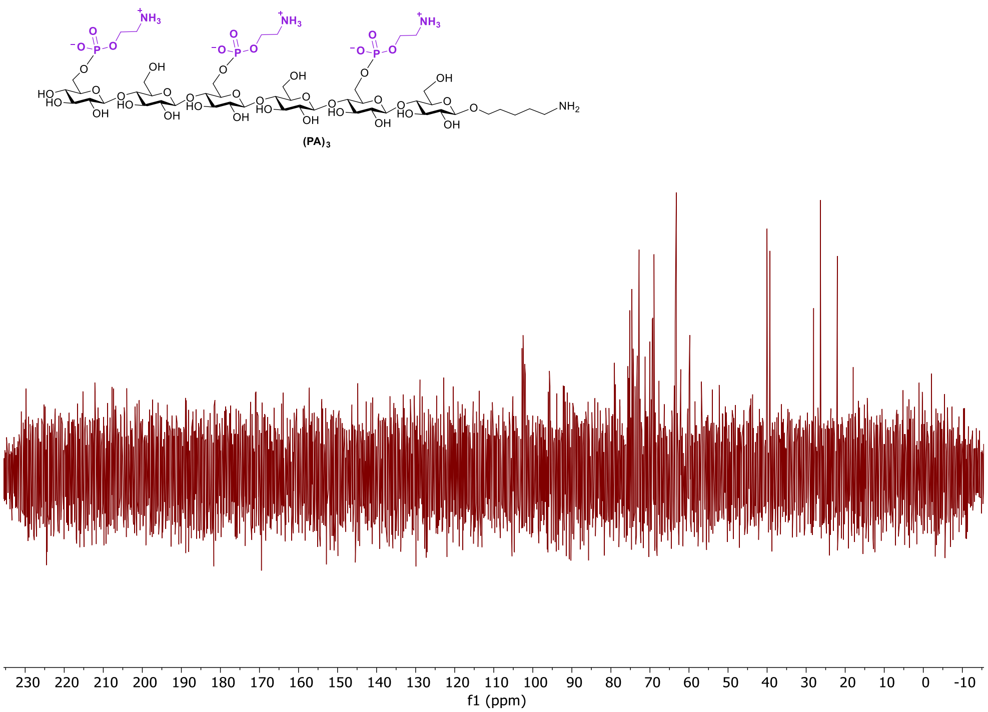
## HSQC NMR of $P_2APA_2$ ($D_2O$)



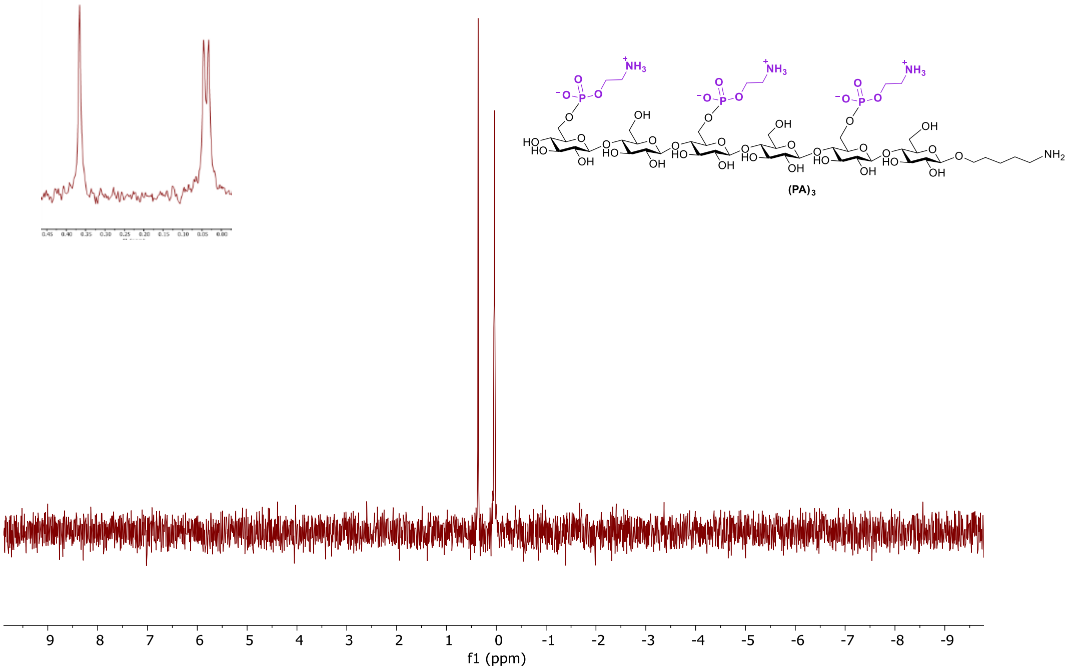
## ¹H NMR of (PA)₃ (600 MHz, $D_2O$)



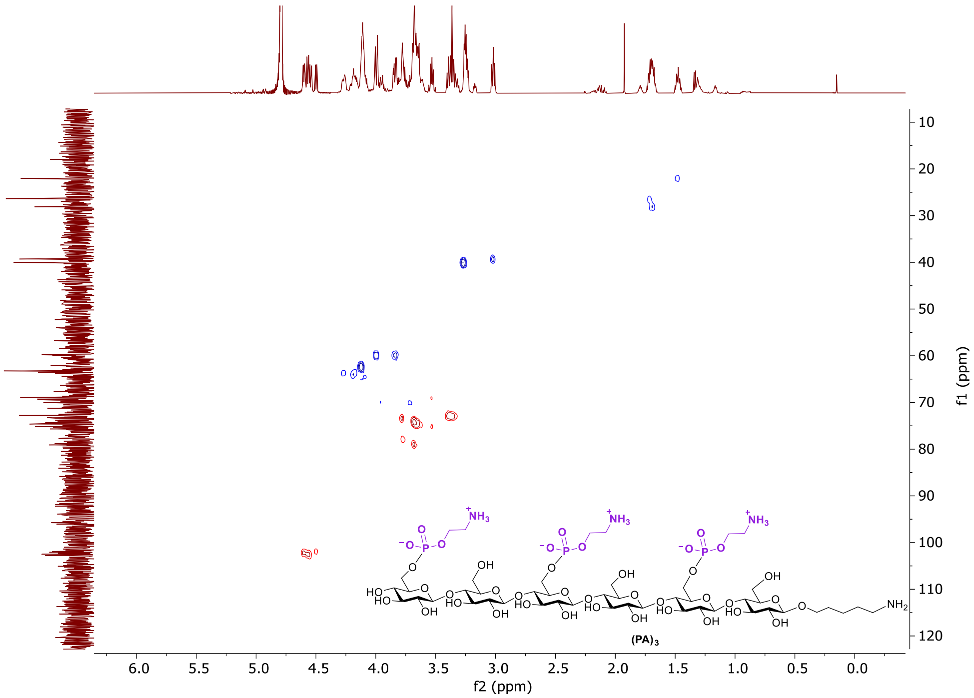
## $^{13}$C NMR of (PA)$_3$ (151 MHz, D$_2$O)



## $^{31}$P NMR of (PA)$_3$ (243 MHz, D$_2$O)



## HSQC NMR of (PA)$_3$ (D$_2$O)


f1 (ppm)
10
20
30
40
50
60
70
80
90
100
110
120
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f2 (ppm)
(PA)3