UNIVERSITÀ DEGLI STUDI DI MILANO

Universiteit Utrecht

UNIVERSITÀ DEGLI STUDI DI MILANO
CORSO DI DOTTORATO IN CHIMICA
XXXIII CICLO

Dipartimento di Chimica

UTRECHT UNIVERSITY
DRUG INNOVATION PhD PROGRAMME

Utrecht Institute for Pharmaceutical Sciences (UIPS)

TESI DI DOTTORATO DI RICERCA

DECOYS FOR DISRUPTING CARBOHYDRATE-LECTIN INTERACTIONS

Settore Scientifico Disciplinare CHIM/06

Dottoranda: Nives HRIBERNIK
Matr. R12136

Tutor: Prof. Anna BERNARDI

Tutor: Prof. Roland J. PIETERS

Coordinatore: Prof. Emanuela LICANDRO

PhD4GlycoDrug
Marie Skłodowska-Curie Innovative Training Network
H2020-MSCA-ITN-2017-EJD-765581

A.A.
2019/2020

Table of contents

CHAPTER ONE

The Sugar Code

Cellular communication in all living organisms involves the coding of information in biomolecules and its subsequent translation into cellular effects. In analogy with the human language, also the molecular language is based on alphabets. The first alphabet of life is written by nucleic acids, the second by proteins and the third by carbohydrates.

The first two alphabets of life (nucleotides and amino acids) are set in direct relation by the genetic code and the message they carry is completely defined by the sequence of the building blocks of each type of oligo- or polymer (nucleic acids and proteins).[1]

The limited space on the cell surface and the multitude of different messages on the other hand require a higher level of density for encoded information, which can be achieved with the compositional and structural variability of glycans.

1.1. Monosaccharides (letters)

The monosaccharides present the letters of the third alphabet of life. Monosaccharides preferentially exist in aqueous solution as cyclic hemiacetals and hemiketals with 5- or 6-membered rings. In the energetically preferred hexopyranose form, each carbon atom of the monosaccharide has two substituents, one in axial and one in equatiorial orientation. The switch in configuration of these two substituents on carbons C_2-C_5 gives a new letter (epimer-e.g. glucose, galactose, mannose, etc.).[2]

The chemically equivalent hydroxyl groups can form the glycosidic bond *via* various linkages and a large set of isomers (disaccharides) can thus be generated, which is in sharp contrast to nucleotides and amino acids, whose chains are built by 5',3'-phosphodiester or by peptide bonds, respectively.[2-4] The difference is summarized in **Fig. 1.1**, showing the difference in variability of possibilities to form linkages when comparing carbohydrates to nucleotides and amino acids.

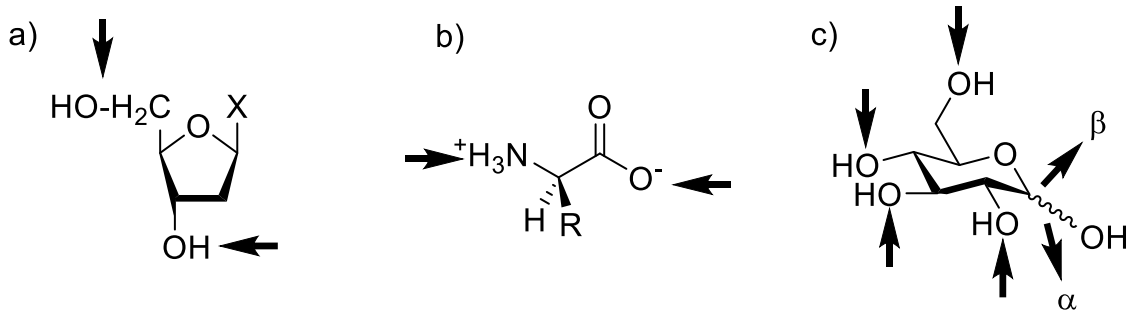


Fig. 1.1 Letters of the three alphabets of life: a) nucleotides, b) amino acids and c) monosaccharides. (arrows present the possible linkage points on each type of biomolecule) - adapted from Ref[5]

As can be read from **Fig. 1.1**, the phosphodiester bond in nucleic acid biosynthesis and peptide bond in protein biosynthesis only give linear oligomers, however various possible linkages through hydroxyl groups in oligosaccharide synthesis can yield linear and branched structures.[5] Linear oligomers form the first dimension of the sugar code and the second dimension comes from branched structures which contributes significantly to the diversity. The third dimension of the sugar code comes from spatial flexibility that is often restricted to few energetically preferred conformers.[2, 6]

Fig. 1.2 summarizes the alphabet of the sugar language, displaying structural representation of the monosaccharides and the known acceptor positions in glycoconjugates which are marked with arrows. Additionally, symbol nomenclature used for monosaccharides in representation of glycans (SNFG)[7-9] is also shown on **Fig. 1.2**. Four sugars have L-configuration: fucose, rhamnose, arabinose and iduronic acid. The others have D-configuration. Neu5Ac is one of more than 50 sialic acids that terminate sugar chains in animal glycoconjugates. Kdo is found in lipopolysaccharides in the cell-walls of Gram-negative bacteria and in cell-wall polysaccharides of green algae and higher plants. Microbial polysaccharides also contain furanose form of D-galactose and D/L-arabinose, which do not occur in mammalian glycochemistry.[2, 3, 5]

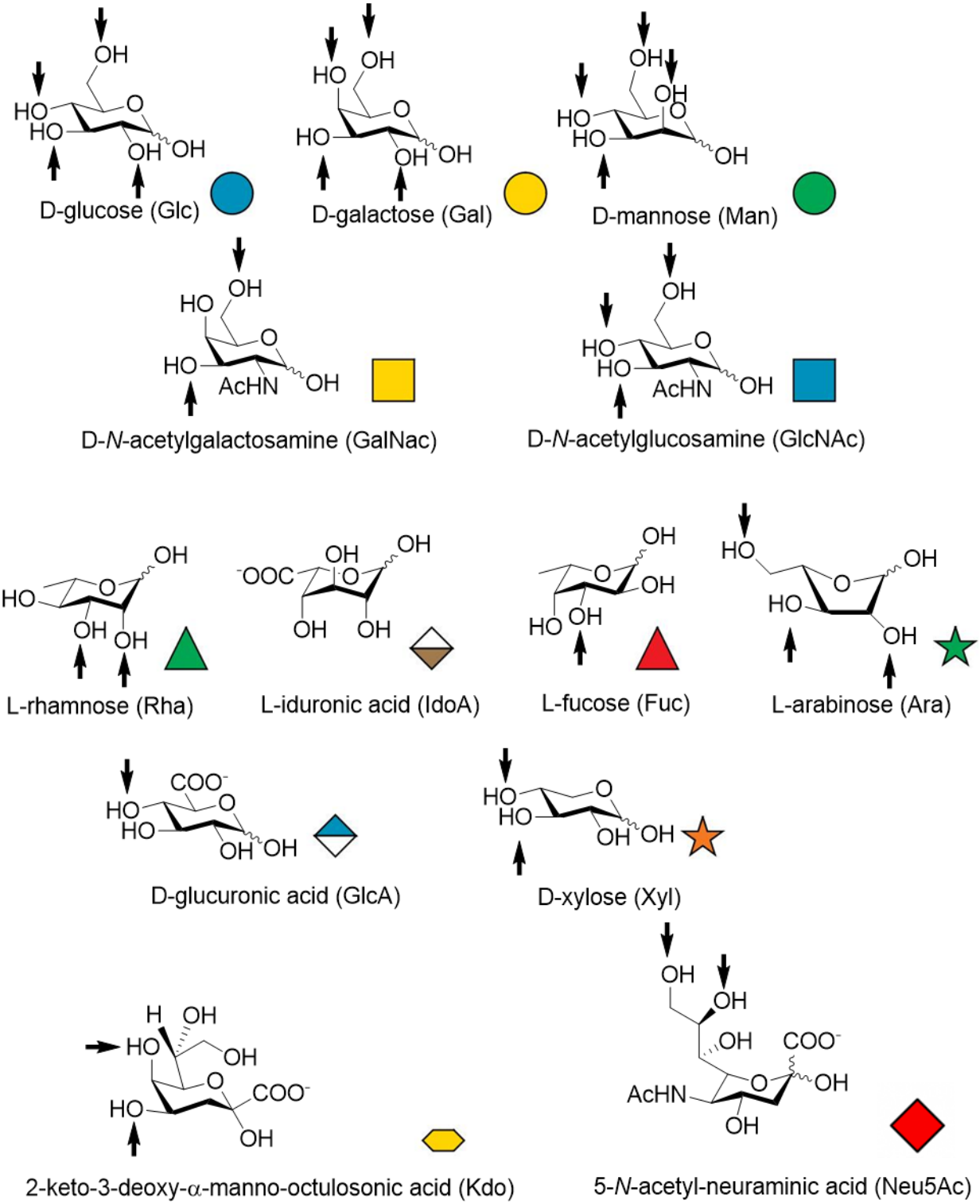


Fig. 1.2 The alphabet of the sugar language, structures of monosaccharides (arrows pointing to known linkage points), symbol nomenclature for monosaccharides[8]-adapted from Ref[2]

1.2. Glycans (words)

Glycan structures can be categorized in different groups based on similarities in their structures. A brief overview of common glycan structures is described below.

N-glycans

N-glycans are covalently attached to protein at asparagine (Asn) residues belonging to Asn-X-Ser sequences by an *N*-glycosidic bond. In prokaryotes different sugars can be attached to Asn, while all eukaryotic *N*-glycans share a common core sequence Manα(1-3)((Manα(1-6))Manβ(1-4)GlcNAcβ(1–4)GlcNAcβ1–Asn-X-Ser/Thr. Based on the extension of the core, *N*-glycans can then be divided in 3 types as shown in **Fig. 1.3**: oligomannose-type where only Man residues extend the core, complex-type where the core extension starts with GlcNAc and is further extended by "antennae" and hybrid-type where Man extends the Manα(1-6) arm of the core and one or two GlcNAc extend the Manα(1-3) arm of the core.[10]

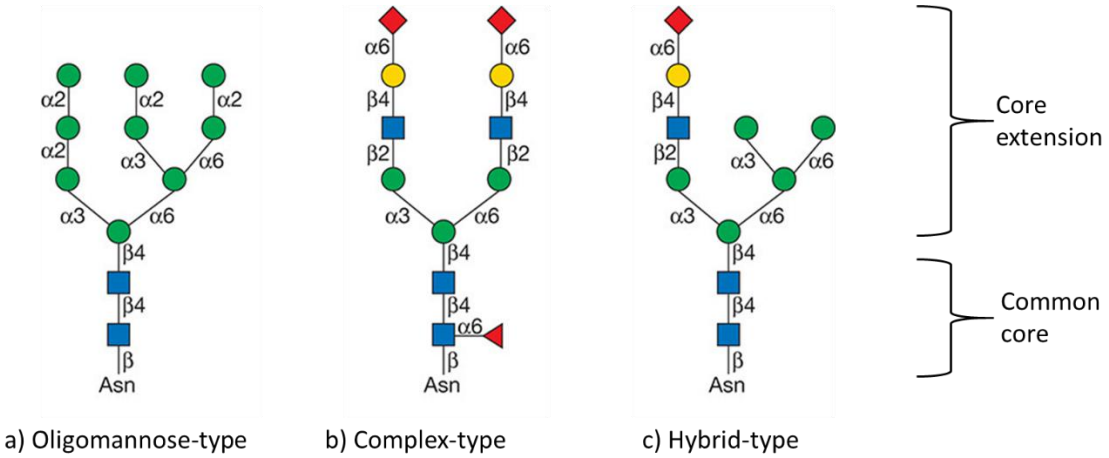


Fig. 1.3 N-linked eukaryotic glycans: a) oligomannose type, b) complex type, c) hybrid type -adapted from Ref[10]

O-GalNAc glycans

These glycan structures are initiated by GalNAc attached to the Ser or Thr (**Fig 1.4**). Mucin glycoproteins (also known as mucin type *O*-glycans) carry the greatest number of *O*-GalNAc glycans. Sugars that are found in *O*-GalNAc glycans are GalNAc, Gal, GlcNAc, Fuc and sialic acids, but no Man, Glc and Xyl can be found. Terminal sialic acids can be modified by *O*-acetylation and Gal and GlcNAc residues by sulfation.[11]

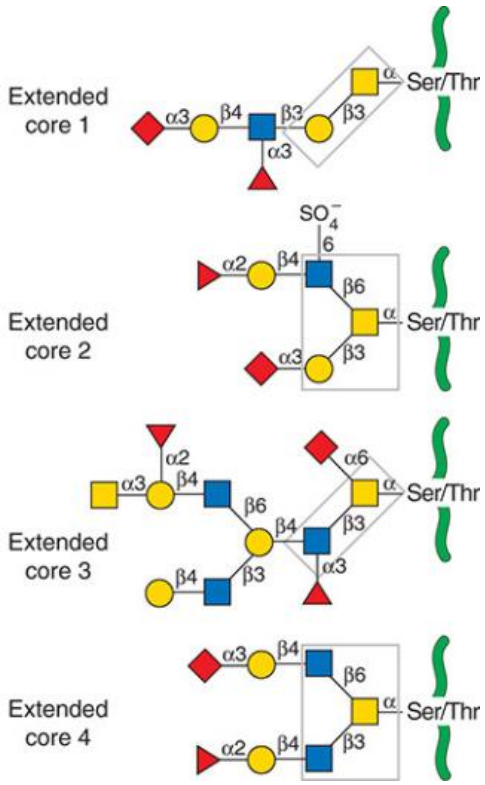


Fig 1.4 Examples of complex *O*-GalNAc glycans with different cores found on mucin glycoproteins. Cores are marked with a grey box, proteins are marked with green lines-adapted from Ref[11]

Glycolipids

Most animal glycolipids are glycosphingolipids that are built on a ceramide lipid moiety, consisting of a long chain amino alcohol (sphingosine) in amide linkage to the fatty acids. Another group of glycolipids are glycoglycerolipids that are built on a diacyl or acylalkylglycerol lipid moiety (**Fig 1.5**).[12]

a) Glycosphingolipids b) Glycoglycerolipids

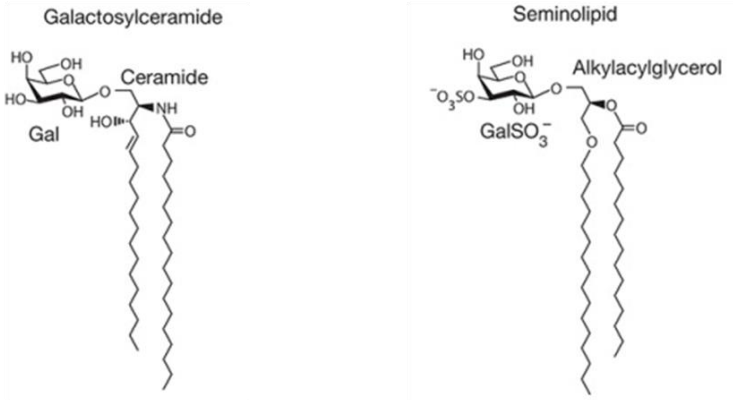


Fig. 1.5 Representatives from the class of: a) glycosphingolipids, b) glyceroglycerolipids- adapted from Ref[12]

Glycosylphosphatidylinositol anchors

Glycosylphosphatidylinositol (GPI) anchors share a common core consisting of ethanolamine-PO_4-6Manα(1-2)Manα(1-6)Manα(1-4)GlcNα(1-6)*myo*-inositol-1-PO_4-lipid. Heterogeneity is then derived from various substitutions on this core structure (**Fig. 1.6**).[13]

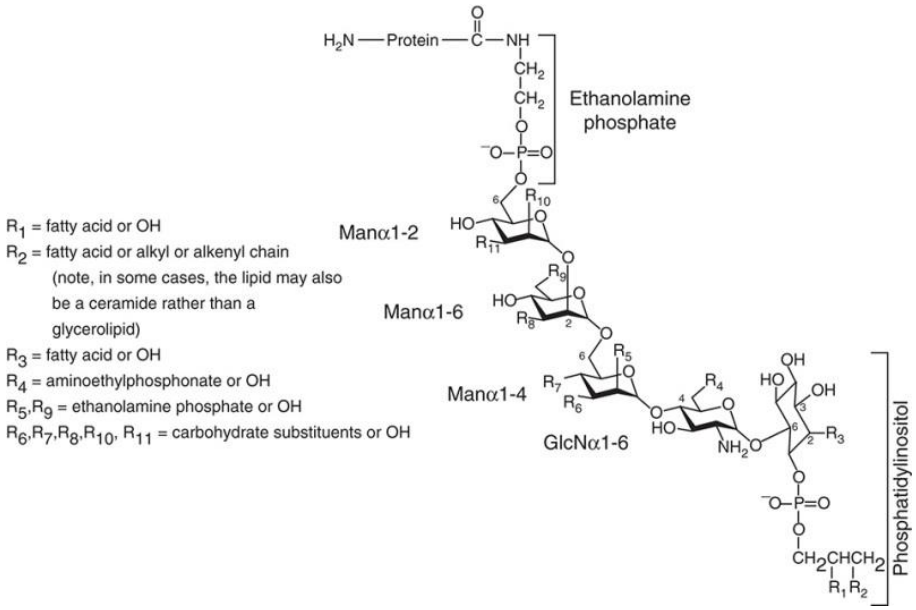


Fig 1.6 General structure of GPI anchors - adapted from Ref[13]

Besides these major classes of glycans, other eukaryotic glycans include for example *O*-linked modifications on epidermal growth factor, *O*-fucosylation of thrombospondin type 1 repeats and *O*-mannosylation of α-dystroglycan.[14]

Proteins are involved in all apects of the sugar code, from writing the coded signals, reading and finally erasing them. Defects in glycosylation can lead to pathological states.[15, 16]

1.3. Proteins involved in sugar code formation and translation

Writers

A set of proteins is necessary to write the sugar code, these include enzymes involved in preparing and using the activated substrates, enzymes that place position-specific substitutions and transport proteins that bring the substrates to the site of glycan assembly.[2]

The letters are supplied by food or synthesized and then activated by reaction with appropriate nucleotide triphosphate and the product. The activated products are then delivered to the site of glycan assembly with the help of transporters. After the first step of protein glycosylation, each subsequent step of glycan chain elongation is carried out by specific glycosyltransferase.

Glycosyltransferases move the carbohydrate from its donor (conjugate with the nucleotide, such as CMP, GDP or UDP) to the acceptor (protein or sphingolipid) to initiate the growth of glycan chains.[17, 18, 19] The synthesis can go through various pathways, allowing for differently branched structures to grow. Nature developed elaborated machinery for glycan biosynthesis, consisting of around 500-700 proteins that can generate more than 7000 different mammalian glycans.[20-22] A major site for glycosylation is Golgi apparatus.[23] The final output depends on the presence and localization of enzymes along the glycan assembly line, availability of sugars and transporters of activated sugars. Consequently, the glycan profile is not directly genetically encoded and transcription can be modulated in response to external factors (metabolic states, etc.).[2, 24]

Erasers

The function of erasers is to make room for new signals when the mission of the outer sugar units presented on the stable core has been accomplished. After letters are removed from the glycan word, the original meaning is abolished and a new meaning may be uncovered. The enzymes responsible for removing the monosaccharides are glycosidases. Additionally, glycans can be modified by other enzyme types to achieve the addition of acetyl, methyl, phosphate, sulfate and other groups.[19]

Readers

Reading of the encoded information begins upon association of glycans with their receptors, called lectins. The word lectin comes from Latin “lego” which means to read and thus indicates that these proteins have a crucial role in deciphering the glycocode. Lectins recognize sugars but do not catalyse their transformation. Criteria for protein to be a lectin include: carbohydrate-binding activity, distinction from immunoglobulins and distinction from enzymes tailoring free saccharides.[25, 26]

Lectins are ubiquitous and their biological functions are very diverse, from host recognition and tissue adhesion of microorganisms to glycoprotein trafficking and clearance, immune defence, development and fertility in humans, and many others.[27-29]

The carbohydrate binding sites of lectins present mainly two topologies with either a cleft (groove) or a pocket (depression) and they are generally shallow. Despite their flexibility, only one conformation of oligosaccharides can usually fit in the binding site.[28]

The interactions between carbohydrates and lectins primarily involve a network of hydrogen bonds and hydrophobic interactions. The energy associated with H-bonding is significantly reduced by competition from bulk solvent and by the flexible nature of the hydroxyl groups, causing significant entropic penalty when they become constrained upon binding.[30] The stacking interactions with aromatic side chains can occur with sugar CH bonds. Some lectins may require coordination of sugar

molecules with metal ions (C-type lectins and related Ca-dependent proteins). Electrostatic interactions are rare and limited to specific monosaccharides such as sialic acid. Side chains and main chain groups can participate in the ligand binding, binding can also be mediated by water molecules.[28, 31]

Shallow pockets of carbohydrate binding sites of lectins result in generally low affinity interactions, with dissociation constants in the mM and µM range. This is however balanced by multivalency that provides high avidity when glycans are presented in multiple copies, e.g. on cell surface.[28]

1.4. Interfering with the sugar code

Due to the importance of information that is encoded in the language of sugars, understanding how the information is read out by lectins and understanding how we can control or alter the flow of information is crucial. Different approaches are used, taking advantage of carbohydrate chemistry, supramolecular chemistry and rational protein engineering deliver test compounds that can be glycans, glycomimetics or lectin-based.

1.4.1. Glycomimetics

Glycomimetics are non-carbohydrate molecules that attempt to reproduce the 3D structure of oligosaccharide binding determinants and compete with the natural ligand for the target lectin. Often they consist of a mono- or disaccharide linked to aglycones. The saccharide moiety works as a lectin anchor, while the aglycone contributes to additional interactions with the lectin and some pharmacologically favourable properties, such as improved lipophilicity and enzymatic stability. [30]

Rational design of glycomimetics is based on the knowledge that only small portions of the oligosaccharides are actually recognized by their receptors, while the rest orients the binding determinants in the appropriate conformation.[30] Second, although oligosaccharides are quite flexible, glycans have their highly favoured conformations.[32, 33] Additionally, different lectins can select different conformations of flexible oligosaccharides, even the conformations that are not populated more than 5-15% in the free state of the ligand, which results in low affinity.[34]

Construction of carbohydrate mimics can therefore follow various approaches.[30, 35-38] Non-interacting functional groups can be removed, while the original glycosidic linkage stays intact. This reduces polarity of sugars, increases affinity by improving hydrophobic interactions and decreases the penalty for polar group desolvation. Hydrophobic or charged groups can be added and enzymatic stability can be improved with replacement of either endo- or exocyclic oxygen atom with another atom. The

chemical and enzymatic stability, polarity, charge, conformation, ring flexibility and hydrogen-bonding pattern of the molecule can change due to these modifications, influencing also the affinity of these glycomimetics towards their target proteins.

Endocyclic oxygen can be replaced by a variety of different heteroatoms as shown in **Fig 1.7**. Among best known examples are iminosugars that have an endocyclic oxygen replaced with a nitrogen and are promising candidates for treatment of various diseases, including cancer, diabetes, viral infections and lysosomal storage disorders, such as Gaucher and Fabry diseases.[33, 34] *Thio*-sugars on the other hand are sugar mimics with endocyclic oxygen replaced by a sulphur atom and can be used for various applications, such as antiviral, antidiabetic and anticancer compounds.[39-41] Endocyclic oxygen can also be replaced by a methylene group.[42, 43] Such modification abolishes the anomeric effects, modifies intramolecular hydrogen bond pattern, changes the flexibility and improves metabolic stability. Endocyclic modification can also be a P=O group, providing a class of oxaphosphinanes[44-46] that is further devided into phospha-sugars, phosphono-sugars and phosphino-sugars. These phosphorus heterocycles are considerably more resistant to the ring-opening/closure processes.

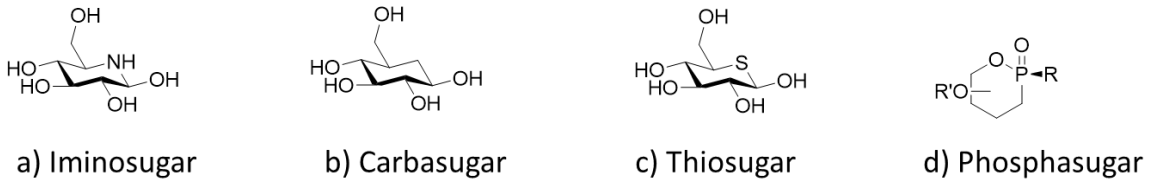


Fig. 1.7 Examples of glycomimetics obtained through endocyclic oxygen replacement

Replacement of exocyclic oxygen atom is also common in design of glycomimetics, examples of this strategy include *C*-glycosides[47], *N*-glycosides[48, 49], *Se*-glycosides[50, 51] and *S*-glycosides[52-54] shown in **Fig. 1.8**. These sugar mimics are hydrolytically and enzymatically stable while preserving biological activity. The conformational behaviour of these mimics can be quite different from their natural counterparts, for example greater flexibility around the interglycosidic linkage or presence of conformations that are not available to the native carbohydrate, which can be well exploited. An example are *S*-glycosides that are conformationally similar to natural *O*-glycosides, however C-S bond is longer than C-O bond by 0.4 Å resulting in enhanced conformational flexibility. In certain cases, therefore thioglycosides can adapt their conformation to better fit in the binding site and result in more potent biological activity. Reduced basicity of sulphur compared to oxygen also provides compounds that are more resistant towards enzymatic and acid-catalyzed hydrolysis.[55]

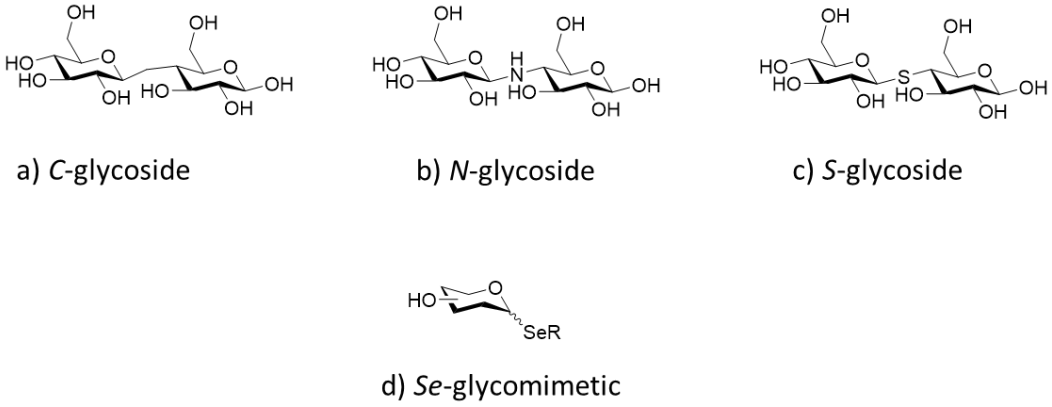


Fig. 1.8 Examples of glycomimetics obtained through exocyclic oxygen replacement

A possibility for sugar modifications that has been explored is also fluorination (**Fig. 1.9**).[56-58] Fluorine atom is the smallest substituent that can be used as replacement on the C-H bond and it gives a strongly polarized C-F bond, which presents a strong dipole moment. Fluorination influences lipophilicity and consequently solubility, permeability and protein binding. Fluorosugars can be used in development of enzyme inhibitors, an interesting application is also a strategic fluorination of antigenic glycans, which gives glycoconjugates with improved metabolic stability and comparable or enhanced immunogenicity, attractive in vaccine development.

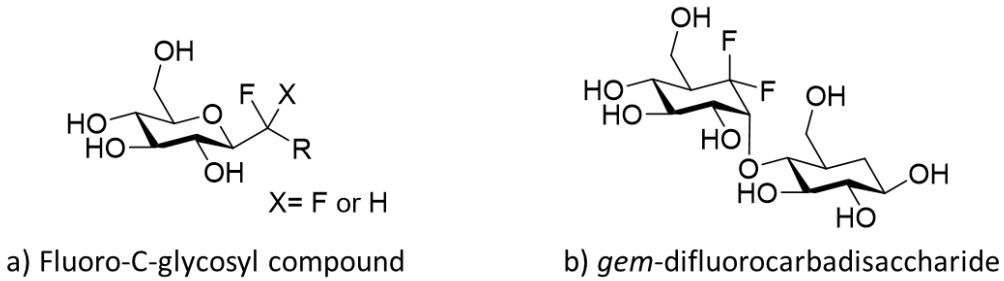


Fig. 1.9 Examples of fluorinated glycomimetics

Alternatively, non-carbohydrate framework can be used, to which required pharmacophoric sugar fragments are tethered so that they maintain same orientation in space as in the bound conformation. Additional groups can be presented to further improve affinity. Replacement of the glycosidic scaffold can facilitate and simplify the otherwise time-consuming complex oligosaccharide synthesis, while improving metabolic stability.[30, 59]

In particular, carbocyclic diols of the appropriate configuration can substitute branching units in complex oligosaccharides. A successful example of such approach is the synthesis of a highly effective mimic of sialyl Lewis^X tetrasaccharide in which the 3,4-disubstituted *N*-acetylglucosamine was replaced by a 1,2-*trans*-cyclohexanediol (**Fig. 1.11, a**).[59] While simple 1,2-cyclohexanediols are

conformationally flexible and lack potential to be conjugated to polyvalent scaffolds, these drawbacks could be overcome with the use of 1,2-*trans*-dicarbohy-4,5-cyclohexanediols. The 1,2-*trans*-dicarboxy substitution acts as a conformational lock on the cyclohexane ring, thus assuring conformational stability of these compounds.[60]

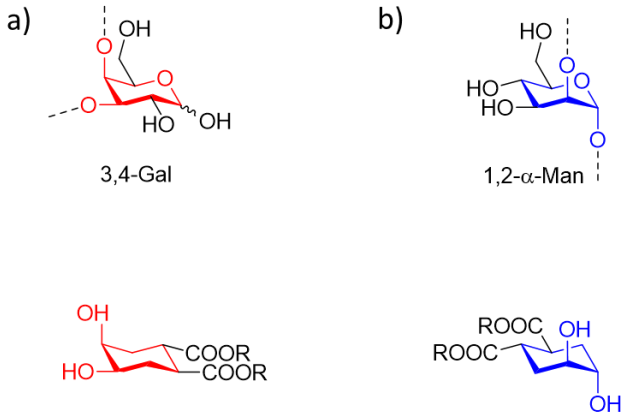
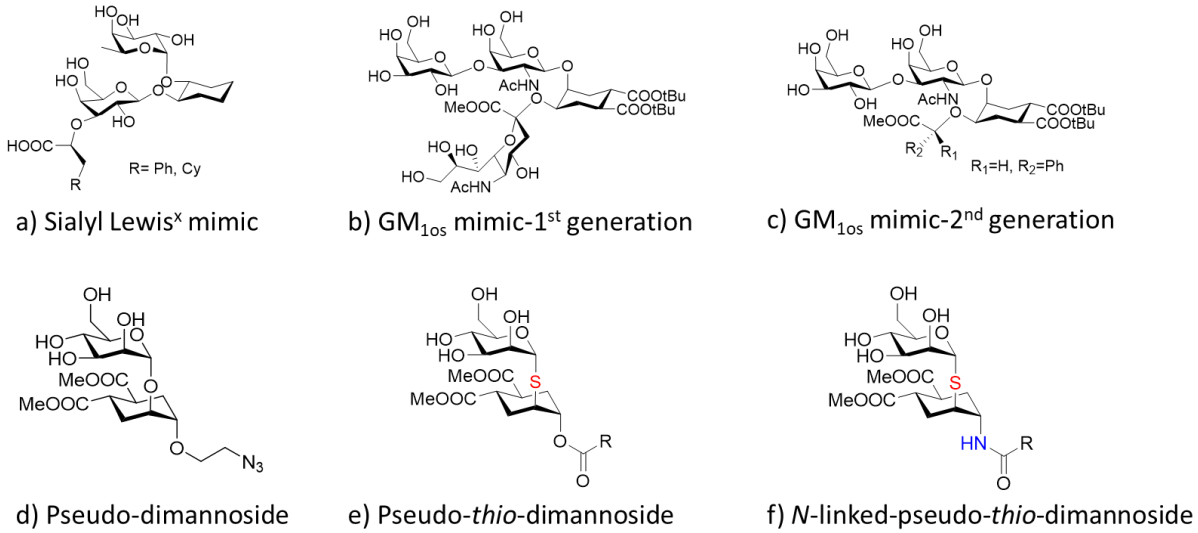


Fig. 1.10 Dicarboxy cyclohexanediol schaffolds designed to mimic a) 3,4-disubstituited galacto and b) 1,2-α-manno motifs

Such scaffolds were successfully used for mimicking the cis diol of the 3,4-disubstituited galactose (**Fig. 1.10, a**) residue in the GM_1 oligosaccharide, accurately replicating its 3D structure as well as biological activity towards cholera toxin (**Fig 1.11 b, c**).[61, 62] Similarly, 1,2-α-manno motif can be replicated (**Fig 1.10, b**), for example in design of DC-SIGN ligands (**Fig. 1.11, d, e, f**), which is discussed to greater details in Chapter Four.[63-65]



a) Sialyl Lewis^x mimic b) GM_{1os} mimic-1st generation c) GM_{1os} mimic-2nd generation

d) Pseudo-dimannoside e) Pseudo-*thio*-dimannoside f) *N*-linked-pseudo-*thio*-dimannoside

Fig. 1.11 Examples of glycomimetics with carbocyclic diol scaffolds: a) Sialyl Lewis^x mimic[59], b) GM_{1OS} mimic-1st generation[61], c) $GM_{1O}S$ mimic-2nd generation[66], d) Pseudo-dimannoside[67], e) Pseudo-*thio*-dimannoside[64], f) *N*-linked-pseudo-*thio*-dimannoside[68]

1.4.2. Lectin analogs

Synthetic chemistry yielded artificial carbohydrate receptors that could potentially be used as anti-infective agents. Such synthetic molecules consist of aromatic moieties forming a hydrophobic roof and floor and pillar-type polar groups that can establish H-bonds to the substrate, thus mimicking the natural H-bonds and hydrophobic interactions between lectins and carbohydrates.

The first example of such construct had a meta-terphenyl structure, five rigid isophthalamides were incorporated and linked by two amide linkages, potentially forming H-bonds with OH groups of the sugar (**Fig. 1.12**). An externally directed tricarboxylate unit on each isophthalamide moiety promotes water solubility and prevent aggregation. This molecule binds D-cellobiose with K_a of 580 M^{-1} and methyl β-D-cellobiose with K_a of 910 M^{-1}.[69]

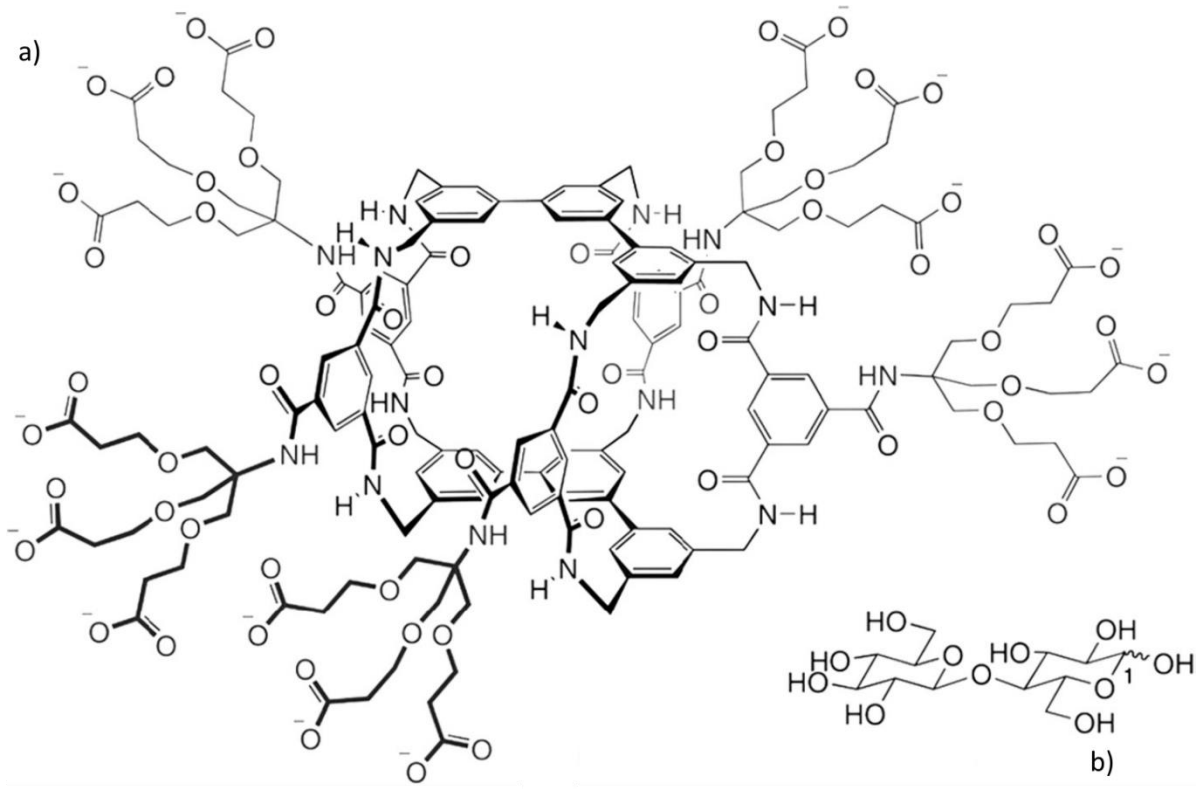


Fig. 1.12 Structure of a) artificial carbohydrate receptor with meta-terphenyl structure and b) cellobiose from Davis et al.[69]

Recently Francesconi et al. reported a simple biomimetic receptor able to recognize the methyl-β-glycoside of $GlcNAc_2$ in water with the affinity of 160 μM.[70] The tweezers shaped acyclic structure (**Fig. 1.13**) of this artificial receptor shows remarkable selectivity, being able to completely discriminate between mono-and disaccharides.

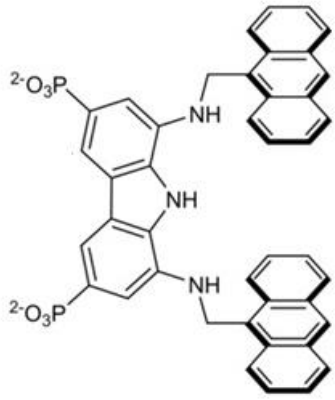


Fig. 1.13 Structure of artificial receptor with a tweezers shaped acyclic structure developed by Francesconi et al.[70]

On the other hand, protein engineering can be used to produce rationally designed peptide-based lectin constructs with improved affinity and glycan selectivity. This approach is further explained in Chapter Six, presenting an example of how lectin constructs can be used to combat viral diseases.

1.5. References

1. M. Barbieri, *The Codes of Life: The Rules of Macroevolution*, Springer Netherlands, 2007.
2. H.-J. Gabius, *Biosystems*, 2018, **164**, 102-111.
3. H. Kaltner, J. Abad-Rodríguez, A. P. Corfield, J. Kopitz and H.-J. Gabius, *Biochemical Journal*, 2019, **476**, 2623-2655.
4. H.-J. Gabius and J. Roth, *Histochemistry and Cell Biology*, 2017, **147**, 111-117.
5. H. Rüdiger and H. Gabius, *The sugar code. Fundamentals of glycosciences*, 2009, **3**, 13.
6. C. A. Bush, M. Martin-Pastor and A. Imbery, *Annual Review of Biophysics and Biomolecular Structure*, 1999, **28**, 269-293.
7. A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J. J. Prestegard, R. L. Schnaar, H. H. Freeze, J. D. Marth, C. R. Bertozzi, M. E. Etzler, M. Frank, J. F. Vliegenthart, T. Lütteke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K. F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi and S. Kornfeld, *Glycobiology*, 2015, **25**, 1323-1324.
8. K. Cheng, Y. Zhou and S. Neelamegham, *Glycobiology*, 2017, **27**, 200-205.
9. S. Neelamegham, K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, T. Lütteke, N. O'Boyle, N. H. Packer, P. Stanley, P. Toukach, A. Varki, R. J. Woods and T. S. D. Group, *Glycobiology*, 2019, **29**, 620-624.
10. P. Stanley, N. Taniguchi and M. Aebi, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015, pp. 99-111.
11. I. Brockhausen, H. Schachter and P. Stanley, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2009.
12. R. L. Schnaar, A. Suzuki and P. Stanley, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2009.
13. M. A. J. Ferguson, T. Kinoshita and G. W. Hart, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2009.
14. R. S. Haltiwanger, L. Wells, H. H. Freeze and P. Stanley, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N.

H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015, pp. 151-160.

15. S. Grünewald, G. Matthijs and J. Jaeken, *Pediatric Research*, 2002, **52**, 618-624.
16. C. Reily, T. J. Stewart, M. B. Renfrow and J. Novak, *Nature Reviews Nephrology*, 2019, **15**, 346-366.
17. T. Hennet, *Cellular and Molecular Life Sciences*, 2002, **59**, 1081-1095.
18. A. Harduin-Lepers, R. Mollicone, P. Delannoy and R. Oriol, *Glycobiology*, 2005, **15**, 805-817.
19. J. Rini, J. Esko and A. Varki, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2009.
20. K. W. Moremen, M. Tiemeyer and A. V. Nairn, *Nature Reviews Molecular Cell Biology*, 2012, **13**, 448-462.
21. S. Neelamegham and L. K. Mahal, *Current Opinion in Structural Biology*, 2016, **40**, 145-152.
22. Richard D. Cummings and J. M. Pierce, *Chemistry & Biology*, 2014, **21**, 1-15.
23. A. Dröscher, *Glycoconjugate Journal*, 1998, **15**, 733-736.
24. F. Bard and J. Chia, *Trends in Cell Biology*, 2016, **26**, 379-388.
25. W. C. Boyd and E. Shapleigh, *Science*, 1954, **119**, 419.
26. W. C. Boyd, *Vox Sanguinis*, 1963, **8**, 1-32.
27. S. André, H. Kaltner, J. C. Manning, P. V. Murphy and H.-J. Gabius, *Molecules*, 2015, **20**, 1788-1823.
28. A. Varrot, B. Blanchard and A. Imberty, in *Carbohydrate Recognition*, 2011, DOI: https://doi.org/10.1002/9781118017586.ch13, pp. 329-347.
29. H. J. Gabius, S. André, J. Jiménez-Barbero, A. Romero and D. Solís, *Trends in biochemical sciences*, 2011, **36**, 298-313.
30. A. Bernardi and P. Cheshev, *Chemistry (Weinheim an der Bergstrasse, Germany)*, 2008, **14**, 7434-7441.
31. S. Pérez and I. Tvaroška, in *Advances in Carbohydrate Chemistry and Biochemistry*, ed. D. Horton, Academic Press, 2014, vol. 71, pp. 9-136.
32. S. Roseman, *The Journal of biological chemistry*, 2001, **276**, 41527-41542.
33. P. Brocca, A. Bernardi, L. Raimondi and S. Sonnino, *Glycoconjugate Journal*, 2000, **17**, 283-299.
34. J. Jiménez-Barbero, J. L. Asensio, F. J. Cañada and A. Poveda, *Current Opinion in Structural Biology*, 1999, **9**, 549-555.
35. J. L. Magnani and B. Ernst, *Discovery medicine*, 2009, **8**, 247-252.
36. B. Ernst and J. L. Magnani, *Nature Reviews Drug Discovery*, 2009, **8**, 661-677.

37. A. Tamburrini, C. Colombo and A. Bernardi, *Medicinal Research Reviews*, 2020, **40**, 495-531.
38. L. Cipolla, B. La Ferla, C. Airoldi, C. Zona, A. Orsato, N. Shaikh, L. Russo and F. Nicotra, *Future medicinal chemistry*, 2010, **2**, 587-599.
39. I. Robina, P. Vogel and Z. J. Witczak, *Current Organic Chemistry*, 2001, **5**, 1177-1214.
40. Z. J. Witczak, *Current medicinal chemistry*, 1999, **6**, 165-178.
41. Z. J. Witczak and J. M. Culhane, *Applied Microbiology and Biotechnology*, 2005, **69**, 237-244.
42. B. López-Méndez, C. Jia, Y. Zhang, L. H. Zhang, P. Sinaÿ, J. Jiménez-Barbero and M. Sollogoub, *Chemistry, an Asian journal*, 2008, **3**, 51-58.
43. G. E. McCasland, M. O. Naumann and L. J. Durham, *The Journal of Organic Chemistry*, 1966, **31**, 3079-3089.
44. J. Hernández, R. Ramos, N. Sastre, R. Meza, H. Hommer, M. Salas and B. Gordillo, *Tetrahedron*, 2004, **60**, 10927-10941.
45. W. S. Wadsworth and W. D. Emmons, *Journal of the American Chemical Society*, 1962, **84**, 610-617.
46. J.-N. Volle, D. Virieux, M. Starck, J. Monbrun, L. Clarion and J.-L. Pirat, *Tetrahedron: Asymmetry*, 2006, **17**, 1402-1408.
47. Y. Yang and B. Yu, *Chemical Reviews*, 2017, **117**, 12281-12356.
48. Y. He, R. J. Hinklin, J. Chang and L. L. Kiessling, *Organic Letters*, 2004, **6**, 4479-4482.
49. S. Lian, H. Su, B.-X. Zhao, W.-Y. Liu, L.-W. Zheng and J.-Y. Miao, *Bioorganic & Medicinal Chemistry*, 2009, **17**, 7085-7092.
50. K. Bijian, Z. Zhang, B. Xu, S. Jie, B. Chen, S. Wan, J. Wu, T. Jiang and M. A. Alaoui-Jamali, *European Journal of Medicinal Chemistry*, 2012, **48**, 143-152.
51. K. Sidoryk, L. Rárová, J. Oklešťková, Z. Pakulski, M. Strnad, P. Cmoch and R. Luboradzki, *Organic & Biomolecular Chemistry*, 2016, **14**, 10238-10248.
52. E. Montero, A. García-Herrero, Juan L. Asensio, K. Hirai, S. Ogawa, F. Santoyo-González, F. J. Cañada and J. Jiménez-Barbero, *European Journal of Organic Chemistry*, 2000, **2000**, 1945-1952.
53. K. Bock, J. Ø. Duus and S. Refn, *Carbohydrate Research*, 1994, **253**, 51-67.
54. E. Montero, M. Vallmitjana, J. A. Pérez-Pons, E. Querol, J. Jiménez-Barbero and F. J. Cañada, *FEBS Letters*, 1998, **421**, 243-248.
55. H. Driguez, *ChemBioChem*, 2001, **2**, 311-318.
56. K. L. Kirk, *Journal of Fluorine Chemistry*, 2006, **127**, 1013-1029.
57. A. Baumann, S. Marchner, M. Daum and A. Hoffmann-Röder, *European Journal of Organic Chemistry*, 2018, **2018**, 3803-3815.

58. B. Xu, L. Unione, J. Sardinha, S. Wu, M. Ethève-Quelquejeu, A. Pilar Rauter, Y. Blériot, Y. Zhang, S. Martín-Santamaría, D. Díaz, J. Jiménez-Barbero and M. Sollogoub, *Angewandte Chemie (International ed. in English)*, 2014, **53**, 9597-9602.
59. H. C. Kolb and B. Ernst, *Chemistry – A European Journal*, 1997, **3**, 1571-1578.
60. A. Bernardi, D. Arosio, L. Manzoni, F. Micheli, A. Pasquarello and P. Seneci, *The Journal of Organic Chemistry*, 2001, **66**, 6209-6216.
61. A. Bernardi, A. Checchia, P. Brocca, S. Sonnino and F. Zuccotto, *Journal of the American Chemical Society*, 1999, **121**, 2032-2036.
62. A. Bernardi, D. Arosio, L. Manzoni, D. Monti, H. Posteri, D. Potenza, S. Mari and J. Jiménez-Barbero, *Organic & Biomolecular Chemistry*, 2003, **1**, 785-792.
63. S. Mari, H. Posteri, G. Marcou, D. Potenza, F. Micheli, F. J. Cañada, J. Jimenez-Barbero and A. Bernardi, *European Journal of Organic Chemistry*, 2004, **2004**, 5119-5225.
64. A. Tamburrini, S. Achilli, F. Vasile, S. Sattin, C. Vivès, C. Colombo, F. Fieschi and A. Bernardi, *Bioorganic & Medicinal Chemistry*, 2017, **25**, 5142-5147.
65. L. Medve, S. Achilli, J. Guzman-Caldentey, M. Thépaut, L. Senaldi, A. Le Roy, S. Sattin, C. Ebel, C. Vivès, S. Martin-Santamaria, A. Bernardi and F. Fieschi, *Chemistry – A European Journal*, 2019, **25**, 14659-14668.
66. G. Terraneo, D. Potenza, A. Canales, J. Jiménez-Barbero, K. K. Baldridge and A. Bernardi, *Journal of the American Chemical Society*, 2007, **129**, 2890-2900.
67. J. J. Reina, S. Sattin, D. Invernizzi, S. Mari, L. Martínez-Prats, G. Tabarani, F. Fieschi, R. Delgado, P. M. Nieto, J. Rojo and A. Bernardi, *ChemMedChem*, 2007, **2**, 1030-1036.
68. A. Tamburrini, PhD Thesis, Università degli studi di Milano, 2019.
69. Y. Ferrand, M. P. Crump and A. P. Davis, *Science*, 2007, **318**, 619.
70. O. Francesconi, F. Milanesi, C. Nativi and S. Roelens, *Angewandte Chemie International Edition*, 2021, **60**, 11168-11172.

CHAPTER TWO

Scope of the thesis: Interfering with the sugar code

As explained in Chapter One, cellular communication in all living cells includes coding the information in biomolecules and subsequent translation into cellular effects. Carbohydrates are a class of biomolecules responsible for coding of the information on cell surface. The cellular responses triggered upon interaction between oligosaccharide determinants with protein receptors are involved in many physiological and pathological processes.

Unravelling the aspects of encoding the biochemical information in the language of sugars has been described as one of the most critical challenges of the post-genomic era.[1, 2] Yet, due to complexity of glycoconjugate structures, their demanding synthesis and analysis, the extent to which the sugar code has been deciphered is still limited.

To crack the so called "sugar code" scientists are taking different approaches, particularly carbohydrate chemistry and rational protein engineering are often explored to deliver molecular probes such as glycomimetics and lectin constructs.

In this thesis I was privileged to experience the work in both fields, developing glycomimetic compounds as well as lectin-based peptides.

On one side, our aim was to develop a new class of glycomimetics that could potentially be recognized by various lectins. The constructs with the general structure of *N*-linked-pseudo-*thio*-glycosides consist of a sugar moiety and an aglycone moiety (**Fig. 2.1**). The sugar is responsible for recognition of these compounds by lectins, while the aglycone works as a tether, it allows for conjugation of these glycomimetics to various scaffolds, for multivalent presentation as well as for pseudo-glycopeptide synthesis. The *thio*- and *N*-linkages inserted in the structure improve enzymatic and hydrolytic stability of these glycomimetics compared to natural saccharides. To simplify the synthesis of these compounds, we set out to develop a new synthetic approach, a one-pot aziridine opening reaction and explore it on different sugar substrates (glycosyl thioacetates) (**Fig. 2.1**). This synthetic method is an operationally simpler alternative to more standard glycosylation methods, it can be performed as a one-pot reaction with *in situ* generation of thiolate and subsequent aziridine opening.

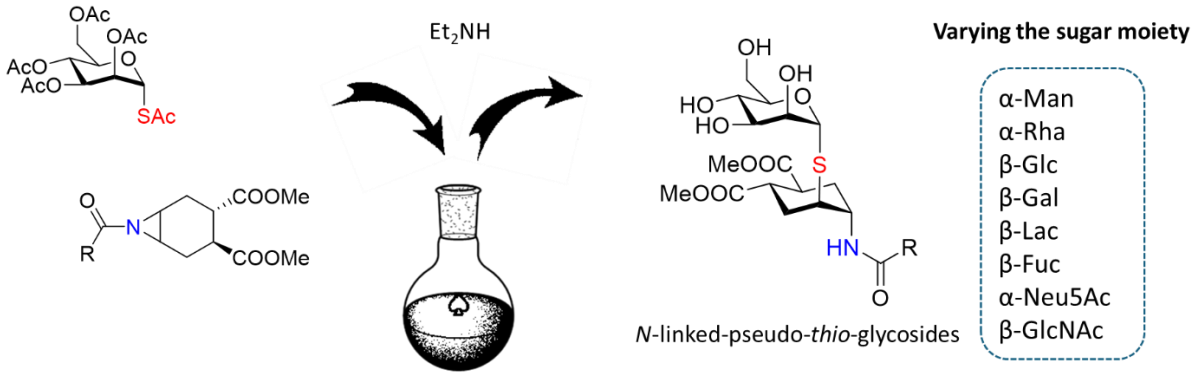


Fig. 2.1 The principle of the one-pot aziridine opening approach for synthesis of glycomimetics with the structure of *N*-linked-pseudo-*thio*-glycosides

On the other hand, we aimed to dive into developing new flexible peptide constructs as antiviral agents. For this purpose, we chose to work with a small lectin Cyanovirin-N which potently inactivates a broad range of viruses including HIV-1, SIV, hepatitis C, Ebola and influenza viruses. Cyanovirin-N consists of two domains, A and B, each one of them contains one carbohydrate binding site (CBS) (**Fig. 2.2**). The carbohydrate binding sites selectively recognize small oligomannose epitopes (Manα(1,2)Man) of mannose type glycans Man_9 and Man_8 found on viral envelope glycoproteins. The multisite binding is of crucial importance for the antiviral activity of this lectin, therefore increasing the number of carbohydrate binding sites may improve the antiviral activity of such peptide construct. Our aim was to work towards building flexible peptide constructs that would include multiple copies of domain B that carries a high affinity carbohydrate binding site. To simplify the construction of a multivalent system we wanted to particularly focus on investigating the possibility of making this domain B peptide synthetically, investigate its folding and binding with the ligand Manα(1,2)Man also in comparison to a peptide consisting of domain B made by recombinant expression from *E. coli*.

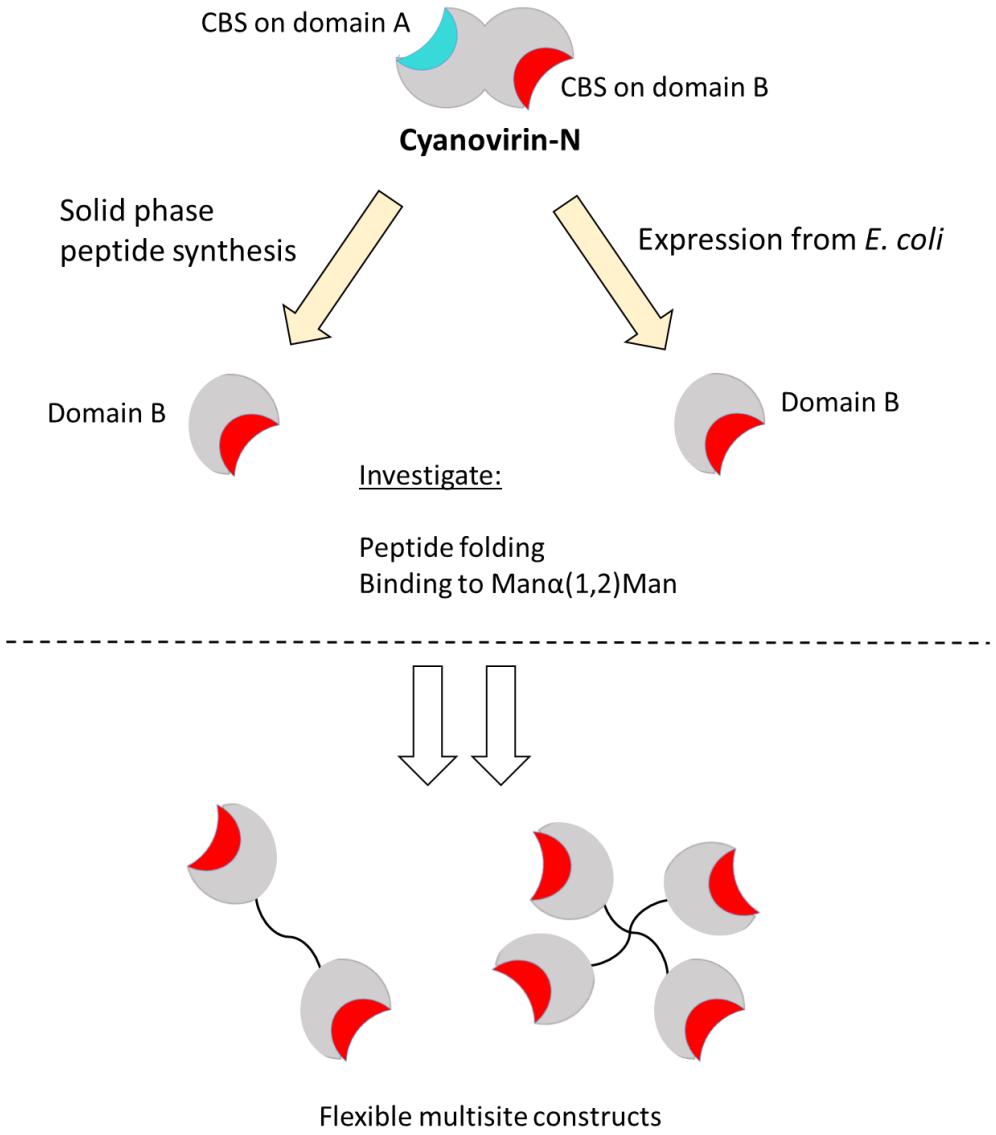


Fig. 2.2 Approach towards developing new flexible multisite constructs containing domain B of Cyanovirin-N

References

1. H.-J. Gabius, H.-C. Siebert, S. André, J. Jiménez-Barbero and H. Rüdiger, *ChemBioChem*, 2004, **5**, 740-764.

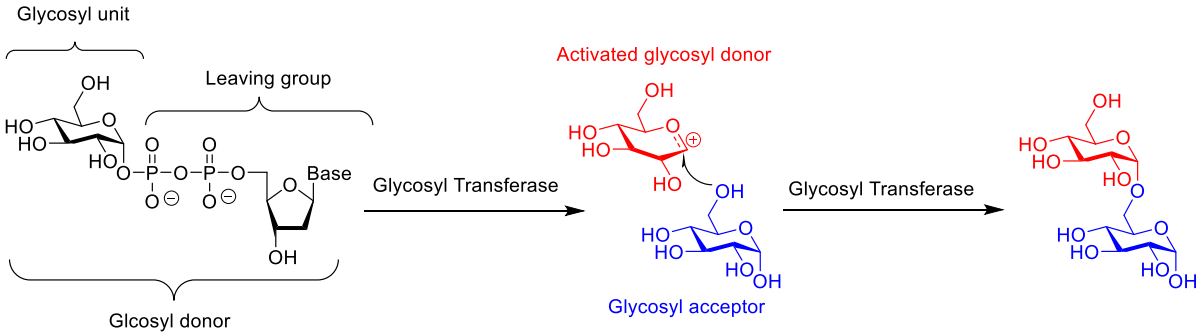
2. J. C. Paulson, O. Blixt and B. E. Collins, *Nature Chemical Biology*, 2006, **2**, 238-248.

CHAPTER THREE

Strategies for synthesis of oligosaccharides and *thio*-oligosaccharides

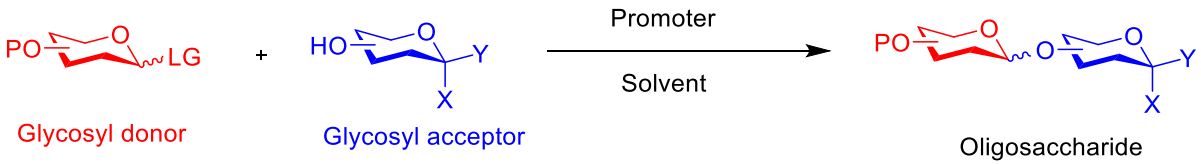
3.1. Introduction

In nature, oligosaccharides are synthesized from activated glycosyl donors (monosaccharide conjugates with nucleotides) and acceptors with free hydroxyl groups (**Scheme 3.1**). Regio- and stereoselectivity is achieved through a specific glycosyl transferase involved in the synthesis.



Scheme 3.1 Nature's approach towards oligosaccharide synthesis

Similarly, the approach taken by organic chemists towards oligosaccharide synthesis uses glycosyl donors equipped with a leaving group and glycosyl acceptors with a free hydroxyl group (**Scheme 3.2**). Glycosyl acceptors can be either sugar substrates or other alcohols. Regio- and sometimes also stereoselectivity is achieved through a careful selection of protecting groups.



Scheme 3.2 Chemists' approach towards oligosaccharide synthesis

Before setting on a demanding journey towards the oligosaccharide synthesis, a chemist has to thoroughly consider several aspects. First, the availability of the starting monosaccharide and preparation of the glycosyl donor and acceptor. While most monosaccharides found in mammalian cells are commercially available, same is often not true for bacterial monosaccharides. A strategy for manipulation of protecting groups has to be designed, to allow selective protections and deprotections, so that regioselectivity of the synthesis will be achieved through masking or exposing

OH or NH_2 groups. Then, one must decide on glycosylation strategy, so that the glycosidic linkage will be formed in (preferably) stereoselective manner. A choice of the leaving group and promoter have to be taken into account, as well as which anomeric configuration is to be synthesized (1,2-*cis* or 1,2-*trans* configuration). When aiming for larger oligosaccharide structures, the order of the glycosidic bond formation also needs to be taken into account. This is necessary to take a decision on temporary and permanent protecting groups that will be employed. Permanent protecting groups are usually ethers (mostly benzyl ethers (Bn), but also pivaloyl (Piv) and benzoate (Bz)). Temporary protecting groups need to be easily cleaved, but stable enough during glycosylation; trityl ethers, fluorenylmethoxycarbonyl (Fmoc), several silyl groups and even acetates can be used as temporary protection groups.[1] Besides regioselectivity, protecting groups may also influence stereochemical outcomes of the glycosylation.

3.1.1. Types of *O*-glycosidic linkages

The two major types of *O*-glycosides are most commonly defined as α- and β- or, more important from the synthetic point of view, 1,2-*cis* and 1,2-*trans* glycosides. 1,2-*cis* glycosyl residues are α-glycosides for D-glucose, D-galactose, L-fucose and β-glycosides for D-mannose and L-rhamnose. Hence, 1,2-*trans* glycosyl residues are β-glycosides for D-glucose, D-galactose, L-fucose and α-glycosides for D-mannose and L-rhamnose. Some glycosides cannot be defined as 1,2-*cis* or 1,2-*trans* derivatives, examples are 2-deoxyglycosides and sialosides (**Fig. 3.1**).

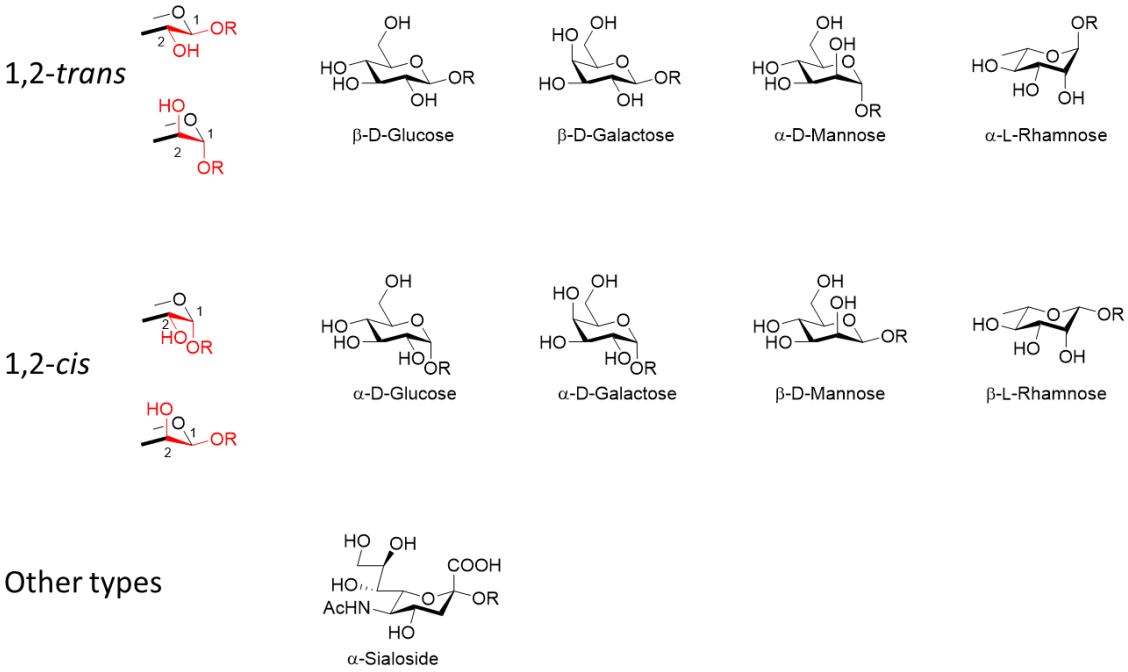
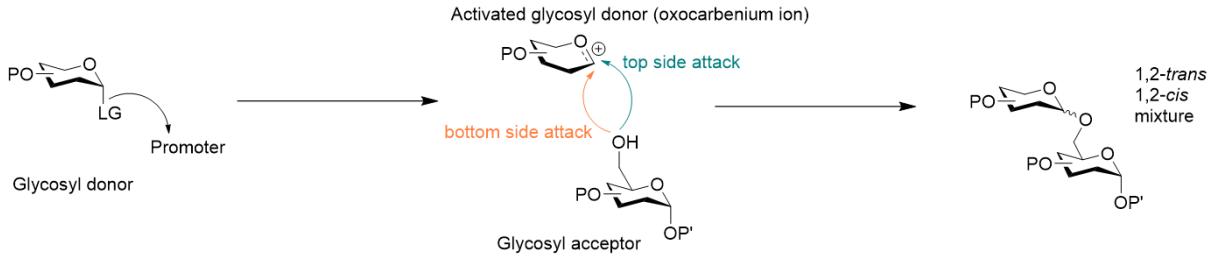


Fig. 3.1 Types of *O*-glycosidic linkages

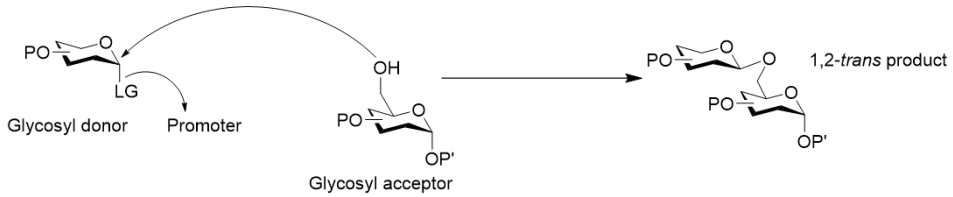
3.2. Mechanistic background of glycosylation reaction

Chemical glycosylation is a coupling of a glycosyl donor with the glycosyl acceptor (monosaccharide or aglycone) and the new linkage that is formed is called the glycoside bond. The donor is activated with the leaving group attached to the anomeric position. Upon the addition of an electrophilic promoter the leaving group is activated and then an oxocarbenium ion is formed. After this nucleophilic attack by the acceptor occurs and leads to glycoside bond formation. The attack on the donor can occur in two different ways, from the bottom of the sugar which leads to 1,2-*cis* glycoside or from the top of the sugar which leads to 1,2-*trans* glycoside (**Scheme 3.3, a**). As the reaction takes place at the secondary carbon atom and usually with the use of weak nucleophiles (acceptors), it often follows unimolecular S_N1 mechanism. However, a bimolecular S_N2 mechanism is also possible with the inversion of anomeric configuration (**Scheme 3.3, b**).[2, 3]

a) Two-step mechanism (via a positively charged intermediate)

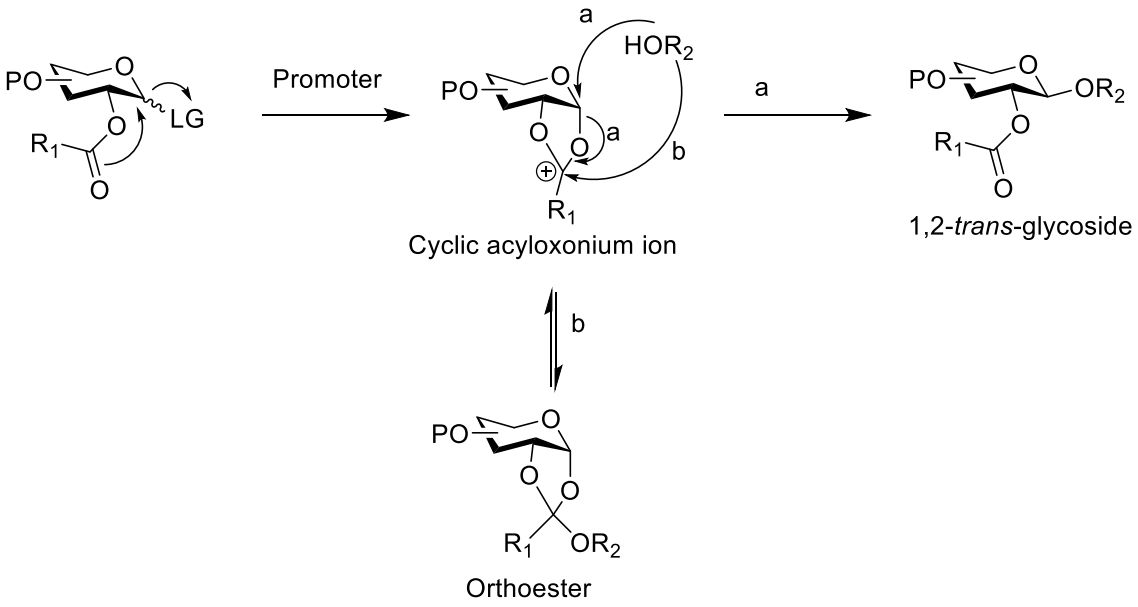


b) One-step mechanism (with inversion of anomeric configuration)



Scheme 3.3 Mechanism of glycosylation reaction: a) S_N1 mechanism leads through oxocarbenium ion formation; b) S_N2 mechanism with the inversion of configuration

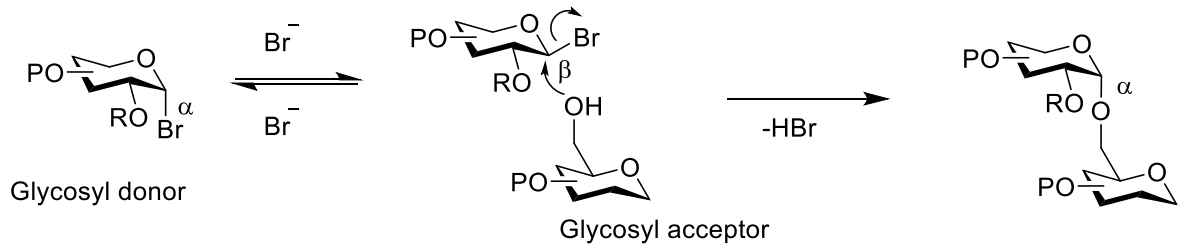
The participating acyl group on C_2 position of the donor directs the stereochemistry. Oxocarbenium ion interacts with the acyl group at the C_2 position to form the acyloxonium ion complex through neighbouring group participation. Therefore, acceptor can only attack from one side to form 1,2-*trans* linkage (**Scheme 3.4**). Common participating groups are *O*-acetyl (Ac), *O*-benzoyl (Bz), 2-phthalimido (NPhth). For 1,2-*cis* glycosidic bond formation, the protecting group at C_2 position has to be non-participating. The most common non-participating groups are benzyl (OBn) for neutral sugars and azide (N_3) for 2-amino-2-deoxysugars.[3]



Scheme 3.4 Glycosylation mechanism with neighbouring group participation

Stereoselectivity is a major obstacle to overcome in glycosylation reactions. For example, in case of 1,2-*trans* glycoside formations, substantial amount of 1,2-*cis*-linked products can sometimes be formed, most often when unreactive alcohols are used as substrates and/or poorly nucleophilic participatory substituents are present on C_2.[3] In case of 1,2-*cis* glycosylations that happen without participation of the C_2 substituent, the nucleophilic attack is almost equally possible from either side (top, *trans* or bottom, *cis*). The *cis* product is thermodynamically favoured due to the anomeric effect, yet substantial amounts of *trans* product can be obtained due to irreversible nature of glycosylation of complex aglycones. Generally, achieving stereoselectivity for 1,2-*cis* glycosylations is more challenging than for 1,2-*trans* glycosylations. Various factors, such as temperature, protecting groups, conformation, solvent, promoter, steric hindrance or leaving groups can all influence the outcome of glycosylation reaction.

A particular solution to stereoselectivity problem of 1,2-*cis* glycosylations was the work of Lemieux on the halide-ion-catalyzed glycosylation reaction. Addition of tetraalkylammonium bromide (Et_4NBr) allowed for rapid equilibrium established between a more stable α-halide and more reactive β-halide. Glycosyl acceptor then preferentially reacts with the more reactive β-glycosyl donor in an S_N2 fashion to give 1,2-*cis* product as the main product (**Scheme 3.5**).[4]

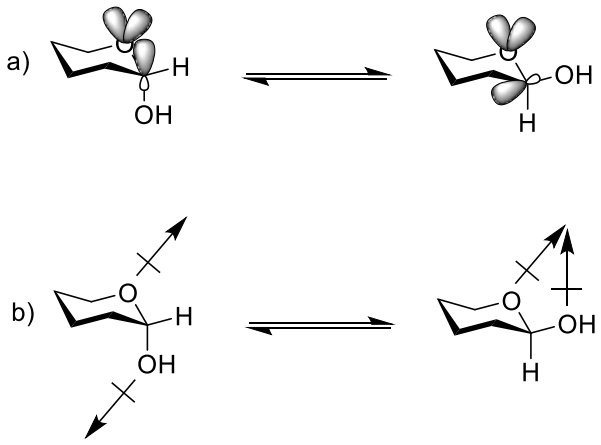


Scheme 3.5 Halide-ion-catalyzed glycosylation leads selectively to 1,2-*cis* product[4]

While a very clever solution was established for selective 1,2-*cis* product formations by Lemieux, there are yet no general solutions to this challenge.

3.2.1. Anomeric effect

Polar substituents such as halide, OR or SR attached to the anomeric center of pyranoses prefer the axial orientation. This phenomenon is known as the anomeric effect or also as Edward-Lemieux effect, due to the pioneer work they have done explaining it.[3, 5] One factor explaining this effect is that the substituent on the atom bonded to the ring at C_1 has lone-pair electrons, that would have repulsive interactions with those of the ring oxygen if the substituent is in equatorial position, but not if it is in axial position. Additionally, electron-withdrawing axial substituent is stabilized by hyperconjugation, due to the periplanar orientation of both nonbonding orbital of ring oxygen and antibonding orbital of C_1. The same is not true for the equatorial substituent, because the two orbitals are in different planes and therefore unable to interact (**Scheme 3.6 a**). Computational studies however showed, that hyperconjugation may only play a minor role in the anomeric effect.[6]



Scheme 3.6 Anomeric effect explained by: a) Molecular orbital theory; b) Dipole moment theory- from Ref[6]

A dipole moment theory on the other hand says that the dipoles of heteroatoms are partially aligned and therefore repel each other in the equatorial isomer. In the axial isomer however the dipoles are

oriented in the opposite direction and stabilize the system by implicating a lower energy barrier (**Scheme 3.6, b**).

The axial stereoisomer is the thermodynamically controlled product, while the equatorial isomer is the kinetically controlled product. This is however not true for sugars with axial 2-OH groups like D-mannose and L-rhamnose, where the axial anomer is both thermodynamically and kinetically controlled.[6]

3.3. Stereoselectivity of glycosylation

A number of factors can play a role in achieving stereoselectivity in glycosylation reactions, some of them are summarized below.

3.3.1. Glycosyl donor structure

Protecting groups

As mentioned above, the main impact on stereoselectivity comes from the participation of the neighboring group on C_2 position. The remote substituents have less effect on stereoselectivity, however there are speculations that a substituent at C_6 position may also direct the stereochemical outcome of the reaction.[7-10] Long-range 6-*O*-acyl or carbonate group assistance may result in the preferential formation of α-glucosides. Additionally, bulkier or strong electron-withdrawing C_6 substituents are beneficial for 1,2-*cis* glycose product formation, presumably due to shielding (steric or electronic) of the top face of the ring and therefore directing the attack from the opposite site.

Participating moiety at C_4 position is important for stereoselectivity in D-galactose. Protecting groups such as p-methoxybenzoyl (anisoyl) and diethylthiocarbamoyl were found to be particularly beneficial for α-galactoside formation.[11]

"Armed-disarmed" strategy was introduced by Fraser-Reid et al. to tune the reactivity of the glycosyl donor through a proper choice of the protecting group.[12] The reactivity of the donor is strongly reduced by ester-type protecting groups ("disarmed" donor) in comparison to the donor with ether-type protecting groups. Such effect can be explained by the electron-withdrawing ability of ester protecting groups, which decrease the electron density of the donor and thus destabilize the resulting oxocarbenium ion. Glycosyl donors with participating groups on C_2 are also more reactive than their analogs with non-participating group at C_2 position (**Fig. 3.2**).

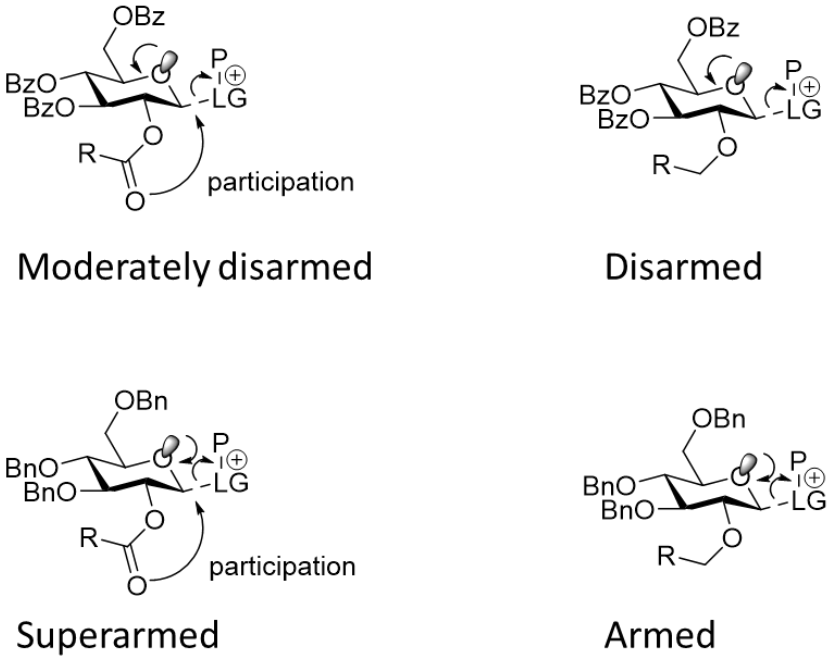


*Fig. 3.*2 "Armed-disarmed" strategy-the influence of protecting groups on reactivity of glycosyl donors

Leaving group

The glycosylation reaction occasionally proceeds via bimolecular S_N2 mechanism with the inversion of the anomeric configuration. 1,2-*trans* glycosyl donors then stereospecifically lead to 1,2-*cis* glycosylation products and vice versa. Examples of such 1,2-*trans* donors are β-glucosyl halides[4], glycosyl thiocyanates[13, 14], anomeric triflates formed *in situ*[15], etc.

3.3.2. Glycosyl acceptor structure

Position of the hydroxyl

Very reactive hydroxyl groups, thus stronger nucleophiles result in faster reaction that is more difficult to control, consequently leading to lower selectivity in terms of α/β ratio. Glycosylation of a more reactive primary hydroxyl provides poorer stereoselectivity than glycosylation of a secondary hydroxyl.[16]

Protecting groups

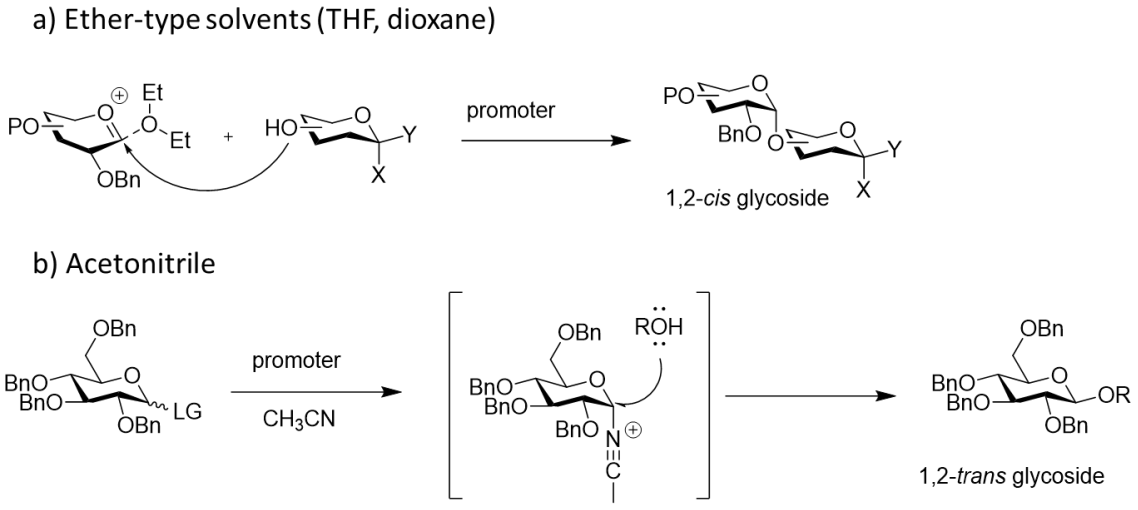
Ester electron-withdrawing substituents reduce electron density of the neighbouring hydroxyl groups by lowering the nucleophilicity.[10, 17, 18] This feature may contribute to the reaction carried out in a more controlled manner, thus improving stereoselectivity. However, reducing the reactivity of hydroxyls may also result in lower glycosylation rates.

3.3.3. Reaction conditions

Solvent effect

From the perspective of solvation, use of polar solvents results in charge separation between O_5 and O_1-β, therefore more β-glycoside is formed. For synthesis of α-glycosides CH_2Cl_2, $ClCH_2CH_2Cl$ or toluene should be beneficial.[3]

There are also more powerful solvent effects that need to be taken into consideration. Acetonitrile diethyl ether, tetrahydrofuran and dioxane are examples of so-called participating solvents (**Scheme 3.7**). In acetonitrile[19], nitrilium cation is formed *in situ* and adopts axial orientation which leads to equatorially substituited 1,2-*trans* glycosides even with glycosyl donors with nonparticipating C_2 substituents. On the other hand, in ether-type solvents (tetrahydrofuran, dioxane) [20, 21] equatorial intermediate is usually present, leading to axial glycosidic bond formation.



Scheme 3.7 Solvent effect: a) Ether-type solvents lead to 1,2-*cis* glycoside formation, b) Acetonitrile leads to 1,2-*trans* glycoside formation

Promoter

Glycosylations with glycosyl halides give best results in halide-ion-catalyzed reactions[4], thioglycosides perform best when activated with mild promoter, such as iodinium dicollidine perchlorate (IDPC)[22, 23] and trichloroacetimidates when activated with strong acidic catalysts, such as trimethylsilyl trifluoromethanesulfonate (TMSOTf)[24]. Various additives can also influence the stereochemical outcome, for example, the use of perchlorate ion additive can strongly influence 1,2-*cis*-glycosylations.[25, 26]

Temperature and pressure

High pressure applied to the reaction system can enhance 1,2-*trans* selectivity of the reactions of donors with participating group on C_2.[21] On the other hand, it does not have a major influence on stereoselectivity of donors with non-participating substituent, but it can increase the yield of the reaction. Lower temperatures generally allow for kinetically controlled glycosylation favouring 1,2-*trans* product formation.[27-29] Cases with the opposite observation were however also reported.[30, 31]

3.3.4. Other factors

Besides the factors described above, the outcome of the glycosylation may also be unexpectedly influenced by other factors and conditions. For example, unfavourable steric interactions may occur between the glycosyl donor and acceptor in the transition state. Double stereo differentiation takes place when interactions between bulky substituents in glycosyl donor and acceptor overshadow the effect of C_2 participating group.[32]

3.4. *O*-glycosylation strategies

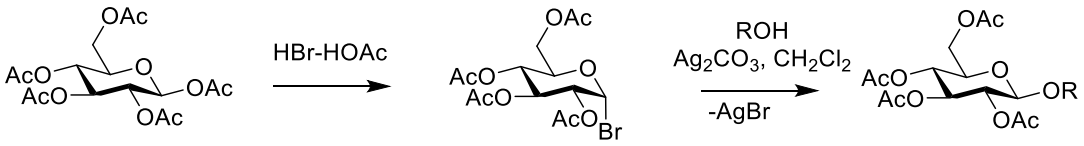
An efficient glycosylation method should in principle meet three main requirements:

- Small amounts of reagents used (glycosyl donor easily synthesized and activated by a catalytic amount of reagent);
- Glycosylation step stereoselective and high-yielding;
- Method applicable on large scale.

3.4.1. Classical methods

Koenigs-Knorr type reactions

The oldest method for stereospecific synthesis of 1,2-*trans* glycosides was developed in 1901 by Wilhelm Koenigs and Eduard Knorr.[33] The glycosyl donors used in this strategy are glycosyl bromides and glycosyl chlorides (**Scheme 3.8**). The promoters in this reaction are silver salts, insoluble Ag_2O and Ag_2CO_3 or soluble AgOTf and $AgClO_3$. The participating group at C_2 position leads to selectively 1,2-*trans* glycosylation products, particularly with the use of insoluble salts as promoters. Reaction in this case presumably occurs on the surface of Ag_2CO_3, through shielding the α-face of the sugar it leads to exclusively *trans* product.



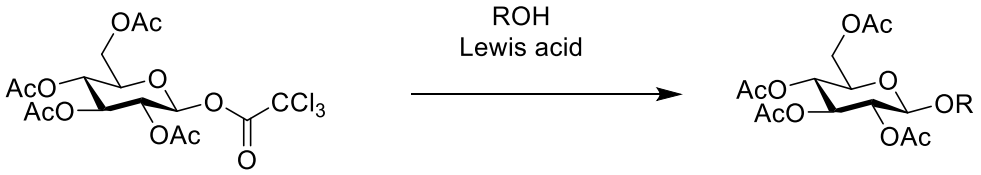
Scheme 3.8 Koenigs-Knorr glycosylation reaction

Thioglycosides

Thioglycosides in the context of *S*-glycosylations will be discussed later in this chapter, however they can also be used as glycosyl donors in *O*-glycosylations. Unlike many other glycosyl donors, including glycosyl halides, thioglycosides are very stable. They can be prepared in a variety of methods and can be converted to all other glycosyl donors directly or by a two-step procedure. All this makes them interesting glycosyl donors that are often used. Different promoters can be used for activation of thioglycosides, for example with iodonium ions obtained from *N*-iodosuccinimide (NIS), iodonium dicollidine perchlorate (ICDP) or (dimethylmethylthiosulfonium)-trifluoromethanesulfonate (DMTST).[2, 6]

The trichloroacetimidate method

Trichloroacetimidates contain a good leaving group on the the anomeric oxygen. These glycosyl donors can be activated under relatively mild conditions by Lewis acid catalyst such as borontrifluoride etherate complex (BF_3xEt_2O) or trimethylsilyl trifluoromethanesulfonate (TMSOTf) (**Scheme 3.9**).[2, 24, 34] The trichloroacetimidate method very often meets the requirements for an efficient glycosidation reaction outlined above. *O*-Glycosyl trichloroacetimidates are formed readily, only require catalytic amounts of Lewis acid for their activation, the method can be scaled up and is applicable to many sugar substrates. The trichloroacetimidate method is therefore widely used and can also be employed in the synthesis of oligosaccharides.



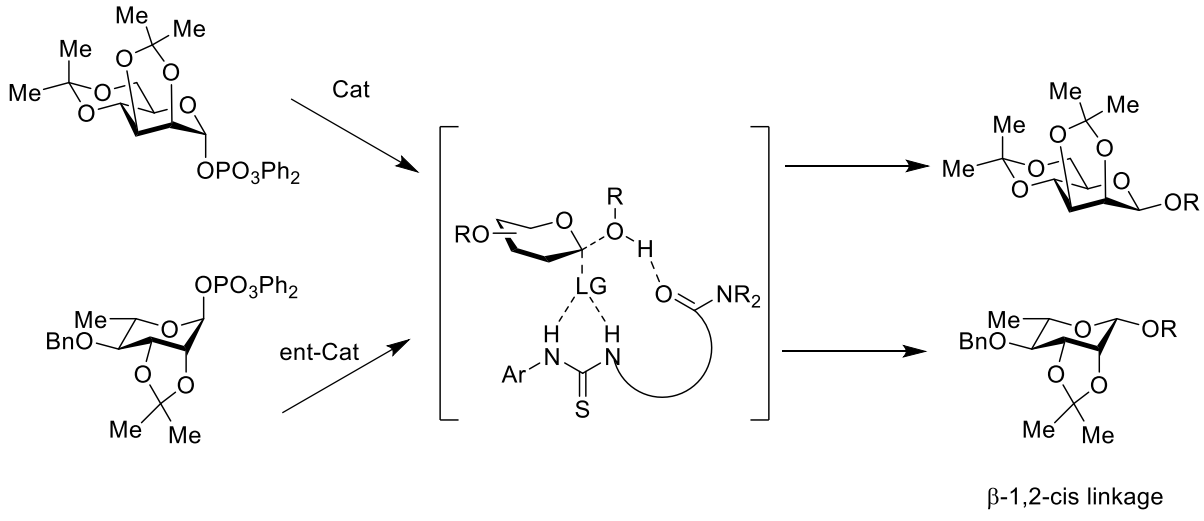
Scheme 3.9 Glycosyation reaction with the trichloroacetimidate method

While the reactivity of the tricholoracetimidates generally presents an advantage for their use in reactions, it also comes with drawbacks such as side reactions or decompositions of the donor even before glycosylation occurs. For example, the donor may rearrange into unreactive trichloroacetamide, thus compromising the yield significantly.

The effect of the participating C_2 group usually leads to 1,2-*trans* glycosylation products. If donor carries a non-participating group, S_N2 type reaction can be achieved by the use of nonpolar solvents, low reaction temperatures and mild promoter (BF_3xEt_2O). On the other hand, harsher conditions such as higher temperature, strong promoter (TMSOTf) and more polar solvent lead to formation of thermodynamically more stable glycosylation product (α-glucose, α-mannose).[2]

3.4.2. Current developments

A number of intriguing glycosidic bond syntheses have been reported and reviewed.[3, 6, 34] However, despite all efforts there are still significant challenges ahead for organic chemists in the field of oligosaccharide synthesis. One of such puzzles is certainly the synthesis of β-1,2-*cis* glycosidic linkages.[35, 36] Due to the anomeric stabilization of α-isomers, a typical mannosylation or rhamnosylation reaction is driven towards a kinetically stabilized α-product. Both β-mannosides and β-rhamnosides are however biologically important building blocks. A significant amount of effort is therefore being devoted to overcoming the α-selectivity and developing β-1,2-*cis* glycosylation methods. Successful strategies include intramolecular aglycone delivery[37, 38], the use of 1,2-anhydropyranose donors with borinic and boronic acid catalysts[39], anomeric *O*-alkylation[40] and protecting group controls such as with 4,6-benzylidene acetals[41] and 2,6-lactones[42]. Recently a method for highly β-selective 1,2-*cis*-O-pyranosylations catalysed by bis-thioureas was reported by Jacobsen et al (**Scheme 3.10**).[43] The method employs easily accessible acetonide protected donors and selectivity is achieved by catalyst control. This method is applicable to a wide, yet still limited, range of alcohol nucleophiles.



Scheme 3.10 The principle of β-1,2-*cis*-glycosylation catalysed by bis-thioureas developed by Jacobsen et al.[43]

With advances in glycoscience, also the field of carbohydrate chemistry is moving towards the synthesis of larger structures, in particular towards synthesis of glycans and glycopeptides. Besides the challenge of multiple selective glycosylations required for constructing larger structures, additional obstacles are also synthesis of branched structures, employing extra functionalities (e.g. uronic acids, amino sugars), substituents (phosphates, sulphates, acetates) and unusual monosaccharide residues.

Learning from other fields of biomolecules synthesis (e.g peptides), a trend in the recent years is also to look for simplifications of glycoside synthesis. Such methods include solid phase synthesis, one-pot glycosylations and enzymatic synthesis.

Solid phase synthesis has become a standard method in the synthesis of peptides, a wide range of solid supports, coupling reagents and protecting groups have been developed and the synthesis can be done in a fully automated fashion with the use of commercially available equipment.

On the other hand, solid phase oligosaccharide synthesis (SPOS) has not yet achieved such bloom. Many common glycosyl donors have been investigated in the acceptor-bound approach for SPOS, such as glycosyl sulfoxides, *O*-glycosyl trichloroacetimidates, thioglycosides, n-pentenyl glycosides and glycosyl phosphates.[34, 44, 45] New polymer supports suitable for SPOS are also being developed.[46] A major effort in this field has been made by Seeberger and co-workers, in 2001 they achieved an automated synthesis of oligosaccharides by using a solid-phase synthesizer with *O*-glycosyl trichloroacetimidates and phosphates as glycosylating agents.[47] While SPOS certainly presents a synthetic advantage, particularly in terms of purification and time[48, 49], it still requires a lengthy preparation of glycosyl donors, optimization of synthetic pathway in solution and does not solve problems related to stereoselectivity.

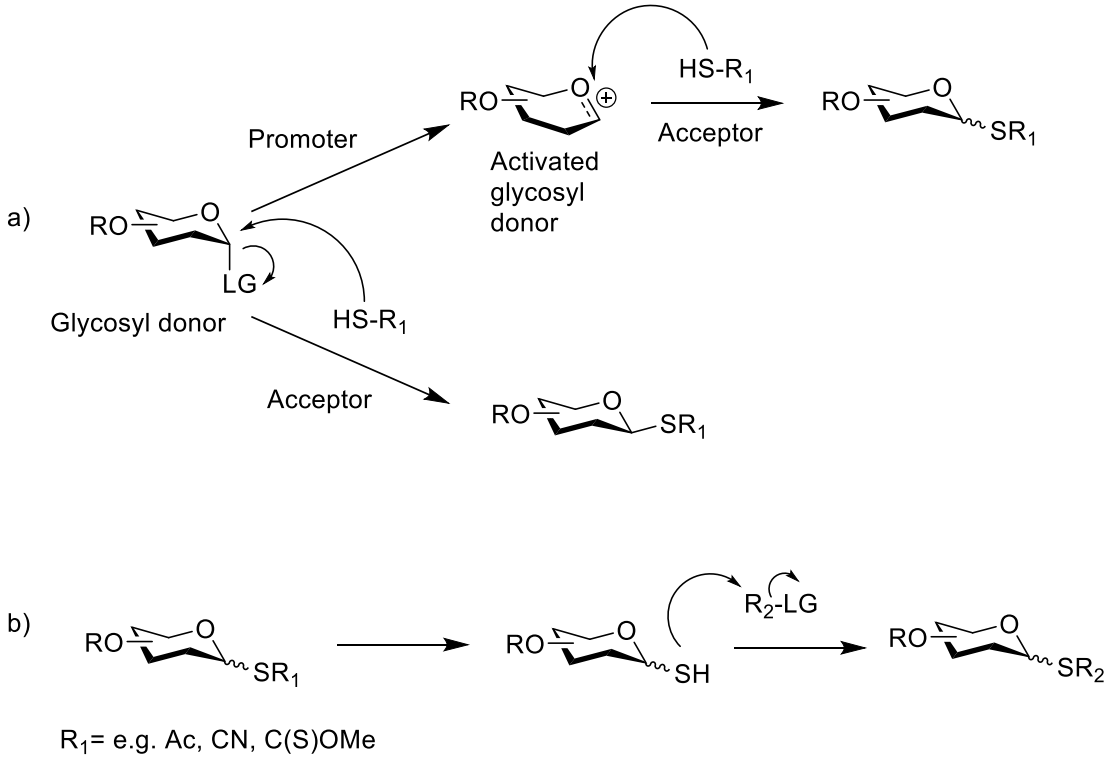
One-pot glycosylation is another useful approach in oligosaccharide synthesis.[50-53] Usually such methodology follows one of three most common strategies:

- On the basis of armed-disarmed approach, chemoselective glycosylation is leaning on different reactivities of glycosyl donors and acceptors;
- Selective activation of the leaving group is the basis for orthogonal glycosylation;
- Preactivation-based glycosylation requires the glycosyl donor to be activated separately, before the addition of the acceptor. The acceptor contains a leaving group for the next step of glycosylation.

While attractive, the one-pot glycosylation approach also does not resolve the stereoselectivity issues, is not applicable to large scale production and is usually limited to construction of two or three different glycosidic linkages.

3.5. *S*-glycosylation strategies

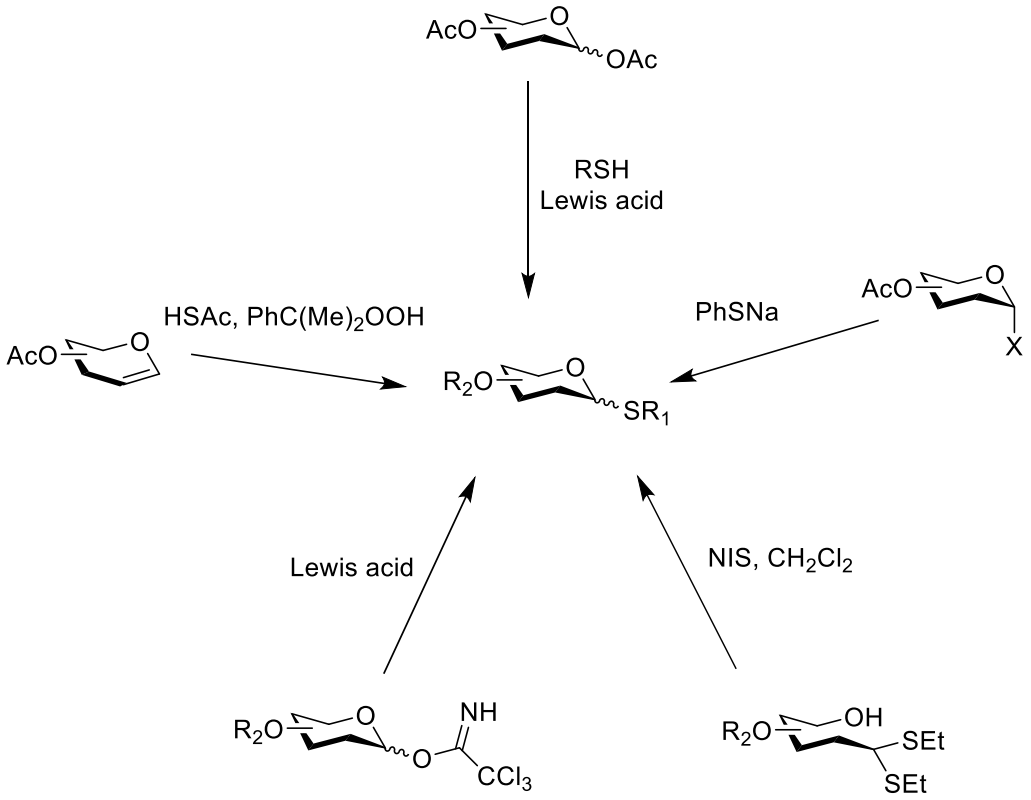
Two general approaches are usually used for synthesis of thioglycosides. A mercaptan group is directly introduced through a displacement reaction of an anomeric leaving group, possibly aided by a promoter. On the other hand, a two (or more)-step procedure is used, in which an anomeric *thio* (non-thiol) group is first introduced and then cleaved or rearranged to give directly a glycoside or an anomeric thiol or thiolate that reacts with an electrophile to give the target thioglycoside (**Scheme 3.11**).[54]



Scheme 3.11 Two general approaches for *S*-glycoside synthesis. a) Thiol group is present on glycosyl acceptor; b) Glycosyl thiol is generated first and then reacts with an electrophile[54]

3.5.1. Synthesis of monothioglycosides

Both of the methods described above are extensively used in the synthesis of *S*-monosaccharides. Different precursors can be used for the synthesis of monothioglycosides, examples include anomeric acetates, glycosy halides, hemiacetals, trichloroacetimidates, glycals, anhydro sugars, etc. (**Scheme 3.12**).



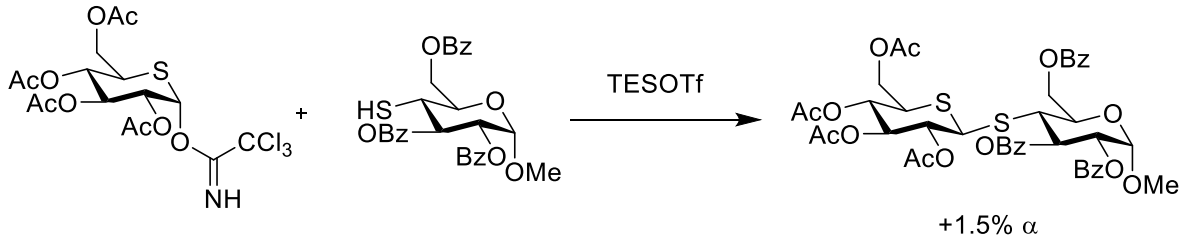
Scheme 3.12 Precursors for synthesis of monothioglycosides

3.5.2. Synthesis of thiooligosaccharides

Both of the above described approaches for thioglycosylation, employing a thiol-bearing acceptor or generating a glycosyl thiol or thiolate donor are commonly used in the synthesis of thioglycosides.[54, 55] The second approach is sometimes pointed out to be advantageous, because in principle anomeric configuration should be more easily controlled, since it is constructed during the formation of the anomeric thiol and then mantained during the glycosylation reaction. However, as explained in Chapters Four and Five, there is clear evidence that there are exceptions to this rule and thus glycosylations with glycosyl thiols should be approached with care.

Glycosylations with thiol acceptors

Glycosylations with glycosyl acceptors bearing a thiol group can sometimes be complicated by the synthesis of the acceptor itself, which therefore cannot be used in large excess. Promoters used in the glycosylation obviously also cannot be thiophilic, which excludes the possibility to use many *O*-glycosylation methods. The trichloroacetimidate method is the most generally applied[56-59] (**Scheme 3.13**), additionally anhydro sugars[60] and glycals[61] can be used as donors.



Scheme 3.13 Synthesis of thiodisaccharide from a trichloroacetimidate glocosyl donor[56]

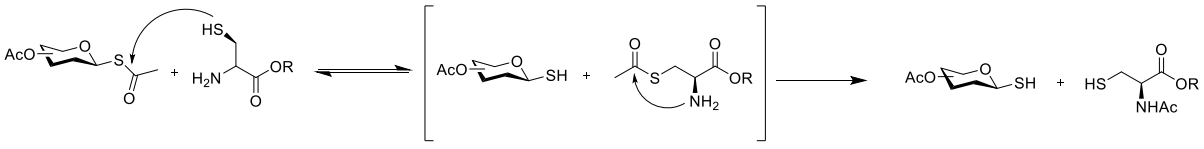
Glycosylations with thiol donors

In this approach the anomeric thiol (or thiolate) is first generated and it then reacts with the electrophile through S_N2 displacement method.

The synthesis of 1,2-*trans*-1-*thio* derivatives is usually straightforward with the participating protecting groups. The synthesis of 1,2-*cis* derivatives is also performed from donors with acetyl protecting groups, in this case from 1,2-*trans* halogeno sugars under conditions favouring S_N2-type reaction to give 1,2-*cis* product.[54]

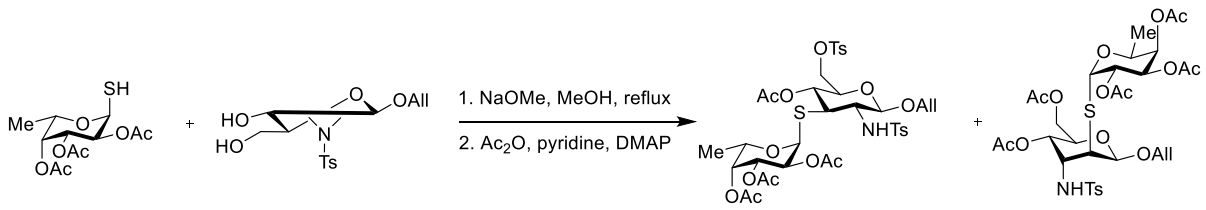
The glycosyl thiols required are usually generated from glycosyl thioacetates, thiourea salts or thiophosphates. The free thiols can be isolated and then activated as thiolates under basic conditions, alternatively the anomeric deprotection and activation are simply achieved *in situ*.

Among more interesting and widely applicable methods to generate free thiols developed in the recent years is the work of Wan et al.[62] Their approach was inspired by native chemical ligation and allows selective anomeric deacetylation of the glycosyl thioacetate with DTT in the presence of basic conditions, addition of $NaHCO_3$ or Et_3NH. The mechanism of this approach is predicted to be similar to native chemical ligation in peptide synthesis and involves a reversible transthioesterification between a thioacetate and the thiol of DTT and subsequent irreversible S to N acyl shift (**Scheme 3.14**).

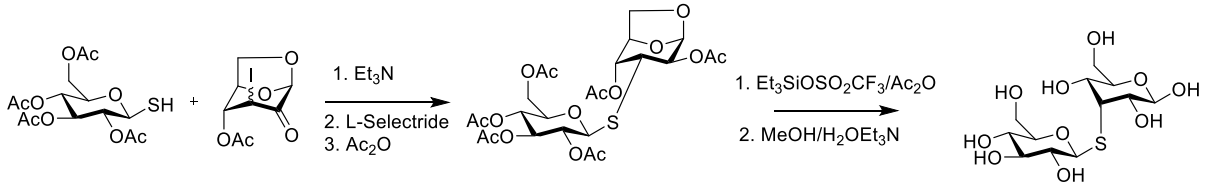


Scheme 3.14 Mechanism of anomeric deacetylation of glycosyl thioacetates inspired by native chemical ligation as established by Wan et al.[62]

Once formed the thiol can be activated to thiolate with a range of different bases, most commonly NaH, NaOMe, Et_3N and DIPEA are used (**Scheme 3.15**, **Scheme 3.16**).[62-68]

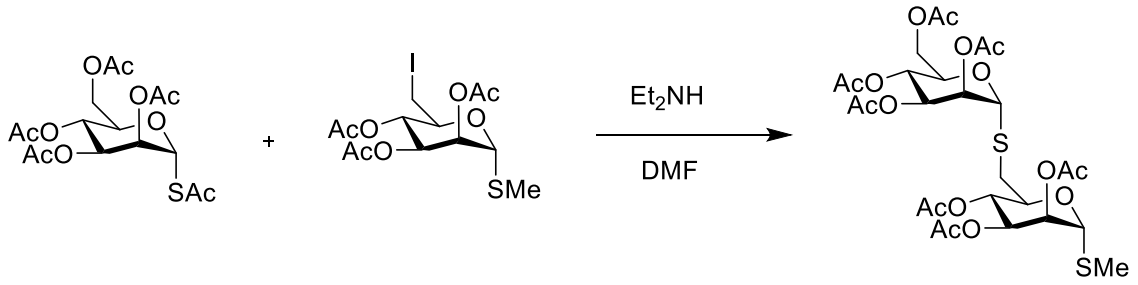


Scheme 3.15 Synthesis of thiodisaccharide by Horito et al. [66]

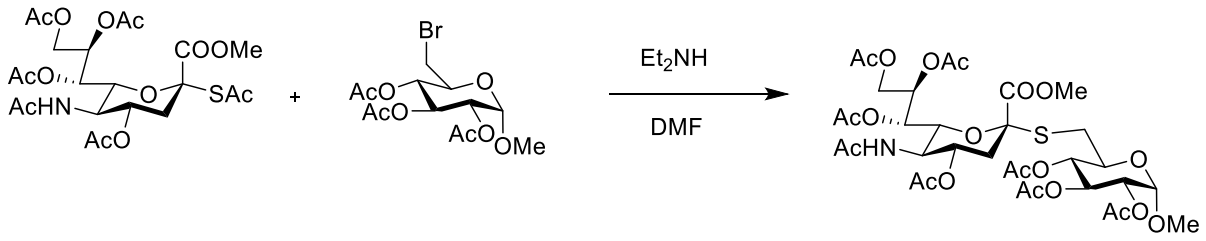


Scheme 3.16 Synthesis of thiodisaccharides by Witczak et al. [67]

Among methods for *in situ* formation of the thiol/thiolate, an elegant and widely useful strategy is deacetylation of glycosyl thioacetate with Et_2NH. This method has been successfully used several times in the past, examples include the work of von Itzstein's and Williams' groups, who successfully applied this method in the synthesis of *S*-linked α-1,6-oligomannosides[65] (**Scheme 3.17**) and *N*-acetylneuraminic acid based rotavirus inhibitors[63] (**Scheme 3.18**), respectively.



Scheme 3.17 Synthesis of *S*-linked α-1,6-oligomannosides by Williams at al.[65]



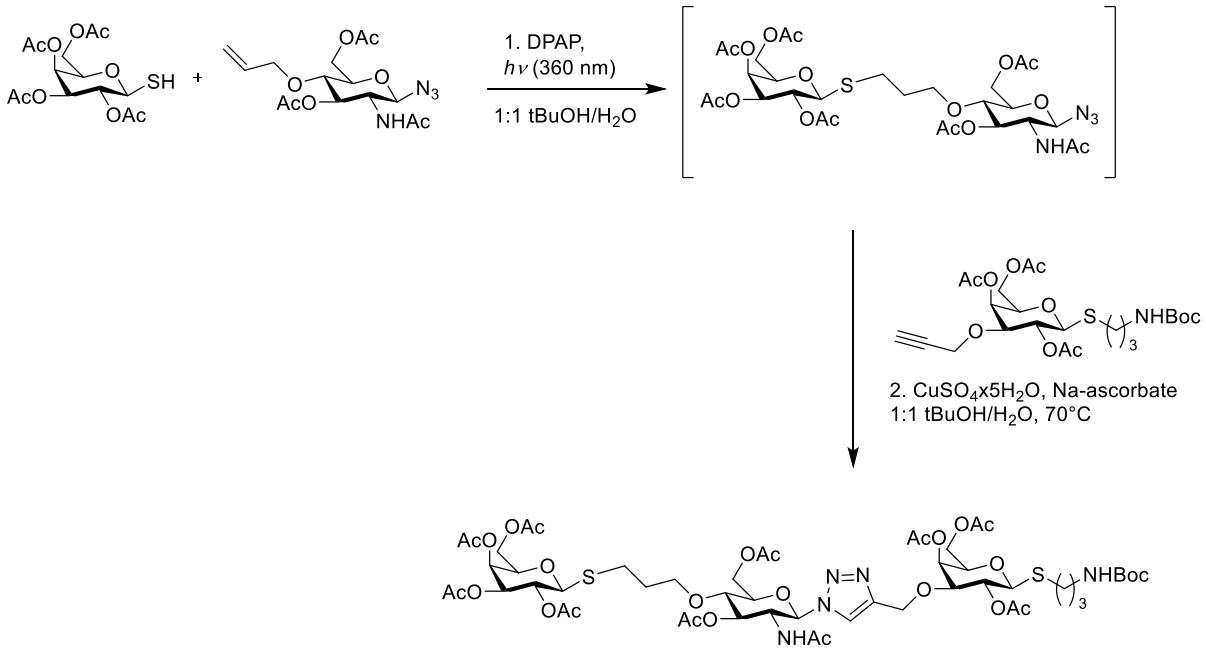
Scheme 3.18 Synthesis of *N*-acetylneuraminic acid containing thiodisaccharides by von Itzstein et al.[63]

Such *in situ* activation approach has also been utilized for the one-pot aziridine opening reaction described in Chapter Four.

3.5.3. Current developments

Similar to *O*-glycoside synthesis, solid-phase synthesis of thiooligosaccharides was described.[69] The approach involved attaching unprotected monosaccharides with an anomeric ethyl disulphide group to a solid support through their primary hydroxyl group, then activation through formation of the sodium thiolate and subsequent reaction with a saccharide triflate to give thiodisaccharides. Finally, the saccharides were cleaved from the resin with TFA. Thiooligosaccharide synthesis could be achieved by using electrophiles containing an anomeric ethyl disulphide.

Simplifications of synthesis of thiooligosaccharides towards one-pot processes are also under development. One of more recent methods was developed by Jensen et al. in 2019 (**Scheme 3.19**).[70] They successfully combined thiol-ene coupling and the copper-catalyzed alkyne-azide cycloaddition in a one-pot procedure to create mimics of trisaccharides and tetrasaccharides. The thiol-ene coupling between a glycosyl thiol and glycoside equipped with an alkene linker and anomeric azide was performed with the use of 2,2-dimethoxy-2-phenylacetophenone (DPAP) and utilizing a UV lamp (365 nm) for thiol-ene radical chemistry. Afterwards an alkyne equipped glycoside was added to an azide-containing product of the first step and a copper-catalyzed alkyne-azide cycloaddition was performed. The approach was also applicable to deprotected building blocks in aqueous media.



Scheme 3.19 One-pot synthesis of trisaccharides containing *S*-linkages developed by Jensen et al.[70]

Recently Kern and Pohl reported a method for automated solution-phase synthesis of *S*-glycosides.[71] Their methodology was applied to the synthesis of mannose-containing thiooligosaccharides. Automated oligosaccharide synthesis process is based on standard glycosylation reactions (e.g. trichloroacetimidate method), the donor and activator are added as solutions to an acceptor containing a tag or solid phase linker. The tag or linker can be used to facilitate the purification of intermediates and final product.

3.6. Conclusions and outlooks

Carbohydrates and glycoconjugates are among the most complex biomolecules, their synthesis is by no means trivial and routine. Some major problems associated with glycoside-bond formation have already been addressed and efficient strategies to combat with this complex task have been developed. Yet there is still more to discover to advance this field.

In the spirit of time and in parallel to other fields of biopolymer synthesis, developing new strategies towards automatic and routine synthesis of larger oligosaccharide complexes is particularly attractive. Yet for this to be feasible, there is still a lot of fundamental knowledge about glycosidic bond formation that needs to be unveiled. Particularly, mechanisms behind achieving complete stereocontrol in glycosylation reactions need to be better understood to facilitate the developments of new methodologies. With better understanding of underlying mechanistic principles, it might become easier to predict the outcomes of glycosylation reactions and develop new methods with a broad application range to various sugar and/or aglycone substrates.

Thus, despite many great advances in synthetic organic chemistry, chemical synthesis of oligosaccharides still remains a notable challenge for synthetic chemists.

3.7. References

1. A. Lipták, A. Borbás and I. Bajza, in *Comprehensive Glycoscience*, ed. H. Kamerling, Elsevier, Oxford, 2007, DOI: https://doi.org/10.1016/B978-044451967-2/00007-6, pp. 203-259.
2. T. K. Lindhorst, *Essentials of Carbohydrate Chemistry and Biochemistry*, Wiley, 2007.
3. A. V. Demchenko, in *Handbook of Chemical Glycosylation*, 2008, DOI: https://doi.org/10.1002/9783527621644.ch1, pp. 1-27.
4. R. U. Lemieux, K. B. Hendriks, R. V. Stick and K. James, *Journal of the American Chemical Society*, 1975, **97**, 4056-4062.
5. G. R. J. Thatcher, in *The Anomeric Effect and Associated Stereoelectronic Effects*, American Chemical Society, 1993, vol. 539, ch. 2, pp. 6-25.
6. R. Das and B. Mukhopadhyay, *ChemistryOpen*, 2016, **5**, 401-433.
7. C. P. Fei and T. H. Chan, *Tetrahedron Letters*, 1987, **28**, 849-852.
8. S. Houdier and P. J. A. Vottero, *Carbohydrate Research*, 1992, **232**, 349-352.
9. K. Fukase, I. Kinoshita, T. Kanoh, Y. Nakai, A. Hasuoka and S. Kusumoto, *Tetrahedron*, 1996, **52**, 3897-3904.
10. L. G. Green and S. V. Ley, in *Carbohydrates in Chemistry and Biology*, 2000, DOI: https://doi.org/10.1002/9783527618255.ch17, pp. 427-448.
11. Y. Nakahara and T. Ogawa, *Tetrahedron Letters*, 1987, **28**, 2731-2734.
12. D. R. Mootoo, P. Konradsson, U. Udodong and B. Fraser-Reid, *Journal of the American Chemical Society*, 1988, **110**, 5583-5584.
13. N. K. Kochetkov, E. M. Klimov, N. N. Malysheva and A. V. Demchenko, *Carbohydrate Research*, 1993, **242**, C7-C10.
14. N. K. Kochetkov, E. M. Klimov, N. N. Malysheva and A. V. Demchenko, *Carbohydrate Research*, 1991, **212**, 77-91.
15. D. Crich and S. Sun, *The Journal of Organic Chemistry*, 1996, **61**, 4506-4507.
16. Q. Chen and F. Kong, *Carbohydrate Research*, 1995, **272**, 149-157.
17. H. Paulsen, *Angewandte Chemie International Edition in English*, 1982, **21**, 155-173.
18. P. Sinaÿ, *Pure and Applied Chemistry*, 1978, **50**, 1437-1452.
19. R. Eby and C. Schuerch, *Carbohydrate Research*, 1974, **34**, 79-90.
20. G. Wulff and G. Röhle, *Angewandte Chemie International Edition in English*, 1974, **13**, 157-170.
21. A. Demchenko, T. Stauch and G.-J. Boons, *Synlett*, 1997, **1997**, 818-820.
22. G. H. Veeneman, S. H. van Leeuwen and J. H. van Boom, *Tetrahedron Letters*, 1990, **31**, 1331-1334.

23. G. H. Veeneman and J. H. van Boom, *Tetrahedron Letters*, 1990, **31**, 275-278.
24. R. R. Schmidt and X. Zhu, in *Glycoscience: Chemistry and Chemical Biology*, eds. B. O. Fraser-Reid, K. Tatsuta and J. Thiem, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, DOI: 10.1007/978-3-540-30429-6_11, pp. 451-524.
25. K. Fukase, A. Hasuoka, I. Kinoshita, Y. Aoki and S. Kusumoto, *Tetrahedron*, 1995, **51**, 4923-4932.
26. T. Mukaiyama and K. Matsubara, *Chemistry Letters*, 1992, **21**, 1041-1044.
27. F. Andersson, P. Fúgedi, P. J. Garegg and M. Nashed, *Tetrahedron Letters*, 1986, **27**, 3919-3922.
28. S. Manabe, Y. Ito and T. Ogawa, *Synlett*, 1998, **1998**, 628-630.
29. H. K. Chenault, A. Castro, L. F. Chafin and J. Yang, *The Journal of Organic Chemistry*, 1996, **61**, 5024-5031.
30. R. R. Schmidt and E. Rücker, *Tetrahedron Letters*, 1980, **21**, 1421-1424.
31. H. Dohi, Y. Nishida, H. Tanaka and K. Kobayashi, *Synlett*, 2001, **2001**, 1446-1448.
32. N. M. Spijker and C. A. A. van Boeckel, *Angewandte Chemie International Edition in English*, 1991, **30**, 180-183.
33. W. Koenigs and E. Knorr, *Berichte der deutschen chemischen Gesellschaft*, 1901, **34**, 957-981.
34. X. Zhu and R. R. Schmidt, *Angewandte Chemie International Edition*, 2009, **48**, 1900-1934.
35. K. Sasaki and K. Tohda, *Tetrahedron Letters*, 2018, **59**, 496-503.
36. E. S. H. E. Ashry, N. Rashed and E. S. I. Ibrahim, *Current Organic Synthesis*, 2005, **2**, 175-213.
37. Y. J. Lee, A. Ishiwata and Y. Ito, *Journal of the American Chemical Society*, 2008, **130**, 6330-6331.
38. A. J. Fairbanks, *Synlett*, 2003, **2003**, 1945-1958.
39. N. Nishi, J. Nashida, E. Kaji, D. Takahashi and K. Toshima, *Chemical Communications*, 2017, **53**, 3018-3021.
40. H. Nguyen, D. Zhu, X. Li and J. Zhu, *Angewandte Chemie International Edition*, 2016, **55**, 4767-4771.
41. K. S. Kim, J. H. Kim, Y. J. Lee, Y. J. Lee and J. Park, *Journal of the American Chemical Society*, 2001, **123**, 8477-8481.
42. Y. Hashimoto, S. Tanikawa, R. Saito and K. Sasaki, *Journal of the American Chemical Society*, 2016, **138**, 14840-14843.
43. Q. Li, S. M. Levi and E. N. Jacobsen, *Journal of the American Chemical Society*, 2020, **142**, 11865-11872.
44. C. S. Bennett, *Organic & Biomolecular Chemistry*, 2014, **12**, 1686-1698.

45. P. H. Seeberger, *Journal of Carbohydrate Chemistry*, 2002, **21**, 613-643.
46. M. Panza, D. Neupane, K. J. Stine and A. V. Demchenko, *Chemical Communications*, 2020, **56**, 10568-10571.
47. O. J. Plante, E. R. Palmacci and P. H. Seeberger, *Science*, 2001, **291**, 1523-1527.
48. R. Van Noorden, *Nature*, 2010, **466**, 1029-1029.
49. M. Panza, S. G. Pistorio, K. J. Stine and A. V. Demchenko, *Chem Rev*, 2018, **118**, 8105-8150.
50. G.-J. Boons, *Contemporary Organic Synthesis*, 1996, **3**, 173-200.
51. J. D. C. Codée, R. E. J. N. Litjens, L. J. van den Bos, H. S. Overkleeft and G. A. van der Marel, *Chemical Society Reviews*, 2005, **34**, 769-782.
52. J. Sun, J. Fang, X. Xiao, L. Cai, X. Zhao, J. Zeng and Q. Wan, *Organic & Biomolecular Chemistry*, 2020, **18**, 3818-3822.
53. T. Polat and C.-H. Wong, *Journal of the American Chemical Society*, 2007, **129**, 12795-12800.
54. S. Oscarson, in *Glycoscience: Chemistry and Chemical Biology*, eds. B. O. Fraser-Reid, K. Tatsuta and J. Thiem, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, DOI: 10.1007/978-3-540-30429-6_14, pp. 661-697.
55. K. Pachamuthu and R. R. Schmidt, *Chem Rev*, 2006, **106**, 160-187.
56. S. Mehta, J. S. Andrews, B. D. Johnston and B. M. Pinto, *Journal of the American Chemical Society*, 1994, **116**, 1569-1570.
57. S. Mehta, J. S. Andrews, B. Svensson and B. M. Pinto, *Journal of the American Chemical Society*, 1995, **117**, 9783-9790.
58. J. S. Andrews and B. Mario Pinto, *Carbohydrate Research*, 1995, **270**, 51-62.
59. J. S. Andrews, B. D. Johnston and B. M. Pinto, *Carbohydrate Research*, 1998, **310**, 27-33.
60. L.-X. Wang, N. Sakairi and H. Kuzuhara, *Journal of the Chemical Society, Perkin Transactions 1*, 1990, DOI: 10.1039/P19900001677, 1677-1682.
61. M. Blanc-Muesser and H. Driguez, *Journal of the Chemical Society, Perkin Transactions 1*, 1988, DOI: 10.1039/P19880003345, 3345-3351.
62. P. Shu, J. Zeng, J. Tao, Y. Zhao, G. Yao and Q. Wan, *Green Chemistry*, 2015, **17**, 2545-2551.
63. M. J. Kiefel, B. Beisner, S. Bennett, I. D. Holmes and M. von Itzstein, *Journal of Medicinal Chemistry*, 1996, **39**, 1314-1320.
64. H. N. Yu, C.-C. Ling and D. R. Bundle, *Journal of the Chemical Society, Perkin Transactions 1*, 2001, DOI: 10.1039/B009626L, 832-837.
65. T. Belz and S. J. Williams, *Carbohydrate Research*, 2016, **429**, 38-47.
66. H. Hashimoto, K. Shimada and S. Horito, *Tetrahedron: Asymmetry*, 1994, **5**, 2351-2366.

67. Z. J. Witczak, P. Kaplon and M. Kolodziej, in *Timely Research Perspectives in Carbohydrate Chemistry*, eds. W. Schmid and A. E. Stütz, Springer Vienna, Vienna, 2002, DOI: 10.1007/978-3-7091-6130-2_11, pp. 171-180.
68. L. L. Morais, K. Bennis, I. Ripoche, L. Liao, F.-I. Auzanneau and J. Gelas, *Carbohydrate Research*, 2003, **338**, 1369-1379.
69. G. Hummel and O. Hindsgaul, *Angewandte Chemie (International ed. in English)*, 1999, **38**, 1782-1784.
70. A. Brinkø, C. Risinger, A. Lambert, O. Blixt, C. Grandjean and H. H. Jensen, *Organic Letters*, 2019, **21**, 7544-7548.
71. M. K. Kern and N. L. B. Pohl, *Organic letters*, 2020, **22**, 4156-4159.

CHAPTER FOUR

Synthesis of *thio*-glycosides *via* one-pot aziridine opening reactions

4.1. Introduction

The Bernardi's group has been focusing on design and synthesis of glycomimetics to target the C-type lectin DC-SIGN (Dendritic Cell-Specific ICAM-3-Grabbing Non-integrin). DC-SIGN is mainly expressed and displayed on the surface of immature dendritic cells (DCs) on dermal and mucosal tissues. It plays a key role in the immune system, it mediates the interaction between dendritic cells and T-cells by binding to glycoproteins Inter Cellular Adhesion Molecules ICAM-2 and ICAM-3.[1, 2] DC-SIGN is able to recognize several pathogens, including viruses (Hepatitis C, Ebola, Dengue, SARS), bacteria (*Mycobacterium tuberculosis*, *Klebsiella pneumonae*), yeasts (*Candida albicans*) and parasites (*Leishmania spp*, *Schistosoma mansoni*) by recognizing specifically the glycosylated structures on their surface. Some pathogens however, hijack DCs to disseminate infections in the human body. A well-known example is Human Immunodeficiency Virus type-1 (HIV-1) which exploits DC-SIGN for its entry pathway.[3] The viral entry occurs through interactions between the HIV's highly mannosylated envelope protein (gp120) with DC-SIGN.[4] The main natural ligand that is recognized by the carbohydrate binding site of DC-SIGN is the high-mannose glycan $(Man)_9(GlcNAc)_2$ and the minimum epitope required for the recognition is the dimannoside (Manα(1,2)Man), shown in **Fig. 4.1**.[5]

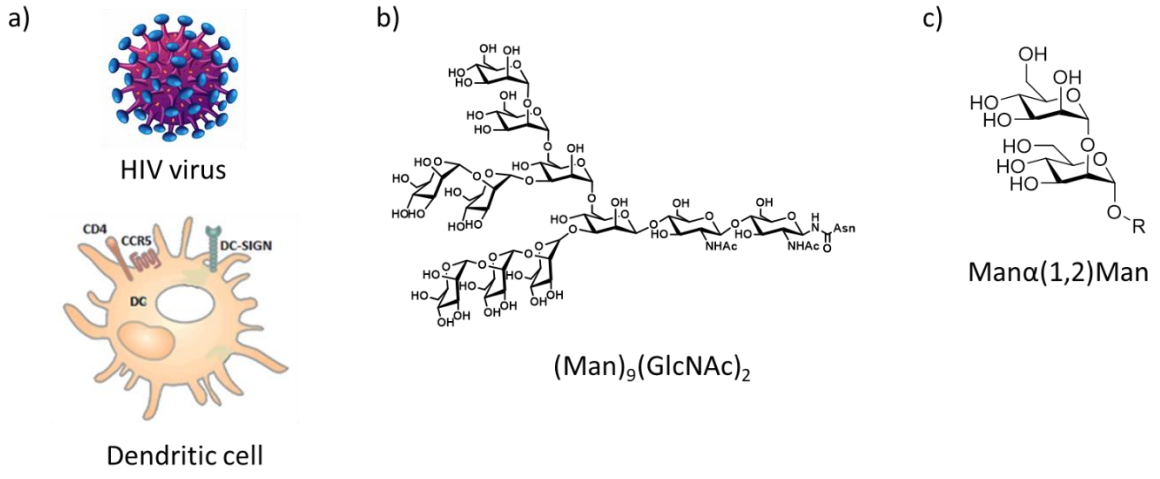


Fig. 4.1 a) HIV virus and dendritic cell displaying DC-SIGN, b) High-mannose glycan $Man_9(GlcNAc)_2$, c) Minimum binding epitope dimannoside Manα(1,2)Man

Because of its role in infections with different pathogens, DC-SIGN has become an interesting target for the design of anti-infective agents. Particularly, a lot of effort was put towards making a molecule that would mimic structural features of the binding epitope Manα(1,2)Man (**Fig. 4.2**), but would have an improved enzymatic and hydrolytic stability. Pseudo-dimannoside **4.1** (**Fig 4.2**) was a promising candidate, because it is significantly more resistant to enzymatic hydrolysis (with jack-bean mannosidase) than the natural dimannoside.[6] Compound **4.1** consists of a mannose moiety that is involved in interaction with a calcium ion in the carbohydrate binding site, and a conformationally stable 1,2-*trans*-diaxial cyclohexanediol moiety that is able to establish additional van der Waals interactions with amino acid residues in the binding site. The cyclohexanediol moiety mimics the reducing end mannose of the dimannoside and greatly contributes to the enzymatic stability of the molecule.

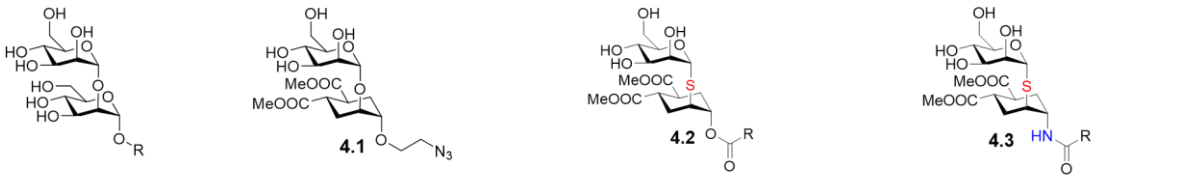


Fig. 4.2 DC-SIGN ligands developed by the Bernardi's group

However, as it is often for many studies utilizing glycomimetics, the preparation of pseudo-dimannoside **4.1** had some limits (**Fig. 4.3**). In particular, the synthesis of glycosyl donor is quite lengthy, the glycosylation reaction between donor and acceptor (**Fig. 4.3**) required anhydrous conditions and low temperatures, additionally the azido linker had to be introduced at a specific point in the synthetic pathway and could not be easily modified. Therefore, new strategies were investigated to simplify the synthesis of glycomimetic DC-SIGN ligands.

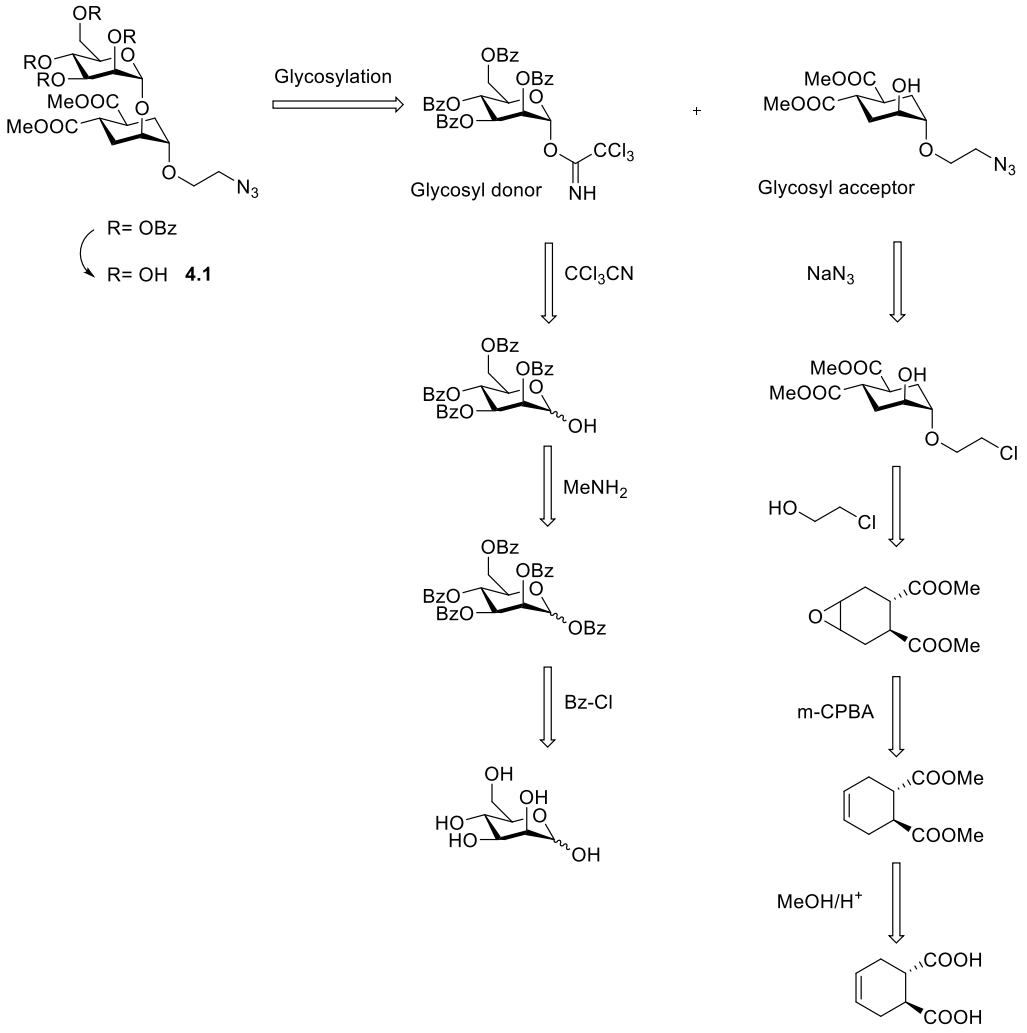


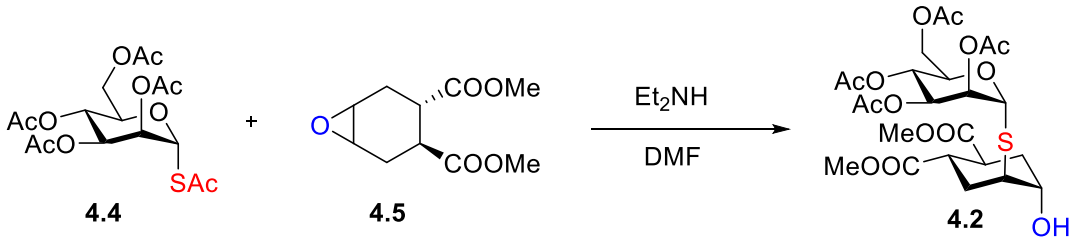
Fig 4.3 Retrosynthesis of the pseudo-dimannoside **4.1**

Thioglycosides were used to simplify the synthesis on the rationale that oxygen and sulphur share similar bonding and geometries, but the lower basicity and better nucleophilicity of the sulphur atom simplifies the synthesis and improves stability of the *thio*-glycomimetics towards acid-catalyzed and enzymatic hydrolysis.

Indeed, the synthesis of pseudo-*thio*-dimannoside analog **4.2** (**Fig 4.2**) could be simplified by using a one-pot synthetic approach. The idea for this approach was inspired by the work of Belz et al.[7] where Et_2NH was used to deacetylate the anomeric sulphur of a mannosyl thioacetate, generating a thiolate able to displace a sugar halide. This approach allowed for the synthesis of a family of *S*-linked α-1,6-oligomannosides.

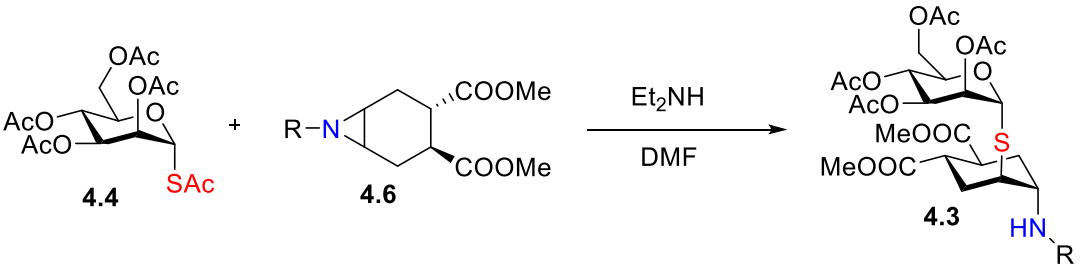
From this starting point, the Bernardi's group devised an approach in which tetra-*O*-acetylmannosyl-α-thioacetate **4.4** is deacetylated *in situ* on the anomeric position by excess amount of Et_2NH. This generates a thiolate able to open an epoxide **4.5**. The product of the reaction (pseudo-*thio*-

dimannoside **4.2**) is formed as a single compound from a selective *trans*-diaxial epoxide opening (**Scheme 4.1**). The reaction is carried out as a one-pot procedure at room temperature and is thus facile and operationally much simpler than classic glycosylation methods. The pseudo-*thio*-dimannoside **4.2** has a comparable affinity towards DC-SIGN as the pseudo-dimannoside **4.1** and an improved enzymatic stability (against jack-bean mannosidase).[8, 9]



Scheme 4.1 One-pot epoxide opening approach towards the synthesis of pseudo-*thio*-dimannoside **4.2**

To further improve the hydrolytic stability of the *thio*-glycomimetic **4.2** and introduce a site for easy conjugation, an *N*-linkage was introduced by opening of an aziridine ring **4.6** instead of the epoxide. Such approach gave metabolically stable DC-SIGN ligands with the structure of *N*-linked-pseudo-*thio*-dimannoside **4.3**.[9] (**Scheme 4.2**)



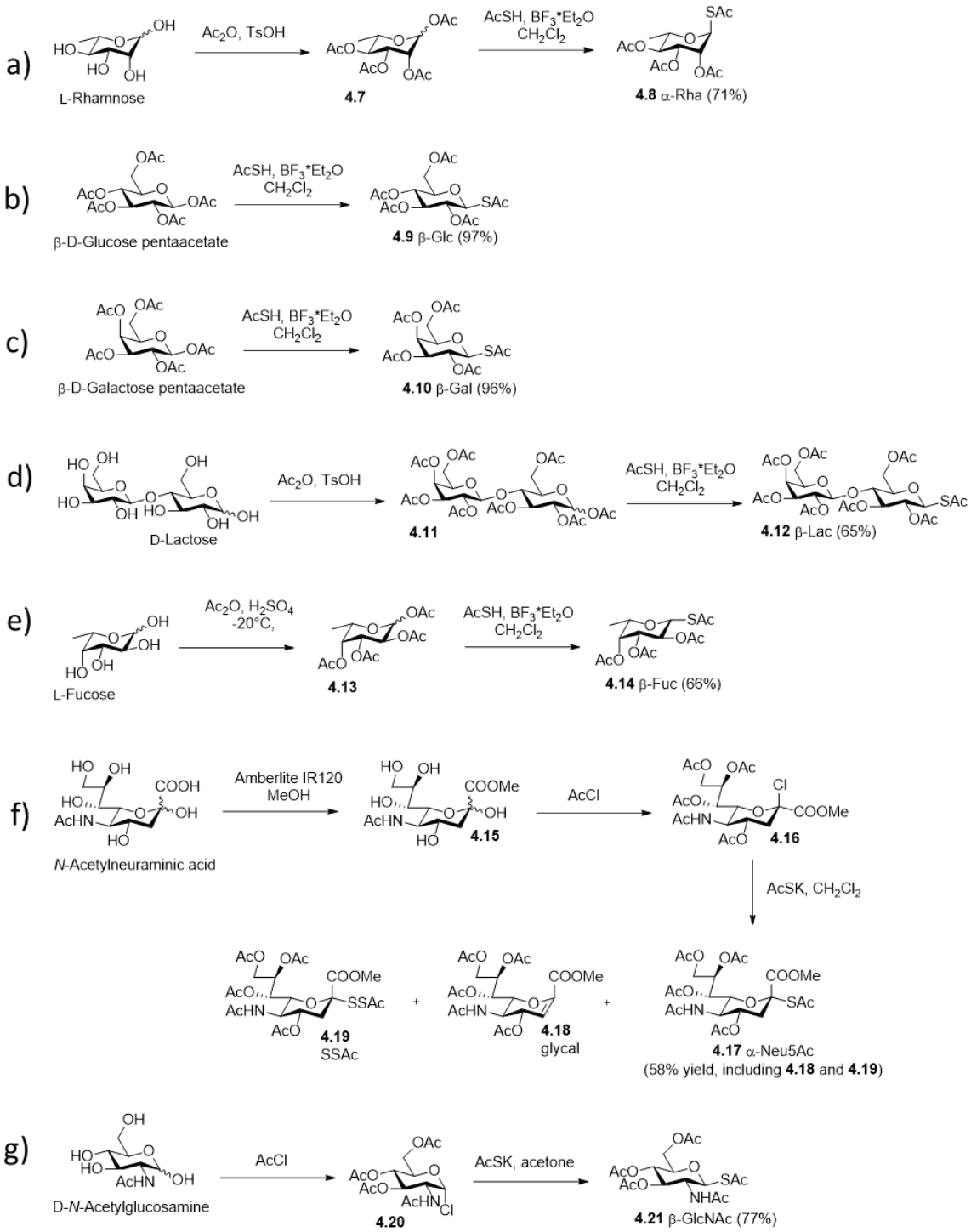
Scheme 4.2 One-pot aziridine opening approach towards the synthesis of *N*-linked-pseudo-*thio*-dimannoside **4.3**

The utility and operational simplicity of such strategy inspired us to further investigate the aziridine opening approach on a series of mono- and disaccharides, creating a new class of glycomimetics with the general structure of *N*-linked-pseudo-*thio*-glycosides.

4.2. Synthesis of peracetylated glycosyl *thio*-acetates

The peracetylated glycosyl *thio*-acetates (**Scheme 4.3**, entries **a-g**) were used to generate the nucleophile in the aziridine opening reaction. Various methods for their synthesis have been reported. They can be easily synthesized from free sugars either *via* peracetylation and subsequent treatment

with AcSH in a Lewis acid catalysed process (**Scheme 4.3**, entries **a-e**) or through S_N2 displacement of sugar halides by AcSK (**Scheme 4.3**, entries **f**, **g**).[10] To simplify the synthesis and at the same time achieve satisfying yields, we selected one-pot procedures with conditions tailored for each sugar individually. The conditions used are summarized in **Scheme 4.3** and described in detail in the experimental section.



Scheme 4.3 Summary of conditions for the synthesis of peracetylated glycosyl *thio*-acetates used in this thesis

Traditionally per-*O*-acetyl glycosides are prepared with the use of excess amount of pyridine, however basic conditions are not compatible with the one-pot procedure, because thioacetylation is an acid-catalyzed process. Mong et al.[11] developed a versatile carbohydrate acetylation protocol that uses common sulfonic acids and can be paired with a subsequent thioglycosidation. We adopted their approach for the synthesis of 1-*S*-acetyl derivatives of rhamnose (Rha) **4.8**, glucose (Glc) **4.9**, galactose (Gal) **4.10**, lactose (Lac) **4.12** and fucose (Fuc) **4.14** (**Scheme 4.3**, entries **a-e**). They were obtained by acid catalysed acetylation of the free sugar, followed by reaction with thioacetic acid (AcSH) under BF_3 catalysis (two steps, one pot).

To determine the configuration of the products, we used NMR analysis. A general rule states that the coupling constant in ^{1}H NMR spectra is larger if the protons are in axial-axial orientation ($J_{1,2}$=8-13 Hz), than if they are axial-equatorial ($J_{1,2}$=1-6 Hz) or equatorial-equatorial ($J_{1,2}$=0-5 Hz). This rule helps determining the configuration of derivatives of Glu **4.9**, Gal **4.10**, Lac **4.12**, Fuc **4.14** and GlcNAc **4.21**. Determining the configuration of rhamnosyl thioacetate **4.8** was not as straightforward, because the configuration of the C_2 implies an equatorial H_2 proton and thus a small and not distinct value of $J_{1,2}$ in either configuration.

The peracetylated *thio*-derivatives of Glc **4.9**, Gal **4.10**, Lac **4.12** and GlcNAc **4.21** were obtained as single β-isomers and characterized by means of the $J_{1,2}$ value in their ^{1}H-NMR spectrum, which is cca. 10 Hz.

The *thio*-acetate of Rha **4.8** was obtained as a 85/15 mixture of isomers were chromatographically separated on a Biotage HP column using 100% iPr_2O.

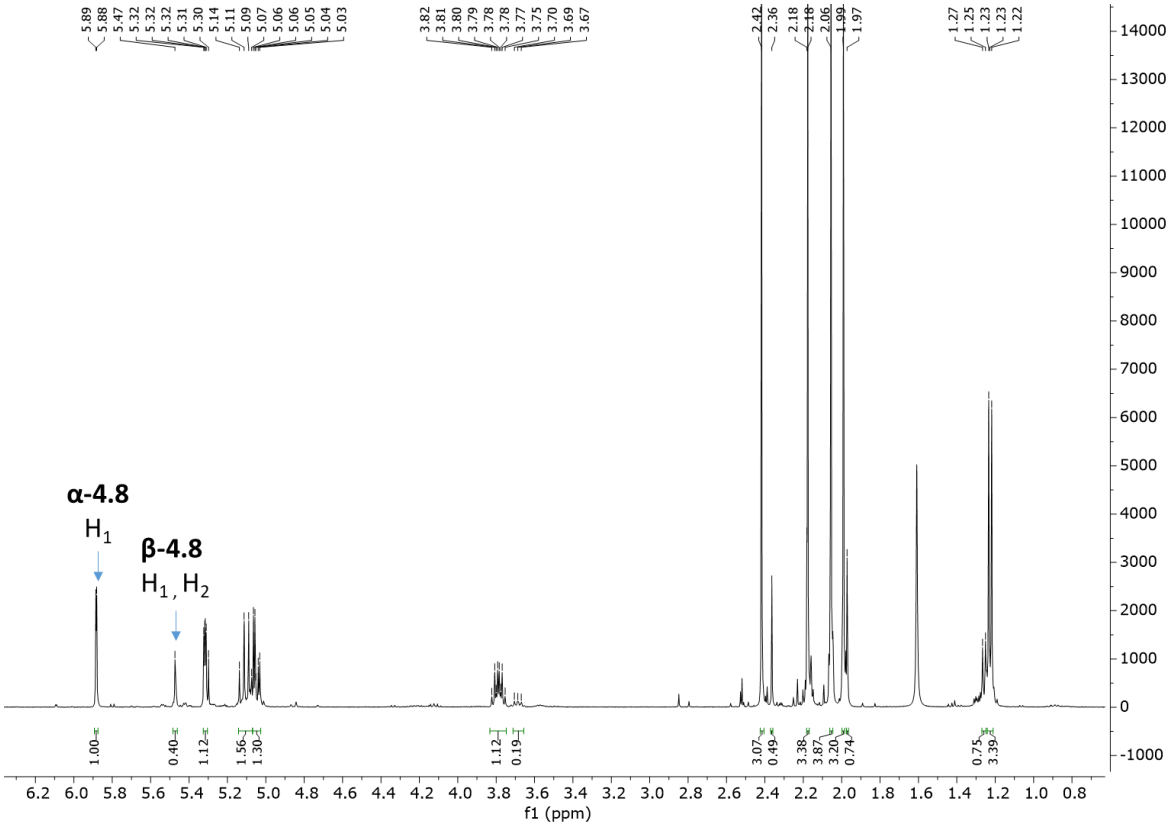


Fig. 4.4 ^{1}H NMR spectra of Rha derivative **4.8** showing the mixture of the two anomers **α-4.8** and **β-4.8** (400 MHz NMR, $CDCl_3$)

The major isomer **4.8** was assigned as the α-anomer by NMR NOESY experiments showing a cross-peak between the signals of H_3 (5.10 ppm) and H_5 (3.79 ppm) and the absence of cross peak between H_1 and both H_3 and H_5. On the contrary, the NOESY spectrum of the minor isomer **β-4.8** clearly displays cross peaks between H_1 (5.50 ppm), H_3 (5.10 ppm) and H_5 (3.75 ppm). The traces of the NOESY spectra are shown in **Fig. 4.5**. The **α-4.8** product was erroneously assigned as the β-anomer in ref [12].

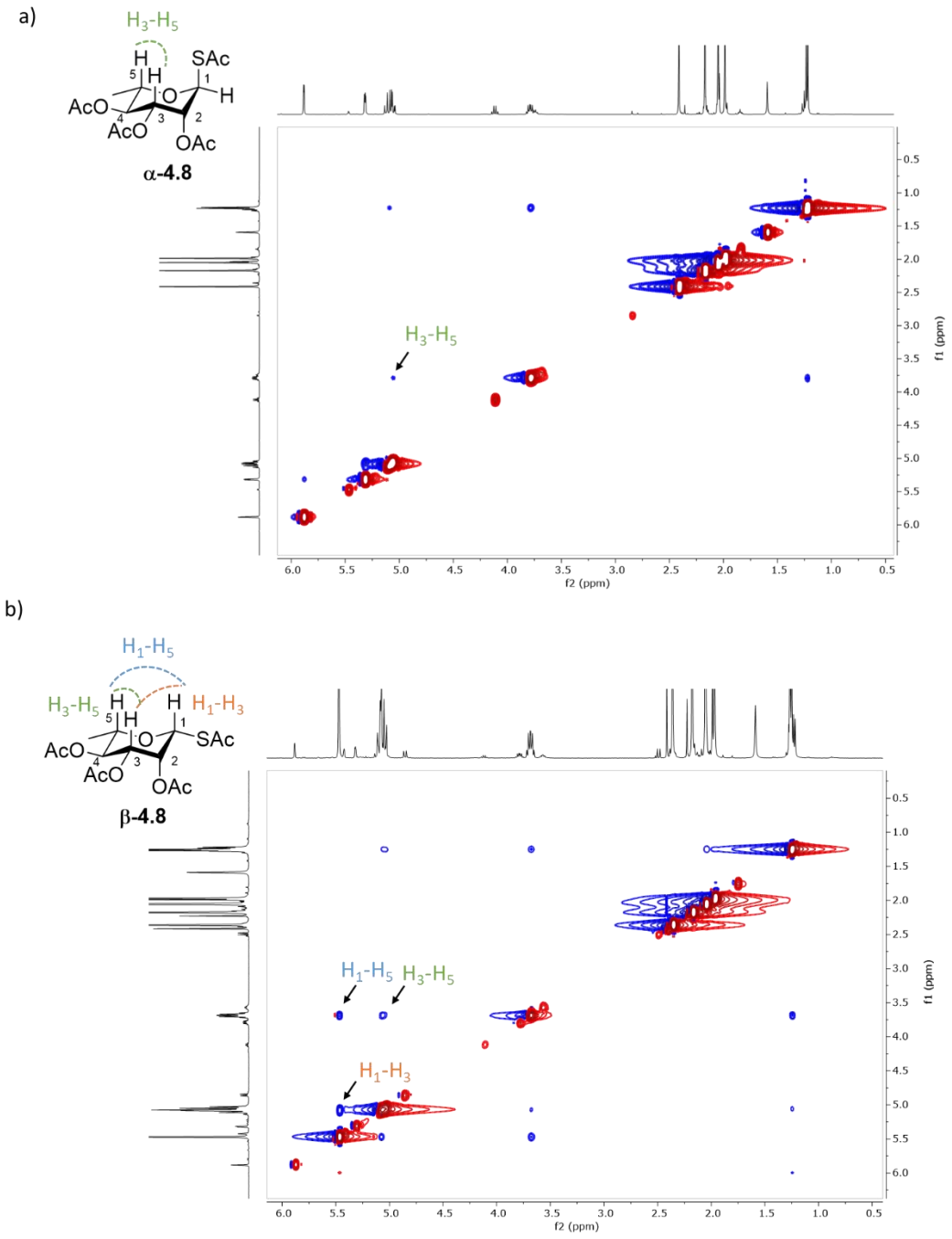


Fig. 4.5 NOESY NMR characterization of the Rha isomers a) **α-4.8** and b) **β-4.8** (400 MHz NMR, $CDCl_3$)

The acetylation of L-fucose required special attention as this sugar is known to form furanosyl acetates besides the desired pyranose acetates. Mong et al.[11] noticed that such furanosyl isomers were also formed with their standard TsOH-catalyzed acetylation protocol (26% of the total peracetyl acetates), therefore the reaction must be carried out at low temperatures (-20°C). At such low temperature a more reactive H_2SO_4 was needed to perform the reaction. Complete removal of the solvent after the acetylation was described to lead to undesired dehydration. This can be solved by quenching the

remaining Ac_2O with 1.2 eq of MeOH and then the following thioacetylation can be achieved using AcSH and BF_3 catalysis.

Nevertheless, the final thioacetylated product **4.14** was obtained as a complex mixture with traces of furanose product. The purification was attempted with column chromatography (Hex:EtOAc 3:1), however the product was not obtained as a completely pure compound **β-4.14** (**Fig. 4.6**). Although some other glycosyl thioacetates (Glc, Rha) could be successfully purified by recrystallization (Hex:EtOAc) on small scales, we could not find suitable conditions for recrystallization of fucose product **4.14**. Thus, we obtained the final product as a mixture of compounds with the ratio product/impurity of approximately 88/12 (**Fig. 4.7**). Although not entirely pure, we were able to use the product **4.14** in the following aziridine opening to generate an idea on how this fucose derivative behaves in the reaction.

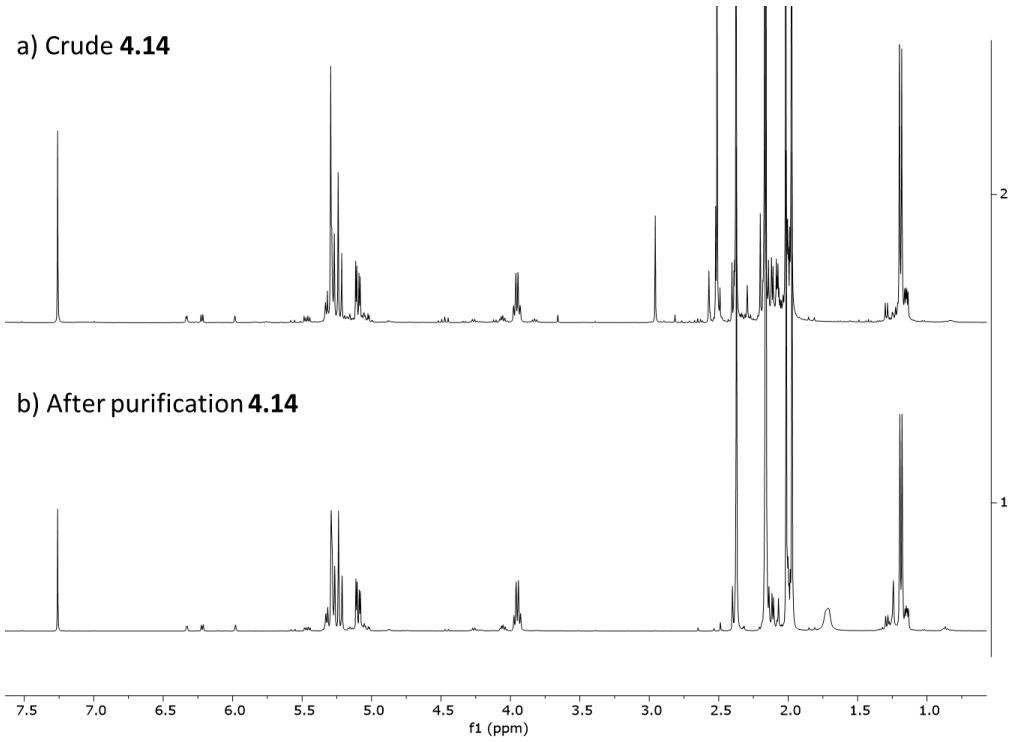


Fig. 4.6 ^{1}H NMR spectra of Fucose thioacetylated product **4.14** before and after purification with column chromatography (400 MHz NMR, $CDCl_3$)

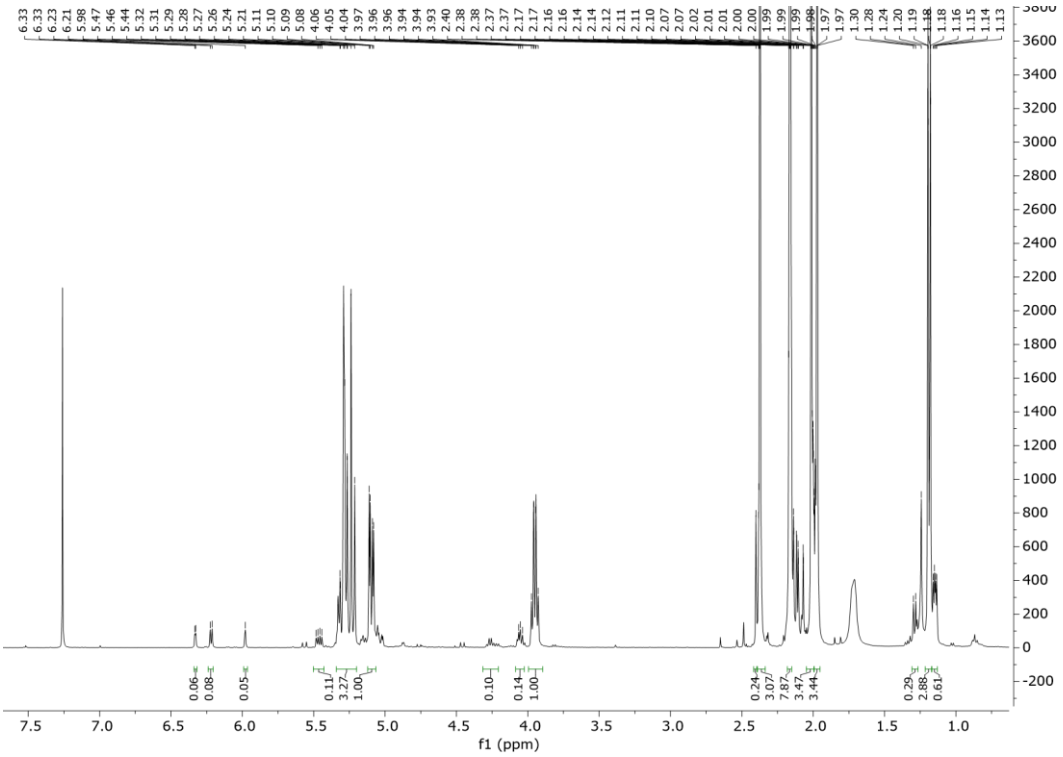


Fig. 4.7 Fucose thioacetylated product **4.14** after purification

The 1-*S*-acetyl derivatives of sialic acid (Neu5Ac) **4.17** and *N*-acetylglucosamine (GlcNAc) **4.21** were synthesized by reaction of AcSK with the corresponding peracetylated anomeric chloride, generated *in situ* by reaction of the free sugar with AcCl (**Scheme 4.3**, entries **f**, **g**)

The thioacetylation of Neu5Ac by acid catalysed substitution of the anomeric acetate is expected to give the β-anomer of Neu5Ac thioacetate.[11] Therefore, to obtain the α-isomer we used as the intermediate the chloride **4.16** (**Scheme 4.3,** entry **f**). It is however known that this procedure gives an almost inseparable mixture of 3 compounds the thioacetate **4.17**, the glycal **4.18** and the bisthioacetate **4.19** in variable ratios, presumably depending on the batch of KSAc (**Fig. 4.8**).[13]

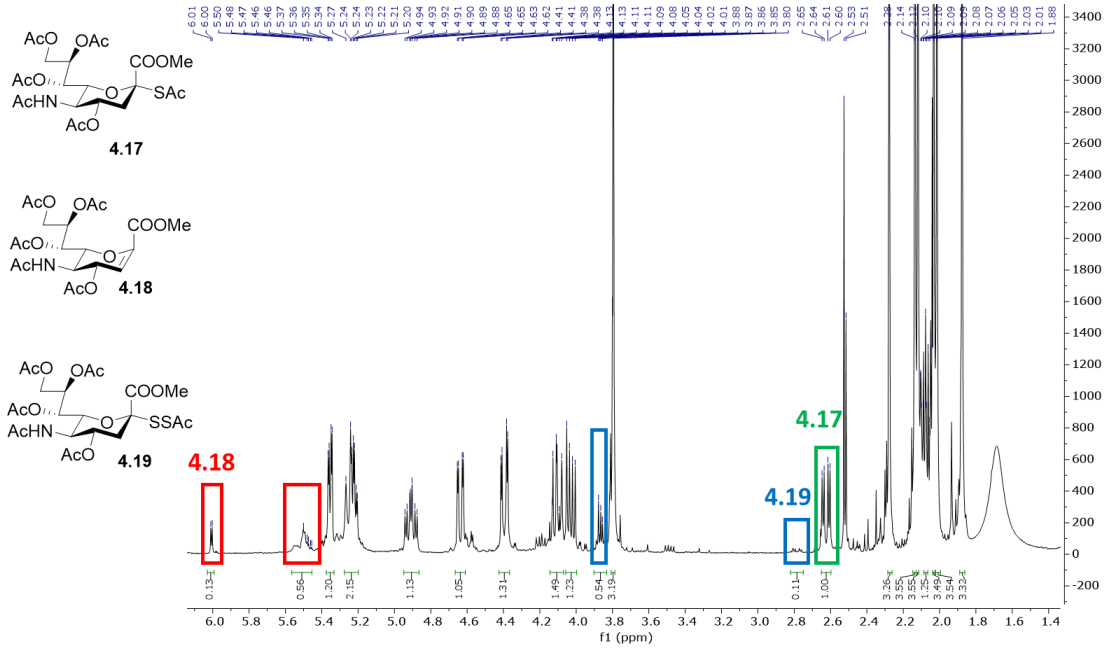


Fig. 4.8 NMR spectra of a mixture of Neu5Ac product α-thioacetate **4.17** (highlighted in green, signal dd at 2.60 ppm H_{3eq}), bisthioacetate **4.19** (highlighted in blue, signal dd at 2.80 ppm H_{3eq}), glycal **4.18** (highlighted in red, signal d at 6.00 ppm =CH) (400 MHz NMR, $CDCl_3$)

The synthetic route of peracetylation and subsequent thioacetylation described above could not be applied to *N*-acetylglucosamine, due to the formation of a stable oxazoline derivative **4.22** (**Fig. 4.9**), that is formed upon acid treatment of tetra-*O*-acetyl-*N*-acetyl-glucosamine and consequently prevents the formation of the thioacetate. One of the possibilities to prevent the oxazoline from forming is to use specific protecting groups on the amino group of glucosamine, such as 2,2,2-trichloroethoxycarbonyl (Troc) group. However, to avoid additional steps of deprotection and subsequent acetylation of the amine, it was simpler to operate through the intermediate anomeric chloride **4.20**. (**Scheme 4.3**, entry **g**)

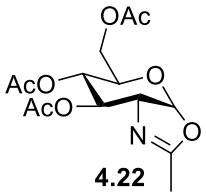


Fig. 4.9 Oxazoline derivative **4.22**

The intermediate chloride **4.20** is formed from GlcNAc using an excess of AcCl in the absence of base. While free OH groups of GlcNAc are acetylated by AcCl, the HCl formed in the process allows chlorination of the anomeric carbon. Understanding of the mechanism is crucial for a proper set up of the reaction: to keep the gaseous HCl in contact with the solution of GlcNAc in AcCl, the reaction

glassware must be well sealed and vigorous stirring or shaking is necessary to facilitate the consumption of the starting material.

4.3. Synthesis of the aziridine

Aziridines are the simplest structures from the class of three-membered saturated nitrogen-containing heterocycles. They are highly reactive due to the bond strain caused by the geometric constraints of the three-membered ring.

The first example of a successful synthesis of a free aziridine was reported by Wenker in 1935.[14] He described the cyclization of amino alcohols, a procedure that gave a free aziridine in low yields.

Afterwards many new strategies to deal with the task of aziridine synthesis were developed and reviewed.[15, 16] Generally, aziridines are obtained through 3 classical approaches:

- Addition of a nitrogen moiety on C=C bond of olefins
- Addition of a carbon moiety on C=N bond
- Intramolecular cyclization of amino derivatives (1,2-aminohalides, 1,2-azidoalcohols).

Among these strategies, aziridination of olefins is the most popular and therefore also the most studied approach. The most used nitrogen-transfer reagents include azides, iminoiodinanes, *N*-aminophthalimide, sulfonamides, sulfonimidamides, sulfamates, and sulfonyloxycarbamates.

The first method for the stereospecific synthesis of unprotected N-H and N-Me aziridines directly from unactivated olefins was reported in 2014 by Falck et al.[17] Falck's N–H aziridination occurs *via* homogenous Rhodium catalysis ($Rh_2(esp)_2$, Du Bois' catalyst **4.26**, **Scheme 4.5**) and uses *O*-(2,4-dinitrophenyl)hydroxylamine (DPH **4.27**, **Scheme 4.5**) as the aminating agent, with no external oxidants. The method was applied on a number of substrates, producing aziridines in good to excellent yields. A plausible Rh-nitrene pathway mechanism was proposed based on quantum mechanics calculations.

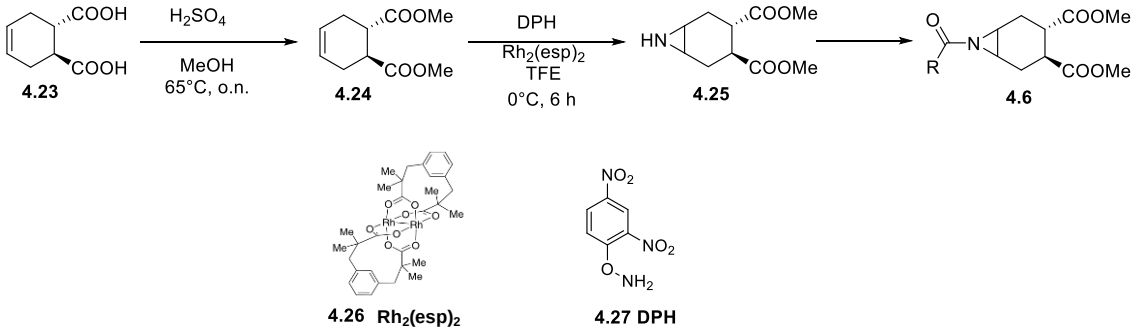
In 2017 Kürti et al. aimed to further optimize the aziridination of unactivated olefins. Although the Falck's aziridination is operationally simple, uses mild conditions and produces good yields, the use of DPH comes with certain drawbacks. In particular, the presence of NO_2 groups in DPH and phenoxide by-product are problematic for working on larger scale (e.g. industrial use). The solution was to employ hydroxylamine-*O*-sulfonic acid (HOSA) to replace DPH in the aziridination. HOSA does not contain a NO_2 group, it is thermally stable, commercially available and gives only an inorganic sulfate by-product that can be easily removed by water extraction. The TFE solvent used by Falck was replaced with a

more acidic 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), that together with the addition of a base and $Rh_2(esp)_2$ catalyst allowed for short reaction time of the aziridination.[18]

Direct preparation of aziridines from olefins could also be achieved by using *O*-(sulfonyl)hydroxylamines as aminating agent in the presence of Du Bois' catalyst in TFE. This procedure also allows NO_2 and additive free aziridination that yields only water soluble by-products.[19]

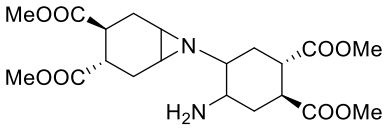
The aziridine **4.25** required in our project was prepared as a single enantiomer by N–H aziridination of (1*S*,2*S*)-1,2-dicarbomethoxy-4-cyclohexene **4.24** using the method introduced by Falck and co-workers.[17]

Upon aziridination the N–H aziridine **4.25** can be transformed in a variety of amides and carbamates with general formula **4.6**. (**Scheme 4.5**)



Scheme 4.5 Aziridination of an olefin **4.24** *via* homogenous rhodium catalysis

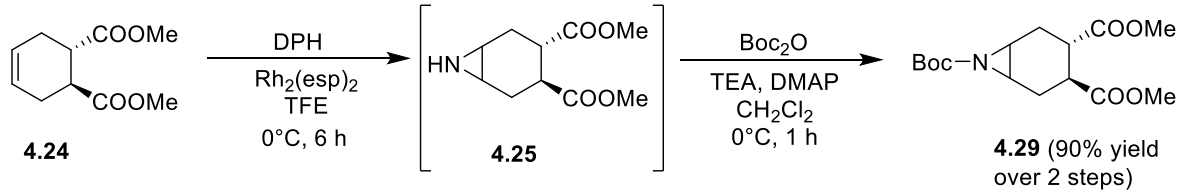
Falck's N–H aziridination method was applied on a number of substrates, including unsubstituted cyclohexene that is structurally similar to our substrate. The cyclohexene derived aziridine was obtained in 71% yield, after 3 h at room temperature (25°C).[17] We found Falck's procedure for aziridination operationally simple, fast and scalable on olefin **4.24** but, careful optimization of the reaction conditions was required to avoid dimerization of **4.25**, a process which occurs rapidly at room temperature to afford **4.28** (**Fig. 4.10**). Previously our group has optimized the procedure for aziridination of **4.24**, proving that the key factor to avoid dimerization is the reaction temperature. On a 50 mg (0.25 mmol) scale the best yield (68%) was obtained under following conditions: 10°C, 8 h, addition of 1 mol% of $Rh_2(esp)_2$ at halftime of the reaction.[9]



Dimeric by-product **4.28**

Fig. 4.10 Dimeric by-product **4.28**

While it is possible to obtain a free aziridine after a fast and careful work up and chromatographic purification, free aziridines are highly unstable. *N*-aceylation directly on the crude product gave better yields. Although a number of *N*-acetyl groups were previously explored, Boc protection on the aziridine was used in this work, because it allows easy deprotection and further functionalization of aziridine opening products. Boc protection was achieved using di-*t*-butyl-dicarbonate and adapting the conditions reported by Mordini and co-workers for a similar substrate.[20]



Scheme 4.6 Aziridination of olefin **4.24** and subsequent Boc protection of free aziridine to form **4.29**

Under the already established conditions (10-15°C, 8 h, 0.1 M, addition of 1 mol% $Rh_2(esp)_2$) the yield obtained for synthesis of **4.29** on 300 mg scale was 73% after Boc protection.

Although the reaction yield was satisfactory, we aimed to further optimize the reaction conditions. Scale up of the reaction resulted in more dimer being produced, additionally we aimed to shorten the reaction time, to make the process of aziridination and immediate protection of the free aziridine operationally more feasible. The overall yield of the aziridination depends on the consumption of the starting material **4.24** and formation of the dimer **4.28**. Parameters we could change were temperature, time, concentration of the olefin substrate and addition of the reagents. Increasing the concentration of the olefin from 0.1 M to 0.15 M resulted in faster consumption of the starting olefin **4.24**, which reduced the reaction time from 8 h to 5.5 h. However, the more concentrated reaction mixture slightly increased the amount of dimer **4.28** formation (roughly the product:dimer ratio changed from 6:1 to 4.5:1). Therefore, the temperature was lowered from 10°C to 0°C and to guarantee a good conversion of the starting material 20 mol% of DPH together with 1 mol% $Rh_2(esp)_2$ was added at the halftime of the reaction. Such conditions resulted in an improved product:dimer ratio of approximately 8:1 and a 67:1 ratio of product: starting material on 200 mg scale in 6 h. (**Table 4.1**)

This careful tuning of the reaction conditions finally gave a Boc-protected aziridine **4.29** in 90% yield on 200 mg scale.

Table 4.1 Optimization of the aziridination conditions for synthesis of **4.25**

Entry	T (°C)	Time	Conc. of **4.24** (M)	Add. reagents	Ratio (**4.25**:**4.24**)	Ratio (**4.25**:**4.28**)	Yield after Boc potection (**4.29**) (%)	Scale (mg)
1	10-15	8 h	0.1	1 mol% $Rh_2(esp)_2$	22.2 : 1	6.3 : 1	63	100
2	10-15	8 h	0.1	1 mol% $Rh_2(esp)_2$	20 : 1	5.3 : 1	73	300
3	10-15	5.5 h	0.15	1 mol% $Rh_2(esp)_2$	22.2 : 1	4.5 : 1	67	50
4	0	6 h	0.15	1 mol% $Rh_2(esp)_2$ + 20 mol % DPH	40 : 1	20 : 1	84	50
5	0	6 h	0.15	1 mol% $Rh_2(esp)_2$ + 20 mol % DPH	66.7 : 1	8.3 : 1	90	200

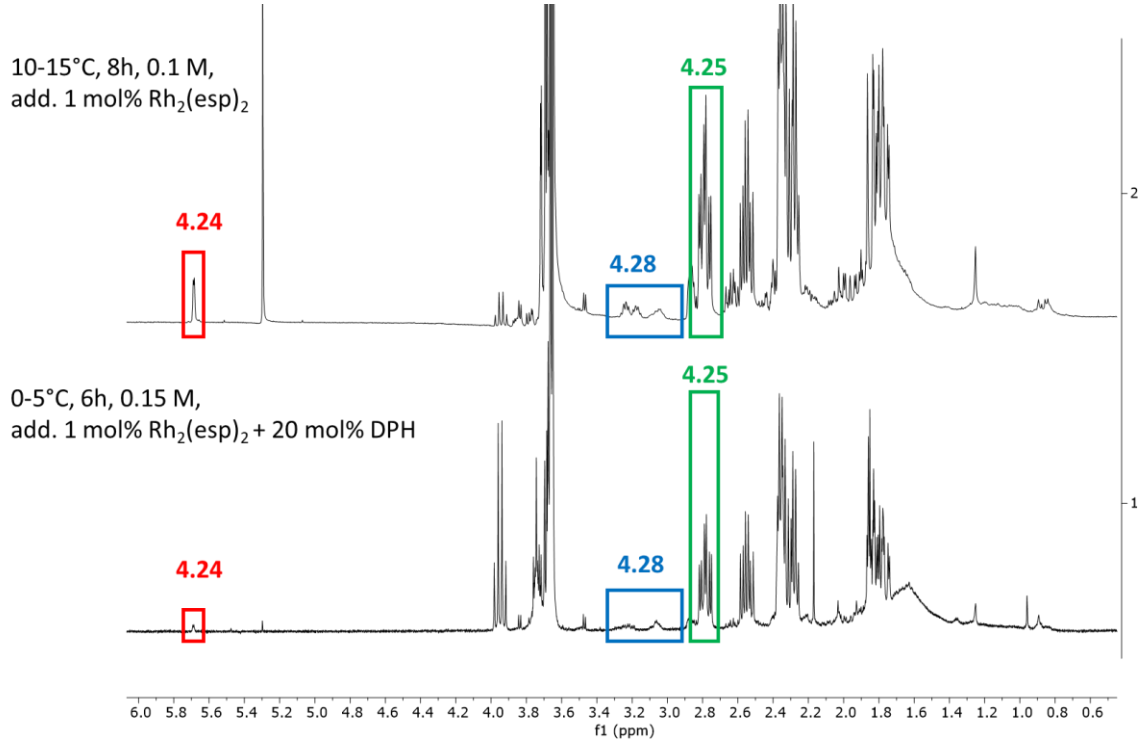
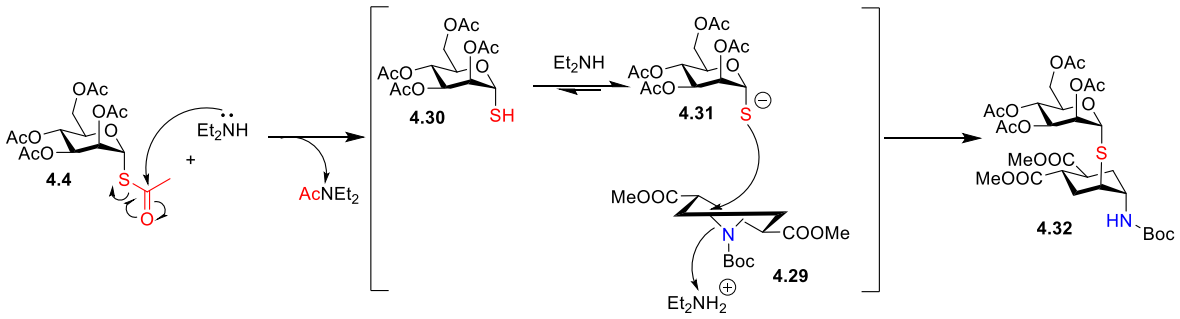


Fig. 4.11 ^{1}H NMR of crude aziridine product **4.25** obtained under different reaction conditions. Starting olefin **4.24** highlighted in red, dimer by-product **4.28** in blue, aziridine **4.25** in green (400 MHz, $CDCl_3$)

4.4. One-pot aziridine opening reaction

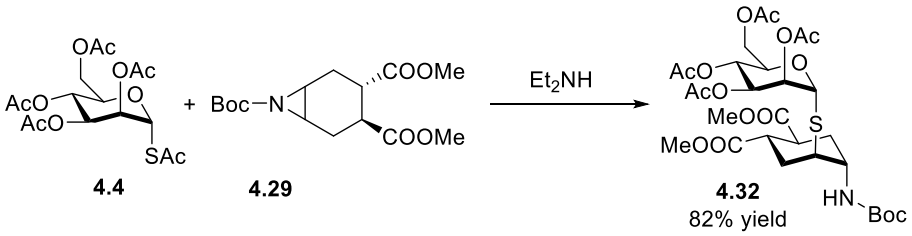
The hypothesized mechanism for the one-pot opening reaction of the N-Boc aziridine **4.29** by glycosyl thioacetates is shown in **Scheme 4.7** for the α-Man thioacetate **4.4**. Et_2NH acts as a nucleophile to selectively deacetylate the sulphur at the anomeric position of the peracetylated thioacetate **4.4**. This gives a thiol **4.30** in equilibrium with the thiolate **4.31** thanks to the excess amount of Et_2NH. The thiolate can then attack the three-membered ring of the aziridine **4.29**, forming the anion which is protonated by the $Et_2NH_2^+$ salt (or by the thiol itself) thus shifting the equilibrium towards the product **4.32**.



Scheme 4.7 Hypothesised mechanism of the one-pot aziridine opening reaction with Man derivate **4.4**

Additionally, because the starting glycosyl thioacetate is added in excess amount (1.3 eq), over time the excess thiol is oxidised to a disulphide by-product. The process can to some extent be controlled by stricter inert conditions and, in particular, by degassing the reaction solvent.

The aglycone moiety with the general formula **4.6** (including Boc protected derivative **4.29**) used in this study was carefully selected for its symmetry properties and high conformational stability, both imparted by the 1,2-*trans*-dicarboxy substituents.[21] Additionally, aziridines **4.6** lack potential chelating groups that have been shown to impinge on the regio- and stereocontrol of nucleophilic substitutions of both cyclitol epoxides and aziridines.[22, 23] Thus it was expected that a single reaction product would be formed by *trans*-diaxial opening of the aziridine.

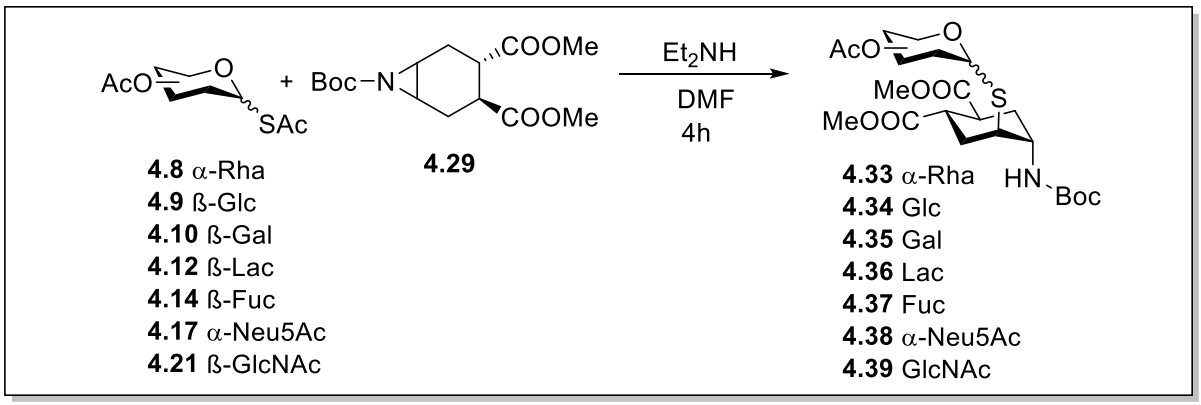


Scheme 4.8 Aziridine opening reaction of mannosyl thioacetate **4.4** with the aziridine **4.29** to give product **4.32**

Indeed, a previously established ring opening reaction of the aziridine **4.29** with mannosyl thioacetate **4.4** (**Scheme 4.8**) proceeded smoothly under the conditions established for opening of the epoxide **4.5** (1.9 mol eq of Et_2NH in DMF [0.65 M] at room temperature for 4 h). Product **4.32** was obtained in 82% yield as a single isomer from a completely selective *trans*-diaxial opening process, preserving, as expected, the α-configuration of mannose. The product configuration was fully confirmed by coupling constant analysis and NOESY, as previously described for **4.2**. [8, 9]

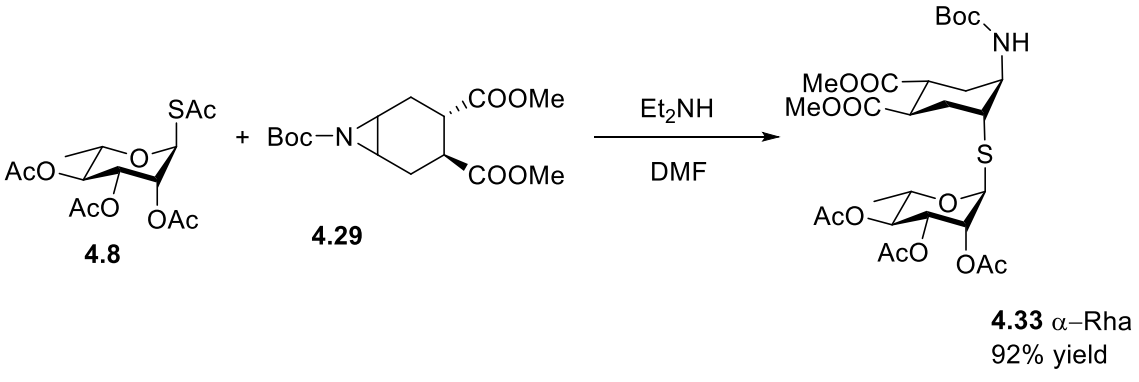
Inspired by the initial success of the one-pot ring opening approach, we were interested in exploring the aziridine opening reaction on a number of mono- and disaccharides, using the aglycone moiety **4.29** as a model aziridine substrate. (**Scheme 4.9**)

Reactions were followed by TLC, NMR and/or LC-MS analysis to provide an insight into the behaviour of individual saccharides under the aziridine opening conditions.



Scheme 4.9 Scope of the aziridine opening reaction

4.4.1. α-Rhamnosyl thioacetate (**4.8**)



Scheme 4.10 Aziridine **4.29** opening reaction of Rha derivative **4.8** to give the product **4.33**

Similar to the Man derivative, reaction of **4.29** with the α-rhamnosyl thioacetate **4.8** under the conditions established for Man (RT, 1.9 eq Et_2NH, DMF [0.65 M], 4 h) afforded **4.33** with 92% isolated yield (0.3 mmol scale) as a single α isomer.

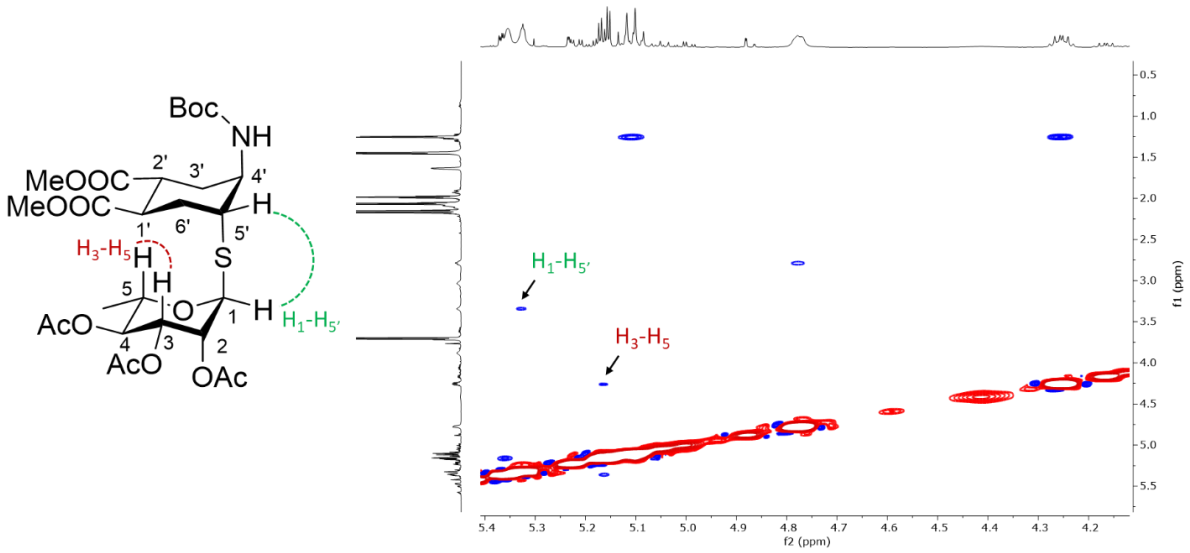


Fig. 4.12 NOESY spectrum of Rha aziridine opening product **4.33** ($CDCl_3$, 600 MHz)

The α anomeric configuration of the Rha derivative **4.33** was established on the basis of the NOESY experiment shown in **Fig. 4.12**, displaying the absence of a cross-peak between the signal of the proton H_1 (at 5.31 ppm) and H_5 (4.23 ppm) on the sugar ring. The presence of a clear H_3-H_5 cross-peak further supports the assignment.

Additionally, the $J_{H1\text{-}C1}$ was measured in a HSQC NMR without ^{13}C decoupling as $J_{H1\text{-}C1}$=170.3 Hz. A general rule states that an equatorial arrangement of the proton (α) typically leads to a higher J_{CH} (~170 Hz) than an axial arrangement (β) (~160 Hz).[24]

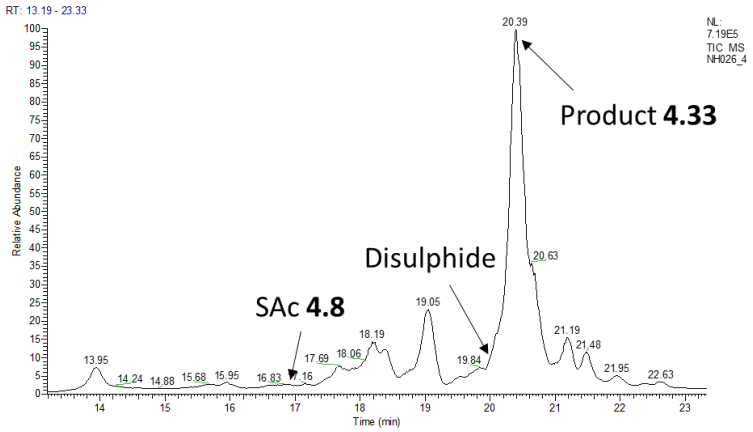
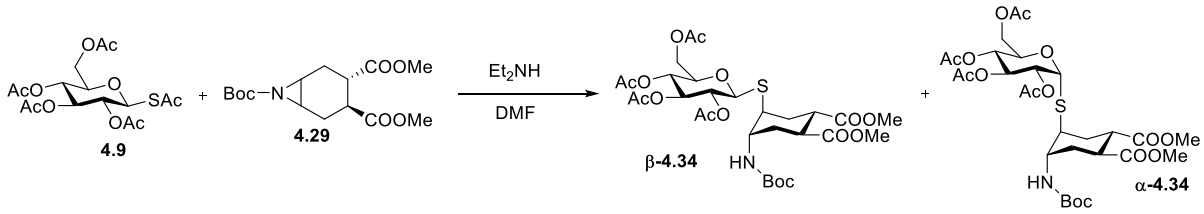


Fig. 4.13 LC-MS analysis of the crude Rha product **4.33** (20°C, 1.9 eq Et_2NH, DMF 0.65 M, 4 h)

LC-MS analysis (**Fig. 4.13**) of the crude product showed a clear peak of the product with T_R 20.39 min and m/z 641.83 corresponding to the product **4.33**. Traces of disulphide by-product with T_R 20.09 min and m/z 632.84 were observed and unreacted thioacetate **4.8** could be detected with T_R 16.88 min and m/z 370.84.

4.4.2. β-Glucosyl thioacetate (**4.9**)



Scheme 4.11 Aziridine **4.29** opening reaction of glucose derivative **4.9** to give the product **4.34**

Under the same conditions established for Man and Rha derivatives (RT, 1.9 eq Et_2NH, DMF [0.65 M], 4 h), however, reaction of the β-glucosyl thioacetate **4.9** afforded both the β- and α-isomers **β-4.34** and **α-4.34** (**Scheme 4.11**) in 2 : 1 ratio, as estimated from crude ^{1}H NMR by integration of the anomeric proton signals at 4.72 and 5.73 ppm, respectively (**Fig. 4.14**).

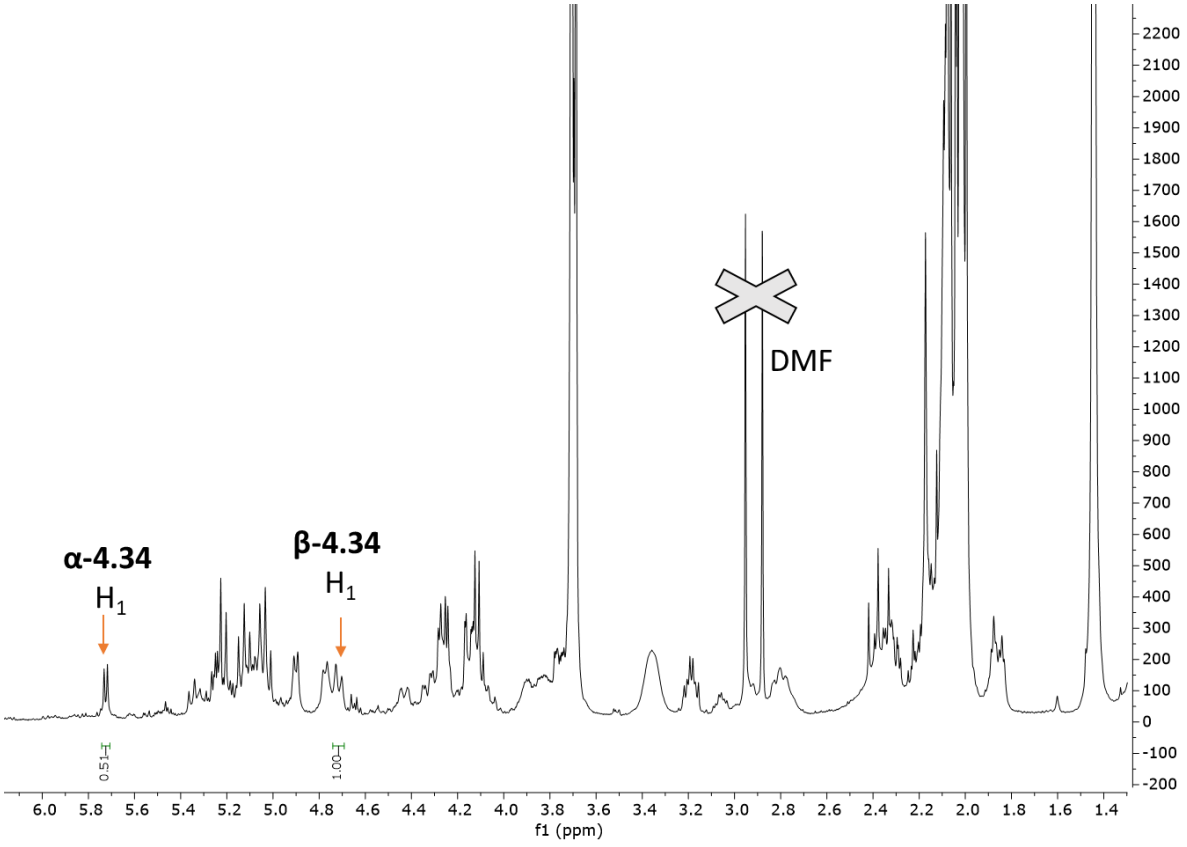


Fig. 4.14 Crude ^{1}H NMR spectra of the glucose aziridine opening product **4.34** (RT, 1.9 eq Et_2NH, DMF [0.65 M], 4 h) (400 MHz, $CDCl_3$)

The two isomers were separated chromatographically (iPr_2O : EtOAc eluent), and analysis of coupling constants ($J_{1,2}$=10.0 Hz corresponding to **β-4.34** and $J_{1,2}$=5.7 Hz corresponding to **α-4.34**) (**Fig. 4.15**) supported by HSQC-NMR (**Fig. 4.16**) and MS data undoubtedly confirmed their structure and anomeric configuration. Besides the characteristic peaks of the anomeric protons, another feature that could be used to distinguish between the two anomers in the ^{1}H NMR spectra are the chemical shifts of the protons on the aglycone moiety. In particular, for the product **β-4.34** the $H_{5'}$ proton appears as a multiplet at 3.42 – 3.26 ppm, the $H_{1'}$ proton at 2.99 – 2.86 ppm and $H_{2'}$ proton at 2.81 – 2.70 ppm. On the other hand, for the product **α-4.34** the chemical shift of $H_{5'}$ proton was at 3.24 – 3.15 ppm, the $H_{1'}$ proton at 3.11 – 2.98 ppm and $H_{2'}$ proton at 2.88 – 2.73 ppm. This detail additionally allowed us to quickly assess whether the aziridine opening products were forming as a single isomer or as a mixture of anomers (**Fig. 4.15**).

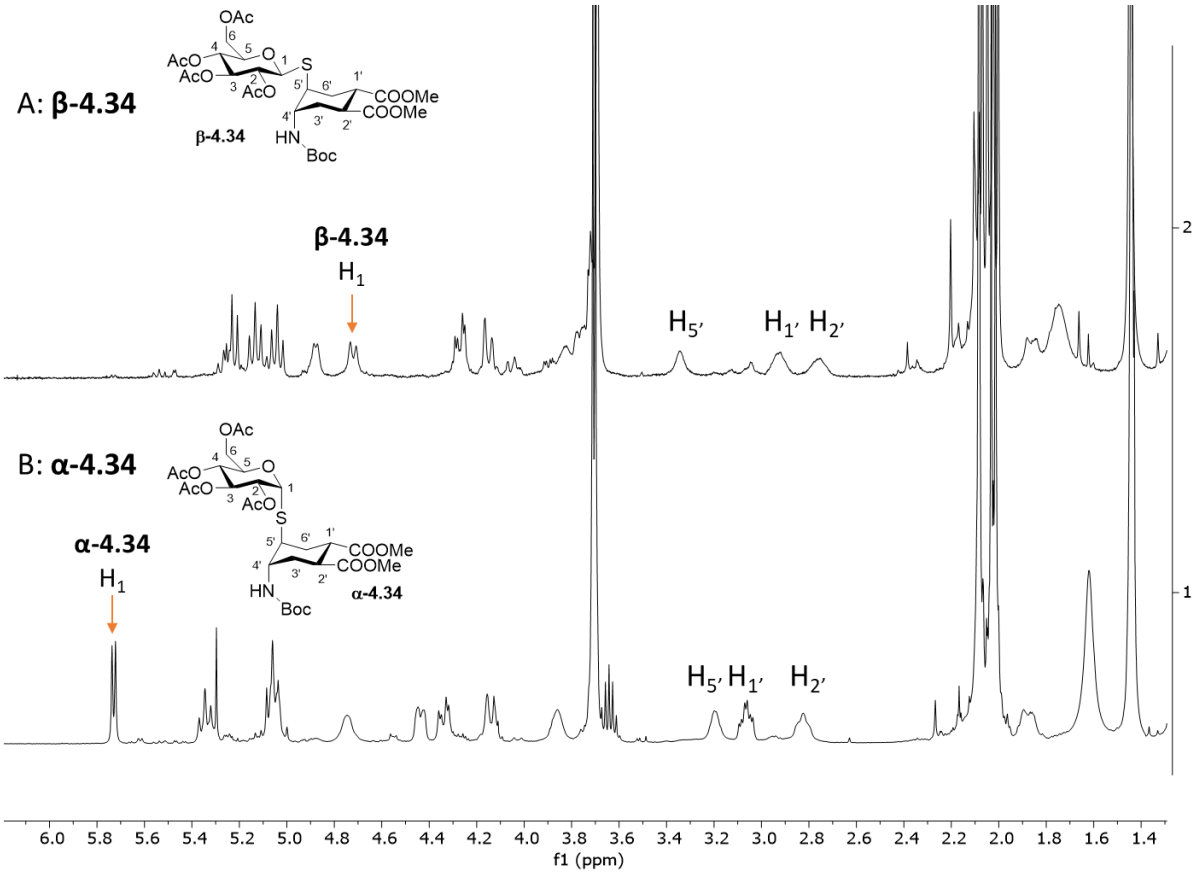


Fig. 4.15 ^{1}H NMR spectra of the glucose aziridine opening products A: **β-4.34** and B: **α-4.34** (400 MHz, $CDCl_3$)

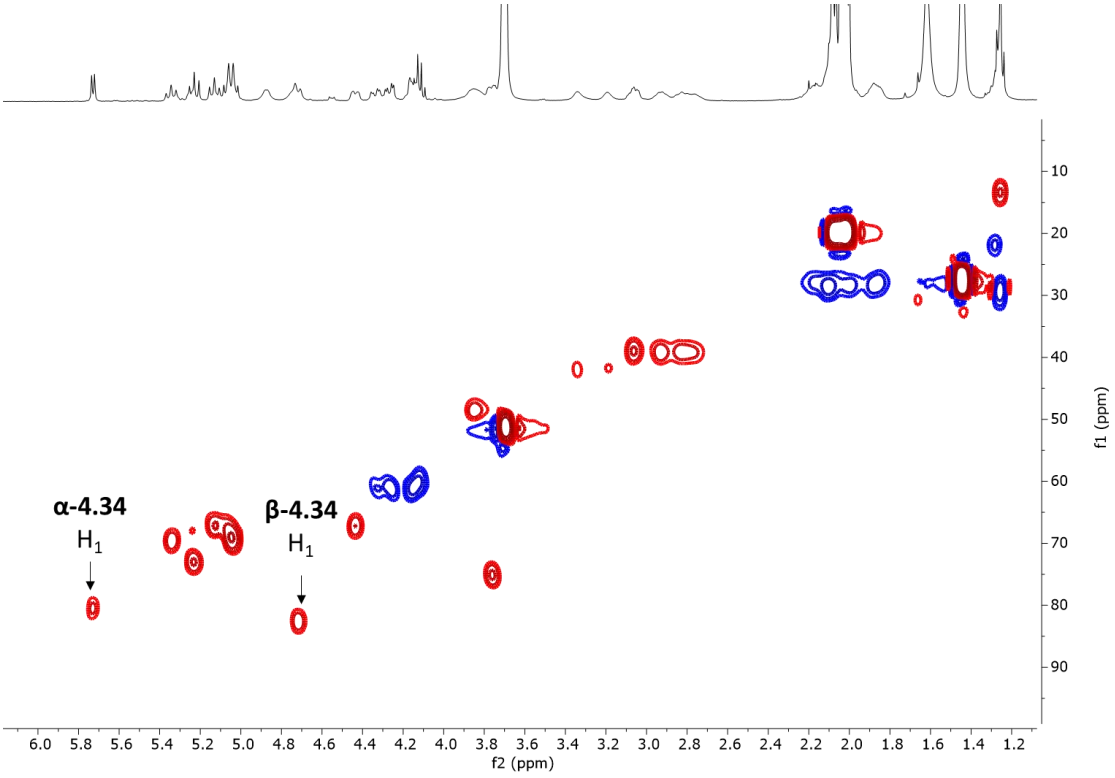
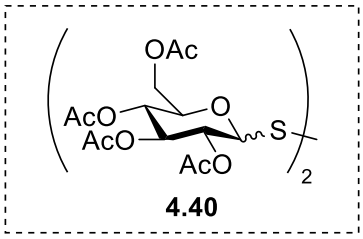


Fig. 4.16 HSQC analysis of the mixture of two isomers of the glucose product **4.34** showing the signals of 2 anomeric protons belonging to **β-4.34** (4.72 ppm) and **α-4.34** (5.73 ppm) ($CDCl_3$)

Similarly, the disulphide by-product **4.40** (**Fig. 4.17**) was seen in the crude ^{1}H NMR spectra as an anomeric mixture.



4.40

Fig. 4.17 Disulphide by-product **4.40** in the synthesis of **4.34**

It was rather unexpected to observe the products being formed as mixtures of anomers, because most sources claim glycosyl thiols are configurationally stable, particularly under basic conditions.[10, 12, 25-27]

We examined the role of reaction temperature, substrate concentration, base concentration and solvent on the relative rate of these reactions. The amount of anomeric isomerization strongly depended on temperature: at 0 °C (**Table 4.2,** entry 3) the ratio changed to 10 : 1 in favour of the product **β-4.34**, but the overall yield of **4.34** decreased. Reducing the amount of base and/or the substrate concentration did not have an effect on the β : α ratio, but also slowed down the S_N2 reaction process, thus increasing the amount of disulphide by-product **4.40** (*e.g.* compare entries 1 and 2, 3

and 4, 3 and 5 in **Table 4.2**). Changing the solvent from DMF to CH_2Cl_2 (**Table 4.2,** entry 6) gave low yields (16%) of almost pure **β-4.34** product, but dimerization of the glycosyl thiol was the major result. Using both acetonitrile or DMF/CH_2Cl_2 mixtures of various composition (from 7/3 to 9/1) the reaction was slow and low yields of β : α = 3 : 1 were obtained (not shown in the table). Thus, it appears that low temperatures (0°C) favour aziridine opening over anomerization. A high concentration of substrate is also beneficial, by increasing the rate of the nucleophilic substitution. A solvent of lower polarity such as CH_2Cl_2 appears to reduce the anomeric isomerization, but it also slows down the S_N2 reaction, thus resulting in low yields and extensive dimerization of the thiol. Reducing the amount of base has no influence on the selectivity, but again favours dimerization over the formation of the aziridine opening product **4.34**.

Table 4.2 [a] Opening reaction of aziridine **4.29** with 1-*S*-acetyl-β-D-glucopyranose **4.9**

Entry	T (°C)	Et_2NH (mol equiv)	**β–4.34** : **α–4.34**[b]	Yield[c] (%)	**4.40** (%)[b]
1	20	1.9	2:1	61	< 5
2	20	1.4	2:1	43	37
3	0	1.9	10:1	34	41
4	0	1.4	10:1	28	61
5[d]	0	1.9	10:1	24	63
6[e]	20	1.9	20:1	16	71

*a) Unless otherwise noted, all reactions were performed on a 0.06 mmol scale, with a 0.65 M concentration of **4.29** in DMF and 1.3 mol equiv of **4.9** for 4h with the amount of base and at the temperature indicated b) as judged by ^{1}H-NMR of the crude; c) isolated, combined yields of the two anomeric products **4.34** d) 0.3 M concentration of **4.29**; e) reaction performed in CH_2Cl_2.*

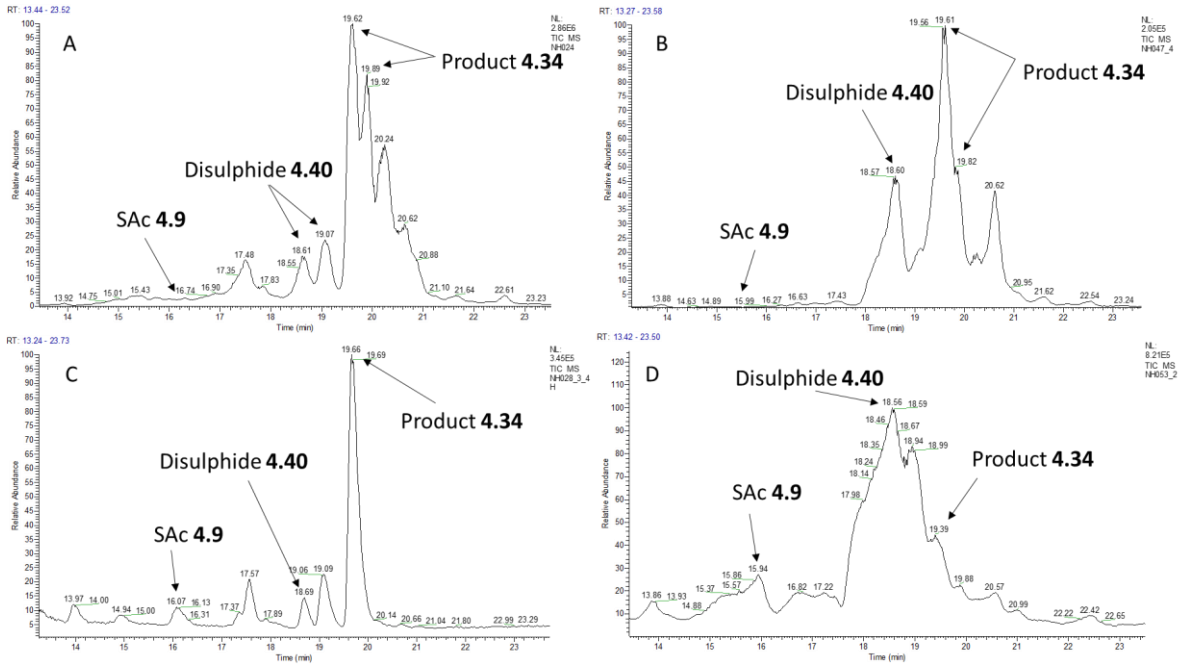


Fig. 4.18 LC-MS analysis of reaction mixtures of the aziridine opening reactions of Glc thioacetate **4.9** to give product **4.34** under different conditions **A**: 20°C, 1.9 eq Et_2NH, 0.65 M DMF (Table 4.2, entry 1), **B**: 0°C, 1.9 eq Et_2NH, 0.65 M DMF (Table 4.2, entry 3), **C**: 0°C, 1.9 eq Et_2NH, 0.3 M DMF (Table 4.2, entry 5), **D**: 20°C, 1.9 eq Et_2NH, 0.65 M CH_2Cl_2 (Table 4.2, entry 6)

Although not quantitative, also the LC-MS analysis of the crude product indicated that in certain cases both the aziridine opening product **4.34** and the disulphide by-product **4.40** were split in 2 signals, indicating the presence of two isomers (**Fig. 4.18**). The peaks of the product **4.34** were observed with T_R 19.56 min and 19.70 min and m/z 699.81. The peaks of the disulphide **4.40** were observed with T_R 18.61 min and 19.07 min and m/z 748.88. This is particularly visible in **Fig. 4.18, spectra A** which corresponds to **Table 4.2, entry 1** where the **β-4.34** and **α-4.34** isomers were formed in 2:1 ratio. On the other hand, **Fig. 4.18, spectra C** does not show any splitting of the signals, as is expected, since the β:α ratio in this experiment was 10:1. **Fig. 4.18, spectra D** that corresponds to **Table 4.2, entry 6** shows a significant peak of disulphide product **4.40** and a much smaller product **4.34** signal. This is due to the aziridine opening being slowed down when using CH_2Cl_2 as a solvent and the presence of O_2 (non-degassed solvent) allowed for oxidation of thiols to disulphide.

LC-MS analysis was also used to follow reactions over time. **Fig. 4.19** shows LC-MS analysis of the reaction under conditions reported in **Table 4.2, entry 3** (0°C, 1.9 eq Et_2NH, 4 h). A sample from the reaction mixture was taken after each hour and quenched with 1 M HCl. This type of analysis showed us that the reactions proceeded very fast, the product **4.34** was clearly formed already after 1 hour of the reaction (**Fig. 4.19, spectra A**).

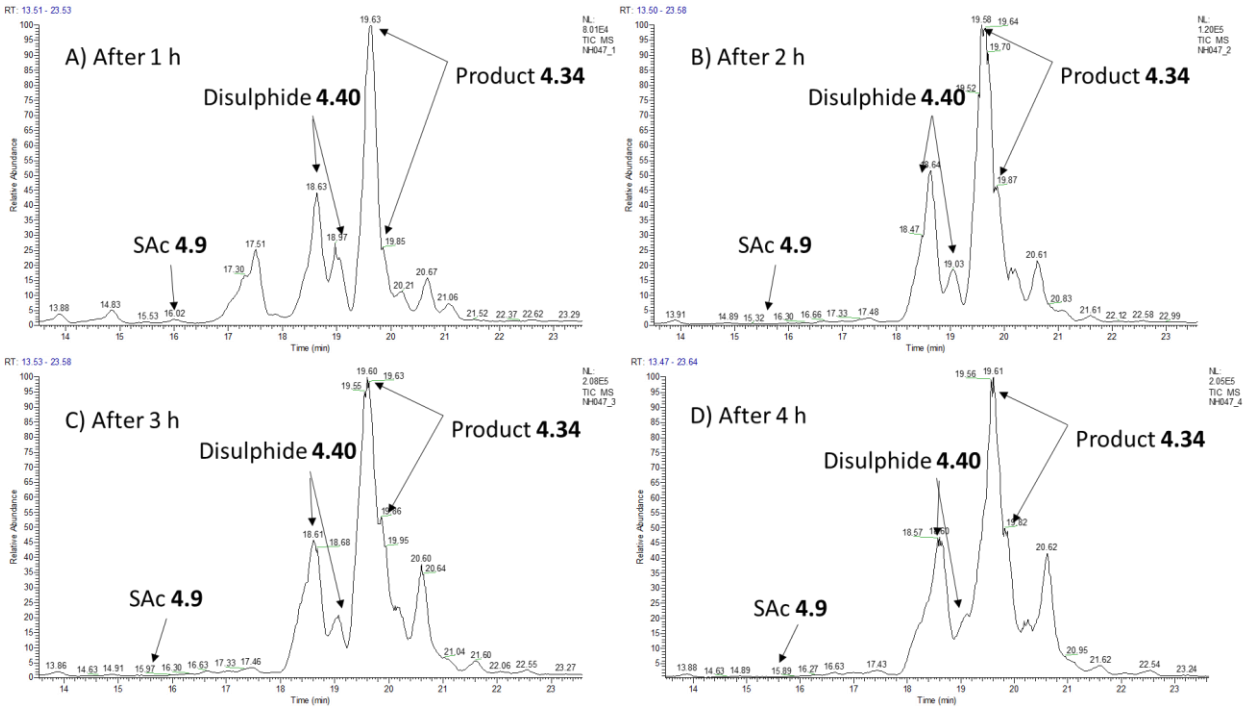
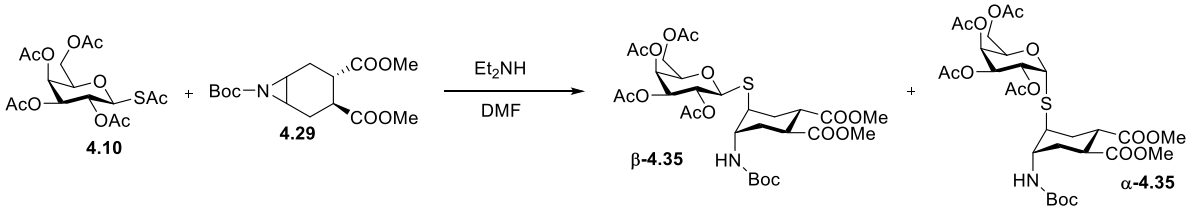


Fig. 4.19 LC-MS analysis of the reaction of Glc derivative **4.9** with aziridine **4.29** to give product **4.34** A) after 1 h, B) after 2 h, C) after 3 h, D) after 4 h (0°C, 1.9 eq Et_2NH, DMF 0.6 M)

In conclusion, the highest yields of the aziridine opening product **4.34** formed in the reaction of **4.9** with the aziridine **4.29** were obtained using the reaction conditions developed for mannose and operating at room temperature. Under these conditions, a high isomerization rate was observed. Performing the reaction at 0 °C, improved the **β-4.34** : **α-4.34** ratio to synthetically useful levels (10 : 1), but reduced the yield to 34%, due to extensive formation of the disulphide **4.40**.

As mentioned above, the extent of the thiolate anomerization starting from β-Glc thioacetate **4.9** and even its presence were rather surprising on the background of previous reports.[10, 12, 25] Therefore, we settled to analyse some mechanistic aspects of this transformation. The experiments performed and their results are described in Chapter Five.

4.4.3. β-Galactosyl thioacetate (**4.10**)



Scheme 4.12 Aziridine **4.29** opening reaction of galactose derivative **4.10** to give the product **4.35**

Similarly, anomeric isomerization was observed also when using the Gal derivative **4.10** in the aziridine opening reaction with the aziridine **4.29**. Under the conditions optimized for Man (RT, 1.9 eq Et_2NH, DMF 0.65 M, 4 h) **β-4.35** and **α-4.35** products were obtained in 3:1 ratio as estimated by crude ^1H NMR (**Fig. 4.20, spectra A**). The NMR analysis confirmed the presence of 2 anomeric protons at 4.69 ppm ($J_{1,2}$= 10.0 Hz) and 5.81 ppm ($J_{1,2}$= 5.6 Hz), corresponding to the products **β-4.35** and **α-4.35**, respectively (**Fig. 4.20**). The two isomers **β-4.35** and **α-4.35** could be chromatographically separated with a Biotage automated column chromatography using a gradient of the eluent iPr_2O:EtOAc from 8:1 to 8:2 (**Fig. 4.20, Spectra B: α-4.35, Spectra C: β-4.35**). The overall yield of **4.35** under these conditions was 44% (**Table 4.3**, **entry 1**).

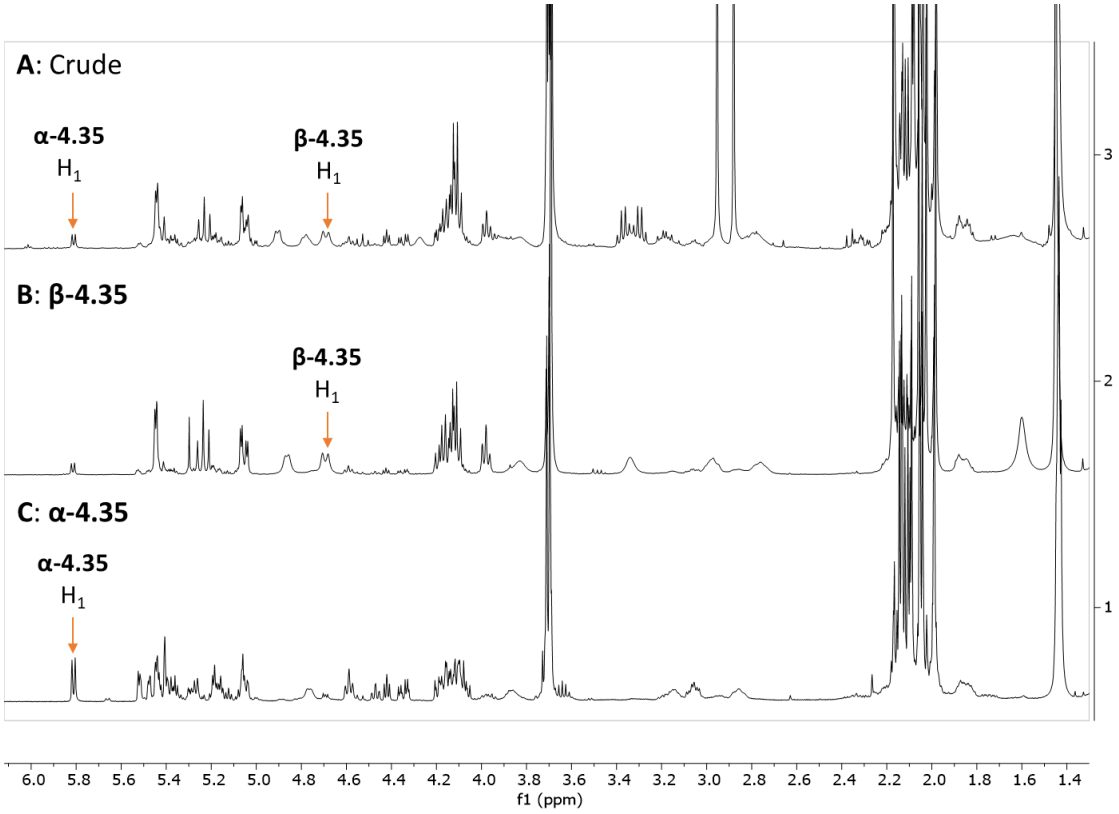


Fig. 4.20 ^1H NMR analysis of Gal **4.10** reaction with aziridine **4.29**. A: Crude product **4.35**, B: Spectra of **β-4.35** (H_1 5.81 ppm), C: Spectra of **α-4.35** (H_1 4.69 ppm) (RT, 1.9 eq Et_2NH, DMF 0.65 M, 4 h) (400 MHz, $CDCl_3$)

Performing the reaction at 0°C only slightly improved the β:α ratio to 4.5:1 (**Table 4.3, entry 2**) and decreased the overall yield of **4.35** to 36%. Furthermore, lower concentration of Et_2NH gave a β:α ratio of 5:1 and a yield of 32% (**Table 4.3**, **entry 3**).

Table 4.3[a] Opening reaction of aziridine **4.29** with galactose thio-derivative **4.10**.

Entry	T (°C)	Et_2NH (mol equiv)	**β-4.35:α-4.35**[b]	Yield[c] (%)
1	20	1.9	3:1	44
2	0	1.9	4.5:1	36
3	0	1.4	5:1	32

*a) Reactions were performed with a 0.65 M concentration of **4.29** in DMF and 1.3 mol equiv of **4.10** for 4h with the amount of base and at the temperature indicated b) as judged by ^{1}H-NMR of the crude; c) isolated, combined yields of the two anomeric products **4.35***

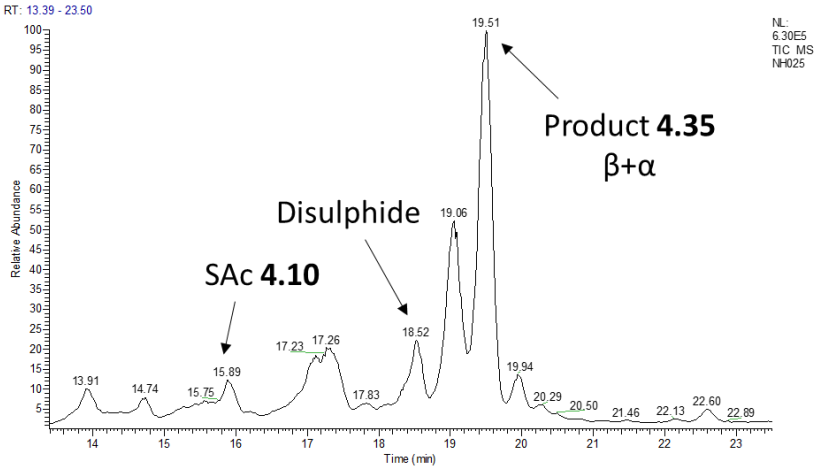
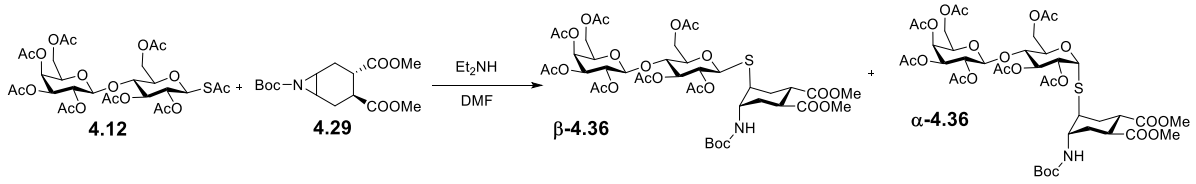


Fig. 4.21 LC-MS analysis of the crude **4.35** from reaction of Gal derivative **4.10** with the aziridine **4.29** (RT, 1.9 eq Et_2NH, DMF 0.65 M, 4 h)

LC-MS analysis of crude from the reaction of Gal derivative **4.10** with the aziridine **4.29** under the conditions reported in **Table 4.3**, entry 1 (RT, 1.9 eq. Et_2NH, 0.65 M DMF, 4 h) showed the presence of the product **4.35** (β+α) with T_R 19.51 min and m/z 699.71, the disulphide by-product with T_R 18.52 min and m/z 748.77 and starting galactosyl thioacetate **4.10** with T_R 15. 89 min and m/z 428.80. (**Fig. 4.21**)

4.4.4. β-Lactosyl thioacetate (**4.12**)



Scheme 4.13 Aziridine **4.29** opening reaction of lactose derivative **4.12** to give the product **4.36**

Lac derivative **4.12** showed a behaviour comparable to Gal thioacetate **4.10** in terms of anomeric isomerisation. At 20°C the estimated β:α ratio in crude ^{1}H NMR spectra of the product **4.36** was 3:1 (**Table 4.4, entry 1**). The isomer **β-4.36** was identified by a characteristic peak of the anomeric proton

at 4.67 ppm ($J_{1\beta\text{-}2\beta}$ = 10.1 Hz) and the isomer **α-4.36** at 5.62 ppm ($J_{1\alpha\text{-}2\alpha}$=5.7 Hz) (**Fig. 4.22**). The anomeric ratio **β-4.36:α-4.36** was only slightly improvemed at 0°C to 5:1 (**Table 4.2, entry 2**).

Table 4.4[a] Reaction of aziridine **4.29** with Lac derivative **4.12**.

Entry	T (°C)	Et_2NH (mol equiv)	**β-4.36:α-4.36**[b]	Yield[c] (%)
1	20	1.9	3:1	39
2	0	1.9	5:1	42

*a) All reactions were performed with a 0.65 M concentration of **4.29** in DMF and 1.3 mol equiv of **4.12** for 4h with the amount of base and at the temperature indicated b) as judged by ^{1}H-NMR of the crude; c) isolated, combined yields of the two anomeric products **4.36***

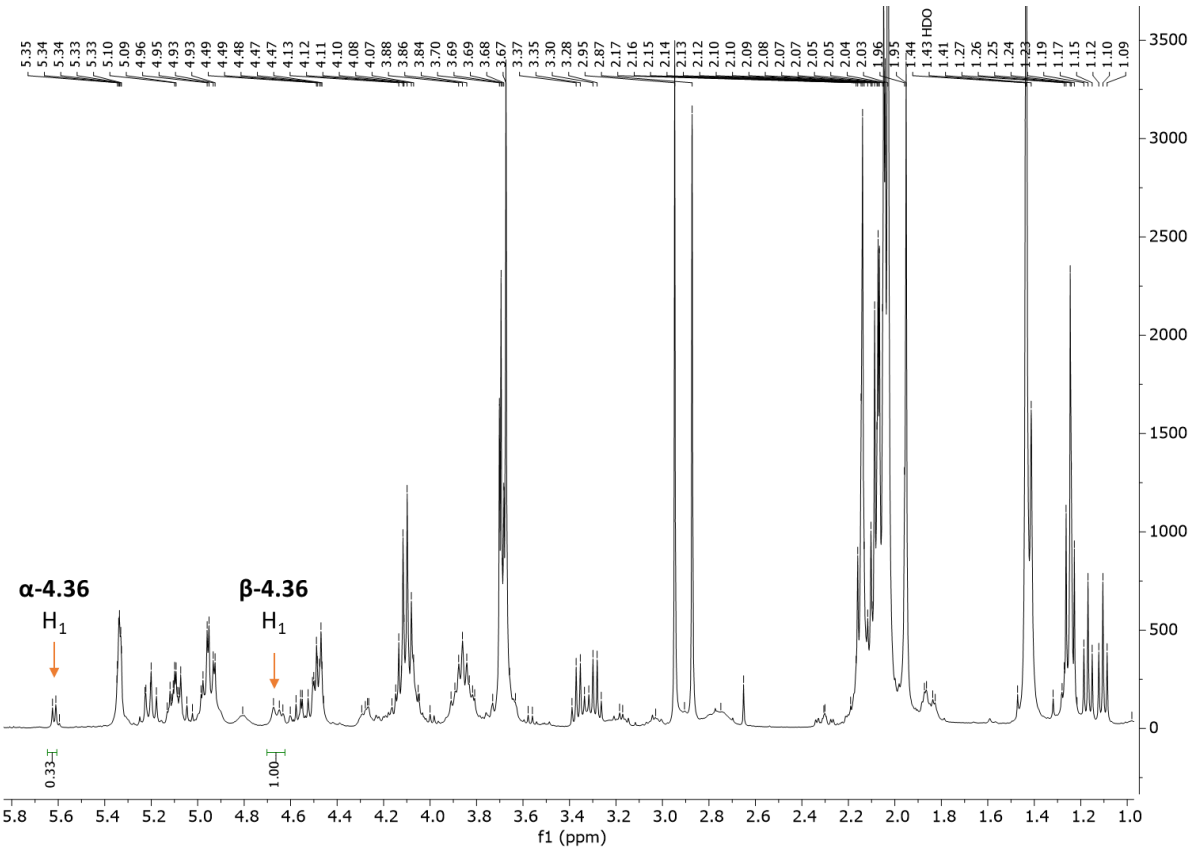


Fig 4.22 Crude ^{1}H NMR spectra of Lac derivative **4.36** showing the presence of 2 anomeric protons belonging to the 2 isomers **β-4.36** and **α-4.36** (RT, 1.9 eq. Et_2NH, 0.65 M DMF, 4 h) (400 MHz, $CDCl_3$)

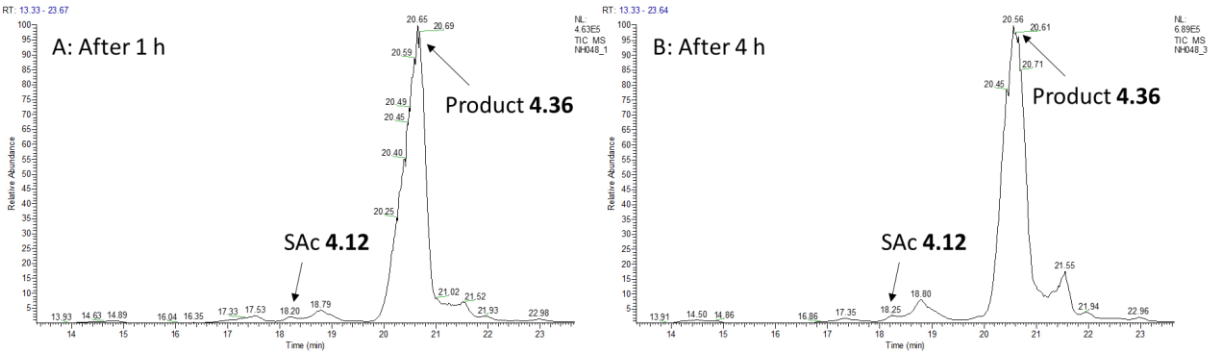
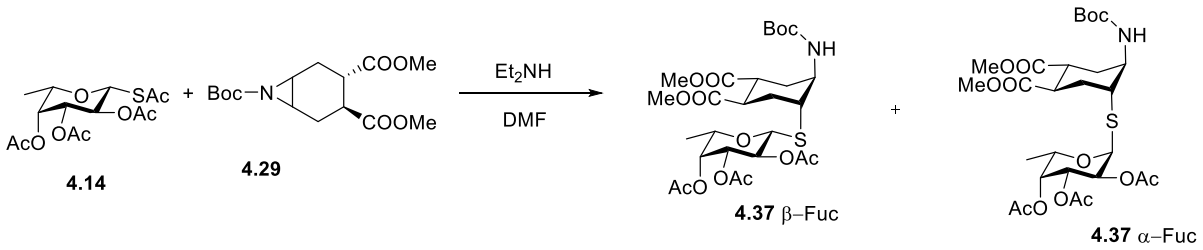


Fig 4.23 LC-MS analysis of the crude **4.36** from the reaction of Lac derivative **4.36** with the aziridine **4.29** A: After 1 h, B: after 4 h (RT, 1.9 eq. Et_2NH, 0.65 M DMF, 4 h)

Comparing the LC-MS analysis of the reaction mixture under conditions reported in **Table 4.4, entry 1** after 1 h and after 4 h showed that the reaction proceeded fast and the product **4.36** with T_R 20.65 min and m/z 987.65 was formed already after 1 h. Additionally, we saw a peak of the starting lactose derivative **4.12** with T_R 18.20 min and m/z 716.76 (**Fig. 4.23)**.

4.4.5. β-Fucosyl thioacetate (**4.14**)



Scheme 4.14 Aziridine **4.29** opening reaction of fucose derivative **4.14** to give the product **4.37**

Due to the impure starting material, the aziridine opening product of the fucosyl thioacetate derivative **4.14** was difficult to quantitatively characterize in terms of anomeric isomerization. The ^1H NMR spectra of the product **4.37** obtained under the reaction conditions used for Man (RT, 1.9 eq Et_2NH, 4 h) show the presence of two anomeric protons at 4.61 ppm ($J_{1,2}$=10.0 Hz) and 5.78 ppm ($J_{1,2}$=5.9 Hz) as well as two sets of $H_{5'}$, $H_{2'}$ and $H_{1'}$ signals that indicate the formation of the products **β-4.37** and **α-4.37**, in approximately 5:1 ratio (**Fig. 4.24**).

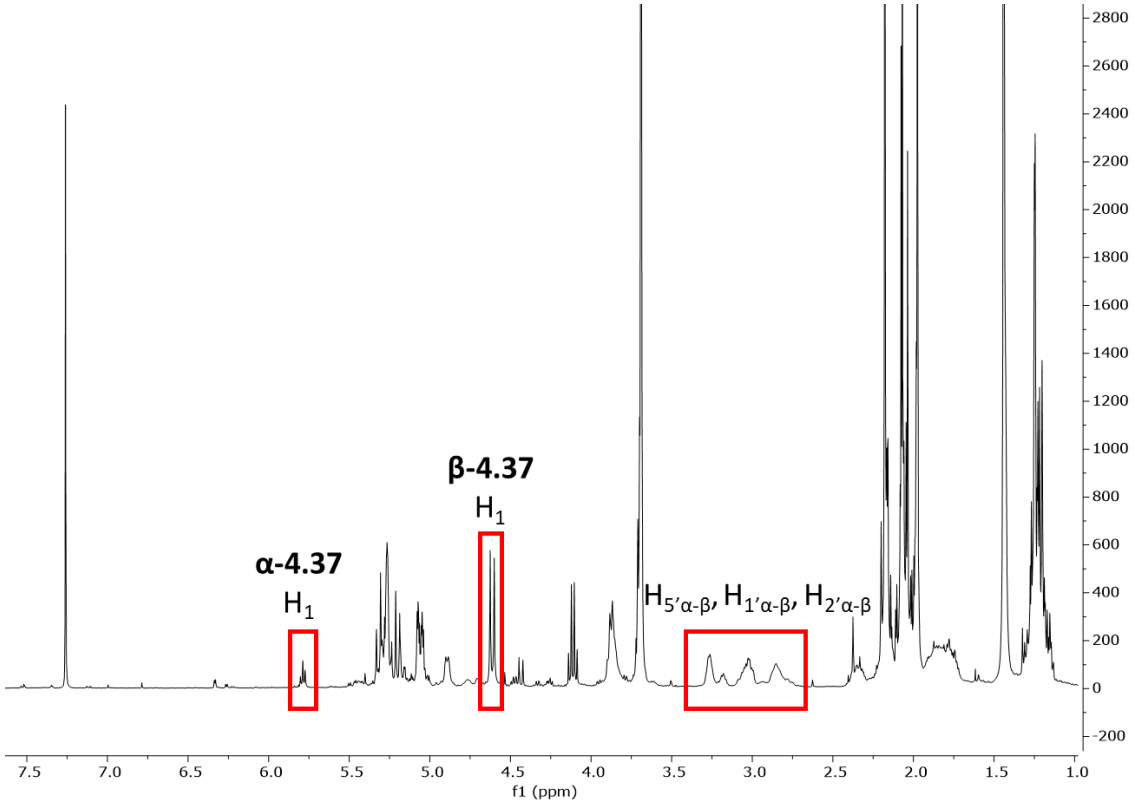


Fig. 4.24 ^{1}H NMR analysis of Fuc product **4.37** (after purification) as a mixture of 2 isomers, characteristic peaks highlighted in red (RT, 1.9 eq. Et_2NH, 0.65 M DMF, 4 h) (400 MHz, $CDCl_3$)

The LC-MS analysis of the crude product **4.37** showed the product **4.37** peak split in two with T_R 20.12 min and 19.67 min and m/z 641.77, further supporting the evidence that the isomerization is occurring. The disulphide by-product was observed at T_R 18.67 min with m/z 632.73. (**Fig. 4.25**)

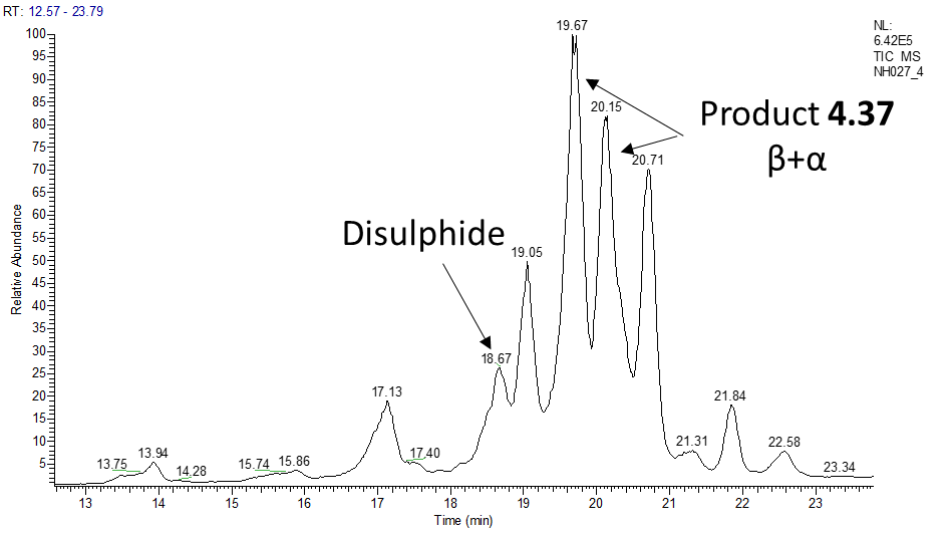
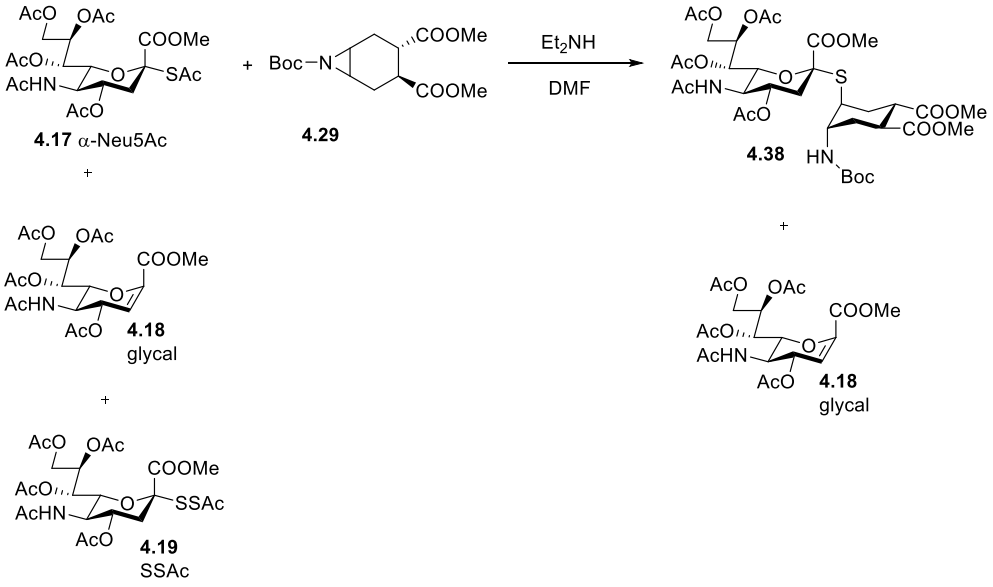


Fig. 4.25 LC-MS analysis of the crude **4.37** from the reaction of the Fuc derivative **4.14** with the aziridine **4.29** (RT, 1.9 eq. Et_2NH, 0.65 M DMF, 4 h)

4.4.6. *N*-Acetylneuraminic acid thioacetate (**4.17**)



Scheme 4.15 Aziridine **4.29** opening reaction of Neu5Ac derivative **4.17** to give the product **4.38**

Thioacetylation of Neu5Ac is known to yield a basically inseparable mixture of products which contains the expected α-thioacetate **4.17**, together with the glycal **4.18** and bisthioacetate **4.19** products. The bisthioacetate product **4.19** is unstable under the Et_2NH conditions used in the aziridine opening reaction and thus gives a thiol that reacts with the aziridine.[13] Using this mixture in the aziridine opening reaction under conditions established for Man (RT, 1.9 eq Et_2NH, DMF 0.65 M, 4 h) afforded a crude which, upon chromatographic purification (CH_2Cl_2:acetone=3:1), yielded the α-thioglycoside **4.38** (56%), together with *ca*. 17% glycal **4.18**.

The mixture was purified by HPLC for analytical purposes and NMR analysis confirmed that **4.38** was obtained as a single α-anomer, as shown by the signal of the H_{3eq} proton, which appears at 2.70 ppm as a doublet of doublet (J_{gem} = 12.8 Hz, J_{3eq-4} = 4.5 Hz)[28] (**Fig. 4.26**) and by the signal of the C_1 carbon in a proton not decoupled ^{13}C NMR spectra, which appears at 168.4 ppm as a doublet of quartets with a coupling constant $J_{C1-H3ax}$ = 3.9 Hz (**Fig. 4.26**)[29] Thus, as for mannose and rhamnose, the Neu5Ac thiol is configurationally stable under the reaction conditions.

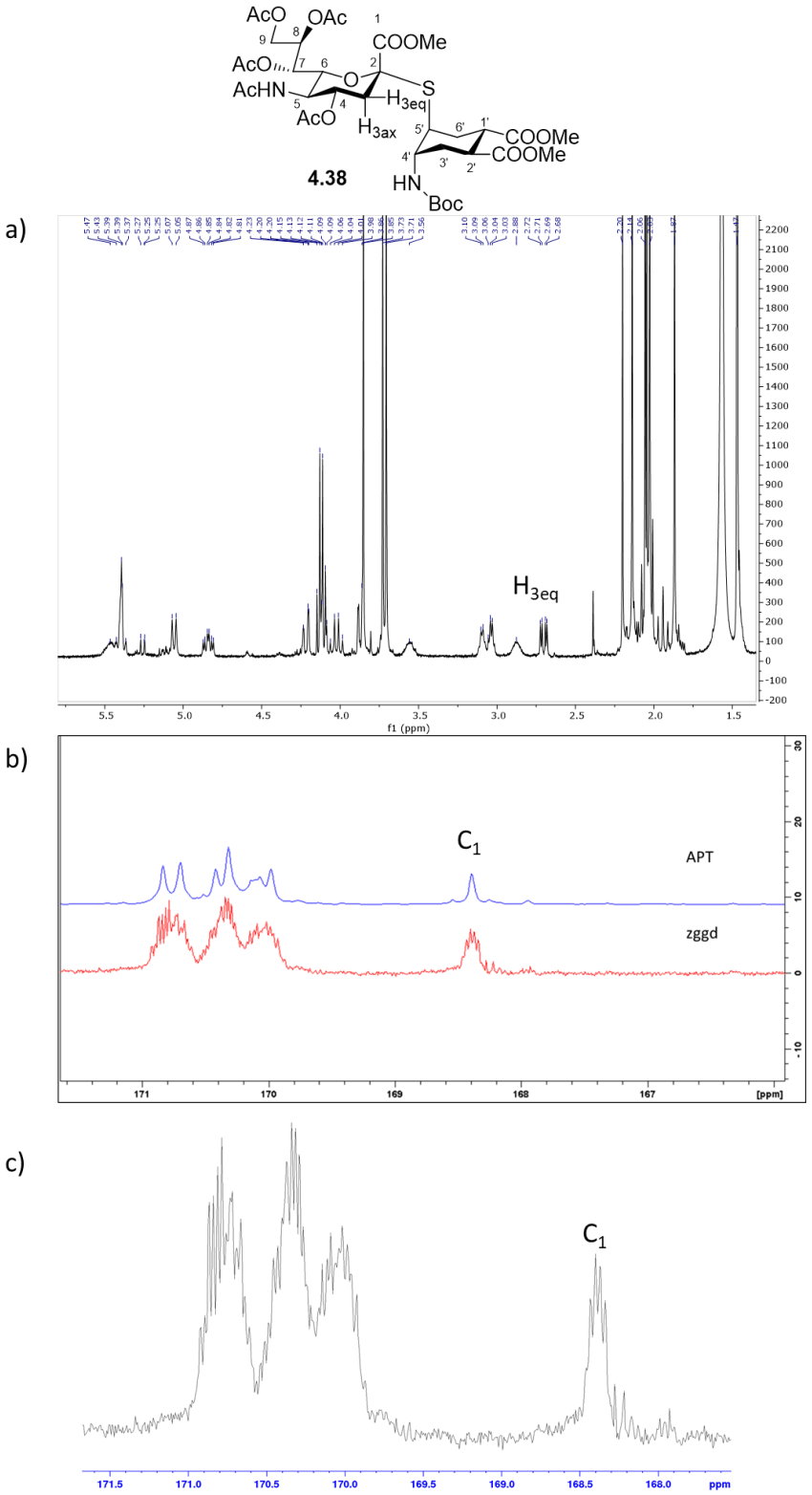


Fig. 4.26 NMR spectra of compound **4.38** ($CDCl_3$, 100 MHz). a)^{1}H NMR the signal of Neu5Ac H_{3eq} appears at 2.70 ppm as a dd, b) ^{13}C NMR the signal of Neu5Ac C_1 appears at 168.4 ppm as a singlet in the APT trace and as a dq in the proton not decoupled spectrum (zggd); c) expansion of the 168-171 ppm spectral region of the zggd spectrum

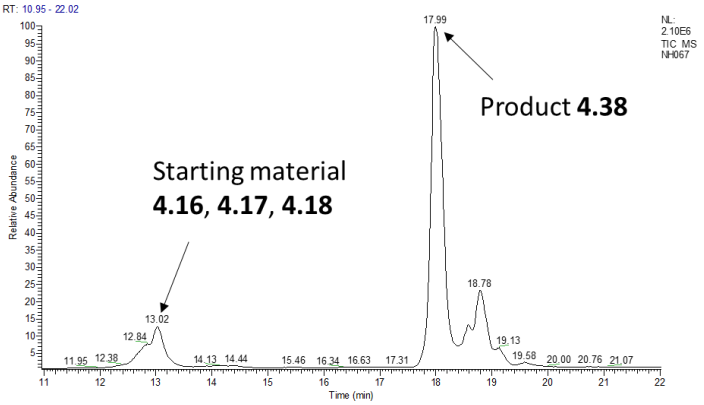
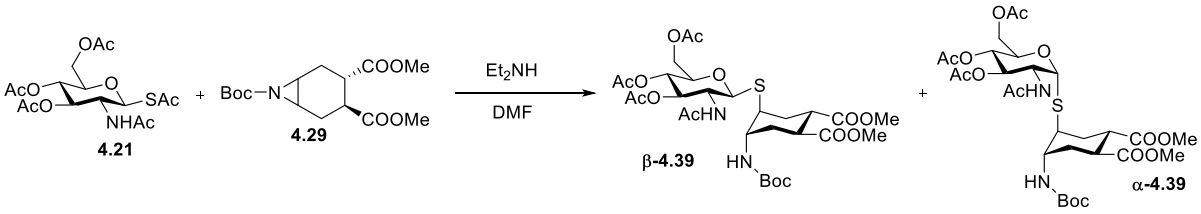


Fig. 4.27 LC-MS analysis of crude **4.38** from reaction of Neu5Ac derivative **4.17** with the aziridine **4.29** (RT, 1.9 eq Et_2NH, DMF 0.65 M, 4 h)

LC-MS analysis of the crude product showed a clear peak of the product **4.38** with T_R 17.99 min and m/z 842.71 and a peak with a T_R 13.02 min containing a mixture of starting material SAc **4.17** (m/z 571.57), SSAc **4.19** (m/z 603.69) and glycal **4.18** (m/z 495.71) (**Fig. 4.27**).

4.4.7. β-*N*-Acetylglucosaminyl thioaceate (**4.21**)



Scheme 4.17 Aziridine **4.29** opening reaction of GlcNAc derivative **4.21** to give the product **4.39**

Further exploring the scope of the reaction, using the GlcNAc derivative **4.21** in the aziridine opening reaction turned into a pretty demanding task. The crude product obtained after reproducing the conditions optimized for Man (RT, 1.9 eq Et_2NH, 0.65 M DMF, 4 h) was a complex mixture of side products and isomers. A careful two step purification with Biotage automated chromatography (first column CH_2Cl_2:MeOH 80:1 and then second column Hex:EtOAc 1:1) allowed us to isolate the product **4.39** as a mixture of two isomers **β-4.39** and **α-4.39**, confirmed by NMR HSQC analysis showing the signals of two anomeric protons at 4.75 ppm ($J_{1,2}$=10.4 Hz) corresponding to **β-4.39** and 5.52 ppm ($J_{1,2}$=5.2 Hz) corresponding to **α-4.39** (**Fig. 4.28** and **Fig. 4.29**). The reaction of **4.21** with **4.29** gave *ca*. 19% overall yield of **4.39** as a 2:1 β:α mixture.

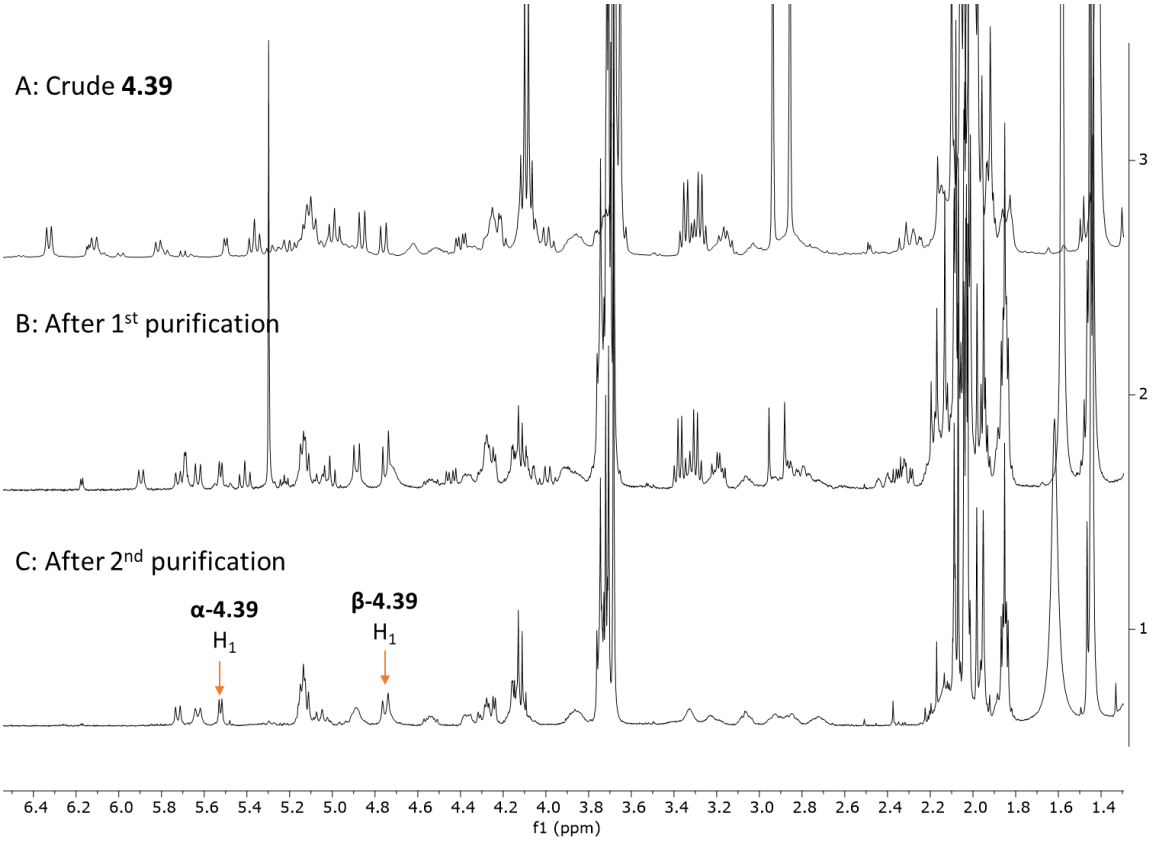


Fig. 4.28 ^{1}H NMR analysis of reaction of aziridine **4.29** with GlcNAc derivative **4.21** to give **4.39**. A: Crude product, B: After 1st purification (CH_2Cl_2:MeOH 80:1), C: After 2nd purification (Hex:EtOAc 1:1) (RT, 1.9 eq Et_2NH, 0.65 M DMF, 4 h) (400 MHz, $CDCl_3$)

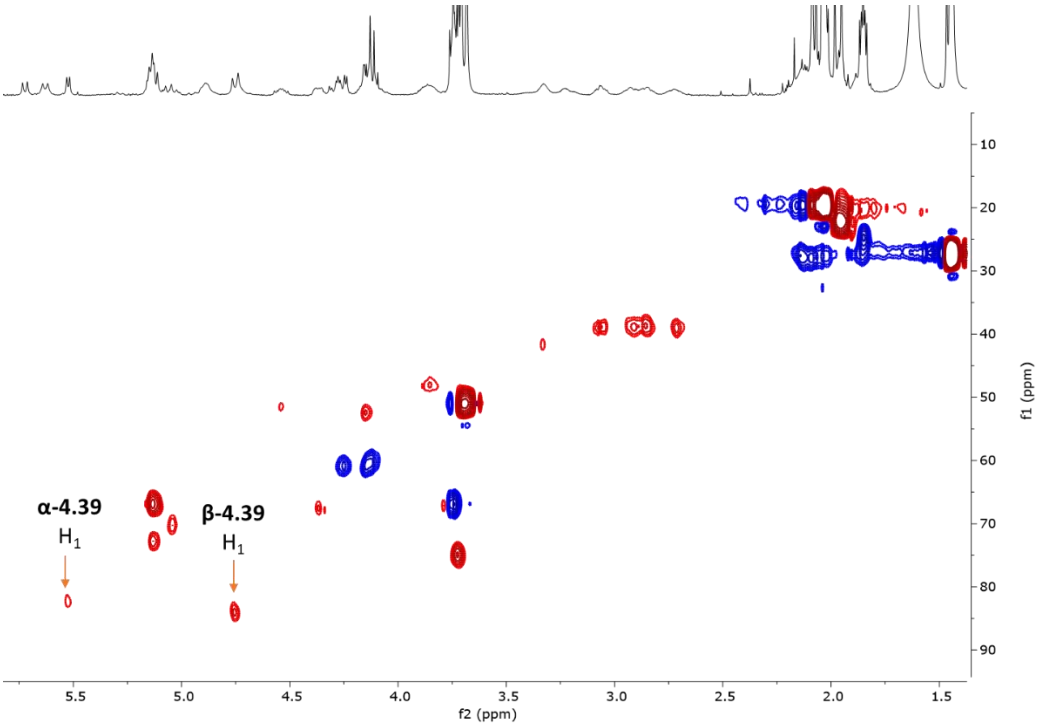


Fig. 4.29 HSQC spectra of GlcNAc product **4.39** showing the presence of 2 isomers: 4.75 ppm **β-4.39** (H_1), 5.52 ppm **α-4.39** (H_1) ($CDCl_3$).

Lowering the reaction temperature to 10°C and 0 °C, the β : α ratio improved marginally (3 : 1), but the yields became so low, that it was not possible to isolate the product anymore, thus it was not worth optimizing the temperature further. (**Table 4.5**)

Table 4.5[a] Reaction of aziridine **4.29** with GlcNAc thioaceate **4.21**

Entry	T (°C)	Et_2NH (mol equiv)	**β-4.39:α-4.39**[b]	Yield[c] (%)
1	20	1.9	2:1	19
2	0	1.9	3:1	n.d.

*a) All reactions were performed with a 0.65 M concentration of **4.29** in DMF and 1.3 mol equiv of **4.21** for 4h with the amount of base and at the temperature indicated b) as judged by ^{1}H-NMR of the crude; c) isolated, combined yields of the two anomeric products **4.39***

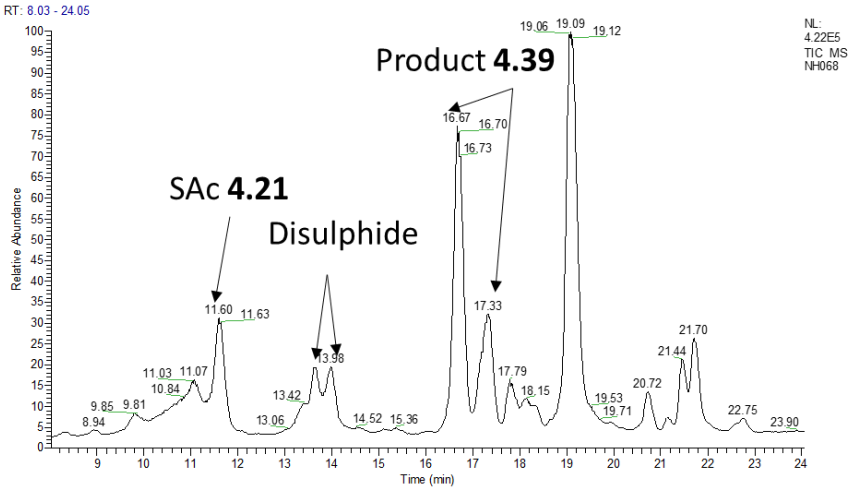


Fig. 4.30 The LC-MS analysis of the crude **4.39** from the reaction of GlcNAc derivative **4.21** with the aziridine **4.29** (RT, 1.9 eq Et_2NH, 0.65 M DMF, 4 h)

The LC-MS analysis of the crude product **4.39** obtained under conditions reported in **Table 4.5, entry 1** (RT, 1.9 eq Et_2NH, 0.65 M DMF, 4 h) showed two peaks at T_R 17.33 min and 16.67 min with m/z 698.75 corresponding the product **4.39**. Additionally, there were two peaks with m/z 746.65 corresponding to the disulphide by-product. A peak of unreacted starting material **4.21** was also observed with T_R 11.60 min and m/z 427.71 (**Fig. 4.30**).

The likely reason for the aziridine opening to be so sluggish is the reactivity of the GlcNAc thioacetate **4.21** that appears to be lower than that of the other glycosyl thioacetates used in this study. We performed an experiment in which we followed the deacetylation of the starting glycosyl thioacetate **4.21** over time and its conversion to the disulphide by NMR. After 1 h most of the fully protected thioacetate was still present and some was still detected after 3.5 h.

We tried to improve the yield of the aziridine opening reaction by adding the aziridine **4.29** (1 eq) after the thioacetate **4.21** (1.3 eq) was pretreated with the Et_2NH (1.9 eq) for 1 h at RT and then an

additional portion (0.3 eq) of the aziridine **4.29** was added 1 h after the first addition. Nevertheless, the analysis of the crude product showed that the main product was the disulphide and only traces of the aziridine opening product **4.39** were obtained.

4.4.8. Scope of the one-pot aziridine opening reaction

After examining the one-pot aziridine opening reaction on eight thioacetates we observed that three of them, α-Man thioacetate **4.4**, α-Rha thioacetate **4.8** and α-Neu5Ac thioacetate **4.17** afforded products with good yield and without isomerizing. A comprehensive overview of the scope of the reaction is presented in **Table 4.6**.

Table 4.6 Scope of the aziridine opening reaction[a]

Entry	Sugar	T (°C)	Product	β/α ratio[b] (yield[c] %)
1	Man **4.4**	20	**4.32**	α only (82)
2	Rha **4.8**	20	**4.33**	α only (92)
3	Glc **4.9**	0	**4.34**	10:1 (34)
4	Glc **4.9**	20	**4.34**	2:1 (61)
5	Gal **4.10**	0	**4.35**	4.5:1 (36)
6	Gal **4.10**	20	**4.35**	3:1 (44)
7	Lac **4.12**	0	**4.36**	5:1 (42)
8	Fuc **4.14**	20	**4.37**	5:1 (34)
9	Neu5Ac **4.17**	20	**4.38**	α only (56)[d]
10	GlcNAc **4.21**	20	**4.39**	2:1 (19)

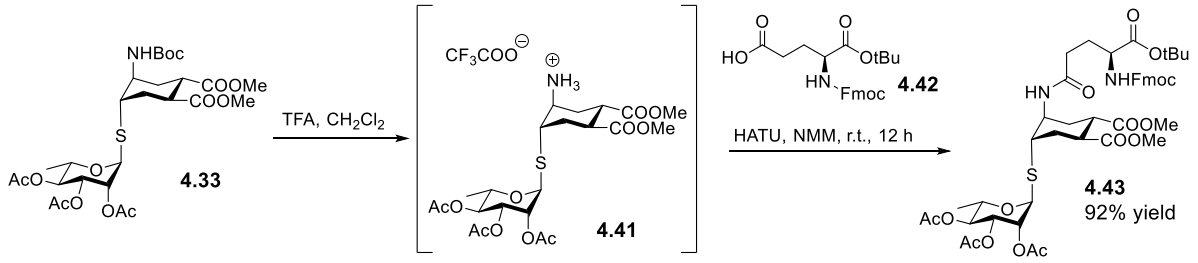
*a) All reactions were performed with a 0.65 M concentration of aziridine **4.29** and 1.9 mol equiv of Et_2NH in DMF, at the indicated temperature; b) evaluated by ^{1}H-NMR; c) isolated, both isomers; d) contains 17% glycal*

4.5. Further functionalization of aziridine opening products

The *N*-linkage installed on the pseudo-anomeric position of the aziridine opening products provides an opportunity for these molecules to be further functionalized. Possibilities for functionalization include synthesis of pseudo-glycosylamino acids to be used in synthesis of pseudo-glycopeptides and synthesis of multivalent constructs.

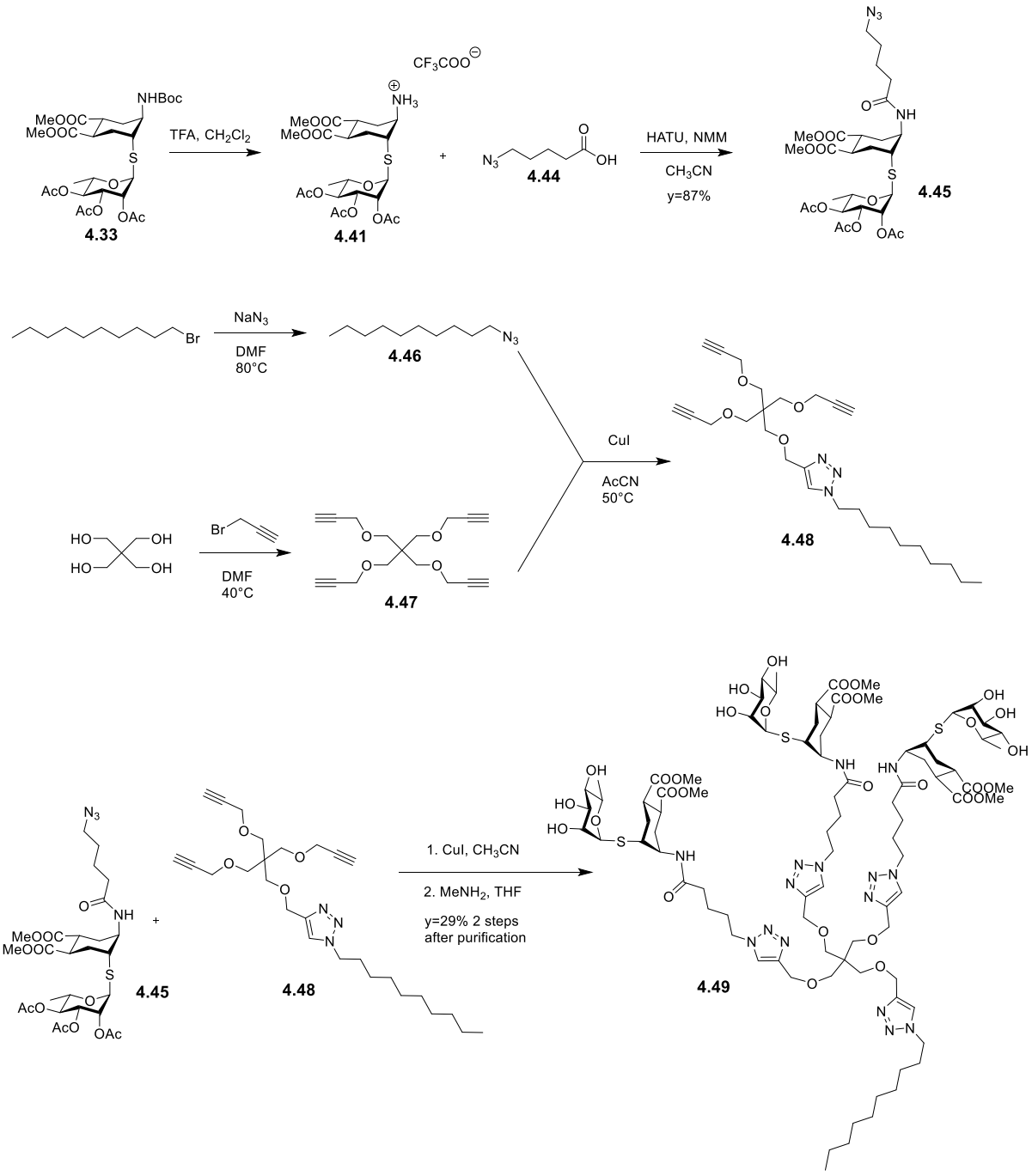
The Man and Rha derivatives performed best in the aziridine opening reaction, with the products **4.32** and **4.33** formed as single α-isomers in 82% and 92% yield, respectively. As the Man derivatives were previously explored in the Bernardi's group as DC-SIGN ligands, in this work we rather focused on exploring Rha derivatives.

The *thio*-rhamno conjugate **4.33** was successfully used in the preparation of an *N*-linked pseudo-glycosylaminoacid scaffold, a viable building block for solid phase pseudo-glycopeptide synthesis (**Scheme 4.18**). After removal of the Boc protecting group (TFA, quant), the resulting amine **4.41** was coupled in solution with the side chain carboxyl group of an appropriately protected glutamic acid **4.42**, using HATU as the coupling agent. The pseudo-glycosylamino acid **4.43** was isolated in 92% yield after chromatography.



Scheme 4.18 Coupling reaction of Rha product **4.33** with glutamic acid **4.42** to give pseudo-glycosylamino acid **4.43**

A multivalent construct was also synthesized from the *thio*-rhamno conjugate **4.33** (Scheme **4.19**)**.** After Boc deprotection (TFA, quant), the resulting amine **4.41** was coupled with 5-azido-valeric acid **4.44** using HATU as the coupling agent to give **4.45** in 87% yield. On the other hand, the tetravalent alkyne **4.48** was synthesized from pentaeritritol with propargyl bromide and treated with 10-azidodecane **4.46** to afford the trivalent alkyne **4.48**. The monovalent rhamnose-azide conjugate **4.45** was then used in a click reaction with **4.48** using CuI/CH_3CN and the final multivalent construct **4.49** was obtained in 29% yield after deacetylation of hydroxyl groups with $MeNH_2$/THF and purification with size exclusion chromatography.



Scheme 4.19 Synthetic strategy towards the multivalent construct **4.49**

4.6. Conclusions and future perspectives

In this part of the thesis we developed a novel class of glycomimetic compounds with a general structure of *N*-linked-pseudo-*thio*-glycosides.[30] To simplify the synthesis of these glycomimetics we aimed to develop a new synthetic approach, a so called one-pot aziridine opening reaction.

The synthetic approach that we devised is a facile and operationally simpler alternative to more standard glycosylation methods that often require very strict reaction conditions. In fact, the aziridine opening reaction can be performed as a one-pot procedure with *in situ* formation of a nucleophile thiolate from glycosyl thioacetate and a subsequent attack on the aziridine ring. To synthesize the glycomimetics we used thioacetate derivatives of different mono- and disaccharides in this reaction.

In some cases, we observed that the final products were formed as a mixture of isomers, which was an unexpected and rather intriguing discovery. We carefully investigated the tuning of the reaction conditions to improve the anomeric ratio to synthetically useful levels.

The most important outlook of this work is certainly that while we developed and optimized this synthetic strategy on a model aziridine substrate, the method could be potentially applied on different aziridine substrates. Aziridines can be found in various biologically active natural products and they have enormous potential in synthesis due to their versatility. An interesting application of our approach could for example be the use of glycosyl thioacetates in ring opening reactions with aziridine-containing peptides to form pseudo-glycopeptides.

The glycomimetics that we synthesized through the aziridine opening approach can be potentially recognized by various lectins based on the sugar moiety employed. The compounds also have an improved hydrolytic and enzymatic stability due to the *S*- and *N*-linkages inserted in the structure. The *N*-linkage on the pseudo-anomeric position can be exploited for further functionalization for example, it was successfully utilized in the synthesis of rhamnose-glutamic acid conjugate and a rhamnose multivalent construct.

L-Rhamnose is a particularly interesting sugar that is not present on human cells, but it is often found on bacterial cells. In the last few years rhamnose has become an interesting compound for possible use in immunotherapy, since anti-rhamnose antibodies were identified as the most abundant anti-carbohydrate antibodies in human serum.[31] These antibodies could be directed towards tumor cells in order to induce antitumor immune response. Recruitment of antibodies could be achieved with the help of an antibody recruiting molecule.[32] Such molecule should consist of two parts: an antibody binding module (antigen) and a tumoral binding module (**Fig. 4.31**). The tumoral binding module should selectively bind to a biomarker present of tumor cells, while the antibody binding module could

be a rhamnoside. Upon binding of the antibodies to rhamnoside now displayed on the tumor cells, antitumor immune response would be induced (**Fig. 4.31**). Several rhamnosides were developed for this purpose[33, 34], including multivalent rhamnose dendron (**Fig. 4.31**).[35] We displayed that our synthesized rhamnose glycomimetic **4.33** could be potentially used for multivalent presentation, as well as for easy conjugation to tumor-binding module, while it also benefits from an improved hydrolytic and enzymatic stability compared to previously developed rhamnosides.

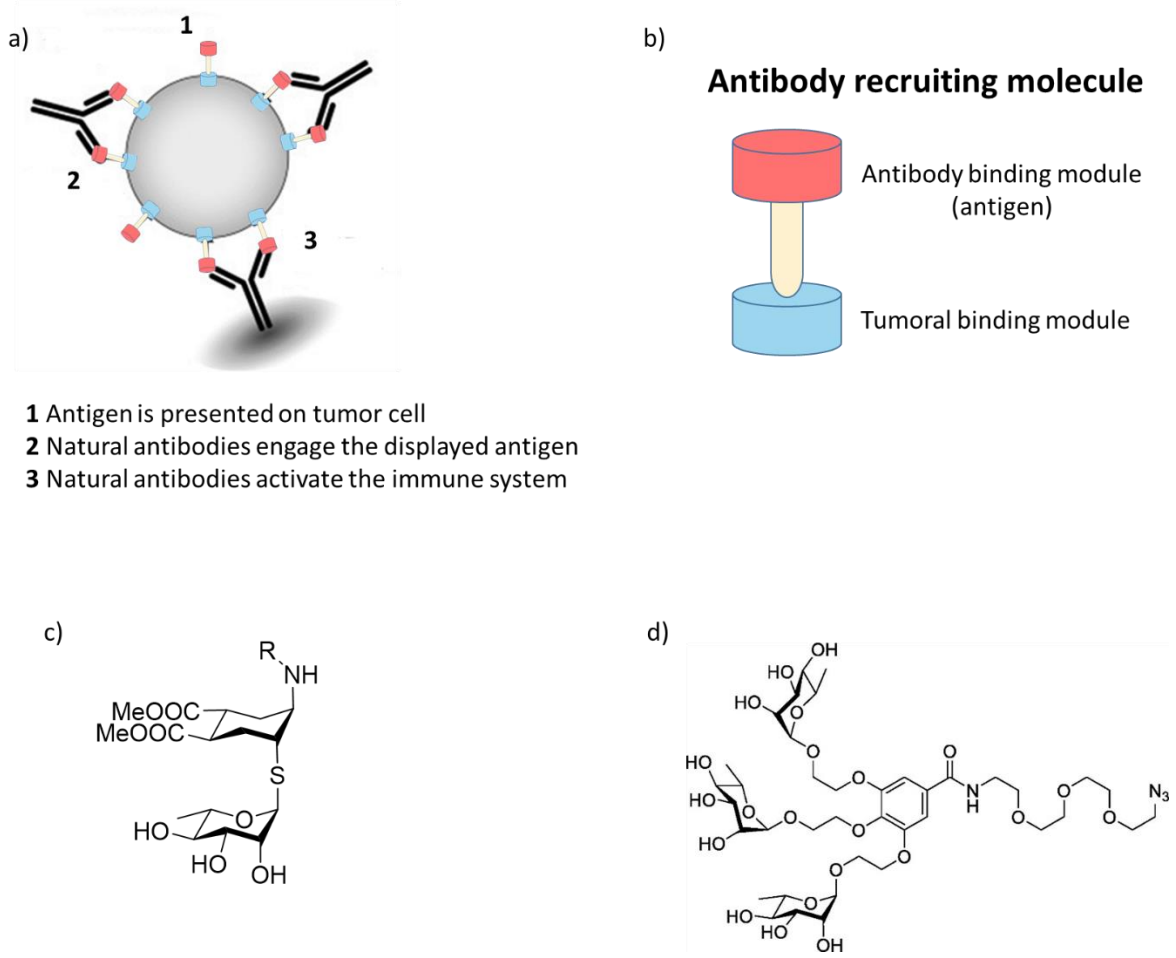


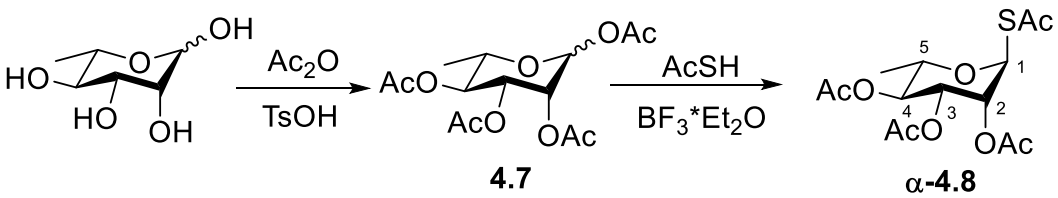
Fig. 4.31 a) Strategy to recruit natural (anti-rhamnose) antibodies towards tumor cells with the help of antibody recruiting molecule[32], b) Composition of antibody recruiting molecule, c) L-Rhamnose glycomimetic developed in our study, d) Multivalent L-rhamnose dendron[35]

4.7. Experimental

Chemicals were purchased from commercial sources and used without further purification, unless otherwise indicated. When anhydrous conditions were required, the reactions were performed under nitrogen or argon atmosphere. Anhydrous solvents were purchased from Sigma-Aldrich® with a content of water ≤0.005%. Triethylamine (TEA), methanol and dichloromethane were dried over calcium hydride, THF was dried over sodium/benzophenone and freshly distilled. *N*,*N*-dimethylformamide (DMF) was dried over 4Å molecular sieves. Reactions were monitored by analytical thin-layer chromatography (TLC) performed on Silica Gel 60 F_{254} plates (Merck), and TLC Silica gel 60 RP-18 F_{254}s (Merck) with UV detection (254 nm and 365 nm) and/or staining with ammonium molybdate acid solution, potassium permanganate alkaline solution or ninhydrin. Flash column chromatography was performed using silica gel 60 (40-63 µm, Merck). Automated flash chromatography was performed with Biotage Isolera Prime system and SNAP ULTRA cartridges were employed. Microwave irradiation was performed by a Biotage Initiator^{+} system. NMR experiments were recorded on a Bruker AVANCE-400 MHz or 600 MHz instruments at 298 K. Chemical shifts (δ) are reported in ppm. The ^{1}H and ^{13}C NMR resonances of compounds were assigned with the assistance of COSY, HSQC and in some cases NOESY experiments. Multiplicities are assigned as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), mult. (for multiplets encompassing more than one proton).

Mass spectra were recorded on Apex II ICR FTMS (ESI ionization-HRMS), Waters Micromass Q-TOF (ESI ionization-HRMS) or Thermo Fischer LCQ apparatus (ESI ionization). The mass spectrometer was operated with electrospray ionization in the positive ion mode. Full-scan mass spectra were recorded in the mass/charge (m/z) range of 50–2000. Liquid chromatography-mass spectrometry (LC-MS) analyses were carried out on a Thermo Fisher LCQ Fleet ion trap mass spectrometer equipped with a UPLC UltiMate™ 3000 system containing UV detector. A Zorbax RX-C18 (2.1x150 mm-5 µm) was used as column. The column oven was maintained at 30 °C. 5 µL of each sample solution were eluted using a binary gradient elution consisting of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) as follows: from 2% to 95% B in 25 min, 95% B kept for 10 min, then the eluent composition was brought at 2% B in 5 min. The flow rate was 0.25mL /min. The mass spectrometer was operated with electrospray ionization in the positive ion mode. Full-scan mass spectra were recorded in the mass/charge (m/z) range of 50–2000. Specific optical rotation values were measured using a Perkin-Elmer 241, at 589 nm in a 1 dm cell.

Synthesis of α-1-*S*-acetyl-2,3,4-tri-*O*-acetyl-L-rhamnopyranose **4.8**

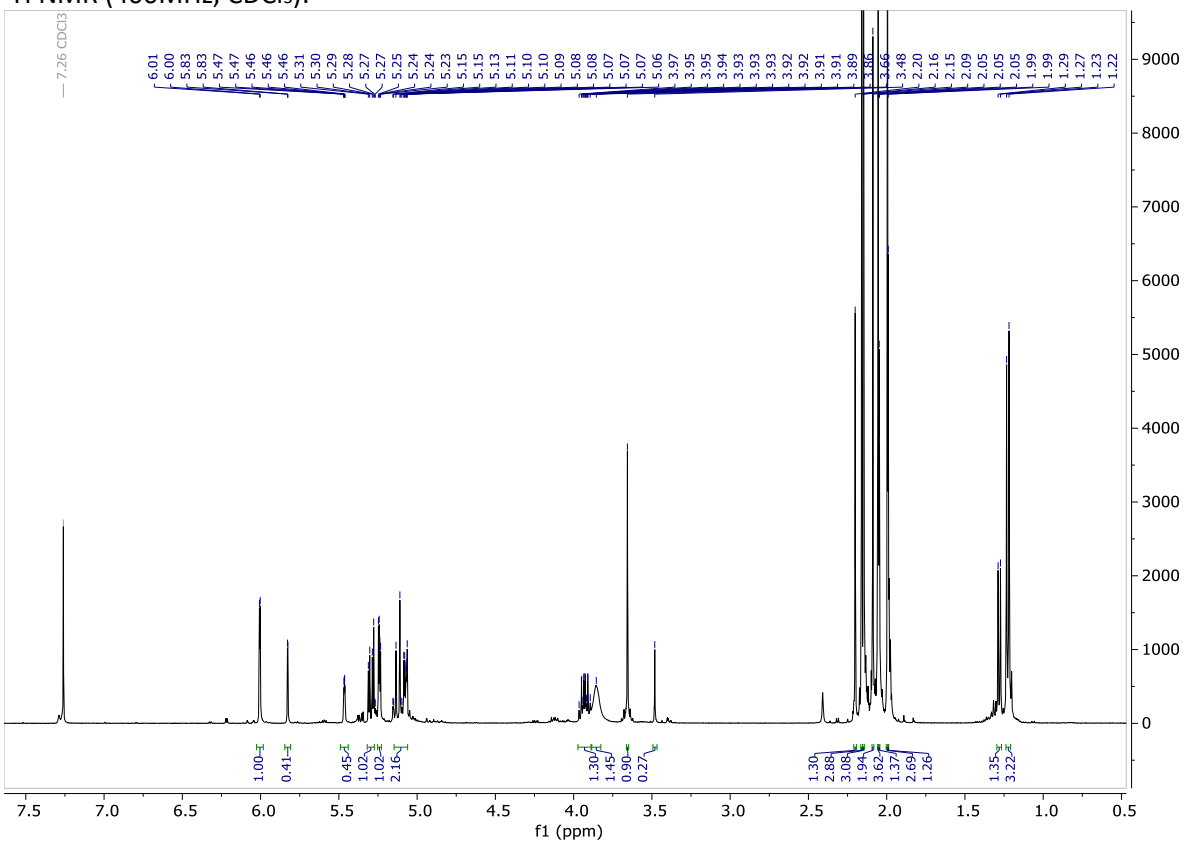


4.7 **α-4.8**

L-Rhamnose monohydrate (1.00 g, 5.49 mmol) and TsOH (0.09 g, 0.44 mmol) were dissolved in Ac_2O (3.11 mL, 32.94 mmol) at 0°C. After 10 minutes the ice bath was removed and the reaction stirred for 4.5 h at 30°C. Afterwards the reaction mixture was concentrated and dried under high vacuum. Crude **4.7**[11] was used in the next step without further purification.

^{1}H NMR (400MHz, $CDCl_3$) δ 6.00 (d, 1H, $J_{1\alpha\text{-}2\alpha}$ = 1.9 Hz, $H_{1\alpha}$), 5.83 (d, 1H, $J_{1\beta\text{-}2\beta}$ = 1.2 Hz, $H_{1\beta}$), 5.46 (m, 1H, $H_{2\beta}$), 5.24 (mult., 2H, $H_{3\alpha}$, $H_{2\alpha}$), 5.10 (mult., 2H, H_4, $H_{3\beta}$), 3.92 (m, 1H, $H_{5\alpha}$), 3.66 (m, 1H, $H_{5\beta}$), 2.20 (s, 3H OAc), 2.16 (s, 3H, OAc), 2.15 (s, 3H, OAc), 2.09 (s, 3H, OAc), 2.05 (s, 3H, OAc), 2.05 (s, 3H, OAc), 1.99 (s, 3H, OAc), 1.99 (s, 3H, OAc), 1.28 (d, 3H, $J_{CH3\beta\text{-}5}$ = 6.2 Hz, $CH_{3\beta}$), 1.23 (d, 3H, $J_{CH3\alpha\text{-}5}$ = 6.3 Hz, $CH_{3\alpha}$).

^{1}H NMR (400MHz, $CDCl_3$):

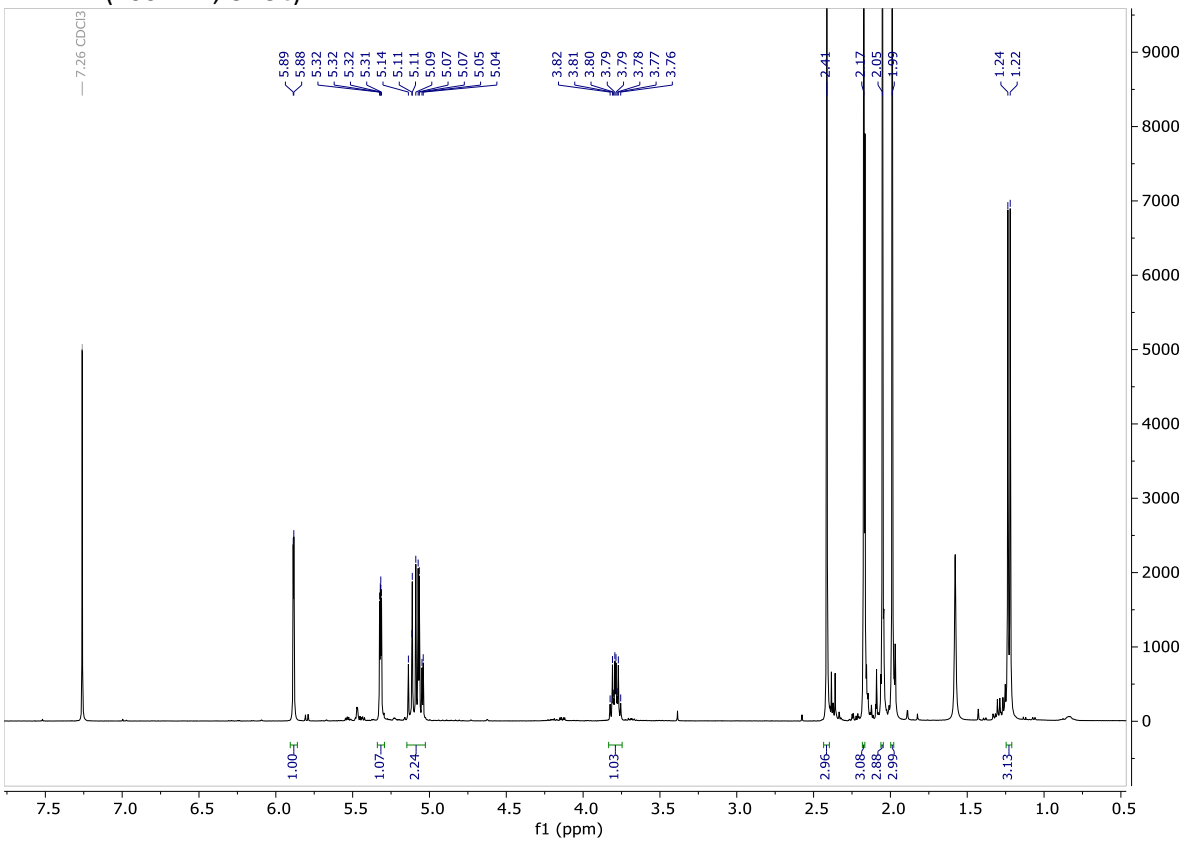


Crude tetra-*O*-acetyl-L-rhamnose **4.7** (5.49 mmol) was then dissolved in dry CH_2Cl_2 (25.0 mL) under N_2 atmosphere. The solution was cooled to 0°C. AcSH (1.64 mL, 23.00 mmol) and then BF_3*Et_2O (4.16 mL, 33.71 mmol) were added dropwise. The ice bath was removed and the reaction was kept stirring at

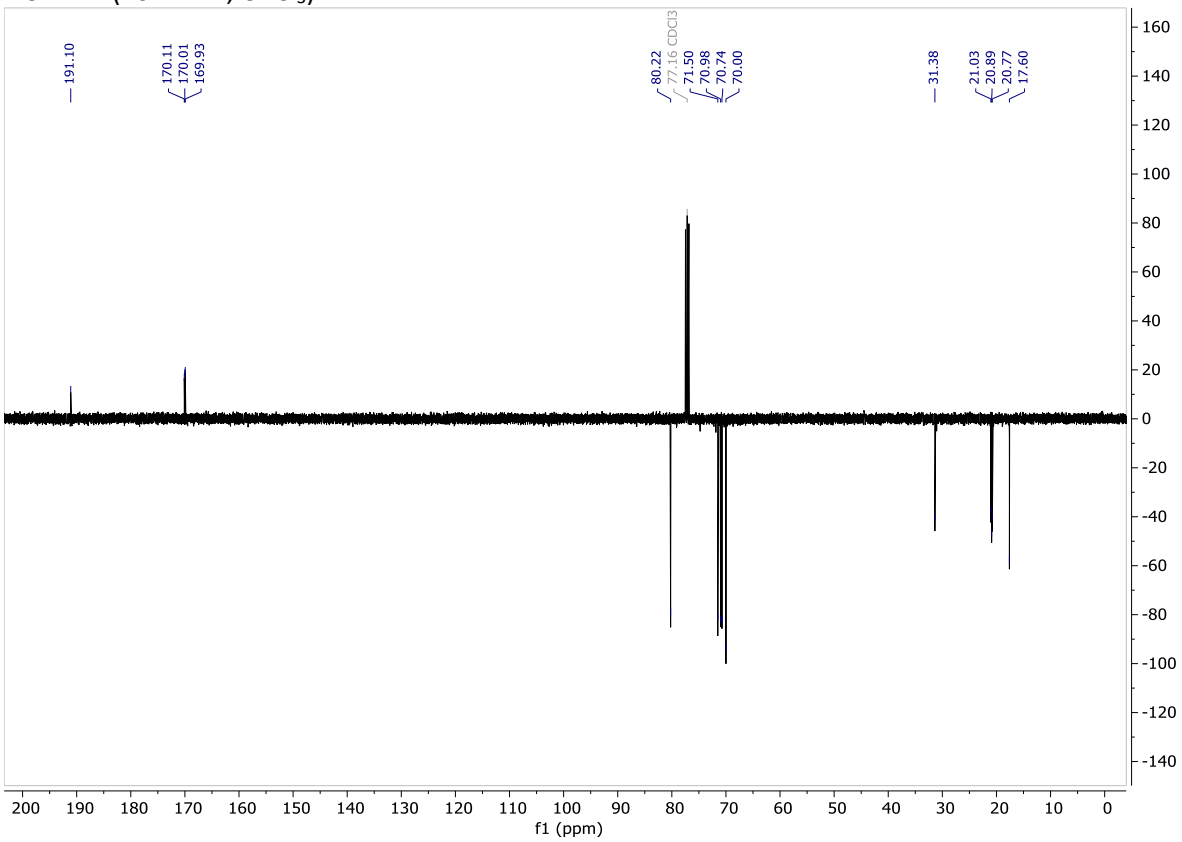
30°C under N_2 atmosphere for 3.5 h. The reaction mixture was diluted with CH_2Cl_2, then washed with H_2O, sat. $NaHCO_3$, H_2O and brine. Aqueous phases were washed with CH_2Cl_2. The combined organic phases were dried over Na_2SO_4 and concentrated under reduced pressure. NMR analysis showed the presence of both α and β isomers in 5:1 ratio, the crude was purified by automated flash chromatography (Biotage Isolera, HP Cartridge, 100% iPr_2O). Product **α-4.8**[12] was obtained as a yellow oil in 71% yield (1.44 g, 4.12 mmol).

R_f = 0.3 (Hex/EtOAc 3:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.88 (d, $J_{1\text{-}2}$ = 2 Hz, 1H, H_1), 5.32 (dd, $J_{2\text{-}3}$ = 3.1, $J_{2\text{-}1}$ = 2 Hz, 1H, H_2), 5.15 – 5.02 (mult., 2H, H_3, H_4), 3.79 (dq, $J_{5\text{-}4}$ = 9.2, $J_{5\text{-}CH3}$ = 6.2 Hz, 1H, H_5), 2.41 (s, 3H, SAc), 2.17 (s, 3H, OAc), 2.05 (s, 3H, OAc), 1.99 (s, 3H, OAc), 1.23 (d, $J_{CH3\text{-}5}$ = 6.2 Hz, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 191.1 (CS), 170.1 (CO), 170.0 (CO), 169.9 (CO), 80.2 (C_1), 71.5 (C_2), 71.0 (C_5), 70.7 (C_4), 70.0 (C_3), 31.4 (SAc), 21.0 (OAc), 20.9 (OAc), 20.8 (OAc), 17.6 (CH_3); MS (ESI) calcd for $C_{14}H_{20}O_8S$ $[M + Na]^+$ m/z: 371.08; found: 370.81.

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

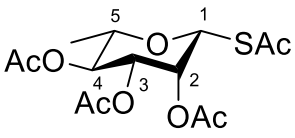


^{13}C NMR (101 MHz, $CDCl_3$)



^{1}H NMR assignment of β-1-*S*-acetyl-2,3,4-tri-*O*-acetyl-L-rhamnopyranose

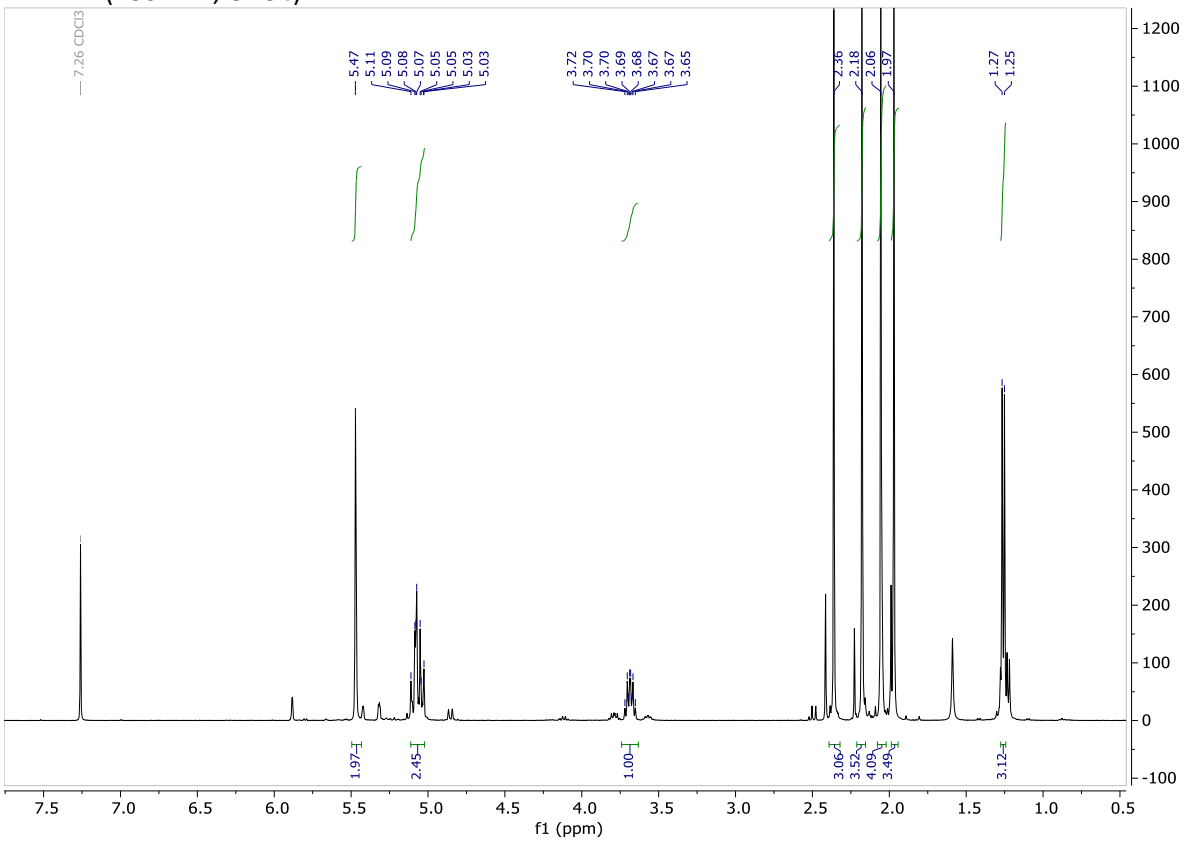
β-4.8



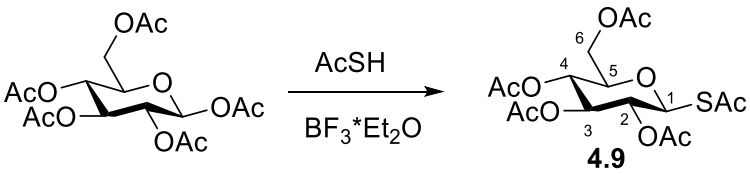
β-4.8

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.47 (s, 2H, H_1, H_2), 5.12 – 5.02 (m, 2H, H_3, H_4), 3.69 (dq, $J_{5\text{-}4}$ = 8.7, $J_{5\text{-}CH3}$ = 6.1 Hz, 1H, H_5), 2.36 (s, 3H, SAc), 2.18 (s, 3H, OAc), 2.06 (s, 3H, OAc), 1.97 (s, 3H, OAc), 1.26 (d, $J_{CH3\text{-}5}$ = 6.1 Hz, 3H, CH_3).

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



Synthesis of β-1-*S*-acetyl-2,3,4,6-tetra-*O*-acetyl-D-glucopyranose **4.9**

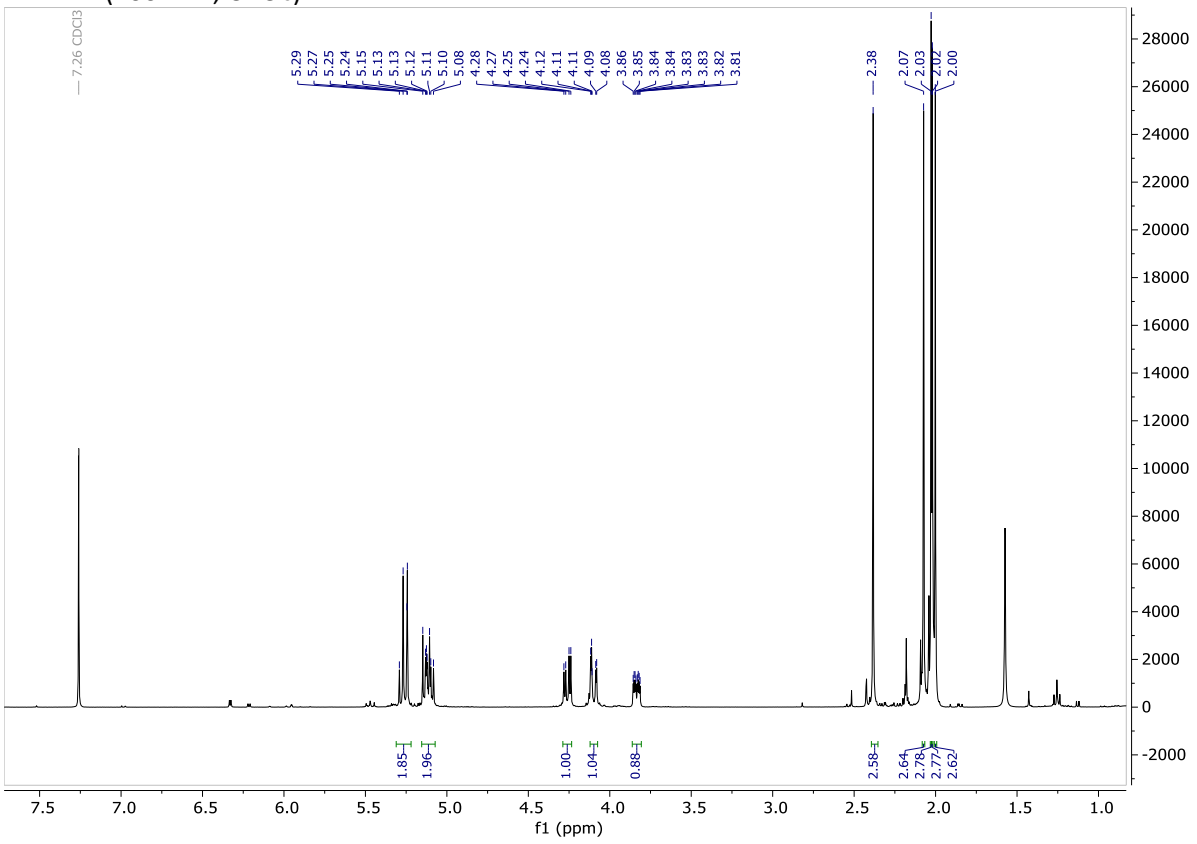


4.9

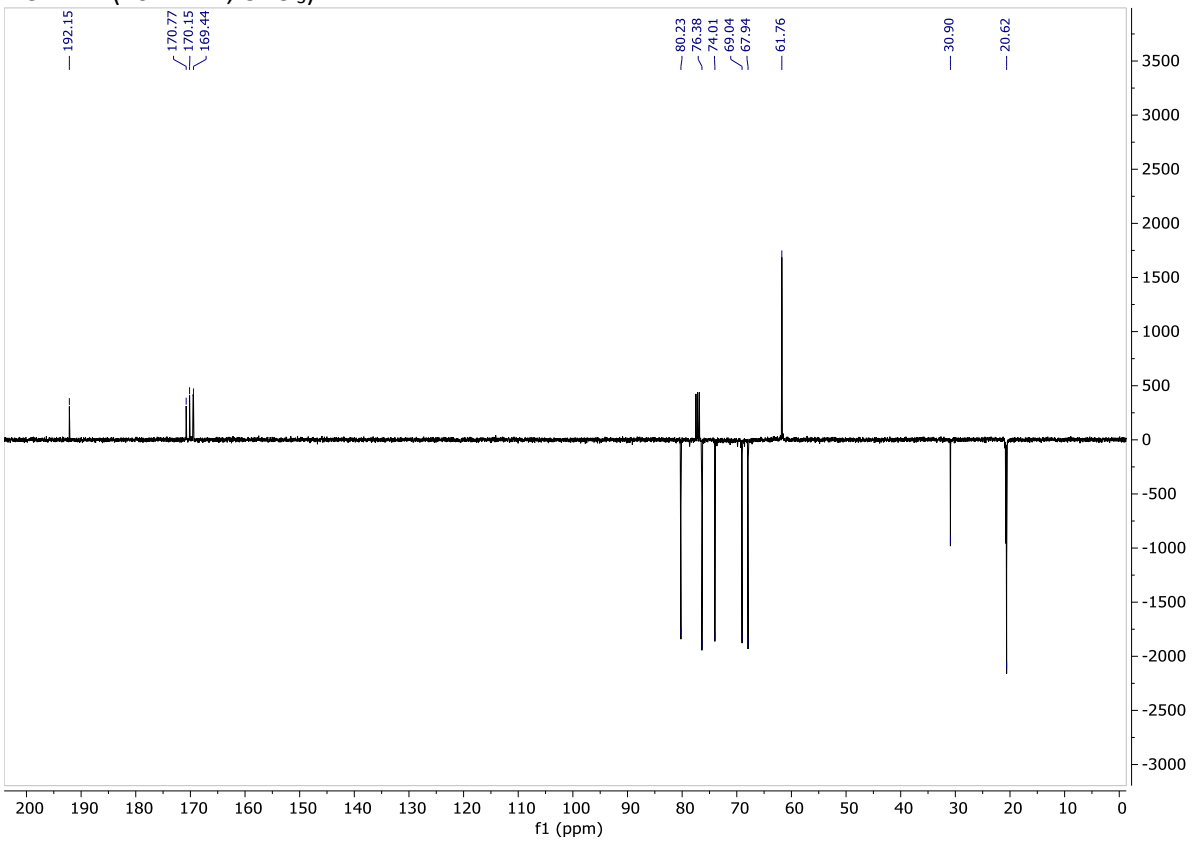
β-glucose pentaacetate (3.00 g, 7.68 mmol) was dissolved in dry CH_2Cl_2 (37.5 mL) under N_2 atmosphere and the solution cooled to 0°C. AcSH (2.3 mL, 32.18 mmol) and then BF_3*Et_2O (5.82 mL, 47.15 mmol) were added dropwise. The ice bath was removed and the reaction was kept stirring at 30°C under N_2 atmosphere for 3h. The reaction mixture was diluted with CH_2Cl_2 and then the organic phase was washed with H_2O, sat. $NaHCO_3$, H_2O and brine. The combined aqueous phases were further extracted with CH_2Cl_2. The combined organic phases were dried over Na_2SO_4 and then concentrated under vacuum. The crude was purified by flash chromatography (iPr_2O/EtOAc=8:2) and the product **4.9**[12] was obtained as a white solid in 97% yield (3.02 g, 7.43 mmol).

R_f = 0.5 (iPr_2O/EtOAc 8:2); ^1H NMR (400 MHz, $CDCl_3$) δ 5.30 – 5.24 (mult., 2H, H_3, H_1), 5.15 – 5.08 (mult., 2H, H_2, H_4), 4.26 (dd, $J_{6a\text{-}6b}$ = 12.5, $J_{6a\text{-}5}$ = 4.5 Hz, 1H, H_{6a}), 4.10 (dd, $J_{6b\text{-}6a}$ = 12.5, $J_{6b\text{-}5}$ = 2.2 Hz, 1H, H_{6b}), 3.84 (ddd, $J_{5\text{-}4}$ = 10.1, $J_{5\text{-}6a}$ = 4.5, $J_{5\text{-}6b}$ = 2.2 Hz, 1H, H_5), 2.38 (s, 3H, SAc), 2.07 (s, 3H, OAc), 2.03 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.00 (s, 3H, OAc); ^{13}C NMR (101 MHz, $CDCl_3$) δ 192.2 (CS), 170.8 (CO), 170.2 (CO), 169.5 (CO), 169.4 (CO), 80.2 (C_1), 76.4 (C_5), 74.0 (C_3), 69.0 (C_2), 67.9 (C_4), 61.8 (C_6), 30.9 (SAC), 20.8 (OAc), 20.6 (OAc), 20.6 (OAc); MS (ESI) calcd for $C_{16}H_{22}O_{10}S$ $[M + Na]^+$ m/z: 429.08; found: 428.76.

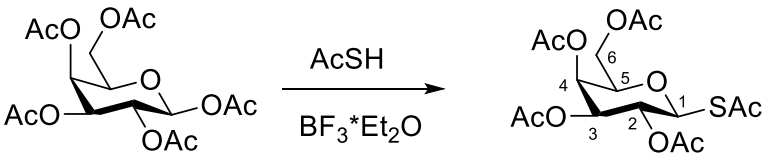
^1H NMR (400MHz, $CDCl_3$)



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



Synthesis of β-1-*S*-acetyl-2,3,4,6-tetra-*O*-acetyl-D-galactopyranose **4.10**



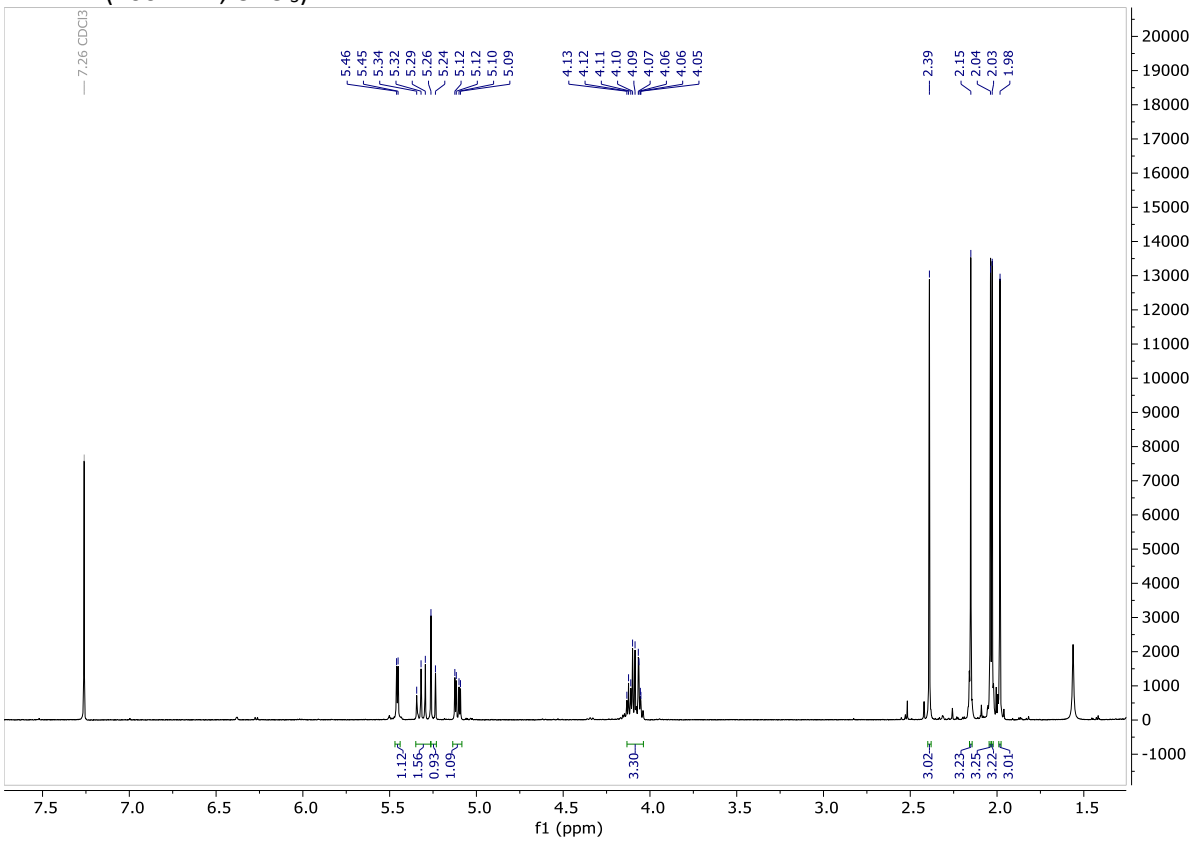
4.10

β-galactose pentaacetate (3.00 g, 7.68 mmol) was dissolved in dry CH_2Cl_2 (37.5 mL) under N_2 atmosphere and the solution was cooled to 0°C. AcSH (2.30 mL, 32.17 mmol) and then BF_3*Et_2O (5.82 mL, 47.15 mmol) were added dropwise. Ice bath was removed and the reaction was kept stirring at 30°C under N_2 atmosphere for 5h.

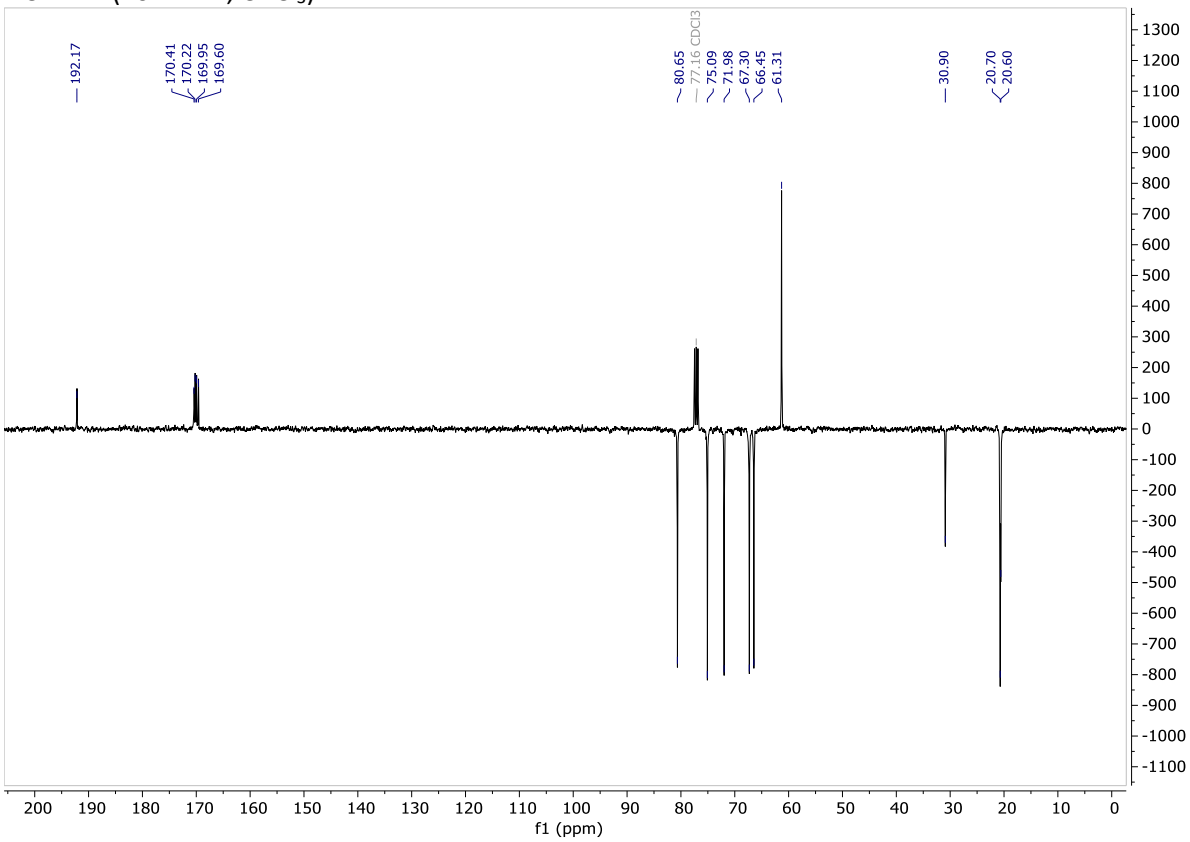
Reaction mixture was diluted with CH_2Cl_2 and then the organic phase was washed with H_2O, sat. $NaHCO_3$, H_2O and brine. Combined aqueous phases were washed with CH_2Cl_2. The combined organic phases were dried over Na_2SO_4 and then concentrated under vacuum. Crude was further purified by filtration on silica (eluent: 100 % EtOAc). The product **4.10**[12] was obtained as a white solid in 96% yield (3.00 g, 7.39 mmol).

R_f = 0.4 (iPr_2O/EtOAc 8:2); ^1H NMR (400 MHz, $CDCl_3$) δ 5.46 (dd, $J_{4\text{-}3}$ = 3.5 Hz, 1H, H_4), 5.36 – 5.26 (m, 1H, H_2), 5.25 (d, $J_{1\text{-}2}$ = 10.3 Hz, 1H, H_1), 5.11 (dd, $J_{3\text{-}2}$ = 9.6, $J_{3\text{-}4}$ = 3.5 Hz, 1H, H_3), 4.15 – 4.02 (mult., 3H, H_5, H_{6a}, H_{6b}), 2.39 (s, 3H, SAc), 2.15 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.03 (s, 3H, OAc), 1.98 (s, 3H, OAc); ^{13}C NMR (101 MHz, $CDCl_3$) δ 192.2 (CS), 170.4 (CO), 170.2 (CO), 170.0 (CO), 169.6 (CO), 80.7 (C_1), 75.1 (C_5), 72.0 (C_3), 67.3 (C_4), 66.5 (C_2), 61.3 (C_6), 30.9 (SAc), 20.7 (2xOAc), 20.6 (2xOAc); MS (ESI) calcd for $C_{16}H_{22}O_{10}S$ $[M + Na]^+$ m/z: 429.08; found: 428.80.

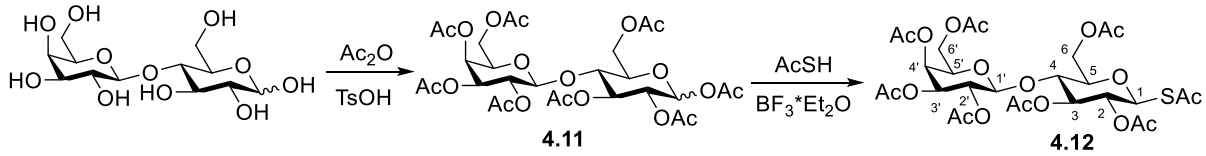
^1H NMR (400 MHz, $CDCl_3$)



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



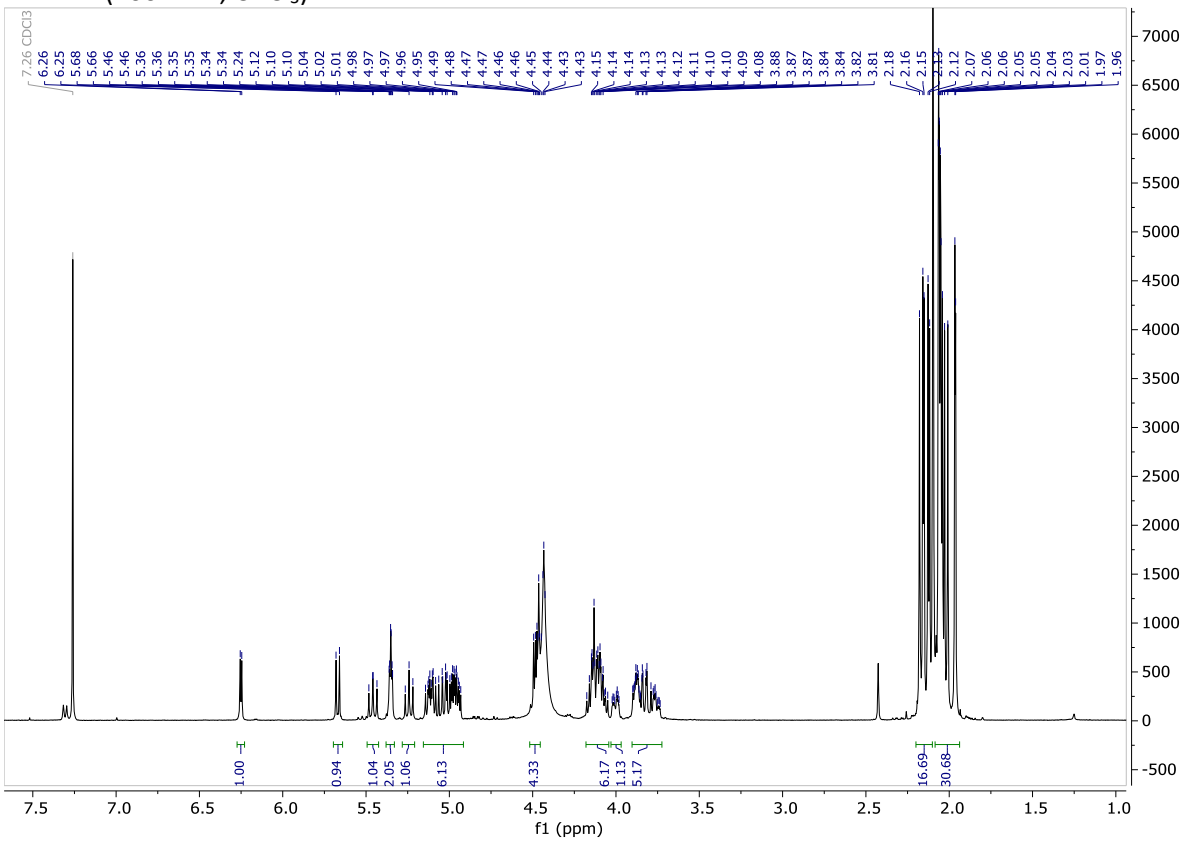
Synthesis of 2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl-(1→4)-β-1-*S*-acetyl-2,3,6-tri-*O*-acetyl-D-glucopyranose **4.12**



D-Lactose (0.100 g, 0.29 mmol) and TsOH (0.009 g, 0.044 mmol) were suspended in Ac_2O (0.30 mL, 3.16 mmol) at 0°C. After 10 min the ice bath was removed and the reaction mixture was kept stirring at 40°C overnight, then concentrated and dried under high vacuum. Crude **4.11**[11] was used further without purification.

^{1}H NMR (400 MHz, $CDCl_3$) δ 6.25 (d, $J_{2\alpha\text{-}1\alpha}$ = 3.7 Hz, 1H, $H_{1\alpha}$), 5.67 (d, $J_{1\beta\text{-}2\beta}$ = 8.3 Hz, 1H, $H_{1\beta}$), 5.46 (dd, $J_{3\alpha\text{-}4\alpha}$ = 10.3, $J_{3\alpha\text{-}2\alpha}$ = 9.2 Hz, 1H, $H_{3\alpha}$), 5.35 (td, $J_{4'\text{-}3'}$ = 3.5, $J_{4'\text{-}5'}$ = 1.2 Hz, 2H, $H_{4'\alpha,\beta}$), 5.24 (t, $J_{3\beta\text{-}4\beta} = J_{3\beta\text{-}2\beta}$ = 9.1 Hz, 1H, $H_{3\beta}$), 5.16 – 4.93 (mult., 6H, $H_{2'\alpha,\beta}$, $H_{2\alpha,\beta}$, $H_{3'\alpha,\beta}$), 4.52 – 4.38 (mult., 4H, $H_{1'\alpha,\beta}$, $H_{6b\alpha,\beta}$), 4.20 – 4.04 (mult., 6H, $H_{6'b\alpha,\beta}$, $H_{6a\alpha,\beta}$, $H_{6'a\alpha,\beta}$), 4.00 (ddd, $J_{5\alpha\text{-}4\alpha}$ = 10.1, $J_{5\alpha\text{-}6a\alpha}$ = 4.3, $J_{5\alpha\text{-}6b\alpha}$ =2.1 Hz, 1H, $H_{5\alpha}$), 3.90 – 3.72 (mult., 5H, $H_{5'\alpha,\beta}$, $H_{4\alpha,\beta}$, $H_{5\beta}$), 2.15 (singlets, 48H, OAc).

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

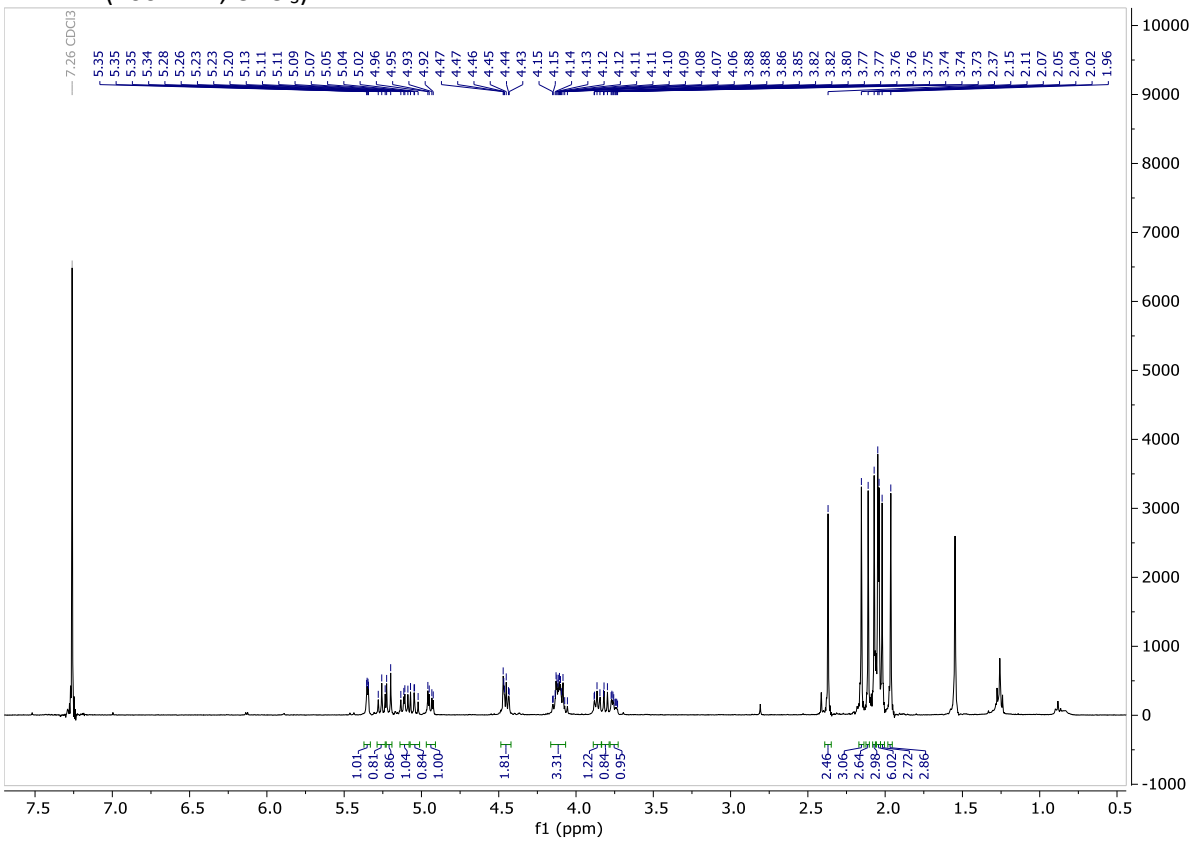


Crude **4.11** (0.29 mmol) was then dissolved in dry CH_2Cl_2 (1.4 mL) at 0°C under N_2 atmosphere, then AcSH (87.5 µL, 1.22 mmol) and BF_3*Et_2O (0.22 mL, 1.79 mmol) were added. Ice bath was removed and the reaction mixture was kept stirring at 30°C overnight.

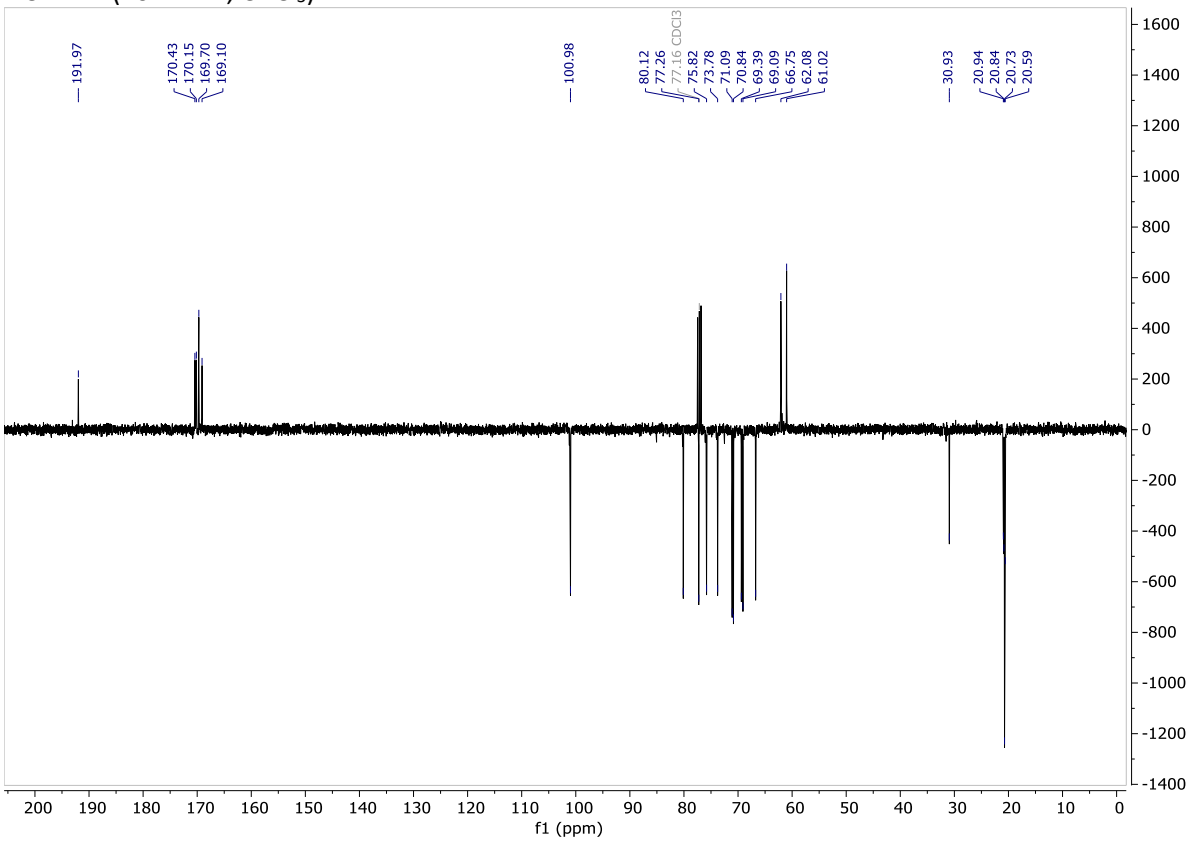
Reaction mixture was diluted with CH_2Cl_2 and washed with H_2O, sat. $NaHCO_3$, H_2O and brine. Aqueous phases were washed with CH_2Cl_2. Combined organic phases were dried over Na_2SO_4 and concentrated under vacuum. Crude product was purified with flash chromatography (eluent Hex/EtOAc 1:1) to give **4.12**[12] as colorless waxy solid in 65% yield (0.130 g, 0.189 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.35 (dd, $J_{4'-3'}$ = 3.5, $J_{4'-5'}$ = 1.2 Hz, 1H, $H_{4'}$), 5.26 (dd, J_{3-2} = J_{3-4} = 9.0 Hz, 1H, H_3), 5.21 (d, J_{1-2} = 10.5 Hz, 1H, H_1), 5.11 (dd, $J_{2'-1'}$ = 10.5, $J_{2'-3'}$ = 7.8 Hz, 1H, $H_{2'}$), 5.05 (dd, J_{2-1} = 10.5, J_{2-3} = 9.0 Hz, 1H, H_2), 4.94 (dd, $J_{3'-2'}$ = 10.4, $J_{3'-4'}$ = 3.5 Hz, 1H, $H_{3'}$), 4.49 – 4.41 (mult., 2H, $H_{1'}$, $H_{6'a}$), 4.17 – 4.05 (mult., 3H, H_{6a}, H_{6b}, $H_{6'b}$), 3.91 – 3.84 (m, 1H, $H_{5'}$), 3.83 – 3.79 (m, 1H, H_4), 3.79 – 3.71 (m, 1H, H_5), 2.37 (s, 3H, SAc), 2.15 (s, 3H, OAc), 2.11 (s, 3H, OAc), 2.07 (s, 3H, OAc), 2.05 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.02 (s, 3H, OAc), 1.96 (s, 3H, OAc); ^{13}C NMR (101 MHz, $CDCl_3$) δ 192.0 (CS), 170.5 (CO), 170.4 (CO), 170.2 (CO), 170.2 (CO), 169.7 (CO), 169.1 (CO), 101.0 ($C_{1'}$), 80.1 (C_1), 77.3 (C_5), 75.8 (C_4), 73.8 (C_3), 71.1 ($C_{5'}$), 70.8 ($C_{3'}$), 69.4 (C_2), 69.1 ($C_{2'}$), 66.7 ($C_{4'}$), 62.1 (C_6), 61.0 ($C_{6'}$), 30.9 (SAc), 20.9 (OAc), 20.8 (OAc), 20.7 (OAc), 20.6 (OAc); MS (ESI) calcd for $C_{28}H_{38}O_{18}S$ $[M + Na]^+$ m/z: 717.17; found: 716.74.

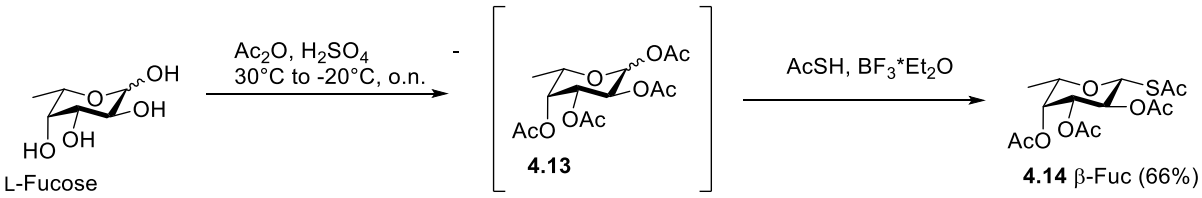
^{1}H NMR (400 MHz, $CDCl_3$)



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



Synthesis of β-1-*S*-acetyl-2,3,4-tetra-*O*-acetyl-L-fucopyranose **4.14**



L-fucose (0.2 g, 1.22 mmol) was dissolved in dry CH_3CN (0.2 mL) and Ac_2O (0.55 mL, 5.85 mmol) at -30°C under N_2 atmosphere, then H_2SO_4 (5.1 µL, 0.096 mmol) was added. Reaction mixture was kept stirring at -30°C for 1 h, then warmed to -20°C and kept stirring overnight at this temperature.

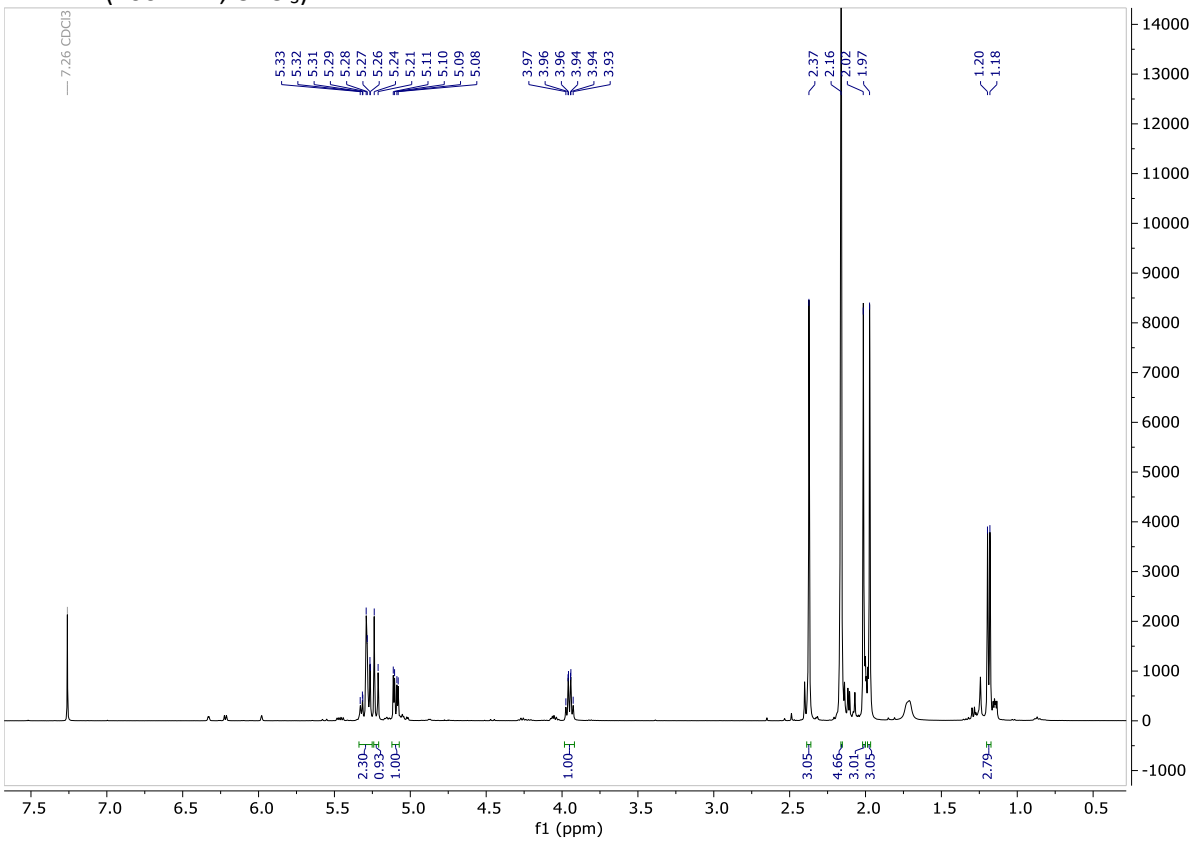
The next day dry CH_3OH (0.059 mL, 1.46 mmol) was added to quench the reaction at -20°C, then allowing the reaction mixture to warm up to 0°C over 1 h.

Dry CH_2Cl_2 (5 mL) was added to the reaction mixture containing **4.13**[11] at 0°C under N_2 atmosphere, followed by the addition of HSAc (0.37 mL, 5.11 mmol) and BF_3xEt_2O (0.92 mL, 7.48 mmol). The reaction mixture was then left to stir overnight at RT.

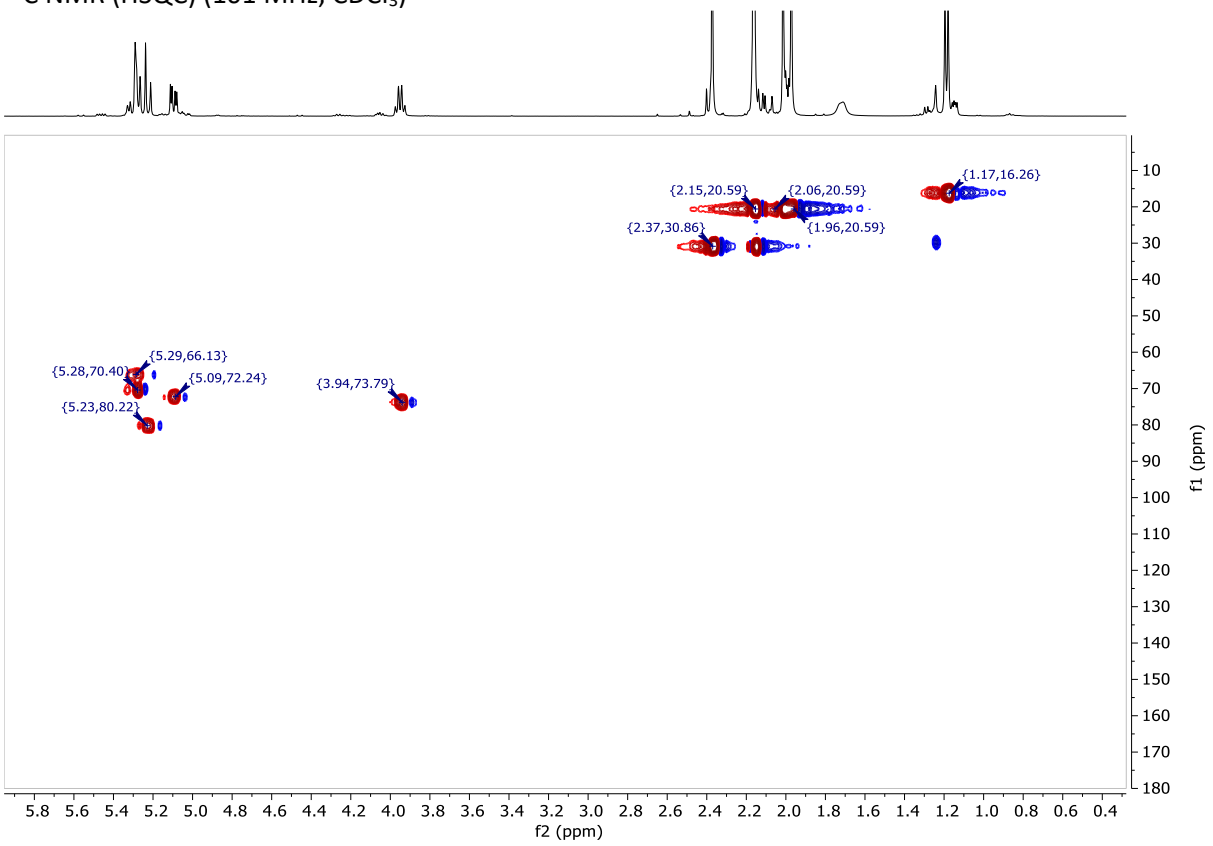
Reaction mixture was poured into 10 mL of cold H_2O and diluted with 5 mL of CH_2Cl_2. Organic layer was washed with $NaHCO_3$, H_2O and NaCl. Aqueous phases were washed with CH_2Cl_2. Combined organic phases were dried over Na_2SO_4 and concentrated under reduced pressure. Crude was further purified with column chromatography (Hexane:EtOAc 3:1) to give product **4.14** as colorless oil in 66% yield (0.28 g, 0.80 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.34 – 5.26 (m, 2H, H_4, H_3), 5.22 (d, $J_{1\text{-}2}$= 10.2 Hz, 1H, H_1), 5.10 (dd, $J_{2\text{-}1}$= 10 Hz, $J_{2\text{-}3}$= 3.4 Hz, 1H, H_2), 3.95 (dt, $J_{5\text{-}4}$= 7.1 Hz, $J_{5\text{-}CH3}$= 6 Hz, 1H, H_5), 2.37 (s, 3H, SAc), 2.16 (s, 3H, OAc), 2.02 (s, 3H, OAc), 1.97 (s, 3H, OAc), 1.19 (d, $J_{CH3\text{-}5}$= 6.4 Hz, 3H, CH_3); ^{13}C NMR (HSQC) (101 MHz, $CDCl_3$) δ 80.2 (C_1), 73.8 (C_5), 72.2 (C_2), 70.4 (C_3), 66.1 (C_4), 30.9 (SAc), 20.6 (3xOAc), 16.3 (CH_3); MS (ESI) calcd for $C_{14}H_{20}O_8S$ $[M + Na]^+$ m/z: 371.08; found: 371.18.

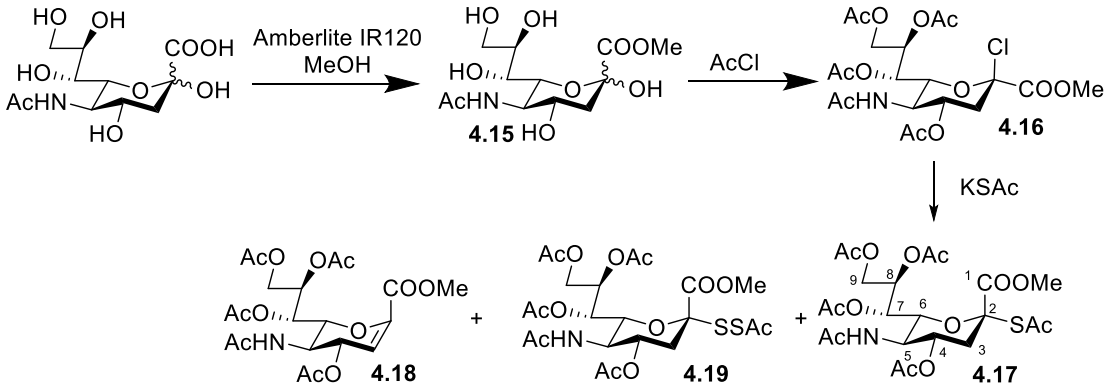
^{1}H NMR (400 MHz, $CDCl_3$)



^{13}C NMR (HSQC) (101 MHz, $CDCl_3$)



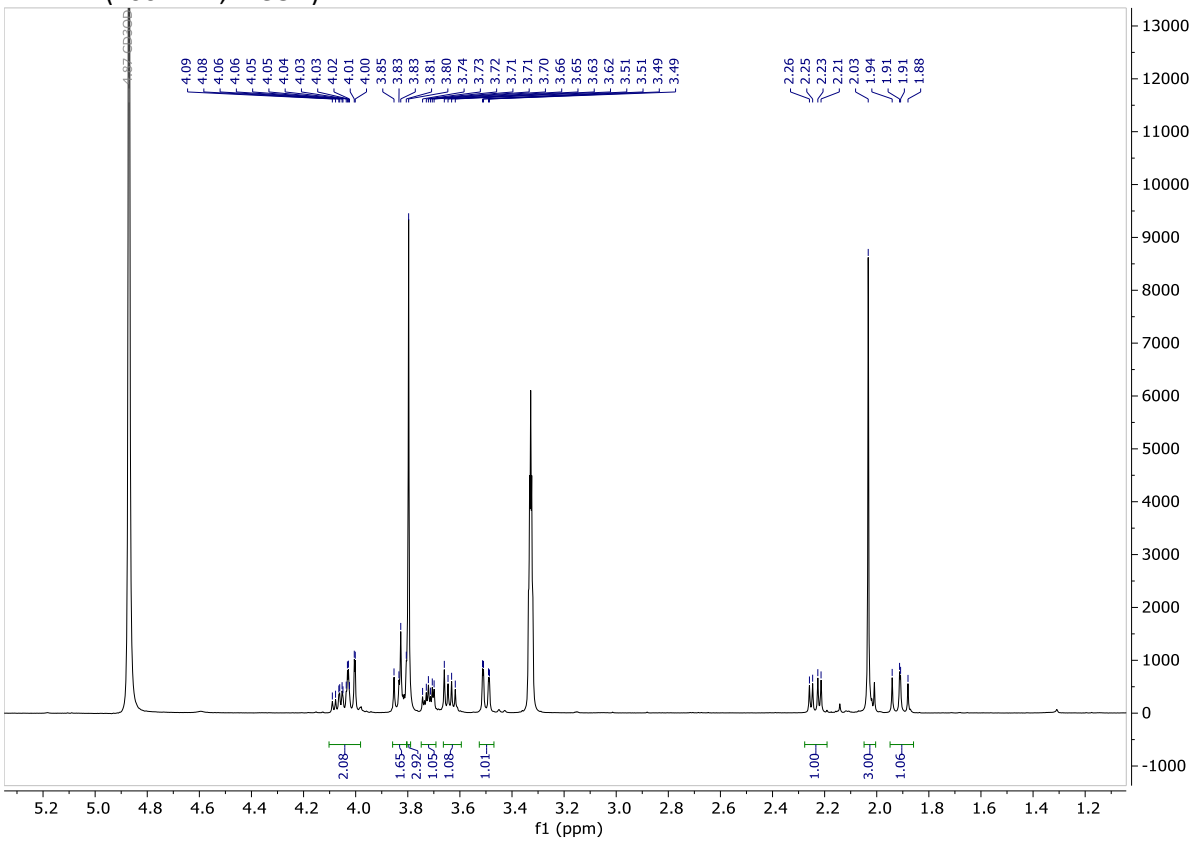
Synthesis of α-2-*S*-acetyl-4,7,8,9-tetra-*O*-acetyl-*N*-acetylneuraminic acid methyl ester **4.17**



N-acetylneuraminic acid (2.00 g, 6.47 mmol) and Amberlite IR120 (5 g) were suspended in dry MeOH (200 mL) under N_2 atmosphere. The reaction was kept stirring overnight at room temperature, the mixture was filtered with Buchner funnel and the filtrate was concentrated under vacuum. The known methyl ester **4.15**[36] was obtained as a white solid in 85% yield (1.79 g, 5.53 mmol) and used in the following reaction without further purification.

R_f = 0.3 (iPr_2O/EtOAc/H_2O 4:1:1); ^{1}H NMR (400 MHz, MeOD) δ 4.12 – 4.00 (mult., 2H, H_4, H_6), 3.89 – 3.80 (mult., 2H, H_5, H_{9a}), 3.81 (s, 3H, COOMe), 3.77 – 3.71 (m, 1H, H_8), 3.65 (dd, $J_{9b\text{-}9a}$ = 11.2, $J_{9b\text{-}8}$ = 5.7 Hz, 1H, H_{9b}), 3.52 (dd, $J_{7\text{-}8}$ = 9.2, $J_{7\text{-}6}$ = 1.5 Hz, 1H, H_7), 2.25 (dd, $J_{3eq\text{-}3ax}$ = 12.9, $J_{3eq\text{-}4}$ = 4.9 Hz, 1H, H_{3eq}), 2.05 (s, 3H, NHAc), 1.93 (dd, $J_{3ax\text{-}3eq}$ = 12.9, $J_{3ax\text{-}4}$ = 11.3 Hz, 1H, H_{3ax}).

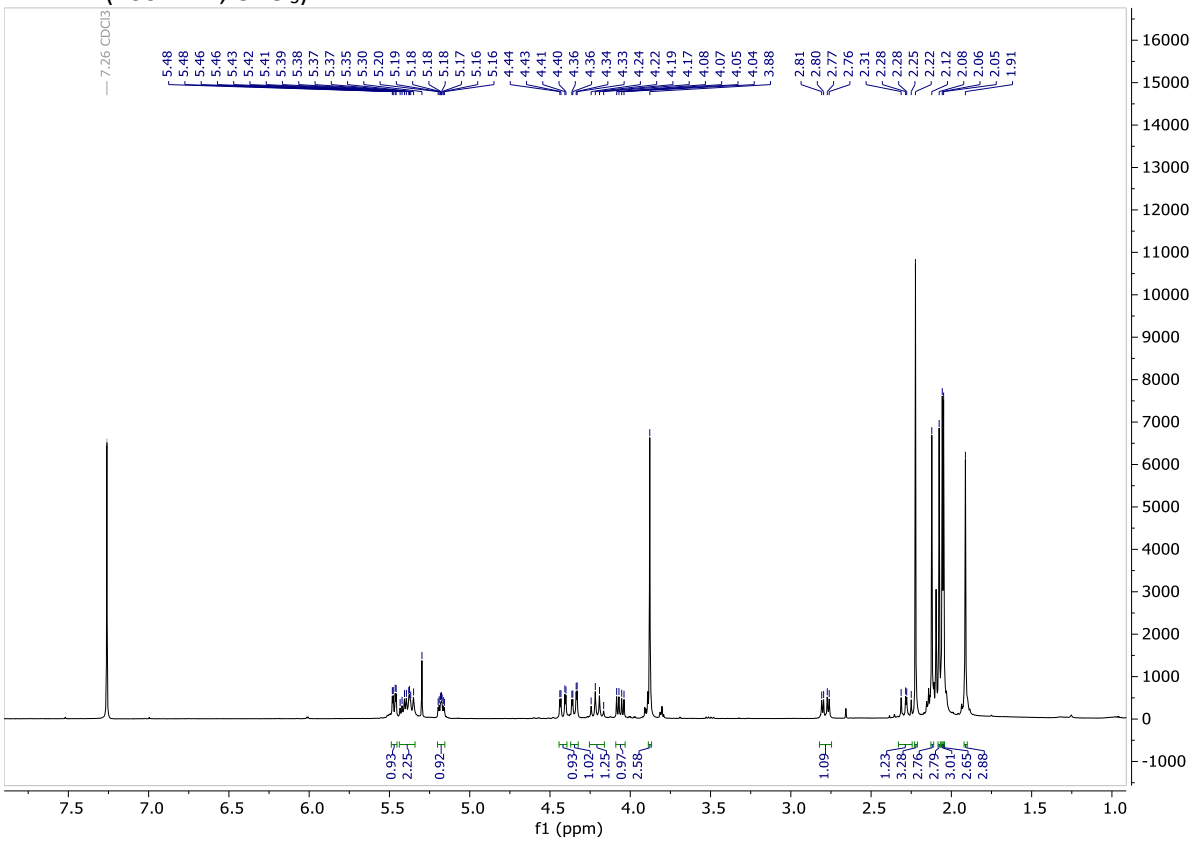
^{1}H NMR (400 MHz, MeOD)



To *N*-acetylneuraminic acid methylester (0.058 g, 0.18 mmol) AcCl (1.15 mL) was added under the N_2 atmosphere. The reaction mixture was kept stirring overnight at room temperature, then concentrated and dried under high vacuum. The crude chloride **4.16**[36] was used without any purification.

R_f=0.67 (CH_2Cl_2/Acetone 3:1); ^{1}H NMR (400MHz $CDCl_3$) 5.47 (m, 1H, H_7), 5.39 (mult., 2H, H_8, NH), 5.18 (m, 1H, H_4), 4.42 (dd, 1H, $J_{9b\text{-}9a}$ = 12.5 Hz, $J_{9b\text{-}8}$ = 2.7 Hz, H_{9b}), 4.35 (dd, 1H, $J_{6\text{-}5}$ = 10.8 Hz, $J_{6\text{-}7}$ = 2.3 Hz, H_6), 4.20 (dd, 1H, $J_{5\text{-}6}$ = 10.8, $J_{5\text{-}4}$ = 9.6 Hz, H_5), 4.06 (dd, 1H, $J_{9a\text{-}9b}$ = 12.5 Hz, $J_{9a\text{-}8}$ = 5.7 Hz, H_{9a}), 3.88 (s, 3H, COOMe), 2.78 (dd, 1H, $J_{3eq\text{-}3ax}$ = 13.9 Hz, $J_{3eq\text{-}4}$ = 4.8 Hz, H_{3eq}), 2.28 (dd, 1H, $J_{3ax\text{-}3eq}$ = 13.9 Hz, $J_{3ax\text{-}4}$ = 11.3 Hz, H_{3ax}), 2.12 (s, 3H, OAc), 2.08 (s, 3H, OAc), 2.06 (s, 3H, OAc), 2.05 (s, 3H, OAc), 1.91 (s, 3H, NHAc).

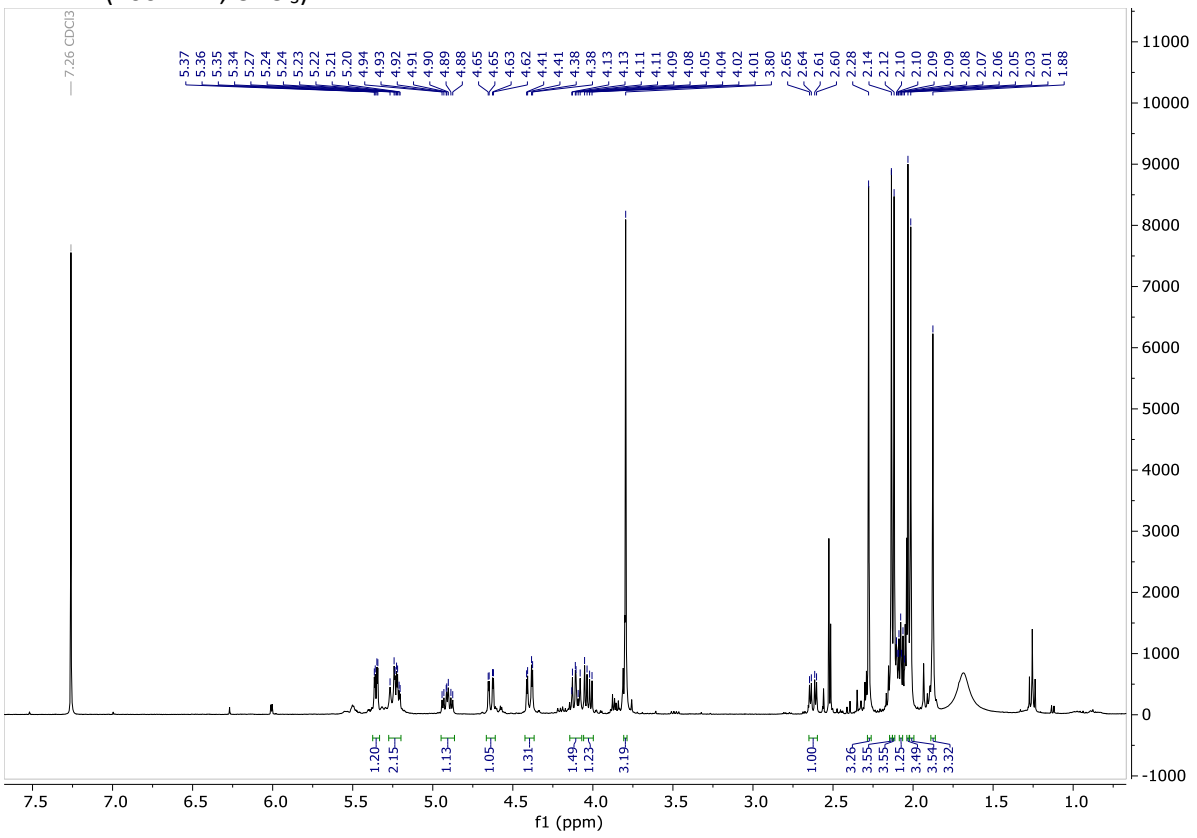
^{1}H NMR (400 MHz, $CDCl_3$)



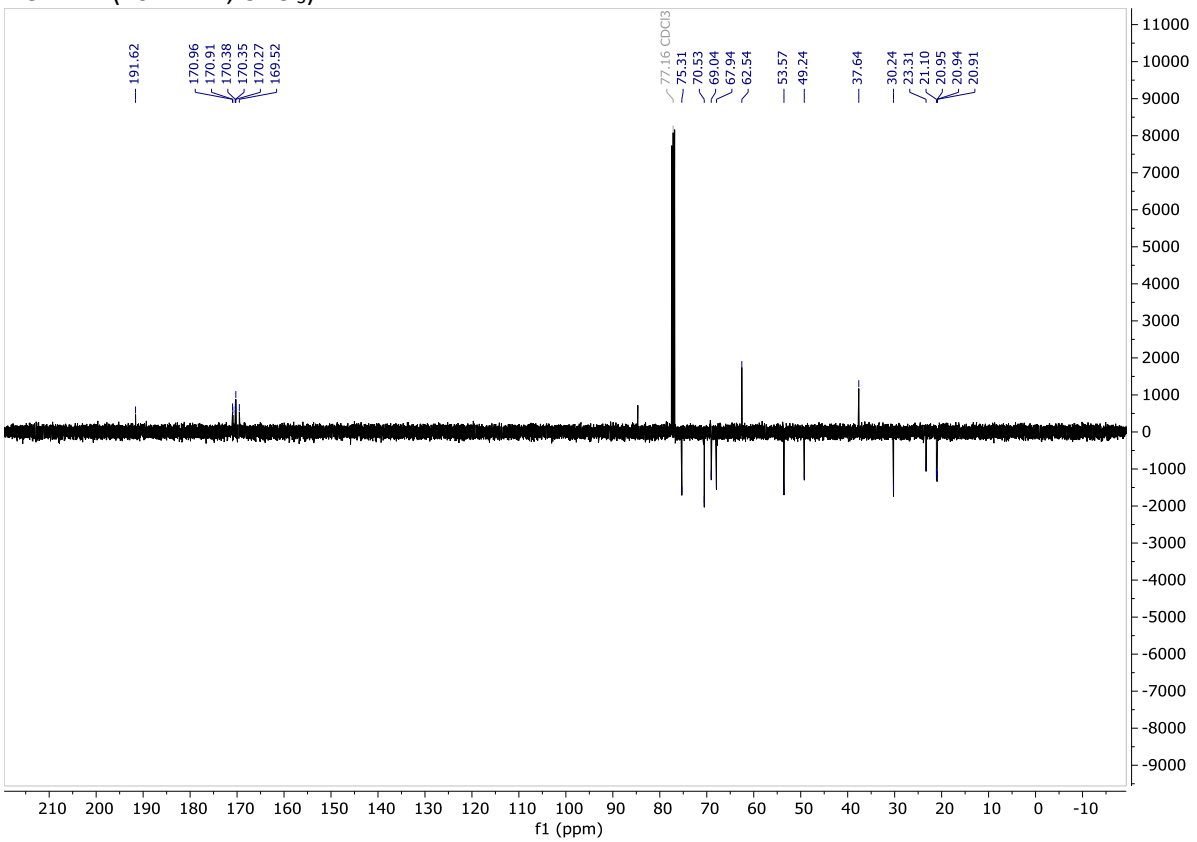
Crude **4.16** from previous step (0.18 mmol) and AcSK (0.095 g, 0.83 mmol) were dissolved in dry CH_2Cl_2 (0.80 mL) under N_2 atmosphere. The reaction mixture was kept stirring at room temperature overnight, then the solution was filtered with Buchner funnel to remove salts. The filtrate was concentrated under vacuum and the crude (which contains **4.17**, the glycal **4.18** and variable amounts of the SSAc product **4.19**, depending on the batch, as described in ref[13]) was purified by flash chromatography (eluent: 100% EtOAc). The product was obtained as a white solid in 58% yield (0.057 g, 0.10 mmol) as a mixture of product **4.17**[13] and glycal **4.18**. When scaled up to 1.3 g scale, an additional chromatographic purification (CH_2Cl_2 : Acetone) was required.

4.17 R_f = 0.4 (EtOAc 100%); ^{1}H NMR (400 MHz, $CDCl_3$) **4.17** δ 5.35 (dd, $J_{7\text{-}8}$ = 6.5, $J_{7\text{-}6}$ = 2.4 Hz, 1H, H_7), 5.27 – 5.19 (mult., 2H, H_8, NH), 4.96 – 4.86 (m, 1H, H_4), 4.64 (dd, $J_{6\text{-}5}$ = 10.8, $J_{6\text{-}7}$ = 2.4 Hz, 1H, H_6), 4.40 (dd, $J_{9b\text{-}9a}$ = 12.4, $J_{9b\text{-}8}$ = 2.6 Hz, 1H, H_{9b}), 4.15 – 4.07 (m, 1H, H_5), 4.03 (dd, $J_{9a\text{-}9b}$ = 12.5, $J_{9a\text{-}8}$ = 6.1 Hz, 1H, H_{9a}), 3.80 (s, 3H, COOMe), 2.62 (dd, $J_{3eq\text{-}3ax}$ = 12.9, $J_{3eq\text{-}4}$ = 4.6 Hz, 1H, H_{3eq}), 2.28 (s, 3H, SAc), 2.14 (s, 3H, OAc), 2.12 (s, 3H, OAc), 2.09 – 2.06 (m, 1H, H_{3ax}), 2.03 (s, 3H, OAc), 2.01 (s, 3H, OAc), 1.88 (s, 3H, NHAc); **glycal 4.18** δ 6.01 (d, $J_{3\text{-}4}$ = 3.2 Hz, 1H, H_3); ^{13}C NMR (101 MHz, $CDCl_3$) **4.17** δ 191.6 (CS), 171.0 (CO), 170.9 (CO), 170.4 (CO), 170.4 (CO), 170.3 (CO), 169.5 (CO), 75.3 (C_6), 70.5 (C_8), 69.0 (C_4), 67.9 (C_7), 62.5 (C_9), 53.6 (COOMe), 49.2 (C_5), 37.6 (C_3), 30.2 (SAc), 23.3 (NHAc), 21.1 (OAc), 21.0 (OAc), 20.9 (2xOAc); MS (ESI) calcd for $C_{22}H_{31}NO_{13}S$ $[M + Na]^+$ m/z: 572.14; found: 571.72.

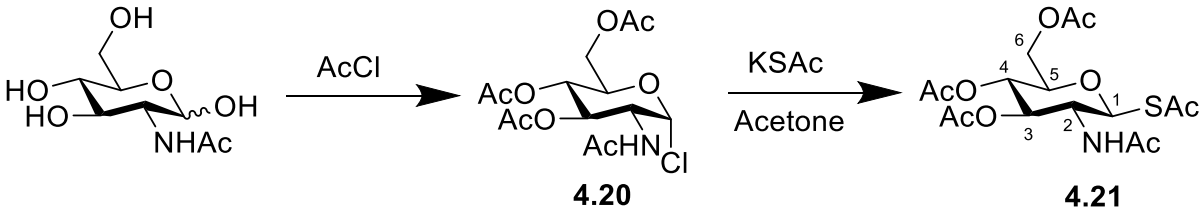
^{1}H NMR (400 MHz, $CDCl_3$)



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



Synthesis of β-1-*S*-acetyl-2-*N*-acetamido-2-deoxy-3,4,6-tri-*O*-acetyl-D-glucopyranose **4.21**

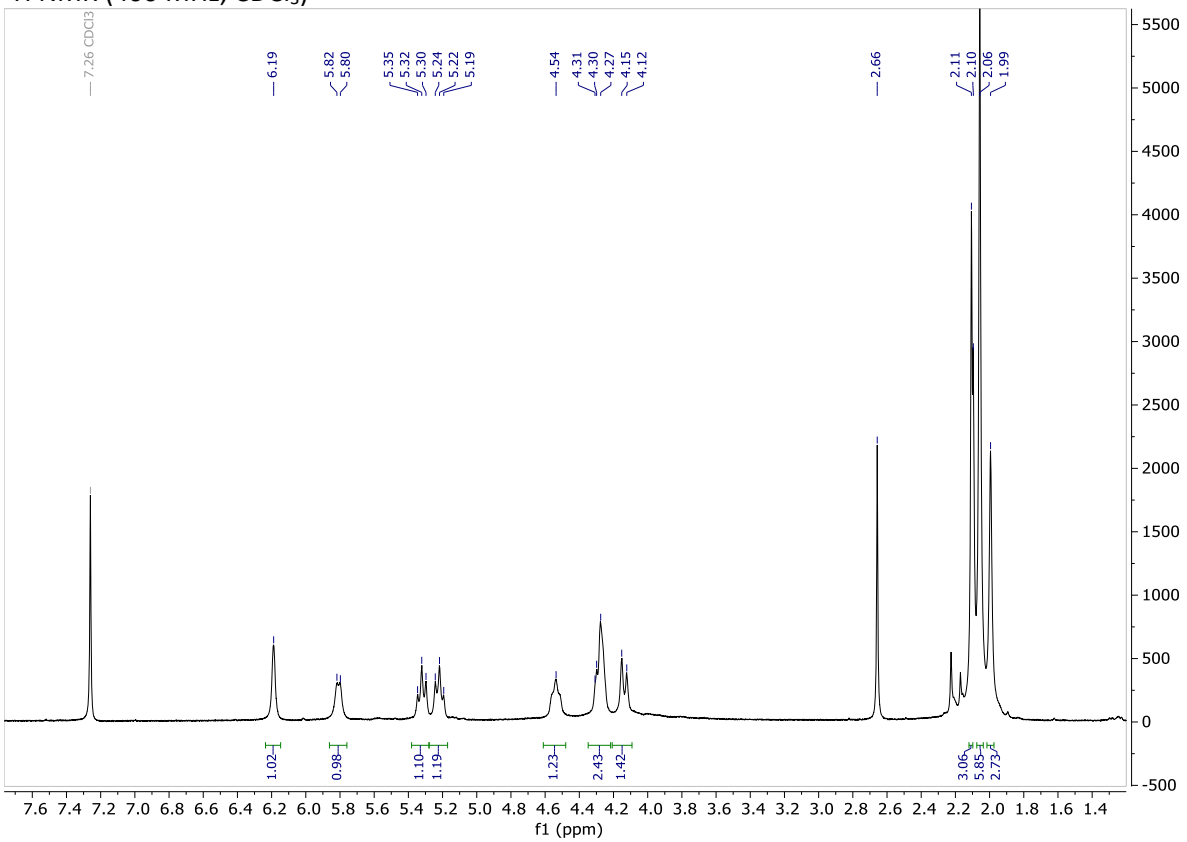


4.20 **4.21**

N-Acetylglucosamine (0.48 g, 2.44 mmol) was suspended in AcCl (2.50 mL) under N_2 atmosphere. The reaction mixture was kept shaking for 5 days at room temperature. The reaction mixture was then concentrated and dried under high vacuum and the crude anomeric chloride **4.20**[37] was used in the following step without purification.

^{1}H NMR (400 MHz, $CDCl_3$) δ 6.19 (s, 1H, H_1), 5.81 (d, J_{NH-2} = 8.1 Hz, 1H, NH), 5.32 (dd, J_{4-3} = 9.9 Hz, 1H, H_4), 5.22 (t, J_{3-4} = 9.9 Hz, 1H, H_3), 4.60 – 4.50 (m, 1H, H_2), 4.32 – 4.22 (mult., 2H, H_{6a}, H_5), 4.14 (dd, J_{6b-a} = 11.8 Hz, 1H, H_{6b}), 2.11 (s, 3H, OAc), 2.06 (s, 6H, 2xOAc), 1.99 (s, 3H, NHAc).

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

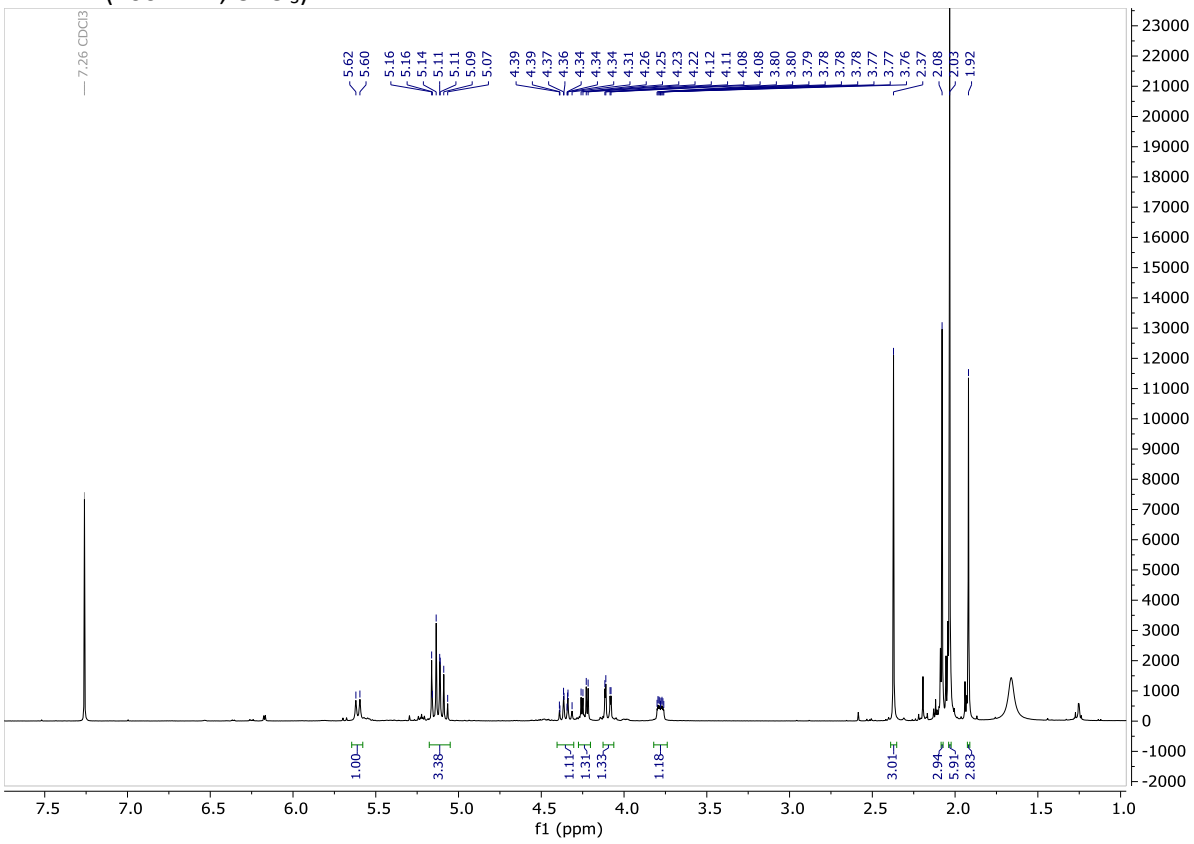


Crude 1-chloro-*N*-acetyl-3,4,6-tri-acetylglucosamine **4.20** (2.44 mmol) and AcSK (0.28 g, 2.44 mmol) were dissolved in dry acetone (8.26 mL) under N_2 atmosphere. The reaction mixture was kept stirring at room temperature overnight, then filtered with Buchner funnel to remove the salts. The filtrate was concentrated and the crude was purified by automated flash chromatography (Biotage Isolera,

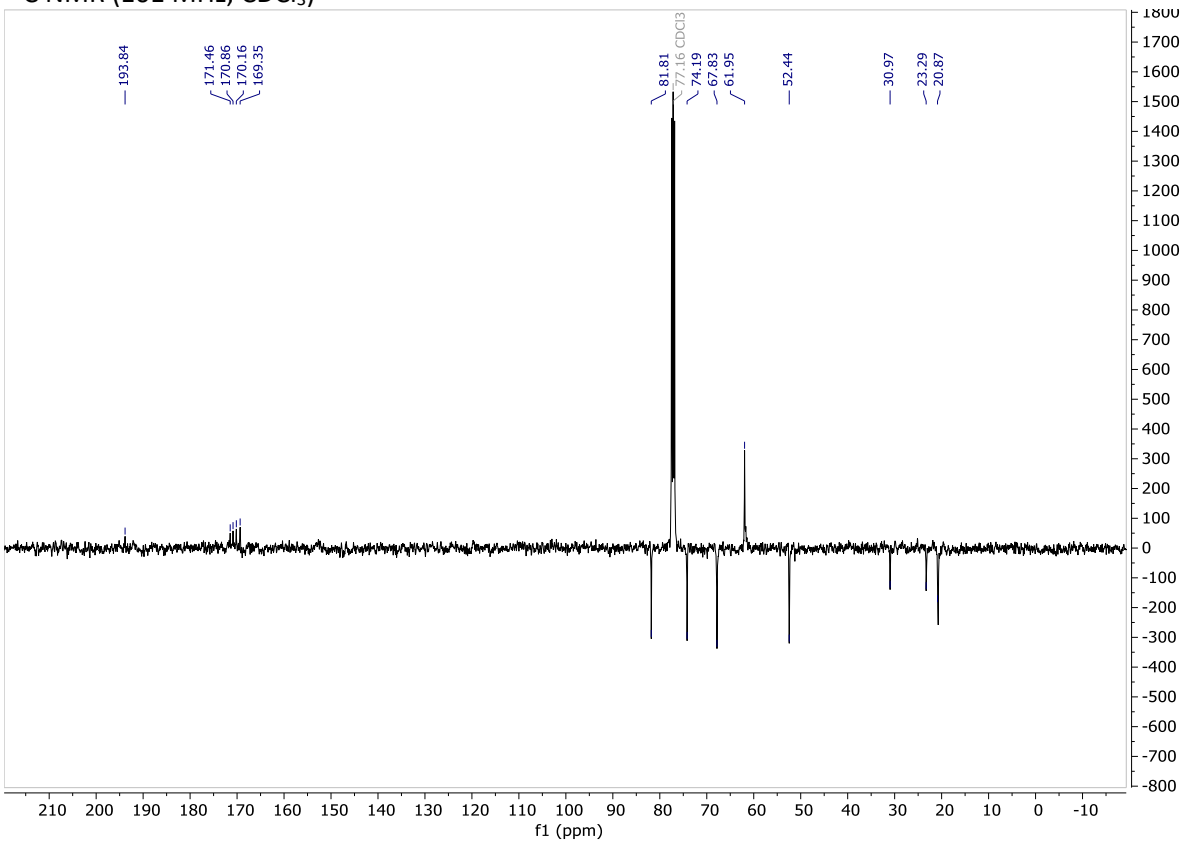
SNAP KP Cartridge, 0-5% gradient MeOH/CH_2Cl_2). The product **4.21**[12, 38] was obtained as a white solid in 77% yield (0.76 g, 1.88 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.61 (d, $J_{NH\text{-}2}$ = 9.9 Hz, 1H, NH), 5.17 – 5.06 (mult., 3H, H_1, H_3, H_4), 4.35 (dd, $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = $J_{2\text{-}NH}$ = 9.9 Hz, 1H, H_2), 4.24 (dd, $J_{6a\text{-}6b}$ = 12.5, $J_{6a\text{-}5}$ = 4.5 Hz, 1H, H_{6a}), 4.10 (dd, $J_{6b\text{-}6a}$ = 12.5, $J_{6b\text{-}5}$ = 2.3 Hz, 1H, H_{6b}), 3.78 (ddd, $J_{5\text{-}4}$ = 9.6, $J_{5\text{-}6a}$ = 4.5, $J_{5\text{-}6b}$ = 2.3 Hz, 1H, H_5), 2.37 (s, 3H, SAc), 2.08 (s, 3H, OAc), 2.03 (s, 6H, 2xOAc), 1.92 (s, 3H, NHAc); ^{13}C NMR (101 MHz, $CDCl_3$) δ 193.8 (CS), 171.5 (CO), 170.9 (CO), 170.2 (CO), 169.4 (CO), 81.8 (C_1), 74.2 (C_5, C_3), 67.8 (C_4), 62.0 (C_6), 52.4 (C_2), 31.0 (SAc), 23.3 (NHAc), 20.9 (OAc), 20.8 (OAc), 20.7 (OAc); MS (ESI) calcd for $C_{16}H_{23}NO_9S$ [M + Na]$^+$ m/z: 428.1; found: 427.71.

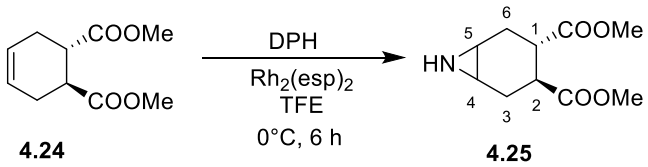
^{1}H NMR (400 MHz, $CDCl_3$)



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



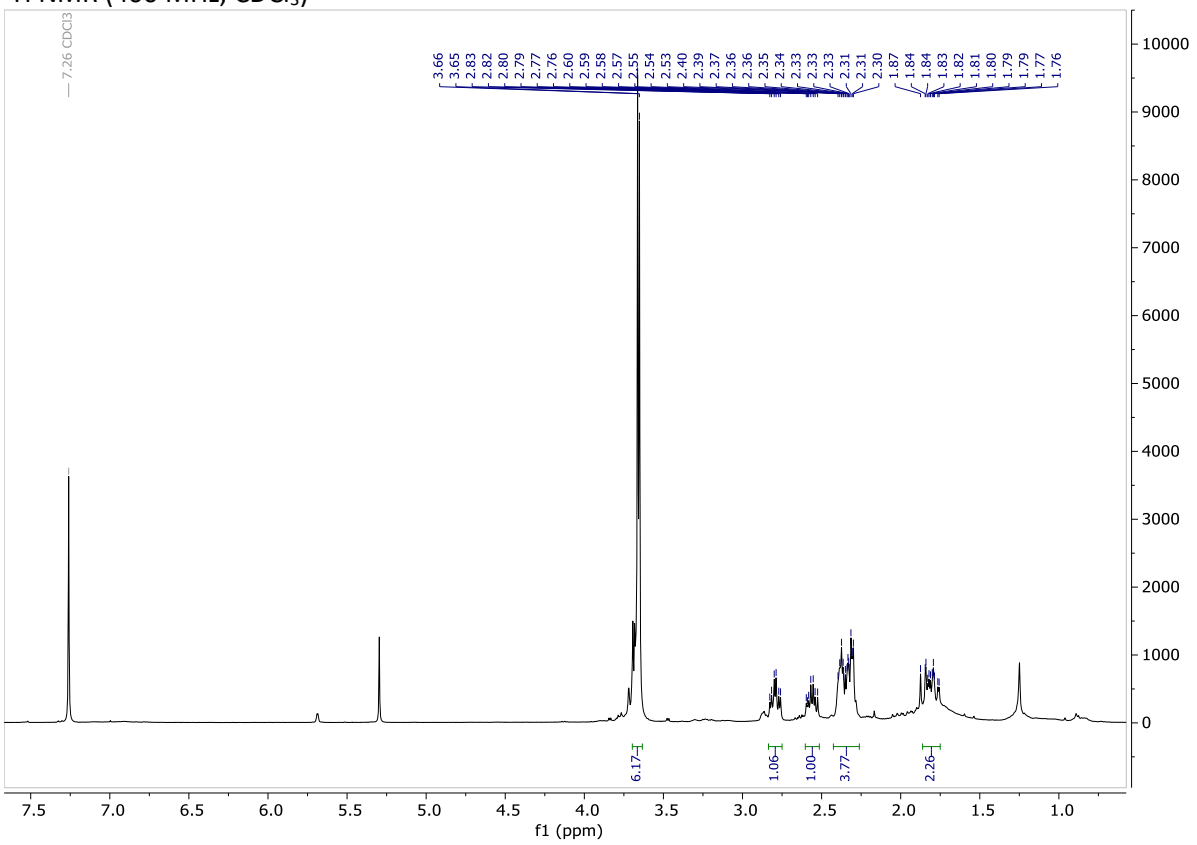
Synthesis of aziridine **4.25**



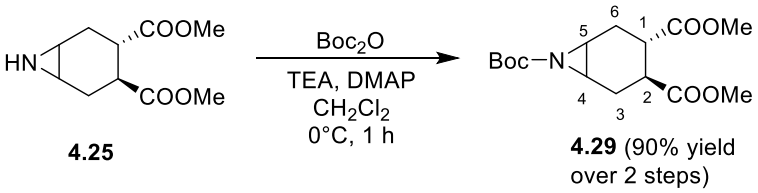
The reaction was performed in previously well-dried glassware flushed with Ar. Olefin **4.24**[39] (200 mg, 1.01 mmol) was dissolved in CF_3CH_2OH (6.7 mL, 0.15 M) and transferred into the flask under Ar. The solution was cooled to 0°C and then $Rh_2(esp)_2$ (7.7 mg, 0.010 mmol) and DPH (241 mg, 1.21 mmol) were added. The reaction mixture was stirred at 0°C (ice bath) under Ar atmosphere for 6 h. After 3 h additional portions of $Rh_2(esp)_2$ (7.7 mg, 0.010 mmol) and DPH (40 mg, 0.20 mmol) were added. After 6 h the reaction mixture was diluted with CH_2Cl_2 (10 mL) and washed with sat. $NaHCO_3$ (10 mL), then H_2O was added (10 mL) to dissolve any forming salts. The aqueous phase was washed twice with CH_2Cl_2. Combined organic phases were washed with brine and dried over Na_2SO_4, filtered and concentrated under vacuum. The crude product **4.25** was used directly and without further purification in the *N*-protection reaction to avoid decomposition and dimerization that occur during chromatographic purification.

R_f=0.2 (CH_2Cl_2/MeOH 95:5); ^{1}H NMR (400 MHz, $CDCl_3$) δ 3.66 (s, 3H, OMe), 3.64 (s, 3H, OMe), 2.78 (td, $J_{1\text{-}6ax}$= $J_{1\text{-}2}$= 11.4 Hz, $J_{1\text{-}6eq}$= 4.6 Hz, 1H, H_1), 2.54 (td, $J_{2\text{-}1}$= $J_{2\text{-}3ax}$= 11.4 Hz, $J_{2\text{-}3eq}$= 6.5 Hz, 1H, H_2), 2.39-2.33 (mult., 4H, H_{3eq}, H_{6eq}, H_4, H_5), 1.86-1.72 (mult., 2H, H_{3ax}, H_{6ax}).

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



Synthesis of the N-Boc-aziridine **4.29**

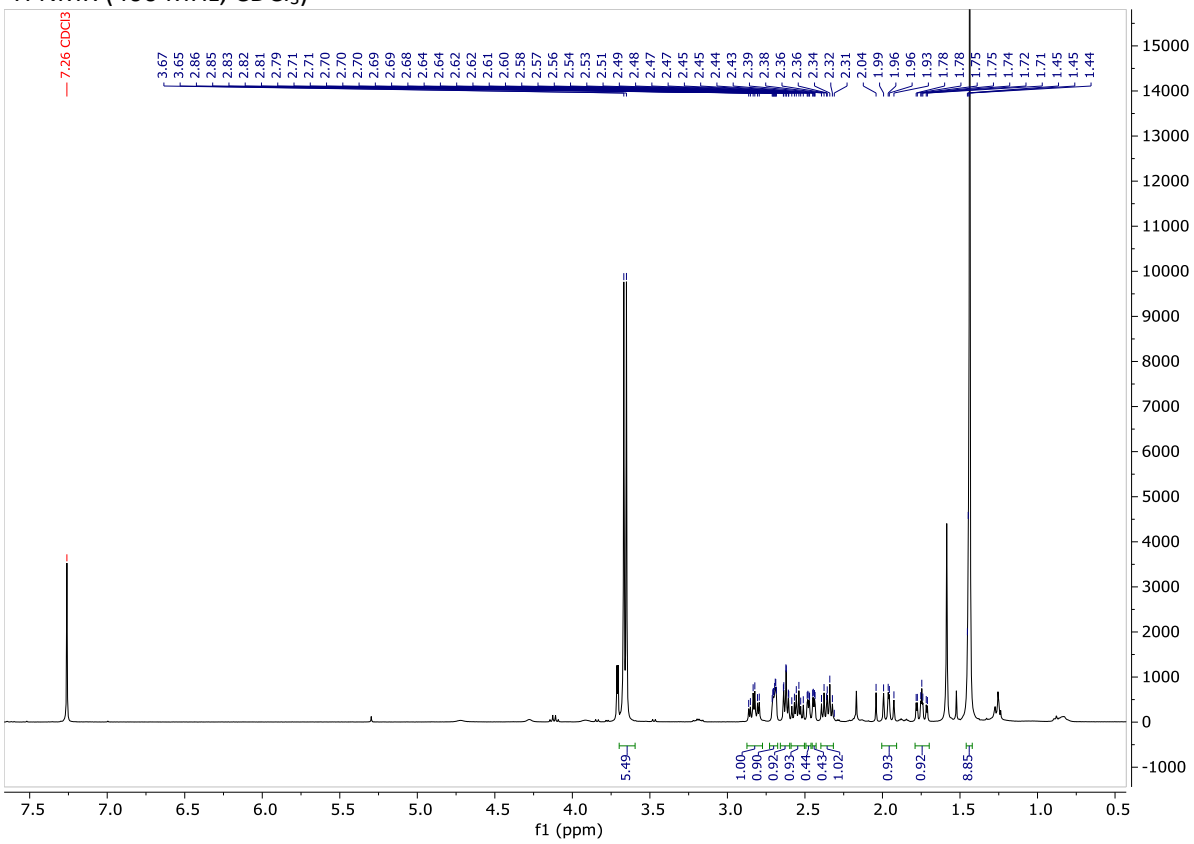


To a solution of the free aziridine **4.25** (crude from the previous step, 1.01 mmol) dissolved in freshly distilled CH_2Cl_2 under N_2 atmosphere, DMAP (a few crystals) were added. The mixture was cooled to 0°C and Et_3N (freshly distilled, 0.84 mL, 6.06 mmol) and Boc_2O (1.2 mL, 5.05 mmol) were added. After 1 h under stirring at 0°C, the reaction was allowed to reach room temperature, then quenched with H_2O. The resulting mixture was diluted with a saturated solution of NH_4Cl; the aqueous phase extracted twice with EtOAc. The organic layers were washed with saturated solutions of $KHSO_4$, Na_2CO_3 and NaCl in this order and finally dried over Na_2SO_4, filtered and concentrated in *vacuum*.

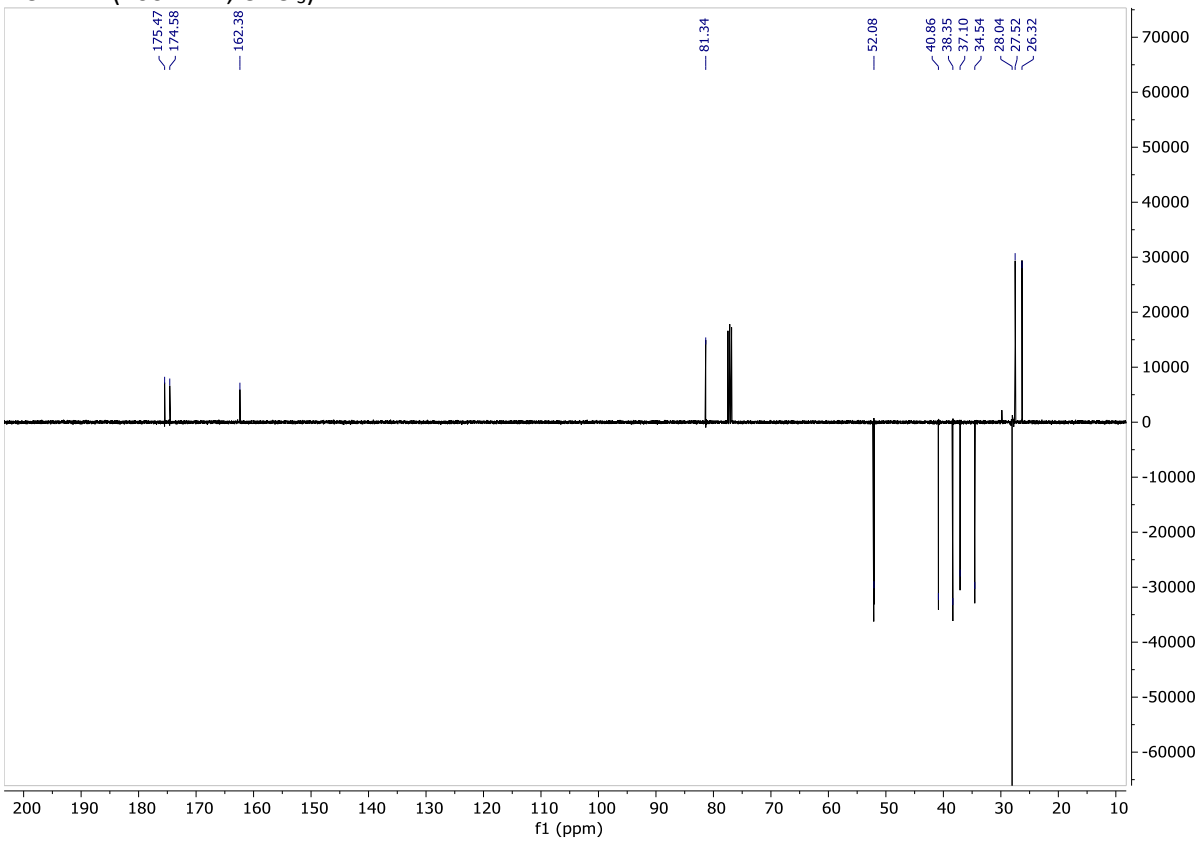
Purification by flash chromatography (7:3 Hex:EtOAc) afforded the *t*-butylcarbamate **4.29** as a yellow waxy solid in 90% yield (0.29 g, 0.91 mmol).

R_f=0.33 (Hex/EtOAc 7:3); $[\alpha]_D^{26}$ ($CHCl_3$, *c* 1.05): + 34; ^{1}H NMR (400 MHz, $CDCl_3$): δ = 3.66 (s, 3H, OMe), 3.65 (s, 3H, OMe), 2.82 (td, $J_{1\text{-}6ax}$= $J_{1\text{-}2}$= 11.5 Hz, $J_{1\text{-}6eq}$= 4.5 Hz, 1H, H_1), 2.72-2.67 (m, 1H, H_5), 2.62 (td, $J_{4\text{-}3eq}$= $J_{4\text{-}5}$= 6.4 Hz, $J_{4\text{-}3ax}$= 1.0 Hz 1H, H_4), 2.54 (td, $J_{2\text{-}1}$= $J_{2\text{-}3ax}$= 11.5 Hz, $J_{2\text{-}3eq}$= 6.4 Hz, 1H, H_2), 2.45 (ddd, $J_{6eq\text{-}6ax}$= 14.2 Hz, $J_{6eq\text{-}1}$= 4.6 Hz, $J_{6eq\text{-}5}$= 1.6 Hz, 1H, H_{6eq}), 2.35 (dt, $J_{3eq\text{-}3ax}$= 14.9 Hz, $J_{3eq\text{-}4}$= $J_{3eq\text{-}2}$= 6.4 Hz, 1H, H_{3eq}), 1.95 (ddd, $J_{3ax\text{-}3eq}$= 14.9 Hz, $J_{3ax\text{-}2}$= 11.8 Hz, $J_{3ax\text{-}4}$= 1.0 Hz, 1H, H_{3ax}), 1.74 (ddd, $J_{6ax\text{-}6eq}$= 14.5 Hz, $J_{6ax\text{-}1}$= 11.4 Hz, $J_{6ax\text{-}5}$= 3.0 Hz, 1H, H_{6ax}), 1.43 (s, 9H, tBu); ^{13}C NMR (100 MHz, $CDCl_3$): δ = 175.4 (CO), 174.5 (CO), 162.3 (CO, carbamate), 81.3 (C_{IV} Boc), 52.0 (2xOMe), 40.8 (C_2), 38.3 (C_1), 37.0 (C_5), 34.5 (C_4), 28.0 (tBu 3xMe), 27.5 (C_6), 26.3 (C_3).; MS (ESI) calcd for $C_{15}H_{23}NO_6$ [M + Na]$^+$ m/z: 336.14; found m/z: 336.18.

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



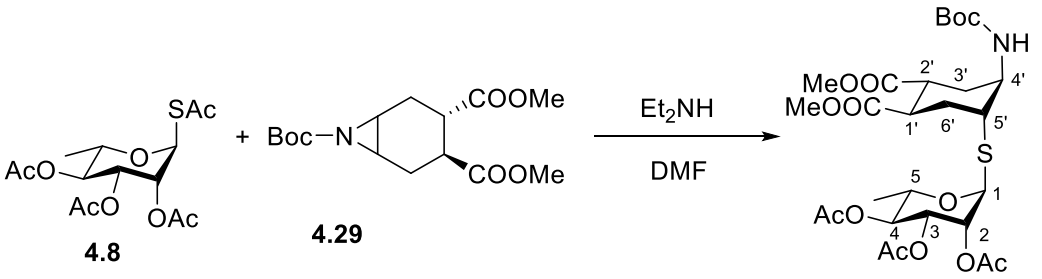
^{13}C NMR (100 MHz, $CDCl_3$)



General procedure for the opening reaction of aziridine **4.29**

Peracetylated *thio*-glycoside (1.3 eq) and aziridine (1 eq) were dissolved in dry DMF (0.65 M) under N_2 atmosphere at 20°C and Et_2NH (1.9 eq) was added. The reaction mixture was stirred for 4 h at 20°C (except for Lac derivative **4.12** at 0°C), then the reaction mixture was diluted with EtOAc and washed with 1 M HCl. The organic phase was washed three times with H_2O and the aqueous phases were additionally extracted with EtOAc. The combined organic phases were dried over Na_2SO_4, filtered and concentrated under vacuum.

Synthesis of the pseudo *thio*-disaccharide **4.33**

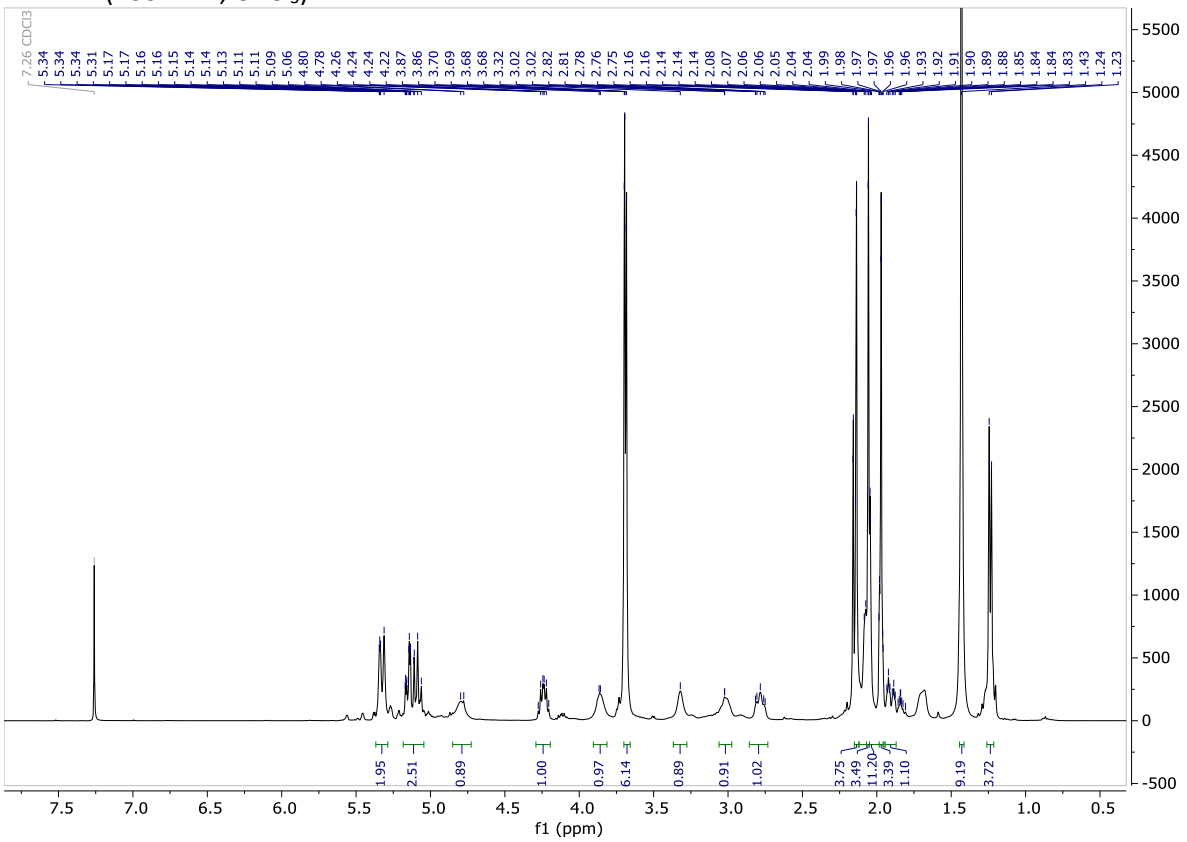


4.33 α–Rha

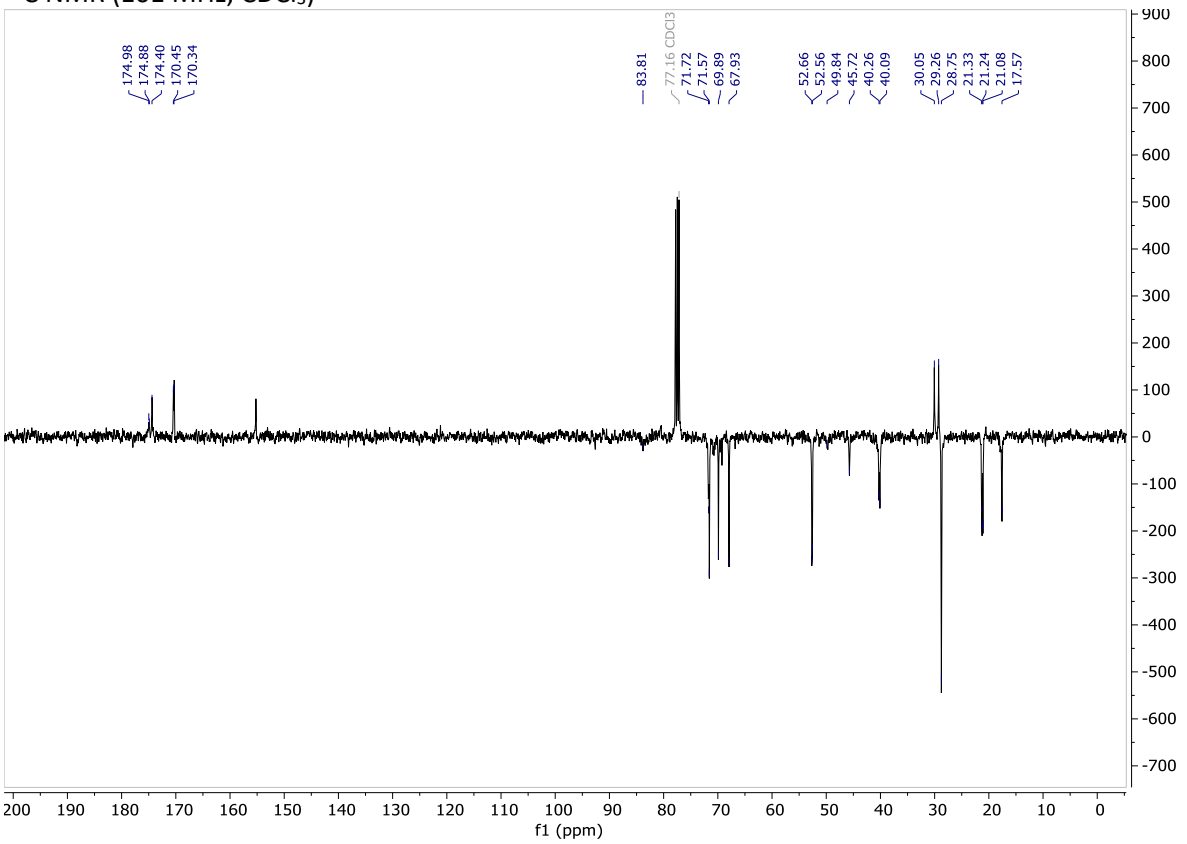
Prepared from peracetylated *thio*-rhamnose **4.8** (126 mg, 0.36 mmol) and aziridine **4.29** (87 mg, 0.28 mmol) according to the general procedure and purified by automated flash chromatography (Biotage Isolera, SNAP Cartridge, gradient 30-50% EtOAc/Hex) to give **4.33** as a single α isomer in 92% yield (158 mg, 0.26 mmol) as colourless oil.

R_f=0.34 (Hex/EtOAc 1:1); $[\alpha]_D^{21}$ ($CHCl_3$, *c* 1.00): - 71; ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.37 – 5.31 (m, 1H, H_2), 5.31 (s, 1H, H_1), 5.15 (m, 1H, H_3), 5.13 – 5.03 (m, 1H, H_4), 4.79 (s, 1H, NH), 4.23 (dt, $J_{5\text{-}4}$= 12.3 Hz, $J_{5\text{-}Me}$=6.3 Hz, 1H, H_5), 3.90 – 3.83 (m, 1H, $H_{4'}$), 3.70 (s, 3H, OMe), 3.68 (s, 3H, OMe), 3.35 – 3.28 (m, 1H, $H_{5'}$), 3.05 – 2.99 (m, 1H, $H_{1'}$), 2.84 – 2.73 (m, 1H, $H_{2'}$), 2.14 (s, 3H, OAc), 2.06 (s, 3H, OAc), 2.17 – 1.93 (mult., 3H, $H_{3'eq}$, $H_{6'ax}$, $H_{6'eq}$), 1.97 (s, 3H, OAc), 1.90 (m, 1H, $H_{3'ax}$), 1.43 (s, 9H, tBu), 1.24 (d, $J_{Me\text{-}5}$= 6.3 Hz, 3H, Me); ^{13}C NMR (101 MHz, $CDCl_3$) δ 175.0 (CO), 174.4 (CO), 170.4 (CO), 170.3 (CO), 170.3 (CO), 83.8 (C_1), 71.7 (C_2), 71.6 (C_4), 69.9 (C_3), 67.9 (C_5), 52.7 (OMe), 52.6 (OMe), 49.8 (C_4'), 45.7 ($C_{5'}$), 40.3 ($C_{2'}$), 40.1 ($C_{1'}$), 30.1 ($C_{6'}$), 29.3 ($C_{3'}$), 28.7 (tBu-3xMe), 21.3 (OAc), 21.2 (OAc), 21.1 (OAc), 17.6 (Me); $J_{H1\text{-}C1}$=170.3 (HSQC without ^{13}C decoupling); LC-MS (Rt=20.41 min) calcd for $C_{27}H_{41}NO_{13}S$ [M + Na]$^+$ m/z: 642.23; found m/z: 641.78. MS (HRMS): calcd for $C_{27}H_{41}NO_{13}S$ [M + Na]$^+$ m/z: 642.2196; found m/z: 642.2195.

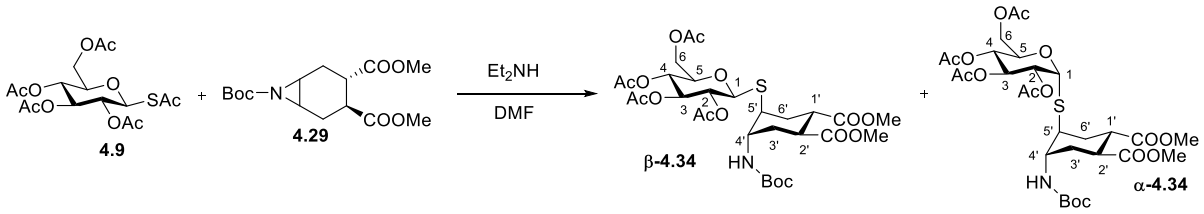
^{1}H NMR (400 MHz, $CDCl_3$)



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



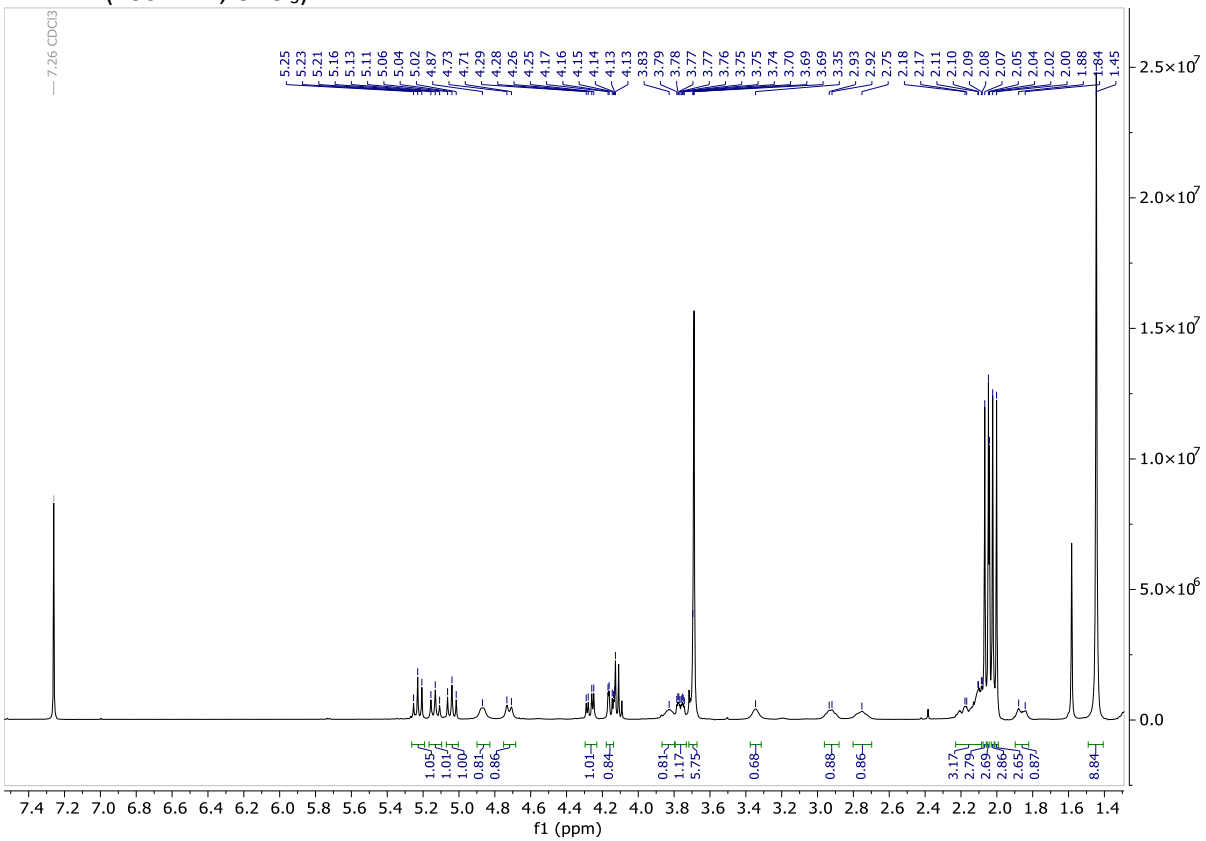
Synthesis of the pseudo *thio*-disaccharide **4.34**



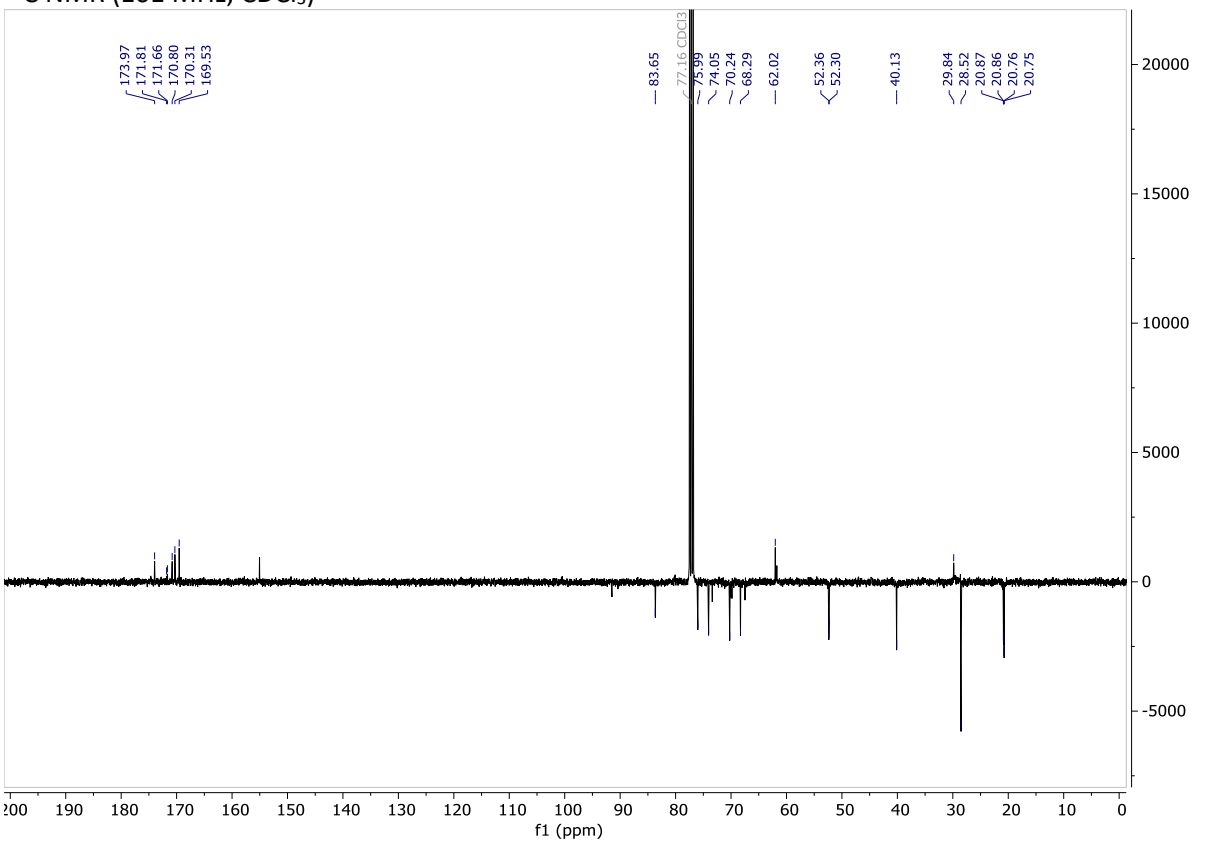
Prepared from peracetylated *thio*-glucose **4.9** (34 mg, 0.083 mmol) and aziridine **4.29** (20 mg, 0.064 mmol) according to the general procedure and purified by flash chromatography (1:1 Hex: EtOAc) to give **4.34** in 61% yield (26 mg, 0.039 mmol) as colourless oil (entry 4 in Table 2). The β and α isomers could be separated by flash chromatography (8:1 iPr_2O:EtOAc).

β-4.34: R_f=0.22 (iPr_2O/EtOAc 8:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.23 (dd, $J_{3\text{-}4}$= $J_{3\text{-}2}$= 10 Hz, 1H, H_3), 5.13 (dd, $J_{4\text{-}3}$= $J_{4\text{-}5}$= 10 Hz, 1H, H_4), 5.04 (dd, $J_{2\text{-}1}$= $J_{2\text{-}3}$= 10 Hz, 1H, H_2), 4.87 (m, 1H, NH), 4.72 (d, $J_{1\text{-}2}$= 10 Hz, 1H, H_1), 4.27 (dd, $J_{6a\text{-}6b}$= 12.4 Hz, $J_{6a\text{-}5}$= 4.5 Hz, 1H, H_{6a}), 4.15 (dd, $J_{6b\text{-}6a}$= 12.4 Hz, $J_{6b\text{-}5}$= 2.4 Hz, 1H, H_{6b}), 3.93 – 3.80 (m, 1H, $H_{4'}$), 3.76 (ddd, $J_{5\text{-}4}$= 10.0 Hz, $J_{5\text{-}6a}$= 4.5 Hz, $J_{5\text{-}6b}$= 2.4 Hz, 1H, H_5), 3.69 (s, 6H, 2xOMe), 3.42 – 3.26 (m, 1H, $H_{5'}$), 2.99 – 2.86 (m, 1H, $H_{1'}$), 2.81 – 2.70 (m, 1H, $H_{2'}$), 2.24 – 2.08 (mult., 3H, $H_{3'eq}$, $H_{6'eq}$, $H_{6'ax}$), 2.07 (s, 3H, OAc), 2.05 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.00 (s, 3H, OAc), 1.91 – 1.81 (m, 1H, $H_{3'ax}$), 1.45 (s, 9H, tBu); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.0 (CO), 171.8 (CO), 171.7 (CO), 170.8 (CO), 170.3 (CO), 169.5 (CO), 83.7 (C_1), 76.0 (C_5), 74.1 (C_3), 70.2 (C_2), 68.3 (C_4), 62.0 (C_6), 52.4 (OMe), 52.3 (OMe), 48.2* ($C_{4'}$), 42.1* ($C_{5'}$), 40.1 ($C_{1'}$, $C_{2'}$), 29.8 ($C_{3'}$, $C_{6'}$), 28.5 (tBu-3xMe), 20.9 (Ac), 20.9 (Ac), 20.8 (Ac), 20.8 (Ac). * These signals are better visible in the HSQC spectrum; LC-MS (Rt=19.64 min) calcd for $C_{29}H_{43}NO_{15}S$ $[M + Na]^+$ m/z: 700.24; found m/z: 699.81; MS (HRMS): calcd for $C_{29}H_{43}NO_{15}S$ $[M + Na]^+$ m/z: 700.2251; found m/z: 700.2249.

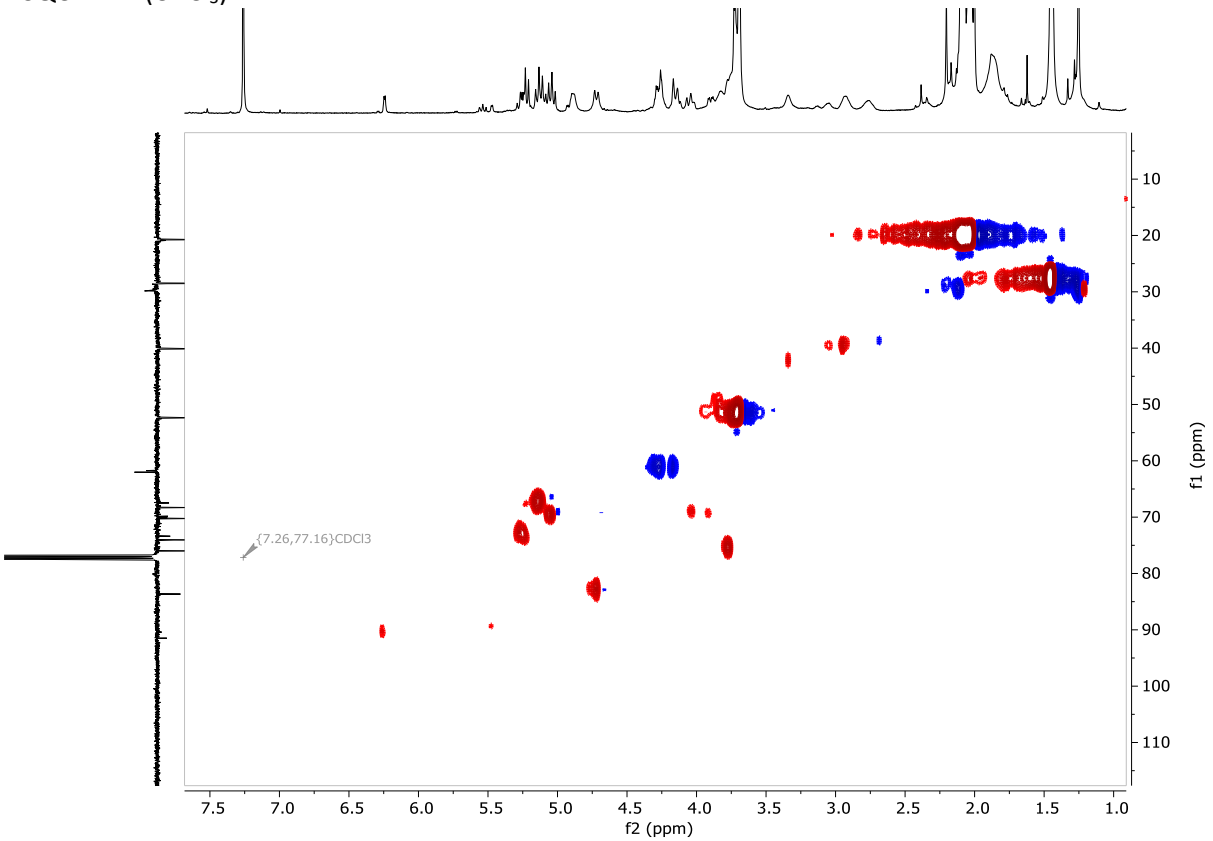
^{1}H NMR (400 MHz, $CDCl_3$)



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

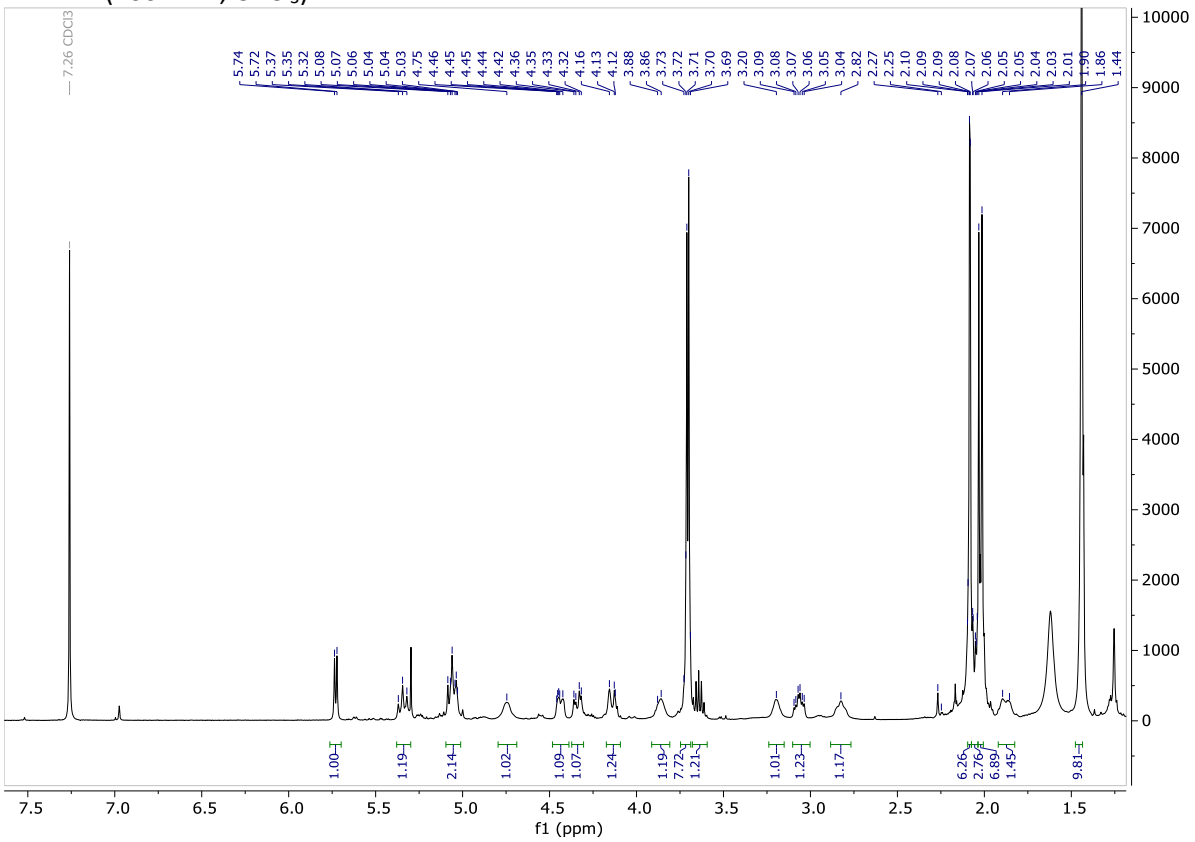


HSQC NMR ($CDCl_3$)

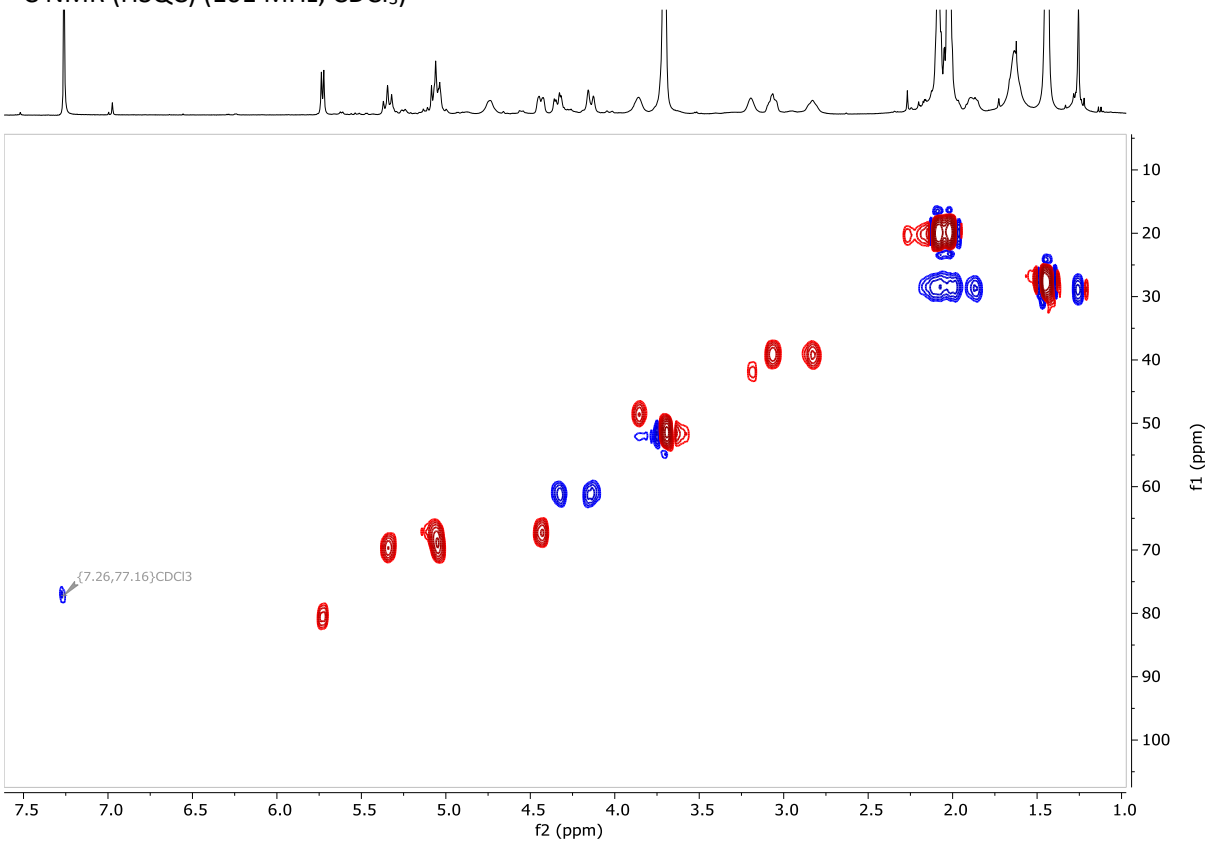


α-4.34: ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.73 (d, $J_{2\text{-}1}$ = 5.7 Hz, 1H, H_1), 5.35 (dd, $J_{3\text{-}2}$ = $J_{3\text{-}4}$ = 9.8 Hz, 1H, H_3), 5.14 – 4.99 (mult., 2H, H_4, H_2), 4.75 (s, 1H, NH), 4.48 – 4.39 (m, 1H, H_5), 4.34 (dd, $J_{6a\text{-}6b}$ = 12.5 Hz, $J_{6a\text{-}5}$ = 4.6 Hz, 1H, H_{6a}), 4.19 – 4.09 (m, 1H, H_{6b}), 3.92 – 3.83 (m, 1H, $H_{4'}$), 3.71 (s, 3H, OMe), 3.70 (s, 3H, OMe), 3.24 – 3.15 (m, 1H, $H_{5'}$), 3.11 – 2.98 (m, 1H, $H_{1'}$), 2.88 – 2.73 (m, 1H, $H_{2'}$), 2.09 (s, 3H, OAc), 2.08 (s, 3H, OAc), 2.10 – 2.00 (mult., 3H, $H_{3'eq}$, $H_{6'eq}$, $H_{6'ax}$), 2.03 (s, 3H, OAc), 2.01 (s, 3H, OAc), 1.92 – 1.84 (m, 1H, $H_{3'ax}$), 1.44 (s, 9H, tBu); ^{13}C NMR (HSQC) (101 MHz, $CDCl_3$) δ 81.2 (C_1), 70.2 (C_3), 67.8 (C_5, C_2, C_4), 61.6 (C_6), 51.9 (2x OMe), 49.2 ($C_{4'}$), 42.5 ($C_{5'}$), 39.7 ($C_{1'}$, $C_{2'}$), 29.0 ($C_{3'}$, $C_{6'}$), 29.0(tBu-3xMe), 20.6 (4xOAc); LC-MS (Rt=19.87 min) calcd for $C_{29}H_{43}NO_{15}S$ $[M + Na]^+$ m/z: 700.24; found m/z: 699.81.

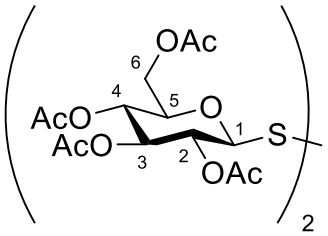
^{1}H NMR (400 MHz, $CDCl_3$)



^{13}C NMR (HSQC) (101 MHz, $CDCl_3$)



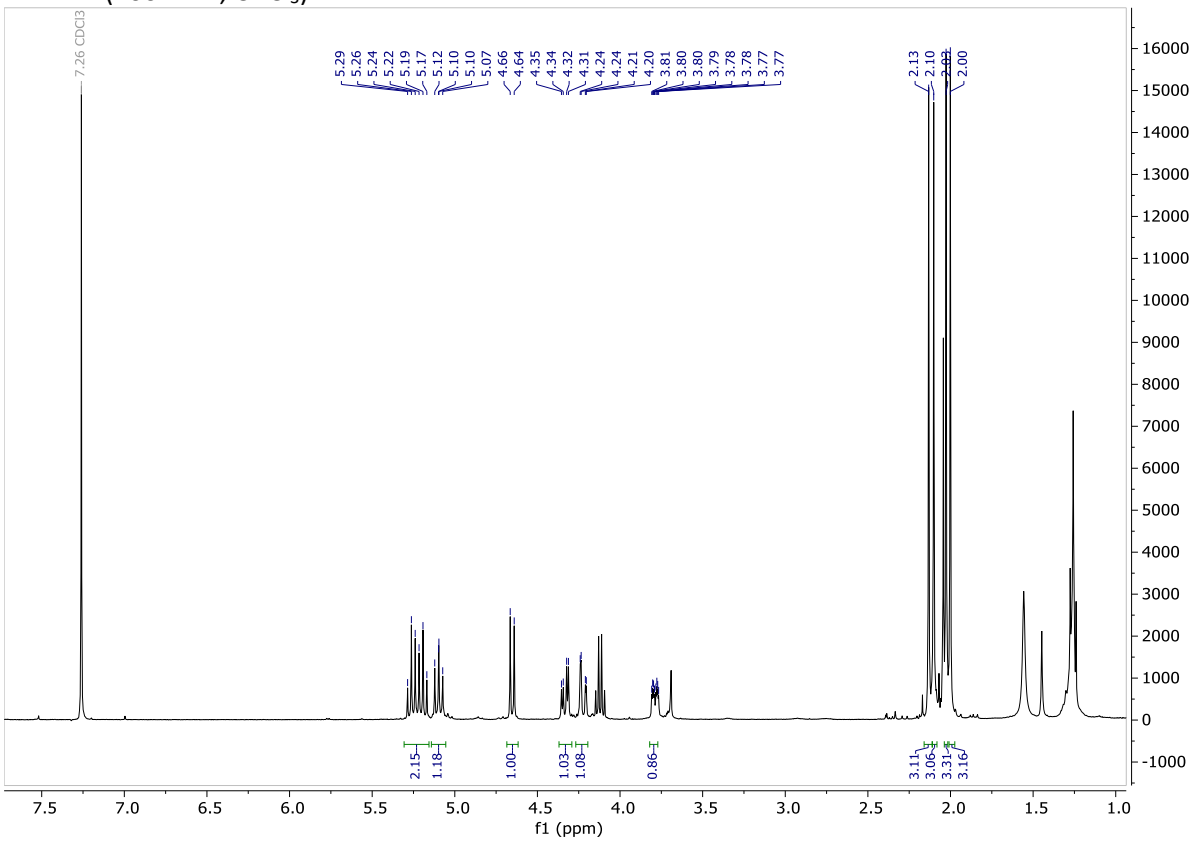
^{1}H NMR assignment of bis-*S*-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl) disulphide (**4.40**)



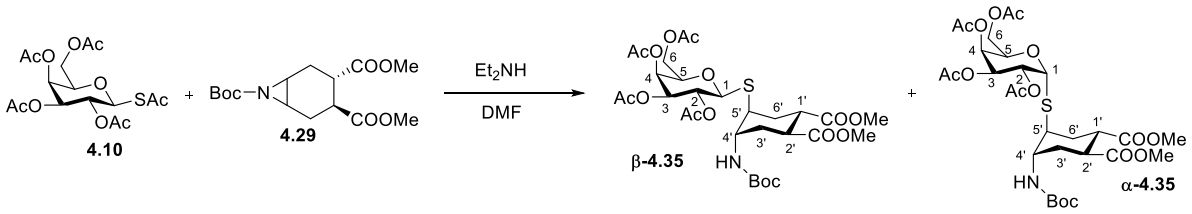
4.40

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.31 – 5.16 (mult., 2H, H_2, H_3), 5.14 – 5.05 (m, 1H, H_4), 4.65 (d, $J_{1\text{-}2}$ = 9.7 Hz, 1H, H_1), 4.33 (dd, $J_{6a\text{-}6b}$ = 12.5, $J_{6a\text{-}5}$ = 4.4 Hz, 1H, H_{6a}), 4.22 (dd, $J_{6b\text{-}6a}$ = 12.5, $J_{6b\text{-}5}$ = 2.3 Hz, 1H, H_{6b}), 3.79 (ddd, $J_{5\text{-}4}$ = 10.0, $J_{5\text{-}6a}$ = 4.4, $J_{5\text{-}6b}$ = 2.3 Hz, 1H, H_5), 2.13 (s, 3H, SAc), 2.10 (s, 3H, OAc), 2.03 (s, 3H, OAc), 2.00 (s, 3H, OAc). Reference for **4.40**[40]

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



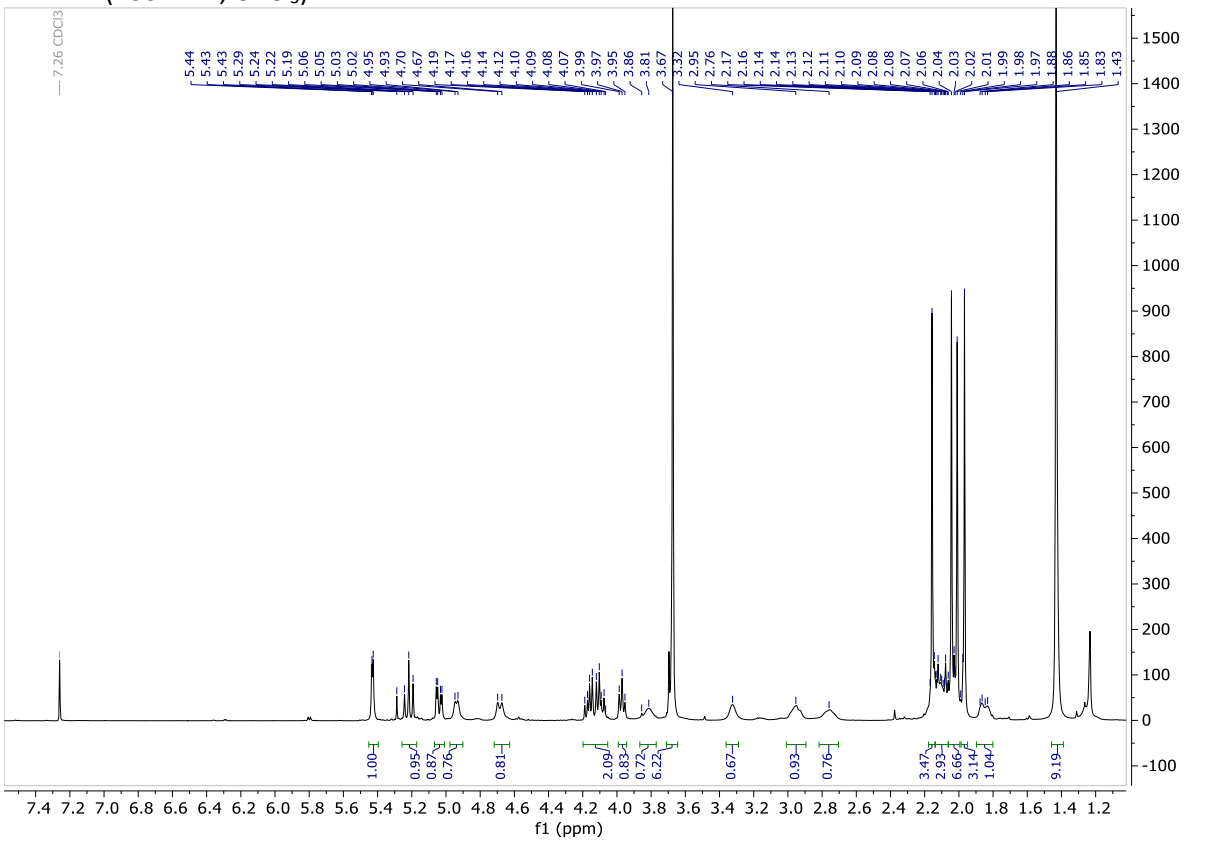
Synthesis of the pseudo *thio*-disaccharide **4.35**



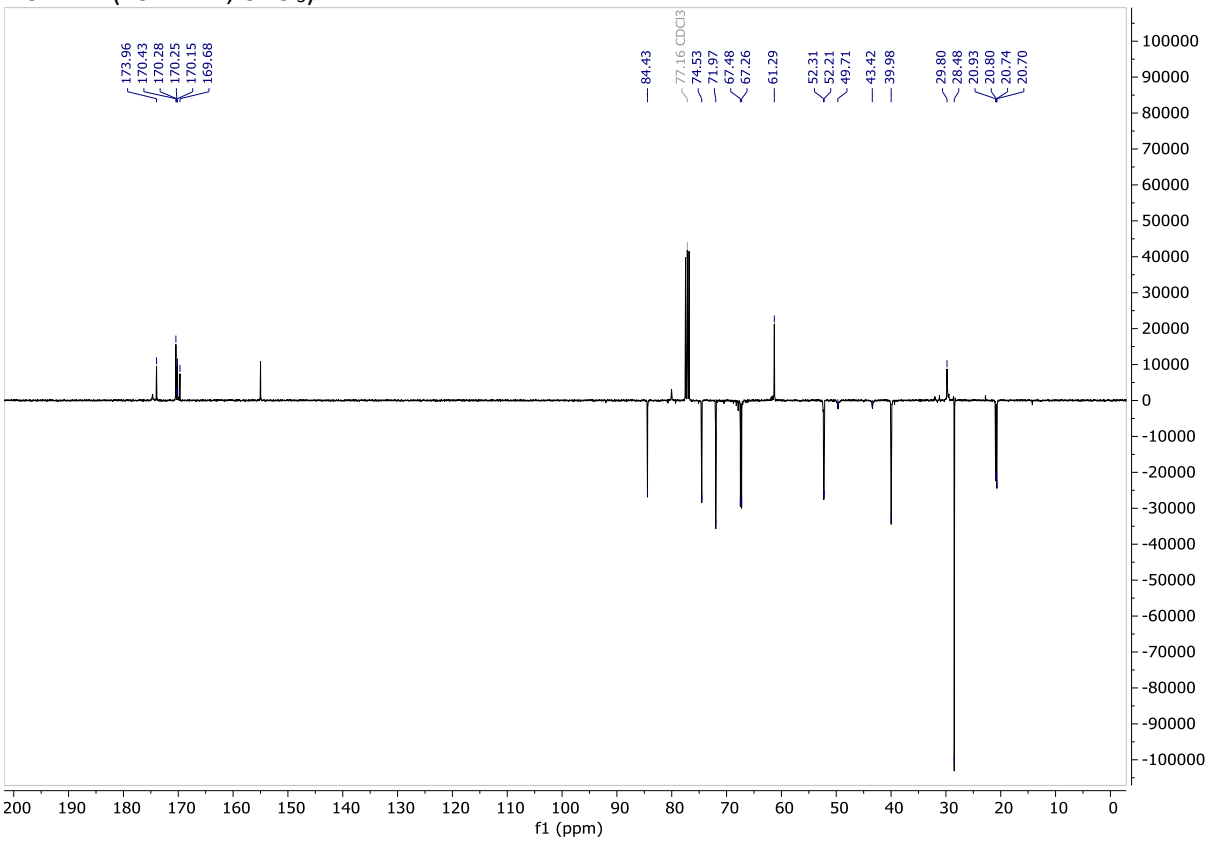
Prepared from peracetylated *thio*-galactose **4.10** (51 mg, 0.13 mmol) and aziridine **4.29** (30 mg, 0.096 mmol) according to the general procedure and purified by flash chromatography (1:1 Hex:EtOAc) to give **4.35** as a mixture of α and β isomers in 44% (29 mg, 0.042 mmol) yield as colourless oil. The β and α isomers were separated by a second flash chromatography (8:1 iPr_2O:EtOAc) to give pure **β-4.35** in 34% yield (22 mg, 0.033 mmol).

β-4.35: R_f=0.33 (iPr_2O/EtOAc 8:2); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.45 (dd, $J_{4\text{-}3} = J_{4\text{-}5} = 3.4$ Hz, 1H, H_4), 5.24 (dd, $J_{2\text{-}1} = J_{2\text{-}3} = 10.0$ Hz, 1H, H_2), 5.05 (dd, $J_{3\text{-}2} = 10.0$ Hz, $J_{3\text{-}4} = 3.4$ Hz, 1H, H_3), 4.86 (m, 1H, NH), 4.69 (d, $J_{1\text{-}2} = 10.0$ Hz, 1H, H_1), 4.13 (mult., 2H, H_{6a}, H_{6b}), 3.98 (m, 1H, H_5), 3.87 – 3.77 (m, 1H, $H_{4'}$), 3.70 (s, 3H, OMe), 3.69 (s, 3H, OMe), 3.39 – 3.28 (m, 1H, $H_{5'}$), 3.02 – 2.93 (m, 1H, $H_{1'}$), 2.81 – 2.69 (m, 1H, $H_{2'}$), 2.17 (s, 3H, OAc), 2.14 – 2.09 (mult., 3H, $H_{3'eq}$, $H_{6'ax}$, $H_{6'eq}$), 2.04 (s, 3H, OAc), 2.03 (s, 3H, OAc), 1.98 (s, 3H, OAc), 1.89 – 1.82 (m, 1H, $H_{3'ax}$), 1.45 (s, 9H, tBu); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.0 (CO), 170.4 (CO), 170.4 (CO), 170.2 (CO), 170.1 (CO), 169.7 (CO), 84.4 (C_1), 74.5 (C_5), 72.0 (C_3), 67.5 (C_4), 67.3 (C_2), 61.3 (C_6), 52.3 (OMe), 52.2 (OMe), 49.7 ($C_{4'}$), 43.4 ($C_{5'}$), 40.0 ($C_{2'}$, $C_{1'}$), 29.8 ($C_{3'}$, $C_{6'}$), 28.5 (tBu-3xMe), 20.9 (OAc), 20.8 (OAc), 20.7 (OAc), 20.7 (OAc); LC-MS (R_t=19.51 min) calcd for $C_{29}H_{43}NO_{15}S$ $[M + Na]^+$ m/z: 700.24; found m/z: 699.74; MS (HRMS): calcd for $C_{29}H_{43}NO_{15}S$ $[M + Na]^+$ m/z: 700.2251; found m/z: 700.2259.

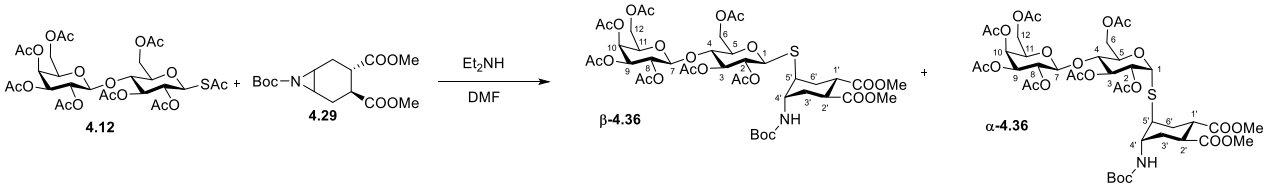
^{1}H NMR (400 MHz, $CDCl_3$)



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



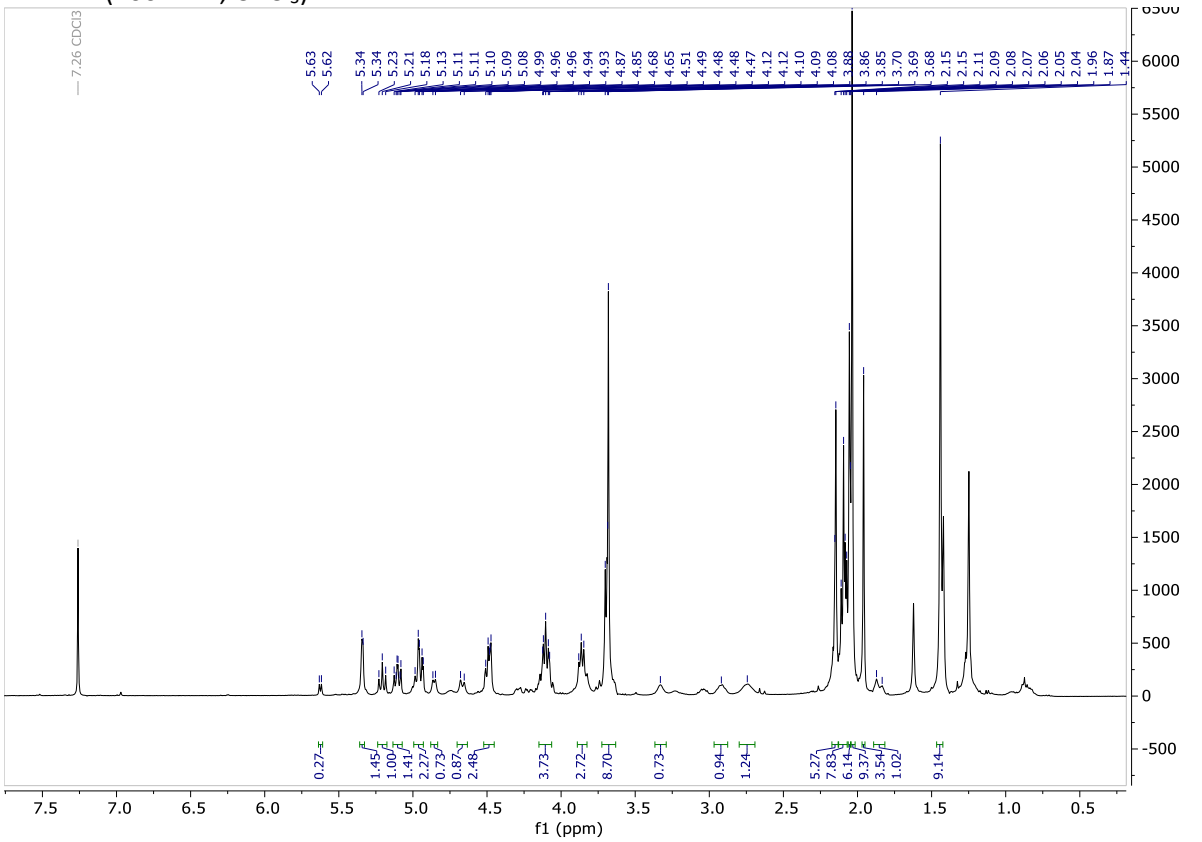
Synthesis of the pseudo *thio*-disaccharide **4.36**



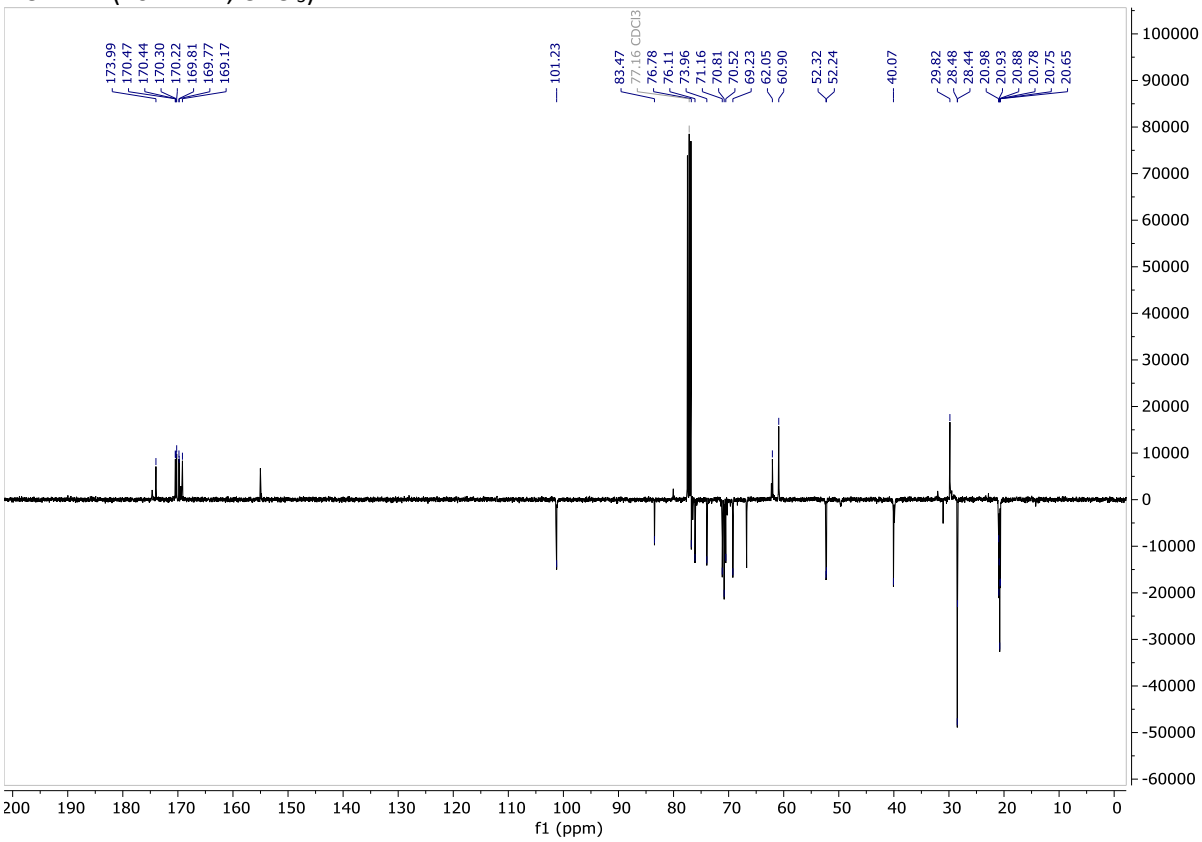
Prepared from peracetylated *thio*-lactose **4.12** (56 mg, 0.080 mmol) and aziridine **4.29** (19 mg, 0.062 mmol) according to the general procedure (at 0°C) and purified by flash chromatography (3:2 Hex:EtOAc) to give **4.36** as a 5:1 β:α anomeric mixture in 42% yield (25 mg, 0.026 mmol) as colourless oil.

R_f=0.14(Hex/EtOAc 1:1); ^{1}H NMR (400 MHz, $CDCl_3$) **β-4.36** 5.35 – 5.33 (m, 1H, H_{10}), 5.21 (dd, $J_{3\text{-}4}$= $J_{3\text{-}2}$= 9.2 Hz, 1H, H_3), 5.10 (dd, $J_{8\text{-}9}$= 10.4 Hz, $J_{8\text{-}7}$= 7.9 Hz, 1H, H_8), 5.02 – 4.97 (m, 1H, H_2), 4.95 (dd, $J_{9\text{-}8}$ = 10.4 Hz, $J_{9\text{-}10}$ = 3.2 Hz, 1H, H_9), 4.86 (d, $J_{NH\text{-}4'}$ = 6.9 Hz, 1H, NH), 4.67 (d, $J_{1\beta\text{-}2\beta}$ = 10.1 Hz, 1H, $H_{1\beta}$), 4.49 (mult., 2H, H_7, H_{12a}), 4.15 – 4.01 (mult., 3H, H_{12b}, H_{6a}, H_{6b}), 3.91 – 3.81 (mult., 3H, H_{11}, H_4, H_5), 3.70 (s, 1H, $H_{4'}$), 3.68 (2x s, 6H, 2xOMe), 3.37 – 3.30 (m, 1H, $H_{5'}$), 2.96 – 2.87 (m, 1H, $H_{1'}$), 2.82 – 2.70 (m, 1H, $H_{2'}$), 2.15 (s, 3H, OAc), 2.09 (s, 3H, OAc), 2.12 – 2.05 (mult., 3H, $H_{3'eq}$, $H_{6'ax}$, $H_{6'eq}$), 2.06 (s, 6H, 2xOAc), 2.04 (s, 9H, 3xOAc), 1.96 (s, 3H, OAc), 1.89 – 1.81 (m, 1H, $H_{3'ax}$), 1.44 (s, 9H, tBu), **α-4.36** δ 5.62 (d, $J_{1\alpha\text{-}2\alpha}$ = 5.7 Hz, 1H, $H_{1\alpha}$); **β-4.36** ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.7 (CO), 174.0 (CO), 170.5 (CO), 170.4 (CO), 170.3 (CO), 170.2 (CO), 169.8 (CO), 169.8 (CO), 169.2 (CO), 83.5 (C_1), 76.8 (C_5), 76.1 (C_{11}), 74.0 (C_3), 71.2 (C_4), 70.8 (C_9), 70.5 (C_2), 69.2 (C_8), 66.7 (C_{10}), 62.1 (C_{12}), 60.9 (C_6), 52.3 (OMe), 52.2 (OMe), 49.7 ($C_{4'}$), 43.1 ($C_{5'}$), 40.1 ($C_{1'}$, $C_{2'}$), 32.1 ($C_{6'}$), 29.8 ($C_{3'}$), 28.5 (tBu-3xMe), 21.0 (OAc), 20.9 (OAc), 20.9 (OAc), 20.9 (OAc), 20.8 (OAc), 20.8 (OAc), 20.6 (OAc); LC-MS (Rt=20.56 min (β), 20.66 min(α)) calcd for $C_{41}H_{59}NO_{23}S$ [M + Na]$^+$ m/z: 988.32; found m/z: 987.62; MS (HRMS): calcd for $C_{41}H_{59}NO_{23}S$ [M + Na]$^+$ m/z: 988.3096; found m/z: 988.3090.

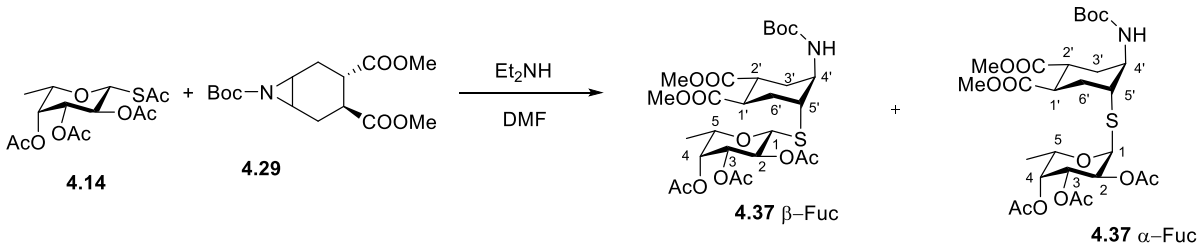
^{1}H NMR (400 MHz, $CDCl_3$)



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



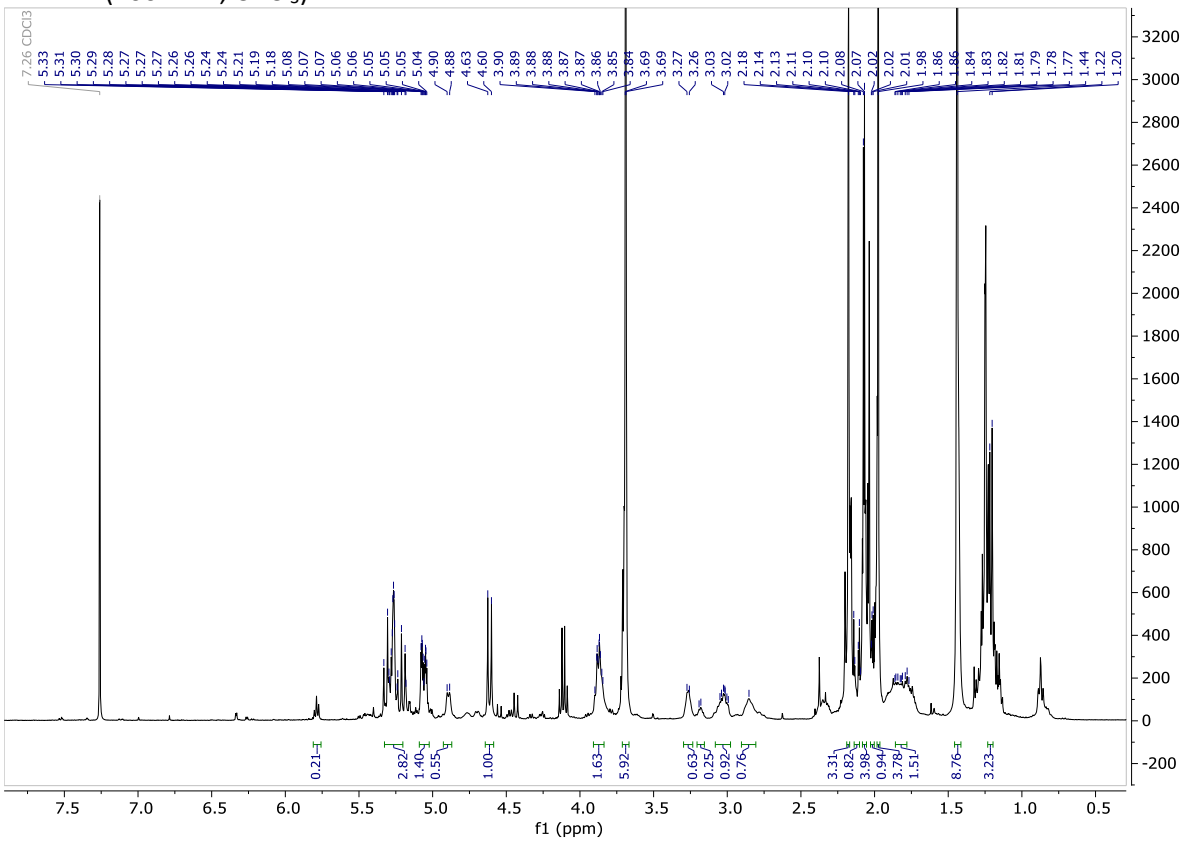
Synthesis of the pseudo *thio*-disaccharide **4.37**



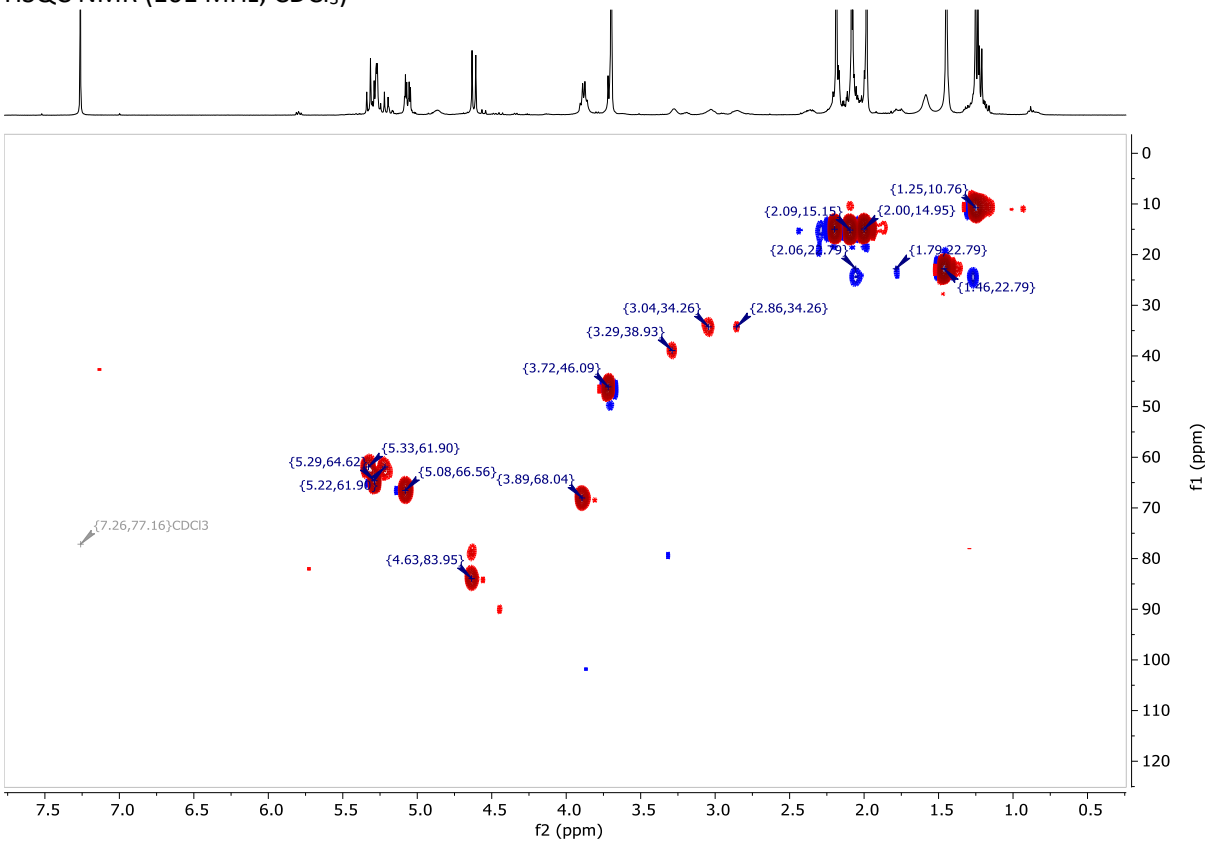
Prepared from *thio*-fucose **4.14** (28.9 mg, 0.083 mmol) and aziridine **4.29** (20 mg, 0.064 mmol) according to the general procedure and purified by flash column chromatography (1:1 hexane:EtOAc) to give **4.37** as a colourless oil in 34% yield (13.6 mg, 0.022 mmol).

R_f=0.35 (Hex/EtOAc 1:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.33 – 5.17 (m, 2H, H_2, H_4), 5.06 (m, 1H, H_3), 4.89 (d, $J_{NH\text{-}4'}$= 6.9 Hz, 1H, NH), 4.61 (d, $J_{1\text{-}2}$= 10.0 Hz, 1H, H_1), 3.93 – 3.84 (m, 2H, H_5, $H_{4'}$), 3.69 (2xs, 6H, 2xCOOMe), 3.27 (m, 1H, $H_{5'}$), 3.12 – 2.98 (m, 1H, $H_{1'}$), 2.85 (m, 1H, $H_{2'}$), 2.18 (s, 3H, OAc), 2.14 – 2.01 (m, 3H, $H_{3'eq}$, $H_{6'ax}$, $H_{6'eq}$), 2.07 (s, 3H, OAc), 1.98 (s, 3H, OAc), 1.87 – 1.76 (m, 1H, $H_{3'ax}$), 1.44 (s, 9H, tBu-3xMe), 1.21 (d, $J_{CH3\text{-}5}$= 6.4 Hz, 3H, CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 84.4 (C_1), 73.2 (C_5), 71.9 (C_4), 70.0 (C_3), 67.3 (C_2), 51.8 (2xCOOMe), 49.1 ($C_{4'}$), 44.3 ($C_{5'}$), 39.6 ($C_{1'}$, $C_{2'}$), 29.7 ($C_{3'}$, $C_{6'}$), 28.1 (tBu-3xMe), 20.5 (OAc), 20.4 (OAc), 20.2 (OAc), 16.0 (CH_3); LC-MS (Rt=19.71 min) calcd for $C_{27}H_{41}NO_{13}S$ $[M + Na]^+$ m/z: 642.23; found m/z: 641.81.

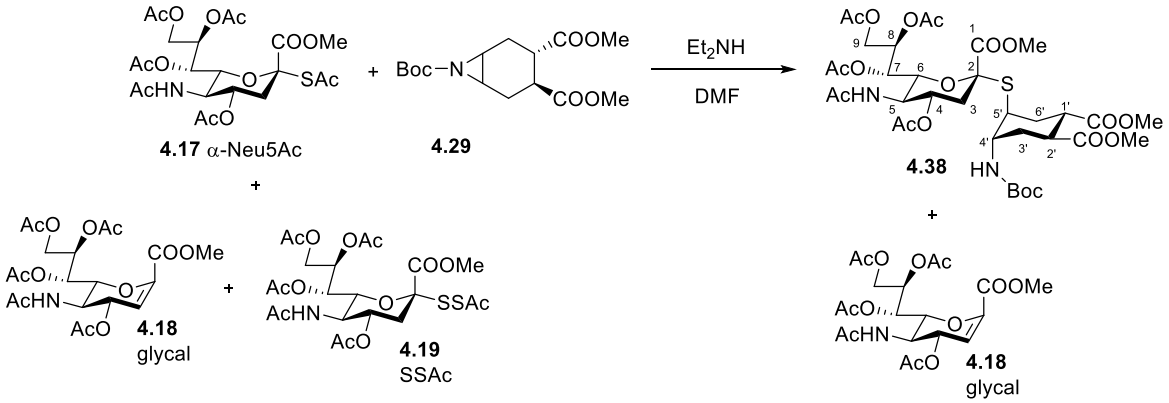
^{1}H NMR (400 MHz, $CDCl_3$)



HSQC NMR (101 MHz, $CDCl_3$)



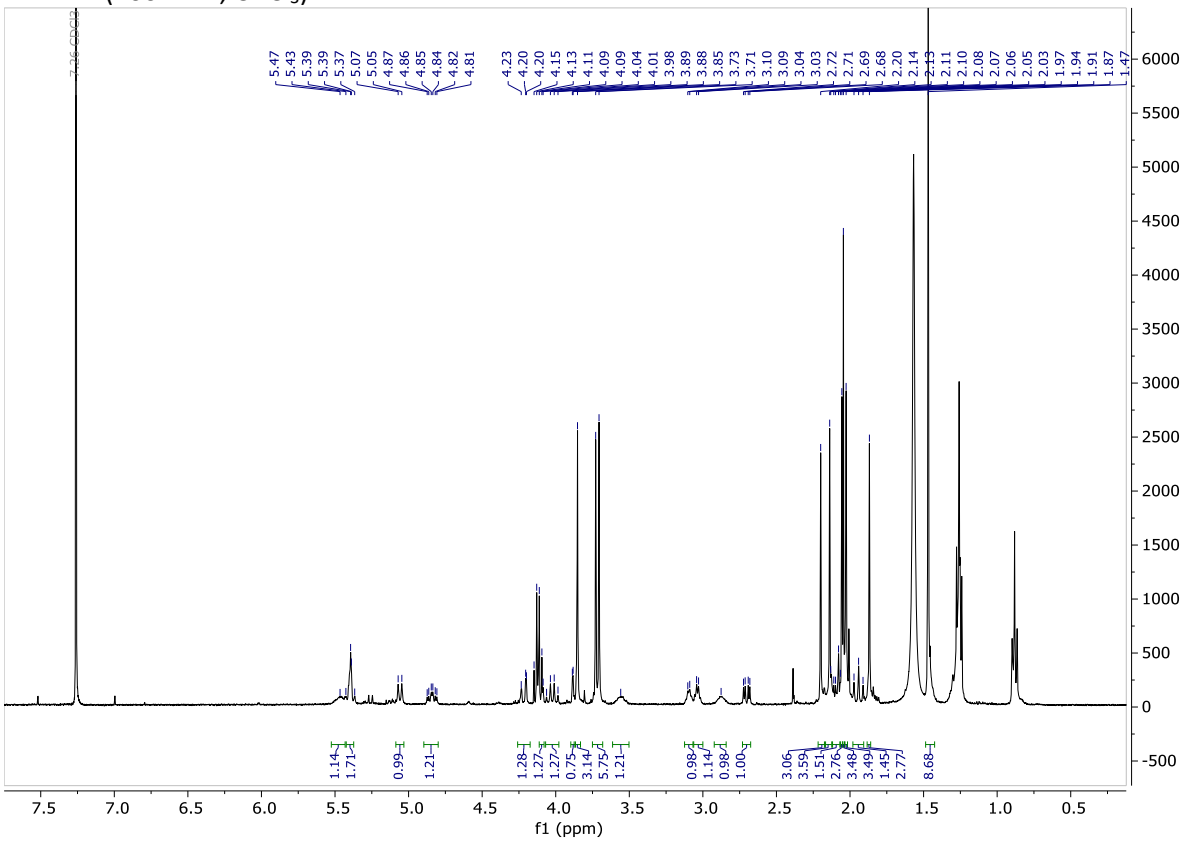
Synthesis of the pseudo *thio*-disaccharide **4.38**



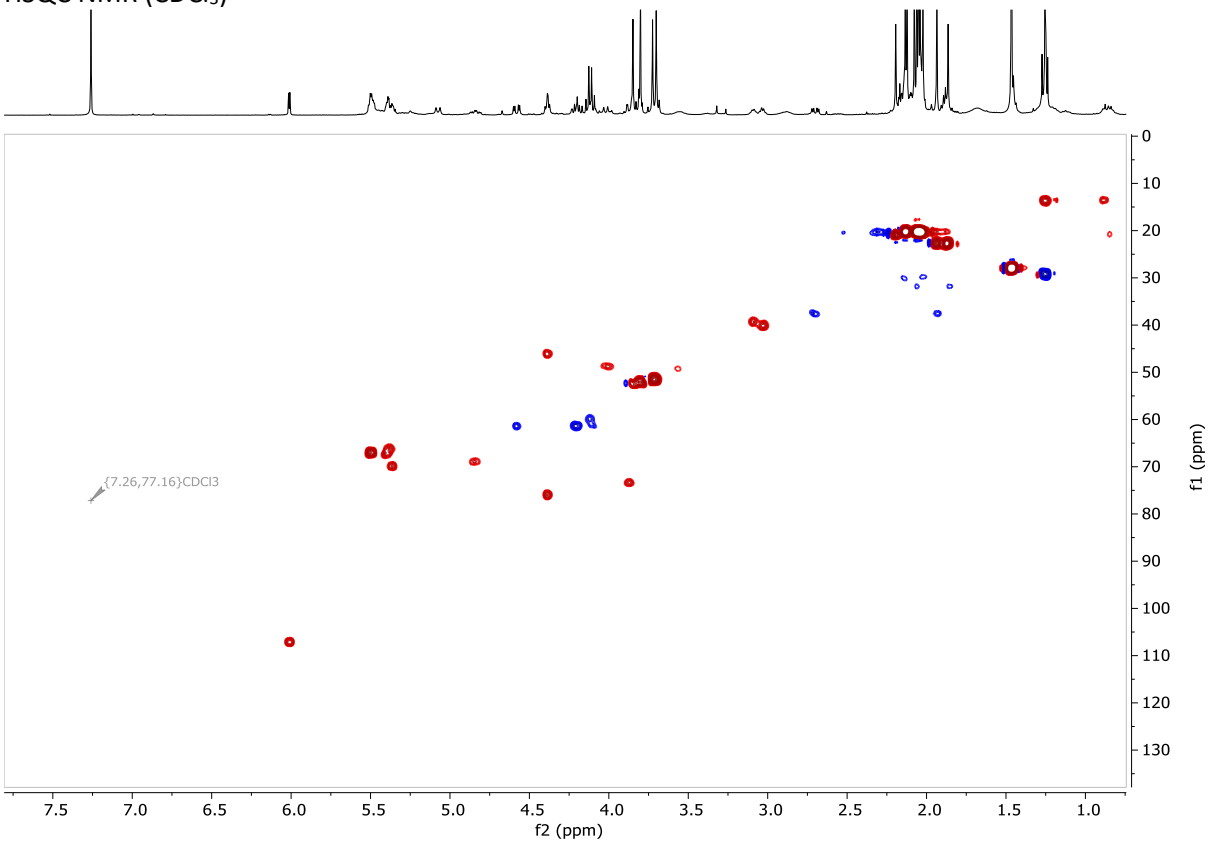
Prepared from peracetylated *thio*-sialic acid **4.17**[13] (68 mg, 0.12 mmol) and aziridine **4.29** (30 mg, 0.096 mmol) according to the general procedure and purified by flash chromatography (3:1 CH_2Cl_2:Acetone) to give **4.38** as a single isomer, in 56% yield (containing 17% glycal **4.18** as estimated by NMR) as yellow waxy solid.

R_f=0.33 (CH_2Cl_2: Acetone=3:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.51 – 5.44 (m, 1H, H_7), 5.43 – 5.36 (mult., 2H, H_8, NH), 5.06 (d, J = 11.0 Hz, 1H, NH), 4.84 (td, $J_{4\text{-}5}$ = $J_{4\text{-}3ax}$ = 11.0 Hz, $J_{4\text{-}3eq}$ = 4.5 Hz, 1H, H_4), 4.28 – 4.18 (m, 1H, H_9), 4.28 – 4.13 (mult., 2H, H_{9a}, H_{9b}), 4.02 (ddd, $J_{5\text{-}NH}$ = $J_{5\text{-}4}$ = $J_{5\text{-}6}$ = 11.0 Hz, 1H, H_5), 3.91 – 3.86 (m, 1H, H_6), 3.85 (s, 3H, COOMe), 3.73 (s, 3H, COOMe), 3.71 (s, 3H, COOMe), 3.60 – 3.51 (m, 1H, $H_{4'}$), 3.13 – 3.07 (m, 1H, $H_{5'}$), 3.07 – 3.01 (m, 1H, $H_{1'}$), 2.91 – 2.84 (m, 1H, $H_{2'}$), 2.70 (dd, $J_{3eq\text{-}3ax}$ = 12.8 Hz, $J_{3eq\text{-}4}$ = 4.5 Hz, 1H, H_{3eq}), 2.20 (s, 3H, OAc), 2.14 (s, 3H, OAc), 2.06 (s, 3H, OAc), 2.18 – 2.04 (mult., 4H, $H_{3'eq}$, $H_{3'ax}$, $H_{6'eq}$, $H_{6'ax}$), 2.03 (s, 3H, OAc), 2.00 – 1.89 (m, 1H, H_{3ax}), 1.87 (s, 3H, NHAc), 1.47 (s, 9H, tBu); ^{13}C NMR (HSQC) (101 MHz, $CDCl_3$) δ 72.3 (C_6), 68.2 (C_4), 66.2 (C_7), 66.0 (C_8), 60.2 (C_9), 51.3 (OMe), 50.6 (OMe), 50.5 (OMe), 47.4 (C_5), 41.3 ($C_{2'}$), 38.6 ($C_{1'}$, $C_{5'}$), 36.6 (C_3), 29.6 ($C_{3'}$, $C_{6'}$), 27.1 (tBu-3xMe), 21.8 (NHAc), 19.8 (Ac), 19.4 (Ac), 19.3 (2xAc); LC-MS (Rt=17.99 min) calcd for $C_{35}H_{52}N_2O_{18}S$ $[M + Na]^+$ m/z: 843.29; found m/z: 842.71; MS (HRMS) calcd for $C_{35}H_{52}N_2O_{18}S$ $[M + Na]^+$ m/z: 843.2834; found m/z: 843.2820.

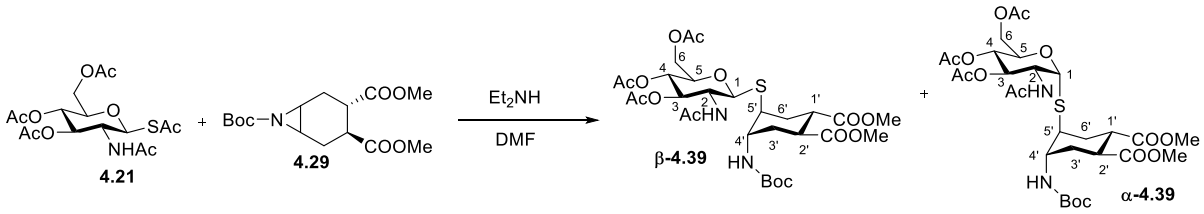
^{1}H NMR (400 MHz, $CDCl_3$)



HSQC NMR ($CDCl_3$)



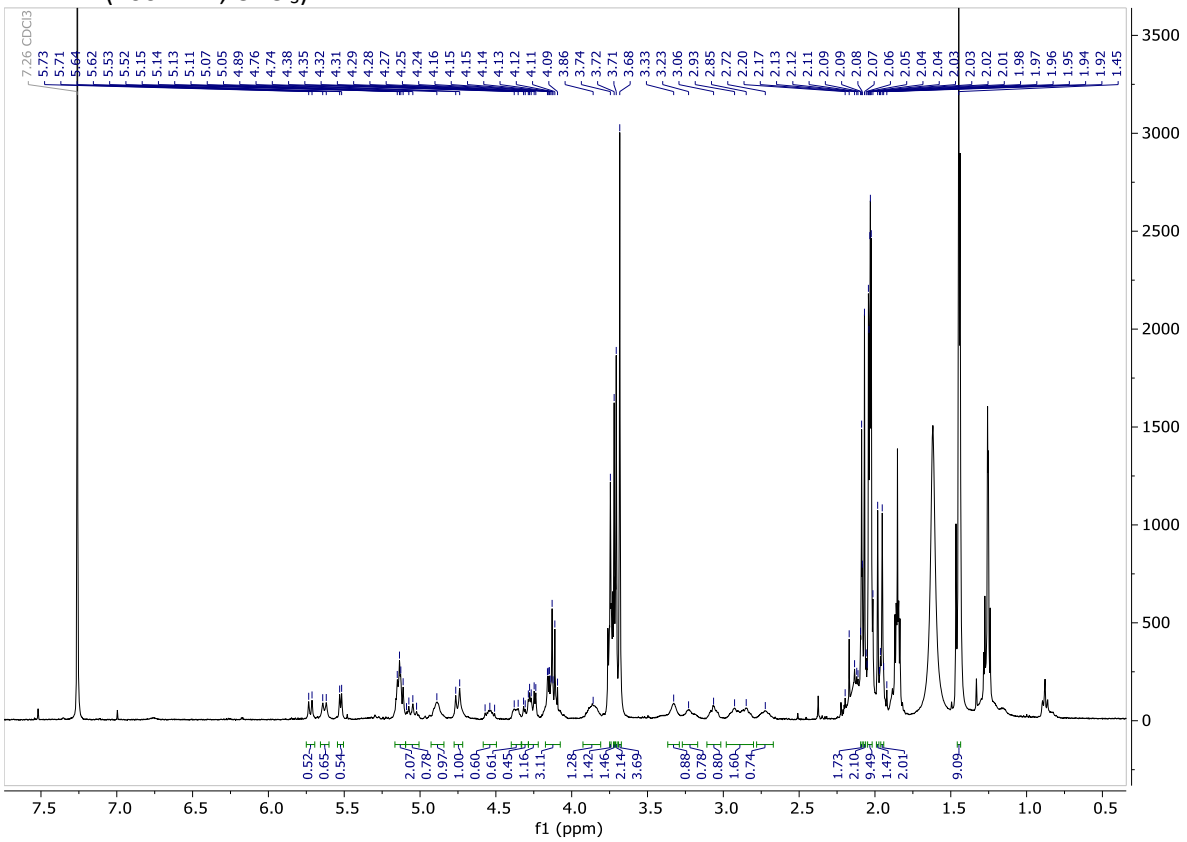
Synthesis of the pseudo *thio*-disaccharide **4.39**



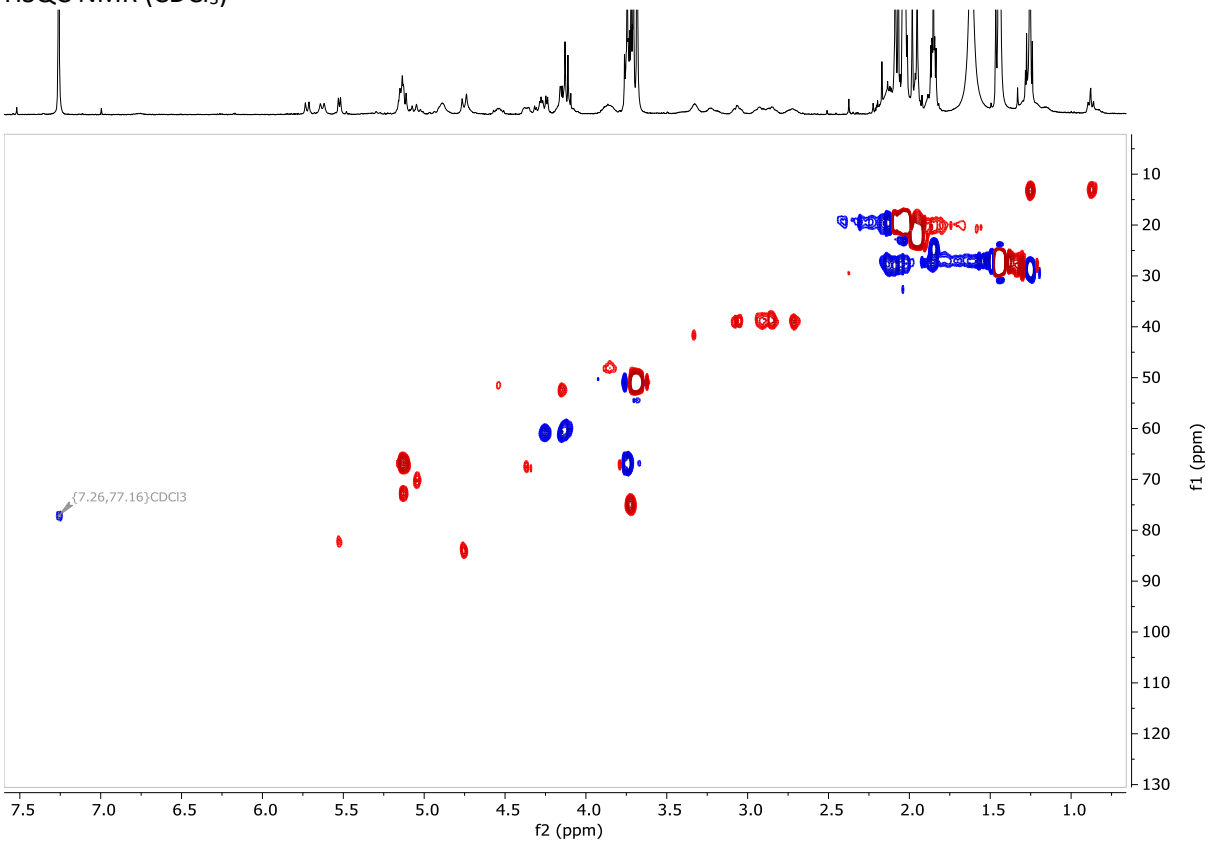
Prepared from peracetylated *thio*-*N*-acetylglucosamine **4.21** (34 mg, 0.083 mmol) and aziridine **4.29** (20 mg, 0.064 mmol) according to the general procedure and purified by flash chromatography (9:1 CH_2Cl_2:MeOH) to give **4.39** as an inseparable anomeric mixture in 19% yield (8 mg, 0.012 mmol) as yellow waxy solid.

R_f =0.08 (Hex/EtOAc 1:1); ^{1}H NMR (400 MHz, $CDCl_3$) **β-4.39** δ 5.63 (d, $J_{NH\text{-}2}$ = 9.3 Hz, 1H, NH), 5.18 – 5.09 (mult., 2H, H_3, H_4), 4.92 – 4.84 (m, 1H, NH), 4.75 (d, $J_{1\beta\text{-}2\beta}$ = 10.4 Hz, 1H, $H_{1\beta}$), 4.40 – 4.34 (m, 1H, H_2), 4.26 (dd, $J_{6a\text{-}6b}$ = 12.1 Hz, $J_{6a\text{-}5}$ = 4.3 Hz, 1H, H_{6a}), 4.18 – 4.13 (m, 1H, H_{6b}), 3.92 – 3.79 (m, 1H, $H_{4'}$), 3.76 – 3.74 (m, 1H, H_5), 3.71 (s, 3H, COOMe), 3.68 (s, 3H, COOMe), 3.35 – 3.30 (m, 1H, $H_{5'}$), 2.97 – 2.90 (m, 1H, $H_{1'}$), 2.75 – 2.68 (m, 1H, $H_{2'}$), 2.13 – 2.01 (mult., 3H, $H_{3'eq}$, $H_{6'ax}$, $H_{6'eq}$), 2.09 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.03 (s, 3H, OAc), 2.02 (s, 3H, OAc), 1.98 (s, 3H, OAc), 1.88 – 1.83 (m, 1H, $H_{3'ax}$) 1.45 (s, 9H, Boc); **α-4.39** δ 5.52 (d, $J_{1\alpha\text{-}2\alpha}$ = 5.2 Hz, 1H, $H_{1\alpha}$); ^{13}C NMR (HSQC) (101 MHz, $CDCl_3$) δ 67.9 (C_3), 73.7 (C_4), 85.2 (C_1), 68.0 (C_2), 61.3 (C_6), 49.0 ($C_{4'}$), 52.1 (OMe), 52.0 (OMe), 41.0 ($C_{5'}$, $C_{2'}$, $C_{1'}$), 29.1 ($C_{6'}$), 28.6 ($C_{3'}$), 28.4 (tBu-3xMe), 23.1 (NHAc), 20.8 (OAc), 20.6 (OAc), 20.7 (OAc); LC-MS (Rt=16.67 min (β), 17.33 min(α)) calcd for $C_{29}H_{44}N_2O_{14}S$ $[M + Na]^+$ m/z: 699.25; found m/z: 698.75; MS (HRMS) calcd for $C_{29}H_{44}N_2O_{14}S$ $[M + Na]^+$ m/z: 699.2411; found m/z: 699.2413.

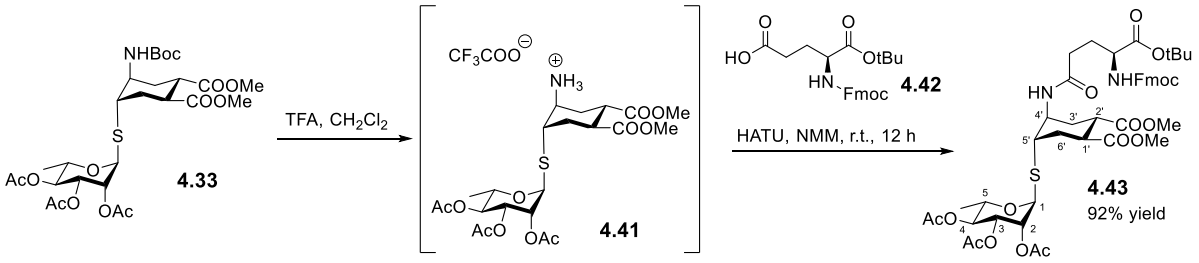
^{1}H NMR (400 MHz, $CDCl_3$)



HSQC NMR ($CDCl_3$)



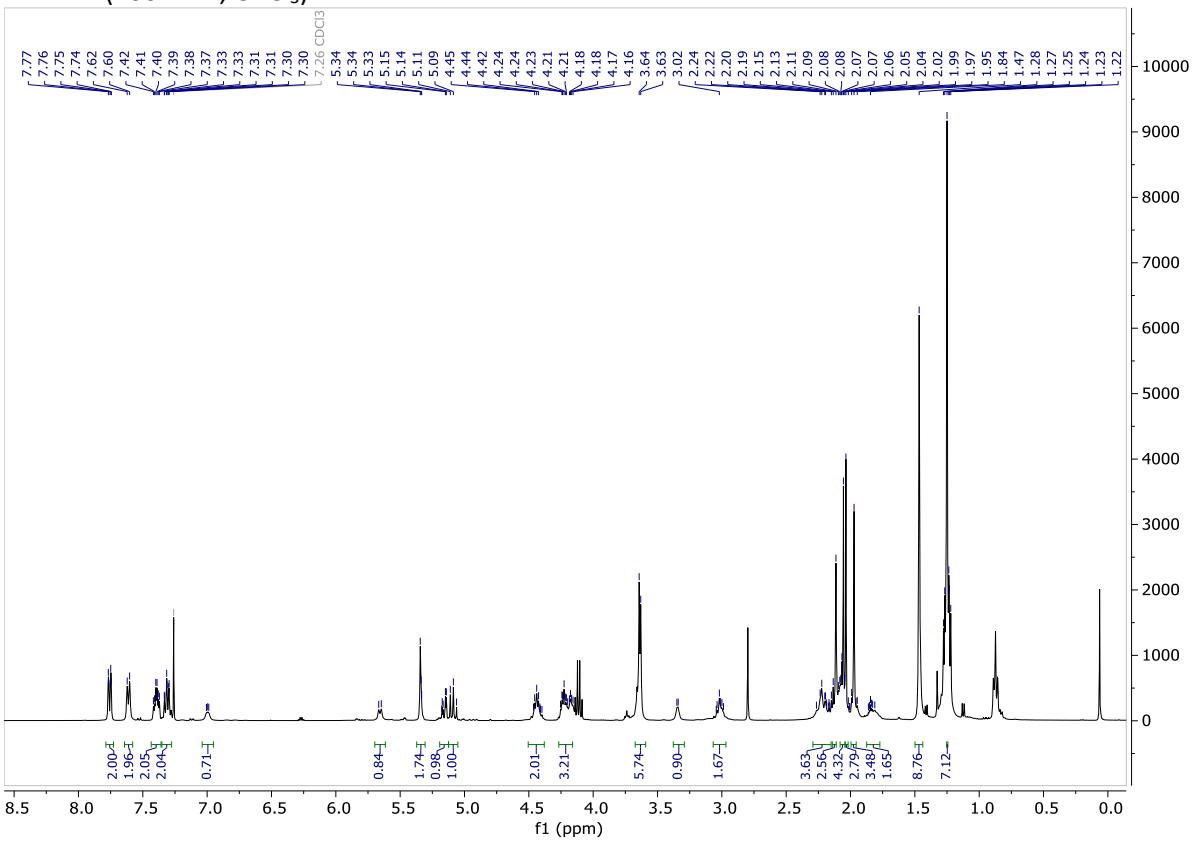
Synthesis of the rhamnosyl aminoacid **4.43**



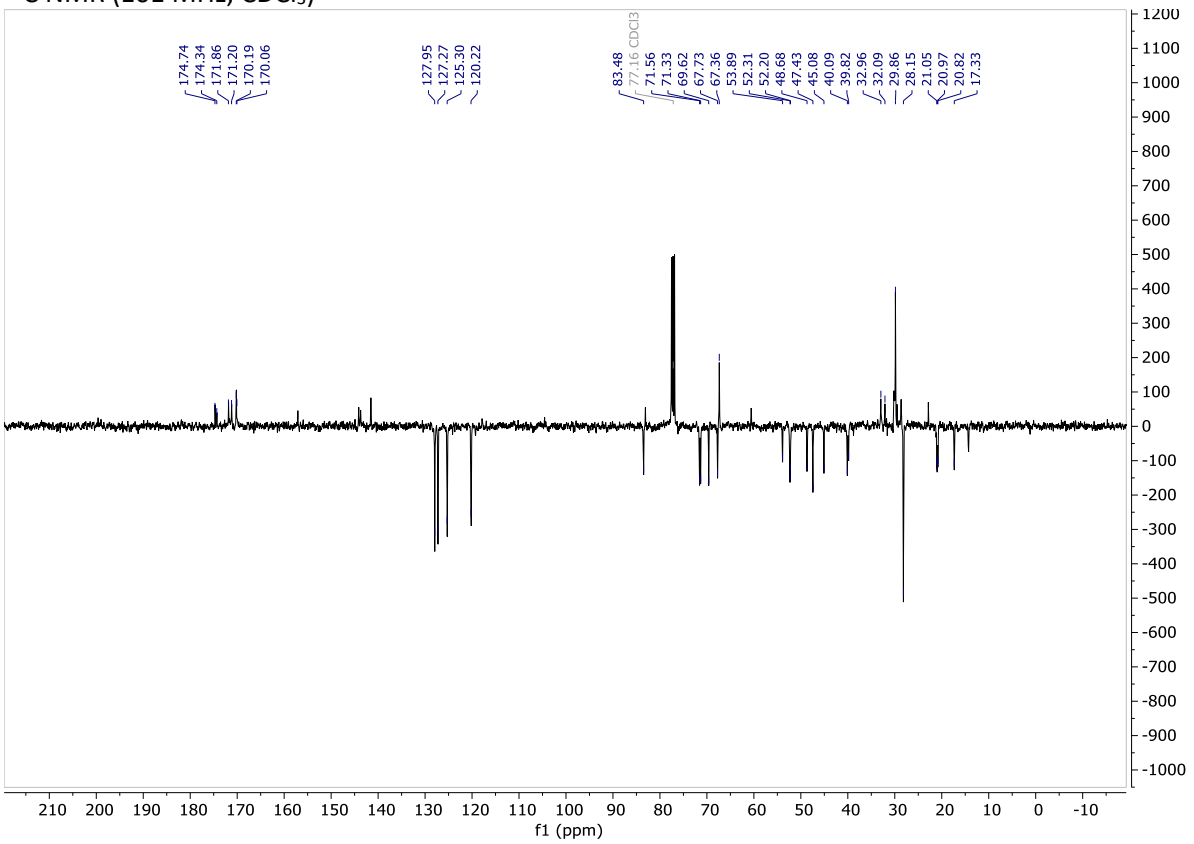
Compound **4.33** (58 mg, 0.093 mmol) was dissolved in CH_2Cl_2 (1.5 mL) and then TFA (0.4 mL) was added. Reaction mixture was kept stirring at RT for 1 h and then concentrated under vacuum. Residue was co-evaporated with toluene 3x. The crude was dissolved in CH_3CN (1.9 mL) and N-methylmorpholine (0.010 mL, 0.093 mmol), amino acid **4.42** (from Fluorochem, 40 mg, 0.093 mmol) and HATU (35 mg, 0.093 mmol) were added. Reaction mixture was kept stirring overnight. Then the reaction mixture was concentrated, dissolved in $CHCl_3$ and washed with water, sat. $KHSO_4$, water, sat. $NaHCO_3$ and water. Combined aqueous phases were additionally washed with $CHCl_3$ and combined organic phases were dried over Na_2SO_4 and concentrated under reduced pressure. The crude was further purified by flash chromatography (1:1 Hex:EtOAc) to give **4.43** in 92% yield (79 mg, 0.086 mmol) as colourless oil.

R_f=0.25 (Hex/EtOAc 1:1); $[\alpha]_D^{23}$ ($CHCl_3$, *c* 1.00): - 37; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (d, *J* = 7.4 Hz, 2H, CH-Ar-Fmoc), 7.62 (d, *J* = 7.5 Hz, 2H, CH-Ar-Fmoc), 7.45 – 7.35 (m, 2H, CH-Ar-Fmoc), 7.36 – 7.26 (m, 2H, CH-Ar-Fmoc), 6.97 (d, *J* = 6.7 Hz, 1H, NH), 5.64 (d, *J* = 8.0 Hz, 1H, NH), 5.35 (s, 2H, H_1, H_2), 5.16 (dd, $J_{3\text{-}4}$ = 10 Hz, $J_{3\text{-}2}$ = 2.9 Hz, 1H, H_3), 5.09 (dd, $J_{4\text{-}3}$ = $J_{4\text{-}5}$ = 10 Hz, 1H, H_4), 4.50 – 4.40 (mult., 2H, CH_2-Fmoc), 4.28 – 4.14 (mult., 4H, CH-Fmoc, H_5, CH-Glu, $H_{4'}$), 3.65 (s, 3H, COOMe), 3.63 (s, 3H, COOMe), 3.37 – 3.32 (m, 1H, $C_{5'}$), 3.07 – 2.96 (mult., 2H, $C_{2'}$, $C_{1'}$), 2.30 – 1.82 (mult., 8H, $C_{3'}$, $C_{6'}$, CH_2-Glu, CH_2-Glu), 2.12 (s, 3H, OAc), 2.06 (s, 3H, OAc), 1.98 (s, 3H, OAc), 1.47 (s, 9H, tBu), 1.23 (d, $J_{CH3\text{-}5}$ = 6.2 Hz, 3H, CH_3-Rha); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.7 (CO), 174.3 (CO), 171.9 (CO), 171.2 (CO), 170.2 (CO), 170.1 (CO), 128.0 (CH-Ar-Fmoc), 127.3 (CH-Ar-Fmoc), 125.3 (CH-Ar-Fmoc), 120.2 (CH-Ar-Fmoc), 83.5 (C_1), 71.6 (C_2), 71.3 (C_4), 69.6 (C_3), 67.7 (C_5), 67.4 (CH_2-Fmoc), 53.9 (CH-Glu), 52.3 (COOMe), 52.2 (COOMe), 48.7 ($C_{4'}$), 47.4 (CH-Fmoc), 45.1 ($C_{5'}$), 40.1 ($C_{1'}$), 39.8 ($C_{2'}$), 33.0 (CH_2-Glu), 32.1 (CH_2-Glu), 29.9 ($C_{3'}$, $C_{6'}$), 28.2 (tBu-3xMe), 21.1 (OAc), 21.0 (OAc), 20.8 (OAc), 17.3 (CH_3-Rha); MS (HRMS) calcd for $C_{46}H_{58}N_2O_{16}S$ $[M + Na]^+$ m/z: 949.3405; found m/z: 949.3400.

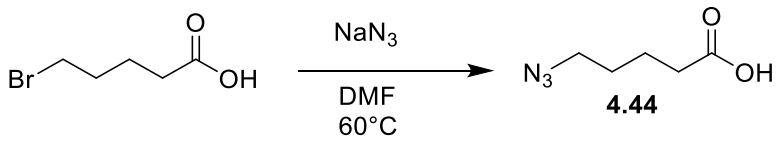
^{1}H NMR (400 MHz, $CDCl_3$)



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



Synthesis of 5-azidovaleric acid **4.44**

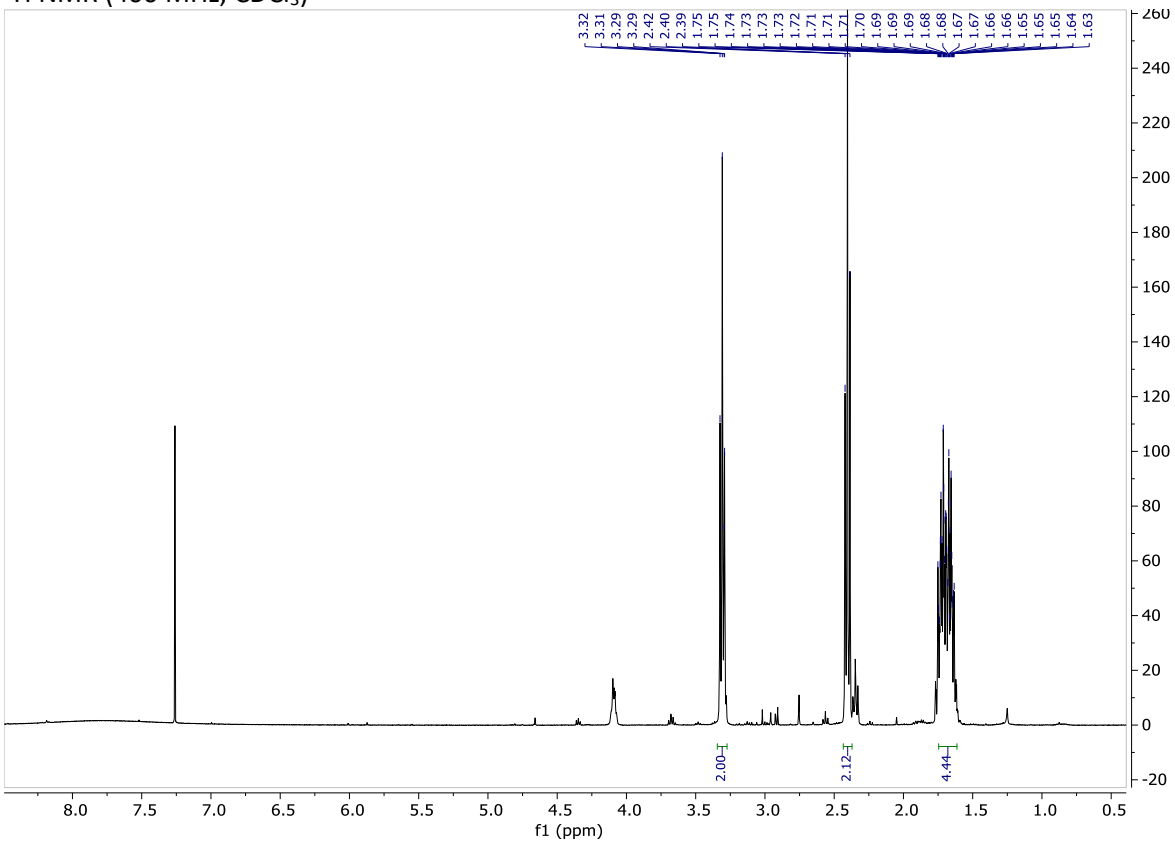


5-Bromovaleric acid (0.5 g, 2.5 mmol) and NaN_3 (0.163 g, 2.5 mmol) were dissolved in DMF (13.75 mL), the reaction mixture was heated to 60°C and kept stirring overnight at this temperature.

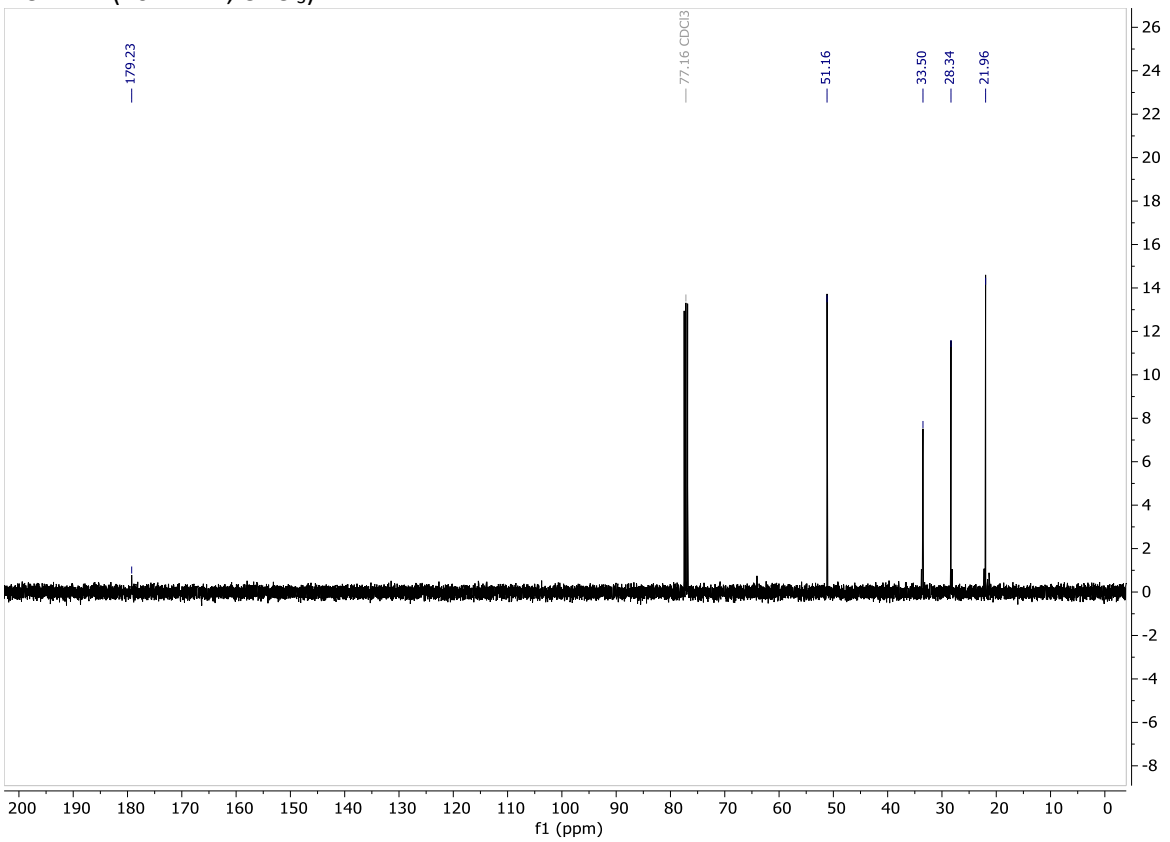
0.1 M HCl was added to the reaction mixture and washed 3x with EtOAc. Combined organic phases were washed with H_2O and brine, then dried over Na_2SO_4 and concentrated under vacuum. Crude product **4.44**[41] (0.30 g) was used without further purification.

R_f=0.74 (100% EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$) δ 3.31 (t, *J* = 6.6 Hz, 2H, CH_2), 2.40 (t, *J* = 7.1 Hz, 2H, CH_2), 1.77 – 1.60 (mult., 4H, 2x CH_2); ^{13}C NMR (101 MHz, $CDCl_3$) δ 179.2 (C=O), 51.2 (CH_2), 33.5 (CH_2), 28.3 (CH_2), 22.0 (CH_2).

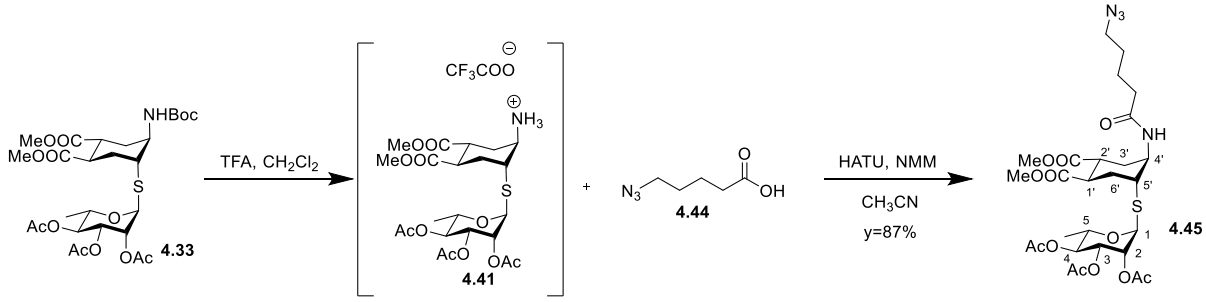
^{1}H NMR (400 MHz, $CDCl_3$)



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



Synthesis of rhamnose-valeric acid conjugate **4.45**



Rhamnose aziridine opening product **4.33** (30 mg, 0.048 mmol) was dissolved in CH_2Cl_2 (0.77 mL) and then TFA (0.19 mL) was added. Reaction mixture was kept stirring at RT for 1 h.

After 1 h the reaction mixture was concentrated under reduced pressure and co-evaporated with toluene to remove TFA. Crude product **4.41** was used immediately and without further purification.

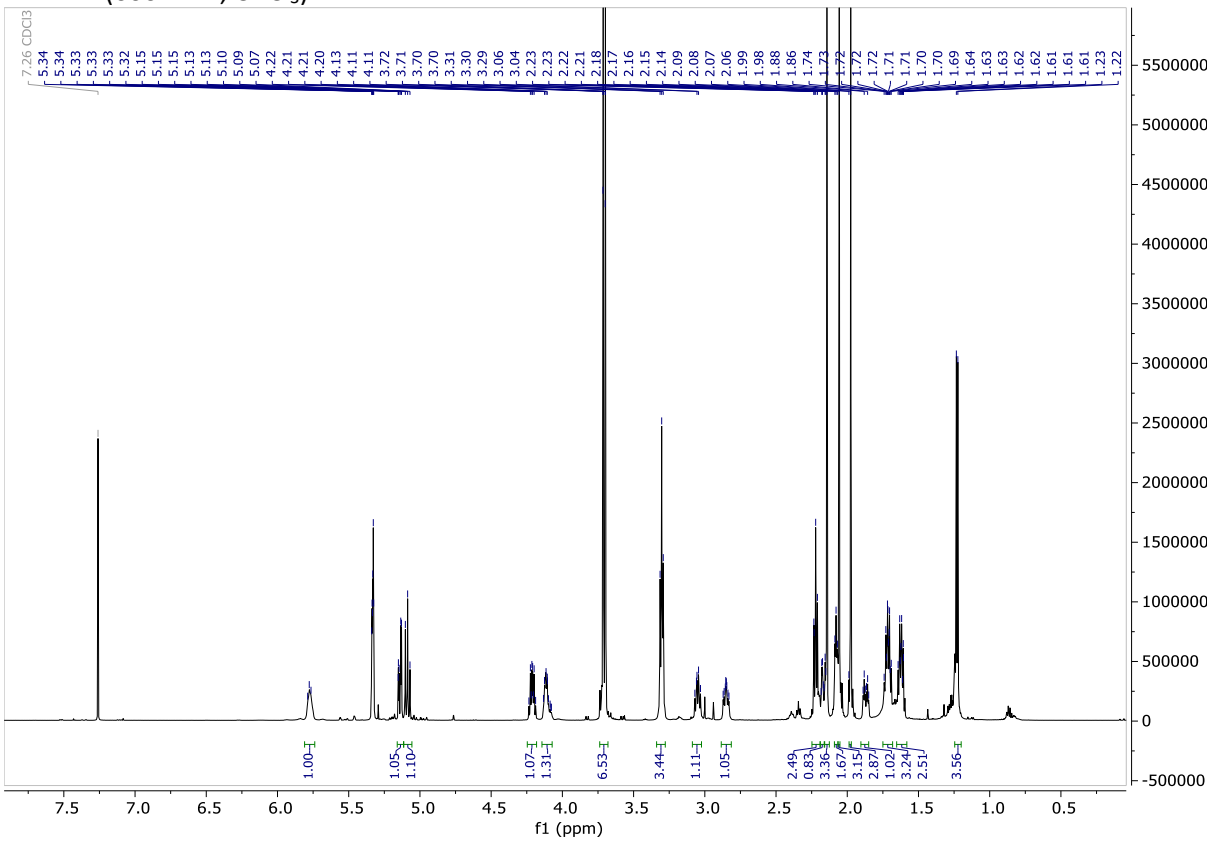
Crude **4.41** from the previous step (0.048 mmol) and 5-azidovaleric acidc **4.44** (14.0 mg, 0.097 mmol) were dissolved in dry CH_3CN (0.97 mL), then *N*-methylmorpholine (10.6 µL, 0.097 mmol) and HATU (27.6 mg, 0.073 mmol) were added. Reaction mixture was kept stirring overnight at RT.

Reaction mixture was concentrated, then the residue was dissolved in $CHCl_3$. Organic phase was washed with H_2O, $KHSO_4$, H_2O, $NaHCO_3$ and H_2O. Combined aqueous phases were washed again with CHCl3. Finally, combined organic phases were concentrated under vacuum.

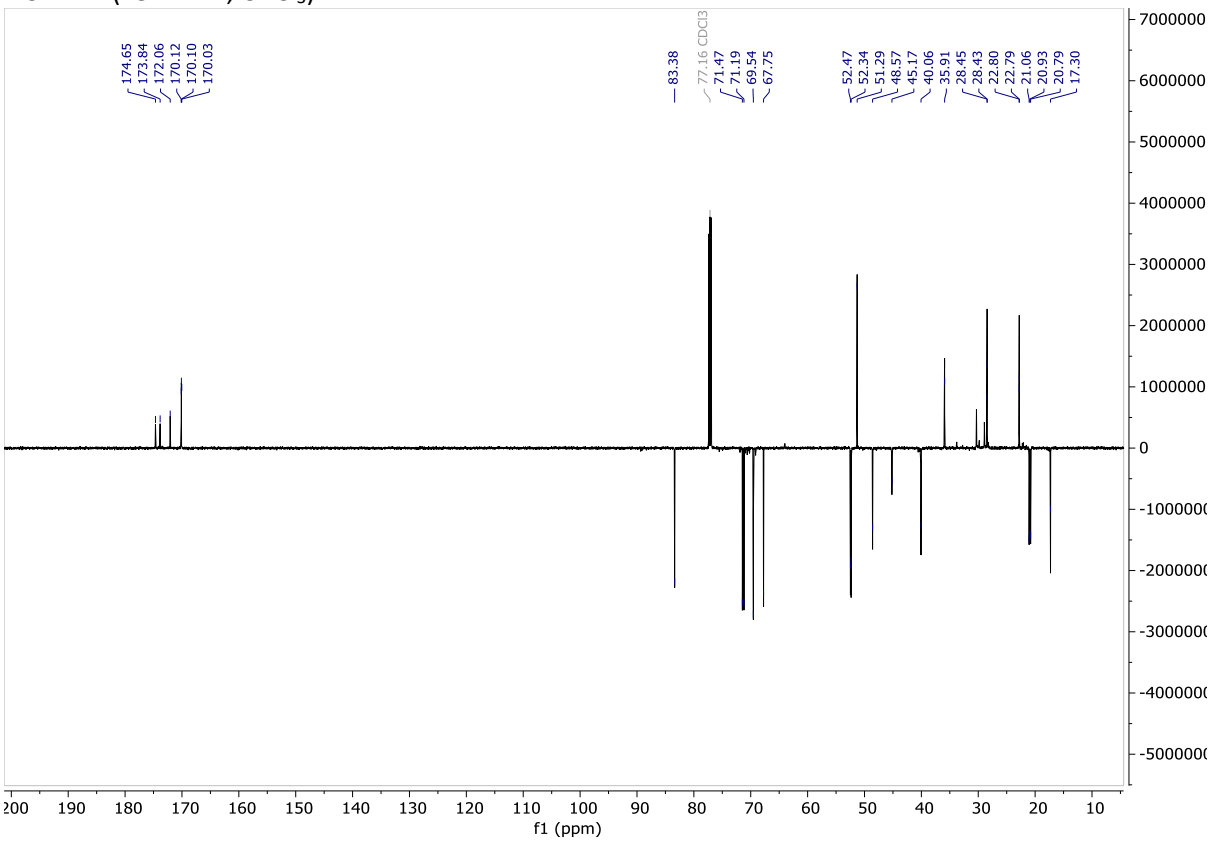
Crude was further purified with column chromatography (petrol ether:EtOAc 1:1). Product **4.45** was obtained as colorless oil in 87% yield (27 mg, 0.042 mmol).

R_f=0.29 (PE:EtOAc 1:1); $[\alpha]_D^{22}$ ($CHCl_3$, *c* 2.6): -45; ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.80 – 5.72 (m, 1H, NH), 5.35 – 5.31 (mult., 2H, H_1, H_2), 5.14 (dd, $J_{3\text{-}4}$= 10.0 Hz, $J_{3\text{-}2}$= 3.1 Hz, 1H, H_3), 5.09 (dd, $J_{4\text{-}3}$= $J_{4\text{-}5}$= 10 Hz, 1H, H_4), 4.21 (dq, $J_{5\text{-}4}$= 10 Hz, $J_{5\text{-}CH3}$= 6.2 Hz, 1H, H_5), 4.14 – 4.09 (m, 1H, $H_{4'}$), 3.71 (s, 3H, COOMe), 3.70 (s, 3H, COOMe), 3.30 (mult., 3H, CH_2, $H_{5'}$), 3.07 – 3.03 (m, 1H, $H_{1'}$), 2.85 (m, 1H, $H_{2'}$), 2.22 (t, *J* = 7.5 Hz, 2H, CH_2), 2.17 (m, 1H, $H_{3'eq}$), 2.14 (s, 3H, OAc), 2.08 (mult., 2H, $H_{6'eq}$, $H_{6'ax}$), 2.06 (s, 3H, OAc), 1.98 (s, 3H, OAc), 1.87 (m, 1H, $H_{3'ax}$), 1.74 – 1.69 (m, 2H, CH_2), 1.65 – 1.59 (m, 2H, CH_2), 1.23 (d, $J_{CH3\text{-}5}$= 6.2 Hz, 3H, CH_3); ^{13}C NMR (151 MHz, $CDCl_3$) δ 174.7 (CO), 173.8 (CO), 172.1 (CO), 170.1 (CO), 170.1 (CO), 170.0 (CO), 83.4 (C_1), 71.5 (C_2), 71.2 (C_4), 69.5 (C_3), 67.7 (C_5), 52.5 (COOMe), 52.3 (COOMe), 51.3 (CH_2), 48.6 ($C_{4'}$), 45.2 ($C_{5'}$), 40.1 ($C_{2'}$, $C_{1'}$), 35.9 (CH_2), 28.4 ($C_{3'}$), 28.4 ($C_{6'}$), 22.8 (CH_2), 22.8 (CH_2), 21.1 (OAc), 20.9 (OAc), 20.8 (OAc), 17.3 (CH_3); MS (HRMS) calcd for $C_{27}H_{40}N_4O_{12}S$ $[M + Na]^+$ m/z: 667.2400; found m/z: 667.2385.

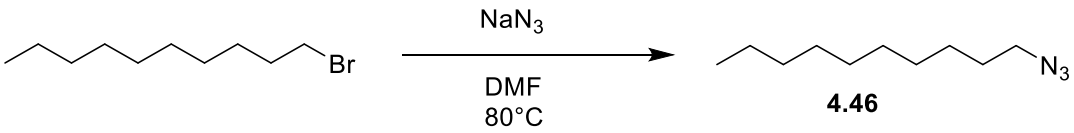
^{1}H NMR (600 MHz, $CDCl_3$)



^{13}C NMR (151 MHz, $CDCl_3$)



Synthesis of 1-azidodecane **4.46**



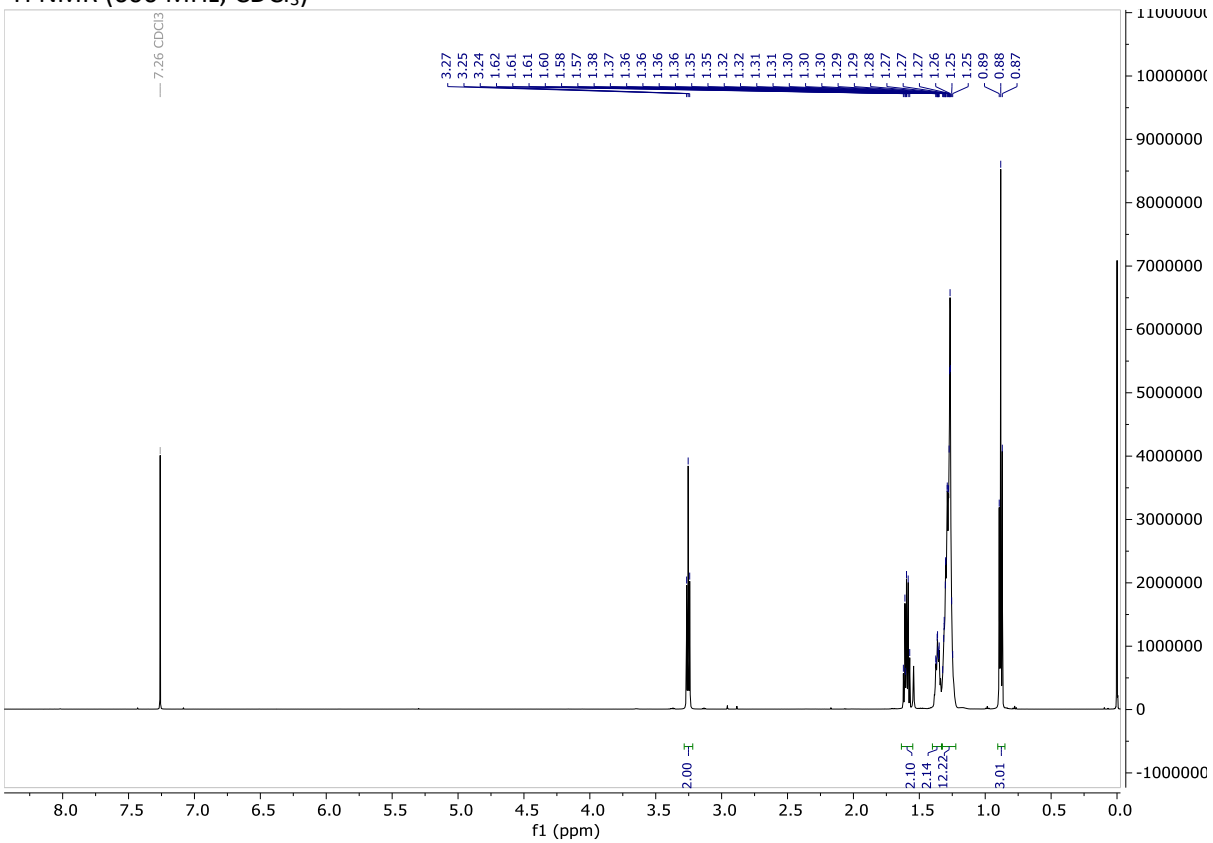
4.46

1-bromodecane (0.1 mL, 0.48 mmol) and NaN_3 (0.16 g, 2.41 mmol) were dissolved in DMF (1.93 mL). The reaction mixture was heated to 80°C and kept stirring overnight at this temperature.

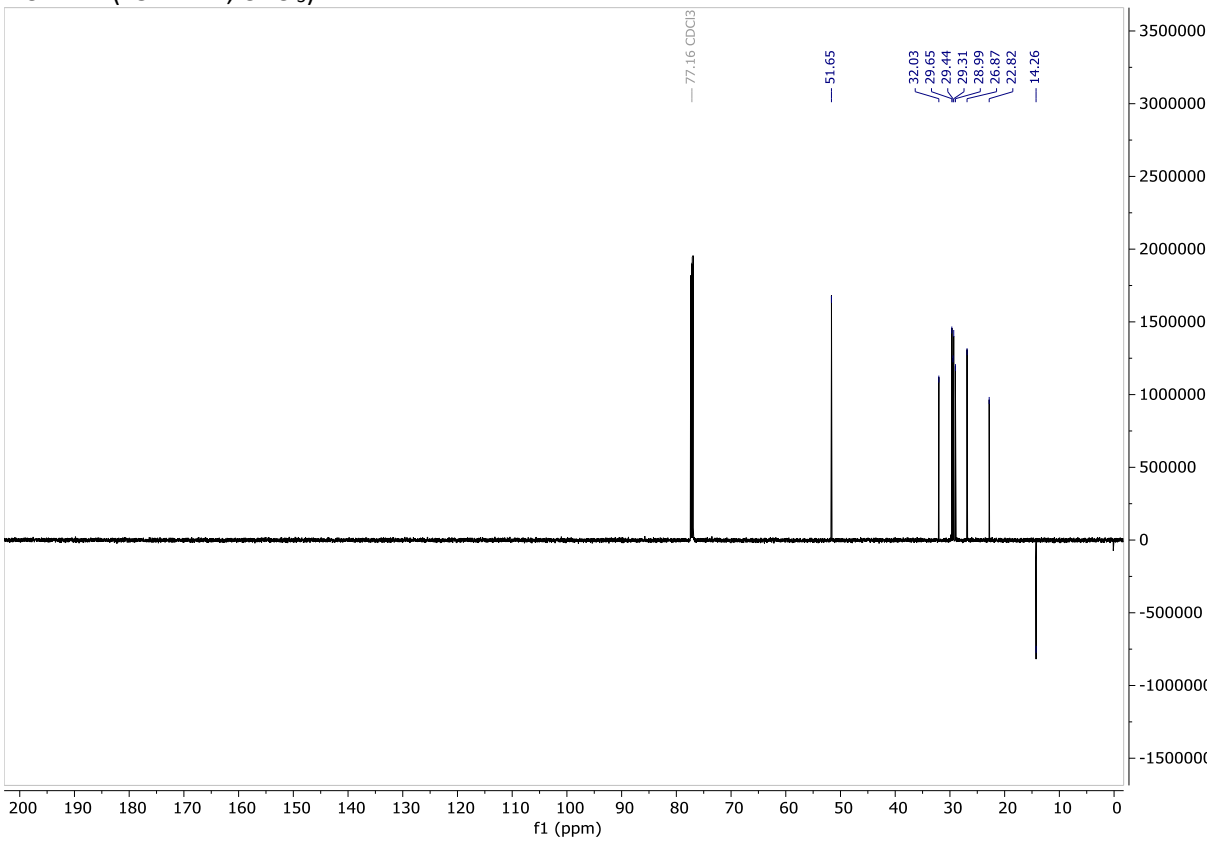
H_2O was added to the reaction mixture and extracted 3x with CH_2Cl_2. Combined organic phases were washed with H_2O and brine and dried over Na_2SO_4, then concentrated under reduced pressure. Crude **4.46**[42] (19 mg, 0.1 mmol) was used without further purification.

^{1}H NMR (600 MHz, $CDCl_3$) δ 3.25 (t, $J_{CH2N3-CH2}$= 7.0 Hz, 2H, CH_2N_3), 1.60 (dt, $J_{CH2-CH2}$= 14.6, $J_{CH2-CH2}$= 7.0 Hz, 2H, CH_2), 1.42 – 1.33 (m, 2H, CH_2), 1.34 – 1.21 (mult., 12H, 6xCH_2), 0.88 (t, $J_{CH3-CH2}$= 7.0 Hz, 3H, CH_3); ^{13}C NMR (151 MHz, $CDCl_3$) δ 51.7 (CH_2N_3), 32.0 (CH_2), 29.7 (CH_2), 29.4 (CH_2), 29.3 (CH_2), 29.0 (CH_2), 26.9 (CH_2), 22.8 (CH_2), 14.3 (CH_3).

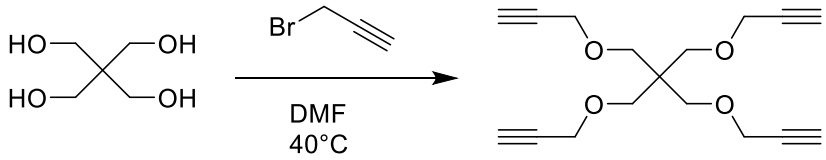
^{1}H NMR (600 MHz, $CDCl_3$)



^{13}C NMR (151 MHz, $CDCl_3$)



Synthesis of tetravalent alkyne **4.47**



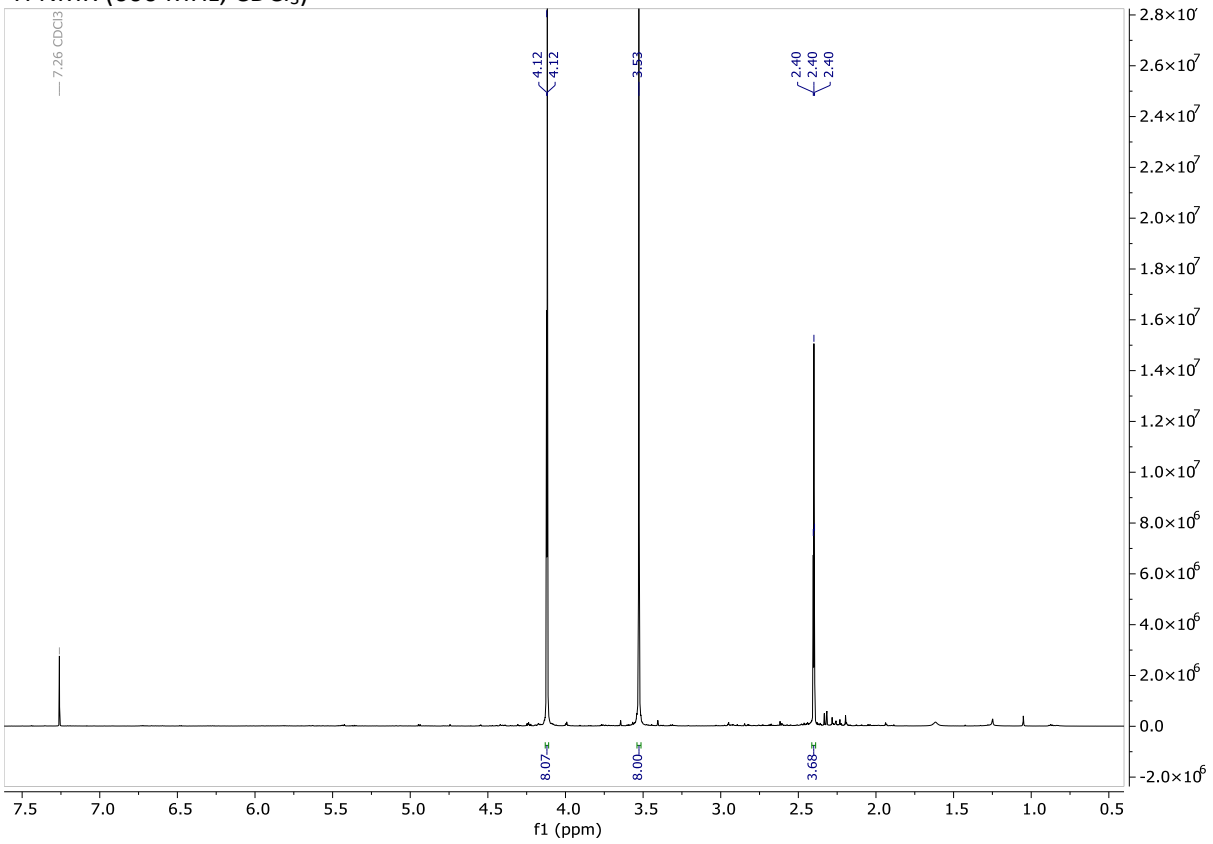
4.47

Pentaerytritol (0.5 g, 3.67 mmol) and KOH (3.1 g, 55.2 mmol) were suspended in DMF at 0°C and kept stirring for 5 min. Then a solution of propargyl bromide (5.8 mL) (80% in toluene) was slowly added at 0°C over a period of 30 min. The reaction mixture was then heated to 40°C and kept stirring overnight at this temperature.

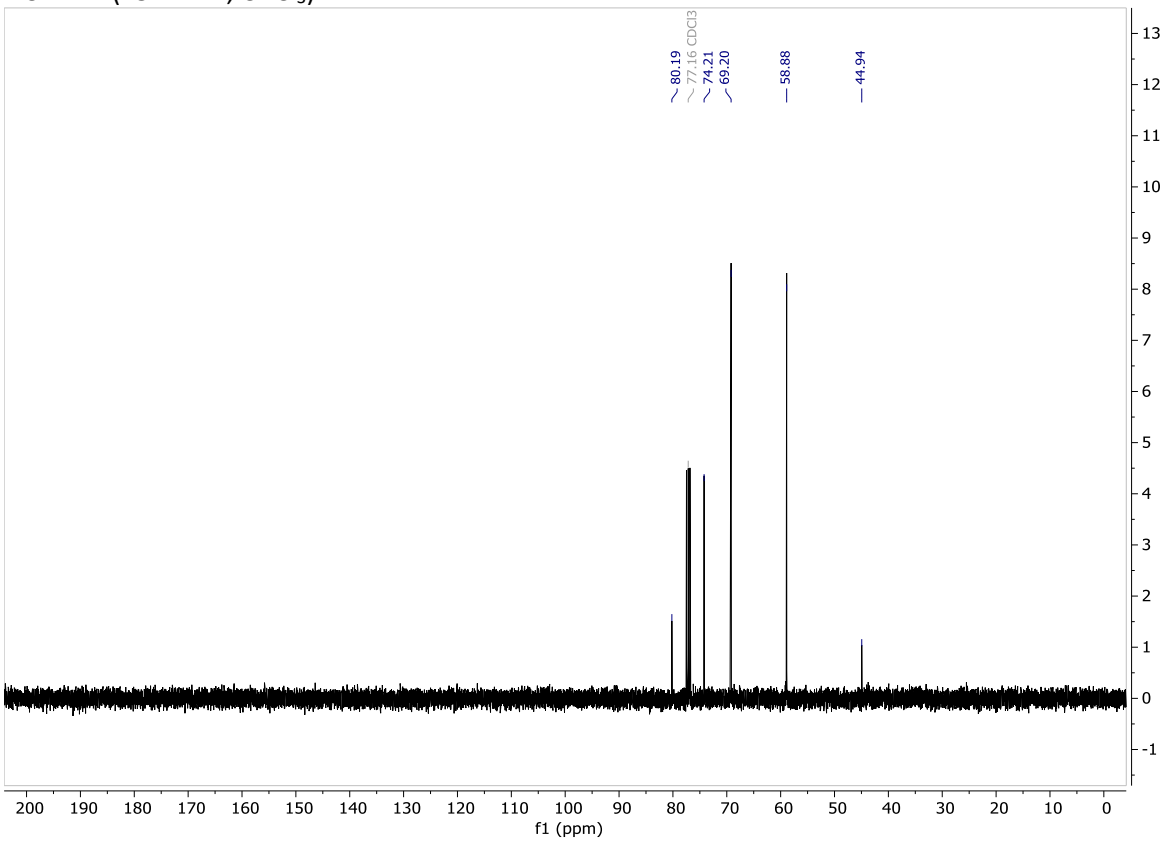
Water was added to the reaction mixture and washed 3x with EtOAc. Combined organic phases were washed with brine and dried over Na_2SO_4, then concentrated under reduced pressure. Crude was further purified by column chromatography (petrol ether: EtOAc 85:15). Product **4.47**[43] was obtained as brown wax in 51% yield (0.541 g, 1.87 mmol).

R_f=0.4 (PE:EtOAc 85:15); ^{1}H NMR (600 MHz, $CDCl_3$) δ 4.12 (d, *J* = 2.4 Hz, 8H, $\underline{CH_2}$-C-CH), 3.53 (s, 8H, C-$\underline{CH_2}$-O), 2.40 (t, *J* = 2.4 Hz, 4H, CH); ^{13}C NMR (101 MHz, $CDCl_3$) δ 80.2 (<u>C</u>-CH), 74.2 (CH), 69.2 (C-$\underline{CH_2}$-O), 58.9 (CH-C-$\underline{CH_2}$), 44.9 (C-O).

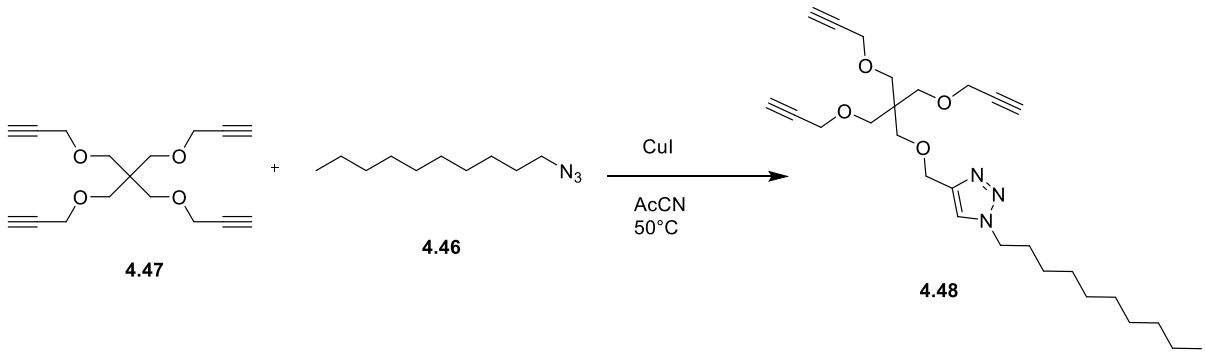
^{1}H NMR (600 MHz, $CDCl_3$)



^{13}C NMR (151 MHz, $CDCl_3$)



Synthesis of trivalent alkyne **4.48**

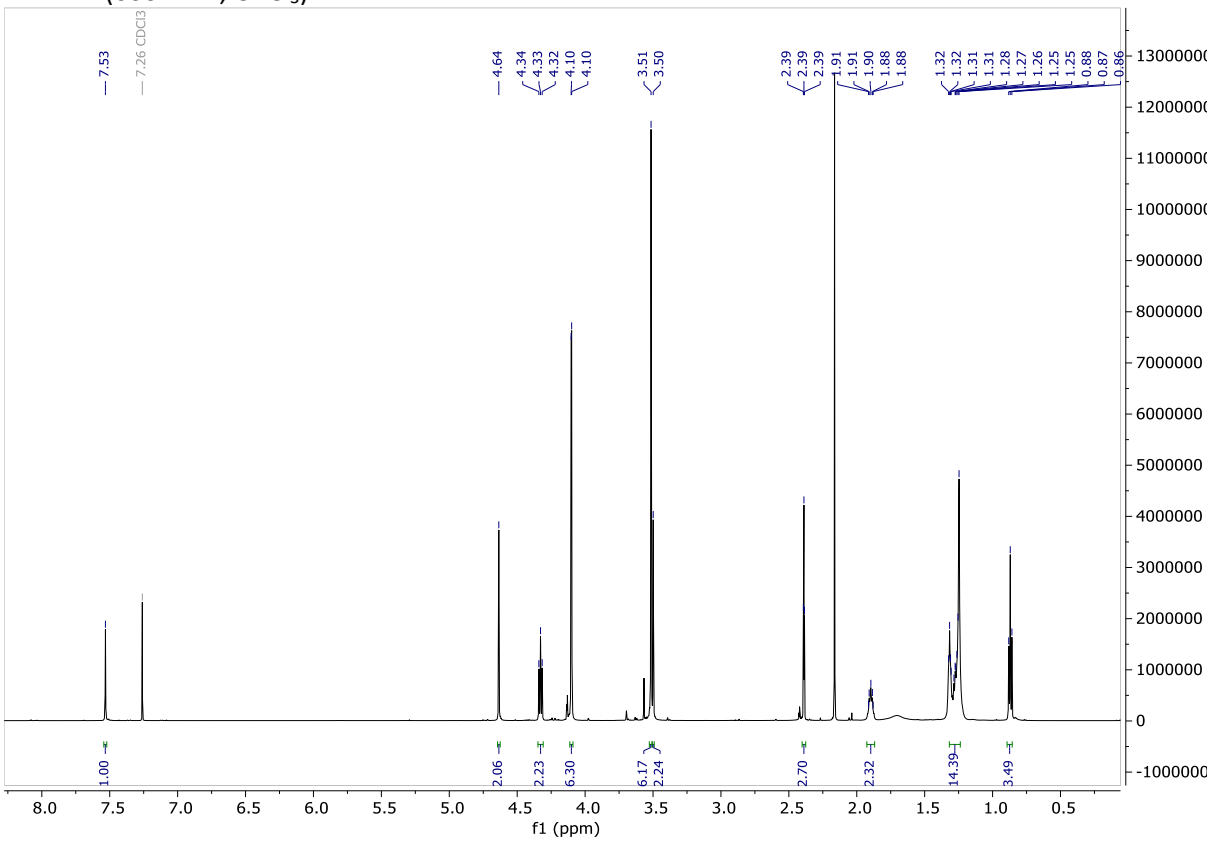


Alkyne **4.47** (94.4 mg, 0.38 mmol) was dissolved in dry CH_3CN (0.4 mL), then CuI (1.6 mg, 0.0082 mmol) was added. Azide **4.46** (15 mg, 0.082 mmol) was dissolved in dry CH_3CN (0.4 mL) and then slowly added to the solution of alkyne over 30 min at RT. Afterwards, the reaction mixture was heated to 50°C and kept stirring overnight at this temperature.

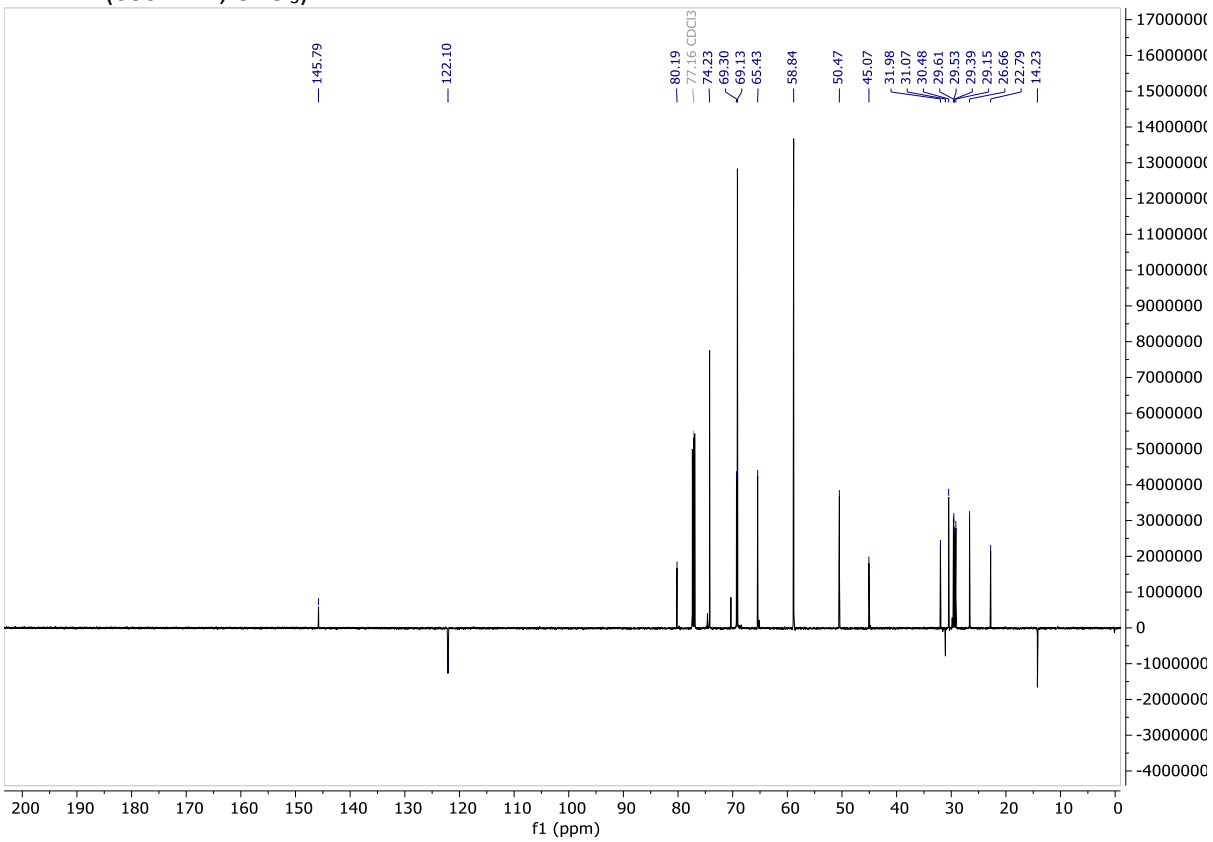
Reaction mixture was concentrated under reduced pressure. Crude product was then purified with column chromatography (petrol ether:EtOAc 1:1). Product **4.48** was obtained as colourless oil in 35% yield (13.6 mg, 28.84 mmol).

R_f=0.37 (PE:EtOAc 1:1); ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.53 (s, 1H, CH-triazole), 4.64 (s, 2H, O-$\underline{CH_2}$-triazole), 4.33 (t, *J* = 7.3 Hz, 2H, CH_2-decane), 4.10 (d, *J* = 2.4 Hz, 6H, 3xCH-C-$\underline{CH_2}$), 3.51 (s, 6H, 3xC-$\underline{CH_2}$-O), 3.50 (s, 2H, C-$\underline{CH_2}$-O), 2.39 (t, *J* = 2.4 Hz, 3H, 3xCH-alkyne), 1.90 (t, *J* = 7.1 Hz, 2H, CH_2-decane), 1.33 – 1.22 (m, 14 H, 7xCH_2-decane), 0.87 (t, *J* = 7.0 Hz, 3H, CH_3-decane); ^{13}C NMR (151 MHz, $CDCl_3$) δ 145.8 (C), 122.1 (CH-triazole), 80.2 (CH-$\underline{C}$-CH_2), 74.2 ($\underline{CH}$-C-CH_2), 69.3 (C-$\underline{CH_2}$-O), 69.1 (C-$\underline{CH_2}$-O), 65.4 (O-$\underline{CH_2}$-triazole), 58.8 (CH-C-$\underline{CH_2}$), 50.5 ($\underline{CH_2}$-decane), 45.1 ($\underline{C}$-CH_2-O), 32.0 (CH_2-decane), 31.1 (CH_2-decane), 30.5 (CH_2-decane), 29.6 (CH_2-decane), 29.5 (CH_2-decane), 29.4 (CH_2-decane), 29.2 (CH_2-decane), 26.7 (CH_2-decane), 22.8 (CH_2-decane), 14.2 (CH_3-decane); MS (HRMS) calcd for $C_{27}H_{41}N_3O_4$ $[M + Na]^+$ m/z: 494.3100; found m/z: 494.3107.

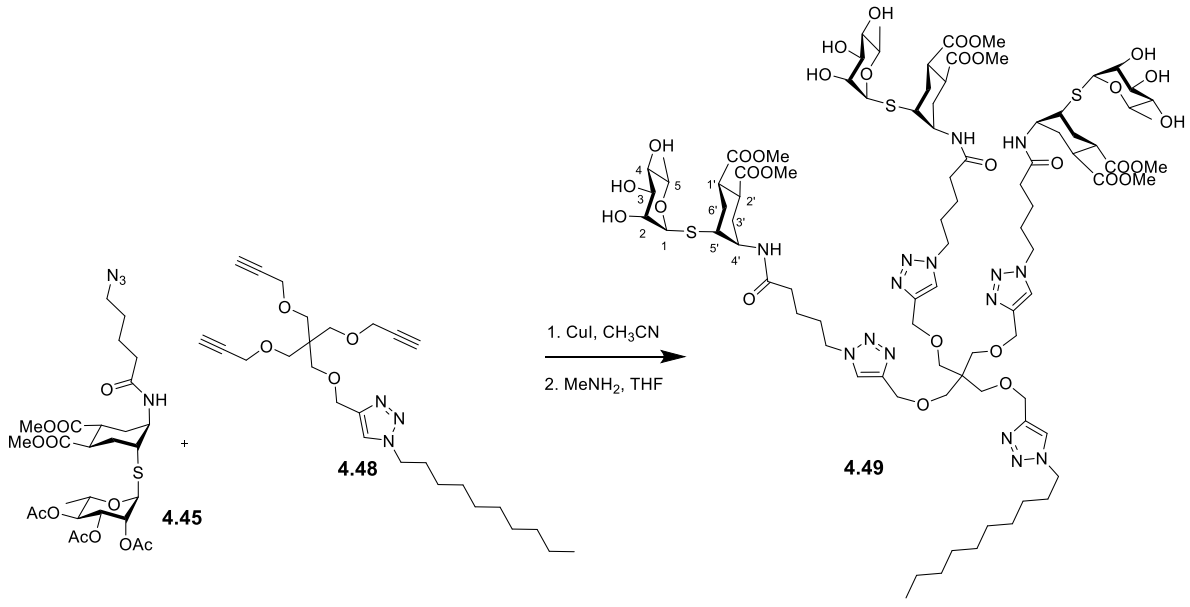
^{1}H NMR (600 MHz, $CDCl_3$)



^{1}H NMR (600 MHz, $CDCl_3$)



Synthesis of multivalent rhamnose construct **4.49**



Rhamnose glycomimetic azide **4.45** (14 mg, 0.022 mmol) was dissolved in dry CH_3CN (0.11 mL) and then CuI (1.25 mg, 0.0066 mmol) was added. Alkyne **4.48** (3.1 mg, 0.0066 mmol) was dissolved in CH_3CN (0.11 mL) and slowly added to the solution of the azide. Reaction mixture was heated to 40°C

and kept stirring overnight at this temperature. The next day additional portion of CuI (1.25 mg, 0.0066 mmol) was added and reaction was kept stirring overnight again at 40°C.

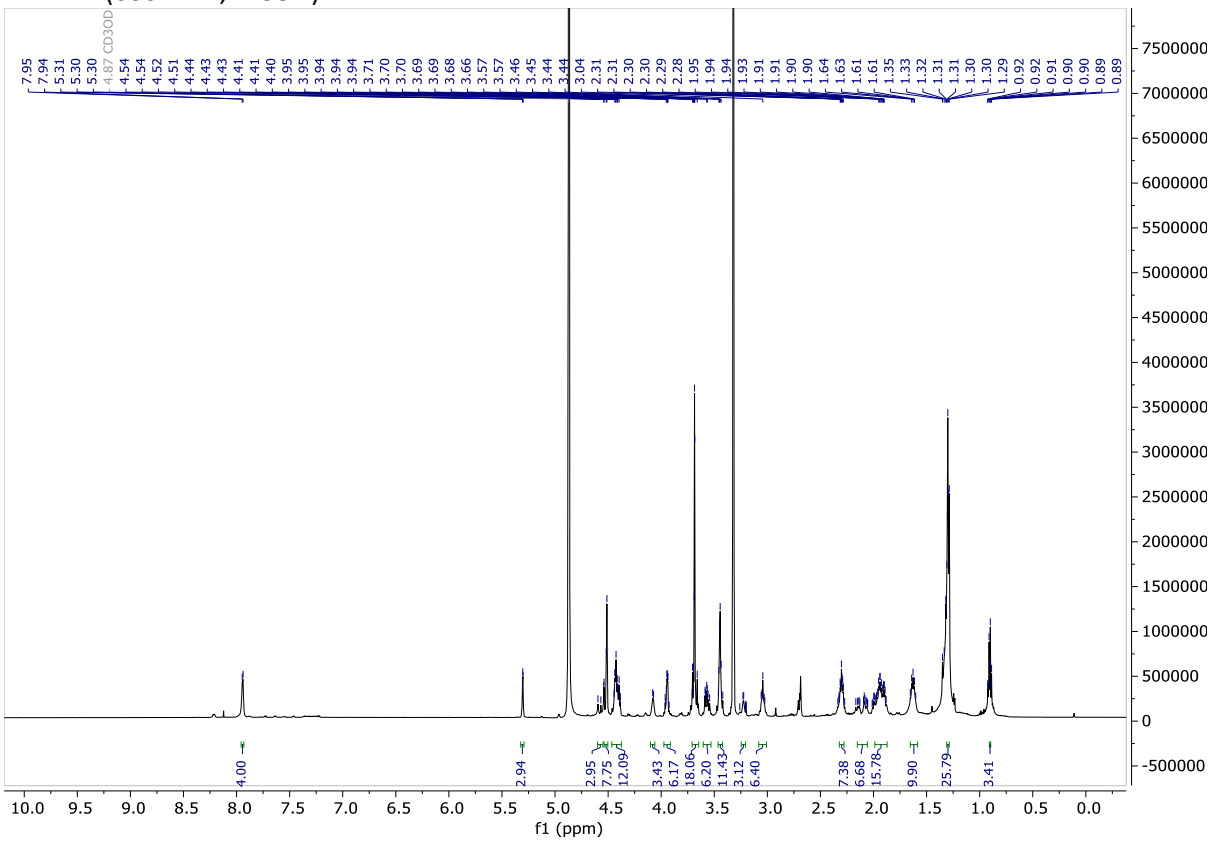
Reaction mixture was concentrated under reduced pressure. The residue was dissolved in CH_3OH and Cu-chelating resin (Cuprisorb) was added and left stirring for 3 h. The mixture was filtered and the filtrate concentrated.

The residue was dissolved in 2 M $MeNH_2$ in THF (0.132 mL) and left stirring overnight at RT.

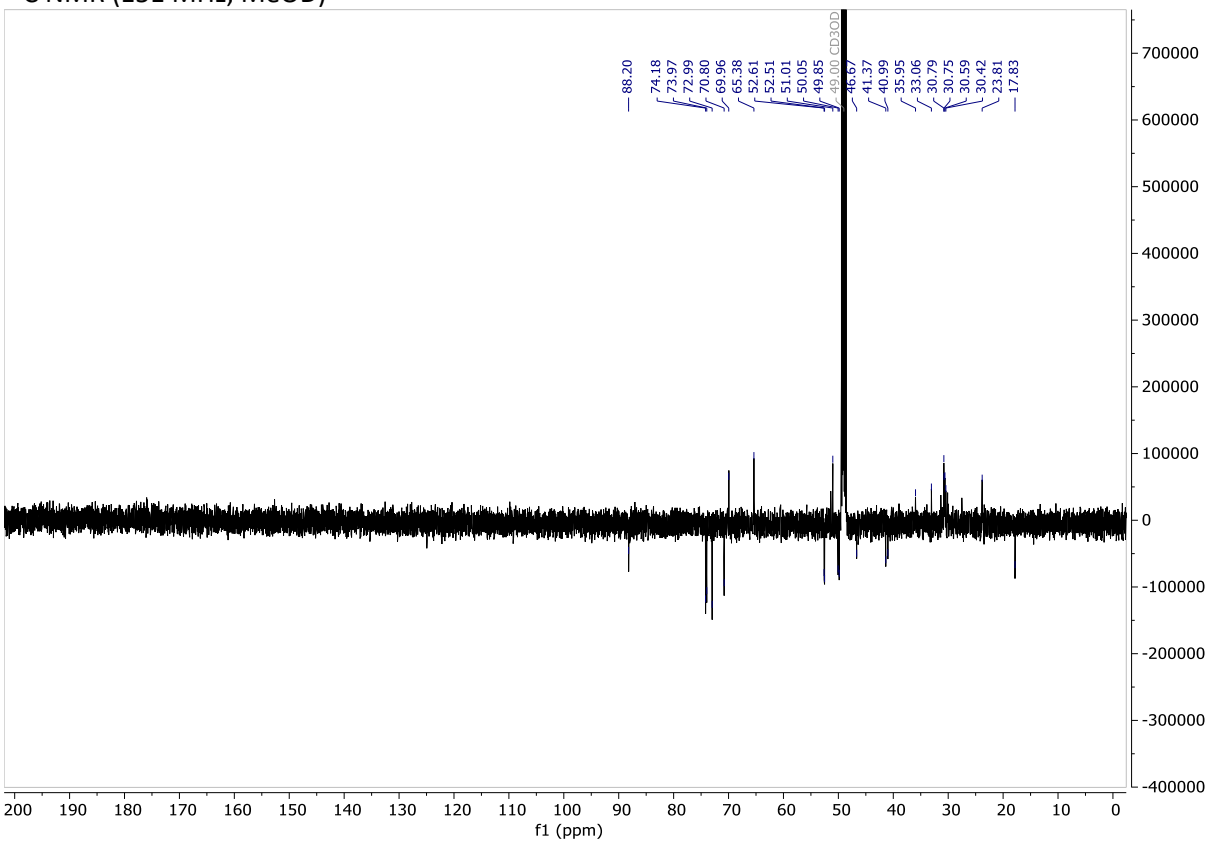
Crude product was purified by size exclusion column chromatography (P2 gel, H_2O:CH_3OH 1:1). Product **4.49** was obtained as colourless oil in 29% yield (3.9 mg, 0.001914 mmol).

$[\alpha]_D^{24}$ (CH_3OH, *c* 0.15): - 84; ^{1}H NMR (600 MHz, MeOD) δ 7.94 (s, 4H, CH), 5.30 (s, 3H, C_1), 4.61 – 4.53 (m, 3H, NH), 4.53 – 4.50 (mult., 8H, $\underline{CH_2}$-triazole), 4.46 – 4.38 (mult. 12H, $\underline{CH_2}$-triazole), 4.08 (d, *J* = 4.8 Hz, 3H, $H_{4'}$), 3.97 – 3.92 (mult., 6H, H_2, H_5), 3.73 – 3.65 (m, 18H, COOMe), 3.60 – 3.54 (m, 3H, H_3), 3.47 – 3.43 (mult., 11H, H_4, C-$\underline{CH_2}$-O), 3.26 – 3.20 (m, 3H, $H_{5'}$), 3.07 – 3.01 (m, 6H, $H_{2'}$, $H_{1'}$), 2.34 – 2.28 (mult., 6H, CH_2-val. ac.), 2.19 – 2.04 (m, 6H, $H_{3'}$), 2.01 – 1.86 (mult., 17H, $H_{6'}$, CH_2-val.ac), 1.63 (mult., 10H, CH_2-val.ac., CH_2-decane), 1.32 – 1.28 (m, 23H, CH_3-Rha, CH_2-decane), 0.92 – 0.88 (m, 3H, CH_3-decane); ^{13}C NMR (151 MHz, MeOD) δ 123.4 (CH) (visible in HSQC), 88.2 (C_1), 74.2 (C_2), 74.0 (C_4), 73.0 (C_3), 70.8 (C_5), 70.0 (C-$\underline{CH_2}$-O), 65.4 ($\underline{CH_2}$-triazole), 52.6 (COOMe), 52.5 (COOMe), 51.0 ($\underline{CH_2}$-triazole), 49.8 ($C_{4'}$), 46.7 ($C_{5'}$) 41.4 ($C_{1'}$), 41.0 ($C_{2'}$), 36.0 (CH_2-val.ac.), 33.1 (CH_2-decane), 30.8 (CH_2-decane), 30.8 (CH_2-decane), 30.6 (CH_2-decane), 30.4 (CH_2-decane), 30.2 ($C_{3'}$, $C_{6'}$), 23.8 (CH_2-val.ac.), 17.8 (CH_3-Rha), 13.2 (CH_3-decane) (visible in HSQC); MS (MALDI) calcd for $C_{90}H_{143}N_{15}O_{31}S_3$ $[M + Na]^+$ m/z: 2048.9200; found m/z: 2049.0829.

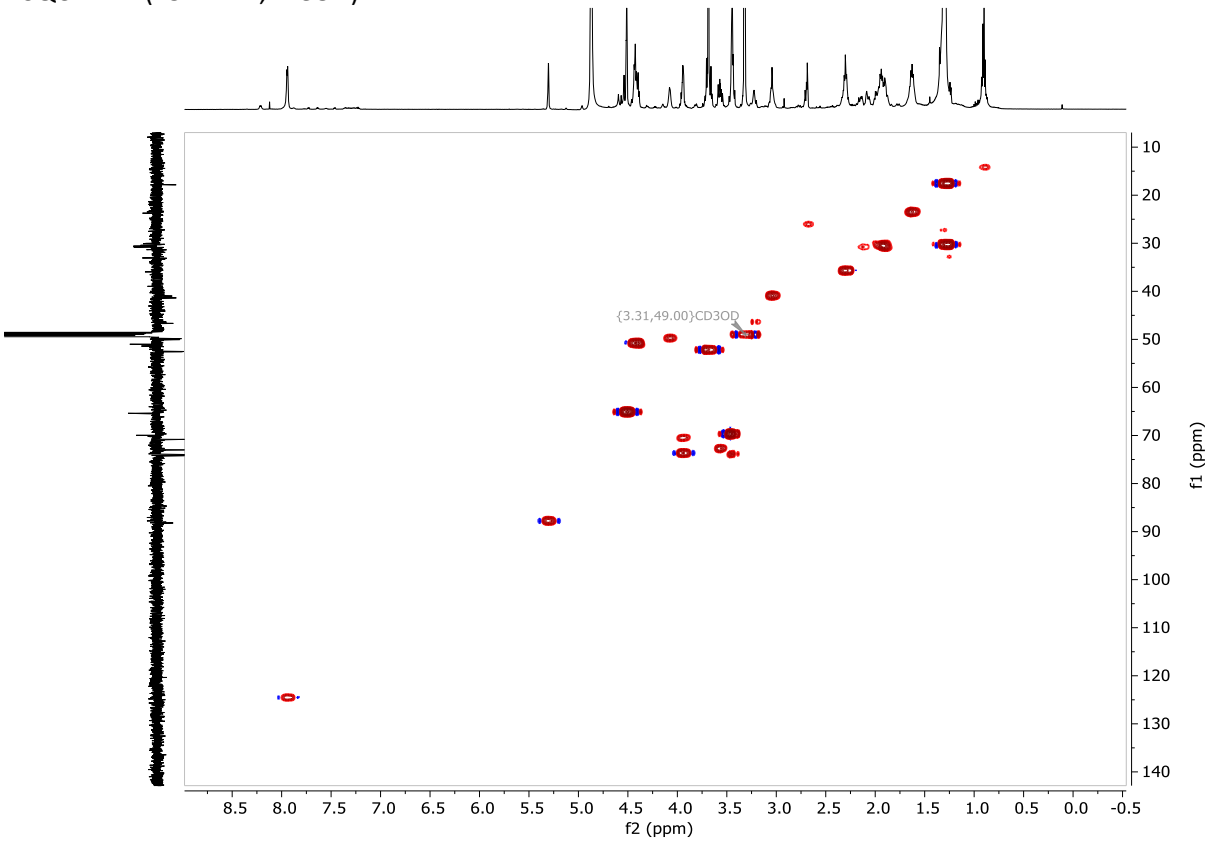
^{1}H NMR (600 MHz, MeOD)



^{13}C NMR (151 MHz, MeOD)



HSQC NMR (151 MHz, MeOD)



4.8. References

1. T. B. Geijtenbeek, R. Torensma, S. J. van Vliet, G. C. van Duijnhoven, G. J. Adema, Y. van Kooyk and C. G. Figdor, *Cell*, 2000, **100**, 575-585.
2. D. A. Mitchell, A. J. Fadden and K. Drickamer, *The Journal of biological chemistry*, 2001, **276**, 28939-28945.
3. R. C. Gallo, *Retrovirology*, 2006, **3**, 72.
4. E. W. Adams, D. M. Ratner, H. R. Bokesch, J. B. McMahon, B. R. O'Keefe and P. H. Seeberger, *Chemistry & biology*, 2004, **11**, 875-881.
5. E. van Liempt, C. M. C. Bank, P. Mehta, J. J. Garcı´a-Vallejo, Z. S. Kawar, R. Geyer, R. A. Alvarez, R. D. Cummings, Y. v. Kooyk and I. van Die, *FEBS Letters*, 2006, **580**, 6123-6131.
6. J. J. Reina, S. Sattin, D. Invernizzi, S. Mari, L. Martínez-Prats, G. Tabarani, F. Fieschi, R. Delgado, P. M. Nieto, J. Rojo and A. Bernardi, *ChemMedChem*, 2007, **2**, 1030-1036.
7. T. Belz and S. J. Williams, *Carbohydrate Research*, 2016, **429**, 38-47.
8. A. Tamburrini, S. Achilli, F. Vasile, S. Sattin, C. Vivès, C. Colombo, F. Fieschi and A. Bernardi, *Bioorganic & Medicinal Chemistry*, 2017, **25**, 5142-5147.
9. A. Tamburrini, PhD Thesis, Università degli studi di Milano, 2019.
10. K. Pachamuthu and R. R. Schmidt, *Chemical Reviews*, 2006, **106**, 160-187.
11. C.-S. Chao, M.-C. Chen, S.-C. Lin and K.-K. T. Mong, *Carbohydrate Research*, 2008, **343**, 957-964.
12. P. Shu, J. Zeng, J. Tao, Y. Zhao, G. Yao and Q. Wan, *Green Chemistry*, 2015, **17**, 2545-2551.
13. G. R. Morais, I. F. Oliveira, A. J. Humphrey and R. A. Falconer, *Carbohydrate Research*, 2010, **345**, 160-162.
14. H. Wenker, *Journal of the American Chemical Society*, 1935, **57**, 2328-2328.
15. J. B. Sweeney, *Chemical Society Reviews*, 2002, **31**, 247-258.
16. G. S. Singh, in *Advances in Heterocyclic Chemistry*, eds. E. F. V. Scriven and C. A. Ramsden, Academic Press, 2019, vol. 129, pp. 245-335.
17. J. L. Jat, M. P. Paudyal, H. Gao, Q.-L. Xu, M. Yousufuddin, D. Devarajan, D. H. Ess, L. Kürti and J. R. Falck, *Science (New York, N.Y.)*, 2014, **343**, 61-65.
18. Z. Ma, Z. Zhou and L. Kürti, *Angewandte Chemie International Edition*, 2017, **56**, 9886-9890.
19. S. Sabir, C. B. Pandey, A. K. Yadav, B. Tiwari and J. L. Jat, *The Journal of Organic Chemistry*, 2018, **83**, 12255-12260.
20. A. Mordini, F. Russo, M. Valacchi, L. Zani, A. Degl'Innocenti and G. Reginato, *Tetrahedron*, 2002, **58**, 7153-7163.

21. A. Bernardi, D. Arosio, L. Manzoni, F. Micheli, A. Pasquarello and P. Seneci, *The Journal of Organic Chemistry*, 2001, **66**, 6209-6216.
22. J. Jiang, M. Artola, T. J. M. Beenakker, S. P. Schröder, R. Petracca, C. de Boer, J. M. F. G. Aerts, G. A. van der Marel, J. D. C. Codée and H. S. Overkleeft, *European Journal of Organic Chemistry*, 2016, **2016**, 3671-3678.
23. P. Serrano, A. Llebaria, J. Vázquez, J. de Pablo, J. M. Anglada and A. Delgado, *Chemistry – A European Journal*, 2005, **11**, 4465-4472.
24. J. A. Schwarcz and A. S. Perlin, *Canadian Journal of Chemistry*, 1972, **50**, 3667-3676.
25. S. Oscarson, in *Glycoscience: Chemistry and Chemical Biology I–III*, eds. B. O. Fraser-Reid, K. Tatsuta and J. Thiem, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, DOI: 10.1007/978-3-642-56874-9_20, pp. 643-671.
26. H. Hashimoto, K. Shimada and S. Horito, *Tetrahedron: Asymmetry*, 1994, **5**, 2351-2366.
27. S. Josse, J. Le Gal, M. Pipelier, J. Cléophax, A. Olesker, J.-P. Pradère and D. Dubreuil, *Tetrahedron Letters*, 2002, **43**, 237-239.
28. U. Dabrowski, H. Friebolin, R. Brossmer and M. Supp, *Tetrahedron Letters*, 1979, **20**, 4637-4640.
29. H. Hori, T. Nakajima, Y. Nishida, H. Ohrui and H. Meguro, *Tetrahedron Letters*, 1988, **29**, 6317-6320.
30. N. Hribernik, A. Tamburrini, E. Falletta and A. Bernardi, *Organic & Biomolecular Chemistry*, 2021, **19**, 233-247.
31. O. Oyelaran, L. M. McShane, L. Dodd and J. C. Gildersleeve, *Journal of Proteome Research*, 2009, **8**, 4301-4310.
32. R. T. C. Sheridan, J. Hudon, J. A. Hank, P. M. Sondel and L. L. Kiessling, *Chembiochem*, 2014, **15**, 1393-1398.
33. M. K. Hossain, A. Vartak, P. Karmakar, S. J. Sucheck and K. A. Wall, *ACS Chemical Biology*, 2018, **13**, 2130-2142.
34. B. Liet, E. Laigre, D. Goyard, B. Todaro, C. Tiertant, D. Boturyn, N. Berthet and O. Renaudet, *Chemistry – A European Journal*, 2019, **25**, 15508-15515.
35. J. Alsarraf, P. Gormand, J. Legault, M. Mihoub and A. Pichette, *Tetrahedron Letters*, 2020, **61**, 151706.
36. S. Buchini, F.-X. Gallat, I. R. Greig, J.-H. Kim, S. Wakatsuki, L. M. G. Chavas and S. G. Withers, *Angewandte Chemie International Edition*, 2014, **53**, 3382-3386.
37. H. P. R. Mangunuru, J. R. Yerabolu and G. Wang, *Tetrahedron Letters*, 2015, **56**, 3361-3364.

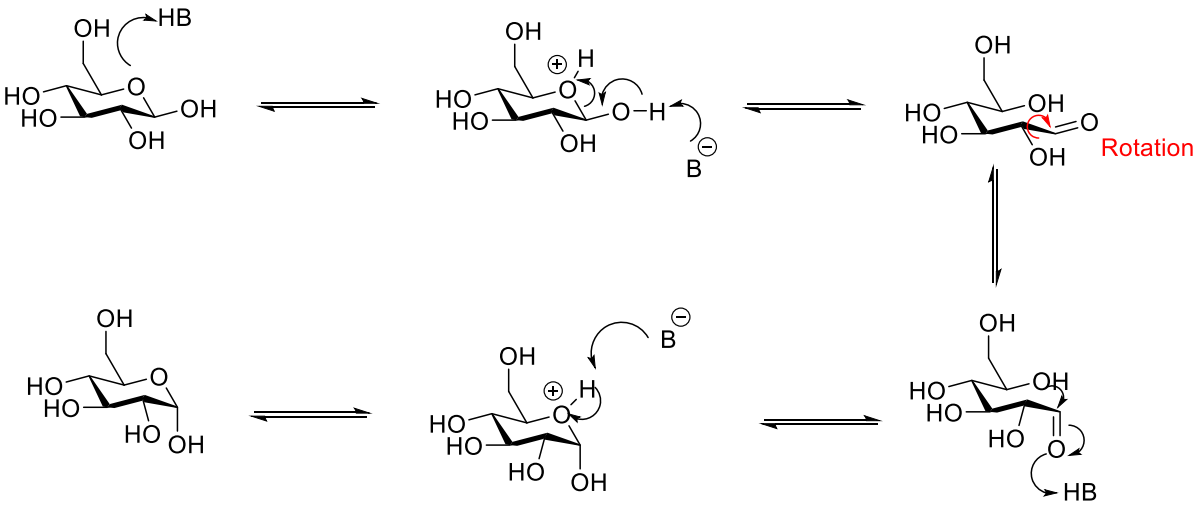
38. G. Yang, R. W. Franck, R. Bittman, P. Samadder and G. Arthur, *Organic Letters*, 2001, **3**, 197-200.
39. S. Mari, H. Posteri, G. Marcou, D. Potenza, F. Micheli, F. J. Cañada, J. Jimenez-Barbero and A. Bernardi, *European Journal of Organic Chemistry*, 2004, **2004**, 5119-5225.
40. N. Floyd, B. Vijayakrishnan, J. R. Koeppe and B. G. Davis, *Angewandte Chemie International Edition*, 2009, **48**, 7798-7802.
41. M. Purwin, J. Hernández-Toribio, C. Coderch, R. Panchuk, N. Skorokhyd, K. Filipiak, B. de Pascual-Teresa and A. Ramos, *RSC Advances*, 2016, **6**, 66595-66608.
42. L. Díaz, J. Bujons, J. Casas, A. Llebaria and A. Delgado, *Journal of Medicinal Chemistry*, 2010, **53**, 5248-5255.
43. E. Ladd, A. Sheikhi, N. Li, T. G. M. Van de Ven and A. Kakkar, *Molecules*, 2017, **22**, 868.

CHAPTER FIVE

Anomeric isomerization

5.1. Mutarotation of reducing saccharides

Augustin-Pierre Dubrunfaut first reported the mutarotation of glucose in 1846. The word mutarotation means "change of rotation" and corresponds to change in specific rotation of carbohydrate derivatives. This phenomenon has been most extensively explored on monosaccharides. Upon dissolution of hexoses in water the hemiacetal ring opens and reforms to give products with different ring sizes and configuration at the anomeric center, such as α- and β-pyranoses as well as α- and β-furanoses (**Scheme 5.1**). It is well known that this process occurs with all reducing saccharides in solutions and can be acid or base catalysed.



Scheme 5.1 Mechanism of acid-catalyzed mutarotation of glucose

The final equilibrium in the mixture of the two anomers is likely determined by a combination of steric hindrance, stereoelectronic and solvent effects.[1] **Table 5.1** shows percentage compositions of sugars in aqueous solution at equilibrium, as determined by NMR and polarimetry.

Table 5.1 Equilibrium compositions of free O-glycosides in aqueous solution[1]

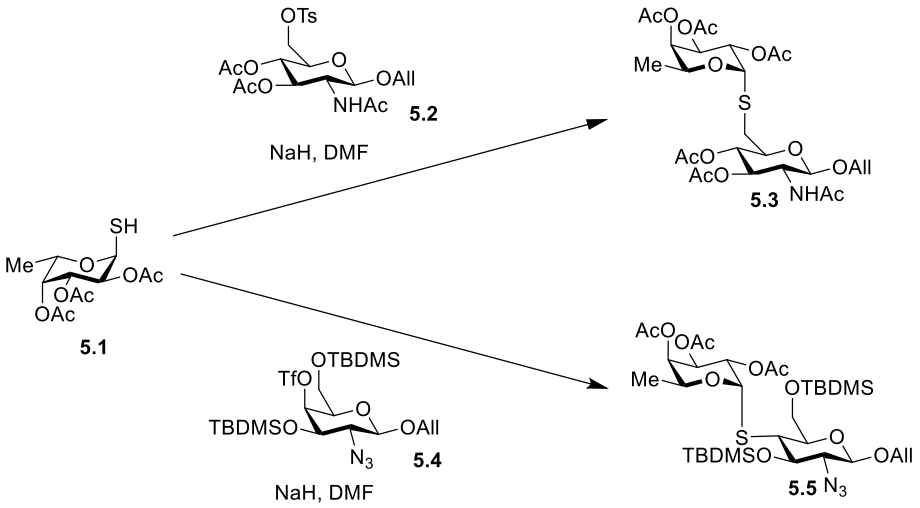
Sugar	T (°C)	α-Pyranose	β-Pyranose	α-Furanose	β-Furanose	Acyclic
Glc	31	38.0	62.0	0.5	0.5	0.002
Man	44	65.5	34.5	0.6	0.3	0.005
Gal	31	30.0	64.0	2.5	3.5	0.02
Rha	44	65.5	34.5	0.6	0.3	0.005

5.2. Anomerization of glycosyl thiols

Unlike reducing saccharides, glycosyl thiols are generally observed to be configurationally stable. This actually presents one of the great advantages for their use in synthesis. Their configurational stability could be partially explained by the poor orbital overlap between the anomeric carbon and the sulphur atom which does not promote the ring opening and subsequent mutarotation.[2, 3]

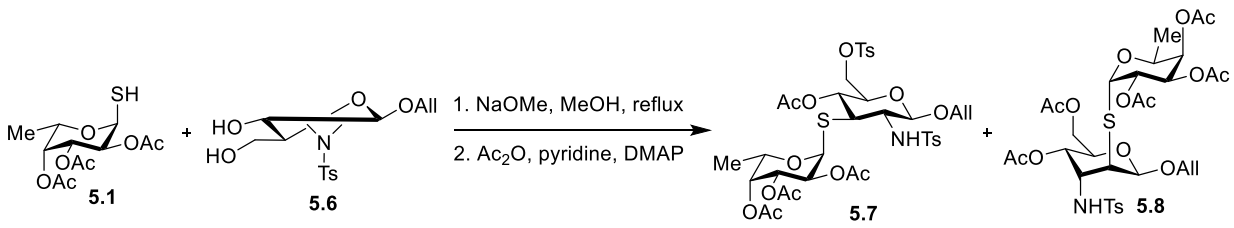
Some literature reports state that glycosyl thiols, once formed, often retain the anomeric configuration in subsequent reactions also under basic conditions, because the anomerization of the glycosyl thiolate is very slow.[2, 4, 5] Sources often cite the work of Hashimoto et al. from 1994 that reports the synthesis of α-L-fucopyranosyl disaccharides with thioglycosidic linkages through S_N2 reactions, including ring-opening reactions of aziridines and epoxides by a glycosyl thiol. They either employed the free thiol 1-*thio*-α-L-fucopyranose tetraacetate **5.1** or generated the thiol *in situ* from 1-thioacetly-α-L-fucopyranose tetraacetate **5.9**. Analogs of α-L-fucopyranosyl disaccharides with different linkage positions α(1,2), α(1,3), α(1,4) and α(1,6) were synthesized using three different approaches.

First, disaccharides with linkages α(1,6) (compound **5.3**) and α(1,4) (compound **5.5**) were synthesized by substituition of a tosylate or a triflate by a thiolate nucleophile formed *in situ* by treatment of glycosyl thiol **5.1** with NaH in DMF. (**Scheme 5.2**)



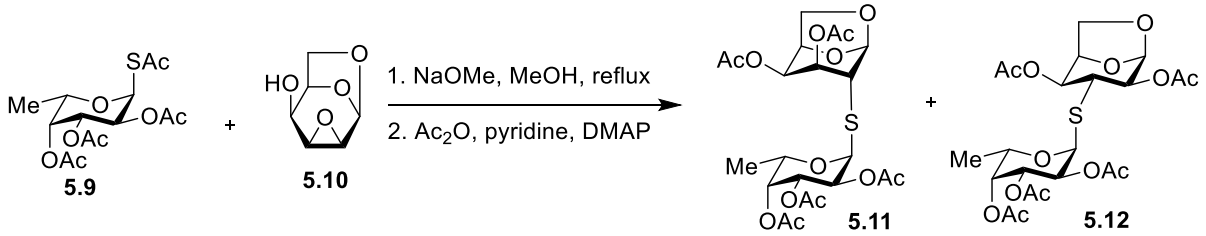
Scheme 5.2 Synthetic approach towards α-L-fucopyranosyl disaccharides with α(1,6) and α(1,4) linkages developed by Hashimoto et al.[4]

Second, the analog **5.7** with α(1,3) linkage was obtained from a ring opening reaction of a tosyl-functionalized aziridine ring **5.6** by a glycosyl thiol **5.1** in the presence of NaOMe in MeOH under reflux. The reaction gave a mixture of the desired disaccharide **5.7** and its D-altro isomer **5.8**. (**Scheme 5.3**)



Scheme 5.3 Synthetic approach towards α-L-fucopyranosyl disaccharide with α(1,3) linkage developed by Hashimoto et al.[4]

Finally, synthesis of the α(1,2) linked analog **5.11** was achieved by opening of an epoxide **5.10** using α-fucosyl thioacetate **5.9** as a highly concentrated solution in MeOH in the presence of NaOMe under reflux (**Scheme 5.4**). In this latter case, the authors noticed that reducing the concentration of the nucleophile **5.9** from 2.6 M to 0.9 M resulted in longer reaction time and in anomerization of 1-*thio*-L-fucose. This led the authors to the conclusion that the anomeric configuration of nucleophiles was mostly maintained during the couplings, because the anomerization of the glycosylthio anion is very slow.



Scheme 5.4 Synthetic approach towards α-L-fucopyranosyl disaccharide with α(1,2) linkage developed by Hashimoto et al.[4]

5.2.1. Anomerization of unprotected glycosyl thiols

Efforts towards clarifying whether unprotected glycosyl thiols undergo mutarotation were made in 2010 by Caraballo et al.[6] They conducted a study to investigate pH-dependent mutarotation of 1-thioaldoses and their equilibrium anomeric ratios in aqueous media. The mutarotation behaviour of 1-*thio*-D-mannopyranose **5.15**, 1-*thio*-D-galactopyranose **5.14**, 1-*thio*-D-glucopyranose **5.13** and 1-*thio*-L-fucopyranose **5.16** under acidic, neutral and basic pH was monitored by ^{1}H NMR spectroscopy (**Table 5.2**). The expected favoured configurational isomers in aqueous media were used (α- for manno and β- for gluco, galacto) except for the fucopyranose that was initially used in the α-form as shown in **Fig. 5.1**.

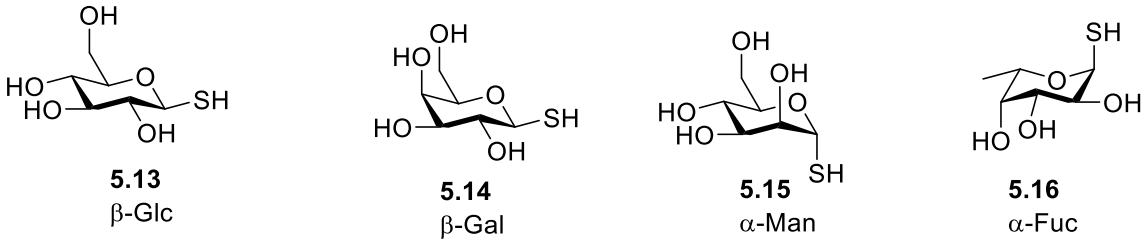


Fig. 5.1 Starting configurations of the glycosyl thiols used in the study by Caraballo et al.[6]

The results displayed in **Table 5.2** show that at RT under acidic and neutral pH, equilibria were reached within 15 h, all thiols were mainly in the β-configuration. Particularly surprising was the behaviour of the Man thiol **5.15** that was also found in the predominantly β-configuration, although its natural analog *O*-mannopyranose has a clear preference for α-configuration (**Table 5.1**). At the acidic pH approximately 78% of β-anomer for the galactopyranose **5.14**, 74% of β-anomer for fucopyranose **5.16** and glucopyranose **5.13** and 66% of β-anomer of mannopyranose **5.15** were found in the equilibria.

Table 5.2 Equilibrium (RT, aqueous media, ~80 h) anomeric composition of 1-thioglycopyranoses in relation to pD. (Caraballo et al.)[6]

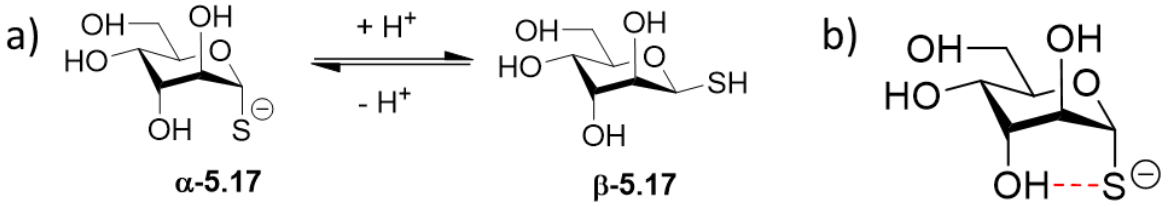
Starting configuration	**1-β-thioaldose (%)**		
	pD 4	pD 7	pD 9
α-Man 5.15	66.2	60.9	24
α-Fuc 5.16	74.3	79.1	<5
β-Gal 5.14	78.1	77.7	>95
β-Glc 5.13	74.1	75.2	93.6

On the other hand, at basic pH mutarotation of unprotected glycosyl thiols proceeded at a much slower rate. For 1-*thio*-D-mannopyranose **5.15** under acidic and neutral conditions half-times of 3.1 h and 3.3 h, respectively, were recorded. Under basic conditions half-time of 14.8 h was recorded. For all other glycosyl thiols the mutarotation under basic conditions was almost completely blocked. Particularly in case of fucose and galactose species **5.16** and **5.14** only very low anomerization from the initial configurations was observed over the time range (approx. 80 h) (**Table 5.2**).

To further validate the influence of pH on the isomerization rate, both β- and α-forms of the fuco-derivative **5.16** were tested under acidic and basic pH. Under acidic pH the anomeric equilibrium compositions (α/β 20/80) were similar regardless to which starting anomeric form was used. In contrast, at basic pH the initial anomeric configuration was preserved for both β- and α-anomer.

This observation can spark a conclusion that, at least in water, the mutarotation of glycosyl thiols occurs at acidic pH and to a much lesser extent at basic pH.

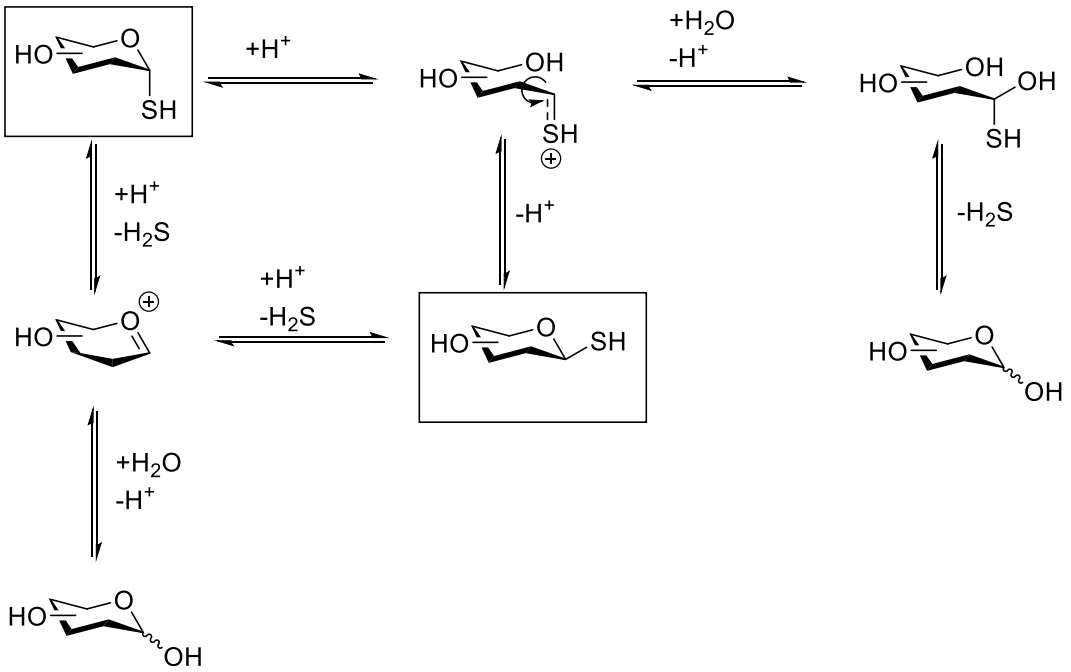
Under acidic or neutral pH, the β-anomer was observed as the major mutarotation product for all 1-thioaldoses, independently of which initial form was used. As mentioned this was particularly surprising for Man thiol **5.15**, because the distribution of its corresponding natural hydroxyl-derivative that is clearly shifted towards the α-form (**Table 5.1**). Based on this, the authors predicted that the strongest influence on the anomeric composition at the equilibrium comes from the orientation of the C_2 hydroxyl group. This was additionally investigated by monitoring the anomeric composition of 1-*thio*-α-D-altropyranose **5.17**, a sugar with an axial C_2 hydroxyl group whose natural derivative has a β-reference at the equilibrium, opposite to mannose. Under acidic and neutral conditions, the altrose **α-5.17** rapidly (<10 min) and entirely converted to β-form and under basic conditions it remained in the α-form.



Scheme 5.5 a) Mutarotation of 1-*thio*-D-altopyranose, b) Internal H-bond between C_3-OH and C_1-thiolate

To rationalize the results observed by NMR studies, calculations were also performed. The results obtained by calculation predictions for altrose **5.17** were in good agreement with the experimental results. For ionic form (thiolate) α-configuration was dominant and for neutral form (thiol) there was clear preference for β-configuration. The α-configuration of **α-5.17** is stabilized under basic pH by an internal H-bond formed between hydroxyl group on C_3 and the negatively charged thiolate group. (**Scheme 5.5**) Such internal H-bond cannot exist in the β-configuration. The internal H-bond found cannot be formed in any other 1-*thio*-aldopyranose, explaining why 1-*thio*-D-altrose has the most strongly favoured α-anomer. Calculations for 1-*thio*-D-mannose **5.15** were not in agreement with the experimental results, most likely due to specific solvent interactions that could not be well reproduced by the implicit solvent model.

The authors proposed that the mechanism of the isomerization could follow two pathways, endocyclic or exocyclic cleavage (**Scheme 5.6**). The first suggests formation of a thiocarbenium ion intermediate through an endo-ring opening mechanism similar to mutarotation. Alternatively, a pathway could lead through the formation of an oxocarbenium ion. Neither of the mechanisms was confirmed or discarded. A trace amount of thiofuranoses was observed during experiments which could support a ring opening to occur. On the other hand, slow formation of aldoses and a characteristic smell of H_2S indicate that a process different from mutarotation occured.

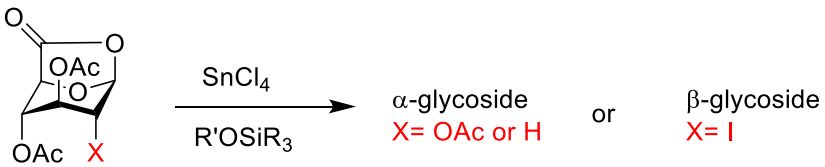


Scheme 5.6 Proposed mechanism for mutarotation of glycosyl thiols by Caraballo et al.[6]

5.2.2. Anomerization of glycosyl thiols with Lewis acids

Contributing to the general belief that the glycosyl thiols are configurationally stable, anomerization during 1-thioglycosylation processes is very rarely reported in the literature. However, there are few examples of anomerization reports, particularly in the presence of Lewis acids.

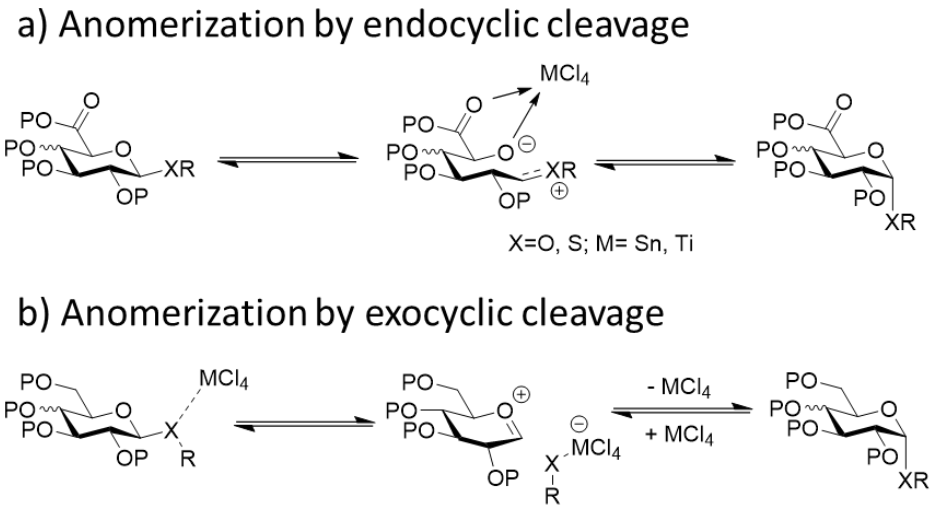
The 1,2-*trans* glycoside is often isolated in reactions with 2-acyl containing donors in the presence of $TiCl_4$ or $SnCl_4$, which can be explained by 2-*O*-acyl group participation.[7, 8] However, occasionally a 1,2-*cis* product or a mixture of two anomers is isolated (**Scheme 5.7**).[9, 10] In such cases the anomerization most likely occurs after the glycosylation reaction, meaning that first the 1,2-*trans* product is formed and it then anomerizes to 1,2-*cis* product.



Scheme 5.7 Conformationally inverted donors derived from glucuronic acid yield 1,2-*trans* or 1,2-*cis* products in $SnCl_4$ catalysed couplings with silyl ethers[9]

The anomerization mechanism can in principle involve either endocyclic or exocyclic cleavage (**Scheme 5.8**). Endocylic cleavage occurs through breaking of the bond between the anomeric carbon atom and the pyranose-ring oxygen. Exocyclic cleavage on the other hand is achieved through cleavage of the

bond between the anomeric carbon atom and the exocyclic oxygen.[11] While ring opening mechanism seems to be a more likely option for anomerization mechanism, exocyclic cleavage cannot be completely ruled out.



Scheme 5.8 Anomerization mechanism through a) endocyclic cleavage or b) exocyclic cleavage[12]

In 2009 Manabe et al. achieved anomerization of *thio*-pyranosides carrying 2,3-*trans*-cabonate and –carbamate groups (**Fig. 5.2**, compounds **5.18**, **5.19**, **5.20**) with the use of Lewis acid BF_3xEt_2O.[11] Additionally, they proved that the anomerization is achieved through endocyclic cleavage mechanism in multiple ways, including reduction of the generated cation as shown in **Scheme 5.9**, summarizing the principle of the anomerization and reduction process in the presence of BF_3xEt_2O and Et_3SiH. The endocyclic cleavage could only be obtained with pyranose rings locked in chair conformation. 2,3-*cis*-carbonate-bearing pyranosides (mannosides and rhamnosides) such as **5.21** had distorted pyranosides and were not anomerized under these conditions, thus did not undergo the endocyclic cleavage.

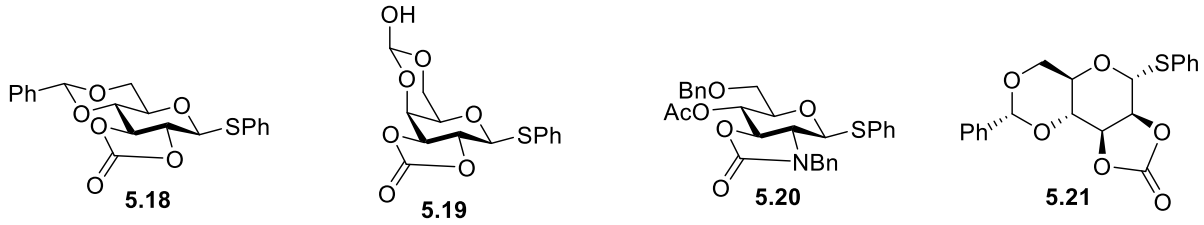
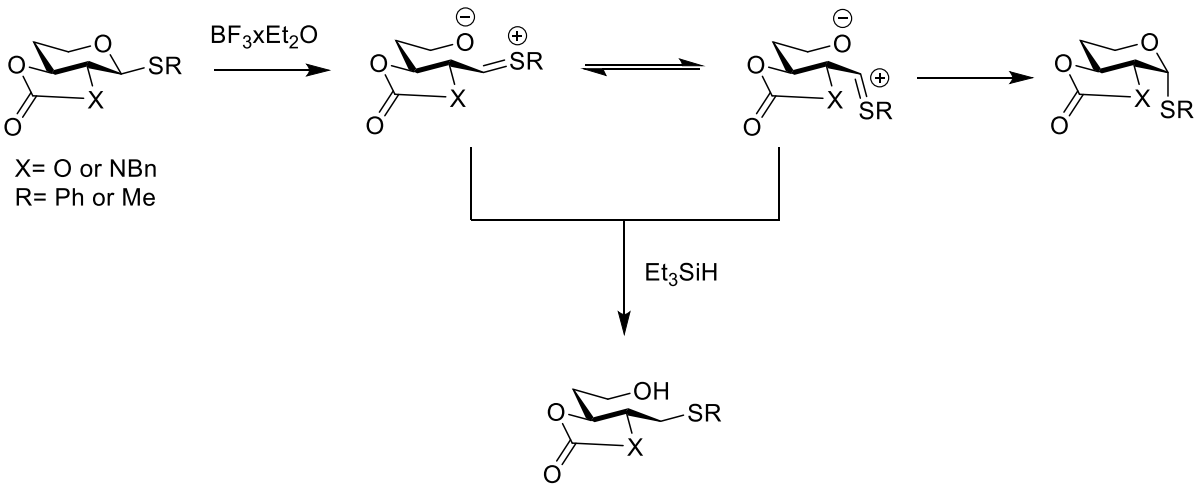


Fig. 5.2 Thio-pyranoses used in the study by Manabe et al.[11]



Scheme 5.9 Anomerization and reduction of thiocarbenium cation by Manabe et al.[11]

Murphy's group investigated the Lewis acid promoted anomerization of *S*-butyl glycosides derived from glucuronic acid (**Fig. 5.3**, compounds **5.22** and **5.23**) as well as from glucose (**5.24**, **5.25**) and galactose (**5.26**, **5.27**).[12]

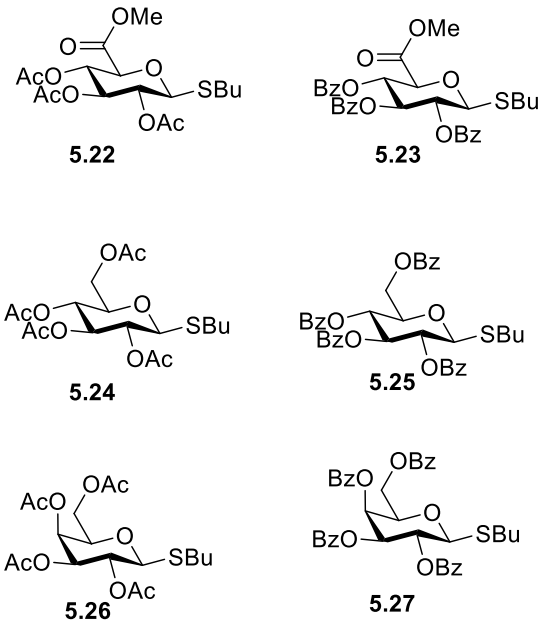


Fig 5.3 *S*-Butyl glycoside substrates used in the study by Murphy et al.[12]

The equilibrium ratio of anomers is based on the complex formed between the saccharide and Lewis acid and not the free glycoside itself. It was established that the rates of anomerization were significantly faster for uronic acid derivatives **5.22** and **5.23** than glucose (**5.24**, **5.25**) and galactose derivatives (**5.26**, **5.27**), probably due to coordination of C_6 carbonyl group and C_1 sulphur to the Lewis acid, which enhances the anomeric effect by increasing the electron-withdrawing ability of the anomeric substituent and leads to an increase in proportion of the α-anomer.

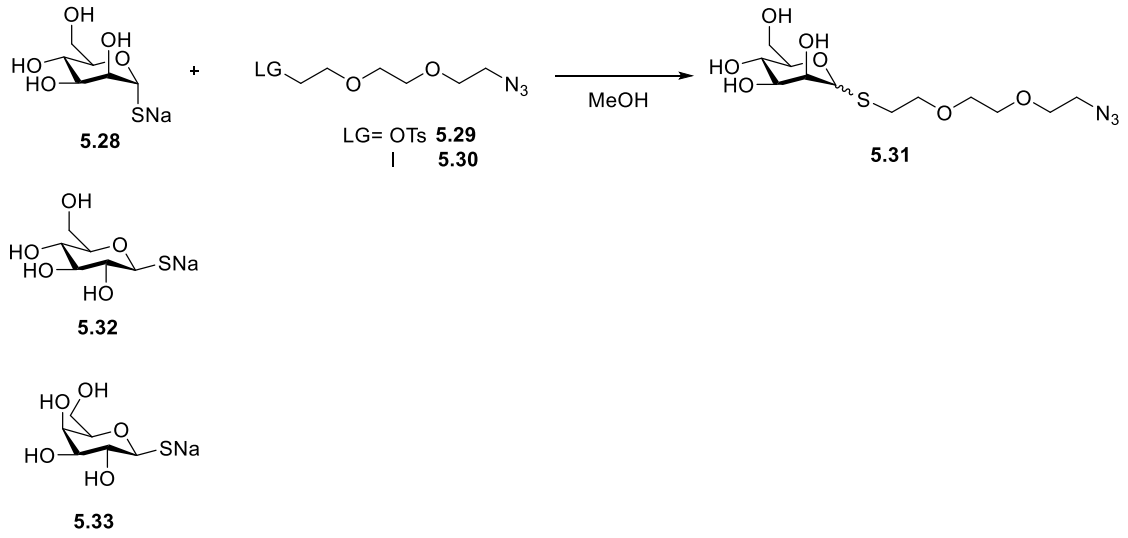
Later Murphy's group expanded their studies also on the anomerization of glycosyl thiols with Lewis acids, for example benzoyl and acetyl protected *thio*-derivatives of glucose, galactose, rhamnose, fucose, lactose, etc.[13] The equatorial to axial (1,2-*trans* to 1,2-*cis*) epimerization of benzoyl protected glycosyl thiols was achieved with $TiCl_4$, while $SnCl_4$ promoted axial to equatorial (1,2-*cis* to 1,2-*trans*) epimerization of both benzoyl and acetyl protected glycosyl thiols.

In summary, their studies showed that anomeric ratios in the equilibrium depend on the saccharide residue, which Lewis acid ($SnCl_4$ or $TiCl_4$) is used and in what quantity, temperature, protecting groups on the sugar and electron withdrawing potential of aglycone.

5.2.3. Anomerization of glycosyl thiols during 1-*S*-glycosylations

While most literature report configurational stability of glycosyl thiols and thiolates used in S_N2 type reactions[14, 15], rare examples exist that observe anomeric isomerization during *S*-alkylation reaction of glycosyl thiols.

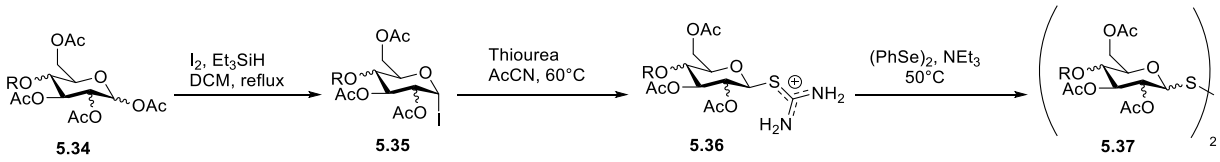
One of such reports comes from Deng et al. using glycosyl thiols to synthesize *S*-linked photoactivatable PFPA-conjugated mono- and disaccharides.[3] They observed a mixture of α- and β-isomers of the product **5.31** formed in the reaction of unprotected α-mannose thiolate **5.28** with tosylate **5.29** in MeOH at room temperature (**Scheme 5.10**). Similar results were obtained also in the reaction of β-glucose thiolate **5.32** and β-galactose thiolate **5.33** in the reaction with tosylate in MeOH/THF at RT.



Scheme 5.10 Anomerization of α-mannosyl thiol **5.28** upon deprotonation and S_N2 reaction by Deng et al.[3]

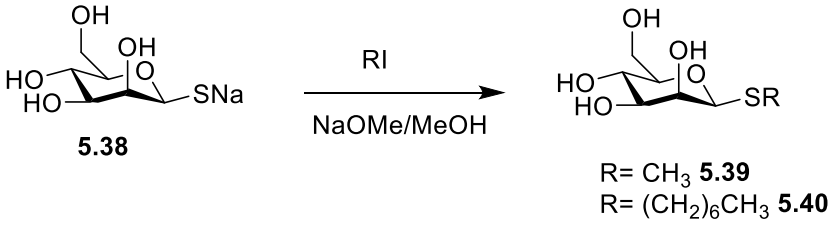
The authors thoroughly investigated the effects of solvent, electrophile, temperature and time on the anomerization to gain control over the stereochemistry in the synthesis. They observed solvent effects, in particular polar solvents (DMSO, DMF) led to more selectivity. Additionally, increasing the rate of S_N2 reaction by employing iodide **5.30** rather than tosylate **5.29** was also beneficial for achieving higher β/α ratio.

Another example of anomerization of glycosyl thiols is a study by Adinolfi et al. reporting a synthetic approach towards formation of symmetrical glycosyl disulfides.[16] In their synthetic route peracetylated glycosides **5.34** were converted to anomeric iodides **5.35** and subsequently to glycosyl thioureas **5.36** (**Scheme 5.11**). The thioureas **5.36** upon treatment with Et_3N gave glycosyl thiolates that were, with the addition of phenyl diselenide, rapidly converted to glycosyl disulphides **5.37**. *Thio*-derivatives of glucose, galactose, mannose, fucose and lactose were used in the study and products were formed as mixtures of isomers in all cases. The stereoselectivity of the process was particularly poor in the case of the manno precursor.



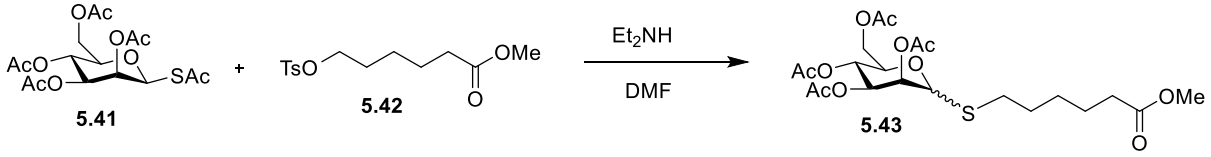
Scheme 5.11 Synthetic approach towards symmetrical glycosyl disulphides by Adinolfi et al.[16]

A particularly interesting example, closely related to our work with glycosyl thioacetates in aziridine opening reactions (Chapter Four) is the work of Bundle et al.[17] They initially investigated the anomerization rate of unprotected β-mannosyl thiolate **5.38** in CD_3ONa/CD_3OD by NMR. They observed that anomerization was very slow, after 3 days almost 90% of the β-anomer was still present. This slow anomerization rate allowed them to synthesize methyl and heptyl glycosides **5.39** and **5.40** as pure β-isomers by addition of the corresponding iodide to the thiolate **5.38** in the presence of NaOMe/MeOH (**Scheme 5.12**).



Scheme 5.12 Synthesis of methyl aand heptyl glycosides **5.39** and **5.40** from unprotected β-mannosyl thiolate **5.38** by Bundle et al.[17]

The authors then used peracetylated β-mannosyl thioacetate **5.41** to generate the thiolate *in situ* with Et_2NH in DMF in reaction with tosylate **5.42** to synthesize **5.43**. The reaction performed at RT gave product **5.43** as an anomeric mixture β:α 1:1 (**Scheme 5.13**). The ratio was improved at lower temperatures, the best result was obtained when the reaction was performed at -55°C for 48 h, the α-anomer was then obtained in 21% yield. This suggests that stabilization of the negative charge on S by H(D) bond can stabilize the thiolate.

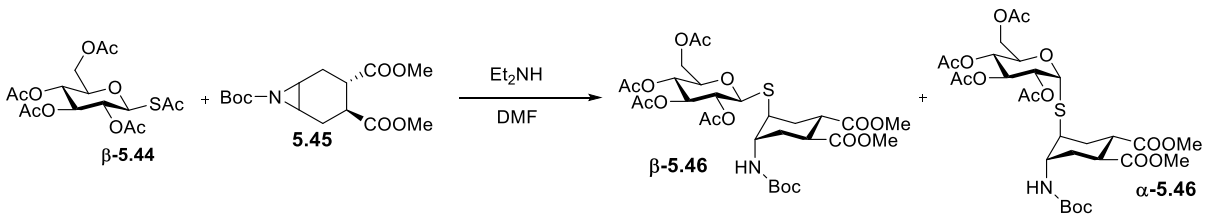


Scheme 5.13 Synthesis of **5.43** from tosylate **5.42** and a thiolate generated *in situ* from **5.41** by Bundle et al.[17]

5.3. Mechanistic aspects of anomerization during the one-pot aziridine opening reaction

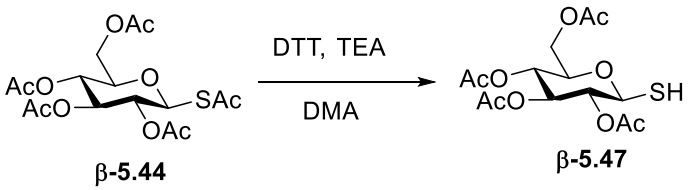
The described overview of the knowledge on anomerization of sugar derivatives with the anomeric sulphur gathered by different research groups shows that this remains a relatively unclear area. Naturally, we were intrigued to further explore the background of the anomerization that was observed under the conditions of the aziridine opening reaction as explained in Chapter Four.

The anomeric isomerization was particularly evident when the peracetylated β-glucosyl thioacetate **β-5.44** was used in the aziridine opening reaction with the aziridine **5.45** (**Scheme 5.14**). The reaction was performed under initial conditions with thioacetate **β-5.44** (1 eq) and aziridine **5.35** (0.77 eq) at RT with 1.5 eq Et_2NH in DMF [0.8 M] for 4 h. Under these conditions the anomeric ratio of the products **β-5.46**/**α-5.46** was approximately 2/1. Therefore, we selected the glucose derivative as a model to further investigate the anomerization.



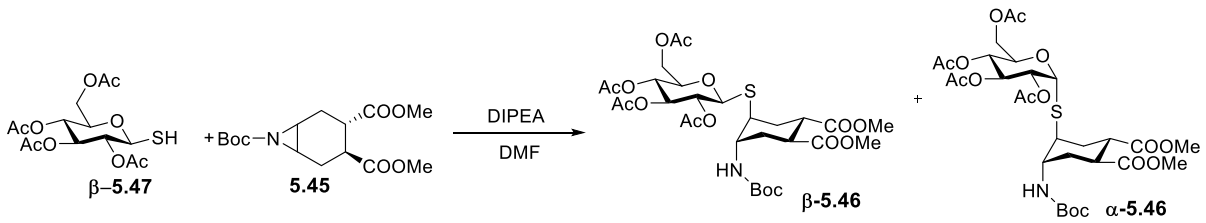
Scheme 5.14 One-pot opening reaction of aziridine **5.45** with glucosyl thioacetate **β-5.44** to give product **5.46**

Initially we speculated, that the anomerization under these specific conditions might be related to the formation of the thiol/thiolate *in situ* with Et_2NH. Therefore, we set to establish whether the isolated free thiol would be able to retain the anomeric configuration in the aziridine opening reaction. Our approach was to split the one-pot procedure in two consecutive reactions, formation of the thiol and subsequent aziridine opening. Thus, **β-5.44** was selectively deacetylated at the anomeric position as previously described by Wan et al.[18] using DTT and Et_3N (**Scheme 5.15**). The isolated thiol **β-5.47** was obtained in β-configuration as confirmed by ^1H NMR analysis of coupling constants of the anomeric proton at 4.54 ppm ($J_{1\text{-}2} = J_{1\text{-}SH} = 9.6$ Hz).



Scheme 5.15 Deacetylation of **β-5.44** to yield thiol **β-5.47**

The free thiol **β-5.47** was then used in the reaction with the aziridine **5.45**. The reaction was performed with the thiol **β-5.47** (1 eq) and aziridine **5.45** (0.77 eq) at RT with 0.23 eq. DIPEA in DMF [0.4 M], overnight (**Scheme 5.16**). When using the free thiol as opposed to the thioacetate **β-5.44**, a nucleophilic base (Et_2NH) is no longer needed and therefore a sub-stoichiometric amount (0.23 mol eq.) of the bulky non-nucleophilic base iPr_2NEt (DIPEA) was used instead.



Scheme 5.16 Aziridine opening reaction with thiol **β-5.47**

Starting from β-configuration of the thiol **β-5.47**, a β-aziridine opening product **β-5.46** was expected. Yet, both the **β-5.46** and **α-5.46** products were still formed in 2.5 : 1 ratio as judged from crude ^1H NMR (**β-5.46** H_1: 4.75 ppm, **α-5.46** H_1: 5.73 ppm) (**Fig. 5.4**, Spectra A). An aliquot was taken from the reaction after 2.5 h and the characteristic peaks of the anomeric protons of **β-5.46** and **α-5.46** were already clearly visible. When the reaction was performed in the absence of base with the thiol **β-5.47** (1 eq) and aziridine **5.45** (0.77 eq) (reaction at RT in DMF [0.4 M], overnight), no product was formed and the β-configuration of the starting thiol **β-5.47** was preserved (**Fig. 5.4** Spectra B). These experiments

indicate that formation of the thiolate is required for both the aziridine opening reaction and the isomerization to occur.

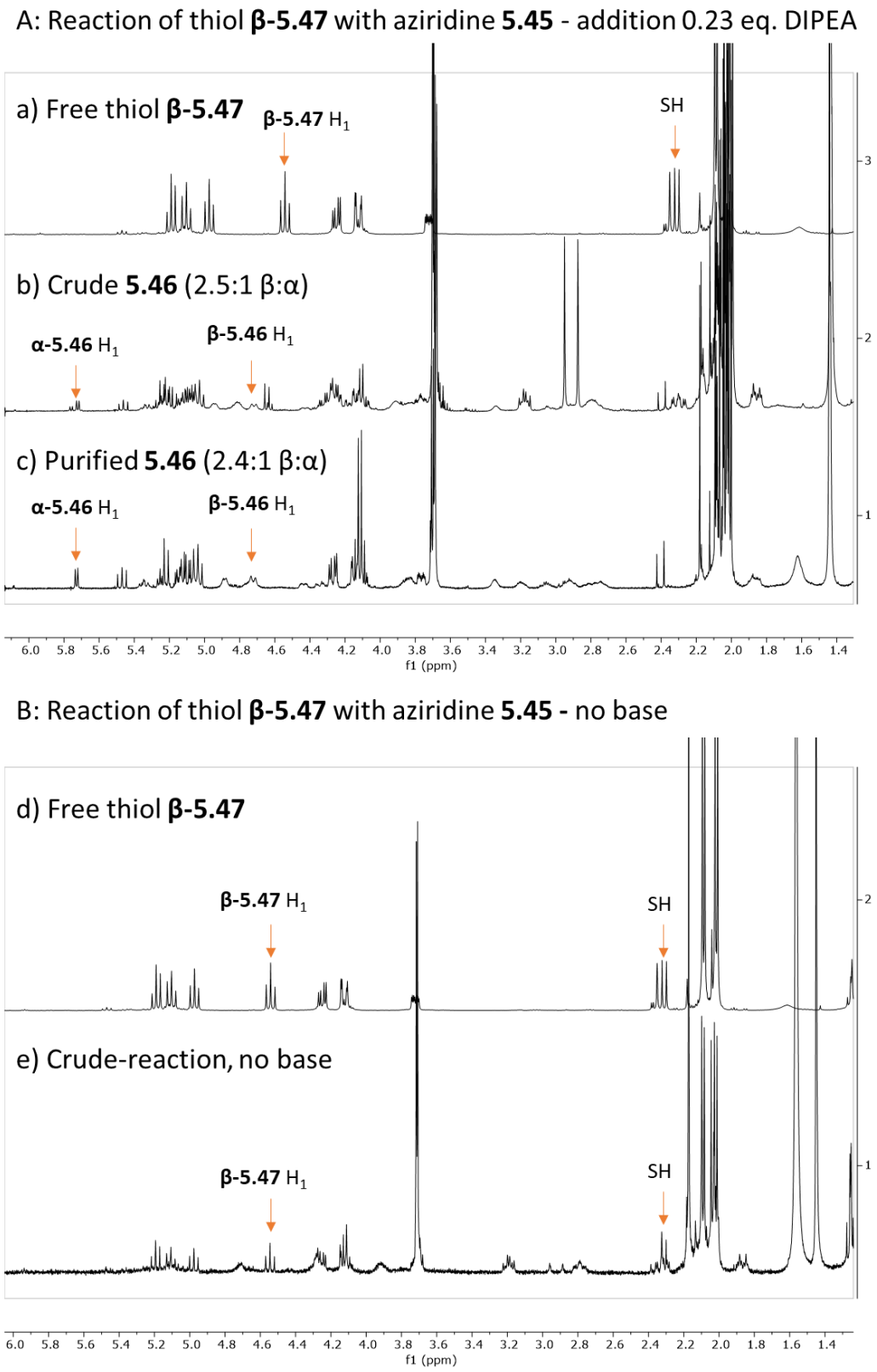
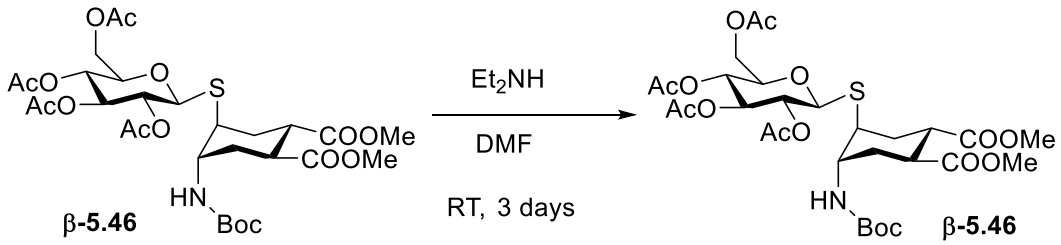


Fig. 5.4 Reaction of thiol **β-5.47** with aziridine **5.45** A: in the presence of base (0.23 eq DIPEA), B: with no base (400 MHz ^{1}H-NMR spectra in $CDCl_3$)

To define the underlying cause for the anomerization, we essentially drew 3 possible hypotheses:

1. Product spontaneously isomerizes under the reaction conditions.
2. Anomerization is influenced by the aziridine substrate.
3. Glycosyl thiols (upon formation of the thiolate) isomerize under basic conditions.

To investigate the first hypothesis about isomerization of the product, we conducted an experiment in which we exposed the pure **β-5.46** product to the reaction conditions (1.5 eq Et_2NH, DMF [0.8 M], RT) for 3 days (**Scheme 5.17**). We analyzed the ^1H NMR spectra of the product **5.46** in $CDCl_3$ before and after the treatment with Et_2NH and did not observe formation of new peaks that could indicate anomerization, particularly the characteristic peak at 5.74 ppm that corresponds to the anomeric proton of **α-5.46** isomer. (**Fig. 5.5**) Furthermore, if the aziridine opening reactions were left to stir overnight, the final anomeric ratio of the products was not different than what was observed after shorter reaction times (1-4 h).



Scheme 5.17 Treatment of the product **β-5.46** with Et_2NH

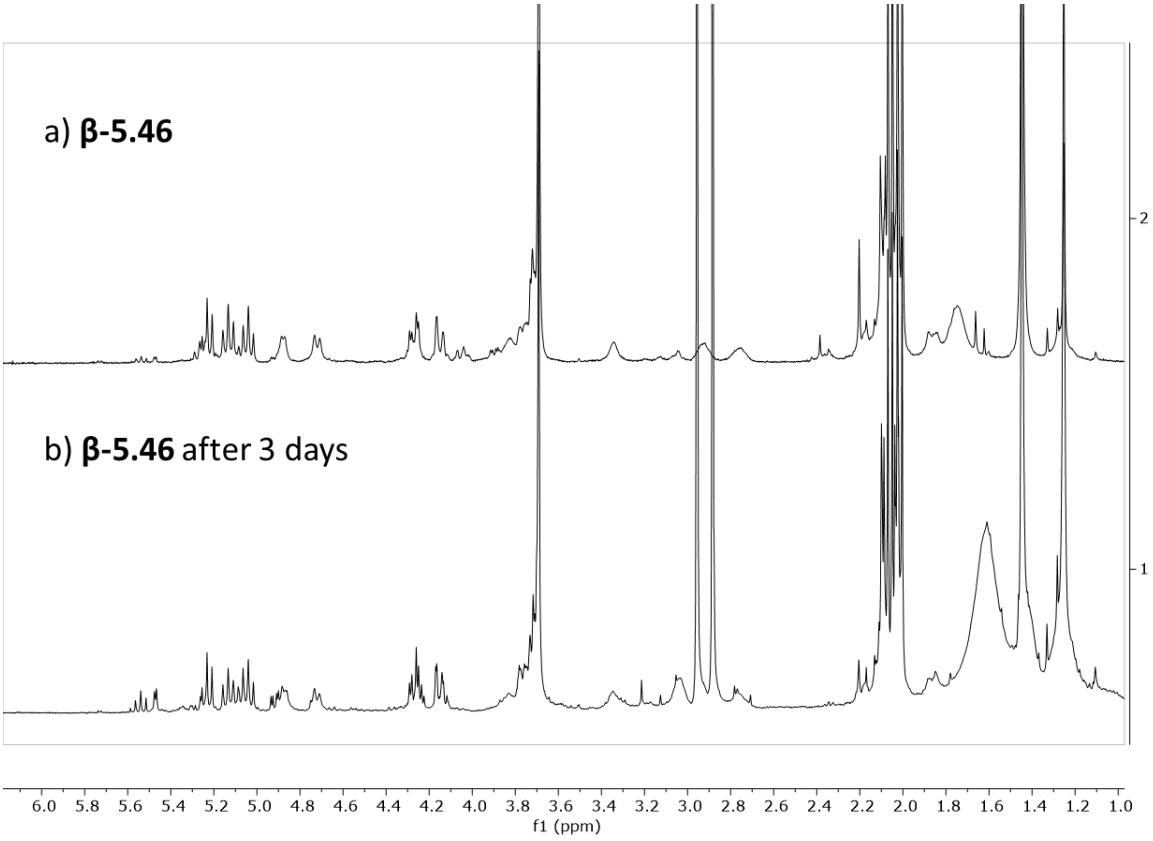
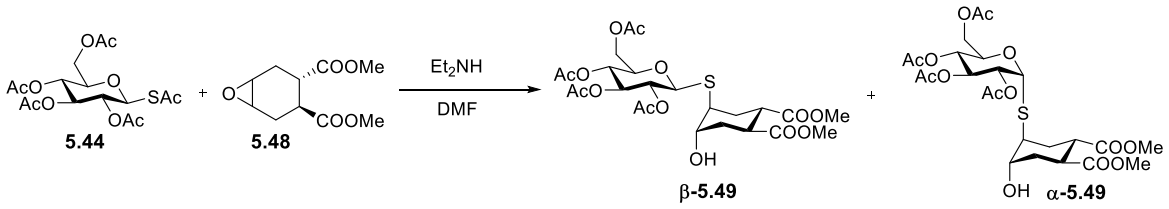


Fig. 5.5 ^{1}H NMR of the product **β-5.46** a) before and b) after treatment with 1.5 eq Et_2NH in DMF [0.8 M] for 3 days (400 MHz ^{1}H NMR in $CDCl_3$)

To examine whether the aziridine substrate could influence and promote the anomerization, we substituted the electrophile in the reaction and performed an epoxide rather than aziridine opening reaction. Epoxide **5.48** was opened by a thiol generated *in situ* from thioacetate **β-5.44** under the initial conditions previously used for the opening of the aziridine: epoxide **5.48** (0.77 eq) with the thioacetate **β-5.44** (1 eq) reaction at RT with 1.5 eq Et_2NH in DMF [0.8 M], 4 h. (**Scheme 5.18**)



Scheme 5.18 Epoxide **5.48** opening reaction with **β-5.44**

Also in this case both **β-5.49** and **α-5.49** products of the epoxide opening reaction were observed in cca. 2.5:1 ratio as judged from crude ^{1}H NMR spectra signals at 4.59 ppm (**β-5.49** H_1) and 5.72 ppm (**α-5.49** H_1) (**Fig. 5.6**). The two isomers were separated by column chromatography and NMR analysis confirmed the configuration of anomeric protons belonging to **β-5.49** product ($J_{1,2}$=10.1 Hz) and **α-**

5.49 product ($J_{1,2}$=5.8 Hz). With this experiment we were able to show that the aziridine substrate is not essential for the anomerization of the thiolate.

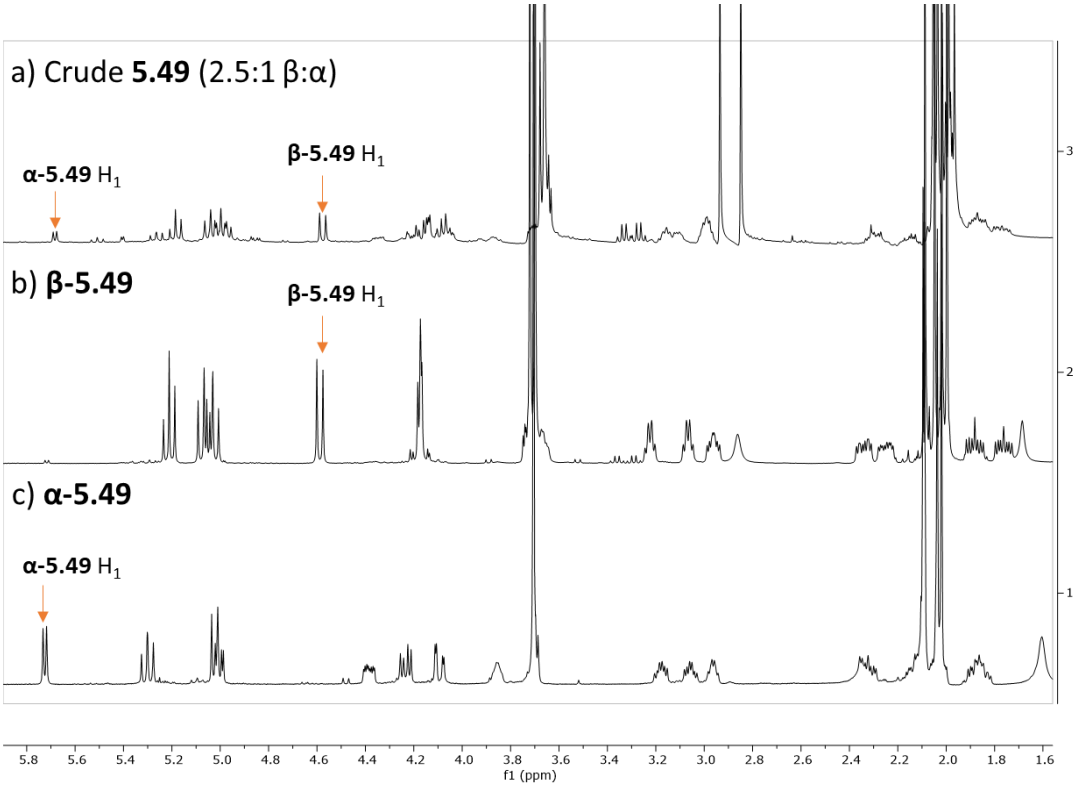
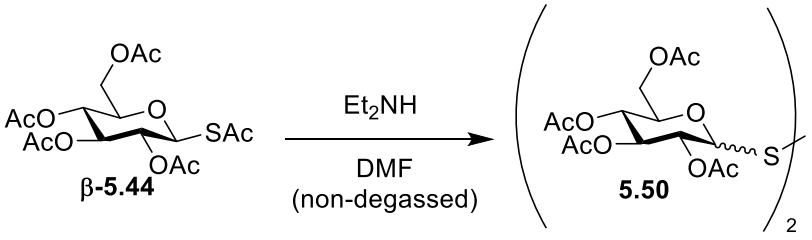


Fig. 5.6 ^{1}H NMR analysis of epoxide **5.48** opening reaction with **β-5.44** to give product **5.49** a) Crude product **5.49**, b) **β-5.49**, c) **α-5.49** (^{1}H NMR 400 MHz in $CDCl_3$)

Finally, we looked into the anomerization of glucosyl thiolate itself under basic conditions. The first experiment we conducted was deacetylation of the starting glucosyl thioacetate **β-5.44** with Et_2NH reproducing the conditions of the aziridine opening, but without the aziridine substrate. Reaction was performed at RT with 1.5 eq Et_2NH in DMF [0.4 M] for 4 h. Operating with non-degassed solvent (DMF) allowed for oxidation of glucosyl thiols to disulphides **5.50** (**Scheme 5.19**).



Scheme 5.19 Deacetylation of glucosyl thioacete **β-5.44** performed in non-degassed solvent (DMF)

In the crude product we were able to see formation of three protons with the chemical shift of 80-90 ppm as judged by the HSQC NMR (**Fig. 5.7**), which could be attributed to the anomeric protons

indicating that anomerization was occurring in the process (**Fig. 5.7**). Considering that formed disulphide products **5.50** have 2 anomeric protons, three different disulphides (β-β, β-α, α-α) can be formed. Additionally, working under not strictly anhydrous conditions can allow for hydrolysis of thiols. While this experiment provided us with the idea that glucosyl thiolate might be anomerizing under employed conditions, the complexity of the mixture of products, prevented us from drawing any definite conclusions.

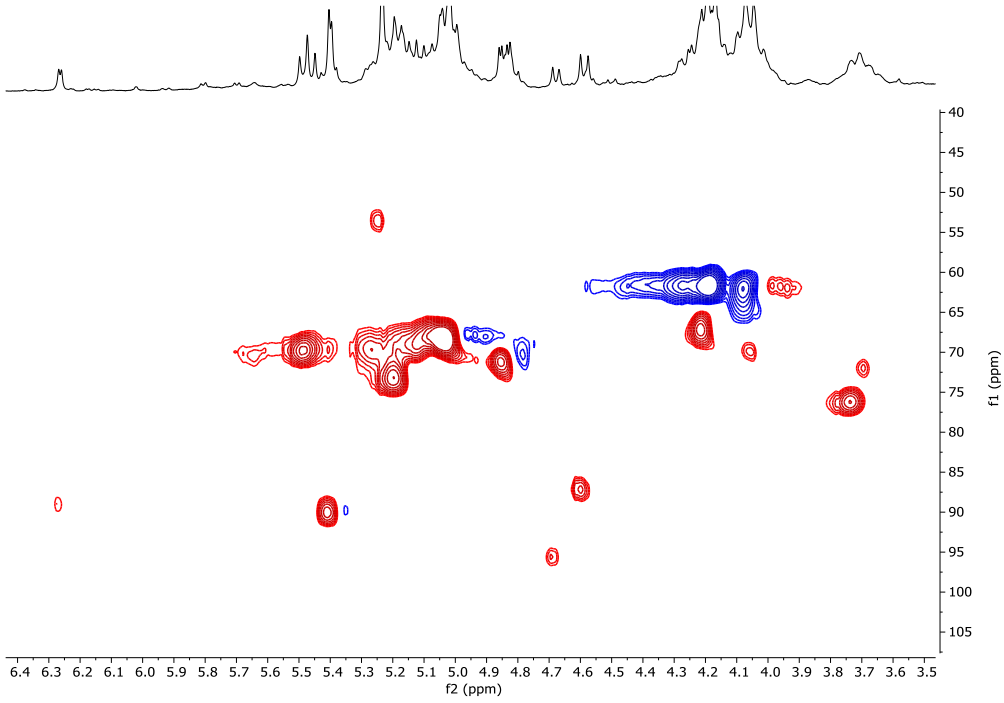
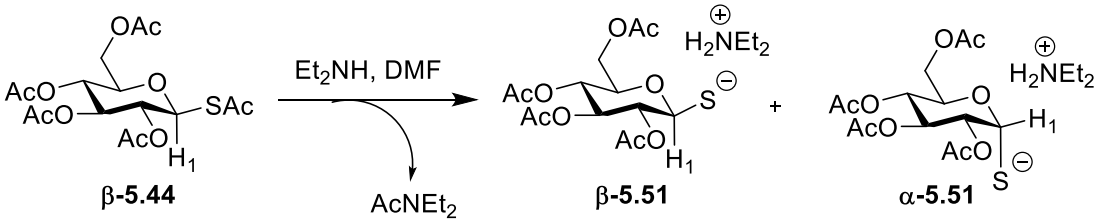


Fig. 5.7 HSQC NMR analysis of crude product **5.50** obtained from the deacetylation reaction of **β-5.44** in non degassed DMF ($CDCl_3$)

Therefore, we performed the experiment again with a more careful set up of the reaction (degassed solvent, Ar atmosphere), this time in deuterated DMF that allowed us to follow the reaction over time by NMR. Reaction was performed in the NMR tube (no stirring) at RT with 1.5 eq Et_2NH in DMF-d_7 [0.4 M] (**Scheme 5.20**).



Scheme 5.20 Deacetylation of glucosyl thioacete **β-5.44** performed in degassed solvent

We recorded the ^{1}H NMR spectra of the reaction mixture every 30 min to follow the deacetylation at the anomeric sulphur of **β-5.44**. We observed that with the consumption of the starting thioacetate **β-5.44** (H_1 proton at 5.51 ppm), the signals of new peaks at 5.16 ppm and 6.19 ppm appeared, which could be attributed to anomeric protons based on HSQC analysis (chemical shift approx. 80 ppm) (**Fig. 5.8**). However, the spectra obtained in DMF-d_7 were not well resolved, additionally deacetylation was slow in the absence of stirring, which again made it difficult to draw any definite conclusions.

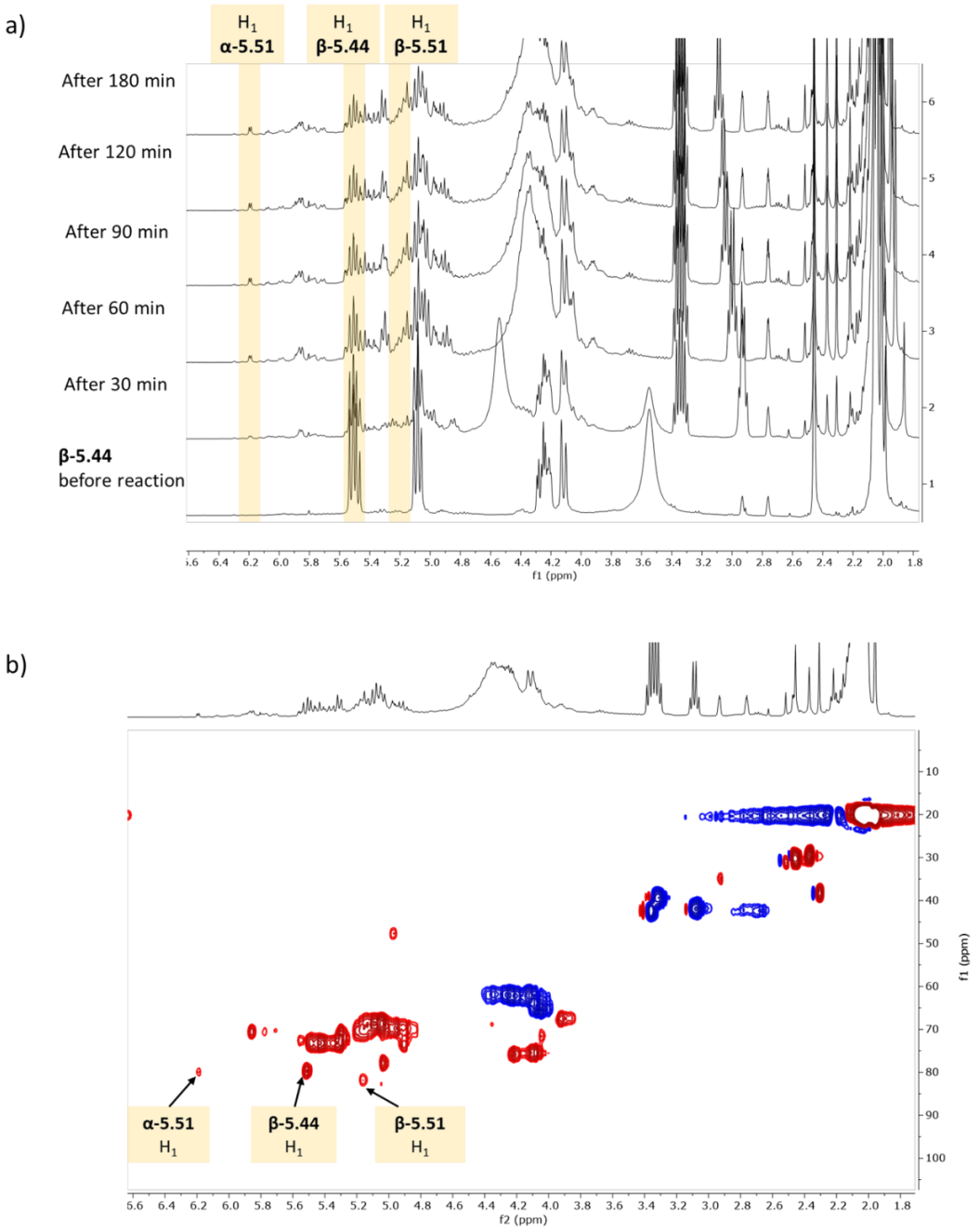


Fig. 5.8 NMR analysis of deacetylation reaction of **β-5.44** performed in degassed DMF-d_7 to give thiolates **5.51** a) ^{1}H NMR spectra recorded over time, b) HSQC NMR after 180 min (400 MHz, DMF-d_7)

Finally, we performed the experiment again in a vial with stirring using degassed dry DMF (non deuterated) under Ar atmosphere. Reaction was performed at RT with 1.5 eq. Et_2NH in DMF [0.8 M], which reproduces the conditions that were initially used for opening or the aziridine **5.45** with the thioacetate **β-5.44**. This time we followed the deacetylation at the anomeric sulphur of **β-5.44** over time by taking small aliquots of the reaction mixture (without quenching) and analysing them by ^1H NMR in $CDCl_3$ immediately. This allowed us to observe the formation of thiolates **5.51** (**Scheme 5.20**). The characteristic peak of anomeric proton of **β-5.51** thiolate was observed at 4.50 ppm ($J_{1,2}$=9.8 Hz) and the peak of anomeric proton of **α-5.51** thiolate at 5.82 ppm ($J_{1,2}$=5.6 Hz) (**Fig. 5.9**). As judged from the ^1H NMR spectra of the reaction mixture after 1 h, the β- and α-thiolates were observed in approximately 2.7:1 ratio, while the starting material was fully converted. The deacetylation of **β-5.44** in the vial with stirring was much faster (1 h) than in the NMR tube without stirring (not fully converted after 3 h), additionally the reaction in the NMR tube was performed at lower concentration, which slowed down the process as well.

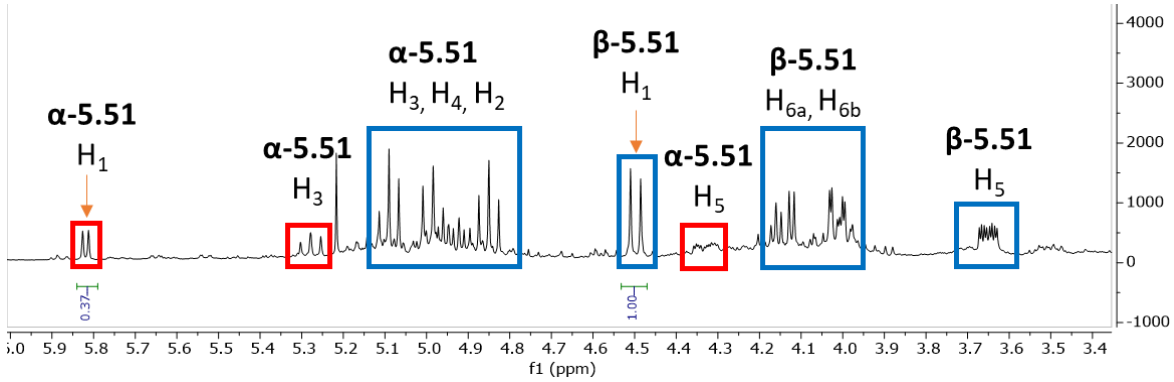


Fig. 5.9 ^1H NMR spectra of crude from deacetylation reaction of **β-5.44** in degassed DMF after 1 h (400 MHz in $CDCl_3$), protons belonging to **β-5.51** are highlighted with blue squares, protons belonging to **α-5.51** are highlighted with red squares

When the deacetylation of **β-5.44** was performed in CH_2Cl_2 as a solvent under the same conditions previously used for the reaction in DMF (reaction at RT with 1.5 eq Et_2NH in CH_2Cl_2 [0.8 M]), the anomeric ratio of thiolates **β-5.51/α-5.51** has changed to 10:1 (**Fig 5.10**), which corresponds well to the result obtained for the aziridine opening reaction performed in CH_2Cl_2 (described in Chapter Four). Deacetylation in CH_2Cl_2 was somewhat slower than in DMF, after 1 h approximately 70% of the starting material was converted.

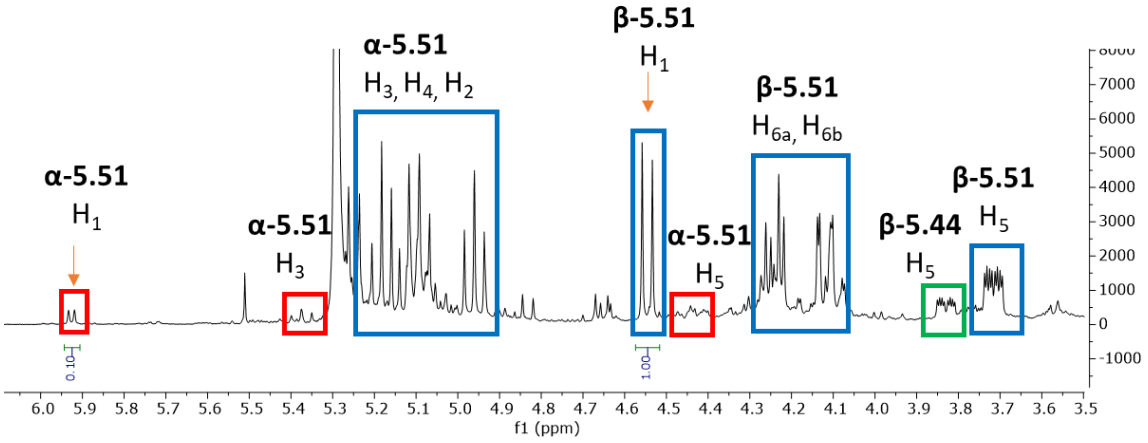
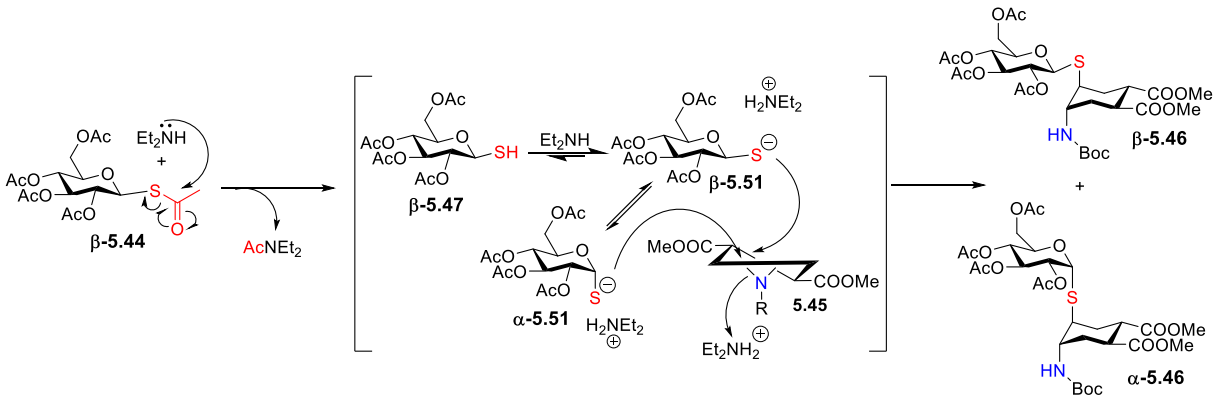


Fig. 5.10 ^{1}H NMR spectra of crude from deacetylation reaction of **β-5.44** in degassed CH_2Cl_2 after 1 h (400 MHz in $CDCl_3$), protons belonging to **β-5.51** are highlighted with blue squares, protons belonging to **α-5.51** are highlighted with red squares, proton belonging to **β-5.44** is highlighted with green square

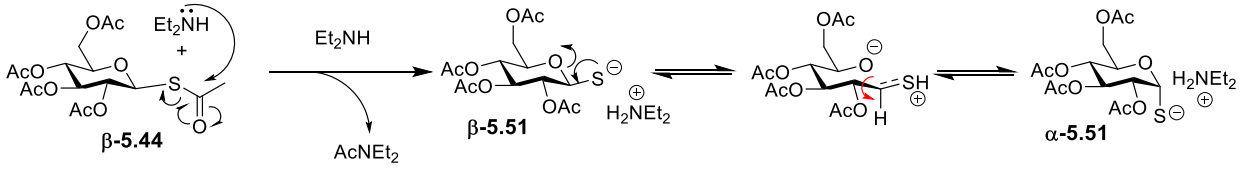
Together with the experiments starting from glucosyl thiol **β-5.47**, these experiments show that the anomeric isomerization occurs through the intermediate thiolate, generated either by deprotonation of **β-5.47** or deacetylation of **β-5.44**. This let us to update the hypothesis of the mechanism of the aziridine opening reaction with glucosyl thioacetate **β-5.44** (**Scheme 5.21**). The mechanism we propose takes into account that the glucosyl thiolate **β-5.51** once generated can attack the aziridine, however it also anomerizes to **α-5.51**, which leads to the product **5.46** being formed as a mixture of β- and α-anomers.



Scheme 5.21 Mechanism of aziridine **5.45** opening reaction with **β-5.44** to give products **β-5.46** and **α-5.46** goes through anomerization of glycosyl thiolate **β-5.51** to **α-5.51**

Most likely the thiolate anomerizes through a mechanism analogous to mutarotation (**Scheme 5.22**) which appears to be very fast in polar aprotic solvents (DMF) and much slower in CH_2Cl_2. Solvents of low polarity, however, also appear to reduce the rate of aziridine opening, so that the overall yields are too low to be synthetically useful. In non-polar solvents, the relative rate of thiolate anomerization

and nucleophilic substitution reaction will determine the final anomeric ratio observed, since the final products are configurationally stable. Therefore, performing the reaction at high concentration is key to favour aziridine opening over the intramolecular anomerization process. *N*-acyl groups able to activate the aziridine towards opening should also help to reduce the extent of anomerization. Finally, performing the opening reaction at low temperature is favourable, but it must be coupled with rigorous exclusion of oxygen to avoid oxidation of the thiolate to disulphide, which depresses the overall yields.



Scheme 5.22 Possible mechanism of anomerization of glycosyl thiolate **5.51** upon deacetylation of **β-5.44**

Table 5.3 ^{1}H NMR peaks of anomeric protons of glucose derivatives used in the study

H_1	δ (ppm)	Shape	J (Hz)	Solvent
β-5.44	5.27	d	9.8	$CDCl_3$
	5.50	mult.	/	DMF-d_7
β-5.46	4.72	d	10.0	$CDCl_3$
α-5.46	5.73	d	5.7	$CDCl_3$
β-5.47	4.54	dd	9.6 ($J_{1\text{-}2} = J_{1\text{-}SH}$)	$CDCl_3$
α-5.47 from Ref[13]	5.94	dd	5.2 ($J_{1\text{-}2} = J_{1\text{-}SH}$)	$CDCl_3$
β-5.49	4.59	d	10.1	$CDCl_3$
α-5.49	5.72	d	5.8	$CDCl_3$
β-5.51	4.55	d	9.8	$CDCl_3$
	5.16	mult.	/	DMF-d_7
α-5.51	5.93	d	5.6	$CDCl_3$
	6.19	d	5.0	DMF-d_7

Table 5.4 Experiments conducted in this study with glucosyl thioacetate **β-5.44** or thiol **β-5.47** and their outcomes

Experiment	Reaction conditions				Outcome (β/α)
	T	Eq base[a]	Conc.[b] (M)	Time (h)	
β-5.44 aziridine **5.45** opening	RT	1.5 Et_2NH	0.8	4	**β-5.46** : **α-5.46** 2 : 1
β-5.47 aziridine **5.45** opening	RT	0.23 DIPEA	0.4	16	**β-5.46** : **α-5.46** 2.5: 1
β-5.47 aziridine **5.45** opening	RT	/	0.4	16	Only **β-5.47** observed (starting material)
β-5.44 epoxide **5.48** opening	RT	1.5 Et_2NH	0.8	4	**β-5.49** : **α-5.49** 2.5 : 1
β-5.44 deacetylation (DMF)	RT	1.5 Et_2NH	0.8	1	**β-5.51** : **α-5.51** 2.7 : 1
β-5.44 deacetylation (CH_2Cl_2)	RT	1.5 Et_2NH	0.8	1	**β-5.51** : **α-5.51** 10 : 1

*[a]Relative to glucosyl thioacetate **β-5.44** or thiol **β-5.47**, [b]Concentration of glucosyl thioacetate **β-5.44** or thiol **β-5.47***

5.4. Conclusions

In Chapter Four we described the one-pot aziridine opening reaction with different glycosyl thioacetates. In certain cases we found that the products of the reaction were formed as mixture of anomers, which was particularly surprising, because literature generally reports configurational stability of glycosyl thiols.

Therefore, to unravel the mystery of the anomeric isomerization of sugar derivatives containing anomeric sulphur we conducted an in depth literature search on this topic and performed mechanistic studies to better understand the cause of anomerization observed during the one-pot aziridine opening reactions.

β-Glucosyl thioacetate **β-5.44** was selected as a model substrate for our studies. Through a series of experiments, we confirmed that the anomerization under the experimental conditions (Et_2NH, DMF) occurs through the intermediate thiolate **5.51**, presumably through endocyclic ring opening mechanism similar to mutarotation.

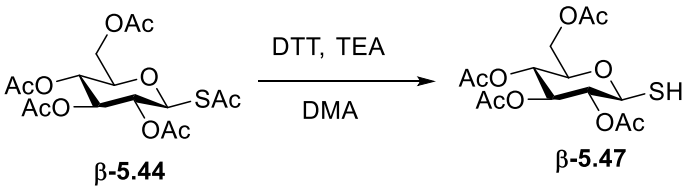
An interesting fact is that anomerization during the one-pot aziridine opening reaction was observed with all glycosyl β-thioacetates (β-glucose, β-galactose, β-fucose, β-lactose, β-*N*-acetylglucosamine), but not with α-thioacetates (α-mannose, α-rhamnose, α-Neu5Ac) used in the study. Additionally,

literature search confirmed that the anomerization under these conditions (Et_2NH, DMF) was observed with β-mannosyl thioacetate[17], but not with α-mannosyl thioacetate[19]. This could potentially indicate that α-thiolates are stabilized by anomeric effect.

Solvent effects also play a significant role in the anomerization of glycosyl thiolates, anomerization is slower in non polar slovents (CH_2Cl_2) than in polar aprotic solvents (DMF). Additionally, stereocontrol is better at lower temperatures as shown by our studies (Chapter Four) and literature[17].

5.5. Experimental

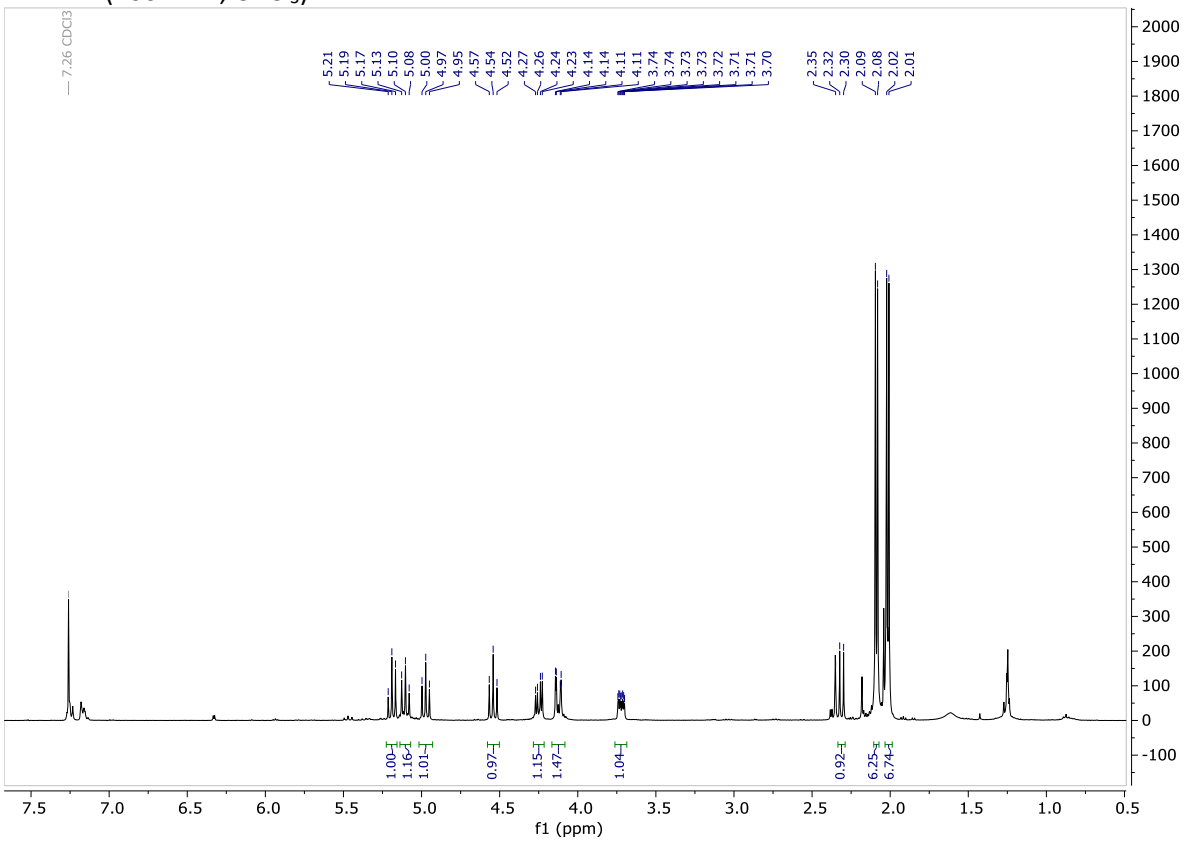
Synthesis of β-1-thiol-2,3,4,6-tetra-*O*-acetyl-D-glucopyranose **β-5.47**



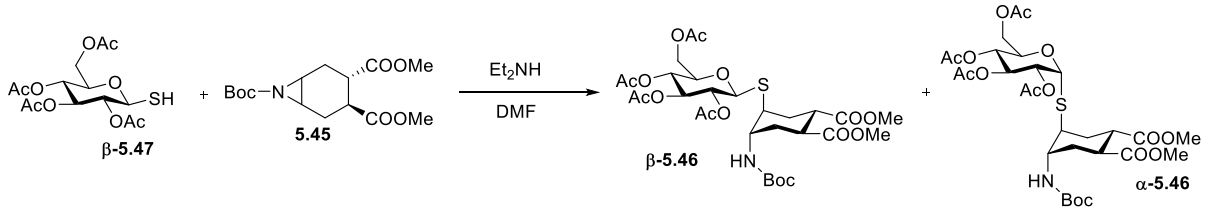
β-**5.44** β-**5.47**

β-1-*S*-acetyl-2,3,4,6-tetra-*O*-acetylglucopyranose **5.44** (0.06 g, 0.148 mmol) and DTT (0.034 g, 0.221 mmol) were dissolved in dry DMA (1 mL) under N_2 atmosphere. Then Et_3N (2 µL, 0.0148 mmol) was added and the reaction was left stirring at room temperature for 1.5 h. Afterwards the reaction mixture was diluted with EtOAc and washed with H_2O. Organic phase was then washed with brine. Combined aqueous phases were washed with EtOAc. Combined organic phases were dried over Na_2SO_4 and concentrated under reduced pressure. The crude was then dissolved in toluene and washed with H_2O (4x). Combined organic phases were then dried over Na_2SO_4, concentrated and dried under high vacuum. The thiol **β-5.47**[18] was obtained as a white solid in 81% yield (0.044 g, 0.120 mmol). ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.19 (dd, $J_{3\text{-}2}$ = $J_{3\text{-}4}$ = 9.6 Hz, 1H, H_3), 5.10 (dd, $J_{4\text{-}3}$ = $J_{4\text{-}5}$ = 9.6 Hz, 1H, H_4), 4.97 (dd, $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.6 Hz, 1H, H_2), 4.54 (dd, $J_{1\text{-}2}$ = $J_{1\text{-}SH}$ = 9.6 Hz, 1H, H_1), 4.25 (dd, $J_{6a\text{-}6b}$ = 12.5, $J_{6a\text{-}5}$ = 4.8 Hz, 1H, H_{6a}), 4.12 (dd, $J_{6b\text{-}6a}$ = 12.5, $J_{6b\text{-}5}$ = 2.3 Hz, 1H, H_{6b}), 3.72 (ddd, $J_{5\text{-}4}$ = 9.6, $J_{5\text{-}6a}$ = 4.8, $J_{5\text{-}6b}$ = 2.3 Hz, 1H, H_5), 2.31 (d, $J_{SH\text{-}1}$ = 9.6 Hz, 1H, SH), 2.09 (s, 3H, OAc), 2.08 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.01 (s, 3H, OAc).

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



Aziridine **5.45** opening reaction with β-glucosyl thiol **β-5.47**



Free thiol **β-5.47** (14 mg, 0.038 mmol) and aziridine **5.45** (9.3 mg, 0.030 mmol) were dissolved in dry DMF (99 µL) at RT under N_2 atmosphere, then DIPEA (1.54 µL, 0.0089 mmol) was added. Reaction mixture was kept stirring overnight at RT.

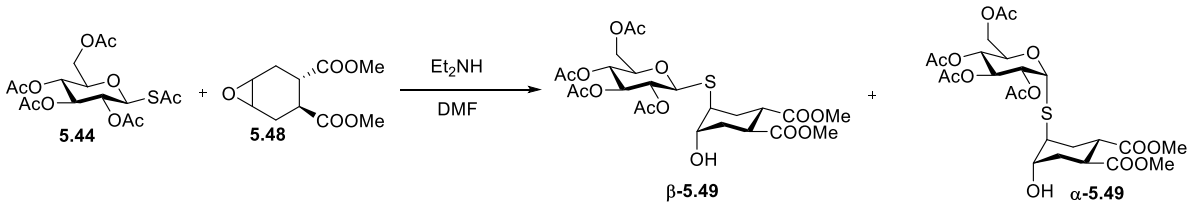
The reaction mixture was diluted with EtOAc and washed with 1 M HCl. The organic phase was washed three times with H_2O and the aqueous phases were additionally extracted with EtOAc. The combined organic phases were dried over Na_2SO_4, filtered and concentrated under vacuum. Crude was further purified by column chromatography (Hex:EtOAc 1:1) and the product **5.46** was obtained as an anomeric mixture in 34% yield (6.9 mg, 0.010 mmol).

β-4.46: R_f=0.22 (iPr_2O/EtOAc 8:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.23 (dd, $J_{3\text{-}4}$= $J_{3\text{-}2}$= 10 Hz, 1H, H_3), 5.13 (dd, $J_{4\text{-}3}$= $J_{4\text{-}5}$= 10 Hz, 1H, H_4), 5.04 (dd, $J_{2\text{-}1}$= $J_{2\text{-}3}$= 10 Hz, 1H, H_2), 4.87 (m, 1H, NH), 4.72 (d, $J_{1\text{-}2}$= 10 Hz, 1H, H_1), 4.27 (dd, $J_{6a\text{-}6b}$= 12.4 Hz, $J_{6a\text{-}5}$= 4.5 Hz, 1H, H_{6a}), 4.15 (dd, $J_{6b\text{-}6a}$= 12.4 Hz, $J_{6b\text{-}5}$= 2.4 Hz, 1H, H_{6b}),

3.93 – 3.80 (m, 1H, $H_{4'}$), 3.76 (ddd, $J_{5\text{-}4}$= 10.0 Hz, $J_{5\text{-}6a}$= 4.5 Hz, $J_{5\text{-}6b}$= 2.4 Hz, 1H, H_5), 3.69 (s, 6H, 2xOMe), 3.42 – 3.26 (m, 1H, $H_{5'}$), 2.99 – 2.86 (m, 1H, $H_{1'}$), 2.81 – 2.70 (m, 1H, $H_{2'}$), 2.24 – 2.08 (mult., 3H, $H_{3'eq}$, $H_{6'eq}$, $H_{6'ax}$), 2.07 (s, 3H, OAc), 2.05 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.00 (s, 3H, OAc), 1.91 – 1.81 (m, 1H, $H_{3'ax}$), 1.45 (s, 9H, tBu); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.0 (CO), 171.8 (CO), 171.7 (CO), 170.8 (CO), 170.3 (CO), 169.5 (CO), 83.7 (C_1), 76.0 (C_5), 74.1 (C_3), 70.2 (C_2), 68.3 (C_4), 62.0 (C_6), 52.4 (OMe), 52.3 (OMe), 48.2* ($C_{4'}$), 42.1* ($C_{5'}$), 40.1 ($C_{1'}$, $C_{2'}$), 29.8 ($C_{3'}$, $C_{6'}$), 28.5 (tBu-3xMe), 20.9 (Ac), 20.9 (Ac), 20.8 (Ac), 20.8 (Ac). * These signals are better visible in the HSQC spectrum; **α-5.46:** ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.73 (d, $J_{2\text{-}1}$ = 5.7 Hz, 1H, H_1)

(Full characterization of **5.46** is reported in Chapter Four, compound **4.34**)

Epoxide **5.48** opening reaction with β-glucosyl thioacetate **β-5.44**

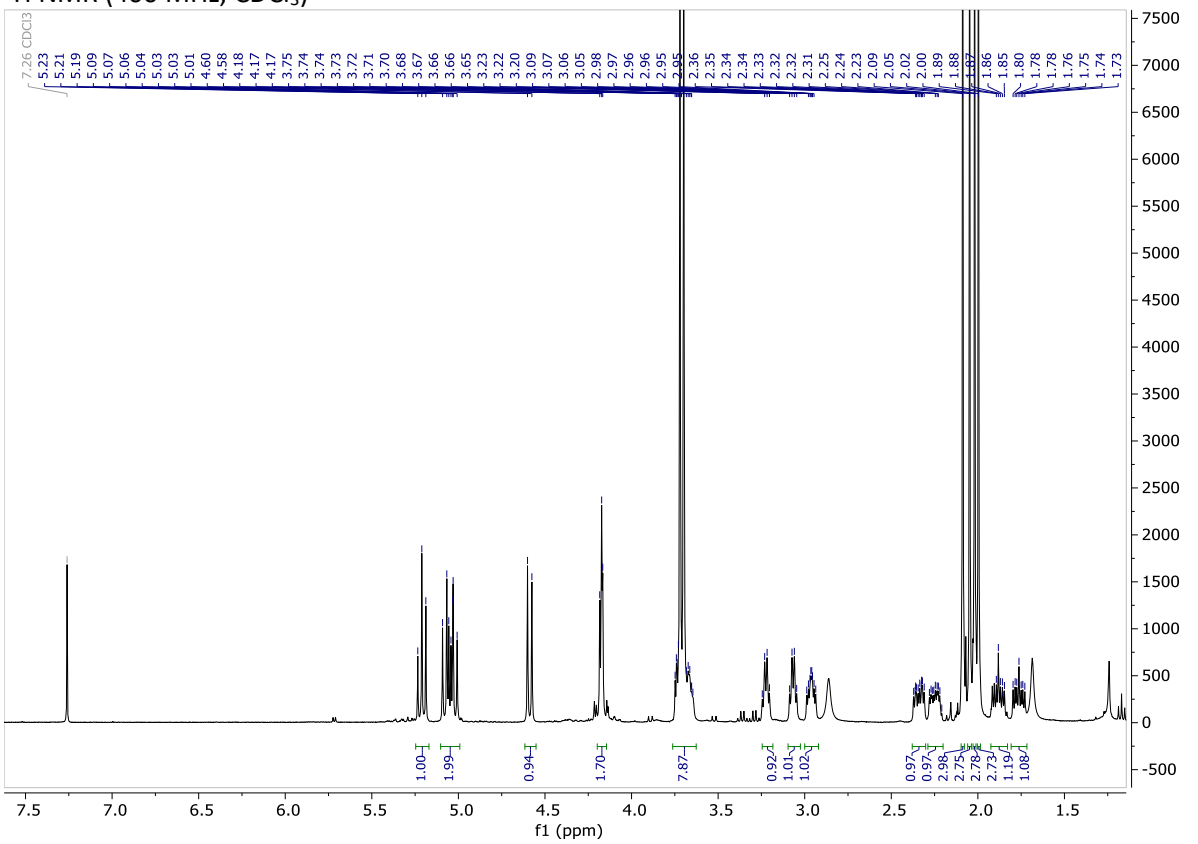


Glucosyl thioacetate **5.44** (49 mg, 0.121 mmol) and epoxide **5.48** (20 mg, 0.093 mmol) were dissolved in dry degassed DMF (0.14 mL) under Ar atmosphere and Et_2NH (18.3 µL, 0.18 mmol) was added. Reaction mixture was kept stirring at RT for 4 h.

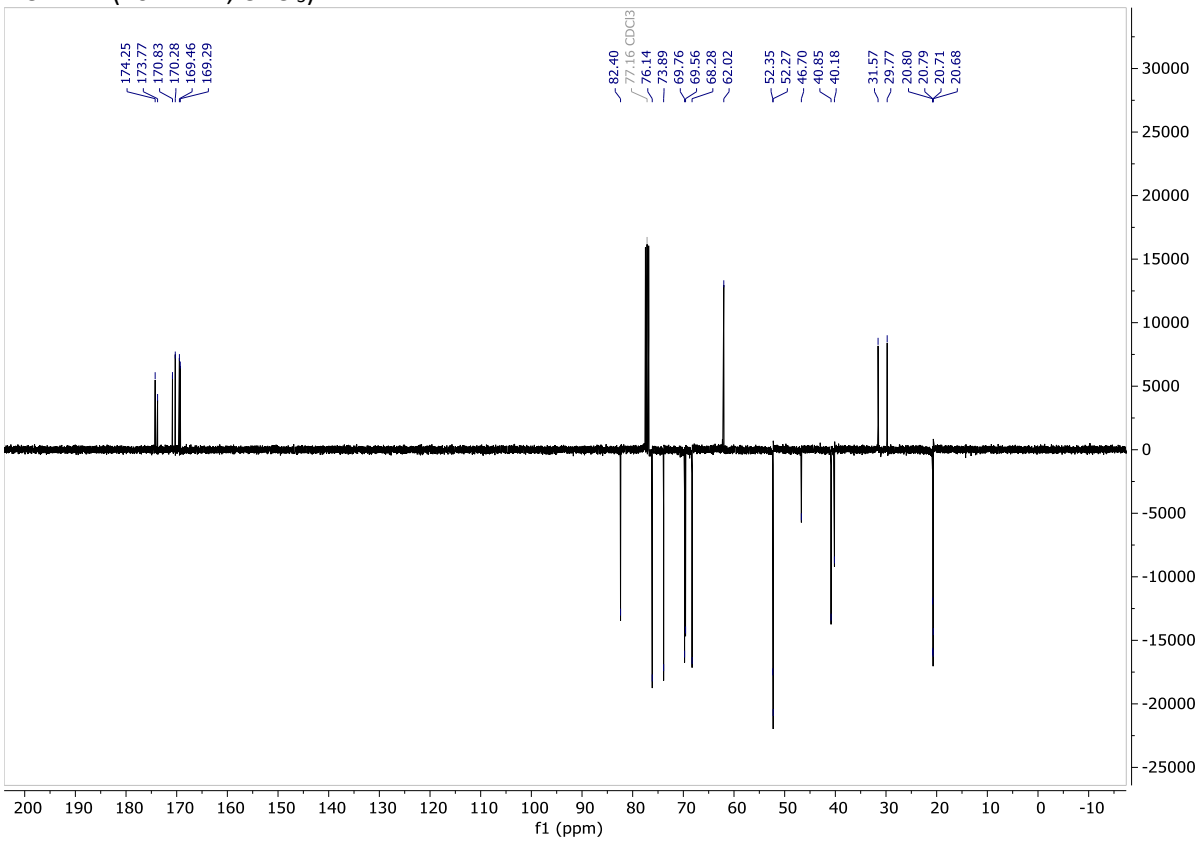
The reaction mixture was diluted with EtOAc and washed with 1 M HCl. The organic phase was washed three times with H_2O and the aqueous phases were additionally extracted with EtOAc. The combined organic phases were dried over Na_2SO_4, filtered and concentrated under vacuum. Crude was further purified by column chromatography (Hex:EtOAc 1:1) and the product **β-5.49** was isolated in 33% yield (18 mg, 0.031 mmol) and **α-5.49** in 15% yield (7.9 mg, 0.014 mmol).

β-5.49 R_f= 0.08 (Hex:EtOAc 1:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.21 (dd, $J_{3\text{-}4}$= $J_{3\text{-}4}$= 9.3 Hz, 1H, H_3), 5.13 – 5.00 (m, 2H, H_4, H_2), 4.59 (d, $J_{1\text{-}2}$= 10.1 Hz, 1H, H_1), 4.17 (mult., 2H, H_{6a}, H_{6b}), 3.77 – 3.63 (mult., 8H, H_5, $H_{4'}$, 2xCOOMe), 3.22 (q, $J_{1'\text{-}2'}$= 5.5 Hz, 1H, $H_{1'}$), 3.07 (q, $J_{2'\text{-}1'}$= 5.5 Hz, 1H, $H_{2'}$), 2.99 – 2.92 (m, 1H, $H_{5'}$), 2.38 – 2.31 (m, 1H, $H_{6'eq}$), 2.25 (m, 1H, $H_{3'eq}$), 2.09 (s, 3H, OAc), 2.05 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.00 (s, 3H, OAc), 1.93 – 1.83 (m, 1H, $H_{6'ax}$), 1.76 (m, 1H, $H_{3'ax}$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.3 (CO), 173.8 (CO), 170.8 (CO), 170.3 (CO), 169.5 (CO), 169.3 (CO), 82.4 (C_1), 76.1 (C_5), 73.9 (C_3), 69.8 (C_2), 69.6 (C_4), 68.3 ($C_{4'}$), 62.0 (C_6), 52.3 (2xCOOMe), 46.7 ($C_{5'}$), 40.8 ($C_{2'}$), 40.2 ($C_{1'}$), 31.6 ($C_{3'}$), 29.8 ($C_{6'}$), 20.8 (OAc), 20.8 (OAc), 20.7 (OAc), 20.7 (OAc); MS (ESI) calcd for $C_{24}H_{34}O_{14}S$ $[M + Na]^+$ m/z: 601.17; found: 601.55.

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

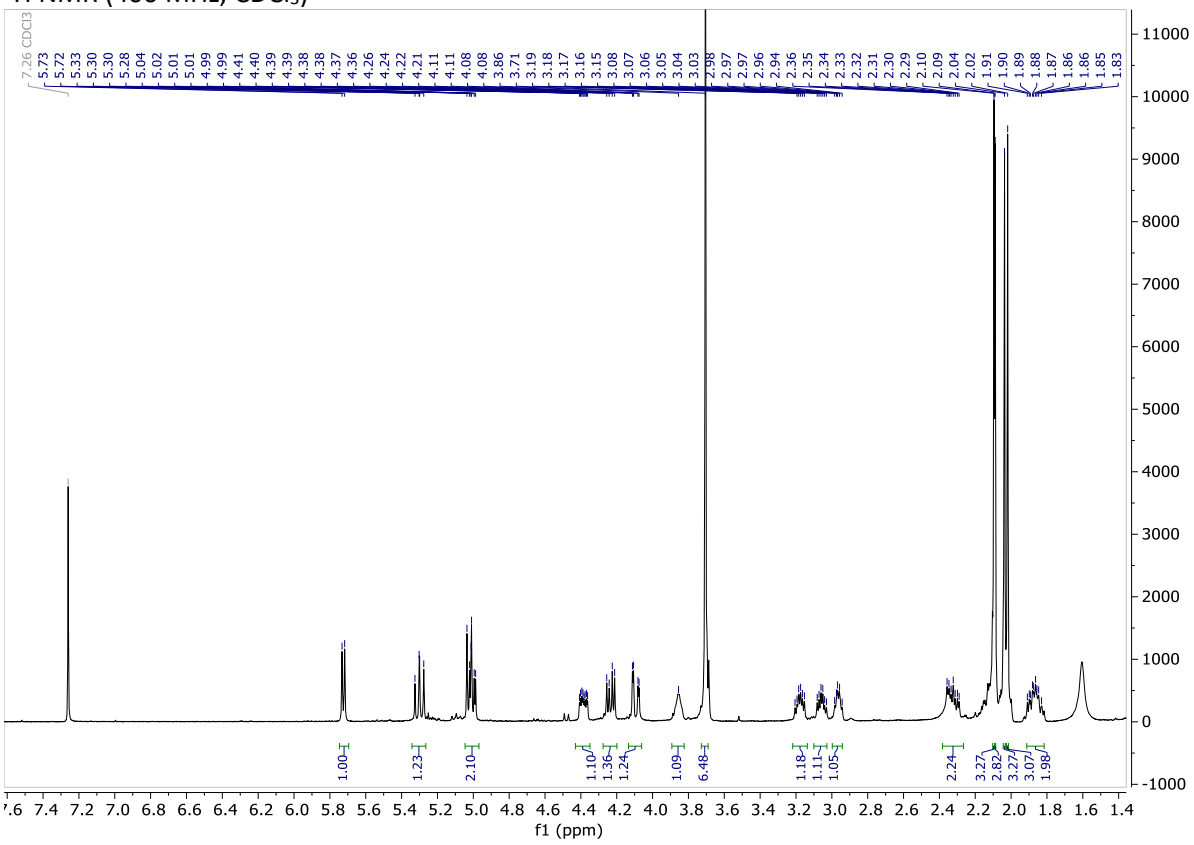


^{13}C NMR (101 MHz, $CDCl_3$)

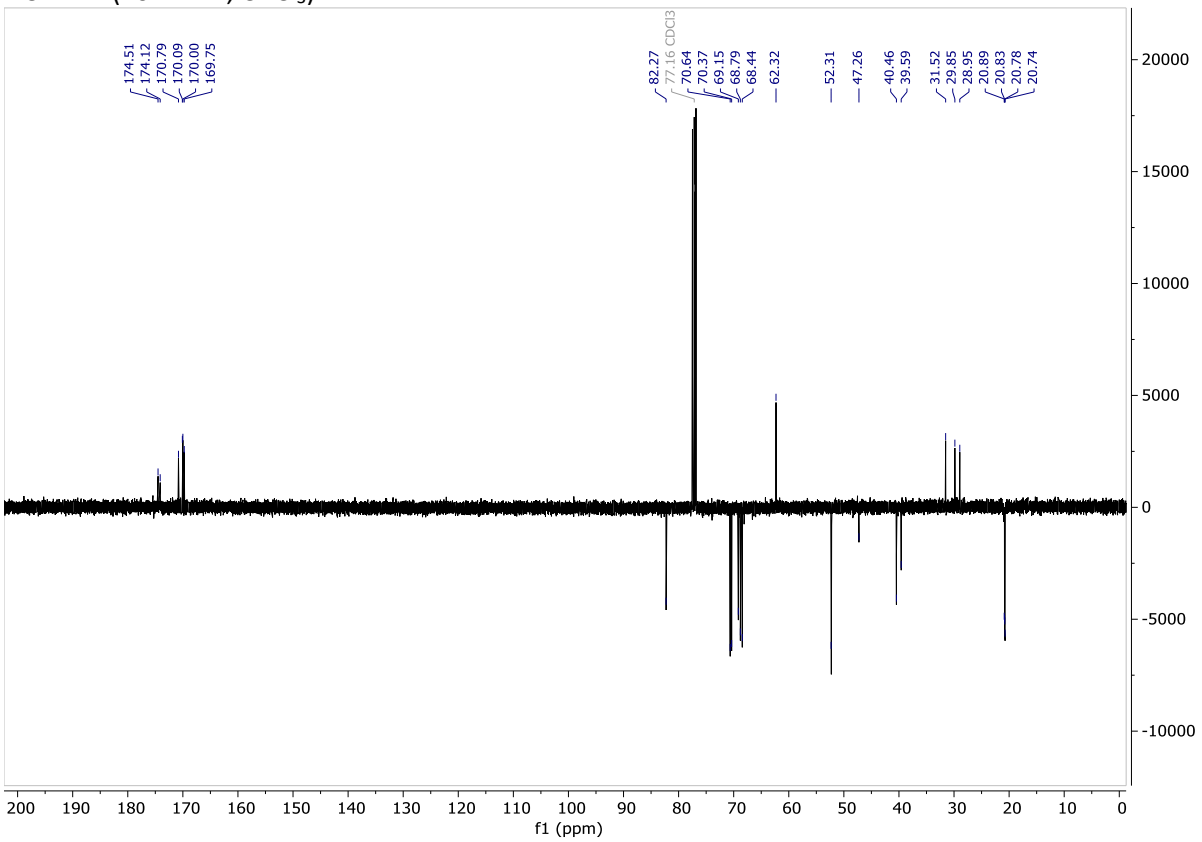


α-5.49 R_f= 0.12 (Hex:EtOAc 1:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.72 (d, $J_{1,2}$= 5.8 Hz, 1H, H_1), 5.30 (dd, $J_{3\text{-}2}$= 10.4, $J_{3\text{-}4}$= 9.2 Hz, 1H, H_3), 5.08 – 4.96 (mult., 2H, H_2, H_4), 4.39 (ddd, $J_{5\text{-}4}$= 10.2, $J_{5\text{-}6a}$= 5.4, $J_{5\text{-}6b}$= 2.5 Hz, 1H, H_5), 4.25 – 4.19 (m, 1H, H_{6a}), 4.09 (dd, $J_{6b\text{-}6a}$= 12.3, $J_{6b\text{-}5}$= 2.5 Hz, 1H, H_{6b}), 3.93 – 3.83 (m, 1H, $H_{4'}$), 3.71 (2x s, 6H, 2xCOOMe), 3.18 (dt, $J_{1'\text{-}2'}$= 8.4, $J_{1'\text{-}6'eq}$= 4.4 Hz, 1H, $H_{1'}$), 3.05 (dt, $J_{2'\text{-}1'}$= 8.4, $J_{2\text{-}3'eq'}$= 4.2 Hz, 1H, $H_{2'}$), 2.97 (dt, $J_{5'\text{-}4'}$= 10.3, $J_{5'\text{-}6'}$= 5.0 Hz, 1H, $H_{5'}$), 2.33 (mult., 2H, $H_{6'eq}$, $_{H3'\ eq}$), 2.10 (s, 3H, OAc), 2.09 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.02 (s, 3H, OAc), 1.92 – 1.81 (mult., 2H, $H_{6'ax}$, $H_{3'ax}$); ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.5 (CO), 174.1 (CO), 170.8 (CO), 170.1 (CO), 170.0 (CO), 169.7 (CO), 82.3 (C_1), 70.6 (C_3), 70.4 (C_2), 69.2 (C_4), 68.8 ($C_{4'}$), 68.4 (C_5), 62.3 (C_6), 52.3 (2xCOOMe), 47.3 ($C_{5'}$), 40.5 ($C_{2'}$), 39.6 ($C_{1'}$), 31.5 ($C_{3'}$), 29.0 ($C_{6'}$), 20.9 (OAc), 20.8 (OAc), 20.8 (OAc), 20.7 (OAc); MS (ESI) calcd for $C_{24}H_{34}O_{14}S$ $[M + Na]^+$ m/z: 601.17; found: 601.43.

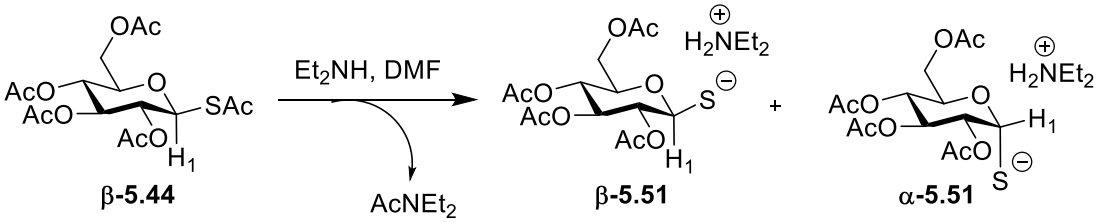
^{1}H NMR (400 MHz, $CDCl_3$)



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



Deacetylation of glucosyl thioacetate **β-5.44** in dry degassed DMF



Starting glucosyl thioacetate **β-5.44** (30 mg, 0.074 mmol) was dissolved in dry degassed DMF (0.84 µL) under Ar atmosphere at RT and treated with Et_2NH (11.4 µL, 0.11 mmol). Reaction mixture was kept stirring at RT under Ar atmosphere and aliquots were taken for NMR analysis keeping the inert atmosphere in the flask.

5.6. References

1. T. K. Lindhorst, *Essentials of Carbohydrate Chemistry and Biochemistry*, Wiley, 2007.
2. S. J. Angyal, in *Advances in Carbohydrate Chemistry and Biochemistry*, eds. R. S. Tipson and D. Horton, Academic Press, 1984, vol. 42, pp. 15-68.
3. L. Deng, X. Wang, S. Uppalapati, O. Norberg, H. Dong, A. Joliton, M. Yan and O. Ramström, *Pure and applied chemistry. Chimie pure et appliquee*, 2013, **85**, 1789-1801.
4. H. Hashimoto, K. Shimada and S. Horito, *Tetrahedron: Asymmetry*, 1994, **5**, 2351-2366.
5. A. Ané, S. Josse, S. Naud, V. Lacône, S. Vidot, A. Fournial, A. Kar, M. Pipelier and D. Dubreuil, *Tetrahedron*, 2006, **62**, 4784-4794.
6. R. Caraballo, L. Deng, L. Amorim, T. Brinck and O. Ramström, *The Journal of Organic Chemistry*, 2010, **75**, 6115-6121.
7. J. Banoub and D. R. Bundle, *Canadian Journal of Chemistry*, 1979, **57**, 2091-2097.
8. J. L. Xue, S. Cecioni, L. He, S. Vidal and J.-P. Praly, *Carbohydrate Research*, 2009, **344**, 1646-1653.
9. M. Poláková, N. Pitt, M. Tosin and P. V. Murphy, *Angewandte Chemie International Edition*, 2004, **43**, 2518-2521.
10. M. Tosin and P. V. Murphy, *Organic Letters*, 2002, **4**, 3675-3678.
11. S. Manabe, K. Ishii, D. Hashizume, H. Koshino and Y. Ito, *Chemistry – A European Journal*, 2009, **15**, 6894-6901.
12. W. Pilgrim and P. V. Murphy, *The Journal of Organic Chemistry*, 2010, **75**, 6747-6755.
13. L. M. Doyle, S. O'Sullivan, C. Di Salvo, M. McKinney, P. McArdle and P. V. Murphy, *Organic Letters*, 2017, **19**, 5802-5805.
14. S. Oscarson, in *Glycoscience: Chemistry and Chemical Biology I–III*, eds. B. O. Fraser-Reid, K. Tatsuta and J. Thiem, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, DOI: 10.1007/978-3-642-56874-9_20, pp. 643-671.
15. K. Pachamuthu and R. R. Schmidt, *Chemical Reviews*, 2006, **106**, 160-187.
16. M. Adinolfi, D. Capasso, S. Di Gaetano, A. Iadonisi, L. Leone and A. Pastore, *Organic & Biomolecular Chemistry*, 2011, **9**, 6278-6283.
17. H. N. Yu, C.-C. Ling and D. R. Bundle, *Journal of the Chemical Society, Perkin Transactions 1*, 2001, DOI: 10.1039/B009626L, 832-837.
18. P. Shu, J. Zeng, J. Tao, Y. Zhao, G. Yao and Q. Wan, *Green Chemistry*, 2015, **17**, 2545-2551.
19. T. Belz and S. J. Williams, *Carbohydrate Research*, 2016, **429**, 38-47.

CHAPTER SIX

Cyanovirin-N as a flexible antiviral platform

6.1. Glycosylation of viral proteins

Viruses can be divided in two groups based on the presence of the lipid bilayer membrane on their outer surface: non-enveloped and enveloped viruses. Enveloped viruses have a lipid bilayer that is derived from the host cell, this membrane incorporates receptor attachment proteins and membrane fusion proteins-all modified by glycosylation, mainly by *N*-glycans and *O*-GalNAc type glycans. Glycans are involved in viral pathogenesis, playing a key role in viral adhesion and epitope shielding from antibody recognition.

For example, the surface of the enveloped virus human immunodeficiency virus (HIV) is dominated by spikes made-up of the glycoproteins gp120 and gp41. Gp120 is heavily glycosylated with *N*-linked glycans, which helps to mask the underlying polypeptide from immune system recognition and additionally supports infection by co-opting host lectins (C-type lectins such as DC-SIGN). Gp120 mediates viral attachment to host cells, primarily CD4+T cells, by binding to host all-surface receptor CD4 and chemokine receptor co-receptor CCR5 and CXCR4.[1]

Another example is the novel type of coronavirus SARS-CoV-2 which has left its mark on the year 2020 with a global threat of severe acute respiratory syndrome COVID-19. SARS-CoV-2 virus has genetic similarities to other types of coronaviruses SARS-CoV and MERS-CoV, but it has higher dissemination rates and has therefore put serious pressure on healthcare systems worldwide. Glycoproteins of SARS-CoV-2 are involved in cell adhesion and invasion, morphogenesis and modulation of immune response processes.[2] The majority of experimental information on the SARS-CoV-2 glycoproteins is currently available for the S-protein that mediates the viral adhesion by binding to the human angiotensin-converting enzyme 2 (hACE2) and interacts with the host immune defence.[3-5] The S-protein has 22 predicted sites for *N*-glycosylation and 3 sites for *O*-glycosylation and experimentally oligomannose- and complex-type glycans were found. The most common configuration of mannose-type glycans on SARS-Cov2 is $Man_5GlcNAc_2$.[6]

6.2. Lectins to combat viral infections

Lectins as glycan binding proteins can potentially be used for multiple purposes in defence against viruses, from biosensors to therapeutics. Some examples how lectins can be utilized in fighting viral infections are described below.

Prevention is the best strategy for managing viral outbreaks and vaccines are therefore of utmost importance. The challenge in vaccine development is induction of long-lasting immunity. Powerful adjuvants can improve the efficacy of immunization, by enhancing the immune response induced by the vaccine. Some lectins are promising candidates for use as adjuvants, because they can promote proliferation of lymphocytes and modulate the release of effector molecules (cytokines and nitric oxide) by immune cells.[2] An example of an adjuvant lectin is the Korean mistletoe lectin (KML-C) isolated from *Viscum album coloratum*. When administered together with inactivated H1N1 influenza virus, it completely protected mice against the virus and significantly increased the levels of anti-influenza antibodies (IgG and IgA) and the population of influenza-specific lymphocytes.[7] However, use of lectins also comes with risks, agglutination of host cells, induction of undesired inflammation and proliferation can occur.

On the other hand, lectin-based biosensors could be an interesting application to detect viral infections, based on detection of glycoconjugates as biomarkers. Lectin-microarrays and lectin-immunoassays could be used for such purposes. Biosensors could detect viral glycoproteins or glycan patterns in infected individuals, but sensitivity and specificity need to be well considered in developing this strategy.[2]

To reduce viral levels in the blood stream, an alternative treatment strategy could be extracorporeal purification of pathogens (plasmapheresis) from blood with the help of lectins. A plasmapheresis apparatus has been designed on the basis of lectin affinity to bind mannose-containing glycans and *in vivo* evaluation showed that association of this device with dialysis effectively reduced the hepatitis C virus (HCV) load in the blood.[8]

Finally, lectins have the potential for therapeutic applications. However, the use of lectins as therapeutics comes with major limitations, such as mitogenicity and proinflammatory properties, as well as rapid clearance and proteolytic digestion. Protein engineering techniques are used to overcome these issues and some examples are described below.

Griffithsin

The Jacalin-related lectin griffithsin (GRFT), first isolated from the red algae *Griffithsia*, contains three carbohydrate binding sites in each monomer and occurs as a domain-swapped dimer with six identical

binding sites. It has shown *in vitro* and *in vivo* activity against HIV, Ebola virus, MERS-CoV and SARS-CoV and it is a promising molecule for SARS-CoV-2 treatment.[9]

With the help of protein engineering, protein constructs containing two, three or four copies of monomeric GRFT were designed and cell-based assays showed a higher antiviral activity against HIV of the trimers and tetramers compared to the dimers and wild type GRFT. [10]

BanLec

Banana lectin (BanLec) isolated from the banana *Musa acuminate* is a mannose-binding lectin that has a tetrameric structure presenting two binding sites in each monomer. To improve its therapeutic potential an engineered derivative with a single-site mutation at position 84 (His->Thr) was designed. This single amino acid replacement yielded a derivative with a better safety profile, while the mannose binding activity stayed intact. This mutant showed antiviral activity against HIV, hepatitis C, influenza and Ebola.[11]

Microvirin

Microvirin (MVN) was first isolated from cyanobacterium *Microcystis aeruginosa*. It is a monomeric lectin with two domains and only one carbohydrate binding site and it potently inactivates HIV-1 and hepatitis C viruses due to binding to high-mannose type *N*-glycans present in the envelope glycoproteins.[12] Tri- and tetrameric versions of this lectin more potently blocked hepatitis C virus infections than dimer and monomer constructs.[13]

A version of MVN with two identical binding sites was also designed recently, thus a MVN based protein with two carbohydrate binding sites. This construct showed lower sugar affinity than wild type, but was still able to block HIV and hepatitis C viral infections and had lower cytotoxicity and immunogenicity.[14]

6.3. Cyanovirin-N

Cyanovirin-N (CVN) was first isolated from cyanobacterium *Nostoc ellipsosporum* in 1997.[15] It has received a great deal of attention, because it potently inactivates diverse strains of HIV at the level of cell fusion.[16] Therefore CVN was considered to become the lead for a new generation of topical microbicides for HIV prevention. Vaginal and rectal topical applications were proposed as the most suitable due to cytotoxic and mitogenic effects observed with the use of CVN.

Later it turned out that CVN has a broad anti-viral activity against enveloped viruses and potently inactivates HIV type 1 and 2, SIV, Ebola virus, influenza, herpesvirus 6 at nanomolar concentrations.[16-18] A particular advantage of CVN in comparison to other lectins is its stability towards a range of

different conditions. CVN is resistant to 0.5% SDS detergents, organic solvents (CH_3CN, MeOH, and DMSO), continuous freezing and thawing, and high temperatures (100°C) without losing antiviral activity. [19, 20]

The anti-HIV activity of CVN is due to its binding to high mannose glycans (Man_8, Man_9) present on the envelope glycoprotein gp120.[21] The lectin-bound gp120 afterwards cannot productively interact with the host cell CD4 and CCR5 receptors, which prevents the necessary conformational changes required for membrane fusion and viral entry into the host cell as shown in **Fig. 6.1**.

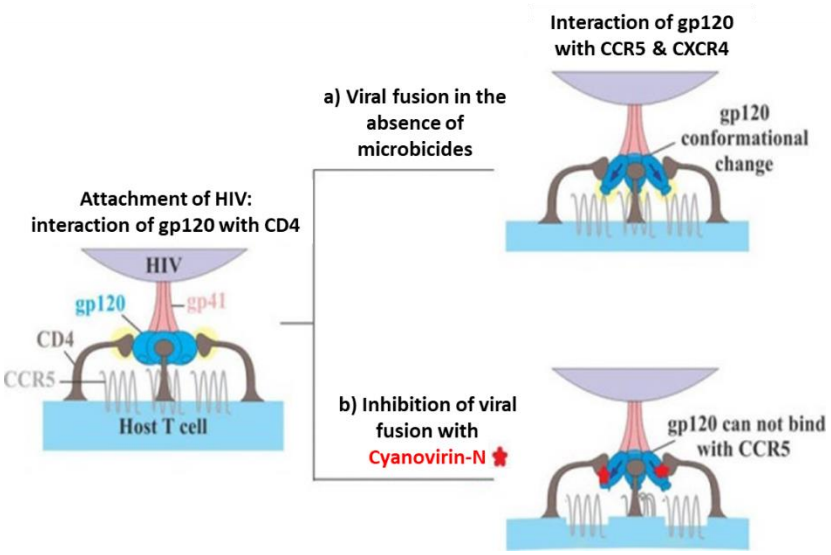


Fig. 6.1 The HIV entry process: a) interaction of HIV-gp120 with host cell CD4 T-cell and co-receptors (CCR5 and CXCR5) leads to conformational changes required for membrane fusion and viral entry; b) CVN prevents gp120 binding to CD4 and CCR5-adapted from Ref[20]

The ability of CVN to bind to high mannose glycans also makes it a promising candidate for use against coronaviruses such as SARS-CoV-2, however experimental evidence is still missing.[2, 22]

6.3.1. Structure

CVN is an 11 kDa peptide with unique sequence and structure. Wild type cyanovirin (wt CVN) is a peptide of 101 amino acids, it has a shape of an elongated prolate ellipsoid with length of ~55 Å and width ~25 Å. The sequence consists of two well-conserved 50 amino acid repeats and the three-dimensional structure forms two pseudosymmetric domains, the protein structure consists altogether of 10 β-strands and 4 short 3_{10} helical turns.[23] The sequence and structure of wt CVN are shown in **Fig. 6.2**.

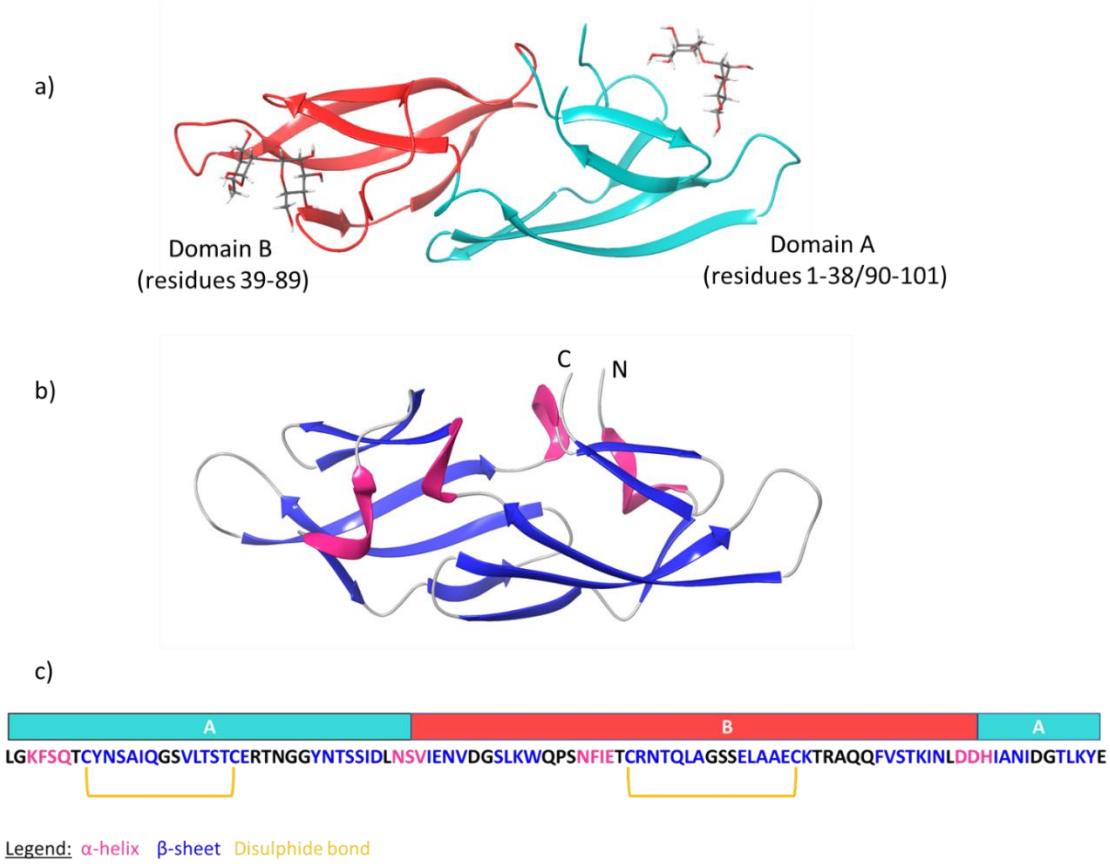


Fig. 6.2 Structure and sequence of wt CVN monomer (PDB: 1IIY) a) The two domains of CVN with bound Manα(1,2)Man highlighted in red (domain B) and cyan (domain A), domain B contains the high affinity carbohydrate binding site and domain A contains low affinity carbohydrate binding site- Manα(1,2)Man is bound in both sites b) Secondary structure of CVN displaying β-strands (blue) and α-helixes (pink), c) Sequence of CVN with structural information

Domain A is formed by amino acid residues 1-38 and 90-101 and thus contains both N- and C-terminals and has a higher degree of flexibility than domain B. There are several interactions important for the proper folding of the domain. Side chains of Phe 4, Leu 18, Ile 34 and Leu 36 form hydrophobic interactions with Leu 87, Ile 91 and Leu 98. The disulphide bridge is formed between Cys 8 and Cys 22 and Sγ atoms are in van der Waals contact with Phe 4, Asn 93 and Leu 98. A hydrogen bond is found between Asn 93 and Thr 7. A tightly bound water molecule serves to bridge hydrogen bonds between the backbone carbonyls of His 90 and Lys 99 and the backbone amide of Ala 92. The N- and C- terminus residues Leu 1 and Glu 101 are in van der Waals contact.[23]

Domain B is formed by amino acids 39-89 and does not contain any terminal.[24] Again, the folding is achieved by multiple interactions between the amino acid residues. Hydrophobic interactions are formed between the side chains of Phe 54, Leu 69, Ile 85 and Leu 87 packed against Leu 36, Ile 40 and Leu 47. Hydrogen bonding occurs between the side chains of Asn 42 and Thr 57. The disulphide bridge is formed between Cys 58 and Cys 73, additionally Sγ atoms form van der Waals contacts with Phe 54,

Asn 42 and Leu 47. Side chains of Thr 61 and Ala 71 are in van der Waals contact and pack against aromatic ring of Phe 54 providing an additional auxiliary element to the core.[23]

The two domains are joined together by two helical turns (residues 37-39 and residues 88-90). At the interface of the two domains there is a cluster of hydrophobic residues, consisting of residues Val 39, His 90, Trp 49 and Tyr 100. Trp 49 and Asp 89 form both a hydrophobic contact as well as a hydrogen bond.[23]

In solution CVN can exist as both a monomer and a domain-swapped dimer and it crystallizes exclusively as a domain-swapped dimer (**Fig. 6.3**). Wild-type dimer in solution is in metastable conformation and slowly converts into the more stable monomeric form.[25] Domain swapping is a term for an oligomerization mechanism in which two or more polypeptide chains exchange identical domains. The exchanged portion can be a single secondary structure element or an entire globular domain. Although the detailed pathway that leads to the domain-swapped dimer formation is not yet known for most proteins, it has been generally observed that destabilizing conditions such as low pH, organic solvents or chaotropic agents promote domain-swapping, especially at high protein concentrations. The domain swapping of CVN occurs through the extension of a hinge region (residues 50-53). Because conversion of the domain-swapped dimer into monomers is fast under conditions used to assay antiviral activity, it is still unclear if the dimeric state of the domain-swapped wild type CVN construct is biologically relevant or just a crystallization artifact. By introducing mutations into the hinge region of the protein it is possible to stabilize either the monomeric or dimeric form of CVN.[26]

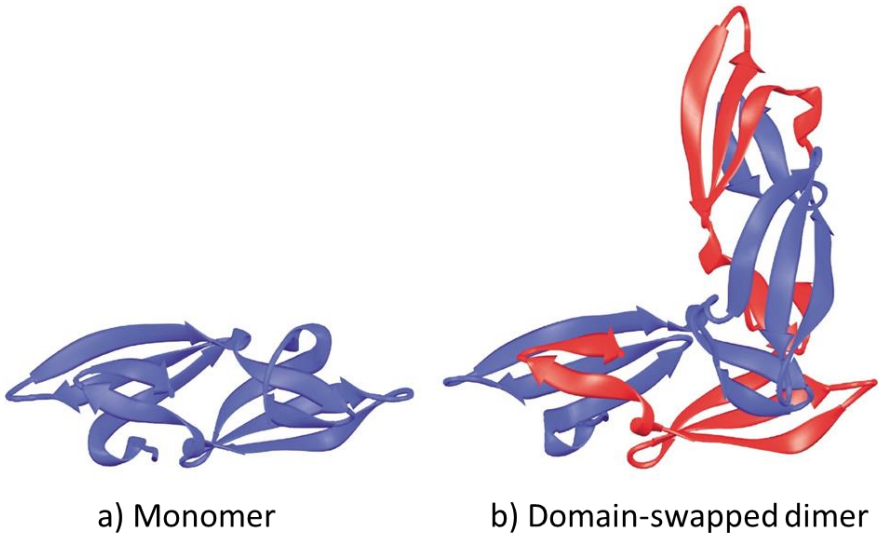


Fig 6.3 Ribbon representation of a) solution structure of monomeric wt CVN, b) X-Ray structure of domain-swapped dimer (different monomer units in the dimer are colored red and blue)- from Ref[25]

6.3.2. Glycan binding

While CVN has a broad anti-viral activity, it has a remarkable glycan specificity. It only binds to Man_9 and the D1D3 isomer of Man_8. The specificity of CVN to bind only to the D1D3 isomer of Man_8 and not to the D1D2 isomer, indicates that Manα(1,2)Man is the minimum epitope required for binding (**Fig. 6.4**).

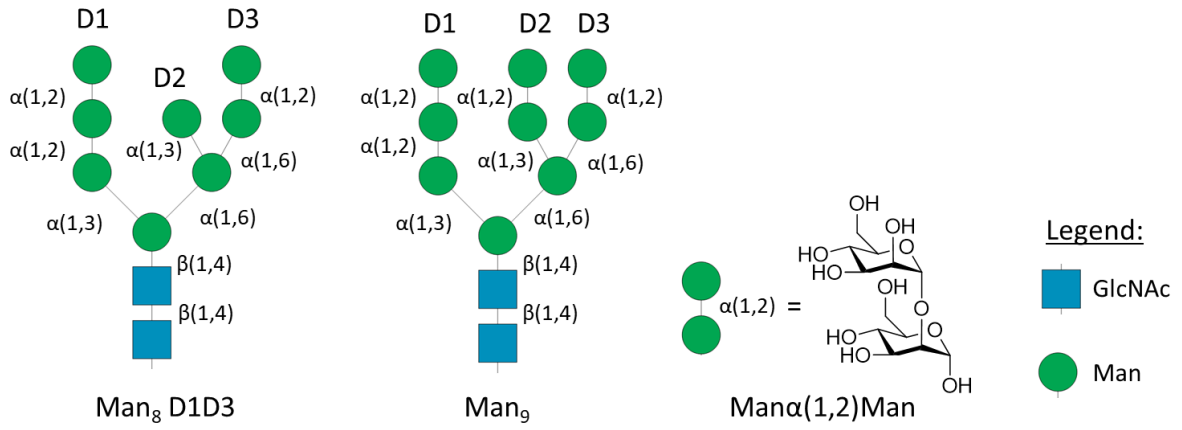


Fig. 6.4 Structures of glycans Man_8 (D1D3) and Man_9 and the minimal binding epitope required for binding of CVN Manα(1,2)Man

The CVN monomer consists of two independent carbohydrate binding sites (CBS), exhibiting slightly different affinities for di- and trisaccharides. Carbohydrate binding site on domain B (residues 39-89), also referred to as the high affinity site, consists of a deep pocket, while the low affinity carbohydrate binding site on domain A (residues 1-38 and 90-101) is a shallow semicircular cleft. The binding site on domain B has about a 10-fold higher affinity towards Manα(1,2)Man than the binding site on domain A. Biding of CVN to oligomannoses occurs with affinities in the μM range, while cell fusion assays showed that CVN binds gp120 in nM range.[16]

As mentioned above, the high affinity CBS on domain B of CVN forms a deep pocket. The Manα(1,2)Man in the wt CVN CBS on domain B is oriented so that the nonreducing mannopyranose ring is located in the top of the binding site as shown in **Fig. 6.5**. The hydroxyl groups of Manα(1,2)Man interact via potential hydrogen bonds (direct or water mediated) with polar or charged amino acid residues. The H-bonds are formed with carboxylate of Glu 41, hydroxyl of Ser 52, amide and carboxylate of Asn 53 and Glu 56. The methyl group of Thr 57 is in van der Waals contact with the C-6 protons of each pyranose ring and with H5' of nonreducing pyranose. Electrostatic interactions are formed with carbonyl oxygen atoms of Lys 74 and Thr 75, the amino group of Lys 74, the guanidinium group of Arg 76 and the amide of the side chain of Gln 78. Additionally, side chains of residues Asn 42 and Asp 44 potentially together with also the side chain of Asn 37 (located on domain A) contribute

to binding indirectly through forming H-bonds with residues that are in direct contact with the disaccharide.[16]

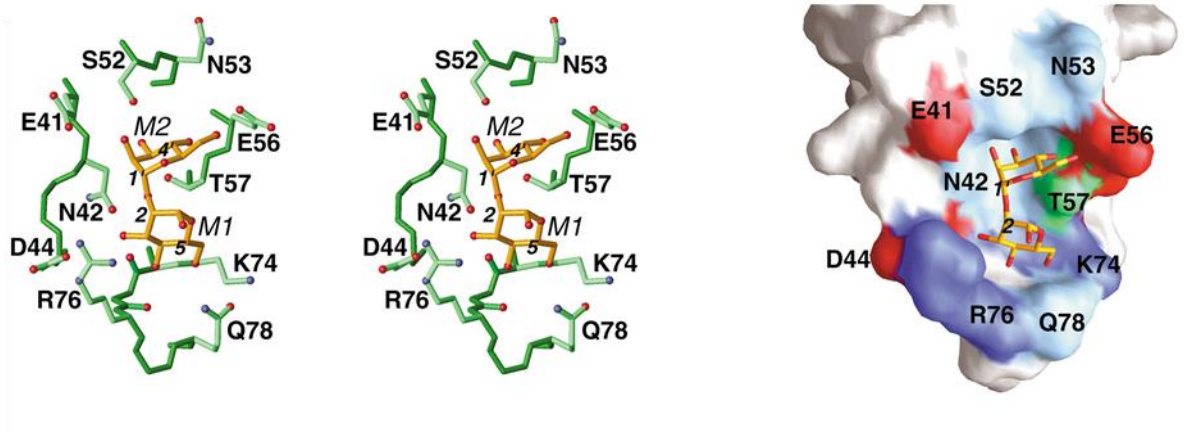


Fig. 6.5 The high affinity carbohydrate binding site on domain B of wt CVN with Manα(1,2)Man.[16]

The approximately 10x lower affinity of CBS on domain A is a consequence of the lack of key polar residues that are present in the high affinity binding site. Specifically, Glu 41, Ser 52 and Gln 78 that form the top of the binding site on domain B, are substituted by nonpolar residues Ala 92, Gly 2 and Gly 27 in the binding site on domain A, as shown in **Fig 6.6**. Additionally, Asp 95 in low affinity site is directed away from the carbohydrate binding site. Asp 95 corresponds to Asp 44 in high affinity site, which is involved in at least three potential hydrogen bonds with the disaccharide.[16]

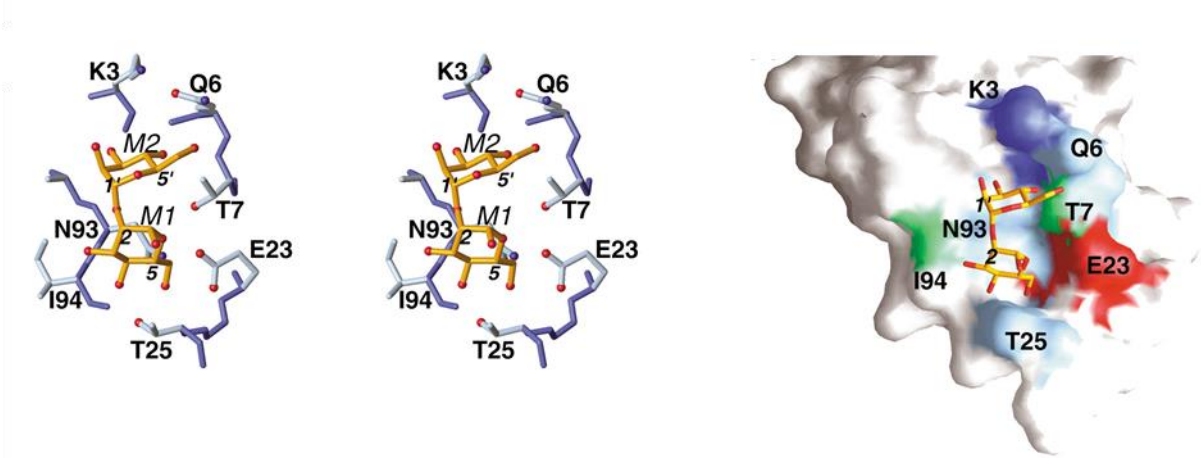


Fig. 6.6 The low affinity carbohydrate binding site on domain A of wt CVN with Manα(1,2)Man.[16]

The binding site on domain A exhibits a slight preference for the trimannose unit of the D1 arm of Man_9, while the binding site on domain B shows a preference for the dimannose unit of D3 arm of Man_9. Therefore, wild-type CVN interacts with Man_9 to create a polymeric, cross-linked aggregate. The distance between the D1 and D3 arms of Man_9 is around ~15 to ~17 Å, which suggests that the CVN

interacts with two different carbohydrate moieties on gp120, because the two CBS are separated by ~40 Å.[21]

Isothermal calorimetric titrations revealed that wt CVN binding to oligomannosides is largely driven by enthalpic contributions (negative ΔH values). This suggests that favourable binding contacts are mediated by polar/electrostatic interactions, van der Waals interactions and hydrogen bonds. On the other hand, binding is entropically disfavoured, because the sum total of binding entropy due to solvation effects and to the rotational, translational, and conformational freedoms of CVN and oligosaccharide is reduced upon binding. Studies focused on binding of CVN to Man_9 are severely limited by aggregation and/or precipitation of the complex. Nevertheless, Shenoy et al.[18] explored the binding of CVN with nonamannoside, an oligosaccharide that shares a close structural similarity to Man_9 apart from lacking the core chitobiose moiety. No precipitation and/or aggregation was observed when employing nonamannoside in the binding studies with CVN. Thus, it was possible to observe that binding of the nonamannoside to CBS binding site on domain B was largely driven by enthalpic contributions. On the other hand, entropy was responsible for binding to CBS on domain A. This was a particularly interesting finding, because binding to linear trimannoside was based on enthalpic contributions for both sites. The authors speculated that the tight binding between Man_9 and CVN that results in aggregation and precipitation, could be due to multisite and multivalent interactions. Multivalent binding is achieved by two molecules of CVN that are able to bind one molecule of Man_9. On the other hand, multisite binding is caused by two molecules of Man_9 binding to one molecule of CVN.[18] The principle of multivalent and multisite binding is summarized in **Fig. 6.7**.

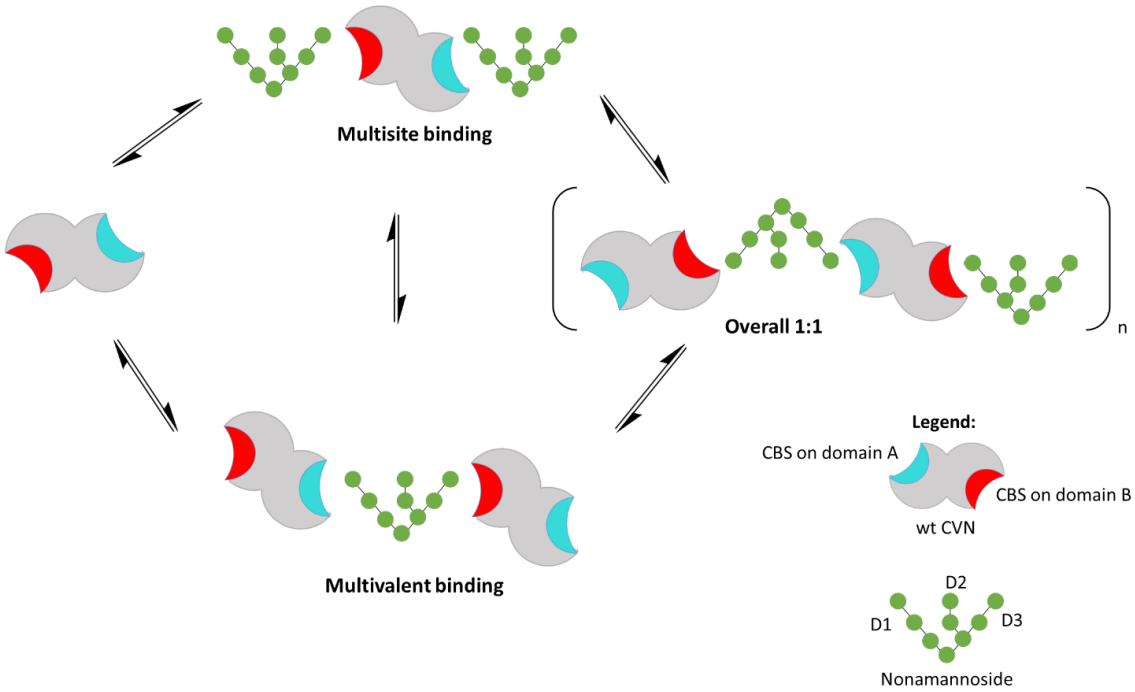


Fig. 6.7 Binding of CVN to nonamannoside, the principle of multivalent and multisite binding

6.3.3. Multisite binding

Multisite binding of CVN is essential for viral inhibition. Monomeric single site mutants of CVN have significantly lower binding affinity to gp120 and completely lose their antiviral activity, regardless to which site is inactivated. However, linking inactive monomeric proteins either via a head-to-head disulphide linkage or by domain swapping restores some of the antiviral activity, demonstrating that involvement of at least two binding sites is critical for anti-HIV activity of CVN.[21]

6.3.4. CVN derivatives

In previous studies, design of CVN derivatives was based on either increasing the number of binding sites to enhance the antiviral activity or reducing the cytotoxicity by preparing PEGylated derivatives.

While two carbohydrate binding sites are sufficient for antiviral activity, more binding sites also result in enhanced neutralization activity.[27] In the study by Keffe et al. CVN dimers containing tandem repeats of CVN were generated by linking the two copies of CVN through a flexible polypeptide linker. These proteins fold into a rigid structure resembling the 3D domain swapped dimer as proven by X-ray crystallography, shown in **Fig. 6.8**. Similarly, also trimers and tetramers were constructed. The addition of a second copy of CVN increased the potency of HIV neutralization, but the addition of the third or fourth copy did not have a significant impact. One possible explanation for this could be that the additional CVN copy may cause steric hindrance to some of the carbohydrate binding sites, since in the wild type CVN the C- and N-terminal are in close proximity and they are also close to the carbohydrate binding site on domain A. Another possible cause may be that nondomain-swapped CVN repeats do not have a significant impact on the activity. Adjustment of linker length may improve the potency by removing the possible steric issues.

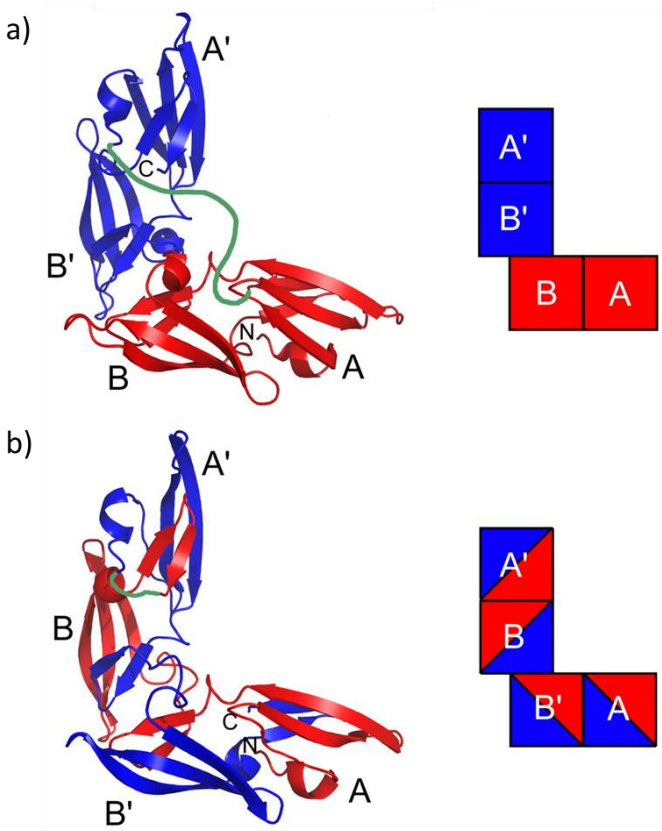


Fig. 6.8 Dimers of CVN created either by a) covalently linking the two copies of CVN or b) domain swapping dimerization-from Ref[27]

Similarly, a covalent dimer of CVN was designed in which two native domains were separated by the "nested" covalent insertion of two additional domains of CVN, yielding four possible carbohydrate binding sites, as shown in **Fig. 6.9**. This construct still displayed similar folding as the wt CVN, but the construct was more flexible, thus avoided steric constrains. This "nested" CVN had a slightly increased anti-HIV activity compared to the wild type CVN, presumably because flexibility allowed for optimal orientation of carbohydrate binding sites on the target gp120.[28]

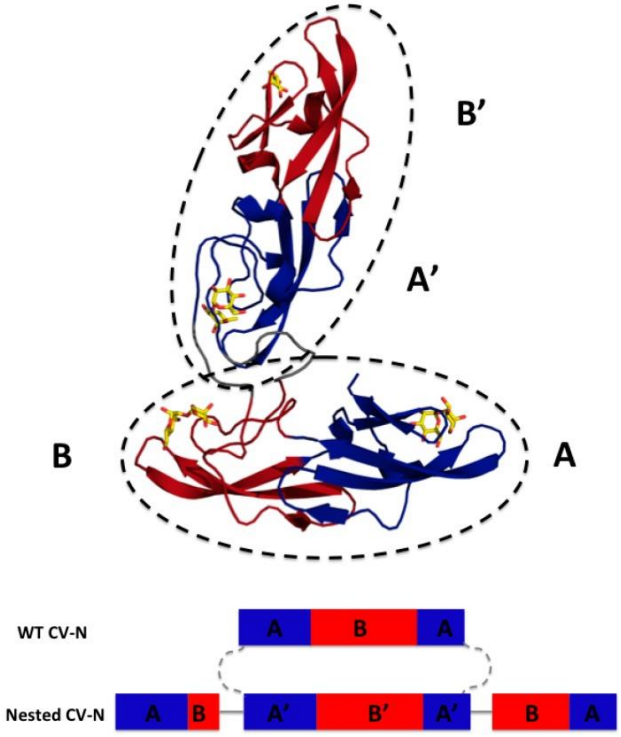


Fig. 6.9 "Nested" dimer of CVN-from Ref[28]

An alternative strategy in developing CVN derivatives was based on improving its safety as a therapeutic agent. Issues related to protein-based pharmaceuticals, such as rapid clearance, proteolytic digestion and immunological reactions all present major drawbacks for the development of CVN or its derivatives as therapeutics. PEGylation of the peptides can address these issues and stabilize the polypeptide structure, reduce immunogenicity, toxicity, renal excretion and prolong the plasma half-life. PEGylation of CVN was achieved through a PEG moiety attached to a flexible linker peptide introduced on the N-terminus of CVN. Because the N- and C-terminal are in close proximity to the carbohydrate binding site on domain A, introducing modifications is risky and in fact it resulted in lowered biological activity of the PEGylated construct. However, the construct also displayed reduced antigenicity and did not induce such strong IgG response as the wt CVN making it a promising and safe microbicide.[29]

6.3.5. Cyanovirin-N homolog Cyt-CVNH

A CVN homolog Cyt-CVNH was isolated from cyanobacterium *Cyanothece*. Its structure and glycan recognition profile are similar to CVN. However, no cross-linking and precipitation was observed upon Man_9 binding of this lectin, in contrast to CVN. Cyt-CVNH also consists of two carbohydrate binding sites, one on domain A and one on domain B. The two CBS however share the same affinities for binding to dimannoside, trimannoside and Man_9. Unlike CVN, both CBS of its homolog Cyt-CVNH

preferentially bind to trimannoside rather than dimannoside, with difference in affinity of about 6-fold. Therefore, Man_9 interacts with both CBS only through the D1 arm and this is the most likely reason that prevents cross-linking of Cyt-CVNH by Man_9. Finally, it was shown that Cyt-CVNH has about 4-times higher potency for HIV-1 inhibition than CVN.[30]

6.4. Designing new multisite constructs

As multisite binding is shown to be of crucial importance for the antiviral activity of CVN, we set out to design new multisite constructs. Our idea for the design of new constructs is based on the fact that previous multisite constructs proved that increasing the number of binding sites can enhance antiviral activity. Additionally, we took into account the structural properties of Cyt-CVNH lectin, able to avoid Man_9 cross-linking and still inhibit HIV-1 more potently than CVN. Therefore, the idea is to build flexible constructs containing multiple copies of a peptide that contains only the high affinity binding site (present on domain B) of CVN.

While single site mutants of CVN that only contain active site B were already investigated, our approach aimed towards finding the minimum peptide sequence sufficient to retain the ability to bind to small oligomannosides. Covalently connecting shorter sequence repeats of the peptide to one another would then allow for smaller peptide constructs, higher flexibility and less steric hindrance. Additionally, we predicted that peptide constructs containing copies of only one binding site would preferentially bind to D3 arm of Man_9 and thus avoid cross-linking upon binding.

The first goal in this study was to shorten the sequence of CVN to isolate domain B of CVN and investigate its folding and carbohydrate binding ability. In order to simplify the construction of the multisite systems, unnatural functional groups can be installed on amino acids (e.g. azido handle), which can be easily achieved with solid phase peptide synthesis. We therefore decided to compare the synthetic approach with recombinant production of the peptides containing domain B of CVN. The new peptide constructs are to be later evaluated for their antiviral activity and selectivity.

6.4.1. Cyanovirin-N production

The production of a recombinant CVN is a challenge that has been addressed many times over the past two decades. As CVN could be a promising microbicide candidate, the focus was usually set on improving the yield and purity of the expressed protein while keeping low cost. A 5 mg dose of CVN as microbicide for topical application was proven to be effective, which means that supplying 10 million women with a twice weekly dose would require production capacity of 5000 kg a year.[31]

Both bacterial and eukaryotes expression systems were investigated (summarized in **Fig. 6.10**) and their advantages and disadvantages were put on the scale.

Bacterial expression systems

E. coli is generally the most widely used organism for recombinant protein production with several advantages: it has a short time of generation, it is easy to handle, the fermentation process is well known and it has high specific yields. It is therefore no surprise that the recombinant CVN (rCVN) was first produced by Boyd at al. in the periplasmic space of *E. coli* using expression vector pFLAG-1.[15] However, yields were low and purity was not sufficient for large-scale operations. The yield was improved significantly by inclusion body expression in the cytoplasm of E. coli.[32] Unfortunately, the final product was obtained as a mixture of full length monomeric and dimeric rCVN and isomers with deleted one or four N-terminal residues. Later, chaperone-fused expression system with hexahistidine (HisTag) and small ubiquitin-related modifier (SUMO) double-tagged CVN gave an intact and native rCVN that could be rapidly purified in a soluble and biologically active form.[33] Recently a new *E. coli* strain was investigated for rCVN production, so called SHuffle® T7 Express lysY. With this strain it was possible to obtain a soluble and bioactive CVN in high yield with easy purification process.[34]

Several attempts were also made to express CVN in other bacteria. For example, CVN was expressed in commensal *Streptococcus gordonii* as excreted protein or attached to the bacterial cell surface to capture HIV-1.[35] *Lactobacillus jensenii* was also investigated as an agent of mucosal drug delivery to secrete CVN. Other attractive commensal bacteria for the excretion of rCVN are also *Lactococcus lactis* and *Lactobacillus plantarum*.[20]

Eucaryotes expression systems

Advantages of yeast expression systems such as *Pichia pastoris* are that they can also be cultured cheaply and rapidly, possess certain posttranslational modification pathways and are able to secrete more efficiently. However, low yield, glycosylation and formation of dimeric aggregates which result in loss of activity make *P. pastoris* a less prospective expression system for production of recombinant CVN.[36]

Transgenic plants on the other hand have more potential to be used for large scale production, as they are easily prepared for production and economically scaled up. *Nicotiana tabacum* was proven to be a suitable system for expression of recombinant CVN, producing the preferred monomeric isomers of rCVN able to bind to HIV-1.[37] An interesting approach was taken by Drake et al. in 2013, they generated a wild type of marshmallow plant (*Althaea officinalis L.*) with transgenic roots by *Agrobacterium rhizogenes* for recombinant CVN expression. Transgenic soybean has also been used, giving rCVN and soybean oil that both met the criteria to be used as anti-HIV microbicides.[20]

Thus, bacteria, yeast and transgenic plants were all explored as expression systems for recombinant CVN production. While plants present the most future potential for a large scale production, *E. coli* soluble expression of rCVN in the cytoplasm remains the most feasible option at the moment.

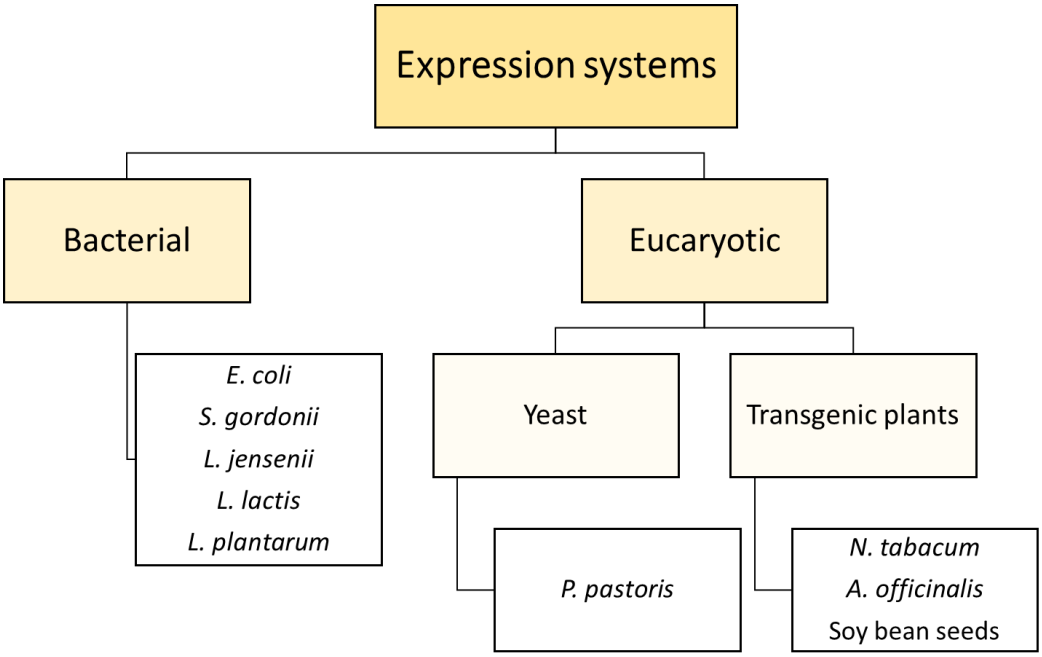


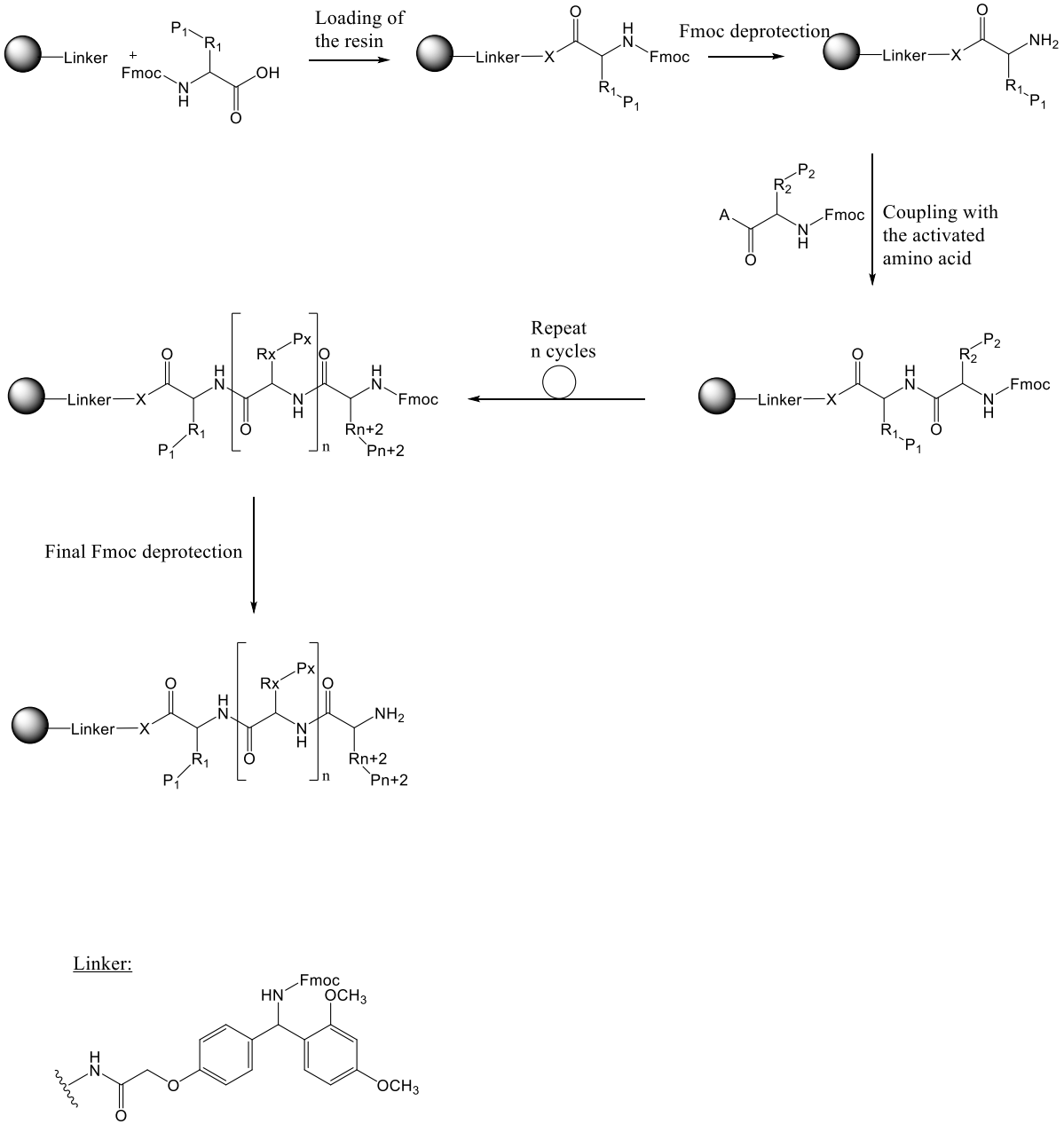
Fig. 6.10 Previously explored expression systems for production of rCVN

Synthesis and expression of CVN domain B

While most work has been focused on optimizing expression of CVN suitable for large scale production, our approach in this study was focused on utilizing the methodology that would allow for simple introduction of mutations as well as installing sites for further conjugation of a peptide construct to scaffold for multisite presentation.

Therefore, we investigated the possibility to make the entire B domain of CVN synthetically with solid phase peptide synthesis. Peptide synthesis has come a long way since 1901 when the first synthesis of a dipeptide (glycyl glycine) was reported by Fischer. The revolution began in 1963 when Merrifield described the first solid-phase synthesis (SPPS) of a peptide and within a few years developed an instrument for automated synthesis of peptides. Nowadays automated SPPS is a common strategy used for synthesis of peptides. Typically, an amino acid with both α-amino group and side chain protection is immobilized on the resin. The α-amino protecting group is usually an acid sensitive tert-butoxycarbonyl (Boc) or a base-sensitive 9-fluorenylmethyloxycarbonyl (Fmoc) group. In a deprotection step the α-amino protecting group is removed and washed from the resin. Then a protected amino acid with an activated carboxyl group can be coupled to the unprotected resin-bound amine, representation is shown in **Scheme 6.1**. To ensure the completion of the reactions the

activated amino acid is used in excess. To complete the sequence, the cycle of deprotection and coupling is repeated. The peptide synthesis is carried from C-terminal to N-terminal. Finally, the peptide is obtained after side chain deprotection and cleavage from the resin.



Scheme 6.1 Principle of solid phase peptide synthesis (Fmoc SPPS approach), structure of Rink Amide linker used in our study

The SPPS technology certainly allows for easy modifications in the amino acid sequence as well as installing unnatural features on amino acids (e.g. azido or alkyne handle). The sequence of the domain B consists of 51 amino acids containing the amino acid residues 39-89 of CVN sequence. The sequence of domain B is shown in **Fig. 6.11**. For the purpose of building constructs presenting multiple copies of domain B, a possible method is to introduce Lys-N_3 in the amino acid sequence. Looking at the

sequence and structure of the domain B peptide, it contains three Lys residues. Two of those are close to the CBS, while Lys 84 (numbering based on the full sequence of wt CVN) is directed away from the binding site and is in close proximity to the C-terminal of the domain B peptide (shown in **Fig. 6.11**). Due to the length of the peptide sequence it is more economical to introduce the unnatural amino acid close to the C-terminal, where the synthesis starts from, rather than at the N-terminal, where couplings might be less effective. Another possibility is to introduce an additional Lys-N_3 in the start of the synthesis, that is on C-terminal.

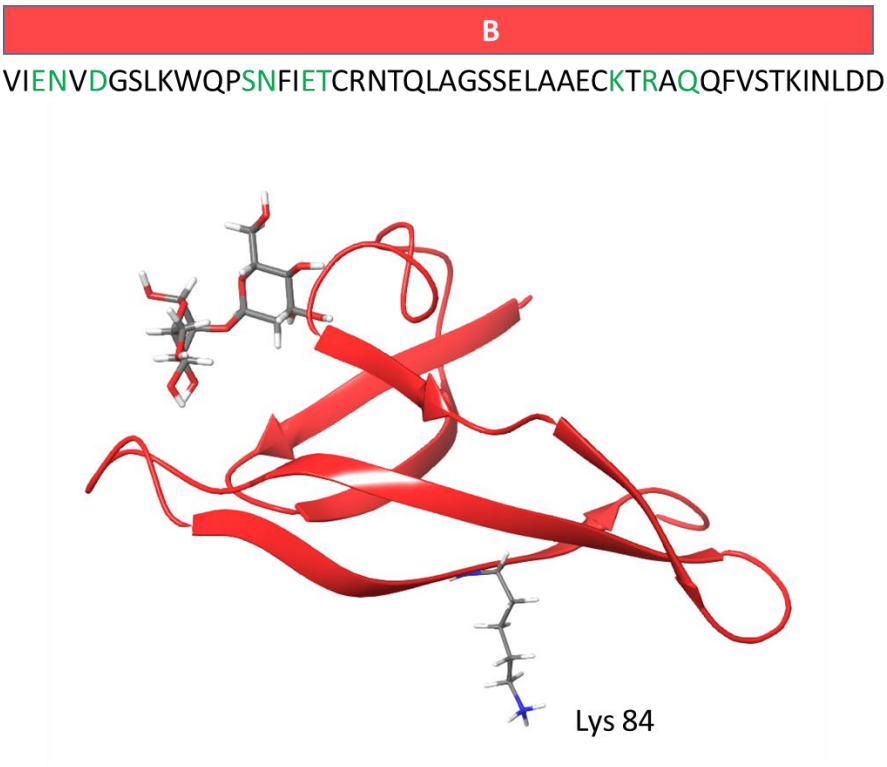


Fig. 6.11 Domain B of CVN: sequence (AA residues involved in binding with Manα(1,2)Man are highlighted in green) and structure (side chain of Lys 84 that could be used for conjugation of multisite constructs is shown)

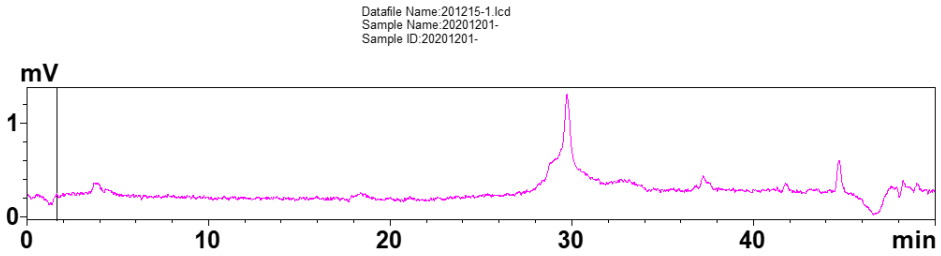
The 51 amino acid peptide **6.1** (sequence of wt CVN domain B, shown in **Fig 6.12**) was synthesized on 25 µmol scale using a microwave-assisted peptide synthesizer (CEM HT12 Liberty Blue) following a Fmoc SPPS approach. As the solid support, RAPP polymere tentagel S RAM resin functionalized with a modified Rink Amide linker (shown in **Scheme 6.1**) was chosen. Coupling was performed with DIC/Oxyma at 90°C and the coupling cycle was repeated twice. After each coupling step the resin was treated with 10% Ac_2O to cap the uncoupled free amine. After cleavage from the resin, the peptide mixture was purified with preparative HPLC and a careful check of the fractions by MS allowed us to find the final peptide (2.7 mg which corresponds to cca. 1.9% yield). The Ac_2O capping after each coupling step proved to be necessary to reduce the number of sequence combinations resulting from

the unfinished couplings. Due to the length of the peptide, without capping it was impossible to find the peptide in a very complex mixture even after HPLC purification. **Fig. 6.12** shows the LC-MS trace of the purified peptide **6.1**, the UV signal and the MS-ESI of the peptide. The found m/z peaks 1419.7 and 1136.0 correspond to $(M+Na)^{+4}$ and $(M+Na)^{+5}$ of the peptide with calculated m/z 5652.79.

a) Sequence

VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD

b) LC-MS UV trace



c) MS

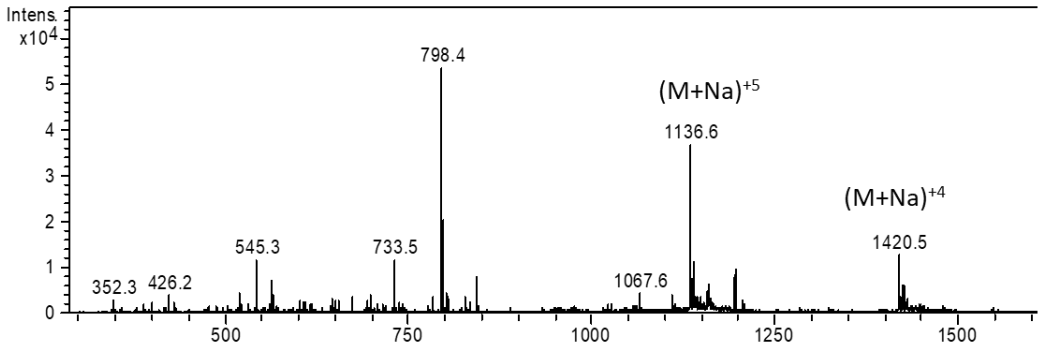


Fig. 6.12 Characterization of peptide **6.1** a) Sequence, b) UV trace of the peptide after preparative HPLC purification and c) MS-ESI profile.

Careful check of the fractions obtained after HPLC with MS-ESI allowed us to also identify motifs in the sequence that are more difficult to couple, these motifs are highlighted in **Fig. 6.13**. The difficult motifs were identified in the middle of the sequence, suggesting that a possible reason could be the formation of a secondary structure on the resin. To further optimize the synthesis therefore a number of different approaches could be tried, such as using chaotropic salts, use of PEG-based polar resins and a choice of amino acids with side-chain protecting groups that disrupt the secondary structure formation, for example substitution of Ser(tBu) or Thr(tBu) with the Trt derivatives of Lys(Boc) with Lys(Tfa).

```
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLDD
```

Fig. 6.13 Sequence of the synthesized peptide **6.1** with motifs that were identified as more difficult to couple highlighted in yellow

As a comparison, we also produced the B domain of CVN with protein expression from *E. coli*. While somewhat more complicated through the eyes of a chemist, also recombinant production allows incorporation of modifications to facilitate conjugation of peptides, an example is a SpyTag/SpyCatcher system.[38]

Recombinant production of the peptide containing only domain B of CVN could be achieved by either expressing only domain B itself or by expressing the entire peptide with cleavage sites between the two domains and later cleave out the B domain. We decided to use the second approach, assuming that the proximity of the amino acid residues from A domain would help proper folding of the peptide. The construct of the peptide was designed so that it would also simplify the purification process. Therefore, we inserted a short sequence for two cleavage sites between domains A and B in the sequence of wt CVN (sequence of the full peptide construct **6.2** shown in **Fig. 6.14**). Additionally, both the N- and C-terminal were labeled with a HisTag to allow purification with the affinity column, the principle of purification is shown in **Fig. 6.16** and explained in detail below. The final peptide **6.3** containing domain B obtained by recombinant production included short sequences remaining from the cleavage sites on the N- and C-terminal and was thus 7 amino acids longer than the synthesized peptide, consisting of 58 amino acids in total (6.5 kDa).

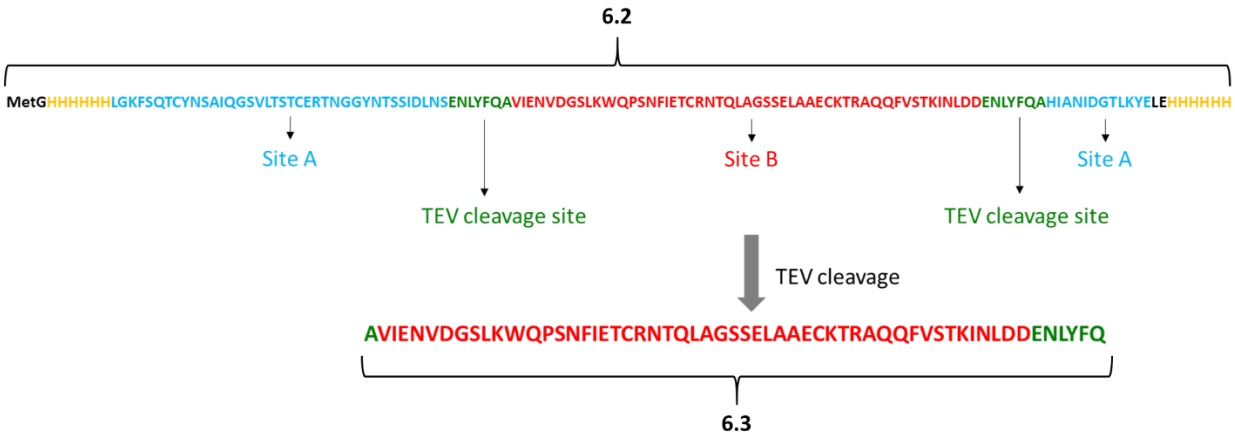


Fig. 6.14 Sequence of full peptide construct **6.2** and expressed domain B peptide **6.3**

We used *E. coli* BL-21 as host strain and pET-26b(+) as vector for expression of the recombinant CVN with two TEV cleavage sites (**6.2**). The composition of the commercial plasmid is shown in **Fig. 6.15**.

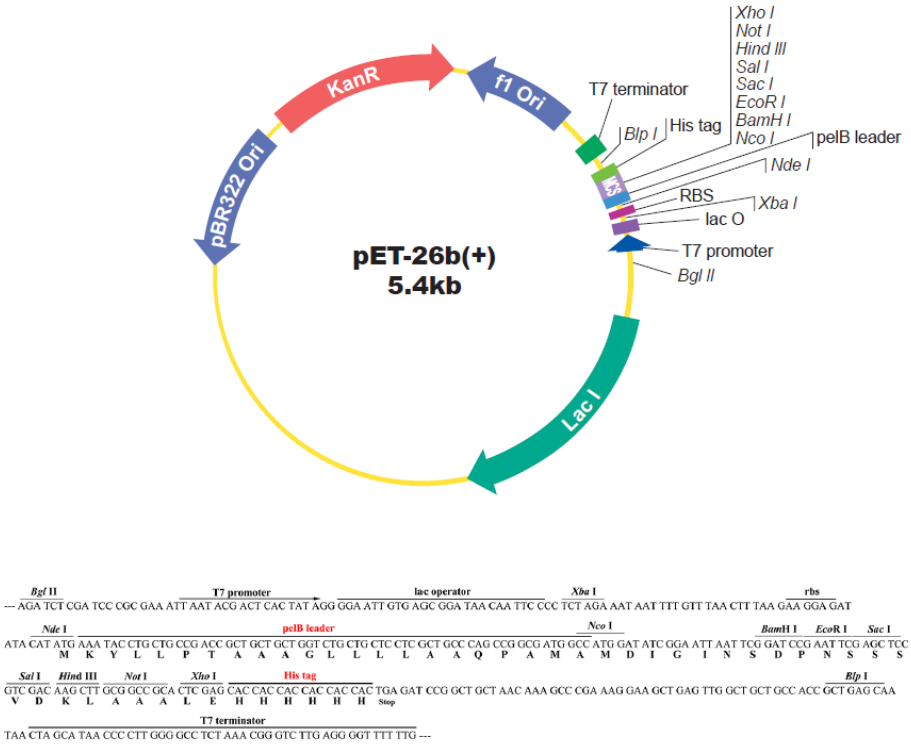


Fig. 6.15 Plasmid pET-26b(+) used for expression of **6.2**

After the expression from bacteria (400 mL cell culture) and cell lysis with sonification, the supernatant containing recombinant **6.2** was purified on Ni sepharose column. Due to the HisTags, the protein **6.2** bound to the column and was then eluted with 50 mM imidazole. Three 2 mL fractions containing 0.753 mg/mL, 0.801 mg/mL and 0.697 mg/mL of peptide/protein were obtained and analyzed by SDS-PAGE electrophoresis (10% acrylamide, Coomassie Blue stain), the result is shown in **Fig. 6.17, gel A**.

The SDS-PAGE gel shows that the three fractions contained mainly the peptide of the appropriate size (15 kDa).

In the second step, a cleavage was performed to "cut" the full peptide construct **6.2** into a peptide containing domain B (no HisTag) and two shorter peptides containing the two parts of domain A (both labeled with HisTags). The cleavage was performed by incubating the peptide (fraction 2, 2 mL, concentration 0.801 mg/mL) with tobacco etch virus (TEV) protease for 24 h at 30°C. Then the peptide containing domain B was separated from the mixture with another Ni sepharose affinity column. The N- and C- terminal parts containing the sequence of domain A were labeled with HisTag and therefore bound to the column. On the other hand, the peptide consisting of domain B did not have a HisTag label and was eluted from the column immediately with a wash with PBS buffer, as shown in **Fig. 6.16**. The TEV protease was also labeled with HisTag and was thus separated from the domain B peptide in the same purification step. The fraction collected from Ni-sepharose column containing the B domain peptide **6.3** was then concentrated and extensively washed with 10 mM phosphate buffered saline (PBS). The washing was performed with ultrafiltration using the Amicon Ultra spin filters with 3 kDa cut-off. Finally, 400 µL of the peptide sample was obtained with the concentration 0.425 mg/mL.

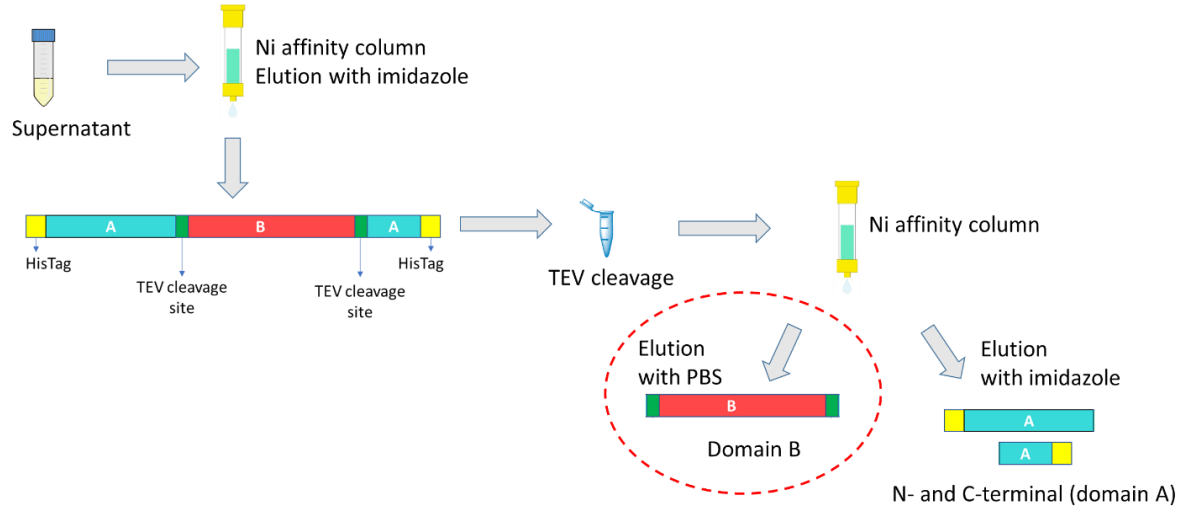


Fig 6.16 Principle of purification of the full peptide construct **6.2** and domain B peptide **6.3**

The SDS-PAGE gel (15% acrylamide, Coomassie blue stain) shown in **Fig. 6.17** (gel B) showed that the fraction eluted form the Ni affinity column with 10 mM PBS (loading 3) in fact contained a peptide of similar size as the synthesized peptide **6.1** (comparing loadings 1 and 3). On the same gel shown in **Fig. 6.17**, loading 4 contained fraction eluted with 50 mM imidazole from the second affinity column and it is visible that there were smaller peptides present corresponding to the cleaved-off N- and C-terminal peptides containing domain A. On the other hand, comparing loading 4 with loading 2 that

contained full peptide **6.2** before cleavage shows that there was still some non-cleaved peptide **6.2** present in loading 4, suggesting that TEV protease cleavage was not complete.

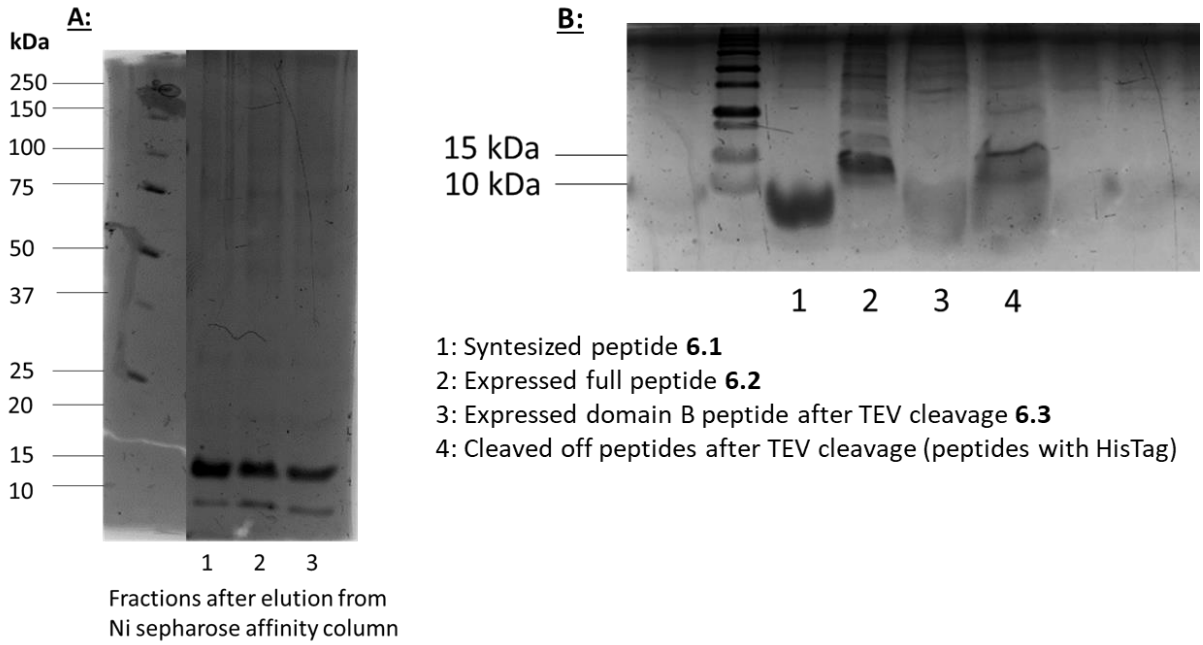


Fig. 6.17 SDS-PAGE gels: A (10% acrylamide, Coomassie Blue stain): Three fractions eluted from Ni affinity column of full peptide construct **6.2**, B (15% acrylamide stain, Coomassie Blue): All peptide constructs 1: synthesized peptide **6.1** (5.7 kDa), 2: expressed full construct **6.2** (15 kDa), 3: expressed domain B peptide **6.3** (6.5 kDa), 4: fraction eluted from Ni affinity column with imidazole after TEV protease cleavage

When comparing the two approaches from the operational point of view, the synthesis of the peptide was faster (coupling steps were finished overnight, cleavage and HPLC purification finished in half-day), with better purity and easier characterization of the final product.

6.4.2. Secondary structure studies

Folding of proteins is achieved through folding pathways that direct the protein into the lowest energy thermodynamic state.[39] The thermodynamic hypothesis of protein folding states that "*the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature and other) is the one in which the Gibbs free energy of the whole system is the lowest; that is, the native conformation is determined by the totality of the interatomic interactions and hence by the amino acid sequence, in a given environment*".[40]

The hydrophobic effect is the main thermodynamic driving force of protein folding, it leads to non-polar interactions within the protein core. The unfolded state is stabilized by a high conformational

entropy and the loss of it is the major opposing force in the folding process. The unfolded peptide exposes a high surface area of non-polar side chains to water molecules which decreases the water H-bonding network and creates an energetically unfavorable state. To minimize this, the water molecules then become more ordered around the exposed hydrophobic groups and thus there is a loss of energy of the protein-water system and the solvent's entropy around the protein. This is then counteracted by associating the nonpolar groups and separating them from the aqueous environment. The hydrophobic cores are shielded from the interactions with water by wrapping the polypeptide chain around them. In a nutshell, the hydrophobic effect is of entropic nature and it is a decrease in an unfavorable energetic state that involves water molecules interacting with polypeptide, which results in its folding.[41]

Different techniques can be used to investigate protein folding, such as X-ray crystallography, nuclear magnetic resonance (NMR), circular dichroism (CD), etc.

We investigated the secondary structure of produced peptides by circular dichroism. The principle of this technique is the measurement of the difference in absorbance for left and right circularly polarized light. Absorbance is related to the molar extinction coefficient, the path length and the protein concentration by Beer-Lambert law.[42]

In the far UV region (240 nm to 190 or 180 nm), the peptide bond is the principal absorbing group and studies in this region can give information on the secondary structure of the proteins. CD is also a good technique to monitor conformational changes of proteins that result from changes in experimental conditions, such as changes in pH, temperature, additives, etc.[43]

The advantages of using CD compared to X-ray crystallography and high resolution NMR are the speed and convenience of the technique. It requires only small amounts of non-labeled material that can be recovered, preparation of samples and measurements are fast and easy. The limitation of the technique is however that it only provides low resolution structural information. The CD measurement can give estimation on the secondary structure content of a protein, but does not indicate which regions of a protein are of which structural type.[43]

For our purposes CD was the technique of choice, because it allowed for a rapid evaluation of the secondary structure of the produced peptides compared to reference wild type CVN protein and to monitor the structural changes derived from a change of experimental conditions. The results of the CD measurements can be analyzed by simply comparing them with the reference peptide spectra and analyzing the characteristic feature of the spectra. Alternatively, different tools for analysis of CD spectra were developed.

Fig. 6.18 shows an ideal representation of characteristic features of three common secondary structure motifs (α-helix, β-sheet and random coil). Characteristic features of all-α proteins are an intense negative band with two peaks (208 nm and 222 nm) and a strong positive band (191-193 nm). All-β proteins have a single negative band (210-225 nm) and a stronger positive band (190-200 nm) with intensities significantly lower than for α-helixes. For proteins with a combination of α-helix and β-sheet structures, spectra are generally dominated by the α-helical component and often show bands at 222, 208 and 190-195 nm. In some cases, however there may be a single broad minimum between 210 and 220 nm. Unordered and denatured proteins have a strong negative band at 195-200 nm and a much weaker band (positive or negative) between 215 and 230 nm.[42]

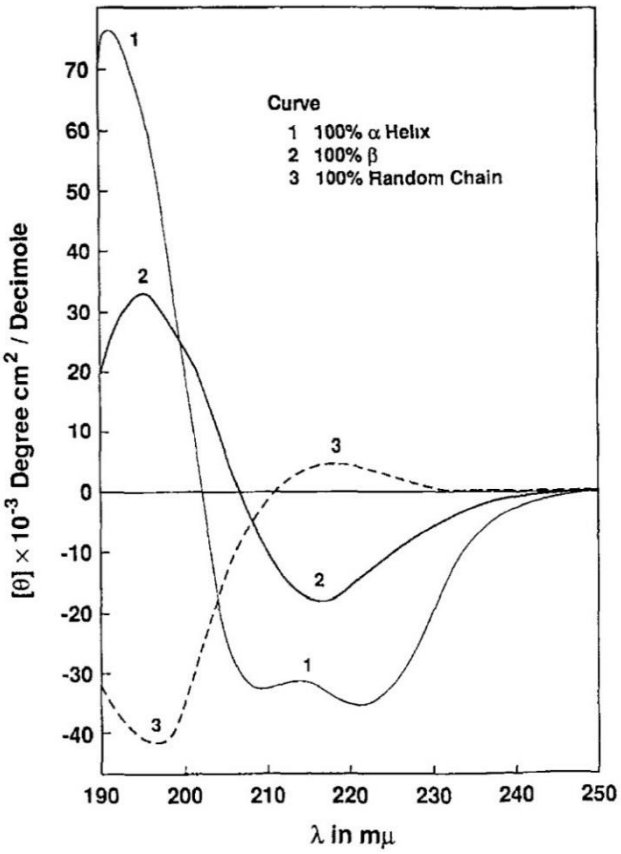


Fig. 6.18 Representation of CD spectra of 3 common secondary structure motifs: 1: All-α helical protein, 2: All β-sheet protein, 3: Random coil structure-from ref[44]

The dominant secondary structure elements of CVN are β-strands and the CD spectra of the wt CVN as well as some of its derivatives have been previously reported. One of such spectra is shown in **Fig. 6.19** and it clearly displays a minimum at around 212 nm and maximum at around 192 nm.[28]

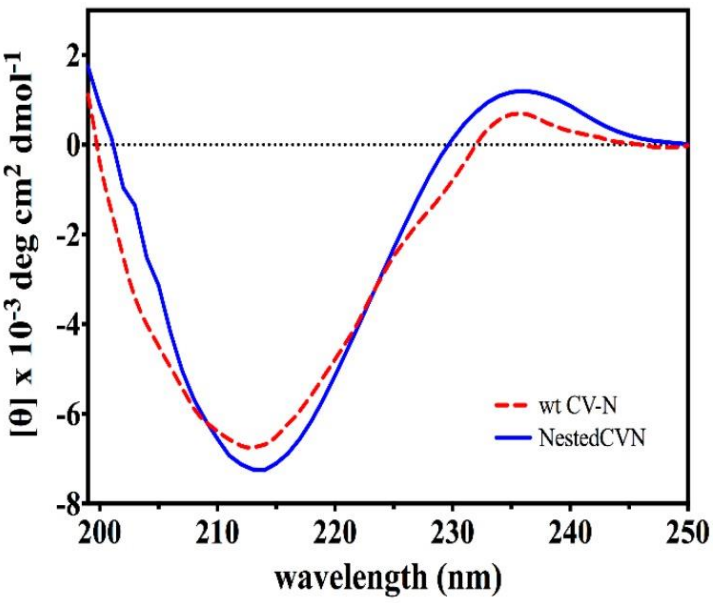


Fig. 6.19 CD-spectra of wt CVN (red line) and Nested dimer CVN (blue line)[28]

We first looked into the secondary structure of the synthesized peptide **6.1**. Initially we recorded a CD spectrum of the sample of **6.1** peptide after HPLC purification and lyophilization simply dissolved in 10 mM PBS (peptide concentration 63.6 µM). The spectrum shows a strong minimum with two peaks at ~208 nm and ~220 nm, which indicate presence of α-helical structure (**Fig. 6.20, entry 1**).

Different procedures for preparation of CVN are reported in literature, therefore we then set to investigate the influence of different conditions and refolding procedures on the secondary structure of the synthesized peptide **6.1**.

To achieve the refolding of a peptide, the secondary structure has to first be disrupted. Unfolding of the peptide can be induced physically (temperature, hydrostatic pressure) or chemically (by addition of chaotropic agents such as urea, GdnHCl, MgCl, alcohols and detergents).[39]

Chemical denaturation with GdnHCl works on a principle that GdnHCl acts as a chaotropic agent by disrupting the network of hydrogen bonds between water molecules, weakening the hydrophobic effect and consequently reducing the stability of the folded state. We tried refolding the peptide by first completely denaturing the structure with 6 M GdnHCl and 5 mM DTT to reduce any possible disulphide bonds already formed. This was followed by a buffer exchange to 2M GdnHCl and 50 mM TrisHCl, then buffer exchange to 10 mM TrisHCl and 100 mM NaCl (refolding procedure 1)[45]. Due to the size of the peptide and quantity of the material, the refolding was performed with ultrafiltration with Amicon Ultra spinfilters with 3 kDa cut off. The CD spectra of **6.1** peptide sample (peptide concentration 15 µM) obtained after this refolding procedure however still showed mainly the α-helical form with two strong minimums at ~205 nm and ~225 nm (**Fig. 6.20, entry 2**).

Since the β-sheet formation and thus wt CVN-like secondary structure was not induced with this complete refolding approach (refolding procedure 1), we tried another approach in which we left out the first step of complete denaturation and treated the peptide only with 2M GdnHCl and 50 mM TrisHCl, followed by buffer exchange to 100 mM NaCl and 10 mM TrisHCl (refolding procedure 2). Low concentrations of GdnHCl (1-2 M) can be used in disulphide-coupled peptide and protein folding to give the correct conformation of proteins. GdnHCl in low concentrations stabilizes the protein conformation by contributing to charge shielding and suppression of inter- and intramolecular interactions.[46] However, also in the sample of **6.1** peptide after the refolding procedure 2 (peptide concentration 26.5 µM) we saw a strong minimum at ~208 nm and somewhat less intense negative peak at ~224 nm, showing that the structure still remained as mainly α-helical (**Fig. 6.20, entry 3**).

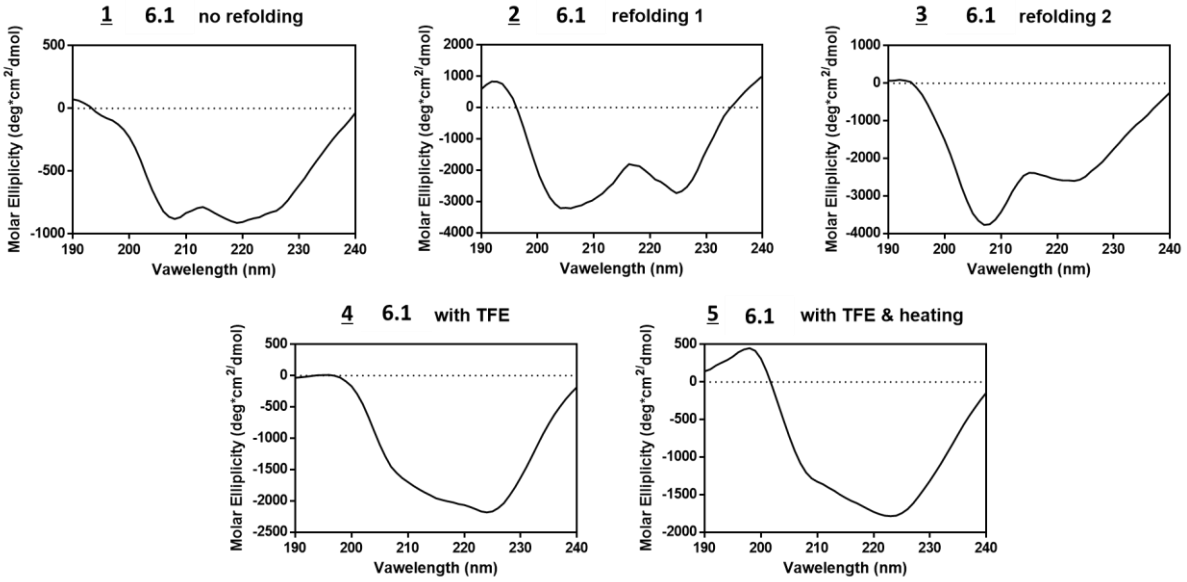


Fig. 6.20 CD spectra of the synthesized peptide **6.1** in 10 mM PBS at 20°C obtained with different preparation procedures 1: Prepared by dissolving peptide in 10 mM PBS (63.6 µM), 2: After refolding procedure 1 (15.0 µM), 3: After refolding procedure 2 (26.5 µM), 4: Peptide dissolved in 10 mM PBS, addition of 20% TFE (63.6 µM), 5: Peptide dissolved in 10 mM PBS, addition of 20% TFE, then heating to 45°C for 30 min and cooling to RT (63.6 µM)

To confirm that the structures of the peptide **6.1** with no refolding procedure and after the refolding procedure 2 are essentially very similar, we used ion mobility-mass spectrometry to compare the overall shape of the peptides. In fact, the result obtained with both samples was identical, as shown in **Fig. 6.21**. In addition, in this step we used HRMS to check m/z of the peptides in these two samples. As expected, the experimental m/z was the same for both samples, found m/z 1139.5609 and 1139.5614. The experimental m/z seems to correspond to $(M+Na+H_2O+H)^{+5}$ of a peptide with a disulphide bond between the two Cys residues (calculated m/z 5650.78). Additionally m/z 1424.2010 was found in both samples which also corresponds to $(M+Na+H_2O+H)^{+4}$ of a peptide with calculated

m/z 5650.78. We expected that the disulphide bond between the two Cys residues would form spontaneously with peptide being exposed to O_2 from the atmosphere, while it was stored for prolonged time (overnight) in PBS or H_2O or during refolding procedures.

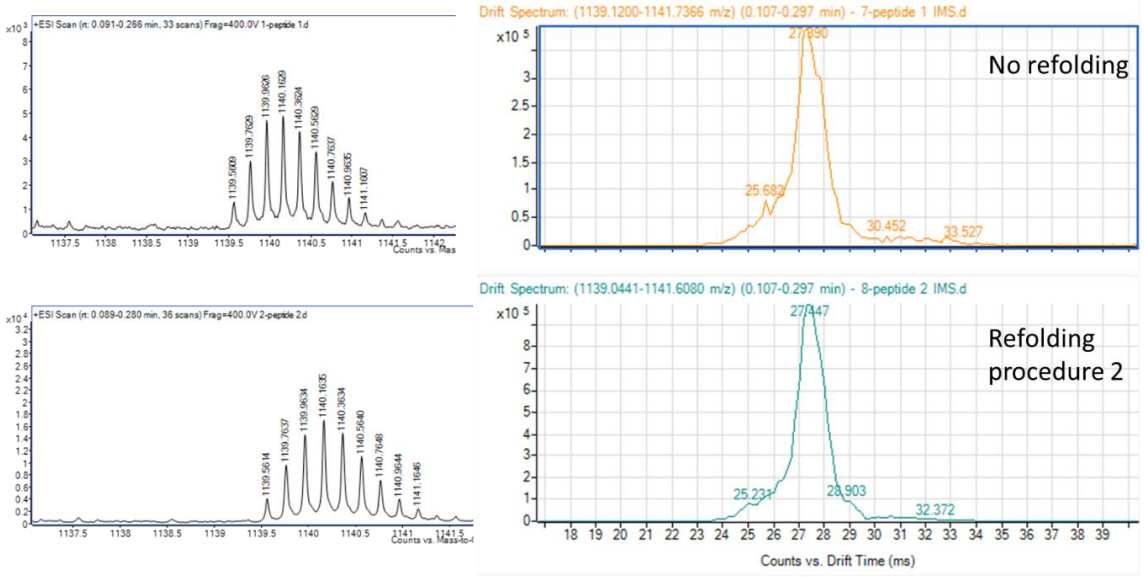


Fig. 6.21 Left: HRMS (M^{+5}) of the peptide **6.1** in the samples without refolding and after refolding procedure 2, Right: IM-MS spectra of the peptide **6.1** without refolding (upper right panel) and after the refolding procedure 2 (bottom right panel)

Addition of alcohols, such as 2,2,2-trifluoroethanol (TFE), promotes local H-bond interactions and can induce secondary structure formation, either α-helixes or β-sheets. Therefore, we investigated the effect of the addition of 20% TFE. The CD spectra of the **6.1** peptide sample with 20% TFE displayed (peptide concentration 63.6 µM) a strong negative band between 210 nm and 225 nm which could indicate a higher β-sheet content in the structure (**Fig. 6.20, entry 4**).

Finally, we also tried to denature the peptide sample physically, by elevating the temperature of the peptide sample (peptide concentration 63.5 µM) with added TFE to 45°C for 30 min and then cooling it down to room temperature, the CD spectra observed was almost identical to that of the spectra without heating (**Fig. 6.20, entry 5**).

As a comparison we also investigated the secondary structure of the expressed full peptide **6.2** and cleaved off B domain peptide **6.3**.

The CD spectrum of full peptide construct **6.2** obtained after affinity column with no refolding (only buffer exchange to 10 mM PBS, protein concentration 8 µM) shows a mixture of α-helical and β-sheet structure, displaying minimums at 208 nm and 218 nm (**Fig. 6.22, entry 1**). After treatment with 2M GdnHCl and 50 mM TrisHCl and then buffer exchange to 10 mM TrisHCl and 100 mM NaCl (refolding

procedure 2, protein concentration 7.3 µM) the structure seems to have more β-sheet form, displaying a minimum at 219 nm (**Fig. 6.22, entry 2**).

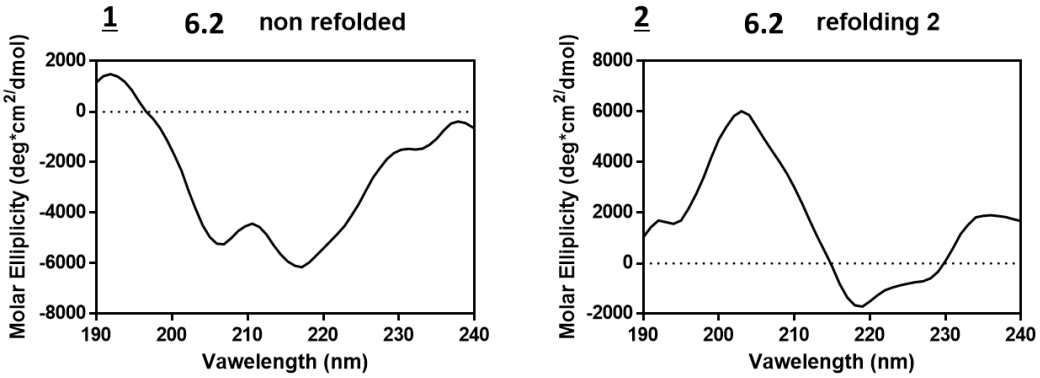


Fig. 6.22 CD spectra of the full peptide construct **6.2** in 10 mM PBS at 20°C obtained with different preparation procedures 1: Sample prepared from fraction of the peptide after Ni affinity column with buffer exchange to 10 mM PBS (8 µM), 2: Sample prepared with refolding procedure 2 (7.3 µM)

To confirm that refolding procedures did not degrade the peptides, we checked them with SDS-PAGE electrophoresis (15% acrylamide, Coomassie Blue stain), the results are shown in **Fig. 6.23**. All peptides after refolding procedures appear to have the same size as before refolding, confirming that the peptides stayed intact throughout refolding procedures.

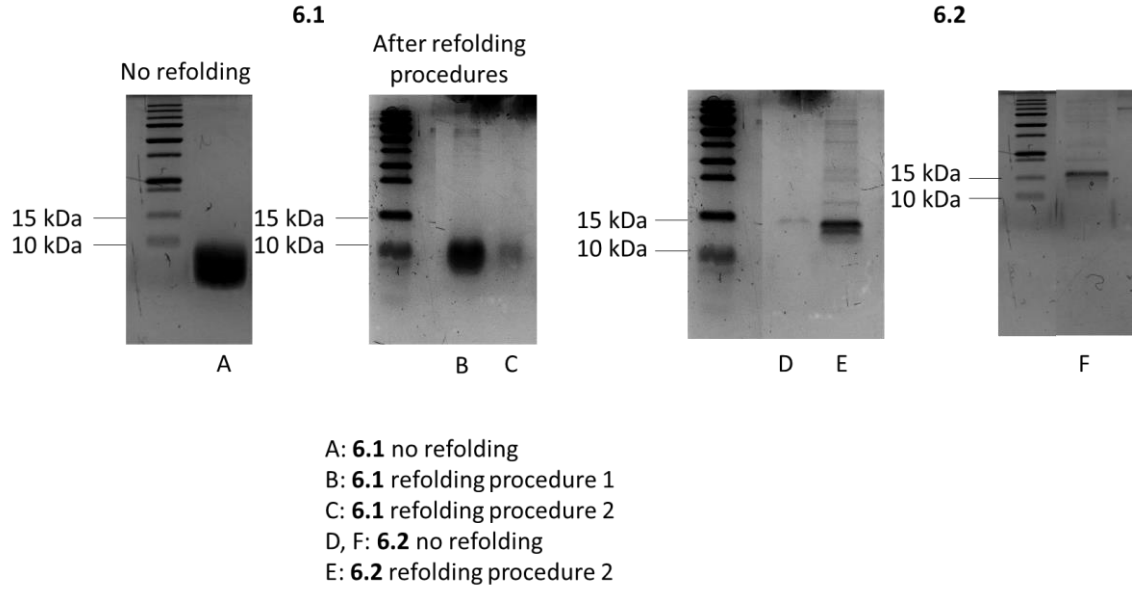


Fig. 6.23 SDS-PAGE gels (15% acrylamide, Coomassie Blue stain) of peptides **6.1** and **6.2** before and after different preparation procedures, confirming that the peptides did not degrade during the preparation

Finally, we looked into the structure of the B domain peptide **6.3** obtained after TEV protease cleavage of the expressed peptide **6.2** prepared in 10 mM PBS, peptide concentration 26.1 µM). The spectrum

shows a negative band at 215-230 nm and a strong positive band at 191-212 nm, indicating that there may be more β-sheet content and also β-turns compared to the peptide **6.1** samples (**Fig. 6.24**). This may result from fold stabilization imparted by the additional 7 amino acids in **6.3** (6 at the C-terminus and 1 at the N-terminus) suggesting that the presence of at least a portion of the A domain is essential to stabilize the fold of B.

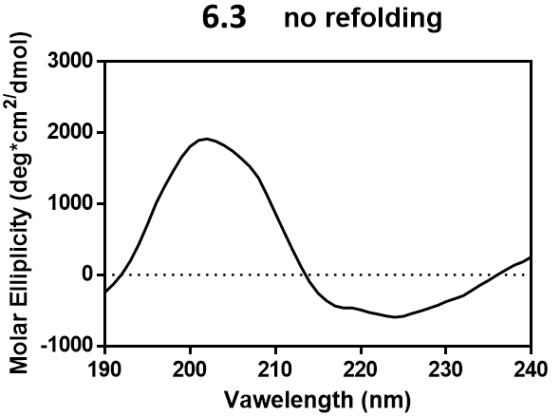


Fig. 6.24 CD spectra of expressed domain B peptide **6.3** prepared in 10 mM PBS at 20°C (26.1 µM)

We supported the empirical observations about the CD spectra also with the use of online prediction tools.

Different online tools and software are available to analyze the CD spectra and estimate structural content, such as CONTIN, VARSLC, SELCON, BestSel, CAPITO, K2D3.

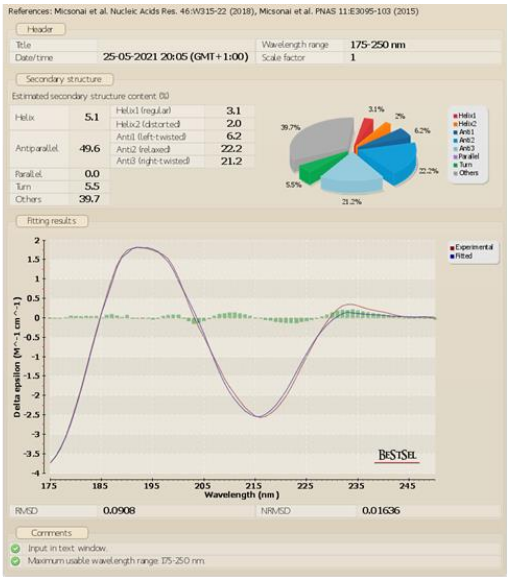
We used BeStSel to predict the structure content from CD spectra and the results obtained with this prediction tool are shown in **Fig. 6.25**. BeStSel is a general tool for a quick structural analysis. A so called β-structure selection method takes into account the twist of β-sheets and can distinguish between parallel and antiparallel β-sheets. The method however is optimized for proteins and therefore the results obtained with peptides can be less reliable, particularly because peptides can have unique structures, distinct from that occurring in proteins.[47]

PDB2CD tool[48] on the other hand can convert the PDB stuctures into CD spectra. We used this tool to calculate the CD spectrum of wt CVN from reported solution NMR structure (PDB code: 1IIY). BeStSel was then used to calculate estimated secondary structure content from the calculated CD spectrum of the wt CVN. The calculated CD spectrum obtained with PDB2CD and estimated secondary structure content obtained with BeStSel closely matched the experimental data for wt CVN (shown in **Fig. 6.25**), therefore we were able to use these results as a reference for analysis of CD spectra of peptides **6.1**, **6.2** and **6.3**.

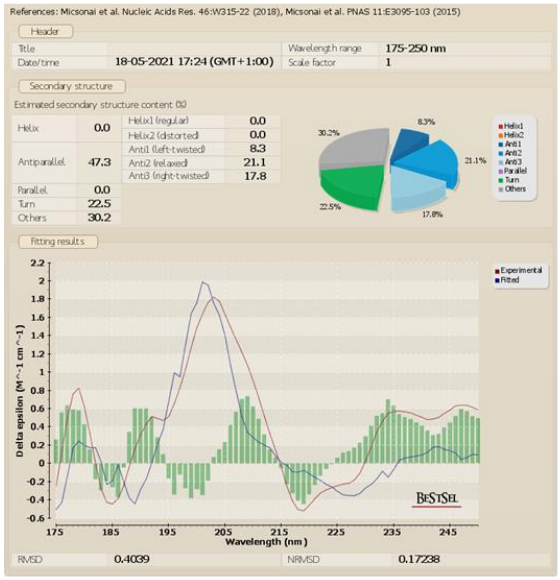
First we compared the estimated secondary structure content of the synthesized peptide **6.1** with the wt CVN reference using the BeStSel prediction tool (**Fig. 6.25**). The structure of synthesized peptide without any refolding procedure is predicted to contain more random coil and β-turn than the wt CVN and less antiparallel β-sheet. The structure of the peptide after the refolding procedure 2 is predicted to be very similar to the non-refolded structure, only the α-helix content is predicted to be slightly higher. The TFE addition with or without heating does not seem to have a significant influence on secondary structure content as estimated from the BeStSel prediction tool.

On the other hand, looking into the structure of full peptide construct **6.2** after refolding procedure 2, it seems to more closely match the wt CVN structure based on the secondary structure content. The percentage of predicted content of antiparallel β-sheets is closely matched, however there is significantly more β-turn and random coil content predicted in the structure. A high β-turn content is also predicted for **6.3** structure, however this structure seems to have a slightly higher antiparallel β-sheet content than the synthesized peptide (**Fig. 6.25**).

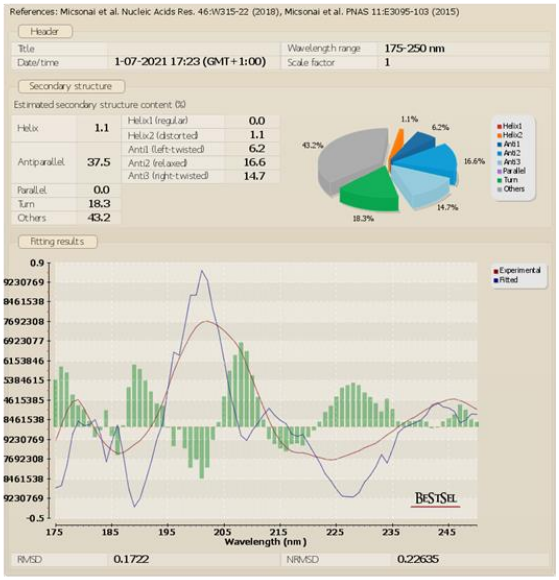
As already mentioned, the method for predicting structural content with BeStSel is optimized for proteins and may therefore be less reliable on shorter peptides. However, taking into account both the empirical observations from CD spectra and results of the prediction tool can provide a clearer idea. Secondary structures of all peptide samples seem to contain less antiparallel β-sheet than the wt CVN and significantly more uncategorized structural motifs. The latter may be attributed to the particular unique structure occurring in smaller peptides or to the fact that a large portion of the peptides is in unordered state. The content of α-helixes in the structures of peptides estimated with the prediction tool does not really seem to be much higher than in the wt CVN, although it might seem so from simply looking at the CD spectra. This is due to the fact that α-helical structure dominates the signal in the CD spectra and it is therefore easy to overestimate its content.



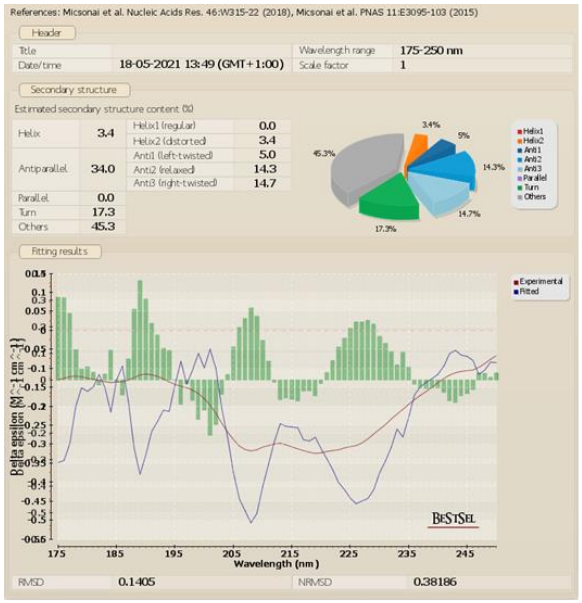
wt CVN prediction



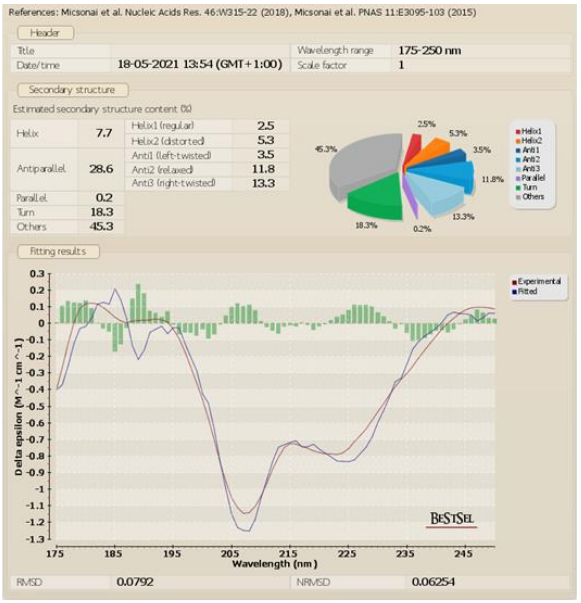
6.2 refolding 2



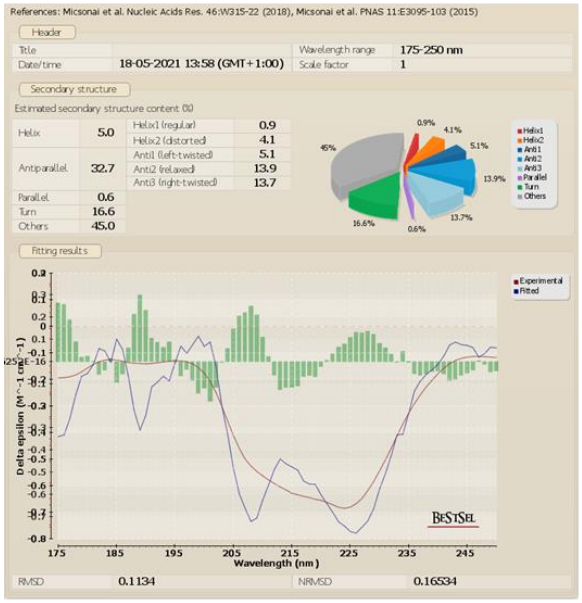
6.3



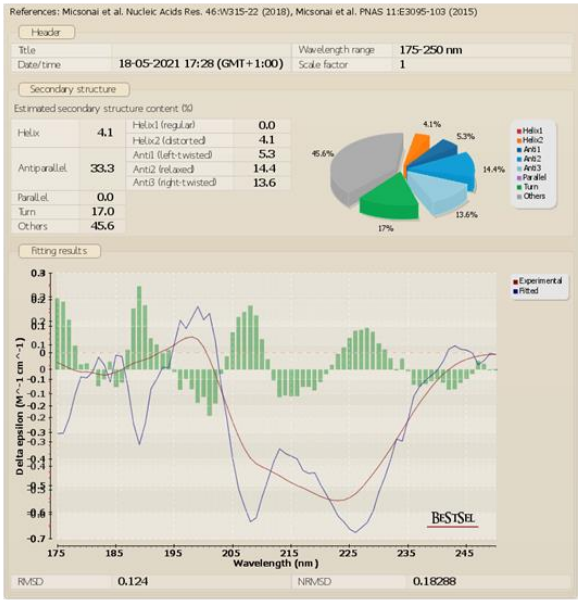
6.1 no refolding



6.1 refolding 2



6.1 TFE

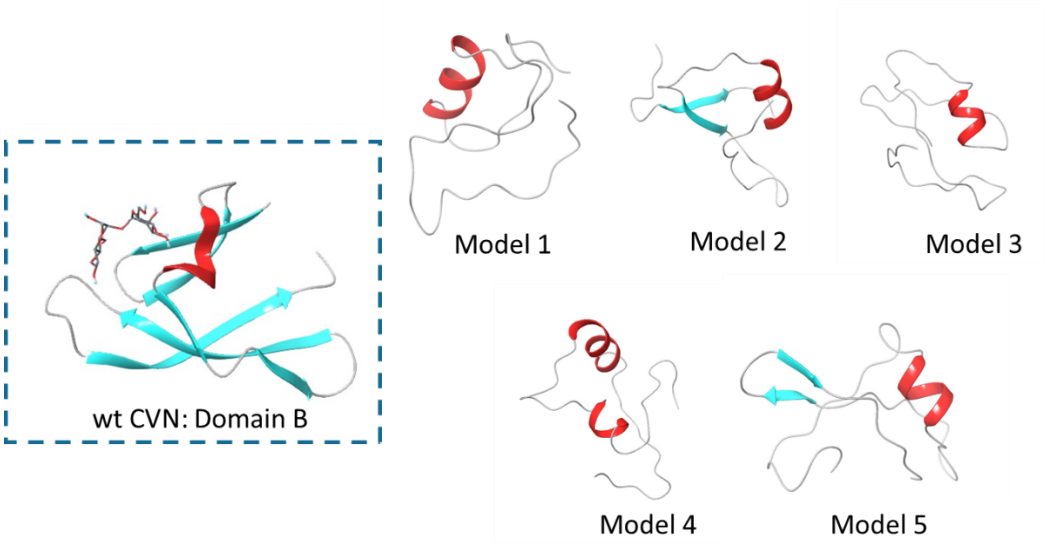


6.1 TFE & heating

Fig. 6.25 Prediction of secondary structure content from CD spectra obtained with BeStSel prediction tool

In the second step of analysis of the secondary structure of the peptides with the help of prediction tools, we used PEP-FOLD 3.5 to predict the secondary structure of the **6.1** peptide from its sequence.[49] This prediction tool gives multiple models of predicted secondary structures of the peptide based on its sequence, a limit is however that the length of the peptide can be only up to 50 amino acids. Therefore, we had to use the sequence that was 1 amino acid shorter on C-terminal for the prediction calculations. As a comparison we also used a sequence that was 1 amino acid shorter on N-terminal

and the predicted models closely matched. Another limit of this prediction tool is that it does not predict the disulphide bond formation, therefore the results need to be interpreted with caution. We generated pdb files of 5 different models and the schematic representation is shown in **Fig. 6.26**.



Sequence: VIENVDGSLKWQPSNFIETCRNTQLAGSSELAAECKTRAQQFVSTKINLD

Fig. 6.26 Left: Secondary structure of wt CVN domain B (PDB 1IIY), Right: Secondary structure predictions of domain B from its sequence, 5 models generated with PEPFold 3.5.

With the use of PDB2CD tool we then converted the secondary structure characteristics of these models into predicted CD spectra and compared them to our experimental results. The spectra are shown in **Fig. 6.27**. The experimental spectra best match the predicted spectra for models 1, 3 and 5 generated with PEP FOLD 3.5 (shown in **Fig. 6.26**). In fact, all models generated with PEPFold 3.5 predict an α-helix in the region Asn 53-Thr 61 (numbering based on the wt CVN) and in this region an α-helix is also found in the wt CVN structure. An additional α-helix or a short β-sheet could be present in the peptide according to the prediction tool, while the rest of the peptide seems to be a random coil. For comparison NPS consensus tool was also used to predict the secondary structure from the sequence and it also suggests that the α-helix could be present in the region Asn 53-Arg 59 (numbering based on the wt CVN) and additionally an α-helix or a β-sheet could be present in the region Ser67-Lys84 (numbering based on the wt CVN), where there are two β-sheets present in the structure of wt CVN. The rest of the peptide is according to both prediction tools mainly in a disordered state.

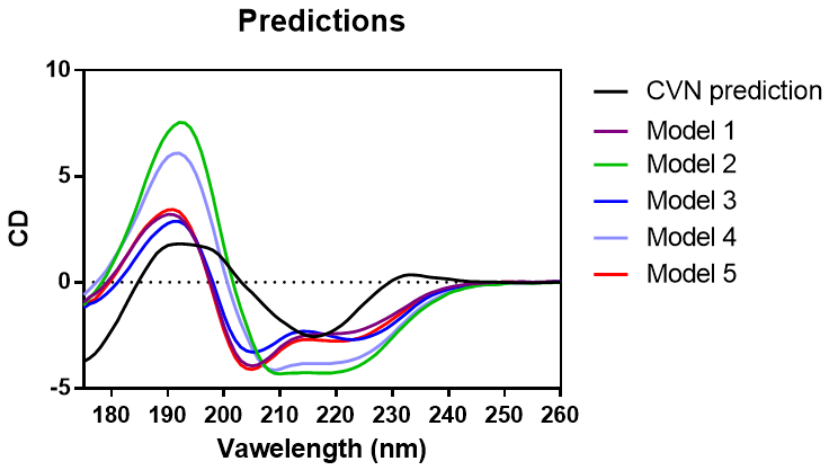


Fig. 6.27 Predicted CD spectra (obtained with PDB2CD prediction tool) of the 5 models obtained from sequence to secondary structure prediction (PEPFold 3.5)

Of course the prediction tools to either estimate the structural content from the CD or to predict the secondary structure from the sequence are not completely reliable. However, a combination of different prediction tools with the observed experimental results can give a good idea on the secondary structure of the peptides.

The folding comparable to the wt CVN was not observed with any of the peptide samples. Looking closely into the structure of isolated domain B, the part of the peptide sequence that most likely retains the structure as the domain B in wt CVN is the Asn 53-Thr 61 (numbering regarding the wt CVN) region that forms an α-helix. The formation of β-sheets and consequently correct overall structure of the synthesized peptide could not be induced by chemical refolding methods. Additionally, conformations of the peptides containing B domain obtained with synthesis (**6.1**) or recombinant production (**6.3**) were different, suggesting that most likely a different thermodynamic minimum is achieved with bacterial expression system that cannot be successfully reproduced with chemical refolding methods. The folding of the CVN might also be driven by the A domain. In this case we expected that the peptide **6.3** obtained with recombinant production would still retain most of the original structure even after the cleavage. However, even the starting full construct **6.2** did not display folding as the wt CVN, which might be attributed to the fact that the short sequence of TEV cleavage sites inserted between the two domains was interfering with the interface of the two domains. Additionally, the C- and N-terminal that form the domain A were labeled with HisTags, which most likely also disrupted the structure of the peptide.

6.4.3. Binding studies

The studies on the secondary structure gave an idea that the peptide constructs were not folding correctly, that is form the same secondary structure as the wt CVN. To confirm these results we evaluated their binding with the minimum binding epitope Manα(1,2)Man. The technique of choice was isothermal titration calorimetry (ITC) that was previously used to quantify the binding affinity of wt CVN with small oligosaccharides. Reported K_a value for high affinity CBS on domain B is 7.2 (±4) x 10^6 M^{-1} and for low affinity CBS on domain A 6.8 (±4) x 10^5 M^{-1}.[50]

ITC is a quantitative technique to study biomolecular interactions. It is often considered a method of choice to analyze protein-carbohydrate interactions, because it allows precise determinantion of the thermodynamic parameters (ΔG, ΔH, ΔS, Kd) associated with the binding process in a single experiment. During the experiment usually the ligand is titrated into a protein solution, although the opposite is also possible. The apparatus directly measures the heat that is released or absorbed during a binding event.

Because the structures of the synthesized peptide **6.1** did not show significant change regardless to which treatment procedure we used, we performed the ITC experiment with the **6.1** peptide prepared only by dissolving in 10 mM PBS buffer. No binding was detected between the peptide (96 µM) and Manα(1,2)Man (20 mM) at 27°C. As can be seen from **Fig. 6.28**, the heat of injection observed in the ITC titration of Manα(1,2)Man in the solution of the peptide was around 0.02 kcal/mol, this corresponds to the heat of injections observed in the titration of Manα(1,2)Man in the 10 mM PBS buffer without the peptide. We also performed an ITC experiment titration of MeαMan (20 mM) in the solution of peptide **6.1** (56 µM) as a negative control. No binding was observed, as expected.

The expressed peptide constructs prepared without any refolding procedure were also evaluated for their binding with Manα(1,2)Man. The full peptide construct **6.2** (65 µM) did not show any binding with Manα(1,2)Man (10 mM), with the heat of injection around 0.03 kcal/mol (**Fig 6.28**).

Finally, cleaved off peptide **6.3** (31 µM) also did not show any binding with Manα(1,2)Man (10 mM), with the heat of injection around 0.03 kcal/mol (**Fig 6.28**).

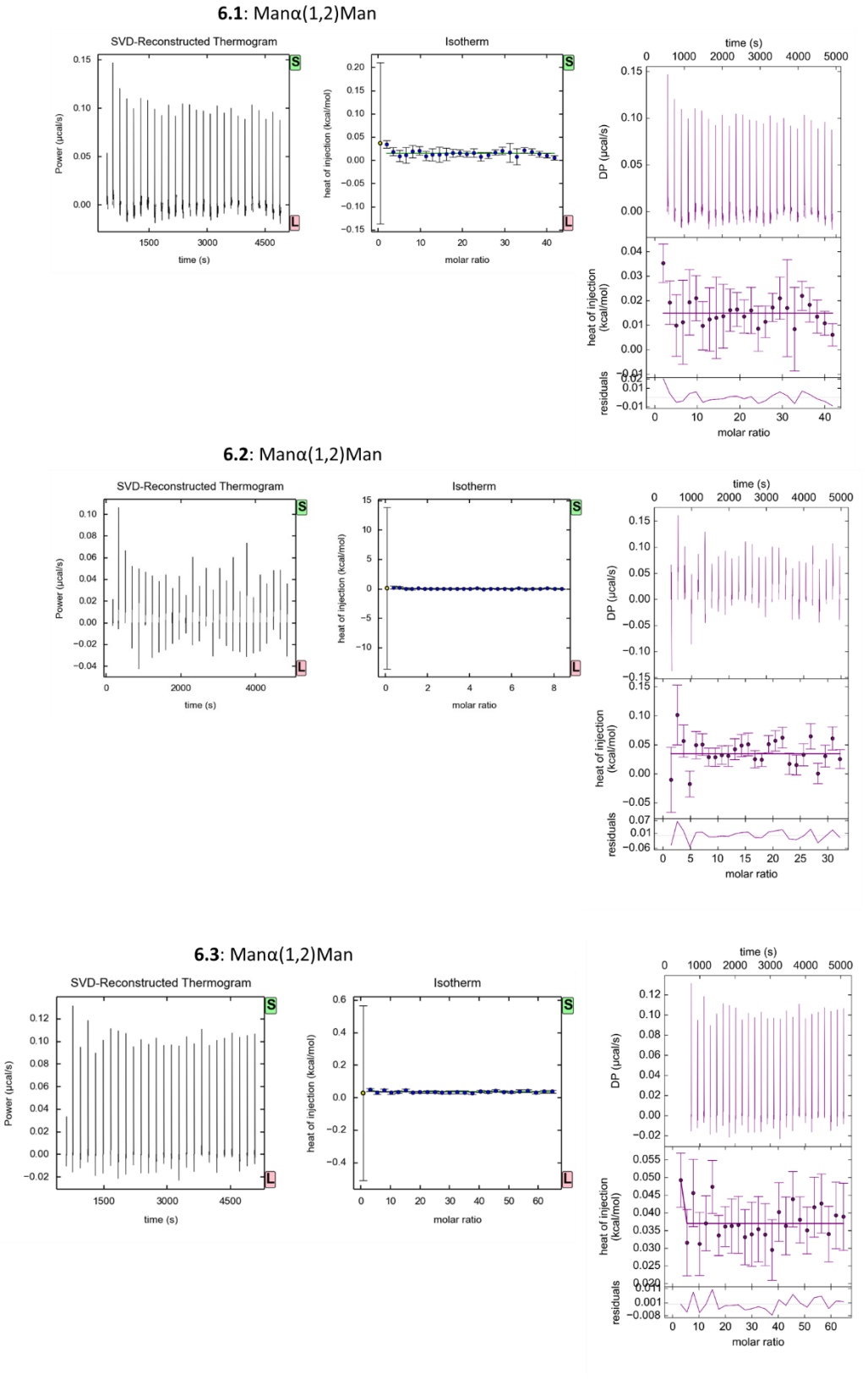


Fig. 6.28 ITC binding experiments of the peptides with Manα(1,2)Man at 27°C in 10 mM PBS. A: **6.1** (96 µM): Manα(1,2)Man (20 mM), B: **6.2** (65 µM): Manα(1,2)Man (10 mM), C: **6.3** (31 µM): Manα(1,2)Man (10 mM).

As control experiments we performed the titration of a well-known mannose binding lectin concanavalin A (Con A) with Manα(1,2)Man and MeαMan.[51] The titrations were performed by injecting the solution of the saccharide ligand into the solution of the protein at 26°C with both the ligand and the protein dissolved in 0.1 M HEPES buffer, pH 7.2, containing 0.9 M NaCl, 1 mM $MnCl_2$ and 1 mM $CaCl_2$. The measured K_a for the titration of Manα(1,2)Man (27 mM) into solution of protein (0.390 mM) the calculated K_a value was $2.995*10^4$ ($\pm1.45*10^3$) M^{-1} (**Table 6.1**). For the titration of MeαMan (46 mM) into the solution of the protein (0.483 mM) the K_a value was $1.01*10^4$ (±515) M^{-1} (**Table 6.1**).

Table 6.1 Thermodynamic parameters obtained with ITC experiments with ConA and saccharides Manα(1,2)Man and MeαMan.

	Con A (0.390 mM) **: Manα(1,2)Man** (27 mM)	**Con A** (0.483 mM) **: MeαMan** (46 mM)
N	0.877 (±0.0685)	1.04 (±0.051)
Ka	$2.995*10^4$ ($\pm1.45*10^3$) M^{-1}	$1.01*10^4$ (±515) M^{-1}
ΔH	-7519 (±414) cal/mol	-5037 (±202) cal/mol
ΔS	-4.67 (±1.29) cal/mol/deg	-1.47 (±0.772) cal/mol/deg

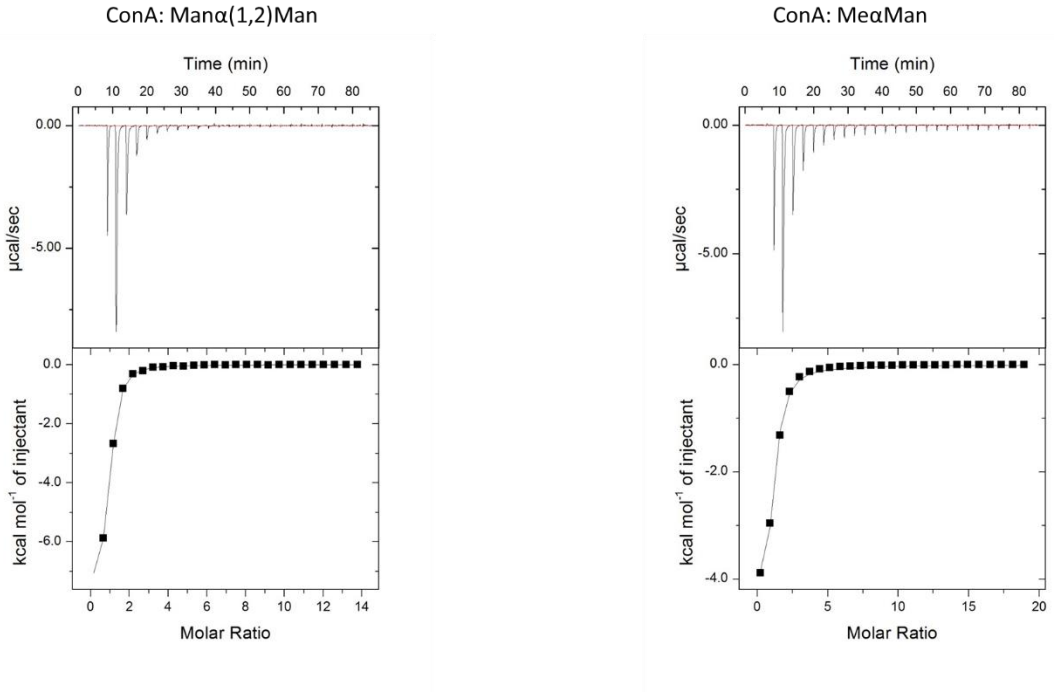


Fig. 6.29 ITC titration curve obtained from ITC experiments with Con A at 26°C in 0.1 M HEPES buffer, pH 7.2, containing 0.9 M NaCl, 1 mM $MnCl_2$ and 1 mM $CaCl_2$. Left: Con A (0.390 mM) and Manα(1,2)Man (27mM). Right: Con A (0.483 mM) and MeαMan (46 mM).

6.5. Conclusions and future perspectives

We focused our attention on the mannose binding protein Cyanovirin-N, a lectin with broad antiviral activity and relatively low toxicity. The lectin consists of two domains (A and B), each one of them carries a carbohydrate binding site. Multisite binding is crucial for antiviral properties of this lectin, thus increasing the number of binding sites could potentially result in enhanced antiviral activity. The approach we explored was to reduce the sequence of the protein to domain B that contains the high affinity carbohydrate binding site, multiple copies of this monovalent peptide could then be conjugated to a number of scaffolds and linkers to obtain constructs with various valencies and spacings. Such an approach is often applied to sugars or glycomimetics, however it is more challenging to reproduce the idea with peptides.

While the CVN proteins are usually produced in a recombinant form, for the purpose of building constructs with multiple copies of domain B we explored a possibility to synthesize the entire B domain. As a comparison we also produced a peptide containing domain B recombinantly from *E. coli*.

We investigated the folding of synthesized peptide under different preparation procedures and compared it with the folding of the recombinant peptides. The studies on secondary structure were supported also by binding studies of the peptides with the ligand Manα(1,2)Man.

We observed that the changes we introduced into the structure of the peptides by either shortening the sequence to a single domain or inserting additional amino acids between the two domains resulted in an altered folding and complete loss of binding affinity of the peptide constructs towards Manα(1,2)Man. This indicates that most likely at least some parts of the domain A have to be included to stabilize the fold of domain B of CVN. The experimental observations will be used for further design of short mannose binding peptides derived from CVN, supported by computational tools to predict the folding.

An interesting possibility would also be to use protein engineering to remodel the low affinity carbohydrate binding site on domain A, so that it would be more similar to high affinity carbohydrate binding site on domain B. Such construct should not cause aggregation when binding to Man_9. Protein engineering may then possibly be used to design oligomers of such protein, similar to previous studies.[27, 28]

6.6. Experimental

General methods

Concentrations of peptides were measured with Thermo Scientific™ NanoDrop 2000, using Protein A280 module and molar extinction coefficient E (M^{-1} cm^{-1}) and molecular weight MW (kDa). E and MW were calculated with ExPASy online tool: **6.1** 5625 M^{-1} cm^{-1} and 5.7 kDa, **6.2** 13200 M^{-1} cm^{-1} and 15.0 kDa, **6.3** 7115 M^{-1} cm^{-1} and 6.5 kDa. Manα(1,2)Man and MeαMan were purchased from Biosynth Carbosynth.

Peptide **6.1** synthesis, cleavage and purification

The peptide was synthesized using a microwave-assisted peptide synthesizer (CEM HT12 Liberty Blue). RAPP polymere tentagel S RAM resin (112 mg, 25 µmol scale) was swollen in 10 mL of a 1:1 mixture of DMF/DCM for 5 min, drained, and then treated with 20 vol.% piperidine (10 mL) in DMF for 65 seconds at 90°C, drained and washed with DMF (3 x 5 mL). The resin was then treated with a solution of Fmoc-Xaa-OH (0.2 mol/L, 2.5 mL, 5 eq), DIC (1 mol/L, 1 mL, 10 eq) and Oxyma (1 mol/L, 0.5 mL, 5 eq) in DMF (4 mL) at 76°C for 15 s before the temperature was increased to 90°C for an additional 110 s before being drained. The resin was then treated again with the same amount of Fmoc-Xaa-OH, DIC and Oxyma in DMF (4 mL) at 76°C for 15 s before the temperature was increased to 90°C for an additional 110 s before being drained. The cycle was performed twice. After each coupling step the resin was treated with 10% acetic anhydride in DMF and heated to 90°C 110 s.

The peptide was cleaved from the resin using a mixture of TFA/water/TIPS/EDT (92.5:2.5:2.5:2.5) shaking for 3 hours at room temperature. The crude peptide was precipitated in a mixture of cold Et_2O/Hexane. The peptide was pelleted by centrifugation (10 min at 4500 rpm), the pellet was washed three times with cold Et_2O, centrifugated (5 min at 4500 rpm) and dried.

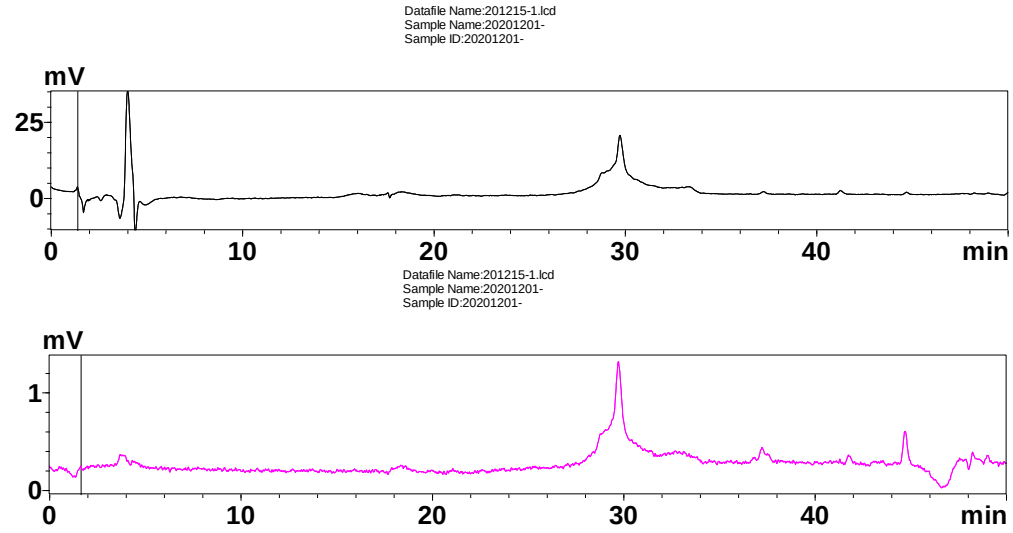
The crude peptide after cleavage from the resin was dissolved in 0.5 mL DMSO and purified by reverse-phase HPLC using a C18 column using a gradient of 0 to 70% over a time of 120 minutes (Buffer A: 5% CH_3CN, 0.1% TFA; buffer B: 95% CH_3CN, 0.1% TFA). The fraction containing the peptide **6.1** was lyophilized and finally 2.7 mg of peptide was obtained.

LC-MS analysis of **6.1**

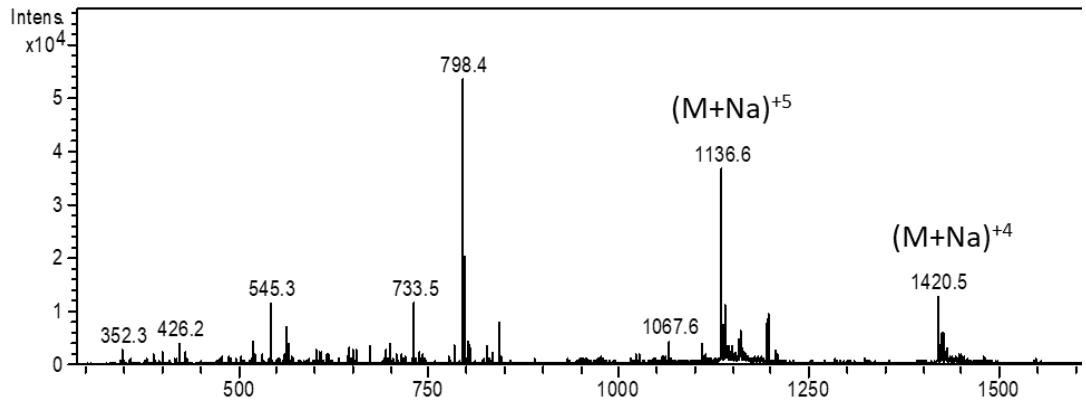
Samples were analyzed on HPLC-MS using a gradient of 0 to 70% over a time of 40 minutes (Buffer A: 5% CH_3CN, 0.1% TFA; buffer B: 95% CH_3CN, 0.1% TFA).

MS-ESI ($C_{242}H_{390}N_{70}O_{82}S_2$): Calculated (m/z): 5675.79 (M+Na), found: 1136.0 $(M+Na)^{+5}$, 1419.7 $(M+Na)^{+4}$

LC (UV detector):



MS (ESI):



Protein **6.2** expression

The plasmid for expression of **6.2** was purchased from GenScript. Protein was expressed from a synthetic gene of CVN with two TEV cleavage sites using a pET26b(+) as vector and *E. coli* BL-21 as host strain. Cells were incubated overnight at 37°C in Luria-Bertani medium. Cell culture was induced with 1 mM isopropyl-β-thiogalactoside and incubated with shaking at 37°C overnight. The cells were harvested and cell lysis was performed with sonification in buffer 6 M GdnHCl, 100 mM NaH_2PO_4, 10 mM TrisHCl, pH=8.0. Suspension was cleared by centrifugation. The supernatant containing the

soluble **6.2** protein with 2 HisTag labels was isolated after Ni sepharose binding column, eluting with 50 mM imidazole, 20 mM TrisHCl, 0.5 M NaCl. 3 fractions of 2 mL containing **6.2** peptide were collected with concentrations 0.753 mg/mL, 0.801 mg/mL, 0.697 mg/mL.

Fasta sequence:
CCATGGGCCATCATCATCATCATCATCTGGGCAAGTTCAGCCAAACCTGCTACAACAGCGCGATTCAAGGCAG
CGTTCTGACCAGCACCTGCGAACGTACCAACGGTGGCTACAACACCAGCAGCATCGACCTGAACAGCGAGAA
CCTGTATTTCCAGGCGGTGATTGAAAACGTTGATGGTAGCCTGAAGTGGCAGCCGAGCAACTTTATCGAAACC
TGCCGTAACACCCAGCTGGCGGGCAGCAGCGAGCTGGCGGCGGAATGCAAGACCCGTGCGCAGCAATTCGT
GAGCACCAAAATTAACCTGGACGACGAAAATCTGTACTTCCAAGCGCACATTGCGAACATTGATGGCACCCTG
AAATACGAACTCGAG

TEV cleavage to obtain **6.3**

To the peptide **6.2** in 10 mM PBS buffer (2 mL, 0.8 mg/mL) 80 µL of TEV buffer and ProTEV Plus enzyme (320 UI, 64 µL) were added and incubated at 30°C for 24 hours. Afterwards the peptide was purified with Ni Sepharose column, eluting the domain B peptide **6.3** (without HisTag) with 10 mM PBS buffer and the domain A portions of the peptide (with HisTag) was eluted with 50 mM imidazole, 20 mm TrisHCl, 0.5 M NaCl.

SDS-PAGE

Peptide samples were prepared with 5x or 4x Laemmli sample buffer, containing 125 mM or 100 mM DTT (final concentration of DTT 25 mM). Samples were heated to 95°C for 5 minutes and resolved by 10% or 15% SDS-PAGE.

After SDS-PAGE the gel was quickly rinsed with demineralized water and peptides were stained with Coomassie Blue for 2 h. The gel was washed with demineralized water and left to destain in water overnight. The gel was imaged with ChemiDoc™ Touch Imaging System (BIO-RAD).

Refolding procedures

The peptides were refolded at room temperature with use of Amicon ultra 0.5 centrifugal filters with 3 kDa cutoff (**6.1** and **6.3** peptides) or Milipore centrifugal filters with 5 kDa cutoff (**6.2** peptide). The speed of centrifugation was 6000-8000 rpm.

Refolding procedure 1:

The protein was treated with 6 M GdnHCl and 5 mM DTT and refolded by dialysis with spin filtration against 2 M GdnHCl and 50 mM TrisHCl, pH 8.0, at room temperature for 4 h and then against 10 mM TrisHCl and 100 mM NaCl, pH 8.0, for 4 hours at room temperature. The peptide was kept at 4°C in the buffer overnight before buffer exchange to 10 mM PBS for CD or ITC studies.

Refolding procedure 2:

The protein was treated with 2 M GdnHCl and 50 mM TrisHCl, pH 8.0, at room temperature for 4 h and then dialyzed against 10 mM TrisHCl and 100 mM NaCl, pH 8.0, for 4 hours at room temperature. The peptide was kept at 4°C in the buffer overnight before buffer exchange to 10 mM PBS for CD or ITC studies.

Addition of TFE and heating:

To the peptide in 200 µL 10 mM PBS 50 µL of TFE was added. The sample was heated to 45°C for 30 min and then cooled down to RT over 30 min.

Circular dichroism

Circular dichroism spectra were obtained by a double beam DSM 1000 CD spectrometer (Online instrument system, Bogart, GA, USA) using quartz cuvettes with path length 1.0 mm. All samples were prepared in 10 mM PBS, pH=7.4. Measurements were performed at 20°C, a wavelength range from 175 to 250 nm. 3-5 measurements were recorded for each sample and the final spectra were plotted with the use of GraphPad Prism 6.

Concentrations of the peptide samples:

Sample	Concentration (mg/mL)	Concentration (µM)
6.1 no refolding	0.36	63.6
6.1 refolding 1	0.085	15.0
6.1 refolding 2	0.15	26.5
6.1 TFE	0.36	63.6
6.1 TFE & heating	0.36	63.6
6.2 no refolding	0.12	8
6.2 refolding 2	0.11	7.3
6.3 no refolding	0.17	26.1

Isothermal titration calorimetry

ITC with peptide samples

ITC experiments were performed on MicroCal Auto ITC200 instrument. Both, the solution of the ligand and the peptide were prepared in 10 mM PBS, pH=7.4. In a typical ITC experiment 26 1.5 µL injections of ligand were injected into a 200 µL cell containing peptide. The ITC experiment was performed at 27°C. Stirring speed was 750 RPM, spacing between injections 150 s. Data were visualized and inspected with NITPIC, Sedphat and Origin softwares.

Concentrations

- **6.1** (96 µM): **Manα(1,2)Man** (20 mM)
- **6.1** (56 µM): **MeαMan** (20 mM)
- **6.2** (65 µM): **Manα(1,2)Man** (10 mM)
- **6.3** (31 µM): **Manα(1,2)Man** (10 mM)

ITC experiments with ConA

ITC experiments were performed on MicroCal Auto ITC200 instrument. Both, the solution of the leigand and the peptide were prepared in 0.1 M HEPES buffer, pH 7.2, containing 0.9 M NaCl, 1 mM $MnCl_2$ and 1 mM $CaCl_2$. In a typical ITC experiment 26 1.5 µL injections of ligand were injected into a 200 µL cell containing peptide. The ITC experiment was performed at 26°C. Stirring speed was 750 RPM, spacing between injections 150 s. Data were processed with MicroCal Origin software.

Concentrations:

ConA (0.483 mM): **MeαMan** (46 mM)

ConA (0.390 mM): **Man(1,2)Man** (27 mM)

6.7. References

1. V. Nizet and J. D. Esko, in *Essentials of Glycobiology*, eds. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2009.
2. L. C. Nascimento da Silva, J. S. P. Mendonça, W. F. de Oliveira, K. L. R. Batista, A. Zagmignan, I. F. T. Viana and M. T. dos Santos Correia, *Glycobiology*, 2021, **31**, 358-371.
3. X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang and Z. Qian, *Nature Communications*, 2020, **11**, 1620.
4. N. Vankadari and J. A. Wilce, *Emerging Microbes & Infections*, 2020, **9**, 601-604.
5. A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire and D. Veesler, *Cell*, 2020, **181**, 281-292.e286.
6. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan and M. Crispin, *Science*, 2020, **369**, 330.
7. S. K. Song, Z. Moldoveanu, H. H. Nguyen, E. H. Kim, K. Y. Choi, J. B. Kim and J. Mestecky, *Vaccine*, 2007, **25**, 6359-6366.
8. R. H. Tullis, R. P. Duffin, H. H. Handley, P. Sodhi, J. Menon, J. A. Joyce and V. Kher, *Blood Purification*, 2009, **27**, 64-69.
9. J. S. Mani, J. B. Johnson, J. C. Steel, D. A. Broszczak, P. M. Neilsen, K. B. Walsh and M. Naiker, *Virus Research*, 2020, **284**, 197989.
10. T. Moulaei, K. B. Alexandre, S. R. Shenoy, J. R. Meyerson, L. R. H. Krumpe, B. Constantine, J. Wilson, R. W. Buckheit, J. B. McMahon, S. Subramaniam, A. Wlodawer and B. R. O'Keefe, *Retrovirology*, 2015, **12**, 6.
11. Michael D. Swanson, Daniel M. Boudreaux, L. Salmon, J. Chugh, Harry C. Winter, Jennifer L. Meagher, S. André, Paul V. Murphy, S. Oscarson, R. Roy, S. King, Mark H. Kaplan, Irwin J. Goldstein, E. B. Tarbet, Brett L. Hurst, Donald F. Smee, C. de la Fuente, H.-H. Hoffmann, Y. Xue, Charles M. Rice, D. Schols, J. V. Garcia, Jeanne A. Stuckey, H.-J. Gabius, Hashim M. Al-Hashimi and David M. Markovitz, *Cell*, 2015, **163**, 746-758.
12. D. Huskens, G. Férir, K. Vermeire, J. C. Kehr, J. Balzarini, E. Dittmann and D. Schols, *The Journal of biological chemistry*, 2010, **285**, 24845-24854.
13. Y.-Q. Min, X.-C. Duan, Y.-D. Zhou, A. Kulinich, W. Meng, Z.-P. Cai, H.-Y. Ma, L. Liu, X.-L. Zhang and J. Voglmeir, *Bioscience Reports*, 2017, **37**.
14. M. Shahid, A. Qadir, J. Yang, I. Ahmad, H. Zahid, S. Mirza, M. P. Windisch and S. Shahzad-ul-Hussan, *Viruses*, 2020, **12**.
15. M. R. Boyd, K. R. Gustafson, J. B. McMahon, R. H. Shoemaker, B. R. O'Keefe, T. Mori, R. J. Gulakowski, L. Wu, M. I. Rivera, C. M. Laurencot, M. J. Currens, J. H. Cardellina, 2nd, R. W.

Buckheit, Jr., P. L. Nara, L. K. Pannell, R. C. Sowder, 2nd and L. E. Henderson, *Antimicrobial agents and chemotherapy*, 1997, **41**, 1521-1530.

16. C. A. Bewley, *Structure*, 2001, **9**, 931-940.
17. L. G. Barrientos and A. M. Gronenborn, *Mini reviews in medicinal chemistry*, 2005, **5**, 21-31.
18. S. R. Shenoy, L. G. Barrientos, D. M. Ratner, B. R. O'Keefe, P. H. Seeberger, A. M. Gronenborn and M. R. Boyd, *Chemistry & biology*, 2002, **9**, 1109-1118.
19. S. Xiong, J. Fan and K. Kitazato, *Applied Microbiology and Biotechnology*, 2010, **86**, 805-812.
20. H. Lotfi, R. Sheervalilou and N. Zarghami, *BioImpacts : BI*, 2018, **8**, 139-151.
21. E. Matei, R. Basu, W. Furey, J. Shi, C. Calnan, C. Aiken and A. M. Gronenborn, *The Journal of biological chemistry*, 2016, **291**, 18967-18976.
22. A. Barre, E. J. M. V. Damme, M. Simplicien, H. Benoist and P. Rougé, *Marine Drugs*, 2020, **18**, 543.
23. C. A. Bewley, K. R. Gustafson, M. R. Boyd, D. G. Covell, A. Bax, G. M. Clore and A. M. Gronenborn, *Nature Structural Biology*, 1998, **5**, 571-578.
24. C. Sandström, B. Hakkarainen, E. Matei, A. Glinchert, M. Lahmann, S. Oscarson, L. Kenne and A. M. Gronenborn, *Biochemistry*, 2008, **47**, 3625-3635.
25. L. G. Barrientos, J. M. Louis, I. Botos, T. Mori, Z. Han, B. R. O'Keefe, M. R. Boyd, A. Wlodawer and A. M. Gronenborn, *Structure*, 2002, **10**, 673-686.
26. I. Botos, T. Mori, L. K. Cartner, M. R. Boyd and A. Wlodawer, *Biochemical and Biophysical Research Communications*, 2002, **294**, 184-190.
27. J. R. Keeffe, P. N. P. Gnanapragasam, S. K. Gillespie, J. Yong, P. J. Bjorkman and S. L. Mayo, *Proceedings of the National Academy of Sciences*, 2011, **108**, 14079.
28. B. W. Woodrum, J. Maxwell, D. M. Allen, J. Wilson, L. R. H. Krumpe, A. A. Bobkov, R. B. Hill, K. V. Kibler, B. R. O'Keefe and G. Ghirlanda, *Viruses*, 2016, **8**.
29. Y. Lei, W. Chen, H. Liang, Z. Wang, J. Chen, H. Hong, L. Xie, H. Nie and S. Xiong, *Archives of virology*, 2019, **164**, 1259-1269.
30. E. Matei, R. Basu, W. Furey, J. Shi, C. Calnan, C. Aiken and A. M. Gronenborn, *Journal of Biological Chemistry*, 2016, **291**, 18967-18976.
31. R. J. Shattock and J. P. Moore, *Nature Reviews Microbiology*, 2003, **1**, 25-34.
32. D. M. Colleluori, D. Tien, F. Kang, T. Pagliei, R. Kuss, T. McCormick, K. Watson, K. McFadden, I. Chaiken, R. W. Buckheit and J. W. Romano, *Protein Expression and Purification*, 2005, **39**, 229-236.
33. X. Gao, W. Chen, C. Guo, C. Qian, G. Liu, F. Ge, Y. Huang, K. Kitazato, Y. Wang and S. Xiong, *Applied Microbiology and Biotechnology*, 2010, **85**, 1051-1060.

34. R. Agarwal, J. Trivedi and D. Mitra, *Process Biochemistry*, 2020, **93**, 1-11.
35. B. Giomarelli, R. Provvedi, F. Meacci, T. Maggi, D. Medaglini, G. Pozzi, T. Mori, J. B. McMahon, R. Gardella and M. R. Boyd, *AIDS (London, England)*, 2002, **16**, 1351-1356.
36. H. Lotfi, R. Sheervalilou and N. Zarghami, *BioImpacts : BI*, 2018, **8**, 139-151.
37. A. Sexton, P. M. Drake, N. Mahmood, S. J. Harman, R. J. Shattock and J. K. C. Ma, *The FASEB Journal*, 2006, **20**, 356-358.
38. B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth, *Proceedings of the National Academy of Sciences*, 2012, **109**, E690.
39. G. Wei, Z. Su, N. P. Reynolds, P. Arosio, I. W. Hamley, E. Gazit and R. Mezzenga, *Chemical Society Reviews*, 2017, **46**, 4661-4708.
40. C. B. Anfinsen, *Science*, 1973, **181**, 223-230.
41. C. M. Gomes and P. F. N. Faísca, *Protein Folding: An Introduction*, Springer International Publishing, 2019.
42. S. R. Martin and P. M. Bayley, in *Calcium-Binding Protein Protocols: Volume 2: Methods and Techniques*, ed. H. J. Vogel, Springer New York, Totowa, NJ, 2002, DOI: 10.1385/1-59259-184-1:043, pp. 43-55.
43. S. M. Kelly and N. C. Price, *Current protein & peptide science*, 2000, **1**, 349-384.
44. N. J. Greenfield and G. D. Fasman, *Biochemistry*, 1969, **8**, 4108-4116.
45. R. Fromme, Z. Katiliene, B. Giomarelli, F. Bogani, J. Mc Mahon, T. Mori, P. Fromme and G. Ghirlanda, *Biochemistry*, 2007, **46**, 9199-9207.
46. M. Okumura, S. Shimamoto and Y. Hidaka, *The FEBS Journal*, 2012, **279**, 2283-2295.
47. A. Micsonai, F. Wien, L. Kernya, Y. H. Lee, Y. Goto, M. Réfrégiers and J. Kardos, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, E3095-3103.
48. L. Mavridis and R. W. Janes, *Bioinformatics*, 2017, **33**, 56-63.
49. A. Lamiable, P. Thévenet, J. Rey, M. Vavrusa, P. Derreumaux and P. Tufféry, *Nucleic acids research*, 2016, **44**, W449-454.
50. C. A. Bewley and S. Otero-Quintero, *Journal of the American Chemical Society*, 2001, **123**, 3892-3902.
51. D. K. Mandal, N. Kishore and C. F. Brewer, *Biochemistry*, 1994, **33**, 1149-1156.

CHAPTER SEVEN
Conclusions

Carbohydrates are biomolecules used for presentation of biochemical signals on the cell surface. Their structural diversity is advantageous for their function, because it allows for high-density information coding in a minimum space. On the other hand, this complexity of carbohydrate structures makes the attempts of scientists to decipher the sugar code significantly more challenging.

Common approaches to interfere with the interactions between saccharides and their receptors (lectins) involve the use of sugar mimics (glycomimetics) and lectin-based peptide constructs. This thesis was designed to explore both strategies, we worked towards the development of a new class of glycomimetics as well as new peptide constructs derived from a lectin, Cyanovirin-N.

Use of natural saccharides in research studies as well as in therapy comes with several disadvantages. Namely, their poor pharmacokinetic properties prevent passive absorption at the intestinal level; at the same time their high renal clearance causes a low plasma half-life even after parental administration. Additionally, the complex structures of oligosaccharides are very often difficult to synthesize, they require complex multistep syntheses which makes it very difficult to create libraries of compounds for drug discovery campaigns. All these are the reasons why glycomimetics come into play. They are structural and functional mimics of natural saccharides, yet they can have an improved methabolic stability, very often they can also be obtained through simpler synthetic procedures.

Our goal was to synthesize a library of glycomimetic compounds with the structure of *N*-linked-pseudo-*thio*-glycosides. These molecules can be potentially recognized by various lectins but have an improved resistance towards enzymatic and acid-catalyzed hydrolysis. To simplify the synthesis of these compounds we developed a new synthetic procedure, so called one-pot aziridine opening reaction, that combines the generation of a glycosyl thiolate with the subsequent aziridine opening. Although our goal was to simplify the synthetic procedure for these glycomimetics, we got a first-hand experience of the complexity of carbohydrate chemistry. Unexpectedly and in contrast to many literature reports, we observed anomeric isomerization of some glycosyl thiolates under the experimental conditions. This clearly proves that despite many efforts some fundamental mechanistic aspects of carbohydrate chemistry still need to be better understood.

I believe that my experience with carbohydrate chemistry during this thesis can be best summarized with the words of Prof. Dr. Hans Paulsen[1]:

«Although we have now learned to synthesize oligosaccharides, it should be emphasized that each oligosaccharide synthesis remains an independent problem, whose resolution requires considerable systematic research and a good deal of know-how. There are no universal reaction conditions for oligosaccharide syntheses.»

Shifting the point of view from molecules that carry the biochemical information (carbohydrates) towards the molecules that read and translate this information into cellular effects (lectins), we also worked towards developing peptide constructs derived from Cyanovirin-N. This small cyanobacterial lectin is a promising antiviral agent, thus our goal was to work towards developing new flexible multisite constructs to further improve its antiviral activity. In this context the first question that we wanted to answer with this research project was whether we could possibly shorten the sequence of this lectin to a single domain and produce it synthetically, all with the goal to facilitate the development of multisite constructs. It turned out that folding of the peptide consisting of a single domain is significantly altered compared to the wild type Cyanovirin-N and therefore new peptide constructs have to be more carefully designed with the use of prediction tools and computational techniques.

As for a comparison between sugar derivatives and lectin-based constructs, there are more similarities than differences between these two types of biomolecules. Both these classes of compounds have long been overlooked because their complexity and diversity were perceived as drawbacks from a pharmaceutical point of view. This view is slowly changing, the researchers have recognized the opportunities to create a variety of structures with an incredible biological and therapeutic potential. Both carbohydrate- and lectin-based compounds are being evaluated for their use in treatment of bacterial, viral and micotic infections, tumors, etc. While these two types of biomolecules are waiting for their boom on the market of therapeutics, they have found their place as valuable research tools. Glycomimetics are used in development of functional probes of protein-carbohydrate interactions to elucidate on physiological and pathological consequences of these interactions. A very interesting application of lectins on the other hand is their use as biosensors, for example to detect viral infections.

At last, this thesis is a proof that the language of sugars requires a huge effort to be learned, but I strongly believe that it is meant to become the most spoken one in biomedical fields.

Ref.: 1. H. Paulsen, *Angewandte Chemie International Edition in English*, 1982, **21**, 155-173.

Acknowledgements

I am greatly thankful to my supervisors prof. Anna Bernardi and prof. Roland J. Pieters for their support and guidance through these years of my PhD. I would like to thank prof. Bernardi for all the knowledge and inspiration she provided, being a true role model to me and everyone in her group. I am most grateful that I was able to learn from her not only how to be a great scientist, but also how to be a great person. I would like to thank prof. Pieters for the opportunity to work on a very multidisciplinatory project and his trust that I was able to juggle with the unknown scientific fields. I am thankful for his insights and valuable discussions on science as well as future career prospects.

These years would certainly not be the same without my network PhD4GlycoDrug that provided me with so many opportunities. I am thankful to all the members of the consortium, and particularly to PhD students who were a great support system and a wonderful group of friends.

I am most thankful also to all the other amazing people I had the opportunity to meet during these years. The current and previous members of the Bernardi, Gennari, Belvisi and Sattin groups who shared the daily lab struggles with me and encoraged me on each step of this PhD journey. I enjoyed every coffee, aperitivo and most of all every insight into the truly amazing Italian culture and food they shared with me. And of course I would like to thank all the CBDD department in Utrecht for helping me find my way around the labs and also for their support in the pandemic year, all the inspiring discussions and fun moments.

I would like to thank my students Daniele, Matteo and Manuel for their efforts and their contributions to my projects as well as for their eagerness to learn. It was a true pleasure working with each one of them.

Nenazadnje pa naj se zahvalim še svojim staršem Dragici in Francu ter sestri Ines, ki so me podprli na moji poti v tujino, me bodrili in mi stali ob strani ob vseh uspehih in še toliko bolj ob neuspehih, ki sem jih izkusila.