UNIVERSITA' DEGLI STUDI DI MILANO

Facoltà di Scienze Matematiche, Fisiche e Naturali

Scuola di Dottorato in Scienze e Tecnologie Chimiche

Dottorato in Scienze Chimiche XXIV ciclo

PhD Thesis

SYNTHESIS OF UNNATURAL α-*N*-LINKED GLYCOPEPTIDES

Tutor: Prof. Anna Bernardi

Coordinator: Prof. Silvia Ardizzone

Cinzia Colombo

Matr. N. R08329

Anno Accademico 2010-2011

Table of contents

Chapter 5
Synthesis of α-*N*-linked glycosyl amino acids via DeShong methodology

Chapter 6
Coupling conditions and synthesis of α-*N*-linked glycopeptides in solution

Chapter 7
Solid phase Synthesis of α-*N*-linked glycopeptides

Chapter 8

Purification of α-*N*-linked glycopeptides

Chapter 9

Conformational analysis and molecular recognition of α-*N*-linked glycopeptides: Preliminary studies

Chapter 10

Conclusions and outlook

Acronyms and **Abbreviations**

Ac	acetyl
Ala	alanine
Asn	asparagine
Boc	*tert*-butyloxycarbonyl
BOP	Benzotriazole-1-yl-oxy-tris(dimethylamino)phosphonium hexafluorophosphate
Cbz	carboxybenzyl
CTC	chlorotrityl chloride
DBU	1,8-Diazobicyclo[5.4.0]undec-7-ene
DCC	dicyclohexyl carbodiimide
DCM	dichloromethane
DIPEA	diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	*N*,*N'*-dimethylformamide
DMSO	dimethylsulfoxide
EDC.HCl	1-ethyl-3-(3-dimethyllaminopropyl)carbodiimide hydrochloride
Fmoc	9-fluorenylmethyloxycarbonyl
Gly	glycine
HATU	O-(7-azabenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	O-(benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOAT	1-hydroxy-7-azabenzotriazole
HOBT	1-hydroxybenzotriazole
NMP	*N*-methylpyrrolidine
PFP	pentafluorophenyl
Phe	phenylalanine
PyBrop	bromo-tris-pyrrolidino phosphoniumhexafluorophosphate
R.T.	room temperature
SASRIN	super acid sensitive resin
TFA	trifluoroacetic acid
Thr	threonine

Chapter 1
Glycopeptides and Glycoproteins: Structure, Applications and Glycosylation's Effect

1.1 Structure of glycopeptides and glycoproteins:

Glycopeptides belong to a family of molecules which have a carbohydrate domain and a peptide domain. Glycoproteins are larger versions that contain more than about 50 amino acids per peptide component. The carbohydrate can be a single monosaccharide or a complex, potentially branched, oligosaccharide formed of up to about 20 monosaccharide units.

Glycoproteins are essential to many important biological processes including fertilization, immune defence, viral replication, parasitic infection, cell growth, cell-cell adhesion, degradation of blood clots, and inflammation.[1] Glycoproteins are indeed, together with glycolipids the major component of the outer layer of mammalian cells. Carbohydrates displayed on the surface of a protein can function as recognition elements, either at the molecular level, between molecules and cells, or at an intercellular level.[2] The nature and the presence of the carbohydrate can vary during the life cycle of a cell, as a complex battery of glycosyl transfer enzymes operate on it, so that oligosaccharide structures vary dramatically during development. It has also been demonstrated that specific groups of oligosaccharides are expressed at distinct levels of differentiation.[3]

This introductive chapter has essentially the function of describing the structures and properties of glycopeptides and glycoproteins, furnishing a picture of their role in biology and of the recent applications of these molecules in biomedicine. It will also provide information about glycosylation itself as a decisive element in determining the folding, stability and function of many proteins and glycopeptides.

Carbohydrates are different from the other two classes of biological polymers in two important features: they can be highly branched molecules, and their monomeric units may be attached to one another by many different linkage types. Proteins and nucleic acids are almost exclusively linear and they have only a single type of linkage between units (amide bonds for proteins and 3'-5' phospho-diester bonds for nucleic acids). When compared with oligonucleotides and peptides in terms of coding capacity, the number of 'words' (oligosaccharide isomers) created from a set of 'letters' (monosaccharides) is several orders of magnitude larger: the theoretical number of all possible combinations is 4096 for oligonucleotides, $6.4 \cdot 10^6$ for peptides and $1.44 \cdot 10^{15}$ for saccharides.[4] Possibilities deriving from bond formation via different linkage positions (1→1, 2, 3, 4, 6 for hexopyranose), from two anomeric configurations (α/β), from change in ring size (pyranose/furanose) as well as from introduction of branching and additional site specific substitutions such as acetylation, phosphorylation or sulfation, explain this exceptional structural diversity. This complexity allows carbohydrates to provide almost unlimited variations in their structures.[5] The two main classes of glycosidic linkages to proteins (**Figure 1**) involve either

oxygen in the side chain of serine, threonine, or hydroxylysine (*O*-linked glycans) or nitrogen in the side chain of asparagine (*N*-linked glycans). To be glycosylated, an asparagine residue must be part of a tripeptide Asn-X-Ser, where X is any amino acid a part from proline, although the presence of this sequence is not in itself sufficient to ensure glycosylation.[6] The role of the peptide sequence in directing O-glycosylation is less clear, but a Pro residue, at -1 and +3, may make it easier. Besides, a consensus sequence (Cys-XX-Gly-Gly-Ser/Thr-Cys) has been found to correlate with *O*-fucosylation in epidermal growth factor domains.[7] A third type of linkage has been found for an increasing number of cell surface proteins, which are known to be integrated into the lipid bilayer via a glycophosphatidylinositol (GPI) anchor (**Figure 1**).[8] Only six amino acids serve as a GPI attachment site; these are Cys, Asp, Asn, Gly, Ala, and Ser (CDNGAS). The amino acids Gly, Ala, and Ser predominate at the +1 positions and are necessary at +2 positions.[9]

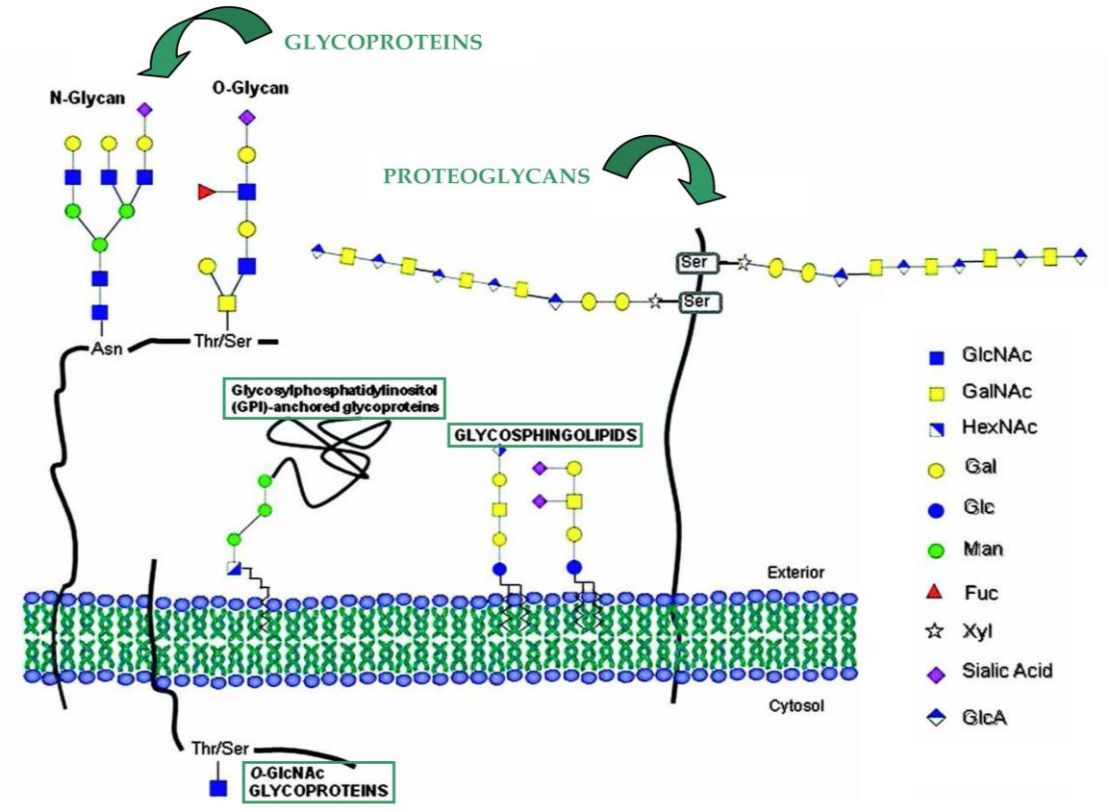


Figure 1. A schematic representation of the main forms of attachment of glycans to polypeptides.

All *N*-linked glycans contain the pentasaccharide Manα1-6(Manα1-3)Man*β*1-4GlcNAc*β*1-4GlcNAc as a common core. On the basis of the structure and the position of glycan residues attached to this mannosyl core, *N*-linked oligosaccharides can be categorized into four main groups: oligomannose (high mannose), complex, hybrid, and poly-*N*-acetyllactosamine (**Figures 2-3**). Oligomannose-type glycans contain only α-mannosyl residues linked to the trimannosyl core (**Figure 2, (A)**). Complex-type glycans contain no mannose residues other than those in the trimannosyl core, but have branches with *N*-acetylglucosamine residues (**Figure 2, (B)**) at their reducing termini attached to the core. The number of antennae normally varies from two (biantennary) to four (tetraantennary), but a pentaantennary structure has been reported in hen ovomucoid.[10]

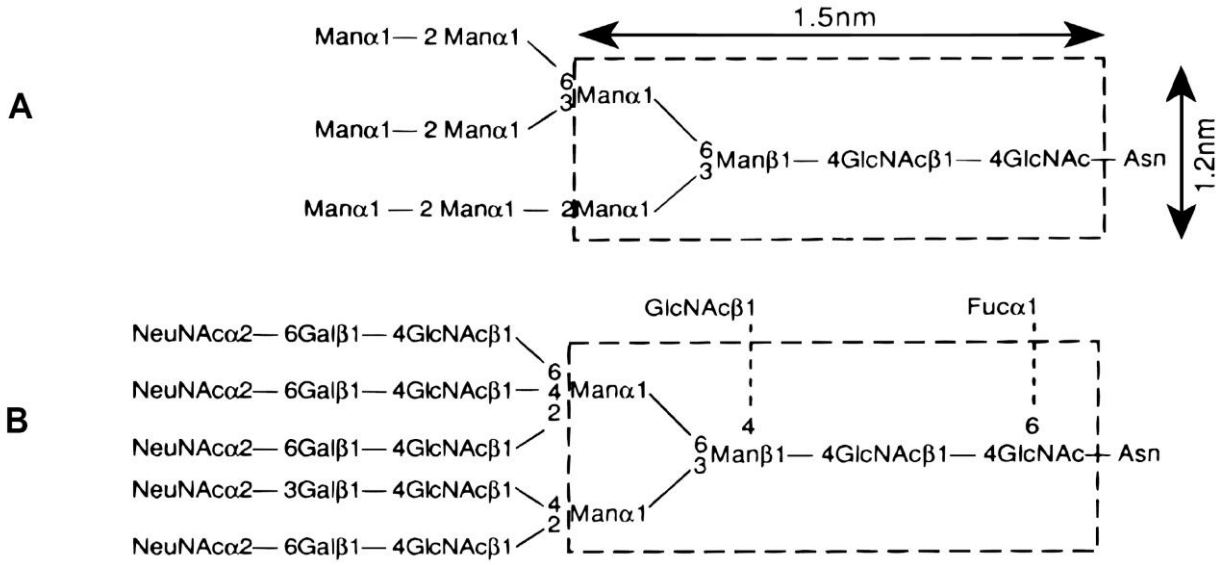


Figure 2. (A) Oligomannose and (B) complex *N*-linked glycans; The structure within the box contains the pentasaccharide core common to all *N*-linked glycans.

While various monosaccharides can be present in the antennae, the presence or absence of fucose and of a "bisecting" GlcNAc on the core contributes to the great structural variation of complex-type glycans (**Figure 2, (B)**). Indeed, complex type *N*-glycans exhibit the largest structural variation in the subgroups resulting mainly from the combinations of different numbers of antennae and kinds of monosaccharides in the outer chains.

The hybrid-type *N*-glycans have the features of both complex and high mannose types (**Figure 3, (C)**). One or two α-mannosyl residues are linked to the Manα1-6 arm of the trimannosyl core, as in the case of oligomannose-type glycans, and generally one or two antennae are linked to the Manα1-3 arm of the core, as in complex type glycans.

The fourth group is the poly-*N*-acetyllactosamine *N*-glycans containing, linked to the core, repeating units of (Galβ1- 4GlcNAcβ1-3). These are not necessarily uniformly distributed on the different antennae and may also be branched (**Figure 3, (D)**). Poly-*N*-acetyllactosamine extensions are most frequently found in tetraantennary glycans.[11]

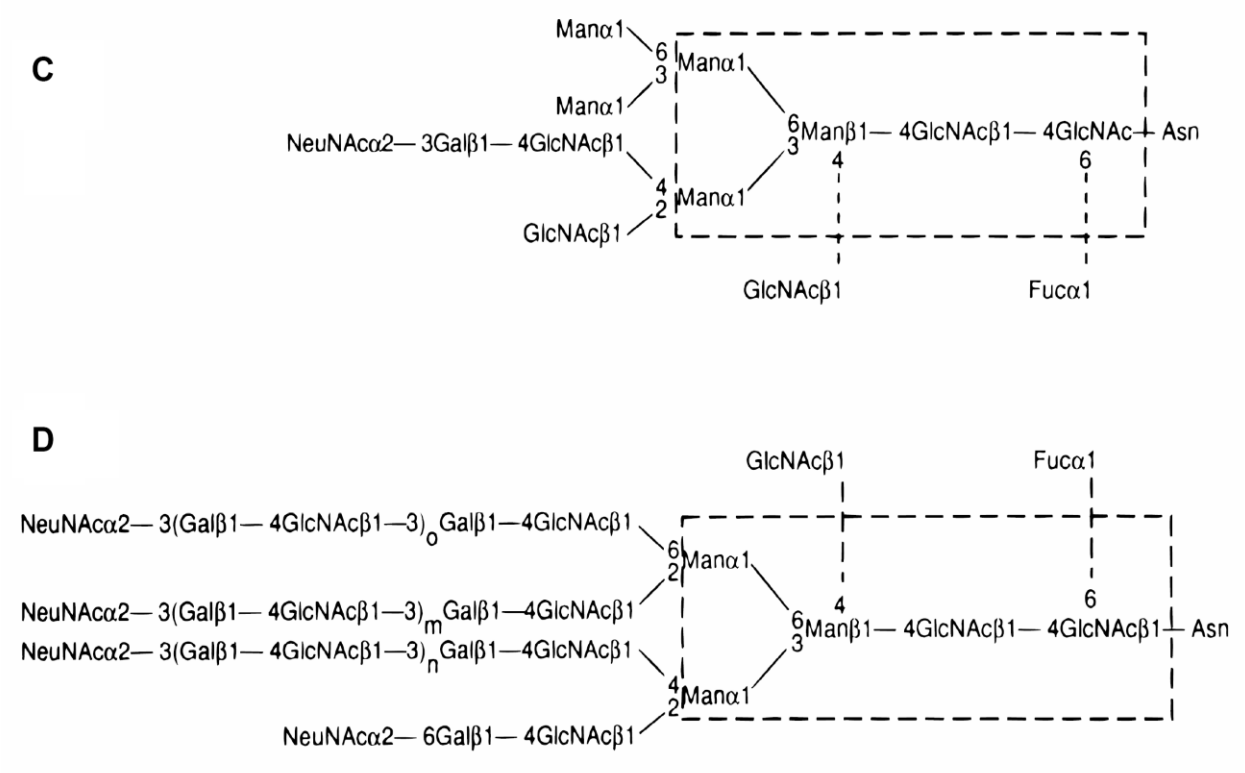


Figure 3. (C) Hybrid; (D) poly-*N*acetyllactosamine N-linked glycans. The structure within the box contains the pentasaccharide core common to all N-linked glycans.

In contrast to *N*-linked glycans, *O*-linked glycans do not present a common core structure and in fact can be classified into at least six groups according to different core structures (**Figure 4**). These cores can be elongated to form the backbone region by addition of Gal in β1-3 and β1-4 linkages, or GlcNAc in β1-3 and β1-6 linkages. Usually, the glycans are linked to serine or threonine residues through GalNAc, but some exceptions have been found in which the linkage is formed through other residues, e.g. fucose. Also single glycans such as fucose or GlcNAc may be *O*-linked to the peptide backbone, and *O*-GlcNAc can also be β-linked as found in cytoplasmic and nucleoplasmic proteins.[12]

Figure 4. Six types of core structures (boxed) among those found in O-linked glycans.

The kind of cell gives a major contribute in determining the extent and type of glycosylation, which is both species and tissue specific.[13] In the biochemical pathway, the initial glycosylation reaction for glycosyl-asparagine-linked structures consist in the transfer of a conserved tetradeca-saccharide ($GlcNAc_2Man_9Glc_3$) from the corresponding dolichyl-pyrophosphate-linked donor (**Figure 5)**

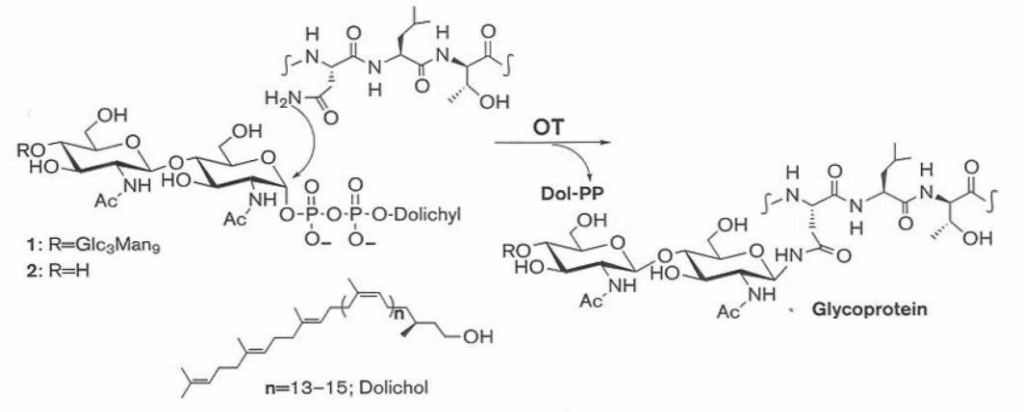


Figure 5. Asparagine-linked glycosylation.

The glycosylation process in which protein-bound and lipid-bound oligosaccharides are formed, takes place on an "assembly line" in the endoplasmic reticulum (ER) and the Golgi apparatus (**Figure 6**).[14] A series of membrane-bound glycosidases and glycosyltransferases act sequentially on the growing oligosaccharide as it moves through the lumen of the ER and Golgi apparatus. The protein-bound glycan structure is then further processed and derivatized through the complementary action of a variety of carbohydrate-specific glycosyl hydrolase and glycosyl transferase enzymes.[15] Many different enzimatic reactions are involved in this processing pathways and each individual enzyme reaction may not go to completion, giving rise to glycosylated variants of the polypeptide, the so called *glycoform*. The type of enzymes, their concentration, kinetic characteristics, and compartmentalization, reflect the external and internal environment of the cell in which the protein is glycosylated. This complex and delicate equilibrium explains why the glycosylation patterns of natural glycoproteins may be influenced by physiological changes such as pregnancy and also by some diseases which may affect one or more of the enzymes in the cell.

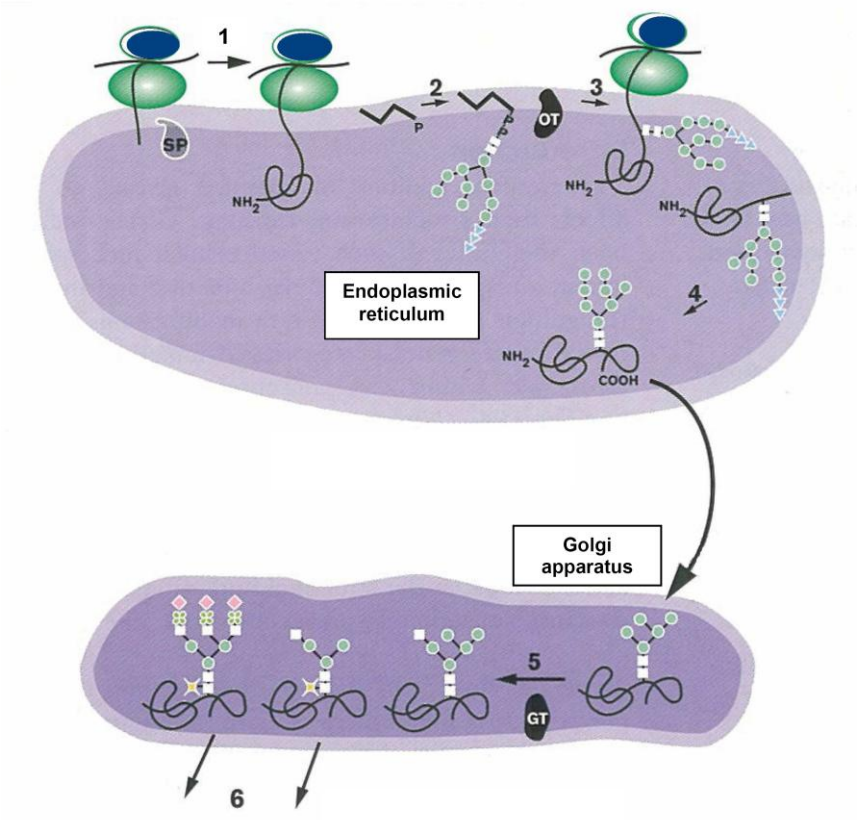


Figure 6. Co-translational protein glycosylation. **1**. A signal peptide is removed from the protein synthesized on membrane-associated ribosomes and is traslocated into the lumen of the ER by signal peptidase (SP). **2.** Biosynthesis of dolichol-linked oligosaccharide, the donor for glycosylation. **3.** Protein glycosylation catalyzed by oligosaccharyl transferase (OT). **4.** Glucose trimming, followed by transfer to the Golgi apparatus. **5.** Modification of the glycan moiety by glycohydrolase and glycosyl transferase enzymes. **6.** Secretion at various stages of glycoprotein products.

1.2 Glycosylation's effect and recent applications of glycopeptides

Although the significance of glycosylation is not completely understood, some of the structural effects of glycosylation have emerged: without glycosylation, immature proteins may misfold, aggregate, and be degraded before leaving the ER.[16] Essentially, glycosylation influences the conformation of nascent polypeptides influencing their folding process. Alterations in oligosaccharides displayed on cell surface are associated with various pathological alterations including malignant transformation. For each glycosylation site on a protein, there is in fact a set of glycoforms that all have the same amino acids but differ in the sequence or position of the attached sugars. It is the populations in this set of glycoforms that change under a variety of conditions such as disease. The carbohydrate-deficient disorders, for example, are a group of multisystemic rare diseases characterized by severe nervous-system disorders, which include growth retardation, abnormal distribution of adipose tissue, abnormal ocular movements, and infertility.[17] In this pathology, a large number of serum glycoproteins (transport proteins, glycoprotein hormones, complement factors, enzymes, and enzyme inhibitors) have been reported to be abnormal with respect to their glycoform populations. In the last twenty years, approximately 1000 patients with multiple organ dysfunctions have been identified and 15 genes' mutations have been found that cause a deficiency of dolichol-linked oligosaccharide biosynthesis (**Figure 5**).[18]

Alteration in mucin expression and aberrant glycosylation are associated with the development of cancer.[19] Mucins, large extracellular proteins which are heavily glycosylated, establish a selective molecular barrier at the epithelial surface and are involved in signal transduction.[20] They have long been implicated in the pathogenesis of cancer. In particular, the earliest pathological examinations and histochemical-staining techniques reported high levels of mucin production by adenocarcinomas.[21] It is possible that tumours use mucins to configure the local microenvironment during invasion and metastasis in sites and conditions that might be inhospitable. Immunohistochemical studies have identified several Tumor-associated antigens (TAAs) on mucins.[22] Most TAAs on mucins were originally found to be oligosaccharide structures, and many turned out to be Blood group antigens;[23] however, some of the antibodies were ultimately shown to recognize protein epitopes that were influenced by glycosylation.[24] Antibodies against TAAs on mucins are widely used clinically as diagnostic tools (serum assays) for different types of cancers. The cell membrane MUC1 mucin for instance is heavily *O*-glycosylated in a large 20-amino acid tandem repeat region, which is aberrantly glycosylated in cancer and can be detected in serum of late-stage patients.[25] Among the most important tumour associated antigen in MUC1 mucin are the

glycopetides T_N and the sialyl T_N antigens which are found in human colon cancer, ovarian cancer and breast cancer (**Figure 7**). [26]

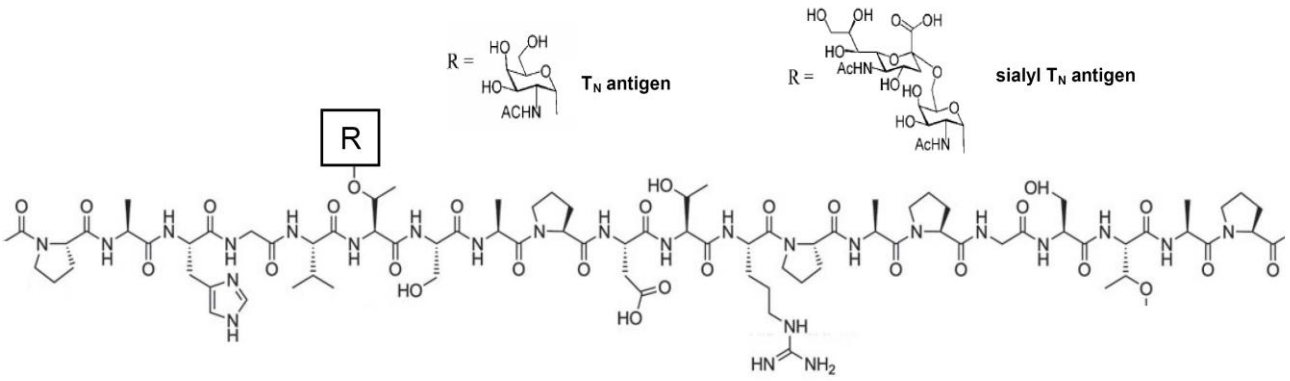


Figure 7. Tumour associated antigens T_N and sialyl T_N.

The immunogenicity of these tumor-associated glycoproteins is too low to overwrite the endogenous tolerance of the immune system. However, it was later demonstrated that synthetic tumor-associated glycopeptides from the tandem repeat region of MUC1, if conjugated to a Tcell epitope peptide from ovalbulmin, afford fully synthetic vaccines which produce a strong, highly specific immune response in transgenic mice.[27] An even stronger and highly specific immune response was induced by immunization of wild-type balb/c mice with a vaccine containing the tumor associated MUC1 glycopeptide bound to tetanus toxoid as the carrier protein.[28] This type of vaccines has the advantage of being applicable to humans. Of course, such MUC1 glycopeptide/tetanus toxoid vaccines also cause immune reactions against tetanus toxoid. To suppress the generation of an anticarrier immune reaction, Kunz et al. recently developed an alternative form of a synthetic vaccine in which the MUC1-glycopeptide is covalently bound to a general immunostimulating structure, the Toll-like receptor ligand $Pam_3CysSer(Lys)_4$.[29]

Furthermore, the overexpressed and aberrantly glycosylated mucin MUC1 are able to generate autoantibodies that might serve as diagnostic biomarkers for cancer-diagnosis.[30] Wandall, Blixt et al., using an antibody-based ELISA assay, documented that aberrant glycoforms cannot be detected in sera of cancer patients, but used the aberrant glycoforms to detect antibodies.[31] A *O*-glycopeptide microarray was developed that detected IgG antibodies to aberrant *O*-glycopeptide epitopes in patients vaccinated with a keyhole limpet hemocyanin–conjugated truncated MUC1 peptide. The synthesis of *O*-glycopeptide was performed chemoenzymatically, using recombinant glycosyl transferases in vitro.[32] Cancer-associated IgG autoantibodies were detected in sera from breast, ovarian, and prostate cancer patients against different aberrant *O*-glycopeptide epitopes derived

from MUC1. These autoantibodies represent a previously unaddressed source of sensitive biomarkers for early detection of cancer.

The application of synthetic glycopeptides as disease-specific autoantibodies biomarkers was also recently applied in the detection of Multiple Sclerosis (MS), the most frequent chronic inflammatory demyelinating disease of the central nervous system and the most common cause of disability in young adults, especially women.[33] Papini et al.[34] showed that the presence of an aberrant *N*-glucosylation triggered an autoantibody response in MS and detected for the first time autoantibodies in MS patients' sera by ELISA, using the synthetic peptide CSF114(Glc) (**Figure 8**). This glycopeptide, characterized by a β-hairpin structure, optimally exposes on the apex of the β-turn the minimal epitope Asn(βGlc).[35] In a further study the possibility of identifying Rheumatoid arthritis specific antibodies through galactosylated peptides has also been demonstrated.[36]

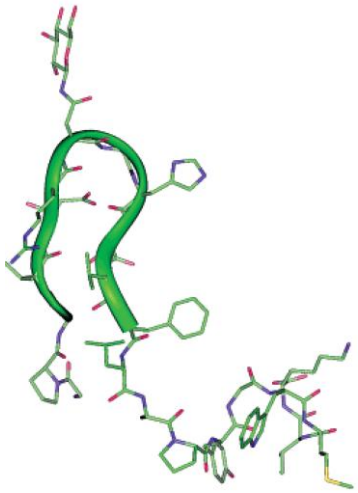


Figure 8. Lowest energy conformer of CSF114(Glc). *β*-hairpin structure is evidenced as a ribbon. (Figure from *Ref 35: J. Med. Chem.* **2006**, *49*, p. 5075)

The fact that glycoproteins are fundamental to viral infection and replication finds a representative example in the HIV infection. The human immunodeficiency virus type 1 (HIV-1) envelope glycoprotein consists of two non-covalently associated subunits, gp120 and gp41, that are generated by proteolytic cleavage of a precursor polypeptide, gp160. In particular, gp120 directs target-cell recognition and viral tropism through interaction with the cell-surface receptor CD4.[37] Notwithstanding enormous scientific effort, the development of a vaccine against HIV has, so far, proven to be evasive. Currently, commonly utilized vaccine formulations have been unable to elicit potent and broadly neutralizing immune responses.[38] The very high rate of viral variation is often cited as the major impediment to vaccine design. Another serious complication is the very low immunogenicity of the protein domain of the viral surface envelope protein gp120 that becomes extensively glycosylated in one of the most effective viral defence mechanism.[39] Indeed, the envelope spike undergoes fast evolution in each individual patient, causing a huge heterogeneity

among individual isolated HIV-1. gp120 is typically modified with carbohydrate motifs that may serve to effectively shield the polypeptide domain from recognition and attack by the immune system.[40] It may be possible that these glycans could themselves serve as targets for an anti-HIV vaccine. In favour of exploring such a course is the fact that some of the glycans are highly conserved, and are located on the outer side of the gp120 trimer, a position that could well enhance their accessibility. Interestingly, 2G12, one of the most potent anti-HIV antibodies currently known, appears to bind to the hybrid- or high-mannose type carbohydrate domains of gp120 (**Figure 9**).[41] With this information in hand, Danishefsky et al. described the synthesis of hybrid- and high-mannose- type nonasaccharide and their incorporation into HIV gp120$^{316–335}$ *N*-linked glycopeptides (**Figure 9**),[42] establishing the ability of their synthetic methodologies to provide access to structurally complex glycopeptide fragments. Antibodies subsequently raised against each of these synthetic glycopeptides would hopefully serve as effective agents against the gp120 envelope protein of HIV. Moreover, identification of antigens that resemble these natural epitopes and could generate similar types of antibodies is an important step toward the development of HIV-1 vaccines.[40] It was recently demonstrated that the composition of HIV-1 envelope *N*-glycans plays a very important role in spreading the infection. In fact, replacement of the heterogeneous *N*-linked glycan composition with a uniform oligomannose *N*- glycan composition increases capture of HIV-1 by immature dendritic cells but results in enhanced viral degradation, decreasing the subsequent transmission to target cells.[43]

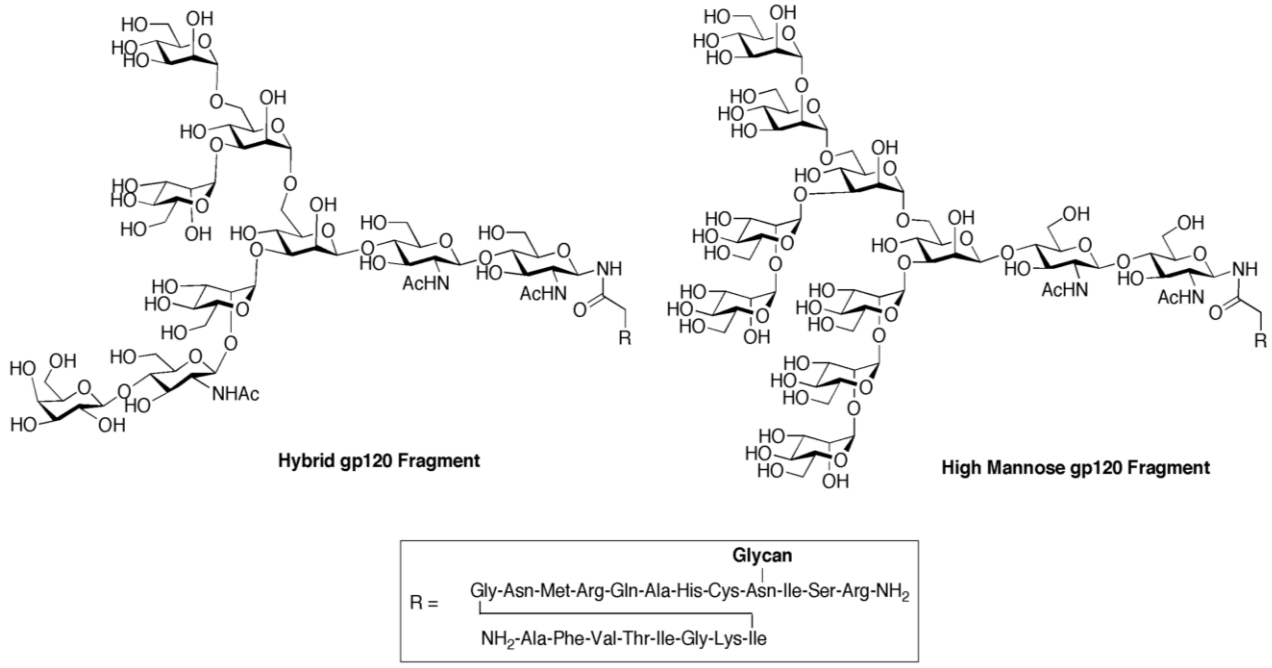


Figure 9. Hybrid and high-mannose gp120 fragments (Figure from *Ref 42b: Pure Appl. Chem.* ***2007****, 79,* p. 2201)

Another useful application of glycopeptides concerns the design of a GPI-based anti-toxic malaria vaccine candidate. GPI anchor is a glycolipid that can be attached to the C-terminus of a protein during post-translational modification. It is composed of a phosphatidylinositol group linked through a carbohydrate-containing linker (glucosamine and mannose glycosidically bound to the inositol residue) and via an ethanolamine phosphate bridge to the C-terminal amino acid of a mature protein. A lipid chain within the hydrophobic phosphatidyl-inositol group anchors the protein to the cell membrane (**Figure 10**). Glycosylphosphatidylinositols are evolutionary conserved glycolipids found in the outer cell membranes of virtually all eukaryotic cells and constitute up to 90% of protein glycosylation in protozoan parasites.[44] Proteins are often modified post-translationally by glycosylation and lipidation.[45] Glycosylphosphatidylinositol (GPI) anchors combine both types of modification and link many proteins to the cell surface. The malaria parasite Plasmodium falciparum has on its cell surface the Glycosylphosphatidylinositol (GPI) glycolipid that seems to be a highly conserved endotoxin of malarial parasite and may contribute to pathogenesis in humans. A recent study suggested that a non-toxic GPI oligosaccharide coupled to a carrier protein was immunogenic and provided significant protection against malarial pathogenesis in a preclinical rodent model.[46] Synthetic GPI is therefore a prototype carbohydrate anti-toxic vaccine against malaria and many efforts were done by the group of Seeberger to synthesize this molecule, grounding on tools derived from recombinant protein engineering and from native chemical ligation.[47] A prominent example of a GPI-anchored protein is the prion protein (PrP) (**Figure 10**).[48] Numerous studies have indicated the strong influence of membrane association through the GPI anchor on the conversion of cellular PrP (PrP^{C}) into its pathogenic isoform PrP scrapie (PrP^{Sc}). However, the speculation that GPI anchoring might contribute to the pathogenicity of PrP is controversial.[49] In vitro and in vivo experiments with this GPI-anchored PrP should be helpful in the elucidation of the influence of GPI-mediated membrane association on the conversion of PrPC into pathogenic PrPSc.

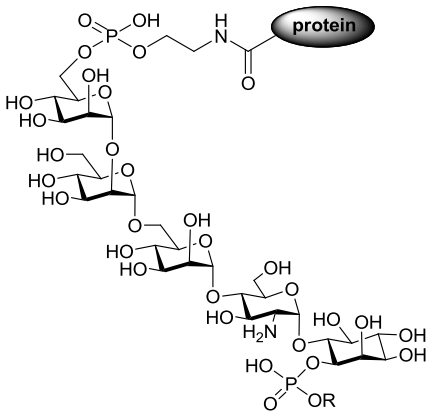


Figure 10. GPI-anchored protein (R=lipid chain, Protein = PrP protein)

Since the application of glycoprotein therapeutics in modern medicine for treatments of various diseases, including cancer, infectious diseases, autoimmune diseases and AIDS/HIV,[50] there was a growing interest in understanding how glycosylation could cause an effect on protein folding. This inspired the application of experimental, computational, and bioinformatic approaches targeted at developing general methods that could define glycosylation-induced effects on protein structure and, further, that could establish a glycosylation code as a predictive tool.[51] Glycosylation seems to have a key role in maintaining the structure and stability of the final folded protein through long range hydrogen-bonding and hydrophobic interactions between the oligosaccharide and the protein.[52] Numerous observations attested for example the importance of asparagine glycosylation for the appropriate folding and assembly of intact protein in the biosynthesis of proteins: in fact the presence of *N*-linked glycosylation inhibitors, such as tunicamycin, often results in extensive aggregation associated with incomplete or incorrect folding.[53] The 3D structure of the individual protein clearly has a role in determining the type and extent of its glycosylation. A number of mechanisms may be involved, including:

a) The position of the glycosylation site in the protein. *N*-Linked sites at the exposed turns of β-sheets, which are sometimes close to proline residues, are normally occupied while those near the *C*-terminus are more often vacant.
b) The interaction of the developing oligosaccharide with the protein surface. This may result in a glycan conformation which may alter the accessibility to specific glycosyltransferases or glycosidases.
c) The interaction of the glycosylenzymes with the protein structure. This can lead to site-specific processing.
d) The interaction of protein subunits to form oligomers. This may prevent glycosylation or restrict the glycoforms at specific sites.

Glycans appear to stabilize tertiary structure[54] also by protecting proteins from proteolysis. Enhanced proteolytic stability, as a result of glycosylation, has been observed, for example, through comparison of RNase A, an unglycosylated pancreatic ribonuclease, and RNase B, which bears a single highmannose oligosaccharide at Asn34.[55] Using proteolysis, the kinetic stability of RNAse B was determined to be approximately 3 kJ/mol higher than RNAse A at 52.5°C. It has been concluded that the increase in stability is due to enthalpic, not entropic, factors. Thermodynamic stability, determined using UV spectroscopy to monitor the unfolding process, indicated that RNAse B is more stable by about 2.5 kJ/mol, independently of temperature.

Five glycoproteins were subjected by Giartosio et al. to enzymatic deglycosylation with various glycosidases to produce either the completely or partially deglycosylated products.[56] Although

circular dichroism measurements suggest that secondary structure was not affected by glycosylation, comparison of the unfolding temperatures suggested enhanced stability in the glycosylated derivatives. The T_m values of avidin and ovotransferrin, glycoproteins that each contain one *N*-linked saccharide, were approximately the same for the glycosylated and non-glycosylated forms. In contrast, the more heavily glycosylated proteins invertase (nine *N*-linked sites), fetuin (three *N*-linked sites) and glucoamylase (thirty *O*-linked sites) exhibited increases in T_m ranging from 1.4°C to 2.3°C depending on the degree of deglycosylation of the derivatives.
Carbohydrates alter other physiochemical properties of proteins: in artic fish,[57] *O*-glycan-rich proteins act as in vivo "antifreeze", preventing nucleation of ice, allowing them to survive low temperatures (Chapter 7.1).

1.3 Structural analysis of glycopeptides

A variety of techniques have been used for the analysis of glycopeptide and glycoprotein structure. Nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopies, molecular-modelling predictions, fluorescence energy transfer (FET) and site-directed mutagenesis studies have provided a structural picture of polypeptides, and have been used to investigate the effect of oligosaccharide modification on the structure or stability of a peptide or protein. X-ray crystallography is not easily applied to glycoproteins, because crystallization is hampered by the heterogeneity of glycan structures and the conformational mobility of the saccharide branches on the surface of the protein. The major impediments to a detailed analysis of the effects of glycosylation on protein folding derive from the fact that most natively glycosylated proteins are large, multisubunit, and/or membrane-associated complexes. Hence, they are currently immensely challenging to study at the level of detail that would provide discrete information on the site-specific effects of glycosylation. Furthermore, biophysical studies are prevented by the limited availability of chemically defined materials for analysis, which is largely due to the intrinsic heterogeneity in natively or heterogeneously expressed glycoproteins. For these reasons, initially, there has been a considerable focus on the conformational analysis of defined glycopeptides, which can be prepared via chemical synthesis.[58] The conformational behaviour of small peptides and glycopeptides provides insight into how glycosylation affects protein secondary structure. Smaller peptide systems are simpler to interpret than large intact proteins; the structure of a glycosylated peptide can be directly compared to that of its non-glycosylated counterpart, providing specific information on the conformational consequences of the modification. Kahne et al. in 1993[59] studied the conformation of the peptide backbone of a linear hexapeptide under the effect of glycosylation

in DMSO**,** a solvent that, like water, does not promote intramolecular hydrogen-bond formation. The hexapeptide sequence studied was Phe-Phe-D-Trp-Lys-Thr-Phe. The sequence contains a threonine, which is a potential glycosylation site. The results showed that glycosylation with a single monosaccharide (GalNAc) has a profound effect on the backbone conformation, limiting the conformational space available to this peptide and favouring conformations in which the backbone bends away from the sugar. Two years later the authors glycosylated the same hexapeptide also with a disaccharide unit[60] and through NMR they evaluated differences in backbone conformation between peptides in different glycosylation states: NOEs between sequential amide protons provide a particularly useful indicator of the ensemble average backbone conformation. As shown in **Figure 11**, the sequential amide-amide ROESY cross peak intensities are very different for glycopeptides **A** (with a monosaccharide) and **B** (with a disaccharide). In peptide **A,** the strongest amide-amide ROESY cross peak is between the amide resonances of Lys and Thr. In contrast, in peptide **B** the NOE cross peak between D-Trp and Lys is very strong. The Lys-Thr cross peak, which is the strongest one in peptide **A,** is absent in peptide **B.** Thus, the hexapeptide with the disaccharide attached has a profoundly different average backbone conformation from the hexapeptide with the monosaccharide attached. The most likely explanation for the observed changes was, in the authors' opinion, the exclusion of many conformations for steric reasons. They supposed that attached sugars may influence protein folding, glycoprotein structure and thermal stability in a similar manner, by restricting conformational space.

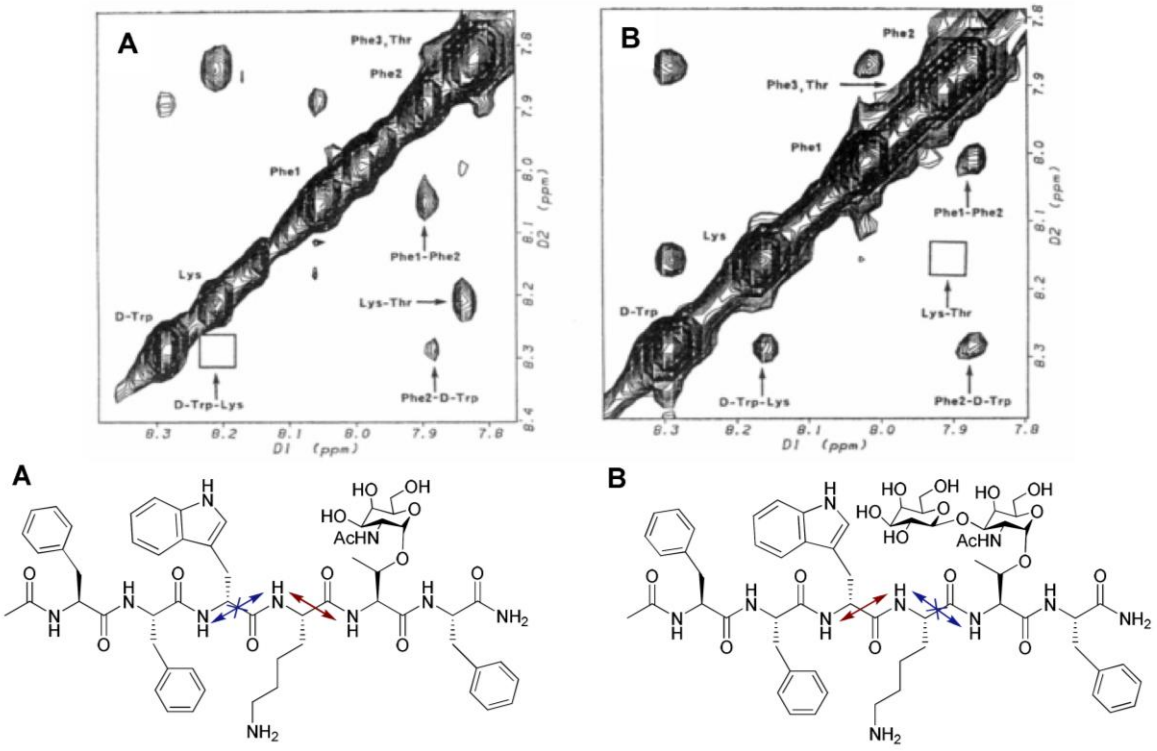


Figure 11. Amide-amide contact regions of the 500 MHz ROESY spectra of **A** (left) and **B** (right).

From early studies on the impact of glycosylation on the conformational dynamics of flexible oligopeptides[61] it emerged that glycosylation could alter the conformational profile of a polypeptide and could allow the polypeptide to sample conformational space not originally accessible to it. The same conclusions were also reached in more recent work:[62] Chemical and chemo-enzymatic methods for the synthesis of glycoconjugates advanced significantly and an increasing number of glycoconjugates derivatized with large, branched oligosaccharides were synthesized in substantial milligram quantities. Powers et al. in 2009 examined the folding energy of the mono-*N*-glycosylated adhesion domain of the human immune cell receptor cluster of differentiation 2 (hCD2ad, **Figure 12**) and studied systematically the influence of the *N*-glycan on the folding energy landscape.[51a] The hCD2ad is a representative member of the immunoglobin superfamily, a small glycoprotein (105 amino acids) with a β-sandwich topology. *N*-glycan structures accelerate folding by 4-fold and stabilize the β-sandwich structure by 3.1 kcal/mol, relative to the non-glycosylated protein. The *N*-glycan's first saccharide unit accounts for the entire acceleration of folding and for 2/3 of the native state stabilization. The remaining third of the stabilization is derived from the next 2 saccharide units. Thus, the conserved *N*-linked triose core, $ManGlcNAc_2$, improves both the kinetics and the thermodynamics of protein folding.

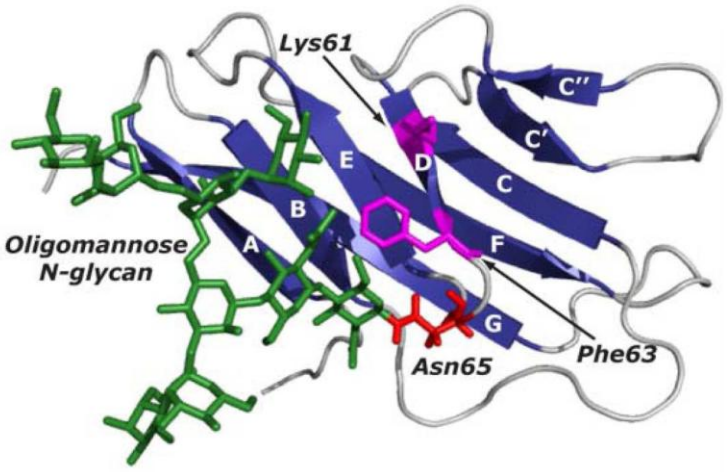


Figure 12. Structure of hCD2ad with *N*-glycan (green). (Figure from *Ref 51a*: *Proc Natl Acad Sci USA* **2009**, *106*, 3131–3136).

Recently, the effect of glycosylation on protein folding was also explored by using computational methods by Levy et al.[51b] They studied *in silico* the folding of the SH3 domain that had been glycosylated with different numbers of conjugated polysaccharide chains at different sites on the protein's surface. The SH3 domain is a small protein (56 amino acids) (**Figure 13**) whose folding was well studied in the past, both experimentally and theoretically. Although the SH3 domain is not a glycoprotein (i.e., it is not a realistic study case), studying a protein whose folding is well

characterized could give an insight into the general effects of the glycan on folding. The authors found that thermodynamic stabilization was correlated with the degree of glycosylation and, to a much smaller extent, with the size of the polysaccharides. The stabilization effect depended on the position of the glycans; thus, the same degree of glycosylation may result in a different thermal effect, depending on the location of the sugars. This article suggests once again that glycosylation can enrich and modulate the biophysical properties of proteins and offers an alternative way to design thermally stabilized proteins.

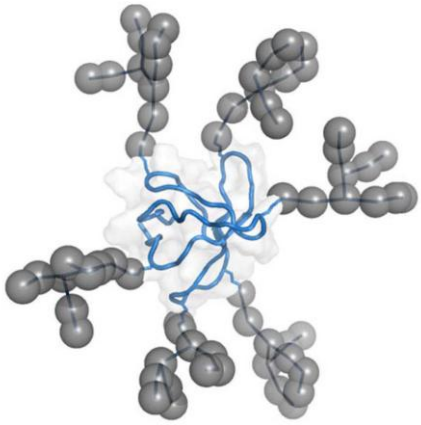


Figure 13. SH3 domain: the polypeptide chain is in blue and the carbohydrate rings are the gray balls. Each glycan contains 11 sugars. (Figure from *Ref 51b*: *Proc Natl Acad Sci USA* **2008**, *105*, 8256–8261).

One of the most systematic work performed with the intent of understanding local structure and stability of glycopeptides was from the group of Imperiali.[63] After having synthesized a family of short glycopeptides in which key molecular elements of the sugar, specifically the *N*-acetyl groups, were modulated, they explored the effect of changes in carbohydrate composition on glycopeptide conformation. The short peptide sequence AcNH-Orn-Ile-Thr-Pro-Asn-Gly-Thr-Trp-Ala-$CONH_2$, based on the essential glycosylation site at residue A285 of the hemagglutinin protein of influenza virus, was synthesized and derivatized with five different carbohydrates (**Figure 15**). The various conformations of the glycopeptides were then assessed using two-dimensional nuclear magnetic resonance (2D NMR) methods. The nonglycosylated peptide is found to be in a predominantly Asx-turn conformation.[64] This turn involves interactions between the asparagine side chain carbonyl and the peptide backbone (**Figure 14a**). β-Chitobiose, on the Asn side chain, induces a native β-turn structure (**Figure 14b**). The Asx and β-turns are topologically similar and are believed to have similar energies.[65] It is possible, therefore, that addition of a bulky substituent to this key asparagine residue could result in the predominance of the β-turn over the Asx turn. The Asx turn could become disfavoured in the glycosylated state for steric reasons; the bulky glycosylated asparagine side chain might no longer be in close proximity with the backbone. Additionally, the flexible sugar

that now adorns the asparagine side chain might disrupt the hydrogen bond between the backbone and asparagine side chain that characterizes an Asx turn.

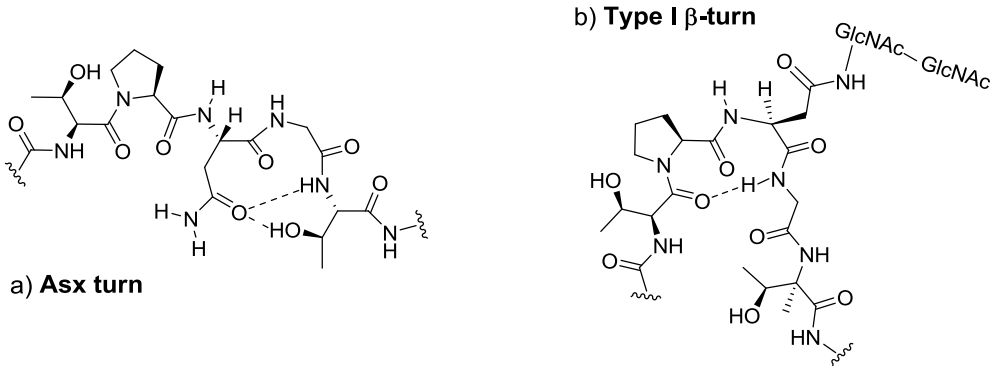


Figure 14. Comparison of the Asx turn and β-turn conformations in a Thr-Pro-An-Gly-Thr-sequence. The Asx turn is formed by the Asn side chain.

Neo-glycoconjugates, including non-natural carbohydrate moieties, were used to probe the role that the sugar plays in stabilizing a β-turn conformation: derivatizations, using β-*N*-linked GlcNAc-GlcNAc, Glc-GlcNAc, GlcNAc, GlcNAc-Glc and Glc-Glc, revealed that the *N*-acetyl group of the proximal sugar is critical for maintaining a β-turn conformation (**Figure 15**). Chitobiose (GlcNAc-GlcNAc) plays a unique role in modulating peptide conformation (**Figure 15-a**); no other saccharide appears to induce such a native β-turn structure. Glycopeptide modified with the disaccharide Glc-Glc, (cellobiose) (**Figure 15-e**), exhibited a half-turn structure with an extended backbone, showing that the cellobiose-derivatized glycopeptide **e**, without *N*-acetyl groups, is more mobile than chitobiose-derivatized glycopeptide **a**. Surprisingly, the additional *N*-acetyl group in the distal sugar (GlcNAc-Glc **Figure 15-d**) or in the proximal sugar appears to play a key role in changing the conformation. Addition of the *N*-acetyl group to the proximal sugar in glycopeptide **b** (Glc-GlcNAc) promoted formation of a β-turn, analogous to that observed for glycopeptide **a (**GlcNAc-GlcNAc), but less compact than in peptide **a**. Glycopeptide **c** (GlcNAc) was also examined: although the NOE data indicate that β-turn character is present in this peptide, the carboxy-terminal region of the turn is highly disordered. β-turn structure appears to be more prevalent in glycopeptide **b** (Glc-GlcNAc) than in glycopeptide **c** (GlcNAc). The NMR analysis of glycopeptide **d** (GlcNAc-Glc), in which only the distal sugar is derivatized with an *N*-acetyl moiety, supports the notion that the proximal *N*-acetyl group is most critical for β-turn formation. Also in this case NOE studies suggest that peptide **d** adopts a conformation in which the carboxy-terminal residues of the β-turn are disordered but despite this, it still displays β-turn-like character, in contrast to glycopeptide **e** which is derivatized with the cellobiose disaccharide that lacks both *N*-acetyl groups. Taken together, these data suggest that although the *N*-acetyl group of the proximal

sugar is critical for compact β-turn formation, the *N*-acetyl of the distal sugar also plays a role in modulating peptide conformation. It should also be noted that the glycopeptide derivatized with **Gal**-GlcNAc failed to produce a β-turn conformation, and instead adopted an extended structure that resembled that of glycopeptide **e**, suggesting a highly specific carbohydrate conformation.

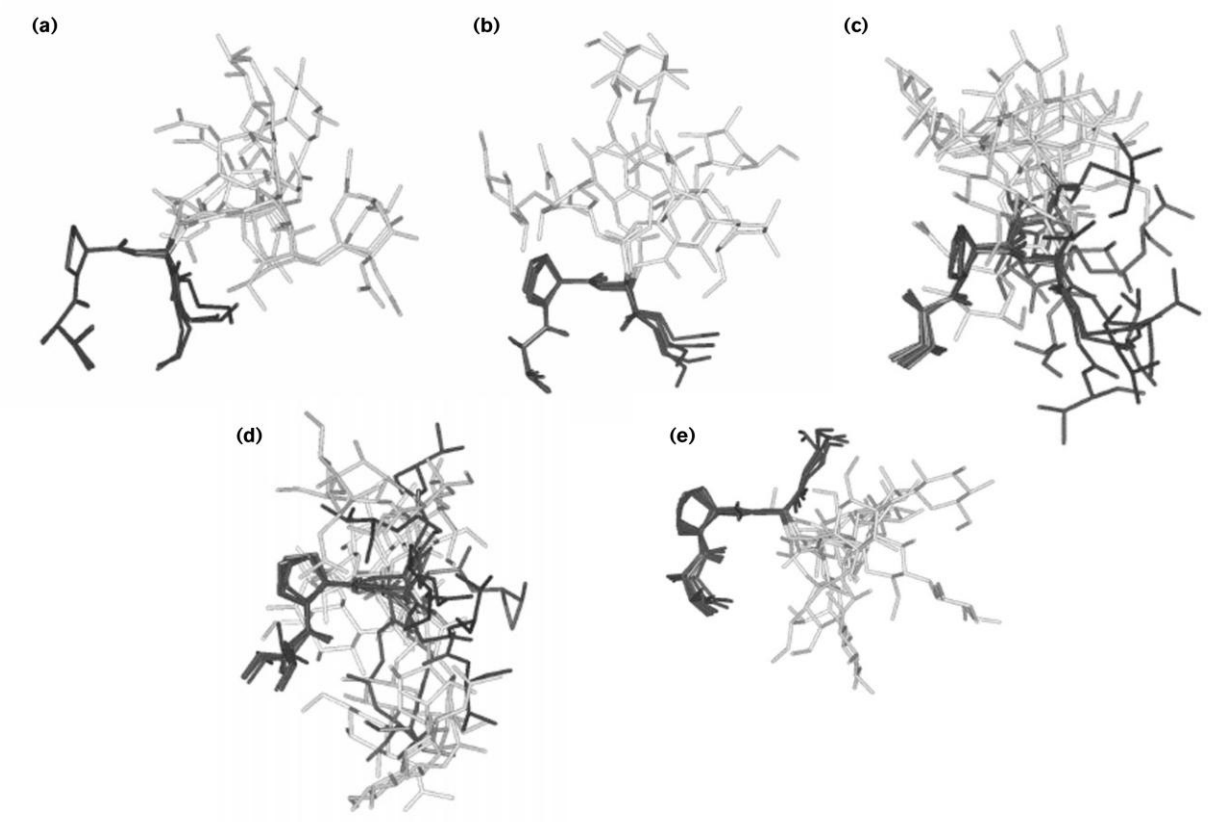


Figure 15. NMR structures of peptides from hemagglutinin protein derivatized with various saccharides (peptide depicted in dark gray, saccharide shown in light gray). **(a)** GlcNAc-(β1-4)-GlcNAc-βAsn. **(b)** Glc-(β1-4)-GlcNAc-βAsn. **(c)** GlcNAc-βAsn. **(d)** GlcNAc-(β1-4)-Glc-βAsn. **(e)** Glc-(β1-4)-Glc-βAsn. (Figure from *Ref 54*: *Curr. Opin. Chem. Biol.* **1999**, *3*, 643-649).

Later on, Imperiali et al.[66] reported a detailed comparison for an α- and a β-linked glycopeptide and the corresponding unglycosylated peptide (**Figure 16**). Two-dimensional NMR studies were used to evaluate the conformational properties of the new α-linked glycopeptide and to compare it with the properties of the unglycosylated peptide, as well as the β-linked glycopeptide, reported previously. The NMR studies revealed that the stereochemistry at the anomeric center of the *N*-linked carbohydrate dramatically affects the backbone conformation of the glycopeptide. Indeed, only the β-linked glycopeptide adopts a compact β-turn conformation.

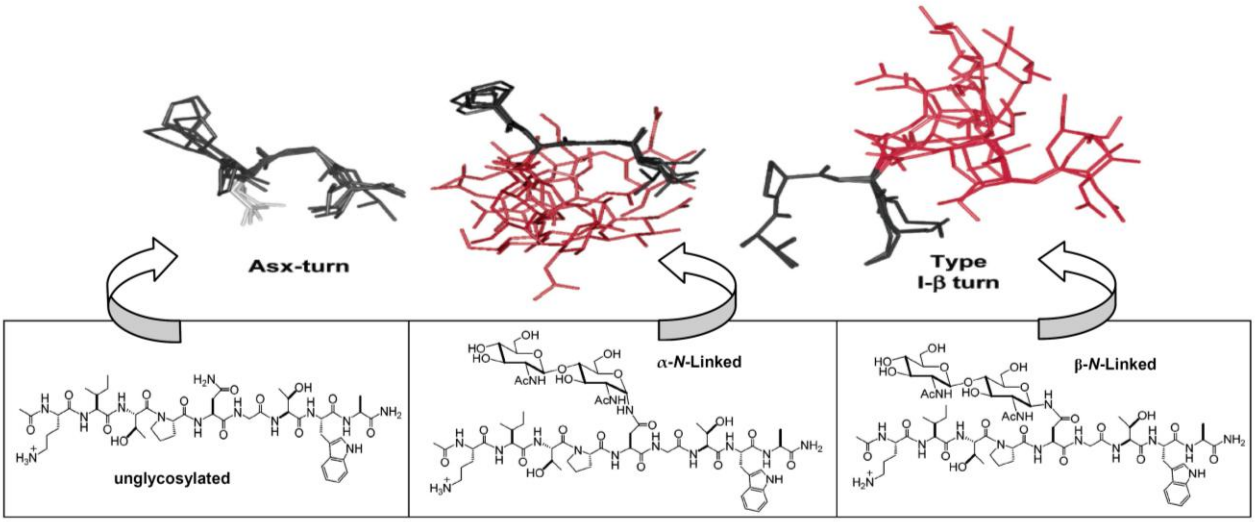


Figure 16. Comparison of the solution-state structures between unglycosylated peptide, α-linked glycopeptide, and *β*-linked glycopeptide.

The conformation of the α-*N*-linked glycopeptide was found to be more similar to that of the unglycosylated peptide, which has an Asx-turn structure, than to that of the *β*-*N*-linked glycopeptide. The three systems have been simultaneously subjected to computational analysis involving MD simulations utilizing explicit water solvation. The findings of these studies are in excellent agreement with the experimental solution-state conformational analysis. The conformational consequences of the stereochemistry of the anomeric center in the carbohydrate-peptide linkage reported in this study may provide valuable information for the design of neoglycopeptides and glycopeptide mimic that may be useful as therapeutic agents.

This is in fact, to our knowledge, the unique study that examined the conformation of an unnatural α-*N*-linked glycopeptide, which was found to behave in a different fashion both from the nonglycosylated and from the natural β-*N*-linked glycopeptide, supporting the idea that unnatural α-*N*-linked glycopeptides could provide new materials and could display new properties. This is what prompted us to establish appropriate methodologies for the synthesis of α-*N*-linked glycopeptides in an effort to establish a viable access to this interesting class of novel glycoconjugates.

1.2 References

[1] a) Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357-2364. b) Rudd, P. M.; Elliot, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291*, 2370-2376.

[2] Taylor, C.M. *Tetrahedron* **1998,** *54*, 11317-11362.

[3] Varki, A. *Glycobiology* **1993**, *3*, 97-130

[4] Laine, R. A. *Glycobiology* **1994**, *4*, 759.

[5] Gamblin, D. P.; Scanian E. M.; Davis, B. D. *Chem. Rev.* **2009**, *109*, 131-163.

[6] Dwek. R. A. *Chem. Rev*. **1996**, *96*, 683-720.

[7] Harris, R. J. *Biochemistry* **1993**, *32*, 6539-6547.

[8] Ferguson, M. A. J.; Williams, A. F. *Annu. Rev. Biochem.* **1988**, *1*, 522-529.

[9] Ferguson, M. A. J. *Curr. Opin. Struct. Biol.* **1991**, *1*, 522-529.

[10] Yamashita, K.; Kamerling, J. P.; A., K. *J. Biol. Chem.* **1982**, *257*, 12809-12814.

[11] Fukuda, M. *Mol. Glycobiol.* **1994**, 1-52.

[12] Haltiwanger, R. S.; Kelly, W. G.; Roquemore, E. P.; Blomberg, M. A.; Dong, L. D.; Kreppel, L.; Chou, T.; Hart, G. W. *Biochem. Soc. Trans.* **1992**, *20*, 264-269.

[13] Parekh, R. B.; Dwek, R. A.; Thomas, J. R.; Rademacher, T. W.; Opdenakker, G.; Wittwer, A. J.; Howard, S. C.; Nelson, R.; Siegel, N. R.; Jennings, M. G.; Harakas, N. K.; Feder, J. *Biochemistry* **1989**, *28*, 7644-7662.

[14] Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631-634.

[15] Hubbard, S. C.; Ivatt, R. J. *Annu. Rev. Biochem*. **1981**, *50*, 555-583.

[16] Wyss, D. F.; Wagner, G. *Curr. Opin. Biotechnol.* **1996**, *7*, 409-416.

[17] a) Jaeken, J.; van Eijk, H. G.; van der Heul, C.; Corbeel, L.; Eeckels, R.; Eggermont, E. *Clin. Chim. Acta* **1984**, *144*, 245-247. b) Jensen, H.; Kjaergaard, S.; Klie, F.; Moller, H. U. *Ophthalmic Genet.* **2003**, *24*, 81–8.

[18] Haeuptle M. A.; Hennet T. *Human Mutation* **2009**, *30*, 1628–1641.

[19] Sörensen A. L.; Reis, C. A.; Tarp, M. A.; Mandel, U.; Ramachandran, K.; Sankaranarayanan V.; Swientek T.; Graham, T.; Taylor-Papadimitriou, J.; Hollingsworth, M. A.; Burchell, J.; Clausen, H. *Glycobiology* **2006**, *16*, 96.

[20] Hollingsworth M. A.; Swanson B. J. *Nature Reviews* **2004**, *4*, 45-60.

[21] Hukill, P. B.; Vidone, R. A. *Lab. Invest.* **1965**, *14*, 1624–1635.

[22] Goldenberg, D. M.; Pegram, C. A.; Vazquez, J. J. *J. Immunol.* **1975,** *114*, 1008–1013.

[23] a) Piller, F.; Cartron, J. P.; Tuppy, H. *Rev. Fr. Transfus. Immunohematol.* **1980**, *23*, 599–611. b) Runge, R. G.; Pour, P. *Cancer Lett.* **1980**, *10*, 351–357. c) Feizi, T. *Med. Biol.* **1982**, *60*, 7–11.

[24] a) Bramwell, M. E.; Bhavanandan, V. P.; Wiseman, G.; Harris, H. *Br. J. Cancer* **1983**, *48*, 177–183. b) Burchell, J.; Durbin, H.; Taylor-Papadimitriou, J. *J. Immunol.* **1983**, *131*, 508–513. c) Gendler, S. J. *Proc. Natl Acad. Sci. USA* **1987**, *84*, 6060–6064. d)Gendler, S. J. *J. Biol. Chem.* **1990**, *265*, 15286–15293.

[25] Rughetti, A.; Pellicciotta, I.; Biffoni, M.; *J Immunol.* **2005**, *174*, 7764–72.

[26] a) Tashiro, Y.; Yonezawa S.; Kim, S. *Human Pathol.* **1994**, *25*, 364-372.

[27] a) Westerlind U.; Hobel, A.; Gaidzik, N.; Schmitt, E.; Kunz H. *Angew. Chem. Int. Ed*. **2008**, *47*, 7551-7556. b) Dziadek, S.; Hobel A.; Schmitt E.; Kunz H. *Angew. Chem. Int. Ed.* **2005**, *44*, 7630-7635.

[28] Kaiser, A.; Gaidzik, N.; Westerlind, U.; Kowalczyk, D.; Hobel, A.; Schmitt, E.; Kunz, H. *Angew. Chem. Int. Ed.* **2009**, *48*, 7551-7555.

[29] Spohn, R.; Buwitt-Beckmann, U.; Brock, R.; Jung, G.; Ulmer, A. J.; Wiesmöller, K.-H. *Vaccine* **2004**, *22*, 2494-2499.

[30] Lu, H.; Goodell, V.; Disis, M. L. *J Proteome Res* **2008**, *7*, 1388–94.

[31] Wandall, H. H.; Blixt, O.; Tarp, M. A.; Pedersen, J. W.; Bennett, E. P,: Mande, U.; Ragupathi, G.; Livingston, P. O.; Hollingsworth, M. A.; Taylor-Papadimitriou, J.; Burchell, J.; Clausen H. *Cancer Res.* **2010**, *70*, 1306-1313.

[32] Tarp, M. A.; Sorensen, A. L.; Mandel, U. *Glycobiology* **2007**, *17*, 197–209.

[33] Whetten-Goldstein, K.; Sloan, F. A.; Goldstein, L. B.; Kulas, E. D. *Mult. Scler.* **1998**, *4*, 419-425.

[34] Lolli, F.; Mulinacci, B.; Carotenuto, A.; Bonetti, B.; Sabatino, G.; Mazzanti, B.; D'Ursi, A. M.; Novellino, E.; Pazzagli, M.; Lovato, L.; Alcaro, M. C.; Peroni, E.; Pozo-Carrero, M. C.; Nuti, F.; Battistini, L.; Borsellino, G.; Chelli, M.; Rovero, P.; Papini, A. M. *Proc. Natl. Acad. Sci.* **2005**, *102*, 10273-10278.

[35] Carotenuto, A.; D'Ursi, A. M.; Mulinacci, B.; Paolini, I.; Lolli, F,; Papini, A. M.; Novellino, E.; Rovero, P. *J. Med. Chem.* **2006,** *49*, 5072-5079.

[36] Alcaro, M. C.; Chelli, M.; Lolli, F.; Migliorini P.; Paolini, I.; Papini, A. M.; Rovero, P. **Patent** EP2050761 A1, **2009**

[37] Chan, D. C.; Fass, D.; Berger, J. M.; Kim P. S. *Cell* **1997**, *89*, 263–273.

[38] McMichael, A. J.; Hanke, T. *Nat. Med.* **2003**, *9*, 874 – 880.

[39] Wei, X.; Decker, J.M.; Wang, S.; Hui, H.; Kappes, J. C.; Wu, X.; Salazar-Gonzalez, J. F.; Salazar, M. G.; Kilby, J. M.; Saag,M. S.; Komarova, N. L.; Nowak, M. A.; Hahn, B. H.; Kwong, P. D.; Shaw, G. M. *Nature* **2003**, *422*, 307–312.

[40] Doores, K. J.; Fulton, Z.; Hong, V.; Patel, M. K.; Scanlan, C N.; Wormald, M. R.; Finn, M. G.; Burton, D. R.; Wilson, I. A.; Davis, B.G. *Proc. Natl. Acad. Sci.* **2010**, *107*, 17107-17112.

[41] Sanders, R. W.; Venturi, M.; Schiffner, J.; Kalyanaraman, R.; Katinger, H.; Lloyd, K. O.; Kwong, P. D.; Moore. J.P. *J. Virol.* **2002**, *76*, 7293-7305.

[42] a) Mandal, M.; Dudkin, V. Y.; Geng, X.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2004**, *43*, 2557 –2561. b) Wilson R. M.; Danishefsky S. J. *Pure Appl. Chem.* **2007**, *79*, 2189–2216.

[43] van Montfort, T.; Eggink, D.; Boot, M.; Tuen, M.; Hioe, C.E.; Berkhout, B.; Sanders, R.W. *J. Immunol*. **2011**, 4676-4685.

[44] a) Ferguson, M. A. J.; Williams, A. F. *Annu. Rev. Biochem.* **1988**, *57*, 285-320. b) Gowda, D. C.; Davidson, E. A. *Parasitol. Today* **1999**, *15*, 147-152.

[45] Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. *Angew. Chem. Int. Ed*. **2005**, *44*, 7342 – 7372.

[46] Schofield, L.; Hewitt, M. C.; Evans, K.; Siomos M.; Seeberger P. H. *Nature* **2002**, *418*, 785-789.

[47] Becker, C. F. W.; Liu, X.; Olschewski, D.; Castelli, R.; Seidel, R.; Seeberger P. H. *Angew. Chem. Int. Ed*. **2008**, *47*, 8215 –8219.

[48] Stahl, N.; Borchelt, D. R.; Hsiao, K.; Prusiner, S. B. *Cell* **1987**, *51*, 229–240.

[49] a) Lewis, P. A.; Properzi, F.; Prodromidou, K.; Clarke, A. R.; Collinge, J.; Jackson, G. S. *Biochem. J.* **2006**, *395*, 443–448. b) Legname, G.; Baskakov, I. V.; Nguyen, H. O. B.; Riesner, D.; Cohen, F. E.; DeArmond, S. J.; Prusiner, S. B. *Science* **2004**, *305*, 673–676.

[50] Sola, R. J.; Griebenow, K. *J. Pharm. Sci.* **2009**, *98*, 1223–1245.

[51] a) Hanson S. R.; Culyba, E. K.; Hsu, T.; Wong, C.; Kelly, J. W.; Powers, E. T. *Proc Natl Acad Sci USA* **2009**, *106*, 3131–3136. b) Shental-Bechor, D.; Levy Y. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8256–8261. c) Petrescu, A. J.;

Milac, A. L.; Petrescu, S. M.; Dwek, R.A.; Wormald, M. R.; *Glycobiology* **2004**, *14*, 103–114. d) Shental-Bechor, D.; Levy, Y. *Curr. Opin. Struct. Biol*. **2009**, *19*, 524–533.

[52] O'Connor, S. E.; Imperiali, B. *Chemistry & Biology* **1996**, *3*, 803-812.

[53] a) Riederer, M. A.; Hinnen, A. *J. Bacteriol*. **1991**, *173*, 3539-3546. b) Marquardt, T.; Helenius, A. *J. Cell Bio*. **1992**, *117*, 505-513. c) Copeland, C. S.; Zimmer, K-P.; Wagner, K. R.; Healey, G. A.; Mellman, I.; Helenius, A. *Cell* **1988**, *53*, 197-209.

[54] Imperiali, B.; O'Connor, S. E. *Curr. Opin. Chem. Biol*. **1999**, *3*, 643-649.

[55] Arnold, U.; Ulbrich-Hofmann, R. *Biochemistry* **1997**, *36*, 2166-2172.

[56] Wang, C.; Eufemi, M.; Turano, C.; Giartosio, A. *Biochemistry* **1996**, *35*, 7299-7307.

[57] Hansen, T. N.; Carpenter, J. F. *Biophys. J.* **1993**, *64*, 1843-1850.

[58] Buskas, T.; Ingale, S.; Boons, G. J. *Glycobiology* **2006**, *16*, 113R–136R.

[59] Hamilton Andreotti, A.; Kahne, D. *J. Am. Chem. Soc.* **1993,** *115*, 3352-3353.

[60] Liang, R.; Hamilton Andreotti, A.; Kahne, D. *J. Am. Chem. Soc*. **1995,** *117*, 10395-10396.

[61] Matthews, C. R. *Annu. Rev. Biochem.* **1993**, *62*, 653-683.

[62] Wormald, M. R.; Dwek, R. A. *Struct. Fold. Des.* **1999**, *7*, R155–R160. b) Yamaguchi, H. *Trends Glycosci. Glycotechnol*. **2002**, *14*, 139–151.

[63] O'Connor, S. E.; Imperiali, B. *Chemistry & Biology* **1998**, *5*, 427-437.

[64] O'Connor, S. E.; Imperiali, B. *J. Am. Chem. Soc.* **1997**, *119*, 2295-2296.

[65] Abbadi, A.; Mcharfi, M.; Aubry, A.; Premilat, S.; Boussard, G.; Marraud, M. *J. Am. Chem. Soc.* **1991**, *113*, 2729-2735.

[66] Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B. *J. Am. Chem. Soc.*, **2004**, *126*, 8421-8425.

Chapter 2

Synthesis of natural and unnatural glycopeptides

2.1. Strategies for the synthesis of natural glycopeptides

Chapter 1 described how the inherent structural diversity of glycoproteins and glycopeptides reflect an enormous range of functions within nature and how variations in oligosaccharide structures cause alteration of many properties of glycopeptides as conformation, stability and interaction with receptors. It is therefore fundamental, also for the design of new glycopeptide structures, the correct identification of associated functions and features of existing natural structures. This process is complicated since naturally expressed glycoproteins often arise as heterogeneous mixtures.[1] In the light of these considerations, there was the necessity to find procedures for having homogeneous synthetic glycoproteins and glycopeptides as model systems. The developed methodologies covered the fields of chemical, enzymatic and molecular biology approaches, used alone or in association with one another.

The synthesis of glycopeptides clearly needs a combination of methods from both carbohydrate and peptide chemistry, with the main scope of forming the integral glycan-amino acid bond. **Figure 1** shows the general retrosynthetic analysis for the formation of this linkage: disconnection A relies on the conjugation of the sugar unit with a preformed full-length peptide and the procedure is called "convergent" assembly. In disconnection B instead, a preformed glycosylated amino acid building block is employed in a stepwise assembly of the glycopeptide and this is exactly called "linear" assembly.

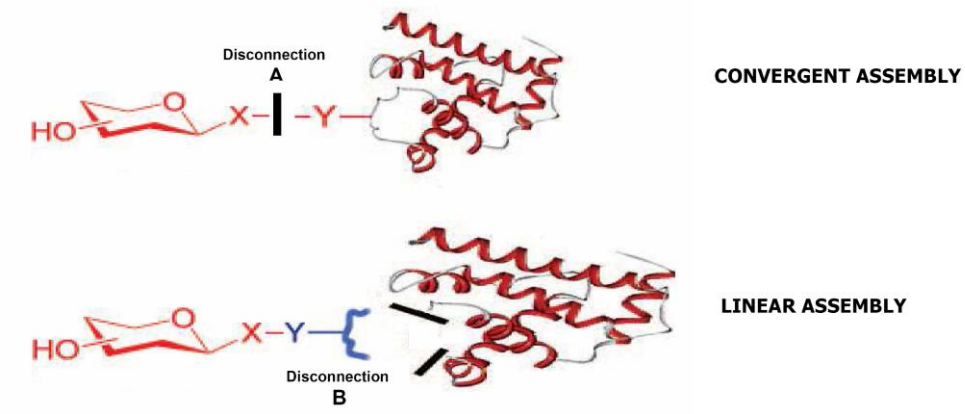


Figure 1. Disconnections strategies for glycopeptides synthesis

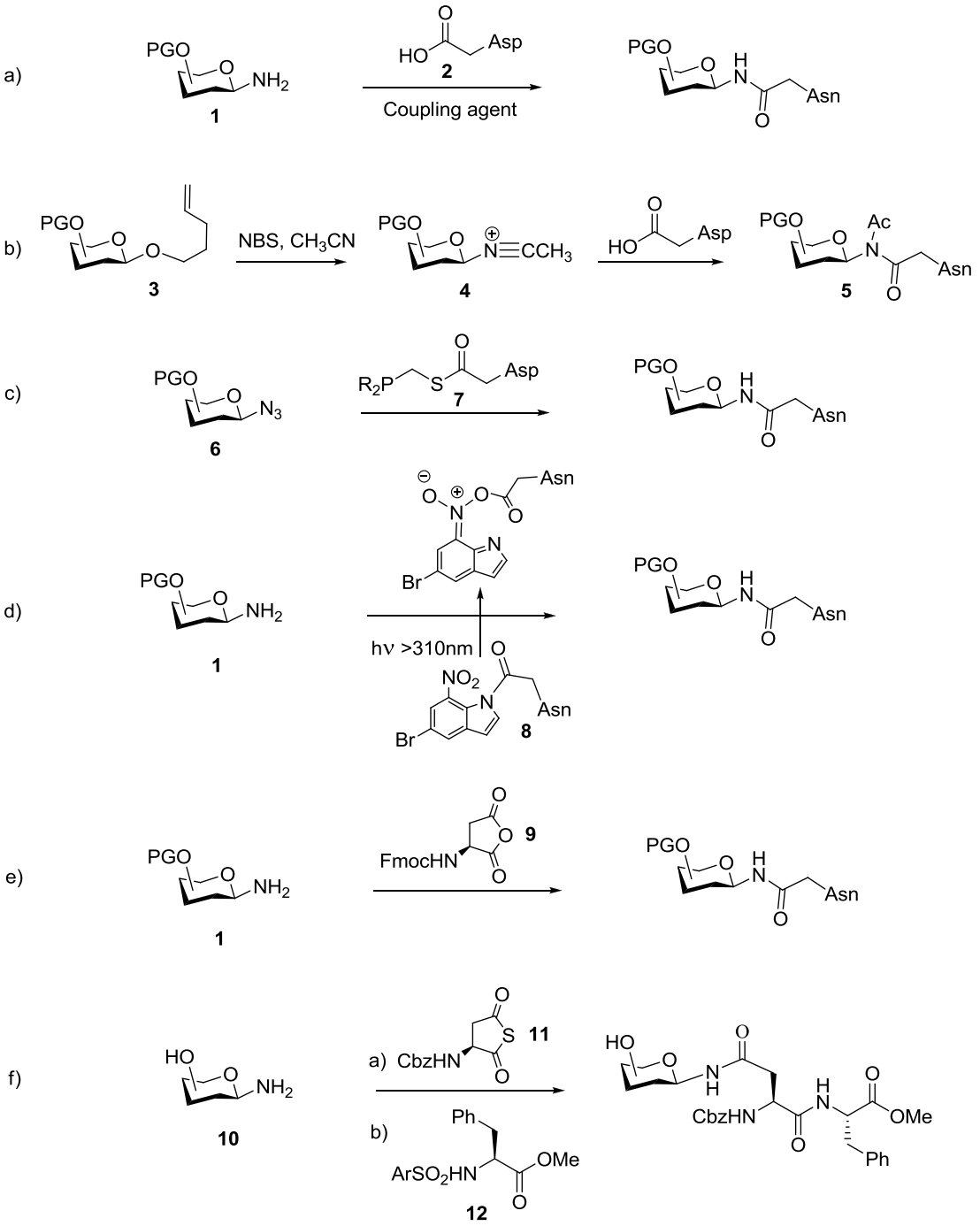
For these two strategies there can be identified common issues as the acid and base lability during global deprotection of the resulting side-chain protected glycopeptides, and individual advantages and disadvantages. Direct convergent peptide glycosylation (Path A), for instance, often suffers from low yields, because of the efficiency of the key coupling step, which involves an oligosaccharyl-amine and a protected polypeptide. Despite this fact, a number of large and complex glycopeptides have been constructed by this strategy.[2] Glycosylamines are accessible by the

Kochetkov reaction where a fully protected reducing sugar or a deprotected sugar is treated with 50 times excess of ammonium bicarbonate for 6 days. Flitsch et al. showed that this procedure can be accelerated using microwave irradiation.[3] Lansbury and coworkers, for the first time, applied glycosylamines and more complex glycans coupling with the aspartate carboxylate side chain in a pentapeptide, which allowed formation of an Asn-β-linked *N*-acetylglucosaminyl-containing glycopeptide.[4] Recently Danishefsky used a sequence consisting of Kochetkov amination,[5] followed by Lansbury aspartylation[6] and then native chemical ligation[7] as a route to complex *N*-linked polypeptides.[8] Because glycosylamines usually equilibrate to the β-anomer, all the approaches that make use of amine intermediates afford β-glycosylated glycopeptides. Davis et al. accessed to a candidate epitope of a tumor antigen through convergent attachment of sugar to polypeptide, employing Staudinger ligation. The carboxylate side chain of the aspartic acid was activated first with HBTU and HOBt followed by addition of GlcNAc azide and Bu_3P.[9]

The linear approach with glycosyl amino acids is for sure the most versatile method, because it is suitable for solid phase synthesis even if it is limited by two factors: the requirement for extensive carbohydrate (and also amino acid) protection and then the acid and base instability of glycosylated amino acid fragments.[10] Especially for *O*-glycopeptides, in fact, the *O*-glycosidic linkages in the oligosaccharide moiety are susceptible to acidic hydrolysis under strong acidic conditions (e.g. TFA) required for final global deprotection and for the release of peptide from the resin. As a consequence, solid phase synthesis technology is typically limited to synthesis of peptide sequences of up to 50 residues and furthermore, linear assembly is generally useful for preparing glycopeptides carrying relatively small oligosaccharides. A potential drawback of this approach is indeed that the bulky glycans attached to the building block may result in low-yield coupling in solid-phase or solution-phase peptide synthesis. Recent works, however, have also demonstrated that, when combined with native chemical ligation, this approach could also be appropriate for constructing some large *N*-glycopeptides.[11]

For the synthesis of *N*-linked glycans, the most common route is the reaction of a protected or unprotected glycosylamine **1** with an Asp derivative **2**, that could be a free amino acid[12] or an activated ester[13] (**Scheme 1a**). This is by far the most popular method even if it is often affected by the ready anomerization of the glycosylamine, and this may lead to the formation of anomeric mixtures of glycopeptides. The Ritter reaction (**Scheme 1b**) also leads to the synthesis of the key GlcNAc*β*-Asn linkage. Reaction of certain nitriles with oxonium ions generated from *O*-pentenyl glycosides **3** followed by reaction of the nitrilium intermediates **4** with the aspartic carboxylate side chain on a *β*-acetonitrilium followed by deacetylation (**Scheme 1b**) afforded *N*-acetyl-*N*-linked Asn derivatives **5**, which can be deacetylated to obtain the target molecule.[14] The Staudinger reaction of

glycosylazides **6** with Asp side chain carboxylates or carboxyl derivatives in the presence of PPh_3[15] also offers access to β-Asn *N*-glycosides.[16] Recently, Kiessling and co-workers described a Staudinger ligation for *N*-glycosylation using glycosylazides and asparagine-derived phosphinothioesters **7** (**Scheme 1c**).[17]

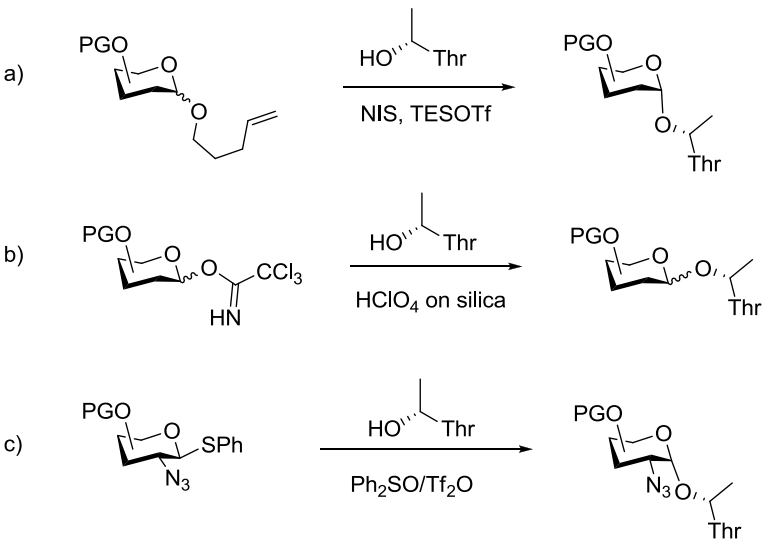


Scheme 1. Methods for *N*-linked glycopeptides preparation.

N-glycosyl asparagines have also been prepared via a mild photochemical coupling in which a photoreactive side chain amide of aspartate **8** was condensed with a glycosyl amine (**Scheme 1d**).[18]

Photolysis activates the γ-carboxylate, rendering it susceptible to nucleophilic attack with glycosyl amines. Another route is the use of *N*-Fmoc aspartic anhydride **9**, which has been employed by Selivanov et al. for the preparation of Fmoc protected β-*N*-linked-glycosyl-asparagine (**Scheme 1e**).[19] The regioselectivity of *N*-Fmoc-aspartic anhydride aminolysis with per-*O*-acetylated glycosyl amines varied greatly depending on the polarity of the reaction media. In less polar solvents, the isoasparagine derivative was the main product, whereas more polar solvents (DMSO was the best) increased the yields of desired glycosylated asparagines. Crich et al. modified this procedure by opening cyclic monothioanhydrides, providing a convenient entry into amido thioacids.[20] In particular, reaction of unprotected glycosyl amines **10** (**Scheme 1f**) with *N*-benzyloxycarbonyl-L-aspartic monothioanhydride **11**, followed by capture of the thioacid intermediate with *N*-sulfonyl amino acid derivatives **12** results in a three-component convergent synthesis of glycosylated dipeptides.

For *O*-linked glycans the *O*-glycosidic bond is normally created through glycosylation of the hydroxyl group of a suitably protected Ser or Thr using glycosyl donors (**Scheme 2**).[21]

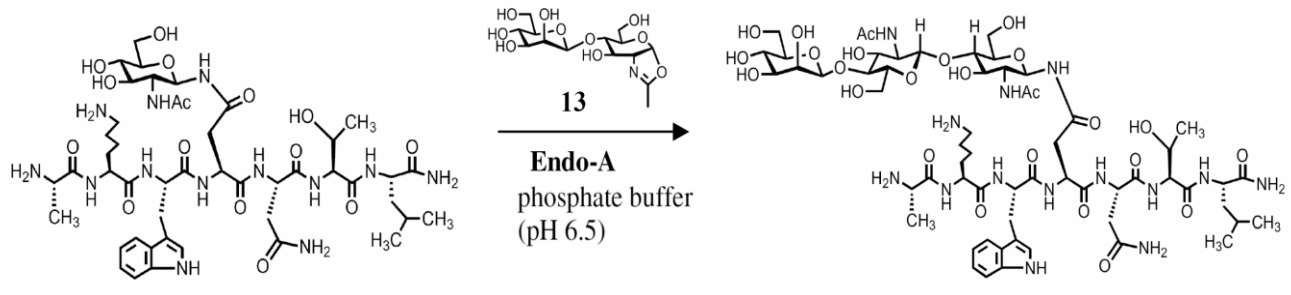


Scheme 2. Methods for *O*-linked glycopeptides preparation.

Use of *O*-pentenyl glycosides as glycosyl donors, for instance, has recently been reviewed (**Scheme 2a**).[22] Field and co-workers describe a recent innovation in the use of trichloroacetimidate (TCA) donors in which, in an 'on-column' approach, the Lewis-acid-catalyzed activation of the TCA occurs with perchloric acid immobilized on silica (**Scheme 2b**).[23] Thioglycosides were also used as donors by Boons et al. and activated with a Ph_2SO/Tf_2O promoter system in a stereospecific preparation of a T_N antigen building block. (**Scheme 2c**).[24]

An alternative strategy for glycopeptide synthesis is to combine enzymatic elaboration of sugar chains with chemical polypeptide synthesis. The chemoenzymatic approaches require the preparation of monosaccharide-tagged polypeptides and the extension of the sugar chain is accomplished using glycosidases and/or glycosyltransferases catalyzed reactions, often performed in aqueous solutions with free (without protecting groups) polypeptides. Thus, this strategy in principle avoids the problems associated with chemical glycopeptide synthesis, such as the incompatibility of protecting group manipulations for glycosylation and lability to final global deprotection. Both glycosyltransferases and endoglycosidases have been taken into consideration for elaboration of the glycan. Common glycosyltransferases can extend sugar chains by appending monosaccharides one by one.[25] In contrast, endoglycosidase catalyzed transglycosylation reactions can attach a large oligosaccharide to a GlcNAc-polypeptide in a single step, affording a highly convergent approach.[26] Endo-β-*N*-Acetylglucosaminidases (ENGases) are a category of endoglycosidases, which release *N*-glycans from glycoproteins by hydrolyzing the β-(1-4)-glycosidic bond in the *N,N'*-diacetylchitobiose core. Beside hydrolytic activity, several enzymes of this class have been found also to have a transglycosylation activity, i.e. the ability to transfer the released oligosaccharyl portion to an appropriate acceptor to form a new glycosidic bond. In this way, a new glycopeptide can be constructed by a two-step approach: First, a GlcNAc-containing polypeptide would be synthesized and then an intact oligosaccharide would be transferred to the acceptor by a suitable ENGase to give the new glycopeptide. The most employed ENGase are the bacterial enzyme Endo-A and the fungal enzyme Endo-M, which have distinct substrate specificity. Endo-A is specific for high-mannose type *N*-glycans while Endo-M can operate on *N*-glycans of the three major types (high-mannose type, hybrid type, and complex type), with preference to the complex type sugars. Previous studies have established that Endo-A can recognise both GlcNAc and Glc moiety, attached to a peptide, as acceptors for transglycosylation.[27] It was also shown that a C-linked GlcNAc-peptide could also serve as an acceptor substrate for Endo-A to form a *C*-glycopeptide.[28] A series of bioactive, large glycopeptides, difficult to obtain by other means, have been synthesized by this chemoenzymatic approach. Typical examples include large HIV-1 envelope glycoprotein fragments,[29] and homogeneous CD52 antigens carrying full-size high-mannose and complex type N-glycans.[30] However, the ENGase-based chemoenzymatic approach has met with several problems that have prevented its broad application: a disadvantage is the relatively low transglycosylation efficiency and the issue of product hydrolysis. In fact, ENGases are inherently glycosyl hydrolases and, in comparison with their hydrolytic activity, their transglycosylation activity is relatively low. Thus product hydrolysis becomes a significant problem when the product is accumulated because the resulting glycopeptide is also a substrate for the

enzyme. Incorporation of organic solvents in the reaction medium can enhance the transglycosylation yield to some extent,[31] but the overall efficiency is nonetheless low (5–20% yields). Another drawback was found to be the restriction to the use of natural *N*-glycans or *N*-glycopeptides as only donor substrates for transglycosylation, since these natural substrates themselves are difficult to obtain. Recent exploration of synthetic sugar oxazolines (enzymatic reaction intermediates) as synthetic substrates for the transglycosylation has partially answered to these problems and made the chemoenzymatic method more efficient.[32] For instance, Wang et al. synthesized a disaccharide oxazoline **13** (**Scheme 3**), corresponding to the core of *N*-glycans, and tested Endo-A and Endo-M catalyzed transglycosylation using a GlcNAc-heptapeptide derived from HIV-1 gp120 as the acceptor.[33] It was found that the synthetic oxazoline could serve as a good substrate for transglycosylation catalyzed by Endo-A to form the corresponding glycopeptide and that the resulting glycopeptides were completely resistant to hydrolysis catalyzed by Endo-A under reaction conditions. The great difference in enzymatic reaction rates between the activated oxazoline substrates and the final glycopeptide substrate favors the accumulation of the transglycosylation product. This chemoenzymatic method was then successfully applied for the synthesis of a novel triazole-linked glycopeptide derived from the HIV-1 envelope glycoprotein, gp41.[29b]



Scheme 3. Use of oxazolines as substrate for Endo–A.

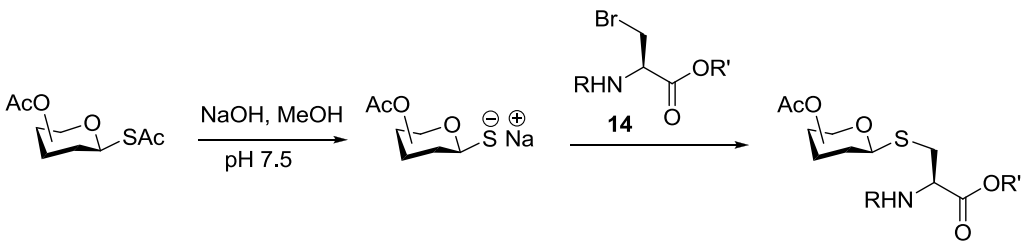
Another class of carbohydrate processing enzymes that also hold great potential for in vitro glycoprotein synthesis are the oligosaccharyl transferases (OST), i.e. the enzymes that transfer an oligosaccharide precursor to the asparagine side chain of the nascent protein during translation in *N*-glycoprotein biosynthesis.[34] However, a practical application of OST for in vitro glycoprotein synthesis has not yet been developed, mainly because of the complexity and instability of the multiple-subunit complex of the enzymes. A promising approach, also able to address the low efficiency and product hydrolysis, which plague glycosidase–based methods, is to create novel glycosidase mutants called glycosynthases. This mutants lack hydrolytic activity because the active site nucleophilic residue has been deleted, but they can still accept an activated glycosyl donor to

form a new glycosidic bond.[35] A series of Endo-M mutants were created by Yamamoto's and Wang's groups by site-directed mutagenesis on residues in, or around, the putative catalytic region of wild type Endo-M.[36] Then their transglycosylation activity was examined using synthetic sugar oxazolines as activated substrates. The experiments led to the discovery of two interesting mutants which presented a much enhanced transglycosylation activity and yet relatively low hydrolytic activity.

2.2 Synthesis of unnatural glycopeptides and glycopeptide mimics

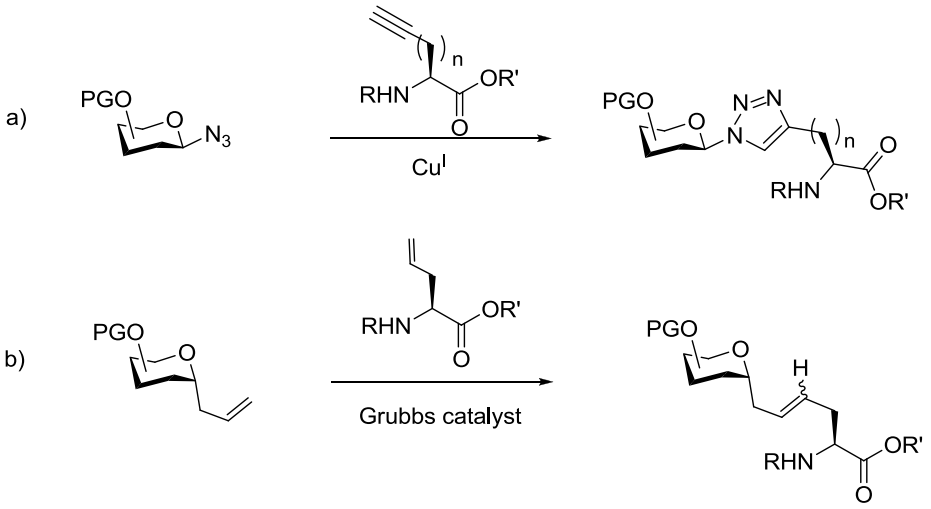
The data described so far show how oligosaccharides are attached to proteins mainly through the hydroxyl group of serine and threonine (*O*-glycans) and the amido group of asparagine (*N*-glycans). Replacement of these linkages with sulfur or carbon linkages, affording *S*- or *C*-linked glycopeptides could lead to modifications that may be tolerated by most biological systems. These mimics therefore, are interesting candidates to develop drugs and biological probes, useful to understand the biological roles of glycopeptides.

Thioglycopetides, which are closely related to *O*-glycopeptides, are relatively easy to access due to the high nucleophilicity of the thiol and thiolate groups compared to the corresponding hydroxyl groups: thioglycoside bond formation can be readily based on SN_2 displacement of halogenoses. This reaction is particularly useful for equatorial thioglycoside bond formation since the halogen atom commonly adopts the axial position on halogenoses.[37] Several methods such as reaction with sulfamidates,[38] Michael addition,[39] aziridine opening,[40] cysteine alkylation[41] were employed for the generation of these mimics. Wong et al.[42] have introduced a two-step one-pot reaction of a bromoalanine derivative **14** with in situ generated sugar thiolate to synthesize a thio-linked glycosyl amino acid product (**Scheme 4**). This method was used in the synthesis of cyclo-glycopeptide mimic of tyrocidin, a cyclic cationic decapeptide antibiotic. The *S*-linked analogue displayed a greater inhibitory activity against *Bacillus subtilis* than the natural antibiotic.[42]



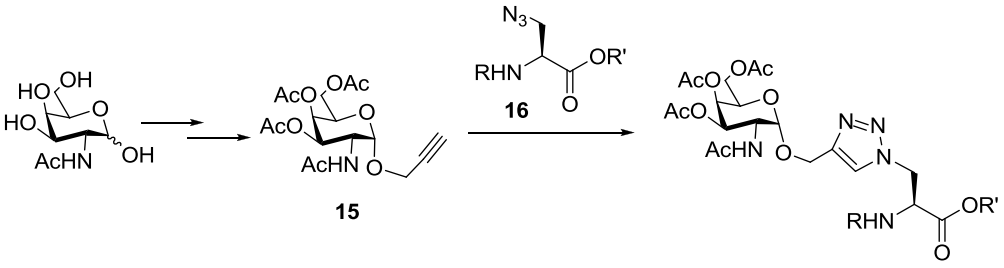
Scheme 4. *S*-linked glycopeptide synthesis with glycosyl thiolates and bromoalanine.

C-linked glycopeptides are also a family of interesting glycopeptide mimics that could provide improved activity and show chemical and metabolic resistance when compared to the natural counterpart. "Click" chemistry [43] (**Scheme 5a**) and olefin cross metathesis[44] (**Scheme 5b**) are only two strategies among the numerous ones [45] that have been employed for the synthesis of these analogues. In the "click" approach, for the formation of triazoles, the triple bond could also be part of the sugar, while the azide could be installed on the peptide part as shown recently by Wang and co-workers.[29b]



Scheme 5. *C*-linked glycopeptide synthesis.

Brimble et co-workers performed a microwave-enhanced click glycoconjugation of a *O*-propargyl α-GalNAc **15** with an azido-functionalized amino acid **16** or multiazido-functionalized peptides (**Scheme 6**).[46] This procedure allowed the synthesis of a T_N-antigen mimic and of click analogues of antifreeze glycopeptides, thus demonstrating a valuable method for the synthesis of biologically relevant neoglycopeptides.

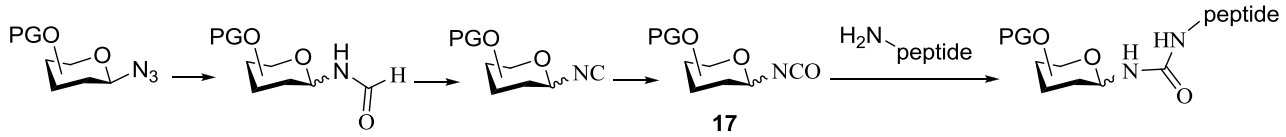


Scheme 6. Synthesis of Neo triazole linked glycosyl Amino Acid Building Block.

C-linked glycopeptides may be poor mimic of the 3D structure of a native glycoside. In a number of cases it has been shown that the conformational flexibility of *C*-linked disaccharides differs from that of *N*-linked sugars.[47-48] *C*-glycosylation may alter the native chair conformation of the sugar,

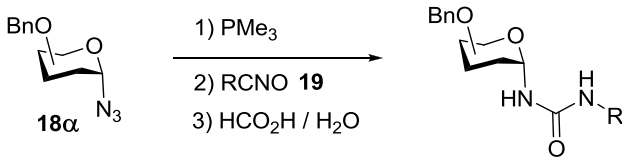
resulting in poor 3D similarity. For instance, a conformational analysis by NMR revealed that a triazolyl-methyl-C-mannoside, reported by Ernst et al., adopted an unusual $^{1}C_4$ chair conformation and, as a consequence, showed a low affinity in the binding with the natural lectin.[48]

Glycosyl ureas have also been used as stable *N*-linked-glycopeptide mimics,[49] for the synthesis of glycosyl-amino acid conjugate.[50] However, only a few methods for the synthesis of glycosyl ureas have been reported.[51] One example is the reaction of an amine with a glycosylisocyanates **17** prepared by oxidation of isonitriles, in turn prepared from anomeric formamides (**Scheme 7**).[52]



Scheme 7. Synthesis of urea-linked glycopeptides.

Another example, developed in our laboratories,[53] allows the stereoselective synthesis of α-glycosyl ureas by treatment of the corresponding azides **18α** with PMe_3 and isocyanates **19** (**Scheme 8**), using a convenient one-pot procedure.



Scheme 8. Synthesis of *N*-tetra-*O*-benzyl-α-D-glucopyranosyl ureas from α-glucopyranosyl azide **18α**.

Preliminary data are also available for the synthesis of α-*N*-linked glycosyl amino acids. They are discussed in the following chapter.

2.3 References

[1] Simanek, E. E.; McGarvey, G. J.; Jablonowski, J. A.; Wong, C. H. *Chem. Rev.* **1998**, *98*, 833-862.

[2] a) Mandal, M.; Dudkin, V. Y.; Geng, X.; Danishefsky, S. J. *Angew. Chem., Int. Ed*. **2004**, *43*, 2557–2561. b) Geng, X.; Dudkin, V. Y.; Mandal, M.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2004**, *43*, 2562–2565. c) Warren, J. D.; Miller, J. S.; Keding, S. J.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 6576–6578.

[3] Bejugam, M.; Flitsch, S. L. *Org. Lett.* **2004**, *6*, 4001-4004.

[4] Ansfield, S. T.; Lansbury, P. T. *J. Org. Chem.* **1990**, *55*, 5560-5562.

[5] Likhosherstov, L. M.; Novikova, O. S.; Derevitskaya, V. A.; Kochetkov, N. K. *Carbohydr. Res.* **1986**, *146*, C1-C5.

[6] Cohen-Anisfeld, S. T.; Lansbury, P. T., Jr *J. Am. Chem. Soc.* **1993**, *115*, 10531-10537.

[7] Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776-779.

[8] Dudkin, V. Y.; Miller, J. S.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 736-738.

[9] Doores, K. J.; Mimura, Y.; Dwek, R. A.; Rudd, P. M.; Elliott, T. E.; Davis. B. G. *Chem. Commun* **2006**, *13*, 1401–1403.

[10] Kunz, H. *Angew. Chem., Int. Ed Engl.* **1987**, *26*, 294-308.

[11] Wan, Q.; Chen, J.; Yuan, Y.; Danishefsky, S. J. *J. Am Chem. Soc* **2008**, *130*, 15814-15816.

[12] a) Deras, I. L.; Takegawa, K.; Kondo, A.; Kato, I.; Lee, Y. C. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1763-1766. b) Van Ameijde, J.; Albada H. B.; Liskamp R. M. J. *J. Chem. Soc., Perkin Trans. 1* **2002**, 1042–1049.

[13] Pratt. M. R.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2003**, *125*, 6149-6159.

[14] Handlon, A. L.; Fraser-Reid, B. *J. Am. Chem. Soc.* **1993**, *115*, 3796-3797.

[15] Inazu, T.; Kobayashi, K. *Synlett* **1993**, 869-870.

[16] Mizuno, M.; Muramoto, I.; Kobayashi, K.; Yaginuma, H.; Inazu, T. *Synthesis* **1999**, 162-165.

[17] He, Y.; Hinklin, R. J.; Chang, J.; Kiessling, L. L. *Org. Lett.* **2004**, *6*, 4479-4482.

[18] a) Vizvardi, K.; Kreutz, C.; Davis, A. S.; Lee, V. P.; Philmus, B. J.; Simo, O.; Michael, K. *Chem. Lett.* **2003**, *32*, 348-349. b) Kaneshiro, C. M.; Michael, K. *Angew. Chem., Int. Ed.* **2006**, *45*, 1077-1081.

[19] Ibatullin, F. M.; Selivanov S. I. *Tetrahedron Letters* **2009**, *50*, 6351–6354.

[20] Crich, D.; Sasaki, K.; Rahaman, M. Y.; Bowers A. A. *J. Org. Chem.* **2009**, *74*, 3886–3893.

[21] Shimawaki, K.; Fujisawa, Y.; Fumihiro, S.; Fujitani, N.; Masaki, K.; Hiroko, H.; Hiroshio, H.; Shin-Ichiro, N. *Angew. Chem., Int. Ed.* **2007**, *46*, 3074-3079.

[22] Svarovsky, S. A.; Barchi, J. J. *Carbohydr. Res.* **2003**, *338*, 1925-1935.

[23] Mukhopadhyay, B.; Maurer, S. V.; Rudolph, N.; van Well, R. M.; Russell, D. A.; Field, R. A. *J. Org. Chem.* **2005**, *70*, 9059-9062.

[24] Cato, D.; Buskas, T.; Boons, G. J. *J. Carbohydr. Chem.* **2005**, *24*, 503-516.

[25] Witte, K.; Sears, P.; Martin, R.; Wong, C. H. *J. Am. Chem. Soc.* **1997**, *119*, 2114–2118.

[26] Yamamoto, K. J. *Biosci. Bioeng.* **2001**, *92*, 493–501

[27] a) Deras, I. L.; Takegawa, K.; Kondo, A.; Kato, I.; Lee, Y. C. *Bioorg. Med. Chem. Lett*. **1998**, *8*, 1763–1766. b) Wang, L. X.; Tang, M.; Suzuki, T.; Kitajima, K.; Inoue, Y.; Inoue, S.; Fan, J. Q.; Lee, Y. C. *J. Am. Chem. Soc.* **1997,** *119*, 11137–11146.

[28] Wang, L. X.; Fan, J. Q.; Lee, Y. C. *Tetrahedron Lett*. **1996**, *37*, 1975–1978.

[29] a) Singh, S.; Ni, J.; Wang, L. X. *Bioorg. Med. Chem. Lett*. **2003**, *13*, 327–330. b) Huang, W.; Groothuijs, S.; Heredia, A.; Kuijpers, B. H. M; Ritjes, F.; van Delft, F.; Wang, L. X. *ChemBioChem* **2009**, *10*, 1234-1242.

[30] Li, H.; Singh, S.; Zeng, Y.; Song, H.; Wang, L. X. *Bioorg. Med. Chem. Lett*. **2005**, *15*, 895–898.

[31] Akaike, E.; Tsutsumida, M.; Osumi, K.; Fujita, M.; Yamanoi, T.; Yamamoto, K.; Fujita, K. *Carbohydr. Res.* **2004**, *339*, 719–722.

[32] a) Li, B.; Song, H.; Hauser, S.; Wang, L. X. *Org. Lett.* **2006**, *8*, 3081–3084. b) Rising, T. W.; Claridge, T. D.; Davies, N.; Gamblin, D. P.; Moir, J. W.; Fairbanks, A. J. *Carbohydr. Res.* **2006**, *341*, 1574–1596.

[33] Li, B.; Zeng, Y.; Hauser, S.; Song, H.; Wang, L. X. *J. Am. Chem. Soc*. **2005**, *127*, 9692–9693.

[34] a) Tai, V. W.; Imperiali, B. *J. Org. Chem*. **2001**, *66*, 6217–6228. b) Gibbs, B. S.; Coward, J. K. *Bioorg. Med. Chem.* **1999**, *7*, 441–447.

[35] a) Perugino, G.; Trincone, A.; Rossi, M.; Moracci, M. *Trends Biotechnol.* **2004**, *22*, 31–37. b) Hancock, S. M.; Vaughan, M. D.; Withers, S. G. *Curr. Opin. Chem. Biol*. **2006**, *10*, 509–519. c) Faijes, M.; Planas, A. *Carbohydr. Res.* **2007**, *342*, 1581–1594.

[36] Umekawa, M.; Huang, W.; Li, B.; Fujita, K.; Ashida, H.; Wang, L. X.; Yamamoto, K. *J. Biol. Chem.* **2008**, *283*, 4469–4479.

[37] Pachamuthu, K.; Schmidt R. R. *Chem. Rev.* **2006**, *106*, 160-187.

[38] Cohen, S. B.; Halcomb, R. L. *Org. Lett.* **2001**, *3*, 405-407.

[39] Zhu, Y.; vanderDonk, W. A. *Org. Lett.* **2001**, *3*, 1189-1192.

[40] Galonic, D. P.; Ide, N. D.; van der Donk, W. A.; Gin, D. Y. *J. Am. Chem. Soc*. **2005**, *127*, 7359-7369.

[41] Zhu, X.; Haag, T.; Schmidt, R. R. *Org. Biomol.Chem.* **2004**, *2*, 31-33.

[42] Thayer, D. A.; Yu, H. N.; Galan, M. C.; Wong, C.H. *Angew. Chem.Int. Ed.* **2005**, *44*, 4596-4599.

[43] Groothuys, S.; Kuijpers, B. H. M.; Quaedflieg, P. J. L. M.; Roelen, H. C. P. F.; Wiertz, R. W.; Blaauw, R. H.; van Delft, F. L.; Rutjes, F. P. J. T. *Synthesis* **2006**, 3146-3152.

[44] a) Biswas, K.; Coltart, D. M.; Danishefsky, S. J. *Tetrahedron Lett.* **2002**, *43*, 6107-6110. b) Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J. L.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 9642-9643.

[45] Dondoni, A.; Marra, A. *Chem. Rev.* **2000**, *100*, 4395-4421.

[46] Miller, N.; Williams, G. M.; Brimble, M. A. *Org. Lett.* **2009**, *11*, 2409-2412.

[47] Dondoni, A.; Carrozzi, N.; Marra, A.; *J. Org. Chem.* **2004**, *69*, 5023-5036.

[48] Schwardt, O.; Rabbani, S.; Hartmann, M.; Abgottspon, D.; Wittwer, M.; Kleeb, S.; Zalewski, A.; Smiesko, M.; Cutting, B.; Ernst, B.; *Bioorg. Med. Chem.* **2011**, *19*, 6454-6473.

[49] a) Ichikawa, Y.; Nishiyama, T.; Isobe, M. *Synlett*, **2000**, 1253-1256. b) Ichikawa, Y.; Nishiyama, T.; Isobe, M. *J. Org. Chem.* **2001**, *66*, 4200-4205.

[50] Ichikawa, Y.; Matsukawa, Y.; Isobe, M. *J. Am. Chem. Soc.* **2006**, *128*, 3934-3938.

[51] a) Benn, M.H.; Jones, A.S. *J. Chem. Soc.* **1960**, 3837-3841. b) Prosperi, D.; Ronchi, S.; Panza, L.; Rencurosi, A.; Russo, G. *Synlett* **2004**, 1529-1532. c) Prosperi, D.; Ronchi, S.; Lay, L.; Rencurosi, A.; Russo, G. *Eur. J. Org. Chem.* **2004**, 395-405. d) Ichikawa, Y.; Nishiyama, T.; Isobe, M. *Tetrahedron* **2004**, *60*, 2621-2627. e) García-Moreno, M.I.; Benito, J.M.; Ortiz Mellet, C.; García-Fernández, J.M. *Tetrahedron: Asymmetry* **2000**, *11*, 1331-1341. f) Somsák, L.; Felföldi, N.; Kónya, B.; Hüse, C.; Telepò, K.; Bokor, E.; Czifràk, K. *Carbohydr. Res*. **2008**, *343*, 2083-2093.

[52] Ichikawa, Y.; Ohara, F.; Kotsuki, H.; Nakano, K. *Org. Lett.* **2006**, *8*, 5009-5012.

[53] Bianchi, A.; Ferrario, D.; Bernardi, A. *Carbohydr. Res.* **2006**, *341*, 1438-1446.

Chapter 3

Strategies for the synthesis of α-*N*-linked glycosyl amino acids and plan of the work

3.1 Introduction

As seen so far, natural *N*-linked glycopeptides are almost invariably β-linked. Therefore the stereoselective synthesis of neo-glycoconjugates in the α *N*-linked configuration is a little explored field, but it could be of great interest as a means of designing glycopeptide mimics with altered metabolic stability and novel physico-chemical properties. It is in fact possible that unnatural α-linked isomers could be stable to hydrolytic enzymes and may be used for in vivo applications.

Our laboratory has been actively exploring the synthesis and biological applications of such unnatural glycoconjugates. For example a small group of α-fucosyl amides derived from β amino acids (**Figure 1**) have been tested for their affinity for the carbohydrate recognition domain of DC-SIGN[1] and of the PA-II lectin.[2] DC-SIGN is a dendritic cell receptor with mannose and fucose specificities. It was shown that many pathogens are recognized by DC-SIGN, which participates in some ways to the corresponding infection process, as in HIV infections.[3] Furthermore, since the detailed molecular mechanisms by which this receptor operates are still unknown, effective modulators of DC-SIGN could help to understand the different biological processes in which it can be involved. Compound **A** was the first reported fucose-based glycomimetics to interact with DC-SIGN with an affinity similar to that of the natural DC-SIGN fucose-based ligand, the Lewis-x trisaccharide.[1] Compounds **B-F** were found to bind to the lectin PA-II in the micromolar range.[2] PA-II lectin is a fucose selective lectin from *Pseudomonas aeruginosa*, involved in the formation and stabilization of microbial biofilms.[4] This soluble bacterial lectin binds with an unusually strong micromolar affinity to L-fucose in a tight binding site which requires two Ca^{2+} ions.[5] Hence, the compounds shown in Figure 1 constitute proof of principle that a α-glycosyl amides can perform as effective mimics of natural fucosyl ligands.

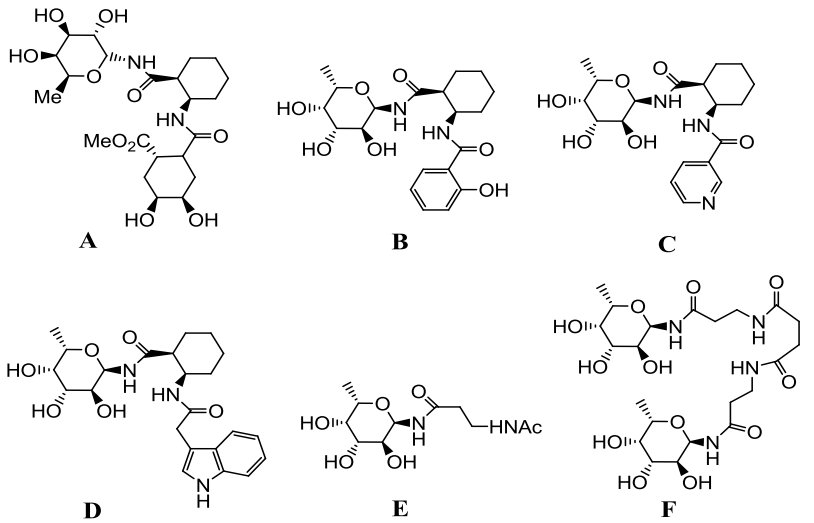


Figure 1. Small library of α-fucosyl amides with affinity for DC-SIGN (**A**) or for PA-II lectin (**B-F**)

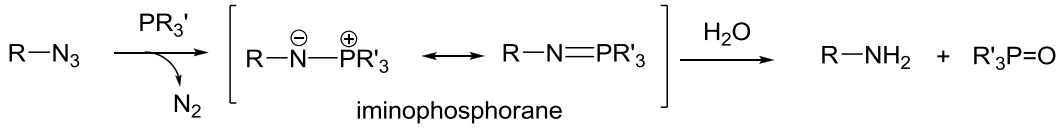
Further studies in our laboratory have been directed to the synthesis of glycosyl amides through the traceless Staudinger ligation using functionalized phosphines. These have allowed to select appropriate reagents and reaction conditions to achieve good stereoselectivity in the ligation of α- and β-glycosyl azides and aminoacids.[6]

Taken these results as a starting point, our new goal becomes, in this thesis, the synthesis of α-*N*-linked glycopeptides as molecules and materials which could behave in an unprecedented fashion. It has already been shown in Section 1.3 that several years ago Imperiali and Woods reported one case whereby the peptide conformation of an *N*-linked glycopeptide was found to depend on the α or β anomeric configuration of the appended glycan.[7] New data, however, have been lagging, mostly because of a lack of methods able to secure a viable synthesis of α-*N*-linked glycopeptides (The α-*N*-linked peptide of Imperiali and Wood was obtained as a by-product in route to the native β-linked isomer). In general, glycopeptides can be prepared using a convergent or a linear approach, as detailed in Chapter 2. However, for α-*N*-linked derivatives, the convergent approach, which involves direct glycosylation of a preformed peptide chain, is precluded by the synthetic difficulties of α-*N*-glycosylation. Several examples of reduction of glycosyl azides by catalytic hydrogenation or phosphine reduction, followed by acylation of the resulting glycosylamines have been reported,[8,9] but, since glycosylamines rapidly and fully equilibrate to the most stable β-anomers, all these approaches, that make use of isolated amine intermediates, afford mostly β-glycosyl amides. An alternative methodology attempts to avoid anomeric equilibration by reducing glycosyl azides in the presence of acylating agents.[10] This approach is successful only in a limited number of cases involving highly reactive acylating agents, such as $(CF_3CO)_2O$. Rather we decided to develop a linear approach which makes use of preformed unnatural α-*N*-glycosyl amino acids that could be obtained by a stereoselective method. Only a few methods have been reported to afford α-glycosyl amides and amino acids, most of which require two steps and have been described for a limited number of substrates.[6, 11, 12, 13]

Two main synthetic methods are available for the synthesis of α-glycosyl amino acids: one, developed primarily by our group,[6,12] is based on the traceless Staudinger ligation of glycosyl azides, using functionalized phosphines. The other, reported by DeShong and co-workers[13] is based on the acylation of a glucopyranosyl oxazoline derived from α or β-glucopyranosyl azides.

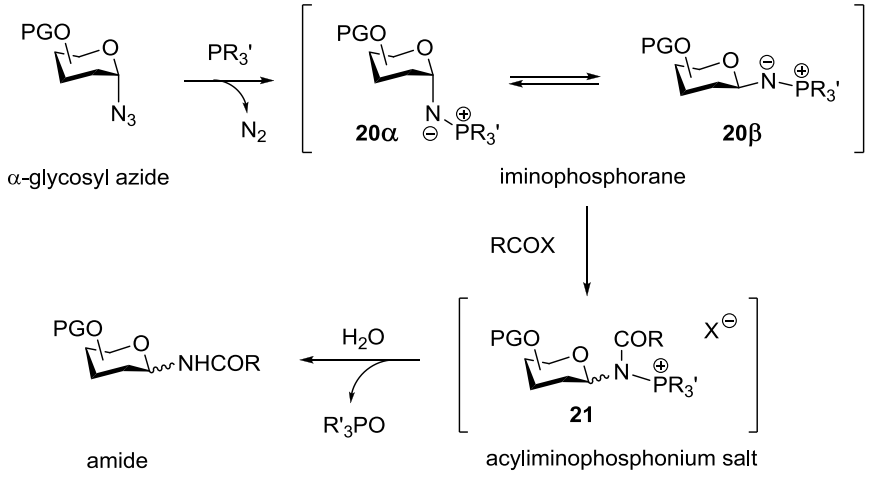
3.2 Traceless Staudinger ligation of glycosyl azides

The Staudinger reaction, originally reported by Staudinger and Meyer in 1919, is the reaction between an azide and a phosphine, which, after loss of nitrogen, results in the formation of an iminophosphorane intermediate, which, through hydrolysis, produces a primary amine and the phosphine oxide (**Scheme 1**).[14]



Scheme 1. Classical Staudinger reaction between a phosphine and an azide.

The Staudinger reduction of α-glycosyl azides affords glycosyl iminophosphoranes **20α** and **20β** (**Scheme 2**), which can be trapped by acylating agents to give configurationally stable acylamino phosphonium salts **21** that, in turn, yield the corresponding amides upon water quenching.

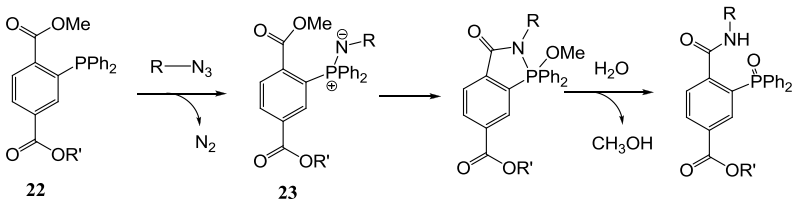


Scheme 2. Mechanism of the Staudinger reduction-acylation of the glycosyl azides

Like glycosylamines, Staudinger's iminophosphoranes are subject to anomeric isomerization, which favours the β-anomers **20β**. Thus, the synthesis of β-glycosyl amides can be easily achieved in this process, while anomerization remains a significant problem during the synthesis of α-glycosyl amides.[9a]

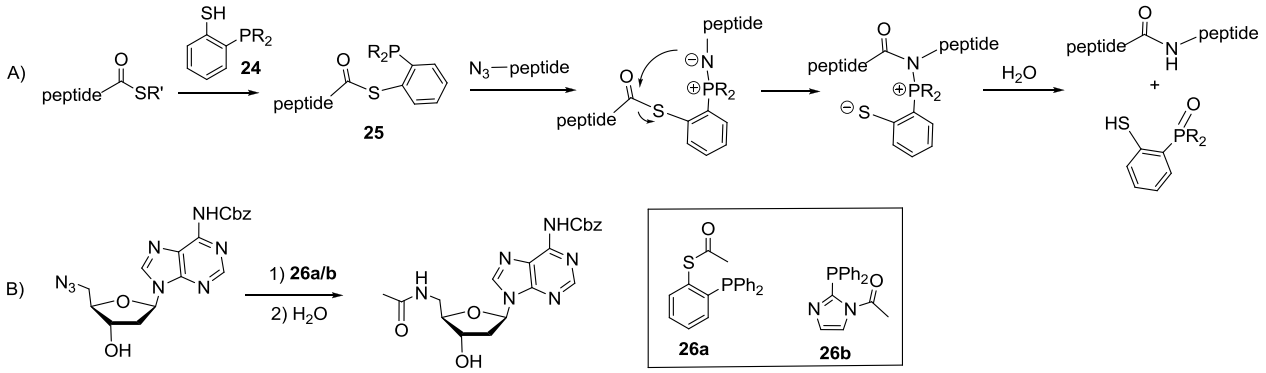
Almost a century after the discovery of the Staudinger reaction, Saxon and Bertozzi exploited this reaction to create an important method for bioconjugation.[15] In their reaction, now called the non-traceless Staudinger ligation, a phosphine reagent that carried an intramolecular electrophilic trap **22** was designed to capture the nucleophilic aza-ylide intermediate **23** by an intramolecular cyclization, so to produce a stable amide bond (**Scheme 3**). This reaction was applied in vivo with

an azide of a sialic acid derivative, incorporated into a glycoprotein and using a phosphine **22** where R' was a fluorescent probe. The result was a selective tagging of glycans with a detectable probe for subsequent in vivo monitoring of glycosylation processes.[16]



Scheme 3. Non-traceless Staudinger ligation developed by Saxon and Bertozzi.

A few years later Kiessling and co-workers[17] as well as the Bertozzi research group[18] simultaneously reported the so-termed traceless Staudinger ligation, which, compared to the nontraceless Staudinger ligation, implied the elimination of phosphine oxide from the final product in the hydrolysis step (**Scheme 4**). Raines/Kiessling applied the Staudinger reaction to peptide synthesis to unite a thioester and an azide (**Scheme 4B**). The ligation begins by transthioesterification with phosphinothiol **24**. Coupling of the resulting phosphinothioester **25** with a peptide azide leads to the formation of a reactive iminophosphorane, which after internal attack from nitrogen and hydrolysis produces the desired amide. The method developed by Bertozzi required the use of acylated phosphine compounds **26a-b**, whose structural prerequisite was the presence of at least two aromatic substituents to prevent excessive oxidation. These phosphines were for instance used for reaction with an azido-nucleoside to afford the corresponding 6-acetamido derivative (**Scheme 4A**).



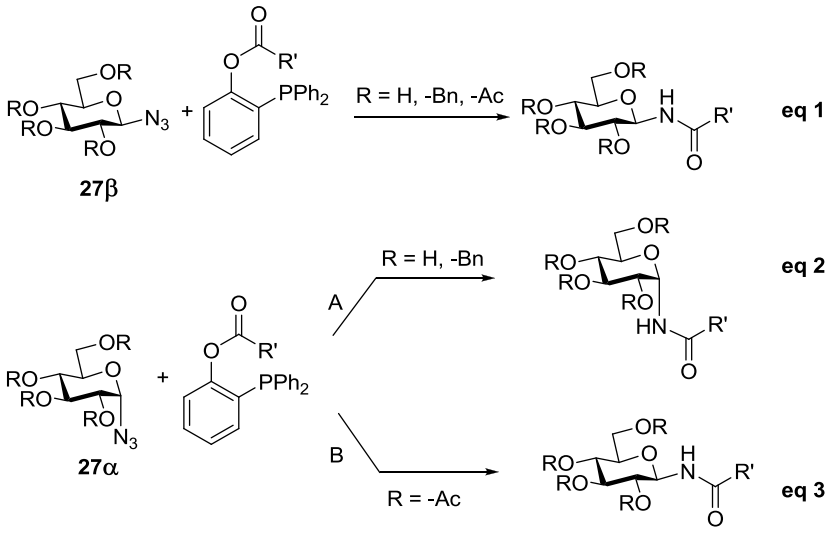
Scheme 4. A) Traceless Staudinger ligation by Raines and co-workers using thiol auxiliaries **24**. B) Traceless Staudinger ligation of azido-nucleoside with Bertozzi's cleavable phosphine derivatives **26a/b**.

These works can be adapted to a method for bioconjugation if one of the reacting functions (phosphine or azide) is present in the biomolecule. The azide is clearly preferred because of its size and stability under physiological conditions and because it can easily be incorporated into glycans, proteins, lipids.

3.3 Previous results by our group on the traceless Staudinger ligation of glycosyl azides

The traceless Staudinger ligation of azides employs a Staudinger-like protocol, such as the one described in **Scheme 2**, but the phosphines used are modified to include an acylating agent. In principle, the reaction allows for reduction of the starting azide and fast intramolecular trapping of the reduction intermediates **20**, resulting in the direct formation of an amide link. In many instances this prevents epimerization and allows retention of configuration at the anomeric carbon.[6,12]

Our laboratory has recently reported a traceless Staudinger ligation of glycosyl azides, whose overall picture is summarized in **Scheme 5**. β-glucosyl azides **27β** can be transformed into the corresponding amides with retention of configuration, irrespective of the presence and nature of the hydroxyl protecting group R (**Scheme 5**, eq 1, R = Bn, Ac, H). On the contrary, the ligation of α-glycosyl azides **27α** depends critically on the nature of the protecting groups: benzyl ethers and free hydroxyl groups allow the reaction to occur with retention of configuration (**Scheme 5**, eq 2), whereas acetates enforce inversion of configuration at the anomeric center and formation of the corresponding β-amide (**Scheme 5**, eq 3).

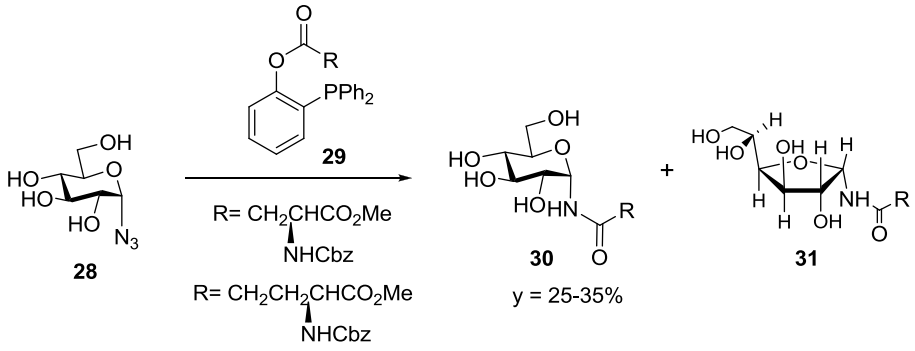


Scheme 5. Traceless Staudinger ligation of glycosyl azides with functionalized phosphines.

The dependence of the ligation stereochemistry on the nature of the sugar protecting group appeared to be related to the electron-withdrawing effect of the acetates,[19, 20] which may reduce the rate of the acylation step and favour anomerization. This effect enforced the use of free hydroxy groups or benzyl ether protecting groups in the synthesis of α-glycosyl azides. The method could be applied to unprotected and *O*-benzyl glycosyl azides in the *fuco*, *gluco*, and *galacto* series to afford the corresponding α-glycosyl amides with good yields and stereoselectivities for a range of acyl chains

with many alkyl and alkenyl groups, both linear and branched, and amino acids with various functional groups.[6,12] The phosphines employed are air stable reagents that can be easily synthesized and purified by flash chromatography, which gives a significant advantage over other ligation reagents. The process described, however, left various synthetic problems unresolved. The reactivity of α-glycosyl azides was uniformly low, and the corresponding amides were obtained generally in modest yields. Moreover, the yields of the ligation appeared to depend critically also on the nature of the acyl group to be transferred and were especially disappointing for the transfer of amino acids to α-azides.

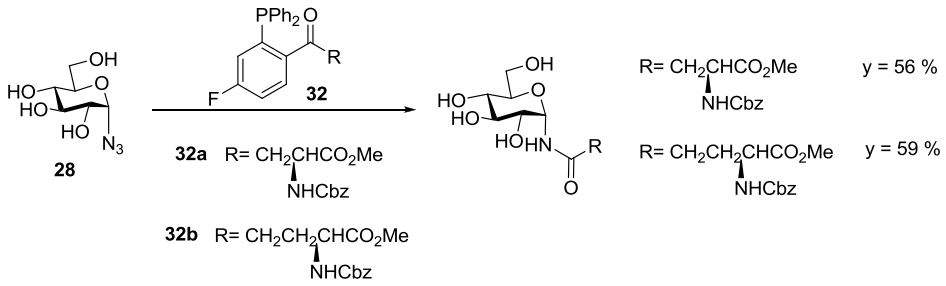
More recently, many efforts have been dedicated to find a protocol for the synthesis of α-glycosyl amino acids. The ligation was applied to unprotected α-glycosylazides **28** using phosphine esters **29** coupled to the carboxylic acid moiety of the amino acid. The amino acids employed were protected as benzyloxycarbammate and methylester. The reaction (**Scheme 6**) often yielded mixtures of the desired pyranosylamide **30** and the isomeric furanosylamide **31**.[6b]



Scheme 6. Poor yields of *N*-glycosyl aspartic and glutamic acid derivatives.

The furanoside **31** must clearly derive from a ring-opening process occurring after the azide reduction step, presumably from the iminophosphorane, which can undergo ring-closure to yield the five-member ring cycle. As noted above, the aspartic and glutamic acid derivatives gave poor yields of the corresponding glucosyl amino acids. Thus, in order to further improve the reactivity and stereoselectivity of the ligation, different acyl phosphines were prepared by our group, trying to vary the basicity of the P atom and the nature of the phenyl ester leaving group. Monofluorophosphines **32** (**Scheme 7**) were found to improve the yields of the amino acids transfer. The fluorinated phosphine was used for the ligation of unprotected β-azides of the *gluco*, *galacto* and *fuco* series with good to excellent yields (60-87%). Lower yields were obtained with β-GlcNAc and for α-glycosyl azides.

However, under these conditions α-glucopyranosyl aspartate was isolated in 56 % yield and 98:2 α/β ratio in the reaction of **28** with **32a** (**Scheme 7**), a significant improvement if compared to 25% yield afforded by the non-fluorinated reagent (**Scheme 6**). Glutamic derivatives could be obtained with a slightly higher yield (**Scheme 7**). [6b]

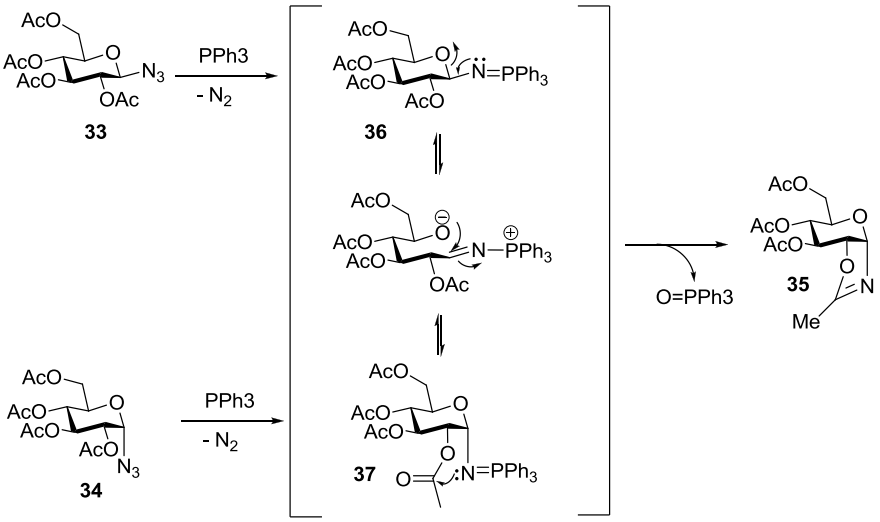


Scheme 7. Synthesis of α-glucosyl aminoacids with fluorinated phosphines.

From these results we decided to investigate the use of amino acids with different protecting groups, so to allow, after their removal, the incorporation of the α-*N*-linked glycosyl aminoacid into a peptide (**Chapter 4**).

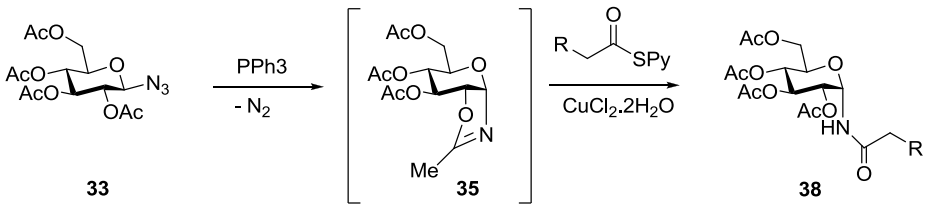
3.4 DeShong methodology

The method reported by DeShong and co-workers[13] starts from α- or β-glucopyranosyl azides **33** and **34** and, by treatment with PPh_3 in refluxing 1,2-dichloroethane, generates oxazoline **35** whose formation from either azide can be explained by the mechanism shown in **Scheme 8**. It involves α/β anomerization of the intermediate iminophosphoranes **36** and **37**,[21, 22] and since oxazoline formation from **36** cannot occur due to strain in the resulting product, epimerization followed by cyclization gives exclusively α-oxazoline **35**.



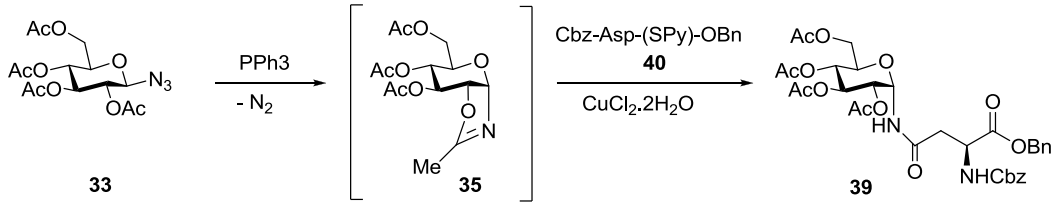
Scheme 8. Mechanism for the formation of isoxazoline.

The second step involves in situ acylation of oxazoline **35** (**Scheme 9**). Acyl chlorides, *N*-hydroxysuccinimidyl or pentafluorophenyl esters were investigated for the acylation of isoxazoline **35.** However, acid chlorides were inappropriate reagents for general glycopeptide synthesis, and the ester derivatives gave a poor α/β ratio. On the contrary, the use of a thiopyridyl ester in combination with copper chloride gave exclusively the α-glucopyranosyl adduct **38** in excellent yield (**Scheme 9**). Metal salts likely coordinate the pyridyl moiety, and presumably increase the electrophilicity of the reagent.



Scheme 9. Coupling reaction of glucopyranosyl isoxazoline **3**.

The method was applied for the synthesis of α-glucopyranosyl asparagine derivative **39**. Acylation with the side chain thiopyridyl ester of *N*-Cbz-protected aspartic benzyl ester **40** in the presence of $CuCl_2 \cdot 2H_2O$ gave exclusively the α-asparagine adduct (**Scheme 10**).[13]



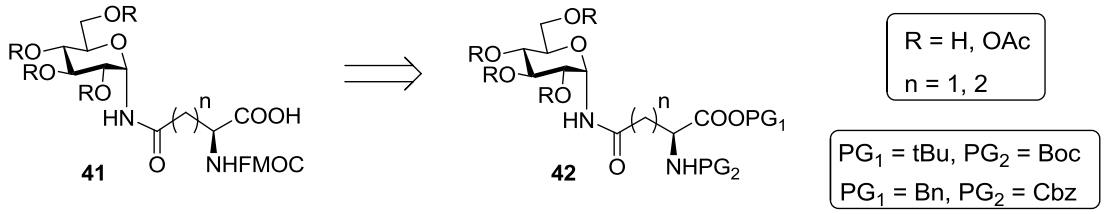
Scheme 10. Synthesis of α-*N*-Aspartyl Glucosylamine **39**.

Taken this methodology as a benchmark for the synthesis of α-*N*-linked glycosyl amino acids, we decided to extend this procedure also for the synthesis of the *galacto* derivative. Then, we devoleped procedures to obtain, after elaboration of the protecting group of the amino acid residue, suitable *gluco*- and *galacto*-building blocks for the linear assembly of glycopeptides (**Chapter 5**).

3.5 Aim of the work and plan of the thesis

The main objective of this thesis is the synthesis of unnatural α-*N*-glycosyl glycopeptides, starting from the individuation and the optimization of a stereoselective synthesis of α-*N*-linked amino acid building blocks. The inclusion of these glycosyl amino acids in synthetic peptides through a solid phase synthesis linear approach, required the synthesis of preformed α-*N*-Fmoc protected glycosyl amino acids **41** (**Scheme 11**). The structures employed are glucose and galactose for the

carbohydrate moiety and aspartic or glutamic acid for the amino acid part. Since *N*-Fmoc protected amino acids were synthesized in poor yields both by Staudinger ligation and by the DeShong approach, we selected combinations of protecting groups for the amino acid (PG_1 and PG_2 in molecules **42**) that could be simultaneously removed and replaced by Fmoc (**Scheme 11**).



Scheme 11. Choice of proper protecting groups

The preliminary studies just summarized (**Section 3.3**) had established that Staudinger ligation works reasonably well with unprotected sugars and better with glutamic than with aspartic derivatives. Therefore, we initially attempted to obtain α-*N*-glucosyl glutamic acid derivatives **42** (R = H, n = 2), through the traceless Staudinger ligation of unprotected glucosyl azide and glutamic acid functionalized phosphines. As illustrated in **Chapter 4**, this method turned out to be not particularly efficient for the synthesis of building blocks **42** (R = H, n = 2). Indeed overall yields remained modest and the method was abandoned in favour of DeShong's approach.

Our efforts in this area are collected in **Chapter 5** which describes: a) The synthesis of α-*N*-glucosyl and galactosyl aspartic acid derivatives **41** (R = OAc, n = 2) by DeShong's method. b) Some initial trials that revealed strong limitations for the synthesis of glutamic acid derivatives under the same conditions.

Couplings condition of α-*N*-glycosyl aspartic acid derivatives were investigated in solution, both from the *C*- and *N*-termini, and are reported in **Chapter 6**. Here we also describe the synthesis in solution of models of a tripeptide and of a pentapeptide that were the subject of computational and NMR conformational studies in a collaboration with the group of Jiménez-Barbero in Madrid and with Fabio Doro, from our lab. The results of these studies are summarized in **Chapter 9**.

The development of solid phase coupling conditions and the solid phase synthesis of α-*N*-linked-galactopeptides, which resemble the structure of antifreeze glycopeptides is described in **Chapter 7**. Finally, a separate chapter, **Chapter 8**, is entirely dedicated to development of purification methods for these compounds**.**

3.6 References

[1] Timpano, G.; Tabarani, G.; Anderluh, M.I.; Invernizzi, D.; Vasile, F.; Potenza, D.; Nieto, P.M.; Rojo, J.; Fieschi, F.; Bernardi, A. *ChemBioChem* **2008**, *9*, 1921-1930.

[2] Andreini, M.; Anderluh, MI.; Audfray, A.; Bernardi, A.; Imberty, A. *Carb. Res.* **2010**, *345*, 1400-1407.

[3] van Kooyk, Y.; Geijtenbeek, T. H. B. *Nat. Rev. Immunol*. **2003**, *3*, 697–709.

[4] Imberty, A.; Wimmerova, M.; Mitchell, E. P.; Gilboa-Garber, N. *Microb. Infect.* **2004**, *6*, 222–229.

[5] Mitchell, E.; Houles, C.; Sudakevitz, D. ; Wimmerova, M.; Gautier, C.; Perez, S. ; Wu, A. M.; Gilboa-Garber, N.; Imberty, A. *Nat. Struct. Biol*. **2002**, *9*, 918–921.

[6] a) Bianchi, A.; Bernardi, A. *J. Org. Chem.* **2006**, *71*, 4565-4577. b) Nisic, F. ; Bernardi, A. *Carb. Res.* **2008**, *343*, 1636-1643. c) Nisic, F. ; Andreini, M. ; Bernardi, A. *Eur. J. Org.Chem*, **2009**, 5744-5751.

[7] Bosques, C. J.; Tschampel, S. M.; Woods, R.; Imperiali, B. *J. Am. Chem. Soc.* **2004**, *126*, 8421-8425

[8] a) Matsuo, I.; Nakahara, Y.; Ito, Y.; Nukada, T.; Nakahara, T.; Ogawa, T. *Bioorg. Med. Chem.* **1995**, *3*, 1455-1463. b) Saha, U. K.; Roy, R. *Tetrahedron Lett.* **1995**, *36*, 3635-3638. c) Sabesan, S. *Tetrahedron Lett.* **1997**, *38*, 3127-3130.

[9] a) Kovàcs, L.; Ósz, E.; Domokos, V.; Holzer, W.; Györgydeàk, Z. *Tetrahedron* **2001**, *57*, 4609-4621. b) Doores, K. J.; Mimura, Y.; Dwek, R. A.; Rudd, P. M.; Elliott, T. E.; Davis. B. G. *Chem. Commun* **2006**, *13*, 1401–1403.

[10] a) Inazu, T.; Kobayashi, K. *Synlett* **1993**, 869-870. b) Bosch, I.; Romea, P.; Urpi, F.; Vilarrasa, J. *Tetrahedron Lett.* **1993**, *34*, 4671-4674. c) Boullanger, P.; Maunier, V.; Lafont, D. *Carbohydr. Res.* **2000**, *324*, 97-106. d) Malkinson, J. P.; Falconer, R. A.; Toth, I. *J. Org. Chem.* **2000**, *65*, 5249-5252. e) Shangguan, N.; Katukojvala, S.; Greenberg, R.; Williams, L. J. *J. Am. Chem. Soc.* **2003**, *125*, 7754-7755.

[11] a) Ratcliffe, A. J.; Fraser-Reid, B. *J. Chem. Soc., Perkin Trans. 1* **1989**, 1805-1810. b) Ratcliffe, A. J.; Konradsson, P.; Fraser-Reid, B. *J. Carbohydr.* 1805-1810. c) Ratcliffe, A. J.; Konradsson, P.; Fraser-Reid, B. *J. Carbohydr. Res.* **1991**, *216*, 323-335.

[12] a) Bianchi A. PhD Thesis, Universita' di Milano, **2004-2005**; b) Bianchi, A.; Bernardi, A. *Tetrahedron Letters* **2004**, *45*, 2231-2234. c) Bianchi, A., Russo, A; Bernardi, A. *Tetrahedron: Asymmetry* **2005**, *16*, 381-386.

[13] Damkaci, F.; DeShong, P. *J. Am. Chem. Soc.* **2003**, *125*, 4408-4409.

[14] Staudinger, H.; Meyer, J. *Hel*v. *Chim. Acta* **1919**, *2*, 635-646.

[15] Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007–2010.

[16] Saxon, E.; Luchansky, H. C.; Hang, C. Y.; Lee, C.R.; Bertozzi, C. R. *J. Am. Chem. Soc* **2002**, *124*, 14893-14902.

[17] Nillson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2000**, *2*, 1939-1941.

[18] Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. *Org. Lett.* **2000**, *2*, 2141-2143.

[19] Mootoo, D. R.; Konradsson, P.; Udodong, U.; Fraser-Reid, B. *J. Am. Chem. Soc.* **1988**, *110*, 5583-5584

[20] Ottoson, H.; Udodong, U.; Wu, Z.; Fraser-Reid, B. *J. Org. Chem.* **1990**, *55*, 6068-6070.

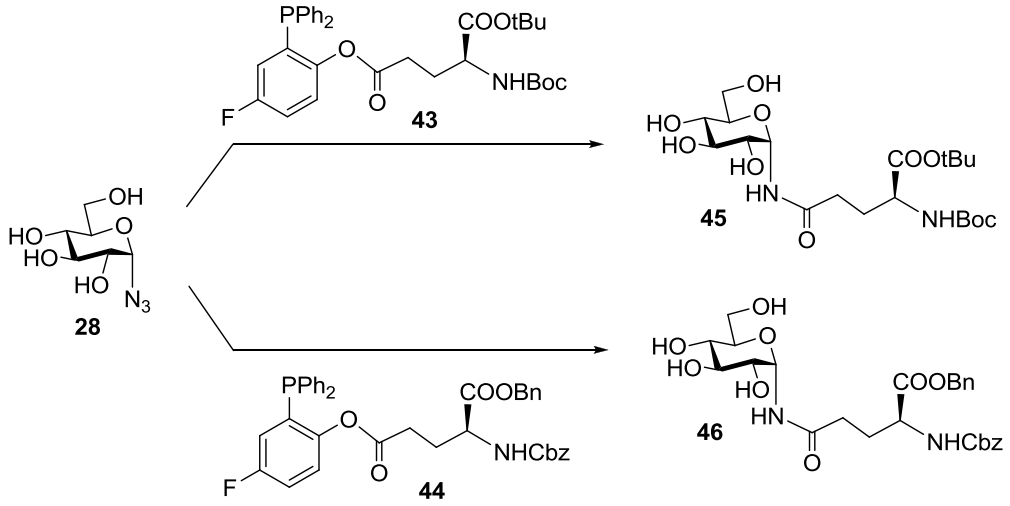
[21] Boullanger, P.; Maunier, V.; Lafont, D. *Carbohydr. Res.* **2000**, *324*, 97-106.

[22] Kovács, L.; Ósz, E.; Domokos, V.; Holzer, W.; Györgydeák, Z. *Tetrahedron* **2001**, *57*, 4609-4621 and references therein.

Chapter 4
Staudinger ligation of unprotected glycosyl azides with glutamic acid derivatives

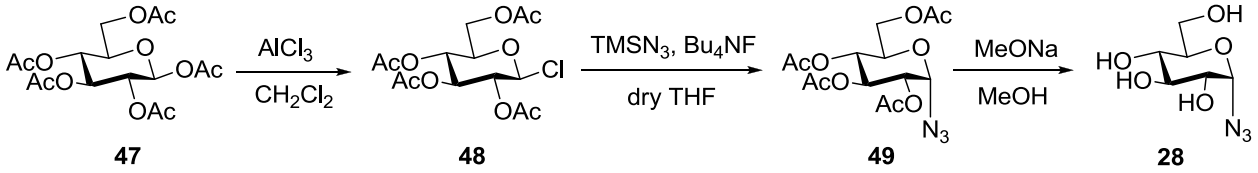
4.1 Synthesis of α-*N*-linked glucosyl amino acids by traceless Staudinger ligation

Previous studies on the use of Staudinger ligation for the synthesis of glycosyl amides have been summarized in **Section 3.3**. Here we describe the traceless Staudinger ligation for the synthesis of α-*N*-glucosyl-(*N*-Boc, *O*-tBu) glutamic acid **45** and α-*N*-glucosyl-(*N*-Cbz, *O*-Bn) glutamic acid **46** (**Scheme 1**), starting from α-glucosyl azide **28** and using the corresponding phosphines **43** and **44**. Fluorinated phosphines were found to be particularly efficient in the transfer of amino acid side chains to iminophosphoranes,[1] but previous work on glutamic acid had been performed using the *N*-Cbz, *O*-methylester derivative, as shown in Chapter 3.



Scheme 1. Synthesis of α-*N*-glucosyl-glutamic acid derivatives **45** and **46**.

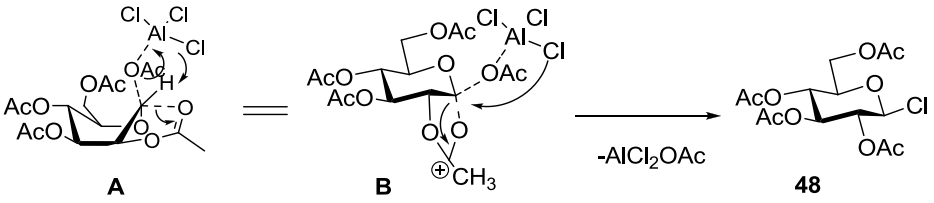
The unprotected α glucosyl azide **28**[2] was obtained by Zemplen hydrolysis of the 2,3,4,6-tetra-*O*-acetyl α-glucosyl azide **49**, which was prepared from the corresponding β-chloride **48**[3] with trimethylsilyl azide and tetrabutylammonium fluoride (**Scheme 2**).[4,5]



Scheme 2. Stereoselective synthesis of α-D-glucopyranosyl azide **28**.

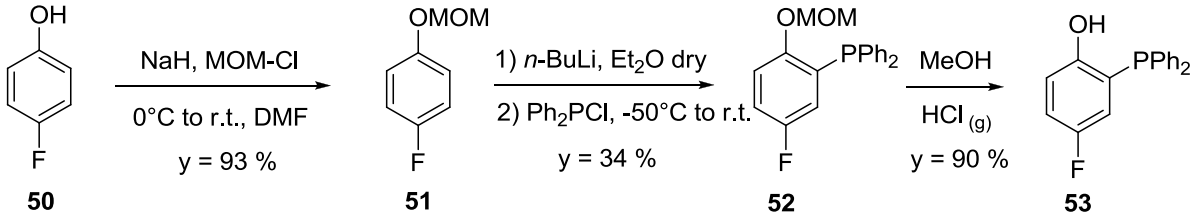
The β-chloride **48** can be obtained from glucose pentaacetate **47** with $AlCl_3$. The proposed mechanism of this reaction (**Scheme 3**) consists in the electrophilic aluminium atom coordination to the ether oxygen of the anomeric acetoxy group, which is transferred to the aluminium, with the assistance of the acetoxy group in position C2 (**A**). Probably the transfer of the chlorine atom from

aluminium to the anomeric position is practically synchronous with fission of the C1-acetoxy bond (**B**).



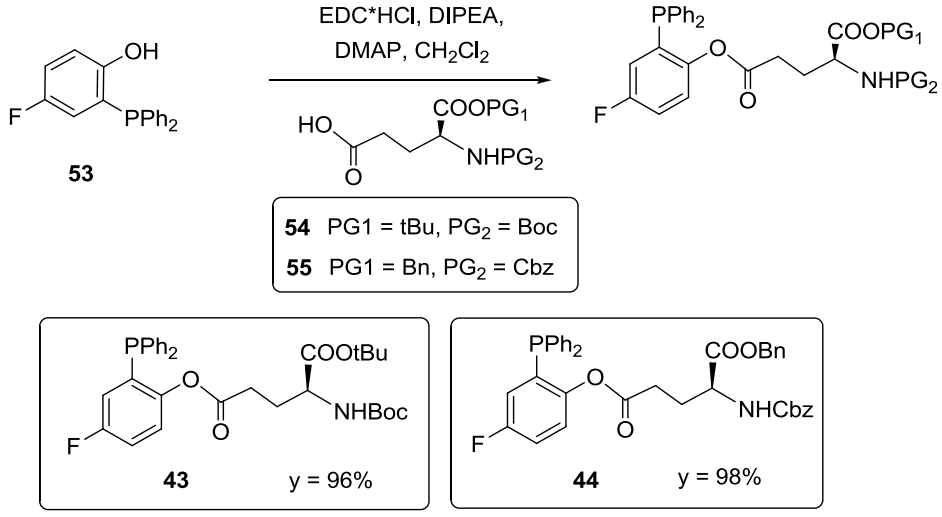
Scheme 3. Mechanism of the formation of the β-chloride **48** [13]

Mono fluorophosphine **53** had been prepared in three steps[1] starting from p-fluoro-phenol **50** and using a procedure introduced by Rauchfuss with some modifications[6] (**Scheme 4**). Initially, the hydroxyl group was protected as the methoxymethyl ether:[7] the MOM-ether protects and, at the same time, activates the phenol ring. Then, the diphenylphosphino group was attached by ortho lithiation of ether **51**, performed according to Fink,[8] using diethyl ether at low temperature -50°C, probably to avoid *ortho* metalation respect to the flourus atom.[9] Finally, the hydroxyl group of **52** was deprotected with HCl in MeOH**.**



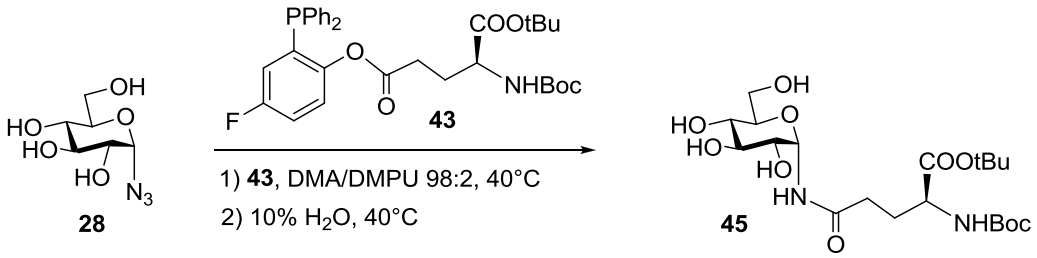
Scheme 4. Synthesis of the fluorinated-*o*-diphenylphosphinophenol **53**

Phosphinophenol **53** was further acylated with protected glutamic acids **54** and **55** activated by *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC·HCl) as condensing agent, to afford functionalized phosphine **43** and **44** respectively (**Scheme 5**).[1]



Scheme 5: Synthesis of phosphines **43** and **44**, functionalized with protected glutamic acid.

During the synthesis of **43** a species with ^{31}P resonance at δ = +26.4 ppm was isolated. This compound has a ^{31}P signal very similar to the corresponding phosphine oxide (^{31}P resonance at δ = +26.7 ppm), but in fact corresponds to a protonated phosphonium salt. Indeed, the signal disappeared when an AcOEt solution of the product was washed with a saturated sodium carbonate solution, while a new ^{31}P signal at δ = -14.1 ppm appeared, which can be assigned to the phosphine structure **43**. Initially, the Staudinger ligation between **28** and **43** was investigated (**Scheme 6**).



Scheme 6: Staudinger ligation for α-*N*-glucosyl *O*-tBu, *N*-Boc glutamic acid **45**.

The reactions were performed in DMA/DMPU (98:2) at 40 °C for 18 h. After completion of the reaction, the crude mixtures were stirred with water for an additional 2 h, then diluted with water and extracted with Et_2O to eliminate the phosphine by-products. The aqueous layer was analyzed by ^1H NMR (typically in D_2O) and the product ratios established, based on the integration of the anomeric protons (**Figure 1**). Typical reaction crudes contained α- and β-pyranosyl amides (**45** and **56**), α-furanosyl amide **57,** free glucose (equilibrium anomeric mixtures), glucosyl amine **58** and unreacted glucosyl azide **28**.

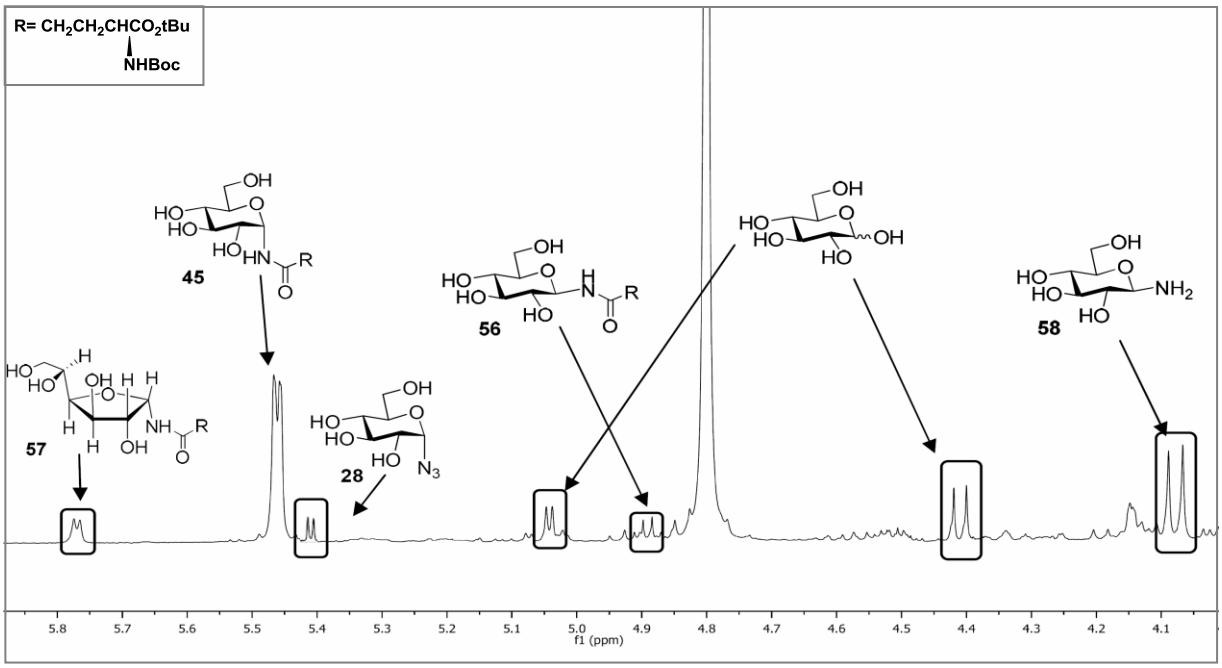
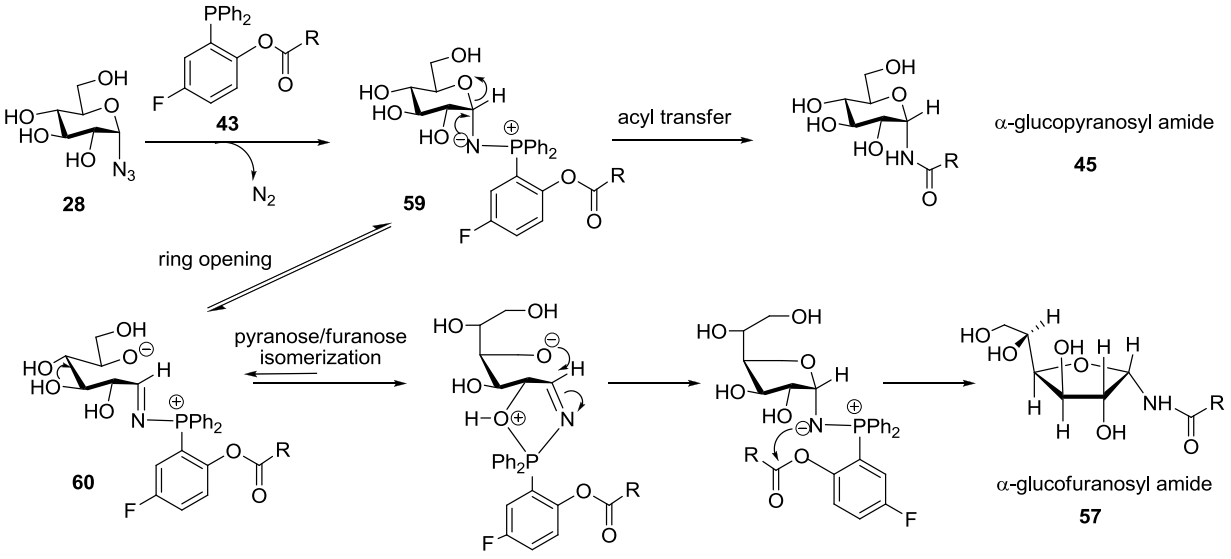


Figure 1: ^1H NMR in D_2O of the aqueous phase derived from extraction of the Staudinger ligation for the synthesis of α-*N*-glucosyl *O*-tBu, *N*-Boc glutamic acid **45**.

The mechanism of the Staudinger ligation, reported in **Scheme 7**, accounts for the formation of the furanosyl compound **57** that was formed as a unique α-furanose isomer. Strikingly, despite the pyranose-furanose isomerization must involve a ring opening step, formation of the β-furanosyl amide, a more obvious candidate product of an equilibration process, is not observed. After the reduction of the azide **28**, immediate intramolecular trapping of the Staudinger aza-ylide intermediate **59** (**Scheme 7**) results in the direct formation of α-glucopyranosyl compound **45** with retention of configuration at the anomeric carbon. The furanosyl compound **57** must derive from a ring-opening process occurring presumably from the iminophosphorane **59** (**Scheme 7**) to afford the phosphinimine **60**, which can undergo ring-closure to yield **57**. This is probably not an equilibrium reaction, because the resulting phosphinimine **60** is blocked in the oxazaphospholane which strongly favours the coordination of the phosphorous atom by the hydroxy group in position 2 of the sugar and the formation of 5-5-fused bicyclic systems. This mechanism explained the unique α-furanose isomer formation. Hydrolysis of the same intermediate **60** accounts for the formation of α and β-D-glucose as by-products in the reaction mixture. If the reaction is particularly slow, unreacted azide **28** and its reduction derivative, the amine **58**, could also be observed as by-product.



Scheme 7. Mechanism for the formation of the **45** and of **57**.

A series of experiments were performed varying the amounts of phosphine **43**, reaction concentration, reaction time, percentage of DMPU and monitoring the formation of the desired product **45** and of the other by-products by means of ^{1}H NMR signal integrations (**Table 1**). The formation of α-furanosyl amide **57** increased with reaction time (entries 5-6). In precedent work[1] it had been noted that formation of α-furanosyl amides was favoured by a temperature increase, and a temperature of 40°C was found to be a good compromise for achieving the acyl-transfer and decreasing ring furanose formation. The nature of the acyl chains transferred obviously plays a

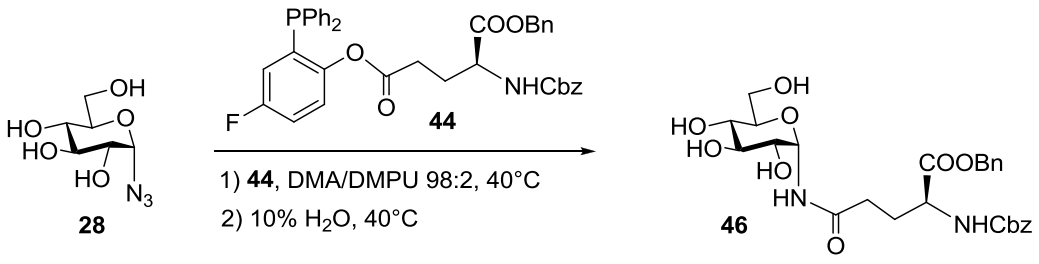
fundamental role for in this process and in this case phosphine **43** turned out to be a poorly reactive substrate, probably due to its steric hindrance. The best reaction conditions are represented by entries 4-5. However, **45** is obtained with a maximum yield of 32%.

Table 1. Reaction conditions of the Staudinger ligation for α-*N*-glucosyl *O*-tBu, *N*-Boc glutamic acid **45**.[a]

Entry	Eq. Phosphine **43**	Conc (M)	Time (h)	DMPU %	α-pyranosyl amide **45** (%)[b]	^{1}H NMR ratio [c]				
						α-furanosyl amide **57**	β-pyranosyl amide **56**	glucose	azide **28**	amine **58**
1	1.5	0.1	18	2	30	0.08	0.07	0.40	0.30	0.10
2	1.5	0.1	22	2	25	0.06	0.07	0.20	0.18	0.40
3	1.5	0.25	18	2	30	0.10	0.20	0.35	-	0.90
4	2	0.1	18	2	32	0.08	0.10	0.35	-	0.30
5	1.5	0.25	48	6	32	0.25	-	0.30	-	0.48
6	1.5	0.25	64	2	25	0.35	-	0.48	-	0.60

[a] All reactions were conducted at 40 °C in DMF. [b] Isolated yield. [c] Relative to the of α-pyranosyl amide **45** anomeric proton, taken equal to 1.

Removal of protecting groups of the amino acid residue of **45** also turned out to be problematic. In fact, using trifluoroacetic acid in CH_2Cl_2 ring opening and anomerization of the glucose ring was observed. A complex mixture of α and β pyranosyl and furanosyl compounds was visible in the ^{1}H NMR spectrum. The Staudinger ligation between **28** and **44** (**Scheme 8**) didn't work better in terms of yields of the α-pyranosyl compound **46**. Various reaction conditions were also investigated (**Table 2**). In this case the ^{1}H NMR signal of α-glucose and of the β-pyranosyl compound are overlapped by the methylene signals of the benzylic protecting groups and weren't reported in the table. The amount of α-furanosyl compound seems to increase with increasing reaction time. The formation of hydrolysis products (β-glucose and amine) is major also in this reaction. Hence, also phosphine **44** is a poorly reactive substrate: ring opening and hydrolysis of the intermediate **60** (**Scheme 7**) are preferred over the acyl transfer. The best result (entry 3) afforded **46** in only 30% yield.



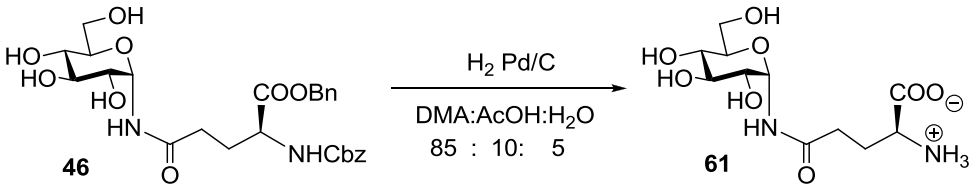
Scheme 8: Staudinger ligation for the synthesis of α-*N*-glucosyl *O*-Bn, *N*-Cbz glutamic acid **46**.

Table 2. Reaction conditions of the Staudinger ligation for α-*N*-glucosyl *O*-Bn, *N*-Cbz glutamic acid **46**.[a]

Entry	Eq. Phosphine **44**	Conc. (M)	Time (h)	α-pyranosyl amide **46** (%)[a]	^{1}H NMR ratio [b]			
					α-furanosyl amide	β-glucose	azide **28**	amine **58**
1	1.5	0.25	64	16	0.90	0.43	-	1.00
2	1.5	0.25	18	25	0.12	0.40	-	0.95
3	1.5	0.5	48	30	0.15	0.30	-	0.50

[a] All reactions were conducted at 40 °C in DMF. [b] Isolated yield. [c] Relative to the of α-pyranosyl amide **46** anomeric proton, taken equal to 1.

Removal of Benzyl ester and Cbz protecting groups of **46** was obtained by hydrogenation on Pd/C. The outcome of the hydrogenation was found to be dependent on the solvents employed. When hydrogenation (Pd/C) was conducted in 80:20:10 MeOH:H_2O:AcOH, a mixture of products was formed, mostly due to ring opening and anomeric equilibration (α and β glucopyranosyl-, α-glucofuranosyl- compounds). This data confirmed the instability of unprotected derivatives towards acidic conditions. On the contrary, hydrogenation on Pd/C in a mixture of 85:10:5 DMA:AcOH:H_2O (**Scheme 9**) afforded unprotected glucosylamino acid **61** in quantitative yield.



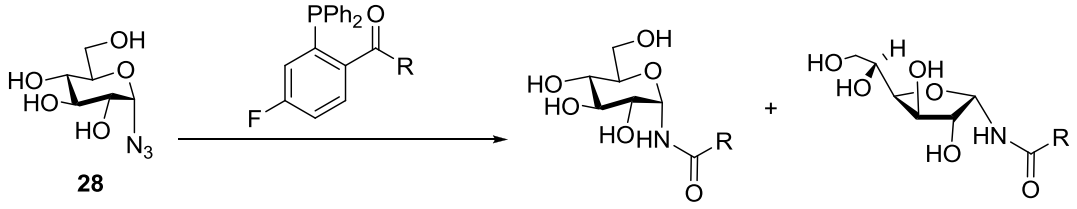
Scheme 9: Hydrogenation of α-*N*-glucosyl *O*-Bn, *N*-Cbz glutamic acid **46**.

Given the poor yields obtained, transfer of aspartic derivatives was not further pursued by this route, but rather we turned our attention to the DeShong protocol.

4.2 Conclusions

The results obtained with the traceless Staudinger ligation of unprotected glucosyl azide **28** and phosphines, functionalized respectively with (*N*-Boc, *O*-tBu) glutamic acid **43** and (*N*-Cbz, *O*-Bn) glutamic acid **44**, weren't promising. The yields of α-*N*-glucosyl amino acids in the pyranosyl form were in fact around 30% in both cases. Fluorophenylphosphines functionalized with protected amino acids have been described in a precedent work[1] for the synthesis of α-*N*-glucosyl amino acids in reasonable yields. In particular phosphine functionalized with (*N*-Cbz, *O*-Me) aspartic acid **32a** and (*N*-Cbz, *O*-Me) glutamic acid **32b** gave respectively 56% and 59% yields of α-*N*-glucosyl

amino acids in the pyranosyl form (**Table 3**, entries 1-2). Comparison of our results (**Table 3**, entries 3-4) with those obtained with phosphines **32a** and **32b** strongly suggests that the lower yields afforded by phosphines **43** and **44** are due to a substantial formation of products derived by ring opening of iminophosphorane **59** (α-*N*-furanosyl amino acid) (**Scheme 7**) and hydrolysis of phosphinimine **60** (glucose and amine). These results could be explained by a scarce reactivity of phosphines **43** and **44**, which are not efficient in the acyl transfer process, probably due to their steric hindrance. This hypothesis is also confirmed by the fact that increasing reaction times led to increased formation of glucose, amine and α-*N*-furanosyl amino acid.



Entry	Phosphine	R	Pyr Isolated yield %	Pyr / fur
1	**32a**	CH_2CHCO_2Me NHCbz	56	85:15
2	**32b**	$CH_2CH_2CHCO_2Me$ NHCbz	59	87:13
3	**43**	$CH_2CH_2CHCO_2tBu$ NHBoc	32	90:10
4	**44**	$CH_2CH_2CHCO_2Bn$ NHCbz	30	85:15

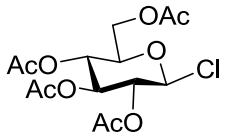
Table 3. Comparison between acyl transfer efficiency of different phosphines. Results with **32a** and **32b** were reported in a previous work.[1]

Since α-*N*-glucosyl glutamic derivatives **45** and **46** were meant to be transformed into Fmoc-protected building blocks and incorporated into *N*-glycopeptide sequence, the method was not considered useful to obtain starting materials in reasonable amounts and we turned our attention to the DeShong approach. The results are described in the following chapter.

4.3 Experimental Section

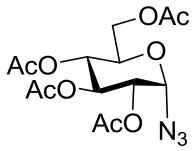
Solvents were dried by standard procedures: dichloromethane, and methanol were dried over calcium hydride; hexane and tetrahydrofuran were dried over sodium; *N*,*N*-dimethylacetamide (DMA), 1,3-dimethyltetrahydro-2(1*H*)pyrimidinone (DMPU), chloroform and pyridine were dried over activated molecular sieves. Reactions requiring anhydrous conditions were performed under nitrogen. ^{1}H and ^{13}C-NMR spectra were recorded at 400 MHz on a Bruker AVANCE-400 instrument. Chemical shifts δ for ^{1}H and ^{13}C are expressed in ppm relative to internal Me_4Si as standard. Signals were abbreviated as s, singlet; bs broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet. Mass spectra were obtained with a Bruker ion-trap Esquire 3000 apparatus (ESI ionization). Optical rotations $[\alpha]_D$ were measured in a cell of 1 dm pathlength and 1 ml capacity with a Perkin-Elmer 241 polarimeter. Thin layer chromatography (TLC) was carried out with precoated Merck F_{254} silica gel plates. Flash chromatography (FC) was carried out with Macherey-Nagel silica gel 60 (230-400 mesh). The Staudinger ligations were carried out on a 15-120 mg scale.

Synthesis of 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl chloride (48) [3]



Aluminium trichloride (512 mg, 3.84 mmol, 0.5 eq) was added, at room temperature and under nitrogen, to a solution of 1,2,3,4,6-penta-*O*-acetyl-β-D-glucopyranose **47** (3 g, 7.68 mmol, 1 eq) in dry CH_2Cl_2 (15.4 mL, 0.5 M). Aluminium trichloride gradually disappeared and was replaced by a fine white precipitate. After 2 h, the mixture was filtered into a large volume (180 mL) of dry hexane. The resulting white precipitate was filtered on celite and washed with dry CH_2Cl_2 and 60:40 hexane/AcOEt. The solvent was evaporated under reduced pressure and the crude product was used without further purification.

Synthesis of 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl azide (49) [4,5]

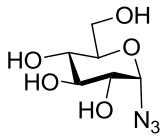


At room temperature and under nitrogen, trimethylsilyl azide (1.41 mL, 10.193 mmol, 1.4 eq) and tetrabutylammonium fluoride 1 M in THF (10.193 mL, 10.193 mmol, 1.4 eq) were added to a

solution of 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl chloride **48** (2.67 g, 7.281 mmol, 1 eq) in dry THF (90 mL, 0.08 M). The solution was heated to 65 °C and stirred for 24 h. The reaction was monitored by TLC (50:50 hexane/AcOEt). The solvent was evaporated under reduced pressure and the crude was purified by flash chromatography using 65:35 hexane/AcOEt as the eluent.

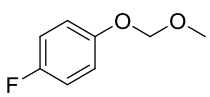
Yield = 67 % (over 2 steps), α/β ratio 82:18. ^{1}H-NMR (400 MHz, $CDCl_3$, 25°C): δ = 5.62 (d, $J_{1,2}$ = 4.3 Hz, 1H, H-1), 5.40 (t, $J_{2,3} = J_{3,4}$ = 9.8 Hz, 1H, H-3), 5.06 (t, $J_{3,4} = J_{4,5}$ = 9.8 Hz, 1H, H-4), 4.96 (dd, $J_{1,2}$ = 4.3 Hz, $J_{2,3}$ = 9.8 Hz, 1H, H-2), 4.29 (dd, $J_{5,6}$ = 4.7 Hz, $J_{6,6'}$ = 12.4 Hz, 1H, H-6), 4.19-4.12 (m, 2H, H-5 and H-6'), 2.11, 2.10, 2.05, 2.03 (4s, 12H, 4xOAc). ^{13}C-NMR (100 MHz, $CDCl_3$, 25°C): δ = 170.5, 169.9, 169.4, 86.2 (C-1), 70.1 (C-5), 69.6 (C-4), 69.5 (C-2), 67.9 (C-3), 61.5 (C-6), 20.6, 20.6, 20.5, 20.5 (4xOAc).

α-D-glucopyranosyl azide (28) [2]



A 1M solution of MeONa in dry methanol (2 mL, 1.967 mmol, 0.5 eq) was added, at room temperature and under nitrogen, to a solution of peracetylated glucosyl azide **49** (1.468 g, 3.933 mmol, 1eq) in dry MeOH (39 mL). The mixture was stirred at room temperature. After 1h TLC monitoring (eluents: hexane/AcOEt 50:50 and $CHCl_3$:MeOH 80:20) showed total consumption of the starting material, the acid resin Amberlyst IRA 120 H^+ was added. The mixture was stirred for 30 minutes (pH = 3). The resin was filtered and washed with MeOH, the solvent was removed under reduced pressure. The product, isolated in quantitative yield, was used without further purification. ^{1}H-NMR (400 MHz, D_2O, 25°C): δ = 5.42 (d, $J_{1,2}$ = 4 Hz, 1H, H-1), 3.81 (dd, $J_{6,6'}$ = 12 Hz, $J_{5,6}$ = 5.5 Hz, 1H, H-6), 3.72 (m, 1H, H-6'), 3.68 (t, J = 5.2 Hz, $J_{3,4}$ = 12 Hz, $J_{4,5}$ = 9.6 Hz, 1H, H-4), 3.58 (dd, J = 4 Hz, $J_{4,5}$ = 9.6 Hz, $J_{5,6}$ = 5.5 Hz, 1H, H-5), 3.51 (t, $J_{1,2}$ = 4 Hz, $J_{2,3}$ = 18.4 Hz, 1H, H-2), 3.33 (t, $J_{2,3}$ = 18.4 Hz, $J_{3,4}$ = 5.2 Hz, 1H, H-3). ^{13}C-NMR (100 MHz, D_2O, 25°C): δ = 89.2 (C-1), 73.8 (C-4), 72.7 (C-5), 70.7 (C-2), 69.2 (C-3), 60.5 (C-6).

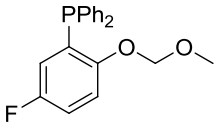
4-fluorophenyl methoxymethyl ether (51) [1]



At 0 °C and with stirring, 4-fluoro phenol **50** (1 eq) was added to a slurry of NaH (1.5 eq) in dimethylformamide (2 M). After 30 minutes, chloromethyl methyl ether (1.5 eq) was added, then the mixture was allowed to warm up to room temperature. The mixture was stirred and the reaction

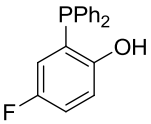
was monitored by TLC (80:20 hexane/AcOEt). After 30 minutes, water was added. The product was extracted with hexane three times. The organic layer was dried over Na_2SO_4 and concentrated. Yield = 93 %.^{1}H-NMR (400 MHz, $CDCl_3$, 25°C): δ = 7.04-6.98 (m, 4H, Ph), 5.16 (s, 2H, CH_2O), 3.51 (s, 3H, OCH_3).

2-(Diphenylphosphanyl)-4-fuorophenyl methoxymethyl ether (52) [1]



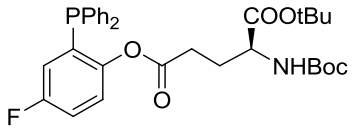
A solution of *n*-BuLi in hexane (1.6 M, 30 mL, 1.1 eq) was added dropwise and under stirring to a solution of 4-fluoro-*O*-methoxymethyl-phenol **51** (7g, 1 eq) in dry Et_2O at -50°C. The mixture was allowed to warm to room temperature. After 3 h the solution was cooled again to -50°C, a solution of Ph_2PCl (10 g, 1.05 eq) in dry Et_2O was added dropwise, and the suspension was stirred overnight at room temperature. The reaction mixture was filtered washing with Et_2O and the solvent evaporated at reduced pressure. Dichloromethane (10 mL) and methanol (100mL) were added. At 0°C white crystals separated that were filtered and dried *in vacuo*. Yield = 34 %. ^{1}H-NMR (400 MHz, C_6D_6, 25°C): δ = 7.40-7.32 (m, 4H, Ph), 7.05-7.00 (m,6H, Ph), 6.91-6.84 (m, 1H, H-6), 6.78-6.70 (m, 2H, H-2, H-3), 4.65 (s, 2H, OCH_2) 2.94 (s, 3H, OCH_3) ppm. ^{13}C-NMR (100 MHz, C_6D_6, 25°C): δ = 158.5, 155.4, 136.8, 134.4, 130.3, 129.1, 128.8, 120.4, 116.5, 115.3, 94.7, 55.7. ^{31}P-NMR (161 MHz, C_6D_6, 25°C): δ = -14.6 ppm

2-(Diphenylphosphanyl)-4-fuorophenol (53) [1]

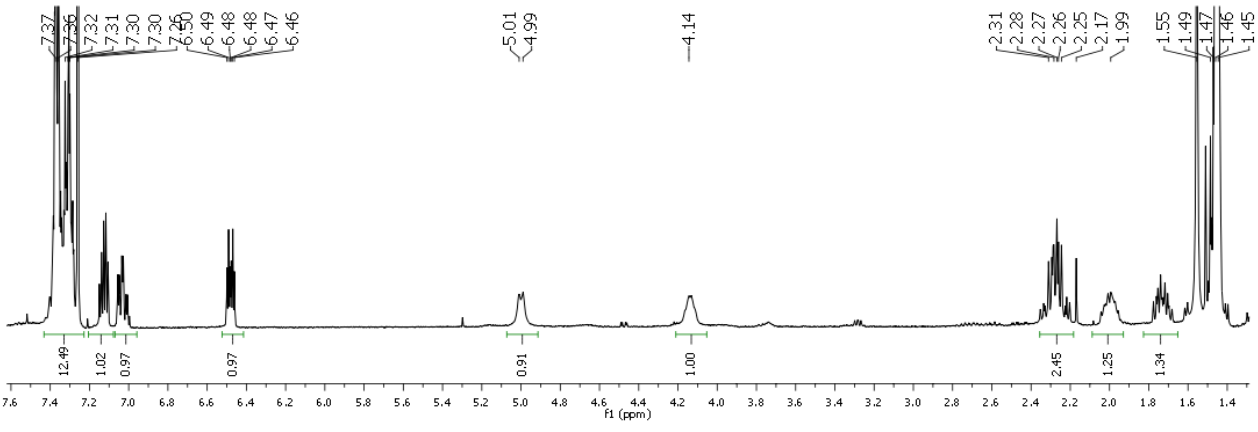


Dry methanol was saturated with $HCl_{(g)}$. Compound **52** was added at room temperature and under nitrogen. The resulting solution was stirred for 1 h then the solvent was evaporated giving a brown oil. The reaction mixture was diluted with AcOEt and extracted with sat. $NaHCO_3$ and water. The organic layer was dried over Na_2SO_4 and concentrated. Compound **53** was purified by flash chromatography (hexane/AcOEt 80:20). Yield = 90 %. ^{1}H-NMR (400 MHz, C_6D_6, 25°C): δ = 7.30-7.21 (m, 4H,Ph), 7.03-6.93 (m, 6H, Ph), 6.84 (m, 1H, H-3), 6.66 (m, 1H, H-3), 6.51 (m, 1H, H-6) 6.10 (bs, 1H, OH) ppm. ^{13}C-NMR (100 MHz, C_6D_6, 25°C): δ = 157.6, 155.7, 135.4, 133.9, 129.3, 129.0, 124.3, 120.1, 118.1, 116.9. ^{31}P-NMR (161 MHz, C_6D_6, 25°C): δ = - 24.8 ppm.

1-*tert*Butyl-5-[2-(Diphenylphosphanyl)-4-fluorophenyl] *N*-(*tert*Butoxycarbonyl)-L-glutamate (43)

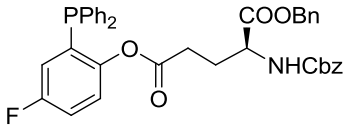


A solution of *o*-diphenylphosphinophenol **53** (250 mg, 0.844 mmol, 1 eq), commercially available *N-tert*butoxy-L-glutamic acid 1-*tert*Butylester **54** (307 mg, 1.012 mmol, 1.2 eq) and *N,N*-dimethylaminopyridine (10 mg, 0.084 mmol, 0.1 eq) in dry CH_2Cl_2 (4 mL) were added, at room temperature and under nitrogen, to a suspension of *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (226 mg, 1.181 mmol, 1.4 eq) and dry *N,N*-diisopropylethylamine (205 μL, 1.181 mmol, 1.4 eq) in dry CH_2Cl_2 (4 mL). The mixture was stirred at RT for 2 h, monitoring by TLC (80:20 hexane/AcOEt). The reaction mixture was diluted with CH_2Cl_2 and extracted with 4 % aqueous HCl, followed by $NaHCO_3$ sat. and water. The organic layer was dried over Na_2SO_4 and concentrated. The crude product was purified by flash chromatography (85:15 hexane/AcOEt) to afford **43** in 96 % yield. If the signal of protonated compound is present [^{31}P = + 26.4 ppm], the product is dissolved in AcOEt and washed with Na_2CO_3 sat. The organic layer was dried over Na_2SO_4 and evaporated. ^{1}H-NMR (400 MHz, $CDCl_3$, 25°C): δ = 7.40-7.27 (m, 15 H, Ph), 7.15 (m, 1H, H-2), 7.05 (m, 1H, H-3), 6.47 (m, 1H, H-1), 5.01 (d, $J_{NH\text{-}CH}$ = 7.9 Hz, 1H, NH), 4.14 (m, 1H, α-H-Glu), 2.28 (m, 2 H, γ-CH_2-Glu), 2.00 (m, 1H, β-CH_2-Glu), 1.75 (m, 1H, β-CH_2-Glu), 1.48 (s, 9H, OCH_3), 1.46 (s, 9H, OCH_3) ppm. ^{31}P-NMR (161 MHz, $CDCl_3$, 25°C): δ = - 14.1 ppm [oxide ^{31}P = + 26.7 ppm], [protonated ^{31}P = + 26.4 ppm].

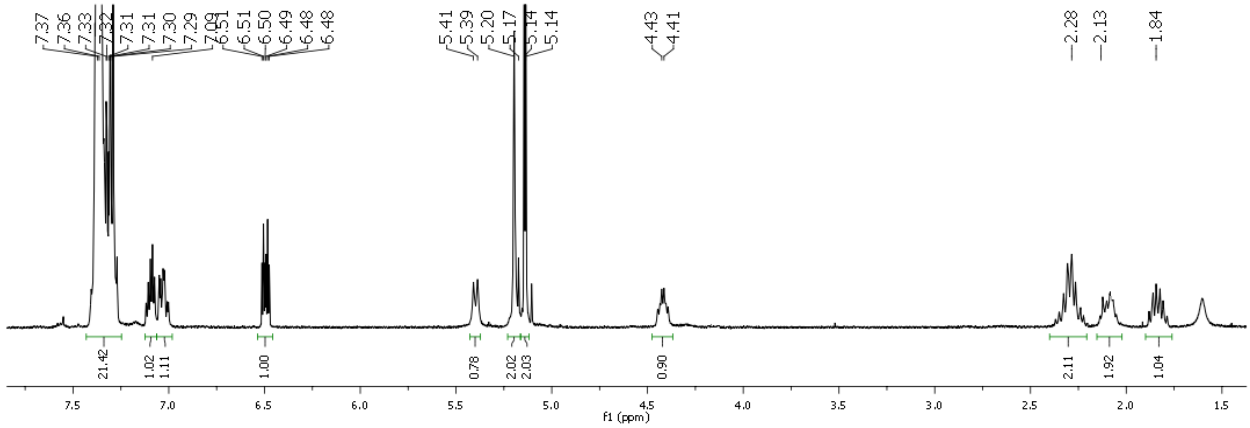


^{1}H NMR spectrum of compound **43** ($CDCl_3$, 400 MHz).

1-Benzyl-5-[2-(Diphenylphosphanyl)-4-fluorophenyl] *N*-(Benzyloxycarbonyl)-L-glutamate (44)

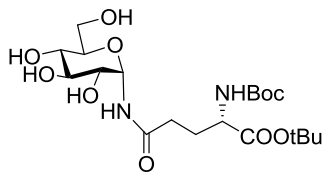


A solution of *o*-diphenylphosphinophenol **53** (500 mg, 1.687 mmol, 1 eq), commercially available *N*-carbobenzyloxy-L-glutamic acid 1-benzyl ester **55** (752 mg, 7.025 mmol, 1.2 eq) and *N,N*-dimethylaminopyridine (21 mg, 0.169 mmol, 0.1 eq) in dry CH_2Cl_2 (4 mL) were added, at room temperature and under nitrogen, to a suspension of *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (453 mg, 2.362 mmol, 1.4 eq) and dry *N,N*-diisopropylethylamine (411 μL, 2.362 mmol, 1.4 eq) in dry CH_2Cl_2. The mixture was stirred at RT for 2 h, monitoring by TLC (80:20 hexane/AcOEt). The reaction mixture was diluted with CH_2Cl_2 and extracted with 4 % aqueous HCl, followed by $NaHCO_3$ sat. and water. The organic layer was dried over Na_2SO_4 and concentrated. The crude product was purified by flash chromatography (85:15 hexane/AcOEt) to afford **44** in 98 % yield. ^{1}H-NMR (400 MHz, $CDCl_3$, 25°C): δ = 7.40-7.25 (m, 15 H, Ph), 7.09 (m, 1H, H-2), 7.02 (m, 1H, H-3), 6.49 (m, 1H, H-1), 5.39 (d, J_{NH-CH} = 7.9 Hz, 1H, NH), 5.18 (d, J = 3.8 Hz, 2H, CH_2-O), 5.13 (d, J = 3.8 Hz, 2H, CH_2-O), 4.41 (m, 1H, α-H-Glu), 2.28 (m, 2 H, γ-CH_2-Glu), 2.08 (m, 1H, β-CH_2-Glu), 1.83 (m, 1H, β-CH_2-Glu) ppm.



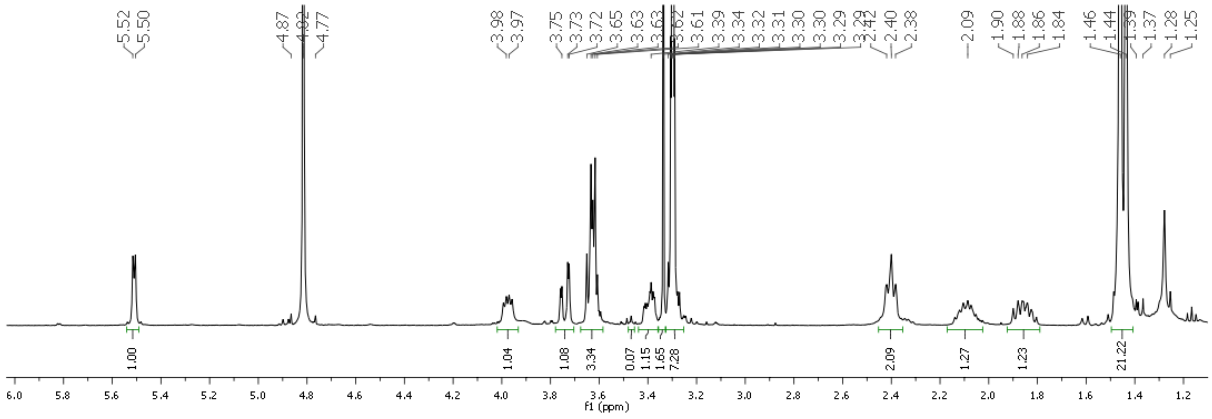
^{1}H NMR spectrum of compound **44** ($CDCl_3$, 400 MHz).

***N*$^{\alpha}$-(*tert*Butoxy-carbonyl)-*N*$^{\delta}$-(α-D-glucopyranosyl)-L-glutamine *O*- *tert*Butyl Ester (45)**



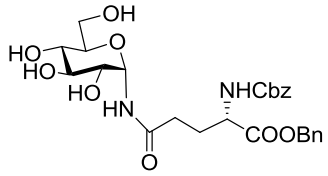
Phosphine **43** (2 eq) was added, at room temperature, to a 0.1 M solution of azide (1 eq) in 98:2 *N,N*-dimethylacetamide and DMPU. The solution was stirred for 18 h at 40 °C, then water was added and the mixture was stirred for an additional 2h at the same temperature. The solvent was

evaporated under reduced pressure, and the residue was diluted with water and extracted with Et_2O. The water layer was evaporated under reduced pressure and the crude was purified by flash chromatography ($CHCl_3$:MeOH:H_2O 90:10:1). Yield = 32 %. ^{1}H-NMR (400 MHz, CD_3OD, 25°C): δ = 5.51 (d, $J_{1,2}$ = 3.6 Hz, 1H, H-1), 3.90-4.10 (m, 1H, α-H-Glu), 3.68 (m, 1H, H-6), 3.60-3.70 (m, 3H, H-5, H-2, H-6'), 3.35-3.41 (m, 1H, H-3), 3.27-3.33 (m, 1H, H-4), 2.40 (m, 2 H, γ-CH_2-Glu), 2.09 (m, 1H, β-CH_2-Glu), 1.85 (m, 1H, β-CH_2-Glu). 1.48 (bs, 9H, OCH_3), 1.46 (bs, 9H, OCH_3) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 176.7–173.4 (CO), 158.3 (C_{quat}), 78.3 (C-1), 74.3 (C-5), 74.2 (C-3), 71.6 (C-4, C-2), 62.8 (C-6), 55.5 (α-C-Glu), 33.2 (γ-CH_2-Glu), 28.9-28.6 (OCH_3), 28.4 (β-CH_2-Glu) ppm.



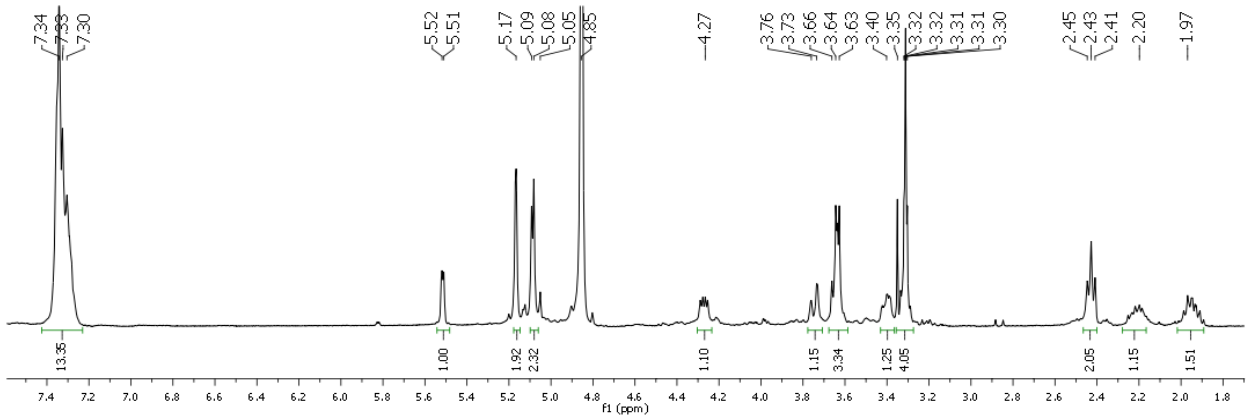
^{1}H NMR spectrum of compound **45** (CD_3OD, 400 MHz).

N^{α}-(Benzyloxycarbonyl)-N^{δ}-(α-D-glucopyranosyl)-L-glutamine *O*-Benzyl Ester (46)



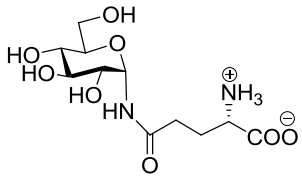
Phosphine **44** (1.5 eq) was added, at room temperature, to a 0.5 M solution of azide (1 eq) in 98:2 *N*,*N*-dimethylacetamide and DMPU. The solution was stirred for 48 h at 40 °C, then water was added and the mixture was stirred for an additional 2h at the same temperature. The solvent was evaporated under reduced pressure, and the residue was diluted with water and extracted with CH_2Cl_2. The water layer was evaporated under reduced pressure and the crude was purified by flash chromatography ($CHCl_3$:MeOH:H_2O 90:10:1). The compound was purified by flash chromatography ($CHCl_3$:MeOH:H_2O 90:10:1). Yield = 30 %. ^{1}H-NMR (400 MHz, CD_3O, 25°C): 7.30-7.42 (m, 10H, Ph), 5.52 (d, $J_{1,2}$ = 4.8 Hz, 1H, H-1), 5.18 (s, 2H, CH_2O), 5.10 (s, 2H, CH_2O), 4.30-4.38 (m, 1H, α-H-Glu), 3.75 (m, 1H, H-6), 3.60-3.70 (m, 3H, H-5, H-2, H-6'), 3.40-3.46 (m, 1H, H-3), 3,27-3.33 (m, 1H, H-4), 2.43 (m, 2 H, γ-CH_2-Glu), 2.22 (m, 1H, β-CH_2-Glu), 1.94 (m, 1H β-CH_2-Glu) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 173.6–171.5 (CO), 135.8 (C_{quat}Bn),

135.1 (C_{quat}Cbz), 142.3 (CH-Bn, CH-Cbz), 73.4 (C-1), 70.3 (C-3), 70.5 (C-5), 67.2 (CH_2-Cbz), 66.3 (C-4, C-2), 66.0 (CH_2-Cbz), 57.5 (C-6), 53.3 (α-C-Glu), 27.7 (γ-CH_2-Glu), 22.4 (β-CH2-Glu) ppm.

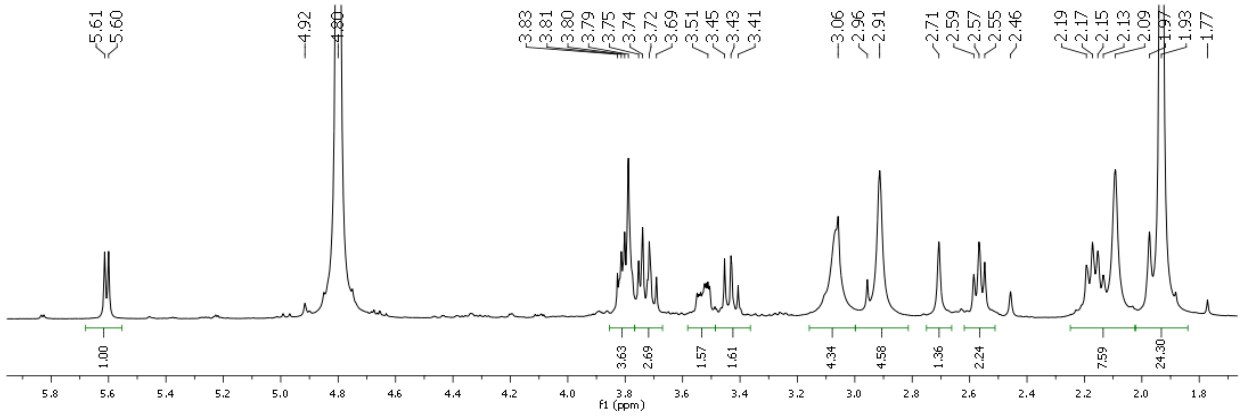


^{1}H NMR spectrum of compound **46** (CD_3OD, 400 MHz).

N^{δ}-(α-D-glucopyranosyl)-L-glutamine (61)



Compound **46** was dissolved in a mixture of DMA:AcOH:H_2O (85:10:5, c = 0.1M) and 10% Pd/C was added. The reaction mixture was stirred under H_2 for 2 h and then was filtered through a pad of Celite and washed with methanol. The solvent was evaporated to afford **61** in quantitative yield.
^{1}H NMR (400 MHz, D_2O, 25 °C): δ = 5.60 (d, $J_{1,2}$ = 5.2 Hz, 1 H, H-1), 3.68-3.84 (m, 5 H, H-2, H-6, H-3, α-H-Glu), 3.50–3.55 (m, 1 H, H-5), 3.41–3.46 (m, 1 H, H-4), 2.53-2.59 (m, 2 H, γ-CH_2-Glu), 2.12–2.20 (m, 1 H, β-CH_2-Glu) ppm. ^{13}C NMR (100 MHz, D_2O, 25 °C): δ = 172.8–170.4 (CO), 73.0 (C-1), 69.4 (C-3), 69.2 (C-5), 65.8 (C-4, C-2), 56.9 (C-6), 50.3 (α-C-Glu), 27.7 (γ-CH_2-Glu), 22.4 (β-CH_2-Glu) ppm.



^{1}H NMR spectrum of compound **61** (CD_3OD, 400 MHz).

4.4 References

[1] Nisic, F. ; Andreini, M. ; Bernardi, A. *Eur. J. Org.Chem,* **2009**, 5744-5751.

[2] Bianchi, A.; Bernardi, A. *J. Org. Chem.* **2006**, *71*, 4565-4577.

[3] Korytnyk, W.; Mills J. A. *J. Chem. Soc.* **1959**, 636-649.

[4] Soli, E. D.; Manoso, A. S.; Patterson, M. C.; DeShong, P. *J. Org. Chem.* **1999**, *64*, 3171-3177.

[5] Dedola, S.; Nepogodiev, S. A.; Hughes, D.L.; Field, R. A. *Acta Cryst*. **2008**, C64, 445-446.

[6] Rauchfuss, T. B. *Inorg. Chem.* **1977**, *16*, 2966-2968.

[7] Jeganathan, S.; Tsukarmoto, M.; Schlosser, M. *Synthesis* **1990**, 109-111.

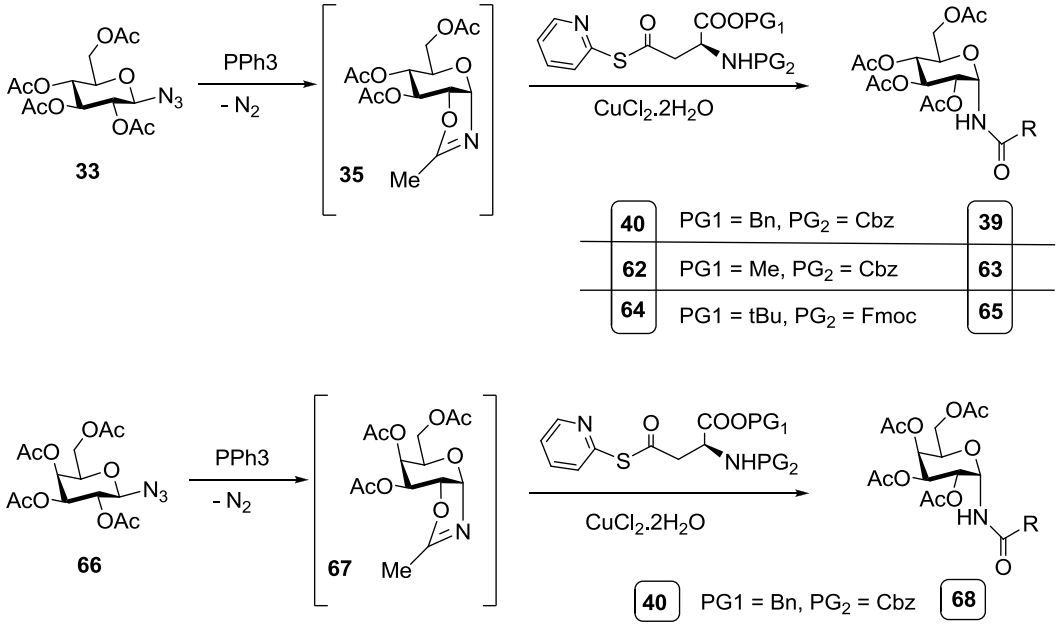
[8] Heinicke, J.; Köhler, M.; Peulecke, M.; He, M.; Kindermann, K.; Keim, W.; Fink, G. *Chem. Eur. J.* **2003**, *9*, 6093-6107.

[9] Furlano, D. C.; Calderon, S.N.; Chen, G.; Kirk, K.L. *J. Org. Chem.* **1988**, *53*, 3145-3147.

Chapter 5
Synthesis of α-*N*-linked glycosyl amino acids via DeShong methodology

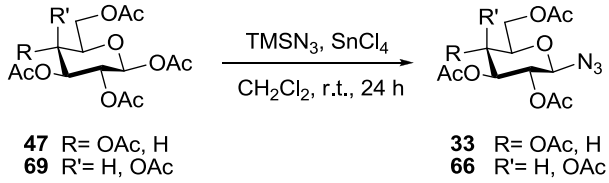
5.1 Synthesis of α-*N*-glucosyl and α-*N*-galactosyl asparagine derivatives

In **Section 3.4** we illustrated the reaction introduced by DeShong[1], explaining its mechanism and outcome. In this chapter we describe our studies on the application of this stereoselective reaction to aspartic and glutamic acid side-chain thioesters with glucosyl and galactosyl azides **33** and **66** (**Scheme 1**), which finally afforded a practical method for the gram-scale synthesis of α-*N*-glucosyl and α-*N*-galactosyl asparagine.



Scheme 1. Synthesis of α-*N*-linked glycosyl asparagine derivatives via DeShong methodology

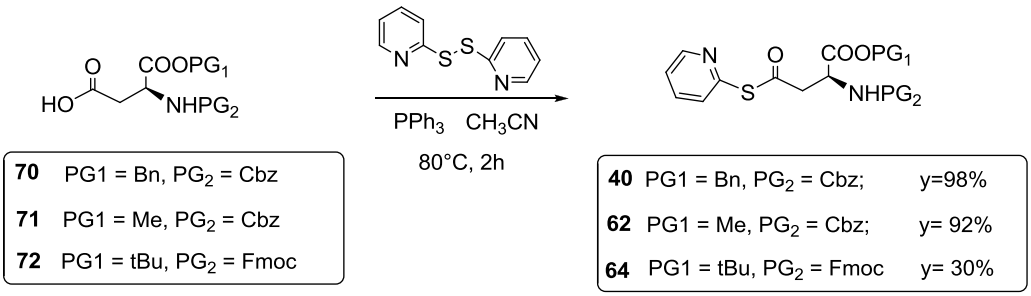
Tetra-*O*-acetyl-β-glucosyl and galactosyl azides **33** and **66** were synthesised from the corresponding tetra-*O*-acetates **47** and **69** with trimethylsilyl azide and tin tetrachloride employing the general procedure described by Paulsen (**Scheme 2**).[2-3]



Scheme 2. Stereoselective synthesis of tetra-*O*-acetyl-β-glycosyl azides **33** and **66** [2, 3]

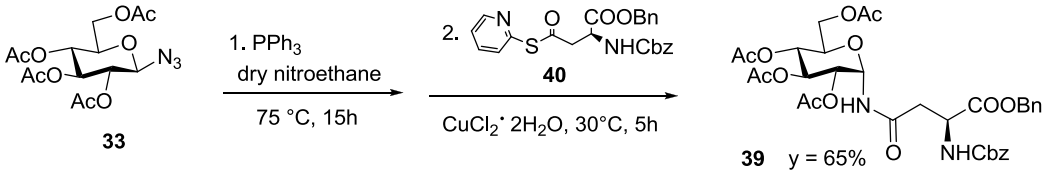
The pyridyl thioesters **40**, **62** and **64** were prepared by activation of protected aspartic or glutamic acids under Mukaiyama conditions (**Scheme 3**)[4] with PPh_3 and 2,2'-dithiodipyridine. The reaction must be refluxed for 2 h (longer reaction times can cause epimerization of the amino acid) and the resulting pyridyl thioesters can be chromatographed on short pads of silica, despite being rather

reactive substrates. They are best stored in inert atmosphere and must be handled in anhydrous conditions. NMR spectra should be recorded using $CDCl_3$ filtered on Al_2O_3.



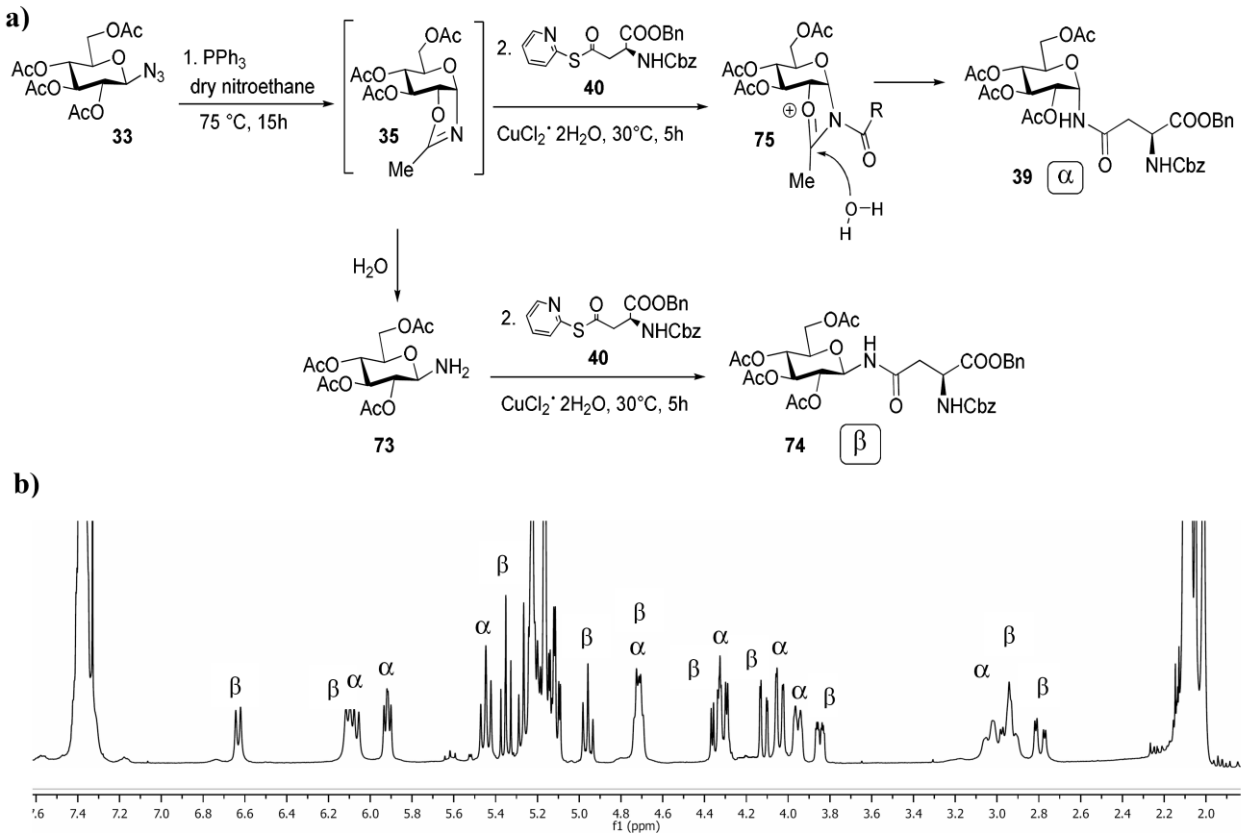
Scheme 3. General procedure for the synthesis of thio pyridyl ester.

Direct synthesis of Fmoc-protected aspartic acid thioester **64** (**Scheme 3**) gave low yields (30 %), presumably as a consequence of Fmoc group instability under the (basic) Mukaiyama conditions. Furthermore, the Fmoc-protected thioester **64** extensively decomposed under the DeShong conditions, so that the corresponding *N*-glycosyl asparagine **65** (**Scheme 1**) was obtained 1ith 10 % only yield. Therefore this synthon was abandoned in favour of the *N*-Cbz benzyl ester thioester **40**, which was obtained in very high yields and allowed simultaneous removal of N- and C- protecting groups. The synthesis of **39** from **33** using thioester **40** was described in 2003 by DeShong.[1] For this reaction we adopted a modified protocol, introduced by our group in 2008,[4b] which replaces dichloroethane with nitroethane as the solvent, and uses 3 Å molecular sieves (**Scheme 4**).



Scheme 4. Synthesis of α-*N*-linked glucosyl asparagine **39**.

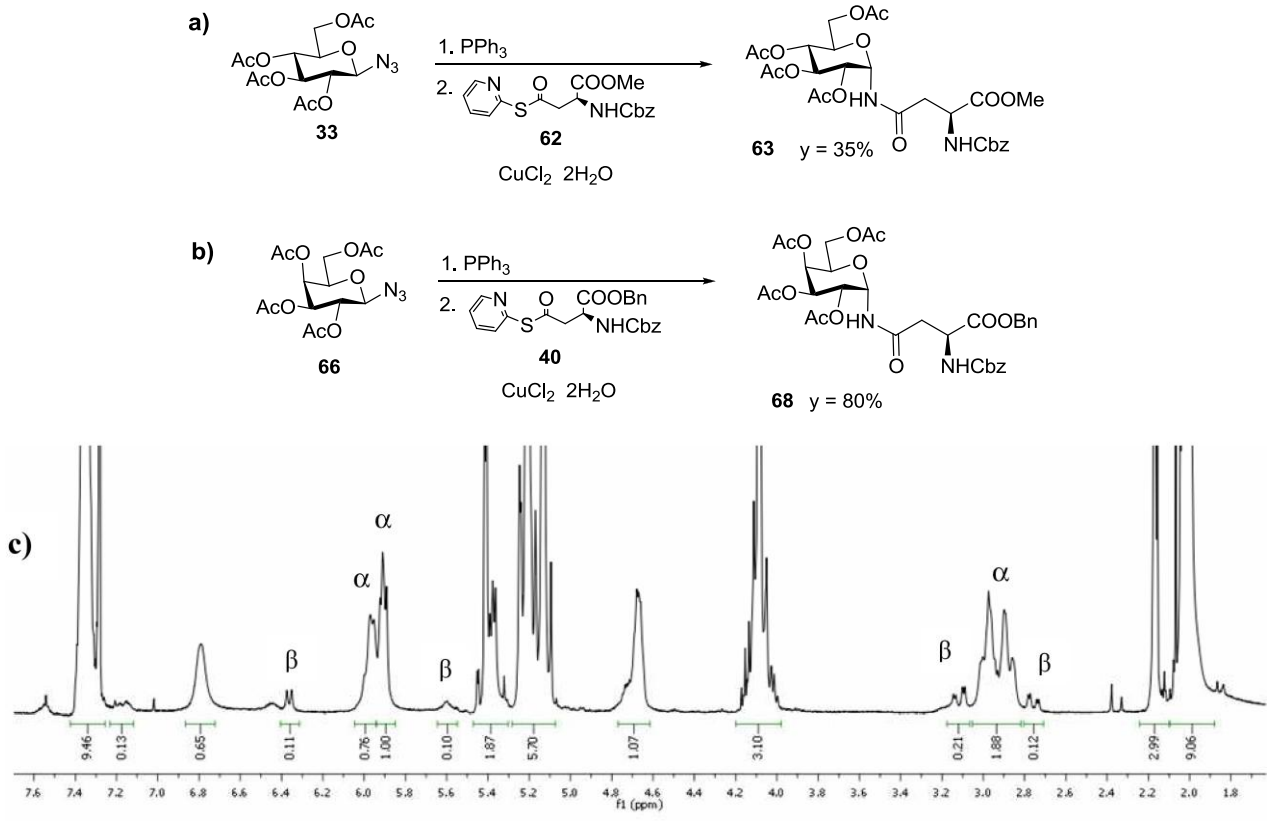
This reaction is rather complex and needs to be carefully worked-out or some by-products can easily be formed. The first step of the DeShong protocol, in fact, consisting in the formation of oxazoline **35** (**Scheme 5**), has to be conducted in severe anhydrous conditions to prevent the hydrolysis of oxazoline. If **35** is hydrolyzed by adventitious water, the generated amine isomerizes to the β anomer **73**, which can itself react with the thioester to give the β-*N*-linked glucosyl asparagine **74** (**Scheme 5**). This compound, once formed, can not be separated from the α-*N*-linked glucosyl asparagine **39** (**Scheme 5b** shows the ^{1}H NMR spectrum of a mixture of α and β products).



Scheme 5. a) Mechanism for the formation of α-*N*-linked and β-*N*-linked glucosyl asparagine **39** and **74**
b) ^{1}H NMR spectrum after purification from the reaction crude, **β** signals of **74**, **α** signals of **39**.

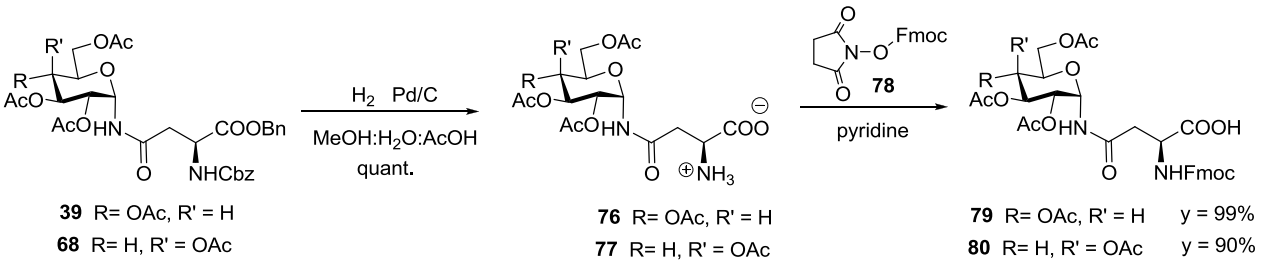
After reaction of **35** with **40**, on the contrary, water is required to hydrolyze intermediate **75** and to form **39** (**Scheme 5a**). Water is furnished by copper chloride, added to the mixture in its hydrated form. Hence the preparation of α-*N*-linked glucosyl asparagine through DeShong reaction requires a delicate equilibrium between water absence and water supply. In particular, we found it useful for the reproducibility of the reaction to maintain a precise ratio between the starting azide and the amount of molecular sieves used (azide: molecular sieves 1:3 in weight). If a small amount of **74** (≤ 10%) is formed, this can be separated by crystallization in a subsequent step, after protecting group removal and Fmoc protection of the nitrogen atom (see next in the text).

This method was further applied to the reaction of **33** with (*N*-Cbz; *O*Me) protected aspartyl thioester **62** to give α-*N*-linked glucosyl asparagine **63** (**Scheme 6a**), and to the reaction of (*N*-Cbz; *O*Bn) **40** with galactosyl azide **66** to give α-*N*-linked galactosyl asparagine **68** (**Scheme 6b**). Also in this case, if some β-*N*-linked galactosyl asparagine **68β** is formed, it can not be separated from the desired α-*N*-linked galactosyl asparagine **68** (**Scheme 6c** shows the ^{1}H NMR spectrum of a mixture of α and β products). Separation by crystallization is possible only in a subsequent step, after protecting group removal and Fmoc protection of the nitrogen atom, and only if less than 10 % of the undesired β-*N*-linked galactosyl asparagine **68β** is present.



Scheme 6. a) Synthesis of α-*N*-linked glucosyl asparagine **63**. b) Synthesis of α-*N*-linked galactosyl asparagine **68**. c) ^{1}H NMR spectrum after purification from a reaction crude in which **68β** is also present.

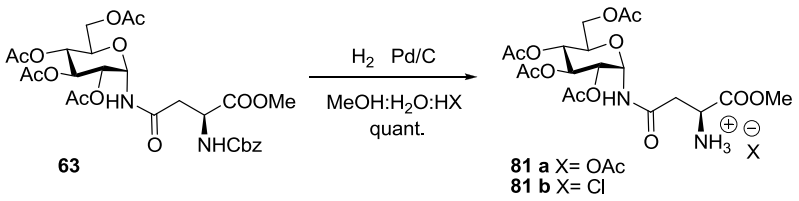
Deprotection of α-*N*-linked glucosyl and galactosyl asparagines **39** and **68** was obtained by catalytic hydrogenation (Pd/C in MeOH/H_2O/AcOH mixtures), which quantitatively afforded **76** and **77** (**Scheme 7**), respectively. They were in turn transformed into the (Fmoc)-protected derivatives **79** and **80** using *N*-(9*H*-fluorenylmethoxycarbonyloxy) succinimide (Fmoc-OSu, **78**) in pyridine[5] (**Scheme 7**). If any β-*N*-linked-glycosyl amino acid was formed in the DeShong phase, it can be separated at this stage by crystallization of **79** and **80** from CH_3CN/H_2O (1:6) which removes the remaining traces of the anomeric β isomer (see experimental section for details).



Scheme 7. Synthesis of Fmoc-α-*N*-linked glycosyl asparagine **77** and **78**.

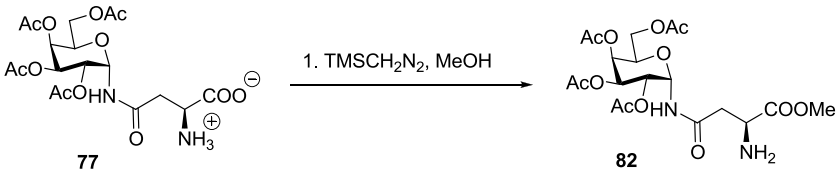
α-*N*- glucosyl asparagine methyl ester **81a** (**Scheme 8**) was obtained as the acetate salt with the same hydrogenation procedure from **63**. The hydrochloride salt **81b** could be obtained by hydrogenation of **63** in MeOH to which a stoichiometric amount of acetyl chloride had been added

(we will next see in **Chapter 6** how coupling yields could be improved starting from **81b** rather than **81a**)



Scheme 8. Synthesis of α-*N*-glucopyranosyl asparagine methyl ester **81**.

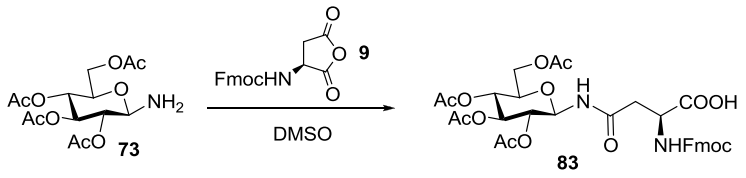
The corresponding α-*N*-galactosyl asparagine methyl ester **82** was obtained treating the free acid **77** with $TMSCH_2N_2$ in MeOH (**Scheme 9**). The compound was not isolated, but used directly for peptide coupling (see Chapter 6)



Scheme 9. Methylation for direct formation of peptide bond from **77**.

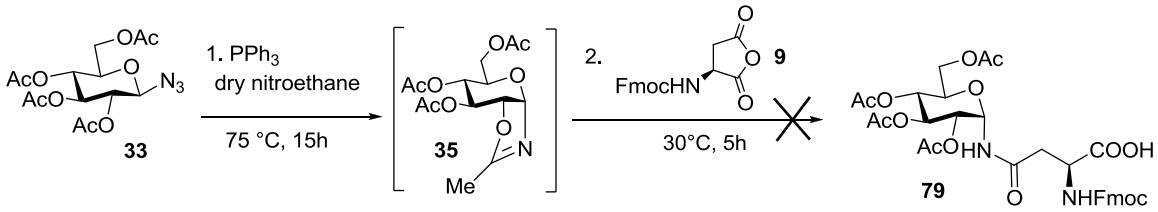
5.2 Attempt of direct synthesis of Fmoc-α-*N*-linked glycosyl asparagine 77

The procedures described above allowed the gram-scale synthesis of α-linked asparagine derivatives in the *galacto* and *gluco* series, but required some protecting group manipulations. In order to simplify the procedure, some alternatives were also explored. In particular, Selivanov et al. have recently reported a method for the preparation of Fmoc protected β-*N*-linked-glycosyl-asparagine **83**,[6] by alkylation of tetra-*O*-acetylglucosylamine **73** with *N*-Fmoc aspartic anhydride **9** (**Scheme 10**). It has been noted that the regioselectivity of *N*-Fmoc-aspartic anhydride aminolysis with tetra-*O*-acetyl glucosyl amines varied greatly depending on the polarity of the reaction media. In less polar solvents, the isoasparagine derivative was the main product, whereas in DMSO the desired β-*N*-linked-glycosyl-asparagine **83** was formed.



Scheme 10. Synthesis of Fmoc protected β-*N*-linked-glycosyl-asparagine

This suggested to modify the DeShong reaction employing *N*-Fmoc aspartic anhydride **9** in place of the pyridyl thioester, in the acylation of oxazoline **35** (**Scheme 11**).



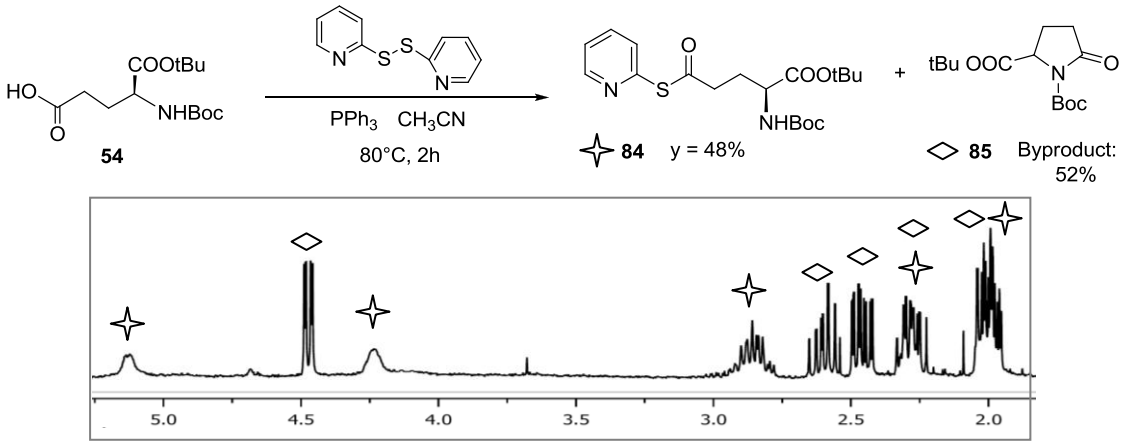
Scheme 11. Trial of DeShong reaction with *N*-Fmoc aspartic anhydride **9**

Three different types of conditions were employed using nitroethane as solvent: a) anhydride **9** without activating agents; b) anhydride **9** and 1% DMAP; c) anhydride **9** and 1% $La(OTf)_3$. No differences in reactivity were noted, neither when, in a subsequent experiment, the solvent was changed from nitroethane to nitroethane:DMSO 1:1 (the anhydride was dissolved in DMSO and added to the reaction mixture dissolved in nitroethane). The reactions were monitored through TLC, using as reference compound **79**, previously obtained. The reaction products corresponding to the two major TLC spots, both different from the reference one, were isolated. Their ^{1}H NMR spectra revealed a mixture of products of difficult identification (maybe some β products or the isoasparagine). No signal of anomeric α-*N*-linked amino acids was noted.

A group of experiments using a monothioanhydride as acylating agent, following the example of Crich et al.[7] also didn't afford the desired product **79**.

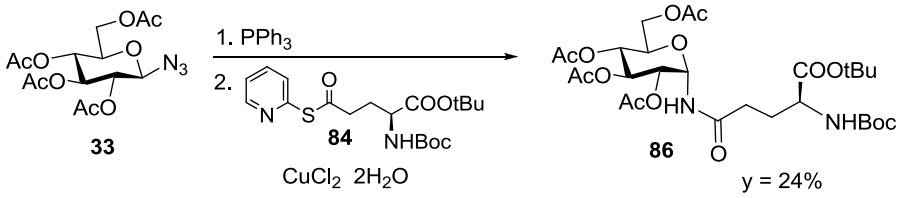
5.3 Synthesis of α-glucosyl glutamines

Synthesis of α-glucosyl glutamine derivatives suffered of low conversions, both in the preparation of the thiopyridyl ester **84** and in the DeShong reaction. During the preparation of activated ester **83** (**Scheme 12**) we assisted to an extensive formation of lactam **85**, due to cyclization of **84**. The lactam was probably formed during the reaction but possibly also during the purification since it could not be separated from **84** (flash chromatography with $CHCl_3$: MeOH 98:2 or hexane:EtOAc 6:4). The presence of both products was confirmed by NMR analysis (**Scheme 12**) and by ESI mass analysis.



Scheme 12.Formation of thiopyridyl ester **84** and lactam **85**. ^{1}H NMR signals of both compounds.

Reaction of **33** with this mixture gave **86** in 24 % yield (**Scheme 13**). A large quantity of lactam **85** was visible in the ^{1}H NMR spectrum of the crude.

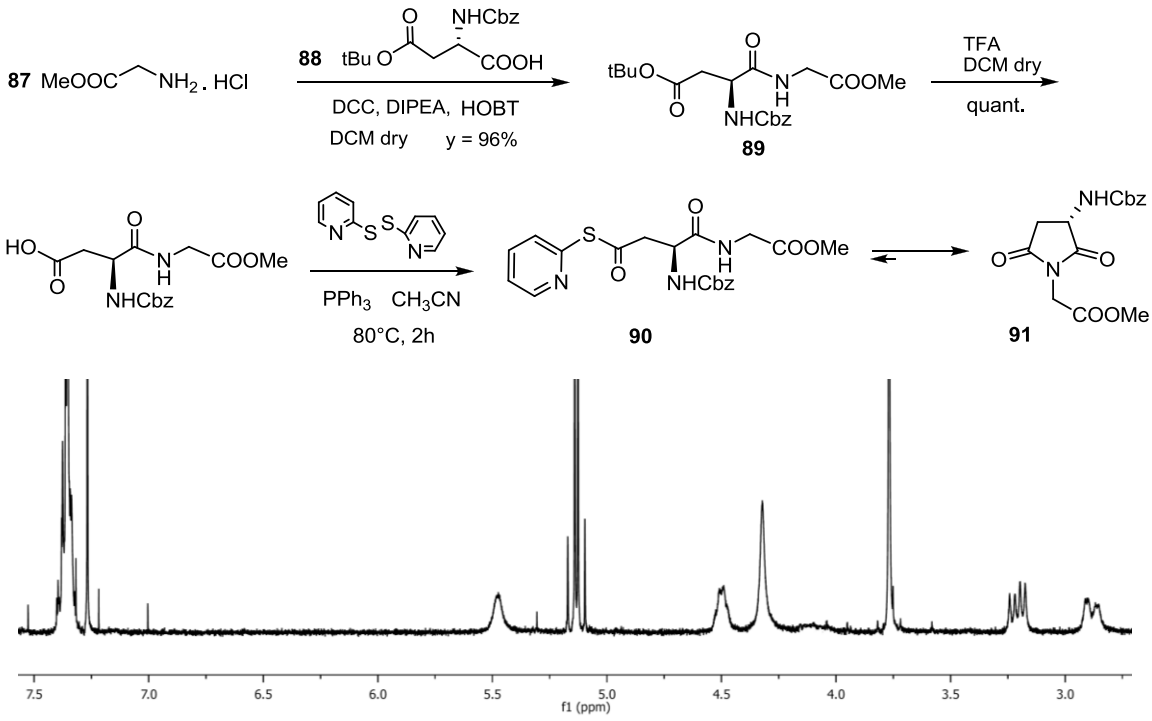


Scheme 13. Synthesis of α-*N*-linked glucosyl glutamine **86**.

Hence, we found that DeShong reaction has a strong limitation for the synthesis of glutamines and we decided to use it for the synthesis of α-*N*-linked glucosyl asparagines only.

5.4 DeShong reaction with a Gly-Asn dipeptide

With the aim of further exploring the scope of the DeShong protocol, we also attempted to use an activated dipeptide as the oxazoline acylating agent, in the hope of obtaining directly a α-*N*-linked glucosyl dipeptide. This procedure was not successful. Coupling reaction between Gly-OMe **87** and Cbz-Asp(OtBu)-OH **88** in solution (DCC, DIPEA, HOBT), afforded dipeptide **89** (**Scheme 14**). After removal of the tertbutyl ester, we attempted to synthesize the thioester of dipeptide **90** following the usual Mukaiyama conditions. The major product of the reaction was instead imide **91**, formed from cyclization of thioester **90** into the favoured and more stable five member-ring product.

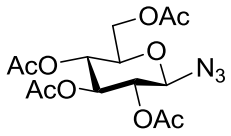


Scheme 14. Attempt of synthesizing pyridyl thioester **90** and ^{1}H NMR spectrum of **91**.

5.5 Experimental Section

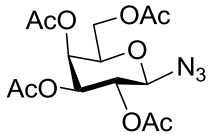
Solvents were dried by standard procedures: dichloromethane, and methanol were dried over calcium hydride; nitroethane, dichloroethane and pyridine were dried over activated molecular sieves. Reactions requiring anhydrous conditions were performed under nitrogen. ^{1}H and ^{13}C-NMR spectra were recorded at 400 MHz on a Bruker AVANCE-400 instrument. Chemical shifts δ for ^{1}H and ^{13}C are expressed in ppm relative to internal Me_4Si as standard. Signals were abbreviated as s, singlet; bs broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet. Mass spectra were obtained with a Bruker ion-trap Esquire 3000 apparatus (ESI ionization). Optical rotations $[\alpha]_D$ were measured in a cell of 1 dm pathlength and 1 ml capacity with a Perkin-Elmer 241 polarimeter. Thin layer chromatography (TLC) was carried out with precoated Merck F_{254} silica gel plates. Flash chromatography (FC) was carried out with Macherey-Nagel silica gel 60 (230-400 mesh).

Synthesis of 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl azide (33)[2, 3]



Tin tetrachloride (179 μL, 1.537 mmol, 0.3 eq) was added at room temperature and under nitrogen, to a solution of 1,2,3,4,6-penta-*O*-acetyl-D-glucopyranose **47** (2 g, 5.124 mmol, 1 eq) in dry CH_2Cl_2 (20 mL, 0.25 M). Then trimethylsilyl azide (952 μL, 7.174 mmol, 1.4 eq) was added and the reaction mixture was stirred at room temperature overnight. The reaction was monitored by TLC (60:40 hexane/AcOEt). After 24 h CH_2Cl_2 was added and the solution was washed with saturated Na_2CO_3 and then with water. The organic layer was dried over Na_2SO_4, filtered and evaporated under reduced pressure. The product was purified by flash chromatography using 60:40 hexane/AcOEt as the eluent. Quantitative yield. ^{1}H-NMR (400 MHz, $CDCl_3$, 25°C): δ = 5.20 (dd, $J_{2,3} = J_{3,4}$ = 9.5 Hz, 1H, H-3), 5.08 (dd, $J_{3,4} = J_{4,5}$ = 9.5 Hz, 1H, H-4), 4.94 (dd, $J_{1,2}$ = 8.9 Hz, $J_{2,3}$ = 9.5 Hz, 1H, H-2), 4.63 (d, $J_{1,2}$ = 8.9 Hz, 1H, H-1), 4.26 (dd, $J_{5,6}$ = 4.7 Hz, $J_{6,6'}$ = 12.5 Hz, 1H, H-6), 4.15 (dd, $J_{5,6'}$ = 2.3 Hz, $J_{6,6'}$ = 12.5 Hz, 1H, H-6'), 3.78 (m, 1H, H-5), 2.08, 2.06, 2.01, 1.99 (4s, 12H, 4xOAc).

Synthesis of 2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl azide (66) 2,3

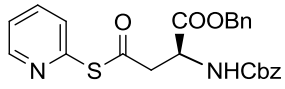


Tin tetrachloride (179 µL, 1.537 mmol, 0.3 eq) was added at room temperature and under nitrogen, to a solution of of 1,2,3,4,6-penta-*O*-acetyl-D-galactopyranose **69** (2 g, 5.124 mmol, 1 eq) in dry CH_2Cl_2 (20 mL, 0.25 M). Then trimethylsilyl azide (952 µL, 7.174 mmol, 1.4 eq) was added and the reaction mixture was stirred at room temperature overnight. The reaction was monitored by TLC (60:40 hexane/AcOEt). After 24 h CH_2Cl_2 was added and the solution was washed with saturated Na_2CO_3 and then with water. The organic layer was dried over Na_2SO_4, filtered and evaporated under reduced pressure. The product was purified by flash chromatography using 60:40 hexane/AcOEt as the eluent. Quantitative yield. ^{1}H-NMR (400 MHz, $CDCl_3$, 25°C): δ = 5.42 (dd, $J_{3,4}$ = 3.2 Hz, $J_{4,5}$ = 1.2 Hz, 1H, H-4), 5.16 (q, $J_{1,2}$ = 8.4 Hz, $J_{2,3}$ = 10.4 Hz, 1H, H-2), 5.03 (dd, $J_{2,3}$ = 10.4 Hz, $J_{3,4}$ = 3.2 Hz, 1H, H-3), 4.59 (d, $J_{1,2}$ = 8.4 Hz, 1H, H-1), 4.20-4.12 (m, 2H, H-6, H-6'), 4.01 (dt, $J_{4,5}$ = 1.2 Hz, 1H, H-5), 2.16, 2.10, 2.07, 1.98 (4s, 12H, 4xOAc). ^{13}C-NMR (100 MHz, $CDCl_3$, 25°C): δ = 170.6, 170.3, 170.2, 169.6, 88.5 (C-1), 73.1 (C-5), 70.9 (C-3), 68.3 (C-2), 67.1 (C-4), 61.4 (C-6), 20.9-20.7 (4xOAc).

General procedure for the synthesis of thiopyridyl esters [4b]

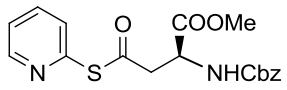
The amino acid (1eq), PPh_3 (1.2 equiv) and 2,2'-dithiodipyridine (1.2 equiv) were dissolved in dry CH_3CN (c = 0.2M). The solution was heated at 80°C for 2 h, then cooled to room temperature. The solvent was evaporated under vacuum and the residue was purified by flash chromatography (hexane/EtOAc) on a short pad of silica.

1-Benzyl-5-[*S*-Pyridinyl]-*N*-(Benzyloxycarbonyl)-L-aspartate (40)



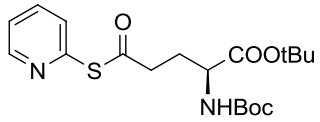
Purification: hexane: EtOAc 6:4. Yield = 98%.$[\alpha]^{20}_{D}$ = +34 (*c* = 0.1, $CHCl_3$ on Al_2O_3). ^{1}H NMR (400 MHz, $CDCl_3$ on Al_2O_3, 25°C): δ = 8.61 (ddd, $J_{a\text{-}b}$ = 4.8 Hz, $J_{a\text{-}c}$ = 1.8 Hz, $J_{a\text{-}d}$ = 0.9 Hz, 1H, H_{Ar}-a), 7.71 (td, $J_{c\text{-}d}$ = 7.8 Hz, $J_{c\text{-}a}$ = 1.8 Hz, 1H, H_{Ar}-c), 7.49 (d, $J_{d\text{-}c}$ = 7.8 Hz, $J_{d\text{-}a}$ = 0.9 Hz, 1H, H_{Ar}-d), 7.37–7.28 (m, 11H, HCbz, HBn, H_{Ar}-b), 5.71 (d, *J* = 8.4 Hz, 1H, NHCbz), 5.17 (s, 2H, CH_2-Bn), 5.11 (s, 2H, CH_2- Cbz), 4.68 (dt, J_{NH} = 8.4 Hz, $J_{\alpha\text{-}\beta}$ = 4.5 Hz, 1H, α-H-Asn), 3.42 (dd, J_{gem} = 17.1 Hz, $J_{\alpha\text{-}\beta}$ = 4.5 Hz, 1H, β-CH_2-Asn), 3.31 (dd, J_{gem} = 17.1 Hz, $J_{\alpha\text{-}\beta}$ = 4.5 Hz, 1H, β-CH_2-Asn) ppm.

1-Methyl-5-[*S*-Pyridinyl]-*N*-(Benzyloxycarbonyl)-L-aspartate (62)



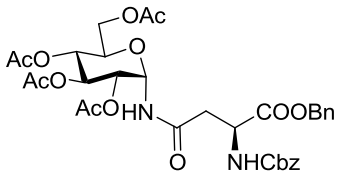
Purification: hexane: EtOAc 6:4. Yield = 92% ^{1}H NMR (400 MHz, $CDCl_3$, 25°C): δ = 8.62 (ddd, $J_{a\text{-}b}$ = 4.9 Hz, $J_{a\text{-}c}$ = 1.8 Hz, $J_{a\text{-}d}$ = 0.8 Hz, 1H, H_{Ar}-a), 7.77 (td, $J_{c\text{-}d}$ = 7.6 Hz, $J_{c\text{-}a}$ = 1.8 Hz, 1H, H_{Ar}-c), 7.60 (d, $J_{d\text{-}c}$ = 7.6 Hz, $J_{d\text{-}a}$ = 0.8 Hz, 1H, H_{Ar}-d), 7.37–7.30 (m, 6H, HCbz, H_{Ar}-b), 5.75 (d, J = 8.1 Hz, 1H, NHCbz), 5.12 (s, 2H, CH_2-Cbz), 4.76-4.52 (m, 1H, α-H-Asn), 3.74 (s, 3H, OMe), 3.42 (dd, J_{gem} = 17.0 Hz, $J_{\alpha\text{-}\beta}$ = 4.6 Hz, 1H, β-CH_2-Asn), 3.32 (dd, J_{gem} = 17.0 Hz, $J_{\alpha\text{-}\beta}$ = 4.6 Hz, 1H, β-CH_2-Asn) ppm.

1-*tert*Butyl-5-[*S*-Pyridinyl]-*N*-(*tert*Butoxycarbonyl)-L-glutamate (83)



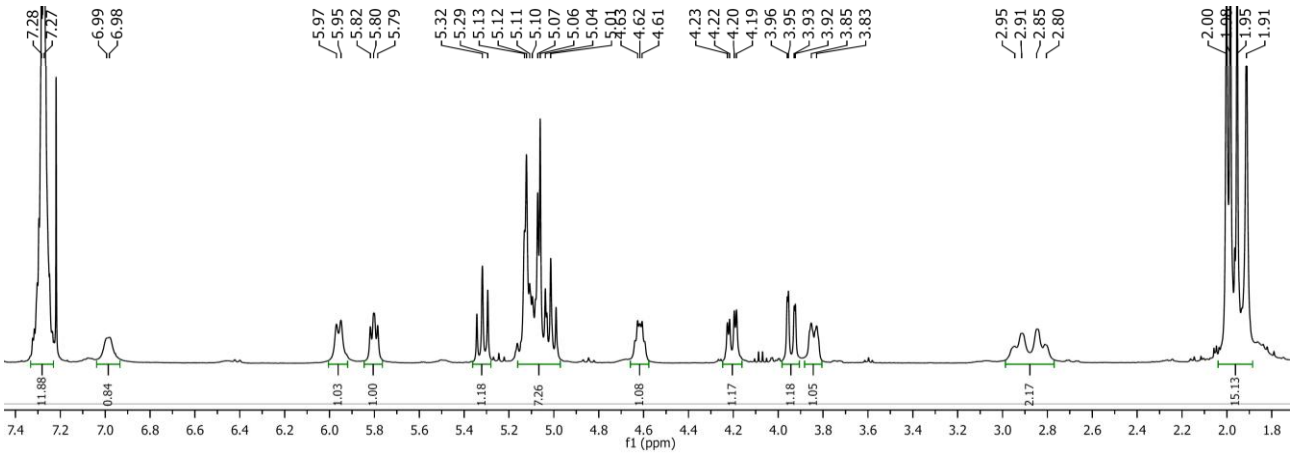
Purification: hexane: EtOAc 1:1. Yield = 48%. ^{1}H NMR (400 MHz, $CDCl_3$, 25°C): δ = 8.65 (d, J = 4.8 Hz, 1H, H_{Ar}-a), 7.75 (t, J = 7.6 Hz, 1H, H_{Ar}-c), 7.64 (d, J = 7.6 Hz, 1H, H_{Ar}-d), 7.28 (td, J = 7.6 Hz, J = 4.8 Hz, 1H, H_{Ar}-b), 5.12 (bs, 1H, NHBoc), 4.23 (m, 1H, α-H-Glu), 2.91-2.79 (m, 1H, γ-CH_2-Glu), 2.32-2.27 (m, 1H, γ-CH_2-Glu), 2.04-1.96 (m, 2H, β-CH_2-Glu), 1.47 (s, 9H, OCH_3). 1.44 (s, 9H, OCH_3) ppm. MS (ESI) *m/z* = 419.1 $[M + Na]^+$

N^{α}-Benzyloxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl)-L-asparagine Benzyl Ester (39)



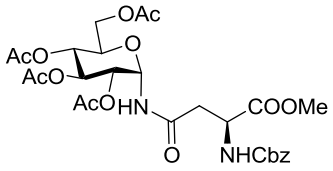
2,3,4,6-Tetra-*O*-acetyl-β-D-glucosyl azide **33** (510 mg, 1.366 mmol, 1 equiv.) and Ph_3P (394 mg, 1.503 mmol, 1.1 equiv.) were dissolved in nitroethane (6 mL) in the presence of ground 3 Å molecular sieves (ca. 1.5 g). The resulting solution was heated at reflux for 15 h under nitrogen and then cooled to room temperature. Cbz-Asp-(SPy)-OBn **40** (820 mg, 1.82 mmol, 1.3 equiv.) and $CuCl_2 \cdot 2H_2O$ (310 mg,1.82 mmol, 1.3 equiv.) were added sequentially and the reaction mixture was stirred at 30 °C for 6 h. After completion (TLC, 1:1 hexane/EtOAc) the reaction mixture was diluted with CH_2Cl_2 and filtered through a celite pad. The solvent was evaporated and the crude was dissolved in EtOAc and washed with a saturated ammonium chloride solution. The organic phase

was dried with sodium sulfate and the solvents evaporated. The crude was purified by flash chromatography (1:1 hexane/EtOAc) to afford **39** (610 mg) in 65% yield. The analytical data were consistent with those reported in the literature.[1] ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 7.30–7.35 (m, 10 H, HArCbz, HArBn), 6.99 (bs, 1 H, Glc-NHAsn), 5.96 (bs, 1 H, 1-H), 5.80 (t, J = 6.2 Hz, 1 H, 3-H), 5.25–5.35 (m, 2 H, 2-H, 4-H), 5.00–5.20 (m, 4 H, CH_2-Cbz, CH_2-Bn), 4.62 (bs, 1 H, α-H-Asn), 4.22 (dd, J_{gem} = 8.0, $J_{6\text{-}5}$ = 4.0 Hz, 1 H, 6-H), 3.95 (dd, J_{gem} = 8.0, $J_{6'\text{-}5}$ = 1.8Hz, 1 H, 6'-H), 3.84 (m, 1 H, 5-H), 3.02–2.80 (m, 2 H, β-CH_2-Asn), 2.04 (s, 3 H, CH_3CO), 2.02 (s, 3 H, CH_3CO), 2.00 (s, 3 H, CH_3CO), 1.96 (s, 3 H, CH_3CO) ppm. MS (ESI): m/z = 709.6 $[M + Na]^+$.



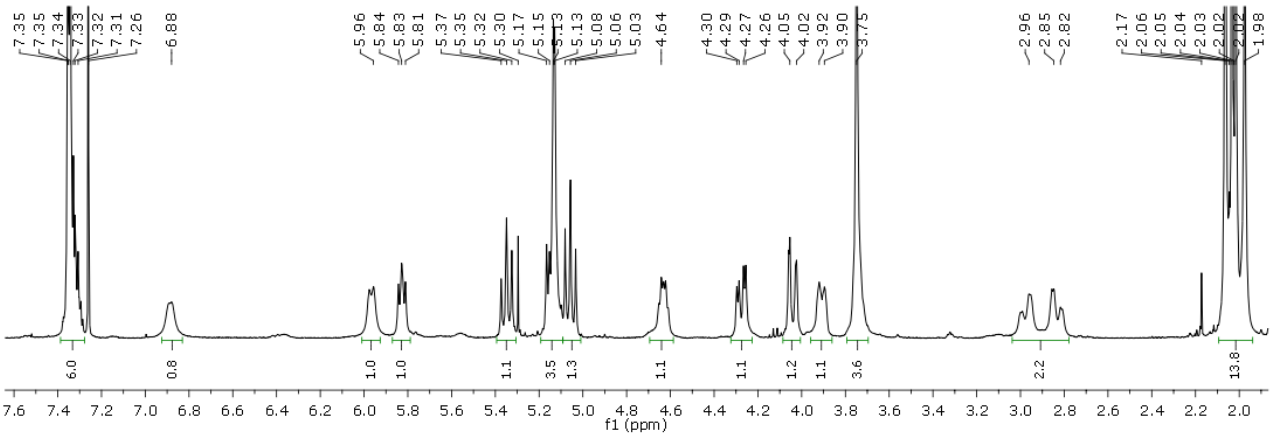
^{1}H-NMR spectrum of **39** ($CDCl_3$, 400 MHz)

N^{α}-Benzyloxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl)-L-asparagine Methyl Ester (63)

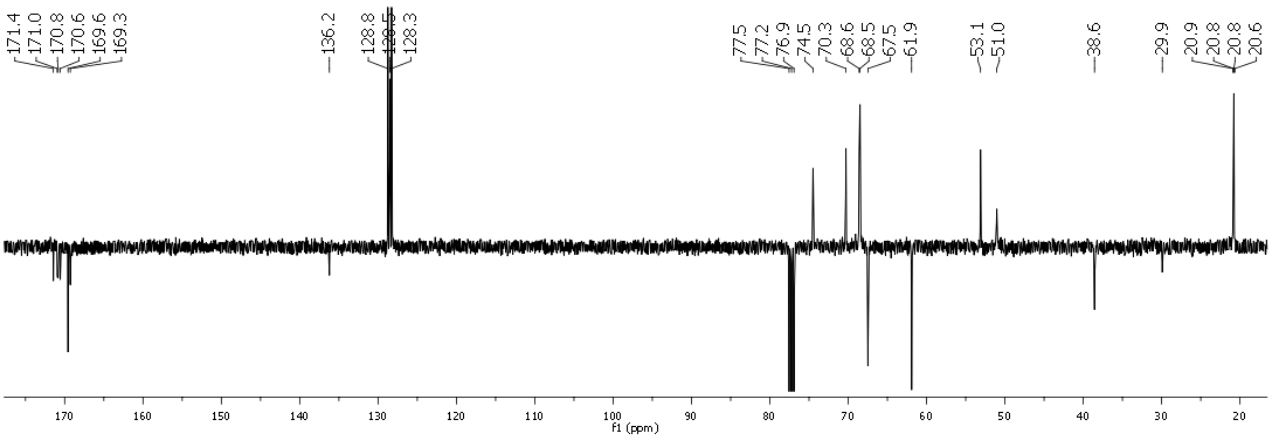


2,3,4,6-Tetra-*O*-acetyl-β-D-glucosyl azide **33** (250 mg, 0.670 mmol, 1 equiv.) and Ph_3P (193 mg, 0.737 mmol, 1.1 equiv.) were dissolved in nitroethane (5 mL) in the presence of ground 3 Å molecular sieves (ca. 900 mg). The resulting solution was heated at reflux for 15 h under N_2 and under stirring, then cooled to room temperature before adding Cbz-Asp-(SPy)-OMe **62** (325 mg, 0.871 mmol, 1.3 equiv.) and $CuCl_2 \cdot 2H_2O$ (148 mg, 0.871 mmol, 1.3 equiv.). The mixture was stirred at 30 °C for 6 h. After completion (TLC, 1:1 hexane/EtOAc) the reaction mixture was diluted with CH_2Cl_2 and filtered through a celite pad. The solvent was evaporated and the crude was dissolved in EtOAc and washed with a saturated ammonium chloride solution. The organic phase was dried with Na_2SO_4 and the solvent removed under reduced pressure. The crude was purified by flash chromatography (1:1 hexane/EtOAc) to afford **63** (143 mg) in 35% yield. $[\alpha]^{20}_{D}$ = +45.6 (c =

0.15, MeOH). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 7.36–7.30 (m, 5 H, HAr-Bn), 6.88 (br. s, 1 H, Glc-NHAsn), 5.96 (d, J = 7.6 Hz, 1 H, NH-Cbz), 5.83 (t, 1 H, 1-H), 5.35 (t, $J_{3,4} = J_{3,2}$ = 9.6 Hz, 1 H, 3-H), 5.18–5.10 (m, 3 H, 2-H, CH_2- Cbz), 5.06 (t, $J_{4,3} = J_{4,5}$ = 9.6 Hz, 1 H, 4-H), 4.64 (br. s, 1 H, α-HAsn), 4.27 (dd, J_{gem} = 12.0, $J_{6,5}$ = 4.0 Hz, 1 H, 6-H), 4.04 (dd, J_{gem} = 12.0, $J_{6'-5}$ = 1.8 Hz, 1 H, 6'-H), 3.90 (m, 1 H, 5-H), 3.75 (s, 3H, $COOCH_3$), 3.02–2.78 (m, 2 H, β-CH_2-Asn), 2.06 (s, 3H, CH_3CO), 2.03 (s, 3H, CH_3CO), 2.02 (s, 3H, CH_3CO), 1.98 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 171.4-169.3 (CO), 136.2 (C_{quat}Cbz), 128.7, 128.4, 128.3 (CH-Cbz), 74.5 (C-1), 70.3 (C-3), 68.6 (C-2, C-4), 68.5 (C-5), 67.5 (CH_2-Cbz), 61.9 (C-6), 53.1 ($COOCH_3$), 51.0 (α-C-Asn), 38.6 (β-CH2-Asn), 20.9, 20.8, 20.6 (4-OAc) ppm. MS (ESI): m/z = 633.3 $[M + Na]^+$.

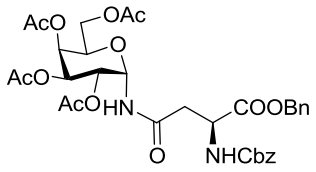


^{1}H-NMR spectrum of **63** ($CDCl_3$, 400 MHz)

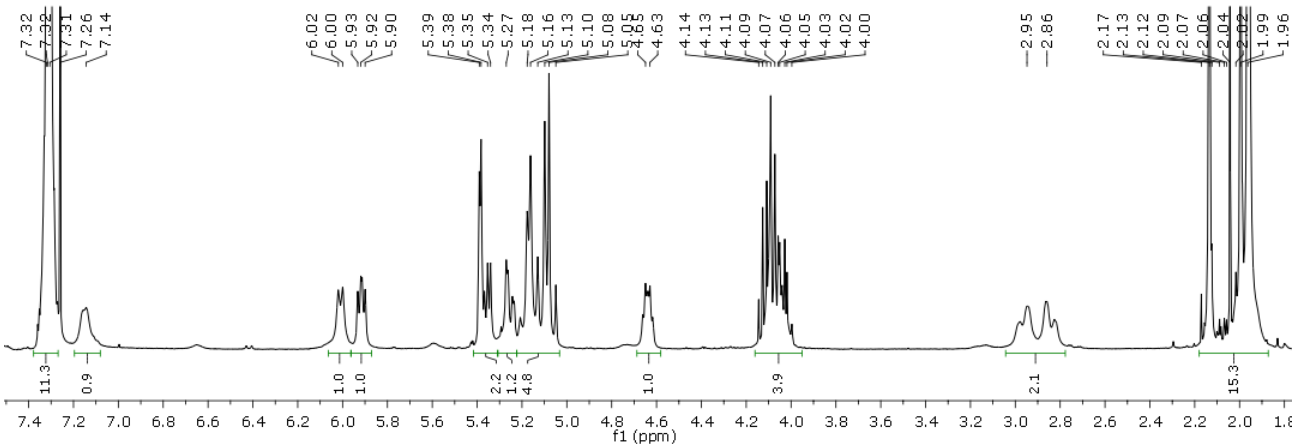


^{13}C-NMR spectrum of **63** ($CDCl_3$, 100 MHz)

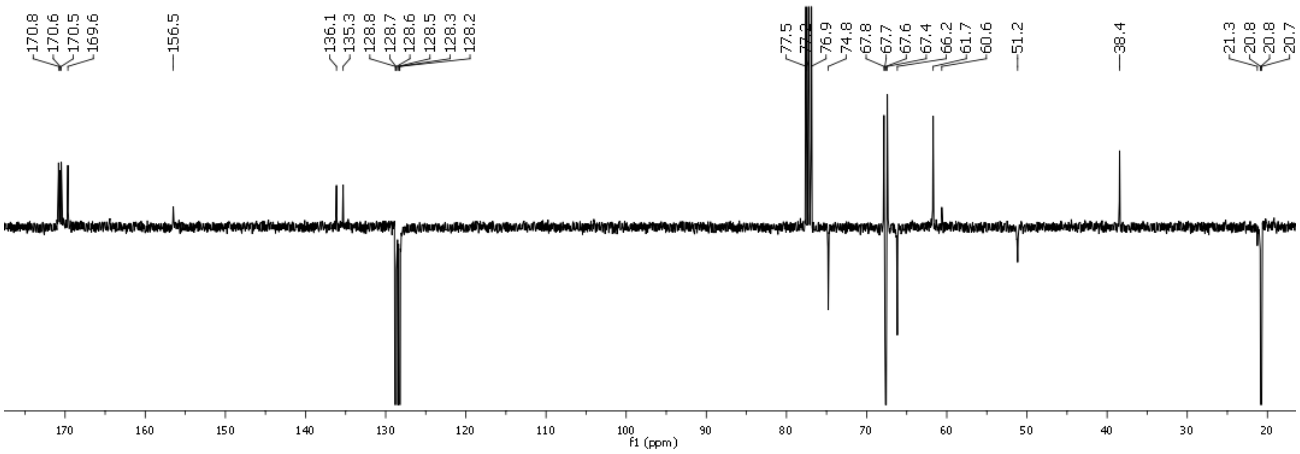
N^{α}-Benzyloxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagine Benzyl Ester (68)



2,3,4,6-Tetra-*O*-acetyl-β-D-galactosyl azide **66** (402 mg, 1.077 mmol, 1 equiv.) and Ph_3P (311 mg, 1.185 mmol, 1.1 equiv.) were dissolved in nitroethane (6 mL) in the presence of ground 3 Å molecular sieves (ca. 1.2 g). The resulting solution was heated at reflux for 15 h under nitrogen and then cooled to room temperature. Cbz-Asp-(SPy)-OBn **40** (630 mg, 1.400 mmol, 1.3 equiv.) and $CuCl_2 \cdot 2H_2O$ (239 mg, 1.400 mmol, 1.3 equiv.) were added sequentially and the reaction mixture was stirred at 30 °C for 5 h. After completion (TLC, 1:1 hexane/EtOAc) the reaction mixture was diluted with CH_2Cl_2 and filtered through a celite pad. The solvent was evaporated and the crude was dissolved in EtOAc and washed with a saturated ammonium chloride solution. The organic phase was dried with sodium sulfate and the solvents evaporated. The crude was purified by flash chromatography (1:1 hexane/EtOAc) to afford **68** (592 mg) in 80% yield. $[\alpha]^{20}_{D}$ = +65.4 (*c* = 1, MeOH). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 7.32–7.28 (m, 10H, H_{Ar}Cbz, H_{Ar}Bn), 7.14 (bs, 1H, Gal-NH-Asn), 6.01 (d, J = 7.6 Hz, 1H, NH-Cbz), 5.91 (dd, $J_{1,NH}$ = 7.6, $J_{1,2}$ = 5.2 Hz, 1H, 1-H), 5.41–5.30 (m, 2H, 2-H, 4-H), 5.25 (dd, $J_{3,4}$ = 3.2, $J_{3,2}$ = 11.0 Hz, 1H, 3-H), 5.18–5.05 (m, 4H, CH_2-Cbz, CH_2-Bn), 4.69 (bs, 1H, α-H-Asn), 4.11–3.98 (m, 3H, 5-H, 6-H), 3.05–2.81 (m, 2H, β-CH_2-Asn), 2.13 (s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO), 2.01 (s, 3H, CH_3CO), 1.96 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 170.8–169.6 (CO), 156.5 (NHCO), 136.1 (C_{quat}Bn), 135.3 (C_{quat}Cbz), 142.7 (CH-Bn, CH-Cbz), 74.8 (C-1), 68.0 (C-3, C-4, C-5), 67.6 (CH_2-Bn), 67.2 (C-2), 66.2 (CH_2-Cbz), 61.7 (C-6), 51.2 (α-C-Asn), 38.5 (β-CH_2-Asn), 21.3, 20.7 (4xOAc) ppm. MS (ESI): m/z = 709.4 $[M + Na]^+$.

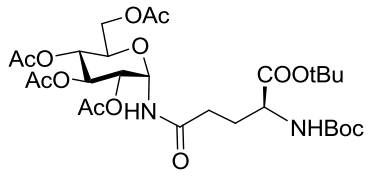


^{1}H-NMR spectrum of **68** ($CDCl_3$, 400 MHz)



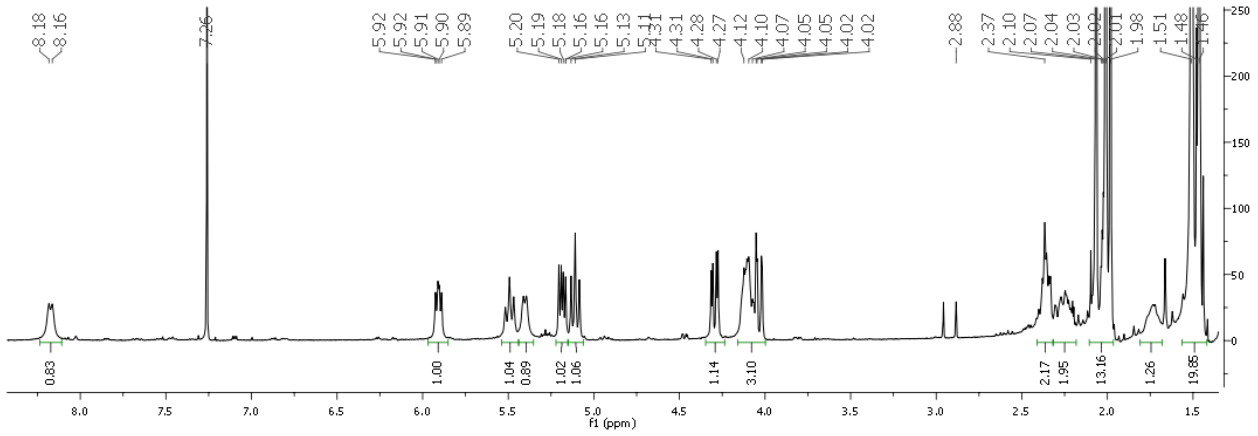
^{13}C-NMR spectrum of **68** ($CDCl_3$, 100 MHz)

N^{α}-*tert*Butoxycarbonyl-N^{δ}-(2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl)- L-glutamine *tert*Butyl Ester (85)

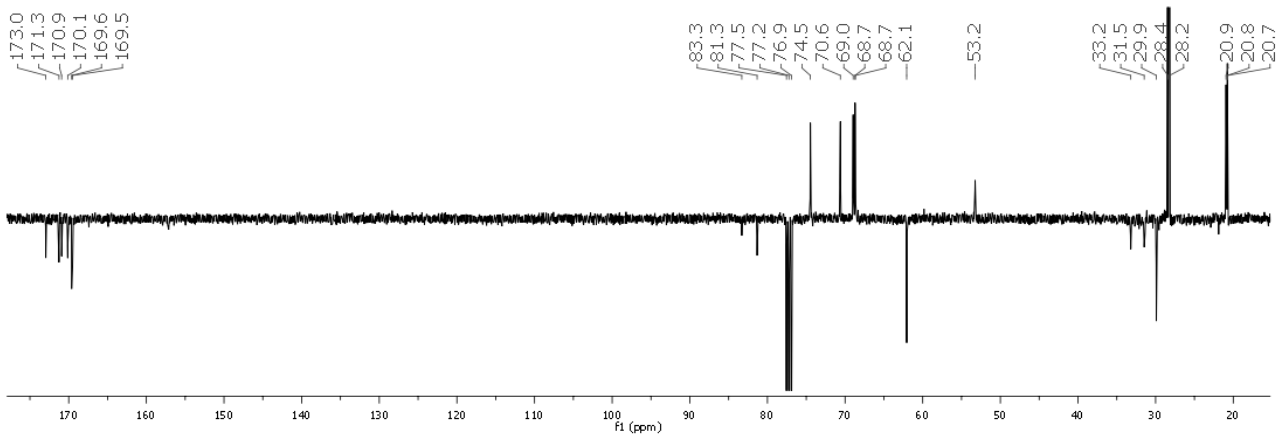


2,3,4,6-Tetra-*O*-acetyl-β-D-glucosyl azide **33** (51 mg, 0.135 mmol, 1 equiv.) and Ph_3P (58 mg, 0.221 mmol, 1.1 equiv.) were dissolved in dichloroethane (5mL) in the presence of ground 4Å molecular sieves. The resulting solution was heated at reflux for 15 h under N_2 and under stirring, then cooled to room temperature before adding Boc-Glu-(SPy)-OtBu **83** (63 mg, 0.160 mmol, 1.2 equiv.) and $CuCl_2{\cdot}2H_2O$ (24 mg, 0.176 mmol, 1.3 equiv.). The mixture was stirred at 30 °C for 24 h. After completion (TLC, 1:1 hexane/EtOAc) the reaction mixture was diluted with Et_2O and filtered through a celite pad. The solvent was evaporated and the crude was dissolved in Et_2O and washed with water two times. The organic phase was dried with Na_2SO_4 and the solvent removed under reduced pressure. The crude was purified by flash chromatography (1:1 hexane/EtOAc) to afford **65** (20 mg) in 24% yield. $[\alpha]^{20}{}_{D}$ = +45.6 (*c* = 0.15, MeOH). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): 8.17 (d, $J_{NH\text{-}1}$ = 7.8 Hz, 1H, Glc-NHGlu), 5.91 (dd, $J_{1\text{-}NH}$ = 7.8 Hz, $J_{1\text{-}2}$ = 5.6 Hz, 1H, 1-H), 5.49 (t, $J_{3\text{-}4}$ = $J_{3\text{-}2}$ = 9.8 Hz, 1H, 3-H), 5.40 (d, *J* = 7.1 Hz, 1H, NHBoc), 5.18 (dd, $J_{3\text{-}2}$ = 9.8, $J_{2\text{-}1}$ = 5.6 Hz, 1H, 2-H), 5.11 (t, $J_{4\text{-}3}$ = $J_{4\text{-}5}$ = 9.8 Hz, 1H, 4-H), 4.29 (dd, J_{gem} = 12.3, $J_{6,5}$ = 3.9 Hz, 1H, 6-H), 4.16-4.06 (m, 2H, 5-H, α-H-Glu), 4.03 (dd, J_{gem} = 12.3, $J_{6'\text{-}5}$ = 2.2 Hz, 1H, 6'-H), 2.35 (m, 2H, γ-CH_2-Glu), 2.32 – 2.19 (m, 1H, β-H-Glu), 2.10 (s, 3H, CH_3CO), 2.06 (s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO), 1.98 (s, 3H, CH_3CO), 1.72 (s, 1H, β-H-Glu), 1.48 (s, 9H, OCH_3), 1.46 (s, 9H, OCH_3). ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 173.0-169.5 (CO), 83.3, 83.1 (C_{quat}Boc-tBu), 74.5 (C-1), 70.6

(C-3), 69.0 (C-2),68.6 (C-4), 68.5 (C-5), 62.1 (C-6), 53.2 (α-C-Glu), 33.2 (γ-CH_2-Glu), 31.5 (β-CH_2-Glu), 28.4 (OCH_3-Boc),, 28.2 (OCH_3-tBu), 20.9, 20.8, 20.7 (4 x OAc) ppm.

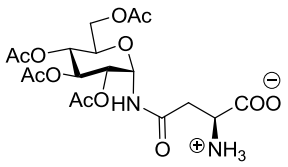


^{1}H-NMR spectrum of **85** ($CDCl_3$, 400 MHz)



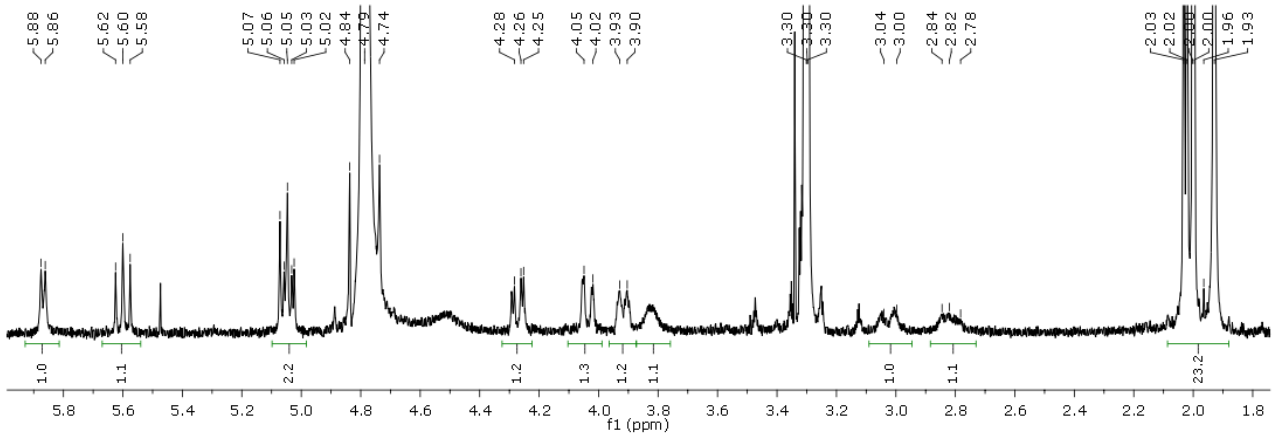
^{13}C-NMR spectrum of **85** ($CDCl_3$, 100 MHz)

N^{γ}-(2,3,4,6-Tetra-*O*-acetyl-α-D-glucopyranosyl)-L-asparagine (76)

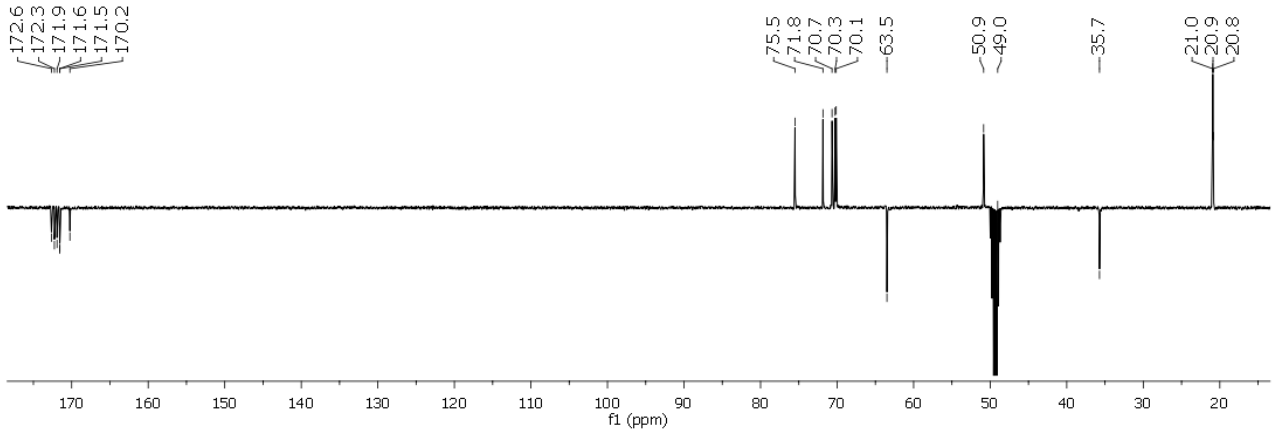


Compound **39** (300 mg, 0.437 mmol, 1eq) was dissolved in a mixture of MeOH:H_2O:AcOH (25:5:3, 33 mL) and 10% Pd/C was added. The reaction mixture was stirred under H_2 for 2 h and then was filtered through a pad of celite and washed with methanol. The solvent was evaporated to afford **76** (228 mg) in quantitative yield. $[\alpha]^{20}_{D}$ = +62.3 (*c* = 0.5, MeOH). ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 5.90 (d, $J_{1,2}$ = 6.0 Hz, 1H, 1-H), 5.62 (t, $J_{3,2}$ = 9.8 Hz, 1H, 3-H), 5.10–5.04 (m, 2H, 2-H, 4-H), 4.29 (dd, J_{gem} = 12.4, $J_{6,5}$ = 3.6 Hz, 1H, 6-H), 4.05 (dd, J_{gem} = 12.4, $J_{6'\text{-}5}$ = 1.6 Hz, 1H, 6'-H), 3.97–3.91 (m, 1H, 5-H), 3.86–3.83 (bs, 1H, α-H-Asn), 3.10–3.01 (m, 1H, β-CH_2-Asn),

2.89–2.80 (m, 1H, β-CH_2-Asn), 2.05(s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO), 2.02 (s, 3H, CH_3CO), 1.99 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 172.6–170.2 (CO), 75.5 (C-1), 71.9 (C-3), 70.7 (C-2), 70.3 (C-4), 70.1 (C-5), 63.5 (C-6), 52.3 (α-C-Asn), 36.7 (β-CH_2-Asn), 21.0, 20.9, 20.8 (4 x OAc) ppm. MS (ESI): m/z = 463.4 $[M + H]^+$, 485.6 $[M + Na]^+$.

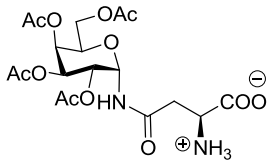


^{1}H-NMR spectrum of **76** (CD_3OD, 400 MHz)



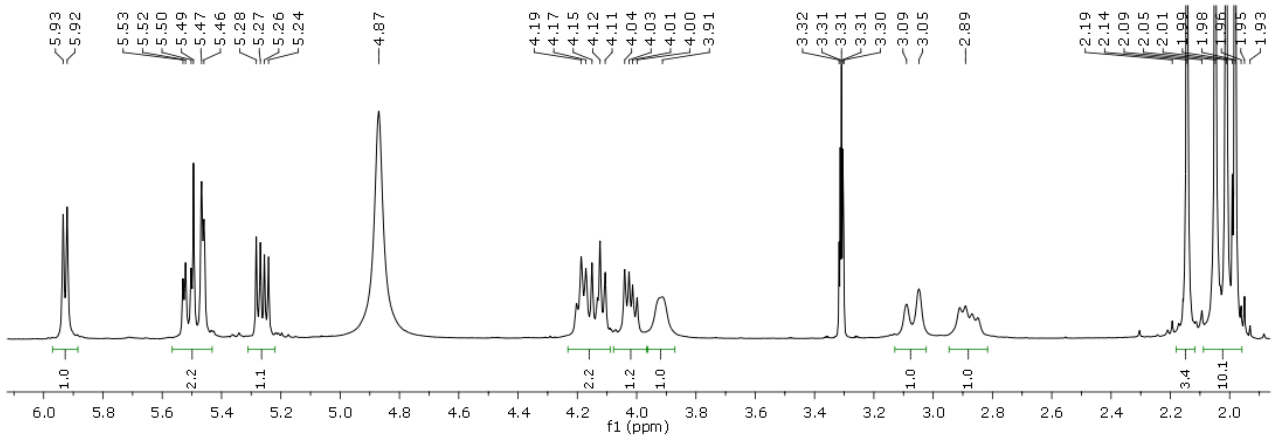
^{13}C-NMR spectrum of **76** (CD_3OD, 100 MHz)

N^{γ}-(2,3,4,6-Tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagine (77)

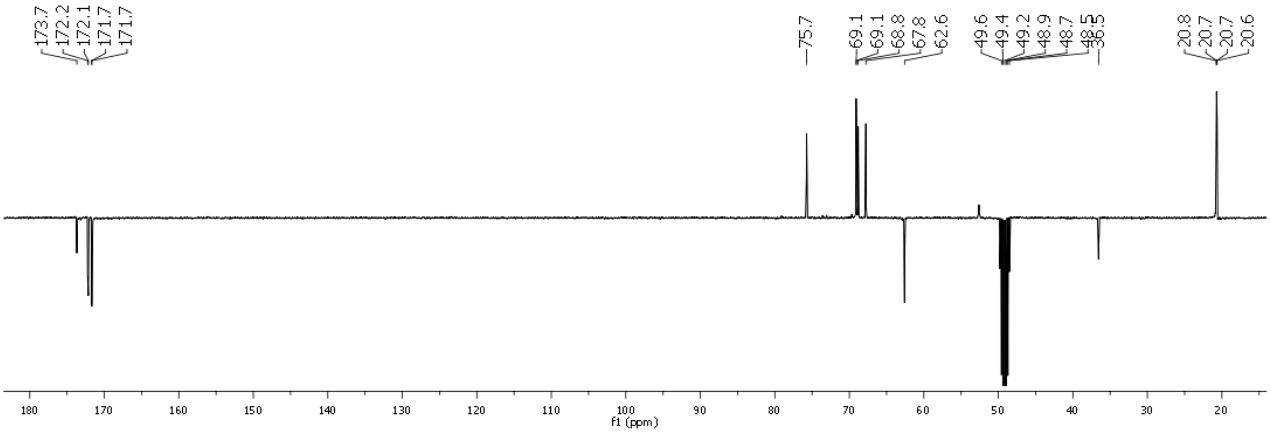


Compound **68** (250 mg, 0.364 mmol) was dissolved in 15 mL of a mixture of MeOH/H2O/AcOH (25:5:3) and 10% Pd/C was added. The reaction mixture was stirred under hydrogen for 2 h, then filtered through a celite pad and washed with methanol. The solvent was evaporated to afford **77** (190 mg) in quantitative yield. $[\alpha]^{20}_D$ = +91.7 (c = 2.25, MeOH). ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 5.93 (d, $J_{1,2}$ = 5.2 Hz, 1H, 1-H), 5.53–5.45 (m, 2H, 3-H, 4-H), 5.26 (dd, $J_{2,1}$ = 5.2, $J_{2,3}$ =

10.8 Hz, 1H, 2-H), 4.21–4.05 (m, 2H, 5-H, 6-H), 4.04–4.00 (m, 1H, 6'-H), 3.92 (bs, 1 H, α-H-Asn), 3.12–3.03 (m, 1H, β-CH_2-Asn), 2.93–2.82 (m, 1H, β-CH_2-Asn), 2.15 (s, 3H, CH_3CO), 2.05 (s, 3H, CH_3CO), 2.01 (s, 3H, CH_3CO), 1.97 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 173.8–171.7 (CO), 75.7 (C-1), 68.8 (C-3, C-4, C-5), 67.8 (C-2), 62.6 (C-6), 52.6 (α-C-Asn), 36.5 (β-CH_2-Asn), 20.8, 20.7, 20.6 (4 x OAc) ppm. MS (ESI): m/z = 463.0 $[M + H]^+$, 485.1 $[M + Na]^+$.

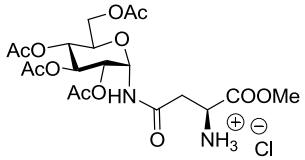


^{1}H-NMR spectrum of **77** (CD_3OD, 400 MHz)



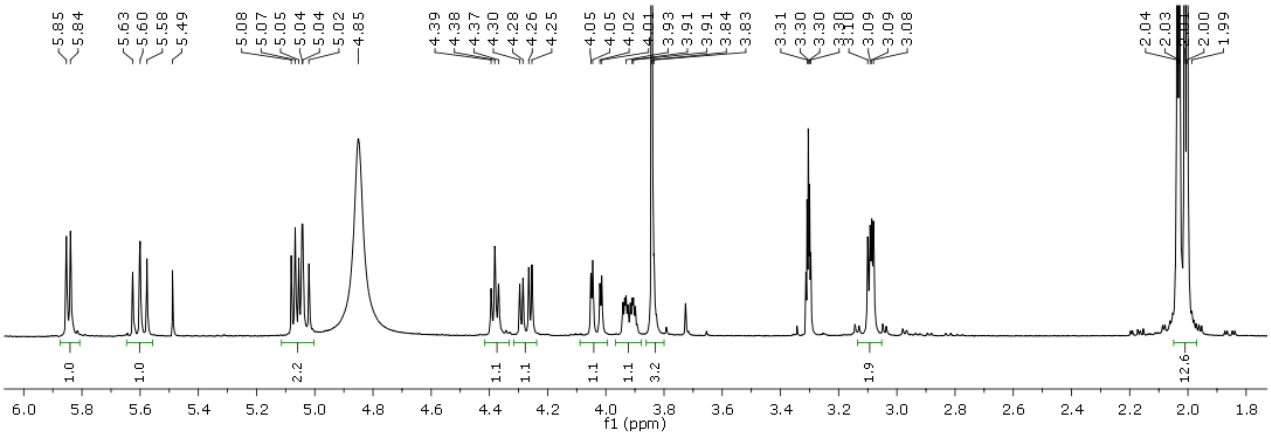
^{13}C-NMR spectrum of **77** (CD_3OD, 100 MHz)

N^{γ}-(2,3,4,6-Tetra-*O*-acetyl-α-D-glucopyranosyl)-L-asparagine MethylEster hydrochloride (81b)

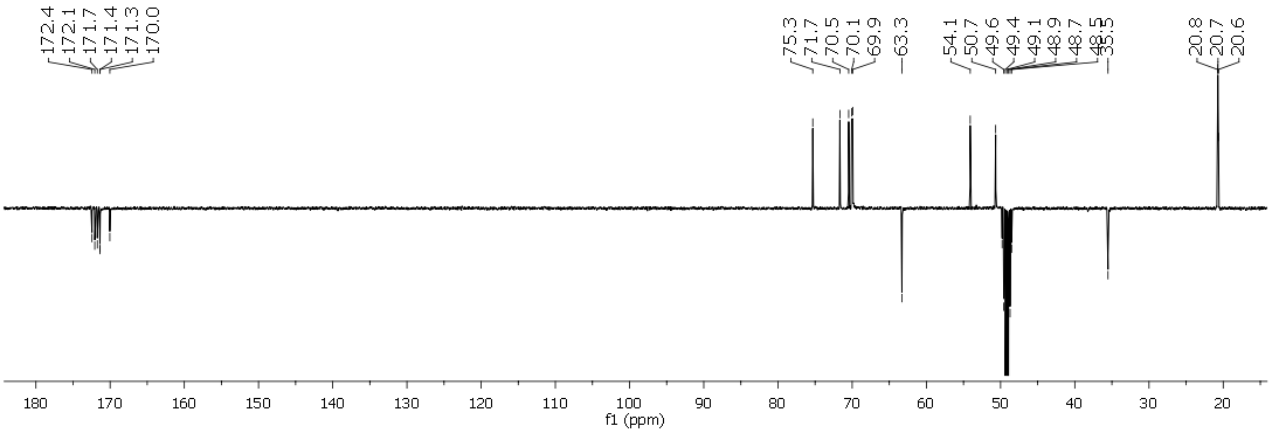


Compound **63** (35 mg, 0.057 mmol) and acetyl chloride (0.057 mmol, 4 µL) were dissolved in MeOH/H_2O (5:1, 6 mL). Pd/C (10%) was added and the mixture was stirred under H_2 for 2 h, filtered through a pad of celite and washed with methanol. The solvent was evaporated to afford **81b** (29 mg) in quantitative yield. $[\alpha]^{20}_D$ = +96.6 (*c* = 1.2, MeOH). ^{1}H NMR (400 MHz, CD_3OD, 25

°C): δ = 5.85 (d, $J_{1,2}$ = 5.6 Hz, 1H, 1-H), 5.60 (t, $J_{3,2}$ = $J_{3,4}$ = 9.6 Hz, 1H, 3-H), 5.09–5.02 (m, 2H, 2-H, 4-H), 4.37 (bs, 1H, α-H-Asn), 4.28 (dd, J_{gem} = 12.0, $J_{6,5}$ = 4.0 Hz, 1H, 6-H), 4.04 (dd, J_{gem} = 12.0, $J_{6'-5}$ = 1.8 Hz, 1H, 6'-H), 3.94–3.90 (m, 1H, 5-H), 3.85 (s, 3H, $COOCH_3$), 3.10–2.97 (m, 2H, β-CH_2-Asn), 2.12 (s, 3H, CH_3CO), 2.07 (s, 3H, CH_3CO), 2.03 (s, 3H, CH_3CO), 2.01 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 172.4–170.0 (CO), 75.3 (C-1), 71.7 (C-3), 70.5 (C-2), 70.1 (C-4), 69.9 (C-5), 63.3 (C-6), 54.1 ($COOCH_3$), 50.7 (α-C-Asn), 35.5 (β-CH_2-Asn), 20.8, 20.7, 20.6 (4 x OAc) ppm. MS (ESI): m/z = 499.3 $[M + Na]^+$.

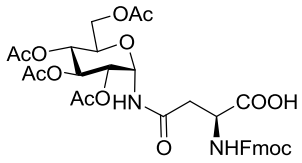


^{1}H-NMR spectrum of **81b** (CD_3OD, 400 MHz)

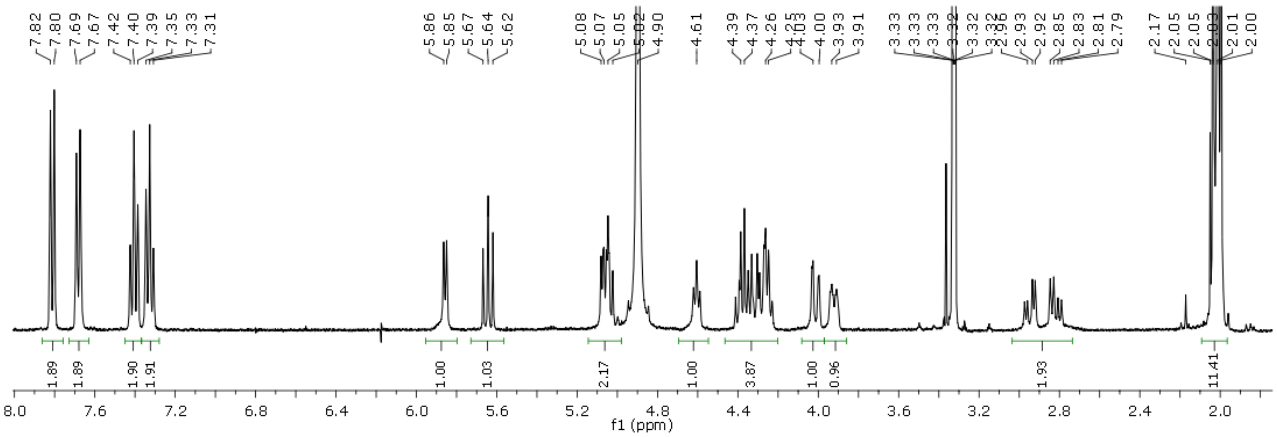


^{13}C-NMR spectrum of **81b** (CD_3OD, 100 MHz)

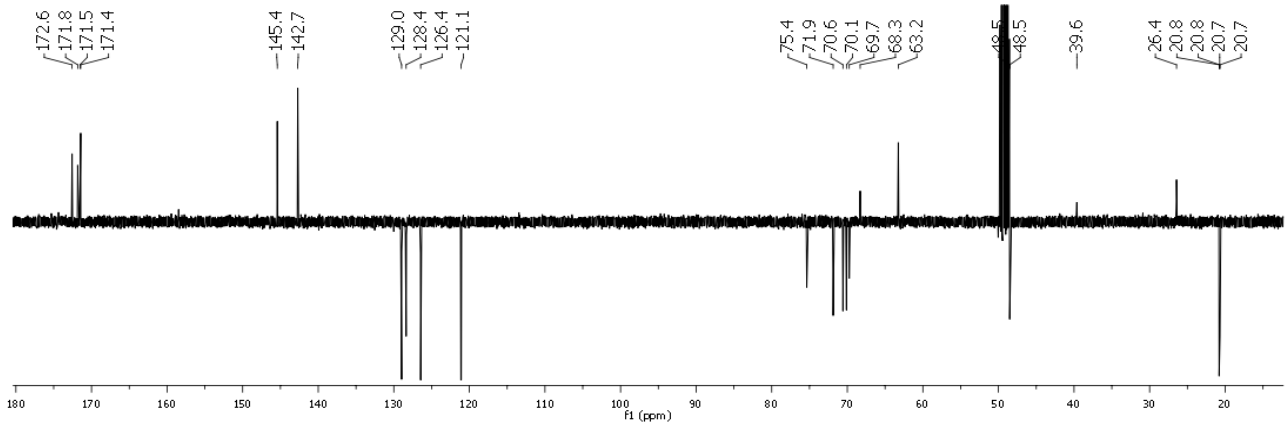
N^{α}-Fluorenylmethoxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl)-L-asparagine (79)



Compound **76** (82 mg, 0.157 mmol, 1 equiv.) and Fmoc-*O*-succinimide **78** (69 mg, 0.204 mmol, 1.3 equiv.) were dissolved in dry pyridine (1.5 mL) under nitrogen. The reaction mixture was stirred at room temperature overnight. After completion (TLC, 85:15 chloroform/methanol and 60:40 chloroform/methanol), the solvent was evaporated, the residue was dissolved in EtOAc and the organic phase was washed with 1M HCl and dried with Na_2SO_4. The solvent was evaporated under reduced pressure to yield 148 mg of the crude product, which was purified by flash chromatography (95:5 chloroform/methanol) to afford **79** (106 mg) in 99% yield. In the case of presence of β-derivative (deriving from DeShong reaction), crystallization was achieved, after chromatography purification, by dissolving the compound in 300 μL of CH_3CN. After warming till complete dissolution of compound, H_2O was added dropwise (ca 1.5 mL). Compound **79** precipitated as a white solid and it was filtered on a buchner and washed with cold methanol. $[\alpha]^{20}_{D}$ = +56.3 (*c* = 0.675, MeOH). ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 7.79 (d, *J* = 7.2 Hz, 2H, 4-HFmoc, 5-H-Fmoc), 7.66 (d, *J* = 7.2 Hz, 2H, 1-H-Fmoc, 8-HFmoc), 7.39 (t, *J* = 7.2 Hz, 2H, 3-H-Fmoc, 6-H-Fmoc), 7.31 (t, *J*= 7.2 Hz, 2-H-Fmoc, 7-H-Fmoc), 5.86 (d, $J_{1,2}$ = 5.6 Hz, 1H, 1-H), 5.64 (t, $J_{3,2} = J_{3,4}$ = 9.2 Hz, 1H, 3-H), 5.09–5.02 (m, 2H, 2-H, 4-H), 4.49 (bs, 1H, α-H-Asn), 4.39–4.18 (m, 4H, 9-H-Fmoc, CH_2-Fmoc, 6-H), 4.01 (dd, J_{gem} = 12.2, $J_{6'\text{-}5}$ = 1.8Hz, 1H, 6'-H), 3.97–3.92 (m, 1H, 5-H), 2.98–2.76 (m, 2H, β-CH_2-Asn), 2.16 (s, 3H, CH_3CO), 2.02 (s, 3H, CH_3CO), 2.00 (s, 3H, CH_3CO), 1.98 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 172.6–171.4 (CO), 145.4 (C_{quat}Fmoc), 142.7 (C_{quat}Fmoc), 129.0 (C-2-, C-7-Fmoc), 128.4 (C-3-, C-6-Fmoc), 126.4 (C-1-, C-8-Fmoc), 121.0 (C-4-, C-5-Fmoc), 75.4 (C-1), 71.9 (C-3), 70.6 (C-2), 70.1 (C-4), 69.7 (C-5), 68.3 (CH_2-Fmoc), 63.2 (C-6), 50.0 (α-C-Asn), 48.5 (C-9-Fmoc), 39.6 (β-CH_2-Asn), 20.8, 20.7, 20.6 (4 x OAc) ppm. FT-ICR MS (ESI): calcd. for $[C_{33}H_{35}O_{14}N_2]^-$ 683.20938; found 683.20968.

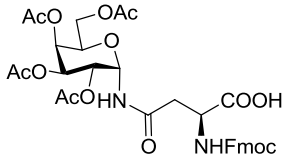


^{1}H-NMR spectrum of **79** (CD_3OD, 400 MHz)



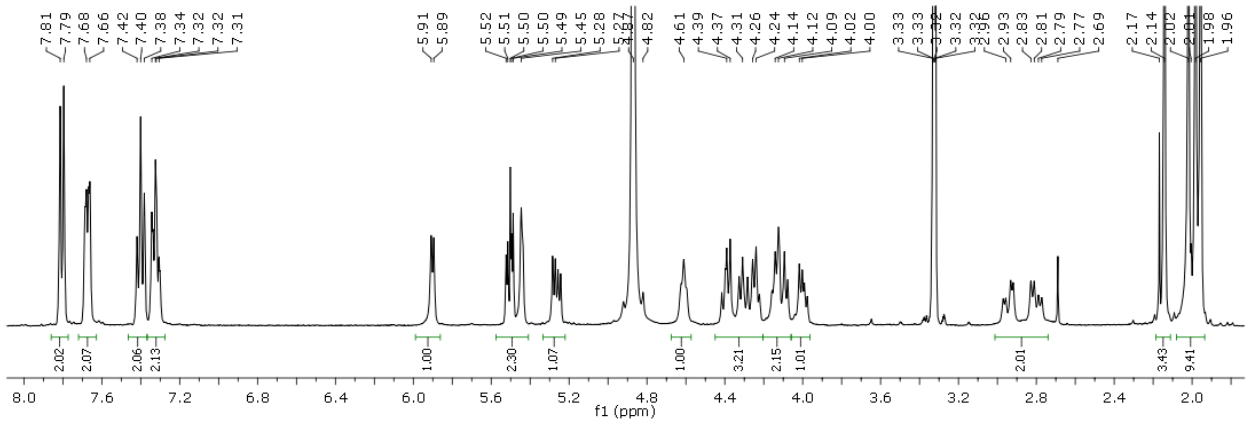
^{13}C-NMR spectrum of **79** (CD_3OD, 100 MHz)

N^{α}-Fluorenylmethoxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-*α*-D-galactopyranosyl)-L-asparagine (80)

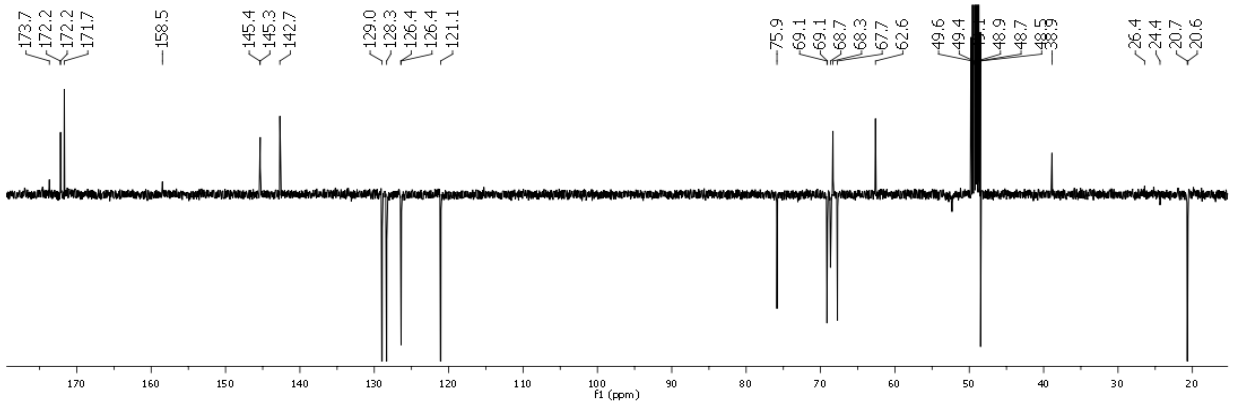


Compound **77** (190 mg, 0.364 mmol, 1 equiv.) and Fmoc-O-succinimide **78** (135 mg, 0.400 mmol, 1.1 equiv.) were dissolved in dry pyridine (1.8 mL) under nitrogen. The reaction mixture was stirred at room temperature overnight. After completion of the reaction (TLC, 8:2 chloroform/methanol and, 6:4 chloroform/methanol) the solvent was evaporated, the residue was dissolved in EtOAc and the organic phase was washed with 1M HCl and then dried with sodium sulfate. The solvent was evaporated under reduced pressure to yield 318 mg of the crude product, which was purified by flash chromatography (95:5 chloroform/methanol) to afford **80** (222 mg) in 90% yield. In the case of presence of β-derivative (deriving from DeShong reaction), crystallization of pure **80** was

achieved, by dissolving the compound in the minimum amount of CH_3CN (ca 600 μL), which, after warming, dissolved completely the compound. Then, H_2O was added dropwise (ca 6:1 H_2O:CH_3CN). Compound **80** precipitated as a white solid and it was filtered on a buchner and washed with cold methanol. $[\alpha]^{20}_D$ = +72.6 (*c* = 0.5, MeOH). ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 7.80 (d, *J* = 7.6 Hz, 2H, 4-H-, 5-H-Fmoc), 7.66 (d, *J* = 7.0 Hz, 2H, 1-H-, 8-H-Fmoc), 7.40 (t, *J* = 7.6 Hz, 2H, 3-H-, 6-H-Fmoc), 7.32 (t, *J*= 7.0 Hz, 2 H, 2-H-, 7-H-Fmoc), 5.90 (d, $J_{1,2}$ = 5.2 Hz, 1H, 1-H), 5.53–5.48 (m, 1H, 3-H), 5.50–5.43 (m, 1H, 4-H), 5.26 (dd, $J_{2,1}$ = 5.2, $J_{2,3}$ = 10.8 Hz, 1H, 2-H), 4.61 (bs, 1H, α-H-Asn), 4.42–4.21 (m, 3H, 9-H-Fmoc, CH_2-Fmoc), 4.17–4.07 (m, 2H, 5-H, 6-H), 4.03–3.96 (m, 1H, 6'-H), 2.94 (dd, J_{gem} = 15.4, $J_{\alpha,\beta}$ = 5.2 Hz, 1 H, β-H-Asn), 2.80 (dd, J_{gem} = 15.4, $J_{\alpha,\beta}$ = 6.8 Hz, 1H, β-H-Asn), 2.14 (s, 3H, CH_3CO), 2.02 (s, 3H, CH_3CO), 1.98 (s, 3H, CH_3CO), 1.95 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 173.6–171.7 (CO), 158.5 (NHCO), 145.4 (C_{quat}Fmoc), 145.3 (C_{quat}Fmoc), 142.7 (C_{quat}Fmoc), 129.0 (C2-, C7-Fmoc), 128.3 (C-3-, C-6-Fmoc), 126.4, 126.3 (C-1-, C-8-Fmoc), 121.1 (C-4-, C-5-Fmoc), 75.9 (C-1), 69.1 (C-3), 69.1 (C-4), 68.7 (C-5), 68.3 (CH_2-Fmoc), 67.7 (C-2), 62.6 (C-6), 52.3 (α-C-Asn), 48.5 (C-9-Fmoc), 38.9 (β-CH_2-Asn), 20.7, 20.6 (4 x OAc) ppm. FT-ICR MS (ESI): calcd. for $[C_{33}H_{35}O_{14}N_2]^-$ 683.20938; found 683.20959.



^{1}H-NMR spectrum of **80** (CD_3OD, 400 MHz)



^{13}C-NMR spectrum of **80** (CD_3OD, 100 MHz)

5.6 References

[1] Damkaci, F.; DeShong, P. *J. Am. Chem. Soc.* **2003**, *125*, 4408-4409.

[2] a) Paulsen H. *Adv. Carbohydr. Chem.* **1971**, *26*, 127. b) Györgydeák, Z.; Szitagyi, L.; Paulsen H. *J. Carb. Chem.* **1993**, *12*, 139-163.

[3] Bianchi, A.; Bernardi, A. *J. Org. Chem.* **2006**, *71*, 4565–4577.

[4] a) Mukaiyama, T.; Matsueda, R.; Suzuki, M. *Tetrahedron Lett.* **1970**, *60*, 1554-1564. b) Timpano, G.; Tabarani, G.; Anderluh, M.I.; Invernizzi, D.; Vasile, F.; Potenza, D.; Nieto, P.M.; Rojo, J.; Fieschi, F.; Bernardi, A. *ChemBioChem* **2008**, *9*, 1921-1930.

[5] Nuti, F.; Paolini, I.; Cardona, F.; Chelli, M.; Lolli, F.; Brandi, A.; Goti, A.; Rovero, P.; Papini, A. M. *Bioorg. Med. Chem.* **2007**, *15*, 3965–3973.

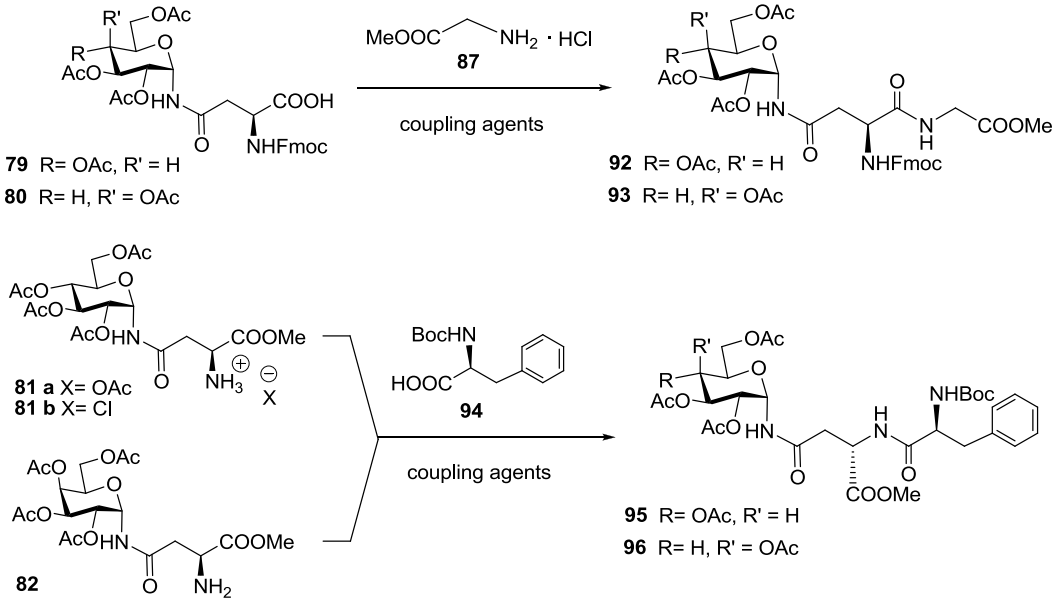
[6] Ibatullin, F. M.; Selivanov S. I. *Tetrahedron Letters* **2009**, *50*, 6351–6354.

[7] Crich, D.; Sasaki, K.; Rahaman, M. Y.; Bowers A. A. *J. Org. Chem.* **2009**, *74*, 3886–3893.

Chapter 6
Coupling conditions and synthesis of α-*N*-linked glycopeptides in solution

6.1 Introduction

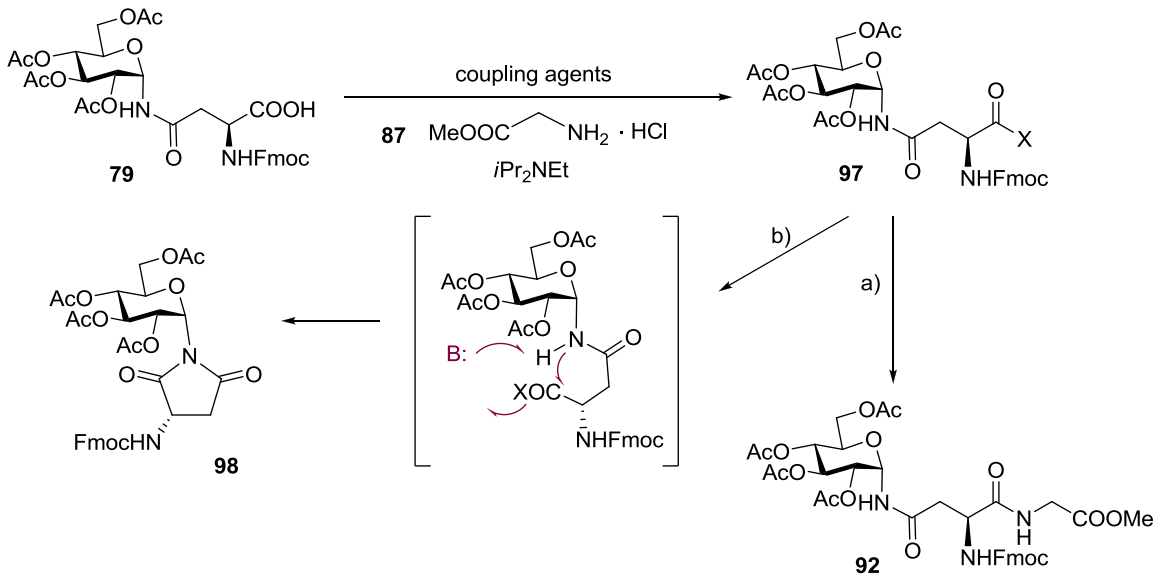
The (Fmoc)-protected α-*N*-linked glycosyl asparagine derivatives **79** and **80** and the α-*N*- glycosyl asparagine methyl esters **81** and **82** (**Scheme 1**) described in **Chapter 5** were employed in coupling reactions in solution, in order to indentify the best conditions to apply further in solid phase synthesis. In particular, compounds **79** and **80** were used to optimize the coupling at the C-terminus in the reaction with glycine-methylester **87**. Compounds **81** and **82** were used for the N-terminus elongation in the coupling reaction with Boc-phenyl-alanine **94** (**Scheme 1**). While N-terminus elongation could be easily achieved, C-terminus activation and elongation required extensive optimization[1]. This Chapter illustrates the results of this screening.



Scheme 1. Synthesis of α-*N*-linked glycosyl dipeptides through coupling reactions in solution.

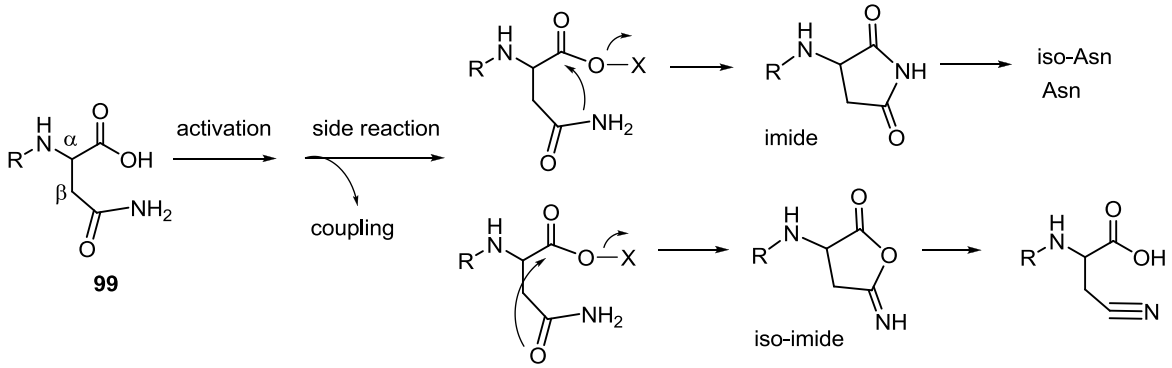
6.2 Elongation at the C-terminus

Activation of (N^{α}-Fmoc)-N^{γ}-glucosyl amino acid **79** with coupling agents under a variety of conditions, followed by addition of Gly-OMe **87**, resulted in unsatisfactory yields of dipeptide **92** (**Scheme 2**). Examination of the ^{1}H-NMR spectra of the crude reaction products showed formation of a by-product, which was isolated and characterized (NMR and MS) as the aspartimide **98** (**Scheme 2**). Apparently **79**, upon activation to form intermediate **97**, undergoes an easy intra-molecular cyclization which was the cause of the modest conversion to **92**



Scheme 2. Coupling conditions for **79**: a) peptide coupling; b) aspartimide formation from activated Asn derivative **97**.

It is widely known from the literature that, notwithstanding the general stability of amides, β- or γ-amides of asparagine or glutamine side chains could give side-reactions in some coupling conditions, especially in basic environment. In particular, during the activation of the α-carboxylic acid in asparagine residues **99** (**Scheme 3**) formation of imide or iso-imide has been observed. These compounds are generated by reaction of the activated carboxy group with the side chain amide. This could result in low coupling yields and in the production of difficult to separate by-products. To avoid this problem, common strategies, especially in solid phase synthesis, rely on the introduction of protecting group on the β-amide.[2]



Scheme 3. Side-reactions of asparagine residues.

To the best of our knowledge, this side-reaction was never described for β-*N*-glycosyl asparagine, which can be activated under standard peptide coupling conditions[3] and also transformed into active esters which can be isolated chromatographically.[4] Aspartimide formation from **79** was found to be favoured by basic conditions (see below) and may be a signal of a rather high acidity of the anomeric N-H group. This side process might also derived from steric factors and might represent a

case in which the α-*N*-glycopeptide behaves similarly to the non glycosylated analogous, rather than the natural β-*N*-linked counterpart (Section 1.3).[5] In order to suppress aspartimide formation and to optimize the coupling conditions, various reagents were screened (**Figure 1**). The results are collected in Table 1.

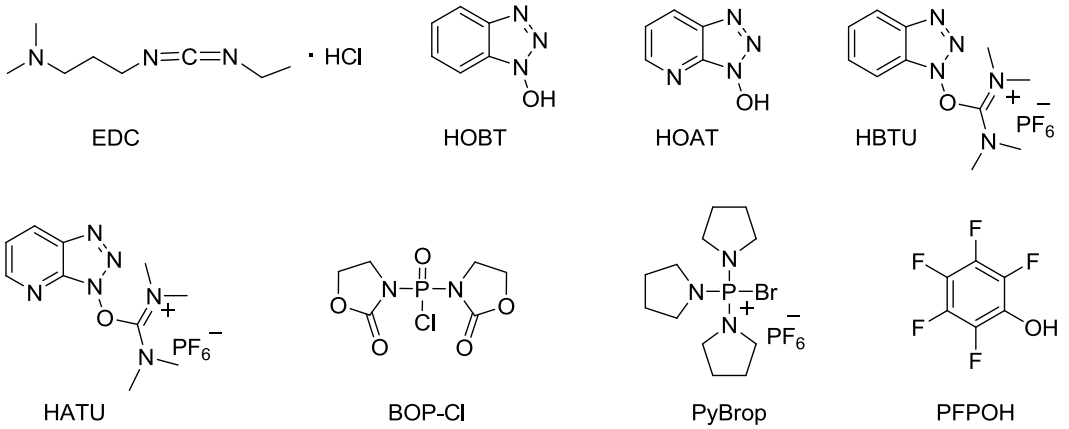


Figure 1. Condensing agents employed for the screening in Table 1.

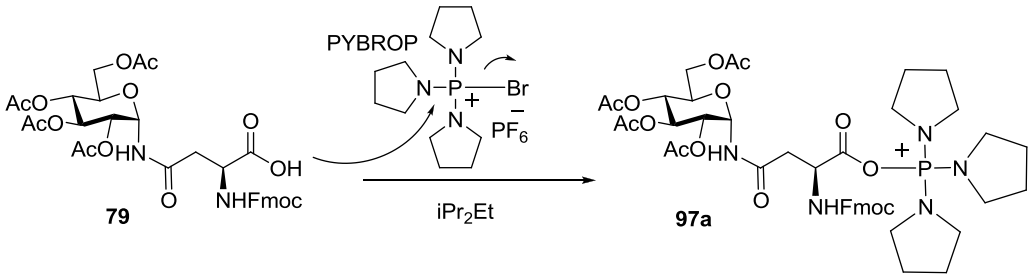
Table 1. C-elongation of **79**. Synthesis of α-*N*-glucosyl dipeptide **92**[a].

Entry	Reagents	Solvent	Reaction time	**92**(%)[b]	**98** (%)[c]
1	EDC	CH_2Cl_2	6 h	-[d]	-[d]
2	EDC , HOBT	CH_2Cl_2	6 h	45	30
3	HBTU, iPr_2NEt	DMF	6 h	40	44
4	EDC , HOAT	CH_2Cl_2:DMF 1:1	18h	-[d]	50
5	HATU, iPr_2NEt	DMF	4 h	-[d]	50
6	BOP-Cl	CH_2Cl_2	18 h	-[e]	-[e]
7	PyBrop, iPr_2NEt	CH_2Cl_2	5 h	69	-
8	DCC, PFPOH [f]	CH_2Cl_2	5 h	70	-

[a] Coupling of Fmoc-protected glucosylasparagine **79** (1eq) with HCl·Glycine-OMe **87** (3eq). iPr_2NEt was used to release the glycine salt and the EDC salt. [b] Isolated yield. [c] ^{1}H-NMR analysis of the crude [d] Not isolated. [e] No reaction occurred. [f] Followed by in situ addition of **87** [e]

EDC·HCl/*i*Pr_2NEt alone (**Table 1**, entry 1) was not sufficient to activate the coupling: after 6h of reaction the starting material **79** was still present. The product was not isolated. EDC used in combination with HOBT (**Table 1**, entry 2) afforded dipeptide **92** in only 45% yield after 6 h. Analysis of the crude (^{1}H-NMR), revealed the presence of 30% aspartimide **98**. Similarly, with HBTU (**Table 1**, entry 3) a 1:1 **92:98** ratio could be evaluated by ^{1}H-NMR of the crude coupling

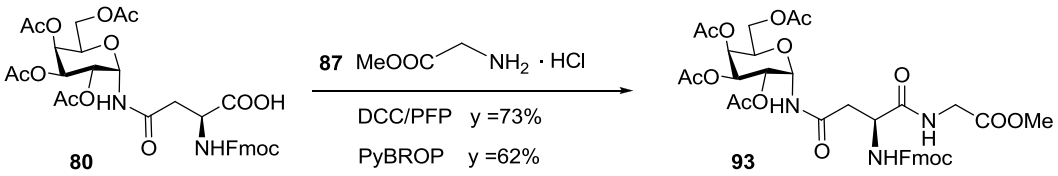
mixture, and only 40% of **92** was isolated after 6 h. Even larger amounts of cyclization product were obtained using EDC/HOAT and a consistent presence of fulvene (from deprotection of the Fmoc group) was observed after 18h (**Table 1**, entry 4). Reaction with HATU led to more than 50% of **98** (**Table 1**, entry 5). These results show that the aza-benzotriazole derivatives (HOAT and HATU) favour aspartimide formation relative to the benzotriazole ones (HOBt and HBTU). This may be related to the higher basicity of the azabenzotriazole condensing agents. Among phosphorus-based activating agents, bis(2-oxo-3-oxazolidinyl)phosphinic chloride (BOP-Cl) gave no reaction in 18 h (**Table 1**, entry 6). However, using bromo-tris-pyrrolidino phosphonium hexafluorophosphate (PyBrop), the desired dipeptide **92** was obtained in 69% yield after 5 h with no trace of the competing aspartimide by-product (**Table 1**, entry 7). This result may depend both on the low basicity of PyBrop and on steric factors. The bulky substituents on the phosphorous atom can prevent intramolecular cyclization of the intermediate **97a** (**Scheme 4**).



Scheme 4. Formation of the activated ester **97** with PyBrop.

Despite the pentafluorophenyl (PFP) ester of β-glucosyl-*N*-Fmoc-asparagine was reported as a stable glycosylated building block for peptide synthesis,[4] the corresponding α-anomer could not be isolated but cyclised spontaneously to aspartimide **98** during chromatography (in hexane: EtOAc 40:60, compound **97b** (x=PFP) (**Scheme 2**) with R_f = 0.58 converts into **98** with R_f = 0.37). However, formation of the PFP ester (with DCC, PFPOH, 1h, 0 °C) and in situ addition of methyl glycine **87** afforded the required dipeptide **92** in 70% yield after 5 h (**Table 1**, entry 8). Thus, both PyBrop and DCC/PFPOH can be used as coupling agents for this reaction with satisfactory yields.

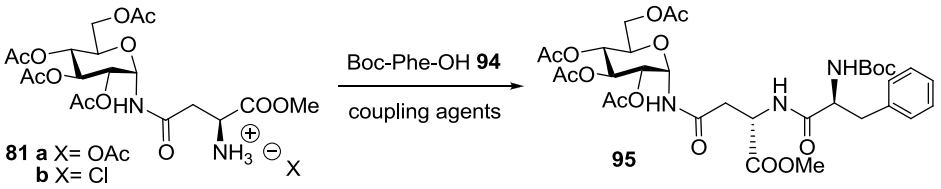
Once the best coupling conditions were identified for α-*N*-linked glucopyranosyl derivative **79**, the same protocols were applied to the α-galactosyl amino acid **80** (Scheme 5). Reaction of **80** with Glycine-OMe **87** using PFP and DCC afforded dipeptide **93** in 73% yield, while reaction with PyBrop gave **93** in 62% yield (**Scheme 5**).



Scheme 5. Coupling in solution of α-*N*-galactopyranosyl asparagine *N*-Fmoc protected **80**.

6.3 Elongation at the N-terminus

Peptide coupling conditions at the amine terminus were tested, starting from α-*N*- glucosyl asparagine methylester **81** and *N*-(*tert*-Butoxycarbonyl)-L-phenylalanine (Boc-Phe-OH) **94** (**Scheme 6**) and turned out to be less problematic. Results are summarized in **Table 2**. The reaction could be performed under standard coupling conditions, but yields were better using the chloride salt **81b** rather than the acetate **81a**.



Scheme 6. N-terminus elongation. Synthesis of **95**.

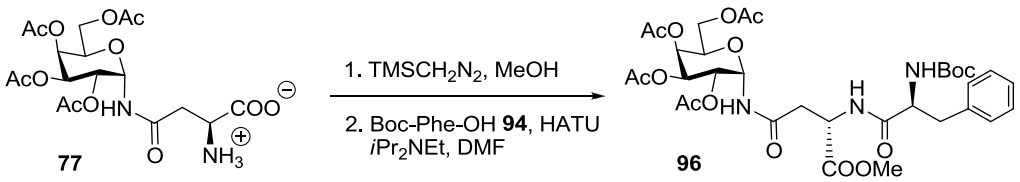
Table 2. C-elongation of **81**. Synthesis of α-N-glucosyl dipeptide **95**.[a]

Entry	Subtrate	Reagents	Solvent	Reaction time	**95**(%)[b]
1	**81a**	EDC · HCl,[c] HOBt	CH_2Cl_2	4h	56
2	**81a**	HATU, iPr_2NEt	DMF	5h	67
3	**81a**	PyBrop, iPr_2NEt	CH_2Cl_2	6h	66
4	**81b**	HATU, iPr_2NEt	DMF	12h	87

[a] Coupling of α-*N*- glucosyl asparagine methylester **81** (1eq) with Phe-Ala-OH (3eq) **94**. [c] iPr_2NEt was used to release the EDC salt. [b] Isolated yield.

From the acetate **81a**, **95** was obtained in 56 % yield, with EDC/HOBT (**Table 2**, entry 1). Reaction with HATU gave 67% yield either with 5h or with 12h of reaction time (entry 2). Coupling with PyBrop proceeded in 66% yield in 6h (entry 3). These yields could be improved starting from the hydrochloride salt **81b**, obtained by hydrogenation of **63** in MeOH to which a stoichiometric amount of AcCl had been added (procedure reported in Chapter 5). Coupling of **81b** with **94** with HATU in DMF afforded dipeptide **16** with 87% yield (entry 4). Most likely, activation of **81a** causes contemporary activation of acetic acid which in turn reacts with **81a**. This by-product wasn't observed starting from **81b** in the same coupling conditions.

For the galactosyl derivative, treatment of **77** with $TMSCH_2N_2$ in MeOH followed by one-pot coupling with Boc-Phenylalanine **94** and HATU afforded **96** in 72% overall yield (**Scheme 7**).



Scheme 7. Coupling in solution of α-*N*-galactosyl asparagine with Boc-Phenylalanine.

Spectroscopic (^{1}H- and ^{13}C-NMR) and chromatographic (HPLC-MS) analysis of dipeptides **92**, **93**, **95** and **96** showed a single isomer, confirming that no epimerization had occurred, neither on the peptide backbone nor at the anomeric carbon.

6.4 Synthesis in solution of model glycopeptides Ac-Asn-(α-*N*-Gal)-NHMe and Ac-Ala-Asn-(α-*N*-Gal)-Ala-NHMe

Having established the couplings conditions from C- and N-termini, we used them to synthesize glycopeptides **100** and **101** in solution (**Figure 2**). These molecules are models of a α-*N*-linked tripeptide and of a α-*N*-linked pentapeptide and were subsequently used in a collaboration with the group of Jiménez-Barbero, for conformational studies, which will be described in Chapter 9.

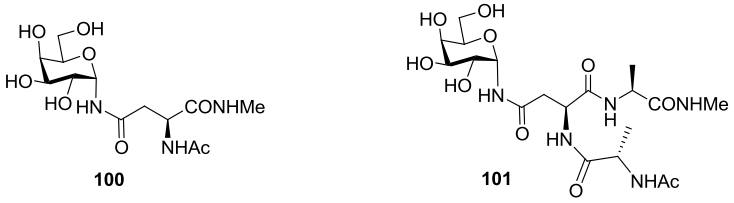
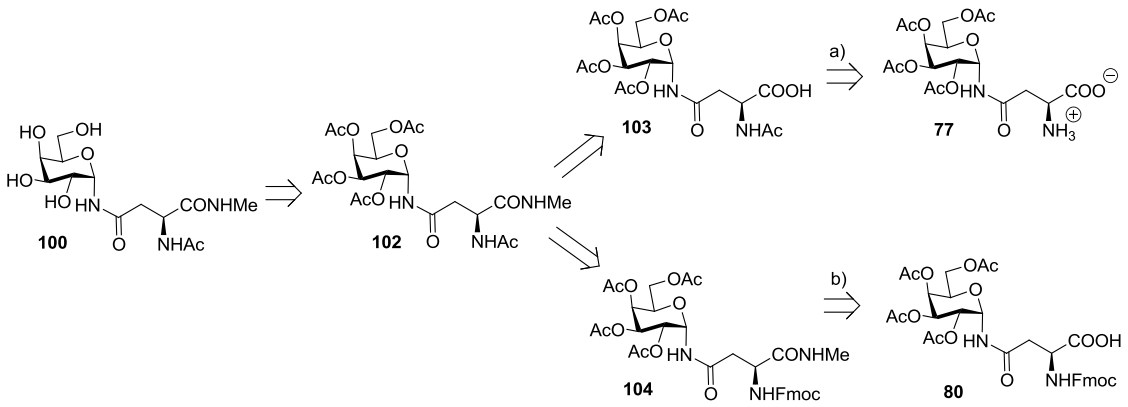


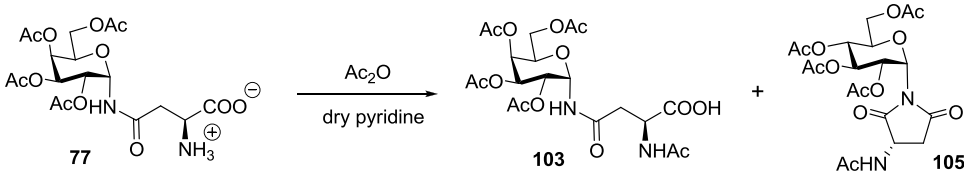
Figure 2. Two α-*N*-linked glycopeptide models described in this section.

The tripeptide model **100** can be synthesized by deprotection of **102**, which can be obtained starting either from **77** (**Scheme 8**, Path a) or **80** (**Scheme 8**, Path b). Path a, avoiding introduction and removal of Fmoc-protecting group clearly represents the faster method. However, it turned out to be unsuccessful, because direct acetylation of **77** afforded poor yields of **103.**



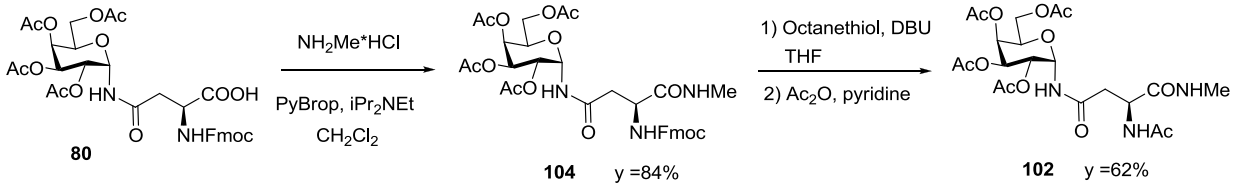
Scheme 8. Retrosynthetic analysis of the model glycopeptide Ac-Asn-(α-*N*-Gal)-NHMe **100**.

Acetylation of α-*N*-galactosyl asparagine **77** was performed initially with 3 equivalents of acetic anhydride for 2h at room temperature in pyridine (**Scheme 9**). After extraction, the crude, examined by ^{1}H NMR, revealed the presence of **103** (R_f = 0.45 in 90:10 $CHCl_3$:MeOH) and of imide **105** (R_f =0.60). Only 55% of **103** was isolated by chromatography. Performing the reaction at 0°C for 1h, then at room temperature for 1h, the cyclization process was enhanced, leading to total conversion of **77** into **105**.



Scheme 9. Acetylation of compound **77**. Formation of aspartimide **105**.

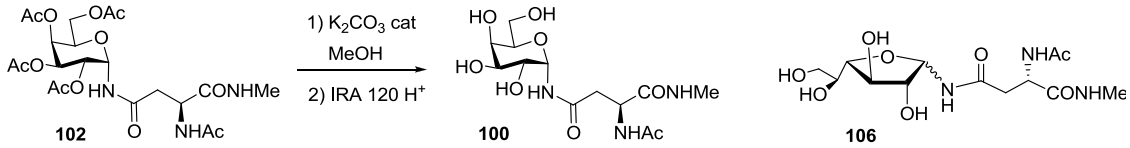
Since the introduction of the Fmoc protecting group proceeds in good yield (90%), we exploited pathway b (**Scheme 8**) for the synthesis of **100**, starting from **80**. *N*-methylamide **104** was obtained treating **80** with methylamine hydrochloride and PyBrop overnight in CH_2Cl_2 (**Scheme 10**). PyBrop, as precedently established, avoided aspartimide formation and afforded **104** in 84% yield. Fmoc removal was performed in solution, using octanethiol (10 eq) and a catalytic amount of DBU (0.5 eq).[6] The crude, washed several times with diethylether, was used without further purification for *N*-acetylation (Ac_2O in pyridine), which gave **102** in 62% yield over two steps.



Scheme 10. Synthesis of α-*N*-galactosyl asparagine derivative **102** .

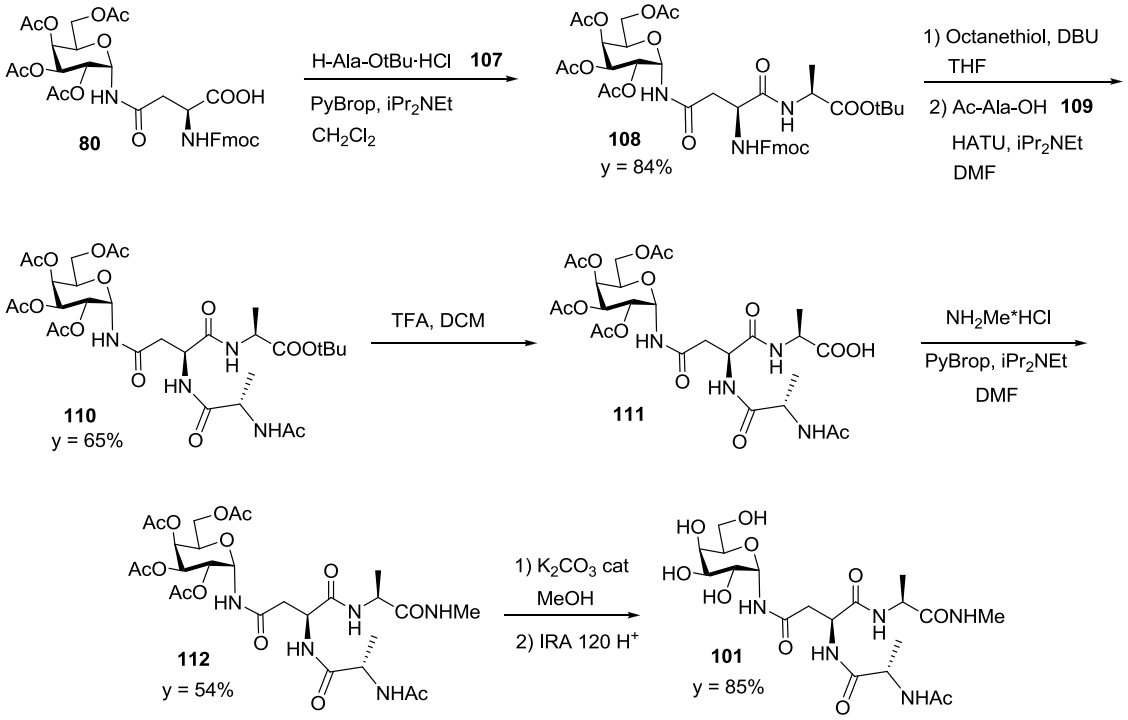
Removal of the *O*-acetyl groups from the carbohydrate moiety to afford **100** was successfully performed with catalytic amounts of K_2CO_3 (0.1 eq) in MeOH (**Scheme 11**), under carefully controlled pH conditions (pH =8.5).[7] The reaction was quenched with IRA 120 H^+ immediately after completion (2h, TLC: $CHCl_3$: MeOH 60:40, R_f **100** = 0.22) to avoid formation of the α-*N*-galacto furanosyl asparagine **106** (R_f =0.25). Formation of the furanose isomer could occur both in the basic condition of the reaction (long contact time with K_2CO_3) and in the acidic condition of the quench (Section 6.5 describes the possible mechanisms for furanosyl formation and solutions to this issue). The furanosyl compound, if formed, could be chromatographically removed by reverse phase purification (Biotage C-18 cartridge; eluant: MeOH/H_2O; gradient: MeOH 0% for 2 min, 0%→20% in 15 min). The mixture of **100** and **106** was practically not retained (t_r=1.2-1.8 min) but

was collected in three different fractions of 3 mL each. The furanosyl compound **106** was found in the last fraction (less polar compound).



Scheme 11. Final deprotection of **102**. Synthesis of **100**.

The model glycopeptide Ac-Ala-Asn-(α-*N*-Gal)-Ala-NHMe **101** was synthesized in a similar manner, starting from *N*-Fmoc-galactopyranosyl asparagine **80** through coupling with alanine tertbutylester hydrochloride **107** and PyBrop, which afforded **108** in 84% yield (**Scheme 12**). Fmoc deprotection (octanethiol and catalytic DBU) followed by reaction with *N*-acetylalanine **109** and HATU gave **110** in 65% yield over two steps.



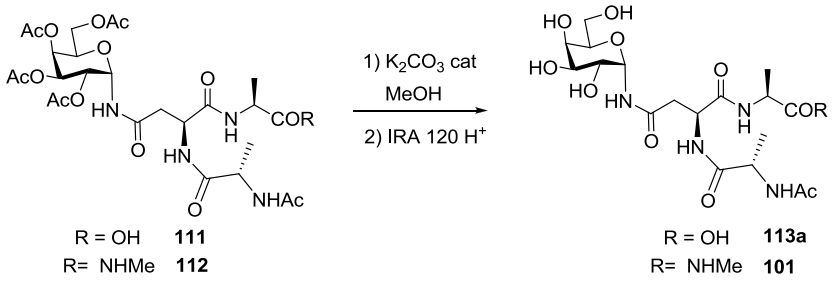
Scheme 12. Synthesis of Ac-Ala-Asn-(α-*N*-Gal)-Ala-NHMe **101**.

Removal of the *tert*-butyl ester from **110** with TFA and subsequent reaction of **111** with methylamine and PyBrop afforded **112** in 54% yield over two steps. Compound **112** showed a marked solubility in water, even if it is protected both on the sugar and on the peptide chain (this property was later observed also for other glycopeptides composed by alanine and *O*-acetyl α-*N*-Gal asparagine described in Chapter 7). This precluded aqueous work-up of the coupling mixture, which was purified, without previous work-up, by reverse phase chromatography (C-18 Biotage cartridge; eluant: CH_3CN/H_2O; gradient: CH_3CN 5→30 % t_r = 5 min). Deprotection on the carbohydrate moiety with catalytic amounts of K_2CO_3 in MeOH under carefully controlled pH conditions (pH

8.5) gave **101** in 85% yield. The optimization of deprotection conditions for this kind of substrates is described in the next section.

6.5 Deprotection conditions for *O*-acetyl groups removal

Common methods for deprotection of *O*-acetyl groups of carbohydrates rely on the use of catalytic amounts of sodium methoxide in methanol (Zemplen conditions) in a pH range = 8.5-9.5.[8] Removal of *O*-acetyl groups is also described using aqueous NaOH solutions in MeOH at pH 10.5 or 11.5 if the sugar portion also contains other groups as methyl esters.[9] Another route, which was recently applied to *O*-acetyl glucose derivatives, employs a catalytic amount of K_2CO_3 (0.1 eq) in MeOH.[10] Initially, we experimented the deprotection reaction on α-*N*-linked glucosyl dipeptide **95** and on α-*N*-linked glucosyl derivative **39** using Zemplen condition (catalytic MeONa in MeOH) with average pH values between 8.5 and 10. Complex mixtures of products were obtained, probably owing to anomerization. Hydrolysis with K_2CO_3 allows easier control of the solution pH. With this method, we initially obtained Ac-Asn-(α-*N*-Gal)-NHMe **100** using 0.1 eq of K_2CO_3 in MeOH (0.1M) under carefully controlled pH conditions (pH =8.5). The reaction was quenched with the sulfonic resin IRA 120 H^+. Performing again this reaction in the same conditions, however, we also observed formation of furanosyl asparagine **106** (**Scheme 11**). The reason for the formation of this by-product was later clarified with further investigations of this reaction on other substrates. In particular, we carried out a detailed investigations of this hydrolysis reaction on Ac-Ala-Asn-(α-*N*-tetra-*O*-acetyl-Gal)-Ala-OH **111** and on Ac-Ala-Asn-(α-*N*-tetra-*O*-acetyl-Gal)-Ala-NHMe **112**, with the aim of understanding the reason of by-product formation and to find a global solution to this problem (**Scheme 13**).



Scheme 13. Deprotection of *O*-acetyl groups of **111** and **112**.

The first trial was performed with **111**, using 0.6 equivalents of K_2CO_3 (1.1 eq, 0.5 eq. of which are necessary for salification of the carboxylic acid in the peptide chain) in MeOH (0.1 M). The reaction, quenched with IRA 120 H^+ after 2h, produced a mixture of products which was analyzed by LC-MS and revealed the presence of signals with the expected molecular mass value of 479 [M +1] and signals with value of 498 [M+18] (Figure 2). ^{1}H- and ^{13}C-NMR spectra of the crude

reaction mixture showed the presence of at least three different products (**Figure 3** shows some of their signals in different colours for the ^{1}H-spectrum).

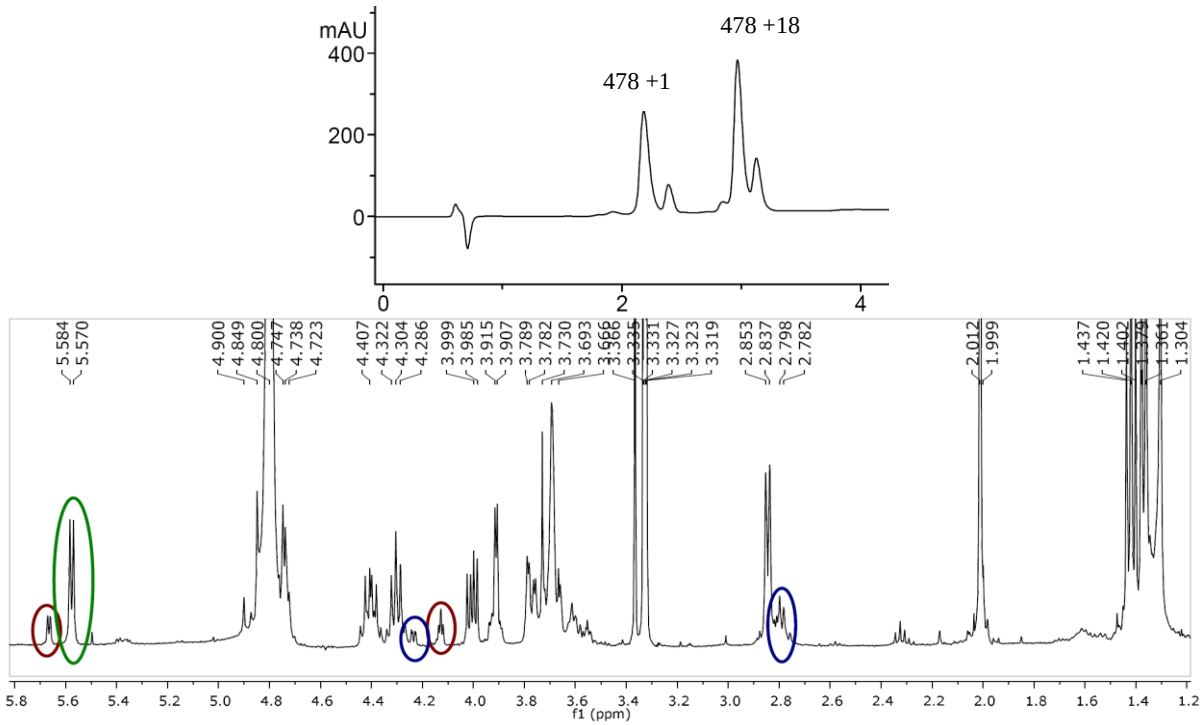
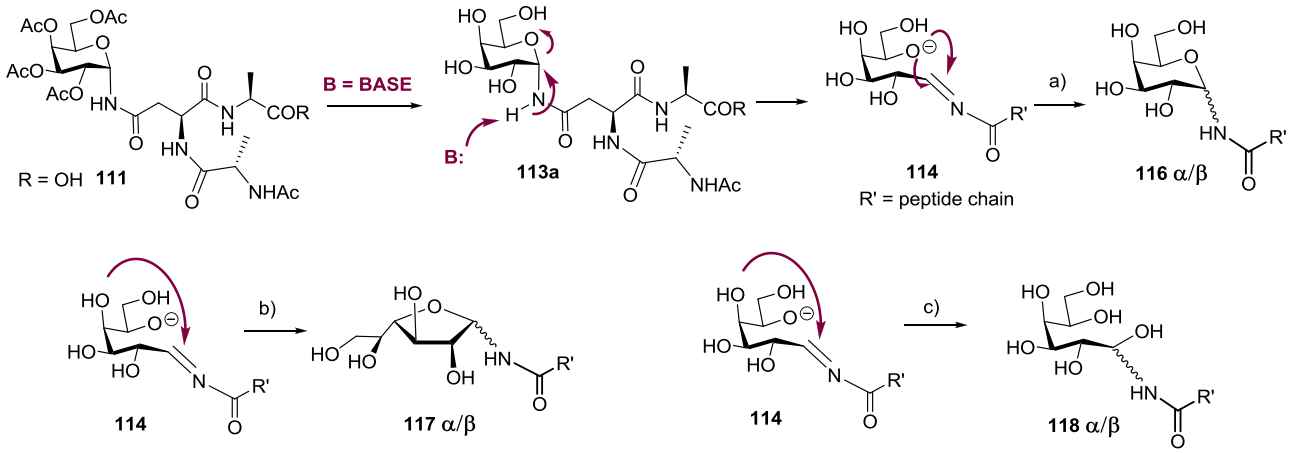


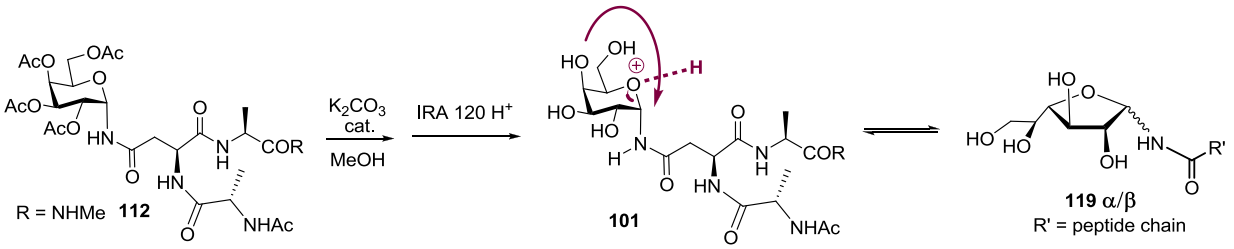
Figure 3. LC-MS and ^{1}H-NMR spectrum of the crude of deprotection of **111** (0.6 eq. of K_2CO_3 in MeOH 0.1M, 2h, quench with IRA 120 H^+).

Having already observed that α-*N*-linked glycosyl molecules have a tendency to undergo anomeric isomerization in basic conditions, we supposed that if the pH is not strictly controlled or for prolonged time contact with bases, the unnatural glycopeptide **113a** could be deprotonated and undergo ring opening to form intermediate **114** (**Scheme 14**). Subsequent ring closure can occur with anomerization of the galactose moiety, leading to compounds of general structure **116** (**Scheme 14**, path a). In addition, as the sugar is deacetylated, a ring contraction process could also take place, involving the 4-OH group and causing formation of furanosyl isomers **117** (**Scheme 14**, path b). The second set of compounds, corresponding to the LC-MS peaks with a [M+18] value, could derive from addition of water to intermediate **114** to generate **118** (**Scheme 14**, path c).



Scheme 14. Possible mechanisms for the formation of by-products in basic conditions during deprotection of acetals.

We reasoned that this side-process could be avoided using a controlled pH (8.5), more diluted solutions and quenching the reaction immediately after completion. Hence, the deprotection was performed with compound **112** using 0.1 equivalents of K_2CO_3 (pH = 8.5) in a 0.075 M solution of MeOH (**Scheme 15**).



Scheme 15. Possible mechanism for formation of furanosyl compounds **119 α/β** during acid quench in deprotection of **112** (0.1 eq. of K_2CO_3 in MeOH 0.075 M, 1h, quench with IRA 120 H^+).

After 1h, TLC monitoring revealed the presence of only one spot corresponding to product **101** (R_f = 0.37; $CHCl_3$:MeOH 60:40) and the reaction was quenched with IRA 120 H^+. However, the ^{1}H-spectrum and the LC-MS of the reaction mixture (**Figure 4**) showed also in this case the presence of two minor byproducts ([M+1]) that probably correspond to the α- and β- furanosyl compounds **119α** (red in **Figure 4**) and **119β** (blue in **Figure 4**). We consequently hypothesized that the deprotected derivative **101** was also unstable towards acidic conditions that could produce a ring contraction to furanosyl compounds with a mechanism as the one reported in **Scheme 15**. This point was later verified when chromatography of **113a** in the presence of 0.1% of TFA induced ring-contraction (see below).

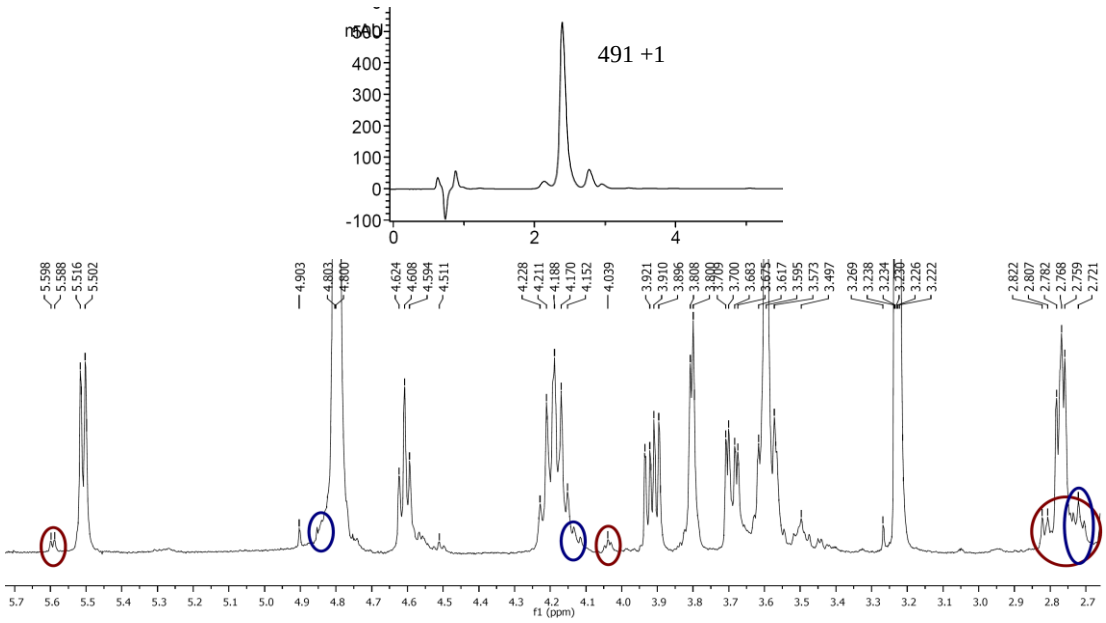
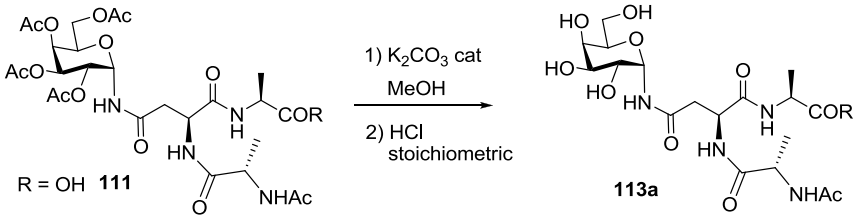


Figure 4. LC-MS and ^{1}H-NMR spectra of the crude of reaction of **Scheme 15**.

The amount of acid added to the solution using the sulfonic resin IRA 120 H^+ is hard to quantify in a precise manner, and we found that addition of stoichiometric amounts of HCl allowed a better pH control after completion of the deprotection step. In fact, reaction of **111** with 0.2 equivalent of K_2CO_3 (0.7 eq., but 0.5 eq. are required for salification of the terminal acid function of the peptide chain) in a 0.05 M solution in MeOH was quenched with a stoichiometric amount of HCl (1M in H_2O) and no formation of furanosyl compounds was observed (**Scheme 16**, **Figure 5**). These conditions were found to be effective even at a lower concentration (0.02M) of the substrate **111** in MeOH.



Scheme 16. Deprotection of **111** to afford only α-pyranosyl compound **113a** (0.2 eq. of K_2CO_3 in MeOH 0.05M or 0.027M, 2h, quench with stoichiometric HCl).

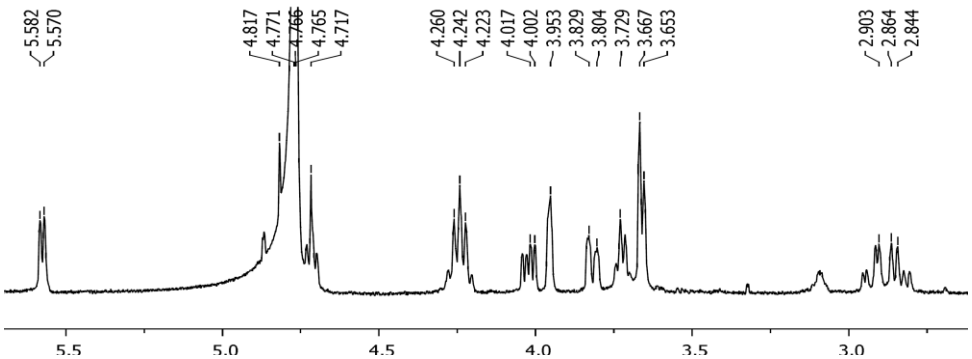


Figure 5. ^{1}H-NMR spectrum of reaction of **Scheme 16**.

The results obtained in the K_2CO_3 acetolysis are summarized in **Table 3**, which connects the formation or absence of the undesired furanosyl derivative with the reaction conditions used (equivalents of K_2CO_3, concentration of the substrate in MeOH, time reaction and type of quenching procedure).

Table 3. Condition for deprotection of *O*-acetyl groups of **111** and **112** with K_2CO_3.

Entry	Eq. K_2CO_3	M in MeOH	Time	Quench	Undesired furanosyl compounds
1	0.6	0.1	2h	IRA 120 H^+	20%
2	0.1	0.075	1h	IRA 120 H^+	10%
3	0.2	0.05	2h	HCl stoich.	-
4	0.2	0.02	3h	HCl stoich.	-

Confirmation of the instability of the synthesized glycopeptides towards acidic conditions was obtained during an attempt of purification of **113a** on silica C-18 (Biotage cartridge). In fact, the 0.1% of trifluoroacetic acid (TFA) added to the eluant (acetonitrile and water), which is commonly used for glycopeptides' and peptides' purification, was sufficient to trigger formation of the furanosyl derivatives **117 α/β** (**Figure 6**, red and blue respectively).

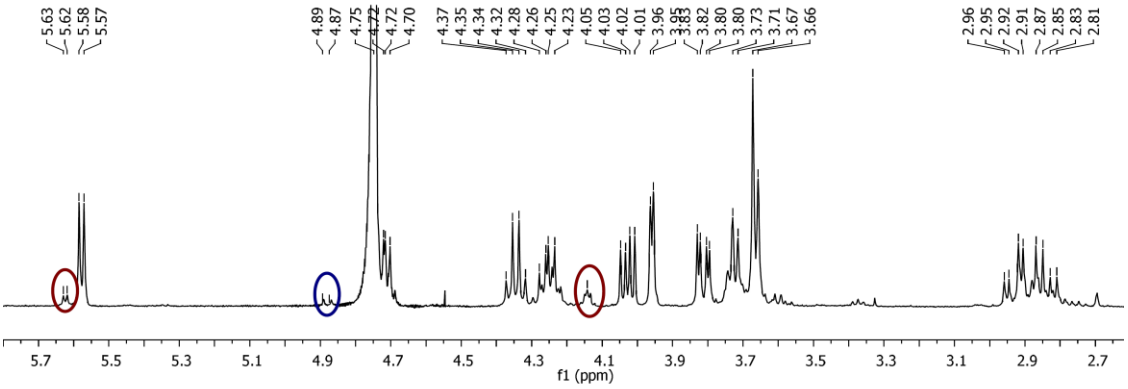
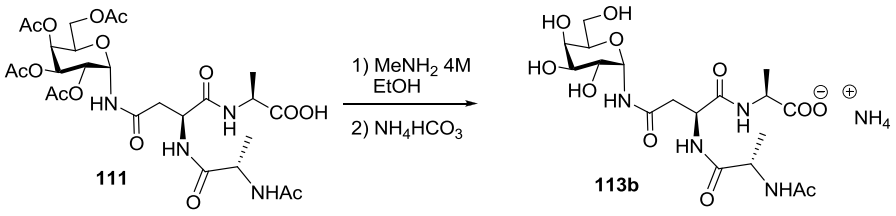


Figure 6. ^{1}H-NMR spectrum of reaction of **Scheme 16** after purification in reverse phase with 0.1% TFA.

An alternative route for both *N*- and *O*-acetyl deprotection was recently reported by Mikhailov et al. using commercially available 8M methylamine-ethanol solution for deacetylation of protected nucleosides.[11] Since **111** contains a *N*-acetyl group, we slightly modified the procedure, employing a 4M $MeNH_2$-EtOH solution with a substrate concentration of 0.05 M for 1h (**Scheme 17**).



Scheme 17. Deprotection of **111** with 4M $MeNH_2$-EtOH solution.

Excess $MeNH_2$ and most of the produced AcNHMe could be removed by evaporation, at the end of the reaction. Treatment with ammonium bicarbonate afforded the ammonium salt **113b**. In this case no trace of furanose formation, nor pyranose anomerization were observed and, as a consequence, this method was further applied for deprotection of all unnatural α-*N*-linked glycopeptides described in Chapter 7. The techniques used for the purification of all glycopeptides are described in Chapter 8.

6.6 Experimental Section

Solvents were dried by standard procedures: dichloromethane, methanol, *N*,*N*-diisopropylethylamine (DIPEA) were dried over calcium hydride; THF was distilled from sodium, N,N-dimethylformamide (DMF) was dried with activated molecular sieves (3Å). Reactions requiring anhydrous conditions were performed under nitrogen. ^{1}H- and ^{13}C-NMR spectra were recorded at 400 MHz on a Bruker AVANCE-400 instrument. Chemical shifts (δ) for ^{1}H and ^{13}C spectra are expressed in ppm relative to internal Me_4Si as standard. Signals were abbreviated as s, singlet; bs, broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet. Mass spectra were obtained with a Bruker ion-trap Esquire 3000 apparatus (ESI ionization) or Ft-ICR Mass spectrometer APEX II & Xmass 4.7 Magnet software (Bruker Daltonics). Thin layer chromatography (TLC) was carried out with pre-coated Merck F_{254} silica gel plates. Flash chromatography was carried out with Macherey-Nagel silica gel 60 (230-400 mesh) or with Biotage® SNAP KP-C18-HS cartridges for reverse phase chromatography. Purification of compound **113a** was performed by preparative HPLC (Phenomenex Jupiter C-18 (5 µm, 300 Å, column 50 x 30 mm), in collaboration with Dr Renato Longhi. HPLC-MS analyses were perfomed with Agilent 1100 with quaternary pump, diode array detector, autosampler, thermostated column holder coupled with MS: Bruker ion-trap Esquire 3000 with ESI ionization. The HPLC column was a Waters Atlantis 50x4.6 mm, 3 µm.

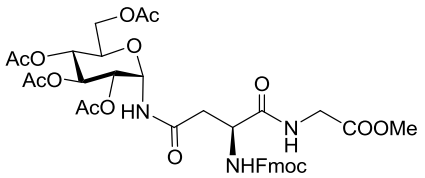
Method: Phase A: Milli-Q water containing 0.05 % (v/v) TFA.

Phase B: Acetonitrile (LC-MS grade) containing 0.05 % TFA.

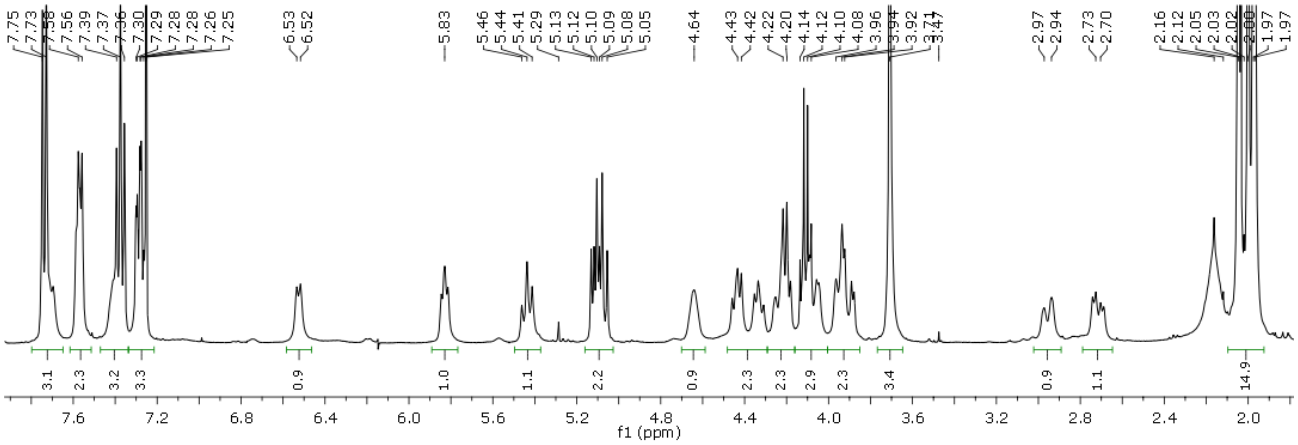
Flow: 1 mL/min, partitioned after UV detector (50 % to MS ESI), Temperature: 40°C.

Gradient: from 0 % B to 30 % B in 6 min, from 30 % B to 90 % B in 1 min, washing at 90 % B for 1 min, equilibration at 0 % B in the next 3 min.

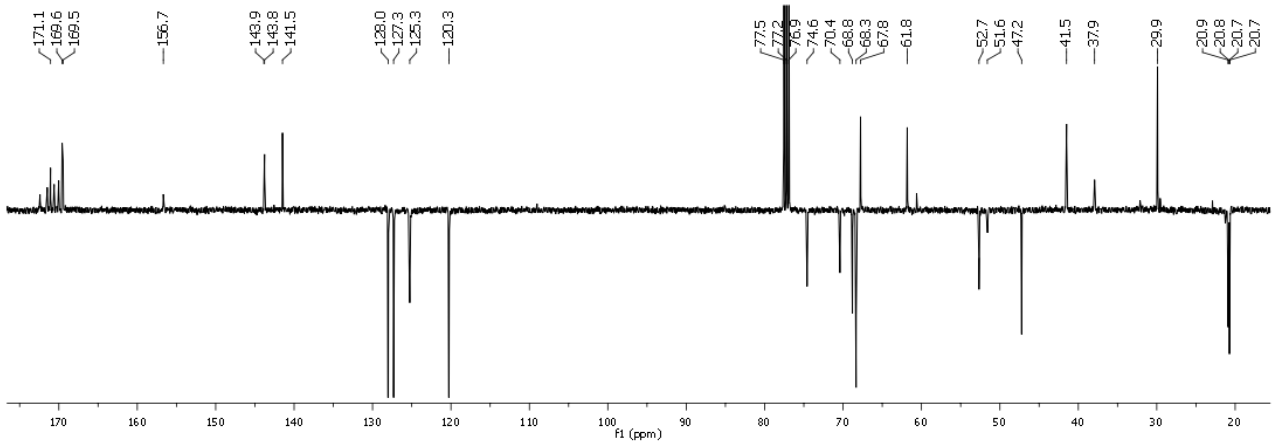
N^{α}-Fluorenylmethoxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl)- L asparagyl-glycine Methyl Ester (92)



Compound **79** (25 mg, 0.036 mmol, 1 equiv.) and PyBrop (37 mg, 0.080 mmol, 2.2 equiv.) were dissolved in dry CH_2Cl_2 (400 μL) under nitrogen at 0 °C. DIPEA (20 μL, 0.117 mmol, 3.2 equiv.) was added. After 10 min, glycine methyl ester hydrochloride **87** (14 mg, 0.109 mmol, 3 equiv.) and DIPEA (19 μL, 0.109 mmol, 3 equiv.) were added and the reaction mixture was stirred at 0 °C for 2 h and then at room temperature for 3 h. After completion (TLC, 85:15 $CHCl_3$: MeOH and 30:70 hexane/EtOAc) the solvent was evaporated, the residue was dissolved in EtOAc and washed with 1M HCl and saturated $NaHCO_3$. The organic phase was then dried with sodium sulfate and filtered. The solvent was evaporated under reduced pressure to yield 33 mg of the crude product, which was purified by flash chromatography (40:60 hexane/EtOAc) to afford **92** (19 mg) in 69% yield. $[\alpha]^{20}_D$=+66.9 (*c* = 0.9, CH_2Cl_2). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 7.75 (d, *J* = 7.6 Hz, 2H, 4-H-, 5-H-Fmoc), 7.72 (m, 1H, Glc-NH-Asn), 7.58 (d, *J* = 7.2 Hz, 2H, 1-H-, 8- H-Fmoc), 7.39 (t, *J* = 7.2 Hz, 2H, 3-H-, 6-H-Fmoc), 7.29 (t, *J* = 7.6 Hz, 2H, 2-H-, 7-H-Fmoc), 6.53 (d, *J* = 7.2 Hz, 1H, NHFmoc), 5.83 (dd, $J_{1,NH}$ = 7.8, $J_{1,2}$ = 5.6 Hz, 1H, 1-H), 5.46 (t, $J_{3,2}$ = $J_{3,4}$ = 9.8 Hz, 1H, 3-H), 5.15–5.05 (m, 2H, 2-H, 4-H), 4.64 (bs, 1H, α-H-Asn), 4.48–4.33 (m, 2H, CH_2-Fmoc), 4.30–4.18 (m, 2H, 9-H-Fmoc, 6-H), 4.15–4.03 (m, 2H, 6'-H, α-H-Gly), 4.00–3.87 (m, 2H, 5-H, α-H-Gly), 3.72 (s, 3 H, $COOCH_3$), 2.99–2.90 (m, 1H, β-H-Asn), 2.74–2.66 (m, 1H, β-H-Asn), 2.05 (s, 3H, CH_3CO), 2.02 (s, 3H, CH_3CO), 1.98 (s, 3H, CH_3CO), 1.96 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 171.5–169.4 (CO), 156.7 (COONHFmoc), 143.8, 143.7 (C_{quat}Fmoc), 141.4 (C_{quat}Fmoc), 128.0 (C-2-, C-7-Fmoc), 127.3 (C-3-, C-6-Fmoc), 125.2 (C-1-, C-8-Fmoc), 120.3 (C-4-, C-5-Fmoc), 74.6 (C-1), 70.4 (C-3), 68.8 (C-2), 68.3 (C-4, C-5), 67.7 (CH_2-Fmoc), 61.8 (C-6), 52.7 ($COOCH_3$), 51.6 (α-CAsn), 47.2 (C-9-Fmoc), 41.5 (CH_2-Gly), 37.9 (β-CH_2-Asn), 20.7, 20.6 (4xOAc) ppm . FT- ICR MS (ESI): calcd. for $[C_{36}H_{41}O_{15}N_3Na]^+$ 778.24299; found 778.24140.

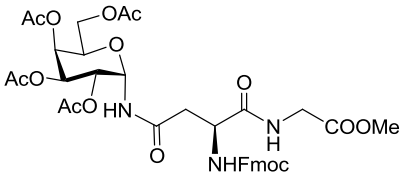


^{1}H-NMR spectrum of **92** ($CDCl_3$, 400 MHz)



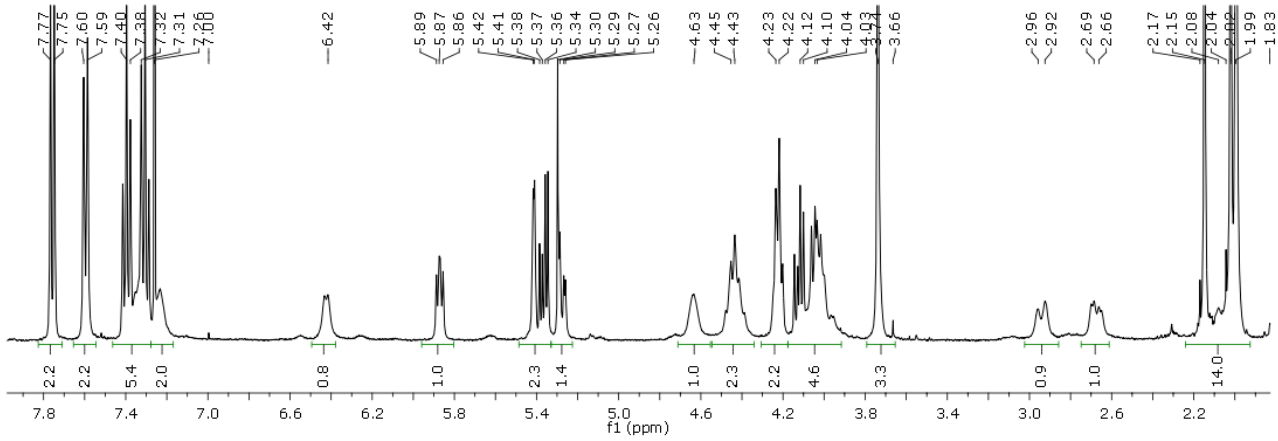
^{13}C-NMR spectrum of **92** ($CDCl_3$, 100 MHz)

N^{α}-Fluorenylmethoxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagyl glycine Methyl Ester (93)

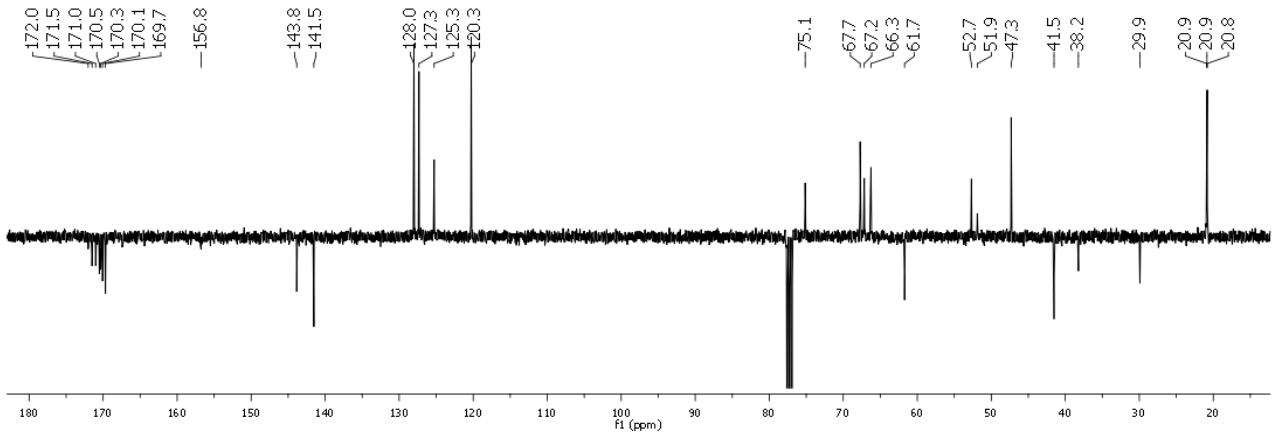


Compound **80** (35 mg, 0.051 mmol, 1 equiv.) and PyBrop (52 mg, 0.112 mmol, 2.2 equiv.) were dissolved in dry CH_2Cl_2 (400 μL) under nitrogen at 0 °C. DIPEA (20 μL, 0.112 mmol, 2.2 equiv.) was added. After 10 min, glycine methyl ester hydrochloride **87** (19 mg, 0.153 mmol, 3 equiv.) and DIPEA (27 μL, 0.153 mmol, 3 equiv.) were added and the reaction mixture was stirred at 0 °C for 2 h and then at room temperature for 5 h. After completion (TLC, 85:15 $CHCl_3$:MeOH and 30:70 hexane/EtOAc) the solvent was evaporated, the residue was dissolved in EtOAc and washed with 1M HCl and saturated $NaHCO_3$. The organic phase was then dried with sodium sulfate and the

solvent evaporated. The crude was purified by flash chromatography (40:60 hexane/EtOAc) to afford **93** (24 mg) in 62% yield. $[\alpha]^{20}_{D}$ = +55.1 (*c* = 0.5, MeOH). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 7.75 (d, *J* = 7.2 Hz, 2H, 4-H-, 5-H-Fmoc), 7.59 (d, *J* = 7.6 Hz, 2H, 1-H-, 8-H-Fmoc), 7.39 (t, *J* = 7.6 Hz, 2H, 3-H-, 6-H-Fmoc), 7.35 (bs, 1H, Gal-NH-Asn), 7.30 (t, *J* = 7.2 Hz, 2H, 2-H-, 7-H-Fmoc), 7.23 (bs, 1H, NH-Gly), 6.41 (d, *J* = 7.4 Hz, 1H, NH-Fmoc), 5.87 (dd, $J_{1,NH}$ = 7.4, $J_{1,2}$ = 5.2 Hz, 1H, 1-H), 5.41 (d, $J_{4,5} = J_{4,3}$ = 3.2 Hz, 1H, 4-H), 5.36 (dd, $J_{2,3}$ = 10.8 Hz, 1H, 2-H), 5.27 (m, 1H, 3-H), 4.63 (bs, 1H, α-H-Asn), 4.50–4.37 (m, 2H, CH_2-Fmoc), 4.26–4.18 (m, 2H, 9-H-Fmoc, 5-H), 4.11 (dd, J_{gem} = 11.4, $J_{6,5}$ = 6.6 Hz, 1H, 6-H), 4.08–3.97 (m, 3H, 6'-H, α-CH_2-Gly), 3.73 (s, 3H, $COOCH_3$), 3.00–2.89 (m, 1H, β-H-Asn), 2.70–2.65 (m, 1H, β-H-Asn), 2.15 (s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO), 2.02 (s, 3H, CH_3CO), 1.99 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 172.0–169.7 (CO), 143.8 (C_{quat}Fmoc), 141.5 (C_{quat}Fmoc), 128.0 (C-2-, C-7-Fmoc), 127.3 (C-3-, C-6-Fmoc), 125.3 (C-1-, C-8-Fmoc), 120.3 (C-4-, C-5-Fmoc), 75.1 (C-1), 67.7 (C-2, C-3), 67.2 (C-4), 67.2 (CH_2-Fmoc), 66.2 (C-5), 61.7 (C-6), 52.7 ($COOCH_3$), 51.9 (α-C-Asn), 47.3 (C-9-Fmoc), 41.5 (CH_2-Gly), 38.2 (β-CH_2-Asn), 20.9, 20.8 (4xOAc) ppm. FT-ICR MS (ESI): calcd. for $[C_{36}H_{41}O_{15}N_3Na]^+$ 778.24299; found 778.24179.

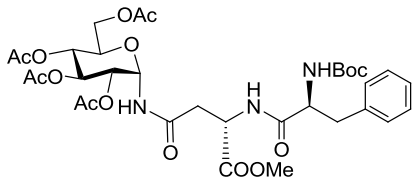


^{1}H-NMR spectrum of **93** ($CDCl_3$, 400 MHz)

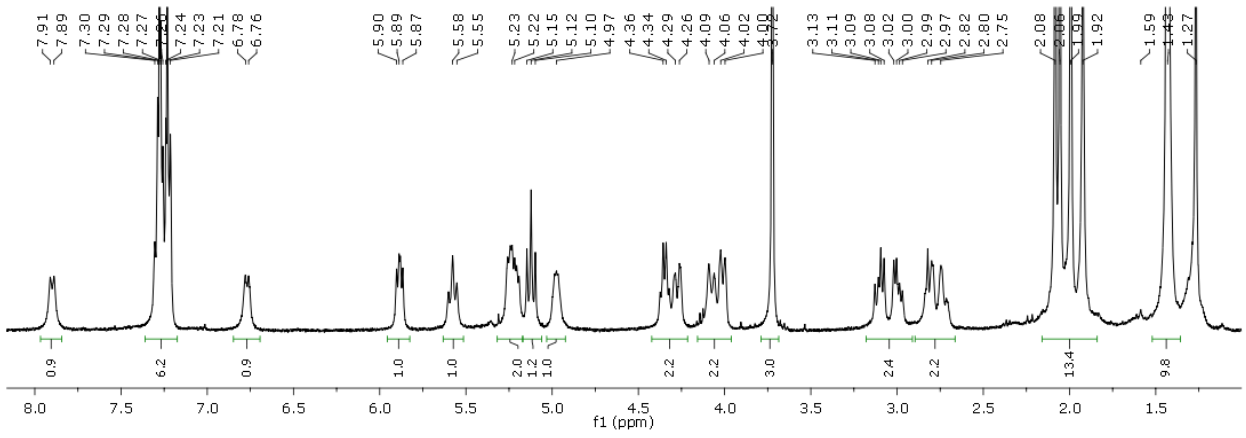


^{13}C-NMR spectrum of **93** ($CDCl_3$, 100 MHz)

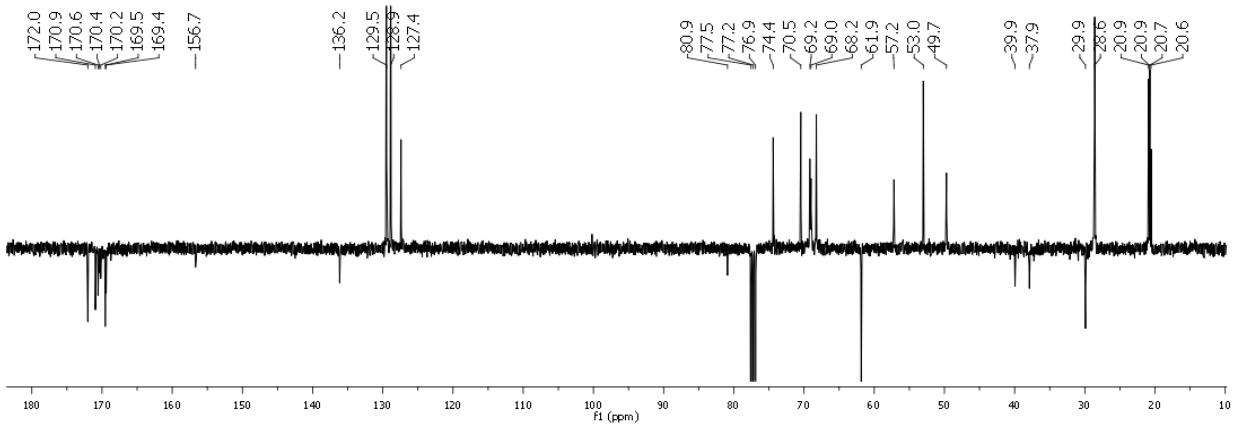
_N_-_tert_-Butoxycarbonyl-L-phenylalanyl-N^{γ}-(2,3,4,6-tetra-_O_-acetyl-α-D-glucopyranosyl)-L-asparagine Methyl Ester (95)



Compound **81b** (18 mg, 0.035 mmol, 1 equiv.) , Boc-Phe-OH **94** (33 mg, 0.123 mmol, 3.5 equiv.) and HATU (47 mg, 0.123 mmol, 3.5 equiv.) were dissolved in dry DMF (400 μL) under nitrogen at 0 °C. DIPEA (28 μL, 0.158 mmol, 4.5 equiv.) was added and the reaction mixture was stirred at 0 °C for 2 h and then at room temperature for 4 h (TLC, 85:15 chloroform/methanol and 30:70 hexane/EtOAc). The solvent was evaporated, the residue was dissolved in EtOAc and the organic phase was washed with 1M HCl and saturated $NaHCO_3$ and then dried with sodium sulfate. The solvent was evaporated and the crude was purified by flash chromatography (40:60 hexane/EtOAc) to afford **95** (23 mg) in 87% yield.$[\alpha]^{D}_{20}$ = +51.9 (*c* = 1.35, MeOH). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 7.90 (d, $J_{NH,1}$ = 8.6 Hz, 1H, Glc-NH-Asn), 7.31–7.22 (m, 5H, CHAr), 6.77 (d, J = 7.2 Hz, 1H, Asn-NH-Phe), 5.88 (dd, $J_{1,NH}$ = 8.6, $J_{1,2}$ = 5.6 Hz, 1H, 1-H), 5.58 (t, $J_{3,2}$ = $J_{3,4}$ = 9.6 Hz, 1H, 3-H), 5.28–5.17 (m, 2H, NHBoc, 2-H), 5.13 (t, $J_{4,3}$ = $J_{4,5}$ = 9.6 Hz, 1H, 4-H), 4.98 (bs, 1H, α-H-Asn), 4.39–4.25 (m, 2H, 6-H, α-H-Phe), 4.10–3.98 (m, 2H, 6'-H, 5-H), 3.73 (s, 3H, $COOCH_3$), 3.14–2.83 (m, 2H, β-H-Phe), 2.84–2.70 (m, 2H, β-H-Asn), 2.06 (s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO), 1.97 (s, 3H, CH_3CO), 1.90 (s, 3H, CH_3CO), 1.41 (s, 9H, COO*t*Bu) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 172.0–169.4 (CO), 156.7 (NHCO), 136.1 (CquatAr), 129.5, 128.9, 127.4 (CHAr), 80.9 (C_{quat}*t*Bu), 74.4 (C-1), 70.5 (C-3), 69.1, 69.0 (C-2, C-5), 68.2 (C-4), 61.8 (C-6), 57.2 (α-C-Phe), 53.0 ($COOCH_3$), 49.7 (α-C-Asn), 39.9 (β-CH_2-Asn), 37.9 (β-CH_2-Phe), 28.6 [$COO(CH_3)_3$], 20.9, 20.8, 20.7, 20.6 (4xOAc) ppm. FT-ICR MS (ESI): calcd. for $[C_{33}H_{45}O_{15}N_3Na]^+$ 746.27429; found 746.27414.

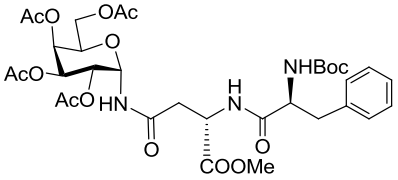


^{1}H-NMR spectrum of **95** ($CDCl_3$, 400 MHz)



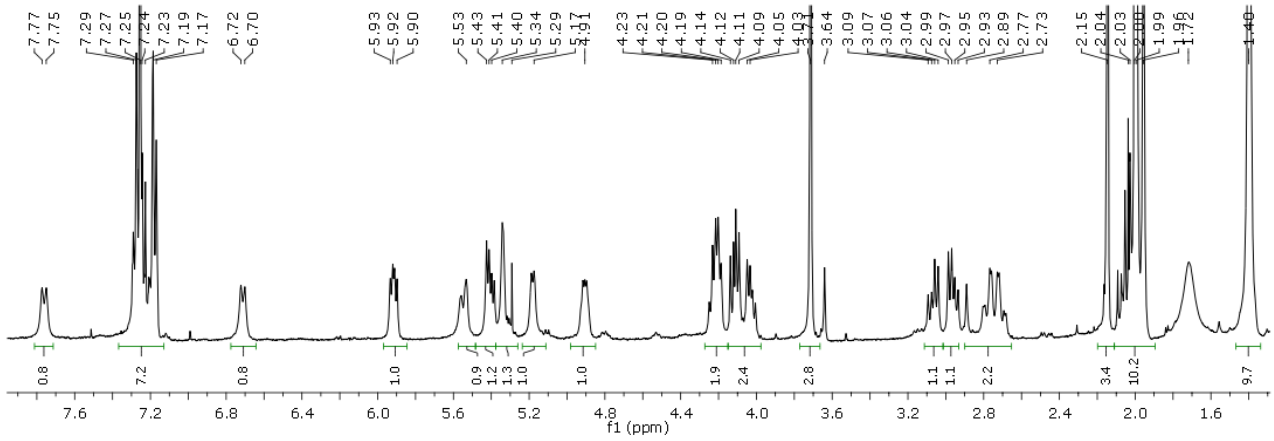
^{13}C-NMR spectrum of **95** ($CDCl_3$, 100 MHz)

***N*-*tert*-Butoxycarbonyl-L-phenylalanyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagine Methyl Ester (96)**

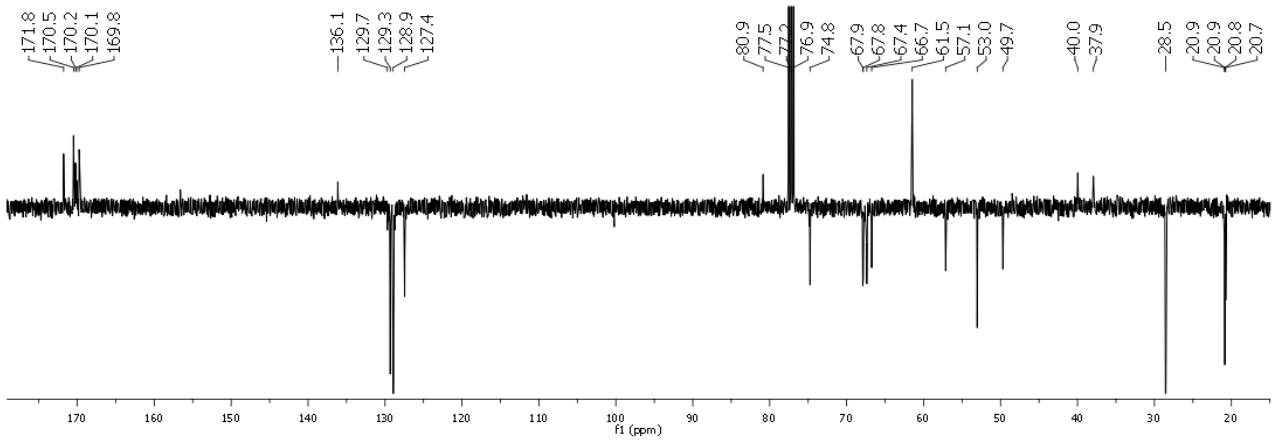


Compound **77** (19 mg, 0.036 mmol, 1 equiv.) was dissolved in dry methanol (400 μL) under nitrogen. A 2M (trimethylsilyl)diazomethane solution in diethyl ether (91 μL, 0.182 mmol, 5 equiv.) was added dropwise. After completion of the reaction (TLC, 6:4 chloroform/methanol and 7:3 chloroform/methanol) the solvent was evaporated. Boc-Phe-OH **94** (29 mg, 0.109 mmol, 3 equiv.) and PyBrop (51 mg, 0.109 mmol, 3 equiv.) were added to the crude and the mixture was dissolved in dry CH_2Cl_2 (350 μL). DIPEA (25 μL, 0.145 mmol, 4 equiv.) was added under nitrogen at 0 °C and the reaction mixture was stirred at 0 °C for 2 h and then at room temperature for 5 h (TLC, 85:15 chloroform/methanol and 30:70 hexane/EtOAc). The solvent was evaporated, the residue was dissolved in EtOAc and the organic phase was washed with 1M HCl and saturated $NaHCO_3$ and then dried with sodium sulfate. The solvent was evaporated and the crude purified by flash chromatography (40:60 hexane/EtOAc) to afford **96** (19 mg) in 72% yield. $[\alpha]^{20}_{D}$ = +51.7 (*c* = 0.6, MeOH). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 7.76 (d, $J_{NH,1}$ = 8.8 Hz, 1H, Gal-NH-Asn), 7.31–7.16 (m, 5H, CH_{Ar}), 6.71 (d, *J* = 8.0 Hz, 1H, Asn-NH-Phe), 5.92 (dd, $J_{1,NH}$ = 8.8, $J_{1,2}$ = 5.2 Hz, 1H, 1-H), 5.59–5.51 (m, 1H, 3 H), 5.41 (dd, $J_{2,1}$ = 5.2, $J_{2,3}$ = 11.0 Hz, 1H, 2-H), 5.35 (m, 1H, 4-H), 5.18 (d, *J* = 4.0 Hz, 1H, NHBoc), 4.91 (bs, 1H, α-H-Asn), 4.26–4.18 (m, 2H, 5-H, α H-Phe), 4.15–4.00 (m, 2H, 6-H), 3.72 (s, 3H, $COOCH_3$), 3.11–2.92 (m, 2H, β-CH_2-Phe), 2.84–2.67 (m, 2H, β-CH_2-Asn), 2.15 (s, 3H, CH_3CO), 2.01 (s, 3H, CH_3CO), 1.99 (s, 3H, CH_3CO), 1.96 (s, 3H, CH_3CO), 1.41 (s, 9H, COO*t*Bu) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 171.8–169.7 (CO), 136.1

(C_{quat}Ar), 129.3, 128.9, 127.4 (CH_{Ar}), 80.8 (C_{quat}*t*Bu), 74.8 (C-1), 67.9 (C-4), 67.8 (C-5), 67.4 (C-3), 66.7 (C-2), 61.5 (C-6), 57.1 (α-C-Phe), 53.0 ($COOCH_3$), 49.7 (α-C-Asn), 40.0 (β-CH_2-Asn), 37.9 (β-CH_2-Phe), 28.5 [$COO(CH_3)_3$], 20.9, 20.8, 20.7 (4xOAc) ppm. FT-ICR MS (ESI): calcd. for $[C_{33}H_{45}O_{15}N_3Na]^+$ 746.27429; found 746.27378.

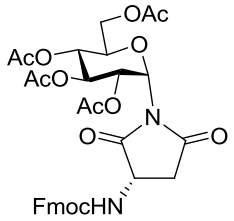


^{1}H-NMR spectrum of **96** ($CDCl_3$, 400 MHz)



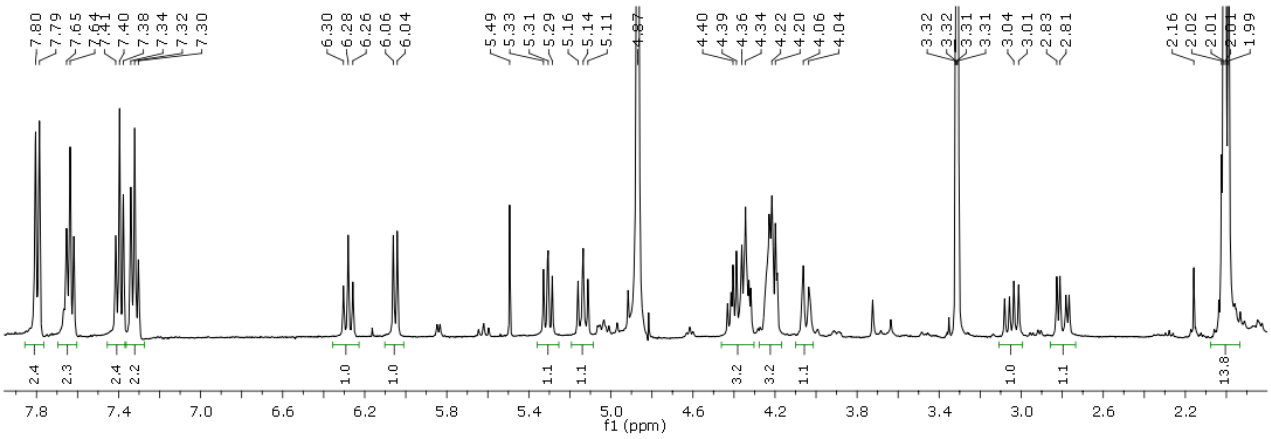
^{13}C-NMR spectrum of **96** ($CDCl_3$, 100 MHz)

(3*S*)-1-(2,3,4,6-Tetra-*O*-acetyl-α-D-glucopyranosyl)-3-(*N*-fluorenylmethoxycarbonyl)-2,5 dioxopyrrolidine (98)

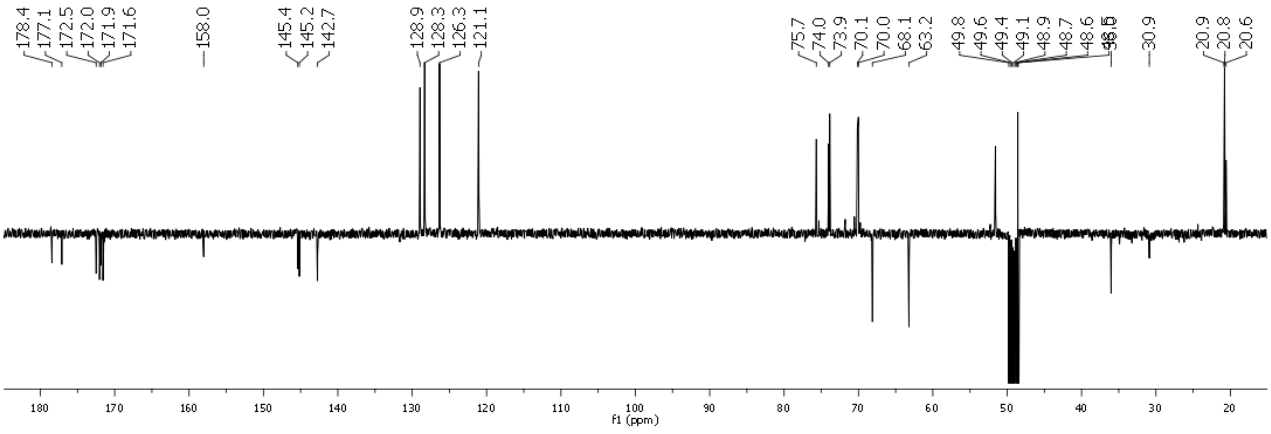


Compound **98** was isolated as a byproduct in some of the coupling reactions leading to **92**, as described in Table 1 (flash chromatography, 40:60 hexane/EtOAc, R_f = 0.37). $[\alpha]^{20}_D$ = +41.2 (*c* = 0.45, CH_2Cl_2). ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 7.79 (d, *J* = 7.2 Hz, 2H, 4-H-Fmoc, 5-H-

Fmoc), 7.66 (t, J = 7.2 Hz, 2H, 1-H-Fmoc, 8-H-Fmoc), 7.40 (t, J = 7.2 Hz, 2H, 3-H-Fmoc, 6-H-Fmoc), 7.32 (t, J = 7.2 Hz, 2H, 2-H-Fmoc, 7-H-Fmoc), 6.28 (t, J = 9.2 Hz, 1H, 3-H), 6.05 (d, $J_{1,2}$ = 7.8 Hz, 1H, 1-H), 5.29 (t, $J_{2,1}$ = 7.8 Hz, 1H, 2-H), 5.13 (t, J = 9.2 Hz, 1H, 4-H), 4.44–4.28 (m, 3H, α-H-Asn, 9-HFmoc, CH-Fmoc), 4.29–4.17 (m, 3H, 6-H, 5-H, CH-Fmoc), 4.08–4.02 (m, 1H, 6'-H), 3.18–3.00 (dd, J_{gem} = 15.6, $J_{\alpha,\beta}$ = 9.2 Hz, 1H, β-H-Asn), 2.84–2.73 (dd, J_{gem} = 15.6, $J_{\alpha,\beta}$ = 6.4 Hz, 1H, β-H-Asn), 2.02 (s, 3H, CH_3CO), 2.01 (s, 3H, CH_3CO), 2.00 (s, 3H, CH_3CO), 1.99 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 175.7–169.9 (CO), 155.7 (COONHFmoc) 143.7, 143.5 (C_{quat}Fmoc), 141.6, 141.5 (C_{quat}Fmoc), 128.1, 128.0 (C-2-, C-7-Fmoc), 127.4 (C-3-, C-6-Fmoc), 125.1 (C-1-, C-8-Fmoc), 120.3 (C-4-, C-5-Fmoc), 74.9 (C-1), 72.9 (C-5), 72.3 (C-3), 68.5 (C-4, C-5), 67.5 (CH2-Fmoc), 61.9 (C-6), 50.7 (α-C-Asn), 47.3 (C-9-Fmoc), 35.3 (β-CH_2-Asn), 21.9, 20.9, 20.7 (4xOAc) ppm. MS (ESI): m/z = 689.4 $[M + Na]^+$.

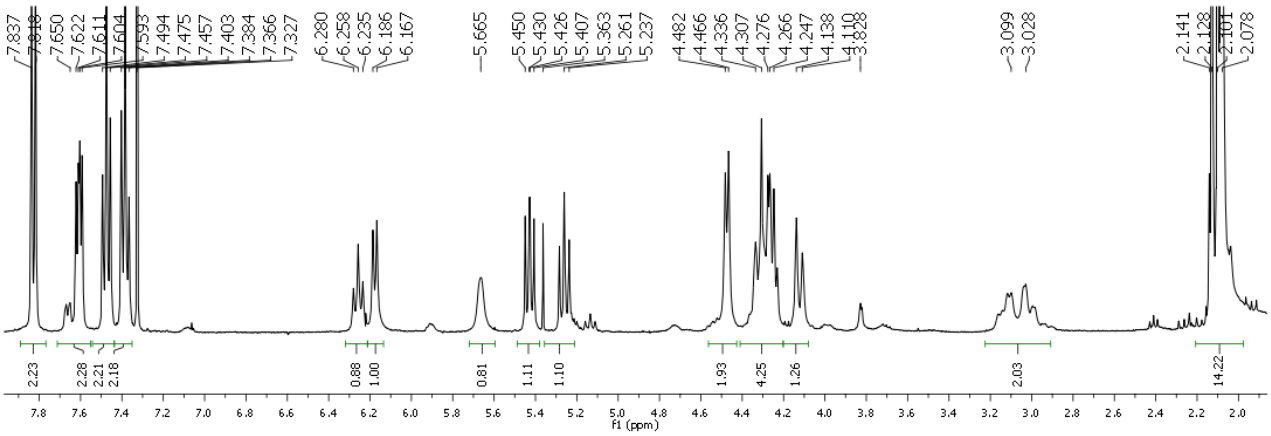


^{1}H-NMR spectrum of **98** (CD_3OD, 400 MHz)

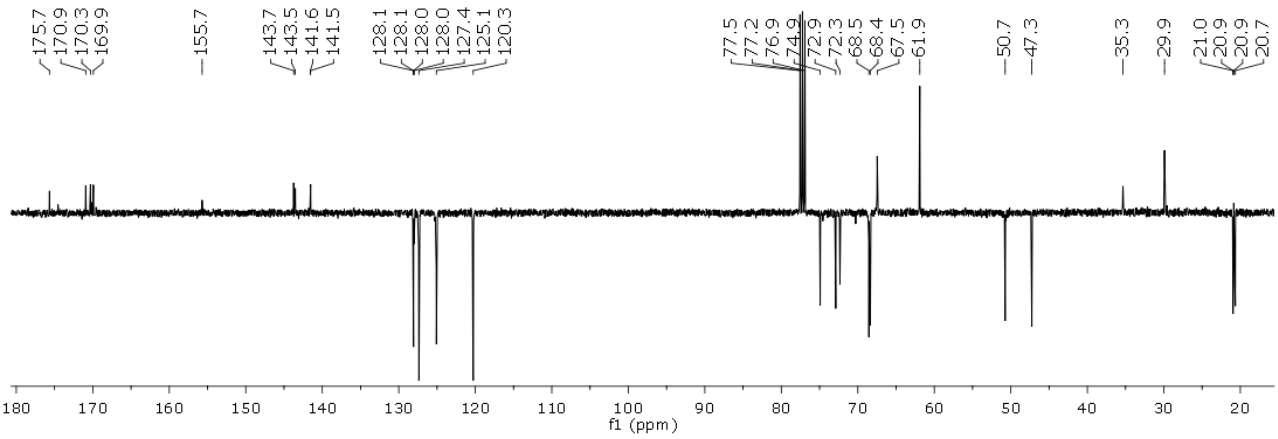


^{13}C-NMR spectrum of **98** (CD_3OD, 100 MHz)

Since spectra of the crude reaction mixtures that contain by-product **98** were recorded in $CDCl_3$, the ^{1}H-NMR spectrum of **98** in $CDCl_3$ is also reported.

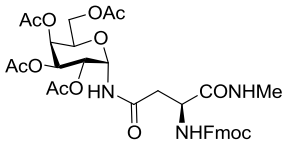


^{1}H-NMR spectrum of **98** ($CDCl_3$, 400 MHz)



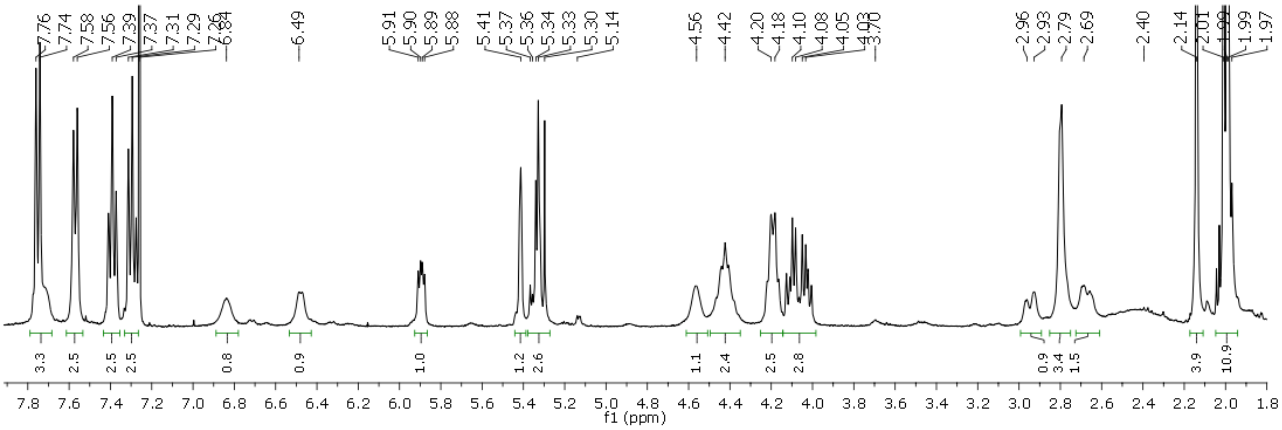
^{13}C-NMR spectrum of **98** ($CDCl_3$, 100 MHz)

N^{α}-Fluorenylmethoxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagine -*N*-methylamide (104)

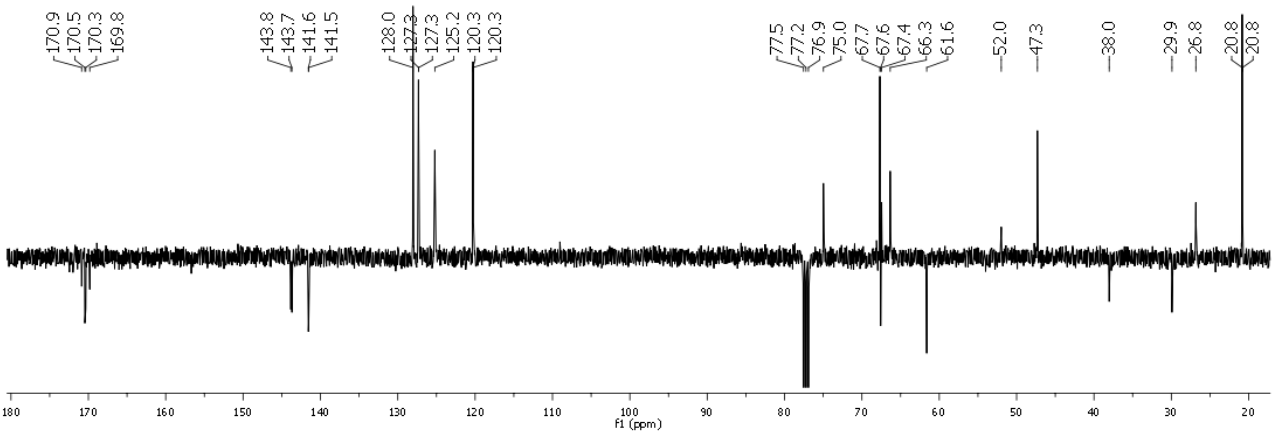


Compound **80** (102 mg, 0.149 mmol, 1 equiv.), *N*-methylamine hydrochloride (30 mg, 0.447 mmol, 3 equiv.) and PyBrop (167 mg, 0.357 mmol, 2.4 equiv.) were dissolved in dry CH_2Cl_2 (1.5 mL) under nitrogen at 0 °C. DIPEA (140 µL, 0.804 mmol, 5.4 equiv.) was added and the reaction mixture was stirred at 0 °C for 2 h and then at room temperature for 4 h (TLC, 8:2 chloroform/methanol and 1:9 hexane/ EtOAc). The solvent was evaporated, the residue was dissolved in EtOAc and the organic phase was washed with 1M HCl and saturated $NaHCO_3$. The organic phase was then dried with sodium sulfate and evaporated. The crude was purified by flash chromatography (1:9 hexane/EtOAc) to afford **104** (87 mg) in 84% yield. $[\alpha]^{20}_{D}$ = +52.8 (*c* = 0.5,

MeOH). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 7.75 (d, J = 7.2 Hz, 2H, 4-H-, 5-H-Fmoc), 7.72 (bs, 1H, Gal-NH-Asn), 7.57 (d, J = 7.6 Hz, 2H, 1-H-, 8-H-Fmoc), 7.39 (t, J = 7.6 Hz, 2H, 3-H-, 6-H-Fmoc), 7.30 (t, J = 7.2 Hz, 2H, 2-H-, 7-H-Fmoc), 6.84 (bs, 1H, NHMe), 6.47 (bs, 1 H, NH-Fmoc), 5.89 (dd, $J_{1,NH}$ = 7.6, $J_{1,2}$ = 4.4 Hz, 1H, 1-H), 5.41 (bs, 1H, 4-H), 5.36–5.28 (m, 2H, 2-H, 3-H), 4.56 (bs, 1H, α-H-Asn), 4.52–4.38 (m, 2H, CH_2-Fmoc), 4.22–4.15 (m, 2H, 9-H-Fmoc, 5-H), 4.12–4.00 (m, 2H, 6-H), 3.00–2.91 (m, 1H, β-CH2-Asn), 2.79 (s, 3H, $NHCH_3$), 2.70–2.63 (m, 2H, β-CH_2-Asn), 2.14 (s, 3H, CH_3CO), 2.00 (s, 3H, CHCO), 1.99 (s, 3H, CH_3CO), 1.96 (s, 3H, CH_3CO) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 171.6–169.8 (CO), 143.8, 143.7 (C_{quat}Fmoc), 141.6, 141.5 (C_{quat}Fmoc), 128.0 (C-2-, C-7-Fmoc), 127.3 (C-3-, C-6-Fmoc), 125.2 (C-1-, C-8-Fmoc), 120.3 (C-4-, C-5-Fmoc), 75.0 (C-1), 67.8, 67.4 (C-2, C-3, C-4), 67.6 (CH_2-Fmoc), 66.3 (C-5), 61.6 (C-6), 52.0 (α-C-Asn), 47.3 (C-9-Fmoc), 38.0 (β-CH_2-Asn), 26.8 ($NHCH_3$), 20.8 (4xOAc) ppm. MS (ESI): m/z = 720.4 $[M + Na]^+$.

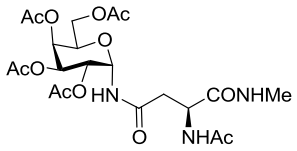


^{1}H-NMR spectrum of **104** ($CDCl_3$, 400 MHz)

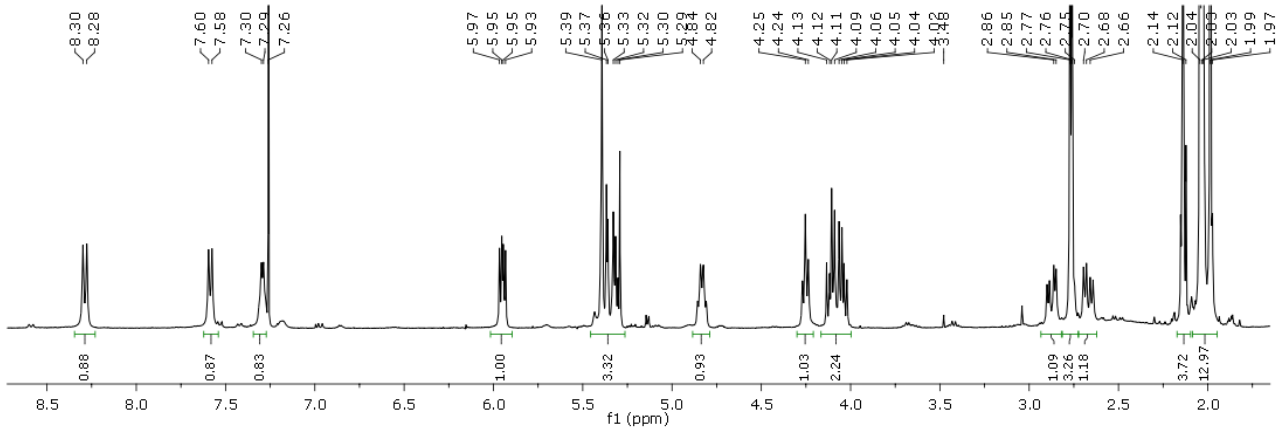


^{13}C-NMR spectrum of **104** ($CDCl_3$, 100 MHz)

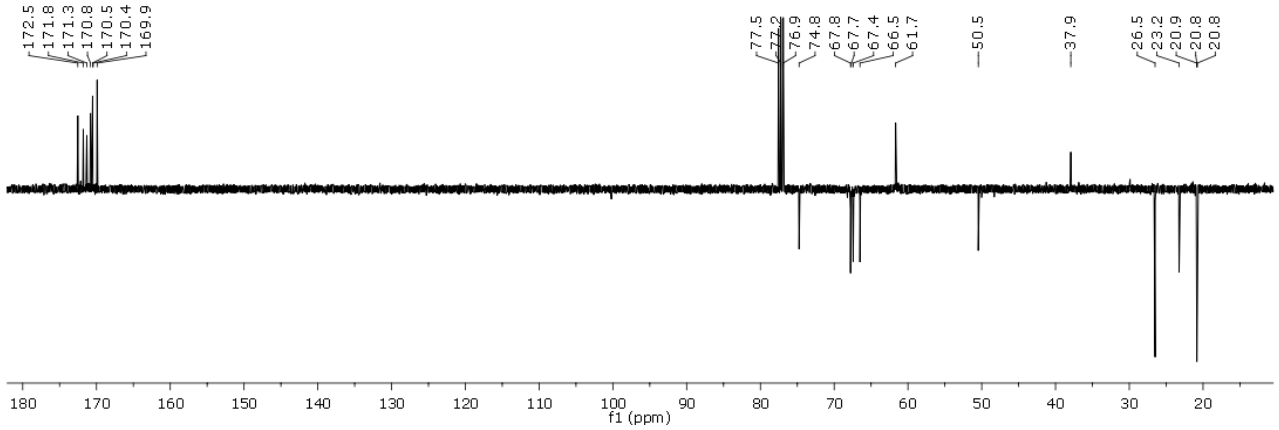
N^{α}-Acetyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagine-*N*-methylamide (102)



Compound **104** (65 mg, 0.093 mmol, 1 equiv.) was dissolved in dry THF (800 µL) under nitrogen. Octanethiol (29 µL, 0.158 mmol, 10 equiv.) and DBU were added sequentially and the reaction mixture was stirred for 1 h (TLC, 9:1 chloroform/methanol and 1:9 hexane/EtOAc). The solvent was evaporated. The residue was washed thoroughly with a mixture of cold diethyl ether and pentane (1:1). The crude solid was then dissolved in Ac_2O (500 µL) and pyridine (15 µL) was added. The reaction mixture was stirred for 3 h at room temperature (TLC, 90:10 chloroform/methanol), then the solvent was evaporated. The reaction mixture was purified by flash chromatography (95:5 chloroform/methanol) to afford **102** (32 mg) in 62% yield over the two steps. $[\alpha]^{20}_{D}$ = +84.2 (*c* = 0.7, MeOH). ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 8.28 (d, *J* = 8.6 Hz, 1H, Gal-NH-Asn), 7.58 (d, *J* = 7.6 Hz, 1H, NHAc), 7.29 (d, *J* = 4.8 Hz, 1H, NHMe), 5.95 (dd, $J_{1,NH}$ = 8.6, $J_{1,2}$ = 5.2 Hz, 1H, 1-H), 5.39 (m, 1H, 4-H), 5.34–5.27 (m, 2H, 2-H, 3-H), 4.83 (m, 1H, α-H-Asn), 4.22–4.18 (m, 1H, 5-H), 4.12–4.01 (m, 2H, 6-H), 2.84 (dd, J_{gem} = 15.4, $J_{\alpha,\beta}$ = 5.4 Hz, 1H, β-H-Asn), 2.79 (d, $J_{Me,NH}$ = 4.8 Hz, 3H, CH_3N), 2.63 (dd, J_{gem} = 15.4, $J_{\alpha,\beta}$ = 6.8 Hz, 1H, β-H-Asn), 2.14 (s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO), 2.02 (s, 3H, CH_3CO), 2.00 (s, 3H, CH_3CO), 1.98 (s, 3H, NHAc) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 172.5–169.9 (CO), 74.7 (C-1), 67.7 (C-2, C-3), 67.5 (C-4), 66.5 (C-5), 61.7 (C-6), 50.5 (α-C-Asn), 37.9 (β-CH_2-Asn), 26.5 ($NHCH_3$), 23.2 ($CONHCH_3$), 20.8, 20.7 (4xOAc) ppm. MS (ESI): *m*/*z* = 540.4 $[M + Na]^+$.

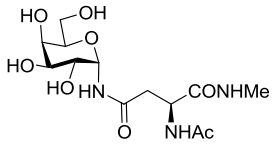


^{1}H-NMR spectrum of **102** ($CDCl_3$, 400 MHz)

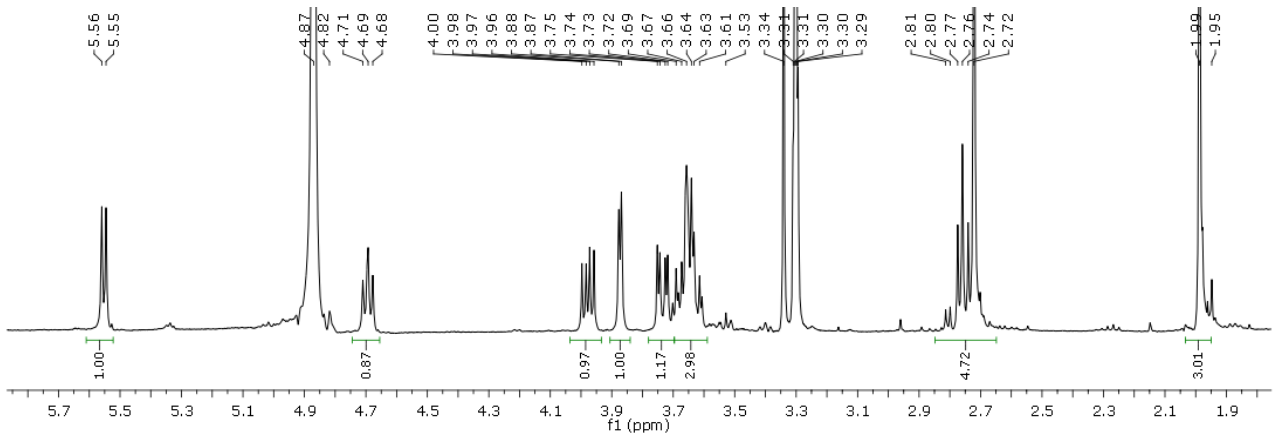


^{13}C-NMR spectrum of **102** ($CDCl_3$, 100 MHz)

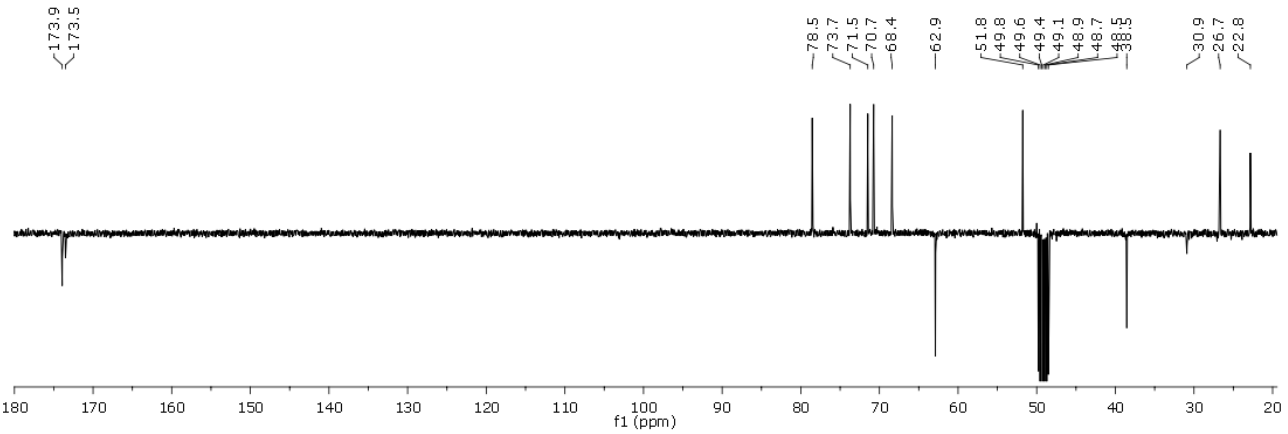
*N*α-Acetyl-*N*γ-(α-D-galactopyranosyl)-L-asparagine-*N*-methylamide (100)



Compound **102** (13 mg, 0.025 mmol, 1 equiv.) was dissolved in dry methanol (250 μL) and a catalytic amount of K_2CO_3 (0.095 equiv., pH 8.5) was added. The reaction mixture was stirred for 2 h (TLC, 6:4 chloroform/methanol and 8:2 chloroform/methanol). IRA H^+ 120 was added to neutral pH. The mixture was filtered and washed with methanol. The solvent was evaporated to afford **100** (8 mg) in 91% yield. $[\alpha]^{20}_D$ = +43.9 (*c* = 0.15, MeOH). ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 5.55 (d, $J_{1,2}$ = 5.4 Hz, 1H, 1-H), 4.69 (t, J = 6.4 Hz, 1H, α-H-Asn), 3.98 (dd, $J_{2,3}$ = 10, $J_{2,1}$ = 5.4 Hz, 1H, 2-H), 3.87 (d, $J_{4,5}$ = $J_{4,3}$ = 3.2 Hz, 1H, 4-H), 3.74 (dd, $J_{3,2}$ = 10, $J_{3,4}$ = 3.2 Hz, 1H, 3-H), 3.70–3.61 (m, 3H, 5-H, 6-H), 2.82–2.72 (m, 5H, β-CH_2-Asn, $NHCH_3$), 1.99 (s, 3H, NHAc) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 173.9–173.5 (CO), 78.5 (C-1), 73.7 (C-5), 71.5 (C-3), 70.7 (C-4), 68.4 (C-2), 62.9 (C-6), 51.8 (α-C-Asn), 38.6 (β-CH_2-Asn), 26.7 ($NHCH_3$), 22.8 ($NHCOO\underline{C}H_3$) ppm. FT-ICRMS (ESI): calcd. for $[C_{33}H_{45}O_{15}N_3Na]^+$ 372.13774; found 372.13771.

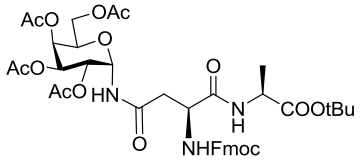


^{1}H-NMR spectrum of **100** (CD_3OD, 400 MHz)



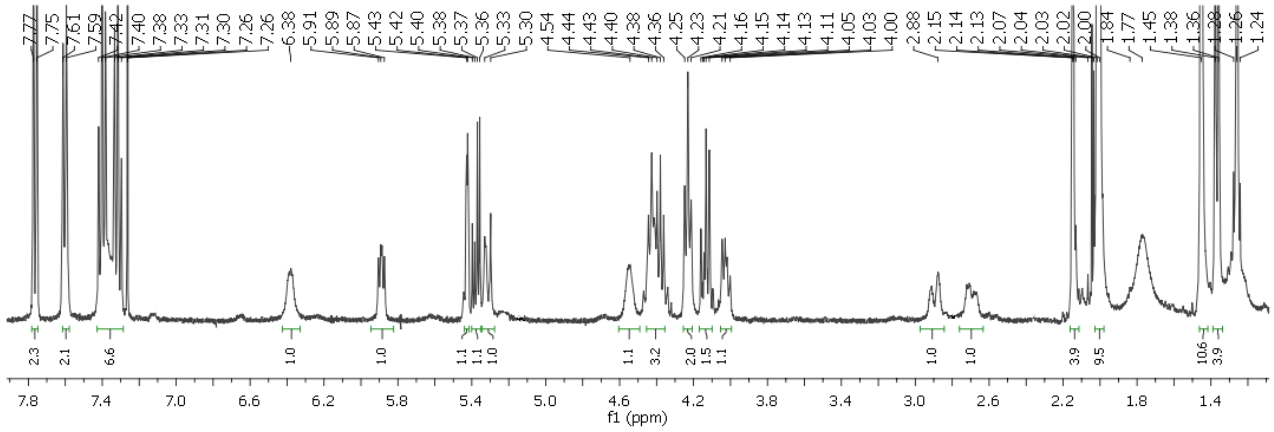
^{13}C-NMR spectrum of **100** (CD_3OD, 100 MHz)

N^{α}-Fluorenylmethoxycarbonyl-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagyl -L-alanine *tert*-butyl ester (108)

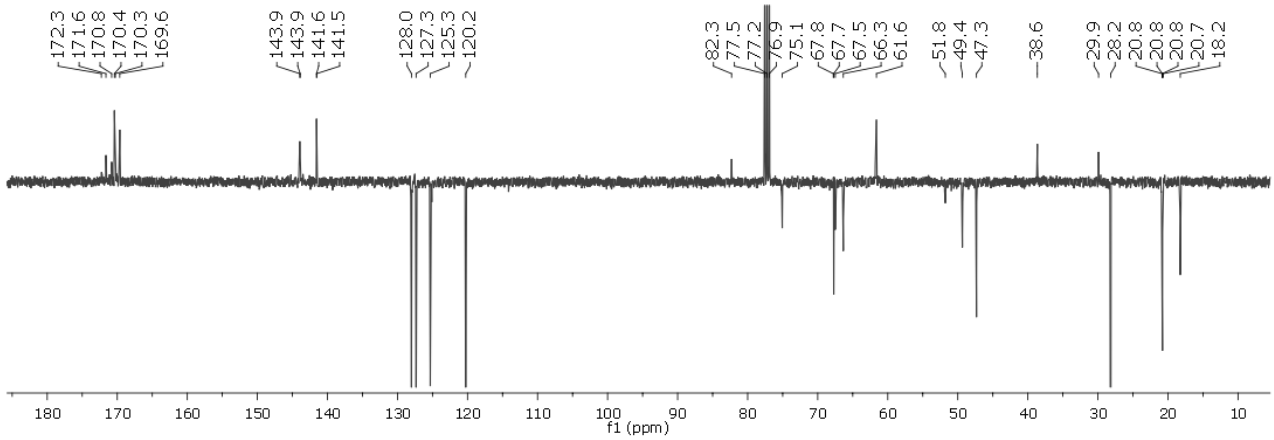


Compound **80** (100 mg, 0.146 mmol, 1 equiv.), H-Ala-OtBu hydrochloride **107** (58 mg, 0.321 mmol, 2.2 equiv.) and PyBrop (150 mg, 0.321 mmol, 2.2 equiv.) were dissolved in dry CH_2Cl_2 (1.5 mL) under nitrogen at 0 °C. DIPEA (132 μL, 0.760 mmol, 5.2 equiv.) was added and the reaction mixture was stirred at 0 °C for 1 h and then at room temperature for 3 h (TLC, 8:2 chloroform/methanol and 4:6 hexane/EtOAc). The solvent was evaporated, the residue was dissolved in EtOAc and the organic phase was washed with 1M HCl and saturated $NaHCO_3$ and then dried with sodium sulfate. The solvent was evaporated and the crude was purified by flash chromatography (4:6 hexane/EtOAc) to afford **108** (99 mg) in 83% yield. ^{1}H NMR (400 MHz, CDCl3, 25 °C): δ = 7.76 (d, J = 7.4 Hz, 2H, 4-H-, 5-H-Fmoc), 7.60 (d, J = 7.3 Hz, 2H, 1-H-, 8-H-Fmoc), 7.40 (t, J =7.4 Hz, 2H, 3-H-, 6-H-Fmoc), 7.36-7.30 (m, 2H, Gal-NH-Asn, Asn-NH-Ala), 7.31 (t, J = 7.4 Hz, 2H, 2-H-, 7-H-Fmoc), 6.38 (bs, 1H, NH-Fmoc), 5.89 (dd, $J_{1,NH}$ = 7.6, $J_{1,2}$ = 5.2 Hz, 1H, 1-H), 5.43-5.40 (m, 1H, 4-H), 5.38 (dd, J_{2-3} = 10.9, $J_{2,1}$ = 5.2 Hz, 1H, 2-H), 5.34–5.28 (m, 1H, 3-H), 4.54 (bs, 1H, α-H-Asn), 4.46–4.34 (m, 3H, CH_2-Fmoc, α-H-Ala), 4.26–4.20 (m, 2H, 9-H-Fmoc, 5-H), 4.16–4.00 (m, 2H, 6-H), 2.94–2.85 (m, 1H, β-CH_2-Asn), 2.75–2.64 (m, 1H, β-CH_2-Asn), 2.14 (s, 3H,CH_3CO), 2.02 (s, 3H, CH_3CO), 2.00 (s, 6H, CH_3CO), 1.45 (s, 3H, OtBu), 1.37 (d, J = 7.2 Hz, 3H, CH_3-Ala) ppm. ^{13}C NMR (100 MHz, CDCl3, 25 °C): δ = 171.6–169.7 (CO), 143.9, 143.8 (CquatFmoc), 141.5 (CquatFmoc), 128.0 (C-2-, C-7-Fmoc), 127.3 (C-3-, C-6-Fmoc), 125.3 (C-1-, C-8-Fmoc), 120.2 (C-4-, C-5-Fmoc), 82.3 (C_{quat}-OtBu), 75.1 (C-1), 67.7 (CH2-Fmoc), 67.7, 67.4 (C-2, C-3, C-4), 66.2 (C-5), 61.6 (C-6), 51.6 (α-C-Asn), 49.3 (α-C-Ala), 47.3 (C-9-Fmoc), 38.5

(β-CH2-Asn), 28.2 (CH_3OtBu), 20.8 (4xOAc), 18.2 (CH_3-Ala) ppm. MS (ESI): *m*/*z* = 834.3 $[M+Na]^+$.

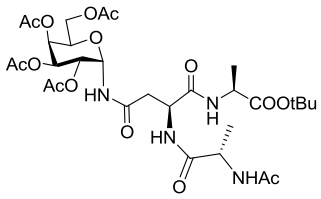


^{1}H-NMR spectrum of **108** ($CDCl_3$, 400 MHz)



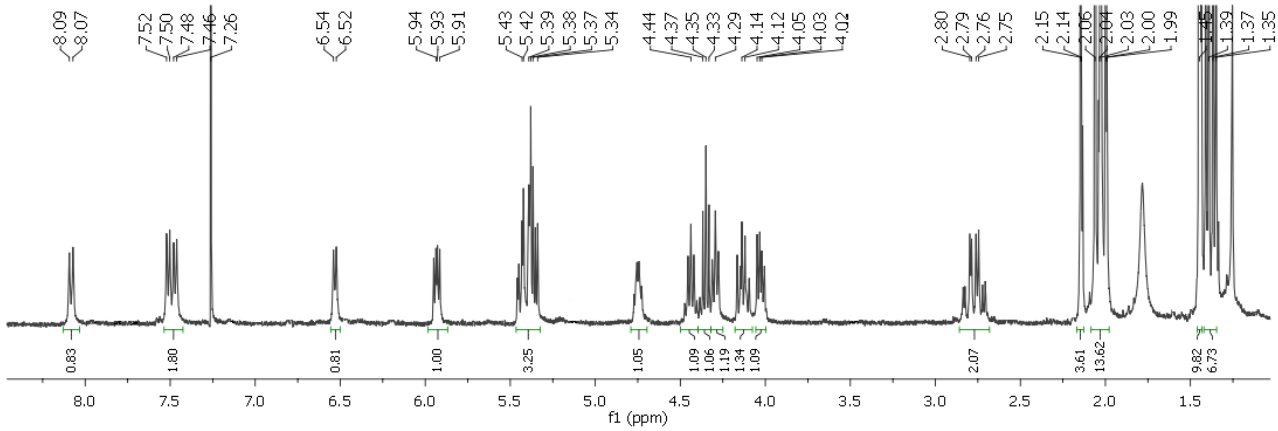
^{13}C-NMR spectrum of **108** ($CDCl_3$, 100 MHz)

N^{α}-(L-*N*-Acetylalanyl)-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagyl-L-alanine *tert*-butyl ester (110)

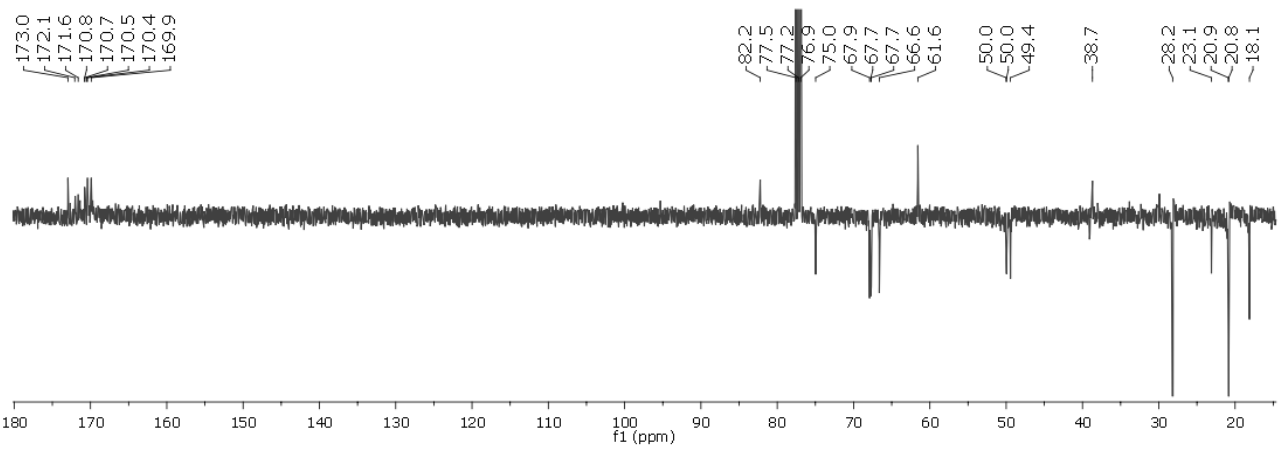


Compound **108** (99 mg, 0.122 mmol, 1 equiv.) was dissolved in dry THF (1.2 mL) under nitrogen. Octanethiol (211 µL, 1.219 mmol, 10 equiv.) and DBU (10 µL, 0.070 mmol, 0.5 equiv.) were added sequentially and the reaction mixture was stirred for 1 h (TLC, 9:1 chloroform/methanol and 2:8 hexane/EtOAc). The solvent was evaporated and the residue was washed thoroughly with a mixture of cold diethyl ether and pentane (1:1). The crude solid was then dissolved in dry DMF (1.5 mL)

under nitrogen at 0 °C. Ac-Ala-OH **109** (40 mg, 0.305 mmol, 2.5 equiv.), HATU (93 mg, 0.244 mmol, 2 equiv.) and DIPEA (75 µL, 0.427 mmol, 3.5 equiv.) were added and the reaction mixture was stirred at 0 °C for 1 h and then at room temperature for 4 h (TLC, 9:1 chloroform/methanol). The solvent was evaporated, the residue was dissolved in EtOAc and the organic phase was washed with 1M HCl and saturated $NaHCO_3$ and then dried with sodium sulfate. The solvent was evaporated and the crude was purified by flash chromatography (95:5 chloroform/methanol) to afford **110** (55 mg) in 65% yield. ^{1}H NMR (400 MHz, $CDCl_3$, 25 °C): δ = 8.28 (d, J = 8.6 Hz, 1H, Gal-NH-Asn), 8.08 (d, J = 8.6 Hz, 1H, Gal-NH-Ala), 7.51 (d, J = 7.4 Hz, 1H, NHAsn), 7.47 (d, J = 7.4 Hz, 1H, NHAc), 6.54 (d, J = 7.4 Hz, 1H, NHAla), 5.93 (dd, $J_{1,NH}$ = 8.6, $J_{1,2}$ = 5.3 Hz, 1H, 1-H), 5.47-5.31 (m, 3H, 4-H, 2- H, 3-H), 4.78-4.70 (m, 1H, α-H-Asn), 4.48–4.39 (m, 1H, α-H-AlaOtBu), 4.38–4.32 (m, 2H, α-H-AcAla, 5-H), 4.14 (dd, $J_{6\text{-}6'}$ = 11.0, $J_{6\text{-}5}$ = 7.3 Hz, 1H, 6-H), 4.03 (dd, $J_{6\text{-}6'}$ = 11.0, $J_{6'\text{-}5}$ = 6.1 Hz, 1H, 6'-H), 2.81 (dd, J_{gem} = 15.4, $J_{\alpha,\beta}$ = 4.1 Hz, 1H, β-HAsn), 2.73 (dd, J_{gem} = 15.4, $J_{\alpha,\beta}$ = 6.2 Hz, 1H, β-H-Asn), 2.14 (s, 3H,CH_3CO), 2.06 (s, 3H, CH_3CO), 2.04 (s, 6H, CH_3CO, $NHCH_3$), 2.00 (s, 3H, CH_3CO), 1.45 (s, 3H, OtBu), 1.40 (d, J = 7.1 Hz, 3H, CH_3-Ala), 1.36 (d, J = 7.2 Hz, 3H, CH_3-Ala) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 172.9–169.9 (CO), 82.2 (C_{quat}-OtBu), 75.0 (C-1), 67.9 (C-2), 67.7 (C-3, C-4), 66.6 (C-5), 61.6 (C-6), 50.0 (α-C-Asn), 50.0 (α-C-Ala), 49.4 (α-C-Ala) 38.7 (β-CH_2-Asn), 28.2 (CH_3OtBu), 23.1 ($NHCH_3$), 20.9-20.8 (4xOAc), 18.1 (CH_3-Ala) ppm. MS (ESI): m/z = 725.2 $[M + Na]^+$.

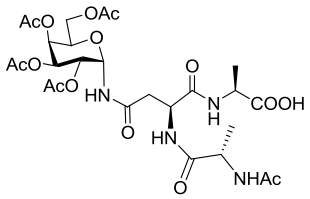


^{1}H-NMR spectrum of **110** ($CDCl_3$, 400 MHz)

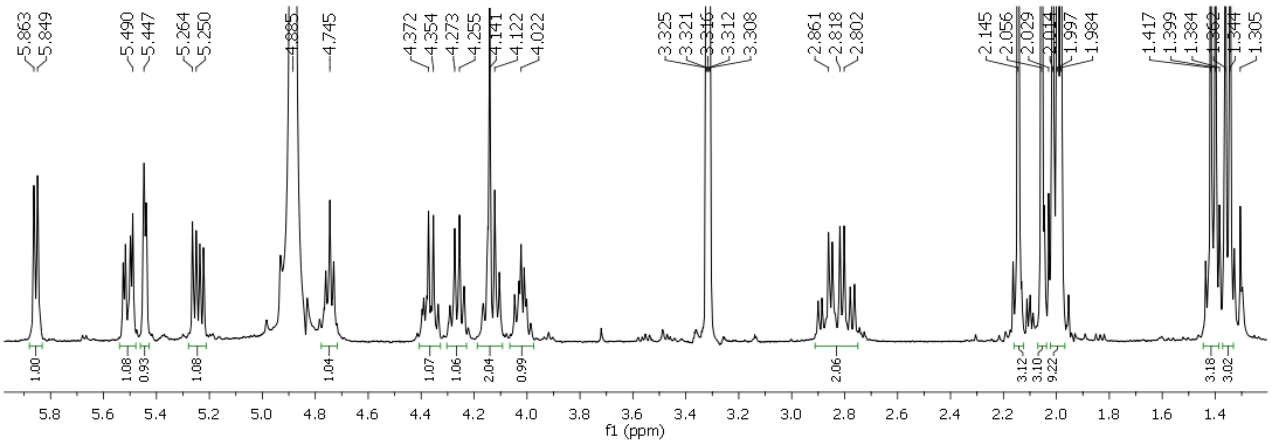


^{13}C-NMR spectrum of **110** ($CDCl_3$, 100 MHz)

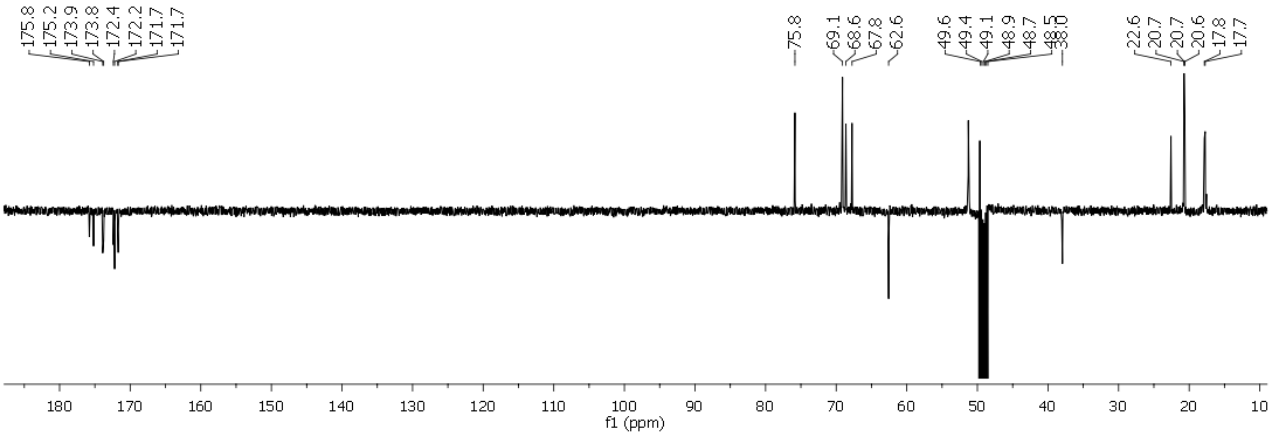
N^{α}-(L-*N*-Acetylalanyl)-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagyl-L-alanine (111)



Compound **110** (20 mg, 0.028 mmol, 1 equiv.) was dissolved in dry CH_2Cl_2 (1.5 mL) under nitrogen**.** TFA (212 µL, 0.280 mmol, 100 equiv.) was added and the reaction mixture was stirred for 2 h at room temperature (TLC, 9:1 chloroform/methanol). The solvent was coevaporated with toluene and with CH_2Cl_2 and compound **111** was used without purification for the subsequent reaction. $[\alpha]^{20}_D$ = +46.3 (*c* = 0.4, MeOH). ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 5.86 (d, $J_{1,2}$ = 5.6 Hz, 1H, 1-H), 5.50 (dd, $J_{3,2}$ = 11.0 , $J_{3,4}$ = 3.4 Hz, 1H, 3-H), 5.44 (d, $J_{4,5}$ = $J_{4,3}$ = 3. Hz, 1H, 4-H), 5.24 (dd, $J_{2,3}$ = 11.0 , $J_{2,1}$ = 5.6 Hz, 1H, 2-H), 4.75 (t, *J* = 6.1 Hz, 1 H, α-H-Asn), 4.36 (m, 1H, α-H-Ala), 4.26 (m, 1H, α-H-AcAla), 4.18-4.09 (m, 2H, 5-H, 6-H), 4.06-3.98 (m, 1H, 6'-H), 2.87 (dd, J_{gem} = 15.7, $J_{\alpha,\beta}$ = 5.8 Hz, 1H, β-H-Asn), 2.79 (dd, J_{gem} = 15.7, $J_{\alpha,\beta}$ = 6.4 Hz, 1H, β-H-Asn), 2.14 (s, 3H, CH_3CO), 2.05 (s, 3H, CH_3CO), 2.01 (s, 3H, NHAc), 2.00 (s, 3H, CH_3CO), 1.98 (s, 3H, CH_3CO), 1.40 (d, *J* = 7.3 Hz, 3H, CH_3-Ala), 1.35 (d, *J* = 7.3 Hz, 3H, CH_3-Ala) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ =175.8–171.7 (CO), 75.8 (C-1), 69.1 (C-4, C-3), 68.6 (C-5), 67.8 (C-2), 62.6 (C-6), 51.3 (α-C-Asn), 51.2 (α-C-Ala), 49.4 (α-C-Ala), 38.0 (β-CH_2-Asn), 22.6 ($NHCH_3$), 20.6 (4xOAc), 17.8 (CH_3), 17.7 (CH_3) ppm. MS (ESI): *m*/*z* = 647.3 $[M + H]^+$.

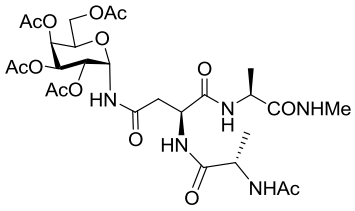


^{1}H-NMR spectrum of **111**(CD_3OD, 400 MHz)



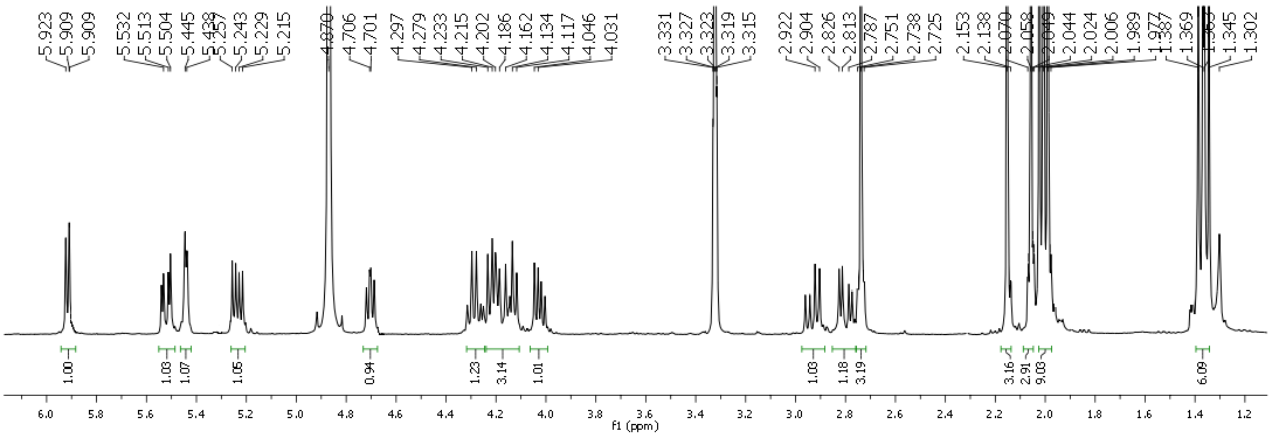
^{13}C-NMR spectrum of **111** (CD_3OD, 100 MHz)

N^{α}-(L-*N*-Acetylalanyl)-N^{γ}-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)-L-asparagyl-L-alanin*e* *N*-methylamide (112)

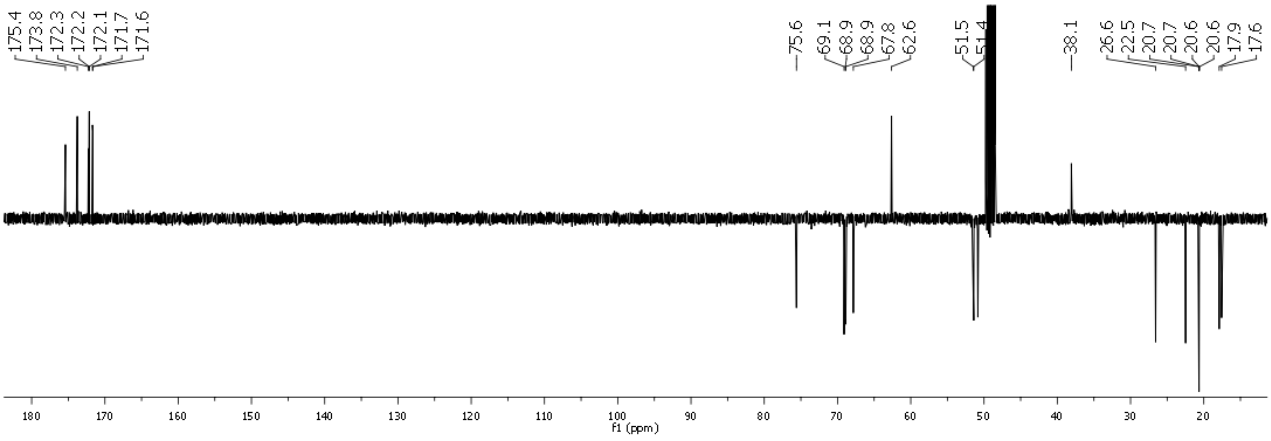


The crude product **111** was dissolved in dry DMF (500 µL) together with PyBrop (29 mg, 0.062 mmol, 2.2 equiv.) and methylamine hydrochloride (8 mg, 0.120 mmol, 4.2 equiv.) under nitrogen at 0 °C. DIPEA (75 µL, 0.427 mmol, 3.5 equiv.) was added and the reaction mixture was stirred at 0 °C for 1 h and then at room temperature for 2 h (TLC, 9:1 chloroform/methanol). The solvent was evaporated and the crude was purified by automated chromatography on a reversed-phase C-18 column (CH_3CN/H_2O 5 to 30 % t_r = 5 min) to afford 10 mg of **112** (54 % over two steps). ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 5.92 (d, $J_{1,2}$ = 5.5 Hz, 1H, 1-H), 5.52 (dd, $J_{3\text{-}2}$ = 11.1, $J_{3\text{-}4}$ = 3.5 Hz,

1H, 3-H), 5.44 (m, 1H, 4-H), 5.24 (dd, $J_{3\text{-}2}$ = 11.1, $J_{2\text{-}1}$ = 5.5 Hz, 1H, 2-H), 4.70 (dd, $J_{\alpha,\beta}$ = 7.2 Hz, $J_{\alpha,\beta}$ = 5.3 Hz, 1H, α-H-Asn), 4.32–4.27 (m, 1H, α-H-AlaOtBu), 4.24–4.11 (m, 3H, α-H-AcAla, 5-H, 6-H), 4.05-3.99 (m, 1H, 6-H), 2.93 (dd, J_{gem} = 15.5, $J_{\alpha,\beta}$ = 7.2 Hz, 1H, β-HAsn), 2.73 (dd, J_{gem} = 15.5, $J_{\alpha,\beta}$ = 5.3 Hz, 1H, β-H-Asn), 2.15 (s, 3H,CH_3CO), 2.05 (s, 3H, CH_3CO), 2.02 (s, 6H, CH_3CO, $NHCH_3$), 1.99 (s, 3H, CH_3CO), 1.38 (d, J = 7.2 Hz, 3H, CH_3-Ala), 1.36 (d, J = 7.2 Hz, 3H, CH_3-Ala) ppm. ^{13}C NMR (100 MHz, CD_3OD, 25 °C): δ = 175.4–171.6 (CO), 75.6 (C-1), 69.1 (C-2), 68.9 (C-3, C-4), 67.8 (C-5), 62.6 (C-6), 51.5 (α-C-Asn), 51.4 (α-C-Ala), 50.9 (α-C-Ala) 38.1 (β-CH2-Asn), 26.6 ($NHCOO\underline{C}H_3$), 22.5 ($NHCH_3$), 20.7-20.6 (4xOAc), 17.9-17.6 (CH_3-Ala) ppm. MS (ESI): m/z = 682.3 $[M + Na]^+$.

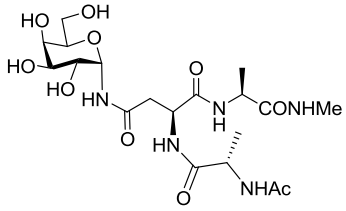


^{1}H-NMR spectrum of **112** (CD_3OD, 400 MHz)

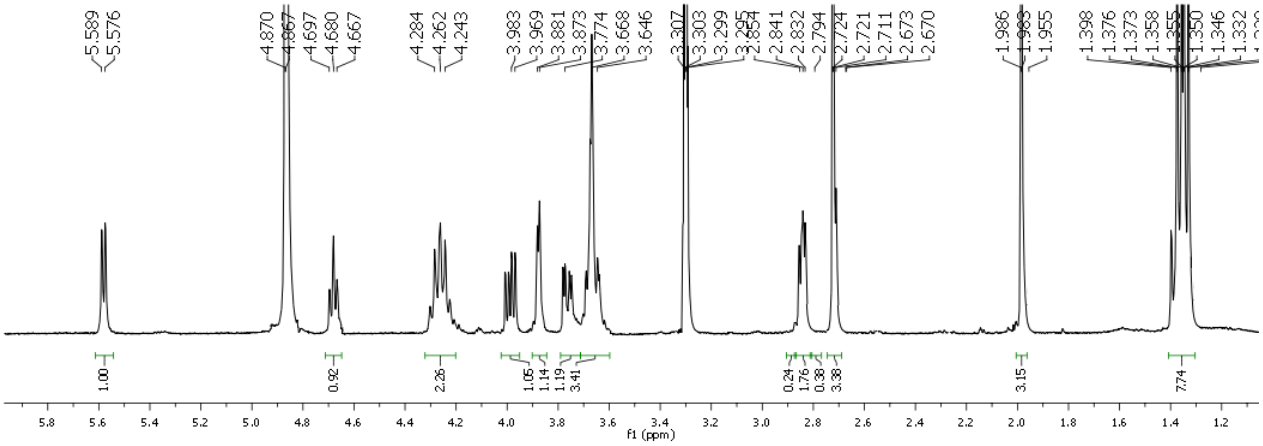


^{13}C-NMR spectrum of **112** (CD_3OD, 100 MHz)

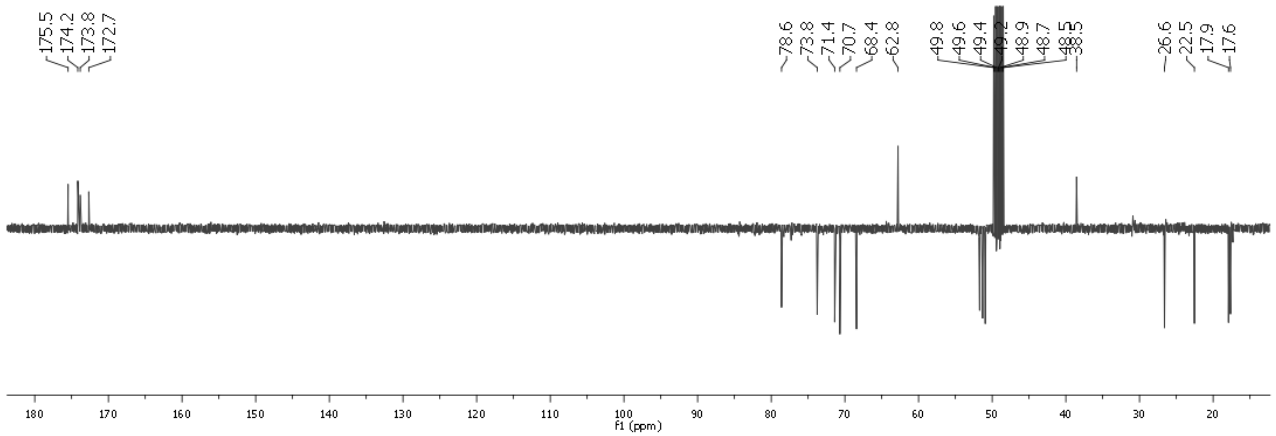
N^{α}-(L-*N*-Acetylalanyl)-N^{γ}-(α-D-galactopyranosyl)-L-asparagyl-L-alanine *N*-methylamide (101)



Compound **112** (10 mg, 0.015 mmol, 1 equiv.) was dissolved in dry methanol (250 μL) and a catalytic amount of K_2CO_3 (0.1 equiv., pH 8.5) was added. The reaction mixture was stirred for 1 h (TLC, 6:4 chloroform/methanol and 8:2 chloroform/methanol). IRA H^+ 120 was added to neutral pH. The mixture was filtered and washed with methanol. The solvent was evaporated and the compound was purified by preparative HPLC (C-18 reverse phase, 100:0 to 60:40 H_2O/CH_3CN in 14 min; t_r = 5 min) to afford **101** (6 mg) in 85% yield. ^{1}H NMR (400 MHz, CD_3OD, 25 °C): δ = 5.58 (d, $J_{1,2}$ = 5.6 Hz, 1H, 1-H), 4.67 (t, J = 6.0 Hz, 1H, α-H-Asn), 4.30-4.21 (m, 2H, α-H-Ala), 3.98 (dd, $J_{2,3}$ = 10.2 Hz, $J_{2,1}$ = 5.6 Hz, 1H, 2-H), 3.87 (d, J = 3.4 Hz, 1H, 4-H), 3.74 (dd, $J_{3\text{-}2}$ = 10.2 Hz, $J_{3\text{-}4}$ = 3.4 Hz, 1H, 3-H), 3.70–3.60 (m, 3H, 5-H, 6- H), 2.85-2.78 (m, 2H, β-CH_2-Asn), 2.72 (s, 1H, $NHCH_3$), 1.98 (s, 3H, NHAc), 1.36 (d, J = 7.3 Hz, 3H, CH_3Ala), 1.33 (d, J = 7.2 Hz, 3H CH_3Ala) ppm. ^{13}C NMR (100 MHz, $CDCl_3$, 25 °C): δ = 175.5–172.6 (CO), 78.6 (C-1), 73.8 (C-5), 71.4 (C-3), 70.7 (C-4), 68.4 (C-2), 62.8 (C-6), 51.7 (α-C-Asn), 51.3 (α-C-Ala), 50.9 (α-C-Ala), 38.5 (β-CH_2-Asn), 26.6 ($NHCH_3$), 22.5 ($NHCOO\underline{C}H_3$), 17.9, 17.6 (CH_3Ala) ppm. MS (ESI): m/z = 492.2 $[M + H]^+$. FT-ICR MS (ESI): calcd. for $[C_{19}H_{33}N_5O_{10}Na]^+$ 514.21196; found 514.21128.

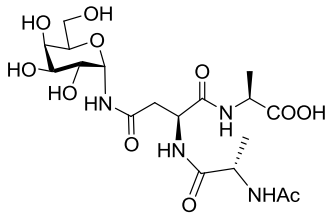


^{1}H-NMR spectrum of **101** (CD_3OD, 400 MHz)

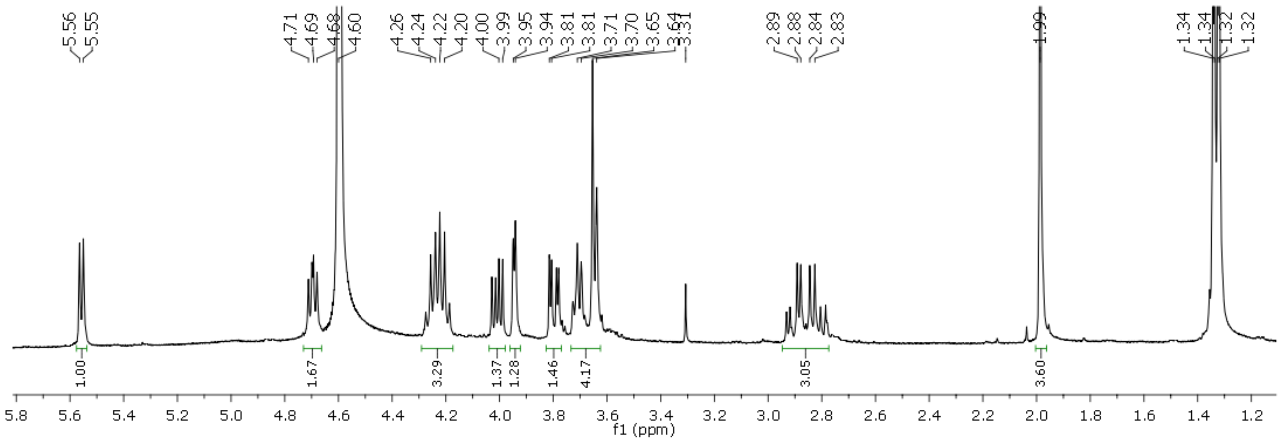


^{13}C-NMR spectrum of **101** (CD_3OD, 100 MHz)

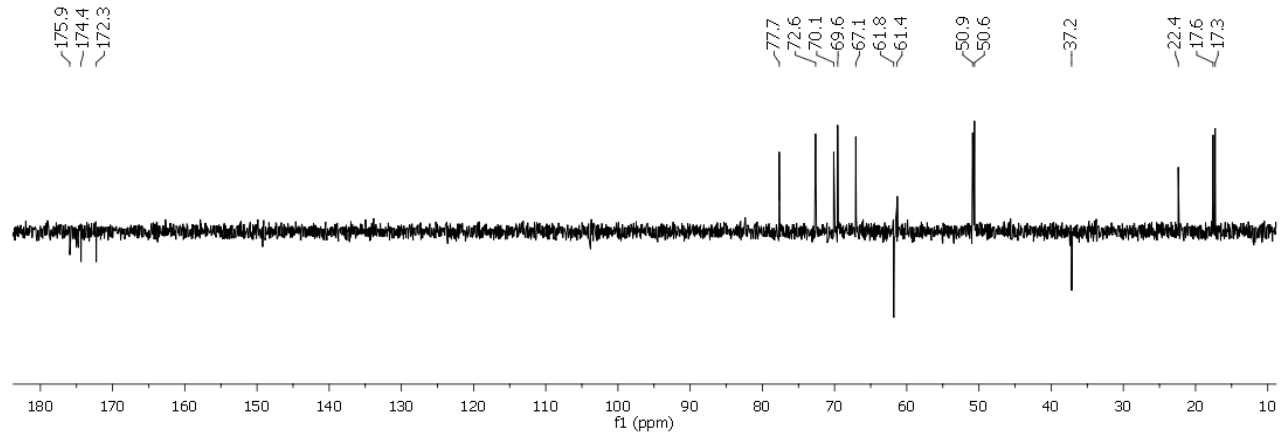
N^{α}-(L-*N*-Acetylalanyl)-N^{γ}-(α-D-galactopyranosyl)-L-asparagyl-L-alanine*N*-methylamide (113a)



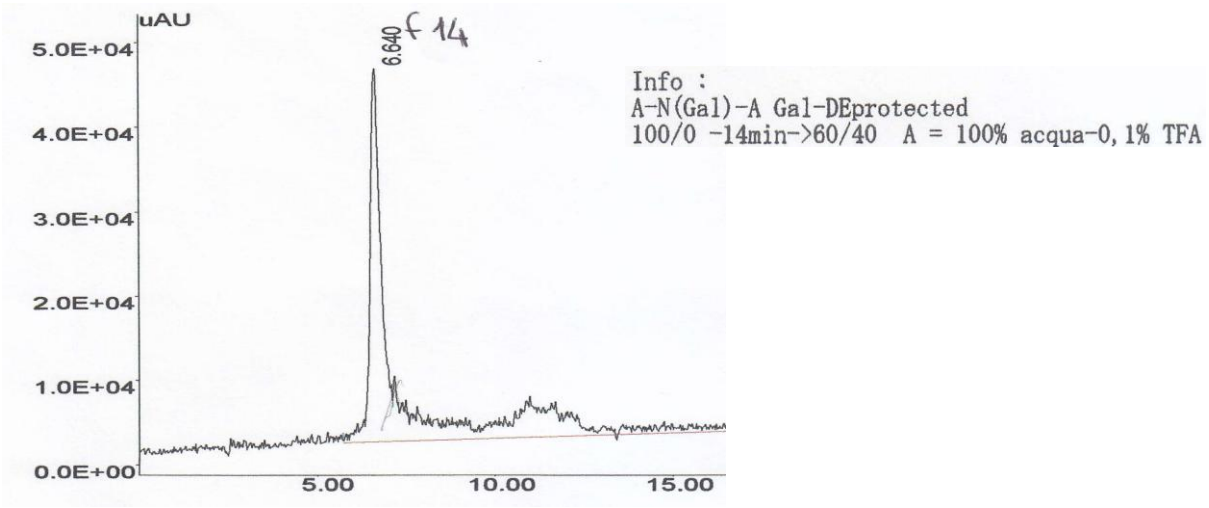
Compound **111** (8 mg, 0.012 mmol, 1 equiv.) was dissolved in dry methanol (250 µL) and K_2CO_3 (0.7 equiv.) was added. The reaction mixture was stirred for 2 h (TLC, 6:4 chloroform/methanol and 8:2 chloroform/methanol). The reaction was neutralized by adding a 0.035M HCl solution (250 µL, 0.7 equiv.). The solvent was evaporated and the compound was purified by preparative HPLC (Phenomenex Jupiter C-18 (5 µm, 300 Å, column 50 x 30 mm), gradient: H_2O/CH_3CN 100:0 to 60:40 in 14 min; t_r = 6 min) to afford **113a** (5 mg) in 84% yield. $[\alpha]^{20}_D$ = +40.4 (*c* = 0.15, MeOH). ^{1}H NMR (400 MHz, D_2O, 25 °C): δ = 5.55 (d, $J_{1,2}$ = 5.6 Hz, 1H, 1-H), 4.70 (dd, *J* = 5.6, *J* = 8.0 Hz, 1H, α-HAsn), 4.23 (m, 2H, α-H-Ala), 4.00 (dd, $J_{2,3}$ = 10.6, $J_{2,1}$ = 5.6 Hz, 1H, 2-H), 3.93 (d, $J_{4,3}$ = 3.4 Hz, 1H, 4-H), 3.80 (dd, $J_{3,2}$ = 10.6, $J_{3,4}$ = 3.4 Hz, 1H, 3-H), 3.73–3.68 (m, 1H, 6-H), 3.66–3.62 (m, 1H, 6'-H), 2.91 (dd, J_{gem} = 16.0, $J_{\alpha,\beta}$ = 5.6 Hz, 1H, β-H-Asn), 2.82 (dd, J_{gem} = 16.0, $J_{\alpha,\beta}$ = 8.0 Hz, 1H, β-H-Asn), 1.99 (s, 3H, NHAc), 1.33 (dd, *J* = 8.8 Hz, 6H, CH_3Ala) ppm. ^{13}C NMR (300 MHz, D_2O, 25 °C): δ = 174.4–172.2 (CO), 77.7 (C-1), 72.6 (C-5), 70.1 (C-3), 69.6 (C-4), 67.0 (C-2), 61.8 (C-6), 61.4 (α-C-Asn), 50.9 (α-CAla), 50.6 (α-C-Ala), 37.2 (β-CH_2-Asn), 22.4 ($NHCH_3$), 17.6 (CH_3Ala), 17.3 (CH_3Ala) ppm. FT-ICR MS (ESI): calcd. for $[C_{18}H_{29}O_{11}N_4]^-$ 477.18383; found 477.



^{1}H-NMR spectrum of **113a** (CD_3OD, 400 MHz)

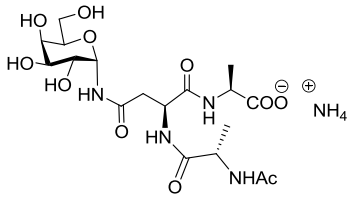


^{13}C-NMR spectrum of **113a** (CD_3OD, 100 MHz)

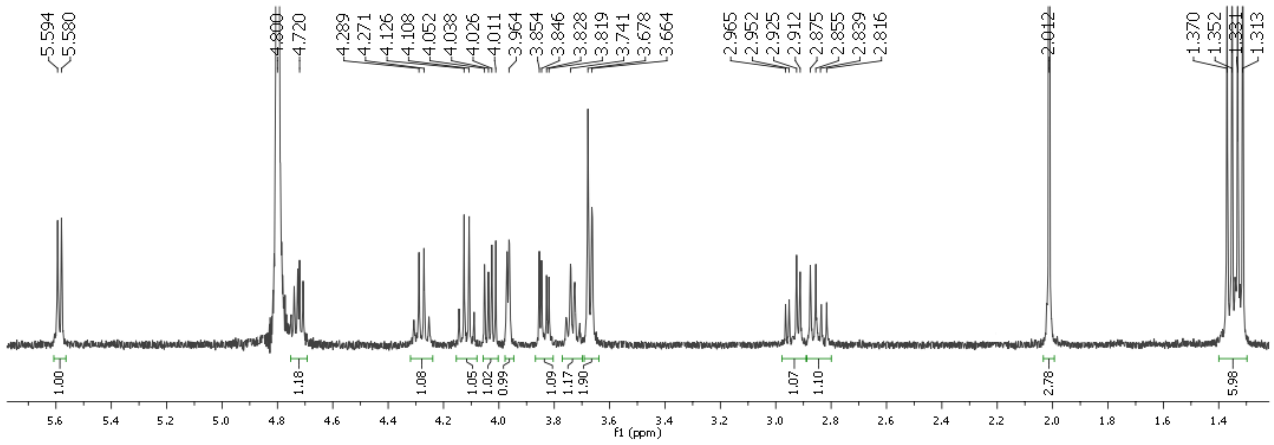


Chromatogram of **113a** purified by preparative HPLC (Jupiter C-18)

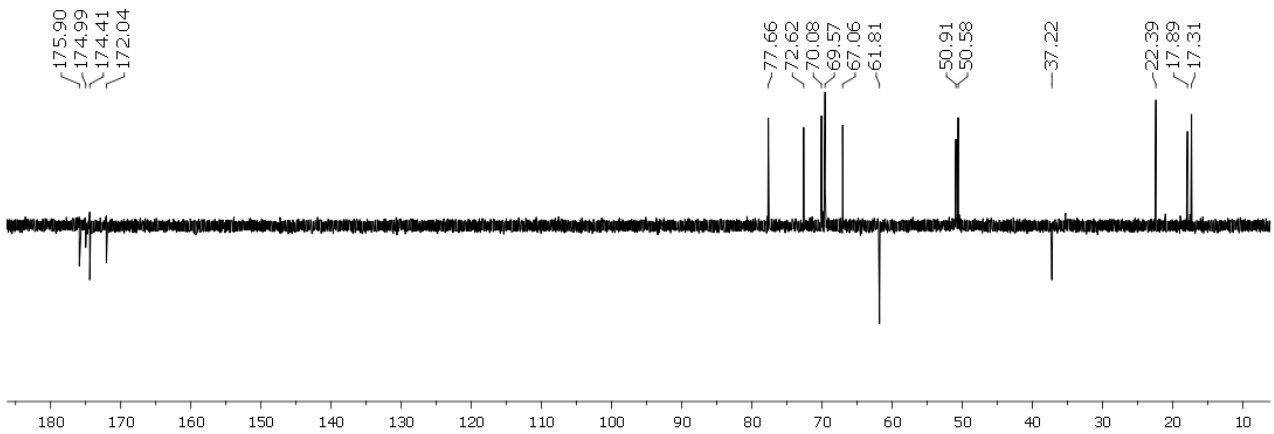
N^{α}-(L-*N*-Acetylalanyl)-N^{γ}-(α-D-galactopyranosyl)-L-asparagyl-L-alanine*N*-methylamide (113b)



Compound **111** (5 mg, 0.0077 mmol, 1 equiv.) was dissolved in a 4M solution of $MeNH_2$ in EtOH (150 µL, c = 0.05M). The reaction mixture was stirred for 2 h (TLC, 6:4 chloroform/methanol and 8:2 chloroform/methanol), then the solvent was evaporated at reduced pressure and stripped three times with methanol. An ammonium bicarbonate solution (1M) in methanol was added and the mixture was stirred overnight. The solvent was evaporated to afford **113b**, which was isolated by precipitation and centrifugation in MeOH:Et_2O 50 µL:250 µL (4mg, 95% yield). ^{1}H NMR (400 MHz, D_2O, 25 °C): δ = 5.59 (d, $J_{1,2}$ = 5.6 Hz, 1H, 1-H), 4.72 (dd, J = 5.6, J = 8.0 Hz, 1H, α-HAsn), 4.30-4.25 (m, 1H, α-H-Ala), 4.15-4.08 (m, 1H, α-H-Ala), 4.03 (dd, $J_{2,3}$ = 10.6, $J_{2,1}$ = 5.6 Hz, 1H, 2-H), 3.96 (d, $J_{4,3}$ = 3.4 Hz, 1H, 4-H), 3.82 (dd, $J_{3,2}$ = 10.6, $J_{3,4}$ = 3.4 Hz, 1H, 3-H), 3.76–3.70 (m, 1H, 6-H), 3.68–3.64 (m, 1H, 6'-H), 2.94 (dd, J_{gem} = 16.0, $J_{\alpha,\beta}$ = 5.6 Hz, 1H, β-H-Asn), 2.85 (dd, J_{gem} = 16.0, $J_{\alpha,\beta}$ = 8.0 Hz, 1H, β-H-Asn), 2.01 (s, 3H, NHAc), 1.35 (dd, J = 8.8 Hz, 6H, CH_3Ala) ppm. ^{13}C NMR (300 MHz, D_2O, 25 °C): δ = 174.4–172.2 (CO), 77.7 (C-1), 72.6 (C-5), 70.1 (C-3), 69.6 (C-4), 67.0 (C-2), 61.8 (C-6), 61.4 (α-C-Asn), 50.9 (α-CAla), 50.6 (α-C-Ala), 37.2 (β-CH_2-Asn), 22.4 ($NHCH_3$), 17.6 (CH_3Ala), 17.3 (CH_3Ala) ppm.



^{1}H-NMR spectrum of **113b** (CD_3OD, 400 MHz)



^{13}C-NMR spectrum of **113b** (CD_3OD, 100 MHz)

6.7 References

1 Colombo, C.; Bernardi, A. *Eur. J. Org. Chem.* **2011**, 3911–3919.

2 Santagada, V.; Caliendo, G. *Peptidi and peptidomimetici* **2003** Ed. Piccin, 98-103.

3 Meinjohanns, E.; Meldal, M.; Paulsen, H.; Dwek, R. A.; Bock K. *J.Chem.Soc. Perkin I* **1998**, 549-560.

4 Christiansen-Brams, I.; Meldal, M.; Bock K. *J. Chem. Soc. Perkin Trans. I* **1993**, 1461-1471.

5 Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B. *J. Am. Chem. Soc.*, **2004**, *126*, 8421-8425.

6 Sheppeck II, J. E.; Kar, H.; Hong, H. *Tetrahedron Letters* **2000**, *41*, 5329-5333.

7 Huang, W.; Groothuijs, S.; Heredia, A.; Kuijpers, B. H. M.; Rutjes, F. P. T. J.; van Delft, F. L.; Wang, L. *ChemBioChem* **2009**, *10*, 1234-1242.

8 Filser, C.; Kowalczyk, D.; Jones, C.; Wild, M. K.; Ipe, U.; Vestweber, D.; Kunz, H. *Angew. Chem. Int. Ed.* **2007**, *46*, 2108 –2111.

9 Peilstçcker, K.; Kunz, H. *Synlett* **2000**, 823–825.

10 Huang, W.; Groothuijs, S.; Heredia, A.; Kuijpers, B. H. M.; Rutjes, F. P. T. J.; van Delft, F. L.; Wang, L. *ChemBioChem* **2009**, *10*, 1234-1242.

11 Drenichev, M. S.; Kulikova, I: V.; Bobkov, G. V.; Tararov, V, I; Mikhailov, S. N. *Synthesis* **2010**, *22*, 3827-3834.

Chapter 7
Solid phase Synthesis of α-*N*-linked glycopeptides

7.1 Mimics of antifreeze glycopeptides

Procedures for coupling of the α-*N*-linked glucosyl and galactosyl building blocks in solution were described in Chapter 6, together with the synthesis of two small glycopeptides. For the synthesis of more complex structures, however, peptide synthesis in solution was not appropriate anymore, and it should be replaced by linear solid phase peptide synthesis (SPPS), employing the Fmoc-protected galactosyl asparagine building block **80**. The development of SPPS procedures and the synthesis of unnatural glycopeptides of general formula **120** that resemble natural antifreeze glycopeptides **119** (**Figure 1**) are described in this Chapter.

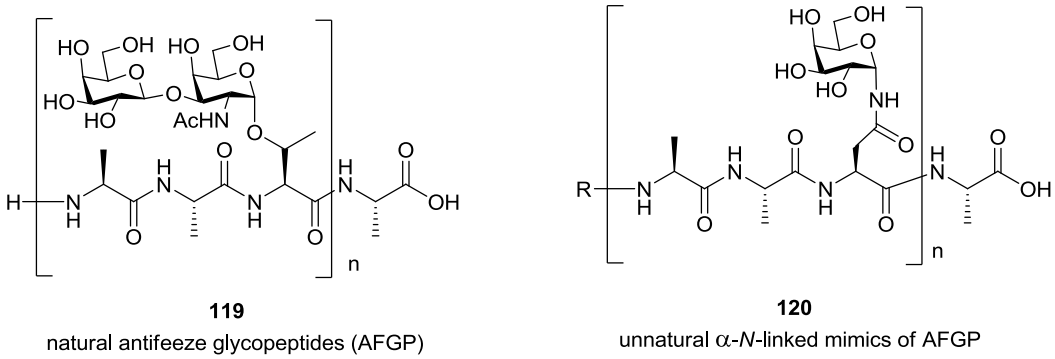


Figure 1. Unnatural α-*N*-linked glycopeptides **120**, described in this Chapter, in comparison to natural AFGPs **119**.

In Nature, antifreeze glycoproteins (AFGPs) **119** are mucin-like, peptide-based structures, composed of a repeating tripeptide unit (Ala-Thr-Ala) in which a disaccharide β-D-Gal-(1→3)-α-D-GalNAc is attached to the threonyl residue. They have a relative molecular mass range from about 2000 to 33000 ($4 \leq n \leq 55$)[1] and are contained in the blood serum of fish that live in the sub-zero Arctic and Antarctic oceans. In fact, these glycopeptides are essential to the survival of polar fish, since their main characteristic consists in the ability to prevent ice crystal growth (IRI = ice recrystallization inhibition)[2] and to decrease the blood freezing point, thus creating a hysteresis between the equilibrium melting point and the observed freezing point (TH = thermal hysteresis).[3] These remarkable molecules have potential applications in many area of agriculture and in frozen-food industry in which ice crystal growth is damaging.[4] The study of the mechanism of action of AFGPs has been challenging because of lack of access to pure samples from natural sources, and because no efficient method exist to produce them in large amounts. The first synthesis of AFGPs as pure glycoforms and the first related structure–activity studies on AFGPs were reported by Nishimura et al. in 2004.[5] These studies have highlighted the importance of hydrophobic interactions, *N*-acetyl group and Ala-Thr-Ala backbone, to antifreeze activity. In particular the TH activity was found to be related to the number of repeating units and hence to the length of the molecules, and reached a good value for n =5, 6, 7. These data were essential to understanding the

mechanism of action and to guiding the design of AFGP mimics. It has been recently hypothesized that this TH properties (i.e. the ability to selectively depress the freezing point of a solution relative to its melting point) could be beneficial in cryopreservation and hypothermic storage of biomedical materials and, thus, biological antifreezes have been explored as cryoprotectants.[6] Unfortunately, in recent experiments with AFGPs, cell damage has been reported at temperatures below the TH gap and large percentages of cells have been found to not survive in the freeze-thawing cycle.[7] However, the other property related to AFGPs, i.e. the inhibition of crystal growth during ice recrystallization (IRI), could show great potential in protecting cells from damage from ice recrystallisation during cryopreservation. In fact, AFGP analogues that are potent inhibitors of ice recrystallization but do not possess thermal hysteresis activity, have been recently reported by Robert Ben's group.[2, 8] Two unnatural *C*-linked galactosyl AFGPs **121a** and **121b** (**Figure 2**) in which the alanine residues have been replaced by glycine residues, have been synthesized and tested for IRI activities. These molecules turned out to be potent inhibitors of ice recrystallization and to protect embryonic liver cells from cryo-injury at millimolar concentration, increasing cell viability after cryopreservation. Furthermore, analogue **121a** showed little or no in vitro cytotoxicity, even at high concentrations.

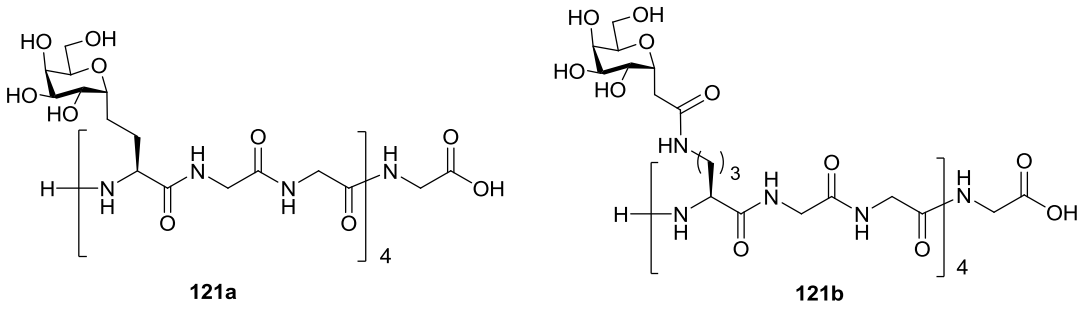


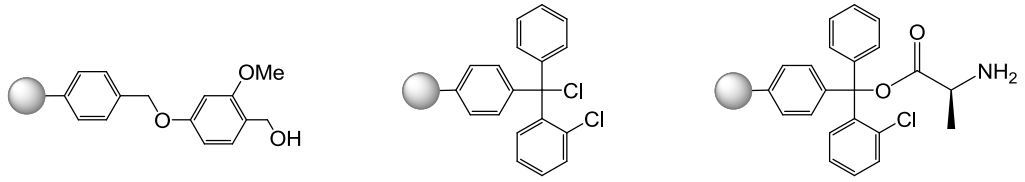
Figure 2. Unnatural *C*-linked glycopeptides **121** described by Ben as cryoprotectants.

Hence, the design and the synthesis of antifreeze mimics appeared to be a challenging field, which could give opportunities to elucidate and predict the antifreeze activity of new molecules. For this purpose, we decided to synthesize glycopeptides of general structure **120** (**Figure 1**) that contain:

a) Ala-Asn-Ala repeating units, with hydrophobicity similar to the natural AFGPs **119**, which also display alanine residues.
b) A galactose moiety instead of a Gal-GalNAc disaccharide, as in the unnatural *C*-linked glycopeptides **121a-b** reported by Ben;
c) An α-*N*-linked galactosyl asparagine in place of the Gal-GalNAc-threonyl residues of natural AFGPs **119**.

7.2 Solid phase synthesis strategies and initial trials

Solid phase peptide synthesis (SPPS), introduced by Merriefield almost fifty years ago,[9] has been significantly optimized and is now a widely employed technique for the synthesis of peptides and glycopeptides, both manually or with automated synthesizers. One of the advantages of this technique is the ease of purification, since the peptide is immobilized on solid resin and is retained during filtration processes, whereas liquid-phase reagents and by-products of synthesis are flushed away. SPPS is based on repeated cycles of coupling-wash-deprotection-wash, where each unit can be added in excess, since washing steps remove unreacted compounds. Currently, two methods are mostly used, the so called Boc protocol and the Fmoc protocol, that take their name from the protecting group used for the *N*-terminal amine of each (glycosyl) amino acid to be coupled. Since the Boc method utilizes acid for both deprotection and cleavage from the resin, we selected the Fmoc method, which was considered more apt to handle acid sensitive structures like α-*N*-linked glycosyl amino acids. In the Fmoc protocol, amino group deprotection is obtained with 20% piperidine solutions in DMF and the final cleavage can be performed using a low concentration of acid, provided that an acid sensitive resin is employed. For this purpose we selected 2-chlorotrityl (CTC) or Super Acid Sensitive (SASRIN) resins (**Figure 3**), two polystyrene-based resins used to immobilize the carboxylic function of C-terminal amino acid, that allow cleavage of the final product under very mild conditions.



SASRIN resin | 2-chlorotrityl resin | H-Ala-2-chlorotrityl resin

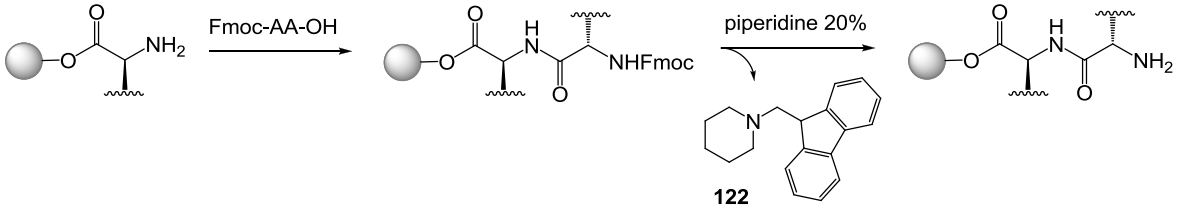
Figure 3. Solid phase synthesis resins employed.

In fact, 2-chlorotrityl resin can be cleaved with 1% TFA in DCM or with a mixture of trifluoroethanol/AcOH/DCM (2:2:6), or with 30% HFIP (hexafluoroisopropanol) in DCM, allowing to retain even acid labile side chain protecting groups. Even the use of scavengers as triethylisopropylsilane (TIS) during this types of cleavage is not mandatory because the amount of acid is low and carbocations are not generated in solution. Scavengers are recommended only when the peptide contains free hydroxy or thiol groups. For storage of this resin, the Fmoc group of the last (glycosyl) amino acid has to be removed since the free amino 2-chlorotrityl resin has better stability. 2-chlorotrityl resin is also commercially available preloaded with alanine (**Figure 3**).

Cleavage from SASRIN resin is carried out similarly to CTC, with a low concentration of acid (1-2% TFA in DCM) and the resin can be stored as Fmoc-protected at the last (glycosyl) amino acid.

These resins appear to be ideally suited for the synthesis of acid-sensitive molecules like glycopeptides. SASRIN, was in fact used for the synthesis of acid-labile cyclic pseudopentapeptides in our laboratories[10] and for the synthesis of linear orthogonally protected decapeptides, which were further applied for conjugation to sugars.[11] CTC has often been used as the support for solid-phase-synthesis of both *O*-linked and *N*-linked glycopeptides: in the first case, for instance, CTC was used for a recent synthesis of MUC1 glycopeptides.[12] In the second case it was employed for the synthesis of the peptide of CD52 antigens, which consists of 12 amino acids and has one *N*-glycosylation site, to which a complex fucosylated glycan is attached.[13]

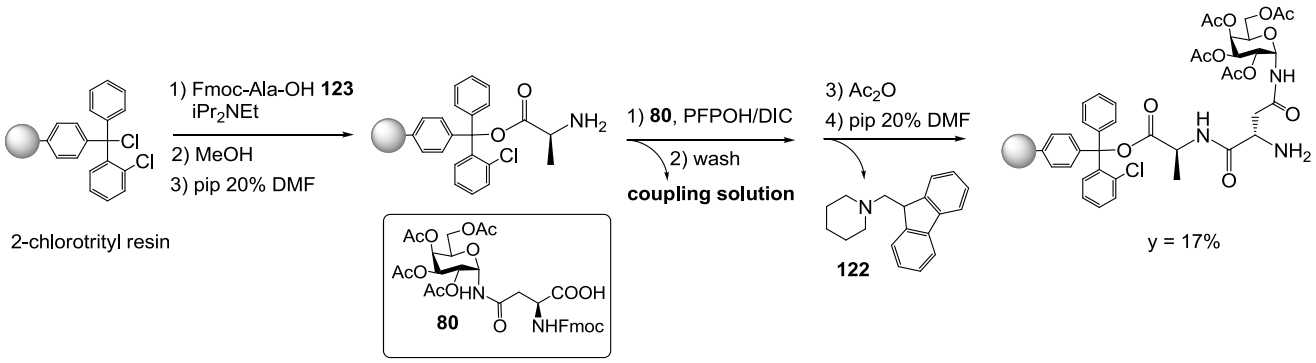
Initially, we performed some trials with both CTC and SASRIN resins, as illustrated below, but since they were found to behave similarly and no particular advantage was observed in the use of one rather than the other, we finally decided to proceed with the CTC resin in the Ala functionalized form (H-Ala-CTC, **Figure 3**). As normal in the Fmoc protocol, yields of each coupling step were estimated by measuring spectrometrically the UV absorbtion of the Fmoc-piperidine adduct **122** (λ = 300 nm, ε = 7800), released after the deprotection of the Fmoc group of the unit just coupled.



Scheme 1. Procedure for measuring coupling yields through UV absorbtion of **122**.

Conditions for the solid phase synthesis were first explored relying on the coupling methodologies identified previously (Chapter 6) and using α-*N*-Fmoc-galactosyl asparagine **80** as the limiting reactant, to establish the efficiency of the coupling conditions and to estimate the effect of the cyclization process (Scetion 6.2) on coupling conditions.

The first trial was performed with CTC, which was loaded with Fmoc-Ala-OH **123** in the presence of *i*Pr_2NEt for 1h (**Scheme 2**) in DCM:NMP 85:15. Methanol was added to cap the unreacted Cl sites. After deprotection with 20% piperidine in DMF, *N*-Fmoc-building block **80** (1 eq.) was preactivated at 0°C with PFPOH/DIC (1eq.) in NMP:DCM 1:1 (0.06M) and coupled at room temperature for 4h (**Scheme 2**). Capping was performed with acetic anhydride 1M in DMF.



Scheme 2. Coupling of **80** with PFPOH/DIC in solid phase synthesis on CTC resin.

The yield estimated using UV absorbtion of **122** after Fmoc-removal was only 17% based on **80**. The coupling solution (**Scheme 2**) was dissolved in Et_2O, washed with water to remove most of the NMP, and examined by ^{1}H NMR (**Figure 4**).

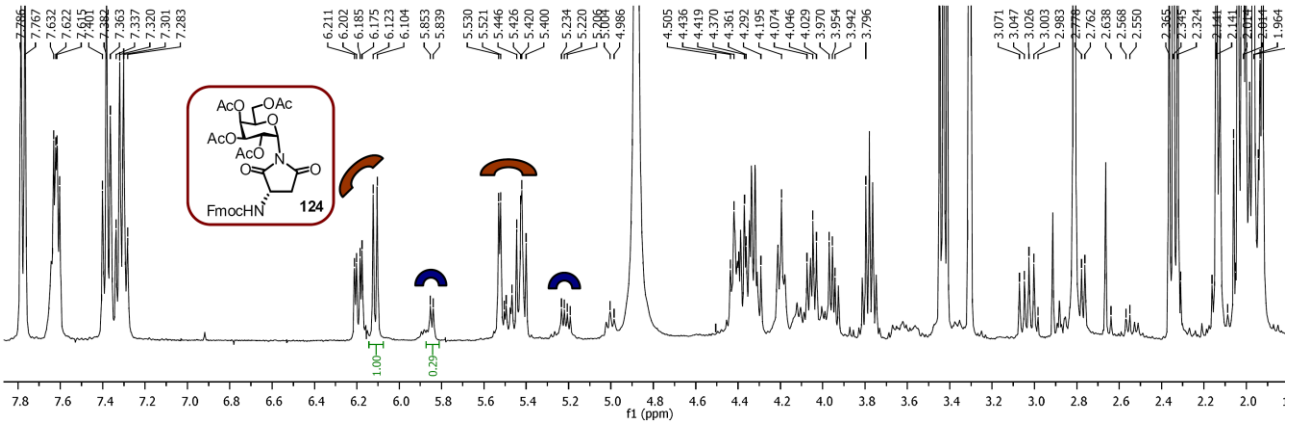
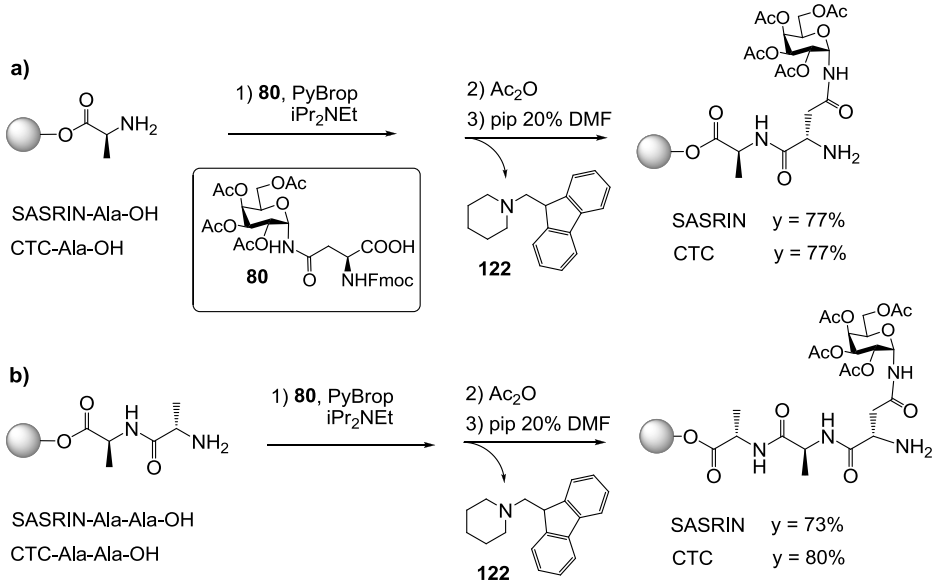


Figure 4. ^{1}H NMR spectrum of the coupling solution of **80** with PFPOH/DIC after 4 h of reaction in solid phase.

A large amount of *N*-galactosyl immide **124** was observed (major signals, in red), together with a minor amount of another by-product (blue) with signals similar to unreacted **80** (possibly the corresponding activated compound). No increase of the yield was observed by extending the reaction time from 4 h to 12 h. Formation of imide **124** under these coupling conditions was enhanced, relative to the solution phase (Chapter 6), by the difficult reaction between the solid-phase attached peptide and the hindered galactosyl amino acid dissolved in solution. Reaction of **80** (limiting reactant) with SASRIN or CTC resins, preloaded with alanine, employing PyBrop/iPr_2NEt (1:1 DMF:CH_2Cl_2, c = 0.1 M, 2h at 0°C, then overnight at room temperature) gave 77% of coupling yield (**Scheme 3a**). So, in this case, the result obtained in solution can be extended to SPPS, confirming the low tendency of the PyBrop derived acyl phosphonium salt to cyclise, and allowing to select PyBrop as the reagent of choice for solid phase coupling of α-*N*-linked galactosyl asparagine **80**.



Scheme 3. Coupling of **80** with PyBrop/*i*Pr$_2$NEt on CTC or SASRIN preloaded resins.

An additional trial, performed using CTC or SASRIN resins preloaded with an Ala-Ala dipeptide, showed that coupling yields were almost the same and no beneficial effect was observed by spacing the reacting amine from the resin (**Scheme 3b**). We also observed that it was not necessary to preactivate the acid for 2 h at 0°C, but mixing the acid with PyBrop at 0°C, just before addition to the resin, worked just as well. Finally, we chose to proceed with the H-Ala-CTC resin, which can be purchased already functionalized, with an average loading of 0.4-0.9 mmol/g. We spectrophotometrically determined the initial loading after reaction with **80** and Fmoc deprotection. Loadings obtained under different reaction conditions are shown in **Table 1**.

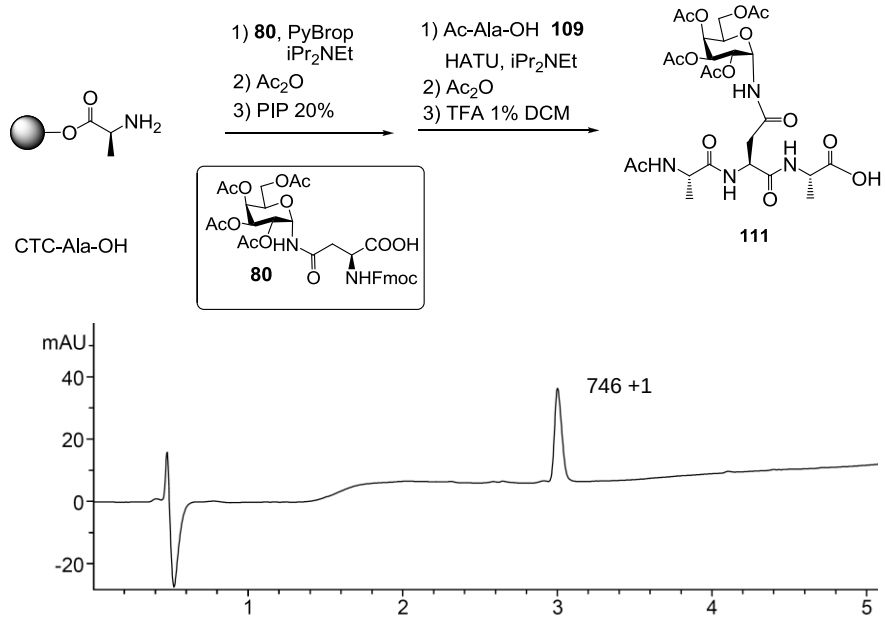
Table 1. Conditions for loading of the first galactosyl amino acid **80** on H-Ala-CTC.

Entry	H-Ala-CTC (mg)[a]	Equivalents (**80**) [b]	Reaction time (h) [N° of Cycles] [c]	Solvent	[M] (**80**)	Loading [d] (mmol/g)
1	200 mg	0.8-1.7	8 +12 [2]	CH_2Cl_2:DMF 1:1	0.07 M	0.4
2	230 mg	0.8-1.4	8 +12 [2]	CH_2Cl_2:DMF 1:1	0.1 M	0.5
3	130 mg	1.2-2.8	12 [1]	CH_2Cl_2:DMF 1:1	0.16 M	0.5

[a] Commercially available with average loading 0.4-0.9 mmol/g. [b] Equivalents relative to the average loading of H-Ala-CTC. The first number refers to a loading of 0.4mmol/g, the second one to a loading of 0.9 mmol/g. [c] The activation was done with **80**/PyBrop/ *i*Pr$_2$NEt at 0°C, then the reactions were conducted at room temperature. [d] Calculated by UV absortion after Fmoc removal.

We determined that, starting from commercially available Ala-CTC, a loading of 0.5 mmol/g could be obtained in a single overnight cycle, using 1.2-2.8 equivalents of **80** in a 0.16 M solution of 1:1 DMF: CH_2Cl_2 (entry 3). Reducing the number of equivalents and/or the concentration of the

activated acid (entries 1 and 2) required two cycles to obtain the same yield. After setting the initial loading of Ala-(α-*N*-Gal)Asn to 0.5 mmol/g, we performed the condensation with Ac-Ala-OH **109** (5 eq.) with HATU in 2 cycles of 4h at room temperature (**Scheme 4**). Capping with Ac_2O and cleavage (1% TFA in CH_2Cl_2) gave glycopeptide **111**.[a] The reaction crude could be purified by automated chromatography on a reverse phase column (C-18 Biotage cartridge; eluant: H_2O:CH_3CN; gradient 0%→40% CH_3CN in 8 min; tr = 6 min) to obtain **111** in 95% yield. LC-MS analysis (gradient: CH_3CN from 0 % to 30 % in 6 min) revealed the presence of a unique compound (**Scheme 4**).

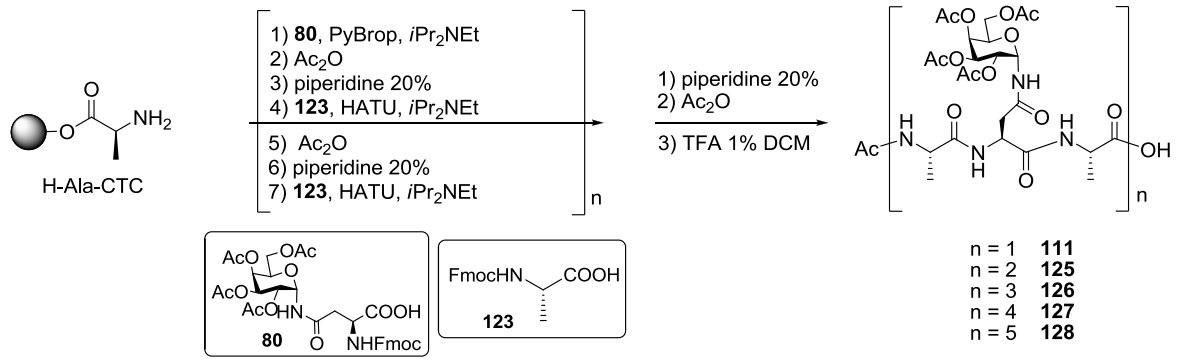


Scheme 4. Synthesis of **111** by solid phase synthesis and LC-MS analysis of the purified peptide.

[a] After cleavage with trifluoroacetic acid, compound **111** should be characterized in D_2O and not in CD_3OD (some formation of methylester was observed).

7.3 Solid phase synthesis of α-*N*-linked galactosyl glycopeptides

Having determined the coupling conditions for the α-*N*-linked galactosyl building block **80** in the solid phase synthesis of the small glycopeptide **111**, we could address the synthesis of the AFGP mimics **120 (Figure 1)**. The general procedure was based essentially on a repeating sequence of reactions, which consisted in coupling (Fmoc protocol) the α-galactosyl amino acid **80**, followed by two alanine residues, up to a maximum of five times (**Scheme 5**).



Scheme 5. Sequence for the solid phase synthesis of glycopeptides **125-128**.

After deprotection of the Fmoc-group from the last alanine residue and acetylation, cleavage with 1% TFA in DCM afforded compounds **111-128** with 1 to 5 repeating units. As seen for **111**, in the first coupling the quantity of galactosyl amino acid **80** was adjusted to obtain a standard loading of 0.5 mmol/g. All subsequent coupling reactions were monitored spectroscopically, thus providing the yield of every condensation step. Generally, couplings with Fmoc-Ala-OH were complete when performed with 5 equivalents of amino acid (c = 0.5 M in DMF) in 2h, for two cycles. Critical steps were instead those involving the galactosyl amino acid **80**, because its hindered structure slowed down the reaction significantly. Furthermore, with increasing peptide length, problems of accessibility or aggregation appear to occur (aggregation usually has a tendency to occur in a sequence of hydrophobic residues, starting from the 5/6th residues from the C-terminus). As a consequence, the reaction time for galactosyl-asparagine couplings became long (8-12h) and double or triple coupling cycles had to be performed to achieve reasonable yields. Reaction conditions are shown in **Table 2**. The data show that completion of the coupling reaction is indeed critical and the efficiency is deeply dependent on concentration, number of equivalents used, number of cycles and above all on the length of the sequence (number of the repeat in which the coupling was performed). For the second repeat, the best conditions, which led to 95%, yield, are reported in entry 3. They required 1.5 equivalents of **80** as a 0.15 M solution for three long cycles of 8 h or 12 h each. It could be noted that even more equivalents and more cycles are required for the coupling of **80** in the third, fourth and fifth repeats (entry 4-7). In these cases the conditions we used weren't

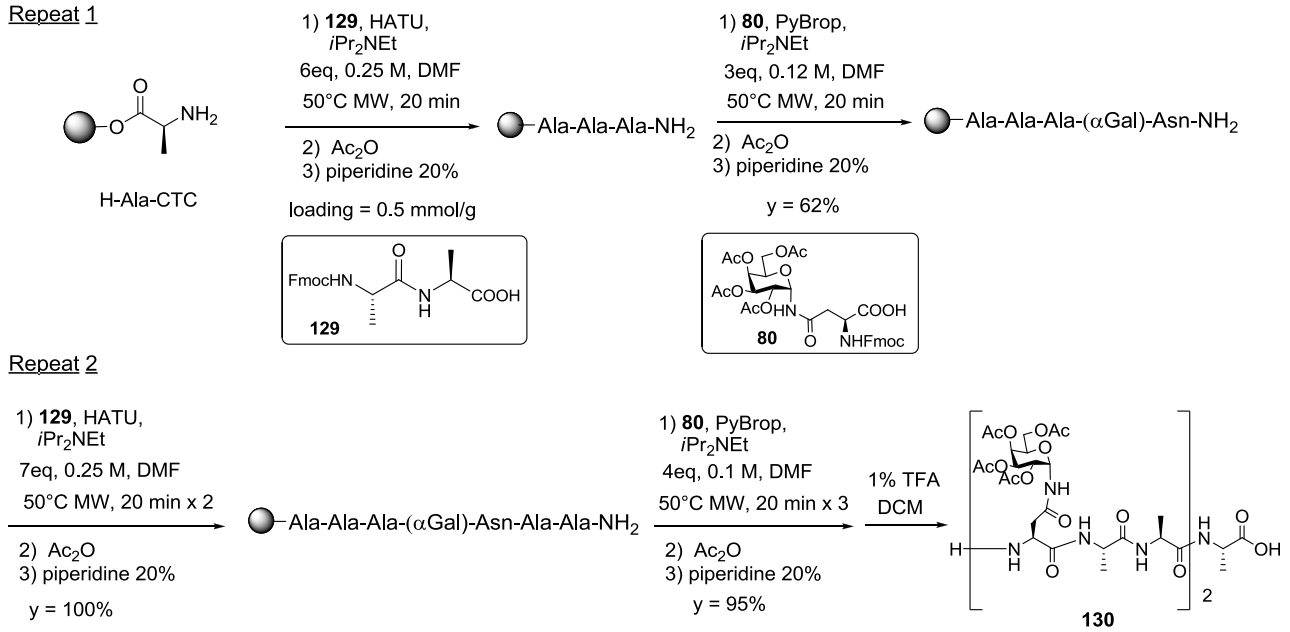
effective to reach completion of the couplings, but yields between 80% and 90% were obtained, depending on the exact conditions. This resulted in complex mixtures of peptide products, in which the truncated sequences were difficult to separate (see Chapter 8 for the purification of all glycopeptides). As a general consideration, probably the coupling performance in the third, fourth and fifth repeats could be optimized further by increasing the number of equivalents and the concentration of **80**.

Table 2. Reaction conditions for the coupling of Fmoc-protected galactosyl asparagine **80**.

Entry	Repeat (n) [a]	Equivalents of **80**	Cycles	Reaction time (h)	[M] in DMF	Yield(%) [b]
1	2	1.2	2	8 + 12	0.1	64
2	2	1.5	2	8 + 12	0.1	80
3	2	1.5	3	8 + 12 + 8	0.15	95
4	3	1.5	3	8 + 12 + 8	0.15	80
5	3	2	2	8 + 12	0.15	83
6	4	2	3	8 + 12 + 8	0.15	80
7	5	3	3	8 + 12 + 8	0.15	90

[a] Identifies the position of the coupled residue. n = 2 means that the α-galactosyl amino acid is part of the repeat sequence n=2. i.e. it has been coupled to the last alanine of the n=1 sequence. [b] Yield determined by UV spectroscopy after Fmoc-removal.

Globally, the major limitations for these sequences of reactions (**Scheme 5**) consist in the efficiency of the coupling of the glycosyl amino acid, in the number of steps, the number of cycles and in the time required for each step. Recently, to improve on these points, we have synthesized a Fmoc-Ala-Ala-OH building block **129**,[14] which should be useful to reduce the number of steps, and we have experimented the use of microwave irradiation on the coupling reactions (**Scheme 6**). The use of microwave irradiation is supposed to accelerate coupling reactions, not only for the increase in temperature, but also due to the alternating electric field of the microwave, to which the polar backbone of the glycopeptides tries to align. This can lead to decreased steric hindrance, prevention of chain aggregation and easier access of the reagents to solid phase matrix.[15] Microwave-assisted solid phase synthesis has been recently reported by Sewald et al. for the synthesis of an antifreeze glycopeptides analogue,[16] on a 2-chlorotrityl resin. In particular, since the chlorotrityl resin is supposed to be quite sensitive, these authors performed all the steps under microwave irradiation at a maximum of 20 W and 40 °C. For our tests on microwave-assisted coupling, we started with H-Ala-2-chlorotritylresin, using Fmoc-Ala-Ala-OH **129** and the galactosyl amino acid **80** as building blocks (**Scheme 6**).



Scheme 6. Solid phase synthesis with microwave irradiation.

For an initial loading of 0.5 mmol/g, 6 equivalents of Fmoc-Ala-Ala-OH **129** were used in a coupling step of 20 minutes at 50°C, under microwave irradiation and vortexing of the vial. The galactosyl amino acid **80** was then coupled (3 equivalents, as 0.12 M solution in DMF x 1 cycle) at 50°C for 20 min affording a coupling yield of 62%. Then, the second Fmoc-Ala-Ala-OH residue was coupled in a total time of 40 minutes (5 equivalents x 2 cycles of 20 minute at 50 °C) with quantitative yield. The second galactosyl amino acid **80** was coupled in 95% yield with a total time of 1 h performing 3 cycles with 4 equivalents, 0.1 M in DMF, at 50°C for 20 min. After deprotection of the Fmoc-group from the last resisue, cleavage with 1% TFA in DCM afforded compounds **130** in 58% yield. LC-MS analysis of the crude reaction mixture (gradient: CH_3CN 0% → 30% in 6 min) shows the predominant presence of the desired compound [M] (**Figure 5**).

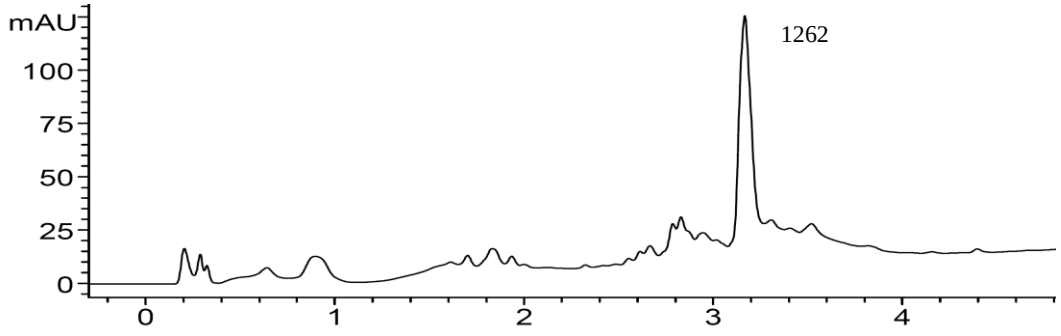


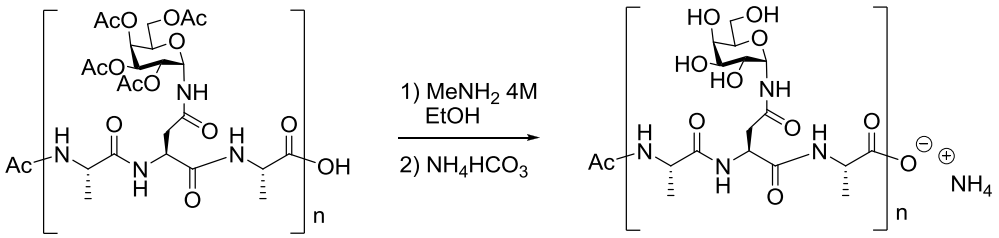
Figure 5. LC-MS analysis of the reaction mixture crude of **130** after microwave-assisted solid phase synthesis

These initial experiments represent a starting point for an improved synthesis of these unnatural compounds which limits the number of steps with the use of the dipeptide Fmoc-Ala-Ala-OH **129** as a building block and shortens the reaction time, improving the coupling yields, using microwave

assisted conditions. The next steps to be addressed and optimized consist in the purification of the *O*-acetyl-glycopeptides synthesized, in the deacetylation of the sugar residues and in the isolation of the final compounds. These steps are somewhat interconnected, and the techniques that can actually be used depend both on the pH stability of the compounds and on the initially purity of the batch released from the resin. In the case at hand, an initial purification of these products was achieved after the cleavage using reverse phase chromatography of the *O*-acetyl-glycopeptides and a second purification was performed by HILIC chromatography after *O*-acetyl removal (Section 7.4). The result of this work, which allowed to establish an efficient isolation protocol, able to rescue pure glycopeptides also from relatively complex mixtures like those obtained from the coupling sequence described in this chapter, are reported in Chapter 8, devoted to the purification and isolation techniques. To conclude this chapter on the synthetic aspects of the project, we outline in section 7.4 the method selected for acetyl groups removal from the sugar moieties.

7.4 *O*-acetyl groups removal from α-*N*-linked galactosyl glycopeptides

Acetyl removal from molecules **125-128** was obtained using a 4M solution of $MeNH_2$ in EtOH and a substrate concentration of 0.05 M for 5h. (**Scheme 7**), as described in Section 6.4.[17] No trace of furanose formation, nor pyranose anomerization were observed. The crude remained contaminated by a large excess of N-methylacetamide (AcNHMe) produced by the reaction. This by-product can generally be removed from deprotection mixtures by co-evaporation with high-boiling solvents.[18] In this case, though, AcNHMe could not be eliminated in this way, possibly because it is stabilized by interaction with the peptides, and was removed chromatographically, as described in the following chapter.



n = 2 **125**
n = 3 **126**
n = 4 **127**
n = 5 **128**

n = 2 **131**
n = 3 **132**
n = 4 **133**
n = 5 **134**

Scheme 7. Deprotection of α-*N*-linked galactosyl glycopeptides with 4M $MeNH_2$-EtOH solution.

7.5 Conclusion

Solid-phase synthesis (Fmoc protocol) conditions were explored using CTC and SASRIN resins. PyBrop was found to be the reagent of choice for the activation of α-*N*-linked galactosyl amino acid **80**. The synthesis of more complex α-*N*-linked glycopeptides has thus become feasible and conditions were optimized for their solid phase synthesis. In particular, we focused on the synthesis of mimic of antifreeze glycopeptides, which displayed repeating units of general structure (Ala-Asn (α-*N*-Gal)-Ala). The major problems were observed for the coupling of our galactosyl building block, which, occurring frequently in the sequence, resulted in inefficient couplings and in long reaction time, due principally to its steric hindrance. The possible solutions to improve the coupling efficiency could be:

a) The use of more equivalents of the galactosy building block, in a more concentrated coupling solution (> 0.15 M)

b) The use of microwave-assisted solid phase synthesis, which seemed promising in accelerating the coupling process.

c) The use of the dipeptide Fmoc-Ala-Ala-OH as a building block, which helps by limiting the number of coupling steps to be performed on the resin.

Despite the fact that for the moment the desired α-*N*-linked glycopeptides were obtained in small amounts, principally due to the low efficiency of the coupling steps, the developed methods allowed understanding the behaviour of these molecules, in terms of reactivity and stability and allowed to determine methods for their purification (Chapter 8).

Moreover, the synthesis of these neo-glycoconjugates allowed the study of some of their properties, as conformation and interaction with lectins, both with computational and NMR techniques (Chapter 9).

7.6 Experimental Section

Solvents were dried by standard procedures: dichloromethane (DCM), methanol, *N*,*N*-diisopropylethylamine (DIPEA) were dried over calcium hydride; THF was distilled from sodium, *N*,*N*-dimethylformamide (DMF) and *N*-methylpyrrolidone (NMP) were dried with activated molecular sieves (3Å). 2-chlorotrityl, SASRIN and H-Ala-2-chlorotrityl resins were purchased from Bachem. UV/Vis spectra were recorded on an Agilent 8453 instrument. ^{1}H- and ^{13}C-NMR spectra were recorded at 400 MHz on a Bruker AVANCE-400 instrument. Chemical shifts (δ) for ^{1}H and ^{13}C spectra are expressed in ppm relative to internal Me_4Si as standard. Mass spectra were obtained with a Bruker ion-trap Esquire 3000 apparatus (ESI ionization) or Ft-ICR Mass spectrometer APEX II & Xmass 4.7 Magnet software (Bruker Daltonics). HPLC-MS analyses were performed with Agilent 1100 with quaternary pump, diode array detector, autosampler, thermostated column holder coupled with MS: Bruker ion-trap Esquire 3000 with ESI ionization. The HPLC column was a Waters Atlantis 50x4.6 mm, 3 µm.

Method: Phase A: Milli-Q water containing 0.05 % (v/v) TFA.
Phase B: Acetonitrile (LC-MS grade) containing 0.05 % TFA.

Flow: 1 mL/min, partitioned after UV detector (50 % to MS ESI), Temperature: 40°C.

Gradient: from 0 % B to 30 % B in 6 min, from 30 % B to 90 % B in 1 min, washing at 90 % B for 1 min, equilibration at 0 % B in the next 3 min.

Solid phase reaction were performed in Biotage Isolute column reservoirs (3mL, 6mL, 15 mL) using Heidolph Vibramax 110 shaker. Microwave-assisted solid phase synthesis was performed with a Biotage Initiator + SPWave apparatus, equipped with a variable vortex mixer (300-1300 rpm). The reactor was a proprietary 2 mL vial (ideal volume: 0.8-1.1 mL; a minimum volume of 600 µL of solvent is required to allow temperature monitoring) and the mixer was set at 1100 rpm.

In this experimental section we have reported the general procedures used for solid phase synthesis of α-*N*-linked glycopeptides. The activation and coupling of Fmoc-Ala-OH **123** and of galactosyl amino acid **80** are reported in **Procedure 5** and **6**. However, Procedure 5 and 6 gave different outcomes, depending on the number of equivalents used, concentration, number of cycles, number of repeat, etc… As a consequence, for each derivative **125-128** we described the actual conditions applied and the yield obtained. The purification of the resulting peptide from the heterogeneous mixtures of truncated sequences is described in **Chapter 8**, together with the characterization of the final compounds.

Procedure 1:

Determination of the amount of coupled (glycosyl) amino acid and of the resin loading

After the (glycosyl) amino acid coupling described below and after washing thoroughly the resin with DMF, Fmoc removal was performed using a 20% piperidine solution in DMF. Generally the operation was repeated twice for 30 minutes. Each time the deprotection solution was collected in a measuring tube. The resin was washed six times with DMF, also collecting the solution. From the resulting solution of deprotection and washes (x mL), 100 μL were taken and added to 2.90 mL of DMF in a 3 mL UV-cuvette. Absorbance (Abs) at 300 nm was measured, using DMF as the blank. The μmoles of (glycosyl) amino acid loaded on the resin are calculated with **Formula 1**.

Formula 1: μmol = (Abs · x · 30)/ 7.8

Abs (300 nm);
x = mL collected from Fmoc-removal and washes;
30 = for 3 mL cuvettes;
7.8 = ε

The resulting loading of the resin is calculated with **Formula 2**.

Formula 2: Loading (mmol/g) = μmol **(Formula 1)**/ mg resin

The yield of the coupling step is the ratio between the loaded μmol of residue n and the μmol of residue n-1 (previous coupling) (**Formula 3**).

Formula 3: y = μmol (n)/ μmol (n-1)

Procedure 2:

Loading of the first amino acid on 2-chlorotrityl resin

2-chlorotrityl resin (300 mg) was swollen in DMF for 30 minutes. Fmoc-Ala-OH · H_2O **123** (55 mg, 0.165 mmol) and DIPEA (171 μL, 1 mmol) were dissolved in 2 mL of a 85:15 CH_2Cl_2:NMP solution and added to the resin. The reaction mixture was shaken at room temperature for 1h, then MeOH was added and the mixture shaken for 10 min. Wash steps were performed with DMF and Fmoc removal was performed with 20% piperidine in DMF (15 min, twice). Absorbance (Abs) at 300 nm, using DMF as blank, was measured as described above and an intial loading of 0.53 mmol/g (160 μmol of alanine/300 mg of resin) was obtained. After Fmoc deprotection the resin must be washed thoroughly with DMF to eliminate all traces of piperidine. The 2-chlorotrityl resin is best stored with the amino acid free at the *N*-terminus.

Procedure 3:

Loading of the first amino acid on SASRIN resin

Fmoc-Ala-OH · H_2O **123** (404 mg, 1.224 mmol, 4 eq) and DMAP (4 mg, 0.030 mmol, 0.1 eq) were dissolved in 6 mL of DMF at -20°C. DIC (194 μL, 1.255, 4.1 eq) was added dropwise and the mixture was shaken for 20 min. SASRIN resin (300 mg, 1.02 mmol/g, 0.300 mmol, 1 eq) was added and the reaction mixture was shaken for 1h at -20°C, then at 0°C for 4h. The resin was filtered and washed with DMF. Capping was performed with a solution of Ac_2O (32 μL, 0.300 mmol, 1eq) and DMAP (4 mg, 0.030 mmol, 0.1 eq) in DMF (2 mL) for 1h, then the resin was filtered and washed with CH_2Cl_2. A loading of 0.70 mmol/g was obtained (see **Formula 1** and **2**). The SASRIN resin can be stored with the last amino acid Fmoc-protected.

Synthesis of *α*-*N*-linked glycopeptides on H-Ala-2-chlorotrityl resin

Procedure 4

Loading of the first galactosyl amino acid 80 on H-Ala-2-chlorotrityl resin

Compound **80** (110 mg, 0.161 mmol, 08-1.4 eq) and PyBrop (150 mg, 0.322 mmol) were dissolved in a DMF/CH_2Cl_2 solution (1:1, 1.6 mL) at 0 °C. DIPEA (168 μL, 0.966 mmol) was added and the reagents were stirred at 0 °C for 15 min, then added to H-Ala-2-chlorotrityl resin (230 mg, loading 0.4–0.9 mmol/g), previously swollen in DMF (30 min). The reaction mixture was shaken for 8h at room temperature, then the resin was filtered and the cycle (activation of **80** with PyBrop and DIPEA at 0°C, then addition to the resin) was repeated. The reaction mixture was shaken overnight at room temperature, then the resin was filtered and washed with DMF. Capping was performed twice with a 1 M Ac_2O solution in DMF (2 mL) for 30 min. After Fmoc removal under standard conditions (20% piperidine in DMF; 30 min twice), loading was estimated by measuring the UV absorption at 300 nm. Loading = 0.5 mmol/g. After Fmoc deprotection the resin must be washed thoroughly with DMF to remove all traces of piperidine.

Procedure 5

General method for Fmoc-Ala-OH 123 coupling

H-Ala-2-chlorotrityl resin was swollen in DMF for 30 minutes. Fmoc-Ala-OH · H_2O **123** (5 eq) and HATU (5 eq) were dissolved in DMF (c = 0.5 M) at 0 °C. DIPEA (7 eq) was added and the reagents were stirred at 0 °C for 15 min, then added to H-Ala-2-chlorotrityl resin. The reaction mixture was shaken at room temperature for 2h, then the resin was filtered and the cycle (activation of **123** with HATU and DIPEA at 0°C, then addition to the resin) was repeated once again. After 2 h of reaction

at room temperature, the resin was filtered and washed with DMF. Capping was performed with a 1 M Ac_2O solution in DMF (twice for 30 min). Fmoc removal was performed under standard conditions (20% piperidine in DMF, twice for 30 min), then the resin was washed thoroughly with DMF to remove all traces of piperidine.

Procedure 6

General method for galactosyl amino acid 80 coupling

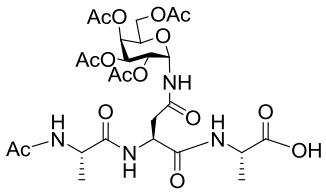
Compound **80** (x mmol) and PyBrop (1.5·x mmol) were dissolved in a solution of DMF at 0 °C. DIPEA (4·x mmol + mmol of resin) was added and the reagents were stirred at 0 °C for 15 min, then added to H-Ala-2-chlorotrityl resin, previously swollen in DMF (30 min). The reaction mixture was shaken for 8h at room temperature, then the resin was filtered. The cycle (activation of **80** with PyBrop and DIPEA at 0°C, then addition to the resin) was repeated and the reaction mixture was shaken overnight at room temperature. If another cycle was necessary, the mixture was shaken for 8h. The resin was filtered and washed with DMF. Capping was performed with a 1 M Ac_2O solution in DMF (twice for 30 min). After Fmoc removal under standard conditions (20% piperidine in DMF; 30 min twice), the resin was washed thoroughly with DMF to remove all traces of piperidine.

Procedure 7

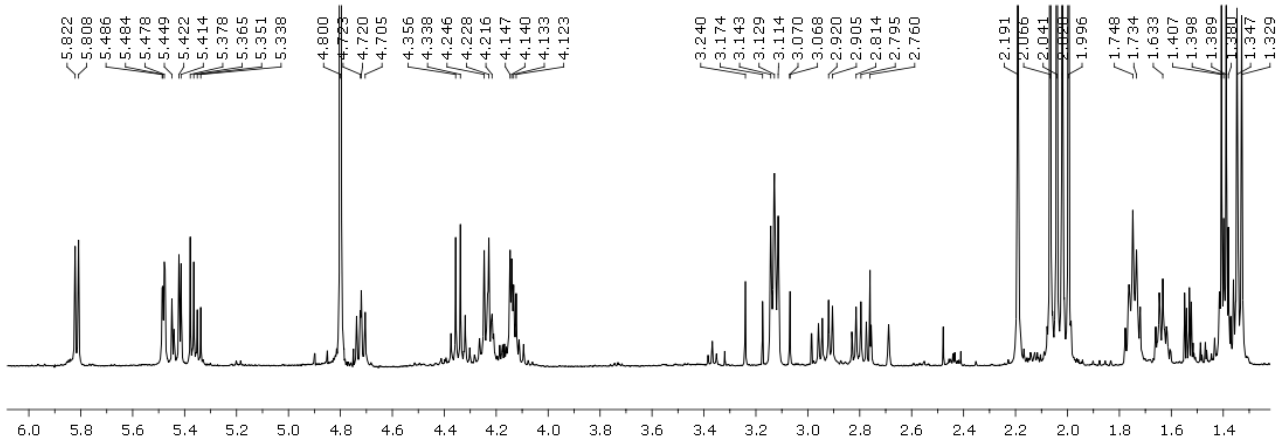
General method for cleavage from 2-chlorotrityl resin

After coupling and capping of the last amino acid, the resin was washed thoroughly with DCM to eliminate DMF traces (minimum 10 times). Then a solution 1% TFA in DCM was added to the resin and the mixture was shaken for 1 minute (the resin becomes light purple), then the solution was collected. The operation was repeated 4-5 times (the resin became dark purple) each time shaking the resin for 5 minutes and collecting the resulting solution. After evaporation, the crude was repeatedly dissolved in toluene and stripped at reduced pressure. Finally the residue was dissolved in water and lyophilized.

_Nα_-(L-_N_-Acetylalanyl)-_Nγ_-(2,3,4,6-tetra-_O_-acetyl-α-D-galactopyranosyl)-L-asparagyl-L alanine (111)



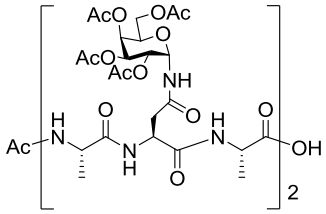
H-Ala-2-chlorotrityl resin (130 mg, loading 0.5–0.9 mmol/g) was swollen in DMF. Then the resin was suspended in a solution of DMF/CH_2Cl_2 (1:1 900 μL) and compound **80** (100 mg, 0.146 mmol) and PyBrop (136 mg, 0.292 mmol) were added. DIPEA (152 μL, 0.876 mmol) was added at 0 °C (Conditions reported in Table 1, entry 3, Section 7.3). The reaction mixture was shaken at 0 °C for 2 h and then overnight at room temperature. Capping was performed with a 1 M Ac_2O solution in DMF (1ml for 30 min, twice). After Fmoc removal under standard conditions (20% piperidine solution in DMF), μmoles of glycosyl amino acid loaded (63 μmoles) and loading of the resin (0.5 mmol/g) were calculated (Formula 1 and 2). Pre-activation of Fmoc-Ala-OH (103 mg, 0.315 mmol), HATU (120 mg, 0.315 mmol) and DIPEA (31 μL, 0.18 mmol) and coupling were performed as reported in **Procedure 5**. After Fmoc removal, *N*-terminus protection was performed with a solution of 1 M Ac_2O in DMF (2h). Cleavage was performed as reported in **Procedure 7**, to obtain 63 mg of crude (^{1}H NMR). After purification on reverse phase (C-18 Biotage cartridge, eluants: H_2O/CH_3CN 0.1% TFA, gradient: 5%→40% CH_3CN in 8 min, 40%→80% CH_3CN in 4 min, flux = 15 mL/min, tr = 6min) 39 mg of **111** (95% yield) were obtained.



^{1}H-NMR spectrum of the crude **111** after the cleavage (D_2O, 400 MHz)

The LC-MS of **111** is reported in **Scheme 4**.

Synthesis of α-*N*-linked glycopeptide 125 (n = 2)



The first alanine residue is already loaded on the commercial resin (H-Ala-2-chlorotrityl resin). Loading of the first galactosyl amino acid **80** was performed as reported in **Procedure 4**. The other steps are reported in **Table A**.

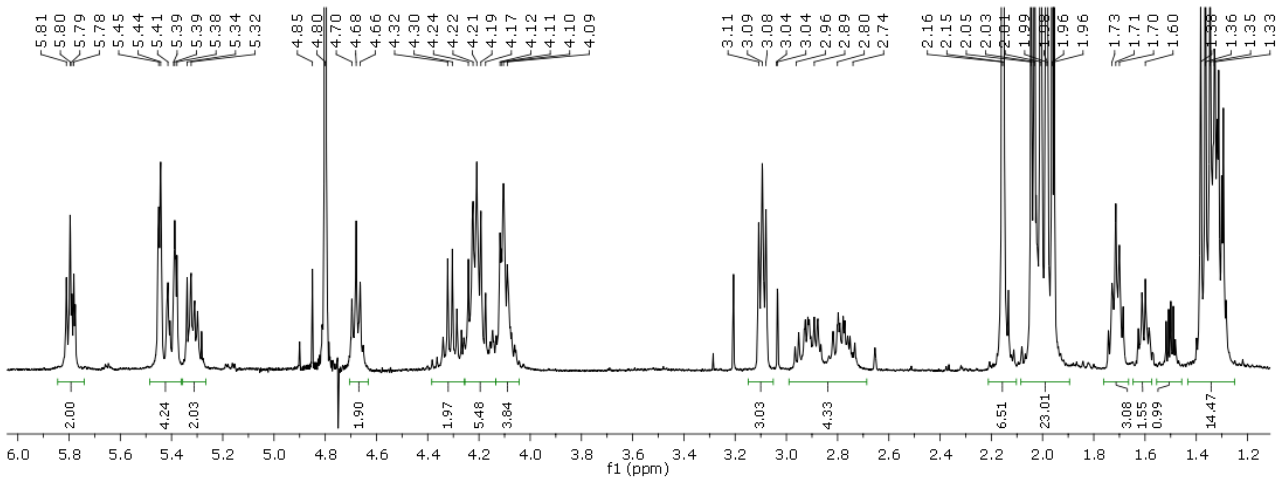
Table A. Reaction conditions for the synthesis of **125**.

Repeat (n)	(Glycosyl) amino acid	Method	Equivalents	Cycles	Reaction time (h)	[M]	Yield[a]
1	**123**	Procedure 5	5	2	2 + 2	0.5	100%
2	**123**	Procedure 5	10	1	3	0.5	85%
	80	Procedure 6	1.5	3	8 + 12 + 8	0.15	95%
	123	Procedure 5	5	2	2 + 2	0.5	100%

[a] As evaluated spectroscopically after Fmoc removal.

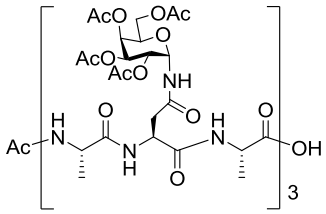
Global Yield = 80% (by UV evaluation).

After Fmoc removal, *N*-terminus protection was performed with a solution of 1 M Ac_2O in DMF (2h). Cleavage was performed as reported in **Procedure 7**, to afford 17 mg of crude, starting from 25 μmol of the first galactosyl amino acid loaded on the resin (maximum yield = 76%). After purification on C-18 silica, reported in Section 8.4, 10 mg of pure **125** were obtained (isolated yield = 50%). The isolated yield suffers from the separation of some deacetylated compounds deriving from **125**, which have the same retention time of truncated sequences.



^{1}H-NMR spectrum of the crude **125** after the cleavage (D_2O, 400 MHz)

Synthesis of α-*N*-linked glycopeptide 126 (n = 3)



The first alanine residue is already loaded on the resin (H-Ala-2-chlorotrityl resin). The loading of the first galactosyl amino acid **80** was performed as reported in **Procedure 4**. The other steps are reported in **Table B**.

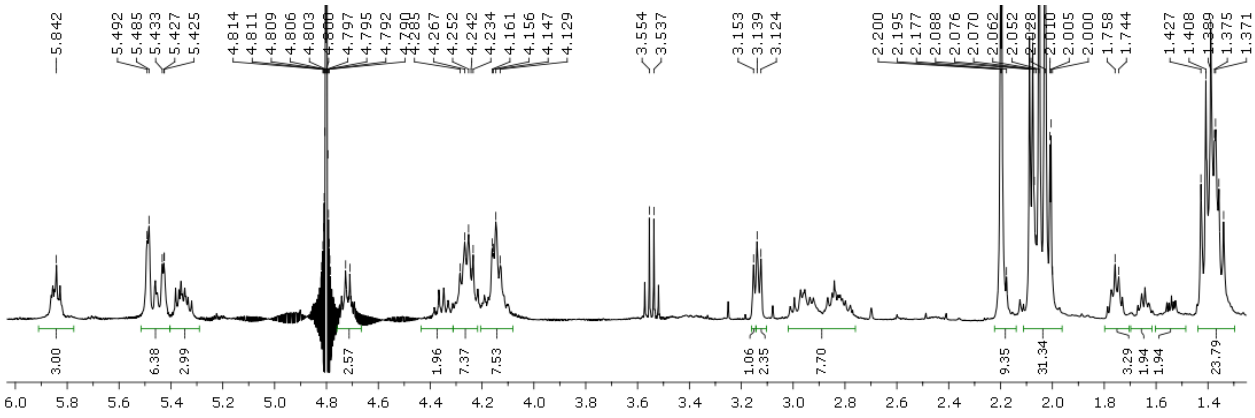
Table B. Reaction conditions for the synthesis of **126**.

Repeat (n)	(Glycosyl) amino acid	Method	Equivalents	Cycles	Reaction time (h)	[M]	Yield[a]
1	**123**	Procedure 5	5	2	2 + 2	0.5	100%
2	**123**	Procedure 5	5	2	2 + 2	0.5	100%
	80	Procedure 6	1.2	2	8 + 12	0.1	64%
	123	Procedure 5	5	2	2 + 2	0.5	100%
3	**123**	Procedure 5	5	2	2 + 2	0.4	94%
	80	Procedure 6	1.5	2	8 + 12	0.15	80%
	123	Procedure 5	10	1	3	0.5	92%

[a] As evaluated spectroscopically after Fmoc removal.

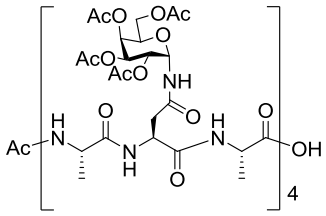
Global Yield = 44 % (by UV evaluation).

After Fmoc removal, *N*-terminus protection was performed with a solution of 1 M Ac_2O in DMF (2h). Cleavage was performed as reported in **Procedure 7**, to afford 35 mg of crude, starting from 25 μmol of the first galactosyl amino acid loaded on the resin (maximum yield = 77%). After purification on C-18 silica, reported in Section 8.4, 11 mg of pure **126** were obtained (isolated yield = 24%). Also in this case, the isolated yield suffers from the separation of some deacetylated compounds deriving from **126**, which have the same retention time of truncated sequences.



[1]H-NMR spectrum of the crude **126** after the cleavage (D_2O, 400 MHz)

Synthesis of α-*N*-linked glycopeptide 127 (n = 4)



The first alanine residue is already loaded on the resin (H-Ala-2-chlorotrityl resin). The loading of the first galactosyl amino acid **80** was performed as reported in **Procedure 4**. The other steps are reported in **Table C**.

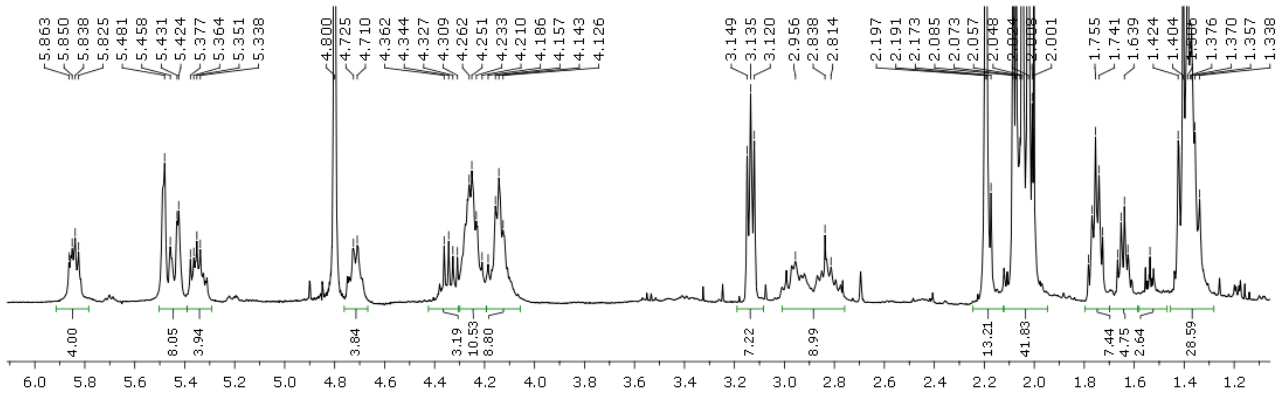
Table C. Reaction conditions for the synthesis of **127**.

Repeat (n)	(Glycosyl) amino acid	Method	Equivalents	Cycles	Reaction time (h)	[M]	Yield[a]
1	**123**	Procedure 5	5	2	2 + 2	0.5	100%
2	**123**	Procedure 5	5	2	2 + 2	0.5	100%
	80	Procedure 6	1.2	2	8 + 12	0.1	64%
	123	Procedure 5	5	2	2 + 2	0.5	100%
3	**123**	Procedure 5	5	2	2 + 2	0.4	94%
	80	Procedure 6	1.5	2	8 + 12	0.15	80%
	123	Procedure 5	10	1	3	0.5	92%
4	**123**	Procedure 5	5	2	2 + 2	0.5	100%
	80	Procedure 6	2	3	8 + 12 + 8	0.15	80%
	123	Procedure 5	5	2	2 + 2	0.5	80%

[a] As evaluated spectroscopically after Fmoc removal.

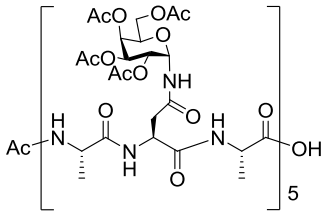
Global Yield = 28 % (by UV evaluation).

After Fmoc removal, *N*-terminus protection was performed with a solution of 1 M Ac_2O in DMF (2h). Cleavage was performed as reported in **Procedure 7**, to afford 43 mg of crude, starting from 25 μmol of the first galactosyl amino acid loaded on the resin (maximum yield = 71%). After purification on C-18 silica, reported in Section 8.4, two different fractions of 8 mg and 10 mg, respectively, were isolated. Since they were found to contain compound **127** together with truncations and deacetylated compounds, the maximum isolated yield was 30%.



^{1}H-NMR spectrum of the crude **127** after the cleavage (D_2O, 400 MHz)

Synthesis of α-*N*-linked glycopeptide 128 (n = 5)



The first alanine residue is already loaded on the resin (H-Ala-2-chlorotrityl resin). The loading of the first galactosyl amino acid **80** was performed as reported in **Procedure 4**. The other steps are reported in **Table D**.

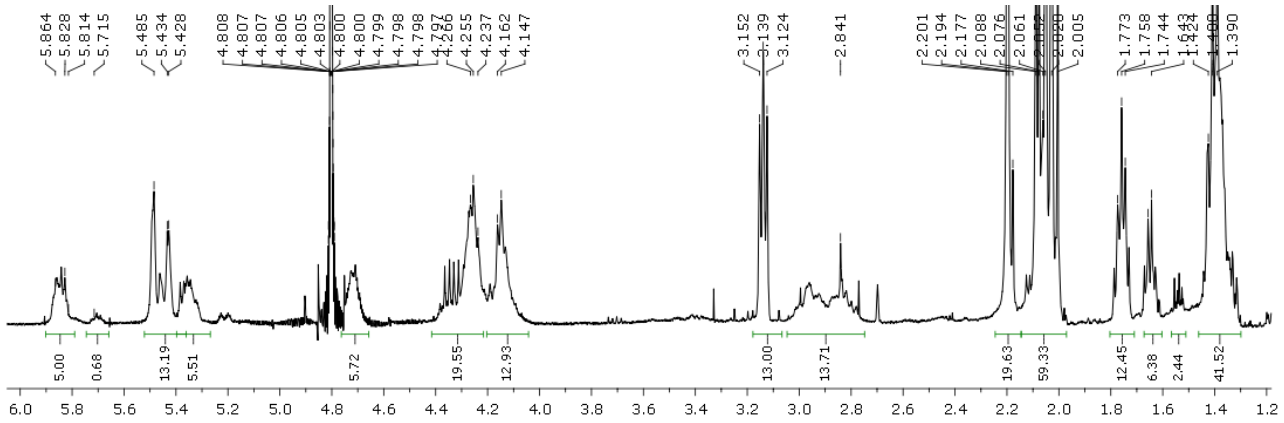
Table D. Reaction conditions for the synthesis of **128**.

Repeat (n)	(Glycosyl) amino acid	Method	Equivalents	Cycles	Reaction time (h)	[M]	Yield[a]
1	**123**	Procedure 5	5	2	2 + 2	0.5	100%
2	**123**	Procedure 5	5	2	2 + 2	0.5	100%
	80	Procedure 6	1.2	2	8 + 12	0.1	64%
	123	Procedure 5	5	2	2 + 2	0.5	100%
3	**123**	Procedure 5	5	2	2 + 2	0.4	94%
	80	Procedure 6	1.5	2	8 + 12	0.15	80%
	123	Procedure 5	10	1	3	0.5	92%
4	**123**	Procedure 5	5	2	2 + 2	0.5	100%
	80	Procedure 6	2	3	8 + 12 + 8	0.15	80%
	123	Procedure 5	5	2	2 + 2	0.5	80%
5	**123**	Procedure 5	5	2	2 + 2	0.5	100%
	80	Procedure 6	3	3	8 + 12 + 8	0.15	90%
	123	Procedure 5	5	2	2 + 2	0.5	80%

[a] As evaluated spectroscopically after Fmoc removal.

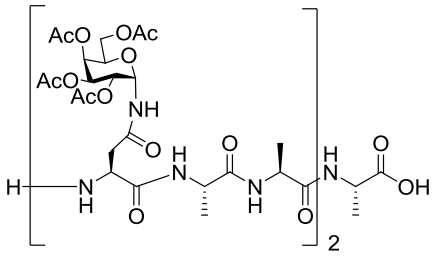
Global Yield = 20 % (by UV evaluation).

After Fmoc removal, *N*-terminus protection was performed with a solution of 1 M Ac_2O in DMF (2h). Cleavage was performed as reported in **Procedure 7**, to afford 54 mg of crude, starting from 25μmol of the first galactosyl amino acid loaded on the resin (maximum yield = 72%). After purification on C-18 silica, reported in Section 8.4, two different fractions of 10 mg and 12 mg, respectively, were isolated. Since they were found to contain compound **128** together with truncations and deacetylated compounds, the maximum isolated yield was 29%.



^{1}H-NMR spectrum of the crude **128** after the cleavage (D_2O, 400 MHz)

Synthesis of compound 130 with microwave-assisted solid phase synthesis



H-Ala-2-chlorotrityl resin (50 mg) was swollen in DMF for 30 minutes at room temperature. Fmoc-Ala-Ala-OH **129** (58 mg, 0.150 mmol) and HATU (57 mg, 0.150 mmol) were dissolved in 600 μL of DMF (c = 0.25 M) at 0 °C. DIPEA (52 μL, 0.300 mmol) was added and the reagents were stirred at 0 °C for 15 min, then added to H-Ala-2-chlorotrityl resin which had been placed in a 2 mL reactor. The reaction mixture was shaken (vortexing) under microwave irradiation at 50 °C for 20 min. Capping was performed twice for 30 min with a 1 M Ac_2O solution in DMF (1 mL). After Fmoc removal with 25% piperidine in DMF (10 min for two times), the loading was estimated by measuring the UV absorption at 300 nm. Loading = 0.5 mmol/g (25 μmol/50 mg). The resin was washed thoroughly with DMF to eliminate any trace of piperidine. Compound **80** (52 mg, 0.075 mmol, 3 eq) and PyBrop (41 mg, 0.087 mmol) were dissolved in 600 μL of DMF (c = 0.125 M) at 0 °C. DIPEA (48 μL, 0.275 mmol) was added and the reagents were stirred at 0 °C for 15 min, then

added to the resin. The reaction mixture was shaken under microwave irradiation at 50 °C for 20 min. Capping was performed twice for 30 min with a 1 M solution of Ac_2O in DMF (1 mL). After Fmoc removal with 25% piperidine in DMF (10 min for two times), the coupling yield was estimated to be 62% (15 μmol).

Fmoc-Ala-Ala-OH **129** (42 mg, 0.110 mmol, 7 eq) and HATU (42 mg, 0.110 mmol, 7 eq) were dissolved in 600 μL of DMF (c = 0.2 M) at 0 °C. DIPEA (38 μL, 0.22 mmol, 14 eq) was added and the reagents were stirred at 0 °C for 15 min, then added to the resin. The reaction mixture was shaken under microwave irradiation at 50 °C for 20 min. This coupling cycle was repeated.

Capping was performed twice for 30 min with a 1 M solution of Ac_2O in DMF (1 mL).). After Fmoc removal with 25% piperidine in DMF (10 min for two times), the coupling yield was estimated to be 100% (15 μmol). The resin was washed thoroughly with DMF. Compound **80** (41 mg, 0.060 mmol, 4 eq) and PyBrop (33 mg, 0.070 mmol, 4.6 eq) were dissolved in 600 μL of DMF (c = 0.1 M) at 0 °C. DIPEA (38 μL, 0.220 mmol, 14 eq) was added and the reagents were stirred at 0 °C for 15 min, then added to the resin. The reaction mixture was shaken under microwave irradiation at 50 °C for 20 min. This coupling cycle was repeated 2 more times. After Fmoc removal with 25% piperidine in DMF (10 min for two times), the coupling yield was estimated to be 95% (14 μmol). The condition are summarized in Table E.

Global Yield = 58 % (by UV evaluation).

Cleavage was performed as reported in **Procedure 7**, to afford 71 mg of crude product, which upon LC-MS analysis shows the predominant presence of compound **130** (**Figure 5**, Section 7.3).

Table E. Reaction conditions for the synthesis of **130** at 50°C with microwave.

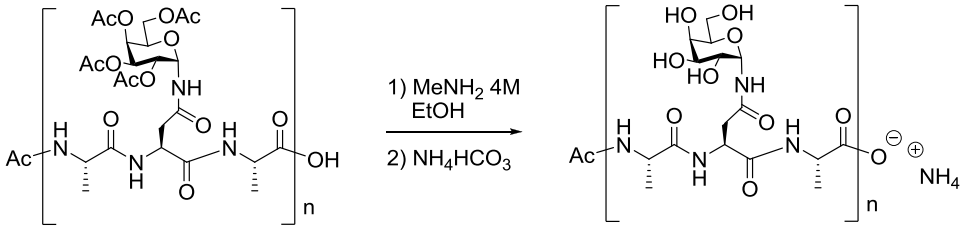
Repeat (n)	(Glycosyl) amino acid	Equivalents	Cycles	Reaction time	[M]	Yield[a]
1	**129**	6	1	20 min	0.25	100%
	80	3	1	20 min	0.12	62%
2	**129**	7	2	20 min	0.2	100%
	80	4	3	20 min	0.1	95%

[a] As evaluated spectroscopically after Fmoc removal.

Procedure 8

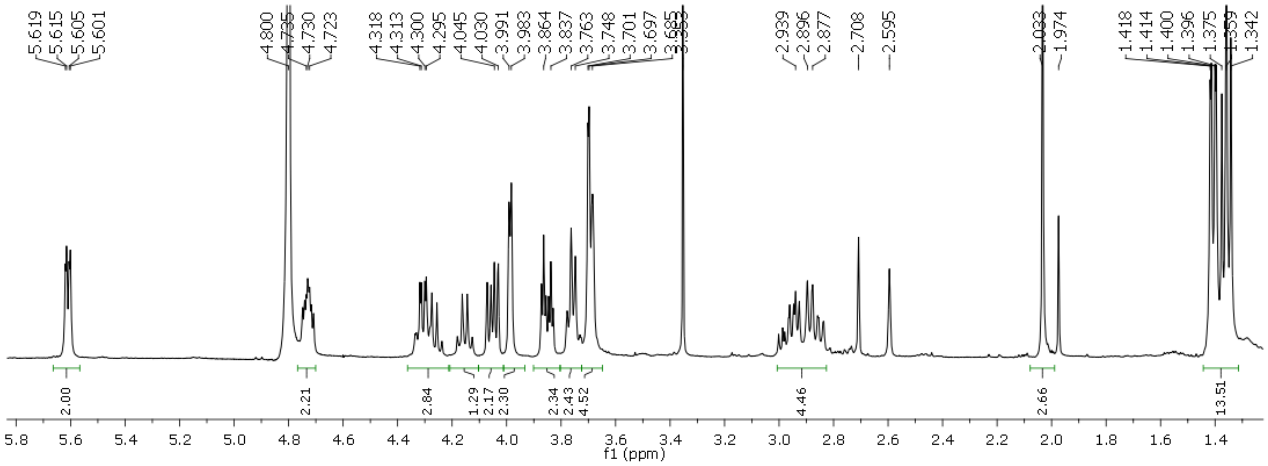
General method for *O*-acetyl removal

After an initial purification of the acetylated glycopeptides (**Chapter 8**), the *O*-acetyl compound was dissolved in a 4M $MeNH_2$-EtOH solution (c = 0.05M). The reaction mixture was stirred for 5 h, then the solvent was evaporated at reduced pressure and stripped three times with methanol. An ammonium bicarbonate solution (1M) in methanol was added and the mixture was stirred overnight. Solvent was evaporated and the crude was lyophilized.

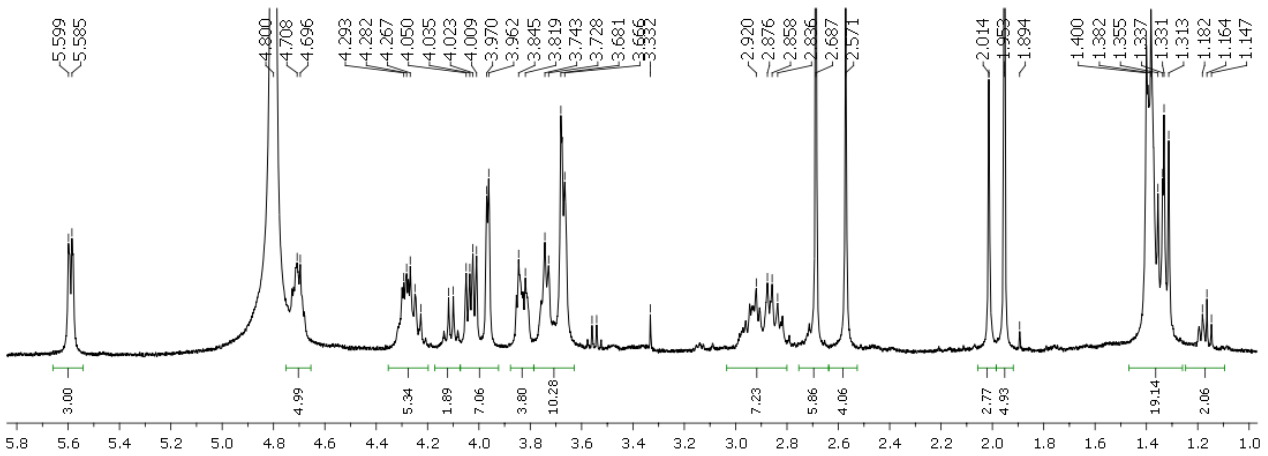


n = 2 **125**
n = 3 **126**
n = 4 **127**
n = 5 **128**

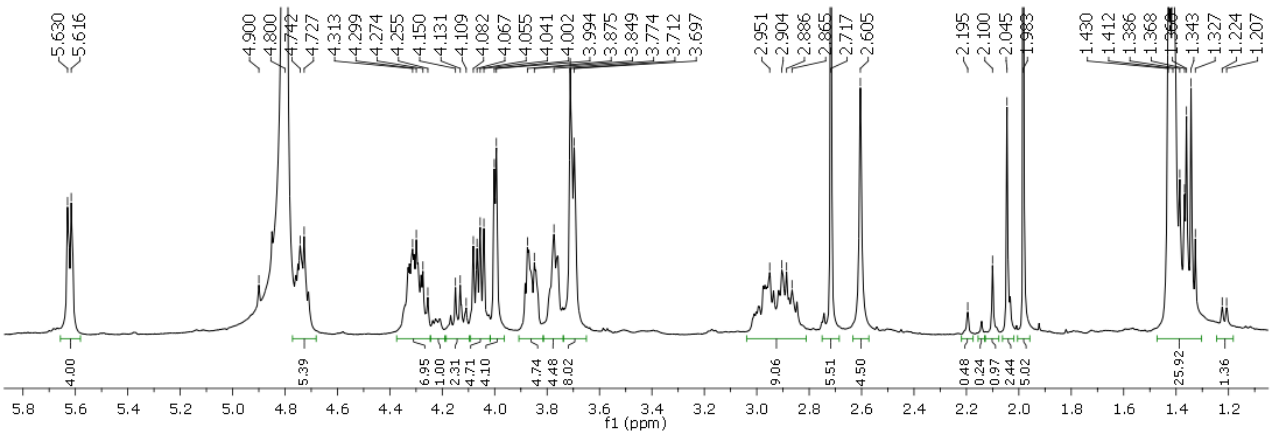
n = 2 **131**
n = 3 **132**
n = 4 **133**
n = 5 **134**



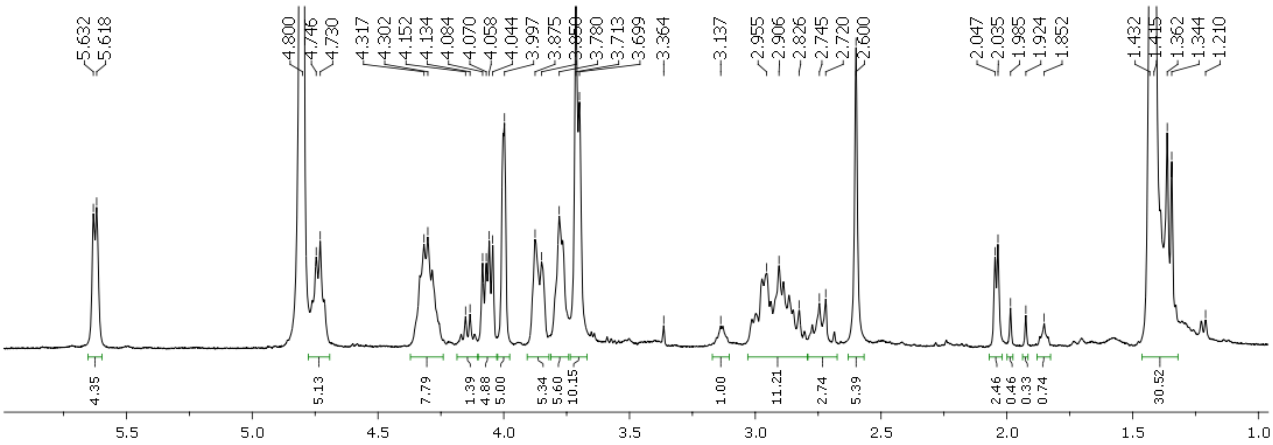
^{1}H-NMR spectrum of the crude **131** (n = 2) after deprotection (D_2O, 400 MHz)



^{1}H-NMR spectrum of the crude **132** (n = 3) after deprotection (D_2O, 400 MHz)



^{1}H-NMR spectrum of the crude **133** (n = 4) after deprotection (D_2O, 400 MHz)



^{1}H-NMR spectrum of the crude **134** (n = 5) after deprotection (D_2O, 400 MHz)

7.6 References

[1] DeVries, A. L,; Vandenheede, J.; Feeney, R. E. *J. Biol. Chem*. **1971**, *246*, 305-308.

[2] Tam, R. Y.; Rowley, C. N.; Petrov, I.; Zhang, T.; Afagh, N. A.; Woo, T. K.; Ben R. N. *J. Am. Chem. Soc* **2009**, *131*, 15745–15753

[3] Garner, J.; Harding, M. M. *ChemBioChem* **2010**, *11* 2489-2498.

[4] Wang, J.-H. *Cryobiology* **2000**, *41*, 1–9.

[5] Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S.-I. *Angew. Chem. Int. Ed.* **2004**, *43*, 856– 862.

[6] a) Hays, L. M.; Feeney, R. E.; Crowe, L. M.; Crowe, J. H.; Oliver, A. E. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6835–6840. b) Inglis, S. R.; Turner, J. J.; Harding, M. M. *Curr. Protein Pept. Sci.* **2006**, *7*, 509– 522.

[7] a) Carpenter, J. F.; Hansen, T. N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8953–8957. b) Chao, H.; Davies, P. L.; Carpenter, J. F. *J. Exp. Biol.* **1996**, *6*, 2071–2076. c) Bouvet, V.; Ben R. N. *Cell Biochem Biophys.* **2003**, **39**, 133-144.

[8] Leclere, M.; Kwok, K. B.; Luke K. W.; Allan, D. S.; Ben, R. N. *Bioconjugate Chem*. **2011**, *22*, 1804-1810.

[9] Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2152.

[10] Belvisi, L.; Bernardi, A.; Checchia, A.; Manzoni, L.; Potenza, D.; Scolastico, C.; Castorina, M.; Cupelli, A.; Giannini, G.; Carminati, P.; Pisano, C. *Org. Lett.* **2001,** 3, 1001-1004.

[11] Renaudet, O.; Dumy P.; *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3619–3622

[12] Wilkinson, B. L.; Chun, C. K. Y.; Payne, R. J. *Peptide Science* **2011**, *96*, 137-146.

[13] Shao, N.; Xue, J.; Guo, Z.; *J. Org. Chem.* **2003**, *68*, 9003-9011.

[14] Zhang, Y.; Gu, H.; Yang, Z.; Xu B. *J. Am. Chem. Soc.* **2003**, *125*, 13680-13681.

[15] Friligou, I.; Papadimitriou, E.; Gatos, D.; Matsoukas, J.; Tselios, T. *Amino Acids* **2011**, *40*, 1431-1440.

[16] Heggermann, C.; Budke, C.; Schomburg, B.; Majer, Z.; Wibrock, M.; Koop, T.; Sewald, N. *Amino Acids* **2010**, *38*, 213-222.

[17] Drenichev, M. S.; Kulikova, I: V.; Bobkov, G. V.; Tararov, V, I; Mikhailov, S. N. *Synthesis* **2010**, *22*, 3827-3834.

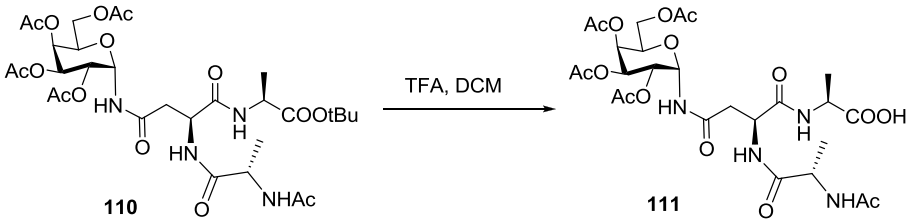
[18] Nisic. F.; Speciale, G.; Bernardi A.; Submitted

Chapter 8
Purification of α-*N*-linked glycopeptides

8.1 Introduction

The particular behaviour of α-*N*-linked glycopeptides in acid or basic environments has been detailed in Chapter 6. Interesting observations were obtained especially during the screening of deacetylation conditions for the sugar moiety. Summarising the main points, we found that:

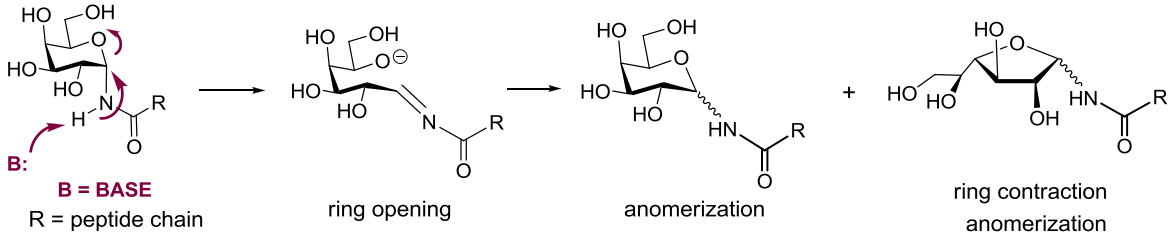
a) *O*-acetyl α-*N*-linked glycopeptides are generally stable to acidic conditions in the absence of nucleophiles; removal of the *tert*-butyl ester in **110** was successfully performed with trifluoroacetic acid in CH_2Cl_2, an aprotic and apolar solvent (**Scheme 1**, see Section 6.4) No isomerization of the sugar moiety was observed.



Scheme 1. Stability of *O*-acetyl α-*N*-linked glycopeptides to acids.

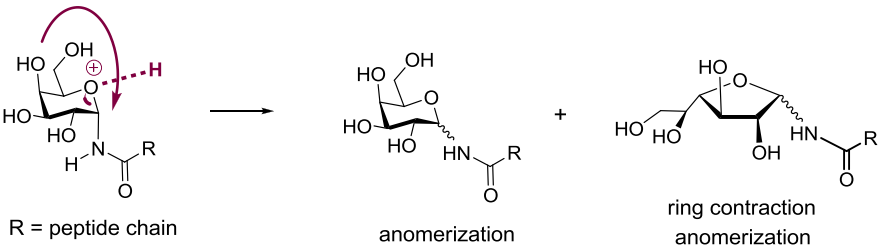
When these substrates were manipulated under acidic conditions and in MeOH or water, partial solvolysis of the acetyl groups was observed, as in the chromatographic purification of **113a** (Section 6.5).

b) Protected and unprotected α-*N*-linked glycopeptides are unstable towards bases, at pH above 9 and for prolonged time contact. Side reactions resulted in ring opening, followed by anomerization and/or ring contraction (formation of the furanosyl compounds) of the galactose moiety (**Scheme 2**, see Section 6.5).



Scheme 2. Instability of unprotected α-*N*-linked glycopeptides in bases.

c) Unprotected α-*N*-linked glycopeptides are unstable towards acids, at pH below 4-5 (**Scheme 3**, see Section 6.5). In particular, ring contraction (furanose formation) and anomerization were observed in deacetylation reactions after the addition of the sulfonic resin IRA 120 H^+.



Scheme 3. Instability of unprotected α-*N*-linked glycopeptides towards acids.

We found a literature precedent for this behaviour in the acid-induced isomerization of *N*-glycineamide-ribofuranose [1]

d) Similarly, unprotected α-*N*-linked glycopeptides are unstable during purification on silica, using 0.1% of trifluoroacetic acid (TFA) added to the eluants acetonitrile and water (see Section 6.5). In these conditions furanoside formation up to 10% was observed.

The propensity of galactose to give furanosides in aqueous and non-aqueous media was described in the literature (data reported in Table 1).[2]

Table 1. Furanose configuration of monosaccharide, in water and DMSO media[a]

Entry	Media	β-pyranose (%)	α-pyranose (%)	β-furanose (%)	α-furanose (%)
1	Water	63.8	30.2	4.3	2
2	DMSO	39	28	24	9

[a] From *ref 1*

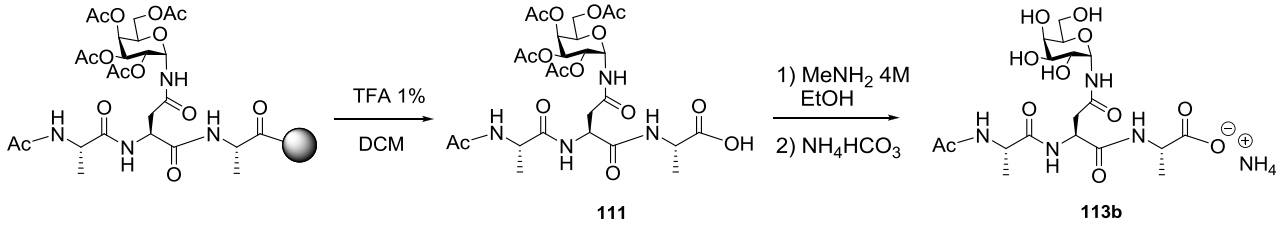
Solutions to avoid anomerization and furanoside formation during sugar deacetylation relied on the use of methylamine in ethanol and were described in Section 6.5. These conditions avoid deprotonation of the anomeric nitrogen during the reaction and do not require an acid quench, thus preserving the structural integrity of the glycopeptide. They were then further applied for the deprotection of complex α-*N*-linked glycopeptides described in Chapter 7.

Eventually, the final problem to solve was the identification of purification methods of the α-*N*-linked glycopeptides, which was aggravated by the complex mixtures obtained for some of the sequences from solid phase synthesis (Chapter 7). The development of purification procedures for the isolation and characterization of the unnatural α-*N*-linked glycopeptides is described in this Chapter. In particular, Section 8.2 describes preliminary studies on the small glycopeptide **111**, which contains only one repeating unit Ala-Asn-(α-Gal)-Ala (n = 1). Section 8.3 is focused on the description of Hydrophilic Interaction Liquid Chromatography (HILIC), a novel chromatographic

method, which allowed the isolation of unprotected glycopeptides. Section 8.4 illustrates the use of these techniques for each glycopeptide (n = 2, 3, 4, 5).

8.2 Preliminary studies: isolation and purification of Ac-Ala-Asn-(α-*N*-Gal)-Ala-OH

Initial purification experiments were performed on glycopeptides **111** and **113b** (**Scheme 4**), whose synthesis is described in Chapter 7. Our original intent was to purify the unprotected compound **113b** directly and to obtain the final product in a single purification step, after cleavage from the resin.



Scheme 4. Cleavage from 2-chlorotrityl resin of **111** and subsequent deprotection to **113b**.

Aware of the instability of these glycopeptides to chromatography in the presence of TFA, we initially attempted to purify the unprotected compound **113b** as the ammonium salt on reverse phase C-18 silica (Biotage cartridge; eluants: acetonitrile and water; gradient CH_3CN: 5% for 1 min, 5%→60% in 6 min, **Figure 1**) but the compound was practically not retained. Compound **113b** was also not retained when eluting with 100% water. The use of 0.1 % TFA in the eluant could obviously generate the free carboxylic acid and lead to increased retention, but TFA cannot be used, as established before (Section 6.5), in order to avoid furanoside formation.

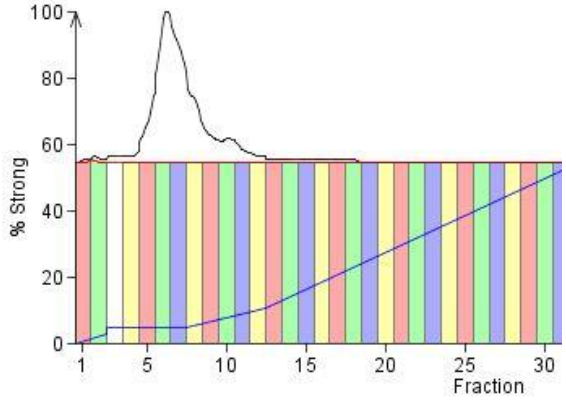


Figure 1. C-18 reverse phase chromatogram of the crude **113b**.

Finally, since direct purification of the final unprotected glycopeptide was not possible, we attempted to purify the *O*-acetyl α-*N*-linked glycopeptide **111**. After cleavage from 2-chlorotrityl resin, ^{1}H NMR spectrum of the crude showed the presence of some alifatic signals (**Figure 2**, indicated in red), which are also found in the ^{1}H NMR spectra of all the α-*N*-linked glycopeptides **125-128** (see Section 7.6) cleaved from 2-chlorotrityl resin.

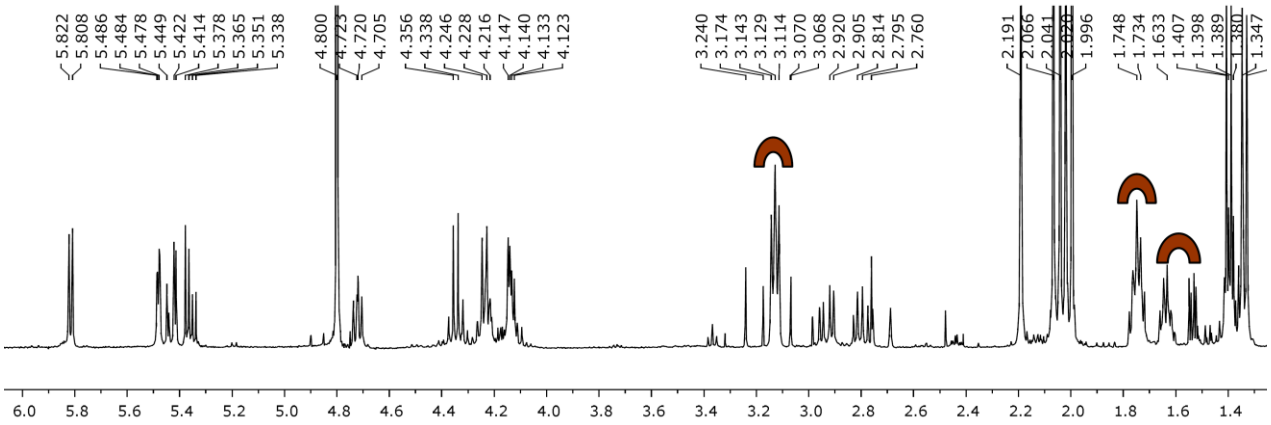


Figure 2. ^{1}H NMR spectrum of crude **111** after clevage from CTC resin.

Purification of **111** on reverse phase C-18 silica, with acetonitrile and water as eluants, without TFA (Biotage cartridge, CH_3CN 5%→40% in 8 min, 40%→80% in 4 min, **Figure 3**) produced two products of the same mass (ESI-MS analysis: $[M]^+$= 646) (from 63 mg of crude, **111s** = 22 mg; **111** = 31mg). The first eluted compound, **111s** (t_r = 1min, fractions 5-8), showed in its ^{1}H NMR spectrum both the presence of the glycopetide's signals and of the signals indicated in red in **Figure 2** and it was tentatively identified as the piperidine salt of **111**. The second eluted compounds (t_r = 6 min, fractions 30-38) was characterized as **111** (**Figure 4**).

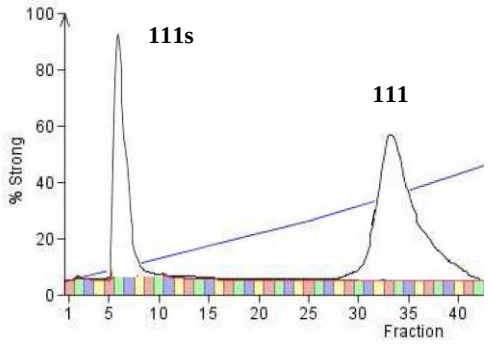


Figure 3. C-18 reverse phase chromatogram of the crude **111**.

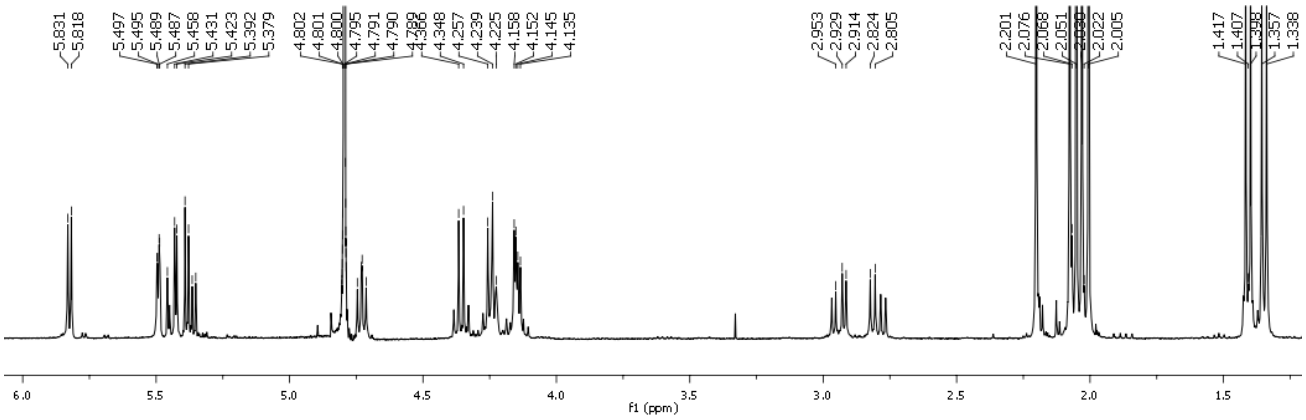


Figure 4. ^{1}H NMR spectrum of the fraction 30-38 of chromatogram in Figure 3.

The identification of **111s** was also supported by another experiment: if pyridine is added to the solution resulting from the cleavage, **111** converts quantitatively in a pyridine salt and it is eluted after 1 minute, like **111s**, in the same conditions (CH_3CN 0%→40% in 8 min, 40%→80% in 4 min). We therefore supposed that piperidine from the Fmoc cleavage steps is retained in the resin, despite the numerous wash cycles, and released in solution when the peptide is cleaved. Addition of 0.1% TFA to the eluant, during C-18 purification, only partially releases the base from **111s** and consequently transformation into **111** is not complete. Longer contact time of the acid with **111s**, however, causes partial deacetylation of the sugar moiety. **Figure 5** shows the chromatogram of **111s**, eluted with 0.1% TFA (C-18 silica, eluants: acetonitrile and water gradient CH_3CN 5%→60% in 8 min). The first eluted fraction (fraction 7-8) contained the alifatic compound alone (^{1}H NMR spectrum showed in **Figure 6**), while the other fractions contained partial-deacetylation products deriving from **111,** as shown by LC/Ms analysis. 7 mg were saved from these fractions, which produced a total yield of **111** of 95%.

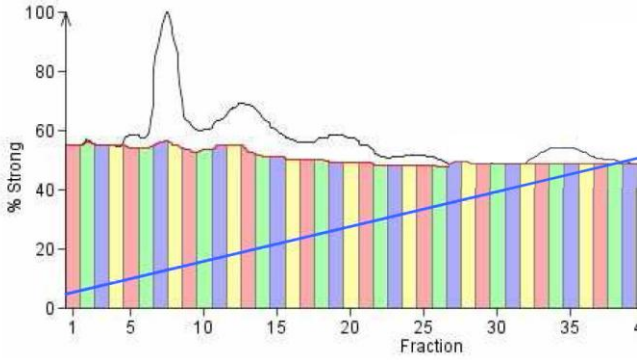


Figure 5. C-18 reverse phase chromatogram of **111s** with 0.1% TFA.

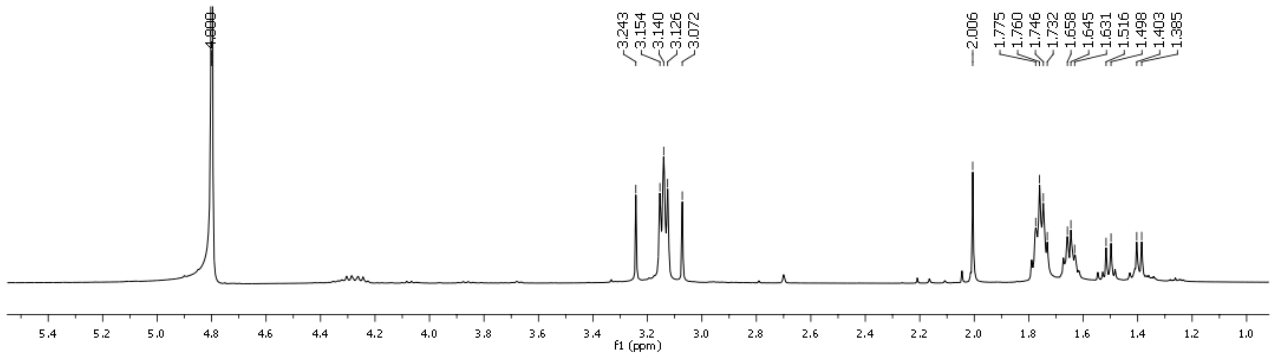


Figure 6. ^{1}H NMR spectrum of the fraction 7-8 of chromatogramm in Figure 5.

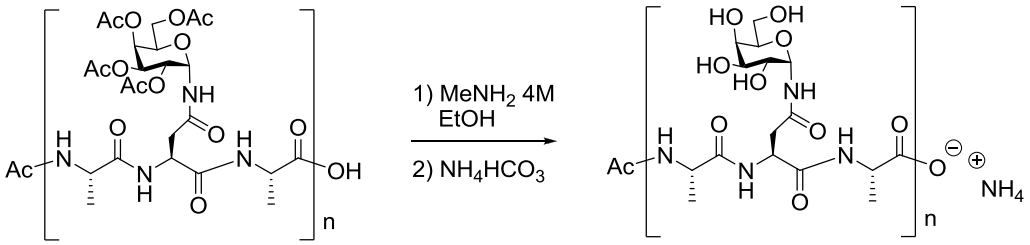
Hence, addition of 0.1% TFA to the chromatography eluants appeared to be necessary to release **111** as a free acid, avoiding salification from adventitious piperidine, thus confering to the cleaved peptide a certain ability of retention on C-18 silica. However, at the same time, the presence of 0.1 % TFA in water was able to cause partial deacetylation of the sugar portion (like in a retro-Fischer reaction), and complicate the purification by creating a sub-group of products which were eluted at different retention times. Nonetheless, after ^{1}H NMR and ESI-MS analysis, the fractions containing the same partially deacetylated glycopeptide could be combined and submitted together to the final deacetylation reaction. This method, despite being tedious and analytically not very elegant, allowed to efficiently separate peptides originating from truncated sequences. Indeed, for the purification of α-*N*-linked glycopeptides reported in **Scheme 5**, which in some cases were obtained as complex mixtures of products containing truncated sequences deriving from partial coupling reactions (Section 7.6), these preliminary experiments allowed us to assess that:

- Unprotected α-*N*-linked glycopeptides **131-134** isolated as ammonium salts, after removal of the *O*-acetyl- groups, cannot be purified on reverse phase C-18 silica, because their retention times are too short. On the other hand use of 0.1 % TFA in the eluant, which can turn the salts into the corresponding free acids must be avoided, to avoid ring contraction and furanoside generation.

- An initial purification of the *O*-acetyl-glycopeptides **125-128** could be achieved after resin cleavage using reverse phase chromatography and 0.1 % TFA in the eluant. Purification at this level is useful to separate part of the truncated sequences. If partial deacetylation occurred, the fractions containing deacetylation products of the same peptide could be identified by LC/MS and reassembled.

- A second purification, after *O*-acetyl removal, of unprotected α-*N*-linked glycopeptides **130-133** could be performed with a different chromatographic technique, HILIC chromatography, which

is specially suited for the purification of highly polar compounds. HILIC chromatography is available only on the semi-preparative scale, and thus cannot be used alone for the full purification of the crude peptides. However, its combination with the pre-purification on C18 of the aceylated glycopeptides described above allowed us to obtain pure samples of the peptides synthesized. The nature of HILIC chromatography will be illustrated in the next section (Section 8.3).



n = 2 **125**
n = 3 **126**
n = 4 **127**
n = 5 **128**

n = 2 **131**
n = 3 **132**
n = 4 **133**
n = 5 **134**

Scheme 5 *O*-acetyl-α-*N*-linked glycopeptides **125-128** and unprotected α-*N*-linked glycopeptides **131-134**

8.3 Hydrophilic Interaction Liquid Chromatography (HILIC)

Generally, adsorption chromatography is based on the interactions between solute molecules and the stationary phase. Normal phase HPLC (NP-HPLC) makes use of polar stationary phases and (relatively) non polar mobile phases so the order of elution is from less polar to more polar analytes. On the contrary, reverse phase HPLC (RP-HPLC), results from the adsorption of hydrophobic solutes onto a hydrophobic solid support in a water-based polar mobile phase. The retention order, opposite to that of NP-HPLC, is from more polar analytes to less polar ones. The matrix for reversed phase chromatography is generally composed of microparticulate silica and chemically grafted hydrophobic ligands consist of linear hydrocarbon chains (*n*-alkyl groups) such as octadecylsilane (C18). Reverse phase-HPLC is widely used with biological polar samples because of its broad compatibility and for the general solubility of these compounds in water, a property which allow the loading of the samples in eluant-like conditions. However, some very polar compounds, such as sugars, oligosaccharides or other biomolecules, are poorly retained on a reversed phase HPLC column even with high aqueous mobile phases. Lack of retention for highly hydrophilic compounds is mainly caused by solvophilic factors. In fact, when polar functional groups of the analyte have the ability to make favourable dipolar bonds with the solvent (becoming solvated) and, in addition, the non-polar stationary phases (C-18) cannot offer similar bonding, the solutes stay in solution and are thus eluted in the void volume.[3] The functional moieties that convey

this property to highly polar compounds are either charged groups or groups capable of entering strong dipolar or hydrogen bonds. In our case, for instance, α-*N*-linked glycopeptides **130-133** are very hydrophilic molecules and in addition have a terminal carboxylic group, which is also present as a salt, that could further reduce retention times, even with highly polar mobile phases.

More than 30 years ago, a new chromatographic technique, which for a long time was mostly employed for carbohydrate analysis, emerged and was called hydrophilic interaction liquid chromatography (HILIC).[4] This method used polar stationary phases and mobile phases containing some water and a higher percentage of acetonitrile and was in fact named HILIC to emphasize the presence of water in the mobile phase as the stronger eluting member, and the partition retention mechanism. Following twenty years of continuous development, HILIC is nowadays accepted as a common separation mode, which can be applied to polar small molecules, oligonucleotides, glycopeptides, amino acids, proteins and highly polar natural products (Figure 7).[5]

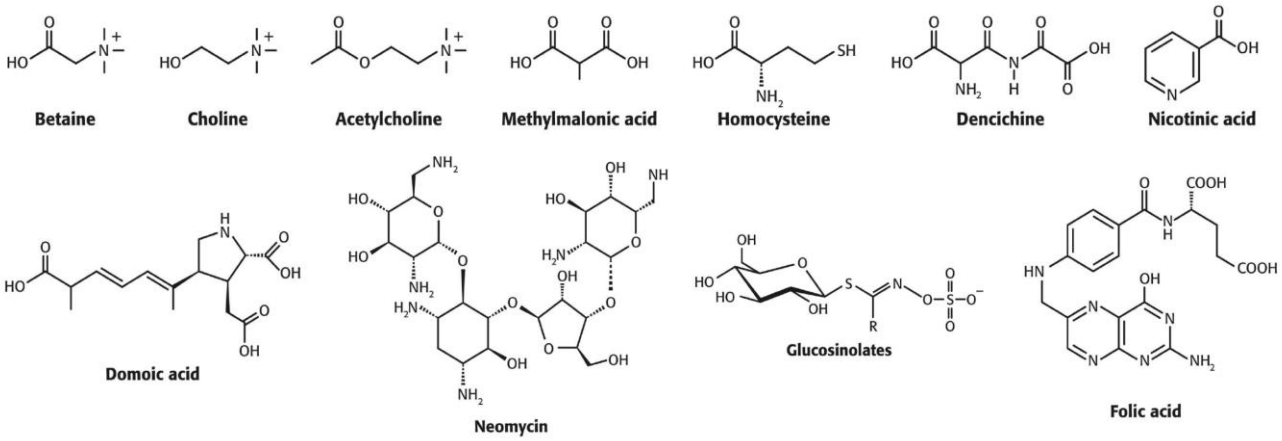


Figure 7. Some representative polar and basic compounds separated by HILIC

The exact retention mechanism for HILIC is still open to debate, but the most widely accepted hypothesis is that the partition mechanism arises from preferential adsorption of water on the polar stationary phase, which results in a relative higher water content in the stagnant liquid phase of the stationary phase support than in the mobile phase eluant. In practice, the partition of the analyte is between a water-enriched layer of stagnant eluant and a relatively hydrophobic bulk eluent, with the main components usually being 5–40% water in acetonitrile. In fact, acetonitrile-water mixtures are the typical mobile phases for HILIC separations. The water content depends on the polarity of both stationary phases and analytes to be separated. In general, the more polar the stationary phase and the analyte, the higher water content is needed for the separation. In other words, water content must be low enough to achieve a separation, but be high enough for the mobile phase to effectively dissolve the analytes and elute them in a reasonable time.[6]

However, retention in HILIC is also supposed to correlate with strong interactions (ionic, ion-dipole, and hydrogen bonding). An adsorptive retention model for peptides in HILIC was recently studied by Yoshida,[7] who concluded that the elution pattern was analogous to non aqueous normal phase chromatography, indicating a similar retention mechanism based on ionic, ion-dipole, and hydrogen bond interactions. At present, all manufacturers of HILIC columns recommend the use of buffered eluants in order to reduce the electrostatic interactions between charged analytes and deprotonated silanol groups of the stationary phase. These interactions are a major factor when separating charged molecules; since, for example, basic analytes can be separated by an ion-exchange mechanism on pure silica. Ammonium salts of formate, acetate or bicarbonate are the most recommended buffer solutions, due to their high solubility in eluants with high CH_3CN content. In our specific case we used:

a) A LUNA-HILIC HPLC column (Phenomenex), with a silica surface covered with cross-linked diol groups (**Figure 8**) The diol silica stationary phase closely resembles naked silica in overall polarity (high polarity and hydrogen bonding properties)

b) Buffer solutions of ammonium bicarbonate and formate at pH ≈ 7.5 to control the ionic strength and pH of the mobile phase, and thus the homogeneity of the sample (see experimental section for the exact composition of these buffers). At this pH, in fact the glycopeptides should exist as ammonium salts.

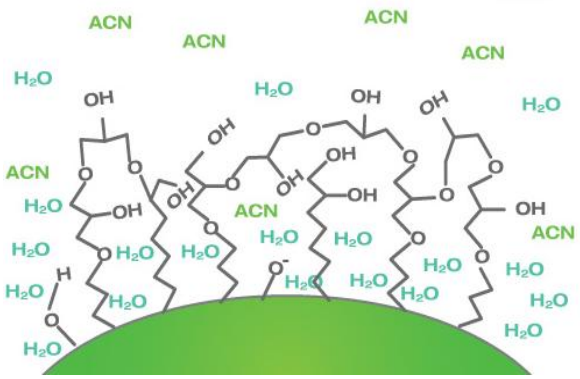


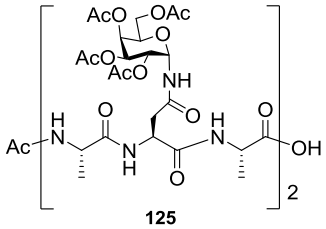
Figure 8. The diol silica stationary phase of LUNA-HILIC HPLC column.

In HILIC chromatography, dependence of the separation performance from the ionic strength of the eluant is the most puzzling feature to the novel user. To the best of our knowledge, there is no clear explanation of the underlying mechanisms which make retention times and separation efficiency highly dependent on the ionic strength parameter. However, we did observe that the efficiency of the purification depended very strictly on the ionic strength of the medium. In the next section, a striking example of this effect will be shown. The manufacturer suggested a range of molarity to be applied which varying from 10 mM to 100 mM, and in the literature 50 mM buffer solutions are the

most used.[6] In our hands, best results were obtained for buffer concentrations ranging from 100 mM to 10 mM. In particular, 100 mM solution was used for the purification of the relatively short glycopeptide **125** (n = 2). This, in turn, clearly implies that the resulting glycopeptide is isolated together with large amounts of inorganic salts (ammonium formates or bicarbonates) from the buffer. They can be removed upon lyophilization of the sample, but clearly reduce the use of HILIC chromatography as a fully preparative method. On the other hand the 100 mM buffer, at the same pH (7.5) and with the same eluant, was not suitable anymore for purification of longer glycopeptides, leading to long retention time and broadening of the peaks (see Section 8.4). These features represent a major problem for samples that are not homogeneous and that showed scarce absorbance even at 215 nm. We noticed that reducing the molarity form 100 mM to 10 mM, maintaining the pH and the percentage of water in the eluant, caused an increase in the relative absorbance of the products and in their resolution, but more surprising, a wide decrease in the retention time.

8.4 Purification of α-*N*-linked glycopeptides

Compound **125** (n = 2) was obtained from solid phase synthesis with a global yield = 80%, determined from UV-absorbance of the Fmoc-piperidine adduct for each coupling reaction.



125

The crude (17 mg, ^{1}H NMR spectrum in Section 7.6) was purified by reverse phase chromatography (**Figure 9**): C-18 Biotage cartridge (12 g); eluant: H_2O/CH_3CN (0.1 % TFA); flux: 15ml/min; gradient: 5%→100% CH_3CN in 11 min.

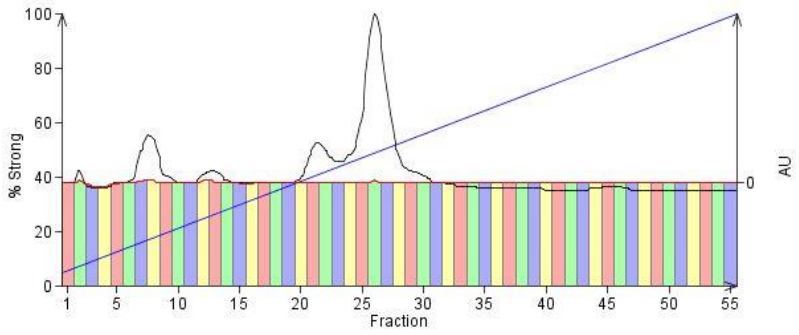


Figure 9. C-18 reverse phase chromatogram of **125**.

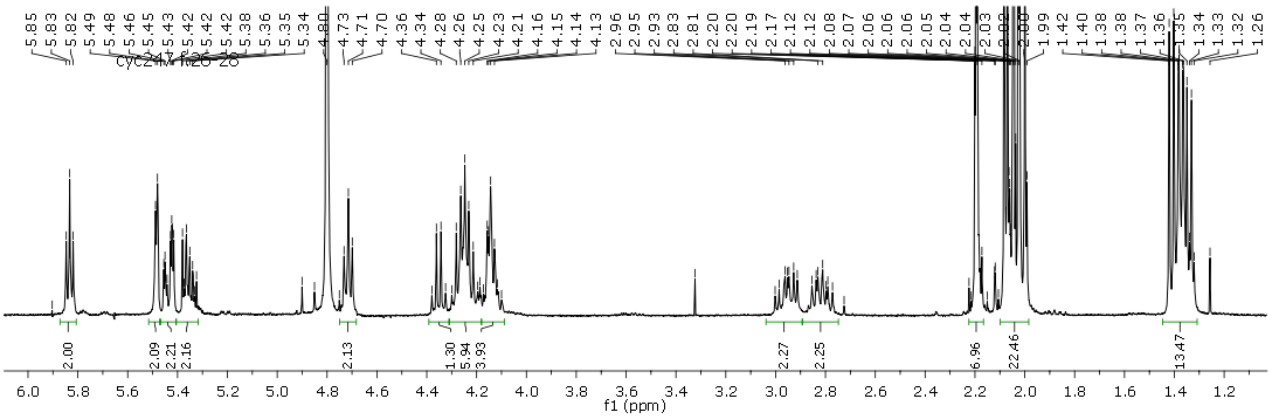
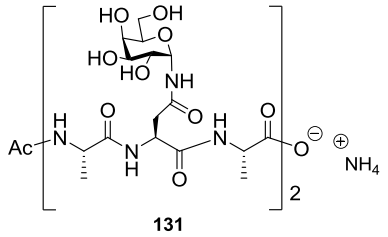


Figure 10. ^{1}H NMR spectrum of fractions 25-28 of chromatogram in Figure 9.

t_r (**125**) = 5 min (fractions 25-28 =10 mg, isolated yield = 50%, **Figure 10**)

Fractions 7-8 contained the alifatic compound showed in **Figure 6** of Section 8.2. Fractions 20-24 contained a truncated sequence and partial-deacetylation products deriving from **125** as established by ESI-MS analysis.

O-acetyl removal from compound **125** afforded compound **131** in 95% yield (7 mg).



131

Its LC-MS chromatogram (gradient: CH_3CN 0%→30% in 6 min) revealed the presence of a single compound (**Figure 11**, [M +1]), while the ^{1}H NMR spectrum showed the presence of *N*-methylacetamide (AcNHMe) produced by the deacetylation reaction.

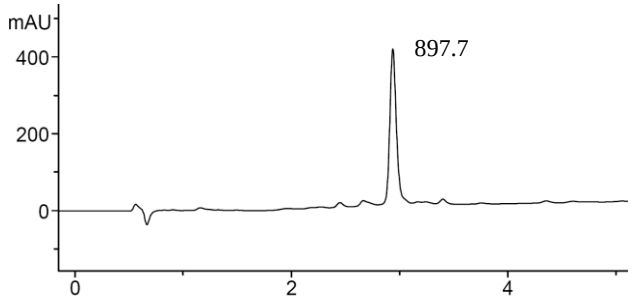


Figure 11. LC-MS chromatogram of crude **131**.

Glycopeptide **131** was therefore purified with semi-preparative HILIC chromatography (**Figure 12**): eluant: CH_3CN:H_2O 7:3 (100 mM ammonium bicarbonate/formate buffer at pH=7.5; the preparation of buffer solution is described in the experimental part), flux: 5ml/min; 215 nm; t_r (**131**) = 12.7 min. Finally, 5 mg of **131** were obtained (isolated yield = 70%). The overall yield of solid

phase synthesis plus deprotection and isolation of **131** (n = 2) was 22%, starting from 25 μmol of the first amino acid loaded on the resin (5 steps, 2 purifications).

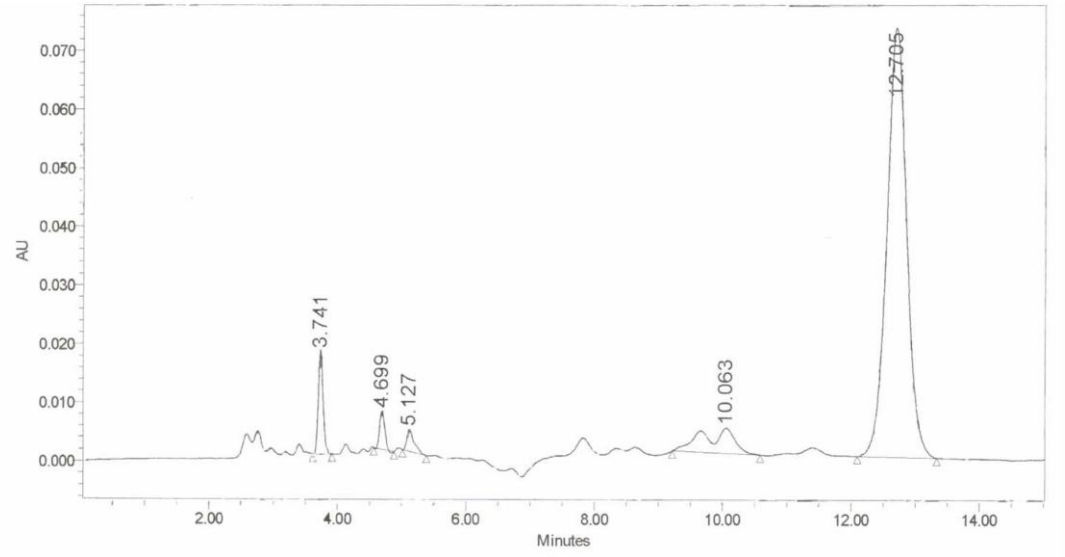
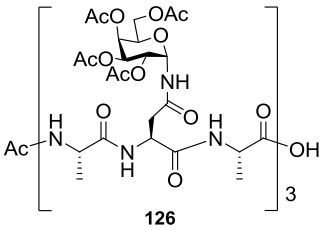


Figure 12:HILIC semipreparative chromatogram of **131**.

Compound **126** (n = 3) was obtained from solid phase synthesis with a global yield = 44%, as estimated by UV measurement after Fmoc-removal for each coupling reaction.



126

The crude (35 mg, ^{1}H NMR spectrum in Section 7.6) was purified on reverse phase chromatography (**Figure 13**): C-18 Biotage cartridge (12 g); eluant: H_2O/CH_3CN (0.1 % TFA); flux: 15ml/min; gradient: 8%→100% CH_3CN in 11 min.

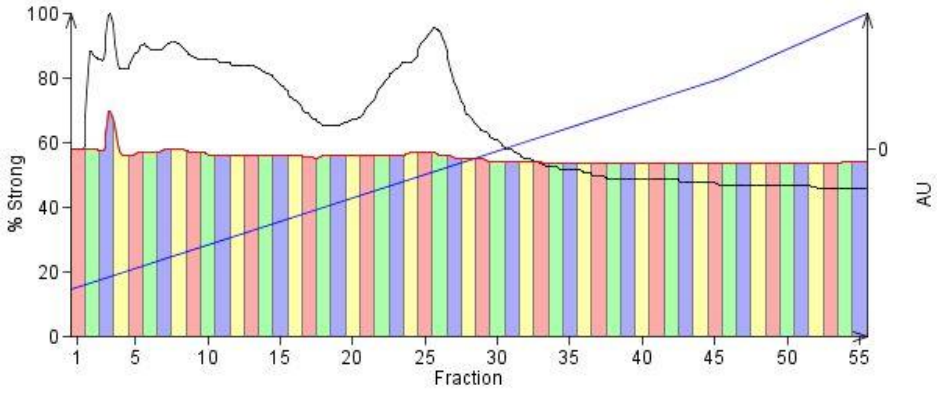


Figure 13. C-18 reverse phase chromatogram of **126**.

Fractions 22-25 and Fractions 26-30 (t_r = 4.5-6 min) were analyzed by LC-MS (gradient: CH_3CN 0%→30% in 6 min) and were found to contain compound **126** and some partial-deacetylation products deriving from it (Figure 14).

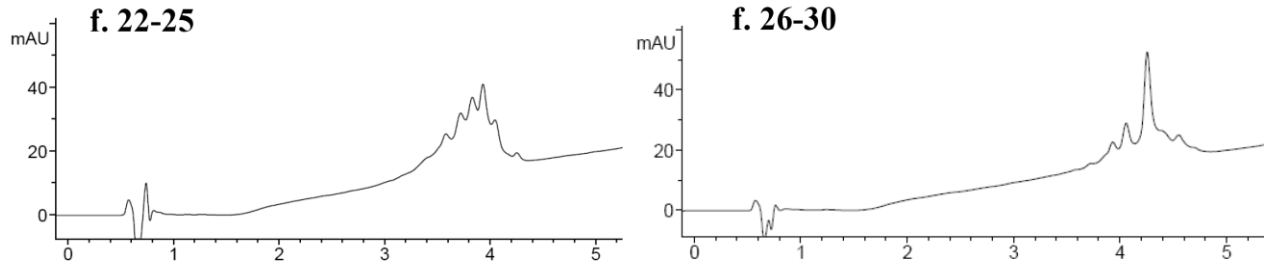


Figure 14. LC-MS chromatogram of f.22-25 and f.26-30 of chromatogram in Figure 13.

These fractions (11 mg) were assembled and subjected to *O*-acetyl removal to obtain **132** (7 mg, 90 % yield, ^{1}H NMR spectrum reported in Section 7.5), which upon LC-MS analysis turned out to be a relatively pure crude (**Figure 15**, [M]), thus confirming the usefulness of the reverse phase purification of the *O*- acetyl glycopeptide.

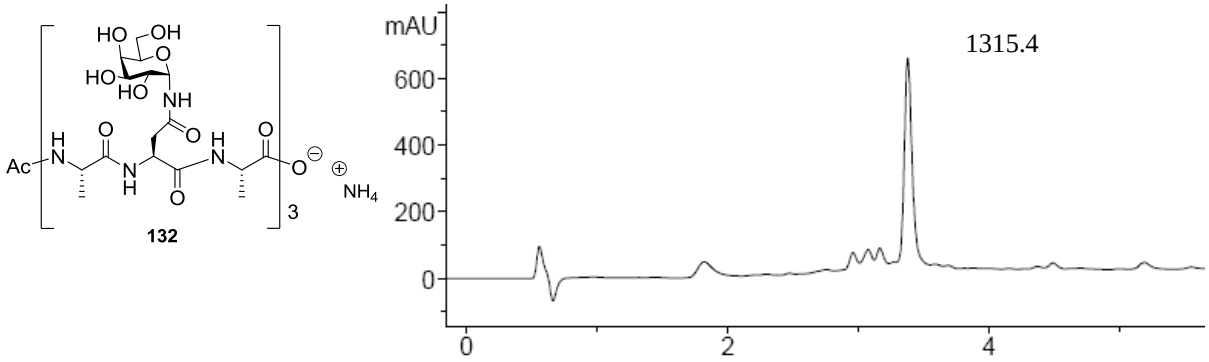


Figure 15. LC-MS chromatogram of crude **132**.

Unfortunately, compound **132** was lost during an attempt of purification by semi-preparative HILIC chromatography. It was in fact at this point that we noticed the inadequacy of the method used previously for purification of glycopeptide **125** (100 mM buffer, CH_3CN:H_2O 7:3, pH 7.5). We suppose that broadening of the peaks together with longer retention time and revealer saturation by the buffer, made compound **131** invisible above the base-line absorbance at 215 nm. This behaviour was further observed for compound **133** and the problem was solved by reducing the concentration of the buffer from 100 mM to 10 mM (see below).

Compound **127** (n = 4) was obtained from solid phase synthesis with a global yield = 28%, determined from UV-absorbance of the Fmoc-piperidine adduct for each coupling reaction. The crude (43 mg, ^{1}H NMR spectrum in Section 7.6) was purified on reverse phase chromatography

(**Figure 16**): C-18 Biotage cartridge (12 g); eluant: H_2O/CH_3CN (0.1 % TFA); flux: 15ml/min; gradient: 15%→100% CH_3CN in 12 min.

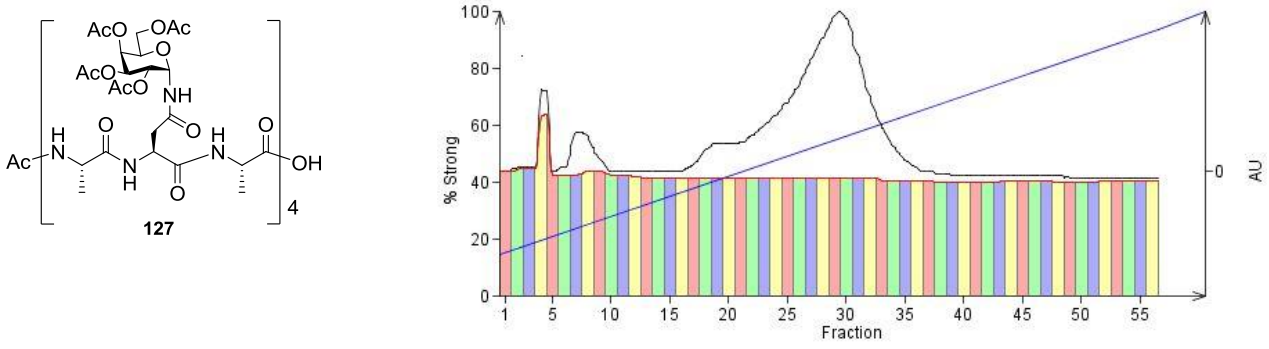


Figure 16. C-18 reverse phase chromatogram of **127**.

Fractions 17-25 were discarded, while fractions 26-28 (8 mg) and fractions 29-34 (10 mg) (t_r = 6-7 min) were found to be very heterogeneous and to contain compound **127** together with some truncated sequences and partial-deacetylation products (LC-MS reported in **Figure 17**, gradient: CH_3CN 0%→30% in 6 min).

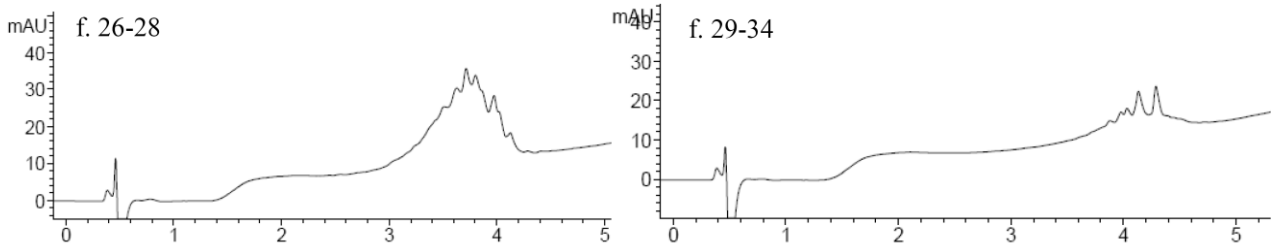


Figure 17. LC-MS chromatogram of f.26-28 and f.29-34 of chromatogram in Figure 16.

These fractions were separately subjected to *O*-acetyl removal to obtain **133a** (4 mg from fractions 26-28, y =70%) and **133b** (6 mg from fractions 29-34, y = 83%). **133a** and **b** were both analyzed with LC-MS analysis (**Figure 18** and **Figure 19**, respectively, gradient: CH_3CN 0%→30% in 6 min), confirming a great heterogeneity for **133a** which contained a consistent amount of truncated sequences, and a relatively superior purity of **133b**.

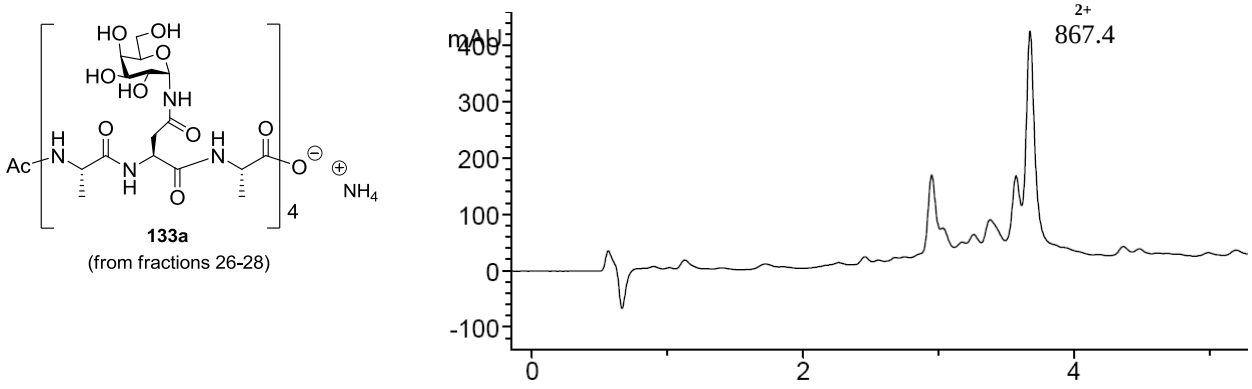


Figure 18. LC-MS chromatogram of crude **133a** [M/2 +2].

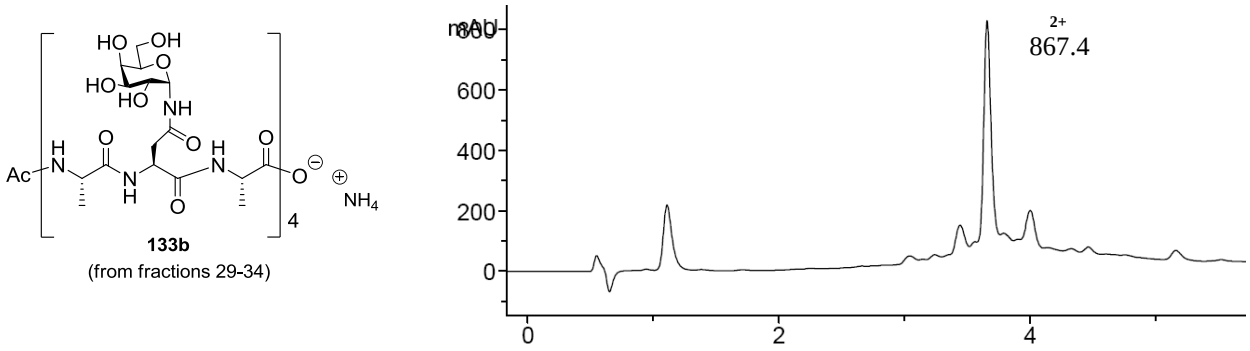


Figure 19. LC-MS chromatogram of crude **133 b** [M/2 +2].

HILIC chromatography was attempted to purify **133a** and **133b**. Starting from the more promising **133b** crude, however, we soon realized that the products' peak (t_r (**133**) = 14 min) was only slightly visible at 215 nm using a 100 mM buffer solution (**Figure 20**, eluant: CH_3CN:H_2O 65:35, 100 mM buffer at pH=7.5 with ammonium bicarbonate/formate; flux: 1ml/min).

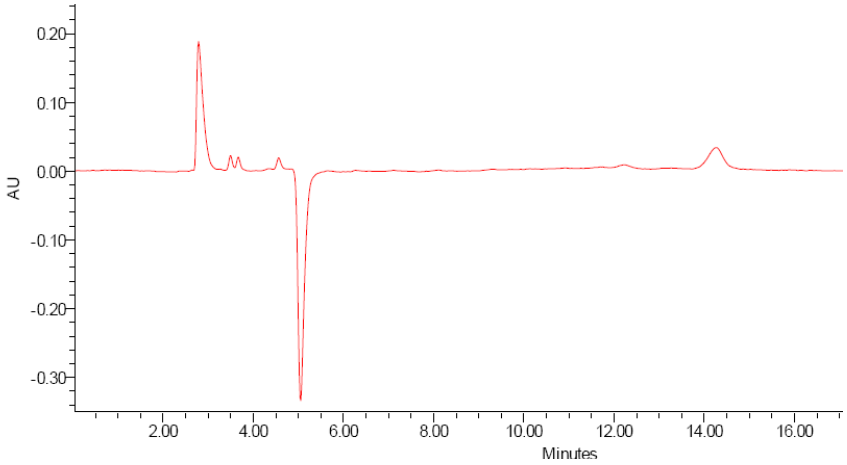


Figure 20. HILIC analytic chromatogram of **133b** (100 mM buffer solution).

This prompted us to examine a lower ionic strength for the buffer and we were surprised to find how effectively a reduction of buffer molarity from 100 mM to 10 mM could modify the retention times and the resolution of the product mixture (t_r (**133**) = 6 min), despite maintaining the same pH and the same percentage of water in the eluant (**Figure 21**, eluant: CH_3CN:H_2O 65:35, 10 mM buffer at pH=7.5 with ammonium bicarbonate/formate; flux: 1ml/min).

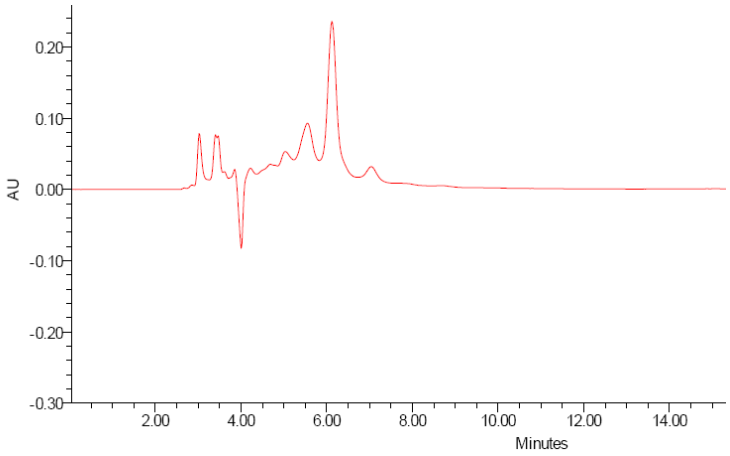


Figure 21:HILIC analitic chromatogram of **133b** (10mM buffer solution).

Finally, **133b** was purified by semi-preparative HILIC cromatography (**Figure 22**): eluant: $CH_3CN:H_2O$ 7:3 (10 mM buffer at pH=7.5 with ammonium bicarbonate/formate); flux: 5ml/min; t_r (**133**) = 9 min. 3 mg of **133**, starting from 6 mg of crude **133b**, were obtained (isolated yield = 42%)

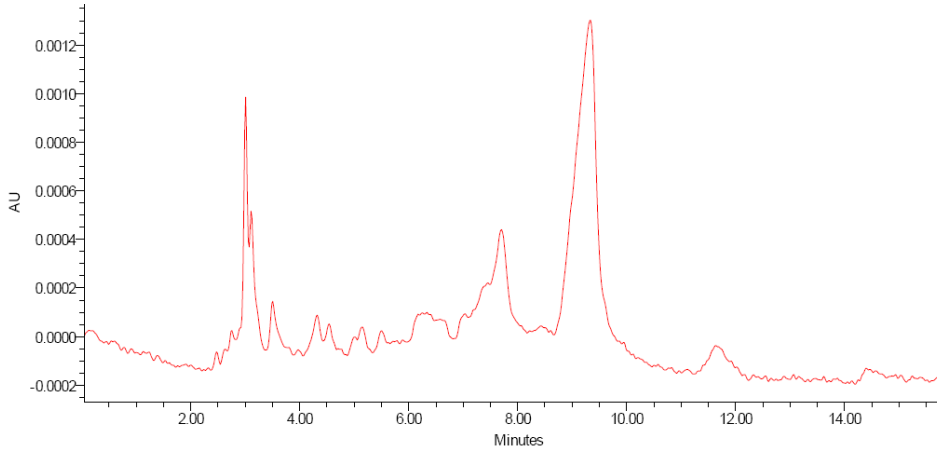


Figure 22: HILIC semipreparative chromatogram of **133b** (10mM buffer solution).

The same conditions were also applied to **133a** (**Figure 23**) which showed a much more complex chromatogram, due to greater amounts of truncated sequences. Only 0.5 mg of **133**, starting from 4 mg of crude **133a**, were obtained.

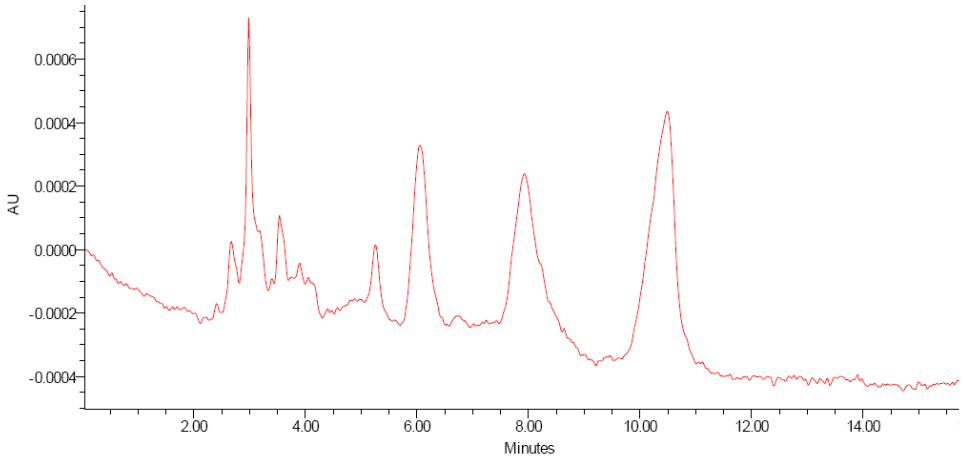


Figure 23: HILIC semipreparative chromatogram of **132a** (10mM buffer solution).

LC-MS analysis for **133** (gradient: CH_3CN 0%→30% in 6 min) is reported in **Figure 24**. The overall yield of solid phase synthesis plus deprotection and isolation of **133** (n = 4) was 8 %, starting from 25μmol of the first amino acid loaded on the resin (11 steps, 2 purifications).

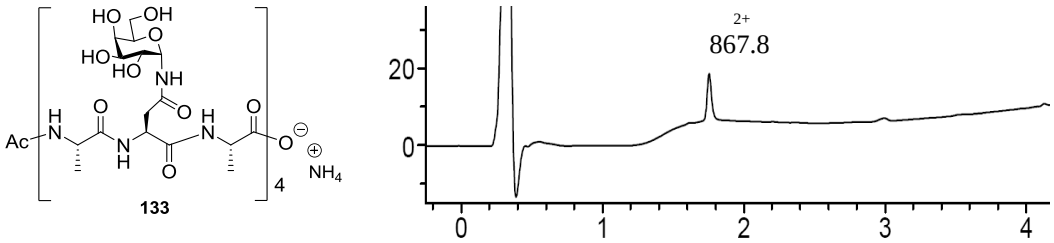


Figure 24. LC-MS chromatogram of **133** [M/2 +2].

Compound **128** (n = 5) was obtained from solid phase synthesis with a global yield = 20%, determined from UV-absorbance of the Fmoc-piperidine adduct for each coupling reaction. The crude (54 mg, ^1H NMR spectrum in Section 7.6) was purified on reverse phase chromatography (**Figure 25**): C-18 Biotage cartridge (12 g); eluants: H_2O/CH_3CN (0.1 % TFA); flux: 15ml/min; gradient: 5%→100% CH_3CN in 14 min.

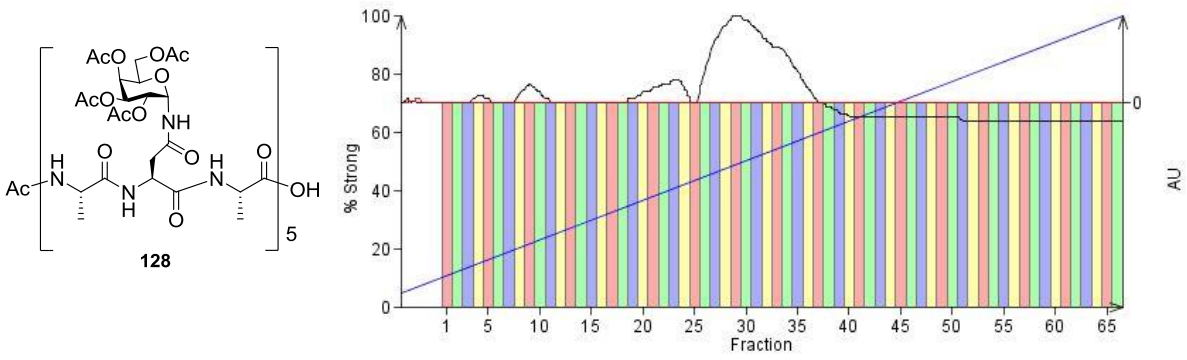


Figure 25. C-18 reverse phase chromatogram of **128**.

Fractions 26-30 (10 mg) and fractions 31-35 (12 mg) were very heterogeneous mixtures, containing compound **128**, truncated sequences and partial-deacetylation products (LC-MS reported in **Figure 26**).

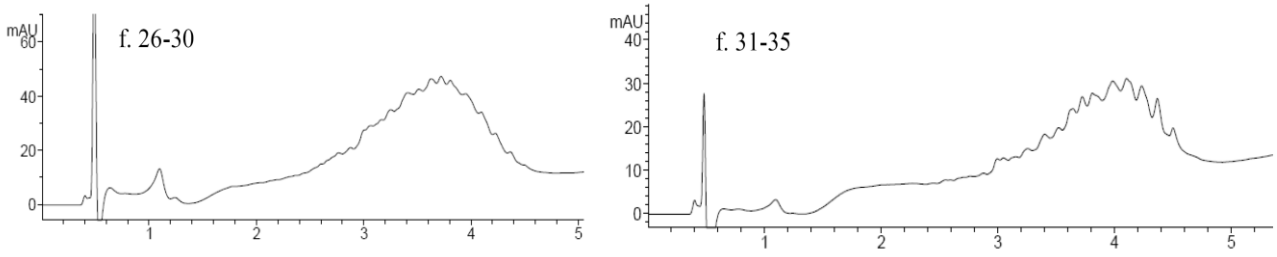


Figure 26. LC-MS chromatogram of f.26-30 and f.31-35 of chromatogram in Figure 25.

These two fractions were separately subjected to *O*-acetyl removal to obtain **134a** (6 mg from fractions 26-30, y =83%) and **134b** (7 mg from fractions 31-35, y =81%). **134a** and **b** were both purified by semi-preparative HILIC chromatography: eluant: $CH_3CN:H_2O$ 7:3 (10 mM buffer at pH=7.5 with ammonium bicarbonate/formate); flux: 5ml/min; t_r (**134**) = 12 min (**Figure 27** and **Figure 28** respectively).

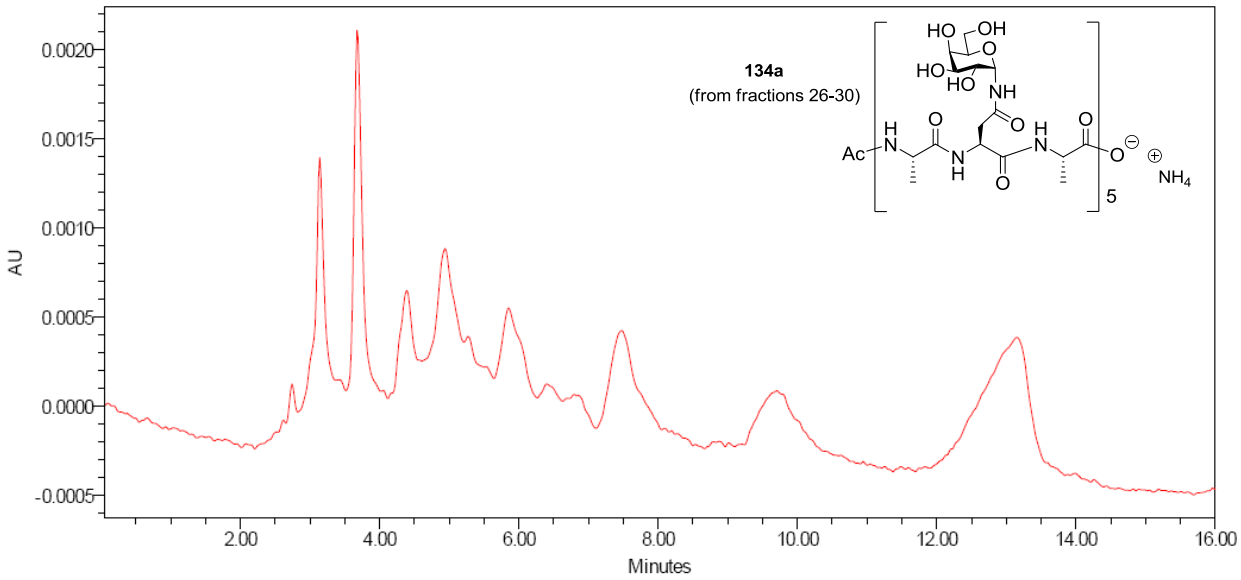


Figure 27. HILIC semi-preparative chromatogram of **134a**.

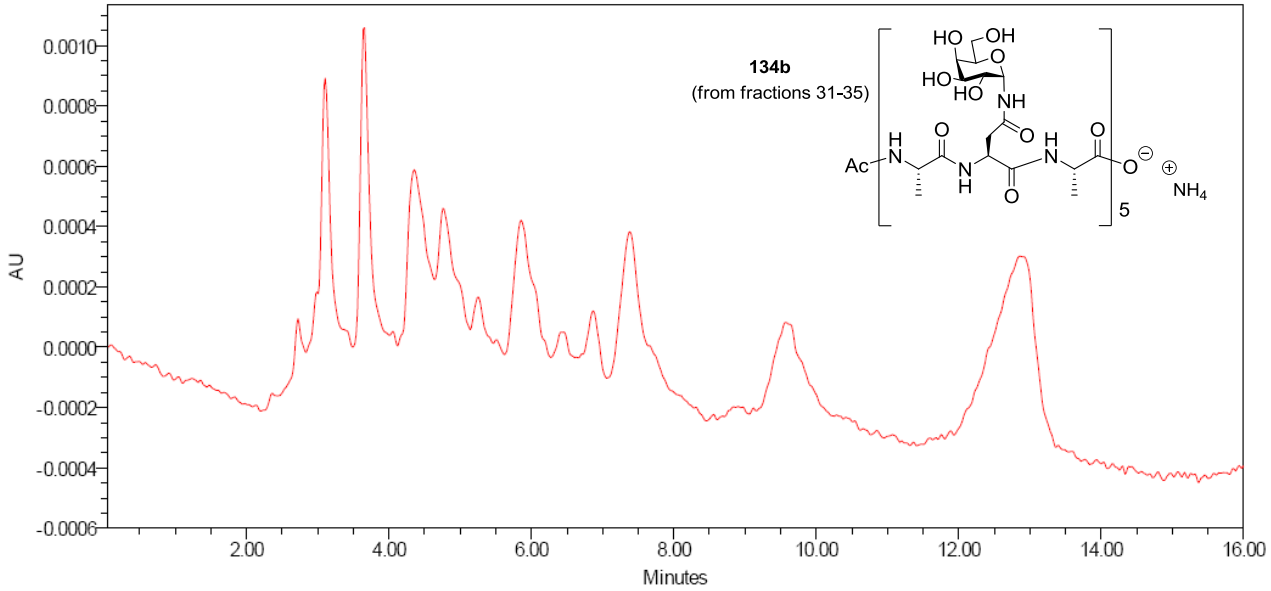


Figure 28. HILIC semi-preparative chromatogram of **134b**.

A total amount of 3 mg of **134** was obtained from **134a** and **b** (**Figure 29** shows the chromatogram of pure compound). The overall yield of solid phase synthesis plus deprotection and isolation of **134** (n = 5) was 6%, starting from 25μmol of the first amino acid loaded on the resin (14 steps, 2 purifications).

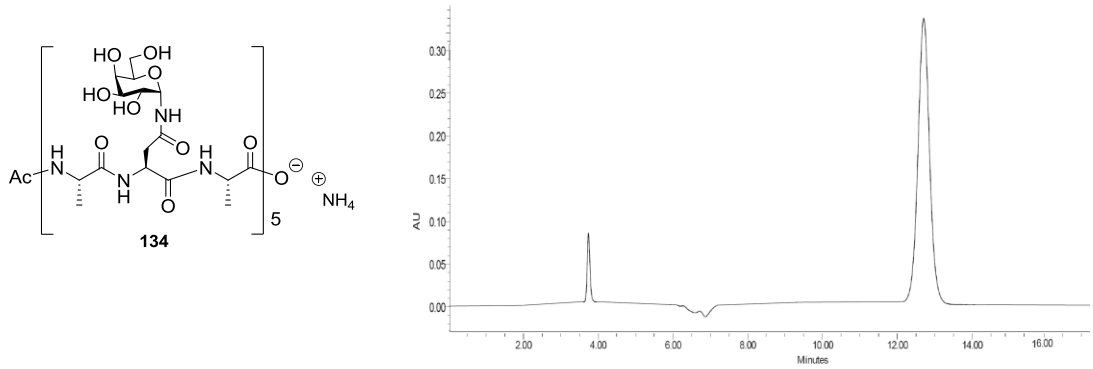


Figure 29. HILIC analytic chromatogram of **134**.

Characterization of all α-*N*-linked glycopeptides is reported in the Experimental Section (Section 8.6).

8.5 Conclusion

Table 1 summarises the results obtained with solid phase synthesis, deacetylation and purification of the α-*N*-linked glycopeptides (n = 2, 3, 4, 5).

Table 1. Reaction conditions for the synthesis of **126**.

Repeat (n)	Reaction steps [a]	UV yield (%) [b]	Isolated yield (%) [c]	Deprotection yield (%) [d]	Isolated deprotection yield (%) [e]	Overall yield /%)
2	5	80	50	95	70	22
3	8	44	24	90	-[f]	-[f]
4	11	28	30 [g]	80	42	8
5	14	20	29 [b]	82	18	6

[a] Number of coupling steps (since the first alanine is preloaded on the resin, the number of steps is equal to (n_{repeat}x3-1). [b] As evaluated spectroscopically after Fmoc removal. [c] After C-18 purification. [d] *O*-acetyl removal. [e] After HILIC purification. The low value, in comparison to deprotection yield, reflects the fact that the deacetylated mixtures contained truncated sequences. [f] not isolated. [g] Maximum isolated yield: the isolated fraction contained the desired compound together with truncations and deacetylated compounds.

Clearly, yields of this process could be improved, probably mostly by improving yields in the synthetic sequence, which would reduce the amount of truncated sequence products in the crude mixture released from the resins. Nonetheless, this work has allowed to establish a reliable protocol for the isolation and purification of glycopeptides up to 15 residues and 5 sugars, and has provided us with samples of most of the desired peptides in mg quantities for conformational studies and for the analysis of their potential antifreeze properties.

8.6 Experimental Section

Biotage® SNAP KP-C18-HS cartridges were used for reverse phase chromatography, performed with a Biotage SP1 Flash purification system with a dual-wavelength detector (254 and 210 nm). HPLC chromatography was performed with a double Waters 515 pump coupled with a photodiode array detector Waters 996. A Phenomenex LUNA HILIC (5micron, 250x4.60 mm) was used for analytic HPLC chromatography. A Phenomenex LUNA HILIC (5micron, 250x10 mm) was used for semi-preparative HPLC chromatography. Absorbance was measured at 215nm. ^{1}H- and ^{13}C-NMR spectra were recorded at 400 MHz on a Bruker AVANCE-400 instrument. Chemical shifts (δ) for ^{1}H and ^{13}C spectra are expressed in ppm relative to internal Me_4Si as standard. Signals were abbreviated as s, singlet; bs, broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet. Mass spectra were obtained with a Bruker ion-trap Esquire 3000 apparatus (ESI ionization) or Ft-ICR Mass spectrometer APEX II & Xmass 4.7 Magnet software (Bruker Daltonics). HPLC-MS analyses were perfomed with Agilent 1100 with quaternary pump, diode array detector, autosampler, thermostated column holder coupled with MS: Bruker ion-trap Esquire 3000 with ESI ionization. The HPLC column was a Waters Atlantis 50x4.6 mm, 3 µm.

Method: Phase A: Milli-Q water containing 0.05 % (v/v) TFA.

Phase B: Acetonitrile (LC-MS grade) containing 0.05 % TFA.

Flow: 1 mL/min, partitioned after UV detector (50 % to MS ESI), Temperature: 40°C.

Gradient: from 0 % B to 30 % B in 6 min, from 30 % B to 90 % B in 1 min, washing at 90 % B for 1 min, equilibration at 0 % B in the next 3 min.

Preparation of 1L of a 100 mM ammonium bicarbonate/formate buffer at pH 7.5

A 1 M solution of ammonium bicarbonate in water was prepared.

For the preparation of 1L of eluant with an acetonitrile:water ratio a:b (typically 65:35 or 70:30):

(a x 10) mL of acetonitrile, 100 mL of 1M ammonium bicarbonate solution and [(b x 10)-100] mL of water were mixed. Example: 1L of acetonitrile:water 65:35 (100 mM) = 650 mL of acetonitrile, 100 mL of 1M NH_4HCO_3 solution, 250 mL of water. Typically the solution has a pH around 8.5-9 and to adjust it to pH 7.5, some drops of formic acid are added, while monitoring the pH with a pHmeter.

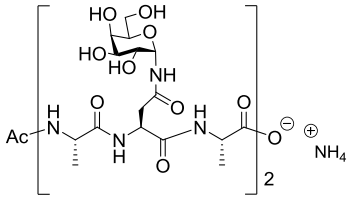
Preparation of 1L of a 10 mM ammonium bicarbonate/formate at pH 7.5

A 100 mM solution of ammonium bicarbonate in water was prepared.

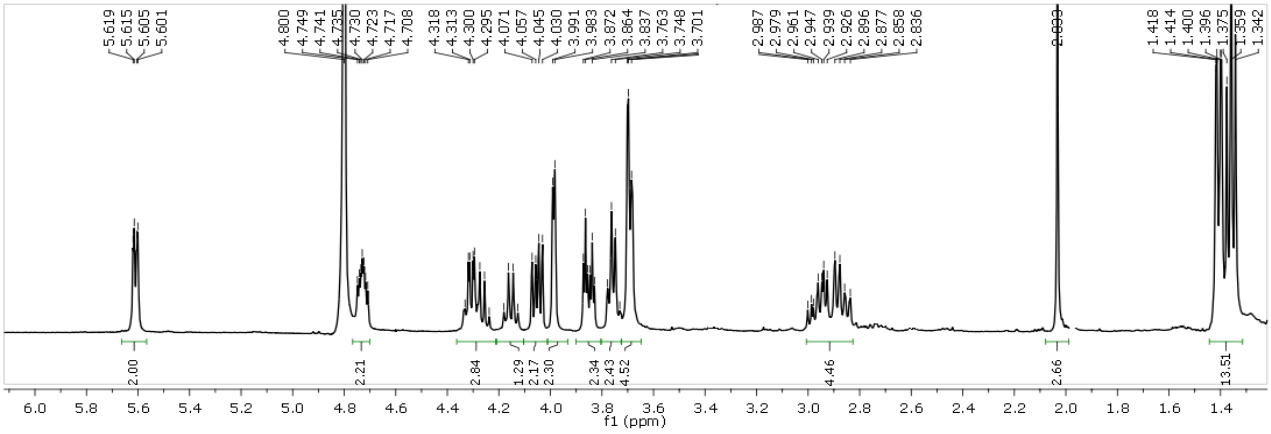
For the preparation of 1L of eluant with an acetonitrile:water ratio a:b (for instance, 70:30):

(a x 10) mL of acetonitrile, 100 mL of the 100 mM ammonium bicarbonate solution and [(b x 10)-100] mL of water were mixed. Example: 1L of acetonitrile:water 70:30 (10 mM) = 700mL of acetonitrile, 100mL of 100 mM solution, 200mL of water. Typically the solution has a pH around 8.5-9 and to adjust it to pH 7.5, some drops of formic acid are added, while monitoring the pH with a pHmeter.

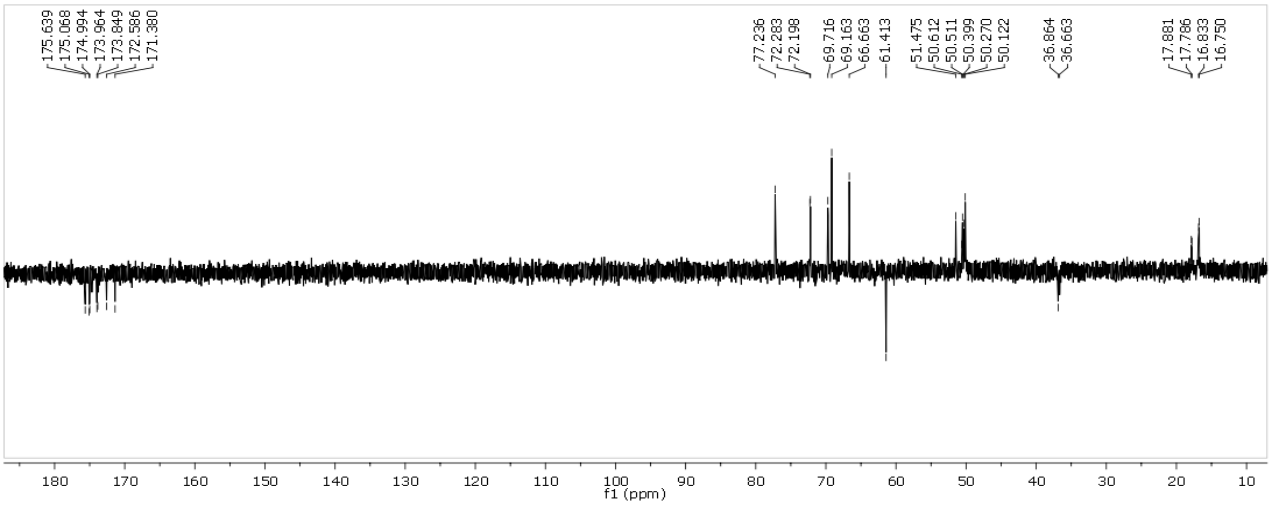
Compound 131 (n = 2)



^{1}H NMR (400 MHz, D_2O, 25 °C): δ = 5.61 (dd, $J_{1,2}$ = 5.6 Hz, 2H, 1-H), 4.72 (m, 2H, α-H-Asn), 4.32-4.21 (m, 3H, α-H-Ala), 4.19-4.10 (m, 1H, α-H-Ala), 4.05 (dd, $J_{2,3}$ = 11.0, $J_{2,1}$ = 5.6 Hz, 2H, 2-H), 3.98 (d, $J_{4,3}$ = 3.2 Hz, 2H, 4-H), 3.87-3.81 (m, 2H, 3-H), 3.78–3.68 (m, 2H, 5-H), 3.68–3.63 (m, 4H, 6-H), 3.00-2.82 (m, 4H, β-H-Asn), 2.03 (s, 3H, NHAc), 1.41-1.33 (m, 12H, CH_3Ala) ppm. ^{13}C NMR (300 MHz, D_2O, 25 °C): δ = 175.7–171.4 (CO), 77.2 (C-1), 72.3, 72.1 (C-5), 69.7 (C-3), 69.2 (C-4), 66.7 (C-2), 61.4 (C-6), 51.5 (α-C-Ala), 50.6, 50.5, 50.4, 50.3, 50.1 (α-C-Asn, α-C-Ala), 36.9, 36.7 (β-CH_2-Asn), 17.9 ($NHCH_3$), 17.7, 16.8, 16.7 (CH_3Ala) ppm. FT-ICR MS (ESI): calcd. for $[C_{34}H_{55}O_{20}N_8]^-$ 895.35381; found 895.35128

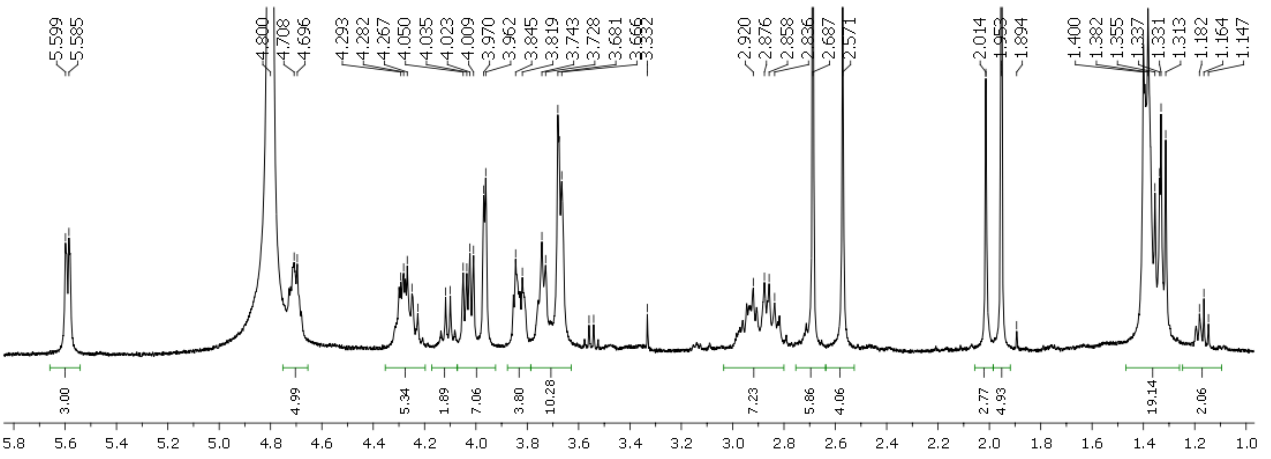


^{1}H-NMR spectrum of **131** (D_2O, 400 MHz)



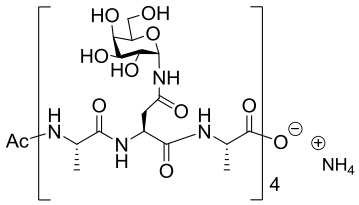
^{13}C-NMR spectrum of **131** (D_2O, 300 MHz)

Compound 132 (n = 3)

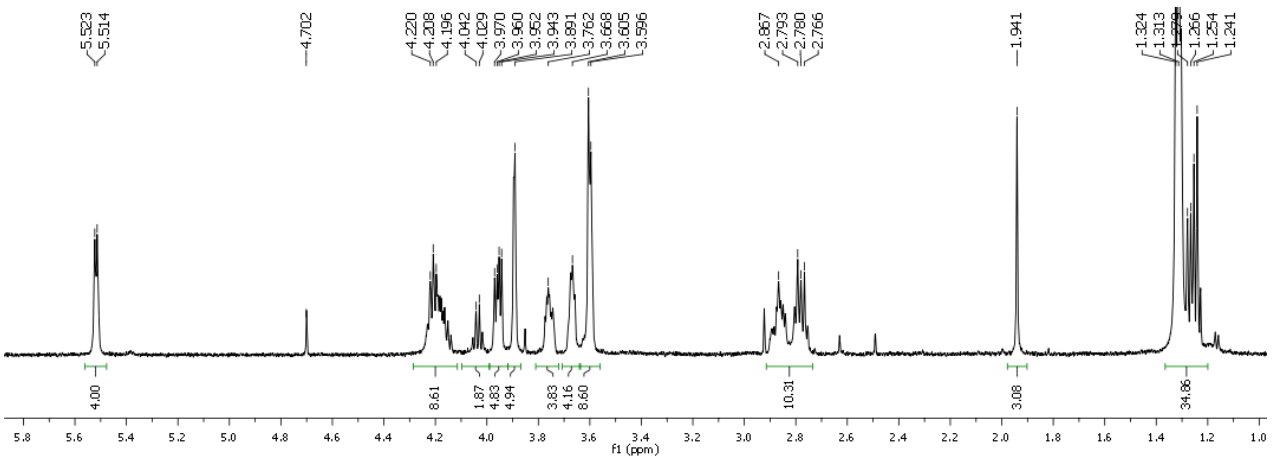


^{1}H-NMR spectrum of the crude **132** (n = 3) (D_2O, 400 MHz) after acetyl removal and before the attempt of purification on HILIC.

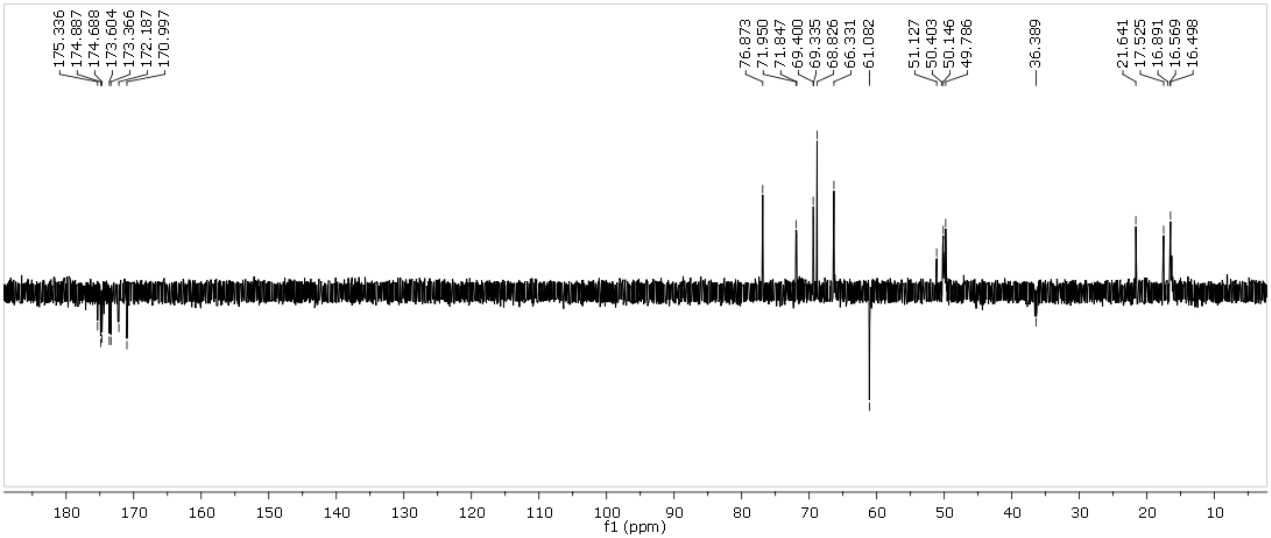
Compound 133 (n = 4)



^{1}H NMR (400 MHz, D_2O, 25 °C): δ = 5.52 (d, J = 3.6 Hz, 4H, 1-H), 4.70 (m, 4H, α-H-Asn), 4.23-4.13 (m, 7H, α-H-Ala), 4.04-4.02 (m, 1H, α-H-Ala), 3.98-3.92 (m, 4H, 2-H), 3.94 (bs, 4H, 4-H), 3.87-3.73 (m, 4H, 3-H), 3.70–3.61 (m, 4H, 5-H), 3.58–3.64 (m, 8H, 6-H), 2.90-2.74 (m, 8H, β-H-Asn), 1.94 (s, 3H, NHAc), 1.32-1.24 (m, 24H, CH_3Ala) ppm. ^{13}C NMR (300 MHz, D_2O, 25 °C): δ = 175.4–171.0 (CO), 76.8 (C-1), 72.0, 71.8 (C-5), 69.4, 69.3 (C-3), 68.8 (C-4), 66.3 (C-2), 61.0 (C-6), 51.1 (α-C-Ala), 50.4, 50.1, 49.8 (α-C-Asn, α-C-Ala), 36.4 (β-CH_2-Asn), 21.6 ($NHCH_3$), 17.5, 16.9, 16.6, 16.5 (CH_3Ala) ppm. LC-MS (ESI): m/z = 867.8 $[M/2 +1 +1]^{2+\cdot}$

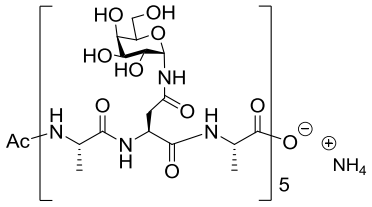


^{1}H-NMR spectrum of **133** (D_2O, 400 MHz)

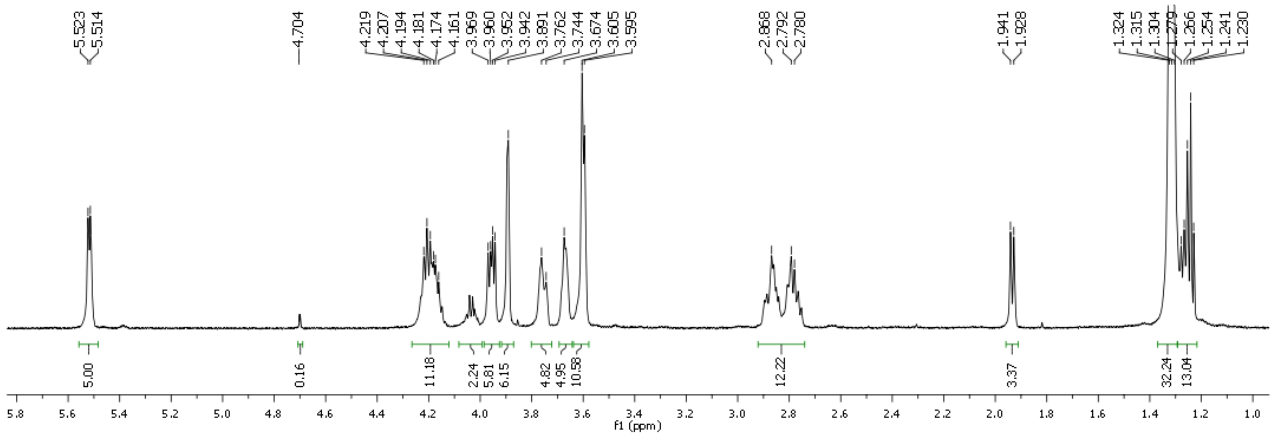


^{13}C-NMR spectrum of **133** (D_2O, 300 MHz)

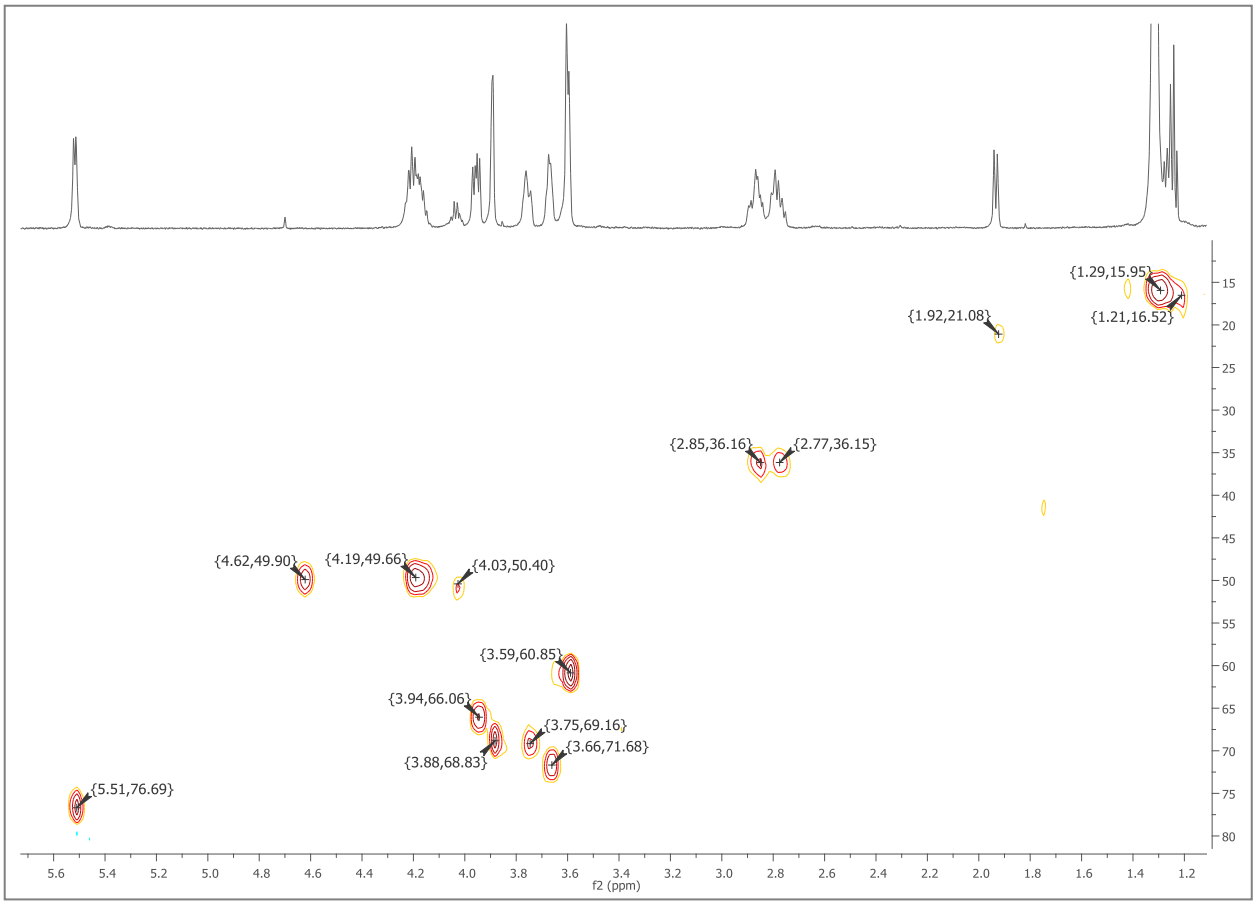
Compound 134 (n = 5)



^{1}H NMR (400 MHz, D_2O, 25 °C): δ = 5.52 (d, J = 3.6 Hz, 5H, 1-H), 4.70 (m, 5H, α-H-Asn), 4.22-4.15 (m, 9H, α-H-Ala), 4.05-4.00 (m, 1H, α-H-Ala), 4.00-3.95 (m, 5H, 2-H), 3.95 (bs, 5H, 4-H), 3.87-3.71 (m, 5H, 3-H), 3.70–3.63 (m, 5H, 5-H), 3.58–3.64 (m, 10H, 6-H), 2.90-2.74 (m, 10H, β-H-Asn), 1.93 (d, 3H, NHAc), 1.32-1.24 (m, 30H, CH_3Ala) ppm. ^{13}C NMR (300 MHz, D_2O, 25 °C): 76.7 (C-1), 72.7 (C-5), 69.2 (C-3), 68.8 (C-4), 66.1 (C-2), 60.8 (C-6), 50.4 (α-C-Ala), 49.9 (α-C-Asn), 49.7 (α-C-Ala), 36.1, 36.2 (β-CH_2-Asn), 21.1 ($NHCH_3$), 16.5, 16.0 (CH_3Ala) ppm. LC-MS (ESI): m/z = 1098.8 $[M/2 +23+23]^{2+}$. MALDI-MS: m/z = 2175.7 $[M+23]^{+}$.



^{1}H-NMR spectrum of **134** (D_2O, 400 MHz)



HSQC-(^{1}H-^{13}C)-NMR spectrum of **134** (D2O, 400 MHz)

8.7 References

1 Boschelli, D. H.; Powell, D.; Sharky, V.; Semmelhack, M. F. *Tetrahedron Lett.* **1989**, *30*, 1599-1600.

2 Angyal, S. J. *Carbohydr. Res.* **1994**, *263*, 1-11.

3 Hemström, P.; Irgum, K. *J. Sep. Sci.* **2006**, *29*, 1784-1821.

4 Alpert, A. J. *J. Chromatogr.* **1990**, *499*, 177–196.

5 a) Garbis, S. D.; Melse-Boonstra, A.; West, C. E.; van Breemen, R. B. *Anal. Chem.* **2001**, *73*, 5358–5364. b) Hsieh, Y. S.; Chen, J. W. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3031–3036. c) Wuhrer, M.; Koeleman, C. A. M.; Hokke, C. H.; Deelder, A. M. *Anal. Chem.* **2005**, *77*, 886–894. d) Mizzen, C. A. *Meth. Enzymol.* **2004**, *375*, 278–297.

6 Nguyen, H. P.; Schug, K. A. *J. Sep. Sci.* **2008**, *31*, 1465-1480.

7 Yoshida, T. *J. Biochem. Biophys. Meth.* **2004**, *60*, 265–280.

Chapter 9
Conformational analysis and molecular recognition of α-*N*-linked glycopeptides: Preliminary studies

9.1 Introduction

Analogs of glycopeptides have been thoroughly investigated throughout the years, due to the relevance of natural glycopeptides and glycoproteins in human health and disease [1] (Chapter 1). We have focused our efforts on the synthesis of unnatural unnatural α-*N*-linked glycopeptides, described in this Thesis, with the scope of studying the behaviours and the properties of these molecules, which represent novel and unknown materials. From a chemical point of view, for instance, we noticed that the α-*N*-linked glycosyl building blocks were prone to cyclization upon activation at the C-terminus with some commonly used coupling reagents,[2] the contrary to he natural β-*N*-linked glycosyl building blocks, which are described in literature to be more easily coupled[3] (Section 6.2). Moreover, α-*N*-linked glycopeptides, when unprotected at the sugar moieties, turned out to be unstable towards basic and acidic condition (Section 6.5), and must be handled in a limited range of pH (4.5≤pH ≤9).

Essentially, α-*N*-linked glycopeptides with repeating-units (Ala-(α-Gal)-Asn-Ala) structure like those reported in **Figure 1**, have been synthesized in order to test their possible anti-freeze activities. Molecule **134,** in fact, has been sent to the group of Robert Ben to evaluate this possible feature.

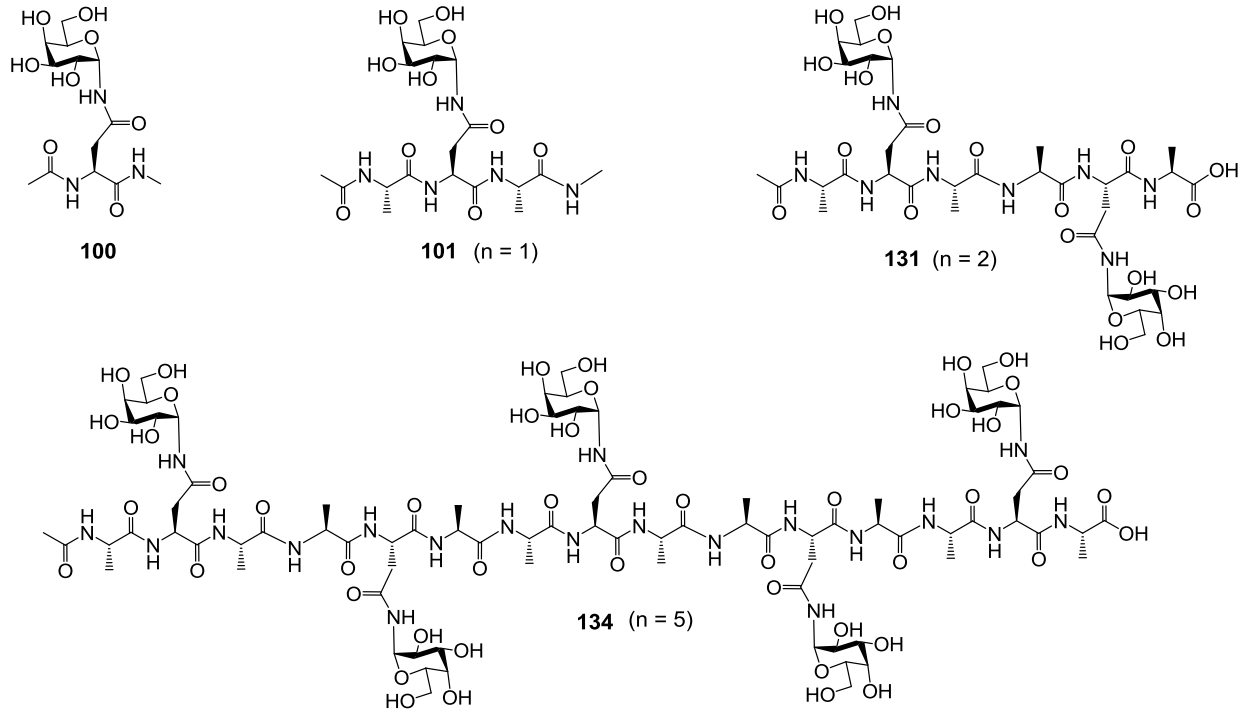


Figure 1. α-*N*-linked glycopeptides examined in this chapter.

However, from a conformational point of view, the only study that examined an unnatural α-*N*-linked glycopeptide was reported several years ago by Imperiali and Woods (Chapter 1, pp 20-21).[4] The α-*N*-linked glycopeptide was found to have a conformation similar to the unglycosylated peptide and different from the *β*-*N*-linked glycopeptide. The peptide conformation of *N*-linked glycopeptides was hence found to depend on the anomeric configuration of the appended glycan. On this basis, some of the recently synthesized α-*N*-linked glycopeptides (**Figure 1**) were conformationally investigated and the results are described in this Chapter.[5] In particular, computational studies and molecular dynamics simulations on glycopeptides **100**, **111**, **131**, **134** were performed by Fabio Doro, from our laboratory, and are described in Section 9.2. The 3D solution structures were determined through NMR-based experiments by Filipa Marcelo (from Jimenez-Barbero's group) for glycopeptides **100** and **101** and by Francesca Vasile, at the University of Milan, for glycopeptide **131**. Section 9.4 is dedicated to the description of the interactions of glycopeptides **100** and **101** with two legume galactose specific lectins, *Viscum album* agglutinin (VAA),[6] and *erythrina cristagalli* lectin (Coral tree)[7], by STD-NMR and TR-NOESY experiments performed by F. Marcelo.

Hence, a multidisciplinary approach for determining the features of non-natural α-Gal-*N*-glycopeptides has been adopted combining synthetic methods, conformational analysis, NMR-based molecular recognition protocols and molecular modeling procedures.

9.2 Computational modelling of α-*N*-linked glycopeptides

A broad literature exists on the peculiar effects of glycosylation on polypeptide conformation. Here, for our peptides which are α-*N*-glycosylated it's interesting to check which changes, if any, are produced by the sugar(s) on the structure of the peptide. In order to understand this, initial studies were performed on molecules **100** and **101** (**Figure 1**) using MC/EM[8]calculations with AMBER* force field using a GB/SA continuum water model,[9] as implemented in Macromodel[10]/Maestro[11]. Moreover a set of *in silico* tests of the same peptides non-glycosylated (where the galactose moiety is replaced by a methyl group) was performed.

For molecule **100** a total number of 41 unique conformers was found in 5 kcal/mol from the global minimum, of which 2 were in the first kcal/mol (**Figure 2**, intramolecular H-bonds are shown in blue).

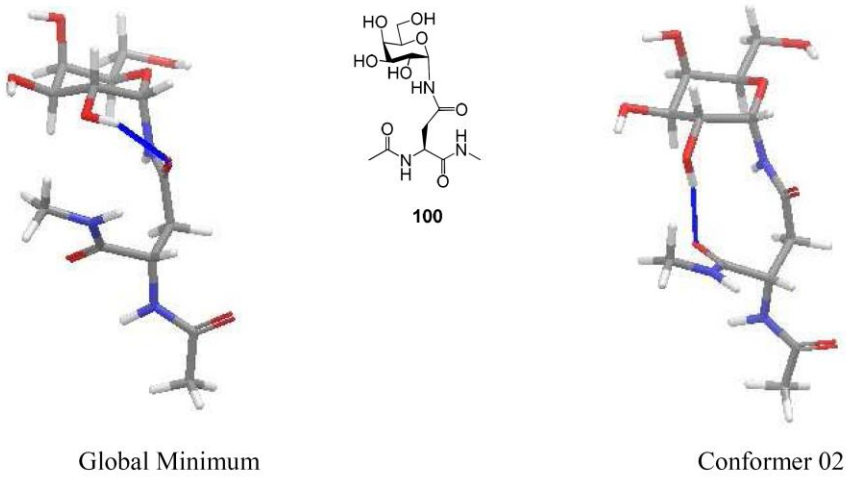


Figure 2. Pictures of conformers 1, 2 of minimum energy for molecule **100**.

For molecule **101**, a total number of 167 unique conformations was found in 5.00 kcal/mol from the global minimum, of which 4 in the first kcal/mol. **Figure 3** shows representative low-energy conformations calculated for **2** and highlights the intramolecular H-bonds predicted

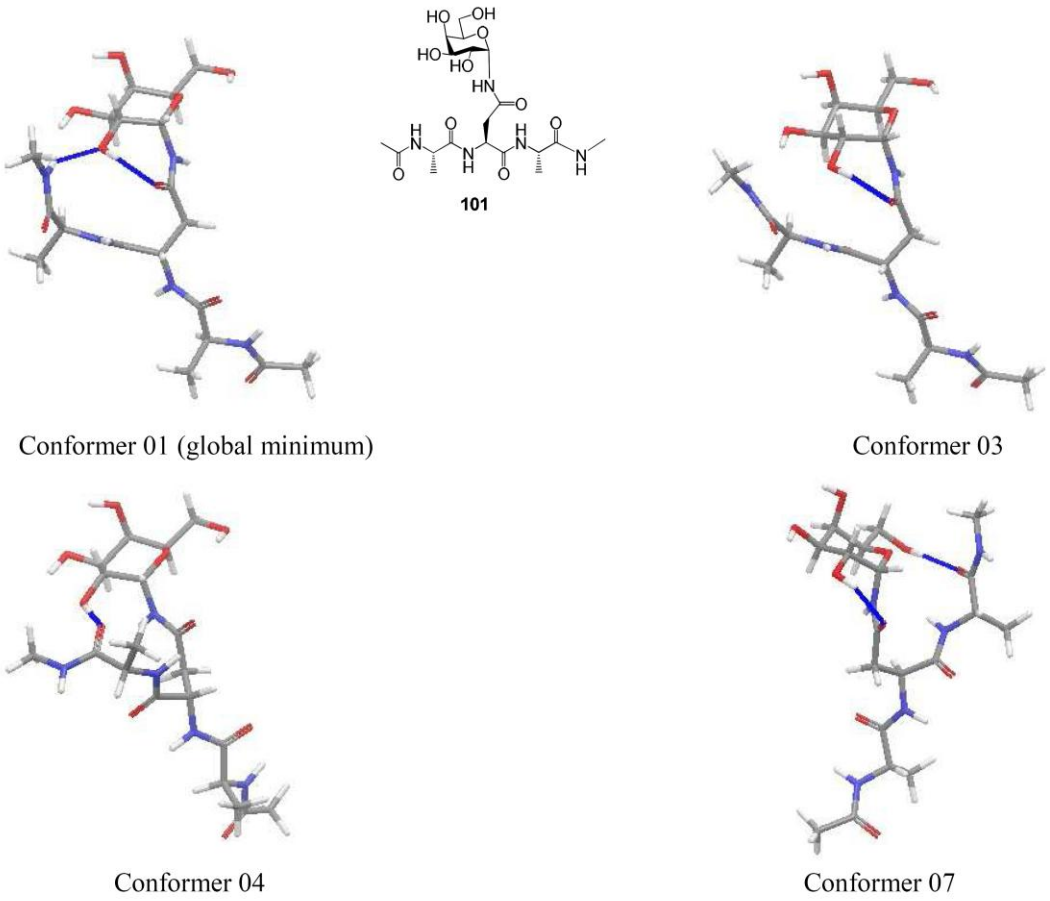


Figure 3. Representative low-energy conformations (within 2 kcal/mol from the global minimum) from the conformational search of **101**.

The calculation predicted mostly extended conformations of the peptide chains for both molecules **100** and **101**, despite the known tendency of the force field to overestimate folded conformations such as γ-turns for small peptides. Dynamic simulations performed both with implicit Macromodel MC/SD,[12] (GB/SA water model) and explicit water (AMBER 9,[13] TIP3P water, periodic boundary conditions) allowed to observe two interesting facts:

a) Extended conformation is greatly preferred.
b) Intramolecular H-bonds between hydroxyl groups of the galactose moiety and acceptor groups in the Asn side chain moiety are present in the most energetically favoured conformations and for the majority of the stochastic dynamics simulation.

These features, which are in agreement with the NMR data illustrated in Section 9.3, may be favored by the formation of H-bond interaction between the sugar and the peptide chain. To the best of our knowledge, no data are available on the preferred conformation of the Ala-Asn-Ala (ANA) sequence (without the sugar moiety), but many AXA tripeptides have been found to exist mostly in extended conformation. A study published in 2004, for example, showed that AXA tripeptides (X

being valine, tryptophan, histidine, and serine) predominantly adopt an extended β-strand conformation while AXA tripeptides for which X is lysine and proline prefer a polyproline II-like (PPII) structure.[14] However, the Ala-Phe-Ala sequence was found to fold as an Inverse γ-Turn in water.[15] Thus, it remains unclear whether the extended conformation observed for **100** and **101** is a direct consequence of peptide glycosylation. However, H-bonds between galactose hydroxyl groups and the Asn side chain could probably encourage this behaviour.

Molecules **131** (n = 2) and **134** (n =5) (**Figure 1, 4** and **5**) possess too many degrees of freedom to analyze them by a full conformational search. Instead, a set of different simulated annealing simulations (where molecules are fast heated and then cooled in a controlled way) was performed, which allows to search for the most stable conformations An implicit water model using MacroModel was used. Simulated annealing experiments for **131** once again reported the extended conformation as the preferred one (**Figure 4**), a feature confirmed by NMR (Section 9.3). MD simulation with explicit water confirms this result, even though here some folded conformations begin to appear. An experiment involving the hexapeptide Ala-Asn-Ala-Ala-Asn-Ala removed of the galactose moiety is yet to be performed.

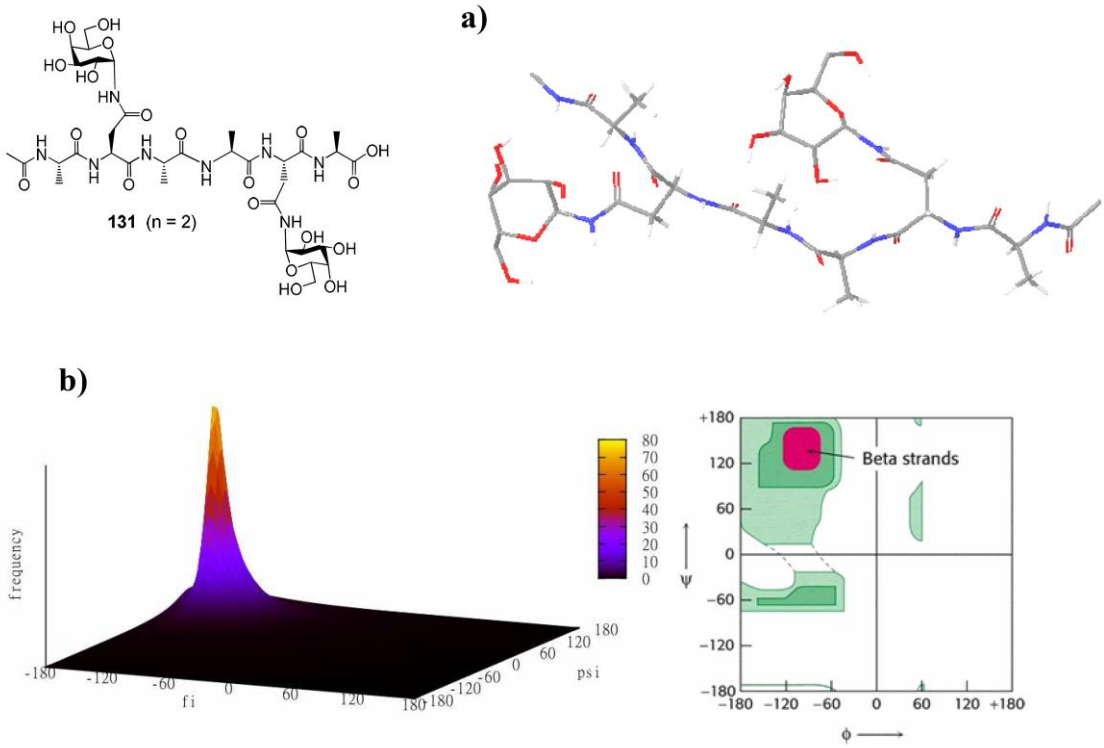


Figure 4. a) Simulating annealing energy minimum for **131** (β-strand arrangement). b) 3D and 2D Ramachandran Plot

On the contrary, simulated annealing and MD simulation experiments of **134** both indicate that now a folded conformation is the preferred one (**Figure 5**). During the MD simulation (which starts from

an extended conformation) a relevant conformational change can be observed (corresponding to a decrease in the radius of gyration of **134**), yielding a folded conformation.

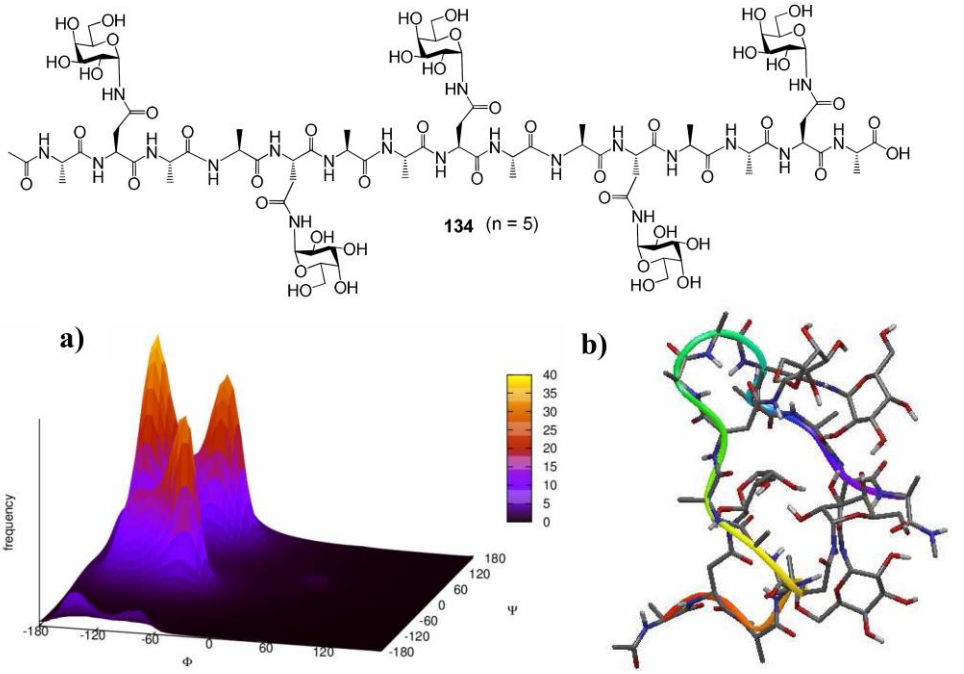


Figure 5. a) 3D Ramachandran Plot for simulating annealing of **134**. b) The lowest energy structure found during the simulated trajectory in TIP3P water.

These predictions cannot be confirmed yet by NMR data. The experimental determination for the preferred solution conformation(s) of **134** is underway in Dr. Vasile's laboratoty.

9.3 Conformational analysis of the free α-N-linked glycopeptides

The solution conformation of the two α-*N*-linked glycopeptides **100** and **101** was first investigated by NMR spectroscopy by F. Marcelo. Coupling constants and NOE data were carefully analyzed and allowed to determine the conformational features of the peptide backbone in water solution (**Figure 6** and **7**).

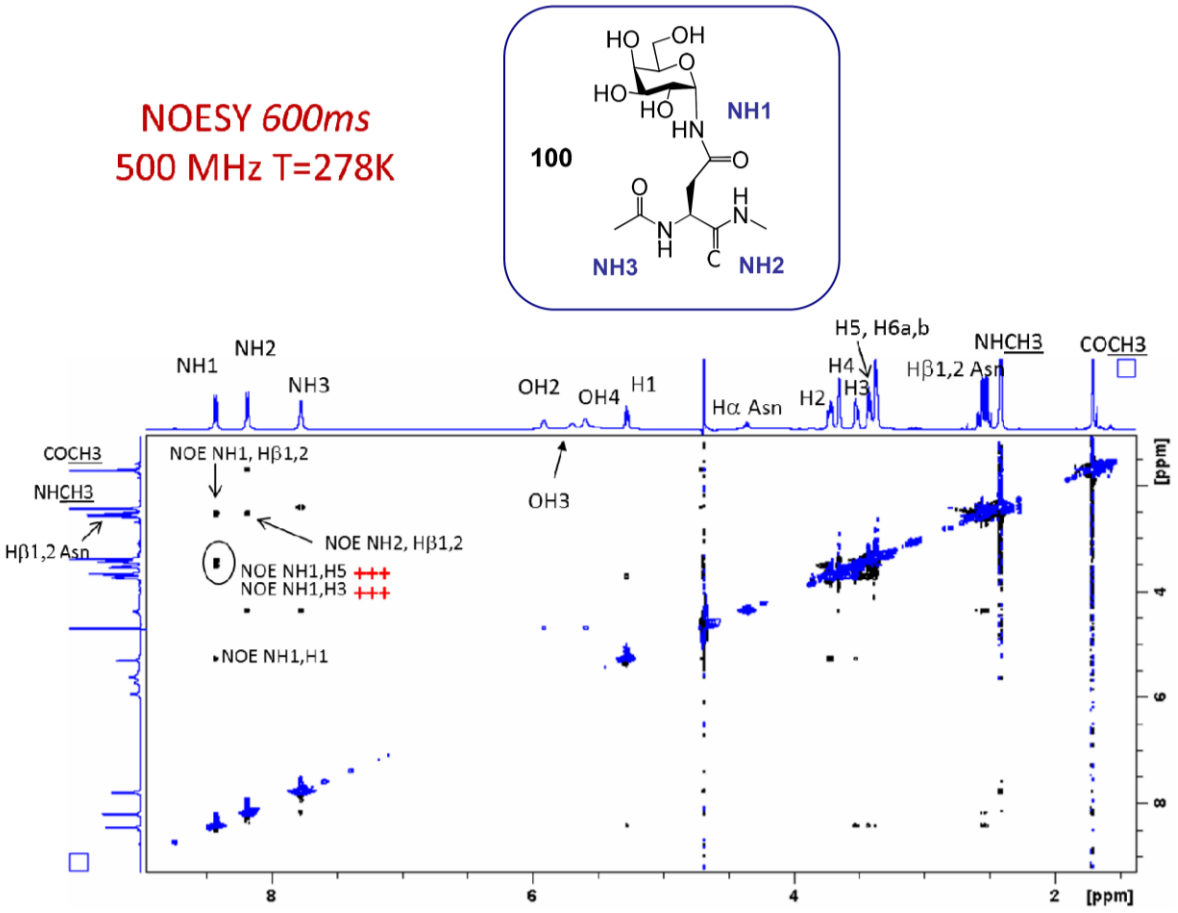


Figure6. 2D-NOESY spectrum obtained for glycopeptide **100** in H_2O/D_20 90:10 recorded at 500MHz with 600ms of mixing time and at 278K.

The experimental $^3J_{NH,H\alpha}$ coupling constants values strongly suggested the presence of an extended conformation for the peptide backbone in water solution, while the $J_{H\alpha,H\beta1}$ / $J_{H\alpha,H\beta2}$ values (5.1/6.5 Hz and 8,1/7.3 Hz, for **100** and **101**, respectively) showed the existence of certain flexibility around χ1 (Hα-Cα-Cβ-Hβ) of both Asn residues. The absence of non-vicinal medium-range NOE contacts supported the notion that glycopeptide **101** adopts, as its main conformation, an extended conformation of the peptide backbone when free in solution (**Figure 7**). Inter-residual NOEs contacts between the side chain NH of Asn amino acid (NH1 on the NMR spectra) and H3 and H5 of the galactose residue were also detected (Figure 7), indicating that the galactose ring adopts the

usual $^4C_1$ chair conformation and that the NH bond is buried below the ring. The analysis of these NOE cross-peaks also revealed the existence of certain degree of flexibility around the glycosidic linkage. Noteworthy, in the case of the **101,** the NOE between NH of Asn and H5 seems to be stronger when compared to that observed between NH of Asn and H3 proton (Figure 7), indicating a preferred orientation of this linkage. In addition, the high value for the glycosidic coupling constant ($^3J_{H1,NH}$ = 8.3 Hz for **100,** and 8.2 Hz for **101**) denotes the existence of a major anti-type orientation between the NH of Asn and H1 of the galactose.

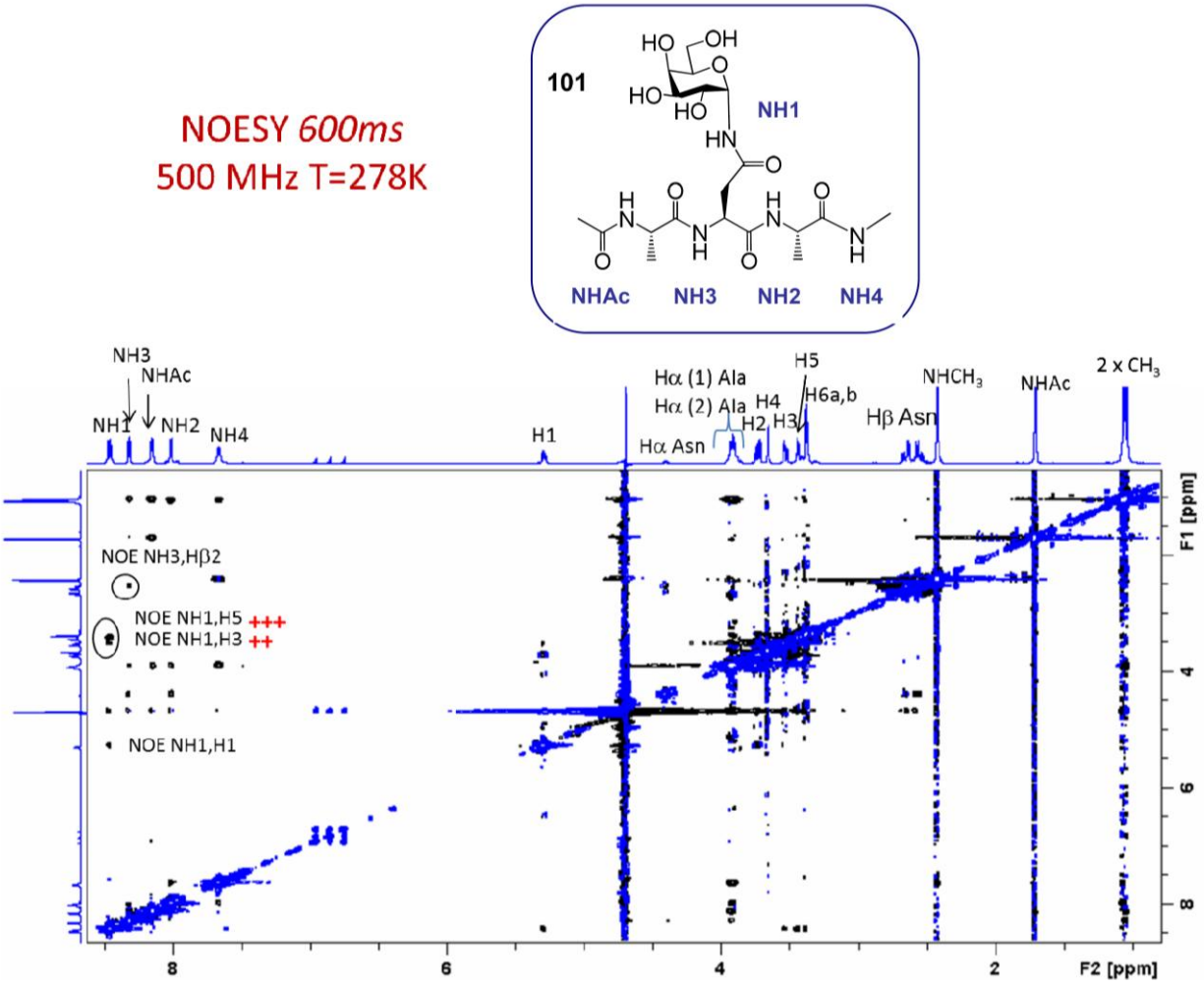
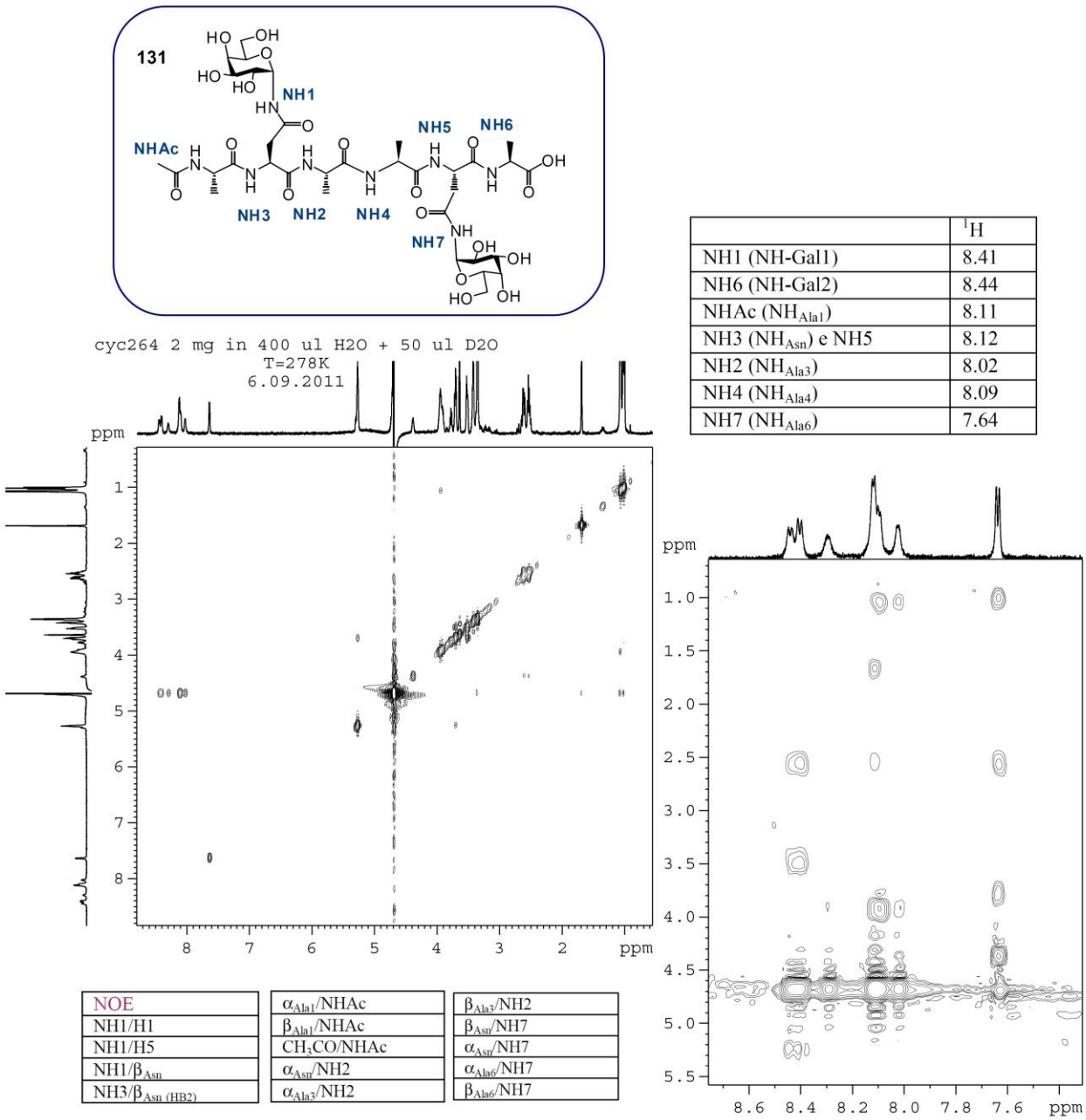


Figure7. 2D-NOESY spectrum obtained for glycopeptide **101** in H_2O/D_20 90:10 recorded at 500MHz with 600ms of mixing time and at 278K.

Molecules **131** (n = 2) was also studied (F. Vasile, **Figure 8**) and the two subunits were found to have overlapping signals. Also for this molecule the extended conformation suggested by MD simulation with explicit water was confirmed. Indeed, as for molecule **101**, the absence of non-vicinal medium-range NOE contacts supported the notion that glycopeptide **131** adopts an extended conformation of the peptide backbone. The galactose rings adopt the usual $^4C_1$ chair conformation

as established by the presence of inter-residual NOEs contacts between the side chain NH of Asn amino acid (NH1/NH6 on the NMR spectrum) and H3 and H5 of the galactose residues (Figure 8).



	^{1}H
NH1 (NH-Gal1)	8.41
NH6 (NH-Gal2)	8.44
NHAc (NH_{Ala1})	8.11
NH3 (NH_{Asn}) e NH5	8.12
NH2 (NH_{Ala3})	8.02
NH4 (NH_{Ala4})	8.09
NH7 (NH_{Ala6})	7.64

Figure 8. 2D-NOESY spectrum for glycopeptide **131** in H_2O/D_20 recorded at 600MHz at 278K.

For the moment, no NMR data are available to confirm the computational findings about glycopeptide **134**, but we count on performing analyses on it as soon as new material will be available.

9.4 Interaction of α-*N*-linked glycopeptides with galactose-binding proteins

The interactions of α-*N*-linked glycopeptides **100** and **101** to the galactose binding lectins VAA[6] and Coral tree[7] was scrutinized. First, STD-NMR data were collected for the individual glycopeptides to establish their respective binding epitopes to the two receptors. Clear STD signals were detected for both molecules with both galactoside specific proteins confirming their specific interaction with both lectins (**Figure 9** showed the STD spectra for interaction with glycopeptide **100, Figure 10-11** for glycopeptide **101**).

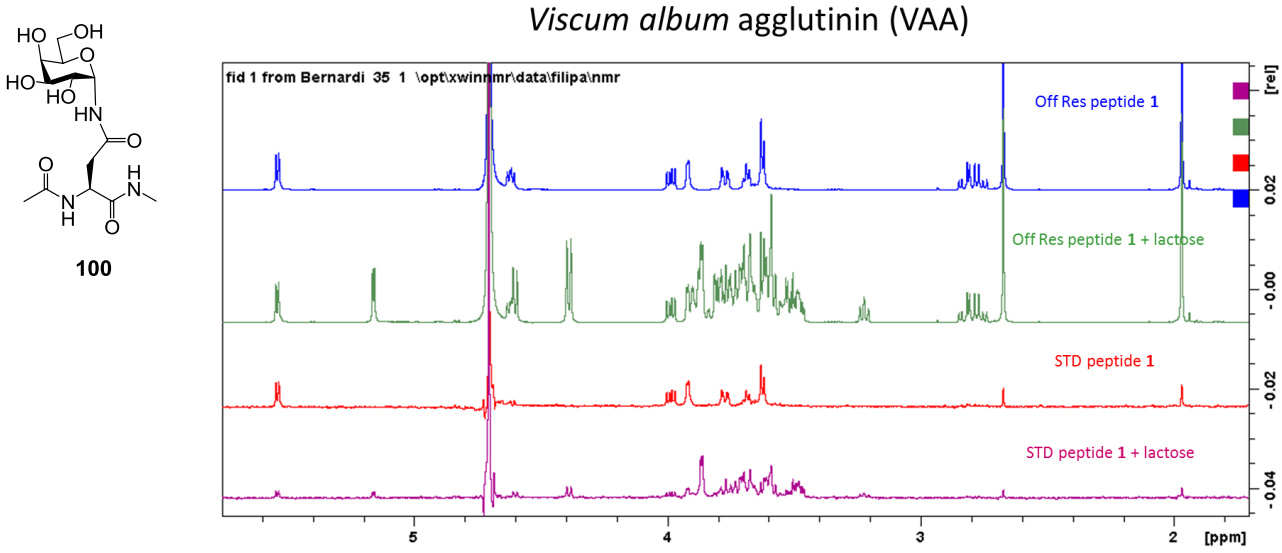


Figure 9. Saturation Transfer Difference spectra for interaction of VAA with glycopeptide **100**.

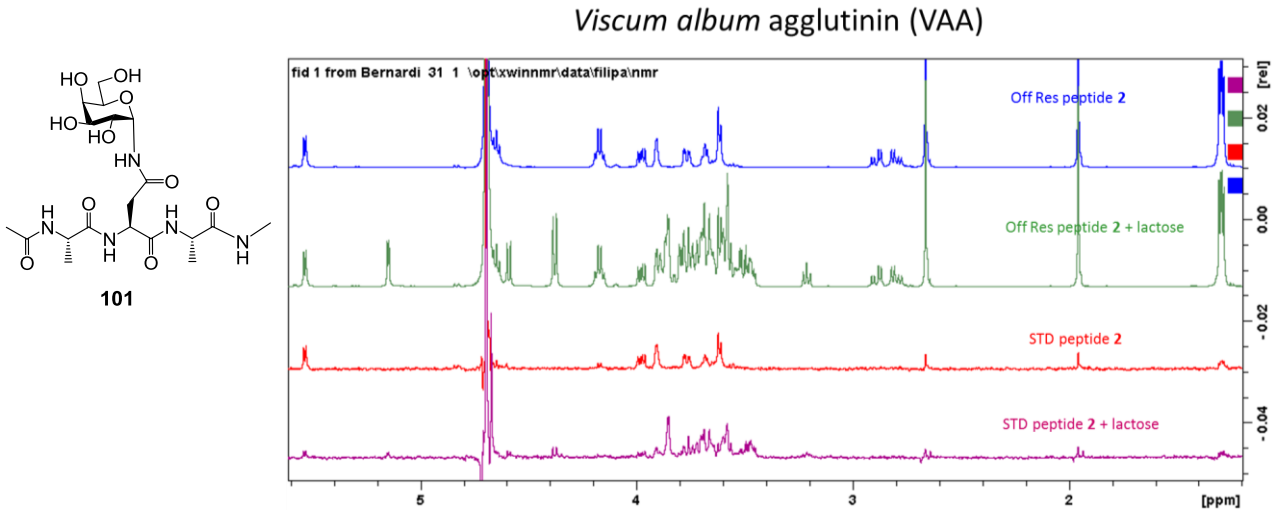


Figure 10. Saturation Transfer Difference spectra for interaction of VAA with glycopeptide **101**.

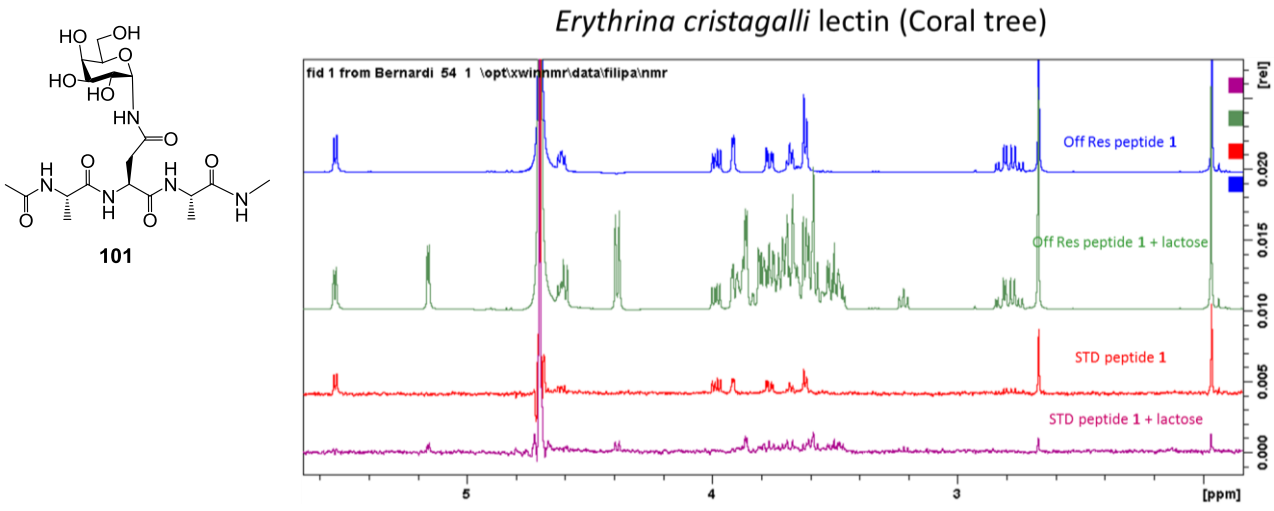


Figure 11. Saturation Transfer Difference spectra for interaction of Coral Tree with glycopeptide **101**.

STD-based competition experiments carried out with both glycopeptides and the proteins, in presence of lactose, clearly indicated that the neoglycopeptides (**100** and **101**) and lactose (chosen as "natural" ligand) indeed compete for the same binding site (Figure 9-10). As can be seen in **Figure 11** and **12**, the most intense signals corresponded, in all cases, to the sugar moiety, highlighting the role of the carbohydrate moiety as the key binding epitope to the lectins.

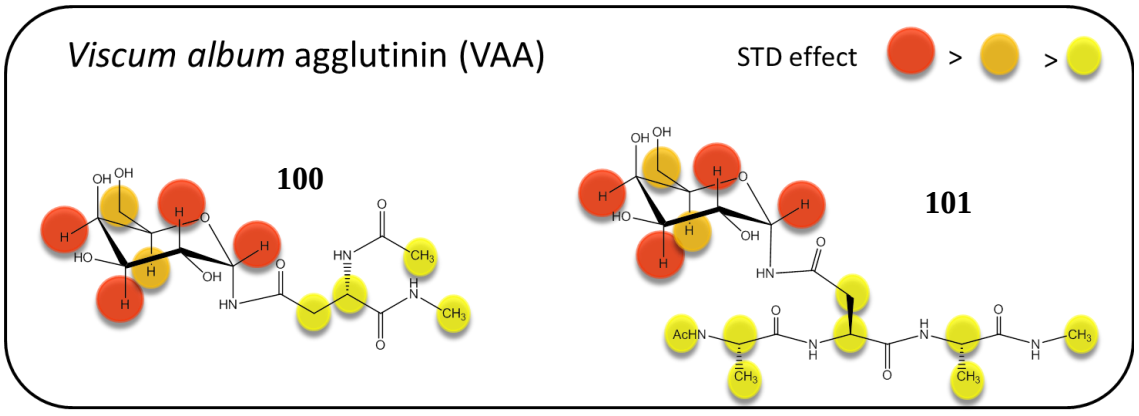


Figure 11. Epitope mapping obtained for glycopeptides **100** and **101** with VAA galactose binding lectin.

Some differences in the molecular recognition features of both α-*N*-glycopeptides by VAA and CoralT lectins were detected. In particular, for VAA, Gal H4 was the proton receiving more percentage of saturation, followed by Gal H2 and Gal H3 (**Figure 11**). In contrast, for Coral tree, Gal H1 and H2 received the highest saturation, followed by Gal H3 (**Figure 12**).

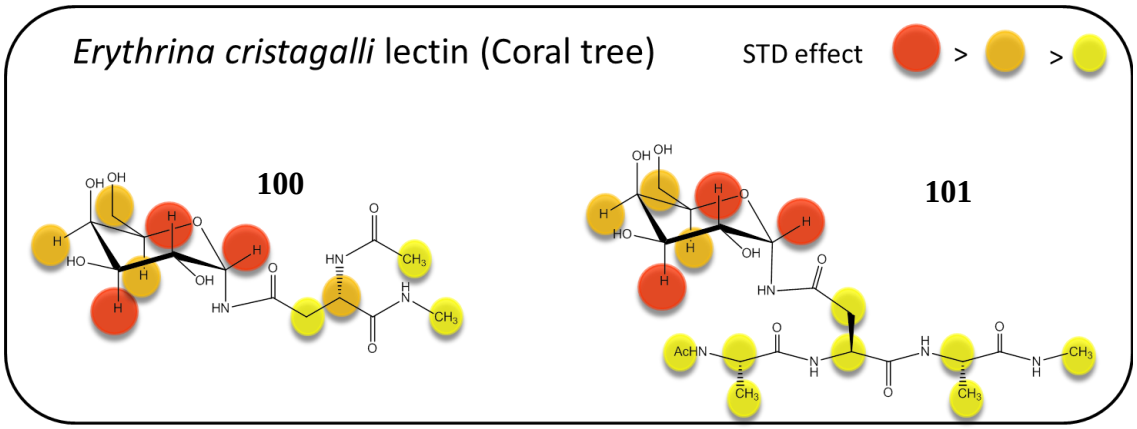


Figure 12. Epitope mapping obtained for glycopeptides **100** and **101** with Coral tree galactose binding lectin.

These results demonstrate that the sugar moiety of these non-natural glycopeptides is properly recognized by the selected galactose model proteins VAA and Coral tree, but with slightly different orientations. Furthermore, for both lectins, the transfer of magnetization was very uniform to all the Hα protons of glycopeptide **101**. This observation suggests that the peptide backbone adopts an extended conformation in the bound state. Regarding the relative recognition of the peptide chain of compound **101** by the two proteins, slight differences in the binding epitope were also appreciated. The STD responses for Asn Hα Hβ1 and Hβ2, as well as for Ala Hα were significantly higher in the presence of Coral tree than with VAA.

9.5 Conclusion

Multidisciplinary studies were applied to some of the new α-N-linked glycopeptides synthesized. In particular the molecules were found to adopt an extended conformations for the peptide backbone in water solution, when the number of (Ala-(α-Gal)-Asn-Ala) subunits is equal to 1 and 2. These findings were indicated both by NMR-based experiments and computational and molecular dynamics simulations. A folded conformation is expected for the glycopeptides with five subunits, as revealed by simulated annealing and MD simulation experiments. This molecule (**134**) has been sent to the group of Robert Ben to evaluate possible antifreeze properties.

Moreover, the two analogues **100** and **101** (n =1) were found to be good mimics of natural galactosyl conjugates since they are properly recognized by two model lectins (VAA and Coral tree). Thus, the employed chemical modification to afford an α-configuration of the Gal-Asn link does not induce major changes in the molecular recognition features of the galactose units for these molecules.

Taking these models as scaffolds, it should be possible to design novel molecules with enhanced recognition abilities, with the final aim of getting enzymatic stable neoglycopeptides and complex glycopeptide mimics which may be useful as therapeutic agents.

9.7 Experimental Section

NMR spectroscopy

For the conformational analysis of the α-N-Linked glycopeptides in solution, the experiments were recorded in H_2O/D_2O 90:10 on a Brüker Avance 500 MHz spectrometer and on a Brüker Avance 600 MHz spectrometer at 278K. 2D-NOESY experiments were carried out with mixing times of 300 and 600 ms. The concentration of glycopeptides for the NMR experiments was set to 10 mM.

STD NMR experiments were recorded at 298K on a Bruker Avance 500 MHz spectrometer. VAA and CoralT were dissolved in D_2O buffer (20 mM NaPi, 100 mM NaCl, pH=7.4) and the final concentration measured by U.V. spectroscopy. For binding studies, the glycopeptides were also suspended in a buffer solution to a final concentration of 40 mM. STD experiments were performed for a molar ratio of 100/1 (glycopeptides 1 or 2 /protein). The final concentration of the protein in the NMR tube was 40μM. A series of Gaussian-shaped pulses of 49 ms each were applied, separated by 1ms delay, with a total saturation time for the protein envelope of 2 s and a maximum *B1* field strength of 50 Hz. An off-resonance frequency of δ=100 ppm (where no proteins signals are present) and on-resonance frequency of δ= 7,2 ppm and -1 ppm (protein aromatic signals region) were employed. No significant differences on the epitope mapping were observed between the two on-resonance frequencies. A total number of 1024 scans were acquired and the spectra were multiplied by an exponential line broadening function of 1Hz prior to Fourier transformation. All experiments were recorded with a 15ms spin lock pulse, which minimizes the protein background resonances.

Conformation analysis and dynamic simulations

MC/EM calculations were performed using MacroModel 9.5 and the Maestro Graphical User Interface. The AMBER* force field with the Senderowitz-Still parameters has been used. Water solvation was simulated using GB/SA continuum solvent model. Extended non-bonded cut off distances (a van der Waals cut off of 8.0 Å and an electrostatic cutoff of 20.0 Å) were used.

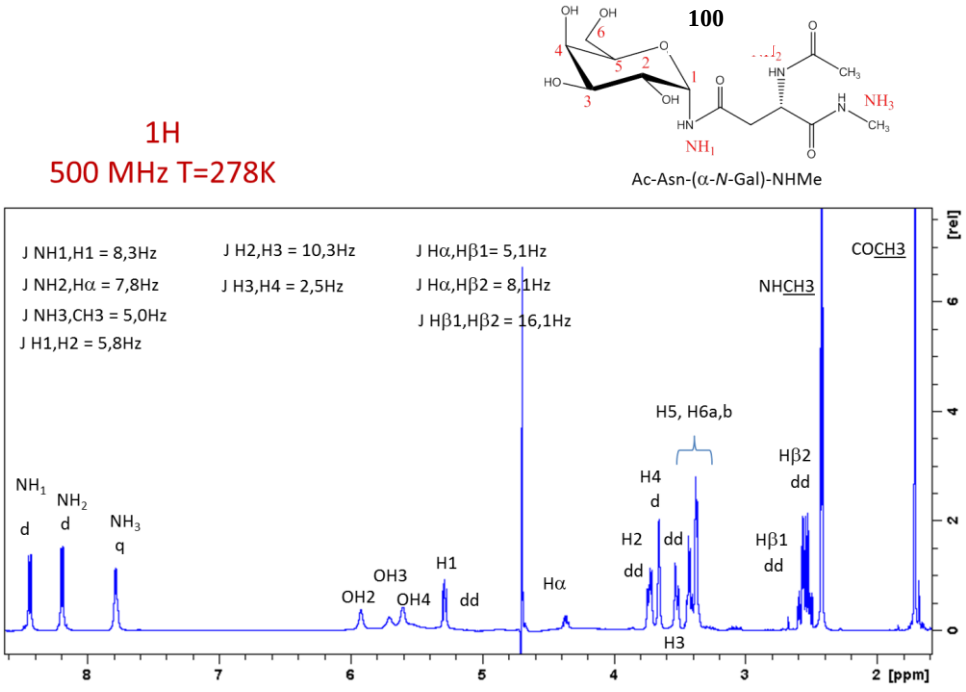
The MC/EM procedure was carried out applying 6000 and 10000 steps for glycopeptides 1 and 2, respectively. Backbone (Φ-Ψ) and sidechain (χ^1- χ^2) dihedral angles were all varied during the simulation, along with the pseudo-Φ anomeric torsion. Only conformers among 5.00 kcal/mol from

the global minimum were analyzed and clustered based on their backbone and sidechain conformation.

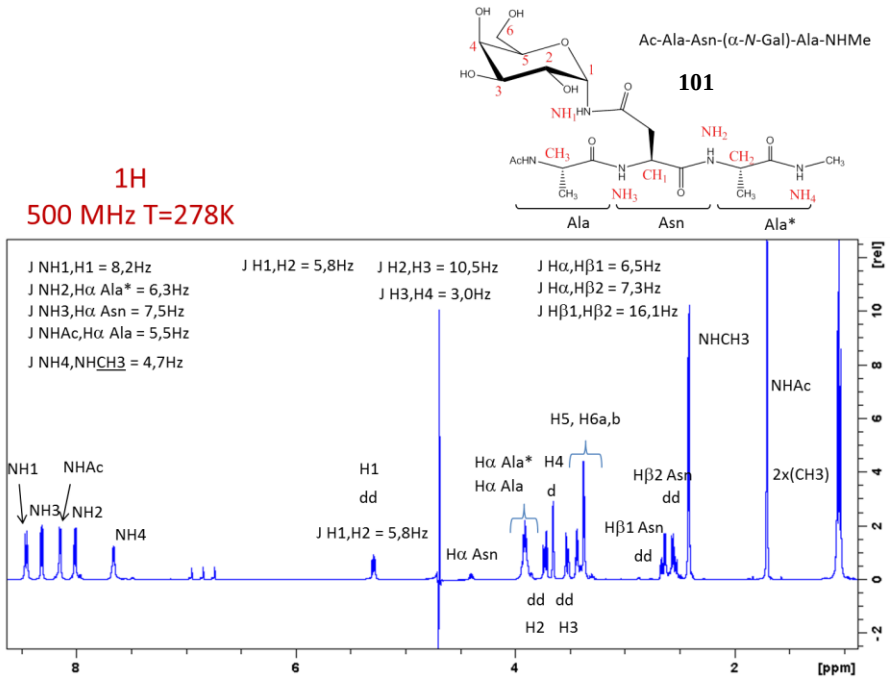
The MC/SD dynamic simulations were run using the AMBER* all-atom force field and van der Waals and electrostatic cutoffs of 25 Å, together with a hydrogen bond cutoff of 15 Å. The same degrees of freedom of the MC/EM searches were used. All simulations were performed at 300 K, with a dynamic time-step of 1.5 fs and a frictional coefficient of 0.1 ps^{-1}. Runs of 5 ns for 1 and 10 ns for 2 were performed, starting from conformations selected from the MC/EM outputs. The acceptance ratios for glycopeptides 1 and 2 were 4.5 and 4.0, respectively.

MD simulations in explicit water with periodic boundary condition were performed using AMBER 9 [21] with ff99SB [28] and glycam04 [29] force fields. TIP3P water model and a truncated octahedron box with 12.0 Å buffer were used. An integration step of 1 fs and a cutoff of 10.0 Å were applied. Glycopeptides 1 and 2 were simulated at 300 K (using the Langevin thermostat) and 1 atm for 20 ns and 50 ns, respectively, after an equilibration time of 450 ps.

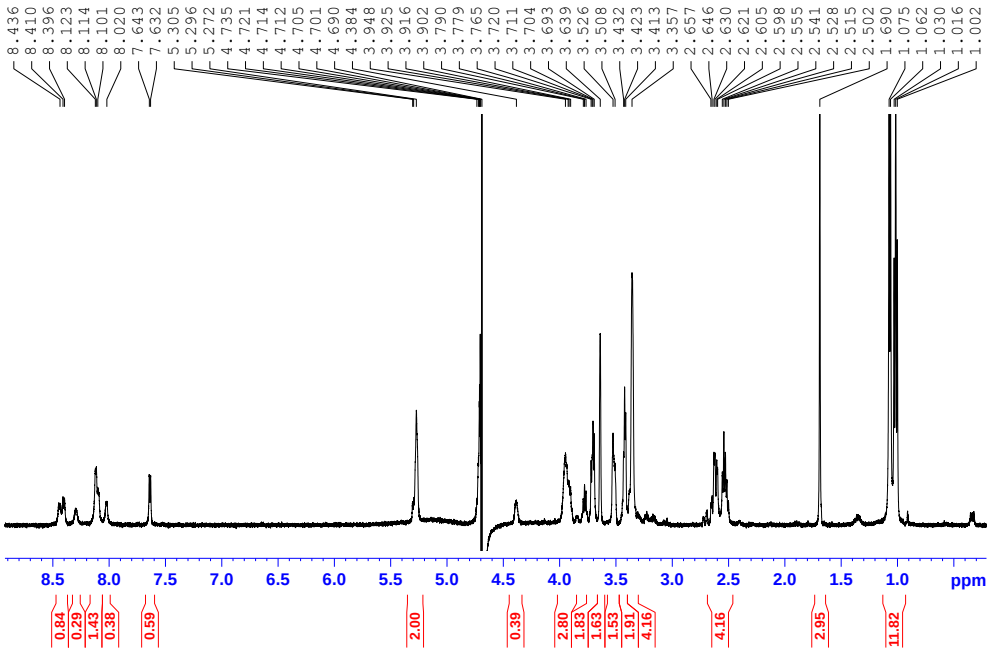
Experimental data



^{1}H NMR spectrum of **100** (H_2O/D_2O, 500 MHz) with coupling constants.



^{1}H NMR spectrum of **101** (H_2O/D_2O, 500 MHz) with coupling constants.



^{1}H NMR spectrum of **131** (H_2O/D_2O, 500 MHz) with coupling constants.

9.6 References

[1] a) Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J. *Essentials of Glycobiology*, New York: Cold SpringHarbor Laboratory Press; **1999**; b) Meyer, B.; Möller, H. *Top. Curr. Chem.* **2007**, *267*, 187–251.

[2] Colombo; C.; Bernardi A. *Eur. J. Org. Chem.* **2011**, 3911–3919.

[3] a) Meinjohanns, E.; Meldal, M.; Paulsen, H.; Dwek, R. A.; Bock K. *J.Chem.Soc. Perkin I* **1998**, 549-560. b) Christiansen-Brams, I.; Meldal, M.; Bock K. *J. Chem. Soc. Perkin Trans. I* **1993**, 1461-1471.

[4] Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali B. *J. Am. Chem. Soc.*, **2004**, *126*, 8421-8425.

[5] Marcelo, F.; Colombo, C.; Doro, F.; Cañada, F. J.; André, S.; Gabius, H-J.; Barbero, J. J.; Bernardi, A. Submitted

[6] Mikeska, R.; Wacker, R.; Arni, R.; Singh, T. P.; Mikhailov, A.; Gabdoulkhakov, A.; Voelter, W.; Betzel, C. *Acta Crystallogr.* **2005**, *F-61*,17-25.

[7] a) Shaanan, B.; Lis, H.; Sharon, N. *Science* **1991**, *254*, 862 – 866; b) Surolia, A.; Sharon, N.; Schwarz, F. P.; *J. Biol. Chem.* **1996**, *271*, 17697 –17703.

[8] Chang, G.; Guida, W. C.; Still, W. C. *J. An. Chem. Soc.* **1989**, *111*, 4379-4386.

[9] Still, W. C.; Tempzyk, A.; Hawley, R.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

[10] MacroModel, version 9.5, Schrödinger, LLC, New York, NY, **2007**.

[11] Maestro, version 8.0, Schrödinger, LLC, New York, NY, **2007**.

[12] Guarnieri, F.; Still, W. C. *J. Comp. Chem.* **1994**, *15*, 1302 – 1310

[13] Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.: Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman P.A. AMBER 9, University of California, San Francisco, **2006**.

[14] Eker, F.; Griebenow, K.; Cao, X.; Nafie, L. A.; Schweitzer-Stenner, R. *Proc. Natl. Acad. Sci. USA* **2004**, *101* **,** 10054–10059.

[15] Motta, A.; Reches, M.; Pappalardo, L.; Andreotti, G.; Gazit, E. *Biochemistry*, **2005**, *44*, 14170–14178

Chapter 10

Conclusions and outlook

We have shown in this thesis our efforts devoted to the synthesis of α-*N*-linked glycopeptides. These *N*-linked glycopeptides are unnatural molecules, since they display an α linkage between the peptide side chain and the sugar moiety, unlike natural glycopeptides which connect the peptide to the glycan through a β-*N*-glycosidic bond. This novel type of glycosylation of peptides might introduce structural diversity, leading to modifications that can mimic and/or interfere with molecular recognition events. Direct glycosylation of peptide chains is not viable for the synthesis of molecules with α-*N*-linked configuration, since the corresponding α-glycosyl amines isomerise to the β-anomers. Hence, forced to employ α-*N*-linked glycosyl amino acids to be linearly incorporated into a peptide sequence, we initially dedicated our efforts to the development of an efficient stereoselective synthesis of preformed *N*α-Fmoc-protected glycosyl amino acids, as suitable building blocks for solid phase synthesis applications. The original purpose was the developments of procedures for the synthesis of both asparagine and glutamine derivatives, with glucose and galactose for the carbohydrate portion.

In fact, α-*N*-linked glycosyl glutamine derivatives were obtained in low yield either:

a) using a traceless Staudinger ligation with fluorophenylphosphines functionalized with protected amino acids as illustrated in **Chapter 4**. This reaction, recently developed in our laboratories for the formation of α glycosyl-amidic bonds, gave low conversion into the desired product with our substrates. Poor reactivity of the phosphines functionalized with (*N*-Boc, *O*-tBu) glutamic acid and (*N*-Cbz, *O*-Bn) glutamic acid was probably due to steric hindrance, which slows the acyl-transfer process.
b) using the DeShong' method, as described in **Section 5.3**. A huge formation of a lactam by-product from the acylating agent (thiopyridyl ester) was the main reason of the low yields observed.

Rather, using DeShong methodology and further elaborations, we succeeded in the synthesis of α-*N*-linked glycosyl asparagine derivatives with good yields in the required scale. In particular, *N*α-Fmoc-protected *gluco* and *galacto* pyranosyl-L-asparagine derivatives were obtained in good yields (**Section 5.1)**. These novel building blocks finally allowed the synthesis of unnatural α-*N*-linked glycopeptides in a linear solid phase peptide synthesis approach.

We further encountered some problems to solve in the investigation of proper coupling conditions for the elongation at C-terminus of the α-*N*-linked glycosyl building blocks, that were found to be prone to cyclization under the usual activation conditions at the C-terminus (**Section 6.2**). We finally found that the reaction could be mastered using PyBROP as the condensing agent and we performed the solution synthesis of models of a galactosyl tripeptide and of a

galactosyl pentapeptide, which were subjected to computational, conformational and molecular interaction studies (**Chapter 9**).

Another obstacle was represented by the selection of conditions for *O*-acetyl removal at the sugar moiety, the general last step to obtain glycopeptides. Commonly used methods turned out to be harmful for our molecules, since side reactions resulted in ring opening, anomerization and ring contraction (formation of the furanosyl compounds) of the galactose moiety (**Section 6.5**). This led to the general conclusion that α-*N*-linked glycopeptides, when unprotected at the sugar moieties, are unstable towards basic and acidic conditions and must be handled in a limited range of pH ($4 \leq pH \leq 9$). Deacetylation with 4M $MeNH_2$ in EtOH was finally identified as the most appropriate approach, which afforded clean glycopeptides in good yields.

The synthesis of more complex α-*N*-linked glycopeptides was performed using solid phase methodology (**Chapter 7**). In particular, we were interested in the synthesis of mimic of antifreeze glycopeptides (**Section 7.1**), which display repeating units of general structure (Ala-Asn (α-*N*-Gal)-Ala). The major problems towards the synthesis of such structures were observed in the coupling reactions of our α-*N*-galactosyl asparagine building block, which, especially with increasing peptide length, reacted with poor yields and in long reaction time, due principally to its steric hindrance. This drawback, inevitably, caused the formation of many truncated sequences and as a consequence, of heterogeneous mixtures of products difficult to separate. The problem was aggravated by the modular structure of the sequence selected, which contain multiple copies of the glycosyl aminoacid. Nonetheless, even in these extreme situations, α-*N*-linked glycopeptides containing up to 15 residues and 5 copies of galactosyl asparagine have been purified with a double step purification (**Chapter 8**). The first step was performed on *O*-acetyl-glycopeptides using reverse phase chromatography, principally with the aim of separating part of the truncated sequences. The second purification was performed on unprotected α-*N*-linked glycopeptides with HILIC chromatography, an efficient technique for the purification of highly polar compounds.

In the end, pure α-*N*-linked glycopeptides were obtained with modest yields, especially the longer ones. However, this work constitutes the first attempt for the synthesis of unnatural α-*N*-linked glycopeptides and has been particularly useful in the understanding of behaviours, sometimes unexpected, of these molecules, in terms of reactivity and stability. Methods for the isolation and purification of glycopeptides up to 15 residues and 5 sugars have been identified, which allowed us to purify even very heterogeneous mixtures.

While the longest glycopeptide produced (with 5 sugars and 15 amino acids) has been sent to Robert Ben's group to test potential antifreeze properties, analyses, both with computational and

NMR techniques, have been performed with some of the synthesized molecules (**Chapter 9**). In particular two analogues were found to be good mimics of the natural molecules since they are properly recognized by two galactose-binding lectins, establishing that the employed chemical modification to afford an α-configuration of the Gal-Asn link does not induce major changes in the molecular recognition features of the galactose units.

These findings encourage the design of novel molecules, using α-*N*-linked glycosyl asparagine derivatives, as glycopeptides with potential enhanced recognition abilities, with the final aim of getting enzymatic stable neoglycopeptides which may be useful as therapeutic agents.

Clearly, much work has still to be done. Yields of the solid phase synthesis process have to be improved, in order to reduce the amount of truncated sequence products.

Possible solutions to improve the coupling efficiency could be:

a) The employment of more equivalents of the galactosyl building block, in more concentrated coupling solutions.
b) The employment of microwave-assisted solid phase synthesis, which seemed promising in our initial investigations to accelerate the coupling process.
c) The employment of dipeptide Fmoc-Ala-Ala-OH as a coupling block, with the aim of limiting the number of coupling steps.
d) The tagging of the glycopeptides with flourous reagents, with the scope of simplifying the isolation of the desired products from truncation mixtures.

Finally, the methods developed can now be used to include α-*N*-glycosylated building blocks in other, possibly less challenging, peptide sequences. The observations accumulated so far indicate that α-*N*-glycosylation may not perturb the 3D structure of the peptide, or of the sugar, and may find a number of applications, for instance to improve the water solubility of peptides without modifying their intrinsic properties, or to improve the carbohydrate affinity for interesting receptors. It remains to be established whether α-*N*-linked glycopeptides are protected against enzymatic hydrolysis of the sugar moiety, a likely possibility, given the absence of this motif from natural glycoproteins. The synthetic availability of these molecules which is the result of my PhD work allows to explore this fascinating class of novel, unnatural glycoconjugates and to answer some of these questions.