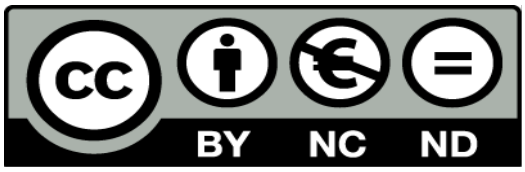
UNIVERSIDAD DE LA RIOJA

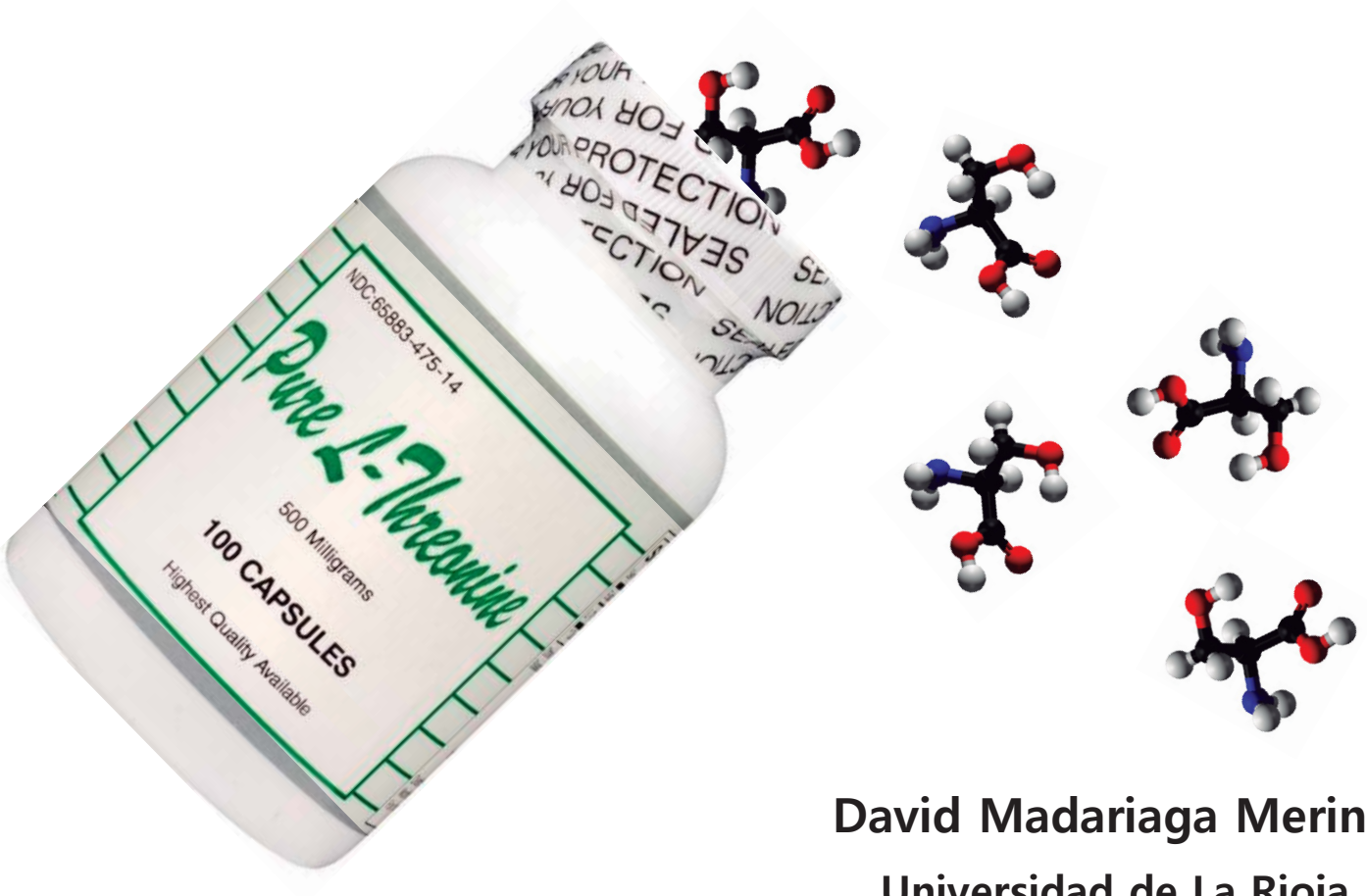
TESIS DOCTORAL

Título
New insights into molecular recognition of epitopes related to MUC1 human mucin by lectins
Autor/es
David Madariaga Merino
Director/es
Jesús Manuel Peregrina García y Francisco Corzana López
Facultad
Facultad de Ciencias, Estudios Agroalimentarios e Informática
Titulación
Departamento
Química
Curso Académico
2014-2015



UNIVERSIDAD DE LA RIOJA

New Insights into Molecular Recognition of Epitopes Related to MUC1 Human Mucin by Lectins



David Madariaga Merino

Universidad de La Rioja

December 2014

FACULTAD DE CIENCIAS, ESTUDIOS
AGROALIMENTARIOS E INFORMÁTICA

UNIVERSIDAD DE LA RIOJA

Departamento de Química
Centro de Investigación en Síntesis Química
Área de Química Orgánica

TESIS DOCTORAL

New Insights into Molecular Recognition of Epitopes Related to MUC1 Human Mucin by Lectins

Memoria presentada en la Universidad de La Rioja para optar al grado de Doctor en Química por

David Madariaga Merino

Diciembre 2014

UNIVERSIDAD DE LA RIOJA

Dr. **JESÚS MANUEL PEREGRINA GARCÍA**, Catedrático de Química Orgánica del Departamento de Química de la Universidad de La Rioja y Dr. **FRANCISCO CORZANA LÓPEZ**, Profesor Contratado Doctor del Departamento de Química de la Universidad de La Rioja

CERTIFICAN:

Que el trabajo presentado en la memoria titulada "***New Insights into Molecular Recognition of Epitopes Related to MUC1 Human Mucin by Lectins***" ha sido realizada por **David Madariaga Merino** (Licenciado en Química) en el Departamento de Química de la Universidad de La Rioja bajo su inmediata dirección y reúne las condiciones exigidas para optar al grado de Doctor en Química.

Logroño, diciembre de 2014

Los directores,

Dr. Jesús Manuel Peregrina García Dr. Francisco Corzana López

Abstract

Tn antigen (α-*O*-GalNAc-Ser/Thr) is one of the most specific human tumor-associated structures. This motif is implicated in HIV infection and it is expressed early in tumor cells. It has been observed that there is a direct correlation between carcinoma aggressiveness and the density of this antigen. For this reason, Tn determinant is a convenient cancer biomarker.

Mucin-like glycopeptides and glycoproteins, in particular MUC1, incorporate this structure in their sequence. While in normal cells this *O*-glycosylated protein carries complex oligosaccharides, in cancer cells, MUC1 expression is increased and its glycans are short and poorly branched. Therefore, Tn antigen is now exposed to the immune system generating an immunological response.

From a structural point of view, Tn antigen is referred to *N*-acetylgalactosamine (GalNAc) α-*O*-linked to serine (Ser) or threonine (Thr), not discriminating the amino acid to which GalNAc is linked. However, in this Thesis, we have synthesized different representative MUC1 epitopes bearing Ser and Thr in their structure and we have detected differences between these two amino acids in terms of affinity to lectins. As a result, it is important to mention specifically the underlying amino acid in Tn antigen.

We have also analyzed the role of flanking amino acids of Tn antigen in the peptide chain, getting surprising results in terms of affinity to lectins. In addition, we have obtained an X-ray structure by the first time of soybean agglutinin (SBA) lectin linked to a MUC1-derived glycopeptide.

All these results may have important implications for better understanding the glycopeptide-lectin interactions and may contribute to engineer new binding sites, allowing the design of novel glycosensors for Tn antigen detection in tumor cells.

Resumen

El antígeno Tn (α-*O*-GalNAc-Ser/Thr) es una de las estructuras asociadas a tumores más específicas en humanos. Esta molécula está implicada en la infección del VIH y en su expresión temprana en células tumorales. También se ha observado una relación directa entre la agresividad de un carcinoma y la concentración de dicho antígeno. Por ello, el antígeno Tn está considerado como un importante biomarcador del cáncer.

Los glicopéptidos y glicoproteínas tipo mucina, en particular la mucina humana MUC1, incorporan en su estructura el antígeno Tn. Mientras que en las células sanas esta *O*-glicoproteína de membrana tiene oligosacáridos complejos, en células cancerosas, la expresión de MUC1 aumenta y sus glicanos son cortos y poco ramificados. Esto hace que el antígeno Tn quede expuesto al sistema inmune generando una respuesta inmunológica.

Desde el punto de vista estructural, el antígeno Tn está formado por *N*-acetilgalactosamina (GalNAc) unida mediante un enlace α-*O*-gliosídico a serina (Ser) o treonina (Thr), sin especificar a qué aminoácido se une el GalNAc. Sin embargo, en esta tesis hemos sintetizado diferentes epítopos representativos de MUC1 que incorporan Ser y Thr en su estructura. Hemos analizado las interacciones que ocurren en el reconocimiento de GalNAc por diferentes lectinas, observando que es importante especificar a qué aminoácido se une el residuo carbohidrato GalNAc.

También hemos analizado la importancia de los aminoácidos que rodean al antígeno Tn en la cadena peptídica, obteniendo resultados sorprendentes en términos de afinidad a lectinas. Además, hemos obtenido por primera vez,

una estructura de rayos X de la lectina *soybean agglutinin* (SBA) unida a un glycopéptido derivado de la mucina humana MUC1.

Todos estos resultados podrían tener importantes implicaciones para entender mejor las interacciones glicopéptido-lectina y podría contribuir al diseño y construción de nuevos sitios de unión, permitiendo el diseño de nuevos *glicosensores* para la detección de antígeno Tn en células tumorales.

Agradecimientos

Me gustaría agradecer a todo mi grupo de investigación, en especial a mis directores Pere y Paco, por su ayuda y apoyo durante el desarrollo de esta Tesis Doctoral. Además, a toda mi familia y amigos por su paciencia y ánimo en todo momento.

Finalmente, quisiera agradecer a todas las instituciones que han hecho posible mediante su apoyo económico este trabajo:

- Universidad de La Rioja, por la beca F.P.I. concedida en el año 2011, por su apoyo en forma de proyectos y ayudas a tesis doctorales (ATUR), así como por conformar el marco humano y tecnológico idóneo para el desarrollo de este trabajo.

- Gobierno de La Rioja, por su aportación económica en forma de proyectos COLABORA.

- Ministerio de Ciencia e Innovación por su aportación económica al proyecto *"Síntesis y análisis conformacional de O-glicopéptidos de interés estructural y biológico"* (CTQ2009–013814).

- Ministerio de Economía y Competitividad por su aportación económica al proyecto *"Diseño racional de glicopéptidos con aplicaciones en química biológica"* (CTQ2012–36365).

Index

Abbreviations

Å	ångström
δ	chemical shift
ε	extinction coefficient
Φ_p	dihedral angle
ʎ	wavelength
μL	microliter
μM	micromolar
°C	Celsius degree
^{1}H NMR	proton nuclear magnetic resonance
^{13}C NMR	carbon-13 nuclear magnetic resonance
3D	three-dimensional
aa	amino acid
Ac	acetyl group
AcOH	acetic acid
Ac_2O	acetic anhydride
Ala, A	alanine
AMBER	assisted model building with energy refinement
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartic acid
Bn	benzyl group
Boc	*tert*-butoxycarbonyl
Boc_2O	di-*tert*-butyl dicarbonate
BPA	*Bauhinia purpurea* agglutinin
br s	broad singlet
CAN	ammonium cerium (IV) nitrate
CD-MPR	cation-dependent mannose 6-phosphate receptor
CESGA	*Centro de supercomputación de Galicia*

COSY	^{1}H-^{1}H correlated spectroscopy
Cys, C	cysteine
d	doublet, day
DBA	*Dolichos biflorus* agglutinin
DCC	*N,N'*-dicyclohexylcarbodiimide
DCM	dichloromethane
dd	doublet of doublets
DFT	density functional theory
DIC	*N,N'*-diisopropylcarbodiimide
DIEA	*N,N*-diisopropylethylamine
DMF	dimethylformamide
DVB	divinylbenzene
ELISA	enzyme-linked inmunosorbent assay
ELLA	enzyme-linked lectin assay
ESI	electrospray ionization
EtOAc	ethyl acetate
EtOH	ethanol
Et_2O	diethyl ether
Fmoc	9-fluorenylmethoxycarbonyl
Fmoc-OSu	*N*-(9-fluorenylmethoxycarbonyloxy)succinimide
g	gram
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GlcNAc	*N*-acetylglucosamine
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
h	hour
HBP	hepatic binding protein

HBTU	*N*,*N*,*N*′,*N*′-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
His, H	histidine
HIV	human immunodeficiency virus
HOAt	1-hydroxy-7-azabenzotriazole
HOBt	1-hydroxybenzotriazole
HPA	*Helix pomatia* agglutinin
HPLC	high-performance liquid chromatography
HRMS	high resolution mass spectrometry
HRP	horseradish peroxidase
HSQC	^{1}H-^{13}C heteronuclear single quantum coherence
IGF-II	insulin-like growth factor II
IgSF	immunoglobulin superfamily
IHC	immunohistochemistry
Ile, I	isoleucine
ITC	isothermal titration calorimetry
J	coupling constant
K_a	affinity constant
K_D	dissociation constant
Leu, L	leucine
Lys, K	lysine
M	molarity
m	multiplet
MALDI	matrix-assisted laser desorption ionization
MBHA	methylbenzhydryl amine
MBPs	human mannan-binding proteins
Me	methyl group
MeOH	methanol
Met, M	metionine
MD	molecular dynamics

MD-tar	molecular dynamics with time-averaged restraints
mg	milligram
MHz	megahertz
min	minute
mL	milliliter
MM	molecular mechanics
mm	millimeter
mmol	millimole
MPR	mannose 6-phosphate receptor
MS	mass spectrometry
ms	milliseconds
MW	molecular weight
m/z	mass-to-charge ratio
nm	nanometer
nmol	nanomole
NAGA	*N*-acetylgalactosaminidase
Neu5Ac	*N*-acetylneuraminic acid or sialic acid
NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
NOESY	nuclear overhauser effect spectroscopy
NPG	N^{α}-protecting group
Pbf	2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl
PBST	phosphate buffered saline with Tween® 20
pdb	protein data bank
PEG	polyethylene glycol
Ph	phenyl group
PHA	phytohemagglutinin
Phe, F	phenylalanine
PNA	peanut agglutinin

ppm	parts per million
Pro, P	proline
PS	polystyrene
PyBOP	(benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
R	alkyl or aryl group, arginine
RHL	rabbit hepatocites lectin
rt	room temperature
s	singlet, second
SBA	soybean agglutinin
Ser, S	serine
Ser*, S*	glycosylated serine
SPG	side chain protecting group
SPPS	solid-phase peptide synthesis
SSA	*Salvia sclarea* agglutinin
STD	saturation transfer difference
t	triplet
TACAs	tumor-associated carbohydrate antigens
TBTU	*N*,*N*,*N*',*N*'-tetramethyl-*O*-(1H-benzotriazol-1-yl)uronium tetrafluoroborate
tBu	*tert*-butyl
tBuOH	*tert*-butanol
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TIS	triisopropylsilane
TMB	3,3´,5,5´-tetramethylbenzidine
TMS	tetramethylsilane
Thr, T	threonine
Thr*, T*	glycosylated threonine

TMS	tetramethylsilane
TOF	time of flight
t_r	retention time
TRIS	tris(hydroxymethyl)aminomethane
Trp, W	tryptophan
Trt	triphenylmethyl or trityl
Tyr, Y	tyrosine
tRNA	transfer ribonucleic acid
U	potential energy
UDP	uridine diphosphate
UV	ultraviolet
Val, V	valine
VVA	*Vicia villosa* agglutinin
WGA	wheat germ agglutinin

INTRODUCTION & GENERAL OBJECTIVES

Cancer is a generic term for a large group of diseases that can affect any part of the body. One defining feature of cancer is the rapid creation of abnormal cells that grow beyond their usual boundaries, and can invade contiguous parts of the body and spread to other organs. This process is termed metastasis. Metastases are the major cause of death from cancer accounting for 8.2 million deaths and 32.6 million people worldwide living with cancer in 2012, according to World Health Organization.[1]

Cancer arises from one single cell. The transformation from a normal cell into a tumor cell is a multistage process, typically a progression from a pre-cancerous lesion to malignant tumors. These changes are the result of the interaction between genetic factors and external agents, including physical (UV and ionizing radiation), chemical (asbestos, components of tobacco, arsenic…) and biological carcinogens (certain viruses, bacteria or parasites). In some specific cancers, the incidence of cancer rises dramatically with age.

Knowledge about the causes of cancer as well as interventions to prevent and manage the disease are widespread. Cancer can be reduced and controlled by implementing evidence-based strategies for cancer prevention, early detection of cancer and management of patients with cancer. Many cancers have a high chance of cure if detected early and treated adequately. In this sense, several studies on tumor cells have revealed that lipoproteins and glycoproteins present in the membrane of cellular surface exhibit important modifications in its structure.[2]

---

[1] Globocan **2012**, IARC

[2] a) R. M. Wilson, S. J. Danishefsky, *J. Am. Chem. Soc.* **2013**, *135*, 14462-14472; b) P. M. Rudd, T. Elliott, P. Cresswell, I. A. Wilson, R. A. Dwek, *Science* **2001**, *291*, 2370-2376.

1.1. Lectins and their importance in cancer treatment

Proteins that interact with carbohydrates in a non-covalent way occur widely in nature. Prominent examples are carbohydrate-specific enzymes and anti-carbohydrate antibodies. The most relevant biological recognition targets for carbohydrates are lectins. The term 'lectin' refers to a Latin word 'to choose' or 'to pick out'. Lectins are absent of any catalytic, enzyme-like activities and antibody-like roles, and as a consequence they are not products of an immune response. These carbohydrate-binding proteins are found in most organisms, ranging from viruses and bacteria to plants and animals.

Classifications of lectins are traditionally based on their monosaccharide specificities.[3] The structures of lectins are as varied as their source of occurrence. Based on the common molecular structure of the lectins, they can be classified into simple, mosaic (or multidomain) and macromolecular assembly (common in bacteria, participate in adhesive processes, heterogeneous group).

1.1.1. Simple lectins

They consist of a small number of subunits, not necessarily identical but homogenous, of molecular weight usually below 40 kDa that may participate in aggregation process forming dimers and tetramers to become active. This class includes practically all known plant lectins (legume, cereal, etc.) as well as the galectins (formerly S-lectins), a family of galactose-specific animal lectins.

[3] H. Lis, N. Sharon, *Chem. Rev.* **1998**, *98*, 637-674.

a) Legume: Legume lectins were first identified as nonenzymatic carbohydrate-binding components that could be isolated from seed extracts.[4] Legume lectins have been intensively studied in the recent past because of their utility in animal and medical histochemistry as probes that recognize specific cell surface carbohydrate epitopes. Most of legume lectins have a relatively uniform three-dimensional protein structure,[5] and their structure has been determined by X-ray crystallography. However, the functional role of legume lectins within plant tissues is not clearly understood. Examples of these lectins are concanavalin A (pdb: 3cna) from jack beans, phytohemagglutinin (PHA, pdb: 1fat) from the red kidney beans, soybean agluttinin (SBA, pdb: 1sbf) or peanut agglutinin (PNA, pdb: 2dvd). Some of them are represented in Figure 1.1.

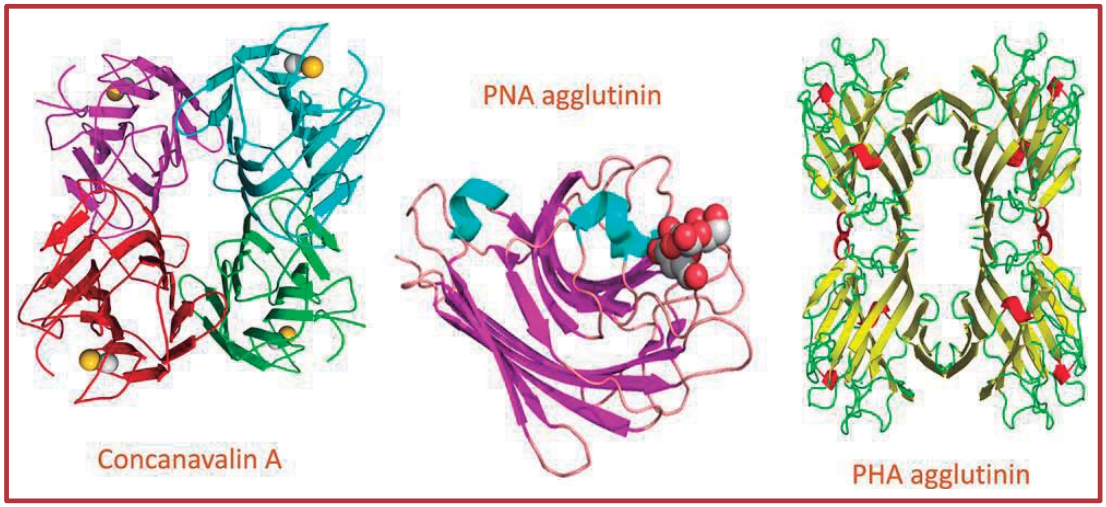


Figure 1.1. *Some representative legume lectins.*

[4] W. J. Peumans, E. J. M. Van Damme, *Plant Physiol.* **1995**, *109*, 347-352.
[5] J. M. Rini, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 551-577.

b) Cereal: *Gramineae* lectins are a homogeneous class of proteins differing from all other simple lectins with respect to their biochemical, physico-chemical and biological properties. Within this extended group of lectins, three subtypes are distinguished, which according to their origin are designated as cereal, rice and barley lectins.[6] All cereal lectins that have been analyzed have outstandingly similar amino acid compositions typified by high contents of cysteine and glycine. Besides their molecular structure, *Gramineae* lectins share their carbohydrate binding specificity too (namely for *N*-acetylglucosamine and *N*-acetylglucosamine oligomers). One of the most studied lectin of this group is wheat germ agglutinin (WGA, Figure 1.2).

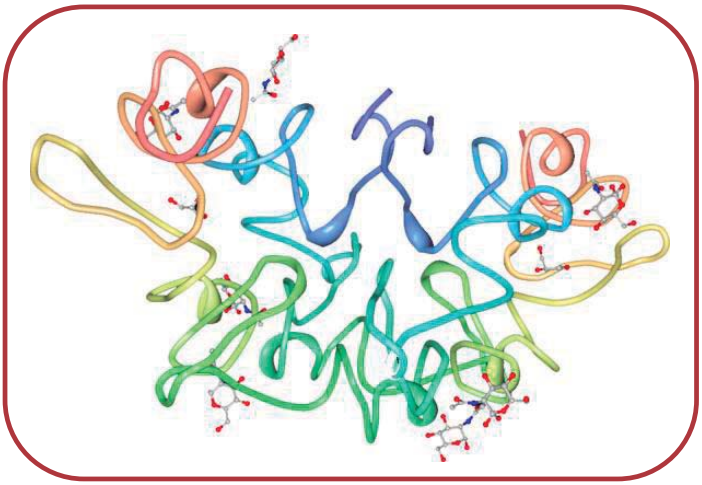


Figure 1.2. *Crystal structure of wheat germ agglutinin in complex with N-acetyl-D-glucosamine (pdb: 2uvo).*

c) Galectins: they are lectins that can bind to specific sugars on other proteins (receptors) to modulate cellular function and communication. Galectins are defined by their affinity to poly-*N*-acetyllactosamine-

[6] B. Cammue, H. M. Stinissen, W.J. Peumans, *Eur. J. Biochem.* **1985**, *148,* 315-322.

enriched glycoconjugates and sequence similarities in the carbohydrate recognition domain. (Figure 1.3). Novel implications of galectins in cell adhesion, cell growth regulation, immunomodulation, apoptosis, inflammation, embryogenesis, metastasis and pre-mRNA splicing are current investigations.[7]

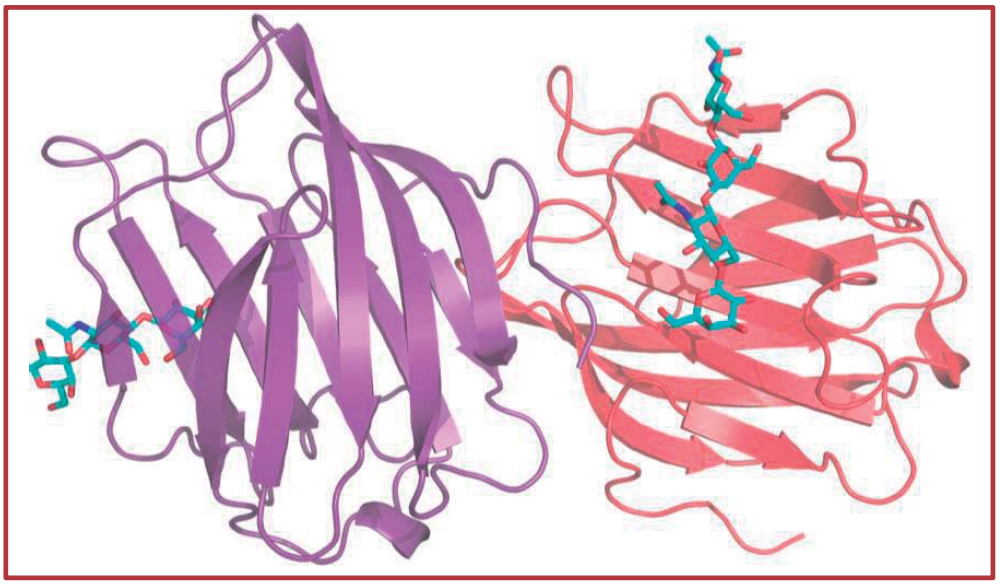


Figure 1.3. *Structure of human galectin-9 in complex with N-acetyllactosamine dimer.*

d) Others: *Amaryllidaceae* families (plans of amaryllis, orchid and garlic families contain lectins that bind mannose), *Moraceae* family (from jackfruit seeds, bind specifically to galactose), *Euphorbiaceae* (lectins from castor tree beans binds specifically to galactose)…

1.1.2. Mosaic or multidomain lectins

Included in this group are diverse proteins from different sources as viral hemagglutinins and animal lectins of the C-, P-, and I-type. They are all

[7] G. A. Rabinovich, *Cell Death Differ.* **1999**, *6*, 711-721.

complex molecules with a wide range of molecular weights, consisting of several kinds of protein modules or domains, but only one of them holds a carbohydrate binding site. Macromolecular assembly structures of lectins are commonly formed in many bacterial species, which express the lectin on their surfaces, through which the organism can bind to cells (Figure 1.4).

a) C-type lectins: They are a family of animal proteins that require Ca^{2+} for activity. C-lectins, as they are normally known, mediate biological processes that include endocytosis of ligands, cell to cell adhesion and serum glycoproteins turnover. In this category are included endocytic lectins (RHL, lectin from rabbit hepatocites or hepatic binding protein, HBP), collectins (MBPs, human mannan-binding proteins) and selectins (they interact with carbohydrate ligands on leukocytes and endothelial cells and mediate the initial adhesion of moving leukocytes to the stationary endothelium in a process called rolling).

b) P-type lectins: The two members of the P-type lectin family, the cation-dependent mannose 6-phosphate receptor (CD-MPR) and the insulin-like growth factor II/mannose 6-phosphate receptor (IGF-II/MPR), are distinguished from all other lectins by their ability to recognize phosphorylated mannose residues. The P-type lectins play an essential role in the generation of functional lysosomes within the cells of higher eukaryotes by directing newly synthesized lysosomal enzymes bearing the mannose 6-phosphate signal to lysosomes.[8]

c) I-type lectins: These are glycan-binding proteins that belong to the immunoglobulin superfamily (IgSF), excluding antibodies and T-cell receptors. The Siglec family of sialic acid–binding lectins is the best well-characterized group of I-type lectins, both structurally and functionally.

[8] N. M. Dahms, M. K. Hancock, *Biochim. Biophys. Acta* **2002**, *1572*, 317-340

These members of the I-type lectin family are predominantly found in immune cells. Their involvement in many facets of innate and adaptive immune cell biology makes them attractive pharmacological targets and it is therefore of great interest to study their glycan specificity.[9]

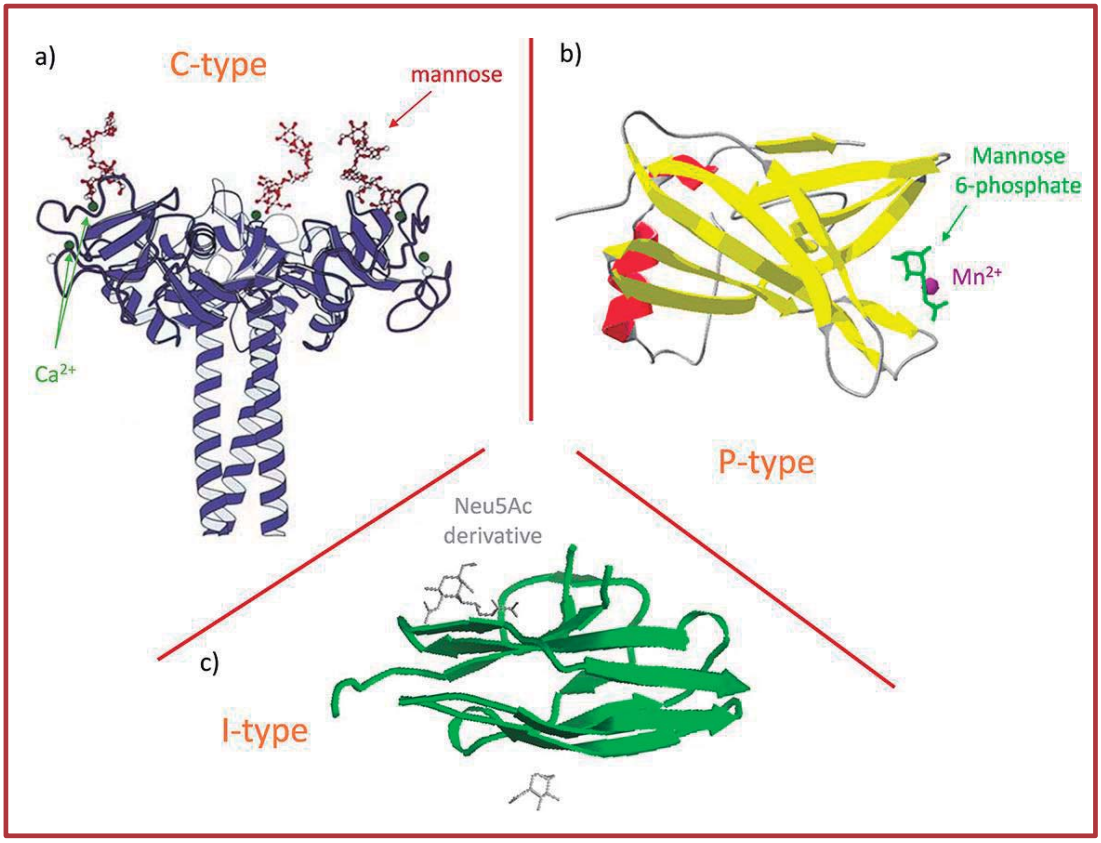


Figure 1.4. *a) C-type trimer represents a cluster of three MPBs bound to mannose, shown as a red stick model. The green spheres represent bound Ca^{2+}. b) 3D structure of CD-MPR. Mannose 6-phosphate is shown in green and Mn^{2+} in purple. c) Crystal structure of Siglec-7 in complex with methyl-9-(aminooxalyl-amino)-9-deoxyNeu5Ac.*

[9] C. Rademacher, T. Bru, R. McBride, E. Robison, C. M. Nycholat, E. J. Kremer, J. C. Paulson, *Glycobiology* **2012**, *22*, 1086-1091.

The ability of lectins to bind carbohydrates serves as a widely used principle in nature for cell recognition and adhesion. Endogenous lectins on the surface of mammalian cells have been gaining attention as antigenic determinants of normal (non-malignant) and various tumor cells. Using labelled glycoconjugates, it can be investigated the identification, localization, and quantitation of surface-bound lectins.

The binding of glycoconjugate ligand to the surface-bound lectins activates internalization or endocytosis of the bound ligand. Specific probes with labeled lectins, with variety of molecules, such as radioisotopes, biotin, enzymes, and flurochromes have been made. The same probes can also be used in targeted therapy for cancer by their ability to serve as carriers of tumoricidal drugs, cytotoxins, and oligonucleotides, which are then internalized upon binding.[10]

One important feature is that lectins stimulate the immune system. Lectins have antitumor and anticarcinogenic activities that could be of benefit in cancer treatment.[11] Recent studies on laboratory mice have shown that plant lectins might be employed in the formulation of novel cancer diagnostic and therapeutic approaches in pre-clinical stages.[12]

[10] R. Mody, S. H. a. Joshi, W. Chaney, *J. Pharmacol. Toxicol. Methods* **1995**, *33*, 1-10.
[11] E. G. De Mejía, V. I. Prisecaru, *Crit. Rev. Food Sci. Nutr.* **2005**, *45*, 425-445.
[12] C. P. Chang, M. C. Yang, H. Y. Lei, *PLoS One* **2011**, *6*, e28323.

1.2. Tn antigen

Tn antigen is a very simple structure: a serine (Ser) or a threonine (Thr) linked to the *N*-acetylgalactosamine (GalNAc) through an α-*O*-glycosidic bond (Figure 1.5).[13]

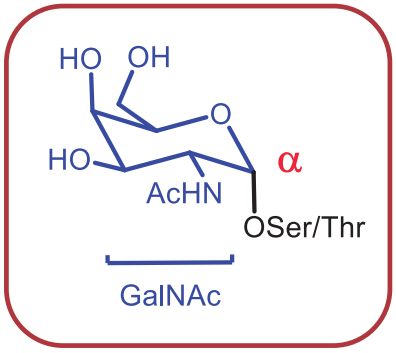


Figure 1.5. *Tn antigen structure.*

The historical naming of the Tn antigen derived from the observation by Moreau [14] who observed that it was similar but different from the T antigen; for that reason, the designation 'T antigen nouvelle' or Tn antigen. In the 'cluster of differentiation' nomenclature, the Tn antigen is also designated as CD175, and its sialylated form (sialyl Tn) CD175s. The disaccharide T antigen is termed CD176 (Figure 1.6).

[13] W. Dahr, G. Uhlenbruck, H. H. Gunson, M. Van Der Hart, *Vox Sang.* **1975**, *29*, 36-50.
[14] R. Moreau, J. Dausset, J. Bernard, J. Moullec, *Bull. Mem. Soc. Med. Hop. Paris* **1957**, *73*, 569-587.

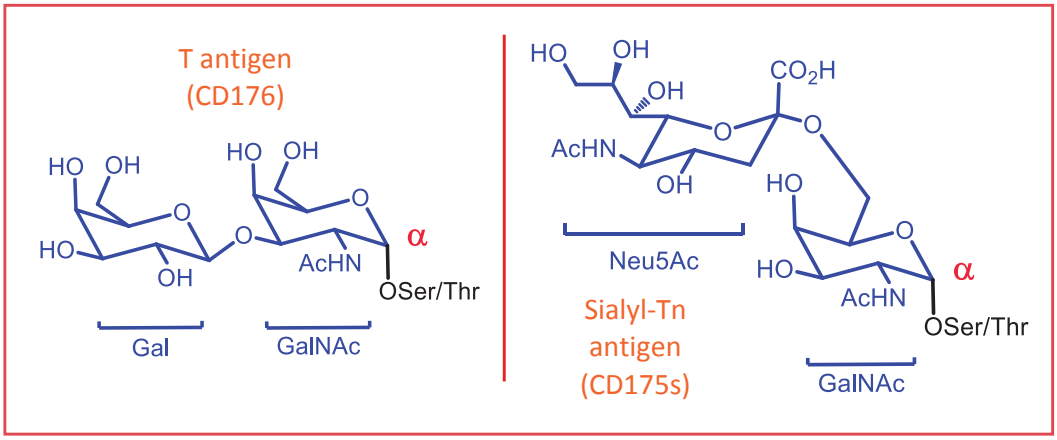


Figure 1.6. *T antigen and sialyl Tn structures.*

From a synthetic point of view, there is several routes to achieve the complete synthesis of Tn antigen.[15]

Tn antigen does not appear in abundance in normal cells and tissues of adult animals, and there is no evidence that it is an oncofetal antigen.
Several approaches for detecting Tn antigen have been developed and can be subdivided into three general categories: chemical, lectin-based and antibody-based detections:

1. Chemical detection: It was very used in the past decades. The direct chemical approach typically involves the β-elimination strategy, where *N*-acetylgalactosaminitol is generated by alkaline degradation of glycopeptides in the presence of NaOH (0.05-0.1 M) or $NaBH_4$ (1 M) with the simultaneous modification of the linking amino acids (Ser to 2-aminoacrylic acid or dehydroalanine and Thr to 2-aminocrotonic acid). [16,17]

[15] a) T. Ju, V. I. Otto, R. D. Cummings, *Angew. Chem. Int. Ed.* **2011**, *50*, 1770-1791; b) M. R. Pratt, C. R. Bertozzi, *Chem. Soc. Rev.* **2005**, *34*, 58-68; c) R. M. Ratcliffe, D. A. Baker, R. U. Lemieux, *Carbohydr. Res.* **1981**, *93*, 35-41.
[16] D. M. Carlson, *J. Biol. Chem*. **1968**, *243*, 616-626.

2. Lectin-based detection: GalNAc-binding lectins from plants and animals have been used to explore expression of the Tn antigen. As some examples of plant lectins we can consider *Dolichos biflorus* agglutinin (DBA),[18] *Salvia sclarea* agglutinin (SSA),[19] *Bauhinia purpurea* agglutinin (BPA),[20] the B4 isolectin of *Vicia villosa* agglutinin (VVA-B4),[21] soybean agglutinin (SBA)[22] and one animal lectin, the snail-derived *Helix pomatia* agglutinin (HPA).[23] For example, VVA-B4 has been used to explore the expression of Tn antigen in many different cancer cells, including those of the pancreas, breast, lung, and prostate.[24] HPA has been widely used since the early 1980s to study Tn antigen expression until nowadays.[25]

3. Antibody-based detection: a number of antibodies with different fine specificities for Tn antigen have been generated and used in several different applications such as western blotting, immuno-histochemistry (IHC) and also for in vivo localization.[26] Although GalNAc residue of Tn antigen is present in the most of *O*-linked glycans, it is normally masked by the attachment of additional sugar

[17] W. R. Alley, M. V. Novotny, *Annu. Rev. Anal. Chem.* **2013**, *6*, 237-265.
[18] M. E. Etzler, S. Gupta, C. Borrebaeck, *J. Biol. Chem.* **1981**, *256*, 2367-2370.
[19] R. A. Newman, G. G. Uhlenbruck, *Eur. J. Biochem.* **1977**, *76*, 149-155.
[20] A. M. Wu, J. H.Wu, J. H. Liu, T. Singh, *Life Sci.* **2004**, *74*, 1763-1779.
[21] S. E. Tollefsen, R. Kornfeld, *J. Biol. Chem.* **1983**, *258*, 5172-5176.
[22] T. K. Dam, T. A. Gerken, B. S. Cavada, K. S. Nascimento, T. R. Moura, C. F. Brewer, *J. Biol. Chem.* **2007**, *282*, 28256-28263.
[23] S. Hammarstrom, L. A. Murphy, I. J. Goldstein, M. E. Etzler, *Biochemistry* **1977**, *16*, 2750-2755.
[24] G. Konska, M. Guerry, F. Caldefie-Chezet, M. De Latour, J. Guillot, *Oncol. Rep.* **2006**, *15*, 305-310.
[25] a) W. Vainchenker, U. Testa, J. F. Deschamps, A. Henri, M. Titeux, J. Breton-Gorius, H. Rochant, D. Lee, J. P. Cartron, *J. Clin. Invest.* **1982**, *69*, 1081-1091; b) S. A. Brooks, D. M. Hall, I. Buley, *Br. J. Cancer* **2001**, *7*, 1014-1022; c) C. Welinder, B. Baldetorp, O. Blixt, D. Grabau, B. Jansson, *PLoS ONE* **2013**, *8*, e61749.
[26] C. Welinder, B. Baldetorp, C. Borrebaeck, B. M. Freudlund, B. Jansson, *Glycobiology* **2011**, *21*, 1097-1107.

residues. In the 1970s, Springer et al. discovered that this truncated form of *O*-glycosylation was present at high levels in ~90% of breast carcinomas.[27] After many years of study, Tn antigen has been also reported to be expressed in 70% to 90% of colon, lung, bladder, cervix, ovary, stomach, and prostate tumors as well as on a variety of pathogens such as parasitic worms and HIV.[28] Because of its exceptional expression profile, Tn antigen has been investigated broadly as a diagnostic marker, a prognostic marker and a therapeutic target for cancer.[29] In fact, several groups have developed Tn-based vaccines for the treatment and prevention of carcinomas.[30] Tn vaccines have produced excellent results in animal models and several clinical trials in humans have been published.[31] Tn antigen is also one of the key structures present on polyvalent cancer vaccines and multiantigen glycoprotein cancer vaccines that have been evaluated in clinical trials.[32]

[27] G. F. Springer, P. R. Desai, I. Banatwala, *J. Nat. Cancer Inst.* **1975**, *54*, 335-339.
[28] Q. Li, M. R. Anver, D. O. Butcher, J. C. Gildersleeve, *Mol. Cancer Ther.* **2009**, *8*, 971-979.
[29] P. R. Desai, *Tranfus. Med. Rev.* **2000**, *14*, 312-325.
[30] a) S. Ingale, M. A. Wolfert, J. Gaekwad, T. Buskas, G. J. Boons, *Nat. Chem. Biol.* **2007**, *3*, 663-667; b) R. Lo-Man, S. Vichier-Guerre, R. Perraut, E. Dériaud, V. Huteau, L. BenMohamed, O. M. Diop, P. O. Livingston, S. Bay, C. Leclerc, *Cancer Res.* **2004**, *64*, 4987-4994.
[31] S. Slovin, G. Ragupathi, C. Fernandez, M. Diani, M. Jefferson, A. Wilton, W. K. Kelly, M. Morris, D. Solit, H. Clausen, P. Livingston, H. Scher, *Cancer Immunol. Immunother.* **2007**, *56*, 1921-1930.
[32] P. J. Sabbatini, G. Ragupathi, C. Hood, C. A. Aghajanian, M. Juretzka, A. Iasonos, M. L. Hensley, M. K. Spassova, O. Ouerfelli, D. R. Spriggs, W. P. Tew, J. Konner, H. Clausen, N. Abu Rustum, S. J. Dansihefsky, P. O. Livingston, *Clin. Cancer Res.* **2007**, *13*, 4170-4177.

1.3. Mucins

A characteristic of mucosal epithelial cells is the production of mucins and other *O*-glycoproteins, whose expression is often altered in different cancers.[33] At least 21 different mucin genes have been identified (MUC1, 2, 3A, 3B, 4, 5AC, 5B, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19, 20, and 21), and although they are not related in sequence, they share the features of having repeating motifs rich in Ser and Thr residues linked to large amounts of *O*-glycans. These mucins can contain 377 to more than 11,000 amino acids residues in a single polypeptide and their 'gel-like' form plays an important role not only in lubrication and protection of epithelial surfaces from either microbial or chemical attack, but also in regulation of cell growth, signaling and differentiation.[34]

The most known and characterized mucin is MUC1.[35] It is a transmembrane glycoprotein, expressed by many glandular epithelial cells. It was first found in human milk as a large molecular weight glycoprotein containing many serines, threonines and prolines and having high content of *O*-linked carbohydrates. MUC1 is expressed in many normal cells but is overexpressed in a variety of adenocarcinomas, such as breast cancer, lung cancer, pancreatic cancer, and prostate cancer.

In normal tissues, MUC1 is found at the apical side of epithelial cells and typically with high glycosylation level. By contrast, in MUC1-overexpressing

---

[33] a) J. Burchell, A. Mungul, J. Taylor-Papadimitriou, *J. Mammary Gland Biol. Neoplasia* **2001**, *6*, 355-364; b) J. Byrd, R. Bresalier, *Cancer Metastasis Rev.* **2004**, *23*, 77-99.

[34] J. Taylor-Papadimitriou, S. J. Gendler, *Int. J. Oncol.* **1992**, *1*, 9-16.

[35] a) S. Nath, P. Mukherjee, *Trends Mol. Med.* **2014**, *20*, 332-342; b) N. Jonckheere, N. Skrypek, I. Van Seuningen, *Biochim. Biophys. Acta* **2014**, *1846*, 142-151; c) N. Gaidzik, U. Westerlind, H. Kunz, *Chem. Soc. Rev.* **2013**, *42*, 4421-4442; d) H. C. Hang, C. R. Bertozzi, *Bioorg. Med. Chem.* **2005**, *13*, 5021-5034.

tumor tissues, it is expressed not only on the apical side but also all around the cell surface. More importantly, MUC1 shows a low glycosylation level in tumor cells (Figure 1.7).

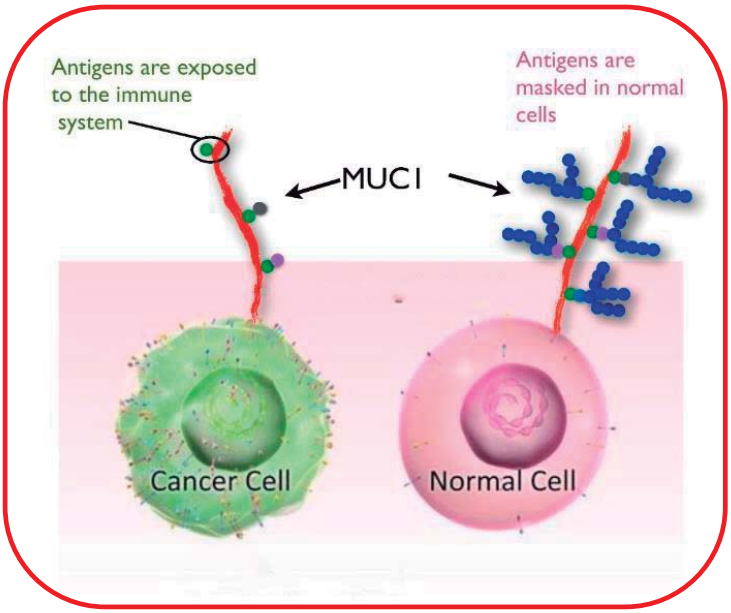


Figure 1.7. *Different glycosylation pattern in healthy and tumor cells.*

From a structural point of view, MUC1 consists of a tandem repeat of 20 amino acids: AHGVTSAPDTRPAPGSTAPP.[36] Each tandem repeat contains five potential *O*-glycosylation sites (three threonines and two serines) and three different important regions: the GVTSA sequence, which is an effective substrate for GalNAc transferases,[37] the PDTR fragment, which is the most immunogenic domain of MUC1, and consequently, a well-known epitope recognized by several anti-MUC1 antibodies[38] and the GSTAP region, which

[36] a) C. L. Hattrup, S. J. Gendler, *Annu. Rev. Physiol.* **2008**, *70*, 431-457; b) M. A. Tarp, H. Clausen, *Biochim. Biophys. Acta Gen. Subj.* **2008**, *1780*, 546-563; c) H. C. Hang, C. R. Bertozzi, *Bioorg. Med. Chem.* **2005**, *13*, 5021-5034.
[37] N. Gaidzik, U. Westerlind, H. Kunz, *Chem. Soc. Rev.* **2013**, *42*, 4421-4442.
[38] U. Karsten, N. Serttas, H. Paulsen, A. Danielczyk, S. Goletz, *Glycobiology* **2004**, *14*, 681-692.

is recognized by different antibodies and represents a potential tool in diagnosis and therapeutic applications.[39]

MUC1 biosynthesis begins in the Golgi apparatus by the addition of *N*-acetylgalactosamine (GalNAc) from UDP-GalNAc to threonines and serines to form the Tn antigen.[40] This glycosylation is performed by a family of different *N*-acetylgalactosaminyltransferases that attach GalNAc and other sugars (sialic acids, fucoses…) forming core structures that protects the cell. Problems in this pathway can degenerate into short and non-branched glycans exposing to the immune system Tn or T antigens.

Based on these facts, glycopeptide vaccines can constitute an important strategy in cancer immunotherapy and, in general, the tandem repeats from mucins become the center of attention. Therefore, mucins that exhibit Tn antigen in their structures are very good candidates to perform an exhaustive research in order to develop new and better vaccines against tumor cells.

---

[39] Z. H. Huang, L. Shi, J. W. Ma, Z. Y. Sun, H. Cai, Y. X. Chen, Y. F. Zhao, Y. M. Li, *J. Am. Chem. Soc.* **2012**, *134*, 8730-8733.

[40] a) D. J. Gill, K. M. Tham, J. Chia, S. C. Wang, C. Steentoft, H. Clausen, E. A. Bard-Chapeau, F. A. Bard, *Proc. Natl. Acad. Sci.* **2013**, *110*, E3152-E3161; b) S. Rottger, J. White, H. H. Wandall, J. C. Olivo, A. Stark, E. P. Bennett, C. Whitehouse, E. G. Berger, H. Clausen, T. Nilsson, *J. Cell Sci.* **1998**, *111*, 45-60.

1.4. General objectives of the research

Bearing in mind the importance of the mechanism of recognition of carbohydrates and glycopeptides by lectins, in particular, Tn antigen in the MUC1 molecule context, in this Thesis we want to study the interactions between some lectins and MUC1-derivatives bearing this Tn determinant.

In consideration of previous works and significance of the field of work, the objectives we have chased in this Thesis are as follows:

- To investigate the importance of the underlying amino acid in MUC1-derivatives recognition when they incorporate Tn antigen. Generally, we talk equally about Tn bearing Ser or Thr but… is it correct? Is there any difference between them in terms of recognition? If the answer is yes, why?

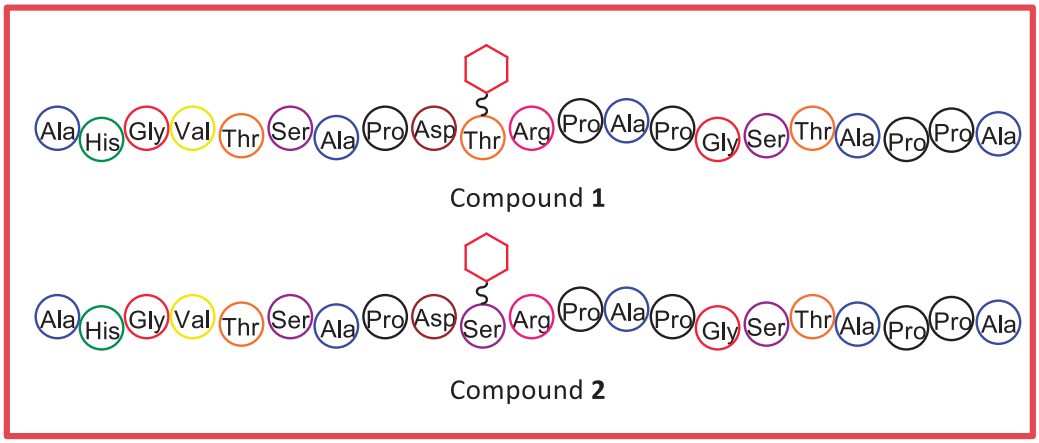


Figure 1.8. *Threonine and serine MUC1-epitopes derivatives target molecules.*

- To understand the key factors that govern the molecular recognition of MUC1-glycopeptides bearing Tn antigen by lectins. Peptide context nearby the glycosylation point and the underlying amino acid are

significant for recognition. Does the lectin recognize only the carbohydrate moiety? What interactions do take place in the molecular recognition process? It is possible to modulate them?

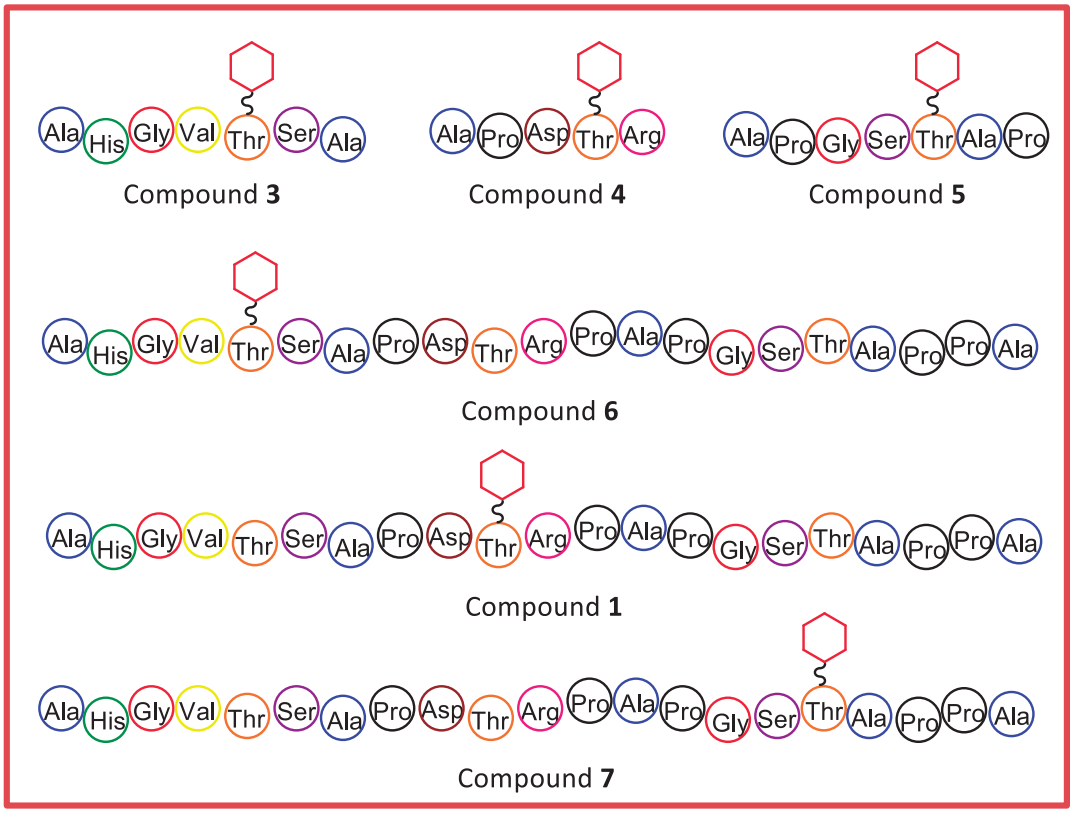


Figure 1.9. *Threonine MUC1-epitopes derivatives target molecules.*

- To characterize the free state in water of short MUC1-glycopeptides as well as their bond-states to some specific lectins. To achieve this goal, we use NMR studies (STDs, NOEs, COSY…), molecular dynamics (MD) simulations, biological probes (ELLA, ITCs…) and X-ray analysis.

As a global and final objective, the detailed analysis of the interactions between MUC1 epitopes and various lectins will open the doors to a better

understand of the molecular recognition process of glycopeptide-lectin system, which in turns will allow us to obtain novel lectins with higher affinity by selective mutation of some crucial residues. As a consequence, we could control the interactions and improve biosensors in cancer research in tumor cells context.

TECHNIQUES & TOOLS

2.1. Solid phase peptide synthesis (SPPS)

2.1.1. Introduction
2.1.2. Protection of amino and side-chain functionalities
2.1.3. Optimal resins for peptide synthesis
2.1.4. Activation and coupling reagents

2.2. NMR experiments

2.2.1. 2D COSY (Correlation Spectroscopy)
2.2.2. 2D NOESY (Nuclear Overhauser Spectroscopy)
2.2.3. Coupling constants
2.2.4. Saturation-transfer difference (STD) NMR Spectroscopy

2.3. Molecular dynamics (MD) simulations and 3D modeling

2.4. Enzyme-linked lectin assay (ELLA)

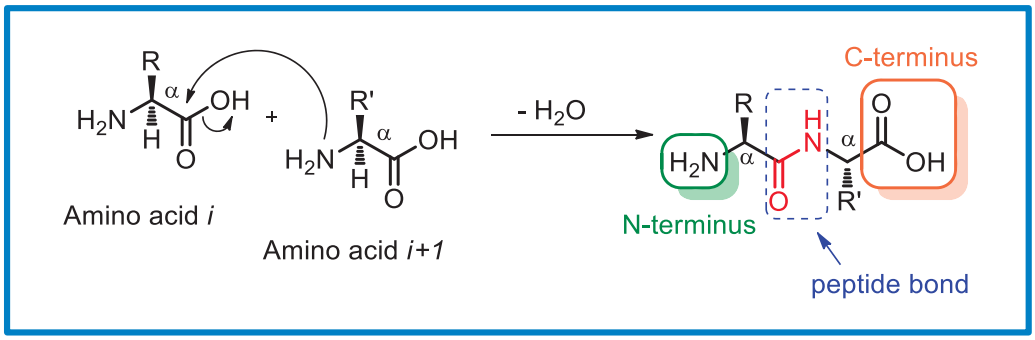
2.5. Isothermal titration calorimetry (ITC)

2.1 Solid phase peptide synthesis (SPPS)

2.1.1. Introduction

As explained in *Chapter One*, proteins, peptides and glycopeptides are important molecules in biochemical and physiological processes and their synthesis has been carried out combining chemical or enzymatic reactions and solid phase strategy. The concept of solid phase synthesis was developed in 1963 by Merrifield.[1]

In cells, peptides and proteins are produced in ribosomes through the involvement of aminoacyl-tRNAs (transfer ribonucleic acid). This process involves a simple reaction of condensation between a carboxylic acid of an α-amino acid and an amine group of the next amino acid. The result is the formation of the amide (or peptide) bond (Scheme 2.1).



Scheme 2.1. *Peptide bond formation.*

In nature, L-enantiomers of α-amino acids are elected for combining and produce (glyco)peptides and proteins. There is a list of 20 natural amino acids that are combined in the ribosomes (Figure 2.1), except in some rare exceptions.

[1] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.

The main goal of the invention of solid-phase peptide synthesis was, from the beginning, to open the path for a faster, simpler and a particularly more automated mode of operation for chemical peptide synthesis.

The instrument should be equipped with a piping system to perform pumping, mixing and removal of solutions. Moreover, the automated peptide robot should contain reservoirs for all chemicals (amino acids, solvents, activators and other agents) and ensure adequate delivery of the solutions.

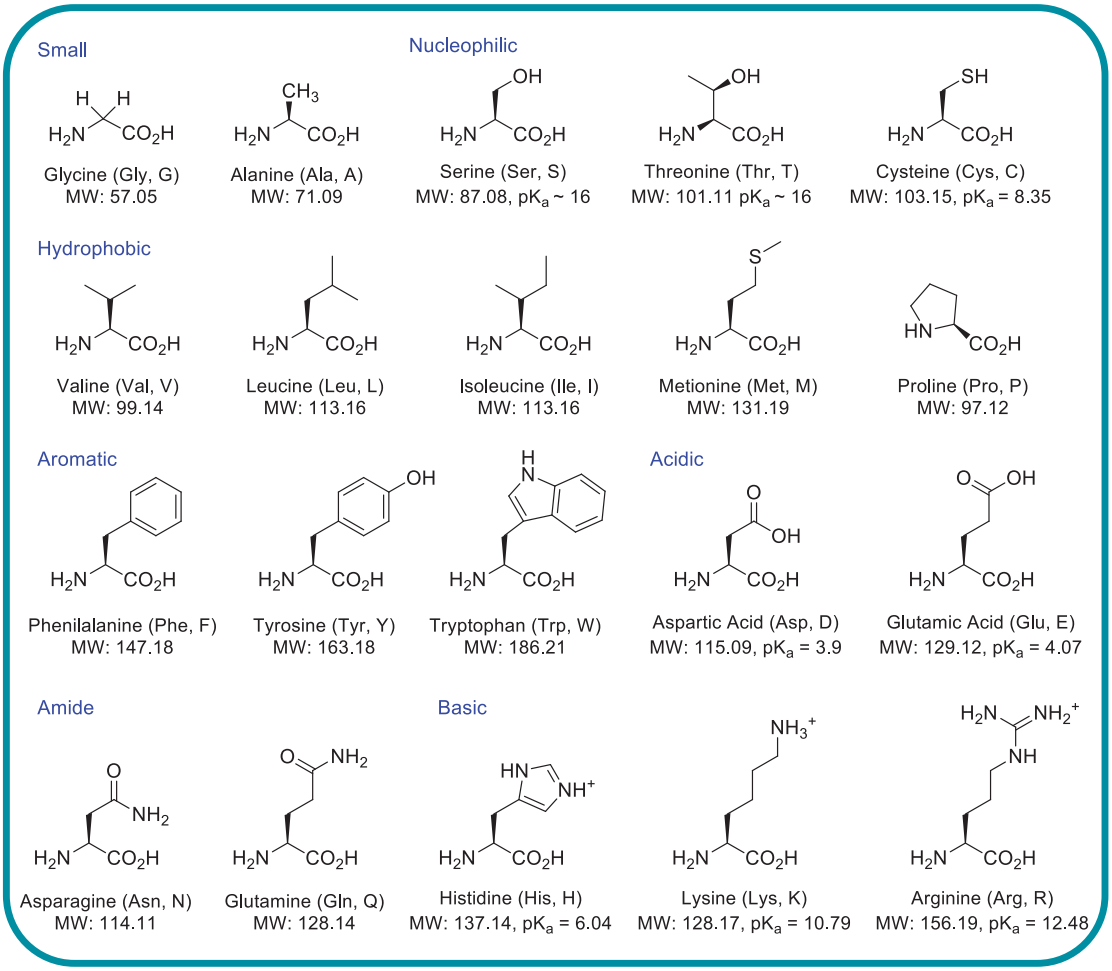


Figure 2.1. *List of natural 20 amino acids. It includes molecular weight and pK_a of the side chain groups.*

The improvement of chemical reactions, solid supports, linkers and especially the development of the Fmoc-based SPPS-strategy have contributed to simplification and many advancement of following instruments.[2]

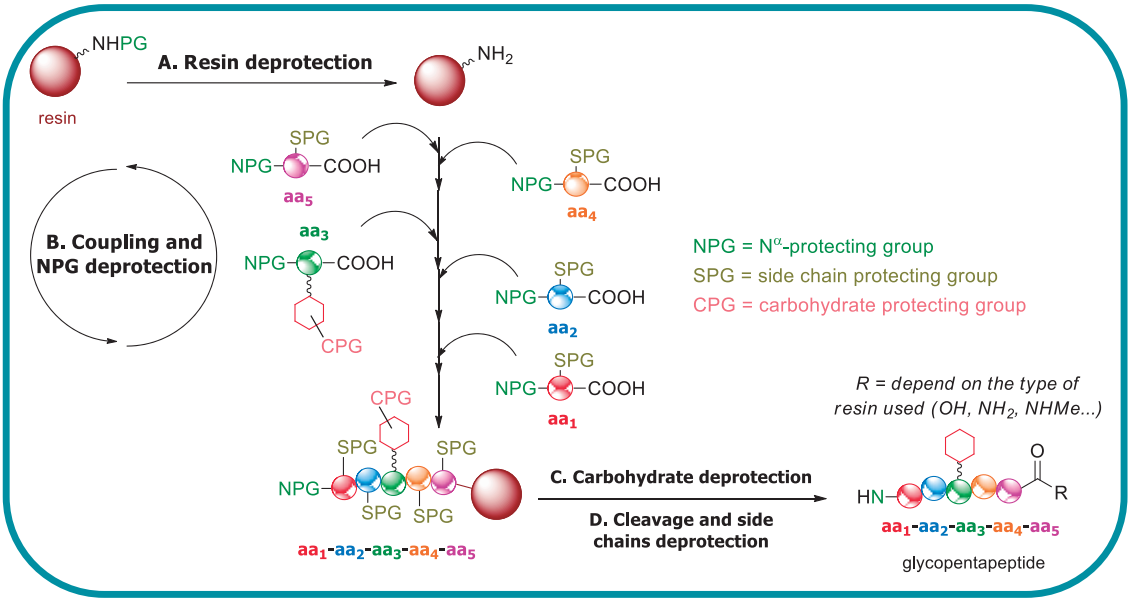
SPPS needs an insoluble solid support, called resin, and liquid soluble reagents in solvents as DMF or NMP and CH_2Cl_2. In this type of synthesis, the carboxylic acid terminus of the first amino acid is always coupled to a resin. Resins consist of polymer particles and protect the C-terminal of generating peptide from side reactions during the process.

In order to get better results and increase the reaction yield, the N-terminal group and reactive side-chain moieties of amino acids have to be blocked. N-protecting groups must be removed specifically after each successful coupling step, while the side-chain protecting groups and the resin keep protected against unwanted side reactions. Besides, the carboxyl groups have to be activated by special auxiliaries to increase the electrophilicity in the peptide bond formation reaction. After loading of the resin, the N-terminal protecting group of the first amino acid is removed and the next activated amino acid can be coupled. These alternating steps of N^{α}-deprotection, activation and coupling are repeated until the desired peptide chain is obtained.

As the next step, it is desirable to remove the side-chain protecting groups and to liberate the peptide from the resin in only one step. This is possible under the same previous conditions and this process is called cleavage. In our case, it will be under acid media. In a glycopeptide, the carbohydrate is

[2] E. Atherton, H. Fox, D. Harkiss, C. J. Logan, R. C. Sheppard, B. J. Williams, *J. Chem. Soc., Chem. Commun.* **1978**, 537-539.

protected conveniently so, as a previous step before the cleavage, it is necessary the deprotection of protecting groups of the carbohydrate (Scheme 2.2).



Scheme 2.2. *Solid phase peptide synthesis (SPPS) route for obtaining a model glycopentapeptide.*

The crude product can be easily separated from the resin and purified by standard analytical methods such as HPLC and identify the correct peak by mass spectrometry (MS) with soft ionization techniques such as MALDI–TOF (matrix-assisted laser desorption ionization-time of flight).

One of the great advantages of SPPS is the possibility of carrying out all reactions in a single vessel. Following a coupling step, unreacted reagents and byproducts can be easily removed by washing, without purification in the intermediate steps. Based on the use of excess amounts of reactants (around 8-10 equivalents of amino acid per resin equivalent), high coupling yields can be obtained (95% per coupling step) and the reaction cycles are very short compared to solution synthesis, which allows faster synthesis.

2.1.2. Protection of amino and side-chain functionalities

Choosing the precise protecting groups is something extremely important for developing a correct synthesis: to protect against side reactions and formation of undesired chemical bonds. Requirements for appropriate protecting groups are always the same: simple incorporation into the desired molecule, high stability against several conditions and easy and safe removal.[3]

In SPPS there are two very common protecting groups for the amino group that have been used many time ago: Boc (*tert*-butoxycarbonyl)[4] and Fmoc (9-fluorenylmethoxycarbonyl)[5]. The initial method developed by Merrifield was based on the Boc/Bn strategy, where the Boc group was used as temporary protecting group of the amino function and Bn (benzyl) of the side chains of amino acids. Standard deprotection of Boc requires treatment with TFA and Bn removal needs strong acids such as HF (Figure 2.2).[6, 7]

Despite this protecting strategy is accepted, the advantages of the Fmoc/^{t}Bu strategy enables the selective removal of the protecting groups using completely different chemical conditions and cleavage mechanisms (Figure 2.2), which ensures milder overall reactions and the use of TFA difficult aggregation of the peptide.[8]

---

[2] T. W. Green, P. G. M. Wuts, *Protective groups in organic chemistry,* 3rd ed.; John Wiley & Sons, Inc.: New York, USA, 1999.
[4] G. W. Anderson, A. C. McGregor, *J. Am. Chem. Soc.* **1957**, *79*, 6180-6183.
[5] a) L. A. Carpino, *Acc. Chem. Res.* **1987**, *20*, 401-407; b) L. A. Carpino, G. Y. Han, *J. Am. Chem. Soc.* **1970**, *92*, 5748-5749.
[6] M. Pennington, in *Peptide Synthesis Protocols, Vol. 35* (Eds.: M. Pennington, B. Dunn), Humana Press, **1995**, 41-62.
[7] V. Mäde, S. Els-Heindl, A. G. Beck-Sickinger, *Beilstein J. Org. Chem.* **2014**, *10*, 1197-1212.
[8] M. Beyermann, M. Bienert, *Tetrahedron Lett.* **1992**, *33*, 3745-3748.

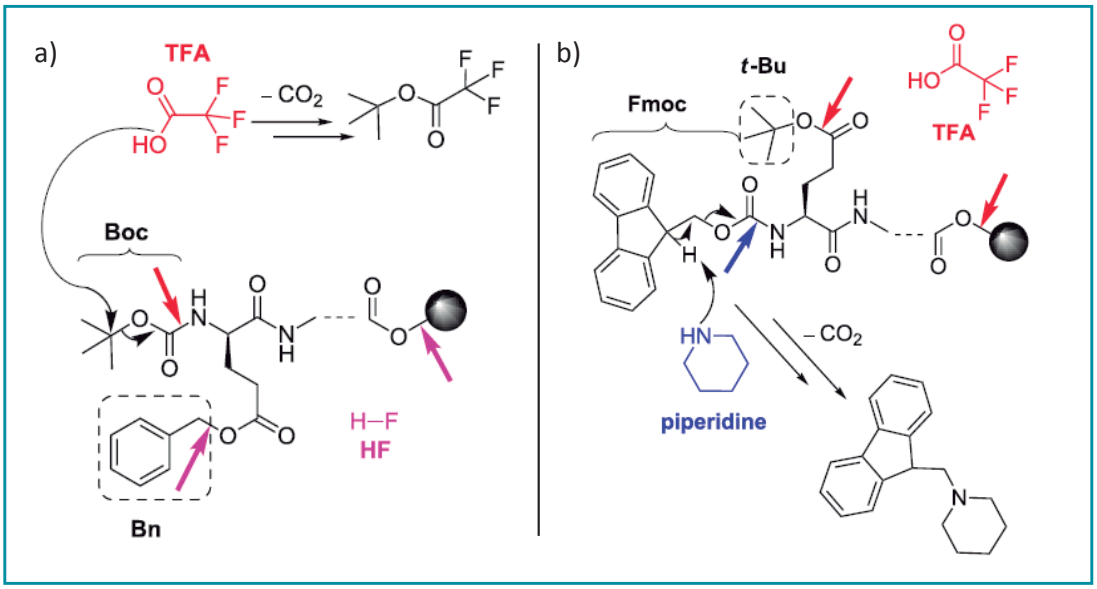


Figure 2.2. *a) Boc/Bn protecting-group strategy used in SPPS with deprotection with HF (purple) and TFA (red). b) Fmoc/^{t}Bu strategy in SPPS with deprotection with basic (in blue) and acid media (red).*

Besides, Fmoc strategy does not require the use of special vessels that have to be stable towards the corrosive and toxicity of HF and, in some cases, the repetitive TFA acidolysis for Boc deprotection could have an important impact on sensitive peptide bonds and acid-catalyzed side reactions.[9]

In the case of the amino acids, there are a remarkable diversity of side-chain protection groups that it has been evolved since the development of SPPS more than 50 years ago. Some of the most common protecting groups in the Fmoc/^{t}Bu strategy are compiled in Figure 2.3 but there is a number of diverse orthogonal protecting groups commercially available depending on the requirements of the synthesis.[10]

---

[9] K. H. Hsieh, M. M. Demaine, S. Gurusidaiah, *Int. J. Pept. Protein Res.* **1996**, *48*, 292-298.
[10] A. Isidro-Llobet, M. Álvarez, F. Albericio, *Chem. Rev.* **2009**, *109*, 2455-2504.

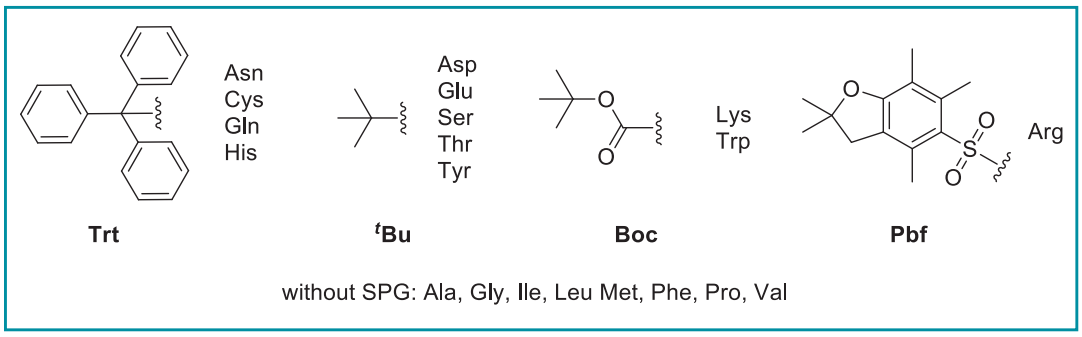


Figure 2.3. *Commonly amino acid side chain protecting groups (SPG) in Fmoc/^{t}Bu strategy.*[6]

We have also followed the Fmoc/^{t}Bu strategy for accomplish the synthesis of our glycopeptides in this thesis.

2.1.3. Optimal resins for peptide synthesis

Other important parameter in SPPS is the resin, which has to overcome a number of requirements to be appropriate for peptide synthesis:

- Insoluble in all solvents used during the synthesis
- Chemically and physically resistant to all conditions
- Mechanically stable to allow filtration

The small resin beads can enlarge up to six times of their original volume in organic solvents, so the solid matrix requires an appropriate solvation, low cross linking for good accessibility and good swelling properties.[11] Besides, the polymer needs to have a functional group for coupling the linker that depends on the desired final peptide. At present, there are different classes

[11] R. B. Merrifield, *Angew. Chem. Int. Ed.* **1985**, *24*, 799-810.

of solid carriers: traditional polystyrene (PS), polyethylene glycol (PEG)-functionalized with PS and pure PEG-based resins.[12]

The linker exhibits a reversible connection between the solid support and the new synthesized peptide. It defines the loading of the resin (in equivalents of amino acid in nmol/g), the distance between resin and peptide, chemical conditions for coupling and release and most importantly, the C-terminal functionality of the synthetic peptide. In most cases, the peptide is released as acid or amide because these are naturally occurring C-terminal functionalities.

Regarding the protocol followed in our lab, we have selected the Rink Amide MBHA resin as solid support (Figure 2.4).

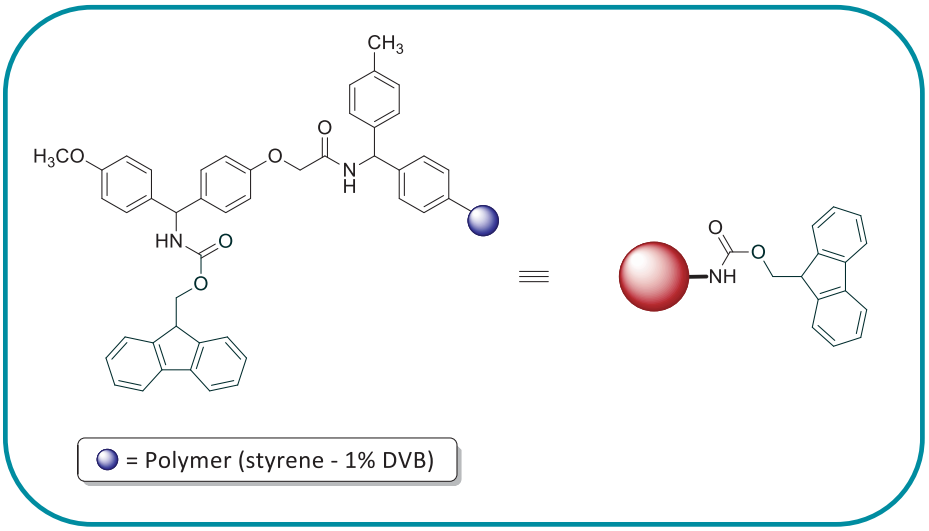


Figure 2.4. *Resin used in SPPS (Rink Amide MBHA resin) and its drawing simplification.*

[12] D. Hudson, *J. Comb. Chem.* **1999**, *1*, 333-360.

2.1.4. Activation and coupling reagents

The free carboxy terminus is not reactive enough to form peptide bonds by itself, so it has to be transformed into an active, more electrophilic group. Thus, coupling reagents are needed in the activation of C-terminal. There are numerous compounds that have been used for many years.[13] Nowadays, a great variety of coupling reagents are commercially available reaching from traditional carbodiimides (DCC, DIC) and classical auxiliary nucleophiles (HOBt, HOAt)[14] to uronium reagents such as HBTU (*N,N,N′,N′*-tetramethyl-*O*-(1H-benzotriazol-1-yl)uronium hexafluorophosphate) or TBTU (*N,N,N′,N′*-tetramethyl-*O*-(benzotriazol-1-yl) uronium tetrafluoroborate),[15] and phosphonium salts such as PyBOP [(benzotriazol-1-yloxy)tripyrrolidino-phosphonium hexafluorophosphate][16] (Figure 2.5).

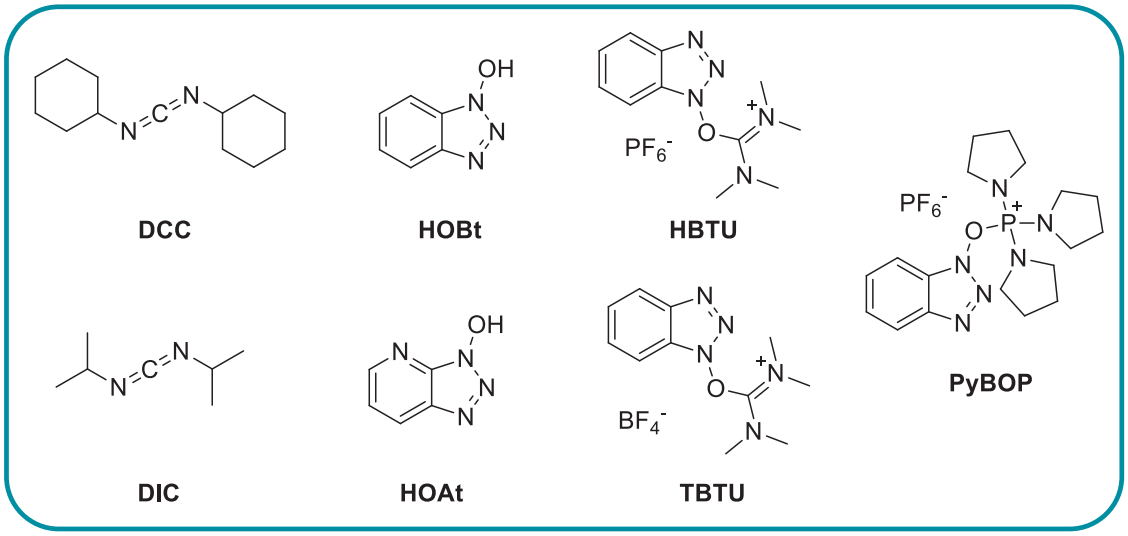


Figure 2.5. *A selection of coupling reagents for SPPS.*

[13] K. V. S. R. G. Prasad, K. Bharathi, B. Haseena-Banu, *Int. J. Pharm. Sci. Rev. Res.* **2011**, *8*, 108-119.
[14] L. A. Carpino, *J. Am. Chem. Soc.* **1993**, *115*, 4397-4398.
[15] J. Hachmann, M. Lebl, *Pept. Sci.* **2006**, *84*, 340-347.
[16] J. Coste, D. Le-Nguyen, B. Castro, *Tetrahedron Lett.* **1990**, *31*, 205-208.

Depending on the number of amino acids to be coupled and the chemical conditions, such as its solubility and its stability, the decision of the proper coupling reagent, offering enhanced reactivity by simultaneous reduction of epimerization, is of high relevance. In our synthesis, we have selected HBTU as coupling agent because it showed a higher reactive than others reagents and gave better results than traditional ones.

2.2. NMR experiments

NMR has been a major technique for the identification and the determination of peptides and glycopeptides structures in solution as well as their interaction with their biological targets under near-physiological conditions. In general, NMR studies with glycopeptides involve three fundamental steps: spectrum acquisition, structural information extraction and molecular modeling based on structural information. Some of the most used and important experiments in the glycobiology field will be explained.

2.2.1. 2D COSY (Correlation Spectroscopy)

2D ^{1}H, ^{1}H COSY correlates all ^{1}H resonances that are scalar coupled. The spectrum is symmetrical with the intensity of the cross peaks dependent upon the coupling constant. In peptides, it can be used to identify which NH^{i} signals are linked to $H\alpha^{i}$ signals and therefore how many backbone NH signals are present in the molecule. All predicted peaks are not necessarily observed due to weak couplings, obscured by solvent, noise, overlap or degenerate peaks (Figure 2.6).

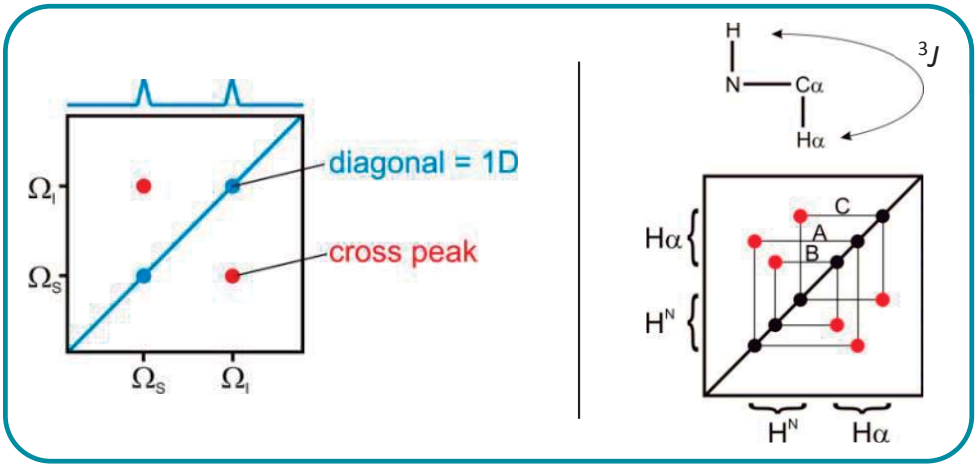


Figure 2.6. *Schematic representation of a 2D COSY experiment.*

2.2.2. 2D NOESY (Nuclear Overhauser Spectroscopy)

The major source of conformational information comes from the nuclear Overhauser effect (NOE). The magnitude of the NOESY depends on the distance separation of the interacting spins (r), but also on its motional properties (τ_c):[17]

$$NOE \propto r^{-6} \cdot f(\tau_c) \qquad \textbf{Equation 2.1}$$

The NOESY experiment correlates protons that are close in space (< 5Å). While a NOESY spectrum can contain cross peaks also seen in a COSY spectra, the mechanism of magnetization transfer is different. NOESY magnetization is via cross relaxation and is a through-space effect, COSY is through *J*-coupling and is a through-bond effect.

NOEs are by far the most important source of information for the structure calculation. However, there is an important problem for larger molecules: NOEs are not only observed between spins close in space but also transfer

[17] J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *J. Chem. Phys.* **1979**, *71*, 4546-4553.

via a relay nucleus may have occurred (spin-diffusion), depending on the mixing time of the NOESY experiment.

Peak volumes, derived from integration of the peaks, will be translated into distances.

In this technique, sequential distances (NOEs between protons of amino acids in regular secondary structure) are chosen such that known NOEs of certain distances will correspond to typically observed values. In fact, the presence of NOEs NH(i), NH(i+1) is typical for α-helix. By contrast, strong sequential NOE NH(i), Hα(i+1) besides absence of NOE NH(i), NH(i+1) suggests β-sheets structures (Figure 2.7).

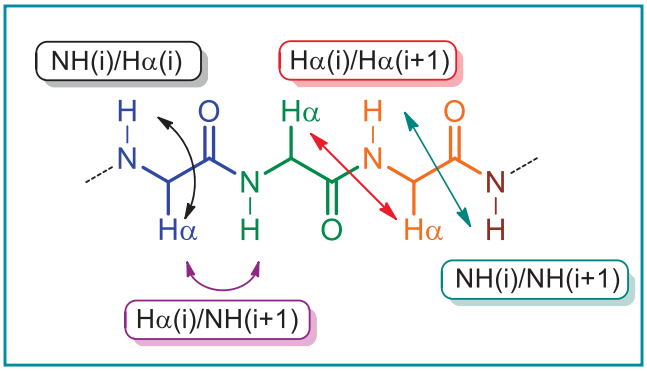


Figure 2.7. *Relevant NOEs in structural analysis of proteins and glycopeptides.*

It is important to highlight that NOEs are measured on a 50 ms to 1 s time scale. This does not matter if the molecule is rigid, but if it is flexible or converts between several conformations then an average NOE will be measured. These cannot even be easily interpreted in terms of an average conformation because of the r^{-6} dependence of the NOE.

2.2.3. Coupling constants

The most useful H-H coupling relationship is related to vicinal protons. The magnitude of $^3J_{H\text{-}H}$ is predictable and provides detailed information about the three-dimensional orientation between the two protons. Almost all $^3J_{H\text{-}H}$ values are positive, but their magnitude varies widely (from close to 0 Hz up to 25 Hz) depending on structural and conformational details. These constants are associated to the corresponding torsion angles by Karplus equation[18] (Equation 2.2) where the constant values are empirically fixed to each particular case.

$$^3J(\theta) = 6.98(cos\theta)^2 - 1.38(cos\theta) + 1.72 \qquad \textbf{Equation 2.2}$$

In glycobiology field, we can obtain important information through the analysis of $^3J_{H\text{-}H}$ constants: $^3J_{NH\text{-}H\alpha}$ constants give information about the peptide chain conformation (Φ_p, dihedral angle) and $^3J_{H\alpha\text{-}H\beta}$ constants are related to the orientation (χ^1 angle) (Figure 2.8).

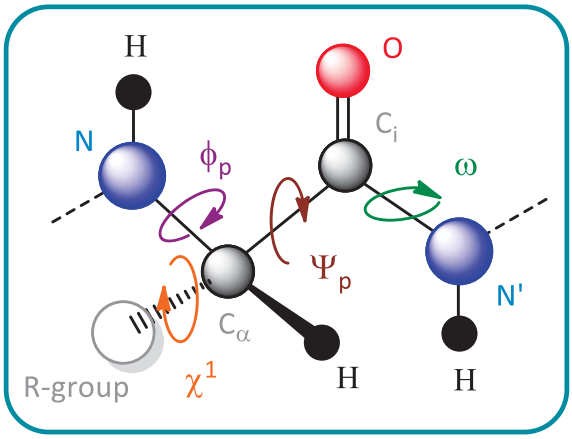


Figure 2.8. *Significant dihedral angles in structural analysis of proteins and glycopeptides.*

[18] a) M. Karplus, *J. Am. Chem. Soc.* **1963**, *85*, 2870-2871; b) R. C. Breton, W. F. Reynolds, *Nat. Prod. Rep.* **2013**, *30*, 501-524.

It is also important to point out that sometimes the target molecule explores different conformational regions, therefore NOE-derived distances as well as coupling constants are an average of all conformations existing in solution. In these cases, computational techniques, such as molecular dynamics can understand the experimental data.[19]

2.2.4. Saturation-Transfer Difference (STD) NMR Spectroscopy

The (STD) NMR experiment has emerged as one of the most popular spectroscopic technique to study the interactions in solution, between a large molecule (protein) and a medium-small sized molecule (ligand), and it is based on the nuclear Overhauser effect and the observation and analysis of the resonances of the ligand protons. [20]

It is very useful to find the binding epitope in a ligand protein interaction because this technique characterizes the hydrogen atoms of the ligand that are closer to the protein in the bound state.

(STD) NMR spectroscopy is based on the different physical properties that exist between a protein and a ligand: intermolecular transfer of magnetization from the protein (selectively irradiated) to the bound molecule (ligand) through the nuclear Overhauser effect (Figure 2.9).[21]

In general, the selective irradiation consists of a cascade of Gaussian-shaped pulses that saturate few protein resonances. This saturation is then spread over the entire protein by spin diffusion and to the bound molecule by

---

[19] a) P. I. Koukos, N. M. Glykos, *J. Comput. Chem.* **2013**, *34*, 2310-2312; b) S. A. Adcock, J. A. McCammon, *Chem. Rev.* **2006**, *106*, 1589-1615.
[20] M. Mayer, B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784-1788.
[21] A. Viegas, J. O. Manso, F. L. Nobrega, E. J. Cabrita, *J. Chem. Educ.* **2011**, *88*, 990-994.

intermolecular NOE. The dissociation of the ligand will then transfer this saturation into bulk solution where it accumulates during the saturation time of the experiment, as result of a much slower relaxation process.

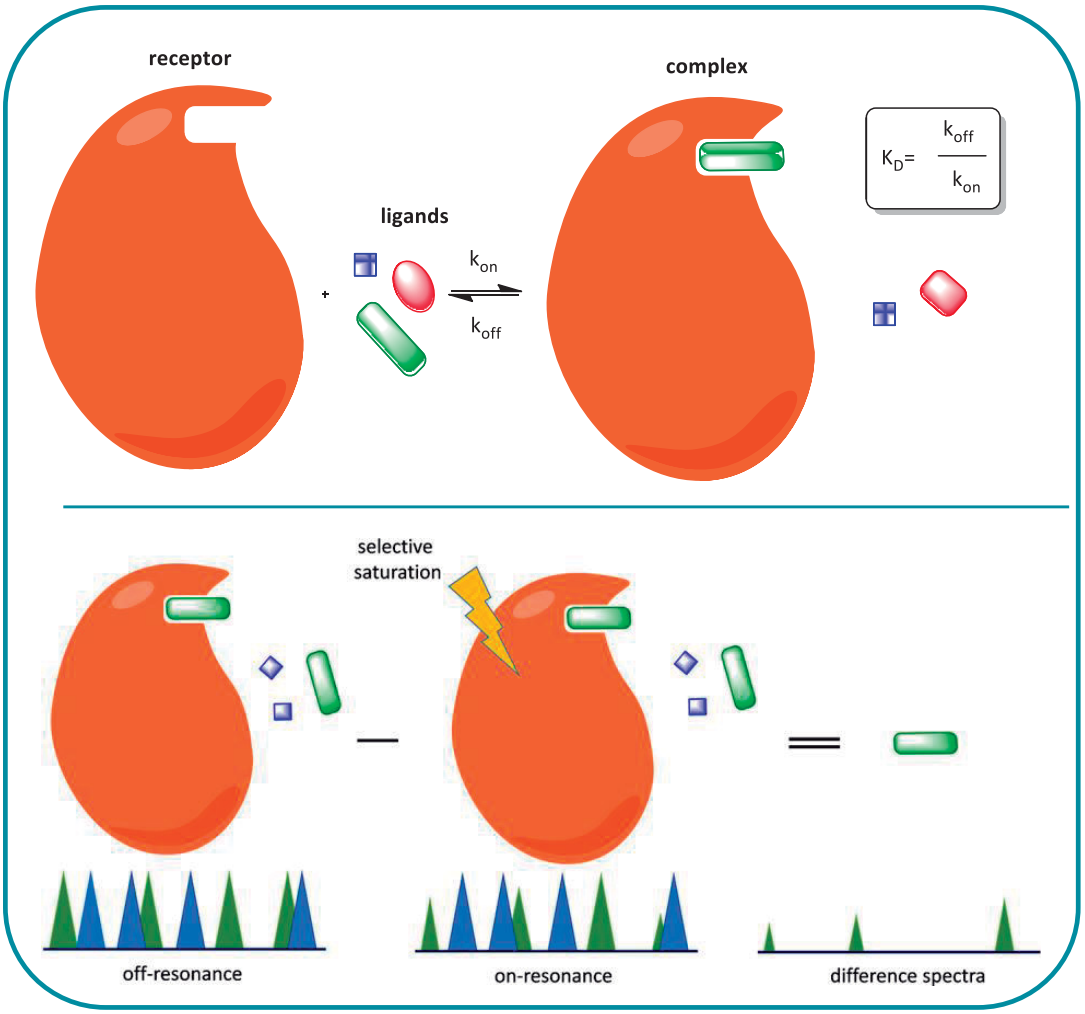


Figure 2.9. *Scheme of the (STD) NMR experiment. The exchange between free and bound ligand allows intermolecular transfer of magnetization from the receptor (orange) to the bound small molecule (green).*

The (STD) NMR experiment relies on the fact that, for a weak binding ligand (dissociation constant, K_D, ranging from 10^{-8} mol/L to 10^{-3} mol/L), there is

exchange between the bound and the free ligand state. Basically, an STD experiment involves subtracting two spectra. Firstly, we have to record a spectrum without protein saturation (off-resonance spectrum), with signal intensities I_0. Second, a spectrum in which the protein was selectively saturated, on-resonance spectrum, obtained by irradiating at a region of the spectrum that contains only resonances of the protein such as 0 ppm to -1 ppm. If the ligands show no resonance signals in the aromatic proton spectral region, the saturation frequency may also be placed there (7 ppm) or even further downfield (δ = 11–12 ppm).

A blank experiment must be carried out to assure the absence of direct irradiation of the ligand. Here we obtain a spectrum with signal intensities I_{SAT}. In the difference spectrum ($I_{STD} = I_0 - I_{SAT}$) only the signals of the ligand(s) that received saturation transfer from the protein (via spin diffusion, through the nuclear Overhauser effect) will remain.

Other compounds that may be present but do not bind to the receptor will not receive any saturation transfer; their signals will be of equal intensity on the on-resonance and the off-resonance spectra and, as a consequence, after subtraction no signals will appear in the difference spectrum from the nonbinding small molecule(s) (Figure 2.9).

The difference in intensity due to saturation transfer can be quantified and constitutes an indication of binding. For a molecule that binds to the receptor, only the signals of the hydrogens that are in close contact to the protein (less than 5 Å) and receive magnetization transfer will appear in the difference spectrum. From those, the ones that are closer to the protein will have more intense signals, owing to a more efficient saturation transfer.

By normalizing all the measured STD signals ($I_0 - I_{SAT}/I_0$) against the most intense signal (which is arbitrarily assigned a value of 100%), and comparing the relative percentages of saturation received by the different ligand protons, it is possible to obtain the so-called 'epitope mapping'. In this way, the resulting STD values expressed in percentages represent a map of the ligand-protein contacts in the bound state. This illustrates the moieties of the ligand that are key for molecular recognition by the receptor in the binding-pocket.[22]

2.3. Molecular dynamics (MD) simulations and 3D modeling

The elucidation of the 3D structure and dynamics properties of a ligand, both in the free state and bound to proteins, is a prerequisite for a better understanding of the molecular basis of their associations and interactions. In this sense, it is well-known the important relationship between structure and function, which is involved in the biochemistry of recognition processes and the subsequent rational design of carbohydrate-derived drugs.

Molecular dynamics (MD) simulations are an important tool for studying dynamics properties of biomolecules. MD simulations have been developed very fast since 50s, following the ideas of molecular mechanics (MM). Here, molecules are simply defined as a set of atoms that is held together by simple elastic (harmonic) forces. Consequently, bond order is related by an elastic constant (Figure 2.10).[23]

---

[22] M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108-6117.

[23] a) G. A. Cisneros, M. Karttunen, P. Ren, C. Sagui, *Chem. Rev.* **2013**, *114*, 779-814; b) J. Wereszczynski, J. A. McCammon, *Q. Rev. Biophys.* **2012**, *45*, 1-25; c) D. Alexander, J. R. Mackerell, *J. Comput. Chem.* **2004**, *25*, 1584-1604.

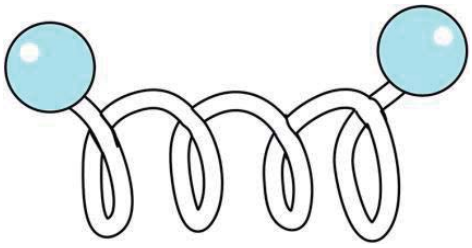


Figure 2.10. *Schematic representation of atoms as spheres and bonds as springs in molecular dynamics.*

The most important aspect of MD simulations is the force field. It is a mathematical expression that describes the dependence of the energy of a system on the coordinates of its particles. It consists of an analytical form of the interatomic potential energy (U) and a set of parameters entering into this form (Figure 2.11). The parameters are typically obtained either from ab initio or semi-empirical quantum mechanical calculations or by experimental data such as neutron, X-ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy, etc.

The force field replaces the true potential with a simplified model valid in the region being simulated. Ideally it must be simple enough to be evaluated quickly, but sufficiently detailed to reproduce the properties of interest of the system studied. There are many force fields available in the literature, having different degrees of complexity, and oriented to treat different kinds of systems.[24] In our case, we used the AMBER force field [25] because it has the optimal parameters for the correct conformational study of (glyco)peptides

[24] a) B. L. Foley, M. B. Tessier, R. J. Woods, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 652-697; b) M. A. González, *Collect. SFN* **2011**, *12*, 169-200.

[25] W. D. Cornell, P. C. Cieplack, I. Bayly, I. R. Gould, K. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

and proteins. Besides, to properly simulate the carbohydrate moiety we have used the GLYCAM[26] parameters combined with AMBER ones.

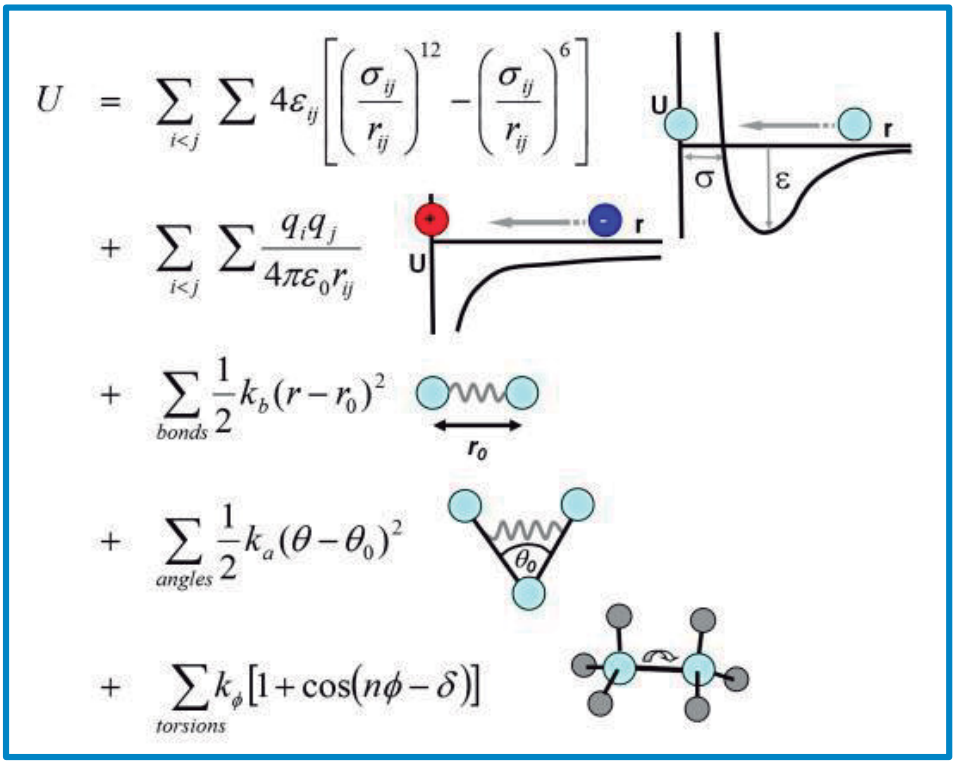


Figure 2.11. *Schematic representation of the interatomic potential energy in MD.*

MD simulations are based on Newton's second law or the equation of motion. From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then allows the description of the positions, velocities and accelerations of the particles as they vary with time. The method is mainly deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past. Molecular dynamics simulations can be time consuming and computationally

[26] R. J. Woods, R.A. Dwek, C. J. Edge, B. Fraser-Reid, *J. Phys. Chem. B* **1995**, *99*, 3832-3846.

expensive. However, computers are getting faster and cheaper and algorithms are getting more efficient. Simulations of solvated proteins are calculated up to the nanosecond time scale, however, simulations into the millisecond regime have been reported.[27]

The accurate modeling of molecules in solution using molecular mechanics (MM) requires realistic models for the interaction of the solvent with the molecule. To treat such a medium (usually water) in a molecular calculation, both explicit (including real molecules of solvent)[28] and implicit (solvent as a fixed dielectric constant)[29] models have been developed. Nowadays, the use of an explicit solvent is preferred in most of the simulations.

MD simulations are very coherent in low energy conformations of single molecules but weak in relative population predictions, particularly with AMBER force field that over-estimate α-helix for small glycopeptides.[30] In order to acquire better results in MD, it is convenient to introduce experimental data obtained by NMR trying to obtain a low-energy conformers distribution useful to replicate NMR data. It is called MD-tar (MD simulations with time-averaged restraints).

Structural restrains (usually, H-H distances from NOE data and $^{3}J_{\text{H-H}}$ coupling constants) are included in the force field equation used in MD constraining the system, allowing to jump potential barriers and performing a deep scan of conformational analysis. Therefore, the new potential function will have two terms (Equation 2.3).

---

[27] G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, P. Zhang, *Nature* **2013**, *497*, 643-646.

[28] A. R. Bizzarri, S. Cannistraro, *J. Phys. Chem. B* **2002**, *106,* 6617-6633.

[29] M. Orzco, F. J. Luque, *Chem. Rev.* **2000**, *100*, 4187-4225.

[30] a) R. B. Best, N. V. Buchete, G. Hummer, *Biophys. J. Biophys. Lett.* **2008**, *95*, L07-L09; b) S. Gnanakaran, A. E. García, *J. Phys. Chem. B* **2003**, *107,* 12555-12557.

$$V_{Total} = V_{theoretical} + V_{experimental} \qquad \textbf{Equation 2.3}$$

The first term, $V_{theoretical}$, is the normal potential of force field while $V_{experimental}$ includes the experimental information and contains the restraints (Equation 2.4).

$$V_{rest} = K_{rest} \cdot (R_{ij} - R_{exp})^2 \qquad \textbf{Equation 2.4}$$

V_{rest} has to satisfy a specific distance by means of energy penalty. If we wanted to obtain the average trajectories of the molecule, the equation has to be substituted by:

$$V_{rest} = K_{rest} \cdot (\langle R_{ij} \rangle - R_{exp})^2 \qquad \textbf{Equation 2.5}$$

As we have explained before, NOEs are R^{-6} dependent so the correct expression for averaged distance will be the following:

$$V_{rest} = K_{rest} \cdot (\langle {R_{ij}}^{-6} \rangle^{-1/6} - R_{exp})^2 \qquad \textbf{Equation 2.6}$$

where:

$$(\langle {R_{ij}}^{-6} \rangle^{-\frac{1}{6}} = \frac{(\sum_0^t R(t)^{-6})^{-\frac{1}{6}}}{N} \qquad \textbf{Equation 2.7}$$

N is the total number of frames. In molecular dynamics, a frame is each particular conformation frozen, in a given time, from the total of conformations in the complete dynamic.

Experimentally, equation 2.7 has one handicap: very long time of simulation are necessary to ensure the satisfaction of all applied restrictions. Hence, trying to minimize the impact in the calculations but reducing the simulation times, the real equation used was:

$$(\langle {R_{ij}}^{-6} \rangle^{-\frac{1}{6}} = \frac{\left(\sum_0^t R(t)^{-6} \cdot e^{\frac{t-t'}{\tau}} \right)^{-\frac{1}{6}}}{N} \qquad \textbf{Equation 2.8}$$

with τ being the characteristic time for the exponential decay or exponential decay constant, and *t'* the total simulation time. The exponential decay constant *helps* the calculation to converge and it is commonly set to a value 10 times smaller than t'.

This MD-tar methodology has been extensively and successfully applied to the study of different biomolecular systems.[31]

2.4. Enzyme-linked lectin assay (ELLA)

Lectins can be employed for glycoanalysis in a number of different formats including lectin microarray techniques.[32] However, the most simplistic format in which lectins can be employed for glycoanaylsis is the enzyme-linked lectin assay (ELLA). This assay has the same basic format as a standard ELISA (enzyme-linked immunosorbent assay) and allows the analysis of glycoproteins, and lectin-glycan interactions, in a standard 96-well microtiter plate format. The most common format of the ELLA requires the immobilization of glycoconjugates on the surface of an ELISA plate,

---

[31] P. M. S. Hendrickx, F. Corzana, S. Depraetere, D. A. Tourwé, K. Augustyns, J. C. Martins, *J. Comput. Chem.* **2010**, *31*, 561-572.

[32] R. Rosenfeld, H. Bangio, G. J. Gerwig, R. Rosenberg, R. Aloni, Y. Cohen, Y. Amor, I. Plaschkes, J. P. Kamerling, R. B. Y. Maya, *J. Biochem. Biophys. Methods* **2007**, *70*, 415-426.

subsequent blocking of the plate surface and then probing with biotinylated lectins.[33]

The glycopeptide in fluid phase is immobilized in the microtiter plates. The antigen is allowed to bind to a specific biotynilated lectin, which is itself subsequently detected by a peroxidase. A chromogenic substrate for the enzyme yields a visible color change, indicating the presence of glycopeptide interacting with the lectin. Quantitative or qualitative measures can be assessed based on such colorimetric reading.

We have optimized a new ELLA methodology that allows the assay to be performed with several lectins, and diverse glycopeptides (see *Experimental section*), thereby extending and enhancing its utility and providing a broadly accessible and simple assay for glycoanalysis (Figure 2.12).

[33] M. Duk, E. Lisowska, J. H. Wu, A. M. Wu, *Anal. Biochem.* **1994**, *221*, 266-272.

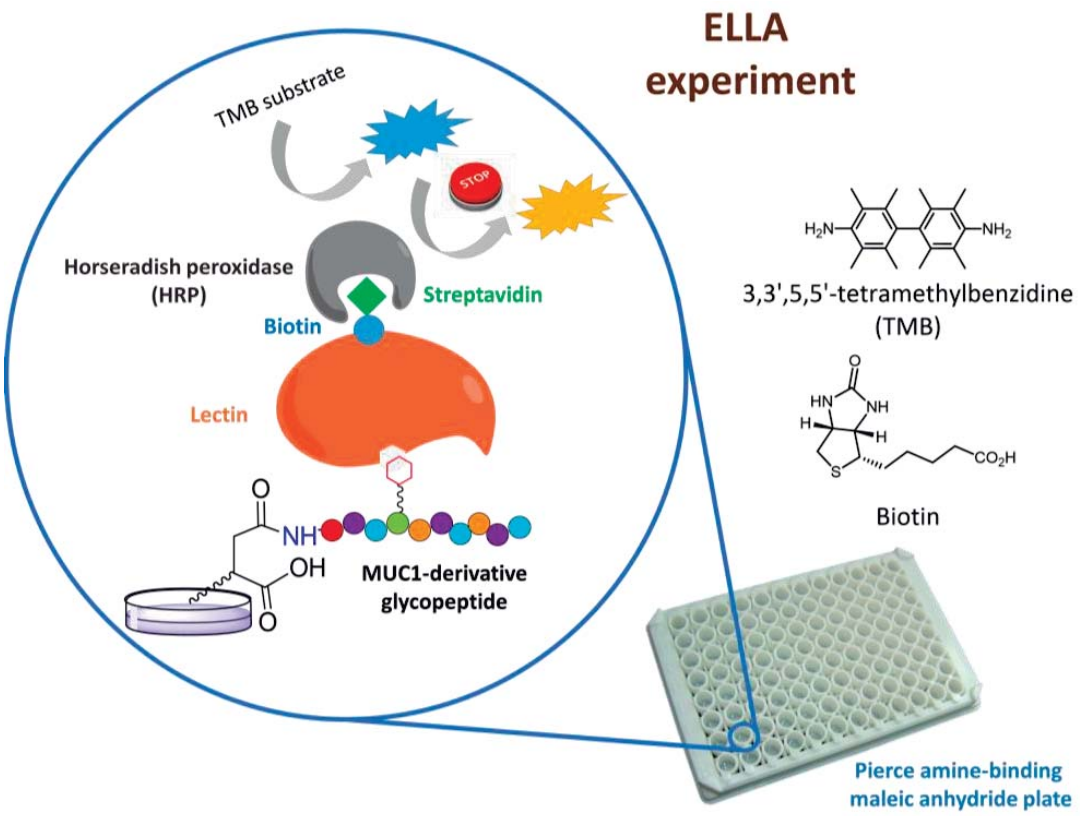


Figure 2.12. *Enzyme-linked lectin assay technique used to detect our MUC1 derivatives.*

2.5. Isothermal Titration Calorimetry (ITC)

Calorimetry is one of the oldest techniques in experimental science, and calorimeters have evolved for several hundred years giving rise to the modern versions.[34] For the past several decades, basically three different types of microcalorimeters, batch, flow, and titration calorimeters, have been used. ITC experiment directly determines the thermodynamics of binding of two molecules in solution at a constant temperature. Typically, this

[34] T. Wiseman, S. Williston, J. F. Brandt, L. N. Lin, *Anal. Biochem.* **1989**, *179*, 131-137.

technique is used for measuring the heat of binding in lectin–carbohydrate interactions in solution. The major advantage of ITC is that it provides simultaneous determination of the thermodynamic binding parameters K_a, ΔG, ΔH, $T\Delta S$ and n in a single experiment.

The range of affinity constants that can be measured is in the mM to nM region. A series of data points of the amount of heat released (exothermic) or absorbed (endothermic) per mole of injectant (ligand) is plotted as a function of the molar ratio (L_T/M_T) of ligand (L_T) and macromolecule (M_T) after each injection to generate the binding isotherm. Thermodynamic binding parameters are determined by nonlinear least-squares analysis of the binding isotherm. ΔH, the change in enthalpy (kcal/mol) on binding; K_a, the association constant (M^{-1}); and n, the number of binding sites per monomer of receptor are the adjustable parameters in the fits. From the equation[35]

$$\Delta G = -RTlnK_a \qquad \textbf{Equation 2.9}$$

G, the free energy of binding (kcal/mol), can be calculated. And from the equation

$$\Delta G = \Delta H - T\Delta S \qquad \textbf{Equation 2.10}$$

$T\Delta S$, the entropy of binding (kcal/mol), can be determined. Determination of the temperature dependence of the enthalpy and entropy changes allows evaluation of the changes in heat capacity (ΔC_p):

$$\Delta H(T_1) = \Delta H(T_0) + \Delta C_p(T_1 - T_0) \qquad \textbf{Equation 2.11}$$

$$\Delta S(T_1) = \Delta S(T_0) + \Delta C_p(T_1 - T_0)/T_0 \qquad \textbf{Equation 2.12}$$

[35] T. K. Dam, C. F. Brewer, *Chem. Rev.* **2002**, *102*, 387–429.

where T_1 and T_0 represent different temperatures (over a narrow range). The number of binding site(s), n, per molecule of the protein is described as follows

$$q = nV\Delta H[ML_n] \qquad \textbf{Equation 2.13}$$

where q is the heat absorbed or evolved, n is the number of binding site(s) per molecule of the receptor, V is the cell volume, ΔH is the binding enthalpy per mole of ligand, and *[ML]* is the concentration of the bound ligand.

Hence, knowledge of the K_a, ΔH, and TΔS values for a carbohydrate or glycopeptide can provide insight into the physical nature of their binding interactions with a lectin and, in certain cases, the size of the binding site of the protein.

DIFFERENCES BETWEEN SERINE AND THREONINE IN Tn ANTIGEN

3

3.1. Introduction

As we have described in *Chapter 1*, Tn antigen (α-*O*-GalNAc-Ser/Thr) is one of the most specific human tumor-associated structures.[1] This motif is implicated in HIV infection[2] and it is early expressed in tumor cells. It has been observed that there is a direct correlation between carcinoma aggressiveness and the density of this antigen. For this reason, Tn determinant is a convenient cancer biomarker.[3] In general, Tn antigen is referred to *N*-acetylgalactosamine (GalNAc) α-*O*-linked to serine (Ser) or threonine (Thr), not discriminating the amino acid to which GalNAc is linked (Figure 3.1).

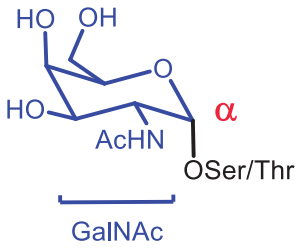


Figure 3.1. *Tn antigen structure.*

Regarding molecular recognition of Tn antigen by biological targets, differences between Ser and Thr have been also described. For instance, in a recent analysis of anti-Tn monoclonal antibodies that bind to synthetic glycopeptides, subtle differences in affinity by these antibodies to glycopeptides carrying Ser or Thr have been reported. The authors stated that

---

[1] a) T. Ju, V. I. Otto, R. D. Cummings, *Angew. Chem. Int. Ed.* **2011**, *50*, 1770-1791; b) G. J. Boons, T. Buskas, P. Thompson, *Chem. Commun.* **2009**, 5335-5349; c) R. M. Wilson, S. J. Danishefsky, *J. Am. Chem. Soc.* **2013**, *135*, 14462-14472; d) R. E. Beatson, J. Taylor-Papadimitriou, J. M. Burchell, *Immunotherapy* **2010**, *2*, 305-327.

[2] J. E. Hansen, C. Nielsen, M. Arendrup, S. Olofsson, L. Mathiesen, J. O. Nielsen, H. Clausen, *J. Virol.* **1991**, *65*, 6461-6467.

[3] K. Terasawa, H. Furumoto, M. Kamada, T. Aono, *Cancer Res.* **1996**, *56*, 2229-2232.

the aglyconic part of the Tn antigen plays a key role in anti-Tn antibodies specificity for breast and colon cancer detection, which may be of important clinical value.[4] Another example is related to the Kanzaki disease, which is attributable to a deficiency in α-*N*-acetylgalactosaminidase (α-NAGA). This enzyme hydrolyzes the glycosidic linkage with an α-GalNAc unit.[5] The hydrolysis rate is quite different for α-*O*-GalNAc-Ser and α-*O*-GalNAc-Thr linkages. More recently, in the context of microtubule-associated protein tau, related with Alzheimer's disease, it has been observed that structural changes caused by phosphorylation of Thr residues are greater than those caused by this modification in Ser.[6]

This observation may have important biological effects. In fact, in the context of antifreeze glycoproteins, it is crucial the presence of Thr residue (instead of Ser amino acid) to show activity (Figure 3.2).[7]

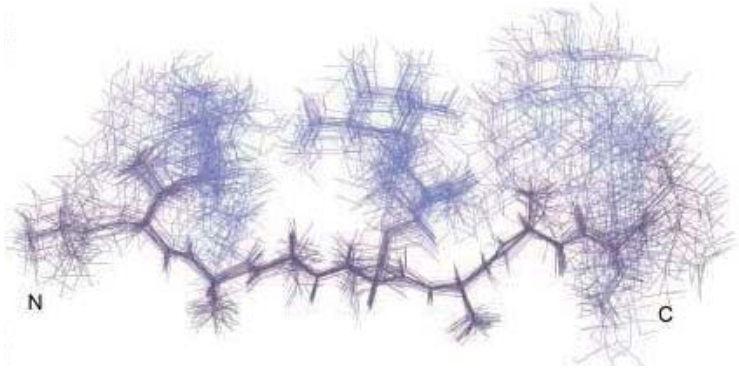


Figure 3.2. *Synthetic antifreeze glycoprotein Ala-Thr*-Ala-Ala-Thr*-Ala-Ala-Thr*-Ala.*

[4] D. Mazal, R. Lo-Man, S. Bay, O. Pritsch, E. Dériaud, C. Ganneau, A. Medeiros, L. Ubillos, G. Obal, N. Berois, M. Bollati-Fogolin, C. Leclerc, E. Osinaga, *Cancer Immunol., Immunother.* **2013**, *62*, 1107-1122.

[5] T. Kanzaki, M. Yokota, N. Mizuno, Y. Matsumoto, Y. Hirabayashi, *Lancet* **1989**, *333*, 875-877.

[6] M. A. Brister, A. K. Pandey, A. A. Bielska, N. J. Zondlo, *J. Am. Chem. Soc.* **2014**, *136*, 3803-3816.

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

Moreover, in our group, we have reported a different conformational behavior of the two corresponding glycosidic linkages α-*O*-GalNAc-Ser and α-*O*-GalNAc-Thr.[8] More precisely, the β-methyl group of Thr residue forces the carbohydrate moiety to adopt an almost perpendicular orientation with respect to the peptide backbone (Figure 3.3).

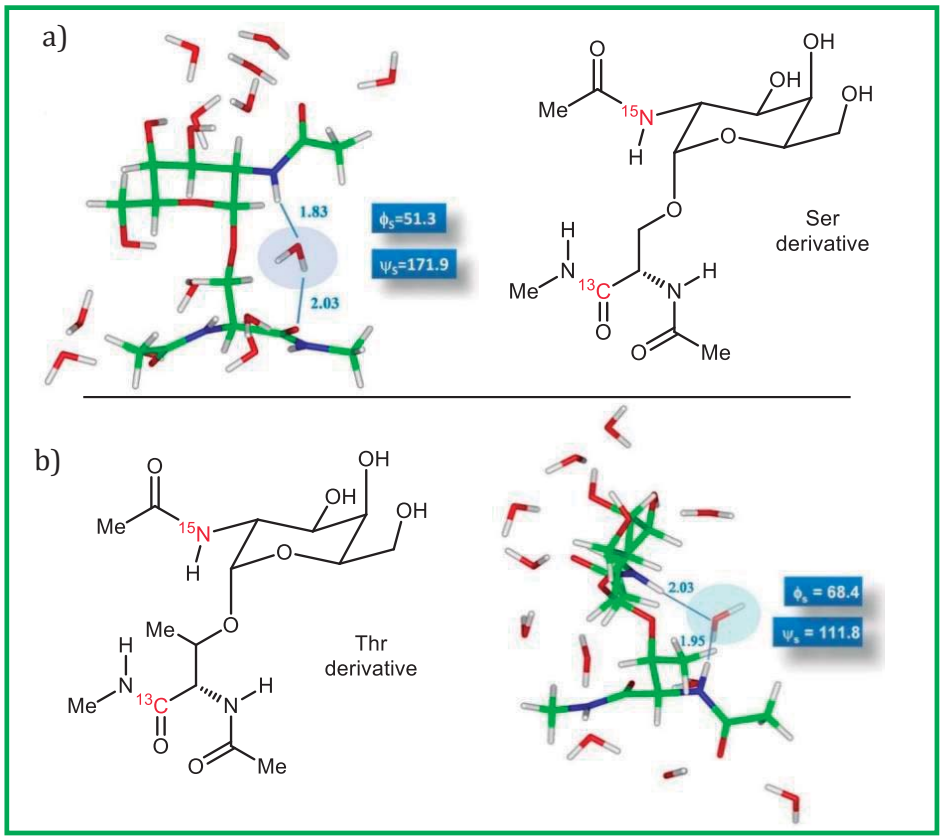


Figure 3.3. *Calculated B3LYP/6-31G(d) geometry of compound a) α-O-GalNAc-L-Ser diamide b) α-O-GalNAc-L-Thr both isotopical labeled with ^{15}N (10%) and ^{13}C (100%) including the surrounding water molecules.*

[8] a) F. Corzana, J. H. Busto, G. Jiménez-Osés, M. García de Luis, J. L. Asensio, J. Jiménez-Barbero, J. M. Peregrina, A. Avenoza, *J. Am. Chem. Soc.* **2007**, *129*, 9458-9467; b) F. Corzana, J. H. Busto, M. García de Luis, J. Jiménez-Barbero, A. Avenoza, J. M. Peregrina, *Chem. Eur. J.* **2009**, *15*, 3863-3874; c) F. Corzana, J. H. Busto, G. Jiménez-Osés, J. L. Asensio, J. Jiménez-Barbero, J. M. Peregrina, A. Avenoza, *J. Am. Chem. Soc.* **2006**, *128*, 14640-14648.

In contrast, in Ser derivatives, the sugar displays a parallel arrangement with respect to the underlying peptide. The different 3D disposition enforces that Ser and Thr derivatives present a completely different first hydration shell (Figure 3.3).

One possible explanation was the orientation of Ψ_s bond. In serine derivative this angle was more flexible and shows values are around 180° so the alternated conformation for H_β-C_β and O_{1s}-C_{1s} was the preferred. By contrast, threonine derivative exhibited values around 120°, so the eclipsed conformation is dominant as it is described in Figure 3.4 with the Newman projection of O_{1s}-C_β bond.

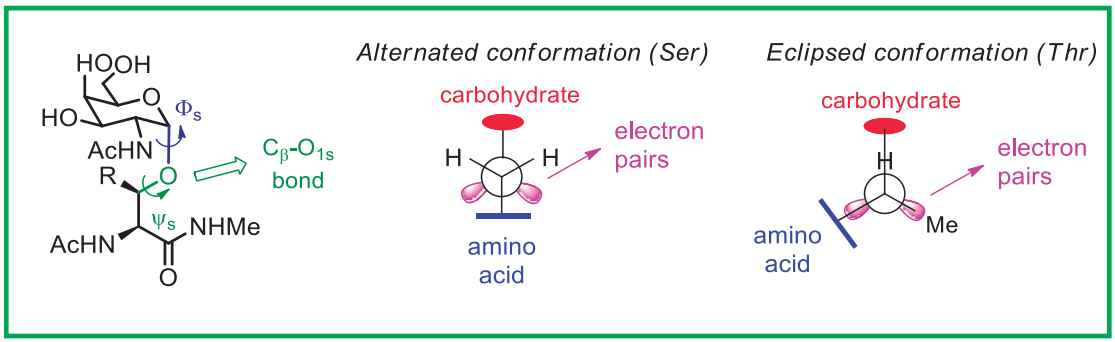


Figure 3.4. *Schematic representation of Ψ_s bond and its Newman projection in the alternated and eclipsed conformations.*

Besides, the synthesis of mimics of natural carbohydrates and glycopeptides helps us to complete our understanding of different biochemical processes and offers a battery of new candidates for biological targets. In previous work of our group,[9] Tn-mimic (α-GalNAc-*S*-linked to Cys) has been synthesized and its conformational analysis has proved that properties and the first hydration

[9] C. Aydillo, I. Compañón, A. Avenoza, J. H. Busto, F. Corzana, J. M. Peregrina, M. M. Zurbano, *J. Am. Chem. Soc.* **2014**, *136*, 789-800.

shell of this compound was similar and comparable to the natural *O*-glycosyl Thr analogue (Figure 3.5).

In addition, other mimics have been achieved and analyzed (Figure 3.6). [10]

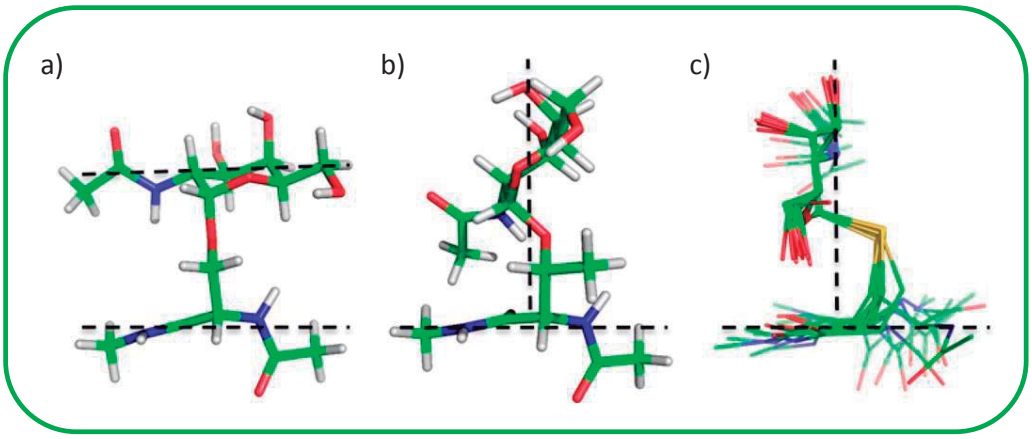


Figure 3.5. *Representative conformations for a) Tn-Ser b) Tn-Thr c) Tn-Cys antigens.*

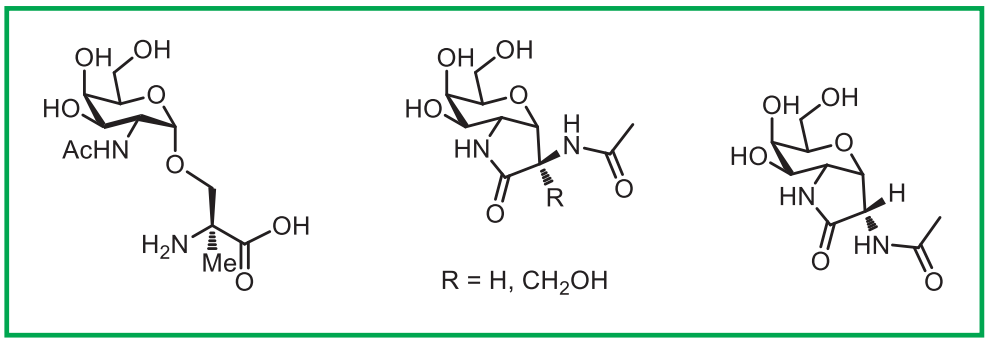
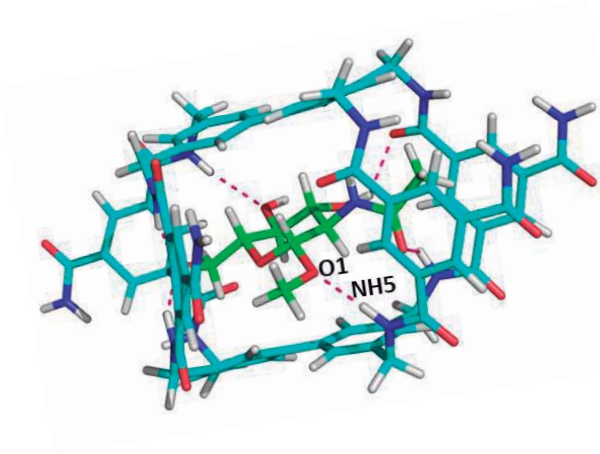


Figure 3.6. *Some Tn derivatives synthesized in our group.*

[10] a) C. Aydillo, C. D. Navo, J. H. Busto, F. Corzana, M. M. Zurbano, A. Avenoza, J. M. Peregrina, *J. Org. Chem*. **2013**, *78*, 10968-10977; b) F. Corzana, J. H. Busto, F. Marcelo, M. García de Luis, J. L. Asensio, S. Martín-Santamaría, J. Jiménez-Barbero, A. Avenoza, J. M. Peregrina, *Chem. Eur. J.* **2011**, *17*, 3105-3110; c) F. Corzana, J. H. Busto, F. Marcelo, M. García De Luis, J. L. Asensio, S. Martín-Santamaría, Y. Sáenz, C. Torres, J. Jiménez-Barbero, A. Avenoza, J. M. Peregrina, *Chem. Commun.* **2011**, *47*, 5319-5321.

On the other hand, lectins have demonstrated to be versatile tools for Tn antigen detection[11] because, in general, they bind to glycopeptides bearing mono-Tn motifs. In this context, the carbohydrate moiety is a key player in the recognition process.[12] However, it is important to note that the peptide fragment around the glycan can also have an influence on the complex stabilization, either by participating in direct or water-mediated interactions with the receptor or by adopting specific conformations. In fact, we have previously reported that the modification of the underlying amino acid of β-*O*-GlcNAc considerably affects to the molecular recognition by synthetic lectins of this carbohydrate. (Figure 3.7).[13]



Name	K_a (M^{-1})	ΔG (Kcal/mol)
βGlcNAcOMe	630	3.82
GlcNAc	156	2.99
αGlcNAcOMe	24	1.88
βGlcNAcSer	91	2.67
βGlcNAcThr	13	1.52
βGlcNAcMeSer	75	2.56
βGlcNAcMeThr	46	2.27

Figure 3.7. *Schematic representation of the different hydrogen bonds obtained from the MD-tar simulations in the receptor–βGlcNAcOMe complex (left) and association constants (K_a) and ΔG values in aqueous solution for the receptor with different substrates (right).*

[11] D. Clark, L. Mao, *Dis. Markers* **2012**, *33*, 1-10.
[12] T. K. Dam, C. F. Brewer, *Chem. Rev.* **2002**, *102*, 387-430.
[13] F. Corzana, A. Fernández-Tejada, J. H. Busto, G. Joshi, A. P. Davis, J. Jiménez-Barbero, A. Avenoza, J. M. Peregrina, *ChemBioChem* **2011**, *12*, 110-117.

3.2. Objectives

Herein, our aim was to study the influence of the underlying amino acid (serine or threonine) in the molecular recognition of Tn antigen (with serine or threonine) by different GalNAc specific lectins. With this purpose, we have selected two glycopeptides derived from the human mucin MUC1, which are characterized by presenting α-GalNAc linked to a serine or a threonine residue. We have synthesized two derivatives from MUC1 tandem repeat (Figure 3.8): Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Asp-Thr(α-*O*-GalNAc)-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala-Pro-Pro-Ala (glycopeptide **1**) and Ala-His-Gly-Val-Thr-Ser-Ala-Pro-Asp-Ser(α-*O*-GalNAc)-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala Pro-Pro-Ala (glycopeptide **2**). Both compounds were synthesized by solid-phase peptide synthesis (SPPS).

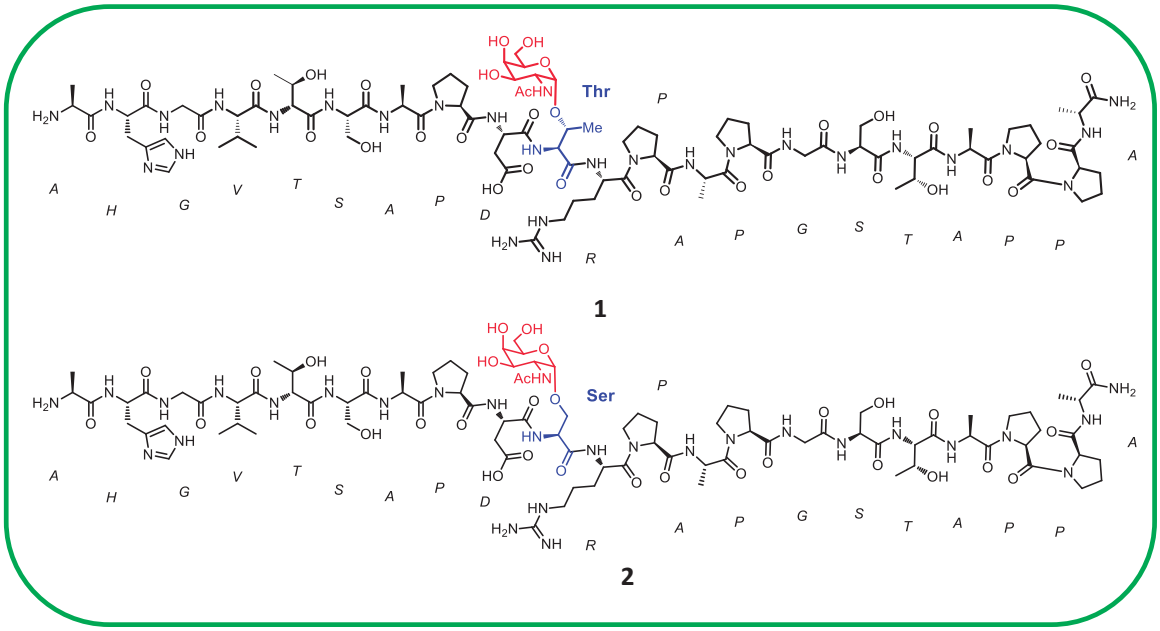


Figure 3.8. *MUC1 derivatives **1** and **2** synthesized by SPPS.*

The compounds were evaluated with three different lectins that recognize Tn antigen (soybean agglutinin (SBA), *Vicia villosa* agglutinin (VVA) and *Helix*

pomatia agglutinin (HPA). All of them are stable and readily-available lectins. In addition, the crystal structures of HPA and VVA complexed to Tn antigen (α-*O*-GalNAc-Ser) as well as the structure of SBA bound to dissacharide Galβ(1-4)GalNAc have already been reported (pdb ids: 2CGZ[14], 1N47[15] and 1SBF,[16] respectively) (Figure 3.9).

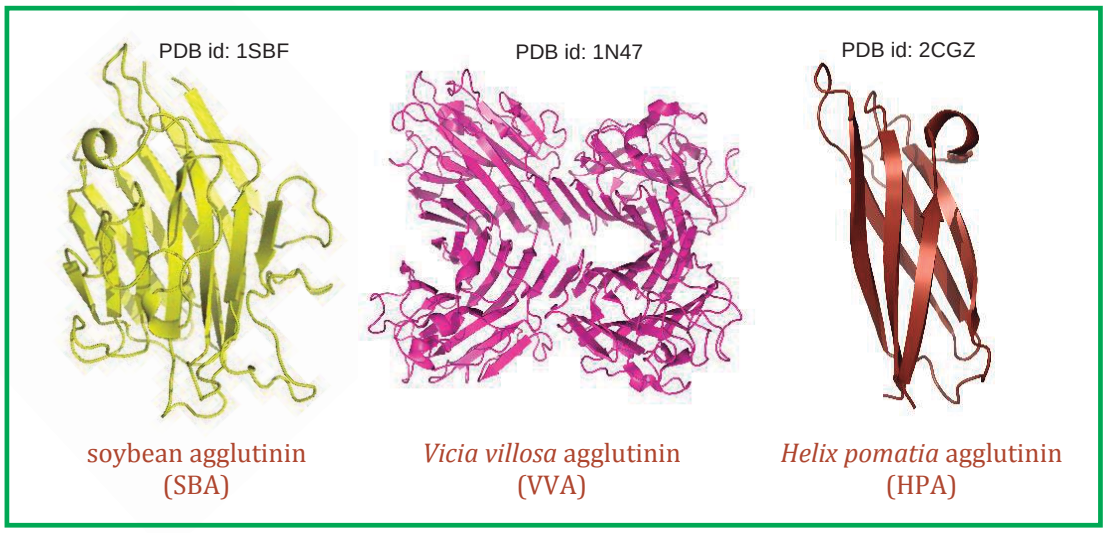


Figure 3.9. *X-ray structures of SBA, VVA and HPA lectins which recognize and bind to α-O-GalNAc.*

In order to analyze the interactions between these three selected lectins and α-*O*-GalNAc glycopeptides **1** and **2**, we used ELLA and ITC experiments, (STD) NMR experiments and MD simulations.

[14] R. Lotan, H. W. Siegelman, H. Lis, N. Sharon, *J. Biol. Chem.* **1974**, *249*, 1219-1224.
[15] S. E. Tollefsen, R. Kornfeld, *J. Biol. Chem.* **1983**, *258*, 5172-5176.
[16] J. F. Sanchez, J. Lescar, V. Chazalet, A. Audfray, J. Gagnon, R. Alvarez, C. Breton, A. Imberty, E. P. Mitchell, *J. Biol. Chem.* **2006**, *281*, 20171-20180.

3.3. Results and discussion

3.3.1. Synthesis

The synthesis of the glycopeptides was performed using the solid phase peptide synthesis (SPPS) protocol with a Rink amide MBHA resin and Fmoc and side chain protected amino acids (see *Chapter 5*).

The protected glycosyl amino acid was manually coupled, following the general procedure described in *Chapter 2* and *Experimental* section.

The first step of the synthetic pathway was the synthesis of both glycosyl amino acid building blocks, with serine and threonine (Figure 3.10) and then they were incorporated into the peptide chain of the MUC1 derivative.

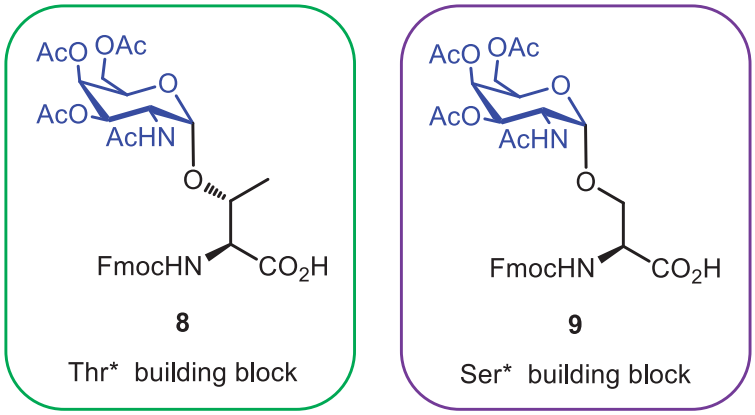
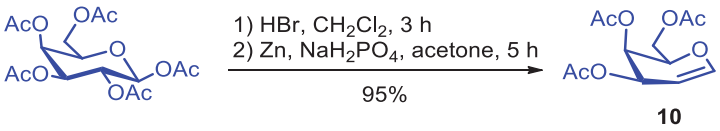


Figure 3.10. *Both building blocks of threonine and serine glycosylated with protected α-O-GalNAc.*

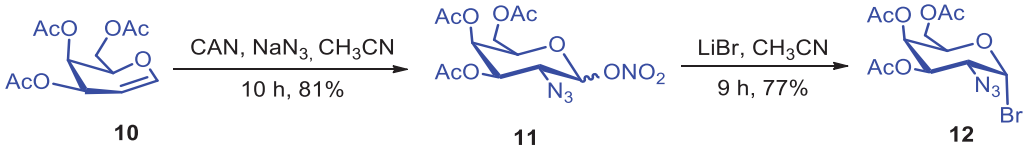
We used two different pathways to obtain the carbohydrate moiety ready to be glycosylated. Both of them required tri-*O*-acetyl-D-galactal **10** as the starting material. For that reason, we started from commercial β-D-galactose pentaacetate, which was treated with hydrobromic acid in dichloromethane,

followed by a second step of reduction with zinc and sodium phosphate monobasic in acetone. This methodology is a modification of the previous reported shynthesis.[17] We obtained the starting material for both pathways in a high yield (Scheme 3.1).



Scheme 3.1. *Synthesis of tri-O-acetyl-D-galactal (**10**).*

a) Azido-bromo derivative:[18] A solution of product **10** in acetonitrile was treated (CH_3CN) with ammonium cerium (IV) nitrate (CAN) and sodium azide (NaN_3) and was stirred under argon atmosphere (Scheme 3.2) for 10 h. The solvent was evaporated and the crude was purified by a silica gel column chromatography. The corresponding stereoisomeric mixture (**11**) was treated with lithium bromide (LiBr) in CH_3CN, to obtain the required product **12**.



Scheme 3.2. *Synthesis of azido-bromo derivative (**12**).*

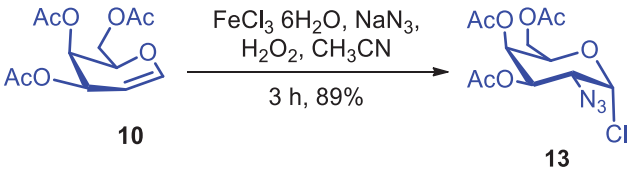
b) Azido-chloro derivative:[19] A solution of product **10** in acetonitrile (CH_3CN) was stirred with iron (III) chloride hexahydrate ($FeCl_3 \cdot 6H_2O$), NaN_3 and hydrogen peroxide (H_2O_2) (Scheme 3.3) for 3 h. The solvent was

[17] P. Levecque, D. W. Gammon, H. H. Kinfe, P. Jacobs, D. De Vos, B. Sels, *Adv. Synth. Catal.* **2008**, *350*, 1557-1568.

[18] a) R. U. Lemieux, R. M. Ratcliffe, C*an. J. Chem*. **1979**, *57*, 1244-1251; b) C. Heggemann, C. Budke, B. Schomburg, Z. Majer, M. Wißbrock, T. Koop, N. Sewald, *Amino Acids* **2010**, *38*, 213-222.

[19] C. Plattner, M. Höfener, N. Sewald, *Org. Lett.* **2011**, *13*, 545-547.

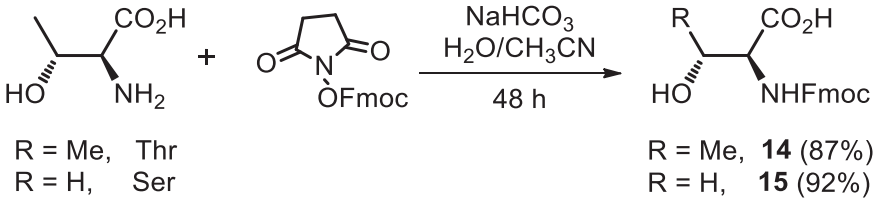
removed and the crude was purified by silica gel chromatographic column obtaining the product **13** in a high yield.



Scheme 3.3. *Synthesis of azido-chloro derivative (**13**).*

Concerning the amino acid moiety, the amino group was conveniently protected with Fmoc and the carboxylic acid as *tert*-butyl ester. We chose this protection thinking in the use of the Fmoc strategy for the solid phase peptide synthesis (SPPS).

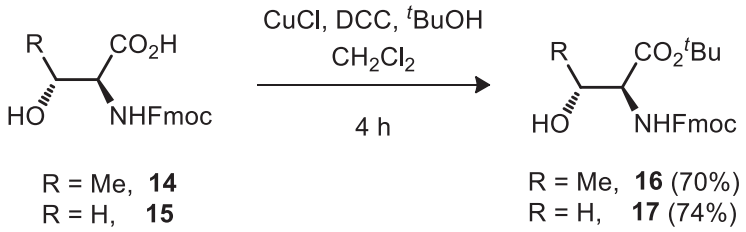
We started with the commercial available L-Thr or L-Ser and we protected the N-terminus with the Fmoc *N*-hydroxysuccinimide ester, as a reagent, in basic medium obtaining the products **14** and **15** (Scheme 3.4).



Scheme 3.4. *Synthesis of amino-protected L-Thr and L-Ser.*

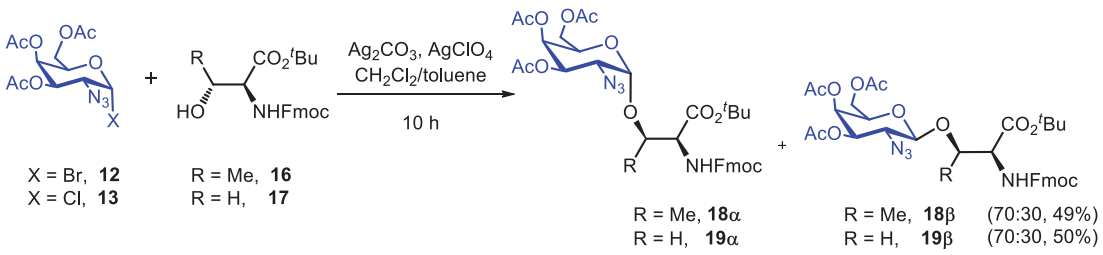
Then, the esters formation of **14** and **15** were prepared using a mixture of CuCl, DCC and ^{t}BuOH and stirring the reaction in methylene chloride under

argon atmosphere.[20] We obtained the protected amino acids **16** and **17**, respectively, in high yields (Scheme 3.5).



Scheme 3.5. *Synthesis of N- and C- protected L-Ser and L-Thr.*

At this point, we were able to construct the glycosidic bond between the conveniently protected amino acid and the carbohydrate moiety following the Koegnis-Knorr reaction.[21] There are two similar pathways depending on the haloderivative used: bromo-azide (**12**) or chloro-azide (**13**). The amino acid **16** (or **17**) was dissolved in a mixture of CH_2Cl_2/toluene and silver salts were added. Next, carbohydrate **12** (or **13**) was also dissolved in toluene and then added to the mixture. We obtained in both cases an α:β ratio (70:30). We purified the mixture to obtain the α-derivative by a silica gel column chromatography (Scheme 3.6).



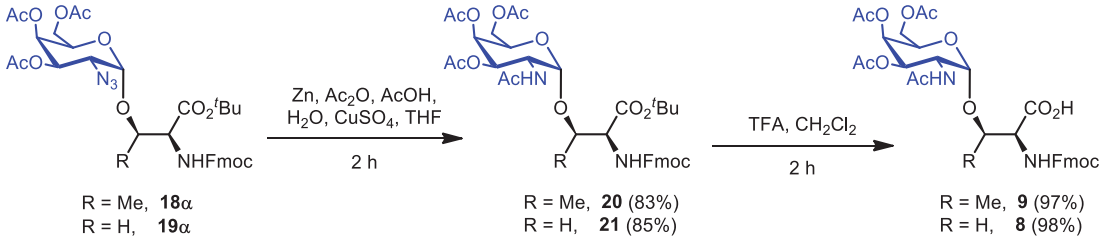
[20] M. Schultz, H. Kunz, *Tetrahedron Asymmetr.* **1993**, *4*, 1205-1220.

[21] a) W. Koenigs, E. Knorr, *Chem. Ber.* **1901**, *34*, 957-981 b) B. Liebe, H. Kunz, *Angew. Chem. Int. Ed. Engl*. **1997**, *36*, 618-621.

Scheme 3.6. *Koegnis-Knorr reaction. Formation of glycosidic bond.*

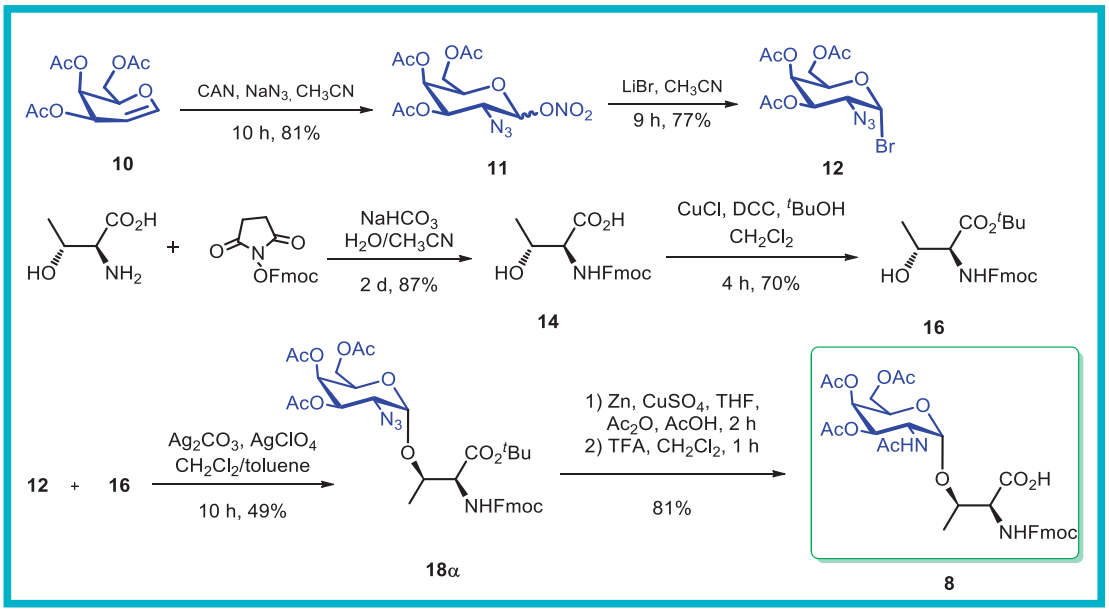
With the pure α-products **18α** and **19α** in our hands, we transformed the azide group into an acetamide group with Zn in a mixture of THF/Ac_2O/HOAc, obtaining compounds **20** and **21** (Scheme 3.7). As a final step in the synthetic route of the building blocks, we deprotected the *tert*-butyl ester from the acid using a mixture of TFA/CH_2Cl_2, to give the corresponding compounds **8** and **9** (Scheme 3.7).

In this way, we obtained the corresponding protected glycosylamino acids **8** and **9** (Scheme 3.7) ready to synthesize the glycopeptides that incorporate these building blocks.

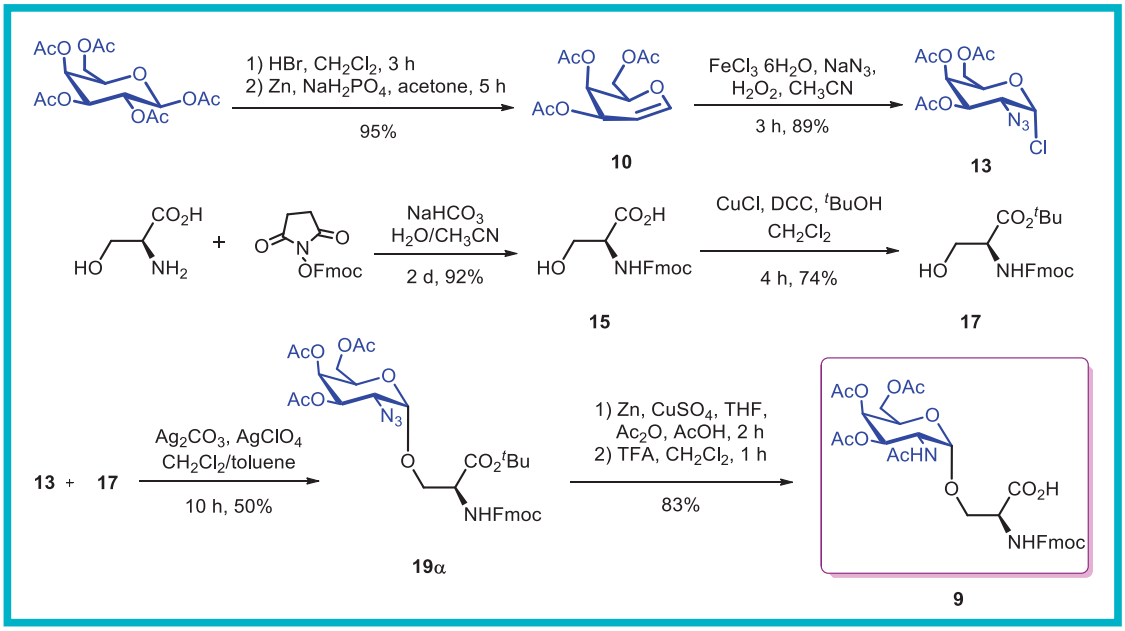


Scheme 3.7. *Synthesis of building blocks **8** (α-O-GalNAc-Thr) and **9** (α-O-GalNAc-Ser)*

Therefore, we have described two similar strategies for obtaining our building blocks (**8**) and (**9**): via azido-bromo derivative (Scheme 3.8) and via azido-chloro derivative (Scheme 3.9). Both ways are appropriate and the yields are almost the same, but it is important to note that in the case of the azido-chloro derivative we obtained compound **13** in just one easy step from derivative **10** (no inert atmosphere is needed, just a flash chromatography column). On the other hand, in the azido-bromo derivative we obtained compound **12** afterwards two reactions under argon conditions and both steps needed a silica gel column chromatography.



Scheme 3.8. *Synthetic pathway to obtain glycosylamino acid **8** (threonine derivative through azido-bromo derivative).*



Scheme 3.9. *Synthetic pathway to obtain glycosylamino acid* ***9*** *(serine derivative through azido-chloro derivative).*

The synthesis of glycopeptides **1** and **2** (Figure 3.8) was carried out following the SPPS (solid phase peptide synthesis) strategy. It is important to note that the coupling of glycosylamino acids **8** and **9** was manually performed to increase the yield and decrease the number of equivalents (see *Chapter 5*). Once synthesized, glycopeptides were purified by preparative HPLC and subsequent lyophilization gave the target compounds in a 70% overall yield.

3.3.2. Biological assays (ELLA and ITC)

With both glycopeptides **1** and **2** in our hands, we performed several affinity studies with SBA, VVA and HPA lectins. As we mentioned before, the X-ray structure of all of them was available and specifically bound to GalNAc moiety. We wanted to study the interaction between those lectins and our glycopeptides using an enzyme-linked lectin assay (ELLA). Different concentrations of glycopeptides **1** and **2** (varying from 0 to 1.5 mM/well) were covalently attached to a maleic anhydride activated surface. Under these conditions, Tn antigen is far from the plastic surface, avoiding the accessibility issues by the lectin. Following the protocol described in *Chapter 5*, biotinylated SBA, VVA and HPA lectins were added into the wells. After washing repeatedly with buffer TWEEN 20, the streptavidin horseradish peroxidase (HRP) conjugate was added. The addition of 3,3′,5,5′-tetramethylbenzidine (TMB) was used as detection system by measuring the absorbance of the wells at 450 nm (Figure 3.11).

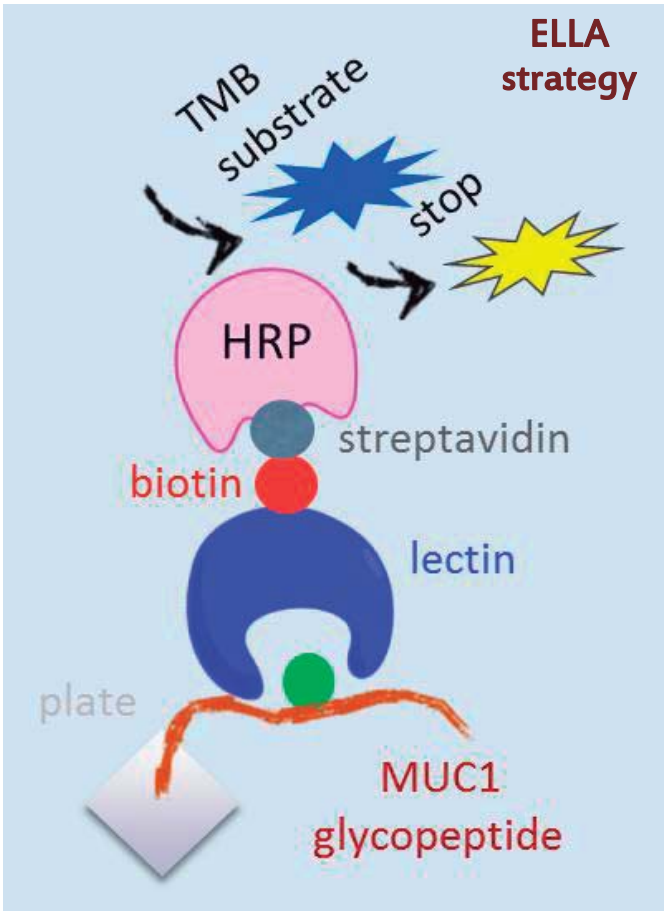


Figure 3.11. *Schematic representation of an ELLA experiment.*

The obtained results are shown in Figure 3.12, where we can see the intensity obtained at different concentrations of glycopeptides **1** and **2** and its recognition by soybean, *Vicia villosa* and *Helix pomatia agglutinin* lectins. It seems to indicate that while SBA and VVA lectins showed a clear preference for the glycopeptide containing Tn-Thr antigen (glycopeptide **1**, pink in Figure 3.12), HPA lectin showed better affinity for glycopeptide with Tn-Ser (glycopeptide **2**, yellow in Figure 3.12). Therefore, these experiments stress the role that the underlying amino acid plays in the molecular recognition process by the lectin.

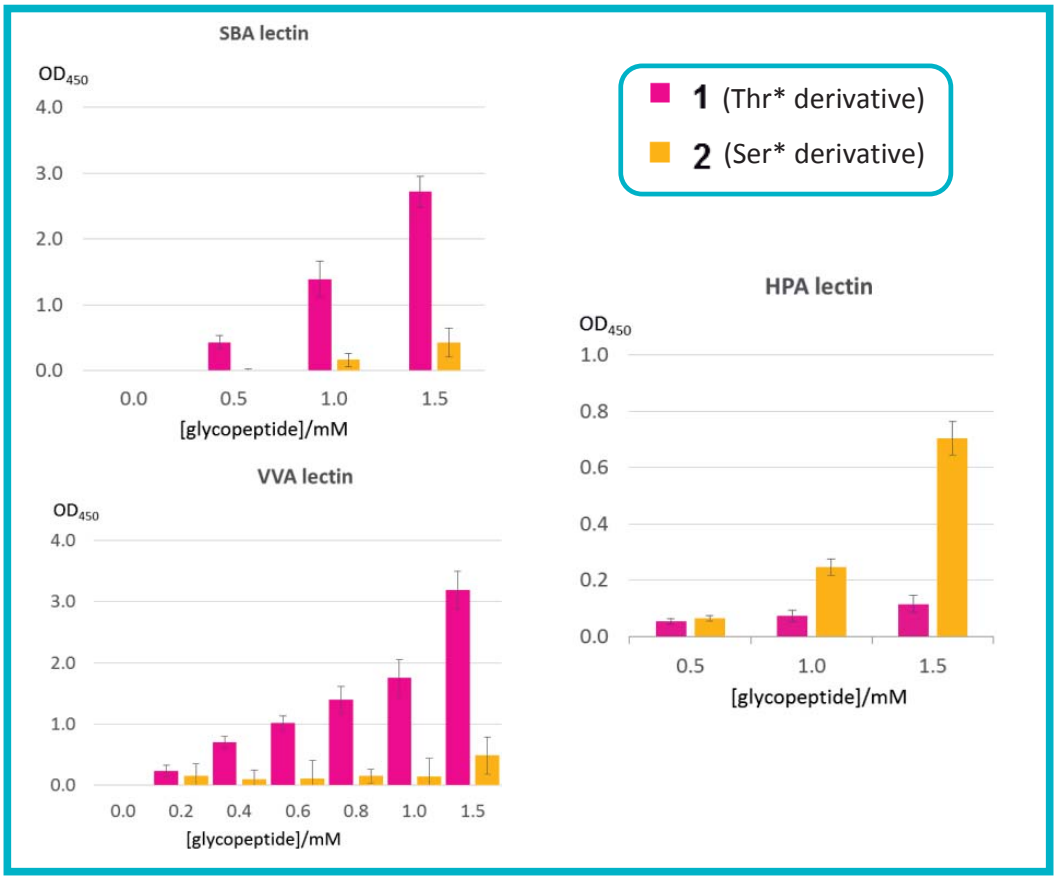


Figure 3.12. *Graphic representation of ELLA experiments of glycopeptides **1** (pink) and **2** (yellow) with three lectins (SBA, VVA, HPA).*

These ELLA tests were subsequently corroborated by isothermal titration calorimetry (ITC) (Figure 3.13 and Tables 3.1 and 3.2).

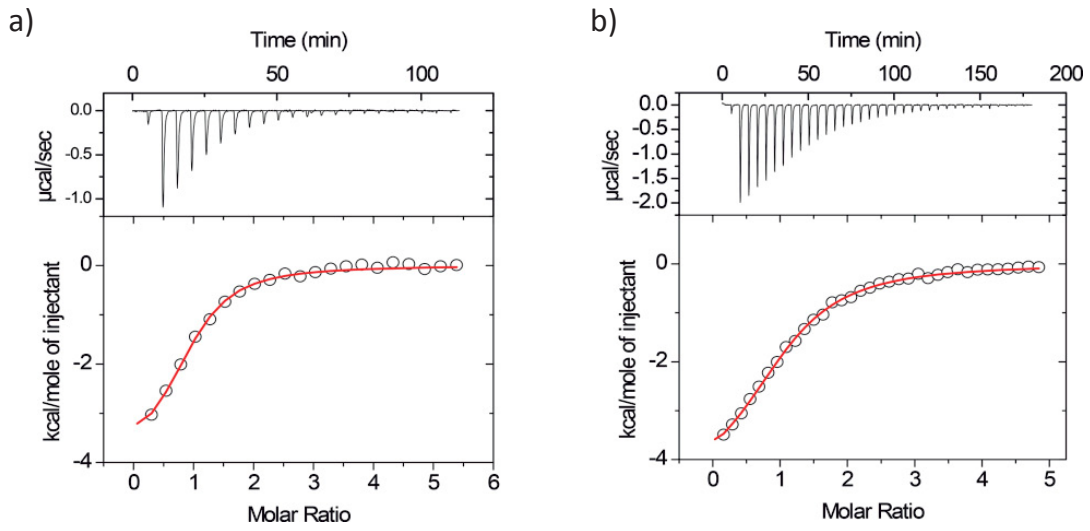


Figure 3.13. *ITC profiles of glycopeptides **1** (a) and **2** (b) titration into 40 µM SBA lectin solutions at 25 °C and pH = 7.5. The solid line represents the least squares fitting of the data to the simplest model (one binding site).*

Table 3.1. *Thermodynamic binding parameters for SBA lectin to glycopeptides **1** and **2** at 25 °C (pH = 7.5).*

	K_D[a] (mM)	ΔG[c] (kcal/mol)	ΔH[b] (kcal/mol)	$T\Delta S$[a] (kcal/mol)	n[b]
1	9.3	- 6.86	- 3.99	2.87	0.89
2	39	- 6.02	- 4.74	1.28	1.06
	[a] Errors in range 1-7% [b] Errors <2%. [c] Errors in range 1-4%.				

Table 3.2. *Thermodynamic binding parameters for HPA lectin to glycopeptides **1** and **2** at 25 °C (pH = 7.5).*

	K_D[a] (mM)	ΔG[c] (kcal/mol)	ΔH[b] (kcal/mol)	$T\Delta S$[a] (kcal/mol)	n[b]
1	~1000.0	- 4.09	-	-	1.00
2	150.2	- 5.22	- 1.79	3.42	1.00
	[a] Errors in range 1-7% [b] Errors <2%. [c] Errors in range 1-4%.				

For SBA lectin (Table 3.1), ITC experiment yielded dissociation constants (K_D) for glycopeptide bearing Tn-Ser antigen (compound **2**) more than 4-fold higher than for the glycopeptide bearing a glycosylated threonine residue (derivative **1**). K_D is the inverted affinity constant (K_a) and describes the affinity between a ligand and a protein, in our case, between the lectin and the glycopeptides.[22]

The results of the ITC were in line with the ELLA experiments (Figure 3.12) and they suggested that the binding with SBA lectin depends mainly on the underlying amino acid (serine or threonine). As it is shown in Table 3.1, the entropic penalty was higher for glycopeptide **2** than for compound **1**. As in the ELLA tests, the scenario is different when the binding with HPA lectin was analyzed (Table 3.2). In fact, this lectin shows an affinity at least 7-fold higher for glycopeptide **2** in comparison to **1**. To rationalize these experimental results, a conformational analysis on the glycopeptides in the free state and bound to SBA, VVA and HPA lectins was performed.

[22] T. K. Dam, C. F. Brewer, *Methods Enzymol.* **2004**, *379*, 107-128.

3.3.3. Free state conformational analysis (NMR and MD)

To a first approximation, we studied the conformational analysis of MUC1 derivatives **1** and **2** combining the information obtained from the NMR experiments with MD simulations.

We obtained the ^{1}H NMR (600 MHz) spectra of both compounds **1** and **2** (Figure 3.14). We tried to assign the signals, but for glycoderivative **2** the signal overlapping did not permit the faithfully assignment of all the signals. In the case of Thr-Tn derivative **1** we also registered the proton spectra in H_2O/D_2O (9:1) to obtain information about the NH signals.

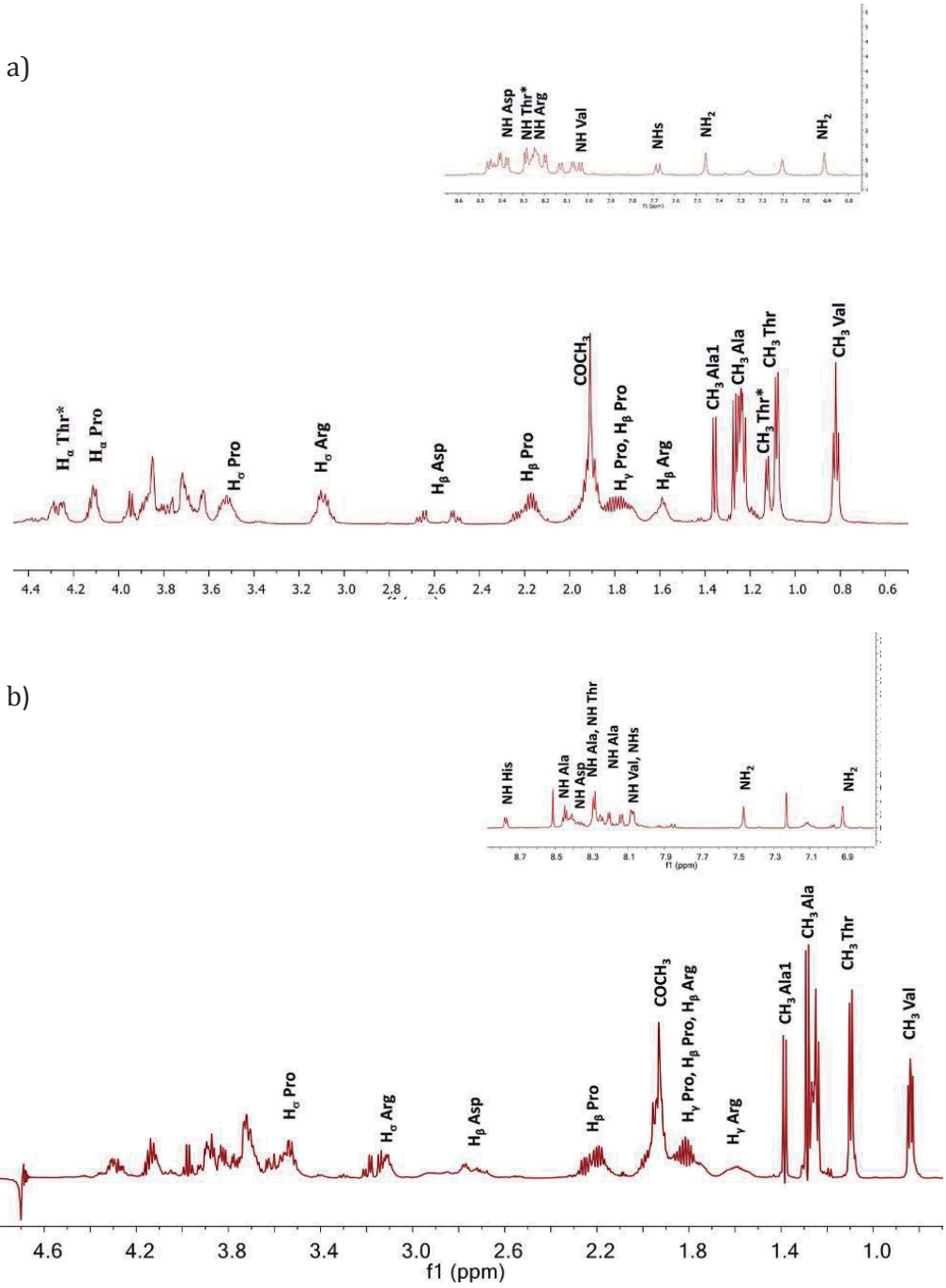


Figure 3.14. *^{1}H NMR (600 MHz) spectra of compound **1** (a) and **2** (b) at 20 ˚C and pH = 5.2 in H_2O/D_2O (9:1).*

In order to get relevant distances between protons in our MUC1-derivatives, we performed 2D-NOESY experiments (Figure 3.15) for which we obtained the more relevant interproton distances for the conformational analysis of glycopeptides **1** (Table 3.3). If we take a look of the experimental distance NH_{Thr}-NH_{GalNAc} we observe that the value (2.4 Å) indicates a NOE between them, therefore, an eclipsed conformation of the molecule is preferred (Figure 3.4).[8a]

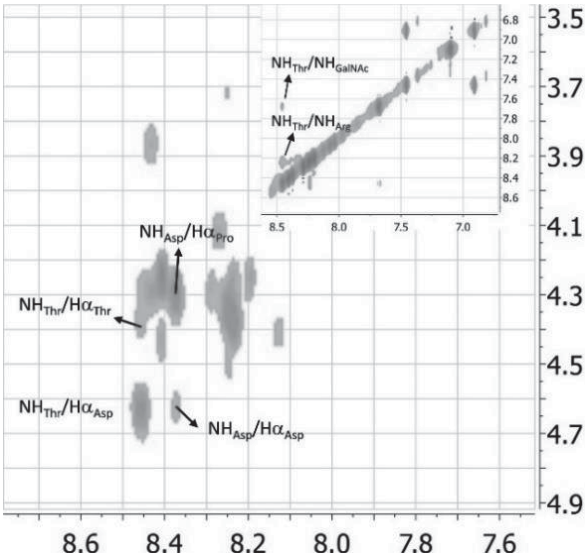


Figure 3.15. *2D-NOESY experiment (600 MHz) of glycopeptide **1** at 20 ˚C and pH = 5.2 in a H_2O/D_2O (9:1) mixture.*

Table 3.3. *Experimental and theoretical H-H distances for derivative **1**.*

	Exp. (Å)	MD_{H2O}-tar (Å)
NH_{Thr}-$H_{\alpha\ Asp}$	2.0	2.1
NH_{Thr}-$H_{\alpha\ Thr}$	2.9	3.0
NH_{Asp}-$H_{\alpha\ Pro}$	2.0	1.9
NH_{Asp}-$H_{\alpha\ Asp}$	2.6	2.7
NH_{Thr}- NH_{GalNAc}	2.4	2.5
NH_{Thr}-NH_{Arg}	2.2	2.4

To rationalize the results obtained from the study of the NMR experiments, we performed molecular dynamics with time-averaged NMR restraints (MD-tar) obtaining the major conformations of glycopeptide **1** in water (Figure 3.16).

In the past, the X-ray structure of monoclonal antibody with a non-glycosylated nonapeptide derived from MUC1 was published and the 3D structure of the PDTR fragment was determined showing two consecutive inverse γ-turns.[23] Our results showed that the 3D structure obtained by molecular dynamics *in the free state* was very similar to that published in solid state for the PDTR sequence and two inverse γ turns were also observed.

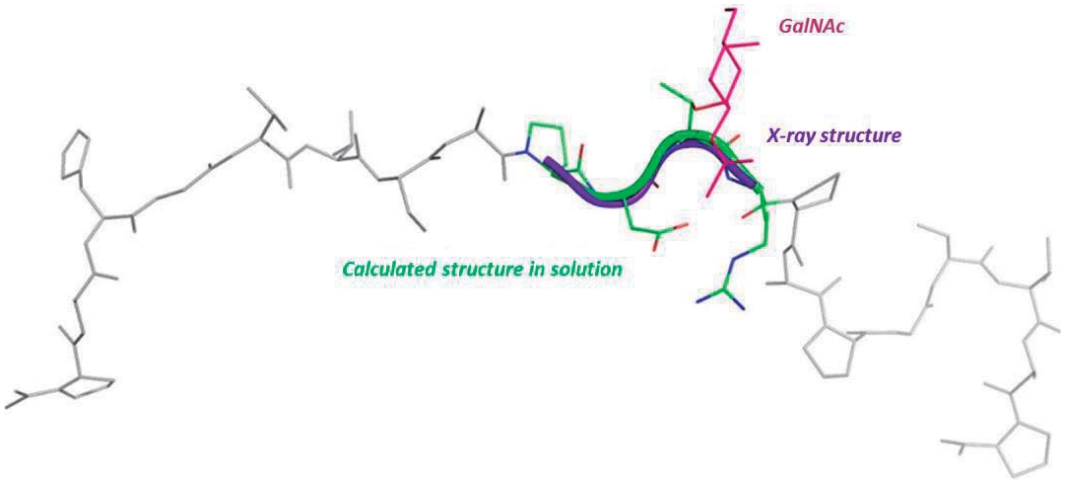


Figure 3.16. *Major conformation for glycopeptide **1** obtained by MD. The sequence PDTR from the X-ray structure (purple) is superimposed with that found for our glycopeptide (green and pink).*

Our results are in agreement with those reported by other groups for glycopeptide **1**, which has been extensively studied in the past, and all the

[23] P. Dokurno, P. A. Bates, H. A. Band, L. M. D. Stewart, J. M. Lally, J. M. Burchell, J. Taylor-Papadimitriou, D. Snary, M. J. E. Sternberg, P. S. Freemont, *J. Mol. Biol.* **1998**, *284*, 713-728.

reports conclude that GalNAc-glycosylation forces the peptide backbone to adopt an extended conformation. In fact, the most immunogenic domain (the PDTRP sequence) displays, upon glycosylation of the threonine residue, an extended geometry.[24] Concerning the glycosidic linkage, the ψ torsional angle takes values around 120° in compound **1,** which is typical for the eclipsed conformation, and χ^1 was quite rigid, with values around 60°, in good agreement with a the low $^3J_{H\alpha\text{-}H\beta}$.

We have also analyzed the main torsional angles by MD-tar for glycopeptide **1** (Figure 3.17). There are obvious differences in the backbone: while proline, aspartic acid and arginine residues exhibit an extended conformation, threonine displays a folded conformation.

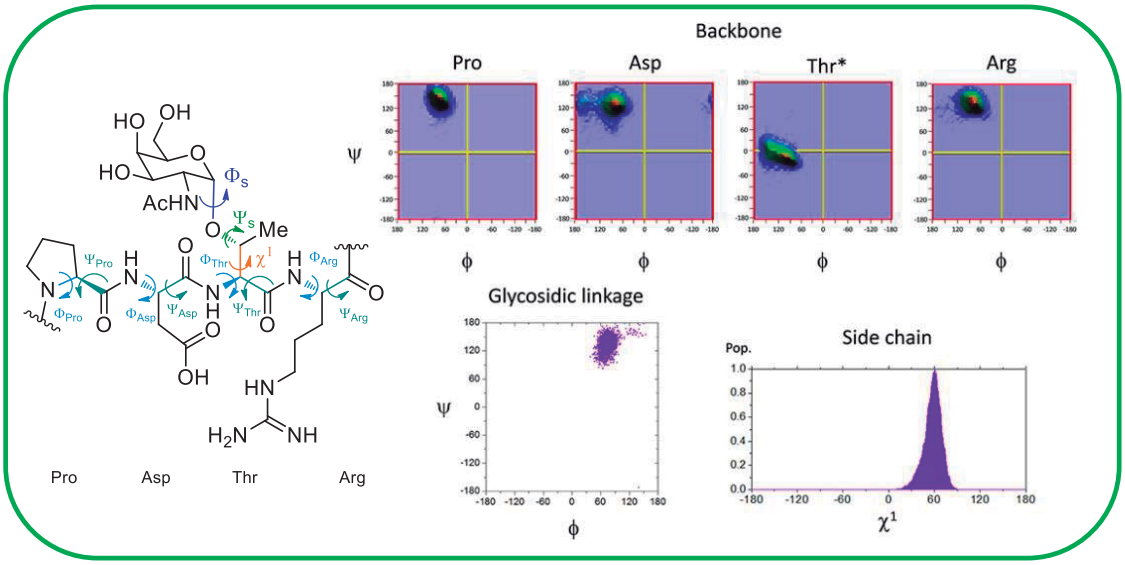


Figure 3.17. *Main torsional angles obtained by MD-tar for PDT*R fragment of glycopeptide **1**.*

[24] a) N. Matsushita, N. Ohyabu, N. Fujitani, K. Naruchi, H. Shimizu, H. Hinou, S. I. Nishimura, *Biochemistry* **2013**, *52*, 402-414; b) U. Karsten, N. Serttas, H. Paulsen, A. Danielczyk, S. Goletz, *Glycobiology* **2004**, *14*, 681-692.

3.3.4. Conformational analysis in the bound state

To check if these different geometrical features of serine and threonine are also present in the bound state, we investigated the conformation of the glycopeptides **1** and **2** bound to the three selected lectins: SBA, VVA and HPA. As a first step, the binding epitopes of glycopeptides **1** and **2** complexed to SBA lectin were studied by saturation-transfer difference (STD) NMR data.[25] STD signals were detected for both molecules (Figure 3.18), validating their specific interaction with the SBA lectin.

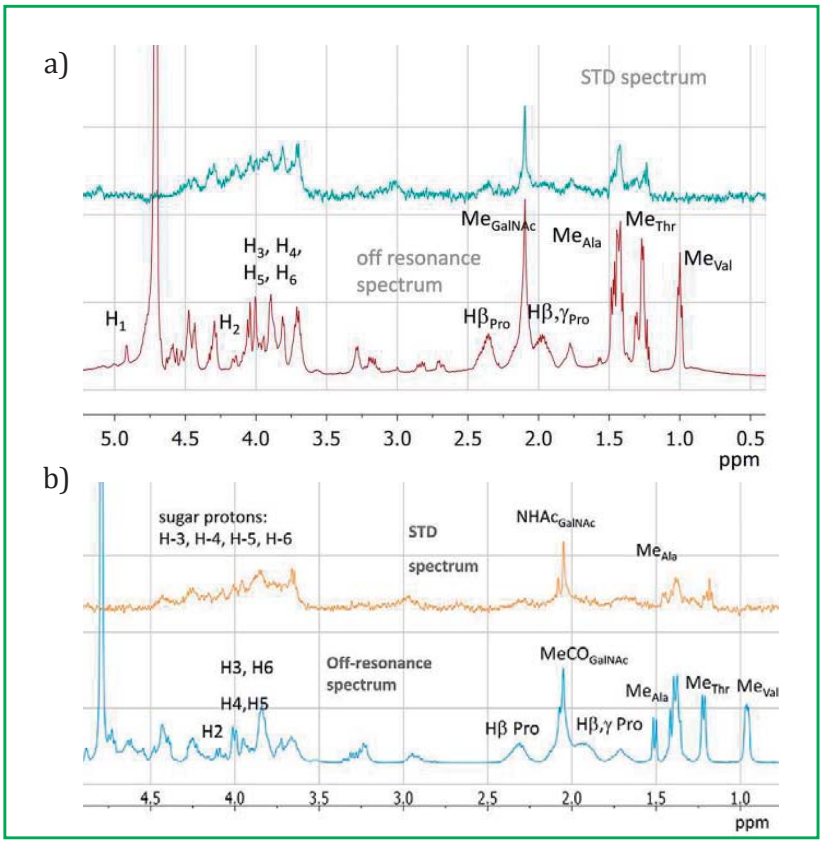


Figure 3.18. *Saturation-transfer difference (STD) spectrum for a sample of SBA (40 μM) and glycopeptide **1** (a) and **2** (b) (0.8 mM).*

[25] M. Mayer, B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784-1788.

If we take a look at the STD spectra, they suggest that the epitope of the both glycopeptides is located not only at the sugar moiety, but also at the residues of the peptide backbone close to the glycosylation point, such as Ala7, Thr10 (in glycopeptide **1**) or Ser10 (in glycopeptide **2**), Pro8 and Pro12. It is clear that the underlying amino acid plays a key role in the lectin-carbohydrate interaction in a glycopeptide.

Trying to rationalize these results and to find a promising answer, 3D models of the complexes between SBA and glycopeptides **1** and **2** were deduced by 100 ns unrestrained MD simulations. We used the X-ray structure of the complex SBA with a disaccharide[26] as starting coordinates. In order to confirm that theoretical binding mode was right, we calculated the theoretical STD effects corresponding to glycopeptides **1** and **2** protons in the complexes with SBA lectin using CORCEMA-STD software[27] and compared to those experimentally obtained (Figure 3.19).

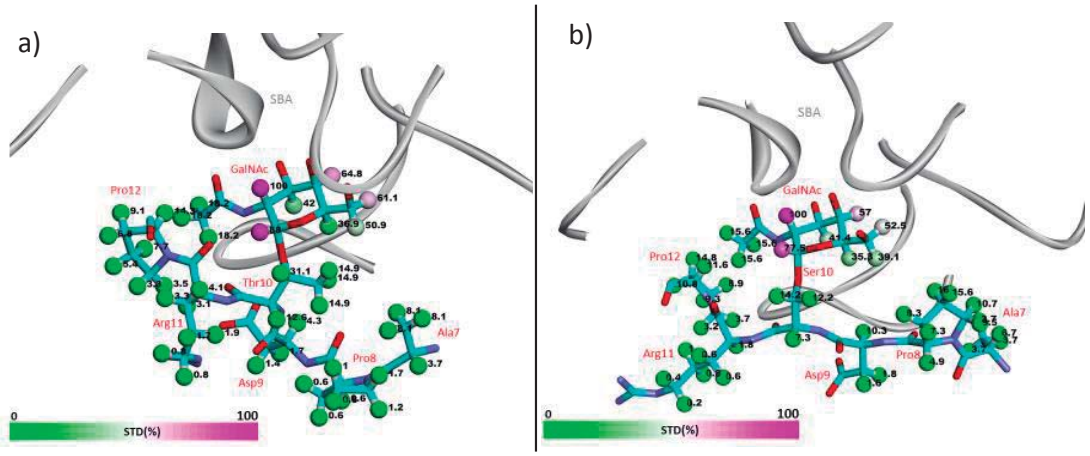


Figure 3.19. *a) Theoretical STD effects calculated for some protons of glycopeptide **1** (a) and **2** (b) in the bound state with SBA lectin. SBA lectin is shown as grey ribbons.*

[26] L. R. Olsen, A. Dessen, D. Gupta, S. Sabesan, J. C. Sacchettini, C. F. Brewer, *Biochemistry* **1997**, *36*, 15073-15080.

[27] V. Jayalakshmi, N. R. Krishna, *J. Am. Chem. Soc.* **2005**, *127*, 14080-14084.

The simulations revealed that although the main epitope is located at the GalNAc moiety, some protons of the peptide backbone have important STD effects. In the case of Thr-Tn derivative (glycopeptide **1**) the STD involves the methyl groups of Thr10 and Ala7 as well as ring protons of Pro12. The agreement between the experimental and theoretical data validates our model in terms of epitope binding.

We performed a molecular dynamics simulation of the bound state with the SBA lectin and both glycopeptides **1** and **2** for trying to understand better the interactions between the lectin and our glycopeptides (Figure 3.20).

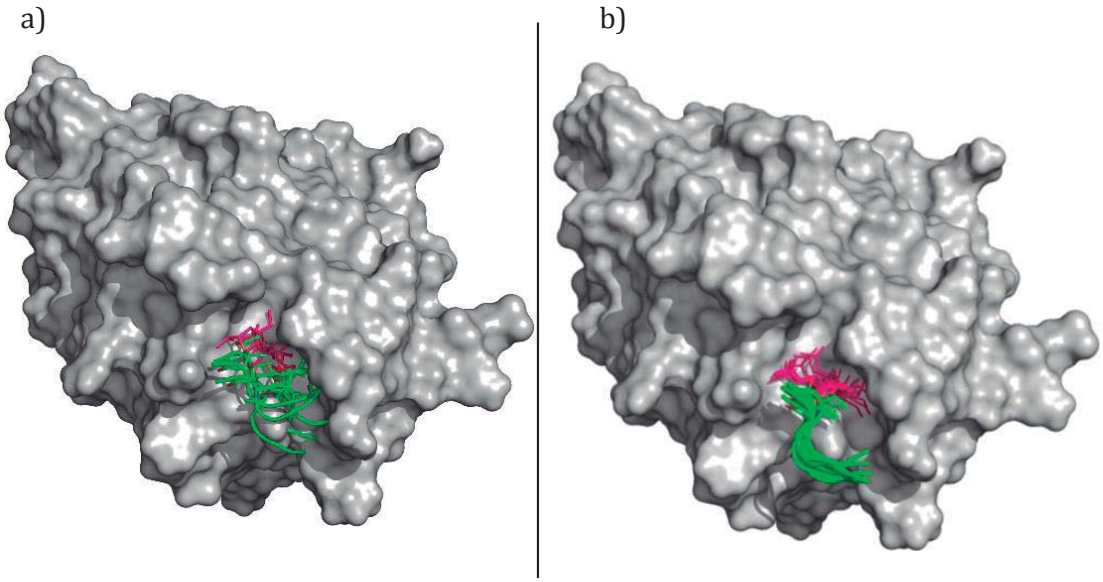


Figure 3.20. *Ensembles obtained from the unrestrained 100 ns simulations for SBA:**1** (a) and SBA:**2** (b) complexes. Only some residues of the peptide backbone are shown for clarity.*

According to these simulations, the hydroxyl groups of GalNAc form a hydrogen-bonding network with the lectin in both complexes (Table 3.4).

Table 3.4. *List of hydrogen bonds (% population) between α-O-GalNAc moiety and SBA lectin for glicopeptides **1** and **2** obtained from unrestrained 100 ns MD simulations.*

	SBA:1	SBA:2
Asn130-O3	74.9	87.7
Asp88-O3	100	100
Asp88-O4	1.3	8.4
Leu214-O4	45.1	51.2
Asp215-O6	13.2	4.3
Gly106-N(GalNAc)	29.1	60.0
Asn130-CO(GalNAc)	5.7	4.3

If we analyze with more detail the interactions that appeared in the simulations, an interesting CH-π interaction between the α-face of the GalNAc and Phe128 of SBA is established in both complexes (Figure 3.21). The methyl groups of Thr10 (complex SBA:**1**) and Ala7 (complex SBA:**1** and SBA:**2**) are also embedded in a hydrophobic pocket formed by Phe128, which is in good agreement with the STD spectrum aforementioned (Figure 3.18).

In glycopeptide **2** complexed to SBA, the hydrophobic contact between the underlying residue (serine) and Phe128 is weaker due to the absence of the methyl group and the interaction between Phe128-Ala7 appears to be less strong that in SBA:**1** complex.

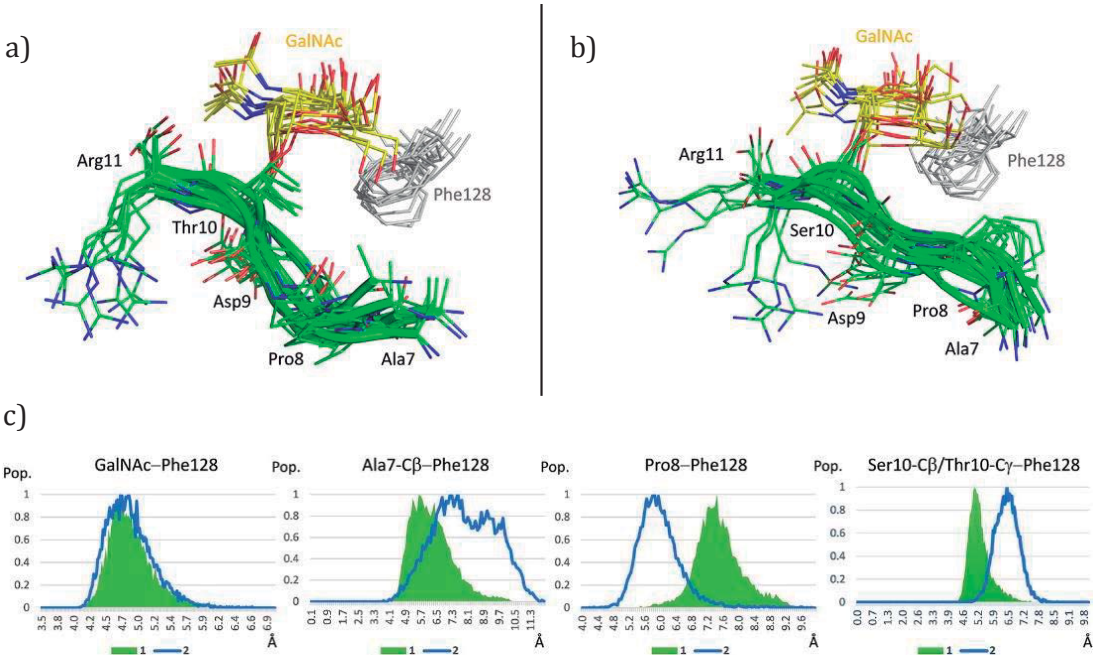


Figure 3.21. *Ensembles obtained from the unrestrained 100 ns MD simulations for SBA:**1** (a) and SBA:**2** (b) complexes. Only some residues of the peptide backbone are shown for clarity. c) Relevant distance distributions obtained from the unrestrained 100 ns MD simulations for SBA:**1** (in green) and SBA:**2** (blue line) complexes.*

By contrast, an interesting contact between Pro8 and Phe128 was also deduced from the MD simulations. As it was expected, the side chains (represented by $\chi 1$ torsional angle) of the glycosidic residue are fixed when these compounds are bound to SBA, with a value of χ^1 close to 60° in both glycopeptides (Figure 3.22).

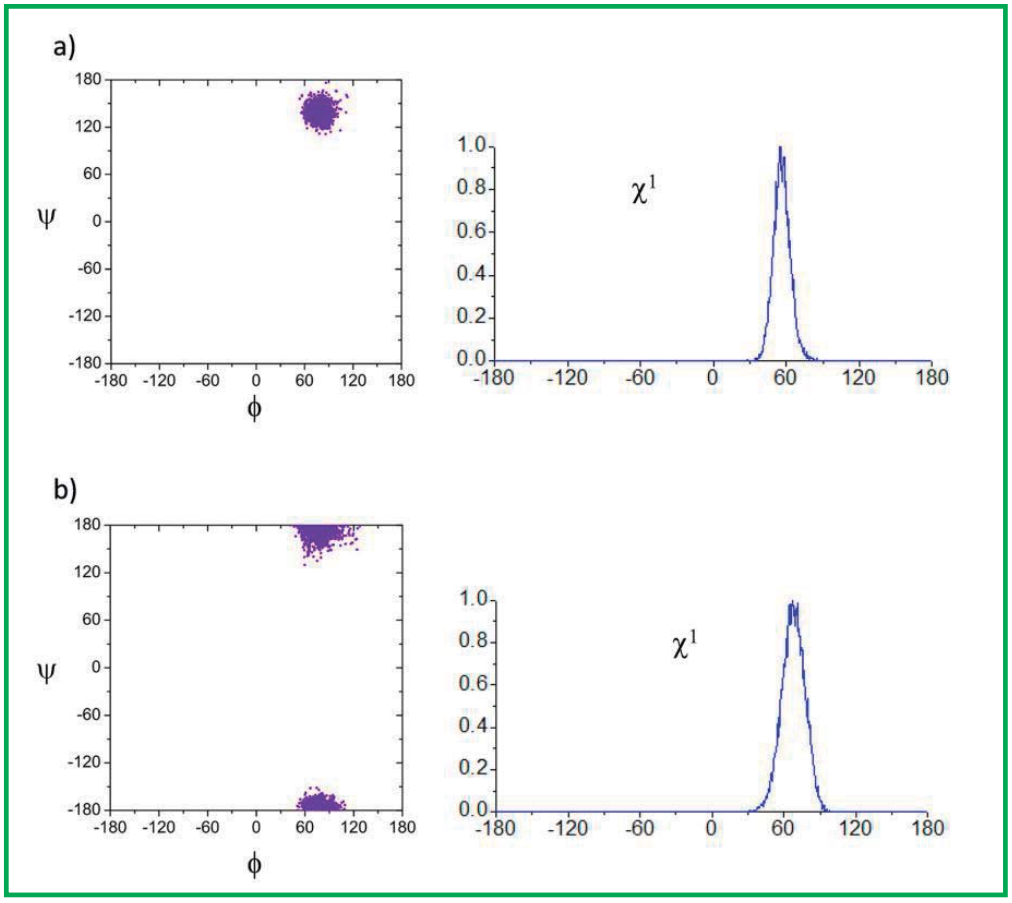


Figure 3.22. *Distribution for the glycosidic linkage (φ/ψ) and the side chain (χ^1) obtained from the unrestrained 100 ns MD simulations in explicit water for glycopeptides **1** (a) and **2** (b) bound to SBA lectin.*

Combining the results obtained for both compounds **1** and **2** as well as the Figure 3.20 and 3.21 showed, the extra flexibility of the GalNAc-Ser fragment in comparison with the GalNAc-Thr might disfavor binding to glycopeptide **2** due to a higher entropy penalty.

This behavior of the side chain could have a negative impact on binding process, explaining the lower affinity of SBA lectin for the Tn-Ser derivative

(glycopeptide **2**). A comparable conclusion could be obtained from the MD simulations performed on VVA:**1** and VVA:**2**, complexes (Figure 3.23). In this sense, the hydrophobic contact of the underlying amino acid is stablished with Tyr127.

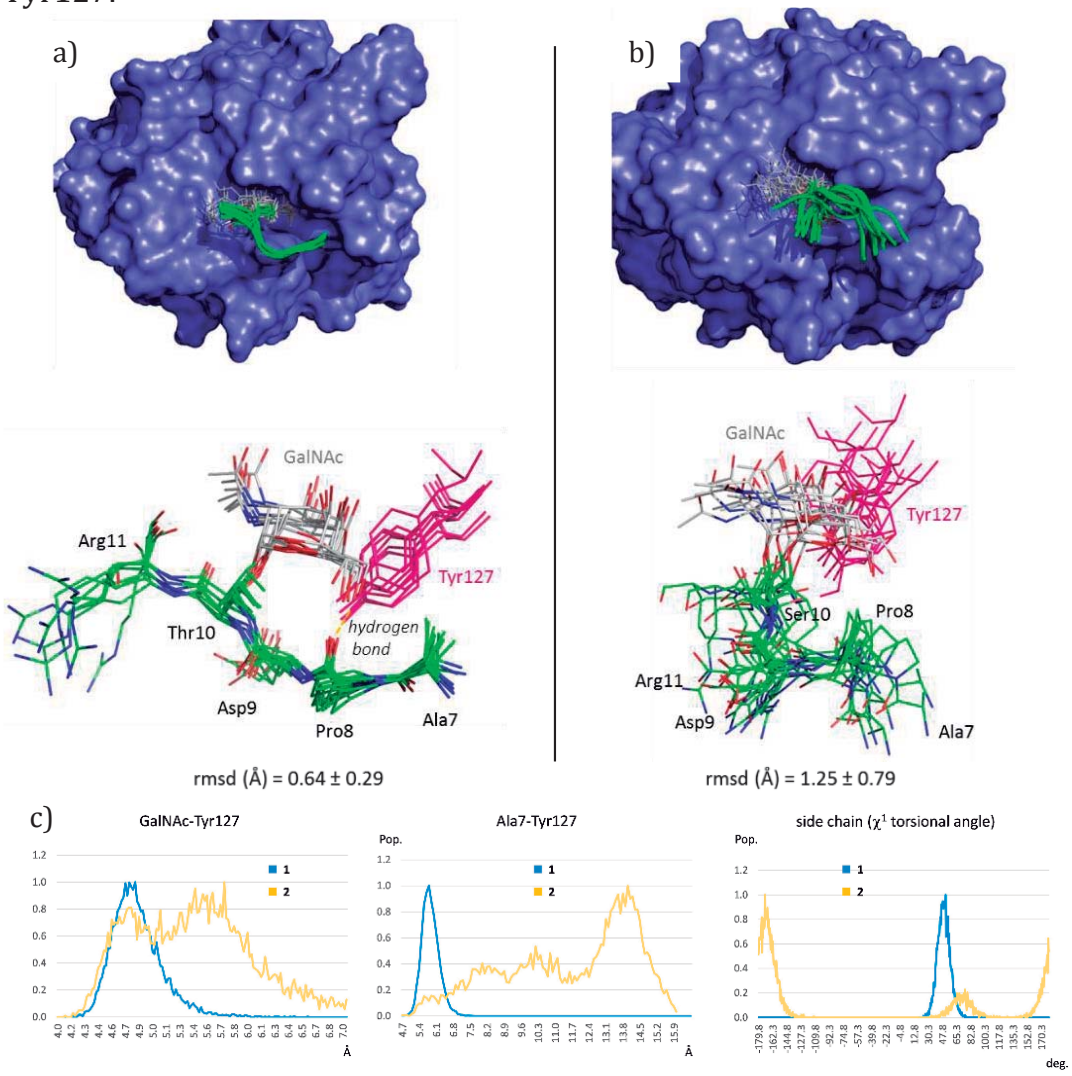


Figure 3.23. *a) Ensembles obtained from the unrestrained 25 ns MD simulations for VVA:**1** (a) and VVA:**2** (b) complexes. Only some residues of the peptide backbone are shown for clarity. c) Some relevant distance distributions as well as the values of χ1 torsional angle obtained from the unrestrained 100 ns MD simulations for VVA:**1** (in blue) and VVA:**2** (in orange) complexes.*

According to these simulations, the hydrophobic contacts are stronger in VVA:**1** complex. In addition, the side chain of glycopeptide **2** bound to VVA adopts a value around 180°, which differs from the expected value in solution (χ^1 close to 60°). These finding may explain the higher affinity of VVA lectin for glycopeptide **1**.

As a next step, we have analyzed the difference in affinity between these two lectins (SBA and VVA) in comparison to HPA lectin. The conformational analysis of glycopeptides **1** and **2** bound to HPA lectin was also performed by MD. The close inspection of the X-ray structure of HPA complexed to Tn antigen[28] revealed that the hydroxyl groups as well as the endocyclic oxygen of the sugar moiety interact with the lectin through various hydrogen bonds. The serine residue of the Tn antigen did not establish hydrogen bonds with the protein (Figure 3.24).

[28] J. Lescar, J. F. Sanchez, A. Audfray, J. L. Coll, C. Breton, E. P. Mitchell, A. Imberty, *Glycobiology* **2007**, *17*, 1077-1083.

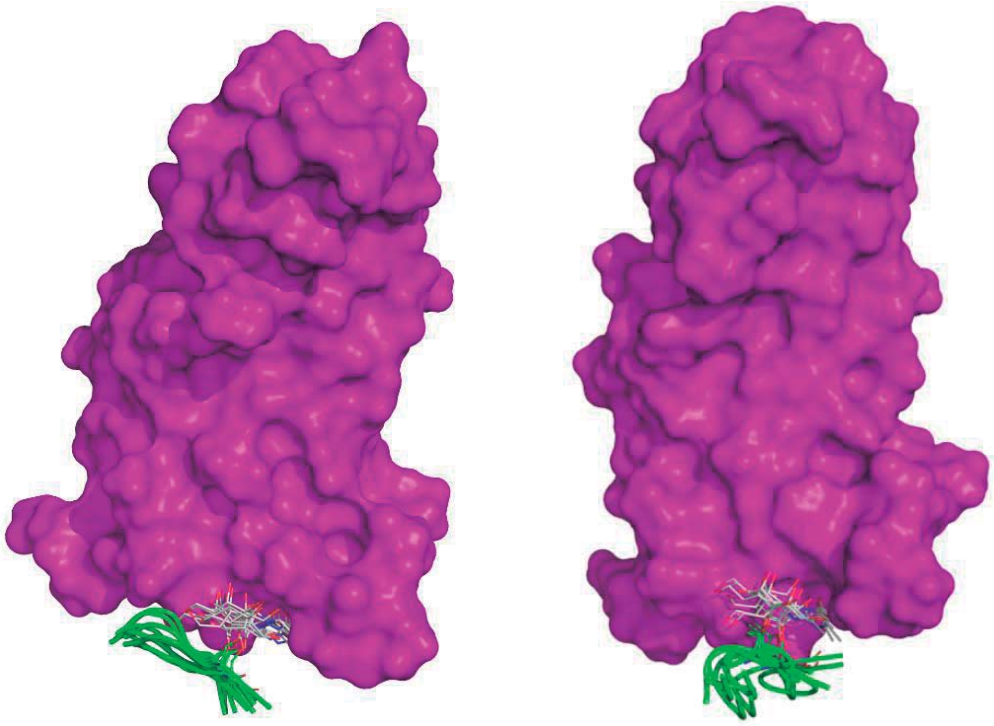


Figure 3.24. *Ensembles obtained from the unrestrained 100 ns MD simulations for HPA:**1** (a) and HPA:**2** (b) complexes. Only some residues are shown for clarity.*

However, a strong hydrophobic interaction is appeared between H_β of the serine of the Tn antigen and C1 of GalNAc with the side chain of His84. The results obtained from the unrestrained MD simulations are summarized in Figure 3.25. According to our MD simulations, the hydrogen bonding network between the GalNAc and the protein are identical to those found in the X-ray structure in both complexes (Table 3.5).

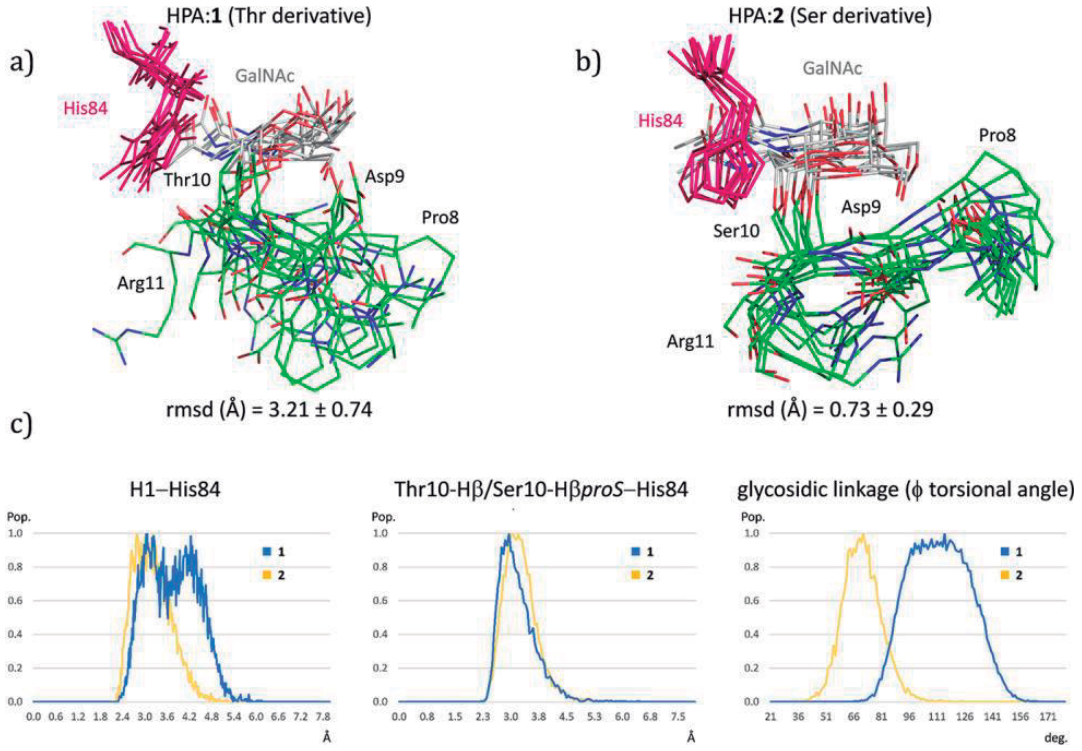


Figure 3.25. *Some relevant distance distributions as well as the values of φ torsional angle obtained from the unrestrained 100 ns MD simulations for HPA:**1** (in blue) and HPA:**2** (in yellow) complexes.*

Table 3.5. *List of hydrogen bonds (% population) between α-O-GalNAc moiety and HPA lectin for glycopeptide **1** and **2** obtained from unrestrained 100 ns MD simulations.*

	HPA:1	HPA:2
Arg63-O5	50.0	38.0
Asp26-CO(GalNAc)	82.5	72.9
Gly24-O3	96.2	91.8
Asn61-O6	34.4	3.1

In HPA:**2** complex, the glycosidic linkage displays the appropriate geometry to establish the hydrophobic pocket with His84. This interaction is also

observed in the complex HPA:**1**. However, to form an effective contact between Hβ of threonine of Tn antigen and His84, the glycosidic torsional angle φ adopts a non-exo-anomeric conformation, with values around 120°. This high-energy geometry of the glycosidic linkage may be forced by the methyl group of the threonine residue that blocks the system to accommodate the typical value of φ (around 80°) in the binding site (Figure 3.24). As shown in Figure 3.24, the flexibility of glycopeptide **1** (rmsd = 3.21 Å) in the bound state is higher than observed for compound **2** (rmsd = 0.73 Å), indicating the difficulty of HPA lectin to properly fit the threonine-containing glycopeptide in the binding site.

3.4. Conclusions

Herein, we have analyzed the molecular recognition of several glycopeptides bearing Tn antigen (α-*O*-GalNAc-Ser or α-*O*-GalNAc-Thr) in their structure by three lectins with affinity for this determinant. The work yields remarkable results in terms of epitope recognition, showing that the underlying amino acid of Tn (serine or threonine) plays a key role in the molecular recognition. In fact, while SBA (soybean agglutinin) and VVA (*Vicia villosa* agglutinin) lectins prefer Tn-threonine, HPA (*Helix pomatia* agglutinin) shows a higher affinity for the glycopeptides carrying Tn-serine.

A reasonable explanation of this result is provided based on a study that involves enzyme-linked lectin assays (ELLA), isothermal titration microcalorimetry (ITC), Saturation-Transfer Difference (STD) NMR experiments, and Molecular Dynamics (MD) simulations. The different conformational behavior of the two Tn biological entities, the residues of the studied glycopeptides in the close proximity to the Tn antigen and the topology of the binding site of the lectins are at the origin of these differences.

Taking into account these results, it seems to be important to refer specifically which is the amino acid underlying in the Tn antigen, rather if contains Ser or Thr linked to the α-*O*-GalNAc.

This work has been published as an article entitled *Serine versus Threonine Glycosylation with α-O-GalNAc: Implications for the molecular recognition by lectins* in the journal *Chemistry – A European Journal*, volume 20, pages 12616–12627.

ROLE OF THE PEPTIDIC SEQUENCE IN THE CARBOHYDRATE RECOGNITION OF MUC1 EPITOPES BY LECTINS

4.1. Introduction

As we have described in *Chapter 1*, lectins are a class of carbohydrate-binding proteins that trigger several important cellular processes. In particular, lectins have been successfully employed to recognize malignant tumors and some of them are indicated for the reduction of treatment-associated side effects as adjuvant agents in chemotherapy and radiotherapy. Recently, it has been observed that certain lectins present anti-tumor activity by inducing apoptosis or autophagy in cancer cells. These features make lectins a current niche of research and can be adopted for an alternative cancer therapy.[1]

Due to the importance of Tn antigen in tumor cells (*Chapter 3*), the premature detection of this antigen is essential to properly treat and eradicate tumors. A variety of lectins bind with moderate to high affinity to glycopeptides bearing Tn motifs. Indeed, certain lectins, as VVA lectin, have been successfully employed as biosensors[2] of these antigens, while other can directly kill human cancerous cells.

From the molecular recognition viewpoint, it is well documented that lectins recognize both the sugar moiety as a free form or linked to peptides. Previous works in our group evidenced that the natural residue (Ser or Thr) bearing the sugar unit does have indeed a clear effect on binding,[3] however, little is

---

[1] a) N. Gaidzik, U. Westerlind, H. Kunz, *Chem. Soc. Rev.* **2013**, *42*, 4421-4442; b) R. M. Wilson, S. J. Danishefsky, *J. Am. Chem. Soc.* **2013**, *135*, 14462-14472; c) R. E. Beatson, J. Taylor-Papadimitriou, J. M. Burchell, *Immunotherapy* **2010**, *2*, 305-327.

[2] a) D. Clark, L. Mao, *Dis. Markers* **2012**, *33*, 1-10; b) M. Kaszowska, A. S. Norgren, P. I. Arvidson, C. Sandström, *Carbohydr. Res.* **2009**, *344*, 2577-2580.

[3] a) D. Madariaga, N. Martínez-Sáez, V. J. Somovilla, L. García-García, M. Á. Berbis, J. Valero-Gónzalez, S. Martín-Santamaría, R. Hurtado-Guerrero, J. L. Asensio, J. Jiménez-Barbero, A. Avenoza, J. H. Busto, F. Corzana, J. M. Peregrina, *Chem. Eur. J.* **2014**, *20*, 12616-12627; b) F. Corzana, J. H. Busto, M. García de Luis, J. Jiménez-Barbero, A. Avenoza, J. M. Peregrina, *Chem. Eur. J.* **2009**, *15*, 3863-3874.

known about the role of amino acids flanking the glycosylation point in the lectin binding.

In addition, to analyze and study the interactions between our glycopeptides and lectins we need to know the conformations in the bound state. For that reason, it will be very important to have an X-ray structure for obtaining real data for analysis and molecular dynamics simulation. Some structures of lectins that recognize GalNAc and Tn antigen are available, for example: HPA lectin with Tn antigen (pdb: 2cgz), VVA-B4 lectin with Tn antigen (pdb: 1n47) and SBA with 2,4- (pdb: 1sbd) or 2,6-pentasaccharide (pdb: 2sba), but no evidence with Tn antigen.

On the other hand, in *Chapter 3*, we have analyzed the difference between Ser and Thr, in a MUC1 glycopeptide context, when they interact with lectins. Following these ideas, we wanted to go one step further by exploring the importance that the peptide sequence nearby the glycosylation site, particularly the residues flanking the α-GalNAc-Thr motif, could have in the molecular recognition process. To this purpose, we have selected soybean and *Vicia villosa* lectins because they show a clear preference for Tn antigen carrying Thr.

4.2. Objectives

In this chapter, our aim is to study the influence of the peptide sequence in the molecular recognition of Tn antigen (with threonine) by two different GalNAc specific lectins (SBA and VVA). We have selected the three important epitopes of MUC1 and we have synthesized both small glycopeptides and 21-mer MUC1 glycopeptides that incorporate the glycosylated epitope.

As aforementioned, there is one X-ray available of the Tn antigen with the VVA and HPA but there is no a reported structure with SBA complexed to Tn; therefore, we also desire to obtain an X-ray of the Tn antigen with the soybean lectin.

To understand the interactions of lectins with the studied glycopeptides is essential to perform a complete analysis in the free and bound states of all derivatives with both lectins.

As a first step, we have synthesized the three derivatives from MUC1 tandem repeat shown in Figure 4.1.

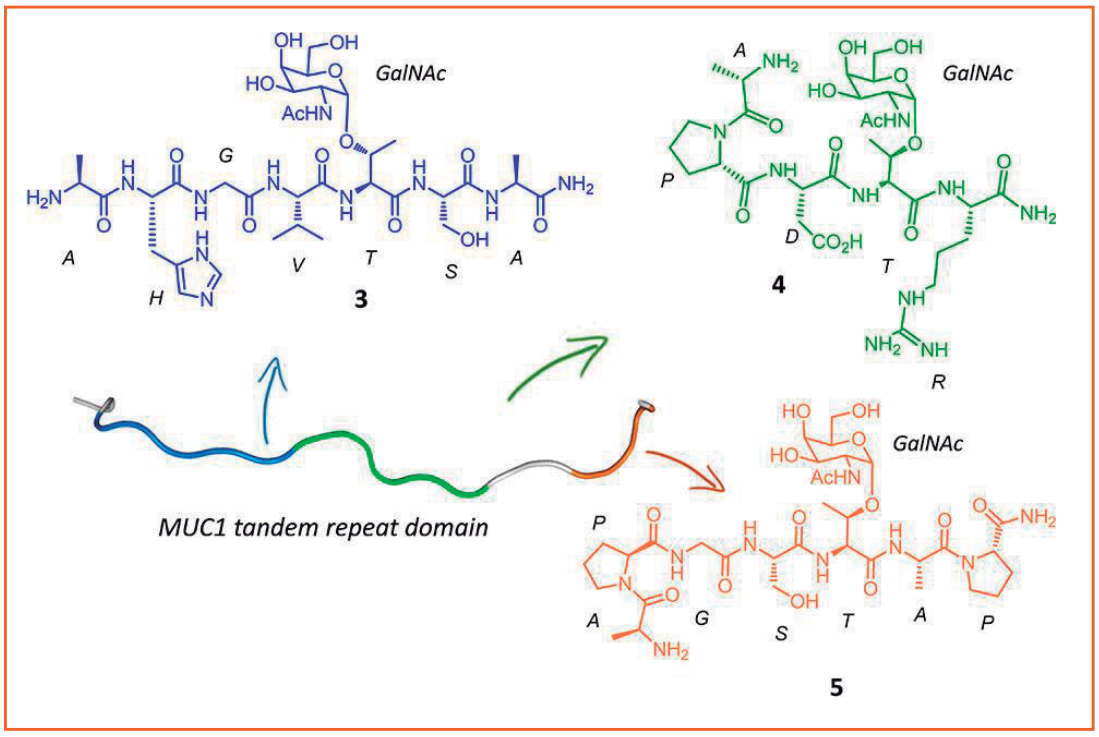


Figure 4.1. *Small glycopeptides synthesized and studied in this work that contain the three relevant epitopes of MUC1 mucin.*

Besides, we have achieved the synthesis of three complete MUC1-like glycopeptides of 21 amino acids that incorporate the three specific regions (Figure 4.2).

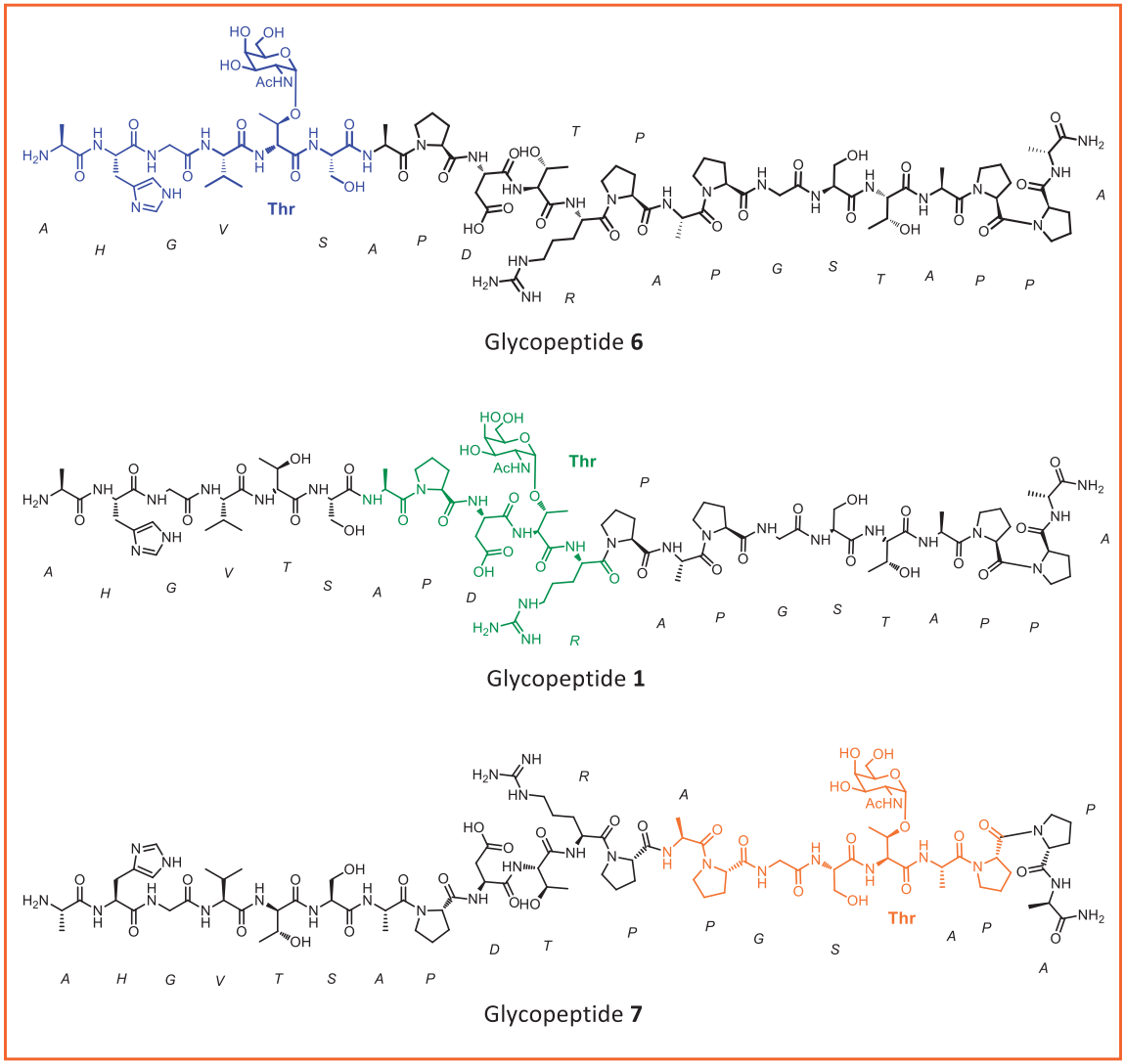


Figure 4.2. *Glycopeptides comprised of 21 amino acids synthesized and studied in this work that contain the three relevant epitopes of MUC1 mucin.*

4.3. Results and discussion

4.3.1. Synthesis

The synthesis of the glycopeptides was performed using the solid phase peptide synthesis protocol explained in *Chapter 2* and with the Koenigs-Knorr methodology described in *Chapter 3*.

4.3.2. Biological assays (ELLA and ITC)

As a first approximation, we investigated the binding properties of small glycopeptides **3**, **4** and **5** to SBA lectin by using an enzyme-linked lectin assay (ELLA) as shown in Figure 4.3.

We also repeated the experiment with the VVA lectin (Figure 4.4). To perform these assays, different amounts of the glycopeptides (from 0 to 1.5 mM) were covalently attached to a maleic anhydride activated surface. Biotinylated SBA and VVA lectins were then selected for estimating the binding. The horseradish peroxidase-3,3′,5,5′-tetramethylbenzidine (TMB) was chosen as detection system by measuring the absorbance of the wells at 450 nm (see *Chapter 5*).

We obtained similar results for both lectins as we summarize in Figure 4.5.

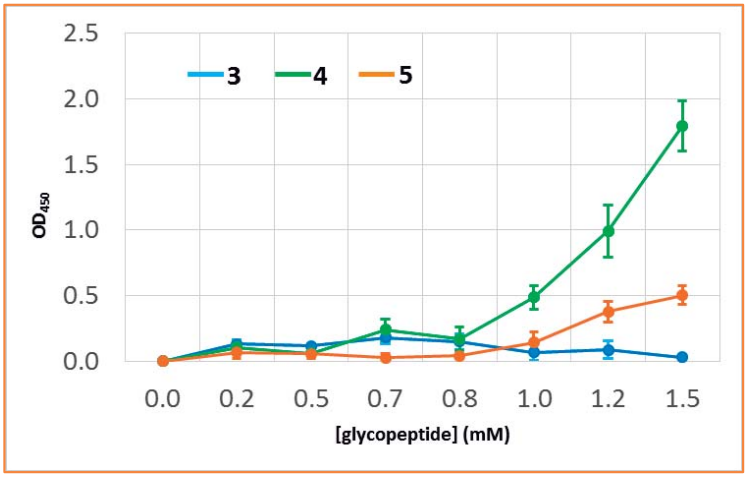


Figure 4.3. *ELLA results in term of affinity of glycopeptides **3**, **4** and **5** to SBA lectin. Absorbance signals are the average of three replicate wells.*

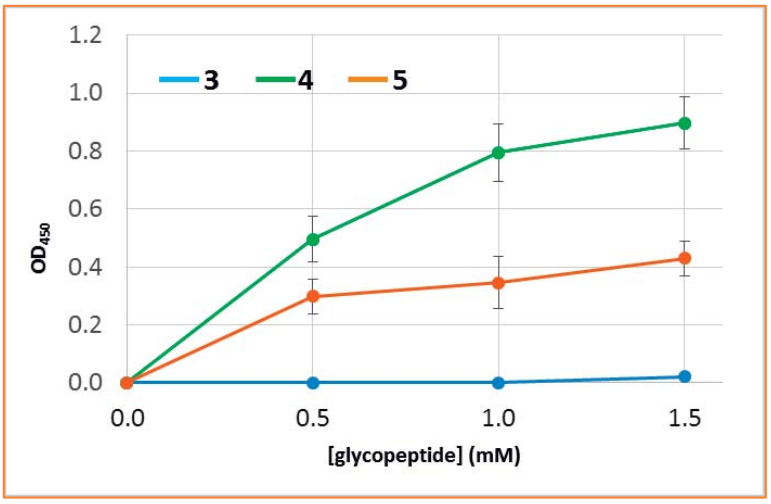


Figure 4.4. *ELLA results in term of affinity of glycopeptides **3**, **4** and **5** to VVA lectin. Absorbance signals are the average of three replicate wells.*

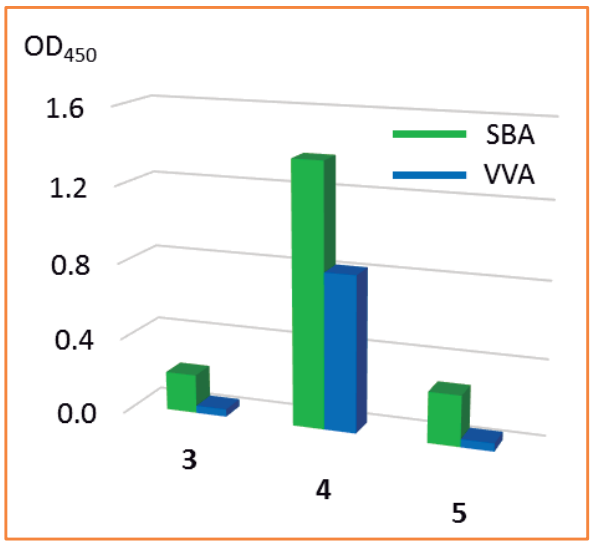


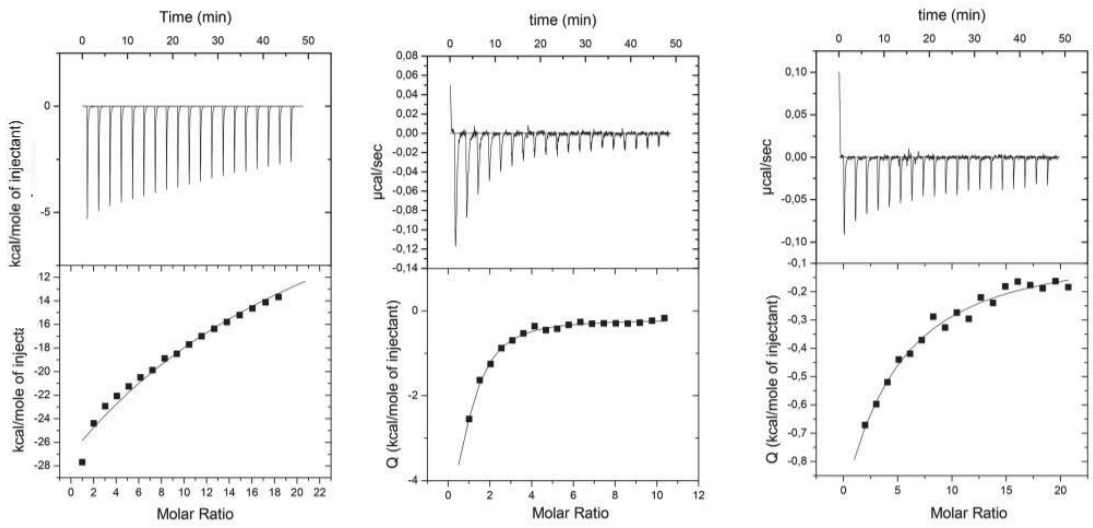
Figure 4.5. *Comparative graphic between SBA (green) and VVA (blue) lectins of glycopeptides **3**, **4** and **5** (concentration = 100 mM).*

SBA lectin showed the maximum affinity for compound **4**. In contrast, the binding of glycopeptides **3** and **5** was very low and negligible, respectively.

The VVA results were in agreement with the previous experiment: glycopeptides **3** and **5** exhibited less affinity to the lectin than **4**.

As in *Chapter 3*, ELLA tests were subsequently corroborated by isothermal titration microcalorimetry (ITC). We chose the SBA lectin for the assays because of its higher affinity to the glycopeptides studied in this work.

The results were in line to those estimated using ELLA assays (Figure 4.6 and Table 4.1) and allowed us to quantify the binding. The affinity of SBA for compound **4** was calculated to be around 15-fold the affinity for glycopeptide **5**. In addition, the binding affinities between compounds **3** and **4** differ by two orders of magnitude. It is important to note that due to the low affinity of compound **3**, indicated by the particular ITC profile shown in Figure 4.6, we could not estimate its K_D for the binding with SBA lectin.



Glycopeptide **3** Glycopeptide **4** Glycopeptide **5**

Figure 4.6. *ITC profiles of glycopeptides **3**, **4** and **5** titration into SBA lectin solutions at 25 °C and pH = 7.2. The solid line represents the least-squares fitting of the data to the simplest model (one binding site).*

Table 4.1. *Thermodynamic binding parameters obtained by ITC experiments for SBA lectin and glycopeptides **3**, **4** and **5** at 25 ˚C and pH = 7.2.*

	K_D (mM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)	n
3	≥ 810	-	-	-	1
4	6.6±1.9	-7.1±0.3	-20.6±2.9	-13.5±3.2	0.9±0.3
5	100.7±20.1	-5.5±0.2	-8.7±1.6	-3.3±0.8	1.02±0.2

Inspired by these results, we decided to investigate whether the affinity of these short molecules could be modified by the presence of the rest of the amino acids that comprise the MUC1 tandem repeat domain.

Subsequently, we synthesized and studied 21-mer glycopeptides **1**, **6** and **7** (Figure 4.2). Preliminary results obtained from the ELLA assay performed on these glycopeptides were equivalent to those obtained for short glycopeptides **3**, **4** and **5**, being the derivative that incorporated the APDT(αGalNAc)R epitope the compound that clearly showed the best binding to SBA and VVA lectins (Figures 4.7 and 4.8).

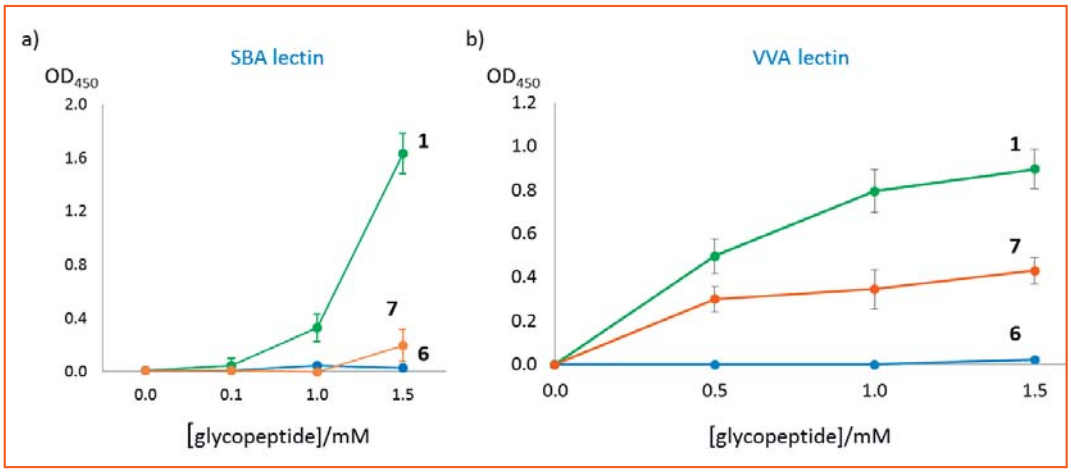


Figure 4.7. *ELLA results in term of affinity of glycopeptides **1**, **6** and **7** to SBA (a) and VVA (b) lectins. Absorbance signals are the average of three replicate wells.*

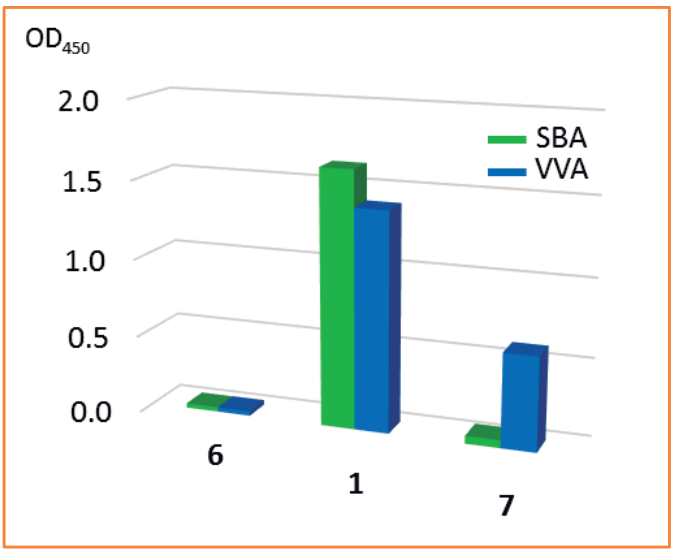


Figure 4.8. *Comparative graphic between SBA (green) and VVA (blue) lectins of glycopeptides **1**, **6** and **7** (concentration = 100 mM).*

Taking into account that compound **6**, with the lower affinity, has the recognition epitope -GVT(αGalNAc)SA- close to the plastic surface in the ELLA tests, and to prevent artifacts due to accessibility issues, we synthesized glycopeptide **6'**, with the epitope situated opposite to the amino acid linked to plastic surface (Figure 4.9).

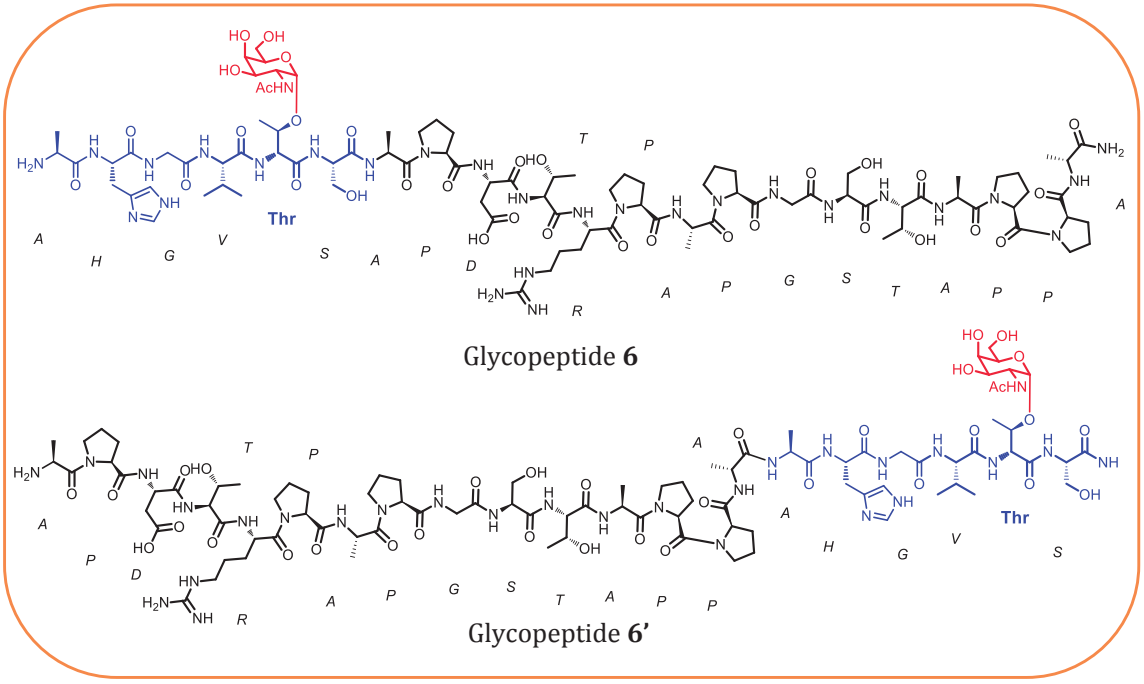


Figure 4.9. *21-mer glycopeptides* ***6*** *and* ***6'****.*

We tested the new synthesized glycopeptide **6'** together with glycopeptide **1**, which exhibited the higher affinity, versus SBA lectin. The results obtained for **6** and **6'** revealed the absence of accessibility difficulties in the recognition of these glycopeptides by SBA lectin (Figure 4.10). Again, glycopeptide **1** showed the highest affinity in all cases.

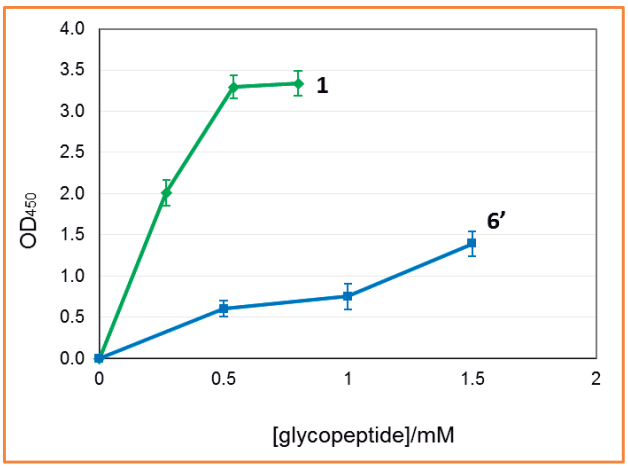


Figure 4.10. *Binding curves for mucin derivatives **1** (green) and **6'** (blue) with SBA lectin. Absorbance signals are the average of three replicate wells and the error bars show the standard deviations for these measurements.*

All these tests disclose that the peptide sequence flanking the glycosylation point does play an active role in the molecular recognition of GalNAc by lectins.

4.3.3. Free state conformational analysis (NMR and MD)

Trying to understand and rationalize these experimental data, and bearing in mind that similar results were found for short (**3**, **4** and **5**) and large glycopeptides (**1**, **6** and **7**), we performed an exhaustive conformational analysis only for glycopeptides **3, 4** and **5** in both the free- and bound-state to SBA lectin.

First, full assignment of the protons of the three compounds was performed using COSY and HSQC experiments (*Chapter 5: Experimental*). Notably, in glycopeptides **4** and **5** a second set of signals (in a small percentage) was

observed, especially in the NH region. They correspond to the *cis* disposition of the amide bond of the proline residues. [4]

As a next step, we used the proton-proton distances with conformational information deduced from 2D-NOESY experiments (Figure 4.11) as restraints in MD simulations. In that case, all the MD simulations were performed using only the set of signals corresponding to *trans* disposition. Due to the presumable flexibility of the glycopeptides, we used MD simulations with time-averaged restraints (MD-tar).

These simulations have been successfully applied to flexible systems and provide a distribution of low-energy conformers able to quantitatively reproduce the NMR data. A good agreement between the experimental and theoretically derived distances was obtained for the three glycopeptides (Tables 4.2, 4.3 and 4.4).

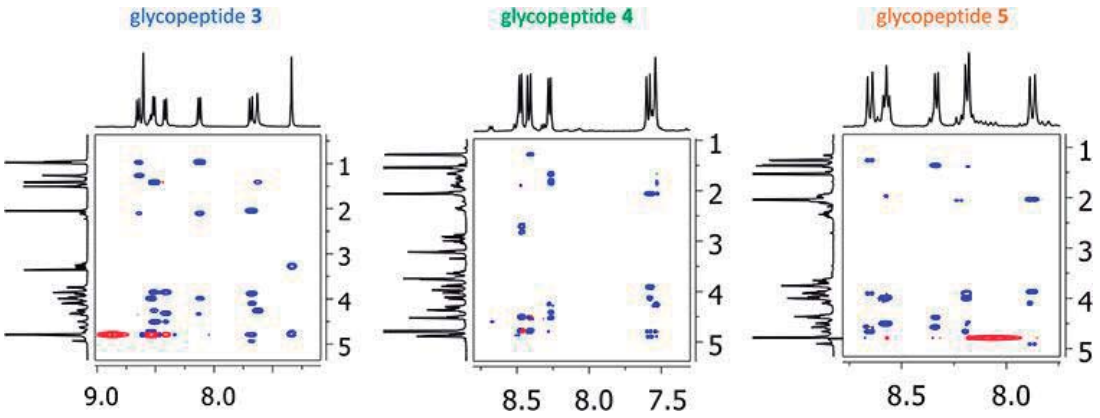


Figure 4.11. *2D-NOESY experiment (400 MHz) of glycopeptides **3**(left) **4** (middle) and **5** (right) at 20 °C and pH = 5.2 in a H_2O/D_2O (9:1) mixture. Diagonal peaks and exchange cross-peaks connecting NH protons and water are negative (blue color). The NOE contacts are represented as positive cross-peaks (red color).*

[4] S. Dziadek, C. Griesinger, H. Kunz, U. M. Reinscheid, *Chem. Eur. J.* **2006**, *12*, 4981-4993.

Table 4.2. *Experimental and calculated distances for glycopeptide **3**.*

	Exp. (Å)	MD_{H2O}-tar (Å)
NH_{Thr} -$H_{\alpha Thr}$	3.0	2.9
NH_{Gly} - $H_{\alpha His}$	2.6	2.4
NH_{Ala7} - $H_{\alpha Ala7}$	2.6	2.9
NH_{Ser} - $H_{\alpha Ser}$	2.2	2.3
NH_{Ser} - $NH_{\alpha Ser}$	2.7	2.9
NH_{Ser} - $NH_{\alpha Thr}$	2.8	2.5
NH_{Val} - $NH_{\alpha Val}$	2.8	2.9
NH_{Ser} - $NH_{\beta Thr}$	2.3	2.3

Table 4.3. *Experimental and calculated distances for glycopeptide **4**.*

	Exp. (Å)	MD_{H2O}-tar (Å)
NH_{Arg} - $H_{\alpha Arg}$	2.7	2.9
NH_{Arg} - $H_{\alpha Thr}$	2.6	2.5
NH_{Thr} - $H_{\alpha Asp}$	2.4	2.6
NH_{Asp} - $H_{\alpha Pro}$	2.2	2.3

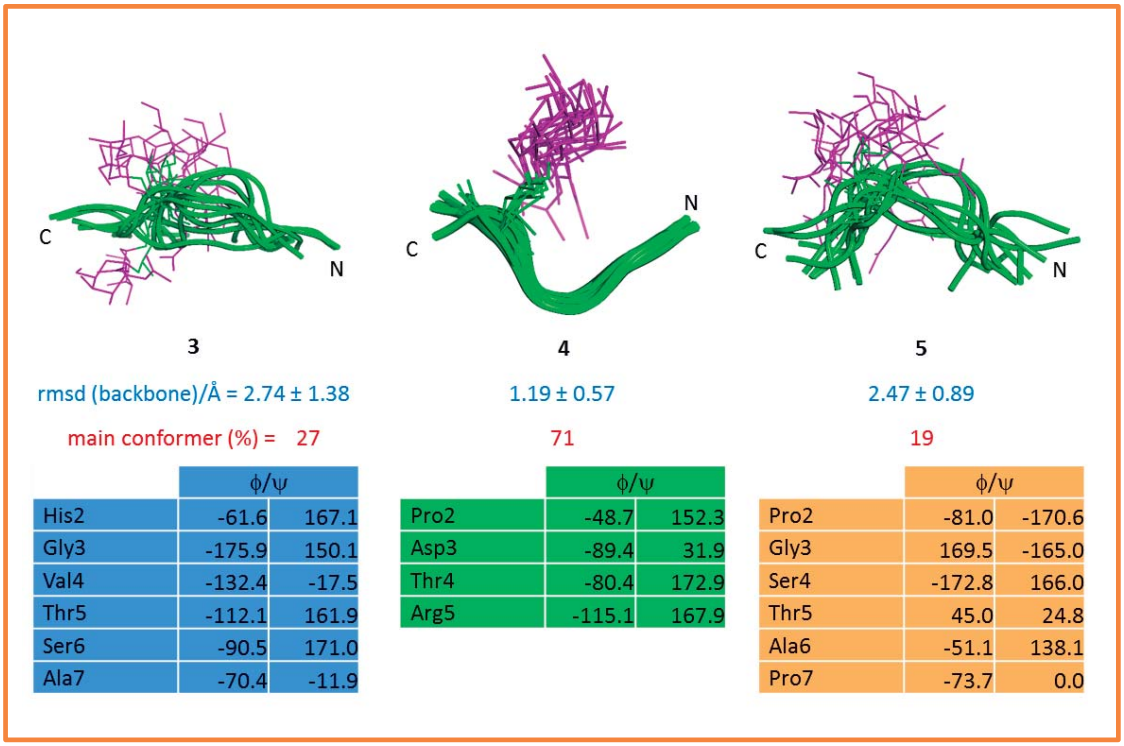
Table 4.4. *Experimental and calculated distances for glycopeptide **5**.*

	Exp. (Å)	MD_{H2O}-tar (Å)
NH_{Thr} -$H_{\alpha Thr}$	2.9	2.6
NH_{Gly} - $H_{\alpha Pro2}$	2.2	2.3
NH_{Ala6} - $H_{\alpha Ala6}$	2.8	2.9
NH_{Ala6} - $H_{\alpha Thr}$	2.6	2.6
NH_{Ser} - $H_{\alpha Ser}$	3.0	2.9

A close look to the molecular structures and ϕ and ψ angles obtained by MD simulations (Figure 4.12), indicates that while glycopeptide **4** is quite rigid in

solution, compounds **3** and **5** are flexible. These data reveal that glycopeptide **4** should have a smaller entropy penalty upon binding to SBA than **3** and **5** derivatives, which is completely well-matched with the experimental thermodynamic parameter measured by ITC.

On the other hand, glycopeptide **4** displays a main conformation (populated around 71%), characterized by an inverse γ-turn that comprises Pro2 and Asp3 residues.



	ϕ/ψ	
His2	-61.6	167.1
Gly3	-175.9	150.1
Val4	-132.4	-17.5
Thr5	-112.1	161.9
Ser6	-90.5	171.0
Ala7	-70.4	-11.9

	ϕ/ψ	
Pro2	-48.7	152.3
Asp3	-89.4	31.9
Thr4	-80.4	172.9
Arg5	-115.1	167.9

	ϕ/ψ	
Pro2	-81.0	-170.6
Gly3	169.5	-165.0
Ser4	-172.8	166.0
Thr5	45.0	24.8
Ala6	-51.1	138.1
Pro7	-73.7	0.0

Figure 4.12. *Ensembles obtained from the 20 ns MD-tar simulations for glycopeptides **3**, **4** and **5**. The peptide backbone is shown in green and the sugar moiety in pink. The rmsd values for the peptide backbone, together with the peptide backbone geometry of the main conformation for each compound present in solution, are shown.*

As deduced from MD simulations, there are more conformers in solution (Figure 4.13). For glycopeptide **4**, we can observe an additional conformation

(29%) with a γ-turn between Asp3 and Thr4 residues, reported for other MUC1-like glycopeptides in aqueous solution.[5] The main conformer of derivative **4** appears to assist the presentation of the GalNAc residue to the target protein. We did not found significant interactions between the side chain of the Asp3 and the peptide backbone or the GalNAc unit. Besides, the electrostatic contact between Arg5 and Asp3 (distance Nε–Arg2/Cγ–Asp3 < 6.0 Å) is almost 30% of the total trajectory time.

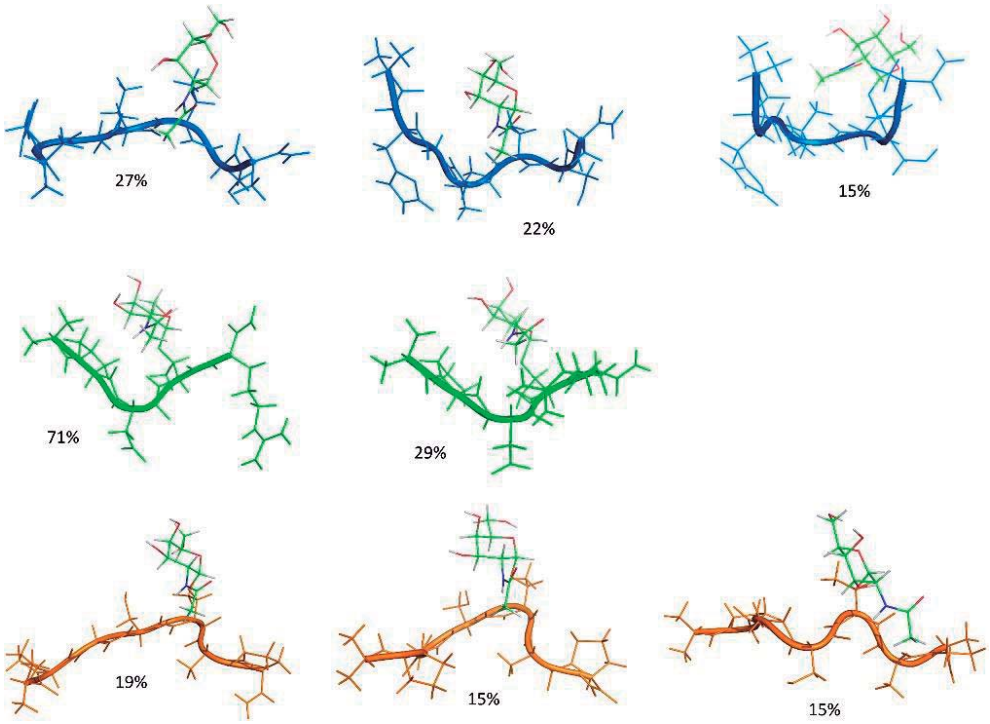


Figure 4.13. *Main conformation of free state glycopeptides in aqueous solution for glycopeptides **3** (upper panel), **4** (middle panel) and **5** (lower panel) obtained from the 20 ns MD-tar in explicit water.*

[5] a) U. Karsten, N. Serttas, H. Paulsen, A. Danielczyk, S. Goletz, *Glycobiology* **2004**, *14*, 681–692. b) T. Matsushita, N. Ohyabu, N. Fujitani, K. Naruchi, H. Shimizu, H. Hinou, S. I. Nishimura, *Biochemistry* **2013**, *52*, 402–414.

The geometry of the glycosidic linkage is equally comparable in the three derivatives. Actually, in all cases, ψ torsional angle takes values around 120º, characteristic of an 'eclipsed' conformation (Figure 4.14).[6]

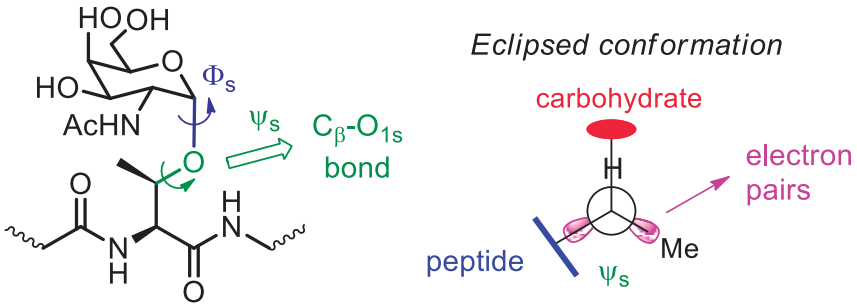


Figure 4.14. *Schematic representation of the glycosidic bond in glycopeptide **4** and its Newman projection of C_β-O_{1s} bond.*

This geometry is supported by a key NOE cross-peak between the NH group of the threonine residue and the NH of GalNAc (Figure 4.15).

[6] F. Corzana, J. H. Busto, G. Jiménez-Osés, M. García de Luis, J. L. Asensio, J. Jiménez-Barbero, J. M. Peregrina, A. Avenoza, *J. Am. Chem. Soc.* **2007**, *129*, 9458-9467.

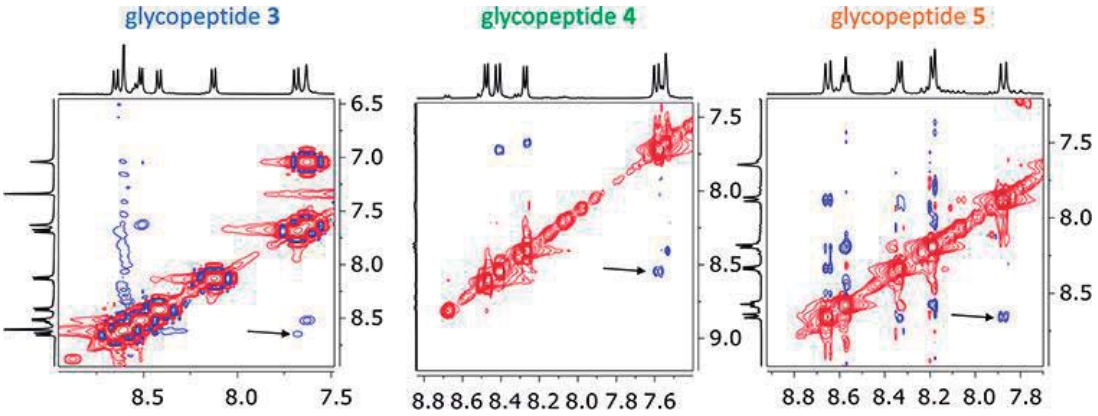


Figure 4.15. *2D-NOESY (H_2O/D_2O, 9:1, pH = 6.5, 293 K) spectra of glycopeptides **3**, **4** and **5** showing the cross-peak between NH of the GalNAc moiety and NH of the underlying Thr. Diagonal peaks and exchange cross-peaks connecting NH protons and water are negative (red color). The NOE contacts are represented as positive cross-peaks (blue color).*

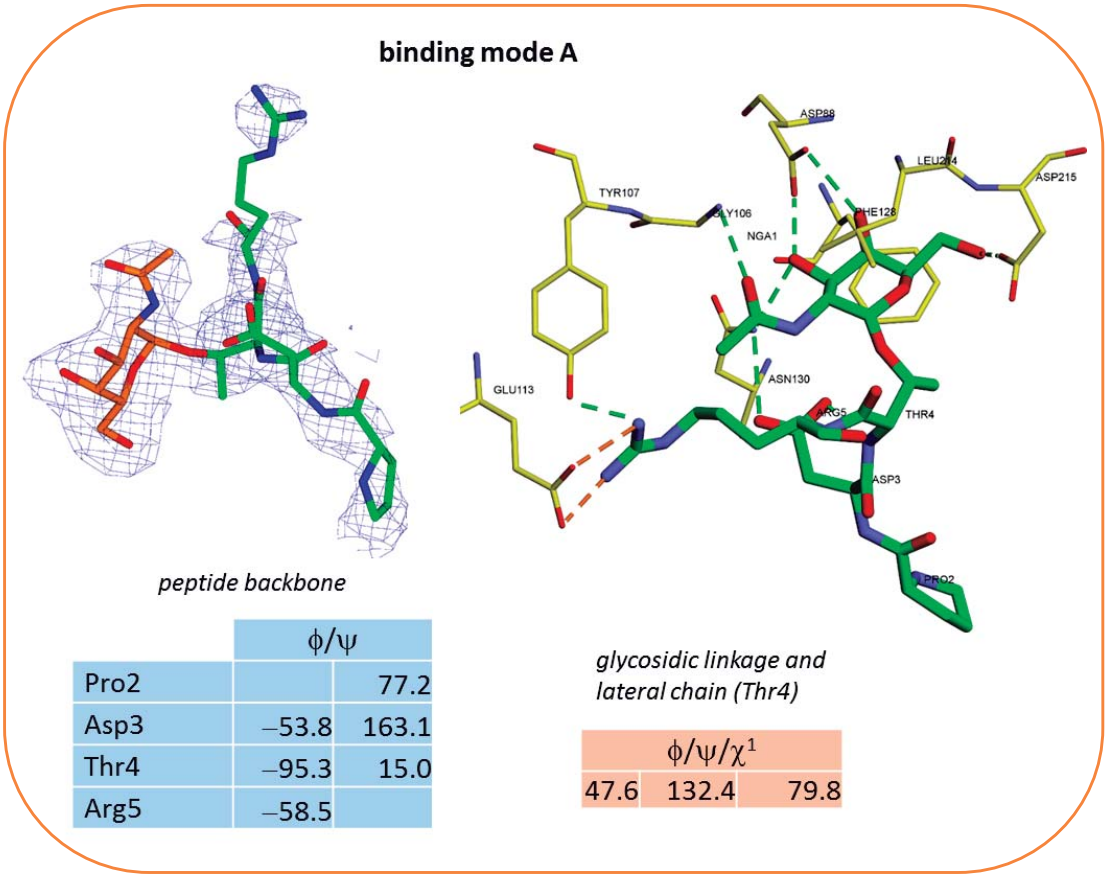
4.3.4. Conformational analysis in the bound state

X-ray analysis of SBA:**4** complex

Trying to rationalize those experimental results, we had performed a study of the interaction between the SBA lectin and glycopeptide **4** because it exhibited the highest affinity to the studied lectins.

As a first step, we tried to obtain the X-ray structure of the soybean agglutinin (SBA) lectin with derivative **4**. Fortunately, in collaboration with Dr. Ramón Hurtado-Guerrero of Institute of Biocomputation and Physics of Complex Systems (BIFI), we could obtain the X-ray structure of SBA:**4** complex, being the first time that a structure between SBA lectin and a Tn antigen derivative is solved and described.

Notably, twelve complexes form the cell unit with two different binding modes (A and B) for glycopeptide **4** complexed to SBA lectin (Figures 4.16, 4.17 and Table 4.5). The electronic density for Ala1 was not observed in any of the complexes, indicating that this residue is exposed to the solvent and shows high flexibility due to the lack of interactions with the lectin.



	ϕ/ψ	
Pro2		77.2
Asp3	–53.8	163.1
Thr4	–95.3	15.0
Arg5	–58.5	

ϕ/ψ/χ^1		
47.6	132.4	79.8

Figure 4.16. *X-ray crystal structure of SBA lectin in complex with compound* ***4*** *(binding mode A) and its schematic representation of the hydrogen bond network. Dihedral angles (ϕ/Ψ) distribution for the backbone, glycosidic linkage and side chain (χ^1) of glycopeptide are also shown.*

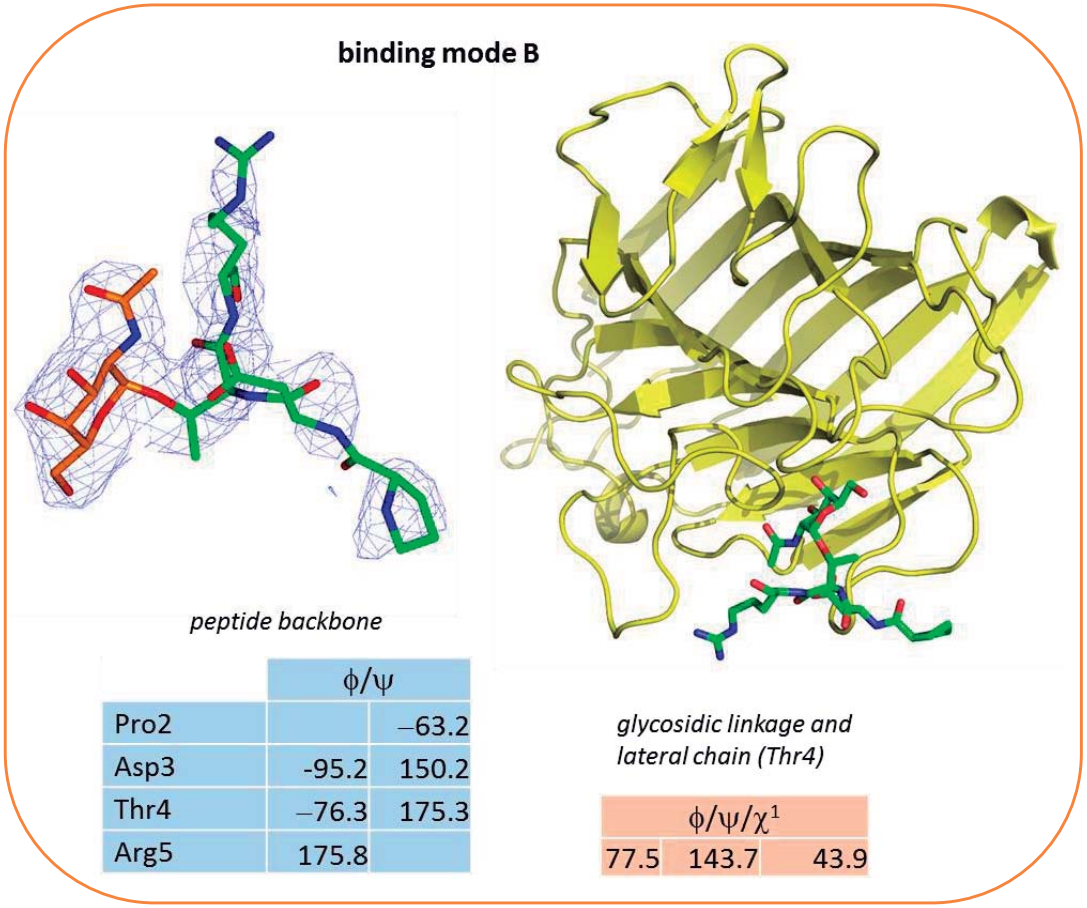


Figure 4.17. *X-ray crystal structure of SBA lectin in complex with compound* ***4*** *(binding mode B) and its schematic representation of the hydrogen bond network. Dihedral angles (ϕ/Ψ) distribution for the backbone, glycosidic linkage and side chain (χ¹) of glycopeptide are also shown.*

Although the 3D dispositions of glycopeptide **4** in binding modes A and B displays almost the same geometry in terms of the glycosidic linkage, the most significant changes are related to the peptide backbone conformation, principally with Thr4 residue.

Table 4.5. *Crystal data and data-collection statistics for SBA:glycopeptide* ***4*** *complex. Values in parentheses refer to the highest resolution shell.*

Space group	$P3_2$	RMSD from ideal geometry, bonds (Å)	0.011
Wavelength (Å)	0.976	RMSD from ideal geometry, angles (°)	1.590
Resolution (Å)	20-2.70 (2.85-2.70)	<*B*> protein (Å^2)	55.76
Cell dimensions (Å)	*a* =114.23 *b*=114.23 *c* =202.89	<*B*> APD-T(GalNAc)-R	59.83
Unique reflections	81148	<*B*> solvent (Å^2)	26.91
Completeness	99.8 (100)	<*B*> Mn^{2+} (Å^2)	58.25
R_{sym}	0.122 (0.739)	Ramachandran plot:	
I/σ(I)	10.8 (2.3)	Most favoured (%)	94.27
Redundancy	4.8 (4.3)	Additionally allowed (%)	3.55
R_{work}/R_{free}	0.230/0.274	Disallowed (%)	2.18

As in solution (Figure 4.14) the ψ torsional angle of the glycosidic linkage of glycopeptide **4** adopts the expected 'eclipsed' conformation in the bound state in both binding modes, with values around 120°. Consequently, the sugar moiety has an almost perpendicular arrangement with respect to the peptide backbone (Figure 4.18).

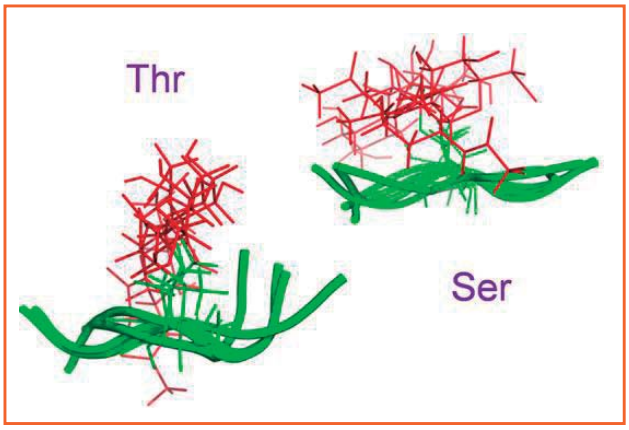


Figure 4.18. *Difference in the orientation of the carbohydrate moiety when the amino acid is a serine or a threonine (see Chapter 3).*

In contrast, while Thr4 adopts a γ-turn-like conformation in binding mode A, it displays an extended geometry in mode B. Nevertheless, the overall shape of both conformations of glycopeptide **4** bound to the lectin are rather similar with a rmsd value for the heavy atoms of 0.94 Å (Figure 4.19). Interestingly, these conformations for glycopeptide **4** are also found in solution.

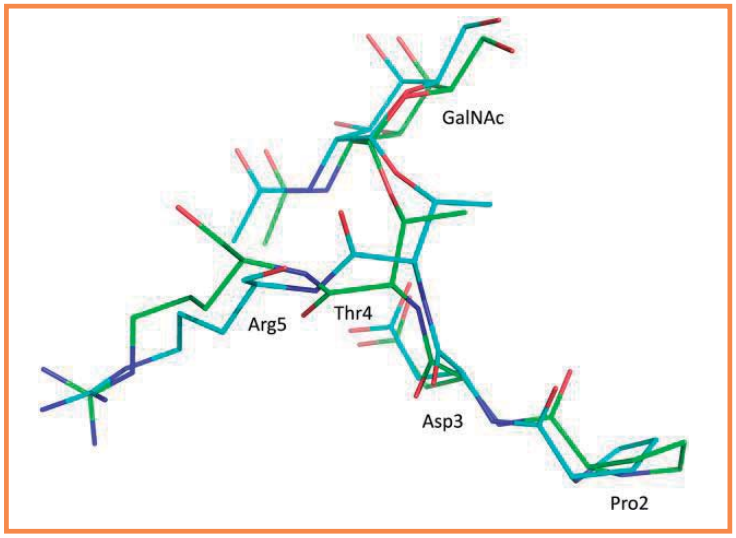


Figure 4.19. *Superposition of the two different conformation found in the X-ray for glycopeptide **4** with the SBA lectin (binding mode A in green and B in blue).*

The analysis of X-ray structures reveals that the interaction pattern between glycopeptide **4** in binding mode A also differs from that observed in binding mode B. If we take a look at mode A, the hydroxyl groups of the GalNAc residue interact with the lectin through different hydrogen bonds. In particular, O3 and O4 form hydrogen bonds with the side chain of Asp88. Moreover, O3 participates in a hydrogen bond with Asn130 and O4 with Leu214. The GalNAc unit establishes an extra hydrogen bond between O6 and Asp251 (Figure 4.20). Besides, the carbonyl group of the *N*-acetyl moiety of GalNAc interacts with the NH of Gly106. This latter interaction may explain the higher affinity of SBA lectin for GalNAc in comparison to galactose.[7] A CH-π contact between the α-face of the sugar and Phe128 is also established (Figure 4.20). Calculated interactions between the sugar and the lectin published in the literature are in agreement with those observed in our crystal structure.[8]

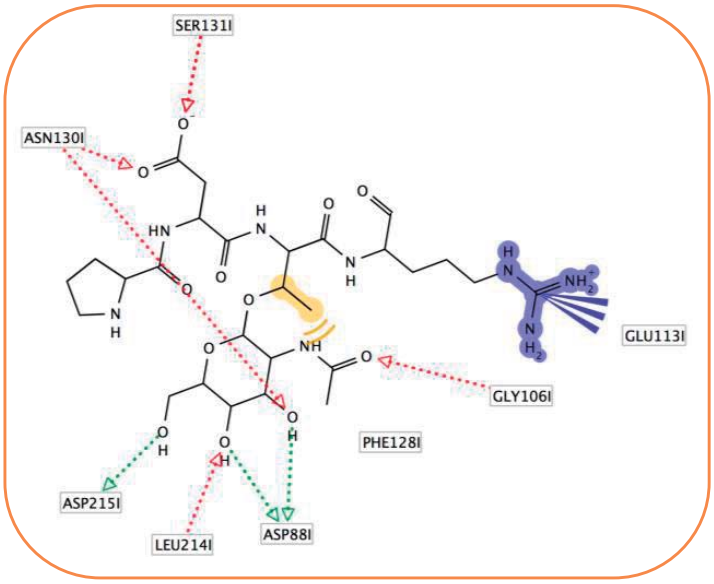


Figure 4.20. *Schematic representation of interactions in SBA:**4** complex*

[7] a) M. E. Pereira, E. A. Kabat, N. Sharon, *Carbohydr. Res.* **1974**, *37*, 89-102; b) H. De Boeck, F. G. Loontiens, H. Lis, N. Sharon, *Arch. Biochem. Biophys.* **1984**, *234*, 297-304.

[8] V. S. R. Rao, K. Lam, P. K. Qasba, *J. Biomol. Struct. Dyn.* **1998**, *15*, 853-860.

A close inspection the interactions between the lectin and the glycopeptide **4**, shows evidences that the methyl group of Thr4 is involved in a hydrophobic pocket with Phe128. This contact may be responsible for the higher affinity observed for SBA lectin to Tn antigen when the underlying amino acid is a threonine residue.[3a] Furthermore, Arg5 participates in a salt-bridge with Glu113 and a hydrogen bond with the hydroxyl of Tyr107. On the other hand, the side chain of Asp3 forms a hydrogen bond with the side chain of Asn130. This aspartic acid residue holds an effective role by constraining the conformation of both the peptide backbone and the sugar moiety and favoring the appropriate presentation of the sugar to the protein receptor (Figure 4.21).

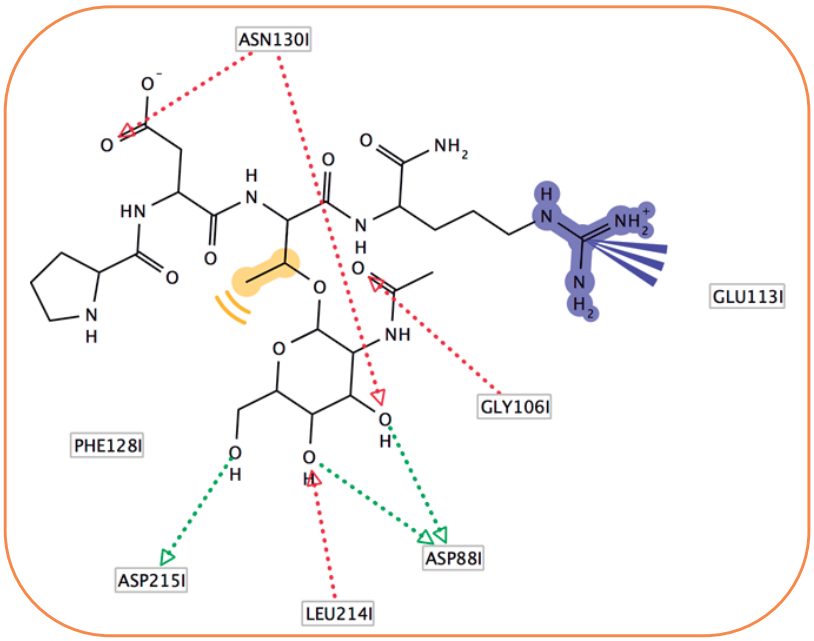


Figure 4.21. *Schematic representation of more interactions in SBA:**4** complex*

In contrast, the interactions of glycopeptide **4** with Ser131 and Tyr107 are negligent in binding mode B. Interestingly, in both binding modes, Asp3 participates in one hydrogen bond with the *N*-acetyl group of GalNAc and in a

second one with the NH group of Thr4 (Figure 4.22a). In addition, Asp3 is involved in an electrostatic interaction with the side chain of Arg5. Of note the oxygen of the carboxylic group of the side chain of Asp3 mimics the role of a bridging water molecule suggested by us between the sugar and the peptide moieties in solution for Tn derivatives (Figure 4.22b).[3]

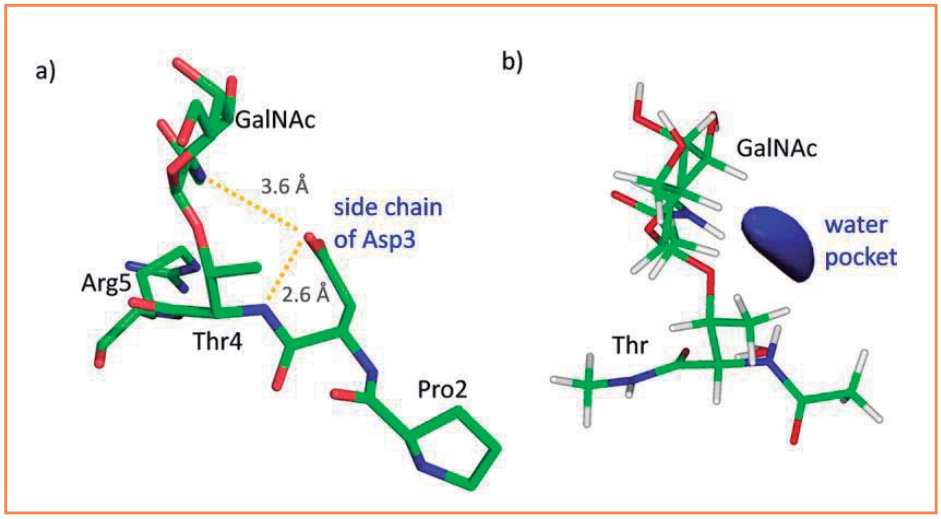


Figure 4.22. *a) Hydrogen bond network found in the X-ray structure for Asp3 in glycopeptide **4** with the sugar and peptide moieties in both (A and B) binding modes. b) Inter-residue water pocket deduced from the MD-tar simulations on Ac-L-Thr(α-D-GalNAc)-NHMe.*

With the aim of testing the importance of Asp3 and Arg5 in the molecular recognition process by SBA lectin, we synthesized the glycopeptide **4'** (Figure 4.23) following the SPPS protocol described in *Chapter 2*. In this new derivative, we replaced Asp3 and Arg5 by Ala residues. As we have predicted, this novel derivative has a low affinity to SBA, validating the hypothesis that these two amino acids fix the conformation of both the peptide and the sugar to maximize the interactions with the protein (Figure 4.23).

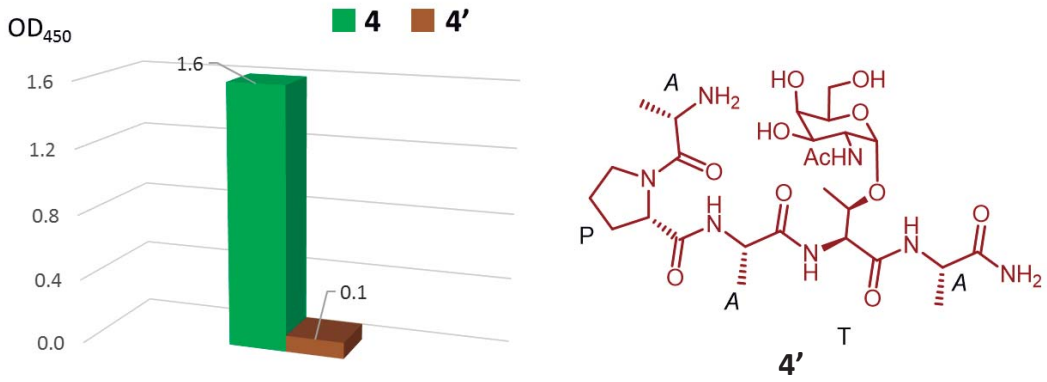


Figure 4.23. *Binding studies of glycopeptides **4** and **4'** with SBA lectin using ELLA tests. Absorbance signals are the average of three replicate wells and the error bars show the standard deviations for these measurements.*

Epitope mapping of glycopeptide **4**

To determine whether the interactions observed in our X-ray structure were maintained in solution, we have also determined the epitope mapping of glycopeptide **4** complexed to SBA by (STD) NMR experiments in collaboration with Dr. Filipa Marcelo from Universidade Nova de Lisboa and Prof. Jesús Jiménez-Barbero from CSIC (Figure 4.24).

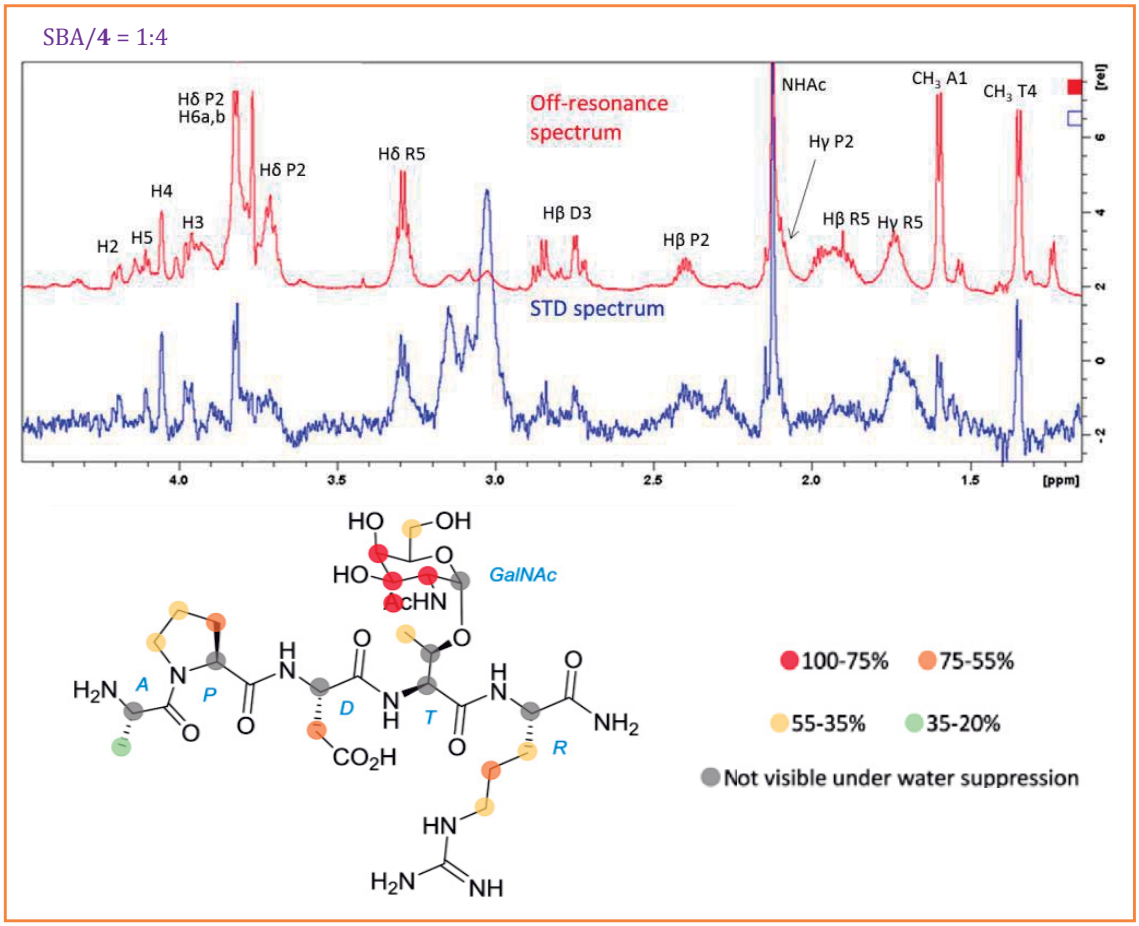


Figure 4.24. *Epitope mapping of glycopeptide **4** complexed to SBA lectin and their STD (blue) and off-resonance spectrum (red). The key proton resonances are showed in the figure. The red values represent the epitope mapping and the back ones are the absolute STD values. (^{1}H NMR 600 MHz with cryoprobe, pH = 7.2, 293 K, molar ratio lectin/glycopeptide 1:4).*

The (STD) NMR experiments showed that the main epitope of glycopeptide **4** is located at the sugar moiety with strong STD signals for H2, H3, H4 and the *N*-acetyl group of GalNAc. Surprisingly, these results showed a clear interaction between the lectin and the peptide backbone. As an example, the methyl group of Thr4, Pro2 ring and side chain of Arg5 are in closer contact with the protein.

Besides, STD experiments are compatible with both binding modes observed in the X-ray structure for glycopeptide **4** (Figures 4.16 and 4.17). The weak STD signal detected for the methyl group of Ala1 (<30% in absolute terms) indicates that this reside should be solvent exposed and explains the poor electronic density found in the X-ray structure for Ala1. This feature is also corroborated by the titrations experiments carry out on a solution of SBA lectin with derivative **4**.

On the other hand, it is also known that in the field of ligand-protein systems, displaying an intermediate exchange ratio (K_D = 6.6±1.9 mM, as for glycopeptide **4**) in the chemical shift scale, the observed resonance frequencies of the signals do not shift when the concentration of the ligand is varied; however, they experience a significant loss in the signal intensity, due to line broadening.[9] Thus, protons that preferentially bind to the protein-binding pocket show a significant decrease of their signal intensities, relative to those far from the receptor recognition site.

Besides, qualitatively our ^{1}H NMR-monitored titration of glycopeptide **4** with SBA (4:1, 4:2, 4:3 1:1 of SBA:**4** ratio) shows that the signal intensity of the CH_3 protons of Ala1 increase slightly faster than that observed for CH_3 protons of Thr4, supporting the idea that methyl group of Thr4 exhibits a stronger interaction with the protein than Ala1 (Figure 4.25).

[9] L. Garrido, N. Beckmann, W. Price, *New Applications of NMR in Drug Discovery and Development*, Royal Society of Chemistry, **2013**.

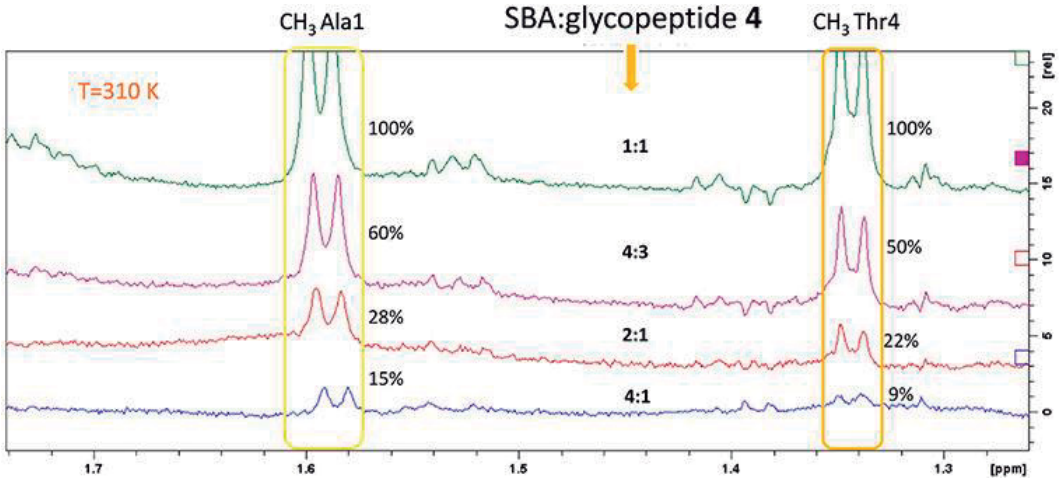


Figure 4.25. *^{1}H NMR spectra of titration experiments performed with a solution of SBA lectin (50 μM) and increasing amounts of glycopeptide* ***4****. The experiments were carried out at 310K and at pH = 7.2 (phosphate buffer).*

MD simulations on SBA:**4** complex

In order to obtain information about dihedral angles of backbone, glycosidic linkage and side chain of the SBA:**4** complex in solution and to compare them to those obtained in solid state, we have analyzed this complex by MD using as starting coordinates data from both X-ray structures: binding mode A (Figure 4.16) and binding mode B (Figure 4.17).

As expected, the main epitope of this glycopeptide was mostly located at the sugar moiety. This result is in agreement with that observed in solid state. Simulations also predicted that both binding modes are stable in solution (Figures 4.26 and 4.27).

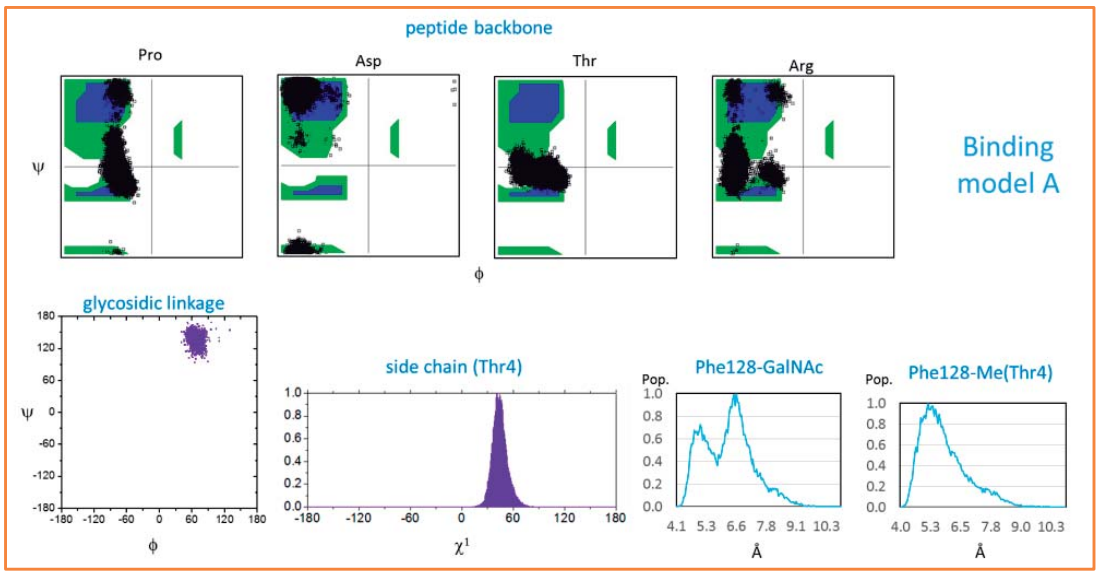


Figure 4.26. *Dihedral angles (ϕ/Ψ) distribution for the backbone, glycosidic linkage and side chain (χ^1) of glycopeptide* ***4*** *bound to SBA lectin obtained from the unrestrained 100 ns MD-tar simulations performed in explicit water (binding mode A).*

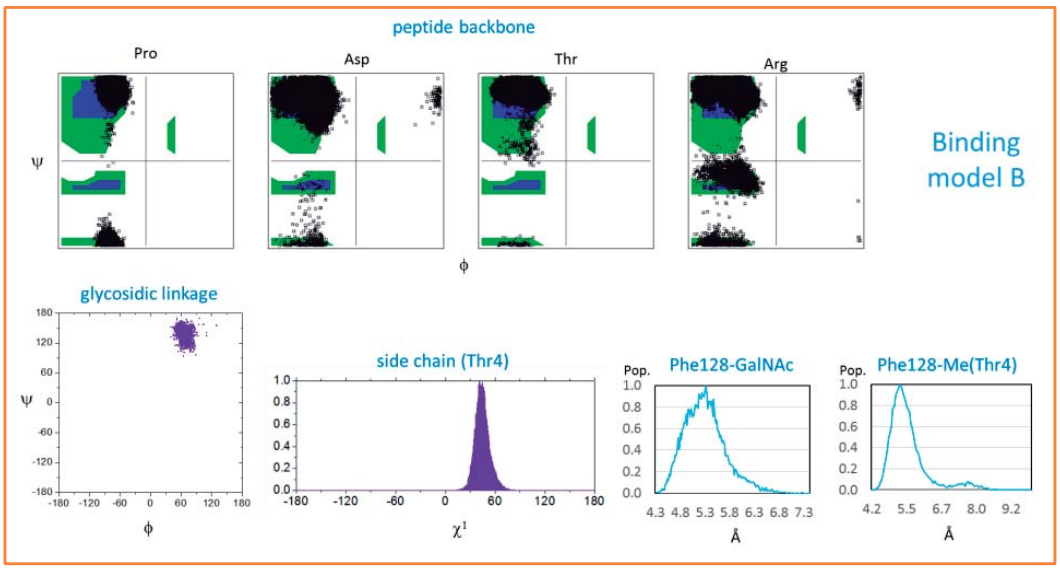


Figure 4.27. *Dihedral angles (ϕ/Ψ) distribution for the backbone, glycosidic linkage and side chain (χ¹) of glycopeptide* ***4*** *bound to SBA lectin obtained from the unrestrained 100 ns MD-tar simulations performed in explicit water (binding mode B).*

The interactions between the sugar and the protein are the same than in the solid state (Figure 4.28), apart from Leu214 that prefers to interact with O6, and are also in good agreement with the STD experiments on SBA:**4** complex commented above (Figure 4.24). The CH-π contact between α-face of GalNAc and Phe128 is also present. Moreover, the simulations predict the characteristic eclipsed conformation of the glycosidic linkage (Figures 4.26 and 4.27). In both simulations, the peptide backbone shows more flexibility at the *N*-terminus region, especially for Ala1 residue. This feature is in good agreement with the lack of electronic density for this residue in the solid state and with the rather weak STD signal observed for Ala1. MD simulations indicate that Pro2 establishes a hydrophobic interaction with the side chain of Arg129. This interaction, not observed in the X-ray structure, tentatively explains the STD effect found for this Pro residue (Figure 4.28).

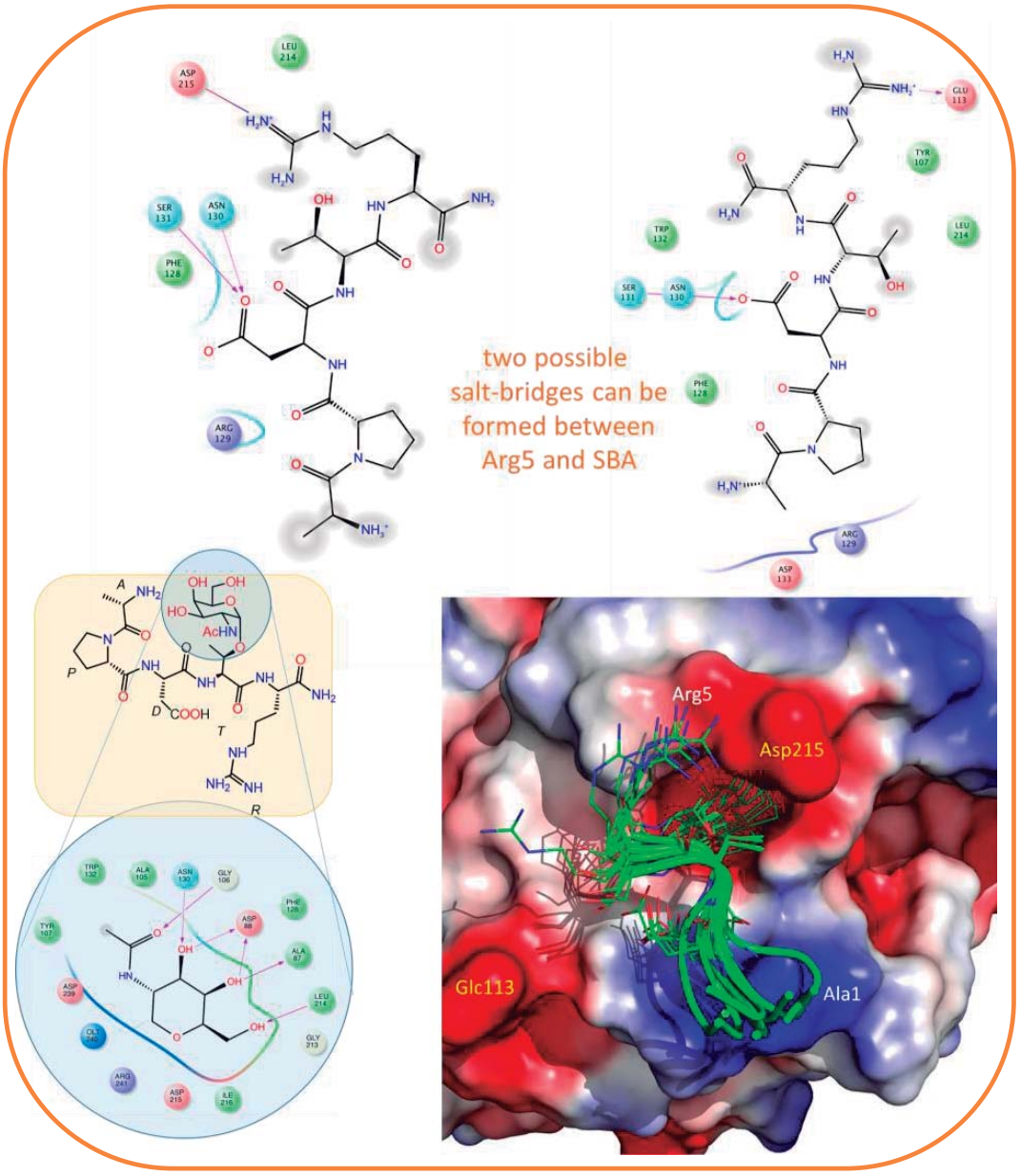


Figure 4.28. *Hydrogen bonds between GalNAc moiety and SBA lectin and peptide interaction diagrams obtained from a representative frame of the unrestrained 100 ns MD simulations carried out on SBA:**4** complex.*

MD simulation starting from binding mode A shows that the hydrogen bond between Asp3 and Asn130 of SBA lectin is also significant, with a population around 46%. An additional hydrogen bond, populated about 40%, is observed between Asp3 and Ser131 of the protein. The hydrogen bonds Asp3-Thr4 and Asp3-GalNAc (Figure 4.22) are only occasionally populated (close to 15% of the total trajectory time).

Besides, the salt bridge Arg5-Glu113, observed in the X-ray structure, has a population very low (around 2%). According to our model, Arg5 establishes also an electrostatic interaction with Asp215 (not observed in the X-ray structure) that is present about 25% of the total trajectory time (Figure 4.28). Therefore, the side chain of Arg5 is fairly flexible and can interact with different residues of the protein (Figure 4.28). Similarly, the interactions observed in the X-ray structure, when binding mode B is present, are also retained in MD simulation (Table 4.6).

Table 4.6. *List of hydrogen bonds (% population) between α-O-GalNAc moiety and SBA (in SBA:**4** complex) obtained from unrestrained 100 ns MD simulations in explicit water.*

	% population
Asp88-O3	93.5
Asp88-O4	88.0
Asp215-O6	13.2
Gly106-CO (GalNAc)	30.1
Asn130-O3	75.6

Study of the SBA:**3** and SBA:**5** complexes

The interaction features of glycopeptides **3** and **5** complexed to SBA lectin were also investigated in solution combining (STD) NMR experiments with unrestrained MD simulations. As in glycopeptide **4**, the main epitope of these glycopeptides is mostly located at the sugar moiety (Figures 4.29 and 4.30). Actually, STD experiments suggest a slightly different binding mode for the sugar unit in both glycopeptides.

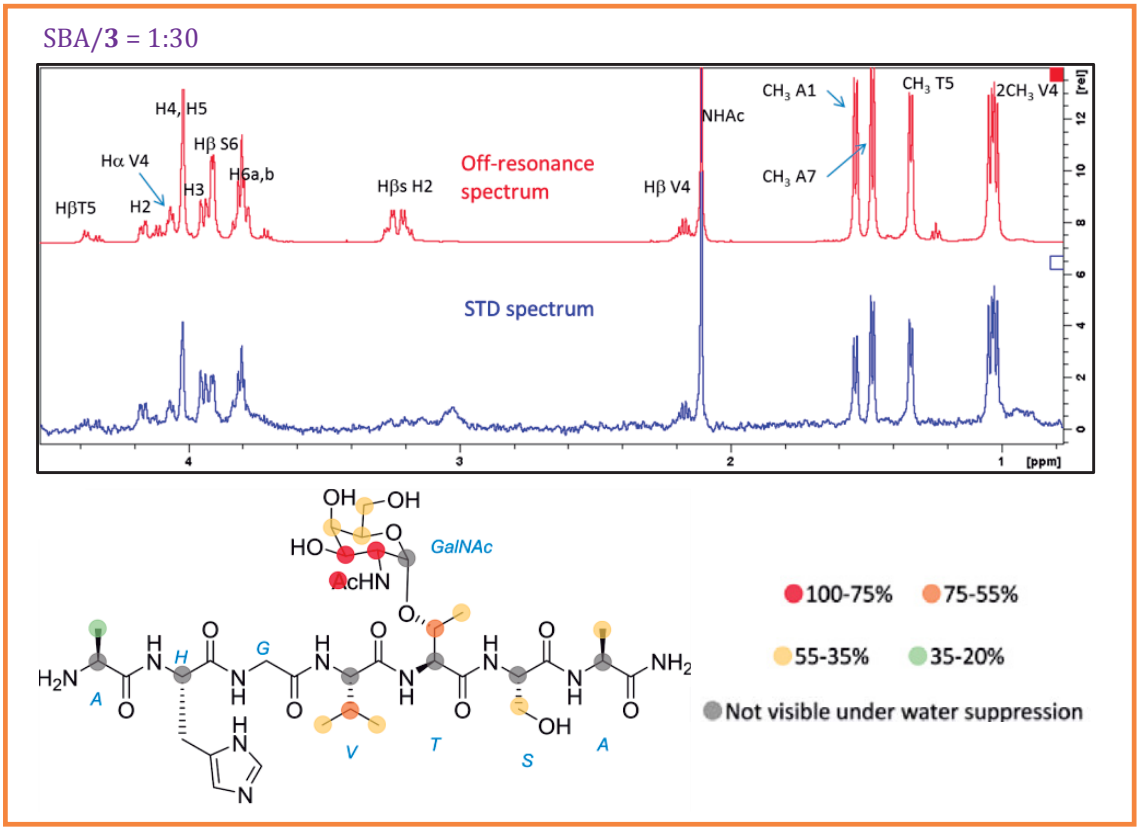


Figure 4.29. *Epitope mapping of glycopeptide **3** complexed to SBA lectin and their STD (blue) and off-resonance spectrum (red). The key proton resonances are showed in the figure. The red values represent the epitope mapping and the back ones are the absolute STD values. (^{1}H NMR 600 MHz with cryoprobe, pH = 7.2, 293 K, molar ratio lectin/glycopeptide 1:30).*

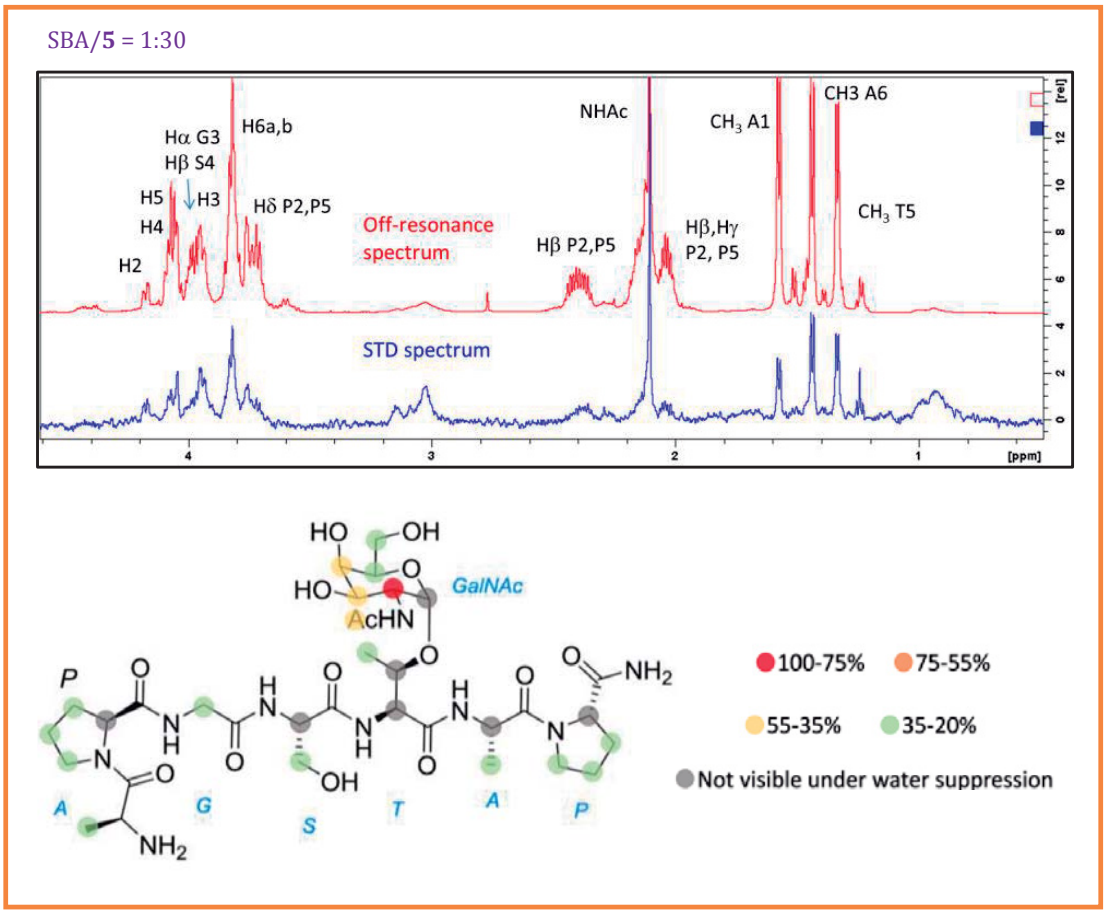


Figure 4.30. *Epitope mapping of glycopeptide **5** complexed to SBA lectin and their STD (blue) and off-resonance spectrum (red). The key proton resonances are showed in the figure. The red values represent the epitope mapping and the back ones are the absolute STD values. (^{1}H NMR 600 MHz with cryoprobe, pH = 7.2, 293 K, molar ratio lectin/glycopeptide 1:30).*

Indeed, the MD simulations indicated that the interactions between glycopeptides **3** and **5** with the surface of SBA are limited to the sugar unit and the glycosylated Thr residue (Figure 4.31). This feature may explain the better affinity and the enthalpy values measured by ITC experiments of glycopeptide **4** for SBA in contrast to low values for **3** and **5** (Figure 4.6).

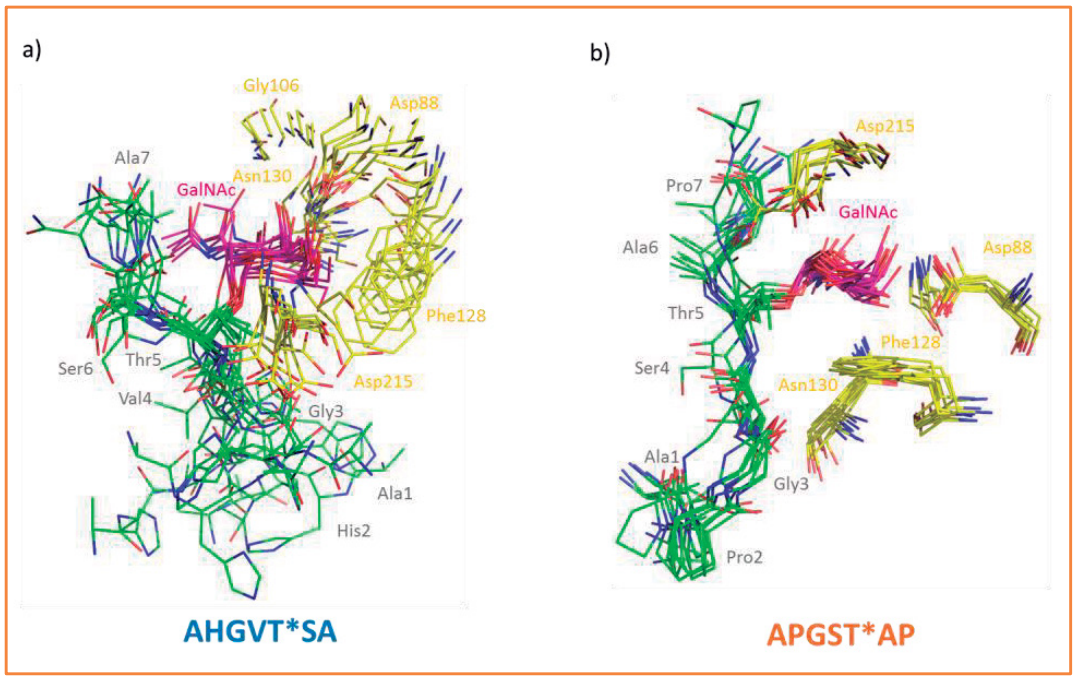


Figure 4.31. *Ensembles obtained from the unrestrained 100 ns MD simulations performed on SBA:**3** (a) and SBA:**5** (b) complexes. Carbon atoms of the protein are in yellow and carbon atoms of GalNAc moiety are in hot pink.*

We also obtained the hydrogen bonds present in both structures. The values are summarized in Table 4.7.

Table 4.7. *List of hydrogen bonds (% population) between α-O-GalNAc moiety and SBA in SBA:**3** and SBA:**5** complexes obtained from unrestrained 100 ns MD simulations in explicit water.*

	%population (SBA:**3**)	%population (SBA:**5**)
Asp88-O3	47.3	97
Asp88-O4	99.8	100
Asp215-O6	7	74.2
Gly106-CO (GalNAc)	6	46
Asn130-O3	22.2	56.8

Additionally, the MD simulations carried out on SBA:**3** and SBA:**5** clearly indicate a higher degree of flexibility for compound **3** when complexed to SBA (Figure 4.31). This fact is in agreement with the poor affinity showed by the lectin for this glycopeptide. Although merely speculative, the presence of the His residue (10% positive charged at pH = 7) could play against binding with SBA lectin due to the proximity of this residue to a positive surface potential at the protein (Figure 4.32).

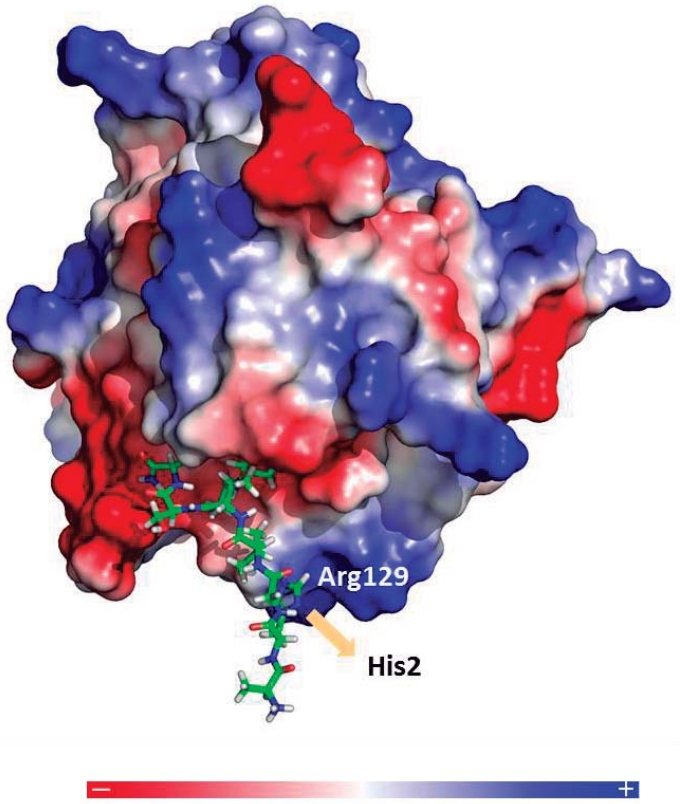


Figure 4.32. *Representative frame obtained from the unrestrained 100 ns MD simulations on SBA:**3**. Red color represents a net negative surface potential while a blue color denotes a net positive surface potential. Glycopeptide **3** is shown as a stick model colored according to atom type.*

On the other hand, for compound **5**, a type II β-turn that comprises the carbonyl group of Ala1 and the NH group of Ser4 may stabilize the peptide backbone in the bound state. This backbone conformation allows the Hα of Ser4 to establish a hydrophobic interaction with Phe128 (Figure 4.31b).

If we take a look at dihedral angles (Φ and Ψ) for glycosidic linkage and the side chain (χ^1) for both glycopeptides **3** (Figure 4.33) and **5** (Figure 4.34) we can observe mostly rigidity values for those terms. For the backbone, the flexibility is higher in derivative **3** than in glycopeptide **5**, as we had observed experimentally. The hydrophobic patch with Phe128 from the SBA is also conserved in both models.

These features, among others factors, may explain the better affinity of SBA for this compound when compared with derivative **3**.

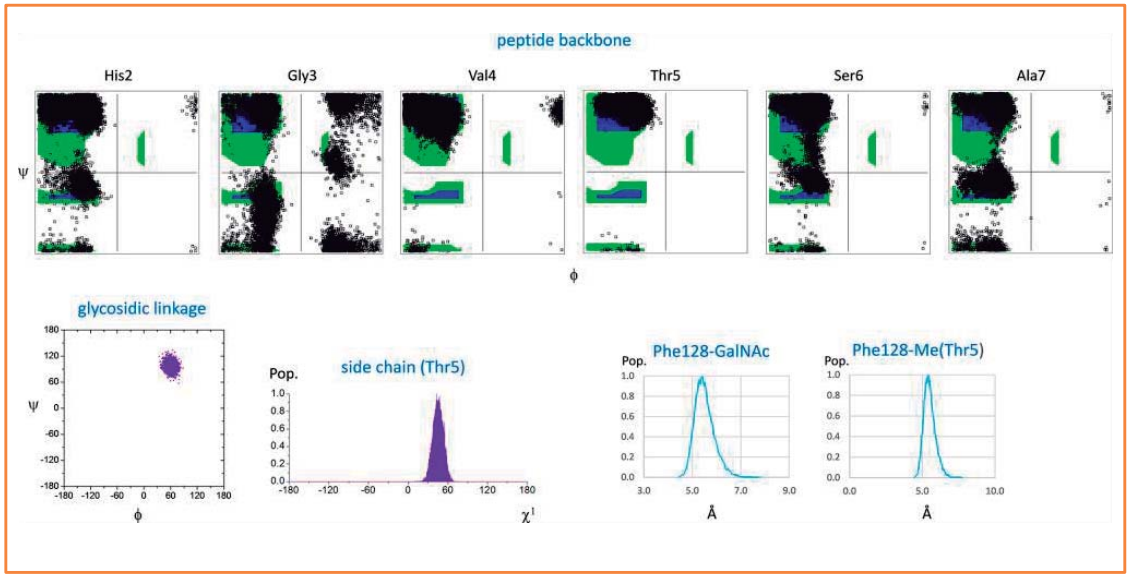


Figure 4.33. *Dihedral angles (Φ/Ψ) distribution of the peptide backbone obtained from the 100 ns MD simulations in explicit water carried out on SBA:**3** complex, together with the geometry of the glycosidic linkage and the side chain of Thr5.*

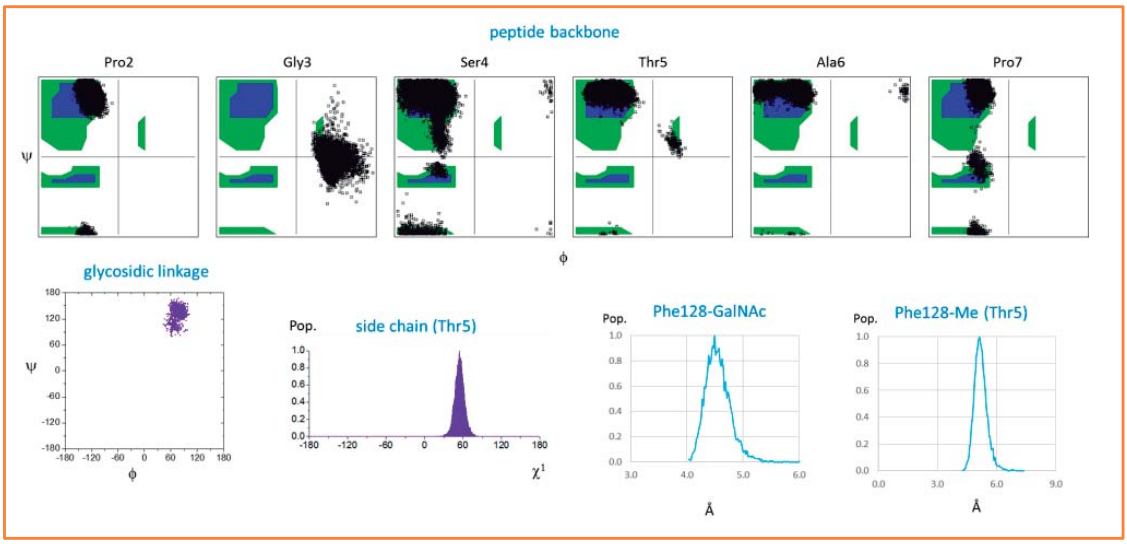


Figure 4.34. *Dihedral angles (Φ/Ψ) distribution of the peptide backbone obtained from the 100 ns MD simulations in explicit water carried out on SBA:**5** complex, together with the geometry of the glycosidic linkage and the side chain of Thr4.*

4.4. Conclusions

The work compiled in this chapter provides remarkable results for the recognition of MUC1-derived peptides by a model lectin, in terms of epitope selection. Although it is well known that lectins recognize sugar moieties, we have shown herein that the recognition process is modulated by the peptide sequence close to the glycosylation point.

In particular, SBA lectin shows higher affinity to the Tn antigen when it is placed at the PDTR peptide region. In contrast, an affinity loss is observed when this antigen is located at the AHGVTSA or GVTSA fragments. Both Asp and Arg residues in the PDTR sequence appear to fix the conformation of both the peptide and the sugar moieties to maximize the interactions with the lectins.

We have also obtained an X-ray structure by first time between the soybean lectin and a MUC1 derivative.

These results could help to better understand the interactions between glycopeptides and lectins contributing to engineer new binding sites and, therefore, new lectins that bind to glycopeptides with higher affinity. Moreover, these features could allow tuning of specificity and the design of novel glycosensors with enhanced detection of Tn antigen.

This work has been published as an article entitled *Detection of Tumor-Associated Glycopeptides by Lectins: The Peptide Context Modulates Carbohydrate Recognition* in the journal *ACS Chemical Biology*, DOI: 10.1021/cb500855x.

CONCLUSIONS

CONCLUSIONES

Several conclusions can be drawn from the work described in this Thesis:

- We have synthesized by an improved SPPS methodology different mucin-like glycopeptides derived from MUC1. This enhanced approach involves the manually coupling of the glycosylamino acids, instead of automated mode, reducing the required equivalents of our building block and getting higher yields.

- We have observed differences between αGalNAc glycopeptides bearing threonine (**1**) or serine (**2**) in their structure. This work reveals that it is important to mention unambiguously the underlying amino acid present in Tn antigen (α-*O*-GalNAc-Thr/Ser) because the recognition by different lectins (SBA, VVA and HPA) reveals that these residues (Thr or Ser) are not equivalent. While soybean and *Vicia villosa* lectins have a preference for the antigen carrying Thr, *Helix pomatia* agglutinin binds better to Tn bearing Ser in its structure. We have also analyzed in detail the interactions that allow the molecular recognition process of MUC1-derivatives by these three lectins. The study indicates that the methyl group of Thr, besides other effects, can modulate recognition.

- We have synthesized and analyzed the three important epitopes of MUC1 protein (**3**, **4** and **5**). This work provides outstanding results for the recognition of these derivatives by SBA and VVA lectins, in terms of epitope selection. Although it is well known that lectins recognize sugar moieties, we have shown herein that the recognition process is modulated by the peptide sequence close to the glycosylation point. In particular, the studied lectin receptors show higher affinity to the Tn antigen when it is placed at the APDTR peptide region. In contrast, an affinity loss is observed when this

antigen is located at the AHGVTSA or APGSTAP fragments. Both Asp and Arg residues in the APDTR sequence appear to fix the conformation of both the peptide and the sugar moieties to maximize the interactions with the lectins.

- The aforementioned outcomes were deduced by a multidisciplinary approach that involves the use of different techniques such as, NMR experiments, ELLAs, ITCs, MD, X-ray diffraction...

- The results compiled in this Thesis could help to better understand the glycopeptide lectin interactions. This information could be used in order to design new binding sites in lectins with higher affinity to MUC1 glycopeptides bearing Tn antigen, allowing the tuning of specificity and the design of novel glycosensors with enhanced detection of Tn antigen in cancer therapy, e.g.

Scientific Publications derived from this Thesis:

Serine versus Threonine Glycosylation with α-O-GalNAc: Implications for the Molecular Recognition by Lectins. D. Madariaga, N. Martínez-Sáez, V. J. Somovilla, L. García-García, A. Berbis, R. Hurtado-Guerrero, J. Valero-Gónzalez, J. L. Asensio, J. Jiménez-Barbero, J. H. Busto, A. Avenoza, F. Corzana, J. M. Peregrina *Chemistry - A European Journal* **2014**, *20*, 12616-12627.

Detection of Tumor-Associated Glycopeptides by Lectins: the Peptide Context Modulates Carbohydrate Recognition. D. Madariaga, N. Martínez-Sáez, V. J. Somovilla, H. Coelho, J. Valero-Gónzalez, J. Castro-López, J. L. Asensio, J. Jiménez-Barbero, J. H. Busto, A. Avenoza, F. Marcelo, R. Hurtado-Guerrero, F. Corzana, J. M. Peregrina *ACS Chemical Biology* **2014**, DOI: 10.1021/cb500855x.

Contribution to Congress derived from this Thesis:

Flash and poster: *Engineering O-Glycosylation Points in Short Tumor-Associated Glycopeptides Derived from Mucins.* D. Madariaga, A. Avenoza, J. H. Busto, F. Corzana, J. M. Peregrina. *I Meeting of the RSEQ Chemical Biology Group*. Santiago de Compostela (Spain). March, 7-8 of 2012.

Poster: *Lectins and mucins: Key elements responsible for molecular recognition.* D. Madariaga, N. Martínez-Sáez, A. Avenoza, J. H. Busto, F. Corzana, J. M. Peregrina. *26th International Carbohydrate Symposium*. Madrid (Spain). July, 22-26 of 2012.

Oral communication: *Insights into Molecular Recognition of Tumour-Associated Glycopeptides by Lectins.* F. Corzana, D. Madariaga, N. Martínez-Sáez, V. J. Somovilla, S. Martín-Santamaría, J. L. Asensio, A. Berbis, J. Jiménez-Barbero, A. Avenoza, J. H. Busto, J. M. Peregrina. *II Meeting of the RSEQ Chemical Biology Group*. Bilbao (Spain). February, 4-5 of 2014.

Oral communication and poster: *Evaluating the Underlying Information in Lectin-Mucins Recognition.* D. Madariaga, A. Avenoza, J. H. Busto, F. Corzana, J. M. Peregrina. *San Diego Glycobiology Symposium*. San Diego (USA). March, 20-22 of 2014.

Flash: *Serine versus threonine glycosylation with α-O-GalNAc: Implications for the molecular recognition.* D. Madariaga, S. Martín-Santamaría, J. L. Asensio, J. Jiménez-Barbero, M. A. Berbis, A. Avenoza, J. H. Busto, F. Corzana y J. M. Peregrina. *XI Carbohydrate Symposium*. Logroño (Spain). May, 28-30 of 2014.

Oral communication: *Glicosilación de Serina versus Treonina con α-O-GalNAc: Implicaciones en el reconocimiento molecular.* D. Madariaga, A. Avenoza, J. H. Busto, F. Corzana y J. M. Peregrina. *V Jornada de Química CISQ.* Logroño (Spain). June, 20 of 2014.

Poster: *Epitope mapping of different MUC1-like glycopeptides with Soybean lectin by Saturation-Transfer Difference (STD)-NMR experiments and Molecular Dynamics simulations.* D. Madariaga, H. Coelho, F. Marcelo, J. L. Asensio, J. Jiménez-Barbero A. Avenoza, J. H. Busto, F. Corzana y J. M. Peregrina. *VII Biennial Meeting of the Nuclear Magnetic Resonance Group of the Spanish Royal Society of Chemistry / IV Iberian NMR Meeting / VI Ibero-American NMR Meeting.* Alcalá de Henares (Spain). September, 22-24 of 2014

Poster: *Recognition of tumor associated glycopeptides by a potential sensor lectin.* D. Madariaga, A. Avenoza, J. H. Busto, F. Corzana y J. M. Peregrina. *XI Simposio de Investigadores Jóvenes RSEQ.* Bilbao (Spain). November, 4-7 of 2014.

Las conclusiones que pueden extraerse del trabajo desarrollado en esta Tesis se resumen a continuación:

- Hemos sintetizado mediante una metodología mejorada de SPPS diferentes glicopéptidos tipo mucina derivados de MUC1. Esta técnica mejorada consiste en la realización del acoplamiento de los glicosil-aminoácidos de forma manual en lugar de automáticamente, reduciendo así el número de equivalentes necesarios de nuestros *building blocks* y obteniendo mayores rendimientos.

- Hemos observado diferencias entre glicopéptidos que incorporan αGalNAc unido a treonina (**1**) o serina (**2**) en su estructura. Este trabajo revela la importancia de mencionar siempre el aminoácido al que está unido el carbohidrato presente en el antígeno Tn (α-*O*-GalNAc-Thr/Ser) ya que el reconocimiento por diferentes lectinas (SBA, VVA y HPA) revela que el comportamiento de estos residuos (Thr o Ser) no son equivalentes. Mientras que las lectinas *soybean* y *Vicia villosa* tienen preferencia por el antígeno Tn unido a Thr, la lectina *Helix pomatia* prefiere el antígeno Tn con Ser en su estructura. También hemos analizado en detalle las interacciones que permiten el proceso de reconocimiento molecular de derivados de MUC1 por estas lectinas. El estudio indica que el grupo metilo de la Thr, junto con otros efectos, puede modular el reconocimiento.

- Hemos sintetizado y analizado tres epítopos importantes de la proteína MUC1 (**3**, **4** y **5**). Este trabajo aporta resultados importantes para el reconocimiento de nuestros derivados por las lectina SBA y VVA en términos de selección de epítopos. Aunque ya es bien sabido que las lectinas reconocen carbohidratos, hemos demostrado que el proceso de reconocimiento está modulado por la secuencia peptídica

cercana al punto de glicosilación. En concreto, los receptores de la lectina estudiada muestran mayor afinidad por el antígeno Tn cuando está situado en la región peptídica APDTR. En cambio, se observa un decrecimiento en la afinidad cuando este antígeno se encuentra localizado en los fragmentos AHGVTSA o APGSTAP. El Asp y la Arg de la secuencia APDTR parecen fijar la conformación tanto del péptido como del carbohidrato maximizando las interacciones con estas lectinas.

- Todas las conclusiones expuestas anteriormente fueron deducidas mediante una aproximación multidisciplinar que involucra el uso de diferentes técnicas como experimentos de resonancia magnética nuclear, ELLAs, ITCs, dinámicas moleculares, difracción de rayos X…

- Los resultados recogidos en esta Tesis podrían ayudar a entender mejor las interacciones entre lectinas y glicopétidos. Esta información podría ser empleada para diseñar nuevos sitios activos en las lectinas que tengan mayor afinidad por glicopéptidos tipo MUC1 que incorporen el antígeno Tn. De esta forma, se podría modular la especificidad de los glicosensores y diseñar otros nuevos más eficaces para la detección, en pequeñas concentraciones, del antígeno Tn. Esto podría ayudar de forma significativa en las terapias contra el cáncer.

EXPERIMENTAL SECTION

I. Reagents and general procedures

II. NMR routine experiments

III. 2D NOESY experiments

IV. Saturation-transfer difference (STD) NMR experiments

V. Unrestrained molecular dynamics simulations

VI. MD simulations with time-averaged restraints (MD-tar)

VII. Isothermal titration calorimetry (ITC)

VIII. General procedure for the enzyme-linked lectin assay (ELLA)

IX. General procedure to obtain glycopeptides by solid phase peptide synthesis (SPPS)

X. Crystallization

XI. Synthesis

XII. NMR spectra and chromatograms

I. Reagents and general procedures

Commercially available reagents were used without further purification. Analytical thin layer chromatography (TLC) was performed on glass plates precoated with a 0.25 mm thickness of silica gel. The TLC plates were visualized with UV light and by staining with Hanessian solution (ceric sulfate and ammonium molybdate in aqueous sulfuric acid) or sulfuric acid-ethanol solution. Column chromatography was performed on silicagel (230-400 mesh).

High-resolution electrospray mass (ESI) spectra were recorded on a microTOF spectrometer, accurate mass measurements were achieved by using sodium formate as an external reference.

II. NMR routine experiments

^{1}H and ^{13}C NMR spectra were measured with a Bruker Avance 400 MHz spectrometer at 298 K with TMS as the internal standard. Magnitude-mode ge-2D COSY spectra were acquired with gradients by using the *cosygpqf* pulse program with a pulse width of 90°. Phase-sensitive ge-2D HSQC spectra were acquired by using z-filter and selection before t1 removing the decoupling during acquisition by use of the *invigpndph* pulse program with CNST2 (J_{HC}) = 145. All NMR chemical shifts (δ) were recorded in ppm and coupling constants (J) were reported in Hz. The results of these experiments were processed with MestReNova software.

III. 2D NOESY experiments

NOESY experiments were recorded on a Bruker Avance 400 MHz spectrometer at 298 K and pH = 6.0-6.5 in H_2O/D_2O (9:1). The experiments

were conducted by using phase-sensitive ge-2D NOESY with WATERGATE for H_2O/D_2O (9:1) spectra. NOEs intensities were normalized with respect to the diagonal peak at zero mixing time. Distances with structural information were semi-quantitatively determined by integrating the volume of the corresponding cross-peaks. The number of scans used was 16 and the mixing time was 500 ms.

IV. Saturation-transfer difference (STD) NMR experiments

STD experiments were recorded on a Bruker Avance 600 MHz spectrometer at 298 K and pH = 7.2 (phosphate buffer) in H_2O/D_2O (9:1). Commercial SBA lectin was purchased from Sigma. The binding of glycopeptides was evaluated by STD experiments performed with a 20:1 molar ratio of the glycopeptide/lectin mixture. The concentration of the lectin was 40 μM. A series of Gaussian-shaped pulses of 50 ms each one was employed with a total saturation time for the protein envelope of 2 s and a maximum B1 field strength of 60 Hz. An *off-resonance* frequency of δ = 40 ppm and *on-resonance* frequency of δ = 7.0 ppm (protein aromatic signals region) and δ = −0.255 ppm were applied. A total number of 1024 scans were acquired.

V. Unrestrained molecular dynamics simulations

All molecular dynamics simulations were carried out on the Finis-Terrae cluster of the *Centro de Supercomputación de Galicia* (CESGA), Spain. The starting geometries for the complexes were generated from the available data deposited in the Protein Data Bank (pdb ids: 1SBF –SBA lectin–, 1N47 –VVA lectin– and 2CGZ –HPA lectin–) and modified accordingly. Each model complex was immersed in a 10 Å-sided cube with pre-equilibrated TIP3P water molecules.

To equilibrate the system, we followed a protocol consisting of 10 steps. Firstly, only the water molecules are minimized, and then heated to 300 K. The water box, together with ions, was then minimized, followed by a short MD simulation. At this point, the system was minimized in the four following steps with positional restraints imposed on the solute, decreasing the force constant step by step from 20 to 5 kcal/mol. Finally, a non-restraint minimization was performed.

The production dynamics simulations were accomplished at a constant temperature of 300 K (by applying the Berendsen coupling algorithm for the temperature scaling) and constant pressure (1 bar). The Particle Mesh Ewald Method, to introduce long-range electrostatic effects, and periodic boundary conditions were also used. The SHAKE algorithm for hydrogen atoms, which allows using a 2 fs time step, was also employed. Finally, a 9 Å cutoff was applied for the Lennard-Jones interactions. MD simulations were performed with the pmemd.cuda module of AMBER 12 (parm99 force field),[1] which was implemented with GLYCAM 06 parameters[2] to accurately simulate the corresponding glycopeptides. A simulation length of 100 ns and the trajectory coordinates were saved each 0.5 ps.

VI. MD simulations with time-averaged restraints (MD-tar)

MD-tar simulations were performed with the pmemd.cuda module of AMBER 12 (parm99 force field), which was implemented with GLYCAM 06 parameters. Distances derived from NOE cross-peaks were included as time-averaged distance restraints. A $\langle r^6 \rangle^{1/6}$ average was used for the distances and a linear average was used for coupling constants. Final trajectories were

[1] J. Wang, P. Cieplak, P. A. Kollman, *J. Comput. Chem.* **2000**, *21*, 1049-1074.
[2] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622-655.

run using an exponential decay constant of 2000 ps and a simulation length of 20 ns in explicit TIP3P water molecules.

VII. Isothermal titration calorimetry (ITC)

ITC was used to evaluate the dissociation constants of SBA and HPA lectins against different peptides (compounds **1** to **5**). All experiments were carried out at 25 °C with concentrations of the lectins between 10 and 20 μM, and concentrations of the glycopeptides between 300 μM and 1 mM, in 10 mM Tris, pH = 7.5. The reactions were performed on an AUTO ITC instrument Microcal Auto-iTC200. Data integration, correction and analysis were carried out using Origin 7 (Microcal) with a single-site binding model.

VIII. General procedure for the enzyme-linked lectin assay (ELLA)

An ELISA plate (Pierce Amine-binding, Maleic anhydride 96-well plate) was coated with 100 μL/well of a solution of the corresponding glycopeptide (0-150 nmol/well) in carbonate/bicarbonate buffer (0.2 M, pH = 9.4) and incubated overnight at 25 °C.

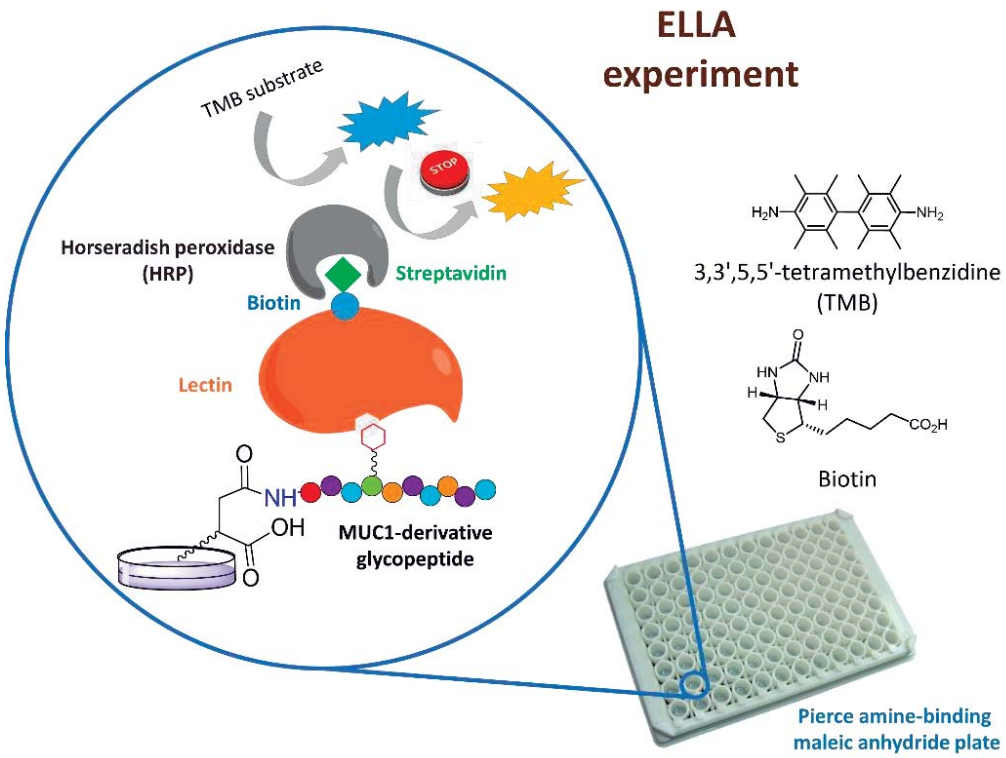
The unbound sites were then blocked by adding 200 μL/well of blocking buffer (commercial available). After 1 h at 25 °C, the blocking buffer was removed and the plate wells were washed 3 × 200 μL/well with PBST [phosphate buffered saline (0.1 M sodium phosphate, 0.15 M sodium chloride, pH = 7.2), containing 0.05% Tween-20 detergent].

As the next step, the wells were incubated with biotin-conjugated SBA, VVA or HPA lectins (100 μL, diluted 1/150 in PBST buffer) for 2 h.

After washing with PBST (3 × 200 µL/well, 2 min/well), the wells were treated with horseradish peroxidase (HRP) conjugated streptavidin (100 µL, diluted 1/3000 in PBST buffer) for 1 h at 25 °C.

The wells were again washed first with PBST (3 × 200 µL/well, 2 min/well) and then with 350 µL of water. 3,3,5,5-tetramethylbenzidine (TMB) was added (90 µL/well) and after incubation for 10 min, the reaction was terminated with the addition of 50 mL/well of stop solution (1M H_2SO_4).

Absorbance detection of the wells was immediately performed at 450 nm using an ELISA plate reader (Bio-Rad, iMark plate reader). Average absorbance intensities of three replicates were plotted against glycopeptides concentration (Scheme 6.1).



Scheme 6.1. *Schematic representation of the ELLA strategy used in this Thesis.*

IX. General procedure to obtain glycopeptides by solid phase peptide synthesis (SPPS)

Rink Amide MBHA resin (178 mg, 0.1 mmol of NH_2) was put into a vessel reactor. Convenient protected amino acids were put into cartridges (1 mmol).

Automatic synthesis started with a piperidine washing during 7 min. Then, the starting cartridge is expelled and the first amino acid was dissolved in DIEA (3 mL of 2 M solution in NMP), HBTU (0.9 mmol of 0.45 M) and DMF (2 g) and added to the reactor for 10 min. Consecutive washes for Fmoc deprotection with piperidine and DMF were also made during 30 min.

The process followed expelling the cartridge and starting the process again with the next cartridge. These steps were repeated as many times as amino acids were needed to couple.

When a glycosylated building block (**8** or **9**) was coupled, the synthesis was carried out manually, obtaining better yields and reducing the equivalents (from 10 to 2). The process was as follows: 1.1 equivalents of HBTU, 0.5 mL of DIEA 2.0 M and 2 mL of DMF was added to the glycosylamino acid. The reaction was stirred until the coupling was completed as deduced by Kaiser test.[3] The resin was then placed again in the synthesizer to obtain the complete sequence of the desired glycopeptide.

As a next step, resin was taken out from the synthesizer and acetate groups of the carbohydrate were deprotected with 10 mL of hydrazine/MeOH (7:3)

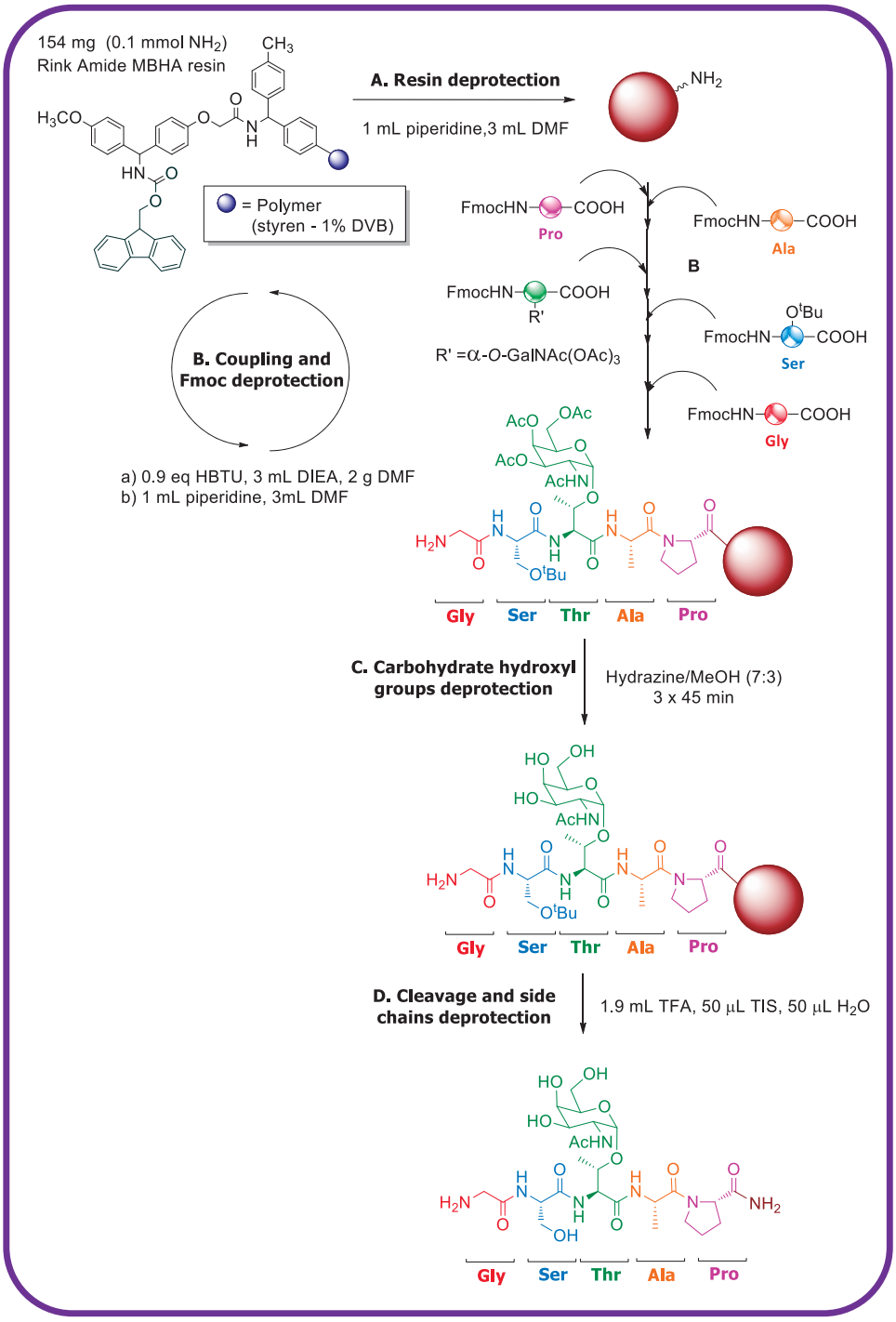
[3] E. Kaiser, R. L. Colescott, C. D. Bossinger, P. I. Cook, *Anal. Biochem.* **1970**, *34*, 595-598.

solution 3 times of 45 min each one. Several washes with DMF y CH_2Cl_2 (DCM) were needed.

After filtration, TFA (1.90 mL), TIS (50 μL) and H_2O (50 μL) were added for the cleavage reaction and removal of the side chain protecting groups. The mixture was stirred for 2 h. At this point, Et_2O (10 mL) was added producing a precipitate that was filtered and redissolved in water. The final glycopeptide was purified by reverse phase HPLC.

Image 6.1 shows the equipment used in the lab where the most important parts are highlighted. Scheme 6.2 represents a schematic approximation of this process.

Image 6.1. *Picture of Model 433A Peptide Synthesizer (Applied Biosystems) used to peptide synthesis in our lab.*



Scheme 6.2. *Solid phase peptide synthesis (SPPS) route for a model pentaglycopeptide.*

X. Crystallization

Soybean agglutinin lectin was dissolved in buffer (25 mM TRIS-HCl, 150 mM NaCl, pH = 7.5). It was further purified by gel filtration chromatography in order to remove aggregates. Then, it was dialyzed in a 25 mM TRIS-HCl and 150 mM NaCl buffer (pH = 7.5) without salt. The concentration of the protein was measured at 280 nm of absorbance using the theoretical extinction coefficient (ε) of 40450 mL mg^{-1} cm^{-1}. Crystals were grown by hanging drop diffusion at 18 °C.

The drops were prepared by mixing 1 μL of protein complex containing 8.5 mg/mL of soybean agglutinin and 10 mM glycopeptide **4** with 1 μL of solutions with 20-30% PEG 3350, 0.2 M NaBr and 0.5 μL of 20-30% PEG 400 as additive. The own mother solution was acted as cryoprotectant.

The data were processed and scaled using the XDS package[4] and CCP4[5] software, relevant statistics are given in Table 4.5 (*Chapter 4*). The crystal structure was solved by molecular replacement using PHENIX[6] and using the PDB entry 1SBF as the template. Initial phases were further improved by cycles of manual model building in Coot[7] and refinement with REFMAC5.[8]

---

[4] W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70*, 2204-2216.

[5] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, K. S. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235-242.

[6] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213-221.

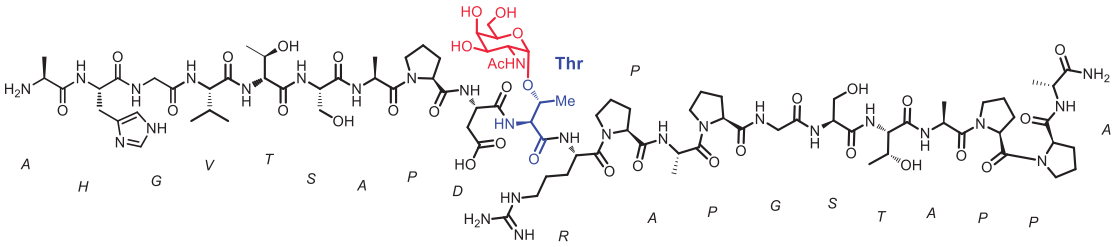
[7] P. Emsley, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126-2132.

[8] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355-367.

The final models were validated with PROCHECK.[9] The asymmetric units of these crystals show 12 molecules of SBA. Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (pdb: 4d69).

XI. Synthesis

Compound 1 (MUC1-APDT*R)



The glycosylated amino acid building block **8** (134 mg, 2 equiv.) was coupled manually, while the other Fmoc amino acids were coupled in the automated mode. After elimination of acetate groups and cleavage, glycopeptide **1** was obtained and then purified by reversed-phase HPLC and lyophilized (94% overall yield).

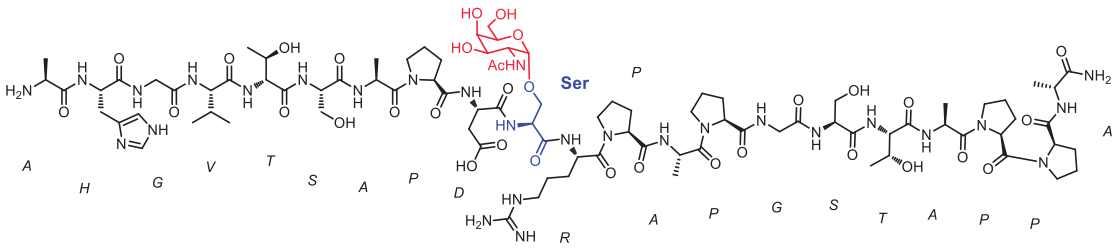
Semi-preparative HPLC gradient (t_R 17.4 min.):

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	5	95
30	10	18	82

HRMS (ESI) (m/z) 720.6973 $[M+3H]^+$, calculated $C_{91}H_{149}N_{28}O_{33}^{3+}$: 720.6942.

[9] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **1993**, *26*, 283-291.

Compound 2 (MUC1-APDS*R)



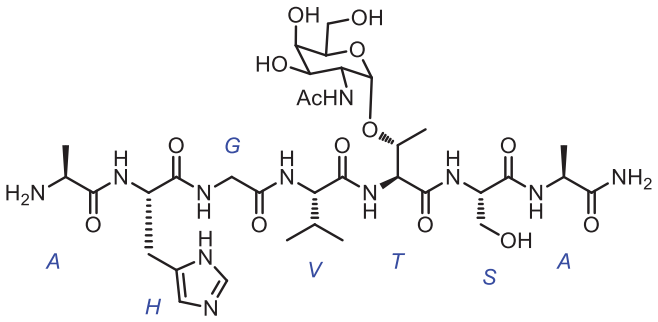
The glycosylated amino acid building block **9** (131 mg, 2 equiv.) was coupled manually, while the other Fmoc amino acids were coupled in the automated mode. After elimination of acetate groups and cleavage, glycopeptide **2** was obtained and then purified by reversed-phase HPLC and lyophilized (95% overall yield).

Semi-preparative HPLC gradient (t_R 18.8 min.):

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O + 0.1% TFA (%)
0	10	5	95
30	10	18	82

HRMS (ESI) (m/z) 716.0234 $[M+3H]^+$, calculated $C_{91}H_{149}N_{28}O_{33}^{3+}$: 716.0223.

L-Ala-L-His -L-Gly- L-Val-L-Thr-(α-*O*-D-GalNAc)-L-Ser-L-Ala-NH_2 (3)



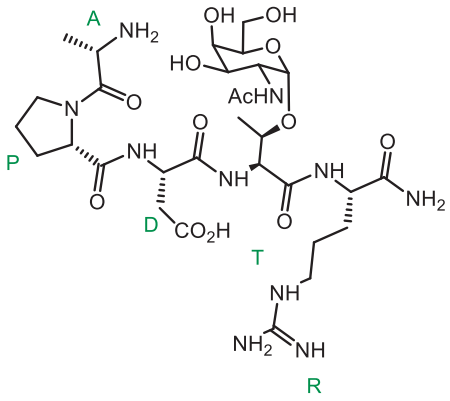
Following SPPS methodology previously described (*Experimental section VII*), with Ala-Fmoc (311 mg, 1 mmol) and Ser(^{t}Bu)-Fmoc (383 mg, 1 mmol) were automatically coupled. Then, Thr building block (compound **8**, 318 mg, 0.474 mmol) was manually coupled. Val-Fmoc (339 mg, 1 mmol), Gly-Fmoc (197 mg, 1 mmol), His(Trt)-Fmoc (620 mg, 1 mmol) and Ala-Fmoc (311 mg, 1 mmol) were then automatically coupled. Glycopeptide **3** was obtained and then purified by reversed-phase HPLC (H_2O/MeCN 98:2) and lyophilized (94% overall yield). HPLC *(A*: MeCN; *B:* 0.1% CF_3COOH/H_2O) within 40 min, t_R 18.7 min.

HRMS (ESI) (m/z) 844.4141 $[M+H]^+$, calculated $C_{34}H_{58}N_{11}O_{14}^+$: 844.4159.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 0.95 (t, 6H, *J* = 6.9 Hz, $2CH_{3\ Val}$), 1.24 (d, 3H, *J* = 6.3 Hz, $CH_{3\ Thr}$), 1.39 (d, 3H, *J* = 7.2 Hz, $CH_{3\ Ala\text{-}7}$), 1.47 (d, 3H, *J* = 7.1 Hz, $CH_{3\ Ala\text{-}1}$), 1.99 – 2.14 (m, 4H, NHCOC$\underline{H}_3$, $H_{\beta\ Val}$), 3.15 – 3.36 (m, 2H, $2H_{\beta\ His}$), 3.65 – 3.77 (m, 2H, $2H_{6S}$), 3.79 – 3.89 (m, 3H, H_{3S}, $2H_{\beta\ Ser}$), 3.90 – 4.03 (m, 4H, H_{4S}, H_{5S}, $2H_{\alpha\ Gly}$), 4.03 – 4.12 (m, 2H, $H_{\alpha\ Val}$, H_{2S}), 4.19 – 4.37 (m, 3H, $H_{\alpha\ Ala\text{-}1}$, $H_{\beta\ Thr}$, $H_{\alpha\ Ala\text{-}7}$), 4.48 (t, 1H, *J* = 5.7 Hz, $H_{\alpha\ Ser}$), 4.61 (d, 1H, *J* = 1.9 Hz, $H_{\alpha\ Thr}$), 4.70 (t, 1H, *J* = 7.0 Hz, $H_{\alpha\ His}$), 4.91 (d, 1H, *J* = 3.7 Hz, H_{1S}), 7.32 (s, 1H, $H_{5\ arom\text{-}His}$), 8.62 (s, 1H, $H_{2\ arom\text{-}His}$).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 16.4, 16.7 ($CH_{3\ Ala-1}$, $CH_{3\ Ala-7}$), 17.7 ($CH_{3\ Val}$), 18.1 ($CH_{3\ Thr}$), 18.5 ($CH_{3\ Val}$), 22.2 ($NHCO\underline{C}H_3$), 26.1 ($C_{\beta\ His}$), 30.3 ($C_{\beta\ Val}$), 42.3($C_{\alpha\ Gly}$), 48.8 ($C_{\alpha\ Val}$), 49.5, 49.7 ($C_{\alpha\ Ala-1}$, $C_{\alpha\ Ala-7}$), 52.7 ($C_{\alpha\ His}$), 54.8 ($C_{\alpha\ Ser}$), 57.2 ($C_{\alpha\ Thr}$), 59.3 (C_{2S}), 61.3, 61.5 ($C_{\beta\ Ser}$, C_{6S}), 68.1, 68.6 (C_{3S}, C_{4S}), 71.3 (C_{5S}), 76.2 ($C_{\beta\ Thr}$), 98.8 (C_{1S}), 117.4 (C_{arom}), 128.2 ($C_{\gamma\ His}$), 133.6 (C_{arom}), 170.8, 170.9, 170.9, 171.0 171.7, 173.8, 174.0, 177.6 (CO).

L-Ala-L-Pro-L-Asp-L-Thr-(α-*O*-D-GalNAc)-L-Arg-NH_2 (4)



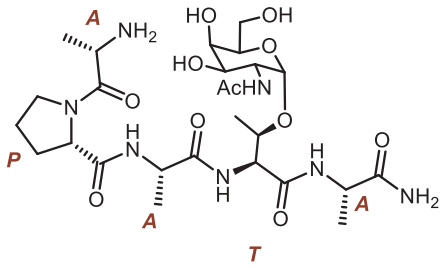
Following SPPS methodology previously described (*Experimental section VII*), Arg(Pbf)-Fmoc (649 mg, 1 mmol) was automatically coupled. Then, Thr building block (compound **8**, 318 mg, 0.474 mmol) was manually coupled. Asp(O^{t}Bu)-Fmoc (411 mg, 1 mmol), Pro-Fmoc (337 mg, 1 mmol) and Ala-Fmoc (311 mg, 1 mmol) were then automatically coupled. Glycopeptide **4** was obtained and then purified by reversed-phase HPLC (H_2O/MeCN 98:2) and lyophilized (94% overall yield). HPLC *(A*: MeCN; *B:* 0.1% CF_3COOH/H_2O) within 40 min, t_R 15.5 min.

HRMS (ESI) (m/z) 761.3792 $[M+H]^+$, calculated $C_{30}H_{53}N_{10}O_{13}^+$: 761.3788.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 1.28 (d, 3H, *J* = 6.4 Hz, $CH_{3\ Thr}$), 1.54 (d, 3H, *J* = 7.0 Hz, $CH_{3\ Ala}$), 1.61 – 1.81 (m, 4H, $2H_{\beta\ Arg}$, $2H_{\gamma\ Arg}$), 1.83 – 1.97 (m, 1H, $H_{\beta\ Pro}$), 1.99 – 2.12 (m, 5H, NHCOC$\underline{H}_3$, $2H_{\gamma\ Pro}$), 2.31 – 2.40 (m, 1H, $H_{\beta\ Pro}$), 2.83 – 3.07 (m, 2H, $2H_{\beta\ Asp}$), 3.21 (t, 2H, *J* = 6.8 Hz, $2H_{\delta\ Arg}$), 3.57 – 3.68 (m, 1H, $H_{\delta\ Pro}$), 3.68 – 3.78 (m, 3H, $H_{\delta\ Pro}$, $2H_{6S}$), 3.88 (dd, 1H, *J* = 11.0, 3.2 Hz, H_{3S}), 3.94 – 3.98 (m, 1H, H_{5S}), 4.03 ('t', 1H, *J* = 6.2 Hz, H_{4S}), 4.13 (dd, 1H, *J* = 11.0, 3.7 Hz, H_{2S}), 4.25 – 4.40 (m, 3H, $H_{\alpha\ Ala}$, $H_{\beta\ Thr}$, $H_{\alpha\ Arg}$), 4.47 – 4.56 (m, 2H, $H_{\alpha\ Thr}$, $H_{\alpha\ Pro}$), 4.83 – 4.93 (m, 2H, $H_{\alpha\ Asp}$, H_{1S}).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 15.1 ($CH_{3\ Ala}$), 18.4 ($CH_{3\ Thr}$), 22.4 ($NHCO\underline{C}H_3$), 24.5, 24.7 ($C_{\gamma\ Pro}$, $C_{\gamma\ Arg}$), 28.6 ($C_{\beta\ Arg}$), 29.4 ($C_{\beta\ Pro}$), 35.0 ($C_{\beta\ Asp}$), 40.6 ($C_{\delta\ Arg}$), 47.8 ($C_{\delta\ Pro}$), 48.7 ($C_{\alpha\ Ala}$), 49.8, 49.9 ($C_{\alpha\ Asp}$, C_{2S}), 53.58($C_{\beta\ Thr}$), 57.5 ($C_{\alpha\ Pro}$), 60.3 ($C_{\alpha\ Arg}$), 61.5 (C_{6S}), 68.1, 68.3, 71.8 (C_{3S}, C_{4S}, C_{5S}), 76.3 ($C_{\alpha\ Thr}$), 98.8 (C_{1S}), 156.7 ($CNH_{\ Arg}$) 169.3, 171.1, 172.7, 173.5, 173.9, 174.2, 175.8 (CO).

L-Ala-L-Pro-L-Ala-L-Thr-(α-*O*-D-GalNAc)-L-Ala-NH_2 (4')



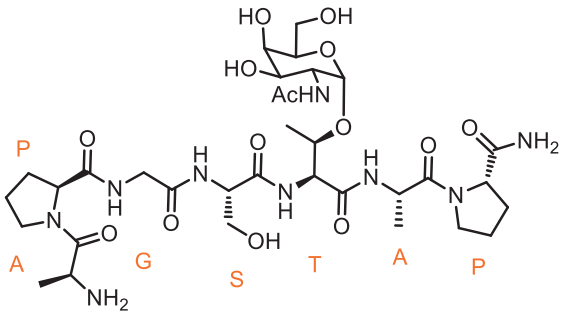
Following SPPS methodology previously described (*Experimental section VII*), Ala-Fmoc (311 mg, 1 mmol) was automatically coupled. Then, Thr building block (compound **8**, 318 mg, 0.474 mmol) was manually coupled. Ala-Fmoc (311 mg, 1 mmol), Pro-Fmoc (337 mg, 1 mmol) and Ala-Fmoc (311 mg, 1 mmol) were then automatically coupled. Glycopeptide **4'** was obtained and purified by reversed-phase HPLC (H_2O/MeCN 98:2) and lyophilized (94% overall yield). HPLC (*A*: MeCN; *B:* 0.1% CF_3COOH/H_2O) within 40 min, t_R 15.3 min.

HRMS (ESI) (m/z) 632.3252 $[M+H]^+$, calculated $C_{26}H_{46}N_7O_{11}^+$: 632.3250.

^{1}H NMR (400 MHz, D_2O) δ (ppm): δ 1.32 (d, 3H, *J* = 6.3 Hz, $CH_{3\ Thr}$), 1.40 – 1.49 (m, 6H, $CH_{3\ Ala-3}$, $CH_{3\ Ala-5}$), 1.57(d, 3H, *J* = 7.0 Hz, $CH_{3\ Ala-1}$), 1.88 – 2.00 (m, 1H, $H_{\beta\ Pro}$), 2.00 – 2.16 (m, 5H, $NHCOC\underline{H}_3$, $2H_{\gamma\ Pro}$), 2.33 – 2.42 (m, 1H, $H_{\beta\ Pro}$), 3.60 – 3.86 (m, 4H, $2H_{\delta\ Pro}$, $2H_{6S}$), 3.87 – 4.02 (m, 2H, H_{3S}, H_{4S}), 4.07 (t, 1H, *J* = 6.1 Hz, H_{5S}), 4.12 – 4.21 (m, 1H, H_{2S}), 4.25 – 4.46 (m, 3H, $H_{\alpha\ Ala-3}$, $H_{\alpha\ Ala-5}$, $H_{\beta\ Thr}$), 4.48 – 4.57 (m, 3H, $H_{\alpha\ Ala-1}$, $H_{\alpha\ Pro}$, $H_{\alpha\ Thr}$), 4.96 (d, 1H, *J* = 3.7 Hz, H_{1S}).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 15.1 ($CH_{3\ Ala-1}$), 15.1, 16.5 ($CH_{3\ Ala-3}$, $CH_{3\ Ala-5}$) 18.3 ($CH_{3\ Thr}$), 22.3 ($NHCO\underline{C}H_3$), 24.8 ($C_{\gamma\ Pro}$), 29.5 ($C_{\beta\ Pro}$), 47.8 ($C_{\delta\ Pro}$), 48.1, 49.2, 49.5, 49.7, 49.9 ($C_{\alpha\ Ala-1}$, $C_{\alpha\ Ala-3}$, $C_{\alpha\ Ala-5}$, $C_{\alpha\ Thr}$, C_{2S}), 57.1 ($C_{\alpha\ Pro}$), 61.3 (C_{6S}), 68.1, 68.6, 71.4 (C_{3S}, C_{4S}, C_{5S}), 76.3 ($C_{\beta\ Thr}$), 98.8 (C_{1S}), 169.1, 170.9, 173.5, 174.4, 175.3, 177.3 (CO).

L-Ala-L-Pro-L-Gly-L-Ser-L-Thr-(α-*O*-D-GalNAc)-L-Ala-L-Pro-NH_2 (5)



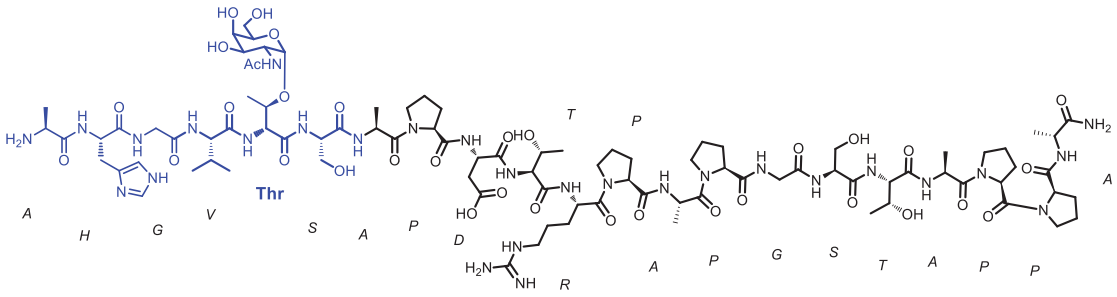
Following SPPS methodology previously described (*Experimental section VII*), Pro-Fmoc (337 mg, 1 mmol) and Ala-Fmoc (311 mg, 1 mmol) were automatically coupled. Then, Thr building block (compound **8**, 318 mg, 0.474 mmol) was manually coupled. Ser(O*t*Bu)-Fmoc (383 mg, 1 mmol), Gly-Fmoc (297 mg, 1 mmol), Pro-Fmoc (337 mg, 1 mmol) and Ala-Fmoc (311 mg, 1 mmol) were then automatically coupled. Glycopeptide **5** was obtained and purified by reversed-phase HPLC (H_2O/MeCN 98:2) and lyophilized (93% overall yield). HPLC (*A*: MeCN; *B:* 0.1% TFA/H_2O) within 40 min, t_R 21.1 min.

HRMS (ESI) (m/z) 802.3957 $[M+H]^+$, calculated $C_{30}H_{53}N_{10}O_{13}^+$: 802.3941.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 1.26 (d, 3H, *J* = 6.4 Hz, $CH_{3\ Thr}$), 1.36 (d, 3H, *J* = 7.1 Hz, $CH_{3\ Ala\text{-}6}$), 1.53 (d, 3H, *J* = 7.0 Hz, $CH_{3\ Ala\text{-}1}$), 1.90 – 2.13 (m, 9H, $NHCOCH_3$, $H_{\beta\ Pro\text{-}2}$, $H_{\beta\ Pro\text{-}7}$, $2H_{\gamma\ Pro\text{-}2}$, $2H_{\gamma\ Pro\text{-}7}$), 2.26 – 2.40 (m, 2H, $H_{\beta\ Pro\text{-}2}$, $H_{\beta\ Pro\text{-}7}$), 3.60 – 3.80 (m, 6H, $2H_{\delta\ Pro\text{-}2}$, $2H_{\delta\ Pro\text{-}7}$, $2H_{6S}$), 3.83 – 4.05 (m, 7H, $2H_{\beta\ Ser}$, $2H_{\alpha\ Gly}$, H_{3S}, H_{4S}, H_{5S}), 4.09 (dd, 1H, *J* = 11.0, 3.8 Hz, H_{2S}), 4.32 – 4.41 (m, 3H, $H_{\alpha\ Ala\text{-}6}$, $H_{\beta\ Thr}$, $H_{\alpha\ Pro\text{-}2}$), 4.45 – 4.62 (m, 4H, $H_{\alpha\ Thr}$, $H_{\alpha\ Ala\text{-}1}$, $H_{\alpha\ Ser}$, $H_{\alpha\ Pro\text{-}7}$), 4.63 – 4.68 (m, 1H, $H_{\alpha\ Ser}$), 4.91 (d, 1H, *J* = 3.8 Hz, H_{1S}).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 15.0 ($CH_{3\ Ala-1}$), 15.5 ($CH_{3\ Ala-6}$), 18.2 ($CH_{3\ Thr}$), 22.2 ($NHCO\underline{C}H_3$), 24.6, 24.7 ($C_{\gamma\ Pro-2}$, $C_{\gamma\ Pro-7}$), 29.3, 29.5 ($C_{\beta\ Pro-2}$, $C_{\beta\ Pro-7}$), 42.3 ($CH_{2\ Gly}$), 47.5, 47.6, 47.7, 48.1 ($C_{\delta\ Pro-2}$, $C_{\delta\ Pro-7}$, $C_{\alpha\ Ala-1}$, $C_{\alpha\ Ala-3}$), 49.7 (C_{2S}), 55.1 ($C_{\alpha\ Ser}$), 57.2 ($C_{\alpha\ Thr}$), 60.1 ($C_{\alpha\ Pro-7}$), 60.7 ($C_{\alpha\ Pro-2}$), 61.2, 61.3 (C_{6S}, $C_{\beta\ Ser}$), 68.1, 68.6, 71.4 (C_{3S}, C_{4S}, C_{5S}), 98.5 (C_{1S}), 169.4, 170.7, 171.2, 172.2, 172.5, 173.8, 174.5, 176.9 (CO).

Compound 6 (MUC1-AHGVT*SA)



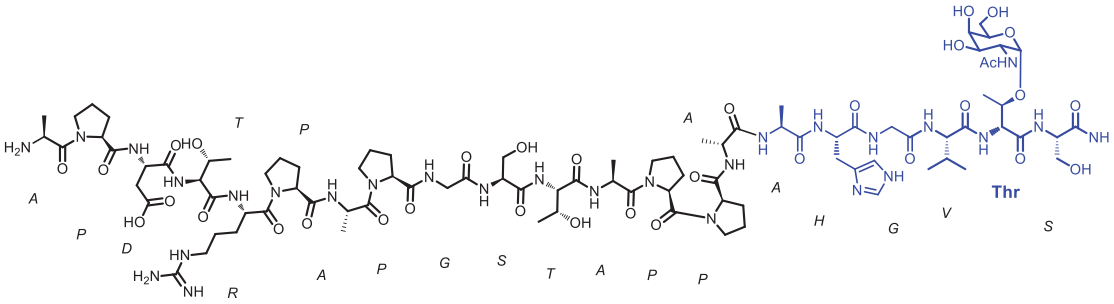
The glycosylated amino acid building block **8** (134 mg, 2 equiv.) was coupled manually, while the other Fmoc amino acids were coupled in the automated mode. After elimination of acetate groups and cleavage, glycopeptide **6** was obtained and then purified by reversed-phase HPLC and lyophilized (90% overall yield).

Semi-preparative HPLC gradient (t_R 33.02 min.):

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O + 0.1% TFA (%)
0	10	5	95
30	10	18	82

HRMS (ESI) (m/z) 720.6946 $[M+3H]^+$, calculated $C_{91}H_{149}N_{28}O_{33}^{3+}$: 720.6942.

Compound 6' (MUC1-AHGVT*SA final)



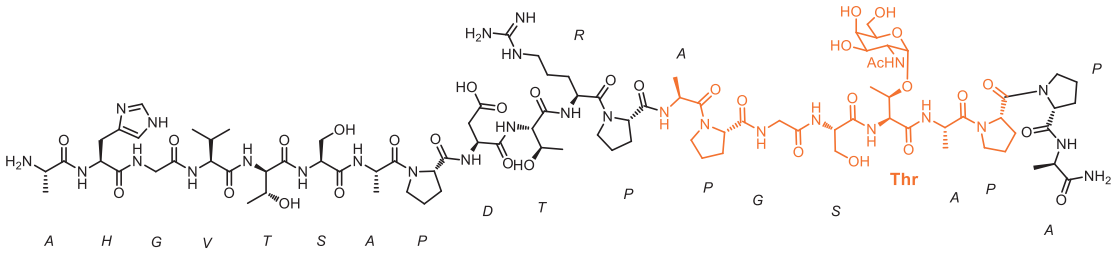
The glycosylated amino acid building block **8** (134 mg, 2 equiv.) was coupled manually, while the other Fmoc amino acids were coupled in the automated mode. After elimination of acetate groups and cleavage, glycopeptide **6'** was obtained and then purified by reversed-phase HPLC and lyophilized (90% overall yield).

Semi-preparative HPLC gradient (t_R 32.69 min.):

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O + 0.1% TFA (%)
0	10	5	95
30	10	18	82

HRMS (ESI) (m/z) 720.6979 $[M+3H]^+$, calculated $C_{91}H_{149}N_{28}O_{33}^{3+}$: 720.6942.

Compound 7 (MUC1-APGST*AP)



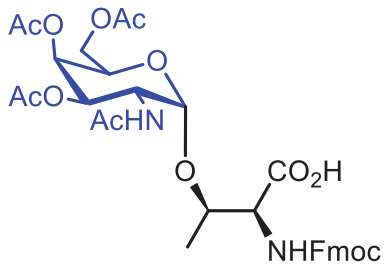
The glycosylated amino acid building block **8** (134 mg, 2 equiv.) was coupled manually, while the other Fmoc amino acids were coupled in the automated mode. After elimination of acetate groups and cleavage, glycopeptide **7** was obtained and then purified by reversed-phase HPLC and lyophilized (92% overall yield).

Semi-preparative HPLC gradient (t_R 21.92 min.):

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O + 0.1% TFA (%)
0	10	5	95
30	10	18	82

HRMS (ESI) (m/z) 720.6948 $[M+3H]^+$, calculated $C_{91}H_{149}N_{28}O_{33}^{3+}$: 720.6942.

Fmoc-L-Thr(α-*O*-D-tri-*O*-acetyl-*N*-acetylgalactosamine) (8)



8

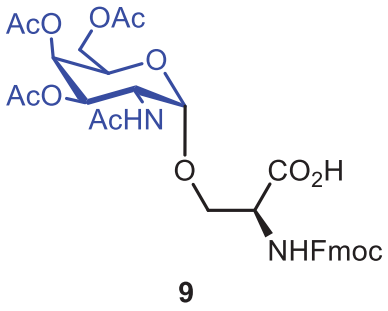
Compound **20** (300 mg, 0.413 mmol) was dissolved in CH_2Cl_2 (2.5 mL) and TFA (2.5 mL). The reaction was then stirred at room temperature for 2 h. After solvent evaporation, helped by Et_2O washes, and further purification by chromatographic column, compound **8** was obtained as a white foam (270 mg, 97%). R_f (CH_2Cl_2/MeOH, 9:1) = 0.80.

HRMS (ESI) (m/z) 671.2443 $[M+H]^+$; calculated $C_{33}H_{39}N_2O_{13}^+$: 671.2447.

^{1}H NMR (400 MHz, DMSO-*d6*) δ (ppm): 1.02 (d, *J* = 6.2 Hz, 3H, Me), 1.85 (s, 3H, CH_3CO), 1.92 (s, 3H, CH_3CO), 1.97 (s, 3H, CH_3CO), 2.09 (s, 3H, CH_3CO), 3.88 (dd, *J* = 9.9, 5.1 Hz, 1H, H_{6s}), 4.07-4.13 (m, 2H, H_{5s}, H_{6s}), 4.21-4.30 (m, 3H, H_β, $CH_{2\ Fmoc}$), 4.30-4.45 (m, 3H, H_{2s}, H_α, CH_{Fmoc}), 4.88 (d, *J* = 3.7 Hz, 1H, H_{1s}), 4.94 (dd, *J* = 11.6, 3.3 Hz, 1H, H_{3s}), 5.25 (m, 1H, H_{4s}), 7.26-7.44 (m, 4H, aryl-H), 7.62 (d, *J* = 8.6 Hz, 1H, NH_s), 7.70-7.77 (m, 2H, aryl-H), 7.87-7.91 (m, 2H, aryl-H), 8.01 (d, *J* = 9.3 Hz, 1H, $N\underline{H}_{Fmoc}$).

^{13}C NMR (100 MHz, DMSO-*d6*) δ (ppm): 14.4 (Me), 20.4, 20.5, 20.5, 22.5 (CH_3CO), 45.4 (C_{2s}), 47.3 (CH_{Fmoc}), 58.3 (C_α), 61.1 (C_{6s}), 65.7 ($CH_{2\ Fmoc}$), 66.2 (C_{5s}), 66.8 (C_{4s}), 67.9 (C_{3s}), 70.6 (C_β), 92.9 (C_{1s}), 120.1, 120.2, 125.1, 125.3, 127.0, 127.0, 127.7, 127.7, 140.8, 140.8, 143.7, 144.0 (C-aryl), 156.8 (CO_{Fmoc}), 169.4, 169.8, 170.0, 170.0 , 172.0 (CO).

Fmoc-L-Ser(α-*O*-D-tri-*O*-acetyl-*N*-acetylgalactosamine) (9)



9

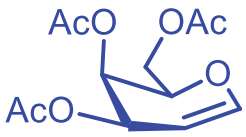
Compound **21** (300 mg, 0.421 mmol) was dissolved in CH_2Cl_2 (2.5 mL) and TFA (2.5 mL). The reaction was then stirred at room temperature for 2 h. After solvent evaporation, helped by Et_2O washes, and further purification by chromatographic column compound **9** was obtained as a white foam (271 mg, 98%). R_f (CH_2Cl_2/MeOH, 9:1) = 0.80.

HRMS (ESI) (m/z) 657.2235 $[M+H]^+$; calculated $C_{32}H_{37}N_2O_{13}^+$: 657.2217.

^{1}H NMR (400 MHz, DMSO-*d6*) δ (ppm): 1.89 (s, 3H, CH_3CO), 1.93 (s, 3H, CH_3CO), 1.99 (s, 3H, CH_3CO), 2.14 (s, 3H, CH_3CO), 3.97-4.04 (m, 2H, H_β), 4.08-4.16 (m, 2H, H_{6s}), 4.29 (m, 1H, CH_{Fmoc}), 4.33-4.40 (m, 4H, H_{2s}, H_{5s}, $CH_{2\ Fmoc}$), 4.47-4.50 (m, 1H, H_α), 5.01 (d, *J* = 3.5 Hz, 1H, H_{1s}), 5.17 (dd, *J* = 11.7, 3.2 Hz, 1H, H_{3s}), 5.41 (d, *J* = 2.3 Hz, 1H, H_{4s}), 7.27–7.45 (m, 4H, aryl-H), 7.74 (dd, *J* = 7.3, 2.3 Hz, 2H, aryl-H), 7.83 (d, *J* = 8.7 Hz, 1H, NH_{Fmoc}), 7.88 (d, *J* = 8.8 Hz, 1H, NH_S), 7.93 (d, *J* = 7.5 Hz, 2H, aryl-H).

^{13}C NMR (100 MHz, DMSO-*d6*) δ (ppm): 24.2, 24.3, 24.3, 26.5 (4xCH_3CO), 51.4 (C_{2s}), 51.8 ($CH_{\ Fmoc}$), 59.0 (C_α), 66.2 (C_{6s}), 70.8 ($CH_{2\ Fmoc}$), 71.2 (C_{5s}), 71.8 (C_{4s}), 72.4 (C_{3s}), 72.8 (C_β), 102.9 (C_{1s}), 124.1, 124.5, 124.5, 124.7, 129.7, 129.7, 131.5, 131.5, 132.0, 132.1, 137.2, 145.5 (C-aryl), 148.6 ($CO_{\ Fmoc}$), 160.1, 161.9, 174.4, 174.6 , 176.1 (CO).

3,4,6-Tri-*O*-acetyl-D-galactal (10)



10

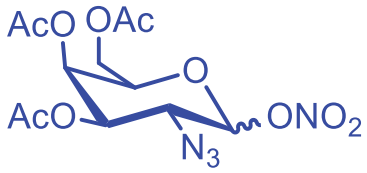
Commercially available β-D-galactose pentaacetate (10 g, 25.6 mmol) was dissolved in CH_2Cl_2 (50 mL). A 33% (w/w) solution of HBr (6 mL, 103.8 mmol) was added dropwise and the reaction was stirred at room temperature for 3 h. The mixture was stopped with $NaHCO_3$ (200 mL) and diluted with CH_2Cl_2 (60 mL). The aqueous layer was extracted with CH_2Cl_2 (2 x 100 mL). Reaction crude was dissolved in acetone (150 mL). Zn dust (20 g, 305.9 mmol), NaH_2PO_4 (5 g, 36 mmol) and NaH_2PO_4 sat. (10 mL, 116.6 mmol) were added to the solution. The mixture was mixed for 5 h. The reaction was then diluted with EtOAc (200 mL) and filtered through celite. The organic layer was extracted with brine (30 mL). The organic phase was collected, dried with anh. Na_2SO_4, filtered, and evaporated. The reaction crude was purified by a flash silica gel chromatographic column to obtain a colorless syrup, compound **10** (6.62 g, 95%). R_f (hexane/EtOAc, 6:4) = 0.56.

HRMS (ESI) (m/z) 295.0851 $[M+Na]^+$, calculated $C_{12}H_{16}O_7Na^+$: 295.0788.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 4.19-4.29 (m, 2H, 2xH_{6s}), 4.31 (m, 1H, H_{5s}), 4.73 (m, 1H, H_{2s}), 5.43 (m, 1H, H_{4s}), 5.56 (m, 1H, H_{3s}), 6.47 (m, 1H, H_{1s}).

^{13}C NMR (100 MHz, $CDCl_3$) δ (ppm): 20.7-20.5 (3 CH_3), 61.8 (C_{6s}), 63.8 (C_{3s}), 63.8 (C_{4s}), 72.8 (C_{5s}), 98.8 (C_{2s}), 145.3 (C_{1s}), 170.0, 170.0, 170.1 (CO).

3,4,6-Tri-*O*-acetyl-2-azido-2-deoxy-α,β-D-galactopyranosyl nitrate (11)



11

Ceric ammonium nitrate (30.20 g, 55.1 mmol) and sodium azide (1.79 g, 27.5 mmol) were dissolved under nitrogen atmosphere at -20 °C. 3,4,6-Tri-*O*-acetyl-D-galactal (5.00 g, 18.6 mmol), compound **10**, was dissolved in dry CH_3CN (80 mL) and added dropwise to the mixture. The reaction was stirred at -20 °C for 10 h. The mixture was diluted with ether (50 mL) and extracted with water (3 × 50 mL) until the organic layer was colorless. The organic phases were collected, dried with anh. Na_2SO_4, filtered, and evaporated. The reaction crude was purified by a flash silica gel chromatographic column to obtain a yellowish syrup, compound **11** (2.62 g, 81%). R_f (hexane/EtOAc, 6:4) = 0.62.

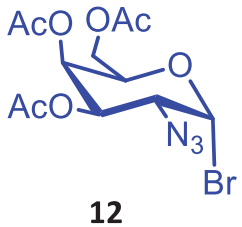
Physical properties and spectroscopic data agreed with earlier reports:[10]

HRMS (ESI) (m/z) 399.0805 $[M+Na]^+$, calculated $C_{12}H_{16}O_{10}Na^+$: 399.0759.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 2.03 (s, 3H, CH_3CO), 2.10 (s, 3H, CH_3CO), 2.18 (s, 3H, CH_3CO), 3.82 (q, 1H, $J_{2,3}$ = 10.8Hz, H_{2s}), 5.08 (q, 1H, J = 3.2 Hz, H_{3s}), 5.42 (m, 1H, H_{4s}), 5.60 (d, 1H, J = 9.0 Hz, H_{1s}).

[10] R. U. Lemieux R. M.Ratcliffe, *Can. J. Chem.* **1979**, *57*, 1244-1251.

3,4,6-Tri-*O*-acetyl-2-azido-1-bromo-2-α-D-deoxygalactose (12)



12

Compound **11** (2.62 g, 6.68 mmol) and LiBr (2.90 g, 33.4 mmol) were dissolved under nitrogen atmosphere at -20 °C in dry CH_3CN (50 mL). The mixture was then stirred at -20 °C for 9 h. The mixture was diluted with CH_2Cl_2 (100 mL) and water (3 × 50 mL) until the organic layer was colorless. The organic phases were collected, dried with anh. Na_2SO_4, filtered, and evaporated. The reaction crude was purified by a flash silica gel chromatographic column to obtain a yellowish syrup, compound **12** (2.03 g, 77%). R_f (hexane/EtOAc, 8:2) = 0.65.

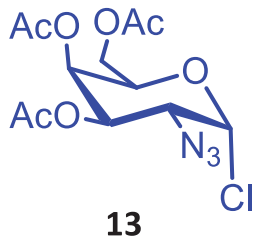
Physical properties and spectroscopic data are the same as described in the literature.[11]

HRMS (ESI) (m/z) 416.0122 $[M+Na]^+$, calculated $C_{12}H_{16}BrN_3O_7Na^+$: 416.0064.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 2.02 (s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO), 2.14 (s, 3H, CH_3CO), 3.96 (q, 1H, *J* = 10.5 Hz, H_{2s}), 4.48 (m, 1H, H_{5s}), 5.33 (q, 1H, *J* = 3.25 Hz, H_{3s}), 5.49 (q, 1H, *J* = 2.7 Hz, H_{4s}), 6.47 (d, 1H, *J* = 4.0 Hz, H_{1s}).
^{13}C NMR (100 MHz, $CDCl_3$) δ (ppm): 20.5, 20.5, 20.7 (CH_3CO), 58.7 (C_{2s}), 60.8 (C_{6s}), 66.6 (C_{4s}), 69.9 (C_{3s}), 71.5 (C_{5s}), 89 (C_{1s}), 169.6, 169.8, 170.4 (CO).

[11] M. Liu, V. G. Young, S. Lohani, D. Live, G. Barany, *Carbohydr. Res.* **2005**, *340*, 1273-1285.

3,4,6-Tri-*O*-acetyl-2-azido-1-chloro-2-α-D-deoxygalactose (13)



13

Compound **10** (5.00 g, 18.6 mmol), $FeCl_3 \cdot 6H_2O$ (14.9 g, 55 mmol) and NaN_3 (2.39 g, 36.7 mmol) were dissolved in CH_3CN (150 mL). H_2O_2 (3.44 mL, 37.2 mmol) was also added to the reaction. Then, the mixture was stirred at -30 °C for 3 h. The mixture was diluted with Et_2O (200 mL) and the organic layer was extracted with H_2O (4 × 70 mL), $NaHCO_3$ sat. (70 mL) and NaCl sat. (70 mL) until the organic layer was colorless. Yellowish syrup, compound **13** (5.79 g, 89%). R_f (hexane/EtOAc, 1:1) = 0.48.

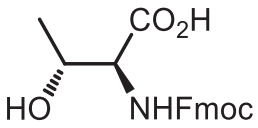
Physical properties and spectroscopic data are in agreement with earlier reports.[12]

HRMS (ESI) (m/z) 372.0603 $[M+Na]^+$, calculated $C_{12}H_{16}ClN_3O_7Na^+$: 372.0569.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 2.07 (s, 3H, CH_3CO), 2.08 (s, 3H, CH_3CO), 2.17 (s, 3H, CH_3CO), 4.17 (dd, 1H, *J* = 11.4, 3.4, H_{2s}), 4.52 (m, 1 H, H_{5s}), 5.38 (dd, 1H, *J* = 10.8, 3.2, H_{3s}), 5.51 (dd, 1H, *J* = 3.1, 1.0, H_{4s}), 6.18 (d, 1 H, *J* = 4 Hz, H_{1s}).
^{13}C NMR (100 MHz, $CDCl_3$) δ (ppm): 20.5, 20.5, 20.7 (CH_3CO), 58.5 (C_{2s}), 60.9 (C_{6s}), 66.8 (C_{4s}), 68.7 (C_{3s}), 69.7 (C_{5s}), 92.6 (C_{1s}), 169.6, 169.9, 170.5 (CO).

[12] C. Plattner, M. Höfener, N. Sewald, *Org. Lett.* **2011**, *13*, 545-547.

Fmoc-L-Thr (14)



14

L-Threonine, commercial available, (2.00 g, 16.8 mmol) and $NaHCO_3$ (2.82 g, 33.6 mmol) were dissolved in H_2O (30 mL) testing the pH was basic. Fmoc-OSu (8.5 g, 25.2 mmol) was dissolved in CH_3CN (60 mL) and then added to the reaction. The mixture was then stirred at 32 °C for 48 h. The CH_3CN was evaporated and the aqueous phase extracted with Et_2O (2 × 100 mL). The aqueous pH was decreased until acidic with HCl 2N. The solution was extracted with a mixture $CHCl_3$/^{i}PrOH (3:1). The organic phases were collected, dried with anh. Na_2SO_4, filtered, and evaporated. Compound **14** (4.99 g, 87%) was obtained as a white syrup.

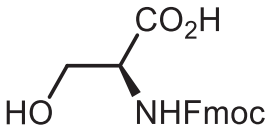
Physical properties and spectroscopic data are in agreement with earlier reports.[13]

HRMS (ESI) (m/z) 364.1201 $[M+Na]^+$, calculated $C_{19}H_{19}NO_5Na^+$: 364.1155.

^{1}H NMR (400 MHz, DMSO-*d6*) δ (ppm): 1.11 (d, 3H, *J* = 6.1 Hz, Me), 4.10 (m, 1H, OH), 4.22-4.32 (m, 3H, H_α, H_β, CH_{Fmoc}), 4.55-4.63 (m, 2H, $CH_{2\ Fmoc}$), 5.72 (br s, 1H, NH), 7.39-7.99 (m, 8 H, aryl-H).

[13] I. Schön, L. Kisfaludy, *Synthesis* **1986**, *1986*, 303-305.

Fmoc-L-Ser (15)



15

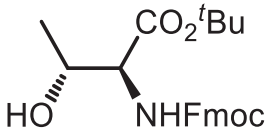
L-serine, commercial available, (2.10 g, 20 mmol) and $NaHCO_3$ (3.36 g, 40.0 mmol) were dissolved in H_2O (30 mL) testing the pH was basic. Fmoc-OSu (12.7 g, 30 mmol) was dissolved in CH_3CN (60 mL) and then added to the reaction. The mixture was then stirred at 32 °C for 48 h. The CH_3CN was evaporated and the aqueous phase extracted with Et_2O (2 × 100 mL). The aqueous pH was decreased until acidic with HCl 2N. The solution was extracted with a mixture $CHCl_3$/^{i}PrOH (3:1). The organic phases were collected, dried with anh. Na_2SO_4, filtered, and evaporated. Compound **15** (5.70 g, 87%) was obtained as a white syrup.

Physical properties and spectroscopic data are in agreement with earlier reports.[5]

HRMS (ESI) (m/z) 350.1025 $[M+Na]^+$, calculated $C_{18}H_{17}NO_5Na^+$: 350.0999.

^{1}H NMR (400 MHz, DMSO-*d6*) δ (ppm): 2.97-3.95 (m, 1H, H_β), 4.09-4.16 (m, 2H, H_β, OH), 4.43 (t, 1H, *J* = 7.0 Hz, CH_{Fmoc}), 4.40-4.55 (m, 3H, H_α, $CH_{2\ Fmoc}$), 5.84 (br s, 1H, NH), 7.40-7.96 (m, 8 H, aryl-H).

Fmoc-L-Thr-^{t}Bu (16)



16

DCC (4.88 g, 23.4 mmol) and CuCl (72.5 mg, 0.73 mmol) was dissolved in ^{t}BuOH (2.85 mL, 30 mmol) under argon atmosphere. Then, the reaction was stirred during 5 days in darkness. Compound **14** (2.50 g, 7.3 mmol) was dissolved in dry CH_2Cl_2 (15 mL) and added dropwise to the reaction. The mixture was stirred for 4 hours. The reaction was then filtered through Celite and extracted with $NaHCO_3$ (3 × 100 mL). The organic phases were collected, dried with anh. Na_2SO_4, filtered, and evaporated. The reaction crude was purified by a flash silica gel chromatographic column to obtain a white syrup, compound **16** (2.03 g, 70%).

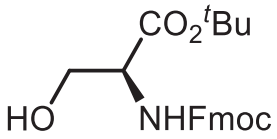
Physical properties and spectroscopic data are in agreement with earlier reports.[14]

HRMS (ESI) (m/z) 398.1993 $[M+H]^+$, calculated $C_{23}H_{28}NO_5^+$: 398.1962.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 1.24 (d, 3H, *J* = 6.3 Hz, Me), 1.46 (s, 9H, ^{t}Bu), 2.08 (bs, 1H, OH), 4.20-4.30 (m, 2H, H_α, H_β), 4.36 (m, 1H, aryl-H), 4.40 (d, 2H, *J* = 7.0 Hz, $CH_{2\ Fmoc}$), 5.63 (br s, 1H, NH), 7.27-7.32 (m, 2 H, aryl-H), 7.38-7.80 (m, 6 H, aryl-H).

[14] a) H. Paulsen, K. Adermann, *Liebigs Ann. Chem.* **1989**, *1989*, 751-769; b) I. Tavernaro, S. Hartmann, L. Sommer, H. Hausmann, C. Rohner, M. Ruehl, A. Hoffmann-Roeder, S. Schlecht, *Org. Biomol. Chem.* **2015**.

Fmoc-L-Ser-^{t}Bu (17)



17

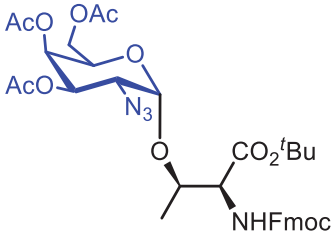
DCC (5.81 g, 27.8 mmol) and CuCl (86.4 mg, 0.87 mmol) was dissolved in ^{t}BuOH (3.39 mL, 36 mmol) under argon atmosphere. Then, the reaction was stirred during 5 days in darkness. Compound **15** (2.85 g, 8.7 mmol) was dissolved in dry CH_2Cl_2 (15 mL) and added dropwise to the reaction. The mixture was stirred for 4 hours. The reaction was then filtered through Celite and extracted with $NaHCO_3$ (3 × 100 mL). The organic phases were collected, dried with anh. Na_2SO_4, filtered, and evaporated. The reaction crude was purified by a flash silica gel chromatographic column to obtain a white syrup, compound **17** (2.47 g, 74%).

Physical properties and spectroscopic data are in agreement with earlier reports.[6]

HRMS (ESI) (m/z) 384.1885 $[M+H]^+$, calculated $C_{22}H_{26}NO_5^+$: 384.1805.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 1.45 (s, 9H, ^{t}Bu), 3.76 (dd, 1H, *J* = 8.5, 5.5 Hz, H_β), 4.02 (dd, 1H, *J* = 8.8, 3.1 Hz, H_β), 4.23 (t, 1H, *J* = 7.0 Hz, CH_{Fmoc}), 4.30-4.46 (m, 3H, H_α, $CH_{2\ Fmoc}$), 5.80 (d, 1 H, *J* = 7.6 Hz, NH), 7.38-8.02 (m, 8 H, aryl-H).

Fmoc-L-Thr(α-*O*-D-tri-*O*-acetyl-2-azido-2-deoxygalactosyl)-^{t}Bu (18α)



18α

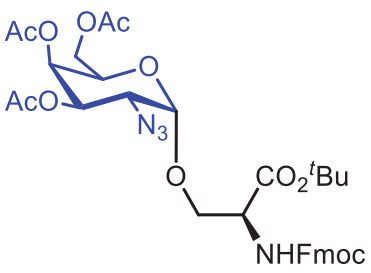
Compound **16** (500 mg, 1.26 mmol) was dissolved in a mixture toluene/CH_2Cl_2 4:6 (20 mL) with 4 Å molecular sieves at inert atmosphere stirring for 1 h. The reaction was then placed into a cold bath and Ag_2CO_3 (381 mg, 1.38 mmol) and $AgClO_4$ (34 mg, 0.164 mmol) were added and the mixture was stirred for 1 h. Compound **12** (496 mg, 1.26 mmol) was dissolved in toluene/CH_2Cl_2 1:1 (20 mL) and added to the mixture. The reaction was stirred overnight. The crude was filtered through Celite and the organic phase was extracted with sat. $NaHCO_3$ (1 × 100 mL) and H_2O (2 × 100 mL). The organic phases were collected, dried with anh. Na_2SO_4, filtered, and evaporated. The reaction crude was purified by a flash silica gel chromatographic column to obtain compound **18α** as a white foam (440 mg, 49%). R_f (toluene/acetone, 10:1) = 0.7, R_f (hexane/EtOAc, 65:35) = 0.65.

HRMS (ESI) (m/z) 733.2778 $[M+Na]^+$, calculated $C_{35}H_{42}N_4O_{12}Na^+$: 733.2691.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 1.29 (d, 3H, *J*= 6.2 Hz, Me), 1.51 (s, 9H, ^{t}Bu), 2.03 (s, 3H,CH_3CO), 2.08 (s, 3H, CH_3CO), 2.14 (s, 3H, CH_3CO), 3.61 (dd, 1H, *J* = 11.2, 3.5 Hz, H_{2s}), 4.03 (m, 1H, H_{5s}), 4.09-4.15 (m, 2H, H_{6s}), 4.21-4.27 (m, 2H, $CH_{2\ Fmoc}$, CH_{Fmoc}), 4.39-4.45 (m, 3H, H_α, H_β, $CH_{2\ Fmoc}$), 5.17 (d, 1H, *J* = 3.6 Hz, H_{1s}), 5.31 (m, 1H, H_{3s}), 5.42 (m, 1H, H_{4s}), 5.54 (d, 1H, *J* = 9.5 Hz, NH), 7.29-7.42 (m, 4H, aryl-H), 7.60-7.63 (m, 2H, aryl-H), 7.75-7.78 (m, 2H, aryl-H).

^{13}C NMR (100 MHz, $CDCl_3$) δ (ppm): 15.0 (Me), 20.8, 20.8 (2×CH_3CO), 21.2 (CH_3CO), 28.1 (^{t}Bu), 47.3 (CH_{Fmoc}), 57.4 (C_{2s}), 59.2 (C_α), 60.5 (C_{5s}), 61.7 (C_{6s}), 67.4 ($CH_{2\ Fmoc}$), 67.5 (C_{3s}), 67.7 (C_{4s}), 72.6 ($C_{\beta\ Thr}$), 83.3 (C_{tBu}), 95.0 (C_{1s}), 120.0, 120.0, 125.3, 125.3, 127.2, 127.2, 127.9, 127.9, 141.5, 141.5, 143.9, 144.0 (aryl), 156.7 (CO_{Fmoc}), 169.3, 169.9, 170.2, 170.5 (CO).

Fmoc-L-Ser(α-*O*-D-tri-*O*-acetyl-2-azido-2-deoxygalactosyl)-^{t}Bu (19α)



19α

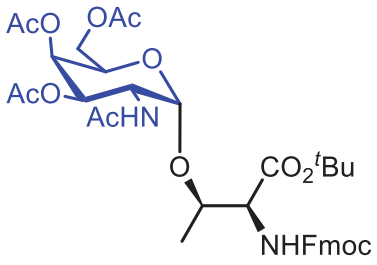
Compound **17** (500 mg, 1.30 mmol) and Ag_2CO_3 (538 mg, 1.95 mmol) were dissolved in a mixture of toluene/CH_2Cl_2 4:6 (20 mL) with 4 Å molecular sieves under inert atmosphere at 0 °C stirring for 30 min. $AgClO_4$ (68 mg, 0.33 mmol) was then added dropwise. Compound **13** (799 mg, 2.28 mmol) was added dissolved in toluene/CH_2Cl_2 1:1 (30 mL). The reaction was stirred overnight in darkness. The mixture was filtered through Celites and the organic phase was washed with sat. $NaHCO_3$ (1 × 100 mL) and H_2O (2 × 100 mL). White foam, compound **19α** (453 mg, 50%). R_f (toluene/acetone, 10:1) = 0.7, R_f (hexane/EtOAc, 65:35) = 0.65.

HRMS (ESI) (m/z) 719.2606 $[M+Na]^+$, calculated $C_{34}H_{20}N_4O_{12}Na^+$: 719.2535.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 1.51 (s, 9H, ^{t}Bu), 1.98 (s, 3H,CH_3CO), 2.06 (s, 3H, CH_3CO), 2.15 (s, 3H, CH_3CO), 3.66 (dd, 1H, *J* = 11.1, 3.5 Hz, H_{2s}), 3.92-3.99 (m, 1H, H_β) 4.01-4.09 (m, 2H, H_{6s}), 4.16-4.21 (m, 2H, H_{5s}), 4.21-4.27 (m, 1H, CH_{Fmoc}), 4.37-4.49 (m, 3H, H_α, $CH_{2\ Fmoc}$), 4.95 (d, 1H, *J* = 3.5 Hz, H_{1s}), 5.28-5.32 (m, 1H, H_{3s}), 5.45 (dd, 1H, *J* = 3.2, 0.8, H_{4s}), 5.90 (d, 1H, *J* = 7.8 Hz, NH), 7.30-7.42 (m, 4H, aryl-H), 7.61-7.65 (m, 2H, aryl-H), 7.74-7.77 (m, 2H, aryl-H).

^{13}C NMR (100 MHz, $CDCl_3$) δ (ppm): 20.7, 20.8 (2×CH_3CO), 21.2 (CH_3CO), 28.0 (^{t}Bu), 47.2 (CH_{Fmoc}), 57.4 (C_{2s}), 57.6 (C_α), 60.5 (C_{5s}), 61.7 (C_{6s}), 67.4 ($CH_{2\ Fmoc}$), 67.4 (C_{3s}), 67.7 (C_{4s}), 69.0 (C_β), 83.2 (C_{tBu}), 98.0 (C_{1s}), 119.9, 125.1, 127.0, 127.7, 141.3, 141.5, 143.7, 143.9 (aryl), 155.7 (CO_{Fmoc}), 168.3, 169.6, 170.0, 170.3 (CO).

Fmoc-L-Thr(α-*O*-D-tri-*O*-acetyl-*N*-acetylgalactosaminyl)-^{t}Bu (20)



20

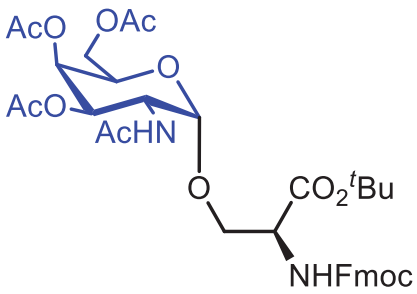
Compound **18α** (440 mg, 0.617 mmol) was dissolved in THF-acetic anhydride-acetic acid mixture (15 mL, 3:2:1) and zinc (411 mg, 6.30 mmol) activated with 2% aq. $CuSO_4$. The mixture was then stirred at room temperature for 2 h before being diluted with THF (50 mL), filtered through Celite and concentrated. The reaction crude was purified by a flash silica gel chromatographic column to obtain compound **20** as a white foam (372 mg, 83%). R_f (hexane/EtOAc, 2:8) = 0.75.

HRMS (ESI) (m/z) 727.3115 $[M+H]^+$, calculated $C_{37}H_{47}N_2O_{13}^+$: 727.3073.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 1.18 (d, 3H, *J* = 7.2 Hz, Me), 1.48 (s, 9H, ^{t}Bu), 1.95 (s, 3H, CH_3CO), 1.99 (s, 3H, CH_3CO), 2.01 (s, 3H, CH_3CO), 2.16 (s, 3H, CH_3CO), 4.03 (m, 1H, *J* = 9.8, 6.6 Hz, H_{5s}), 4.09-4.15 (m, 2H, H_{6s}), 4.18-4.26 (m, 2H, CH_{Fmoc}, H_β), 4.33 (dd, 1H, *J* = 9.3, 3.5 Hz, H_α), 4.48-4.60 (m, 3H, $CH_{2\ Fmoc}$, H_{2s}), 5.01 (d, 1H, *J* = 3.6 Hz, H_{1s}), 5.06 (dd, 1H, *J* = 11.4, 3.2 Hz, H_{3s}), 5.31 (m, 1H, H_{4s}), 5.57 (d, 1H, *J* = 9.2 Hz, NH_{Fmoc}), 5.90 (d, 1H, *J* = 7.4 Hz, NH_s), 7.29-7.44 (m, 4H, aryl-H), 7.58-7.65 (m, 2H, aryl-H), 7.75-7.79 (m, 2H, aryl-H).

^{13}C NMR (100 MHz, $CDCl_3$) δ (ppm):15.5 (C_{Me}), 20.7 (CH_3CO), 20.9, 20.9 (CH_3CO), 23.3 (CH_3CO), 28.1 ($CH_{3\ tBu}$), 47.4 (CH_{Fmoc}), 47.8 (C_{2s}), 59.2 (C_α), 62.0 (C_{5s}), 66.9 (C_{6s}), 67.3 ($CH_{2\ Fmoc}$), 67.4 (C_{4s}), 68.3 (C_{3s}), 72.5 (C_β), 83.3 (C_{tBu}), 94.5 (C_{1s}), 120.2, 120.2, 125.0, 125.1, 127.3, 127.3, 127.9, 127.9, 141.5, 141.5, 143.9, 144.0 (aryl), 156.6, 169.5, 170.3, 170.5, 171.1, 171.3 (CO).

Fmoc-L-Ser(α-*O*-D-tri-*O*-acetyl-*N*-acetylgalactosaminyl)-^{t}Bu (21)



21

Compound **19α** (453 mg, 0.650 mmol) was dissolved in THF-acetic anhydride-acetic acid mixture (15 mL, 3:2:1) and zinc (434 mg, 6.63 mmol) activated with 2% aq. $CuSO_4$. The mixture was then stirred at room temperature for 2 h before being diluted with THF (50 mL), filtered through Celite and concentrated. The reaction crude was purified by a flash silica gel chromatographic column to obtain compound **21** as a white foam (394 mg, 85%). R_f (hexane/EtOAc, 2:8) = 0.75.

HRMS (ESI) (m/z) 713.3002 $[M+H]^+$, calculated $C_{36}H_{45}N_2O_{13}^+$: 713.2926.

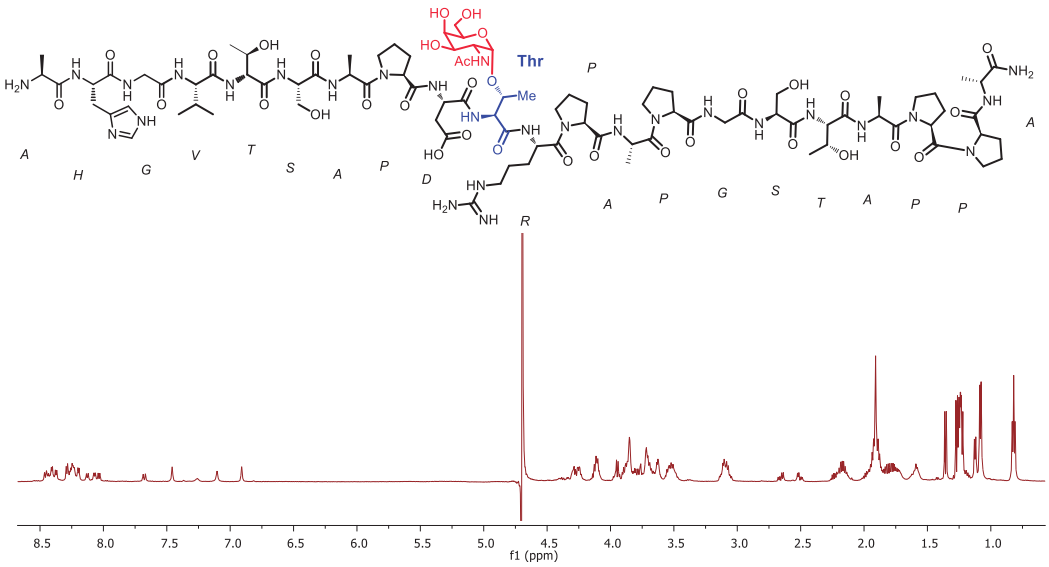
^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 1.49 (s, 9H, ^{t}Bu), 1.94 (s, 3H, CH_3CO), 1.99 (s, 3H, CH_3CO), 2.10 (s, 3H, CH_3CO), 2.16 (s, 3H, CH_3CO), 3.82-3.87 (m, 1H, H_β), 3.95-3.99 (m, 1H, H_β), 4.03-4.10 (m, 2H, H_{6s}), 4.12-4.20 (m, 1H, H_{5s}), 4.24 (t, 1H, *J* = 7.0 Hz, CH_{Fmoc}), 4.58-4.62 (m, 1H, H_{2s}), 4.85 (d, 1H, *J* = 3.6 Hz, H_{1s}), 5.13 (dd, 1H, *J* = 11.4, 3.6 Hz, H_{3s}), 5.38 (dd, 1H, *J* = 3.2, 0.8 Hz, H_{4s}), 5.84 (d, 1H, *J* = 9.6 Hz, NH_s), 5.87 (d, 1H, *J* = 8.0 Hz, NH_{Fmoc}), 7.29-7.42 (m, 4H, aryl-H), 7.60-7.64 (m, 2H, aryl-H), 7.74-7.78 (m, 2H, aryl-H).

^{13}C NMR (100 MHz, $CDCl_3$) δ (ppm): 20.6 (CH_3CO), 20.7, 20.7 (CH_3CO), 23.2 (CH_3CO), 28.1 ($CH_{3\ tBu}$), 47.1 ($CH_{\ Fmoc}$), 47.5 (C_{2s}), 54.4 (C_α), 62.0 (C_{5s}), 66.9 (C_{6s}), 67.3 ($CH_{2\ Fmoc}$), 67.4 (C_{4s}), 68.3 (C_{3s}), 69.5 (C_β), 83.3 (C_{tBu}), 98.5 (C_{1s}), 120.2, 125.0, 127.3, 127.3, 128.9, 128.9, 134.8, 141.5, 143.6 (aryl), 156.4, 169.7, 170.3, 170.5, 171.1, 171.2 (CO).

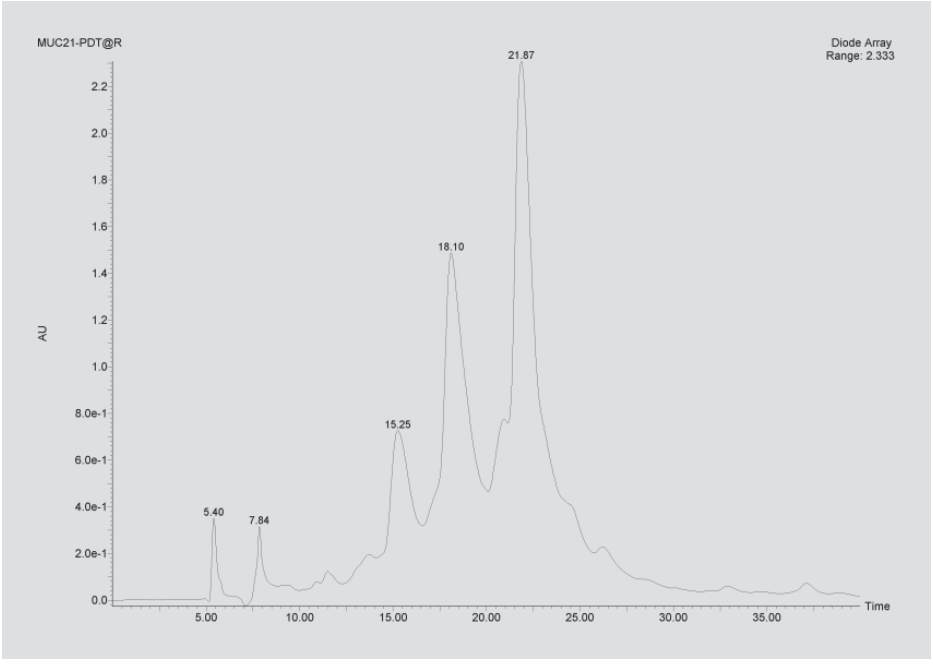
XII. NMR spectra and chromatograms

Compound 1 (MUC1-APDT*R)

^{1}H NMR 400 MHz in D_2O registered at 298K

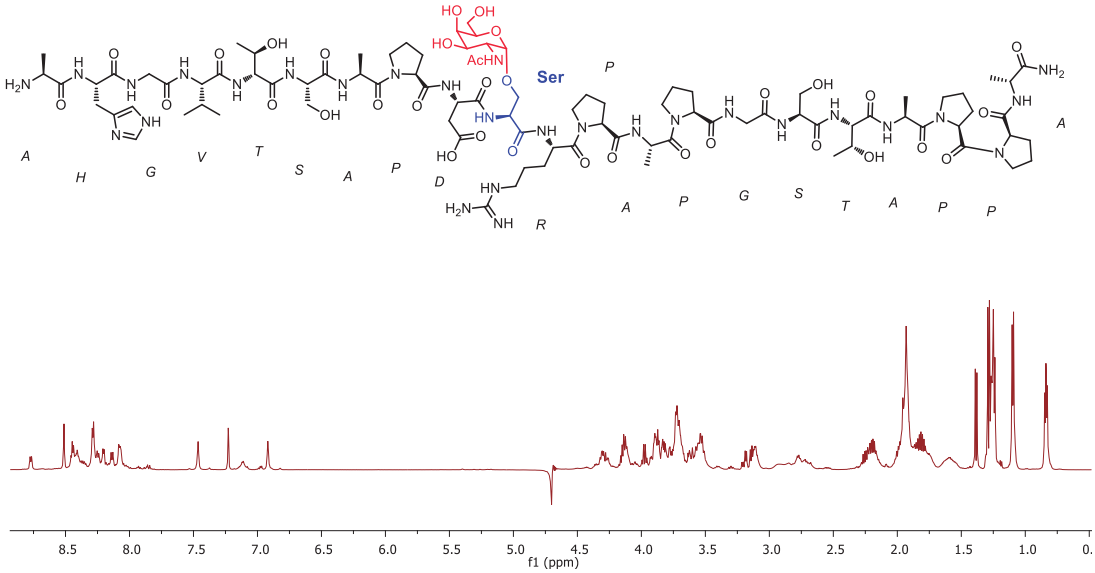


HPLC chromatogram

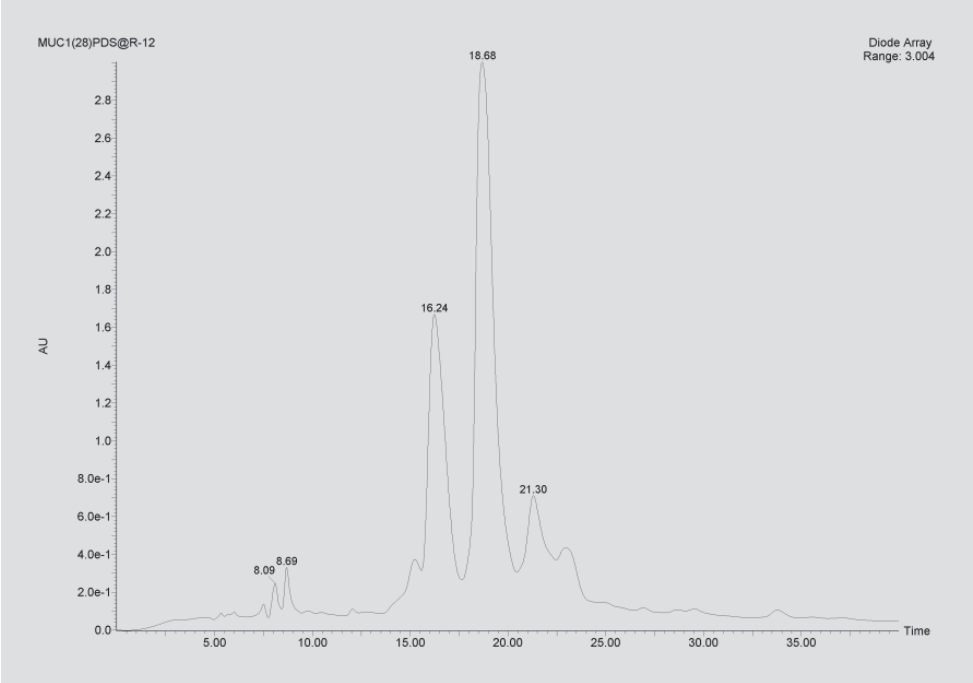


Compound 2 (MUC1-APDS*R)

^{1}H NMR 400 MHz in D_2O registered at 298K

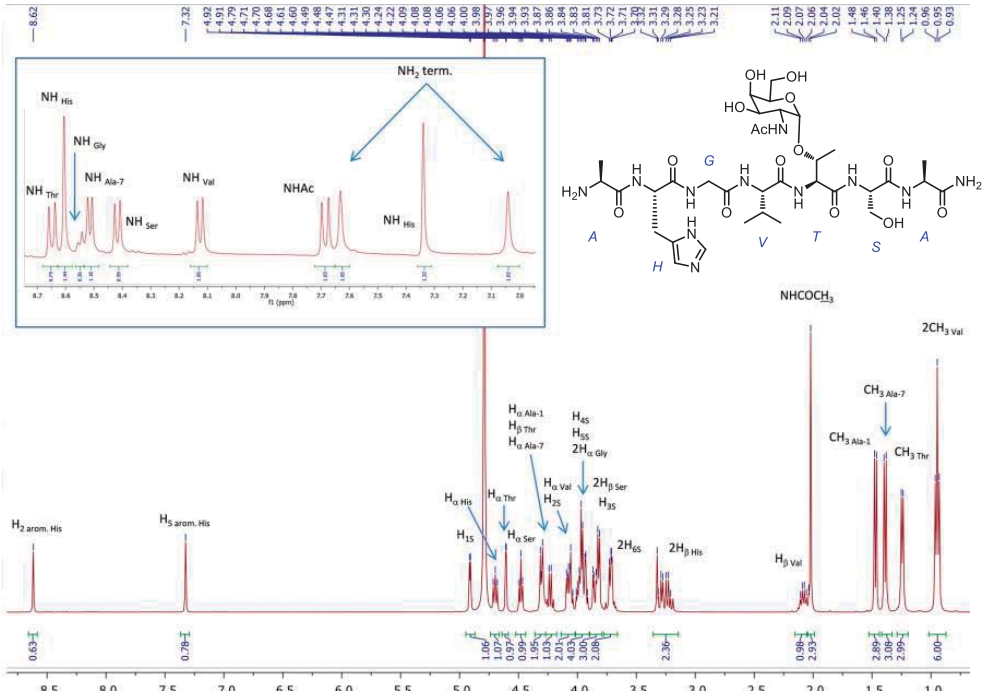


HPLC chromatogram

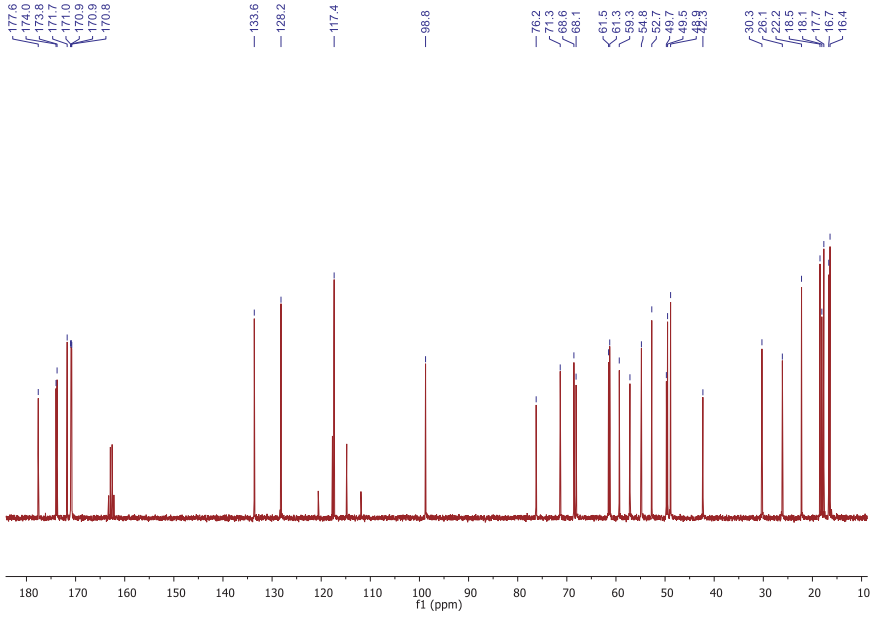


L-Ala-L-His -L-Gly- L-Val-L-Thr-(α-*O*-D-GalNAc)-L-Ser-L-Ala-NH_2 (3)

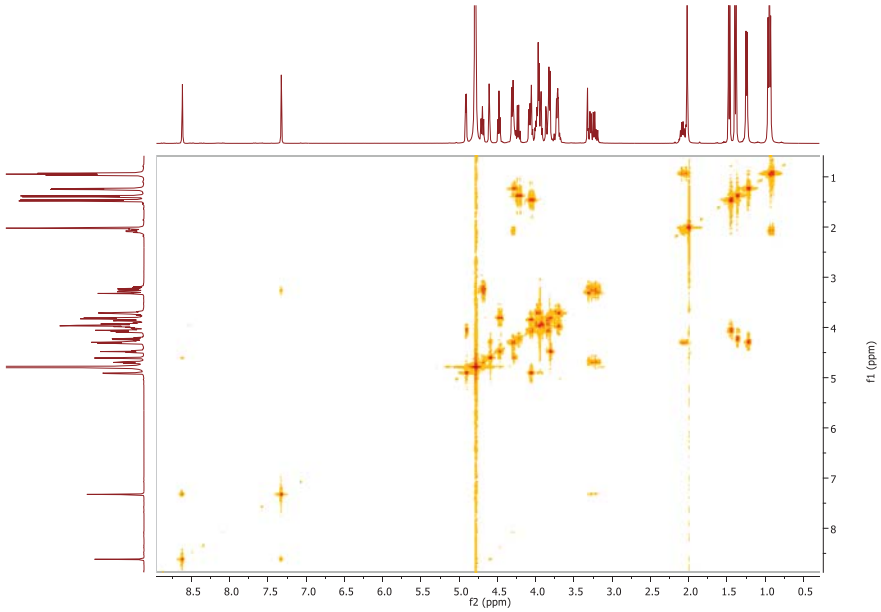
^{1}H NMR 400 MHz in D_2O registered at 298K



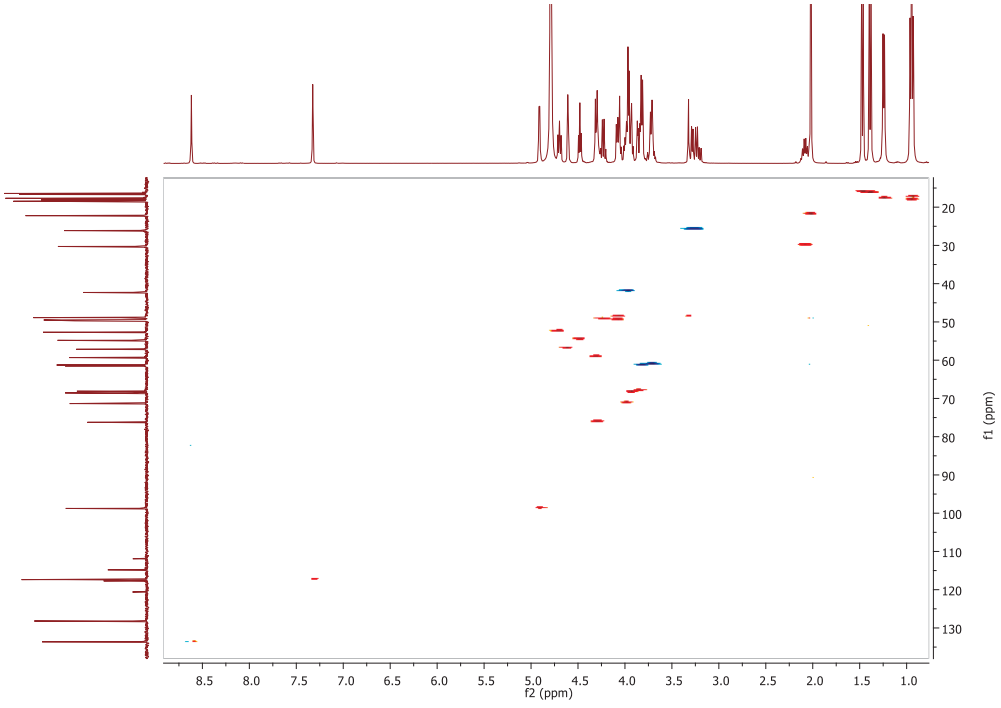
^{13}C NMR 100 MHz in D_2O registered at 298K



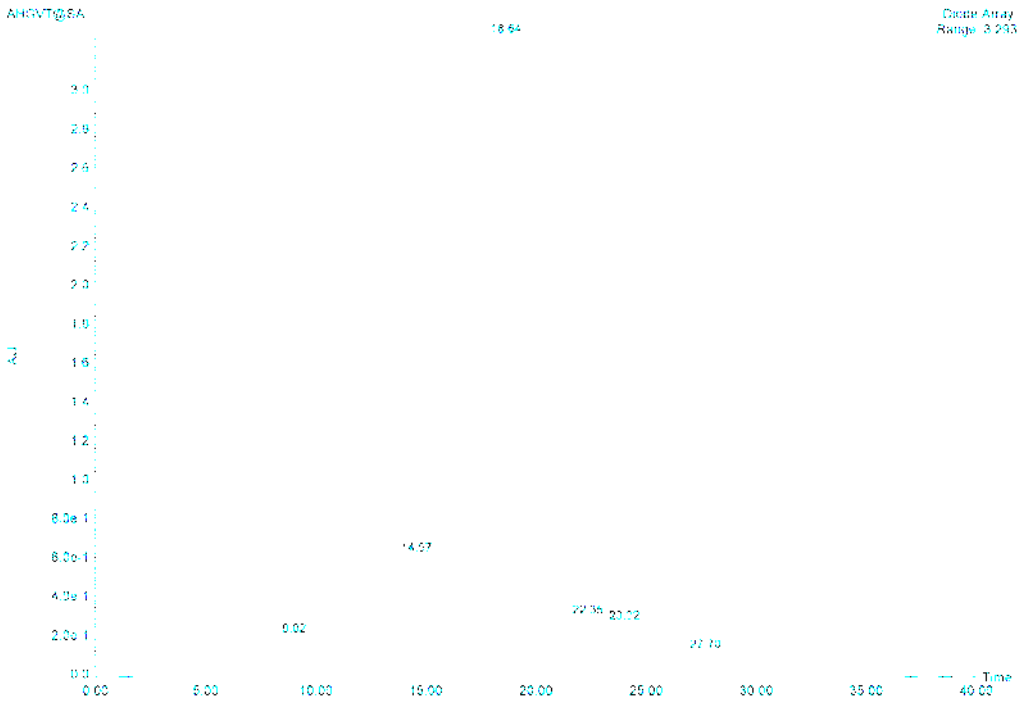
COSY in D_2O registered at 298K



HSQC in D_2O registered at 298K

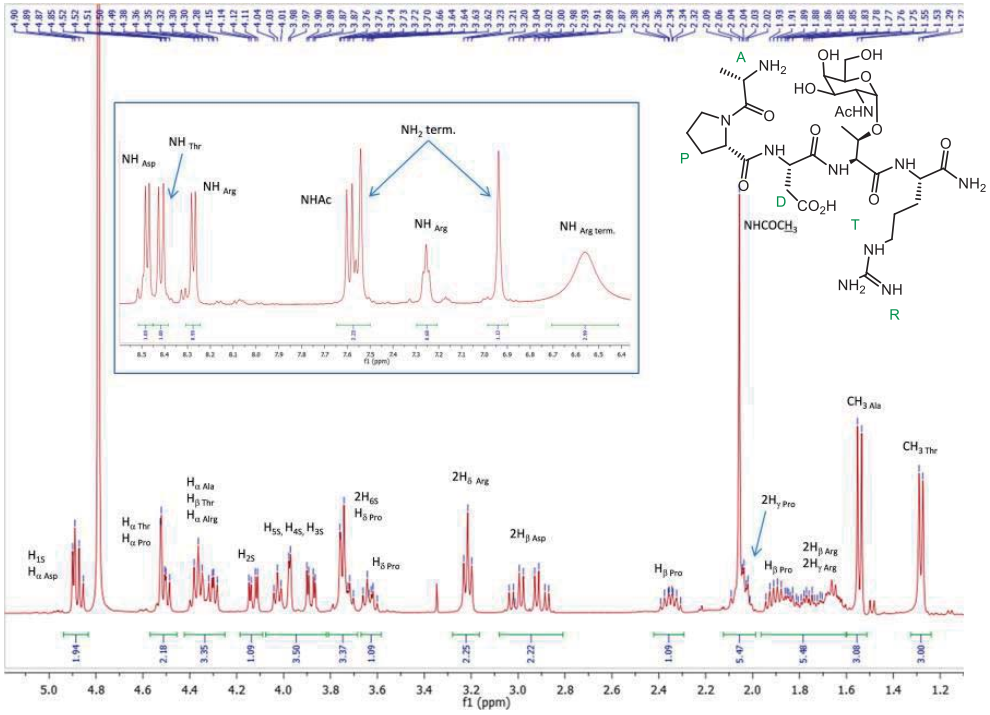


HPLC chromatogram

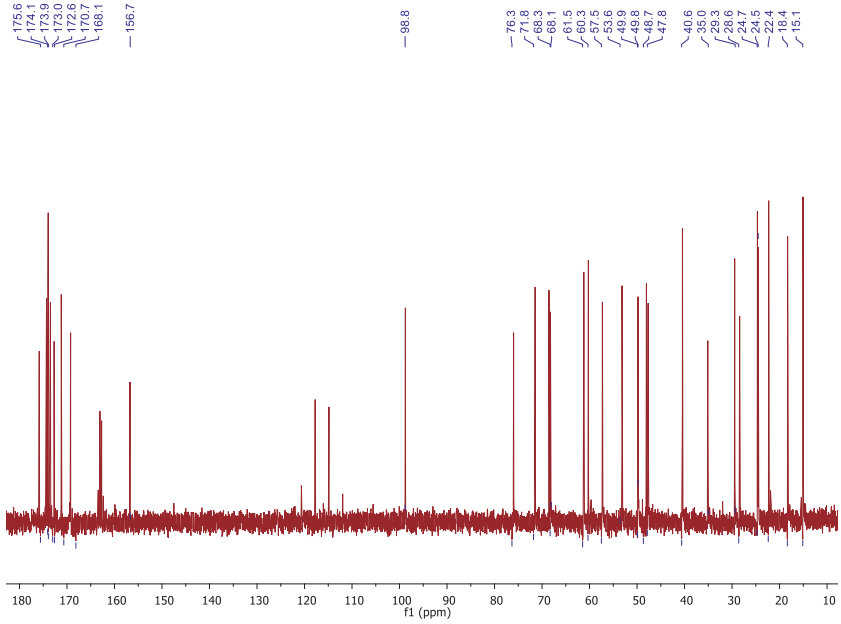


L-Ala-L-Pro-L-Asp-L-Thr-(α-*O*-D-GalNAc)-L-Arg-NH_2 (4)

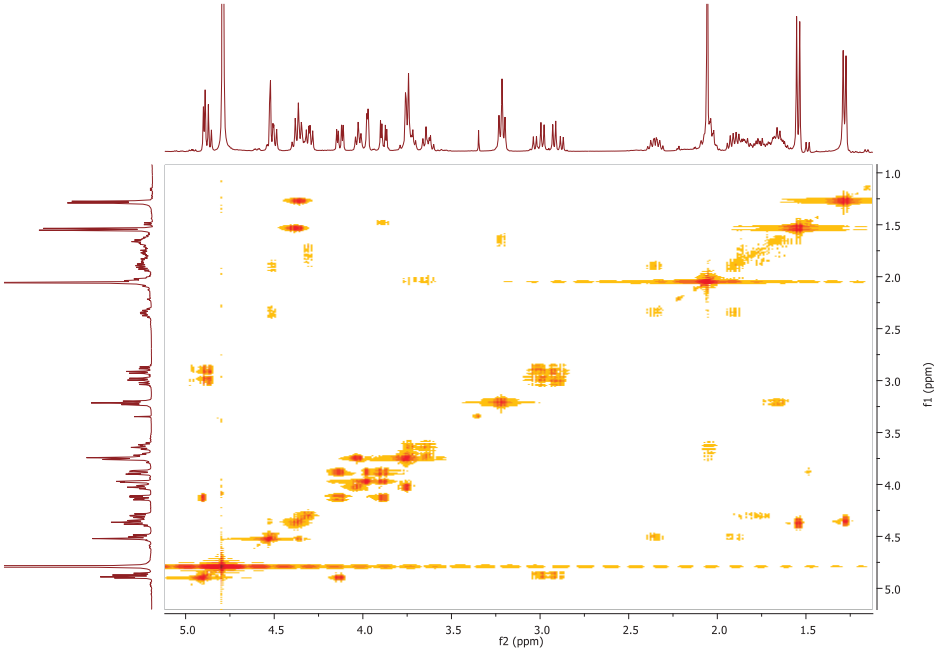
^{1}H NMR 400 MHz in D_2O registered at 298K



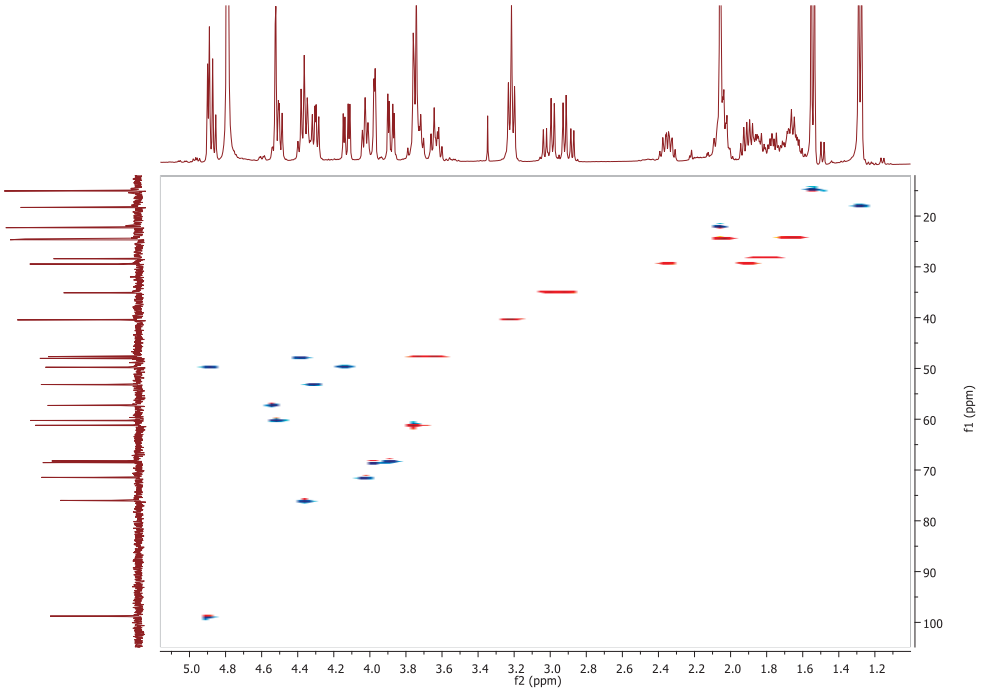
^{13}C NMR 100 MHz in D_2O registered at 298K



COSY in D_2O registered at 298K



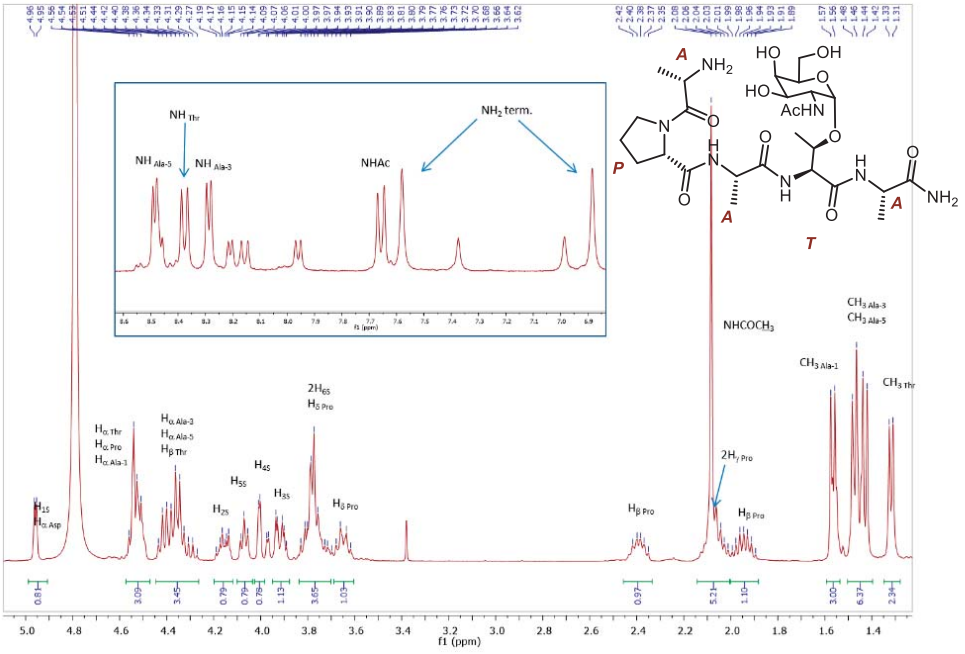
HSQC in D_2O registered at 298K



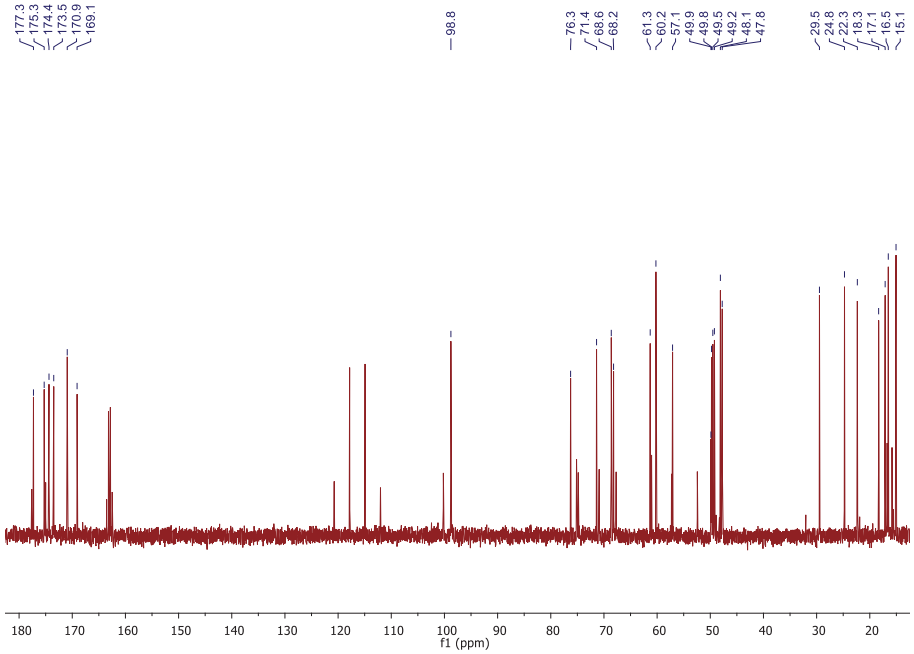
HPLC chromatogram

L-Ala-L-Pro-L-Ala-L-Thr-(α-*O*-D-GalNAc)-L-Ala-NH_2 (4')

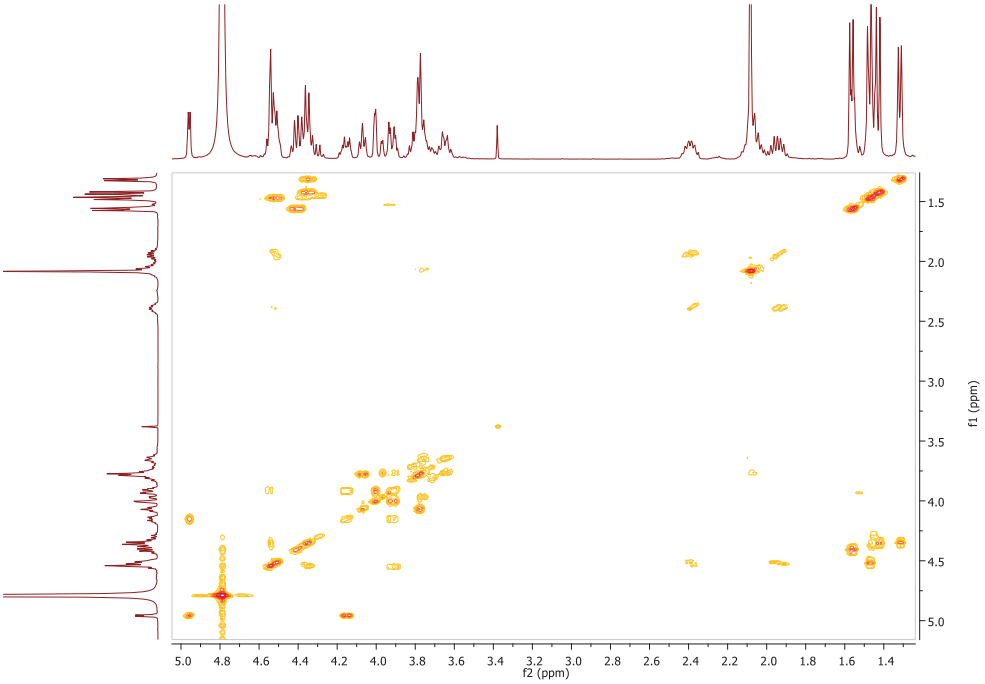
^{1}H NMR 400 MHz in D_2O registered at 298K



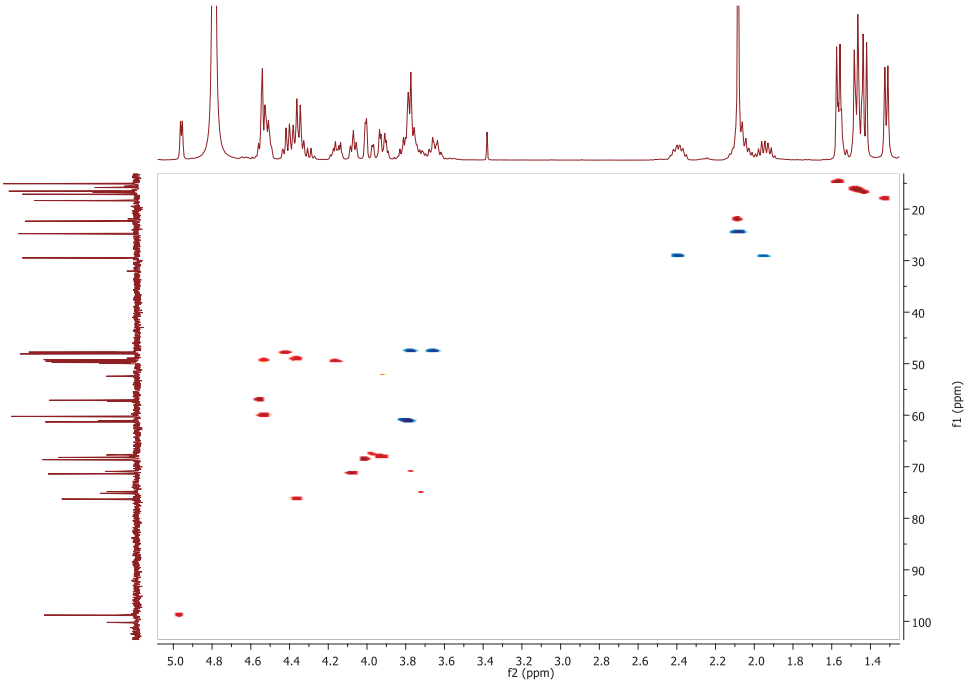
^{13}C NMR 100 MHz in D_2O registered at 298K



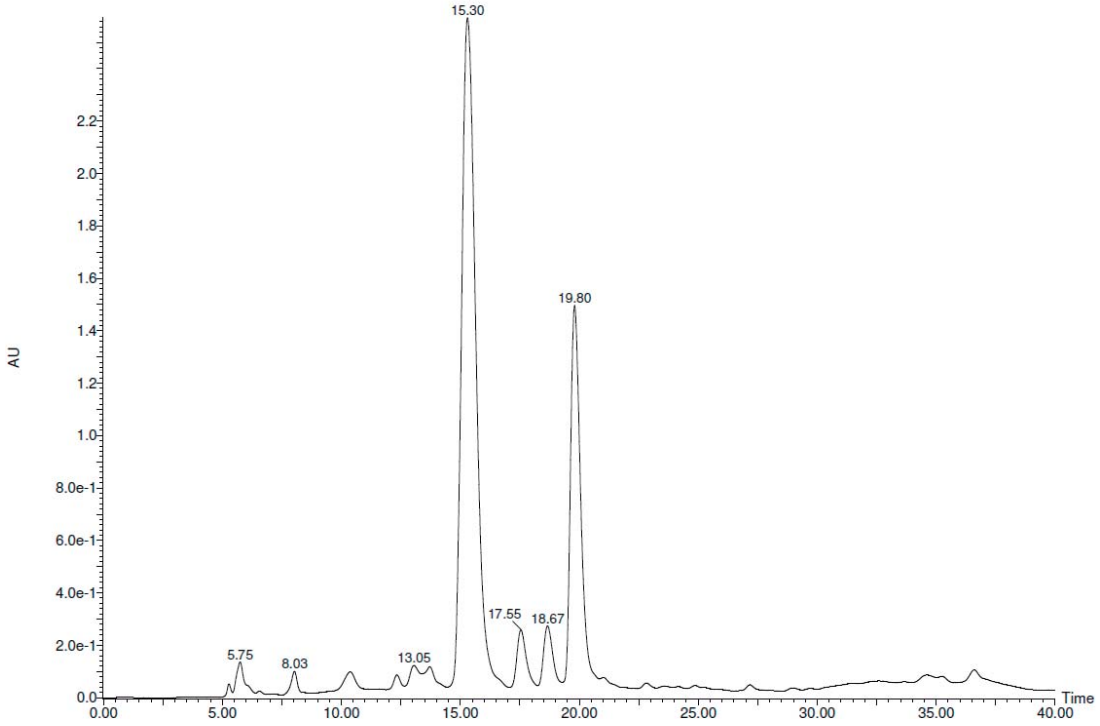
COSY in D_2O registered at 298K



HSQC in D_2O registered at 298K

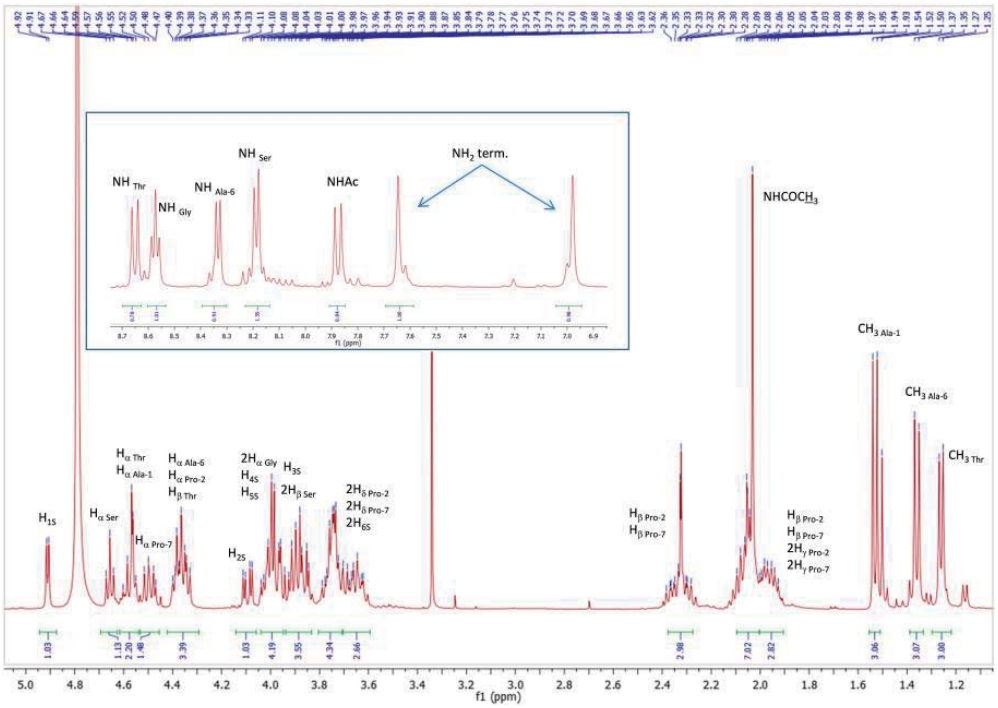


HPLC chromatogram

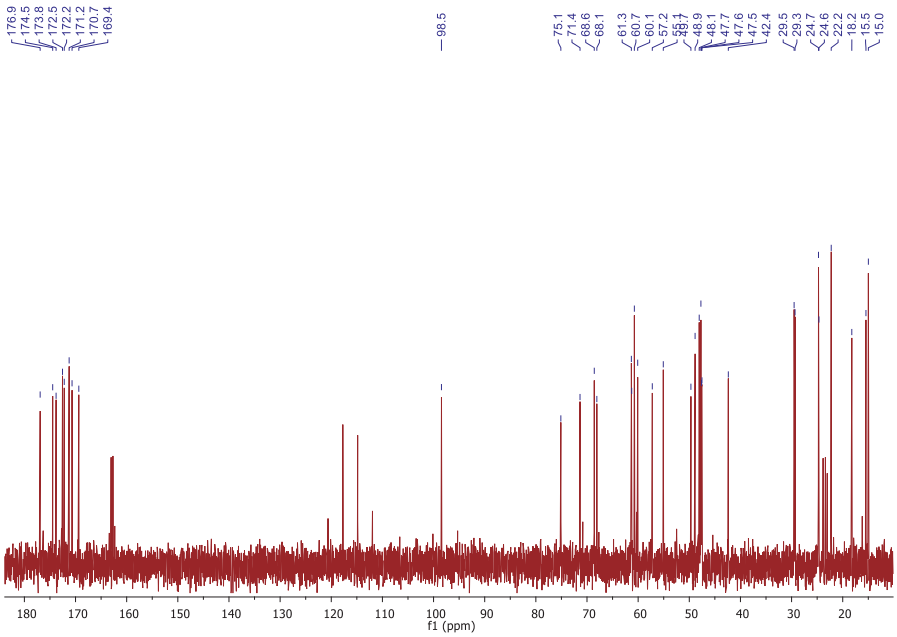


L-Ala-L-Pro-L-Gly-L-Ser-L-Thr-(α-*O*-D-GalNAc)-L-Ala-L-Pro-NH_2 (5)

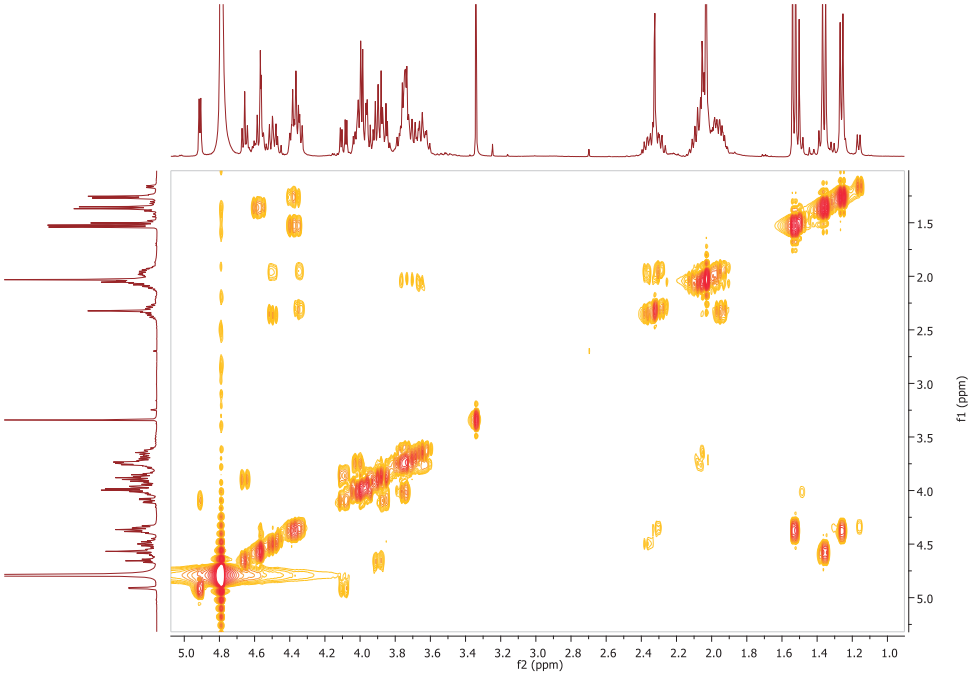
^{1}H NMR 400 MHz in D_2O registered at 298K.



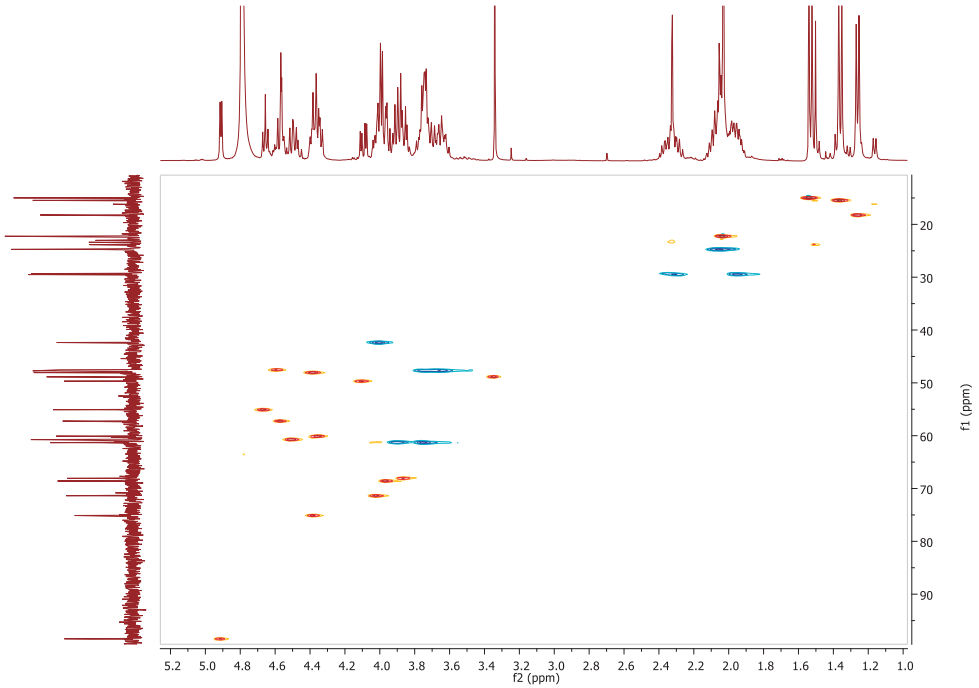
^{13}C NMR 100 MHz in D_2O registered at 298K



COSY in D_2O registered at 298K



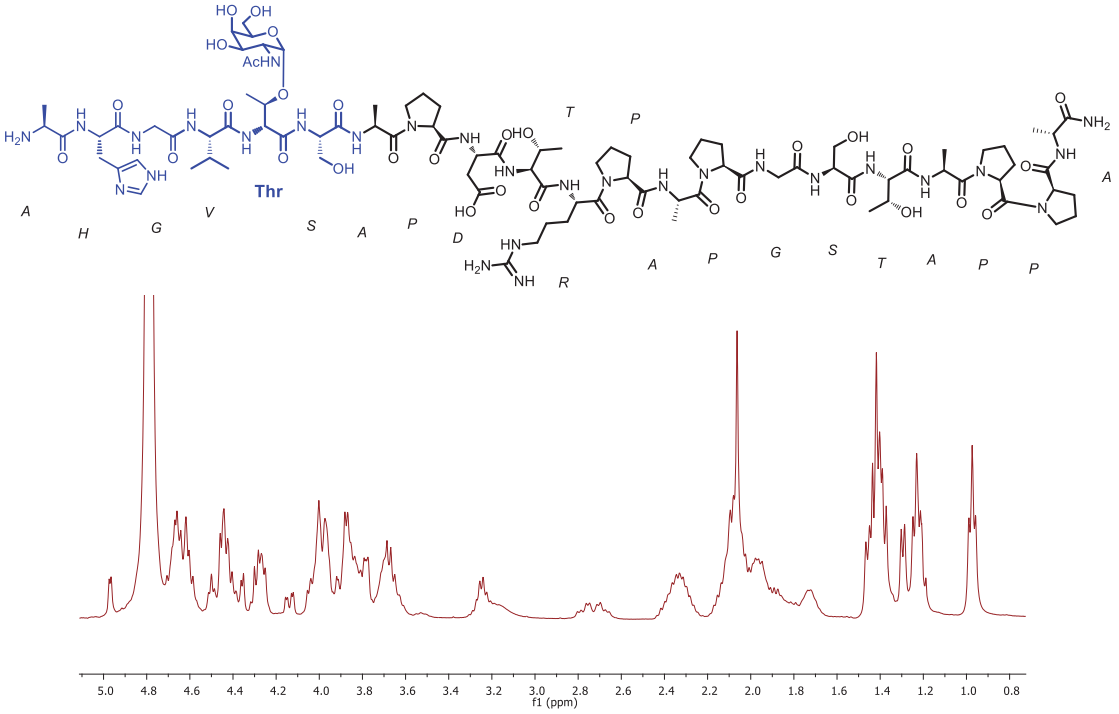
HSQC in D_2O registered at 298K



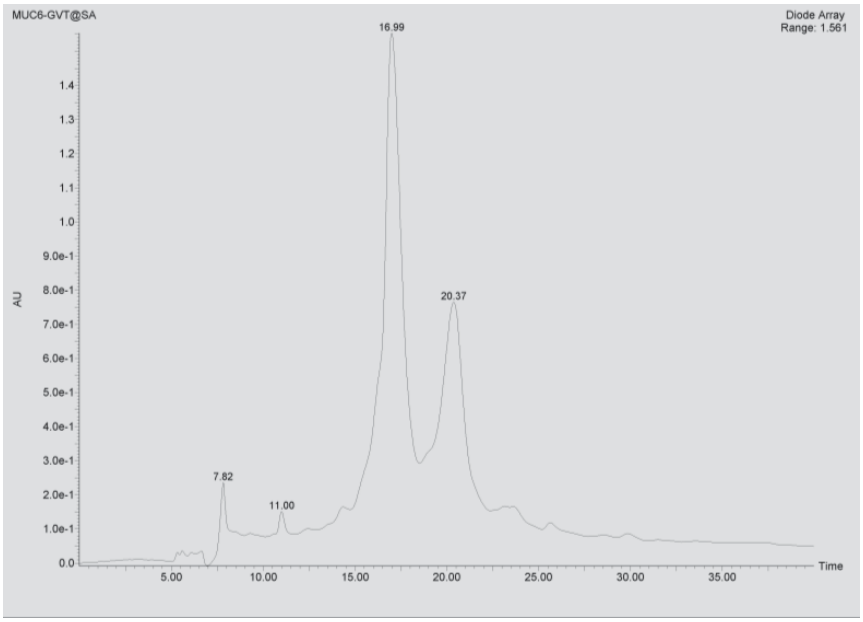
HPLC chromatogram

Compound 6 (MUC1-AHGVT*SA)

^{1}H NMR 400 MHz in D_2O registered at 298K

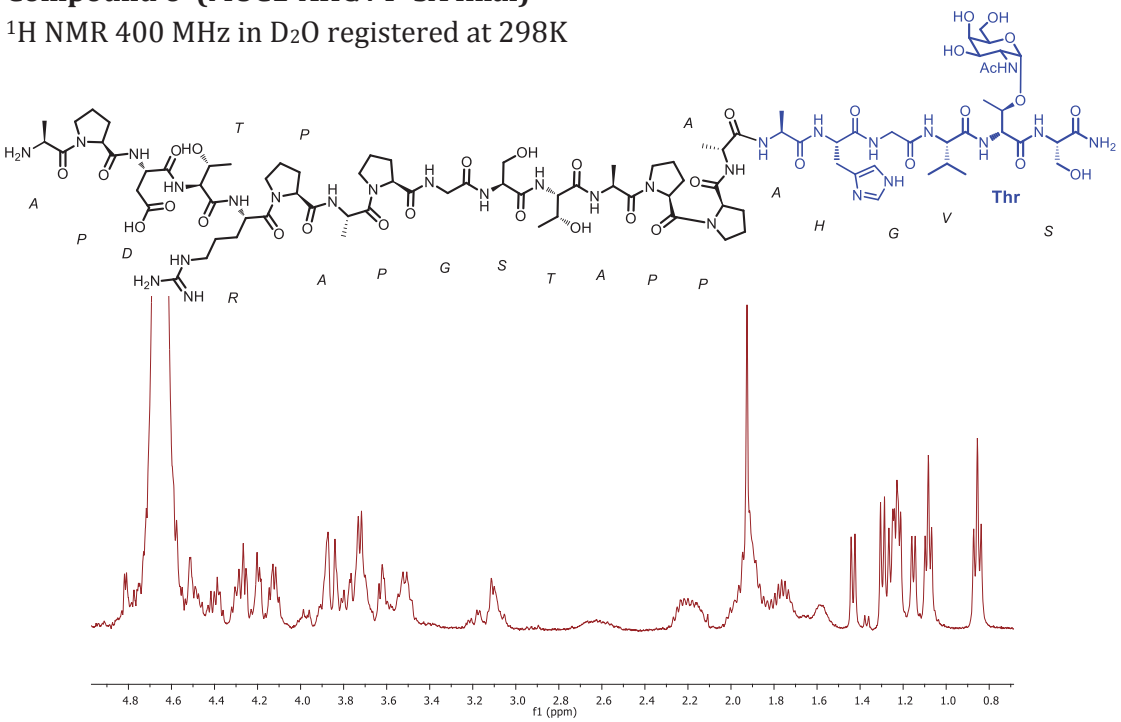


HPLC chromatogram

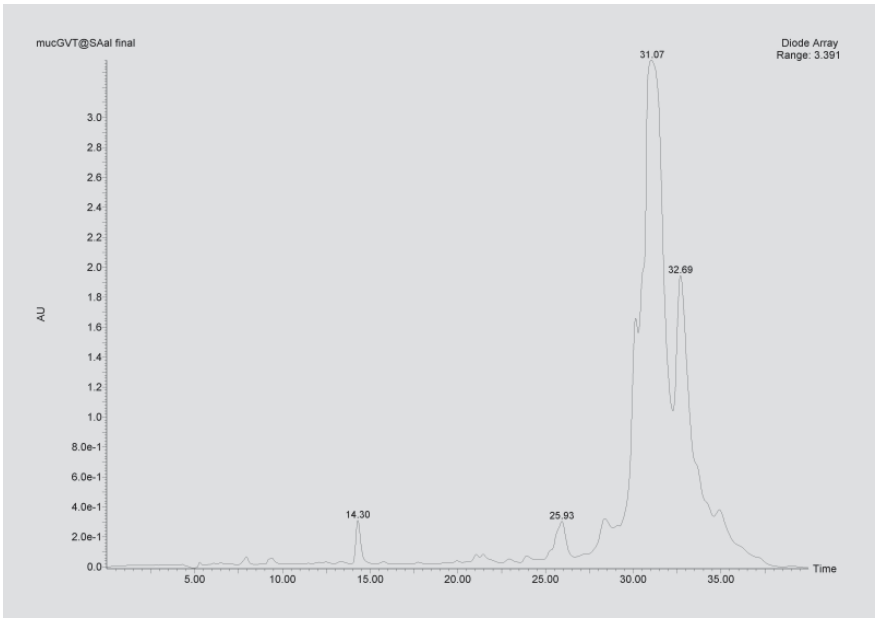


Compound 6' (MUC1-AHGVT*SA final)

^{1}H NMR 400 MHz in D_2O registered at 298K

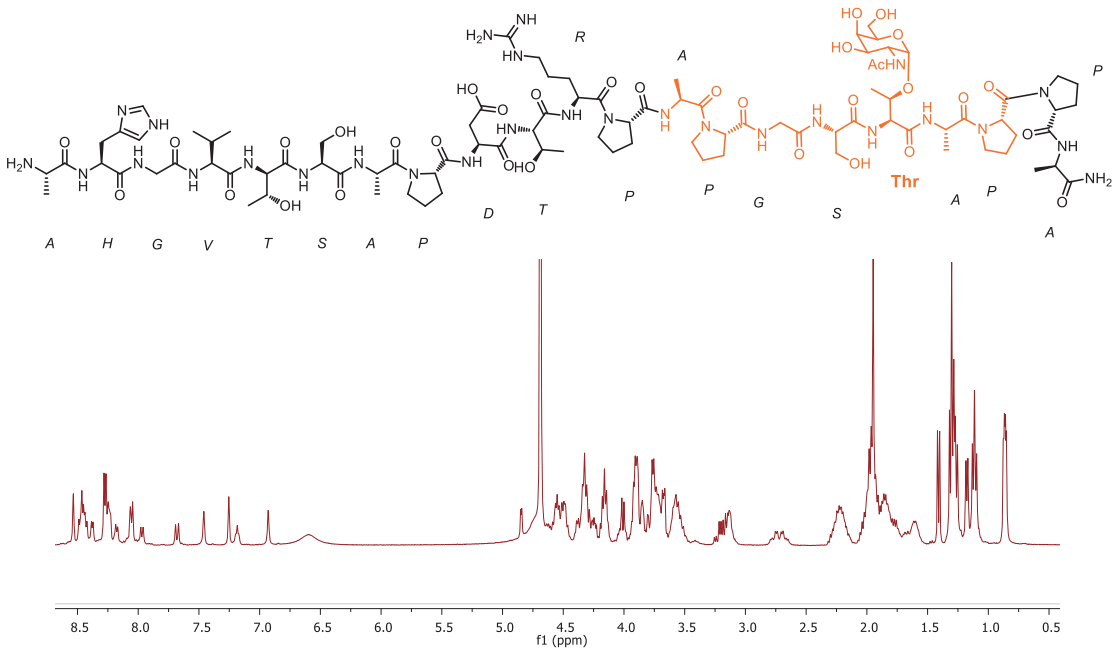


HPLC chromatogram



Compound 7 (MUC1-APGST*AP)

^{1}H NMR 400 MHz in D_2O registered at 298K.



HPLC chromatogram