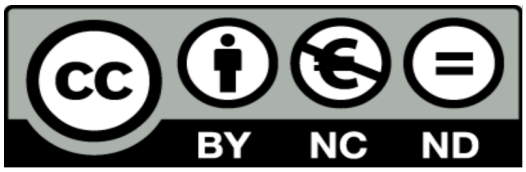
UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL

Título
Unnatural glycopeptides with application on cancer research
Autor/es
Alicia Asín Vicente
Director/es
Francisco Corzana López y Alberto Avenoza Aznar
Facultad
Facultad de Ciencia y Tecnología
Titulación
Departamento
Química
Curso Académico



publicaciones.unirioja.es
E-mail: publicaciones@unirioja.es

UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL 2022

Programa de Doctorado en Química

UNNATURAL GLYCOPEPTIDES WITH APPLICATION ON CANCER RESEARCH

GLICOPÉPTIDOS NO NATURALES CON APLICACIÓN EN LA INVESTIGACIÓN CONTRA EL CÁNCER

Alicia Asín Vicente

Director: Francisco Corzana López

Director: Alberto Avenoza Aznar

Dr. FRANCISCO CORZANA LÓPEZ, Profesor Titular de Química Orgánica del Departamento de Química de la Universidad de la Rioja y

Dr. ALBERTO AVENOZA AZNAR, Catedrático de Química Orgánica del Departamento de Química de la Universidad de La Rioja

HACEN CONSTAR:

Que la memoria 'Unnatural glycopeptides with application on cancer research' realizada por **Alicia Asín Vicente** en el Departamento de Química de la Universidad de La Rioja y bajo su inmediata dirección, reúne las condiciones exigidas para optar al grado de Doctor en Química.

Logroño, Mayo de 2022

Los directores

Francisco Corzana López Alberto Avenoza Aznar

Agradecimientos

Septiembre de 2017 fue la fecha que marcó el inicio de esta aventura, desde entonces he recorrido este camino repleto de experiencias y aprendizajes tanto a nivel profesional como personal. En esta tesis queda reflejada solo una parte de este largo camino, la parte de los logros, pero para llegar a ellos ha hecho falta multitud de tropiezos y de volver a empezar. Para poder superar estos obstáculos, no hay nada mejor que rodearse de los mejores mentores y compañeros. Por ello, quiero agradecer a todas las personas que han sido imprescindibles para la realización de esta tesis y que me han acompañado a lo largo de esta etapa.

Para empezar, quiero dar las gracias a mis directores Paco y Alberto. A **Paco**, por su disponibilidad aun cuando está ocupado con otras mil tareas, y por saber siempre ver el lado positivo de cualquier resultado, lo cual se agradece enormemente. A **Alberto**, por todos los conocimientos y pasión por la Química que sabe transmitir perfectamente. Además, quisiera agradecer al resto de 'seniors' del grupo (**Pere**, **Héctor** y **Mari Mar**) por acogerme en el Quibi y porque cada uno de ellos es indispensable para este grupo.

Por supuesto, esta aventura no habría sido igual sin mis compañeros de laboratorio. En primer lugar, quiero agradecer a los dos mejores compañeros que podría tener y con los que más horas y risas he compartido, Paula y Pablo. **Paula**, hace 10 años que empezamos la carrera y siempre hemos estado juntas en este camino, gracias por todas las cervezas, paseos, congresos, fotolibros, fiestas y cotilleos compartidos y por todos los que nos quedan. Y **Pablo**, gracias por sacarme una sonrisa hasta en los peores momentos y porque contigo todos los planes acaban siendo memorables (echaré de menos que me alumbres con flashes). A **Xhenti**, por todos los buenos planes italianos que hemos compartido tanto en España como en Italia, sé que siempre voy a tener una amiga en Florencia. A **Guille** y a **Iris**, gracias por todas las risas y cotilleos (entre víboras nos entendemos). A **Ismael**, aunque al principio me diera un poco de miedo pedirte ayuda… has acabado siendo fundamental durante estos años y he aprendido mucho gracias a ti y a tu inmensa pasión por los péptidos. A **Fayna** y a **Ester**, por los momentos de desconexión comiendo al solecito juntas. A mi asturiano favorito, **Óscar**, gracias por saber siempre como animarme, aunque sea contándome tus desgracias en el lab para que lo mío parezca menos malo. Espero que continúes regando nuestra plantita (aunque sea con sidra). A **Marina**

y **Carmen** que continúan con el legado de la AECC, gracias por todos los buenos momentos en el despacho amenizando mi escritura.

Tengo que extender el agradecimiento al resto de doctorandos del departamento con los que tantas horas he pasado y sin los cuales esto no habría sido lo mismo. A **Elena**, fiel compañera de viboreos, personal trainer y sobre todo amiga con la que tan buenos momentos he pasado. A **Eduardo** y a **Raúl**, vecinos fotoquímicos con los que siempre se pueden tener conversaciones interesantes. A **Quintana**, otro pilar fundamental a lo largo de estos años, gracias por todos los planes que hemos compartido y por ser siempre tan directo y sincero. A **Mónica**, porque también comenzamos este camino juntas y hemos compartidos multitud de conversaciones camino a casa. A **Gonzalo**, por su buen gusto musical en las casas rurales, sin él no serían lo mismo. A **Nora** y **David**, dos amigos que me llevo de esta aventura y con los que espero seguir disfrutando de vacaciones y cervezas en el Nuevo México juntos.

A los doctorandos que ya terminaron (**Emilio**, **Claudio**, **Nuria**, **Marta**, **Sonia**, **Dani**, **Cintia**, **Rebeca**) y a los que todavía están en ello (**Inés**, **Laura**, **Alba**, **Mónica**, **David**, **Kateryna**, **Nil**, **Alberto**, **Bea**, **Lucien**) gracias por las cervezas, casas rurales, conversaciones en el RMN y fiestas compartidas. Y no puedo olvidarme de **Miguel Ángel** que siempre será parte imprescindible en cenas de San Alberto, cervezas los jueves, vermuts toreros y en cualquier fiesta.

Agradecer también a los miembros del PAS (**Ernesto**, **Jorge**, **Montse**, **Amaya**…) por estar siempre dispuestos a ayudar con cualquier problema. En especial, quiero agradecer a **Nines**, la persona más importante para cualquier doctorando de esta universidad, sin ella estaríamos perdidos. Gracias por todas las conversaciones compartidas y por nuestros martes de manualidades e Ikebanas.

Per quanto riguarda il mio periodo nella Università di Verona, vorrei ringraziare a **Roberto**, per avermi accolto nel tuo gruppo e avermi permesso di ampliare le mie conoscenze sulle nanoparticelle e sugli anticorpi. Apprezzo molto tutti i tuoi consigli e la tua disponibilità sempre con un sorriso sul volto. Vorrei ringraziare le persone che ho conosciuto durante questo periodo (**Silvia**, **Carolina**, **Ricardo**, **Luca**, **Giuliana**, **Francesca**, **Filippo**...), grazie per rendere i mercoledì 'alpino' il miglior giorno della settimana. En especial, gracias **Ander** por acogerme desde el primer día y por estar siempre dispuesto a tomar un spritz conmigo.

Por supuesto, quiero agradecer a mis amigas de la carrera (**María**, **Esther**, **Andrea**, **Sandra**) porque, aunque no nos veamos tanto como nos gustaría, sé que siempre puedo contar con vosotras. Además, quiero agradecer a mis amigas de toda la vida (**Clara**, **Irene**, **Inés**, **Mónica**) porque siempre han estado ahí apoyándome.

Por supuesto, no habría llegado hasta aquí sin el apoyo de mi familia. Gracias por mostraros siempre orgullosos de lo que estoy haciendo. En especial quiero agradecer enormemente a mis **padres**, porque gracias a vosotros he podido seguir mis sueños y habéis estado siempre a mi lado para guiarme y animarme. A mi **hermana**, un pilar fundamental en mi vida, gracias por ponerme la piel de gallina cuando tocas el piano, por sacarme una carcajada con memes estúpidos y por todos los momentos vividos juntas.

Por último, quiero agradecer a **Mattia** porque en los malos momentos sabes siempre cómo animarme y los buenos momentos sabes hacerlos todavía mejores. Gracias por tu paciencia, tus bromas, por las cervezas después de un duro día (aunque sea un lunes) y por hacer piña entre todos los grupos. Esta etapa se cierra y lo que venga ahora todavía es incierto, pero lo que tengo seguro es que siempre vas a estar a mi lado apoyándome.

Finalmente quiero agradecer a la Asociación Española Contra el Cáncer (AECC) en la Rioja, por el contrato predoctoral que me ha permitido desarrollar mi tesis, al Ministerio de Ciencia e Innovación por los proyectos concedidos al grupo (CTQ2015-67727-R y RTI2018-099592-B-C21) y a la Universidad de La Rioja, por las ayudas a tesis doctorales (ATUR), así como por conformar el marco humano y científico idóneo para el desarrollo de este trabajo.

Abstract

Protein glycosylation is the most frequent post-translational modification which mediates a variety of cellular processes. In addition, changes in protein glycosylation can modulate cellular phenotypes such as growth, development, and disease.

Therefore, the main objective of the present PhD dissertation will be focus on the study of two important classes of *O*-glycopeptides with relevant implications to fight against cancer: mucins, such as mucin-1 (MUC1), and antifreeze glycoproteins.

Concerning the first mentioned class, mucins are a family of *O*-glycoproteins with high molecular weight that serve as cell-surface receptors and sensors. Notably, aberration in the glycosylation of this tandem repeats is observed in human carcinomas, which made mucins excellent candidates for the development of novel tools for cancer vaccination or early cancer detection. Despite the vast amount of clinical and immunological data available on anti-MUC1 antibodies, the molecular details by which these antibodies recognize their targets are scarce. This information is essential, for example, to develop new MUC1 glycopeptides with improved binding strength and specificity for anti-MUC1 antibodies that can be used in early cancer detection assays. On this basis, an exhaustive study of the molecular basis by which two important anti-MUC1 antibodies (SM3 and 5E5) interact with their target has been performed. In this sense, rational structure-based modifications on the antigen structure have been carried out achieving more efficient antigens in terms of affinity for anti-MUC1 antibodies. Furthermore, a serologic study based on a nanoparticle dot-blot assay was performed observing significant differences between healthy patients and patients with cancer when using these new unnatural glycopeptides.

On the other hand, antifreeze glycoproteins have recently attracted interest due to their use as crystal inhibitors for applications such as tumor cryotherapy. However, the mechanism of action of these glycoproteins is currently unclear. On this basis, various antifreeze glycopeptide derivatives have been synthesized to understand the mechanism by which natural antifreeze glycoproteins can act as potent inhibitors of ice formation.

Resumen

La glicosilación de proteínas es la modificación postraduccional más frecuente que interviene en diversos procesos celulares. Además, cambios en la glicosilación de proteínas pueden modular fenotipos celulares como el crecimiento, el desarrollo y la enfermedad.

Por ello, el objetivo principal de la presente tesis doctoral es el estudio de dos importantes clases de *O*-glicopéptidos con implicaciones relevantes para la lucha contra el cáncer: las mucinas, como la mucina-1 (MUC1), y las glicoproteínas anticongelantes.

En cuanto a las mucinas, son una familia de *O*-glicoproteínas de alto peso molecular que se encuentran en la superficie de las células y que sirven como receptores y sensores. En células tumorales se producen cambios en la glicosilación, dejando expuestos determinados antígenos normalmente enmascarados en células sanas, lo que convierte a las mucinas en excelentes candidatas para el desarrollo de vacunas contra el cáncer o para el diagnóstico precoz. A pesar de la gran cantidad de datos clínicos e inmunológicos disponibles sobre los anticuerpos anti-MUC1, los detalles moleculares por los que estos anticuerpos reconocen a su diana son escasos. Esta información es esencial para desarrollar nuevos glicopéptidos con mayor afinidad y especificidad y poder así usarlos para desarrollar ensayos de detección precoz del cáncer más sensibles. Por ello, en esta tesis se ha realizado un estudio exhaustivo de las bases por las que dos importantes anticuerpos anti-MUC1 (SM3 y 5E5) interactúan con su diana. Así, se han llevado a cabo modificaciones racionales del antígeno mejorando la afinidad de dichos antígenos con los anticuerpos. Además, se ha realizado un estudio serológico basado en nanopartículas, observando diferencias significativas entre pacientes sanos y pacientes con cáncer al utilizar estos nuevos glicopéptidos no naturales.

Por otro lado, las glicoproteínas anticongelantes han despertado recientemente interés debido a su uso como inhibidores de cristales para aplicaciones como la crioterapia de tumores. Sin embargo, el mecanismo de acción de estas glicoproteínas no está actualmente claro. Por ello, en esta tesis se han sintetizado varios glicopéptidos anticongelantes para comprender el mecanismo por el que

las glicoproteínas anticongelantes pueden actuar como potentes inhibidores de la formación de hielo.

Index

Abbreviations

δ	chemical shift
ΔG	relative Gibbs free energy
ΔH	relative enthalpy
ΔS	relative entropy
^{1}H NMR	proton nuclear magnetic resonance
^{13}C NMR	carbon nuclear magnetic resonance
^{19}F NMR	fluorine nuclear magnetic resonance
AAb	autoantibody
Ab	antibody
Ac	acetyl

Ac_2O	acetic anhydride
AcOEt	ethyl acetate
AcOH	acetic acid
AFGP	antifreeze glycoprotein
Arom.	aromatic
AuNP	gold nanoparticle
Bn	benzyl
BSA	bovine serum albumin
CA	cancer antigen
CAR-T cells	chimeric antigen receptor T cells
calcd.	calculated
CD	circular dichroism
CEA	carcinoembryonic antigen
COSY	1H-1H correlation NMR spectroscopy
d	doublet
DIC	*N,N'*- diisopropylcarbodiimide
DIPEA	*N,N*-diisopropylthylamine
DLS	dynamic light scattering
DMF	*N,N*-dimethylformamide
EDC	1-ethyl-3-(3- dimethylaminopropyl)carbodiimide
ELISA	enzyme linked immunosorbent assay
ESI-MS	electrospray ionization-mass spectrometry
Et_2O	diethyl ether
eq.	equivalents
Fmoc	9-fluorenylmethyloxycarbonyl

Gal	galactose
GalNAc	*N*-acetylgalactosamine
GlcNAc	*N*-acetylglucosamine
HBTU	*N,N,N',N'*-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum correlation spectroscopy
iPrOH	isopropanol
IgG	immunoglobulin G
ITC	isothermal titration microcalorimetry
J	coupling constant
K_D	dissociation constant
m	multiplet
mAb	monoclonal antibody
MALDI	matrix-assisted laser desorption/ionization
MD	molecular dynamics
Me	methyl
MeOH	methanol
MeONa	sodium methoxide
MeSer	α-methylserine
MS	mass spectrometry
MST	microScale Thermophoresis
MW-SPPS	microwave-assisted solid-phase peptide synthesis
Neu5Ac	*N*-acetylneuraminic acid or sialic acid

NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
PBS	phosphate buffered saline
PDB	protein data bank
PEG	polyethylene glycol
PG	protecting group
Ph	phenyl
PPII	polyproline II
ppm	parts per million
PVDF	polyvinylidene difluoride
q	quartet
QLL	quasi-liquid layer
QM	quantum mechanics
R	substituent
ROESY	rotating frame nuclear Overhauser effect spectroscopy
Rt	retention time
r.t.	room temperature
s	singlet
scFv	single-chain variable fragment
SPPS	solid-phase peptide synthesis
SPR	surface plasmon resonance
STn	sialyl Tn antigen
Su	succinimide
t	triplet

TBS	tris-buffered saline
T_f	freezing point
TH	thermal hysteresis
T_m	melting point
TCEP	(tris(2-carboxyethyl)phosphine)
TACA	tumor-associated carbohydrate antigens
TBS	tris-buffered saline
^{t}Bu	*tert*-butyl
TFA	trifloroacetic acid
TH	thermal hysteresis
THF	tetrahydrofuran
TIS	triisopropylsilane
TLC	thin layer chromatography
TMS	tetramethylsilane
UV	ultraviolet
VNTR	variable number of tandem repeats
z	charge

1 Introduction

Protein glycosylation is the most frequent post-translational modification which mediates a variety of cellular processes. It is thought that 20% of all eukaryotic proteins are glycosylated,[1] but this is probably an underestimate. Glycans attachment can confer novel properties to the protein influencing their biological activity. It has been described a wide range of functional aspects such as cellular transport and localization, binding specificity, and thermodynamic stability.[2–4] Cell surface glycoproteins participate in a large number of processes such as cell recognition, signal transduction, and adhesion. Additionally, changes in protein glycosylation, including glycan structure and glycan site occupancy, can modulate cellular phenotypes such as growth, development, and disease.[5]

Characterizing the glycoproteome is a challenging endeavor. The nine monosaccharide building blocks used in vertebrates can be assembled in diverse linear and branched patterns to generate a staggeringly complex collection of carbohydrates in an organism, also known as its "glycome." Additional structural diversity results from the attachment of monosaccharides to specific sites on protein scaffolds.[6,7] These modifications produce different classes of protein glycosylation, including *N*-linked,[8] *O*-linked,[9–11] and *C*-linked, [12–14] with the first two being the most common.

The present PhD dissertation will be focus on the study of two important classes of *O*-glycopeptides with relevant implications to fight against cancer: Mucins, such as mucin-1 (MUC1), and antifreeze glycoproteins. The former is currently used to formulate cancer vaccines and diagnostic tools. The second type is one of the best crystal inhibitors and plays an important role in cryotherapy for the treatment of tumors.

Concerning the first mentioned class, mucins are the mainly component of mucus and therefore their biological properties derived from these glycoproteins. Mucins are a family of *O*-glycoproteins with high molecular weight that serve as cell-surface receptors and sensors and play a critical role coordinating responses as differentiation of epithelium, proliferation, apoptosis and secretion.[15] Human mucins (MUC1 to MUC21) can be classified into two groups, secreted and transmembrane mucins. MUC2, MUC5AC, MUC5B, MUC6 and MUC19 belong to secreted group and serve as a physical barrier protecting epithelial cells interacting exclusively with proteins outside the cells. On the contrary, MUC1, MUC4, MUC11-18, MUC20 and MUC21 are transmembrane mucins composed of a C-terminal tail exposed to the cytoplasm and a *N*-terminal part extracellularly

apparent. MUC1-C is composed by a 58 amino acid extracellular domain, a 28 amino acid transmembrane domain and a cytoplasmic tail of 72 amino acids. This cytoplasmic domain contains five Tyr residues that can be phosphorylated acting as binding motifs for proteins with SH2 domains. In general, MUC1 C-terminal subunit acts as a receptor involved in several signaling pathways based on transformation and tumor progression.[16,17] Conversely, MUC1-N is exposed to the cytoplasm and it is composed by 20 amino acids tandem repeats decorated with *O*-linked glycans. This subunit acts as a physical barrier protecting the epithelial cell layer from external agents, such as microorganisms or toxins. (Figure 1.1).

Primary structure of these glycoproteins consists of tandem repeat domains highly enriched in proline, acting as structural fixative, and threonine and serine residues as possible sites for glycosylation. The first carbohydrate attached to a Thr/Ser residue is a GalNAc, which is linked through an α-*O*-glycosylic linkage. This simple structure is known as the Tn antigen.[18]

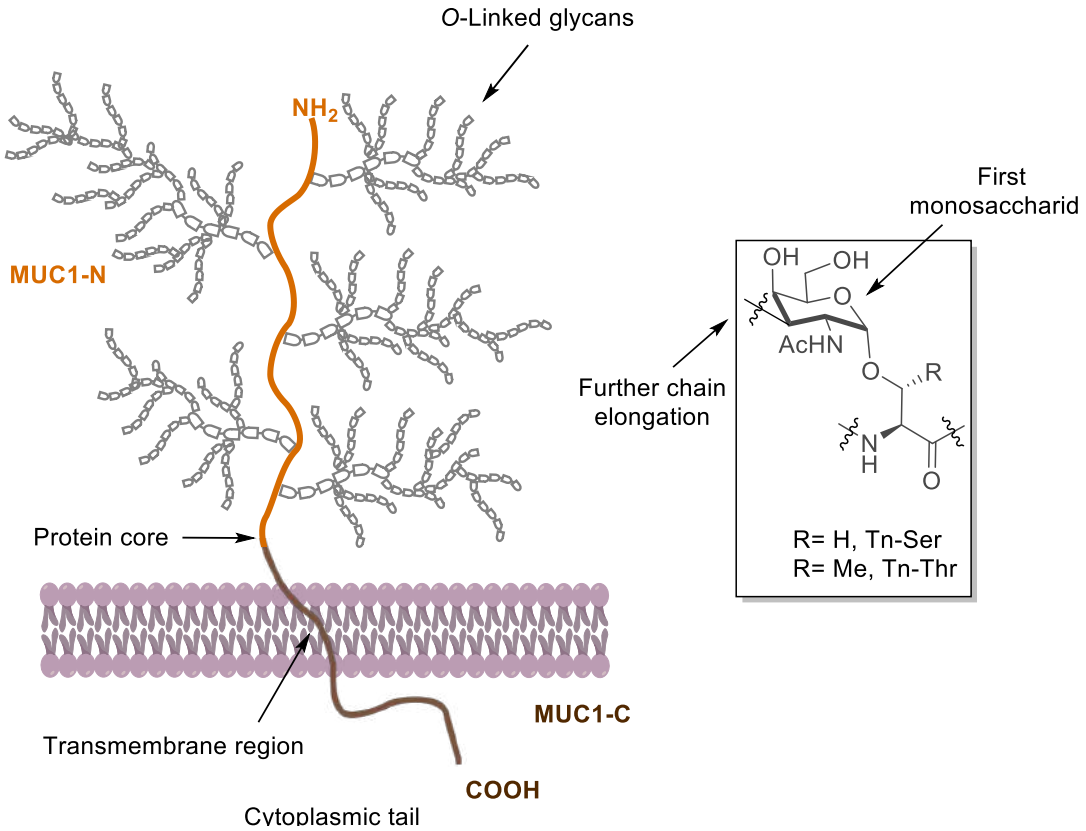


Figure 1.1. Structure of membrane-associated mucins, together with the chemical structure of α-GalNAc-Ser/Thr (Tn antigen).

The aberration in the glycosylation of these tandem repeats is observed in human carcinomas as a consequence of the malfunctions of glycosyltransferases in malignant cells.[19–21] This abnormal glycosylation favors the development and progression of cancer, which made mucins highly attractive targets for the design of vaccines or distinguish normal and disease conditions.[15,22]

One of the most studied mucins is MUC1 which is overexpressed in tumor cells and could be found in more than 80% of carcinomas.[23] This mucin is composed of a tandem repeat region of 20 amino acids (HGVTSAPDTRPAPGSTAPPA) featuring five potential *O*-glycosylation sites, three Thr and two Ser. Significantly, surface of healthy cells is decorated with highly-branched carbohydrate chains. However, aberrant glycosylation in tumors favors the presence of truncated oligosaccharides chains (Figure 1.2).[24]

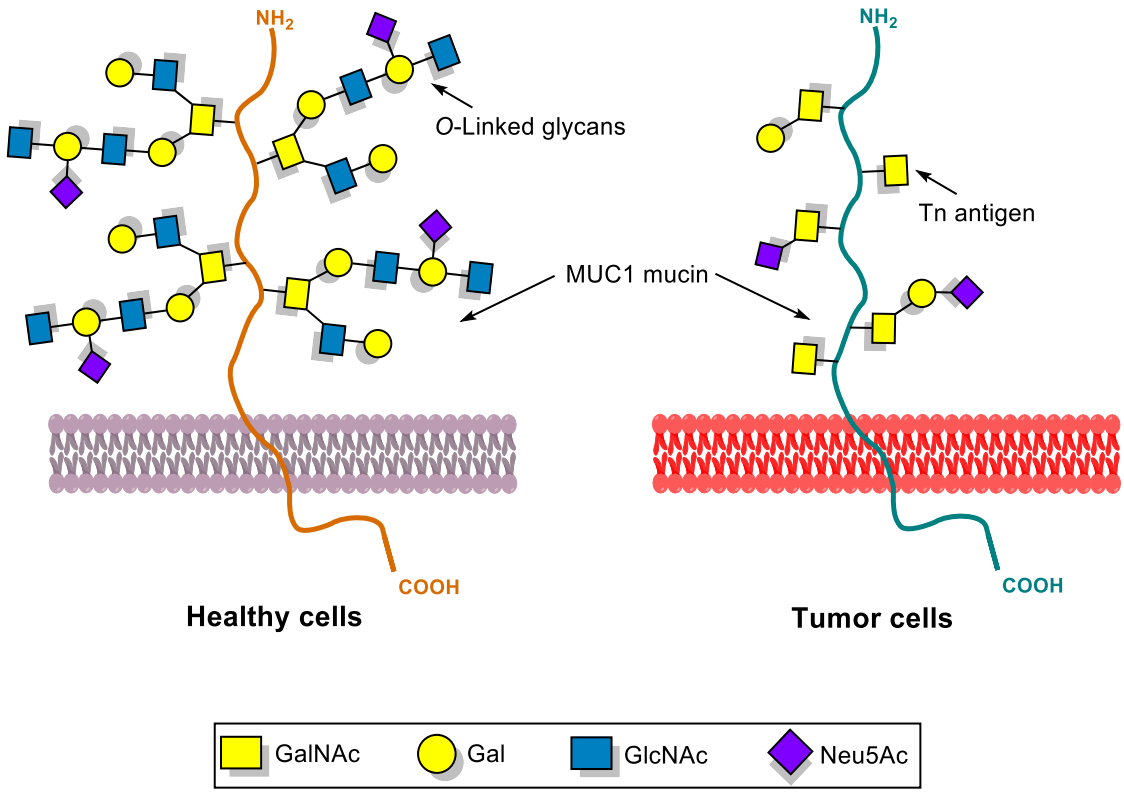


Figure 1.2. Aberrant glycosylation in tumor cells.

In detail, *O*-glycosylation of these sites initiates with the incorporation of the first carbohydrate GalNAc under the action of α-GalNAc transferases. Later, galactose is incorporated to the first carbohydrate generating Core 1 or T antigen.

Subsequently, GlcNAc residue is added to obtain Core 2, which suffers further elongation finishing with the insertion of a fucose or a sialic acid.[25] Notably, down-regulation of glycosyltransferases in cancer cells cause the presence of simple and truncated carbohydrates resulting in the exhibition of Tn and T antigens. Moreover, up-regulation of sialyltransferases causes premature sialylation forming ST and STn antigens (Figure 1.3).

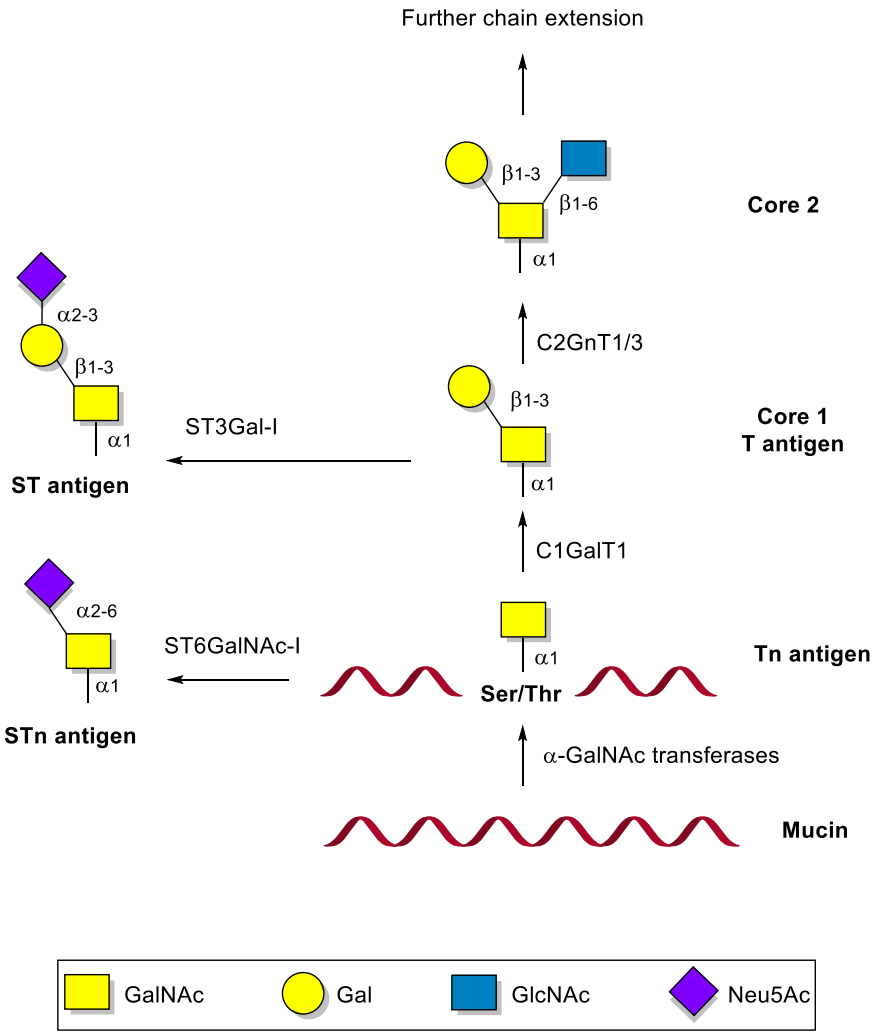


Figure 1.3. Glycosylation biosynthesis pathway.

These antigens are denominated tumor associated carbohydrate antigens (TACAs) and are exposed to the immune system which can elicit immune response. In this regard, several studies have demonstrated the presence of anti-MUC1 antibodies in the sera of cancer patients, which may lead to a favorable prognosis, especially in patients with early-stage breast and pancreatic cancer, as these antibodies may limit tumor growth and spread.[26–29]

Nowadays, cancer is one of the most significant health problems worldwide becoming the second leading cause of death in developed countries accounting for nearly 10 million deaths in 2020.[30] Elimination of risk factors as smoking, unhealthy diet, alcohol, and air pollution could prevent between 30 and 50% of cancer. Moreover, early detection of cancer increases the chances of a good response to treatment increasing survival rates.[31]

Therefore, considering the properties of mucins, these molecules could be excellent candidates for the development of novel tools for cancer vaccination or early cancer detection. In fact, MUC1 has been ranked No. 2 of all 75 tumor-associated antigens as cancer vaccine targets evaluated by National Cancer Institute Translational Research Working Group, based on certain criteria, such as therapeutic function, immunogenicity, cancer cell specificity etc.[32]

As commented above, natural immunity against tumors has been demonstrated owing to the presence of IgG antibodies against tumor-associated epitopes during cancer development. This immunity could be enhanced by anticancer vaccines stimulating the immune response.[33] Generally, cancer vaccines are composed of an antigen like TACA, glycolipids or peptides in combination with an adjuvant.[34] TACA-based antitumor vaccines development is challenging owing to their poor immunogenicity,[24] which could be solved conjugating these derivatives to an immunostimulant capable of activate T-cells, for instance, T-cell epitopes, nanoparticles, protein carriers or lipopeptides ligands. This strategy has been followed to develop therapeutically cancer vaccines.[35–38] For instance, a combination of a tumor-associated-MUC1 (TA-MUC1) glycopeptide with different immunostimulants such as T-cell epitope[37], nanoparticles[39], carrier proteins[40] or lipopeptides,[41] giving access to the so-called multicomponent vaccines, have elicited an immune response (Figure 1.4).

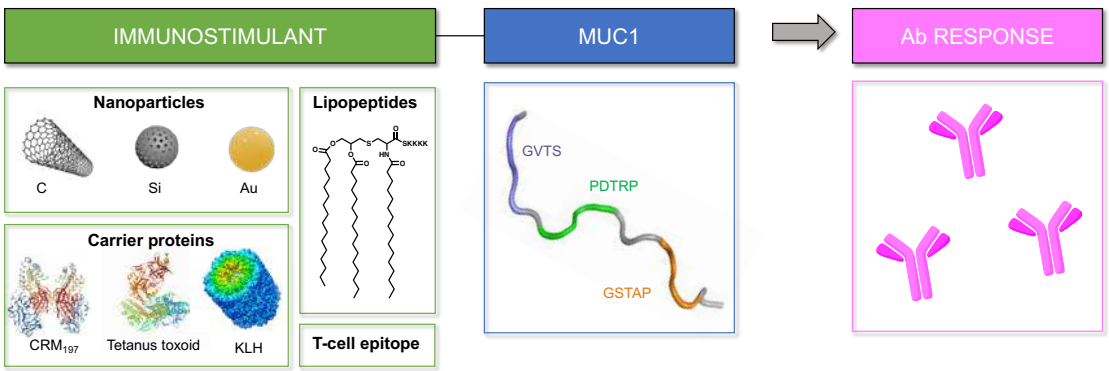


Figure 1.4. Scheme of a cancer vaccine candidate.

In general, weak immune response was obtained for most tested vaccines owing to escape mechanism of the immune system that tolerate these antigens. Therefore, unnatural derivatives could be an ideal alternative to solve this challenge. In principle, these derivatives can be more immunogenic and, at the same time, more resistant against enzymatic degradation.[37,42]

In this regard, our group has developed three vaccine candidates based on MUC1-based unnatural glycopeptides which elicit immune responses in mice generating antibodies that can be recognized by human cancer cells. The first approach was based on the conjugation of an unnatural Tn antigen variant, bearing the α-methylserine amino acid, with a T-helper peptide and a TLR2 agonist. This three components vaccine elicited a strong immune response and exhibited more resistance to enzymatic degradation although its effectivity was not superior to the natural one (Figure 1.5A).[43–46] Another followed strategy was a vaccine composed by a MUC1-based glycopeptide bearing a thiothreonine glycosylated with GalNAc and a fluoroproline conjugated to gold nanoparticles.[47] Oxygen-by-sulfur substitution at the glycosidic linkage favors a preorganized and optimal structure of the antigen for antibody binding whereas the incorporation of a fluorine atom at the proline improve affinity for anti-MUC1 antibodies by an improved CH-π stabilizing interaction.[48] Immunization of mice with this vaccine candidate showed an improved antibody response in comparison with the natural glycopeptide (Figure 1.5B). Lastly, a MUC1-like glycopeptide composed by an unnatural Tn antigen featuring a threonine glycosylated with a sp^2-iminosugar GalNAc mimetic was synthesized to improve the binding with the antibody by an enhanced CH-π interaction. This glycopeptide was then conjugated to a protein carrier and tested in mice resulting in higher levels of antibodies compared to the natural vaccine (Figure 1.5C).[49]

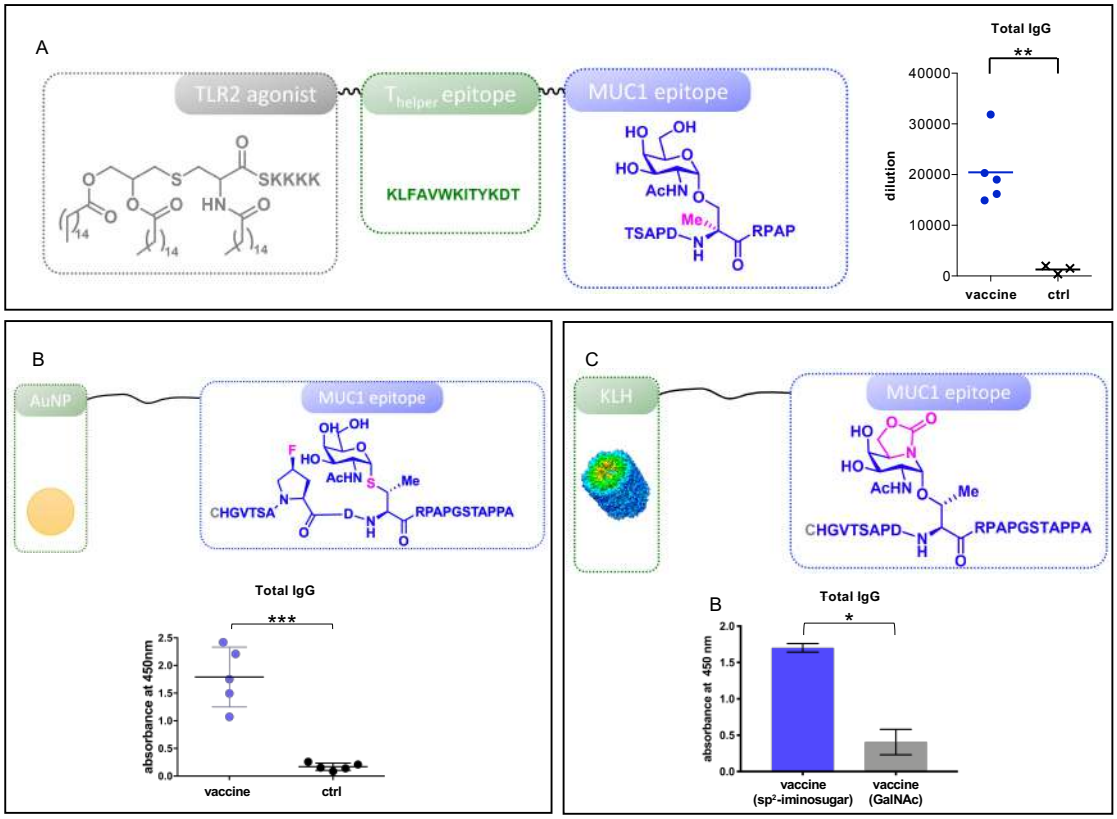


Figure 1.5. Vaccines based on structural changes in the MUC1 sequence.[50]

Another important line followed in the fight against cancer is the development of diagnostic tools for early detection of cancer. For this purpose, released substances from cancer cell to blood and tissues have generated an increasing interest.[51,52] These substances are defined biomarkers and serve as signal of a normal or abnormal condition or disease.[53] Origin of biomarkers is wide ranging, including DNA, RNA or protein.

Traditionally methods of cancer screening consist of invasive techniques such as biopsy and subsequent examination of the tissue by microscopy. The results can be not conclusive, because of the lack of cancer cells in the biopsied tissue. Alternatively, analysis of biomarkers in urine, blood, serum and cerebral spinal fluid could be used as an effective tool to detect cancer at early stages, making treatment more successful and consequently increasing patient survival.[54]

Tumor associated antigens can be used as biomarkers owning aforementioned MUC N-terminal subunit which can be detached from the surface of tumor cells and consequently detected in the sera of cancer patients.[55,56] Therefore, many examples of cancer biomarkers based on antigens are reported. However, most

of them do not demonstrate sufficient sensitivity and specificity for translation into routine clinical use.[51,53,54,56,57] Nevertheless, these tumor-associated antigens are responsible for an immune response promoting the production of autoantibodies, which can be an appealing alternative.[58]

Autoantibodies against MUC1 can be detected in sera from patients with cancer, thus it might be useful as a promising prognostic factor for patients with breast and pancreatic cancer.[27,59–61] Moreover, these antibodies may contribute to limit tumor growth and dissemination. Long persistence of these antibodies in circulating blood and their presence in the first stages of the disease make them ideal for early diagnosis of cancer.[62]

Another relevant family of *O*-glycopeptides are the antifreeze glycoproteins (AFGPs). These compounds consist of the repeating tripeptide unit (Thr*-Ala-Ala)$_{n=4\text{-}55}$, in which the hydroxyl group of Thr is α-*O*-glycosylated with the disaccharide Galβ1-3GalNAc (Figure 1.6A).[63–65] These glycoproteins are an essential class of biomolecules for organisms that inhabit polar waters to survive at subzero temperatures preventing freezing of organism fluids by inhibiting recrystallization of ice.

AFGPs bind to the prismatic plane of incipient ice crystals, inhibiting their growth. This binding results in a lowering of the freezing point without an appreciable change in the melting point. The difference between melting and freezing temperatures, known as thermal hysteresis (TH), is used to detect and quantify the antifreeze activity of these proteins. Furthermore, these compounds alter the morphology of the ice crystal, giving rise to crystals with a hexagonal bipyramid morphology (Figure 1.6B).

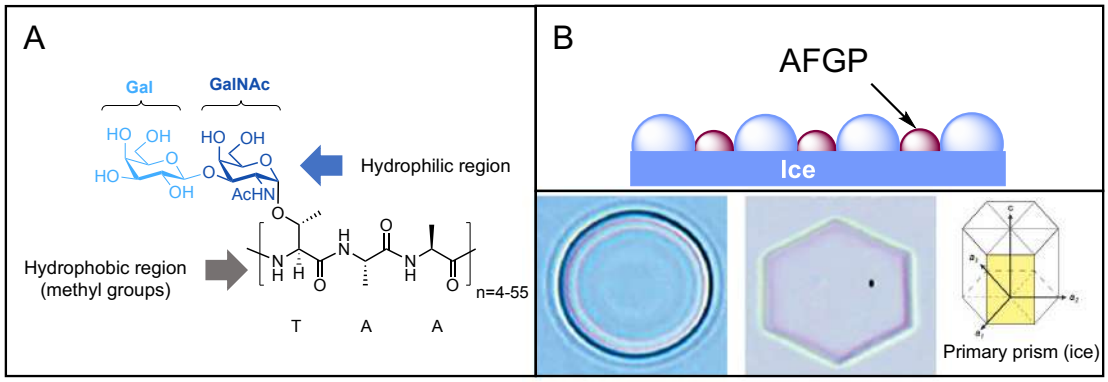


Figure 1.6. (A) Structure of glycoproteins. (B) Adsorption of AFGPs to the ice surface and different morphologies observed for ice crystals (left: pure water; right: water in the presence of an AFGP).

AFGPs have recently attracted great interest due to the wide range of applications presented in fields such as medicine and food, among others. Thus, for instance, these proteins are used to preserve cells, tissues and organs at low temperature.[66] They are also used in aquaculture[67] and for improve the texture of frozen foods.[68] Despite the multiple applications of these glycoproteins, the molecular mechanism by which antifreeze glycoproteins prevent ice crystal growth is currently debated through two opposing models. The classical model[64] states that AFGPs irreversibly bind to the ice surface through the hydroxyl groups of the carbohydrate. In contrast, a recent model[65] assumes that the methyl groups of the peptide and the sugar bind reversibly to the ice surface.

In connection with this dissertation, AFGPs are used in cryotherapy (or cryosurgery) [69–71] a cancer treatment in which extreme cold produced by liquid nitrogen or argon gas is applied to destroy cancer cells and abnormal tissue. These antifreeze proteins are needed to prevent the freezing of surrounding healthy cells. For instance, the treatment of tumors such as prostate and liver cancer by cryosurgery has shown to be more effective when antifreeze proteins are used.[69]

Hence, the study of MUC1 and non-natural variants derived from this glycoprotein to develop biosensors for early detection of cancer and the design of different antifreeze proteins derivatives to clarify the mechanism of action of these biomolecules will be the cornerstones on which this PhD thesis will be based.

References

(1) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. *Sci. Rep.* **2011**, *1*, 1-5.

(2) Mariño, K.; Bones, J.; Kattla, J. J.; Rudd, P. M. *Nat. Chem. Biol.* **2010**, *6*, 713–723.

(3) Raman, R.; Raguram, S.; Venkataraman, G.; Paulson, J. C.; Sasisekharan, R. *Nat. Methods* **2005**, *2*, 817–824.

(4) Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364–2369.

(5) Tabak, L. A. *Semin. Cell Dev. Biol.* **2010**, *21*, 616–621.

(6) Seeberger, P. H. *Monosaccharide Diversity* in Chapter 2 *Essentials of Glycobiology, Fourth Edition*, Cold Spring Harbor, **2022**.

(7) Colley, K. J.; Varki, A.; Haltiwanger, R. S.; Kinoshita, T. *Cellular Organization of Glycosylation* in Chapter 4 *Essentials of Glycobiology, Fourth Edition*, Cold Spring Harbor, **2022**.

(8) Saha, A.; Bello, D.; Fernández-Tejada, A. *Chem. Soc. Rev.* **2021**, *50*, 10451–10485.

(9) Gill, D. J.; Clausen, H.; Bard, F. *Trends Cell Biol.* **2011**, *21*, 149–158.

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

(11) Wandall, H. H.; Nielsen, M. A. I.; King-Smith, S.; de Haan, N.; Bagdonaite, I. *FEBS J.* **2021**, *288*, 7183–7212.

(12) Taniguchi, N.; Endo, T.; Hart, G. W.; Seeberger, P. H.; Wong, C. H. *Glycosci. Biol. Med.* **2015**, 1–1569.

(13) Ihara, Y.; Inai, Y.; Ikezaki, M.; Matsui, I. S. L.; Manabe, S.; Ito, Y. *Glycosci. Biol. Med.* **2015**, 1091–1099.

(14) Hofsteenge, J.; Müller, D. R.; de Beer, T.; Löffler, A.; Richter, W. J.; Vliegenthart, J. F. G. *Biochemistry* **2002**, *33*, 13524–13530.

(15) Hollingsworth, M. A.; Swanson, B. J. *Nat. Rev. Cancer* **2004**, *4*, 45–60.

(16) Rajabi, H.; Kufe, D. *Biochim. Biophys. Acta* **2017**, *1868*, 117-122.

(17) Kufe, D. *Cancer Biol. Ther.* **2008**, *7*, 81–84.

(18) Tongzhong, J.; Vivianne I., O.; Richard D.; Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem. Int. Ed.* **2011**, *50*, 1770–1791.

(19) Finn, O. *J. Immunol.* **2008**, *181*, 1589–1592.

(20) Ichige, K.; Perey, L.; Vogel, C. A.; Buchegger, F.; Kufe, D. *Clin. Cancer Res.* **1995**, *1*, 565–571.

(21) Gendler, S.; Taylor-Papadimitriou, J.; Duhig, T.; Rothbard, J.; Burchel, J. *J. Biol. Chem.* **1988**, *263*, 12820–12823.

(22) Kufe, D. W. *Nat. Rev. Cancer* **2009**, *9*, 874–885.

(23) Hattrup, C. L.; Gendler, S. J. *Annu. Rev. Physiol.* **2008**, *70*, 431–457.

(24) Feng, D.; Shaikh, A. S.; Wang, F. *ACS Chem. Biol.* **2016**, *11*, 850–863.

(25) Nath, S.; Mukherjee, P. *Trends Mol. Med.* **2014**, *20*, 332–342.

(26) Blixt, O.; Bueti, D.; Burford, B.; Allen, D.; Julien, S.; Hollingsworth, M.; Gammerman, A.; Fentiman, I.; Taylor-Papadimitriou, J.; Burchell, J. M. *Breast Cancer Res.* **2011**, *13*, R25.

(27) Hamanakai, Y.; Suehiro, Y.; Fukui, M.; Shikichi, K.; Imai, K.; Hinoda, Y. *Int. J. Cancer* **2003**, *103*, 97–100.

(28) von Mensdorff-Pouilly, S.; Verstraeten, A.; Kenemans, P.; Snijdewint, F.; Kok, A.; Van Kamp, G.; Paul, M.; Van Diest, P.; Meijer, S.; Hilgers, J. *J. Clin. Oncol.* **2000**, *18*, 574–583.

(29) Tang, Z. M.; Ling, Z. G.; Wang, C. M.; Wu, Y. B.; Kong, J. L. *PLoS ONE*. **2017**, *12*, e0182117.

(30) Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A. *CA. Cancer J. Clin.* **2021**, *71*, 7–33.

(31) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. *CA. Cancer J. Clin.* **2021**, *71*, 209–249.

(32) Cheever, M. A.; Allison, J. P.; Ferris, A. S.; Finn, O. J.; Hastings, B. M.; Hecht, T. T.; Mellman, I.; Prindiville, S. A.; Viner, J. L.; Weiner, L. M.; et al. *Clin. Cancer Res.* **2009**, *15*, 5323–5337.

(33) Brockhausen, I.; Melamed, J. *Glycoconj. J.* **2021**, *38*, 459–474.

(34) Lewis, J. J.; Houghton, A. N. *Semin. Cancer Biol.* **1995**, *6*, 321–327.

(35) Wilson, R. M.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2013**, *135*, 14462–14472.

(36) Martínez-Sáez, N.; Peregrina, J. M.; Corzana, F. *Chem. Soc. Rev.* **2017**, *46*, 7154–7175.

(37) Gaidzik, N.; Westerlind, U.; Kunz, H. *Chem. Soc. Rev* **2013**, *42*, 4421–4442.

(38) Miles, D.; Roché, H.; Martin, M.; Perren, T. J.; Cameron, D. A.; Glaspy, J.; Dodwell, D.; Parker, J.; Mayordomo, J.; Tres, A.; Murray, J. L.; Ibrahim, N. K. *Oncologist* **2011**, *16*, 1092–1100.

(39) Parry, A. L.; Clemson, N. A.; Ellis, J.; Bernhard, S. S. R.; Davis, B. G.; Cameron, N. R. *J. Am. Chem. Soc.* **2013**, *135*, 9362–9365.

(40) Palitzsch, B.; Gaidzik, N.; Stergiou, N.; Stahn, S.; Hartmann, S.; Gerlitzki, B.; Teusch, N.; Flemming, P.; Schmitt, E.; Kunz, H. *Angew. Chem. Int. Ed.* **2016**, *55*, 2894–2898.

(41) Wolfert, M. A.; Boons, G.-J. *Nat. Chem. Biol.* **2013**, *9*, 776–784.

(42) Lakshminarayanan, V.; Thompson, P.; Wolfert, M. A.; Buskas, T.; Bradley, J. M.; Pathangey, L. B.; Madsen, C. S.; Cohen, P. A.; Gendler, S. J.; Boons, G.-J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 261–266.

(43) Avenoza, A.; Cativiela, C.; Corzana, F.; Peregrina, J. M.; Sucunza, D.; Zurbano, M. M. *Tetrahedron: Asymmetry* **2001**, *6*, 949–957.

(44) Aydillo, C.; Jiménez-Osés, G.; Busto, J. H.; Peregrina, J. M.; Zurbano, M. M.; Avenoza, A. *Chem. Eur. J.* **2007**, *13*, 4840–4848.

(45) Corzana, F.; Busto, J. H.; Marcelo, F.; De Luis, M. G.; Asensio, J. L.; Martín-Santamaría, S.; Sáenz, Y.; Torres, C.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. *Chem. Commun.* **2011**, *47*, 5319–5321.

(46) Martínez-Sáez, N.; Supekar, N. T.; Wolfert, M. A.; Bermejo, I. A.; Hurtado-Guerrero, R.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Boons, G. J.; Peregrina, J. M.; Corzana, F. *Chem. Sci.* **2016**, *7*, 2294–2301.

(47) Compañón, I.; Guerreiro, A.; Mangini, V.; Castro-López, J.; Escudero-Casao, M.; Avenoza, A.; Busto, J. H.; Castillo, S.; Asensio, J. L.; Jiménez-Osés, G.; Boutureira, O.; Peregrina, J. M.; Hurtado-Guerrero, R.; Fiammengo, R.; Bernardes, G. J. L.; Corzana, F. *J. Am. Chem. Soc.* **2019**, *141*, 4063–4072.

(48) Somovilla, V. J.; Bermejo, I. A.; Albuquerque, I. S.; Martínez-Sáez, N.; Castro-López, J.; García-Martín, F.; Compañón, I.; Hinou, H.; Nishimura, S. I.; Jiménez-Barbero, J.; Asensio, J. L.; Avenoza, A.; Busto, J. H.; Hurtado-Guerrero, R.; Peregrina, J. M.; Bernardes, G. J. L.; Corzana, F. *J. Am. Chem. Soc.* **2017**, *139*, 18255–18261.

(49) Bermejo, I. A.; Navo, C. D.; Castro-López, J.; Guerreiro, A.; Jiménez-Moreno, E.; Sánchez Fernández, E. M.; García-Martín, F.; Hinou, H.; Nishimura, S. I.; García Fernández, J. M.; Ortiz Mellet, C.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L; Hurtado-Guerrero, R.; Peregrina, J. M.; Corzana, F. *Chem. Sci.* **2020**, *11*, 3996–4006.

(50) Asín, A.; García-Martín, F.; Busto, J. H.; Avenoza, A.; Peregrina, J. M.; Corzana, F. *Curr. Med. Chem.* **2022**, *29*, 1258-1270.

(51) Basil, C. F.; Zhao, Y.; Zavaglia, K.; Jin, P.; Panelli, M. C.; Voiculescu, S.; Mandruzzato, S.; Lee, H. M.; Seliger, B.; Freedman, R. S.; Taylor, P. R.; Hu, N.; Zanovello, P.; Marincola, F. M.; Wang, E. *Cancer Res.* **2006**, *66*, 2953–2961.

(52) Somerfield, M. R. *J. Clin. Oncol.* **1996**, *14*, 2843–2877.

(53) Bohunicky, B.; Mousa, S. *Nanotechnol. Sci. Appl.* **2010**, *4*, 1–10.

(54) Tothill, I. *Semin. Cell Dev. Biol.* **2009**, *20*, 55–62.

(55) Boshell, M.; Lalani, E.; Pemberton, L.; Burchell, J.; Gendler, S.; Taylor-Papadimitriou, J. *Biochem. Biophys. Res. Commun.* **1992**, *185*, 1–8.

(56) Smorodinsky, N.; Weiss, M.; Hartmann, M.; Baruch, A.; Harness, E.; Yaakobovitz, M.; Keydar, I.; Wreschner, D. *Biochem. Biophys. Res. Commun.* **1996**, *228*, 115–121.

(57) Meyer, T.; Rustin, G. *Br. J. Cancer* **2000**, *82*, 1535–1538.

(58) Wu, J.; Li, X.; Song, W.; Fang, Y.; Yu, L.; Liu, S.; Churilov, L. P.; Zhang, F. *Autoimmun. Rev.* **2017**, *16*, 1270–1281.

(59) Tang, Y.; Wang, L.; Zhang, P.; Wei, H.; Gao, R.; Liu, X.; Yu, Y.; Wang, L. *Clin. Vaccine Immunol.* **2010**, *17*, 1903–1908.

(60) Tang, Y.; Cui, X.; Xiao, H.; Qi, S.; Hu, X.; Yu, Q.; Shi, G.; Zhang, X.; Gu, J.; Yu, Y.; Wang, L.; Li, Y. *Mol. Med. Rep.* **2017**, *15*, 2659–2664.

(61) Gheybi, E.; Amani, J.; Salmanian, A. H.; Mashayekhi, F.; Khodi, S. *Tumor Biol.* **2014**, *35*, 11489–11497.

(62) Chapman, C. J.; Thorpe, A. J.; Murray, A.; Parsy-Kowalska, C. B.; Allen, J.; Stafford, K. M.; Chauhan, A. S.; Kite, T. A.; Maddison, P.; Robertson, J. F. R. *Clin. Cancer Res.* **2011**, *17*, 1474–1480.

(63) Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S. I. *Angew. Chem. Int. Ed.* **2004**, *43*, 856–862.

(64) Meister, K.; Devries, A. L.; Bakker, H. J.; Drori, R. *J. Am. Chem. Soc.* **2018**, *140*, 9365–9368.

(65) Mochizuki, K.; Molinero, V. *J. Am. Chem. Soc.* **2018**, *140*, 4803–4811.

(66) Heisig, M.; Mattessich, S.; Rembisz, A.; Acar, A.; Shapiro, M.; Booth, C. J.; Neelakanta, G.; Fikrig, E. *PLoS One* **2015**, *10*, e0116562.

(67) Zbikowska, H. M. *Transgenic Res.* **2003**, *12*, 379–389.

(68) Li, B.; Sun, D. W. *J. Food Eng.* **2002**, *54*, 175–182.

(69) Venketesh, S.; Dayananda, C. *Crit. Rev. Biotechnol.* **2008**, *28*, 57–82.

(70) Bouvet, V.; Ben, R. N. *Cell Biochem. Biophys.* **2003**, *39*, 133–144.

(71) Eskandari, A.; Leow, T. C.; Rahman, M. B. A.; Oslan, S. N. *Biomolecules* **2020**, *10*, 1649.

2 Background

2.1. Structure-based design of potent antigens derived from MUC1 to fight cancer

Mucins are essential as cell-surface receptors and sensors, being responsible for responses like renewal and differentiation of the epithelium, proliferation, adhesion and cell signaling.[1] Among them, MUC1 is indubitably one of the most studied mucins due to its aberrantly glycosylation in around 90% of human cancers.[2] In these cases, specific antigens and peptide epitopes are exposed to the immune system, stimulating the production of antibodies against MUC1. These characteristics make tumor-associated MUC1 appealing for the design of vaccines for cancer immunotherapy and biosensors for early diagnosis of cancer. Therefore, understanding the molecular basis of recognition of these antigens by anti-MUC1 antibodies is crucial for structure-based design of MUC1 vaccines and biosensors with improved immune response and sensitivity. In this context, NMR spectroscopy allows the identification of the recognition epitope and their conformations in the bound state with numerous antibodies and provides a dynamic structure related to the solid state. Moreover, X-ray crystallography has become an ideal tool to identify the key elements involved in this process and provides a precise picture of the interactions between antigen-antibody.[3]

Co-crystal structures of glycopeptides and respective monoclonal antibodies are very few. Here, monoclonal antibodies are divided in 3 different groups: antibodies which only bind to sugar portion, antibodies which only bind to peptide portion, and antibodies which bind to both peptide and sugar portions.

Complex antibody-ligand binding through exclusively sugar moiety

This model fits the 'anchored glycan motif' hypothesis that the sugar portion serves as the anchor to bind the antibody.[4] Brooks and co-workers achieved the first X-ray structure of an anti-Tn antibody (237mAb) in a complex with a Tn-glycopeptide (Figure 2.1A). This monoclonal antibody exhibits a high affinity and specificity for Ag104A, an aggressive fibrosarcoma. The structure revealed that the peptide and the carbohydrate moieties of glycodecapeptide antigen ERGT(α-*O*-GalNAc)KPPLEELS are involved in several interactions with the antibody.[5] Thus, the glycopeptide antigen lies within a surface groove (Figure 2.1B) and the carbohydrate moiety is buried in a deep solvent excluded pocket in the center of

antibody-combining site such that every hydroxyl groups of the GalNAc moiety makes at least one hydrogen bond to the antibody (Figure 2.1C).

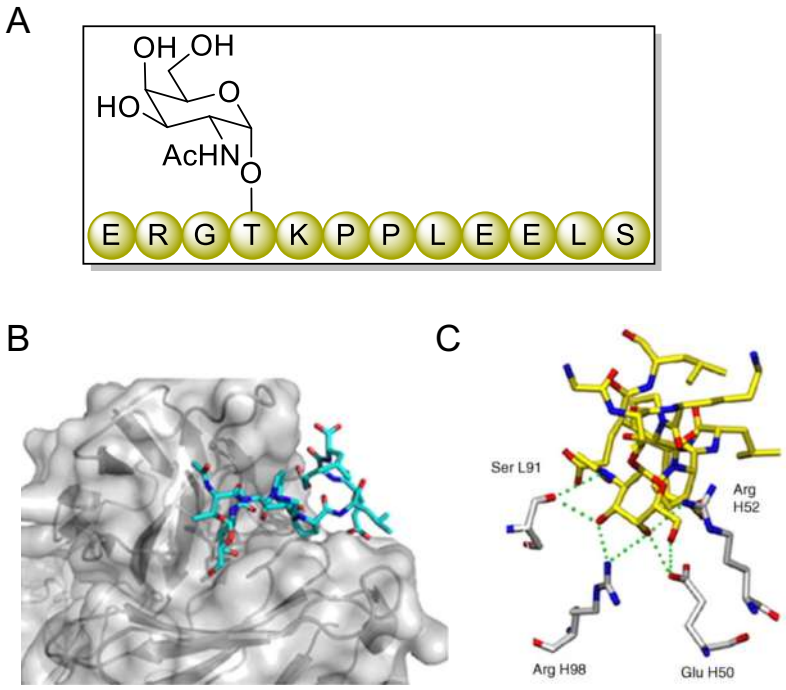


Figure 2.1. X-ray structure of complex of antibody 237 Fab with a Tn-glycopeptide (pdb ID: 3IET). (A) Glycopeptide ERGT(α-*O*-GalNAc)KPPLEELS used in this study. (B) Surface rendering of 237mAb combining site with glycopeptide antigen (blue sticks) bound. (C) Binding interactions of the glycopeptide (shown in yellow) with the 237mAb Fab combining site. Hydrogen bonds are shown in green.[5]

Cummings and co-workers[6] combined NMR studies with microarrays of different mucin-like glycopeptides proving that a panel of seven anti-Tn antibodies (Ca3637, Ca3239, Ca3268, Ca3342, Ca3250, Ca3638 and Ca3749) recognized not only the α-*O*-GalNAc-Thr/Ser structure, but also the surrounding features. This implies a correlation between conformational properties of antigens and their molecular recognition by these antibodies.

Complex antibody-ligand binding through sugar and peptide moieties.

There are very few anti-MUC1 antibodies either alone or in complexed with their cognate glycopeptide epitopes, even though a handful of anti-MUC1 antibodies

have been developed. SM3 was the first anti-MUC1 antibody to be crystallized with its cognate peptide repeat.

Freemont and co-workers obtained the first high-resolution crystal structure of a naked MUC1 peptide (TSAPDTRPAPGST) bound to the Fab fragment of this antibody (Figure 2.2).[7]

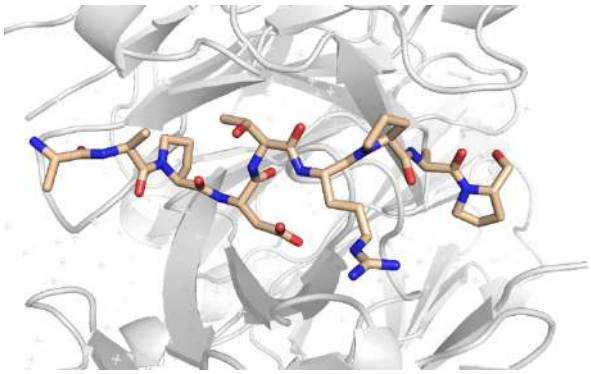


Figure 2.2. MUC1 peptide (SAPDTRPAP) in complex with antibody SM3 (pdb ID: 1SM3). The antibody fragment is shown as a light grey cartoon and the glycopeptide fragment of the antigen is shown as brown carbon atoms.

In this seminal study, the extended conformation of the peptide in the bound state was confirmed, with PDTRP being the main recognition epitope for this antibody which is in good agreement with STD-NMR and tr-NOE studies performed in solution (Figure 2.3).[8] These results will be discussed in more detail in Chapter 5. Additionally, Meyer *et al.*[8] demonstrate by these NMR techniques that the incorporation of a carbohydrate in the sequence of the antigen only weakly affects the interactions whereas its presence confers a conformational stabilization of the peptide in the bound state with this antibody.

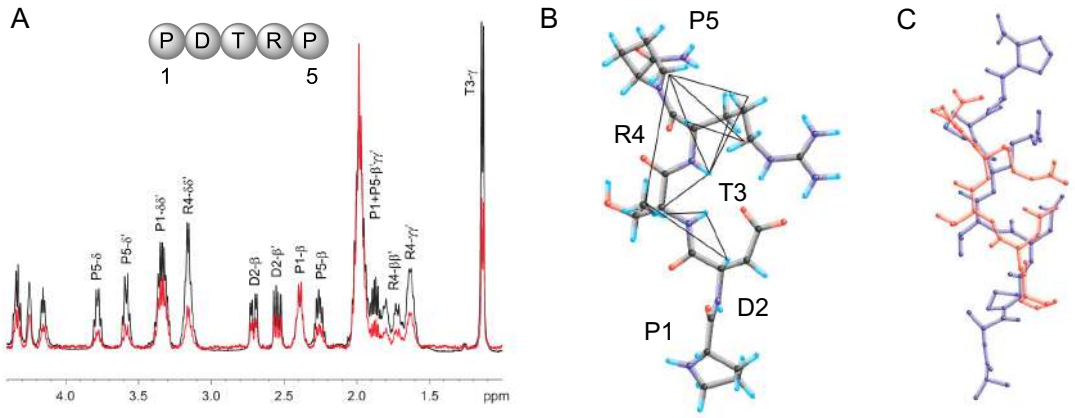


Figure 2.3. PDTRP spectra and structures. (A) Superposition of a 1D STD spectrum (red) and a 1H-spectrum (black) of PDTRP in complex with the antibody SM3. (B) The tr-NOE-derived structure of the peptide PDTRP. (C) Superposition of trNOE-derived structure of PDTRP (red) with the peptide SAPDTRPAP (blue) as determined from the X-ray structure analysis.[8]

In 2015, our group accomplished the first crystallographic analysis of SM3 antibody bound to a glycosylated MUC1 peptide that comprises the APDTRP epitope and the threonine glycosylated with GalNAc.[9] It was confirmed that the peptide fragment is the most recognized part, whereas the carbohydrate moiety is not essential for the recognition. However, a stacking interaction of GalNAc with a tryptophan residue of the antibody provides the driving force for the selectivity of SM3 antibody for GalNAc-containing antigens (Figure 2.5A). This structure will be studied in detail in Chapter 5.

Additionally, B27.29, which is an antibody with specificity for MUC1 and found in ovarian and breast cancer, has been widely studied. NMR studies performed by Campbell and co-workers[10,11] confirmed that the PDTRPAP sequence is involved in the binding, whereas double Tn-glycosylation of the GV**TS**A sequence enhances the recognition, which suggest that B27.29 recognizes these two regions of the glycopeptide. tr-NOE analysis of the antigen glycosylated at Thr3 and Ser4 in complex with the antibody indicates that the glycosylation does not affect the I β-turn conformation observed in solution for the PDTRP fragment when it is bound to the B27.29 antibody (Figure 2.4).[12]

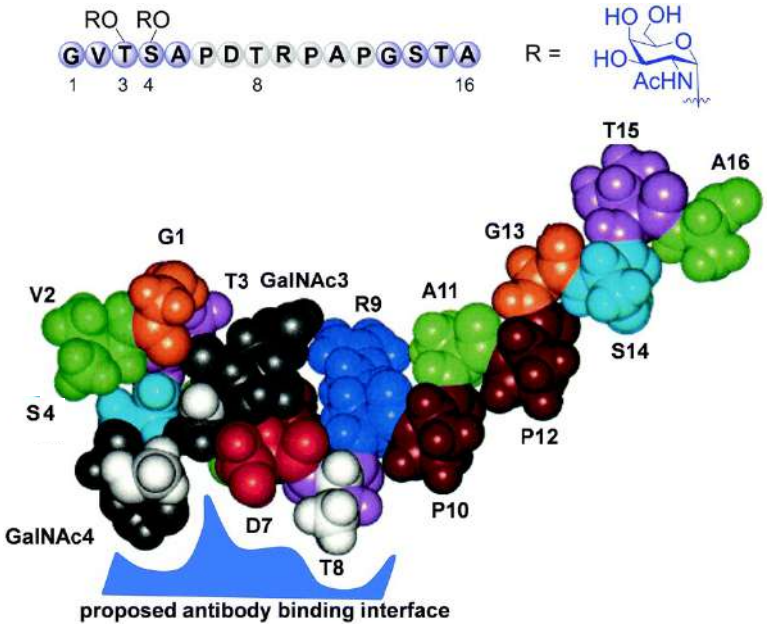


Figure 2.4. Space-filled model of the MUC1 peptide glycosylated at Thr3 and Ser4 showing a proposed antibody binding interface based on peptide-antibody and GalNAc-antibody NOEs observed in the presence of Fab B27.29. To note, contacts of *N*-acetyl groups of the sugars (shown in black) and T8 (shown in white) with the antibody. [10]

Complex antibody-ligand binding through peptide moiety, while the sugar moiety may alter peptide conformation.

Boons *et al.* reported the crystal structure of a MUC1 glycopeptide in complex with AR20.5 (Figure 2.5B).[13] This antibody is a murine anti-MUC1 monoclonal antibody (IgG1) generated with MUC1 from an ovarian cancer patient. This antibody has generated great interest for its therapeutic potential and has been successfully tested in a phase I clinical trial with no observed toxicity. The main recognized epitope of this antibody comprises DTRPAP region. Although the sugar notably improves the binding (around 20-fold) the GalNAc unit is not engaged in contacts with the antibody, suggesting that the sugar moiety could favor the conformation recognized by the antibody in solution.

Significant differences are observed between SM3 and AR20.5 antibodies when X-ray structures of these two anti-MUC1 antibodies in complex with two MUC1 glycopeptides are analyzed. SM3 antibody forms interactions with both the peptide and the carbohydrate moieties, whereas AR20.5 binds exclusively with the peptide moiety of the antigen (Figure 2.5).

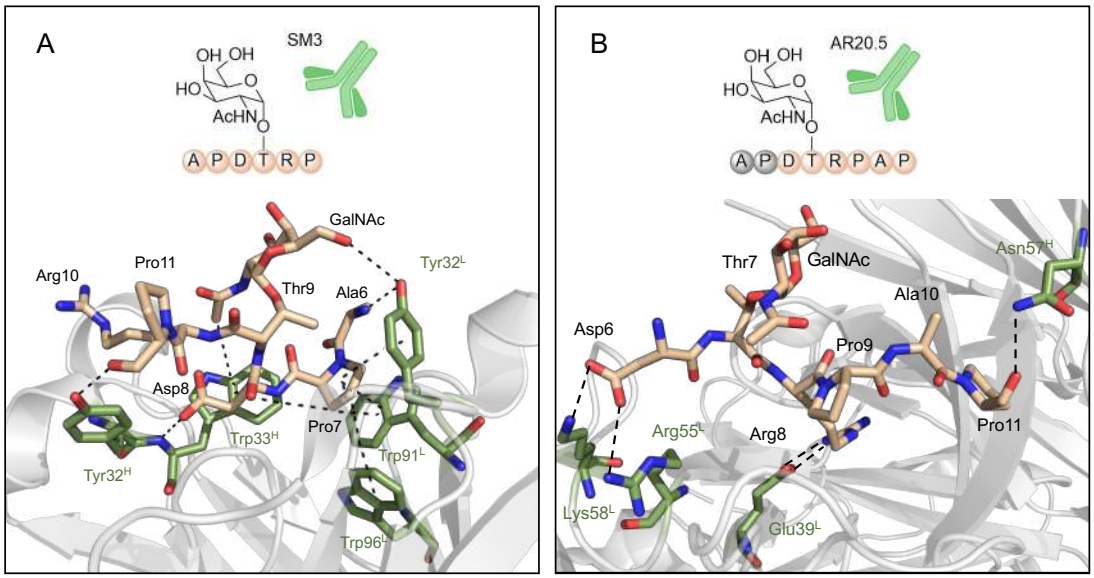


Figure 2.5. (A) Key binding interactions of glycopeptide APDT(Tn)RP in complex with antibody SM3 (pdb ID: 5A2K). (B) Key binding interactions of glycopeptide DT(Tn)RPAP in complex with antibody AR20.5 (pdb ID: 5T78). The antibody fragment is shown as a light grey cartoon and the glycopeptide fragment of the antigen is shown as brown carbon atoms. Interactions between the antibody and the antigen are shown as dashed black lines.

Recently, Marcelo *et al.*[14] have achieved the characterization of a MUC1 library composed by peptides and glycopeptides comprising GalNAc, towards two families of cancer-related monoclonal antibodies by a combination of microarrays, NMR assays and computational calculations (Figure 2.6). The first family comprises two anti-MUC1 antibodies (VU-3C6 and VU-11E2) that recognize naked MUC1-derived peptides through the PDTRP epitope. In the case of glycopeptides bearing Tn antigen in this region, recognition of both antibodies with these derivatives is enhanced whereas the conformation in the bound state is not significant modified related to that observed in solution for the antigens. The second group of antibodies involves two anti-Tn monoclonal antibodies, 8D4 and 14D6, which mostly recognize the GalNAc moiety and do not bind naked peptides.

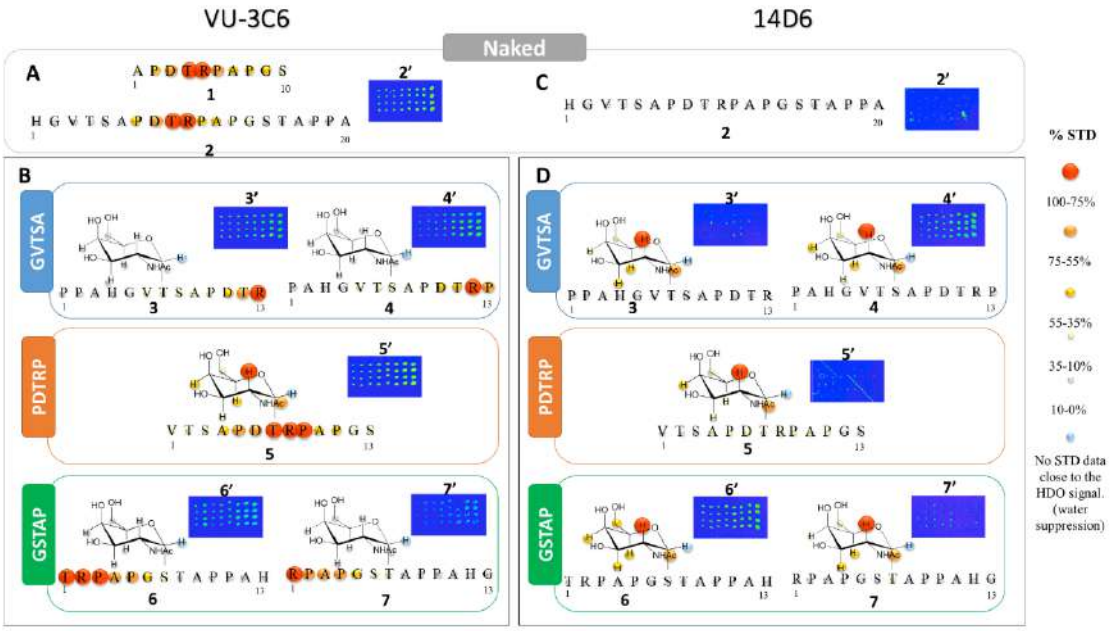


Figure 2.6. STD-derived epitope and microarray fluorescent scan for MUC1-derived (glyco)peptides with antibodies VU-3C6 and 14D6. (A) Naked peptides and (B) glycopeptides with VU-3C6. (C) Naked peptide and (D) glycopeptides with 14D6.[14]

All this structural information can be very useful (*vide infra*) for the structure-based design of novel antigens.

2.2. The use of unnatural MUC1 derivatives to design cancer vaccines

As aforementioned, MUC1 derivatives have attracted great interest for applications as cancer vaccines and diagnostic tools. However, their poor immunogenicity demands an alternative that can be overcome by chemical modifications of the natural tumor-associated MUC1.[15]

2.2.1. Unnatural Tn antigen

In this regard, it has been followed different pathways to enhance the Tn antigen properties, present in these tumor-associated MUC1 (Figure 2.7).

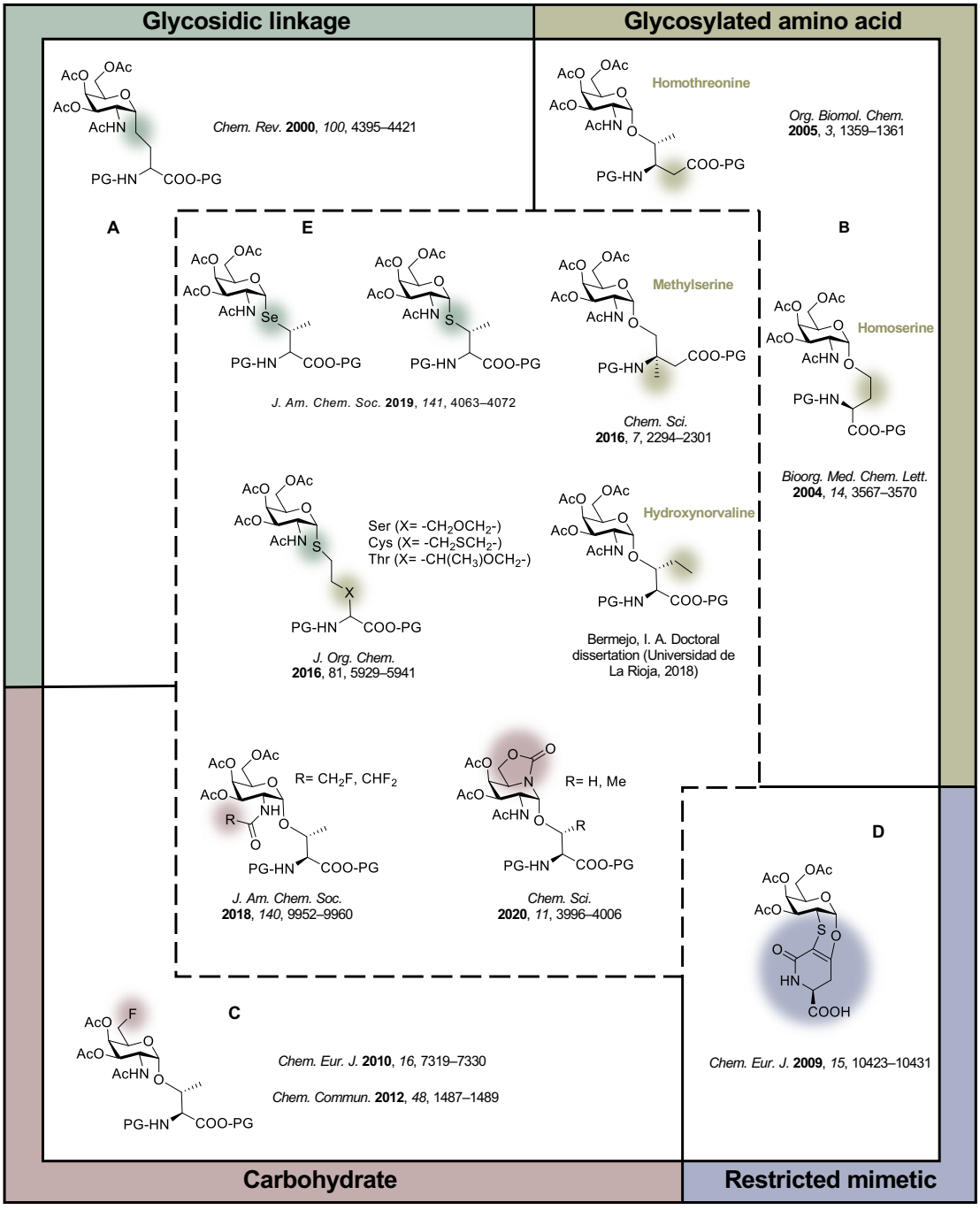


Figure 2.7. Summary of the most important chemical modifications in the antigen found in the bibliography. Modifications reported by our group are limited within the dashed lines.

Firstly, diverse modifications in the Tn antigen moiety were proposed to solve the problem with the low immunogenicity of the natural one. For instance, Dondoni *et al.*[16–18] proposed the substitution of oxygen atom by carbon at the glycosidic linkage giving rise to *C*-glycoside analogs of the Tn antigen (Figure

2.7A). Additionally, different modifications at the glycosylated amino acid have been reported, like its substitution by a homoserine[19] or a homothreonine[20] in order to obtain less enzyme degradation and enhanced immunogenicity (Figure 2.7B).

Various modifications of the carbohydrate moiety have been reported with the same objective.[21–26] For instance, substitution of the hydroxyl group at position C6 of the GalNAc moiety with a fluorine atom[27] allowed the development of specific and promising carbohydrate-based vaccine (Figure 2.7C). Lastly, Nativi and co-workers synthesized a mimetic of the Tn antigen that retains the $^4C_1$ chair conformation (Figure 2.7D) which was used in the preparation of a vaccine obtaining significant levels of IgG/IgM antibodies.[28,29]

It is important to note that, in general, these examples do not follow a rational design of the antigen. Therefore, given the lack of research in this area and the clear impact that new advances could have on cancer fight, our group is focused on rational structure-based design of Tn antigen derivatives. In this regard, understanding the conformational behavior of these antigens in the free and bound states is crucial for the advance in the design of efficacious vaccines and diagnostic tools. *Subtle chemical changes that enhance the stabilizing contacts between the antigen and the antibody should theoretically lead to potent antigens that are more immunogenic and resistant to degradation, as they comprise unnatural residues, and could be used to develop a new generation of potent cancer vaccines. We consider that the high affinity of unnatural antigens for anti-MUC1 antibodies is required to allow cross-reactivity. That is, the antibodies elicited by the vaccine containing these artificial antigens can also recognize the natural antigens expressed by human tumor cells. In addition, these improved antigen-antibody interactions should lead to more efficient antigens in terms of affinity to anti-MUC1 antibodies, allowing us to develop new diagnostic tools to efficiently detect antibodies that are mainly present in early cancer stages.*

In this context, our first strategy in the rational design of new Tn antigens was the use of a quaternary amino acid α-mehylserine (MeSer, Figure 2.7E).[15] This non-natural amino acid retains the bioactive α-helix conformation observed in glycopeptides with SM3 antibody[9,30–32] and, in addition, the methyl group in the Cα should stablish a CH/π interaction with the antibody. In a first step, the natural glycopepide and the corresponding unnatural one bearing the MeSer were studied *in vitro*, observing an enhanced resistance to enzymatic

degradation of this vaccine than the natural variant. Then, a three-component vaccine, like that previously reported by Boons and co-workers,[33] was synthesized and tested in mice. However, the immune response, in terms of total number or antigens elicited by the mice was lower in comparison to those observed with a similar vaccine containing the natural antigen. In this case, the higher flexibility of the glyosidic linkage between the MeSer and the GalNAc residues appears to negatively affect the immune response of this mimic. Of note, in this thesis, we have determined the affinity of the glycopeptide bearing this unnatural residue, observing that its affinity is lower related to the natural variant (Figure 2.8).

Following with the modification of the glycosylated amino acid, we proposed the replacement of Tn-threonine antigen by Tn-hydroxynorvaline which displays a β-ethyl group which could enhanced the CH/π interaction between the antigen and the aromatic ring of Trp33H and Tyr32L. This modification results in similar binding affinity related to the corresponding natural glycopeptide, whereas the humoral response on mice of a vaccine comprising this antigen is currently under study.[34]

Afterward, in 2018, our group observed by X-ray crystallography, a bridging water molecule between the sugar and the amino acid in α-*O*-GalNAc-Thr by the incorporation of a fluorine atom into the *N*-acetyl group of the carbohydrate.[23,35] This water pocket resembled the one previously proposed by our studies in solution, based on MD simulations and NMR experiments. [23,36,37] In this sense, the key role of water in modulating the conformation of the antigen and, therefore, its recognition by antibodies could serve for the structure-based design of new mimics. On this basis, we proposed a rational approach based on the substitution of oxygen atom at the glycosidic bond by sulfur[38–41], or selenium[42] which could confer conformational alterations on the peptide, leading to improvements in binding affinity toward anti-MUC1 antibodies. These minimal modifications increase the distance between the carbohydrate and the peptide fragment, preventing the carbohydrate to interact through bridging water molecules with the peptide. This change impedes that the peptide backbone adopts an extended conformation in solution, thus, favoring the conformer recognized by the SM3 antibody. As a consequence, the affinity towards SM3 antibody is improved and the immune response in mice is enhanced with a nano-vaccine prepared by

conjugation of the unnatural glycopeptide shown in Figure 2.8 to gold nanoparticles (AuNPs).[42]

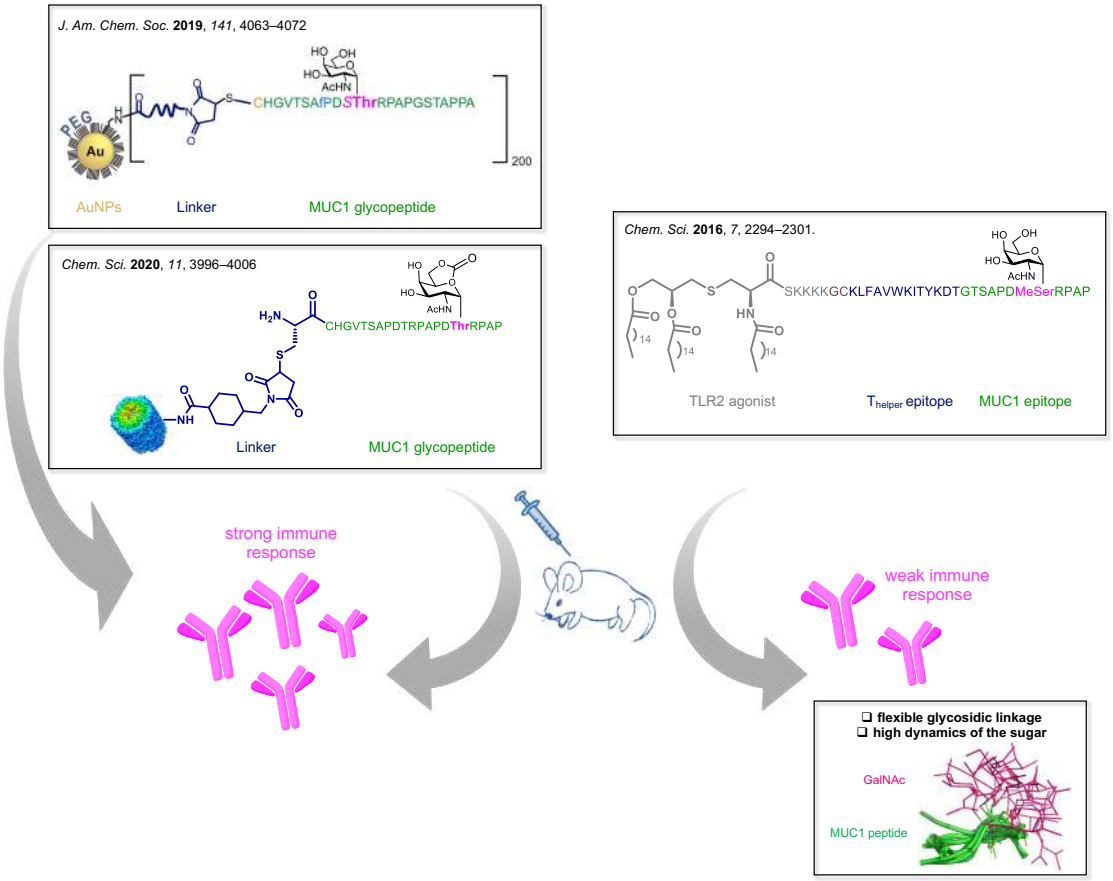


Figure 2.8. Schematic representation of three anti-cancer vaccines based on structural changes in the MUC1 sequence.

Following a similar strategy, our group synthesized several sulfa-Tn antigens featuring a variable linker between the peptide backbone and the sugar which prevents the intramolecular interactions between these two moieties, favoring a helix-like conformation of the glycosylated moiety and the appropriate presentation of the carbohydrate for molecular recognition.[43]

Lastly, we introduced a modification in the carbohydrate part by the replacement of the GalNAc moiety by a sp^2-imminosugar mimetic, which can mimic the main conformation observed in the natural antigen in solution resulting in higher affinity values.[44] A cancer vaccine based on this unnatural glycopeptide was tested in mice proving the higher immune response in comparison to the analog bearing the natural Tn antigen (Figure 2.8).

2.2.2. Modifications of the peptide sequence of MUC1

Based on crystallographic studies, it is possible to establish which amino acids in the MUC1 sequence should be modified to stabilize interactions within the anti-MUC1 antibodies and, consequently, improve the recognition of the peptide.

Thus, based on the X-ray structure of APDT(α-*O*-GalNAc)RP glycopeptide in complex to the SM3 antibody, our group proposed the replacement of a specific hydrogen atom of the Pro ring (in this case Pro2, Figure 2.9) by a fluorine atom. The incorporation of this highly electronegative atom, which withdraws electronic density making hydrogens more electropositive, will lead to an increase of the polarization of the C-H bonds involved in the antigen-antibody interaction, and therefore, the CH/π interaction will be enhanced.

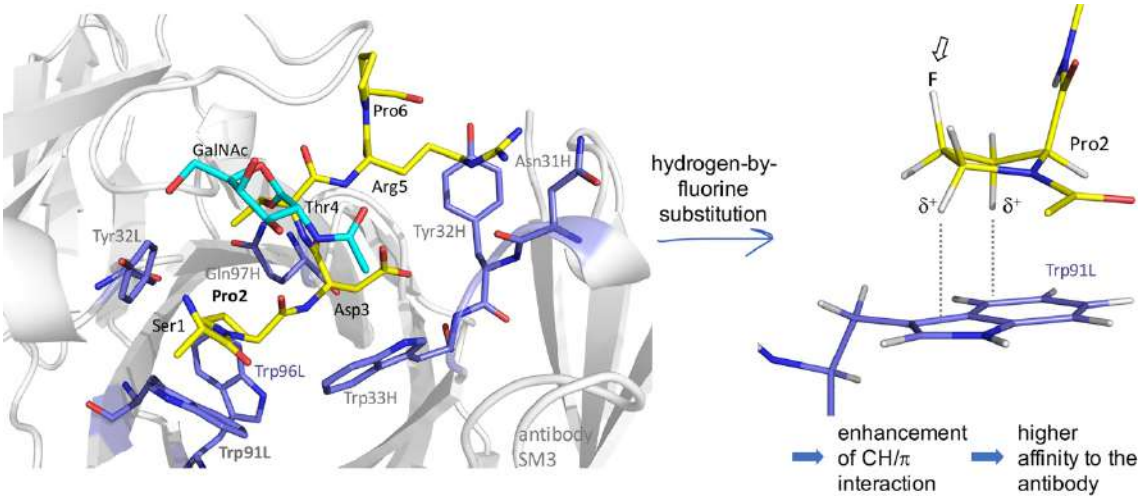


Figure 2.9. Key binding interactions of APDT(α-*O*-GalNAc)RP glycopeptide in complex with SM3 antibody (pdb ID: 5A2K). Hydrogen-by-fluorine substitution to enhance the CH/π interaction between Pro2 and Trp91L.

In this context, our group has recently reported the replacement of the proline in the APDTRP epitope by a (4*S*)-4-fluoro-L-proline or a 4,4-difluoro-L-proline. Affinity assays showed the enhancement of antigen-antibody binding affinity owing to the improved staking interaction (Figure 2.10).[45]

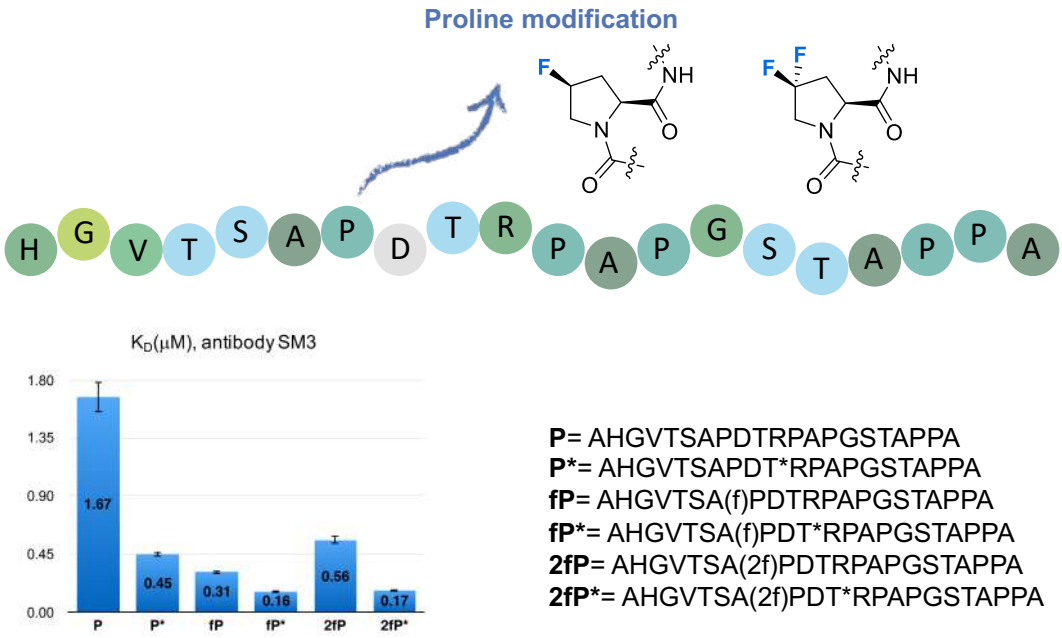


Figure 2.10. Chemical modifications in the proline of A**P**DTRP epitope and K_D constant derived from BLI experiments with SM3 antibody.

2.3. The use of MUC1-like glycopeptides as a potential diagnostic tool

As mentioned earlier, tumor-associated antibodies can be used as a tool for early diagnosis. The presence of these antibodies in the first stages of the disease, long persistence, and stability are the main features responsible for the interest in their use as biomarkers.[46,47] It has been reported many studies using MUC1 for the detection of antibodies in human sera. In terms of detection, enzyme-linked immunosorbent assay (ELISA) is one of the most used techniques for the detection of antibodies in sera using MUC1-like peptides immobilized in plates.[48] Here, we summarize the most important and recent studies using MUC1 and ELISA assays as a diagnostic tool.

In this context, Wang and co-workers[49] have made significant advances, demonstrating that a biosensor, made up of six units of the MUC1 sequence, can be successfully used to determine low levels of antibodies in the sera of patients with breast cancer.

Similarly, a biosensor consisting of a fragment of MUC1 linked to an epitope of the HER2 receptor has recently been developed for diagnostic applications.[50] Overexpression of the latter is associated with the aggressiveness of breast cancer and therefore this biosensor allows the simultaneous detection of antibodies against MUC1 or HER2 in sera.

Furthermore, Chapman *et al.* reported an ELISA assay to detect autoantibodies to different cell surface and internal antigens, such as p53, BRCA1, HER2, MUC1 and so on, obtaining reproducibly elevated levels of autoantibodies in sixty-four percent of primary breast cancer patients. *Interestingly, MUC1 antigen was able to detect the presence of autoantibodies in ≈70% of women diagnosed with breast cancer between 7 and 27 months before the diagnosis made on the screening mammograms. In addition, using a larger group of patients, it was found that ≈60% of patients showed detectable autoantibodies up to 4 years before the diagnosis.*[51]

Most of these studies use the unglycosylated form of the mucin for the detection of antibodies in the sera of breast or ovarian cancer patients.[52–55] Nevertheless, Taylor-Papadimitriou *et al.* performed a novel *O*-glycopeptide microarray-based assay to confirm the presence of autoantibodies in early breast cancer patients.[56] This analysis obtained better results in detecting IgG autoantibodies when MUC1 glycoforms were used in contrast with unglycosylated MUC1. The results of this study showed that antibody levels were significantly higher in patients with Stage I and Stage II breast cancer than in healthy women. Additionally, a strong antibody response was found to be associated with a reduction in the incidence and a delay in metastasis, which reveals the important role of antibodies in inhibiting disease progression.

It should be noted that these diagnostic tools use natural MUC1 antigens, therefore, affinity for anti-MUC1 antibodies and the sensitivity of the test could be improved by using artificial MUC1 antigens. In this regard, our group has reported the use of a MUC1 glycopeptide featuring difluoroproline in the first proline of the A**P**DTRP epitope in an ELISA assay. As previously commented, hydrogen-by-fluorine substitution will reinforce staking interactions of this proline with the aromatic ring of the antibody, and therefore, CH/π interaction will be enhanced, resulting in improved antigen-antibody binding affinity.

This MUC1 glycopeptide derivative was used to check the levels of anti-MUC1

antibodies in different sera of patients with prostatic adenocarcinoma and prostatic benign hyperplasia, as well as healthy controls. From the ELISA experiments, it was confirmed the notable improvement in the sensitivity of the antibodies detection for the difluorinated glycopeptide in comparison with the natural MUC1 mucin (Figure 2.11).[45]

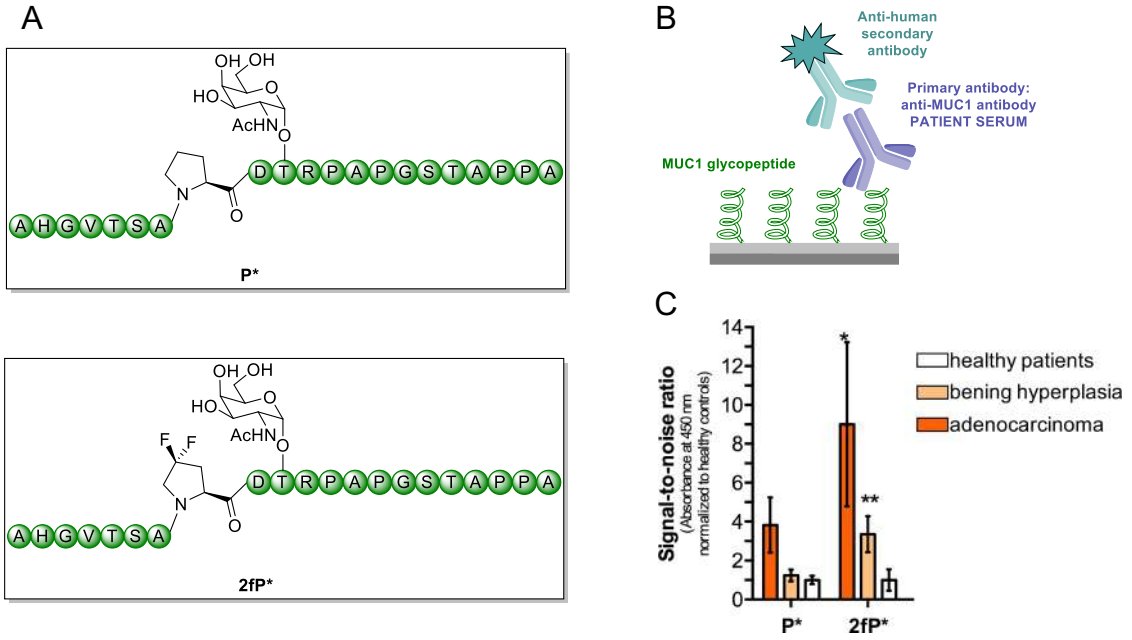


Figure 2.11. Detection of circulating anti-MUC1 antibodies in sera of patients presenting benign and malignant prostate cancer. (A) MUC1-like glycopeptides synthesized and used in this ELISA assay. (B) Schematic design of the assay used to detect anti-MUC1 antibodies in human sera. (C) Binding affinity of human circulating antibodies against these two glycopeptides.[45]

Lastly, although ELISA assays are widely used and offer good selectivity, some problems limit its application, such as a low concentration of the sample to be determined or the difficulty in performing the measurement. In this context and thanks to nanotechnology, conveniently functionalized nanomaterials have proven to be effective in detecting different tumor markers, giving rise to tests with great sensitivity and selectivity.[57] This aspect will be described in more detail in Chapter 5.

2.4. Structural analysis and physical properties of antifreeze glycoproteins

As previously mentioned, another relevant family of *O*-glycopeptides are the antifreeze glycoproteins (AFGPs), which bind to incipient ice crystals, inhibiting their growth. Consequently, the freezing point is reduced without an appreciable change in the melting point. The difference between melting and freezing temperatures is used to measure the antifreeze activity of these proteins and is called thermal hysteresis (TH). Additionally, AFGPs alter the morphology of the ice forming hexagonal bipyramidal ice crystal (Figure 2.12B).[58]

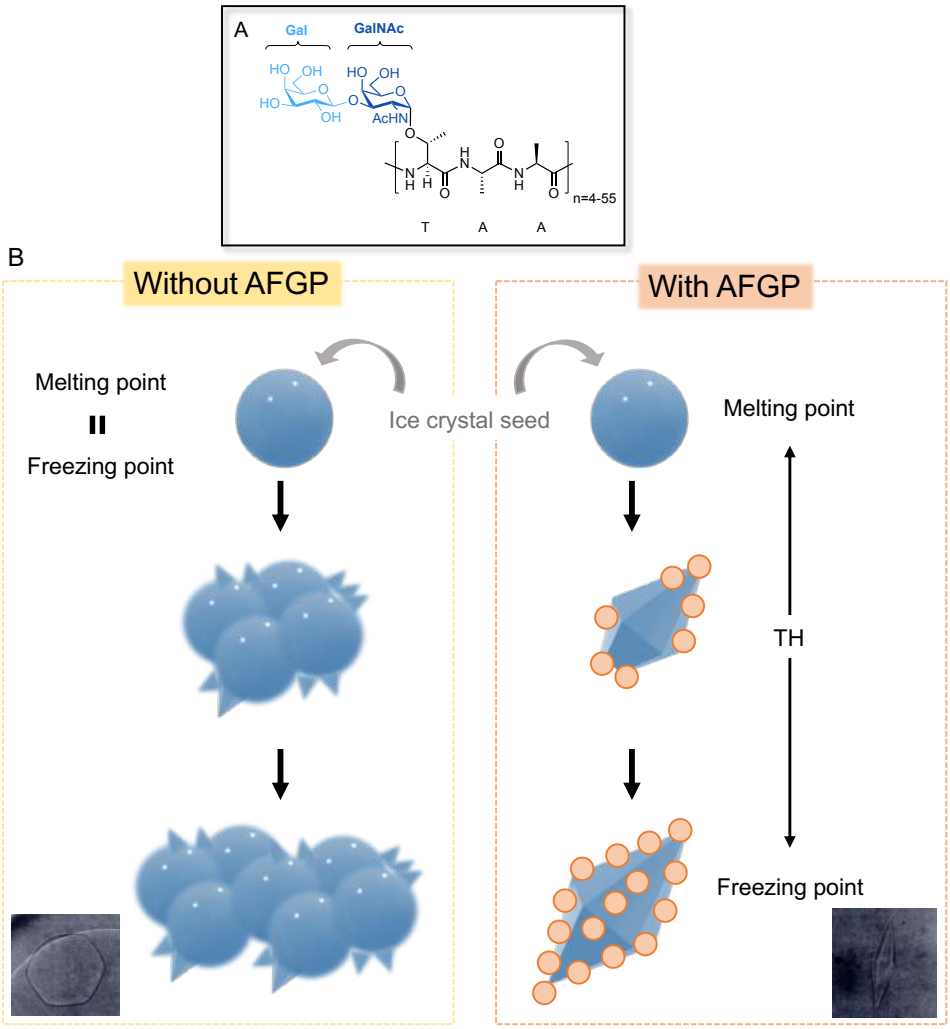


Figure 2.12. (A) Structure of AFGPs. (B) Schematic illustration of ice crystal growth in absence or presence of AFGP.

Antifreeze glycoproteins have recently attracted great interest due to the wide range of applications presented in fields such as medicine and food, among others. For instance, AFGPs can be used in cryotherapy[59] which consists of a cancer treatment in which extreme cold produced by liquid nitrogen or argon gas is applied to destroy cancer cells and abnormal tissue. However, AFGPs are not always suitable for cryopreservation because of potential immunogenicity and toxicity concerns.[60] As a result, several synthetic materials with antifreeze protein-like functions have been developed. These new materials generally have a lower capacity for crystal inhibition than the natural derivatives.[61,62] Therefore, there is a clear motivation to develop synthetic mimics with tunable and tailored functions. Furthermore, the development of these synthetic molecules, capable of mimicking the structures and functions of natural AFGPs, can serve as models to understand the mechanism of action of the latter. [63]

Although the synthesis of AFGPs with relatively large molecular masses and high carbohydrate density is not straightforward, several research groups have developed different synthetic pathways to prepare these derivatives.[64–66] Nishimura's group has designed a procedure for the preparation of natural AFGPs by polymerizing glycopeptides in solution in the presence of diphenylphosphorylazide (DPPA) as a promoter (Figure 2.13A).[67] Subsequently, the original amino acid sequence of the repeat (A-A-T) was changed to (A-T-A) to reduce the steric hindrance associated with threonine-alanine couplings.[68,69] Similarly, the use of other promoters in the polymerization reactions led to better yields.[70] In general, this strategy results in uncontrolled fragment condensation and product inhomogeneity. Hence, the preparation of glycosylated amino acid building block to be used in solid phase peptide synthesis (SPPS) entails a great advantage for the synthesis of antifreeze glycoproteins analogs with pre-defined length and sequence. In this sense, Tseng and co-workers carried out the synthesis of AFGPs (n= 4 or 8) for the first time demonstrating the efficacy of this approach (Figure 2.13B).[71] In addition, Nagel *et al.* enhanced the coupling efficiency of the glycosylated amino acid and reduced the reaction time by using microwave irradiation for SPPS.[72] Lastly, an elegant synthesis based on the combination of SPPS and chemical ligation has been described by Payne's group (Figure 2.13C).[73]

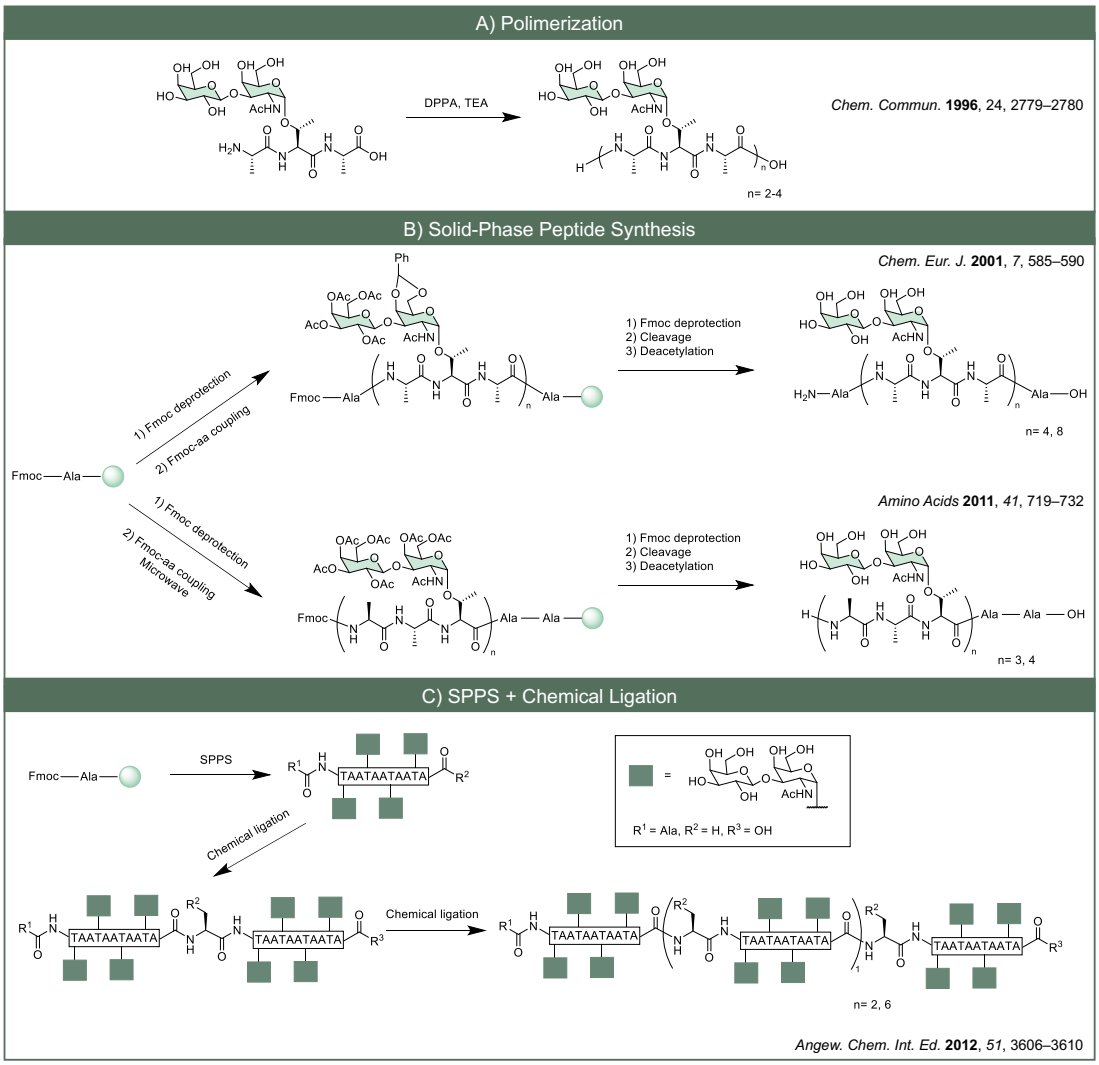


Figure 2.13. Diverse strategies followed for the synthesis of AFGPs.

From the point of view of the activity of these glycoproteins, one of the pioneering works that studies the relationship between their structures and their activity as inhibitors of ice crystal growth was carried out by Nishimura's group.[63] Antifreeze glycoproteins with chain length between two and seven tandem repeats units were synthesized by polymerization as previously described. For derivatives with n = 2 to 5, a positive correlation between chain length and thermal hysteresis was confirmed. However, from n = 5 to 7, a maximum value was observed for this magnitude (Figure 2.14).

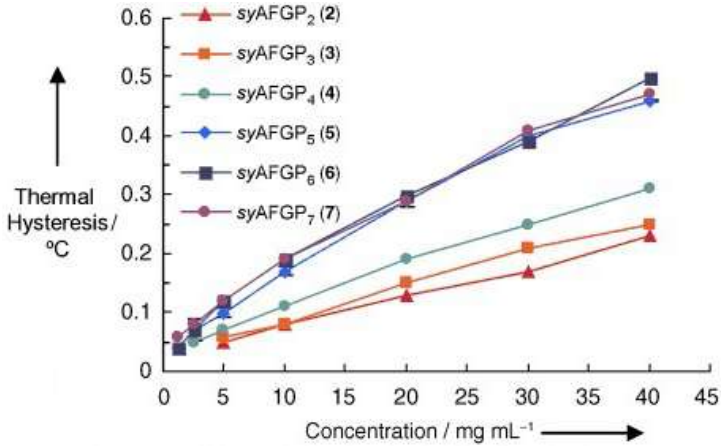


Figure 2.14. Influence of chain length of AFGPs (from 2 to 7 tandem repeats units) in thermal hysteresis activity.[63]

In parallel, glycopeptides were synthesized and studied in which both the peptide part and the nature of the sugar and the configuration of the glycosidic bond (from α- to β-*O*-glycosidic) were modified. Considering the results shown in Table 2.1, it can be concluded that the following structural aspects are key for the antifreeze activity of glycopeptides:

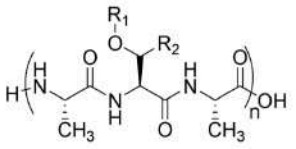
• The amino acid threonine, which provides the methyl group at C_β, is essential for TH activity.

• The first carbohydrate attached to the peptide must be GalNAc due to the importance of the acetamido group at carbon 2, as well as the configuration of the hydroxyl groups of this sugar.

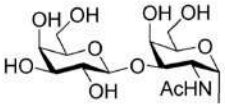
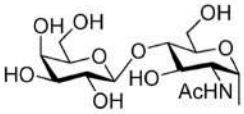
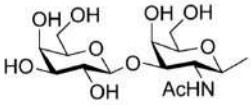
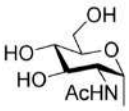
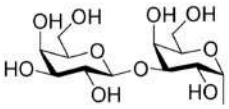
• The α-*O*-glycosidic bond of the residue attached to the peptide chain.

• Finally, although the terminal Gal does not seem to be necessary for freezing activity, the disaccharide Galβ(1-3)GalNAc slightly improves the activity.

Table 2.1. Synthetic AFGPs analogs and their antifreeze properties.[64]



,n= 8

Entry No.	R_1	R_2	Ice crystal morphology	TH	Inference
1.		H	Hexagonal	No	γ-methyl of Thr is essential
2.		CH_3	Hexagonal bipyramidal	Yes	Intra glycosidic linkage in the disaccharide does not affect TH activity
3.	H	CH_3	No change	No	Carbohydrate is essential
4.		CH_3	Hexagonal	No	α-glycosidic linkage between GalNAc and Thr is necessary
5.		CH_3	Hexagonal bipyramidal	Yes	Galactose is not necessary
6.		CH_3	No change	No	NHAc is essential

The molecular mechanism by which the antifreeze glycoproteins act as potent inhibitors of ice formation is currently unclear. In fact, two opposing models have been reported to explain the mechanism of inhibition. The 'classical' model[74] states that these glycoproteins bind *irreversibly* to the ice surface through the hydroxyl groups of the sugar (Figure 2.15A). The alternative model[75] assumes that they bind *reversibly* to the ice via methyl groups of the peptide and the saccharide and the driving force of the binding is the entropy of dehydration of the hydrophobic groups as they settle into the cavities of the ice surface, as well as the returning of some water molecules to the bulk water (Figure 2.15B).

In chapter 6, it will be discussed in more detail the molecular basis for the activity of these glycoproteins via the synthesis of AFGPs with several changes in their chemical structure and the study of their thermal hysteresis, ice crystal morphology, conformational analysis and computational studies.

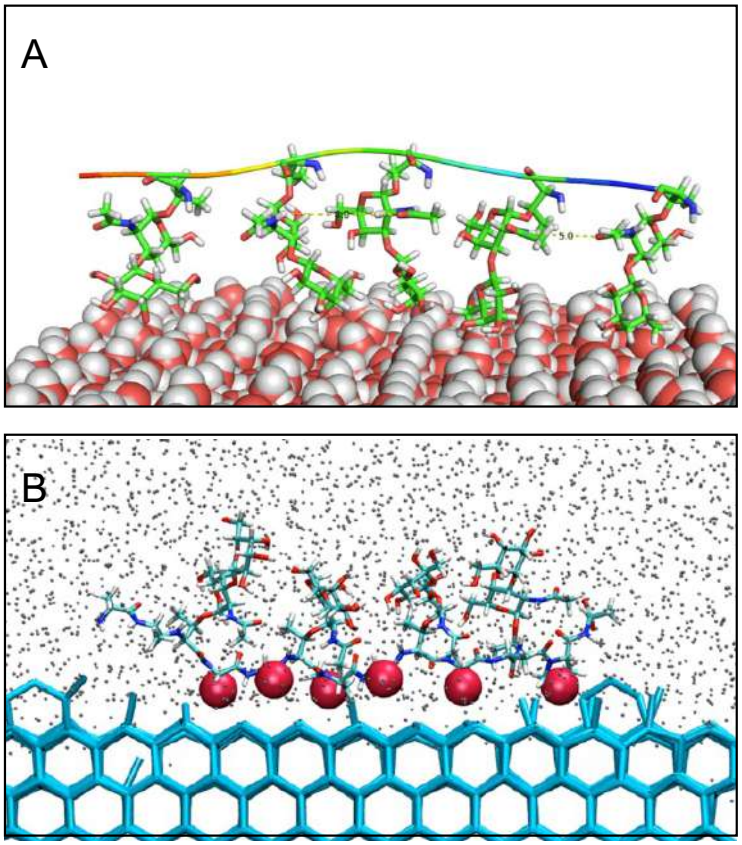


Figure 2.15. Two opposing models to explain the mechanism of inhibition of ice growth by antifreeze glycoproteins. (A) In this model, hydroxyl groups of the galactose residue interact irreversibly with the ice surface. (B) In this model, the methyl groups of the peptide chain and of the carbohydrate residue interact reversibly with the ice surface.

2.5. References

(1) Hollingsworth, M. A.; Swanson, B. J. *Nat. Rev. Cancer* **2004**, *4*, 45–60.

(2) Hattrup, C. L.; Gendler, S. J. *Annu. Rev. Physiol.* **2008**, *70*, 431–457.

(3) Zhou, D.; Xu, L.; Huang, W.; Tonn, T. *Molecules* **2018**, *23,* 1326–1353.

(4) Soliman, C.; Yuriev, E.; Ramsland, P. A. *Curr. Opin. Struct. Biol.* **2017**, *44*, 1–8.

(5) Brooks, C. L.; Schietinger, A.; Borisova, S. N.; Kufer, P.; Okon, M.; Hirama, T.; Mackenzie, C. R.; Wang, L.-X.; Schreiber, H.; Evans, S. V. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10056–10061.

(6) Borgert, A.; Heimburg-Molinaro, J.; Song, X.; Lasanajak, Y.; Ju, T.; Liu, M.; Thompson, P.; Ragupathi, G.; Barany, G.; Smith, D. F.; Cummings, R. D.; Live, D. *ACS Chem. Biol.* **2012**, *7*, 1031–1039.

(7) Dokurno, P.; Bates, P. A.; Band, H. A.; Stewart, L. M. D.; Lally, J. M.; Burchell, J. M.; Taylor-Papadimitriou, J.; Snary, D.; Sternberg, M. J. E.; Freemont, P. S. *J. Mol. Biol.* **1998**, *284*, 713–728.

(8) Möller, H.; Serttas, N.; Paulsen, H.; Burchell, J. M.; Taylor-Papadimitriou, J.; Meyer, B. *Eur. J. Biochem.* **2002**, *269*, 1444–1455.

(9) Martínez-Sáez, N.; Castro-López, J.; Valero-González, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem. Int. Ed.* **2015**, *54*, 9830–9834.

(10) Grinstead, J. S.; Koganty, R. R.; Krantz, M. J.; Longenecker, B. M.; Campbell, A. P. *Biochemistry* **2002**, *41*, 9946–9961.

(11) Grinstead, J. S.; Schuman, J. T.; Campbell, A. P. *Biochemistry* **2003**, *42*, 14293–14305.

(12) Schuman, J.; Campbell, A. P.; Koganty, R. R.; Longenecker, B. M. *J. Pept. Res.* **2003**, *61*, 91–108.

(13) Movahedin, M.; Brooks, T. M.; Supekar, N. T.; Gokanapudi, N.; Boons, G.-J.; Brooks, C. L. *Glycobiology* **2017**, *27*, 677–687.

(14) Coelho, H.; Matsushita, T.; Artigas, G.; Hinou, H.; Cañada, F. J.; Lo-Man, R.; Leclerc, C.; Cabrita, E. J.; Jiménez-Barbero, J.; Nishimura, S. I.; García-Martín, F.; Marcelo, F. *J. Am. Chem. Soc.* **2015**, *137*, 12438–12441.

(15) Martínez-Sáez, N.; Supekar, N. T.; Wolfert, M. A.; Bermejo, I. A.; Hurtado-Guerrero, R.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Boons, G. J.; Peregrina, J. M.; Corzana, F. *Chem. Sci.* **2016**, *7*, 2294–2301.

(16) Dondoni, A.; Marra, A. *Chem. Rev.* **2000**, *100*, 4395–4421.

(17) Kuberan, B.; Sikkander, S. A.; Tomiyama, H.; Linhardt, R. J. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 2073–2075.

(18) Awad, L.; Madani, R.; Gillig, A.; Kolympadi, M.; Philgren, M.; Muhs, A.; Gérard, C.; Vogel, P. *Chem. Eur. J.* **2012**, *18*, 8578–8582.

(19) Vichier-Guerre, S.; Lo-Man, R.; Huteau, V.; Dériaud, E.; Leclerc, C.; Bay, S. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3567–3570.

(20) Norgren, A. S.; Arvidsson, P. I. *Org. Biomol. Chem.* **2005**, *3*, 1359–1361.

(21) Oberbillig, T.; Mersch, C.; Wagner, S.; Hoffmann-Röder, A. *Chem. Commun.* **2012**, *48*, 1487–1489.

(22) Johannes, M.; Reindl, M.; Gerlitzki, B.; Schmitt, E.; Hoffmann-Röder, A. *Beilstein J. Org. Chem. 1115* **2015**, *11*, 155–161.

(23) Bermejo, I. A.; Usabiaga, I.; Compañón, I.; Castro-López, J.; Insausti, A.; Fernández, J. A.; Avenoza, A.; Busto, J. H.; Jiménez-Barbero, J.; Asensio, J. L.; Peregrina, J. L; Peregrina, J. M.; Jiménez-Osés, G.; Hurtado-Guerrero, R.; Cocinero, E. J.; Corzana, F. *J. Am. Chem. Soc.* **2018**, *140*, 9952–9960.

(24) Hoffmann-Röder, A.; Kaiser, A.; Wagner, S.; Gaidzik, N.; Kowalczyk, D.; Westerlind, U.; Gerlitzki, B.; Schmitt, E.; Kunz, H. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 8498–8503.

(25) Yang, F.; Zheng, X. J.; Huo, C. X.; Wang, Y.; Zhang, Y.; Ye, X. S. *ACS Chem. Biol.* **2011**, *6*, 252–259.

(26) Song, C.; Zheng, X. J.; Guo, H.; Cao, Y.; Zhang, F.; Li, Q.; Ye, X. S.; Zhou, Y. *Glycoconj. J.* **2019**, *36*, 399–408.

(27) Wagner, S.; Mersch, C.; Hoffmann-Röder, A. *Chem. Eur. J.* **2010**, *16*, 7319–7330.

(28) Jiménez-Barbero, J.; Dragoni, E.; Venturis, C.; Nannucci, F.; Ardá, A.; Fontanella, M.; André, S.; Cañada, F. J.; Gabius, H. J.; Nativi, C. *Chem. Eur. J.* **2009**, *15*, 10423–10431.

(29) Richichi, B.; Thomas, B.; Fiore, M.; Bosco, R.; Qureshi, H.; Nativi, C.; Renaudet, O.; BenMohamed, L. *Angew. Chem. Int. Ed.* **2014**, *53*, 11917–11920.

(30) Aydillo, C.; Jiménez-Osés, G.; Busto, J. H.; Peregrina, J. M.; Zurbano, M. M.; Avenoza, A. *Chem. Eur. J.* **2007**, *13*, 4840–4848.

(31) Avenoza, A.; Cativiela, C.; Corzana, F.; Peregrina, J. M.; Sucunza, D.; Zurbano, M. M. *Tetrahedron: Asymmetry* **2001**, *6*, 949–957.

(32) Corzana, F.; Busto, J. H.; Marcelo, F.; De Luis, M. G.; Asensio, J. L.; Martín-Santamaría, S.; Sáenz, Y.; Torres, C.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. *Chem. Commun.* **2011**, *47*, 5319–5321.

(33) Lakshminarayanan, V.; Thompson, P.; Wolfert, M. A.; Buskas, T.; Bradley, J. M.; Pathangey, L. B.; Madsen, C. S.; Cohen, P. A.; Gendler, S. J.; Boons, G.-J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 261–266.

(34) Bermejo, I. A. 'Structure-Based Design of Glycopeptides Featuring Unnatural Tn Antigens and Their Applications to Cancer Vaccination and Diagnosis'. Doctoral dissertation (Universidad de La Rioja 2018).

(35) Xiao, A.; Zheng, X. J.; Song, C.; Gui, Y.; Huo, C. X.; Ye, X. S. *Org. Biomol. Chem.* **2016**, *14*, 7226–7237.

(36) Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. *J. Am. Chem. Soc.* **2006**, *128*, 14640–14648.

(37) Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; García de Luis, M.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. *J. Am. Chem. Soc.* **2007**, *129*, 9458–9467.

(38) Wu, X.; Lipinski, T.; Paszkiewicz, E.; Bundle, D. R. *Chem. Eur. J.* **2008**, *14*, 6474–6482.

(39) Bundle, D. R.; Rich, J. R.; Jacques, S.; Yu, H. N.; Nitz, M.; Ling, C. C. *Angew. Chem. Int. Ed.* **2005**, *44*, 7725–7729.

(40) Huo, C. X.; Zheng, X. J.; Xiao, A.; Liu, C. C.; Sun, S.; Lv, Z.; Ye, X. S. *Org. Biomol. Chem.* **2015**, *13*, 3677–3690.

(41) Bousquet, E.; Spadaro, A.; Pappalardo, M. S.; Bernardini, R.; Romeo, R.; Panza, L.; Ronsisvalle, G. *J. Carbohydr. Chem.* **2000**, *19*, 527–541.

(42) Compañón, I.; Guerreiro, A.; Mangini, V.; Castro-López, J.; Escudero-Casao, M.; Avenoza, A.; Busto, J. H.; Castillo, S.; Asensio, J. L.; Jiménez-Osés, G.; Boutureira, O.; Peregrina, J. M.; Hurtado-Guerrero, R.; Fiammengo, R.; Bernardes, G. J. L.; Corzana, F. *J. Am. Chem. Soc.* **2019**, *141*, 4063–4072..

(43) Rojas-Ocáriz, V.; Compañón, I.; Aydillo, C.; Castro-Lopez, J.; Jiménez-Barbero, J.; Hurtado-Guerrero, R.; Avenoza, A.; Zurbano, M. M.; Peregrina, J. M.; Busto, J. H.; Corzana, F. *J. Org. Chem.* **2016**, *81*, 5929–5941.

(44) Bermejo, I. A.; Navo, C. D.; Castro-López, J.; Guerreiro, A.; Jiménez-Moreno, E.; Sánchez Fernández, E. M.; García-Martín, F.; Hinou, H.; Nishimura, S. I.; García Fernández, J. M.; Ortiz Mellet, C.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L.; Hurtado-Guerrero, R.; Peregrina, J. M.; Corzana, F. *Chem. Sci.* **2020**, *11*, 3996–4006.

(45) Somovilla, V. J.; Bermejo, I. A.; Albuquerque, I. S.; Martínez-Sáez, N.; Castro-López, J.; García-Martín, F.; Compañón, I.; Hinou, H.; Nishimura, S. I.; Jiménez-Barbero, J.; Asensio, J. M.; Avenoza, A; Busto, J. H.; Hurtado-Guerrero, R.; Peregrina, J. M.; Bernardes, G. J. L.; Corzana, F. *J. Am. Chem. Soc.* **2017**, *139*, 18255–18261.

(46) Wu, J.; Li, X.; Song, W.; Fang, Y.; Yu, L.; Liu, S.; Churilov, L. P.; Zhang, F. *Autoimmun. Rev.* **2017**, *16*, 1270–1281.

(47) Disis, M. L.; Pupa, S. M.; Gralow, J. R.; Dittadi, R.; Menard, S.; Cheever, M. A. *J. Clin. Oncol* **2016**, *15*, 3363–3367.

(48) Kurtenkov, O.; Klaamas, K.; Rittenhouse-Olson, K.; Vahter, L.; Sergejev, B.; Miljukhina, L.; Shljapnikova, L. *Exp. Oncol.* **2005**, *27*, 136–140.

(49) Tang, Y.; Wang, L.; Zhang, P.; Wei, H.; Gao, R.; Liu, X.; Yu, Y.; Wang, L. *Clin. Vaccine Immunol.* **2010**, *17*, 1903–1908.

(50) Gheybi, E.; Amani, J.; Salmanian, A. H.; Mashayekhi, F.; Khodi, S. *Tumor Biol.* **2014**, *35*, 11489–11497.

(51) Chapman, C.; Murray, A.; Chakrabarti, J.; Thorpe, A.; Woolston, C.; Sahin, U.; Barnes, A.; Robertson, J. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2007**, *18*, 868–873.

(52) Desmetz, C.; Bascoul-Mollevi, C.; Rochaix, P.; Lamy, P. J.; Kramar, A.; Rouanet, P.; Maudelonde, T.; Mangé, A.; Solassol, J. *Clin. Cancer Res.* **2009**, *15*, 4733–4741.

(53) Hermsen, B. B. J.; Verheijen, R. H. M.; Menko, F. H.; Gille, J. J. P.; van Uffelen, K.; Blankenstein, M. A.; Meijer, S.; van Diest, P. J.; Kenemans, P.; von Mensdorff-Pouilly, S. *Eur. J. Cancer* **2007**, *43*, 1556–1563.

(54) Pinheiro, S. P.; Hankinson, S. E.; Tworoger, S. S.; Rosner, B. A.; McKolanis, J. R.; Finn, O. J.; Cramer, D. W. *Cancer Epidemiol. Biomarkers Prev.* **2010**, *19*, 1595–1601.

(55) von Mensdorff-Pouilly, S.; Verstraeten, A.; Kenemans, P.; Snijdewint, F.; Kok, A.; Van Kamp, G.; Paul, M.; Van Diest, P.; Meijer, S.; Hilgers, J. *J. Clin. Oncol.* **2000**, *18*, 574–583.

(56) Blixt, O.; Bueti, D.; Burford, B.; Allen, D.; Julien, S.; Hollingsworth, M.; Gammerman, A.; Fentiman, I.; Taylor-Papadimitriou, J.; Burchell, J. M. *Breast Cancer Res.* **2011**, *13*, R25.

(57) Zhang, Y.; Li, M.; Gao, X.; Chen, Y.; Liu, T. *J. Hematol. Oncol.* **2019**, *12*, 1–13.

(58) Davies, P. L.; Hew, C. L. *FASEB J.* **1990**, *4*, 2460–2468.

(59) Venketesh, S.; Dayananda, C. *Crit. Rev. Biotechnol.* **2008**, *28*, 57–82.

(60) Nishijima, K.; Tanaka, M.; Sakai, Y.; Koshimoto, C.; Morimoto, M.; Watanabe, T.; Fan, J.; Kitajima, S. *Cryobiology* **2014**, *69*, 22–25.

(61) Sumii, Y.; Hibino, H.; Saidalimu, I.; Kawahara, H.; Shibata, N. *Chem. Commun.* **2018**, *54*, 9749–9752.

(62) Biggs, C. I.; Bailey, T. L.; Ben Graham; Stubbs, C.; Fayter, A.; Gibson, M. I. *Nat. Commun.* **2017**, *8*, 1–12.

(63) Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S. I. *Angew. Chem. Int. Ed.* **2004**, *43*, 856–862.

(64) Haridas, V.; Naik, S. *RSC Adv.* **2013**, *3*, 14199–14218.

(65) Peltier, R.; Brimble, M. A.; Wojnar, J. M.; Williams, D. E.; Evans, C. W.; Devries, A. L. *Chem. Sci.* **2010**, *1*, 538–551.

(66) Garner, J.; Harding, M. M. *ChemBioChem* **2010**, *11*, 2489–2498.

(67) Tsuda, T.; Nishimura, S. I. *Chem. Commun.* **1996**, No. 24, 2779–2780.

(68) Tachibana, Y.; Matsubara, N.; Nakajima, F.; Tsuda, T.; Tsuda, S.; Monde, K.; Nishimura, S. I. *Tetrahedron* **2002**, *58*, 10213–10224.

(69) Hachisu, M.; Hinou, H.; Takamichi, M.; Tsuda, S.; Koshida, S.; Nishimura, S. I. *Chem. Commun.* **2009**, No. 13, 1641–1643.

(70) Tachibana, Y.; Monde, K.; Nishimura, S. I. *Macromolecules* **2004**, *37*, 6771–6779.

(71) Tseng, P.-H.; Jiaang, W.; Chang, M.-Y.; Chen, S.-T. *Chem. Eur. J* **2001**, *7*, 585–590.

(72) Nagel, L.; Plattner, C.; Budke, C.; Majer, Z.; Devries, A. L.; Berkemeier, T.; Koop, T.; Sewald, N. *Amino Acids* **2011**, *41*, 719–732.

(73) Wilkinson, B. L.; Stone, R. S.; Capicciotti, C. J.; Thaysen-Andersen, M.; Matthews, J. M.; Packer, N. H.; Ben, R. N.; Payne, R. J. *Angew. Chem. Int. Ed.* **2012**, *51*, 3606–3610.

(74) Meister, K.; Devries, A. L.; Bakker, H. J.; Drori, R. *J. Am. Chem. Soc.* **2018**, *140*, 9365–9368.

(75) Mochizuki, K.; Molinero, V. *J. Am. Chem. Soc.* **2018**, *140*, 4803–4811.

3 Objectives

Given the relevance of *O*-glycopeptides in the fight against cancer from different viewpoints, two important classes of glycosylated peptides will be studied in the present dissertation: MUC1-like glycopeptides, together with some non-natural derivatives and antifreeze-related glycopeptides containing the natural peptide sequence but in which the carbohydrates attached to the threonine residue are altered compared to the naturally occurring ones.

As for the first class, the mucin MUC1 is abnormally glycosylated and overexpressed on the surface of various types of cancer cells, promoting an immune response with the formation of autoantibodies. This peculiarity will be useful for the development of a biosensor for the early diagnosis of cancer.

To achieve this goal, the following specific objectives will be proposed. As a first step, the **molecular basis of recognition between the 5E5 anti-MUC1 antibody with different mucin-like glycopeptides will be studied by X-ray crystallography.** This information is essential to know the structural basis of how antibodies recognize their target and will help us to design MUC1-glycopeptides with improved binding properties and specificity towards anti-MUC1 antibodies which, in turn, will enhance the sensitivity of the biosensor.

With this information in hand, a **structure-guided design of unnatural MUC1-like glycopeptides will be carried out and the affinity towards anti-MUC1 antibodies will be then experimentally determined. Those glycopeptides showing enhanced binding will be conjugated to gold nanoparticles for use as biosensors in serologic assays.**

On the other hand, antifreeze glycoproteins have recently attracted interest due to their use as crystal inhibitors for applications such as tumor cryotherapy. However, the mechanism of action of these glycoproteins is currently unclear. On this basis, **various antifreeze glycopeptide derivatives will be synthesized to contribute to elucidating the mechanism by which natural antifreeze glycoproteins can act as potent inhibitors of ice formation.**

4 Structural and biophysical insights into the broad recognition of glycopeptides by a cancer-specific anti-MUC1 antibody

4.1. Introduction

Despite the vast amount of clinical and immunological data available on anti-MUC1 antibodies, the molecular details by which these antibodies recognize their targets are scarce.[1] This information is essential, for example, to develop enhanced no natural anti-MUC1 antibodies to advance in cancer therapy.[2] Moreover, this structural knowledge may permit the design of new MUC1 glycopeptides with improved binding strength and specificity for anti-MUC1 antibodies to develop early cancer detection assays.[3]

As we have shown in Chapter 2, the peptide sequence PDTRP of MUC1 is the most recognized motif by the majority of anti-MUC1 antibodies. However, GSTAP and GVTS sequences can also be recognized by some of these antibodies.[4–6] In turn, when the PDTRP motif is glycosylated with GalNAc, this modification enhances the binding to the antibodies. Interestingly, the X-ray structure obtained for the complex between glycopeptide APDT*RP (in which T* = Tn-Thr, GalNAc-α-1-*O*-Thr) and AR20.5 antibody confirms the lack of contacts between the carbohydrate and the antibody. In this case, it is proposed that the presence of the sugar favors in solution the conformation recognized by the antibody, thus improving the binding.[1] A different scenario was observed for the SM3 antibody. The sugar is now involved in two stabilizing interactions with the protein, which may explain the moderate enhancement of affinity when the PDTR is glycosylated with GalNAc (see Chapter 2).[7]

Very recently, Nishimura and co-workers have reported the crystal structure of an anti-MUC1 antibody, SN-101, that recognizes the most immunogenic fragment of MUC1 and the GalNAc moiety.[8] Specifically, OH-4 and OH-6 of GalNAc interact through hydrogen bonds with two residues of the antibody, whereas interactions with the *N*-acetyl group of GalNAc are not observed. In addition, the GalNAc unit is inserted into the binding cavity allowing contacts between the glycopeptide and the antibody which precludes the recognition of other tumor-associated carbohydrate antigens such as T or STn antigens that cannot fit in this cavity (Figure 4.1).

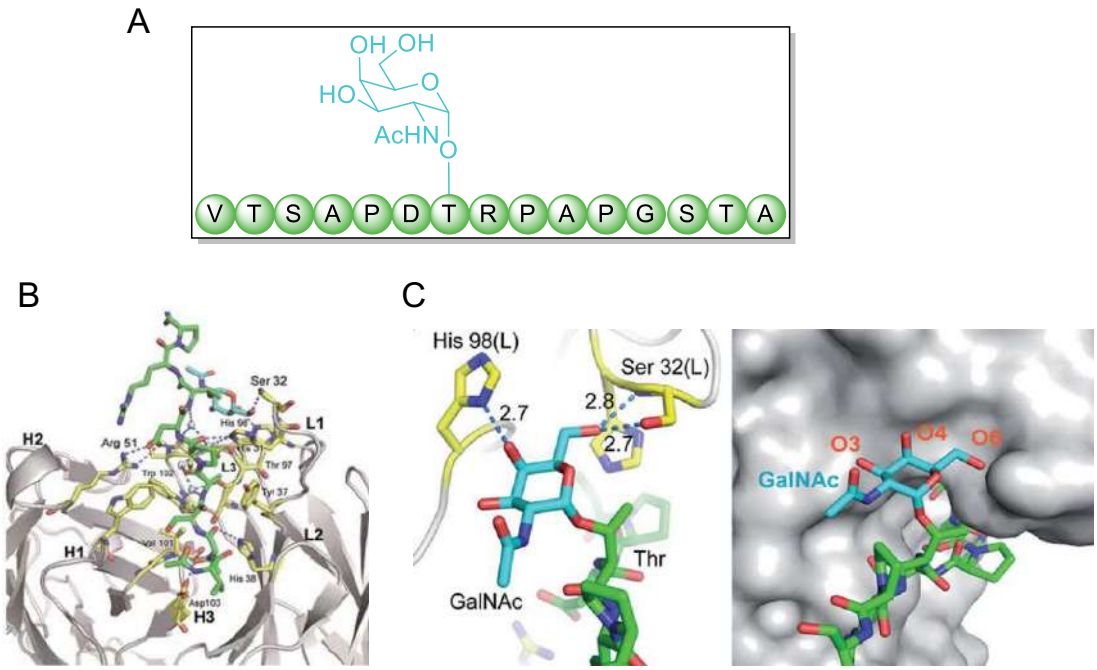


Figure 4.1. X-ray structure of SN-101 mAb in complex with a glycopeptide (pdb ID: 6KX1). (A) Glycopeptide VTSAPDT*RPAPGSTA used in this study. (B) Binding site of SN-101 Fab complexed with MUC1 glycopeptide. (C) Structures focusing on GalNAc recognition by SN-101. Binding interactions of the GalNAc unit by hydrogen bonds between O-4 and O-6 with His98L and Ser32L, respectively (dashed line in blue).[8]

Another interesting example of anti-MUC1 antibody that recognizes both the peptide sequence and the GalNAc moiety is the 5E5 antibody. As a result, this antibody does not bind to unglycosylated MUC1 peptides[9] whereas it presents a great selectivity for Tn and STn moieties located at the GSTAP region of MUC1 (Figure 4.2). Interestingly, recent studies reveal a degree of promiscuity of this antibody due to the fact that it also recognizes tumor cell lines with no MUC1-like glycopeptides expression.[10,11]

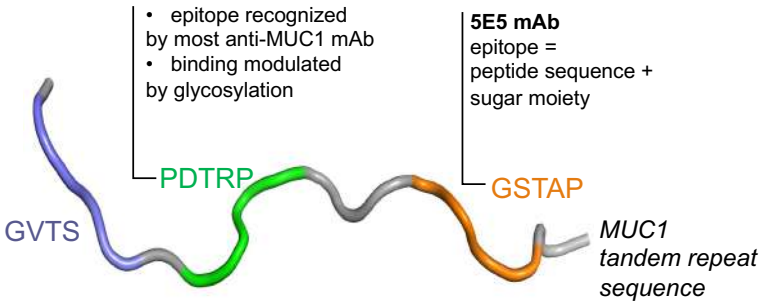


Figure 4.2. Main epitopes of MUC1 against anti-MUC1 mAb.

From a clinical perspective, 5E5 antibody has been successfully used to develop chimeric antigen receptor T cells (CAR-T cells) with target-specific cytotoxicity, demonstrating reduction of tumor growth in some tumors that express Tn-MUC1 on the cell surface.[6,10] 5E5 antibody has demonstrated its importance to be studied owing to its possibilities as a vaccine candidate or as a target for diagnosis of cancer.

4.2. Objectives

The main objective of this chapter is to disclose how anti-MUC1 5E5 antibody recognize mucin-like glycopeptides at the atomic level.

Initially, Tn-Thr (GalNAc-α-1-*O*-Thr), Tn-Ser (GalNAc-α-1-*O*-Ser) or STn-Thr (Neu5Ac-α-(2-6)-GalNAc-α-1-*O*-Thr) antigens will be incorporated in different peptidic fragments of MUC1. Then, combining X-ray crystallography, analytical biochemistry, peptide synthesis, molecular modeling and microarray assays we will determine the minimal epitope for this antibody and try to explain the degree of promiscuity for different glycopeptides.

Finally, by using an approach based on gold nanoparticles, we will perform a serologic assay with prostate cancer patients to attempt the detection of tumor-associated autoantibodies using this minimal epitope. In parallel, we will perform a hydrogen-by-fluorine substitution within the MUC1 mucin backbone to study how this modification could affect the antigen-antibody binding affinity.

4.3. Results

4.3.1. Synthesis of (glyco)peptides

Unnatural amino acids Fmoc-Thr[GalNAc(Ac)$_3$-α-D]-OH (protected Tn-Thr), Fmoc-Ser[GalNAc(Ac)$_3$-α-D]-OH (protected Tn-Ser) and Fmoc-Thr[Neu5Ac(Ac)$_4$(Bn)-α-(2-6)-GalNAc(Ac)$_3$-α-D]-OH (protected STn-Thr) were synthesized in our lab following the procedures described in the literature.[12,13] The different (glyco)peptides were then prepared by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer with Rink Amide MBHA resin using the Fmoc strategy (Figure 4.3).

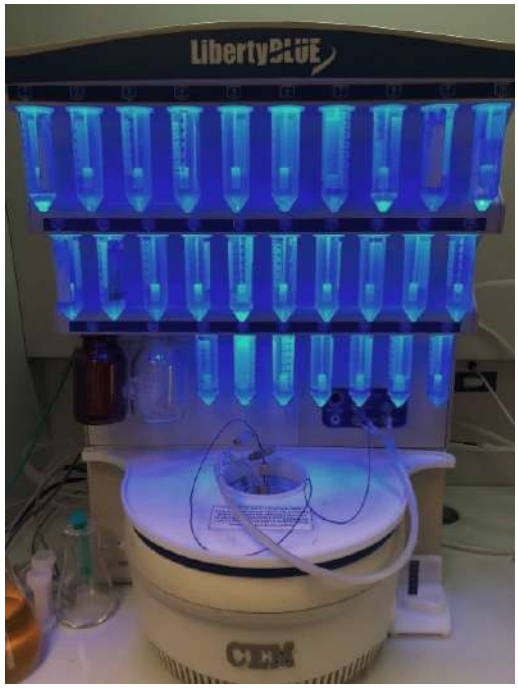


Figure 4.3. Liberty Blue synthesizer used for this work.

The building blocks Tn-Thr, Tn-Ser and STn-Thr were manually coupled as described in Experimental section 4.5.1. The number of equivalents was reduced and the reaction time was extended to increase the yield of this coupling step.[14] The *O*-acetyl groups of GalNAc moiety were then removed in a mixture of NH_2NH_2/MeOH (7:3) for the Tn antigen. Afterward, (glyco)peptides were released from the resin, and all acid sensitive protecting groups were

simultaneously removed using a cleavage cocktail consisting of a mixture of TFA 95%, TIS (triisopropylsilane) 2.5% and H_2O 2.5%, followed by precipitation with cold diethyl ether. In the case of the STn antigen derived glycopeptide, it was first detached from the resin using the aforementioned cleavage cocktail, which was further evaporated, purified through C18 cartridge and lyophilized. Afterward, benzyl group of the sialic acid moiety was removed through hydrogenolysis (Pd/C in MeOH). Finally, remaining *O*-acetyl groups were removed using a MeONa/MeOH solution. After completion of the reaction, solution was neutralized and evaporated. The crude products were purified by HPLC, lyophilized and characterized. Figure 4.4 shows the (glyco)peptides prepared in this chapter. All these derivatives present the GSTAP fragment recognized by 5E5 antibody.

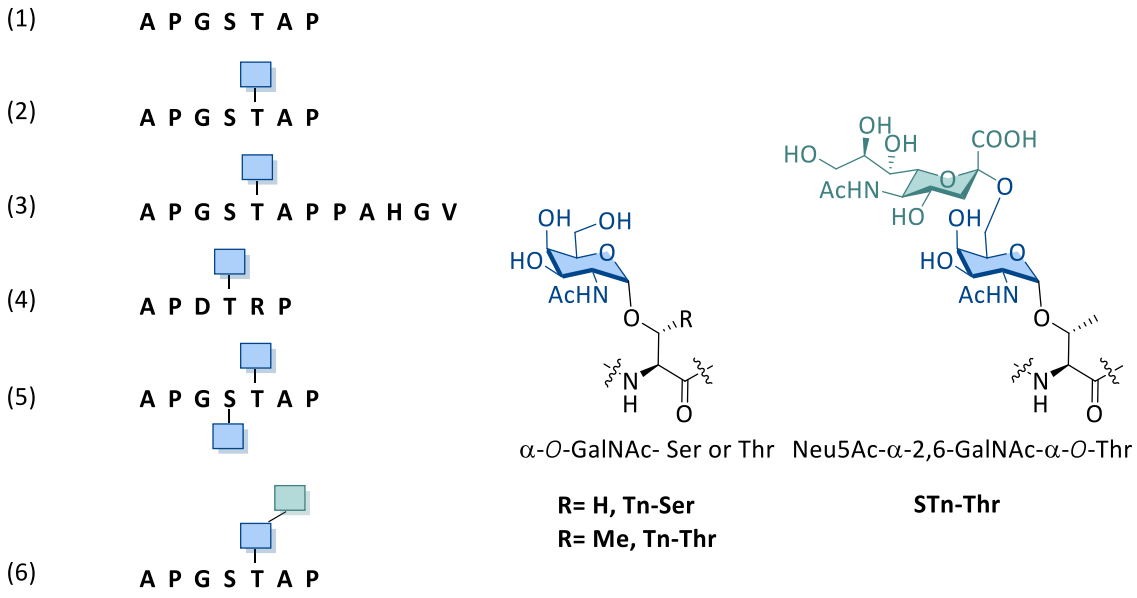


Figure 4.4. (Glyco)peptides synthesized and studied in this work. Note that all (glyco)peptides described in this thesis contain a *C*-terminal primary amide.

4.3.2. Affinity assays

As a first step, we determined the affinity of these (glyco)peptides toward 5E5 antibody. To this purpose, isothermal titration calorimetry (ITC) was used with a scFv-5E5 of this antibody. Thermodynamic binding parameters obtained by ITC for scFv-5E5 with these (glyco)peptides are summarized in Table 4.1.

Table 4.1. Thermodynamic binding parameters obtained by ITC for scFv-5E5 with several (glycol)peptides at 25 °C and pH 7.5.

Compound	Peptide	K_D (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	n
1	APGSTAP	n.d.[a]	--	--	--	--
2	APGST*AP	0.96 ± 0.21	−8.20	−0.72 ± 0.31	−7.47	1.2
3	APGST*APPAHGV	1.28 ± 0.40	−8.03	−5.38 ± 0.28	−2.65	1.2
4	APDT*RP	22.5 ± 3.3	−6.34	−1.89 ± 0.12	−4.45	1.2
5	APGS*T*AP	4.2 ± 0.5	−7.32	−10.72 ± 0.4	3.40	1.0
6	APGST**AP	19.0 ± 4.9	−6.44	−0.47 ± 0.08	−5.97	0.7

[a] n.d. stands for binding not detected under our experimental conditions. T* = GalNAc-α-1-*O*-Thr and S* = GalNAc-α-1-*O*-Ser (Tn antigens) and T** = Neu5Ac-α -(2-6)-GalNAc-α -1-*O*-Thr (STn antigen).

As can be inferred from Table 4.1, the highest affinity of scFv-5E5 was found against glycopeptide **2** APGST*AP whereas it did not bind to unglycosylated peptide **1**. A similar result was obtained for the longer glycopeptide **3**, which implies that the minimal epitope is already comprised in glycopeptide **2**. Concerning the glycopeptide **5**, the presence of two Tn antigens in the peptide discreetly affected binding, while in glycopeptide **6** the presence of STn antigen provoked a 20-fold decrease in affinity (relative to peptides **2** and **6**). This is in agreement with the preferences reported earlier[5] by ELISA experiments where 5E5 binds preferentially to the MUC1 tandem repeat glycopeptide containing

Thr*16 and weaker than the same glycopeptide containing STn antigen, Thr**16, ($H_1GVT_4S_5APDT_9RPAPGS_{15}\mathbf{T_{16}}APPA_{20}$, Figure 4.5)

The ITC assays indicate that the binding of these glycopeptides was entropically driven, with little enthalpic contribution. The only exception was glycopeptide **5**, where binding was enthalpically driven, likely indicating a more rigid preorganized structure.

Interestingly, under our experimental conditions, scFv-5E5 bound to a different MUC1 epitope, such as the APDT*RP fragment though to a lesser extent. This clearly suggests that the antibody recognizes other regions of the tandem-repeat sequence of MUC1 and possibly other proteins as described before.[11]

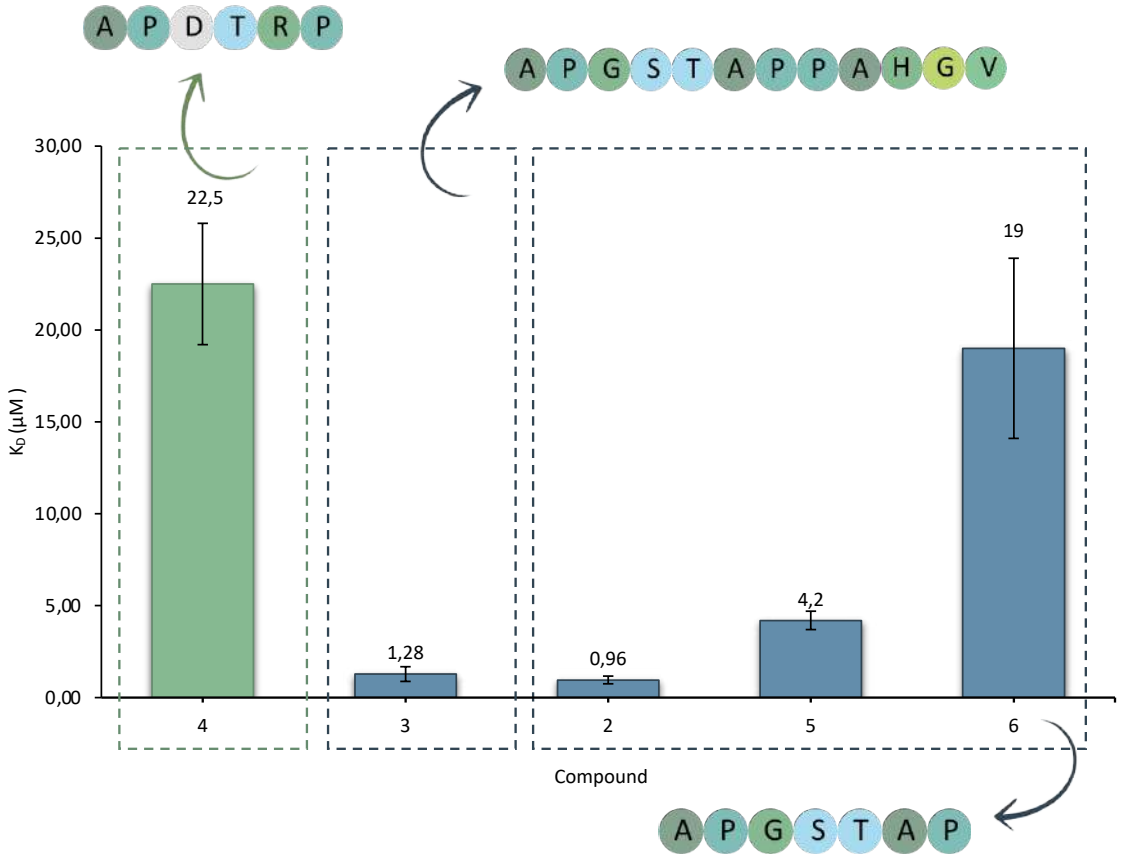


Figure 4.5. K_D (µM) of glycopeptides **2-6** with scFv-5E5 determined by ITC.

4.3.3. Crystallographic studies

Fortunately, we could obtain high-quality crystals of scFv-5E5 in a complex with glycopeptide **2** (Figure 4.6A and 4.6B; pdb entry: 6TNP) which allowed to solve the structure at a resolution of 3.0 Å. The glycopeptide lies within a surface groove formed by the light (L) and heavy (H) chains. GalNAc moiety remains in a deep solvent excluded pocket, while most of the peptide fragment is exposed to the solvent (Figure 4.6A).

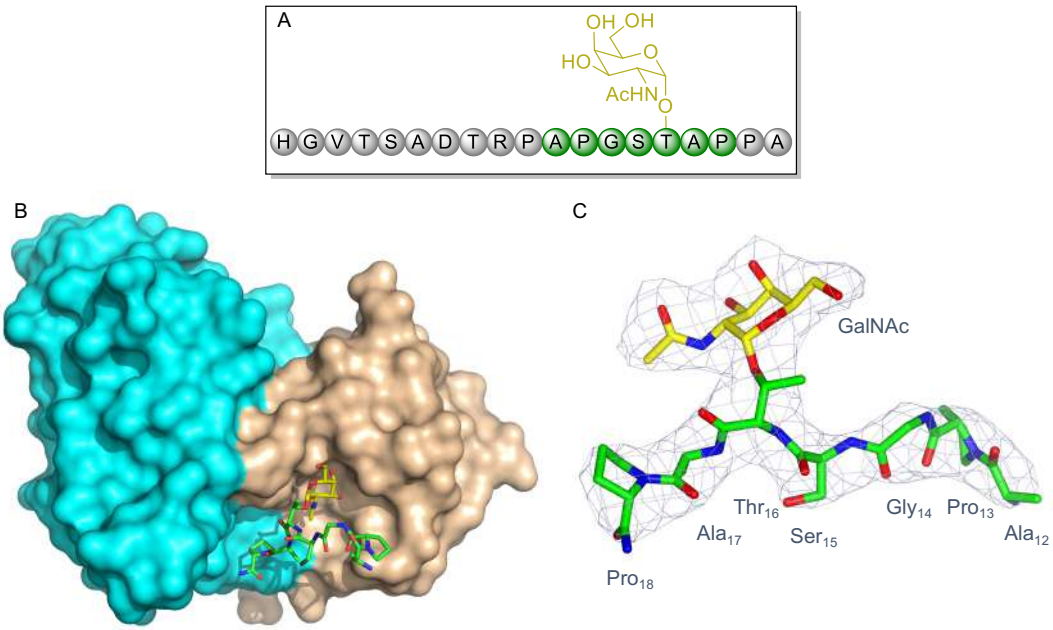


Figure 4.6. X-ray structure of glycopeptide **2** bound to scFv-5E5. (A) Surface representation of scFv-5E5 complexed with glycopeptide **2**. (B) The F_O–F_C electron density map (blue) is contoured at 2.2 σ for glycopeptide **2**.

The most recognized fragment of the glycopeptide is the carbohydrate which is predominantly interacting with the heavy chain (shown in brown in Figure 4.6A). The structure shows that all hydroxyl groups of GalNAc are involved in hydrogen bonds with exception of the OH-6. In detail, hydroxyl group OH-3 interacts with the NH group of $Ala33^H$ and OH-4 with the side chains of $His32^H$ and $Ser99^H$. (Figure 4.7). Three additional hydrogen bonds were found between the endocyclic oxygen of the sugar with $Ser99^H$ and among carbonyl group of GalNAc with the side chain of $His35^H$ and $Thr100^H$. The selectivity of 5E5 for GalNAc-containing glycopeptides could be explained by a CH-π interaction

between the methyl group of NHAc group of the sugar and the aromatic ring of $His50^{H}$. Furthermore, this antibody can also recognize STn-containing glycopeptides due to the completely solvent-exposition of the hydroxymethyl group of the sugar.[5]

The 'eclipsed' conformation of the glycosidic linkage was adopted for the Tn-Thr motif in solution,[15,16] with ϕ/ψ values of $68.1^{0}/151.0^{0}$. A single hydrogen bond was found between the peptide backbone and the antibody, involving $CONH_2$ group in the *C*-terminal region of the glycopeptide (P18) and the carbonyl group of $Tyr98^{L}$. On the contrary, the *N*-terminal region (A12) is fully exposed to the solvent, which entails no contact of this part with the antibody. This result is in contrast to the SN-101 antibody, which does not recognize other peptide moeities.[8] In the complex 5E5/**2**, the glycosylated Thr residue of the glycopeptide adopts a helix-like conformation in the bound state which could be caused by an intramolecular hydrogen bond between the NH of the contiguous Ala (Ala17) and the hydroxyl group of Ser15 (Figure 4.7). In addition, this hydrogen bond provides the appropriate shape complementarity with the scFv-5E5 antibody. Finally, CH-π interactions of peptide moiety with the antibody provide the selectivity of the peptide sequence (APGSTAP). One interaction between the methyl group of threonine with the ring of $Phe102^{H}$ and other between the ring of P18 and the aromatic moiety of $Tyr100^{L}$.

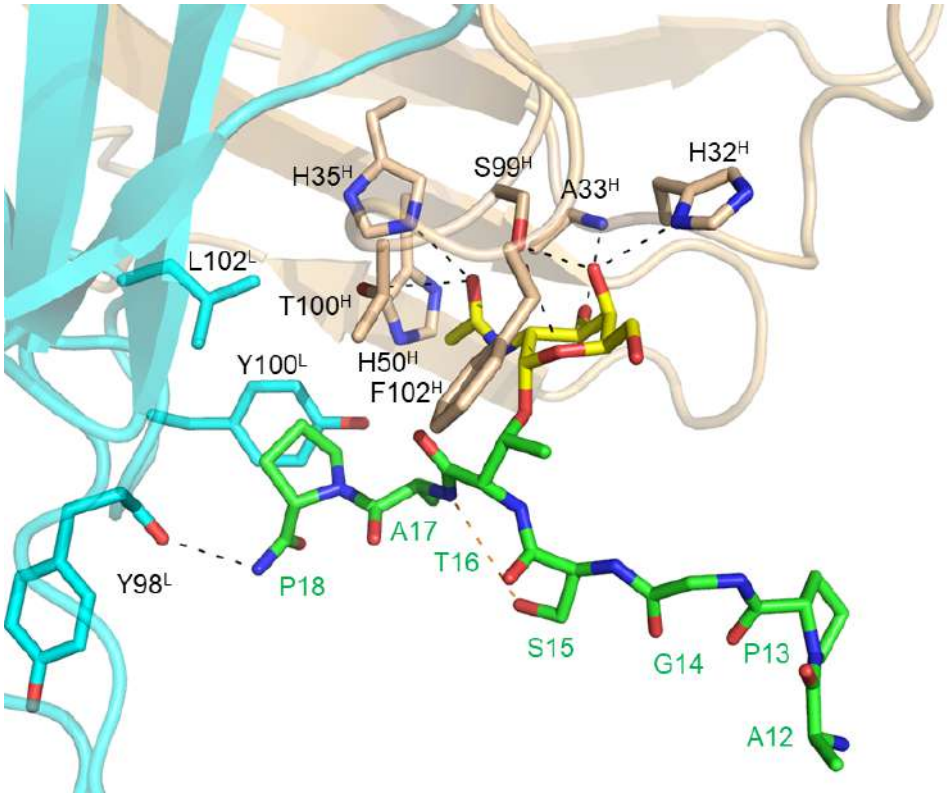


Figure 4.7. Interactions between glycopeptide **2** and scFv-5E5 in the X-ray structure. The antibody fragment is shown as a light blue (light chain) and brown (heavy chain) cartoon. The peptide fragment of the antigen and the GalNAc moiety are shown as green and yellow carbon atoms, respectively. Hydrogen bond interactions between the antibody and the antigen are shown as dotted black lines. The intramolecular hydrogen bond in the peptide backbone of **2** is shown (dotted orange lines). The PDB entry for this complex is 6TNP.

4.3.4. Affinity assays with antibody mutations and glycoform substitutions

To validate the X-ray structure, we expressed three mutants of scFv-5E5 antibody by site-directed mutagenesis and we determined their affinity with glycopeptide **2**. Initially, we performed the substitution of $F102^{H}$, $Y100^{L}$, $H32^{H}$ or $H35^{H}$ by Ala residues. We could not detect a clear binding for the first mutant, while it was found a decrease of ≈790 and ≈1330-fold in binding of mutants $Y100A^{L}$ and $H32A^{H}$-$H35A^{H}$ respectively towards peptide **2**. Then we replaced the *C*-terminal Pro of glycopeptide **2** with an Ala residue, observing a ≈400-fold decrease in affinity. These results indicate that the most important part of 5E5 epitope is the GalNAc-α-1-*O*-Thr moiety (Table 4.2)

Table 4.2. Thermodynamic binding parameters obtained by ITC for scFv-5E5 and three mutants with several glycopeptides at 25 °C and pH 7.5.

Compound	K_D (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	n
APGST*AP (**2**)	0.96 ± 0.21	–8.20	–0.72 ± 0.31	–7.47	1.2
APGST*AA (**7**)	382.0 ± 44.7	–4.66	6.54 ± 0.85	1.88	1.0
Mutant scFv-5E5 H32A-H35A + **2**	1275.4 ± 517.5	–3.95	–16.38 ± 8.55	12.44	1.2
Mutant scFv-5E5 Y100A + **2**	757.6 ± 99.7	–4.26	–1.34 ± 0.22	–2.91	1.2
Mutant scFv-5E5 F102A + **2**	n.d.[a]	--	--	--	--

[a] n.d. stands for binding not detected under our experimental conditions. T* = GalNAc-α-1-*O*-Thr.

In collaboration with the group of Prof. Clausen (University of Copenhagen), we performed a series of microarray experiments in order to identify the minimal recognition epitope. Firstly, it was carried out the substitution of each amino acid for an Ala in a MUC1-like glycopeptide sequence (Figure 4.8A). This simple study shows that the affinity only drastically decreases when Thr16* or Pro18 are mutated in the GST*AP sequence which implies that the minimal epitope should be -T*-A-P-. This result is in agreement with a previous work in which GST*AP sequence was mutated with Ala.[17] Furthermore, Ala of the -T*-A-P- epitope was substituted by different amino acids without significant differences in binding affinities which establish the 5E5 epitope as -T*-X-P- (Figure 4.8B), which complements the crystal analysis and previous data.[17]

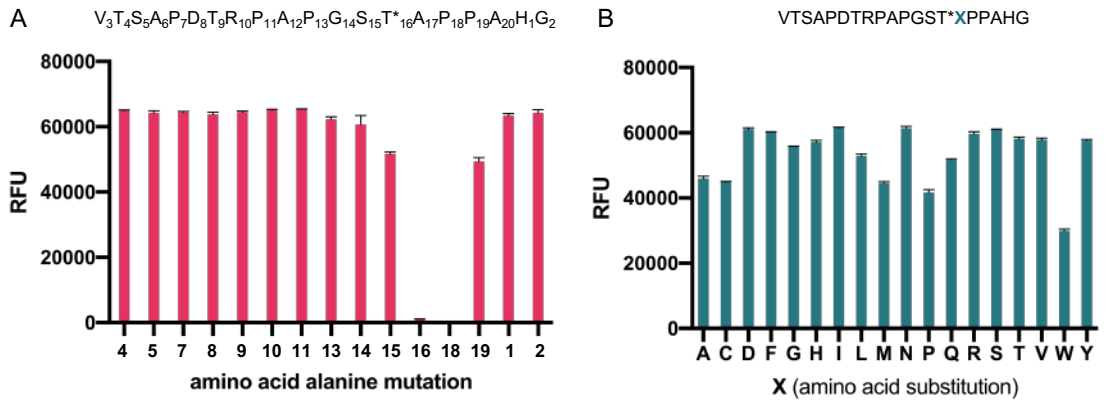


Figure 4.8. Microarray MUC1 epitope mapping analysis of 5E5 mAb. (A) Microarray data for an Ala mutation walk through the entire 20-mer MUC1-Tn glycopeptide (number corresponds to amino acid sequence position in sequence which was substituted with Ala). (B) Microarray data for the T*XP mutations in the GST*AP region (X amino acid substitution with different amino acids indicated as single letter code).

To determine whether the antibody can also recognize Cores 2, 3 and 4, we performed the synthesis of the glycopeptides shown in Figure 4.9. The microarrays studies shown that the antibody does not recognize these analogs. According to our crystal structure, these cores hinder the recognition of the GalNAc unit, probably due to steric issues.

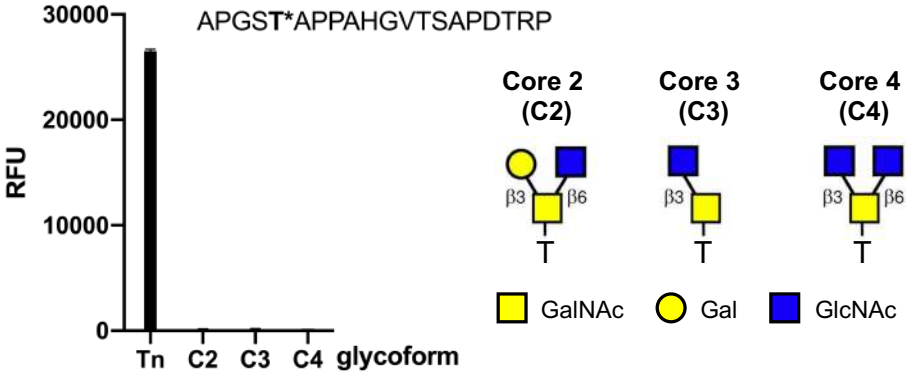


Figure 4.9. Microarray data for the additional glycoform isomers Core-2 (**C2**), Core-3 (**C3**) and Core-4 (**C4**) for MUC1-Tn glycopeptide.

4.3.5. MD simulations of glycopeptides in complex to scFv-5E5

To get insight into the molecular recognition of glycopeptides **2**, **4**–**6** in solution, we performed 0.5 μs Molecular Dynamics (MD) simulations of these derivatives complexed to scFv-5E5 (Figure 4.10). It is important to note that in accordance with the X-ray structure, hydrogen bonds between the sugar and the antibody were observed in all simulated complexes.

According to these simulations, complexes scFv-5E5/**2** (Figure 4.10A) and scFv-5E5/**6** (Figure 4.10C) exhibit flexibility in the *N*-terminal region, which confers a lack of stabilizing contacts between the peptide and the antibody. Stability of complex scFv-5E5/**2** was corroborated through the complete simulation, and it was detected a high population of hydrogen bonds and CH-π interactions previously observed in the X-ray structure. In the case of diglycopeptide **5** (Figure 4.10B), two significant differences were evident in the bound state with respect the other studied complexes. The first one was the rigidity of the *N*-terminal moiety of the peptide and on the other hand, the Ser residue is forced to adopt a helix-like conformation, which is low populated in solution.[18] Additionally, the extra sugar implies more contacts with the antibody which could counterbalance the entropy penalty associated to binding and could explain the decrease of binding observed by the ITC assays. For the scFv-5E5/**6** complex (Figure 4.10C), the Neu5Ac unit diminishes the recognition of the peptide fragment. After 120 ns, a considerable attenuation of the CH-π interaction between the *C*-terminal Pro and $Tyr100^{L}$ as a consequence of the increasing distance between them was observed. This characteristic may explain the moderately worse binding of this glycopeptide to the antibody compared with derivative **2**.

In order to demonstrate the relevance of methyl group of Thr16, MD calculations of glycopeptide **APGSS*AP** bound to scFv-5E5 were accomplished. The replacement of the Thr residue by Ser impaired the CH-π interactions of $F102^{H}$ and $Y100^{L}$ at the beginning of the simulation while it was conserved the stabilizing interactions GalNAc-scFv-5E5. This result was also observed in previous data in which MUC1-like mucin with TS*AP-motif in its sequence did not assist as a ligand for 5E5 antibody.[19] This result revalidated that the minimal epitope recognized by 5E5 mAb is -T*-X-P-.

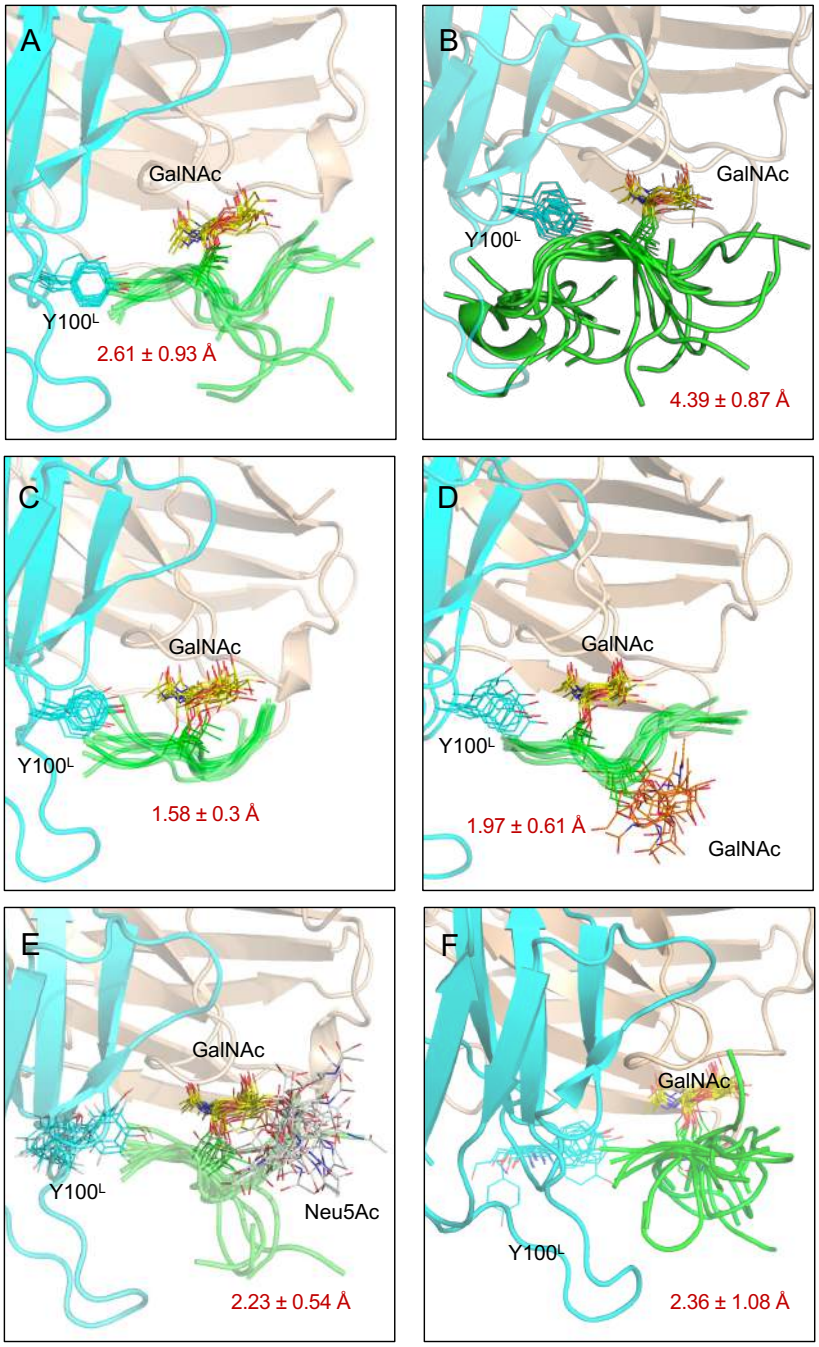


Figure 4.10. 0.5 μs MD simulations on the different complexes. Overlay of 10 frames of glycopeptides **2** (A), **3** (B), **4** (C), **5** (D), **6** (E) and **APGSS*AP** (F) in complex to scFv-5E5, sampled from 0.5 μs MD simulations performed in explicit water. The root-mean-square deviation (RMSD) values (± S.D.) of the peptide backbone are shown in red. Only the structure of the first frame of the antibody is shown for clarity. The antibody is shown as light blue (light chain) and brown (heavy chain) cartoon. The peptide backbone of the antigens is shown in green. The carbon atoms of the sugar moiety of GalNAc-α-1-*O*-Thr and GalNAc-α-1-*O*-Ser are in yellow and orange, respectively. The carbon atoms of Neu5Ac moiety are in grey.

Is the conformation of the epitope APGST*AP observed in the X-ray structure with 5E5 antibody the main conformer in solution? In order to answer the question, we carried out the synthesis and the conformational analysis of the glycopeptide **8** featuring the tandem repeat sequence of MUC1 GalNAc-glycosylated at threonine-16 (HGVTSAPDTRPAPGST*APPA). To this end, we combined NMR experiments (NOE experiments) with MD simulations. From the analysis of NMR spectra, we obtained 68 NMR-derived distances that were used as restraints in experimentally guided molecular dynamics simulations.

In parallel, Microscale thermophoresis (MST) assays of compound **8** revealed that this glycopeptide is recognized by both scFv-5E5 antibody and the complete antibody with similar K_D values (3.42 and 2.59 μM, respectively, Figure 4.11B).

These experimentally guided MD simulations showed that the peptide backbone is rather flexible in water, and in particular those amino acids further away from the glycosylation point exhibit a higher degree of flexibility. Moreover, glycopeptide **8** is forced to adopt an extended conformation due to the GalNAc moiety, which agrees with previous studies (Figure 4.11D, 4.11E).[8,10] Thus, 5E5 antibody recognized a conformation of the peptide backbone which is poorly populated in aqueous solution. This result is in line with those found for SM3 antibody.[7] However, it contrasts to that obtained for the SN-101 antibody,[8] for which NMR-derived structure in solution is similar to that recognized by the antibody.

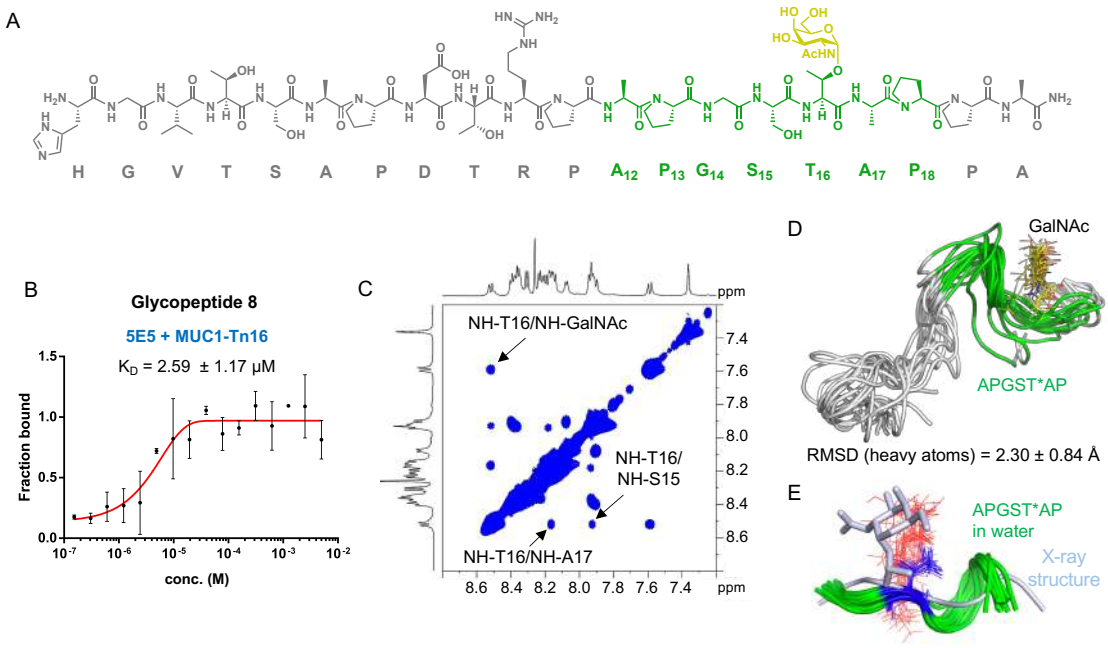


Figure 4.11. (A) Glycopeptide **8** synthesized and studied in this work. (B) MST binding curve used to determine K_D for the whole antibody 5E5 with glycopeptide **8**. Error represents standard deviation between two separate experiments; K_D values ± SD are shown. (C) Section of the 300 ms 2D-NOESY spectrum (600 MHz) in H_2O/D_2O (9:1) at 275 K and pH 5.5 for glycopeptide **8**. (D) Structural ensembles derived from 200 ns experiment-guided MD simulations for glycopeptide **8** in water. (E) Superposition of the APGST*AP fragment of glycopeptide **8** conformers found in solution, together with the conformation of glycopeptide **2** obtained from the X-ray structure of scFv-5E5/**2**.

4.3.6. Nanoparticle-based dot-blot assay with sera of cancer patients

The data reported herein may serve as a framework to design antigen mimics for the development of new diagnostic tools to detect cancer at an early stages of the disease. Previous studies have demonstrated that both MUC1 vaccinated patients[17] and cancer patients do indeed generate antibodies to these epitopes but they vary significantly among individuals and also depend on the type of glycosylation.[20]

Within this context, we have developed a flexible, convenient, and easily adaptable dot-blot assay, based on the conjugation of the artificial antigens on

gold nanoparticles (AuNP), for the detection of anti-MUC1 antibodies in human sera. (Figure 4.12)

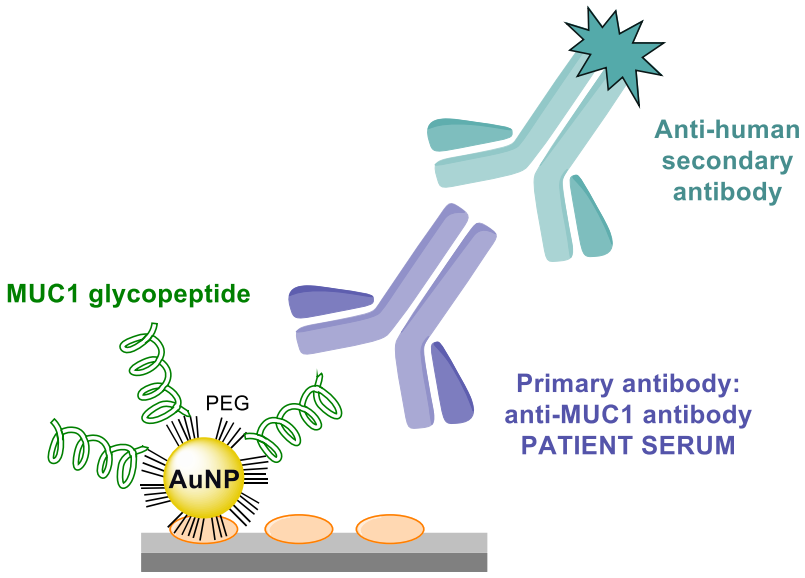


Figure 4.12. Schematic design of the assay used to detect anti-MUC1 antibodies in human sera.

To this purpose, gold nanoparticles (AuNPs) were functionalized with two synthetic glycopeptides whose sequences correspond to one complete tandem repeat domain of MUC1 but glycosylated either at Thr9 or at Thr16 (glycopeptide **9** and glycopeptide **8**, respectively, Figure 4.13B). The peptides were then coupled to PEG-AuNPs (Figure 4.13A) according to a previously reported procedure,[3] yielding conjugates AuNP-glycopeptide **9** and AuNP-glycopeptide **8**. The success of the conjugation reactions was confirmed through dynamic light scattering (DLS) and gel electrophoresis assays. As can be observed in Figure 4.13C, both conjugates were characterized by a reduced electrophoretic mobility relative to their precursor. The glycopeptide loading was determined to be ≈150 glycopeptides/AuNP for both conjugates by applying standard amino acid analysis.[21–23]

As a next step, AuNP-glycopeptide **9**, AuNP-glycopeptide **8** and a 1:1 mixture of AuNP-glycopeptide **9** and AuNP-glycopeptide **8** were spotted in triplicate onto PVDF membranes and incubated with sera of patients presenting prostate cancer. The presence of anti-MUC1 antibodies immobilized on membranes by

means of the specific binding to the glycopeptide antigens presented on the AuNPs was revealed by a fluorescently labeled secondary antibody. The membranes were imaged by using a laser scanner and the intensity of each spot was quantified by densitometry. Notably, 7 patients out of 9 with the diagnosis of prostatic adenocarcinoma showed detectable levels of anti-MUC1 antibodies, while sera from healthy patients did not give any detectable signal (Figure 4.13D). Moreover, there were no significant differences in signal/background intensity depending on the epitope used in the assays, except for one patient for which AuNP-glycopeptide **9** afforded ≈2-fold stronger signals than AuNP-glycopeptide **8**. It is important to note that it was also included 12 control sera, 5 from healthy men and 7 from healthy women. No detectable signals for anti MUC1 antibodies were observed in these cases. Therefore, the data reported here confirm that patients with prostate cancer do raise anti-MUC1 antibodies that recognize the same epitope as 5E5.

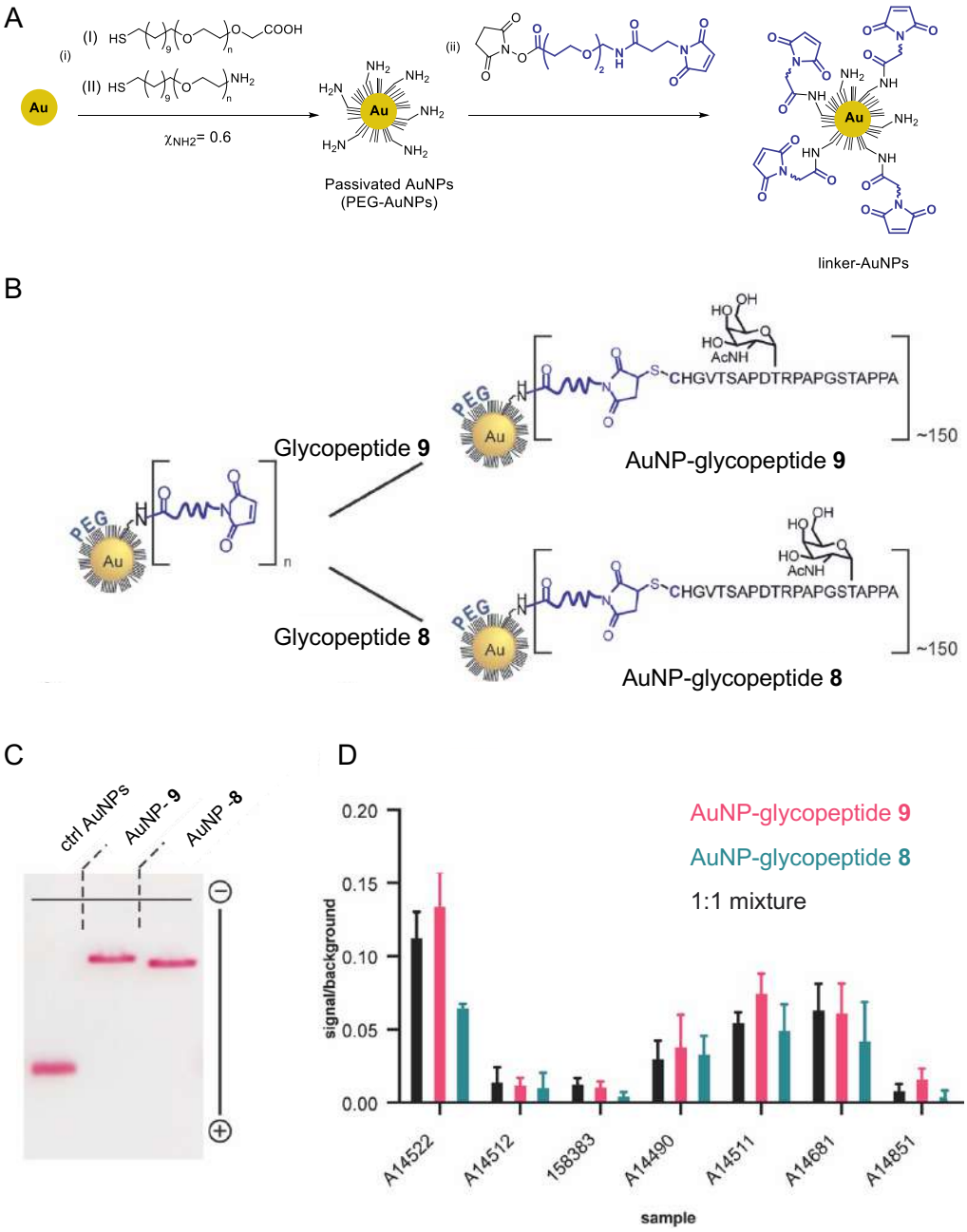


Figure 4.13. Detection of anti-MUC1 antibodies in cancer patients. (A) Synthetic route for the linker-AuNPs. (B) Schematic representation of conjugates containing glycopeptide **8** or glycopeptide **9** attached to the surface of AuNPs. (C) Agarose gel electrophoresis of precursor AuNPs (ctrl AuNP, loaded with the SM(PEG)$_2$-linker); and the glycopeptide-AuNP conjugates. (D) Signal/background obtained from the sera of different prostate cancer patients by using AuNP-glycopeptide **9**, AuNP-glycopeptide **8** or a 1:1 mixture of these conjugates as a detection system.

4.4. Conclusions

5E5-CART is currently under clinical trials[26] for different types of cancers including pancreatic ductal adenocarcinoma, which is a devastating condition with poor prognosis and rising incidence. Despite the translational significance of this approach, the molecular aspects of how 5E5 mAb recognizes MUC1-Tn epitopes have not been determined yet. In this chapter, by conducting a multidisciplinary approach it was uncovered that 5E5 mAb is able to bind to -T*-X-P-. This motif is present in a large number of proteins going through the secretory pathway. Interestingly, this promiscuity was suggested in a previous study revealing that 5E5-CART killed some cancer cell lines that did not express MUC1, as long as these lacked COSMC function and therefore presented Tn/STn antigens.[11]

Our studies indicate that the recognition of the GalNAc moiety is the main driving force of 5E5 mAb binding and the neighboring downstream proline residue also plays a key role. This implies that 5E5 antibody recognizes selectively the Tn antigen and can be useful as a diagnostic tool to detect Tn- and STn- expressing cancers and as a therapeutic and specific anticancer drug. Finally, our finding that epitope APGST*AP is sufficient to detect prostate cancer together with our structural and peptide engineering studies offer clues of how to modify glycopeptides containing the -T*-X-P- motif for cancer diagnosis.

4.5. Experimental section

4.5.1. General procedure for solid-phase peptide synthesis (SPPS).

All (glyco)peptides were synthesized by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol).

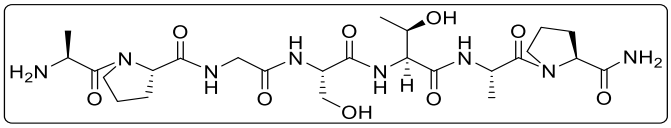
Fmoc-Thr[GalNAc(Ac)$_3$-α-D]-OH (Tn-Thr), Fmoc-Ser[GalNAc(Ac)$_3$-α-D]-OH (Tn-Ser) and Fmoc-Thr[Neu5Ac(Ac)$_4$(Bn)-α-(2-6)-GalNAc(Ac)$_3$-α-D]-OH (STn-Thr) (2.0 equiv) were synthesized as described in the literature,[12,13] and manually coupled using HBTU [(2-(1H-benzotriazol-1-yl)- 1,1,3,3-tetramethyluronium hexafluorophosphate], 0.9 equiv., and 0.25 mL of DIPEA (2.0 M in NMP) dissolved in 1 mL of DMF, while all other Fmoc amino acids (5.0 equiv) were automatically coupled using oxyma pure/DIC (N,N'-diisopropylcarbodiimide).

The *O*-acetyl groups of GalNAc moiety were removed in a mixture of NH_2NH_2/MeOH (7:3) for the Tn antigen. (Glyco)peptides were then released from the resin, and all acid sensitive protecting groups were simultaneously removed using TFA 95%, TIS (triisopropylsilane) 2.5% and H_2O 2.5%, followed by precipitation with cold diethyl ether.

In the case of the STn antigen derived glycopeptide, the derivative was first detached from the resin using the aforementioned cleavage cocktail, which was further evaporated, purified through C18 cartridge and lyophilized. Afterward, benzyl group of the sialic acid was removed through hydrogenolysis (Pd-C in MeOH) overnight. Once this reaction is completed, Pd-C is filtered and solvent evaporated. Finally, remaining acetyl groups are removed using a MeONa/MeOH solution until pH 9.5 is reached. After completion of the reaction, solution is neutralized and evaporated.

The crude products were purified by HPLC on a Phenomenex Luna C18(2) column (10 μm, 250 mm × 21.2 mm) and a dual absorbance detector, with a flow rate of 10 or 20 mL/min.

Synthesis of peptide 1 ($A_1P_2GSTA_6P_7$)



Following SPPS methodology with the adequately protected amino acids, peptide **1** was obtained and purified by semi-preparative HPLC.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.52 (q, 1H, *J* = 7.0 Hz, $H\alpha_{Ala6}$), 4.45 (t, 1H, *J* = 5.6 Hz, $H\alpha_{Ser}$), 4.37 – 4.43 (m, 1H, $H\alpha_{Pro2}$), 4.22 – 4.32 (m, 3H, $H\alpha_{Thr}$, $H\alpha_{Ala1}$, $H\alpha_{Pro7}$), 4.09 – 4.17 (m, 1H, $H\beta_{Thr}$), 3.91 (s, 2H, $2H\alpha_{Gly}$), 3.50 – 3.84 (m, 6H, $2H\beta_{Ser}$, $2H\delta_{Pro2}$, $2H\delta_{Pro7}$), 2.15 – 2.32 (m, 2H, $H\beta_{Pro2}$, $H\beta_{Pro7}$), 1.79 – 2.05 (m, 6H, $H\beta_{Pro2}$, $H\beta_{Pro7}$, $2H\gamma_{Pro2}$, $2H\gamma_{Pro7}$), 1.43 (d, 3H, *J* = 7.0 Hz, CH_{3Ala1}), 1.28 (d, 3H, *J* = 7.1 Hz, CH_{3Ala6}), 1.11 (d, 3H, *J* = 6.4 Hz, CH_{3Thr}).

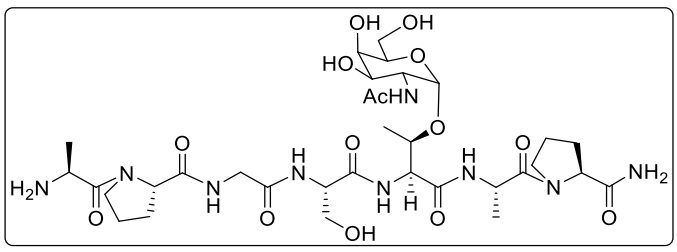
^{1}H NMR (400 MHz, D_2O/H_2O, 1:9, amide region) δ (ppm): 8.49 (t, *J* = 5.8 Hz, NH_{Gly}), 8.25 (d, *J* = 5.8 Hz, NH_{Ala6}), 8.19 (d, *J* = 7.9 Hz, NH_{Thr}), 8.13 (d, *J* = 6.8 Hz, NH_{Ser}), 7.53 (s, NH_{term}), 6.91 (s, NH_{term}).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 176.9, 174.5, 172.8, 172.0, 171.4, 171.2, 169.3 (CON), 67.0 ($C\beta_{Thr}$), 61.1 ($C\beta_{Ser}$), 60.7, 60.3 ($C\alpha_{Pro2}$, $C\alpha_{Pro7}$), 58.8 ($C\alpha_{Thr}$), 55.4 ($C\alpha_{Ser}$), 48.0 ($C\alpha_{Ala1}$), 47.9, 47.8 ($C\delta_{Pro7}$, $C\delta_{Pro2}$), 47.7 ($C\alpha_{Ala6}$), 42.4 ($C\alpha_{Gly}$), 29.3, 29.5 ($C\beta_{Pro2}$, $C\beta_{Pro7}$), 24.7, 24.6 ($C\gamma_{Pro2}$, $C\gamma_{Pro7}$), 18.8 (CH_{3Thr}), 15.3, 14.9 (CH_{3Ala1}, CH_{3Ala6}).

Semi-preparative HPLC: Rt = 11.0 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (7:93) → (13:87), 12.0 min, 10 mL/min, λ = 212 nm).

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{25}H_{43}N_8O_9$: 599.3147; Found: 599.3146.

Synthesis of glycopeptide 2 ($A_1P_2GST^*A_6P_7$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **2** was obtained and purified by semi-preparative HPLC.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.90 (d, 1H, *J* = 3.8 Hz, H_{1S}), 4.53 – 4.67 (m, 3H, $H\alpha_{Ser}$, $H\alpha_{Ala6}$, $H\alpha_{Thr}$), 4.44 – 4.52 (m, 1H, $H\alpha_{Pro2}$), 4.31 – 4.40 (m, 3H, $H\alpha_{Pro7}$, $H\alpha_{Ala1}$, $H\beta_{Thr}$), 4.08 (dd, 1H, *J* = 11.0, 3.8 Hz, H_{2S}), 3.82 – 4.03 (m, 7H, $2H\beta_{Ser}$, $2H\alpha_{Gly}$, H_{3S}, H_{4S}, H_{5S}), 3.58 – 3.79 (m, 6H, $2H\delta_{Pro7}$, $2H\delta_{Pro2}$, $2H_{6S}$), 2.25 – 2.38 (m, 2H, $H\beta_{Pro2}$, $H\beta_{Pro7}$), 1.90 – 2.12 (m, 9H, $H\beta_{Pro2}$, $H\beta_{Pro7}$, $2H\gamma_{Pro2}$, $2H\gamma_{Pro7}$, $NHCOCH_3$), 1.52 (d, 3H, *J* = 7.0, CH_{3Ala1}), 1.35 (d, 3H, *J* = 7.0 Hz, CH_{3Ala6}), 1.25 (d, 3H, *J* = 6.4 Hz, CH_{3Thr}).

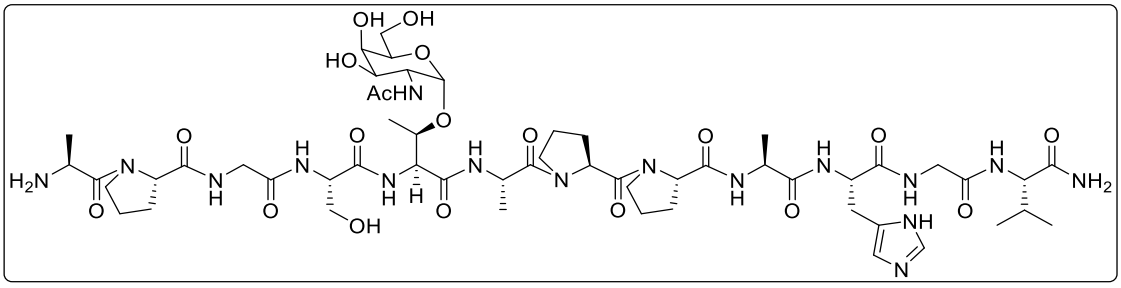
^{1}H NMR (400 MHz, D_2O/H_2O, 1:9, amide region) δ (ppm): 8.56 (d, *J* = 9.0 Hz, NH_{Thr}), 8.48 (t, *J* = 5.8 Hz, NH_{Gly}), 8.23 (d, *J* = 6.1 Hz, NH_{Ala6}), 8.08 (d, *J* = 6.8 Hz, NH_{Ser}), 7.78 (d, *J* = 9.6 Hz, $NHAc_S$), 7.55 (s, NH_{term}), 6.87 (s, NH_{term}).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 176.9, 174.5, 173.7, 172.5, 172.2, 171.2, 170.7, 169.3 (CON), 98.4 (C_{1S}), 75.1 ($C\beta_{Thr}$), 71.3, 68.5, 68.0 (C_{4S}, C_{5S}, C_{3S}), 61.3 (C_{6S}), 61.2 ($C\beta_{Ser}$), 60.7 ($C\alpha_{Pro2}$), 60.1 ($C\alpha_{Pro7}$), 57.2 ($C\alpha_{Thr}$), 55.1 ($C\alpha_{Ser}$), 49.6 (C_{2S}), 48.0 ($C\alpha_{Ala1}$), 47.7, 47.6 ($C\delta_{Pro7}$, $C\delta_{Pro2}$), 47.5 ($C\alpha_{Ala6}$), 42.3 ($C\alpha_{Gly}$), 29.5, 29.3 ($C\beta_{Pro2}$, $C\beta_{Pro7}$), 24.7, 24.6 ($C\gamma_{Pro2}$, $C\gamma_{Pro7}$), 22.2 ($NHCOCH_3$), 18.2 (CH_{3Thr}), 15.4 (CH_{3Ala6}), 15.0 (CH_{3Ala1}).

Semi-preparative HPLC: Rt = 10.8 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (7:93) → (13:87), 13.3 min, 10 mL/min, λ = 212 nm).

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{33}H_{56}N_9O_{14}$: 802.3941; Found: 802.3947.

Synthesis of glycopeptide 3 ($A_1P_2GST^*A_6P_7P_8A_9HGV$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **3** was obtained and purified by semi-preparative HPLC.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 8.63 (s, 1H, Hε1$_{His}$), 7.34 (s, 1H, Hδ2$_{His}$), 4.95 (d, 1H, *J* = 3.9 Hz, H$_{1S}$), 4.64 – 4.74 (m, 3H, Hα$_{Ser}$, Hα$_{Pro8}$, Hα$_{His}$), 4.54 – 4.63 (m, 2H, Hα$_{Ala6}$, Hα$_{Thr}$), 4.52 (t, 1H, Hα$_{Pro2}$), 4.33 – 4.43 (m, 3H, Hα$_{Pro7}$, Hα$_{Ala1}$, Hβ$_{Thr}$), 4.25 (q, 1H, *J* = 7.2 Hz, Hα$_{Ala9}$), 4.15 (d, 1H, *J* = 6.8 Hz, Hα$_{Val}$), 4.11 (dd, 1H, *J* = 11.0, 3.8 Hz, H$_{2S}$), 3.63 – 4.05 (m, 17H, 2Hδ$_{Pro7}$, 2Hδ$_{Pro2}$, 2Hδ$_{Pro9,}$ 2H$_{6S}$, 2Hβ$_{Ser}$, 2Hα$_{Gly3}$, 2Hα$_{Gly11}$, H$_{3S}$, H$_{4S}$, H$_{5S}$), 3.19 – 3.34 (m, 2H, 2Hβ$_{His}$), 2.25 – 2.42 (m, 3H, Hβ$_{Pro2}$, Hβ$_{Pro7}$, Hβ$_{Pro8}$), 1.85 – 2.16 (m, 13, Hβ$_{Pro2}$, Hβ$_{Pro7}$, Hβ$_{Pro8}$, 2Hγ$_{Pro2}$, 2Hγ$_{Pro7}$, 2Hγ$_{Pro8}$, NHCOC<u>H</u>$_3$, Hβ$_{Val}$), 1.55 (d, 3H, *J* = 7.0 Hz, CH$_{3Ala1}$), 1.32 – 1.40 (m, 6H, CH$_{3Ala6}$, CH$_{3Ala9}$), 1.28 (d, 3H, *J* = 6.5 Hz, CH$_{3Thr}$), 0.93 – 1.03 (m, 6H, 2CH$_{3Val}$).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 176.1, 174.9, 174.4, 174.0, 173.6 , 172.1, 172.0, 172.0, 172.0, 171.2, 171.1, 170.7, 169.3 (CON), 163.2, 162.8, 133.6 (Cδ2$_{His}$, Cε1$_{His}$, Cγ$_{His}$), 98.7 (C$_{1S}$), 75.8 (Cβ$_{Thr}$), 71.3, 68.6, 68.2 (C$_{4S}$, C$_{5S}$, C$_{3S}$), 61.4 (C$_{6S}$), 61.3 (Cβ$_{Ser}$), 60.7 (Cα$_{Pro2}$), 60.1 (Cα$_{Pro7}$), 59.2 (Cα$_{Val}$), 58.4 (Cα$_{Pro8}$), 57.1 (Cα$_{Thr}$), 55.2 (Cα$_{Ser}$), 52.4 (Cα$_{His}$), 49.7 (C$_{2S}$), 49.6 (Cα$_{Ala9}$), 48.1 (Cα$_{Ala1}$), 47.8, 47.8, 47.7 (Cδ$_{Pro7}$, Cδ$_{Pro2}$, Cδ$_{Pro8}$), 47.5 (Cα$_{Ala6}$), 42.3 (Cα$_{Gly}$), 29.8 (Cβ$_{Val}$), 29.3, 29.3, 28.0 (Cβ$_{Pro2}$, Cβ$_{Pro7}$, Cβ$_{Pro9}$), 26.4 (Cβ$_{His}$), 24.8, 24.7, 24.7 (Cγ$_{Pro2}$, Cγ$_{Pro7}$, Cγ$_{Pro8}$), 22.3 (NHCO<u>C</u>H$_3$), 18.3 (CH$_{3Thr}$), 17.3 (CH$_{3Val}$), 16.4, 15.3 (CH$_{3Ala6}$, CH$_{3Ala9}$), 15.0 (CH$_{3Ala1}$).

Semi-Preparative HPLC: Rt = 13.0 min (Phenomenex Luna C18 (2), 10 µm, 21.2x250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (12.5:87.5), 15 min, 20 mL/min, λ = 212 nm)

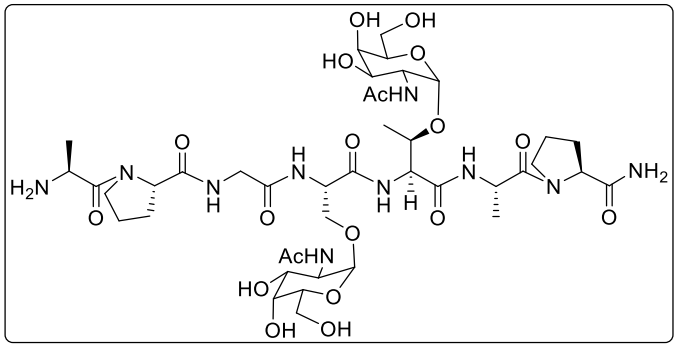
Analytical HPLC: Rt = 10.6 min (Phenomenex Luna C18 (2), 5 µm, 4.6x250mm, Grad: acetonitrile/water+0.1% TFA (7:93) → (13:87), 12 min, 1 mL/min, λ = 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{54}H_{87}N_{16}O_{19}$: 1263.6328; Found 1263.6303.

Synthesis of glycopeptide 4 (APDT*RP)

Synthesis and characterization of glycopeptide **4** has been previously described.[7]

Synthesis of glycopeptide 5 ($A_1P_2GS^*T^*A_6P_7$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **5** was obtained and purified by semi-preparative HPLC.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.95 (d, 1H, *J* = 3.8 Hz, H_{1SSer}), 4.90 (d, 1H, *J* = 3.8 Hz, H_{1SThr}), 4.83 (t, 1H, *J* = 4.7 Hz, $H\alpha_{Ser}$), 4.57 – 4.63 (m, 2H, $H\alpha_{Ala6}$, $H\alpha_{Thr}$), 4.49 – 4.57 (m, 1H, $H\alpha_{Pro2}$), 4.28 – 4.42 (m, 3H, $H\alpha_{Pro7}$, $H\alpha_{Ala1}$, $H\beta_{Thr}$), 3.85 – 4.16 (m, 12H, $2H\beta_{Ser}$, $2H\alpha_{Gly}$, H_{3SThr}, H_{4STHr}, H_{5STHr}, H_{2STHr}, H_{3SSer}, H_{4SSer}, H_{5SSer}, H_{2SSer}), 3.62 – 3.80 (m, 8H, $2H\delta_{Pro7}$, $2H\delta_{Pro2}$, $2H_{6STHr}$, $2H_{6SSer}$), 2.28 – 2.41 (m, 2H, $H\beta_{Pro2}$, $H\beta_{Pro7}$), 1.93 – 2.15 (m, 12H, $H\beta_{Pro2}$, $H\beta_{Pro7}$, $2H\gamma_{Pro2}$, $2H\gamma_{Pro7}$, NHCOC$\underline{H}_{3STHr}$, NHCOC$\underline{H}_{3SSer}$), 1.55 (d, 3H, *J* = 7.0 Hz, CH_{3Ala1}), 1.39 (d, 3H, *J* = 7.2 Hz, CH_{3Ala6}), 1.28 (d, 3H, *J* = 6.4 Hz, CH_{3Thr}).

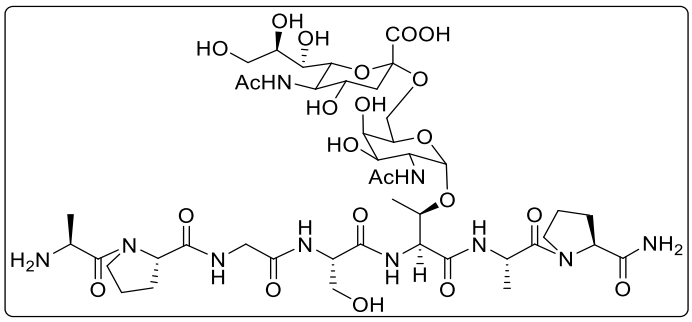
^{13}C NMR (100 MHz, D_2O) δ (ppm): 176.9, 174.3, 174.1, 173.7, 172.4, 171.4, 171.0, 170.1, 169.3 (CON), 98.6 (C_{1SSer}), 98.1 (C_{1SThr}), 76.1 ($C\beta_{Thr}$), 71.3 ($C\beta_{Ser}$), 71.3, 68.6, 68.4, 68.2, 67.8, 67.7 (C_{4STHr}, C_{5STHr}, C_{3STHr}, C_{4SSer}, C_{5SSer}, C_{3SSer}), 61.3, 61.2 (C_{6STHr}, C_{6SSer}), 60.5 ($C\alpha_{Pro2}$), 60.0 ($C\alpha_{Pro7}$), 56.9 ($C\alpha_{Thr}$), 53.4 ($C\alpha_{Ser}$), 49.7, 49.6 (C_{2STHr}, C_{2SSer}), 48.0 ($C\alpha_{Ala1}$), 47.7, 47.6 ($C\delta_{Pro7}$, $C\delta_{Pro2}$), 47.3 ($C\alpha_{Ala6}$), 42.2 ($C\alpha_{Gly}$), 29.5, 29.4 ($C\beta_{Pro2}$, $C\beta_{Pro7}$), 24.7, 24.6 ($C\gamma_{Pro2}$, $C\gamma_{Pro7}$), 22.4, 22.1 (NHCO$\underline{C}H_{3STHr,}$ NHCO$\underline{C}H_{3SSer}$), 18.3 (CH_{3Thr}), 15.6 (CH_{3Ala6}), 15.0 (CH_{3Ala1}).

Semi-Preparative HPLC: Rt = 16.0 min (Phenomenex Luna C18 (2), 10 µm, 21.2x250mm, Grad: acetonitrile/water+0.1% TFA (0:100) → (10:90), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 5.3 min (Phenomenex Luna C18 (2), 5µm, 4.6x250mm, Grad: acetonitrile/water+0.1% TFA (7:93) → (13:87), 12 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{41}H_{69}N_{10}O_{19}$: 1005.4735; Found 1005.4737.

Synthesis of glycopeptide 6 ($A_1P_2GST^{**}A_6P_7$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **6** was obtained and purified by semi-preparative HPLC.

^{1}H NMR (400 MHz, D_2O) δ: 4.93 (d, 1H, *J* = 4.0 Hz, H_{1S}), 4.68 (t, 1H, *J* = 6.0 Hz, $H\alpha_{Ser}$), 4.57 – 4.63 (m, 2H, $H\alpha_{Thr}$, $H\alpha_{Ala6}$), 4.52 (t, 1H, *J* = 7.5, $H\alpha_{Pro2}$), 4.33 – 4.43 (m, 3H, $H\beta_{Thr}$, $H\alpha_{Ala1}$, $H\alpha_{Pro7}$), 3.55 – 4.15 (m, 21H, $2H\delta_{Pro2}$, $2H\delta_{Pro7}$, $2H\beta_{Ser}$, $2H\alpha_{Gly}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H_{6S}$, H_{4S}', H_{5S}', H_{6S}', H_{7S}', H_{8S}', $2H_{9S}$'), 2.71 (dd, 1H, *J* = 12.7, 4.7 Hz, H_{S3}'eq), 2.28 – 2.42 (m, 2H, $H\beta_{Pro2}$, $H\beta_{Pro7}$), 1.92 – 2.16 (m, 12H, NHCOC$\underline{H}_3$', NHCOC$\underline{H}_3$, $H\beta_{Pro2}$, $H\beta_{Pro7}$, $2H\gamma_{Pro2}$, $2H\gamma_{Pro7}$), 1.75 (t, 1H, *J* = 12.2 Hz, H_{S3}'ax), 1.55 (d, 3H, *J* = 7.0 Hz, CH_{3Ala1}), 1.38 (d, 3H, *J* = 7.0 Hz, CH_{3Ala6}), 1.30 (d, 3H, *J* = 6.6 Hz, CH_{3Thr}).

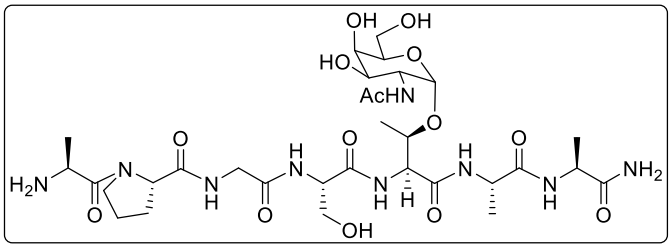
^{1}H NMR (400 MHz, D_2O/H_2O, 1:9, amide region) δ (ppm): 8.63 (d, *J* = 9.2 Hz, NH_{Thr}), 8.57 (t, *J* = 5.8 Hz, NH_{Gly}), 8.35 (d, *J* = 5.9 Hz, NH_{Ala6}), 8.20 (d, *J* = 6.9 Hz, NH_{Ser}), 8.05 (d, *J* = 9.2 Hz, $NHAc_S$'), 7.82 (d, *J* = 9.7 Hz, $NHAc_S$), 7.66 (s, NH_{term}), 6.98 (s, NH_{term}).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 176.9 (COO), 174.9, 174.4, 173.7, 172.5, 172.2, 172.1, 171.2, 170.6, 169.3 (CON), 126.2 (C_{2S}'), 98.7 (C_{1S}), 75.8 ($C\beta_{Thr}$), 72.7, 71.3, 70.8, 69.9, 68.6, 68.3, 68.0, 67.8, 62.8 (C_{3S}, C_{4S}, C_{5S}, C_{6S}, C_{4S}', C_{6S}', C_{7S}', C_{8S}', C_{9S}'), 61.2 ($C\alpha_{Ser}$), 60.7 ($C\alpha_{Pro2}$), 60.0 ($C\alpha_{Pro7}$), 57.2 ($C\alpha_{Thr}$), 55.1 ($C\alpha_{Ser}$), 51.8 (C_{S5}'), 49.6 (C_{2S}), 48.0 ($C\alpha_{Ala1}$), 47.7, 47.6 ($C\delta_{Pro7}$, $C\delta_{Pro2}$), 47.5 ($C\alpha_{Ala6}$), 42.4 ($C\alpha_{Gly}$), 39.6 (C_{S3}'), 29.5, 29.3 ($C\beta_{Pro2}$, $C\beta_{Pro7}$), 24.7, 24.6 ($C\gamma_{Pro2}$, $C\gamma_{Pro7}$), 22.2, 22.0 (NHCO$\underline{C}H_3$', NHCO$\underline{C}H_3$), 18.2 (CH_{3Thr}), 15.4 (CH_{3Ala6}), 15.0 (CH_{3Ala1}).

Semi-preparative HPLC: Rt = 13.8 min (Phenomenex Luna C18 (2), 10 μm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (13:87), 16.0 min, 10 mL/min, λ = 212 nm).

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{44}H_{73}N_{10}O_{22}$: 1093.4895; Found: 1093.4876.

Synthesis of glycopeptide 7 ($A_1PGST^*A_6A_7$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **7** was obtained and purified by semi-preparative HPLC.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.94 (d, 1H, *J* = 3.9 Hz, H_{1S}), 4.67 (t, 1H, *J* = 6.0 Hz, $H\alpha_{Ser}$), 4.58 (d, 1H, *J* = 2.2 Hz, $H\alpha_{Thr}$), 4.48 – 4.55 (m, 1H, $H\alpha_{Pro}$), 4.31– 4.43 (m, 3H, $H\alpha_{Ala7}$, $H\alpha_{Ala1}$, $H\beta_{Thr}$), 4.22 (q, 1H, *J* = 7.2 Hz, $H\alpha_{Ala6}$), 4.12 (dd, 1H, *J* = 11.0, 3.8 Hz, H_{2S}), 3.85 – 4.07 (m, 7H, $2H\beta_{Ser}$, $2H\alpha_{Gly}$, H_{3S}, H_{4S}, H_{5S}), 3.61 – 3.79 (m, 4H, $2H\delta_{Pro}$, $2H_{6S}$), 2.32 – 2.42 (m, 1H, $H\beta_{Pro}$), 1.94 – 2.15 (m, 6H, $H\beta_{Pro}$, $2H\gamma_{Pro}$, $NHCOCH_3$), 1.55 (d, 3H, *J* = 7.0, CH_{3Ala1}), 1.37-1.43 (m, 6H, CH_{3Ala6}, CH_{3Ala7}), 1.29 (d, 3H, *J* = 6.4 Hz, CH_{3Thr}).

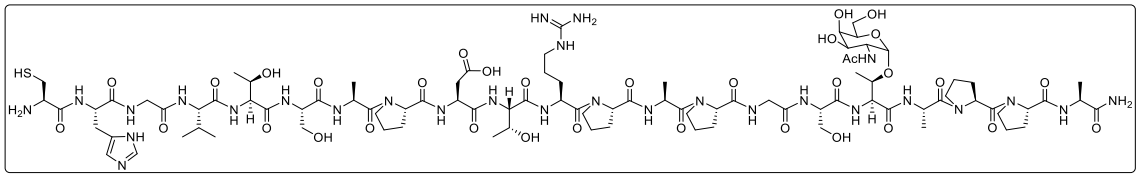
^{13}C NMR (100 MHz, D_2O) δ (ppm): 177.7, 174.5, 174.3, 174.0, 172.3, 171.2, 170.9, 169.4 (CON), 98.6 (C_{1S}), 75.4 ($C\beta_{Thr}$), 71.4, 68.6, 68.2 (C_{4S}, C_{5S}, C_{3S}), 61.3 (C_{6S}), 61.2 ($C\beta_{Ser}$), 60.7 ($C\alpha_{Pro}$), 57.4 ($C\alpha_{Thr}$), 55.1 ($C\alpha_{Ser}$), 49.7 (C_{2S}), 49.3 ($C\alpha_{Ala1}$), 49.2 ($C\delta_{Pro}$), 48.1, 47.7 ($C\alpha_{Ala6}$, $C\alpha_{Ala7}$), 42.4 ($C\alpha_{Gly}$), 29.3 ($C\beta_{Pro}$), 24.7 ($C\gamma_{Pro}$), 22.2 ($NHCOCH_3$), 18.2 (CH_{3Thr}), 16.9, 16.6 (CH_{3Ala6}, CH_{3Ala7}), 15.0 (CH_{3Ala1}).

Semi-preparative HPLC: Rt = 13.7 min (Phenomenex Luna C18 (2), 10 µm, 21.2x250mm, Grad: acetonitrile/water+0.1% TFA (0:100) → (10:90), 20 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 4.1 min (Phenomenex Luna C18 (2), 5µm, 4.6x250mm, Grad: acetonitrile/water+0.1% TFA (7:93) → (13:87), 12 min, 1 mL/min, λ = 212 nm).

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{31}H_{54}N_9O_{14}$: 776.3785; Found 776.3799.

Synthesis of glycopeptide 8 (CHGVTSAPDTRPAPGST*APPA)



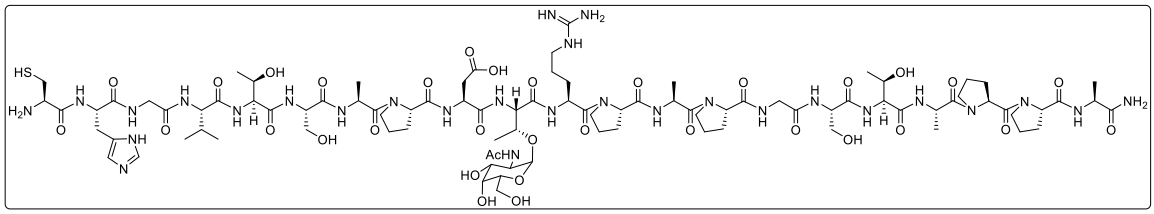
Following SPPS methodology with the adequately protected amino acids, glycopeptide **8** was obtained and purified by semi-preparative HPLC.

Semi-preparative HPLC: Rt = 19.8 min (Phenomenex Luna C18 (2), 10 µm, 21.2x250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 21.6 min (Phenomenex Luna C18 (2), 5 µm, 4.6x250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ = 212 nm).

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{91}H_{148}N_{28}O_{33}S$: 1096.5237; Found 1096.5240.

Synthesis of glycopeptide 9 (CHGVTSAPDT*RPAPGSTAPPA)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **9** was obtained and purified by semi-preparative HPLC.

Semi-preparative HPLC: Rt = 15.6 min (Phenomenex Luna C18 (2), 10 µm, 21.2x250mm, Grad: acetonitrile/water+0.1% TFA (10:90) → (15:83), 20 min, 20 mL/min, λ= 212 nm).

Analytical RP-HPLC: Rt = 8.3 min (Phenomenex Luna C18 (2), 5 µm, 4.6x250mm, Grad: acetonitrile/water+0.1% TFA (10:90) → (17:83), 17 min, 1 mL/min, λ = 212 nm).

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{91}H_{148}N_{28}O_{33}S$: 1096.5237; Found 1096.5197.

4.5.2. Expression and purification of scFv-5E5

Expression and purification of scFv-5E5 were carried out by the group of R. Hurtado-Guerrero (Instituto de Biocomputación y Fisica de Sistemas Complejos and the University of Zaragoza).

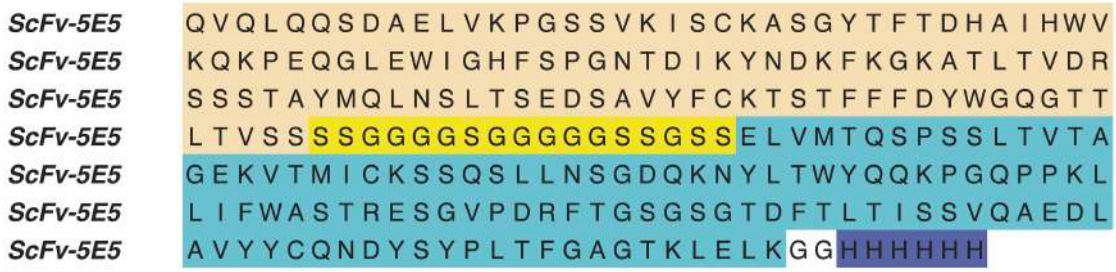


Figure 4.14. Sequence of scFv-5E5. The heavy chain is colored in brown and the light chain in blue. The linker and the histidine tag are indicated in yellow and dark blue, respectively.

4.5.3. Isothermal titration microcalorimetry (ITC)

ITC were carried out by the group of R. Hurtado-Guerrero to characterize the interaction of ScFv-5E5 and its corresponding mutants with different ligands. All experiments were carried out in an Auto-iTC200 (Microcal, GE Healthcare) at 25ºC with scfv-5E5 and its mutants at 20-30 μM and ligands ranging from 300 μM to 2 mM in 25 mM TRIS pH 7.5 150 mM NaCl. The experiments were performed in duplicate. Data integration, correction and analysis were carried out in Origin 7 (Microcal). The data were fitted to a one-site equilibrium-binding model.

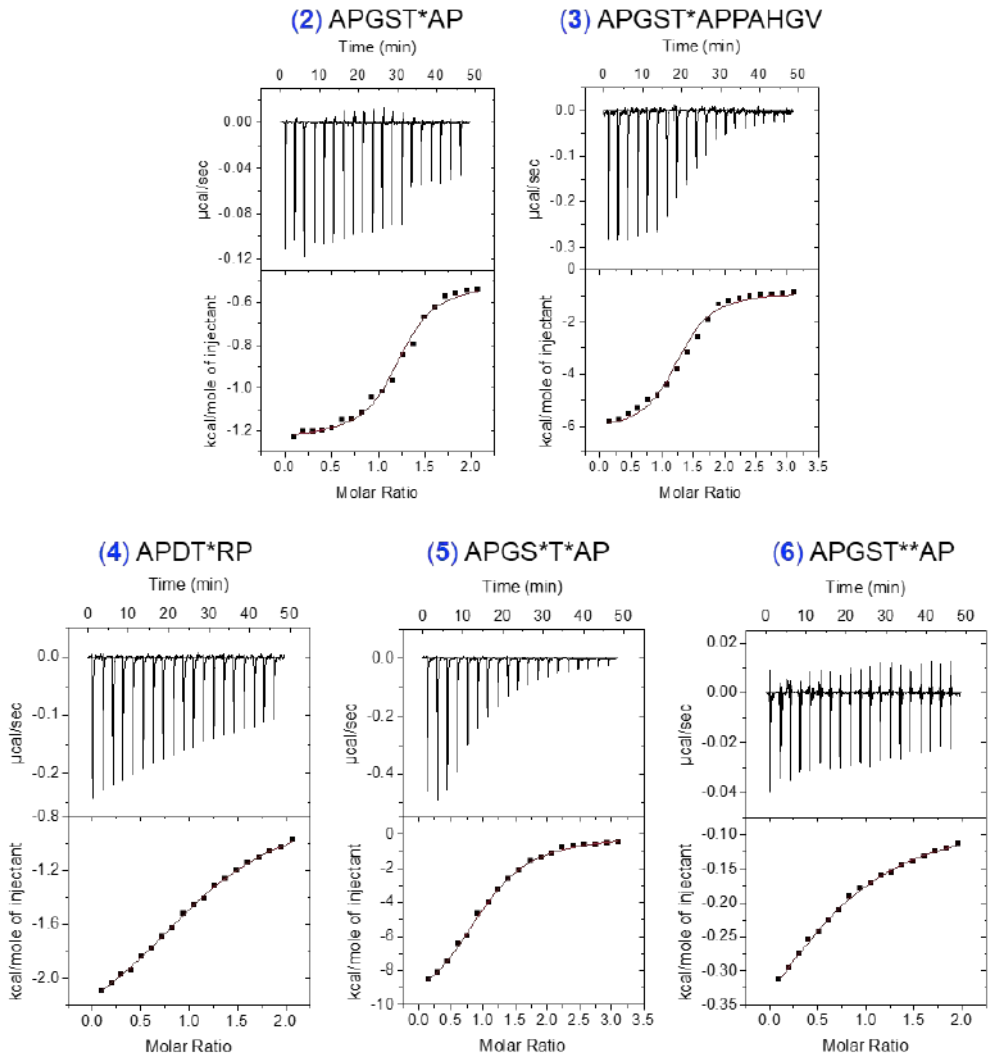


Figure 4.15. ITC profiles of glycopeptides **2-6** with scFv-5E5 at 25 °C and pH 7.5. The solid line represents the least-squares fitting of the data to the simplest model (one binding site). The interaction of glycopeptide **4** with scFv-5E5 is weak and consequently the K_D value derived from these assays should be considered in a qualitative manner ($K_D \geq 19$ µM). T* = GalNAc-α-1-*O*-Thr (Tn antigen) and T** = Neu5Ac-α-(2-6)-GalNAc-α-1-*O*-Thr (STn antigen).

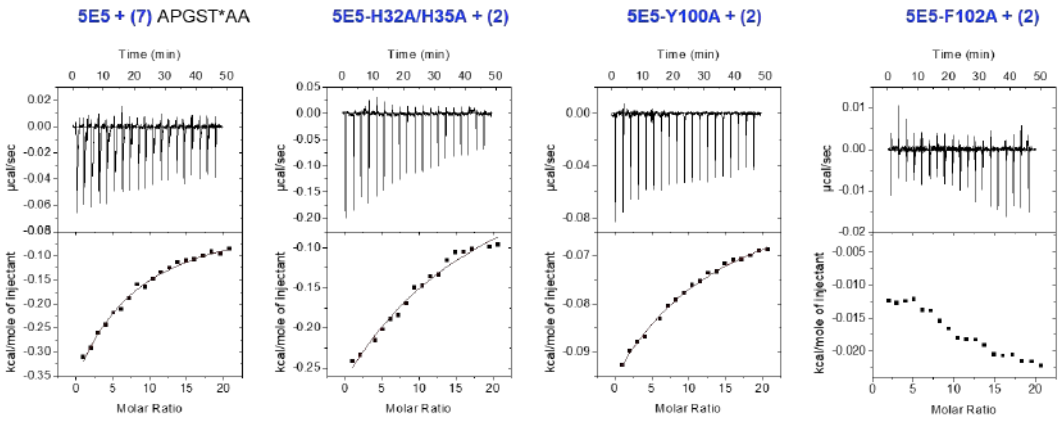


Figure 4.16. ITC profiles of glycopeptide **7** with scFv-5E5 and glycopeptide **2** with several 5E5 mutants at 25 °C and pH 7.5. The solid line represents the least-squares fitting of the data to the simplest model (one binding site). These antibody-ligand interactions shown here are weak and consequently the K_D values derived from these assays should be considered in a qualitative manner

4.5.4. MicroScale Thermophoresis (MST)

Binding assays of antibody 5E5 or scFv-5E5 to the peptides were studied by the group of R. Hurtado-Guerrero with MST using a Monolith NT.115 Pico instrument (NanoTemper Technologies). Prior to the binding experiments, the enzymes were labeled with reactive dyes using *N*-hydroxysuccinimide (NHS)–ester chemistry, which reacts efficiently with the primary amines of proteins to form highly stable dye–protein conjugates (Monolith NTTM Protein Labeling kit RED – NHS 2nd generation, NanoTemper Technologies). For protein labeling, the proteins concentration was adjusted to 20 µM using the labeling buffer (supplied by the kit). Lower concentrations may result in loss of coupling efficiency. The solid fluorescent dye was dissolved in 100% DMSO at a concentration of about 600 µM and mixed thoroughly. Before mixing the proteins and the dye, the concentration of the dye was adjusted to 3-fold concentration of the protein using the labeling buffer. Then, the protein and the fluorescent dye solutions were mixed in 1:1 ratio and incubated for 30 min at room temperature in the dark. Unreacted 'free' dye was eliminated by spin desalting columns (Zeba™ Spin Desalting Columns, ThermoFisher). All experiments were conducted in binding buffer (50 mM HEPES pH 7.5, 50 mM NaCl) using a constant concentration of the two proteins (5 nM) and serial dilutions of each ligand spanning from 10 mM to 150 nM of the glycopeptides. Duplicate measurements were performed in

premium capillaries at 25ºC using 40% laser power and 20% LED power in order to obtain optimized results. Dissociation constants (K_D) were calculated using MO Affinity analysis.

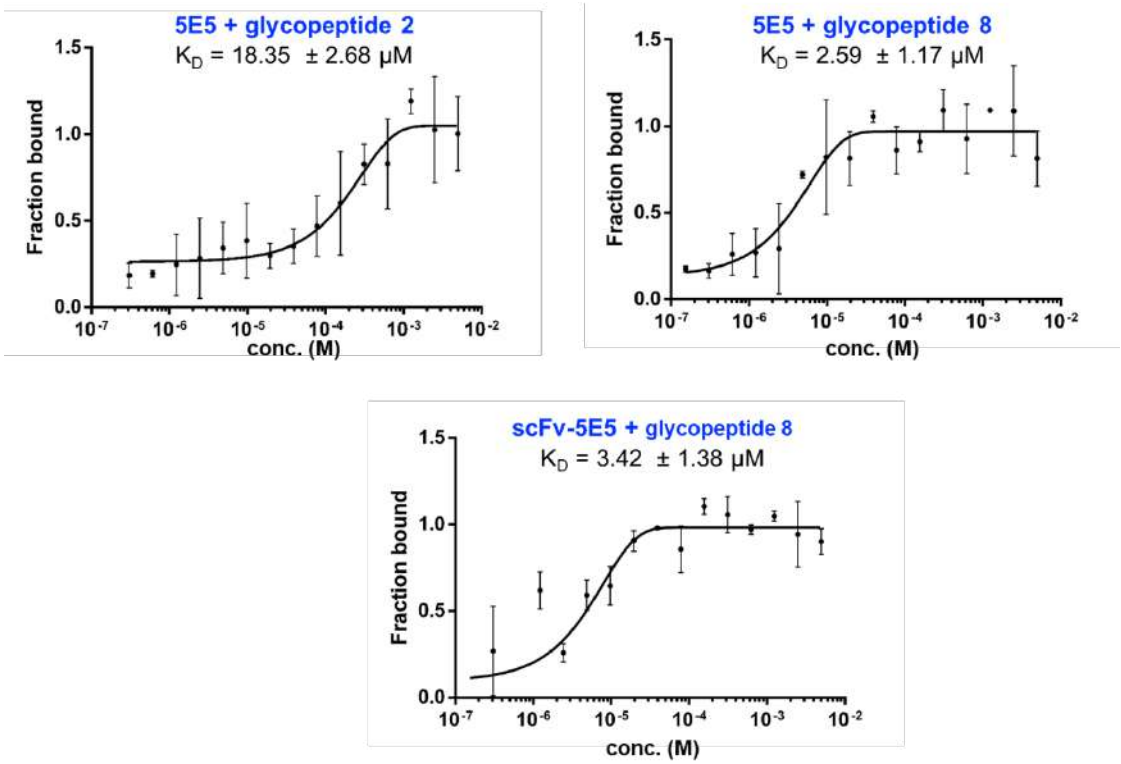


Figure 4.17. MST binding curve used to determine K_D for 5E5 or scFv-5E5 with different glycopeptides. Error represents SD between two separate experiments; K_D values ± SD are shown.

4.5.5. Crystallization of complex scFv-5E5/2 and data collection

Crystal structure was successfully obtained by the group of R. Hurtado-Guerrero. Crystals of the ScFv-5E5 were grown by sitting drop experiments at 18 ºC by mixing 0.4 µl of protein solution (6.7 mg/mL ScFv-5E5 and 5 mM glycopeptide **2** in buffer C) with an equal volume of a reservoir solution (0.2 M ammonium chloride, 0.1M HEPES 7.5, 25% v/v glycerol ethoxylate). The crystals were cryoprotected in mother liquor containing 20% glycerol and flash frozen in liquid nitrogen.

4.5.6. MD simulations of glycopeptides in complex to scFv-5E5

The simulations were carried out by F. Corzana with AMBER 18 package[27] implemented with ff14SB[28] and GLYCAM06j[29] force fields. The X-ray structure of complex **scFv-5E5/2** was used as the starting structure. This structure was conveniently modified with PyMOL to obtain the rest of the complexes. In all cases, the complex was immersed in a water box with a 10 Å buffer of TIP3P water molecules.[30] The system was neutralized by adding explicit counter ions (Cl^-). A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then gently heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 30 kcal·mol^{-1} were applied to the solute, and the Andersen temperature-coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method.[31] An 8 Å cut-off was applied to Lennard-Jones interactions. Each system was equilibrated for 2 ns with a 2-fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for additional 0.5 μs under the same simulation conditions.

4.5.7. Nanoparticle-based dot-blot assay

Synthesis of nanoparticles and dot-blot assays were performed by the group of R. Fiammengo (University of Verona). The assay exploits gold nanoparticles (AuNPs) functionalized with two synthetic glycopeptides (**8** and **9**) whose sequences correspond to one complete tandem repeat of the VNTR domain of MUC1. The peptides were coupled to PEG-AuNPs accordingly to a previously reported procedure.[3] AuNP-glycopeptide **9**, AuNP-glycopeptide **8**, ctrl AuNPs, and a 1:1 mixture of AuNP- glycopeptide **9**, AuNP- glycopeptide **8** were spotted in triplicate onto methanol-activated PVDF membranes (1.5 μL, 4 nM AuNPs) and allowed to dry at room temperature. Each membrane was blocked with 5% BSA in tris-buffered saline (TBS), washed in TBS with 0.1% Tween 20 (3×10 min) and incubated for 3 h at room temperature with patient serum (1:100 dilution in TBS with 1% BSA). After washing with TBS containing 0.2% Tween 20 (4×10 min),

the membranes were incubated with goat anti-human IgG (H+L) cross-adsorbed Alexa Fluor 647 secondary antibody (A-21445, Invitrogen), diluted 1:5000 in TBS for 1 h at room temperature. Finally, membranes were washed in TBS with 0.1% Tween 20 (3×10 min) and visualized using a Typhoon Trio Variable Mode Imager System (GE Healthcare). The intensity of the spots was quantified via densitometry with ImageQuant TL software (GE Healthcare), subtracting the background. The data are reported as signal/background intensity to account for the variations in background observed for different membranes. All sera from healthy patients did not give any detectable signal. The limit of detection (LoD) of the assay was estimated considering the variance of the signal intensity/background ratio for 10 samples with low signal intensity/background (<0.020) on independent membranes. On each membrane, it was deposited 3 dots for each AuNP type (a triplicate). Signal/background ratios were calculated considering the average signal intensity from the three dots measured via densitometry divided by the average background intensity around the three dots. From these 10 samples: average signal intensity/background = 0.0084, standard deviation = 0.0040, s.e.m. = 0.0013. LoD = 3 × s.e.m. = 0.0039; LoQ = 3.3 × LoD = 0.013.

Table 4.3. Physicochemical characterization of AuNPs used in this work.

	DLS[a]		ζ-Potential[a]	$N_{peptide}$/AuNP[c]
AuNP	Ø(nm)[b]	PDI	(mV)	
AuNP-glycopeptide **9**	35 ± 1	0.199	- 37 ± 1	~150
AuNP-glycopeptide **8**	34 ± 1	0.170	- 37 ± 2	~150
ctrl AuNPs	28 ± 1	0.179	- 39 ± 2	--

[a] Measurements in 10 mM $NaHCO_3$. [b] Derived from intensity distributions. [c] Amino acid analysis.

4.6. References

(1) Movahedin, M.; Brooks, T. M.; Supekar, N. T.; Gokanapudi, N.; Boons, G.-J.; Brooks, C. L. *Glycobiology* **2017**, *27*, 677–687.

(2) Martínez-Sáez, N.; Peregrina, J. M.; Corzana, F. *Chem. Soc. Rev.* **2017**, *46*, 7154–7175.

(3) Compañón, I.; Guerreiro, A.; Mangini, V.; Castro-López, J.; Escudero-Casao, M.; Avenoza, A.; Busto, J. H.; Castillo, S.; Asensio, J. L.; Jiménez-Osés, G.; Boutureira, O.; Peregrina, J. M.; Hurtado-Guerrero, R.; Fiammengo, R.; Bernardes, G. J. L.; Corzana, F. *J. Am. Chem. Soc.* **2019**, *141*, 4063–4072.

(4) Karsten, U.; Serttas, N.; Paulsen, H.; Danielczyk, A.; Goletz, S. *Glycobiology* **2004**, *14*, 681–692.

(5) Yoshimura, Y.; Denda-Nagai, K.; Takahashi, Y.; Nagashima, I.; Shimizu, H.; Kishimoto, T.; Noji, M.; Shichino, S.; Chiba, Y.; Irimura, T. *Sci. Rep.* **2019**, *9*, 16641.

(6) Tarp, M. A.; Sørensen, A. L.; Mandel, U.; Paulsen, H.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. *Glycobiology* **2007**, *17*, 197–209.

(7) Martínez-Sáez, N.; Castro-López, J.; Valero-González, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem. Int. Ed.* **2015**, *54*, 9830–9834.

(8) Wakui, H.; Tanaka, Y.; Ose, T.; Matsumoto, I.; Kato, K.; Min, Y.; Tachibana, T.; Sato, M.; Naruchi, K.; Martin, F. G.; Hinou, H.; Nishimura, S. I. *Chem. Sci.* **2020**, *11*, 4999–5006.

(9) Sørensen, A. L.; Reis, C. A.; Tarp, M. A.; Mandel, U.; Ramachandran, K.; Sankaranarayanan, V.; Schwientek, T.; Graham, R.; Taylor-Papadimitriou, J.; Hollingsworth, M. A.; Burchell, J.; Clausen, H. *Glycobiology* **2006**, *16*, 96–107.

(10) Posey, A. D.; Schwab, R. D.; Boesteanu, A. C.; Steentoft, C.; Mandel, U.; Engels, B.; Stone, J. D.; Madsen, T. D.; Schreiber, K.; Haines, K. M.; Cogdill, A. P.; Chen, T. J.; Song, D.; Scholler, J.; Kranz, D. M.; Feldman, M. D.; Young, R.;

Keith, B.; Schreiber, H.; Clausen, H.; Johnson, L. A.; June, C. H. *Immunity* **2016**, *44*, 1444–1454.

(11) He, Y.; Schreiber, K.; Wolf, S. P.; Wen, F.; Steentoft, C.; Zerweck, J.; Steiner, M.; Sharma, P.; Michael Shepard, H.; Posey, A.; June, C. H.; Mandel, U.; Clausen, H.; Leisegang, M.; Meredith, S. C.; Kranz, D. M.; Schreiber, H. *JCI Insight* **2019**, *4*, e130416.

(12) Plattner, C.; Höfener, M.; Sewald, N. *Org. Lett.* **2011**, *13*, 545–547.

(13) Liebe, B.; Kunz, H. *Tetrahedron Lett.* **1994**, *35*, 8777–8778.

(14) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595–598.

(15) Bermejo, I. A.; Usabiaga, I.; Compañón, I.; Castro-López, J.; Insausti, A.; Fernández, J. A.; Avenoza, A.; Busto, J. H.; Jiménez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Jiménez-Osés, G.; Hurtado-Guerrero, R.; Cocinero, E. J.; Corzana, F. *J. Am. Chem. Soc.* **2018**, *140*, 9952–9960.

(16) Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; García de Luis, M.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. *J. Am. Chem. Soc.* **2007**, *129*, 9458–9467.

(17) Kračun, S. K.; Cló, E.; Clausen, H.; Levery, S. B.; Jensen, K. J.; Blixt, O. *J. Proteome Res.* **2010**, *9*, 6705–6714.

(18) Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. *J. Am. Chem. Soc.* **2006**, *128*, 14640–14648.

(19) Blixt, O.; Cló, E.; Nudelman, A. S.; Sørensen, K. K.; Clausen, T.; Wandall, H. H.; Livingston, P. O.; Clausen, H.; Jensen, K. J. *J. Proteome Res.* **2010**, *9*, 5250–5261.

(20) Blixt, O.; Bueti, D.; Burford, B.; Allen, D.; Julien, S.; Hollingsworth, M.; Gammerman, A.; Fentiman, I.; Taylor-Papadimitriou, J.; Burchell, J. M. *Breast Cancer Res.* **2011**, *13*, R25

(21) Alkaitis, M. S.; Nardone, G.; Chertow, J. H.; Ackerman, H. C. *Biomed. Chromatogr.* **2016**, *30*, 294–300.

(22) Perucho, J.; Gonzalo-Gobernado, R.; Bazan, E.; Casarejos, M. J.; Jiménez-Escrig, A.; Asensio, M. J.; Herranz, A. S. *Amino Acids* **2015**, *47*, 963–973.

(23) Molnár-Perl, I. *J. Chromatogr. B* **2011**, *879*, 1241–1269.

(24) Somovilla, V. J.; Bermejo, I. A.; Albuquerque, I. S.; Martínez-Sáez, N.; Castro-López, J.; García-Martín, F.; Compañón, I.; Hinou, H.; Nishimura, S. I.; Jiménez-Barbero, J.; Asensio, J. L.; Avenoza, A.; Busto, J. H.; Hurtado-Guerrero, R.; Peregrina, J. M.; Bernardes, G. J. L.; Corzana, F. *J. Am. Chem. Soc.* **2017**, *139*, 18255–18261.

(25) Madariaga, D.; Martínez-Saéz, N.; Somovilla, V. J.; Coelho, H.; Valero-Gonzaíez, J.; Castro-Lo, J.; Asensio, J. L.; Busto, H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Corzana, F.; Peregrina, J. M. *ACS Chem. Biol.* **2015**, *10*, 747–756.

(26) A Study of CART-TnMUC1 in Patients With TnMUC1-Positive Advanced Cancers https://clinicaltrials.gov/ct2/show/NCT04025216?term=tmunity (accessed Sep 29, 2021).

(27) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S.A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York D. M.; Kollman, P. A. (2020), AMBER 2020, University of California, San Francisco.

(28) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

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

(30) Kiyohara, K.; Gubbins, K. E.; Panagiotopoulos, A. Z. *Mol. Phys.* **1998**, *94*, 803–808.

(31) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

5 | The use of unnatural glycopeptides as diagnostic tool

5.1. Introduction

As previously commented, MUC1 mucin is aberrantly glycosylated and overexpressed on the mucosal surface of several types of carcinoma cells and it is composed of two subunits, MUC *C*-terminal and MUC *N*-terminal.[1,2] Later subunit can be detached from the surface and consequently detected in the sera of cancer patients.[3,4] In this sense, it has been found that a low level of MUC1 expression could be found in healthy human sera,[5,6] however, levels over 40 U/mL when using CA 27-29[7] or levels over 25-30 U/mL for CA 15-3[8] may generally suggest the presence of malignancies. Even though in the bibliography it is possible to observe many examples of biomarkers of cancer based on antigens, most of them do not exhibit satisfactory sensitivity and specificity.[4,9–12]

Nevertheless, circulating anti-MUC1 antibodies could represent an appealing alternative as, in general, they persist in the circulating blood for a longer time than the corresponding antigens and are present in the early stages of the disease. Notably, in the bibliography it has been reported multiples studies using MUC1 as a tool to detect and measure these circulating antibodies. In this context, particular attention should be paid to the system developed by Wang and co-workers. They used a MUC1-like peptide comprising six tandem repeat units in the detection of anti-MUC1 antibodies in human sera.[13] In this line of research, Gheybi and co-workers developed a chimera composed of MUC1 and a growth factor receptor associated with breast cancer when it is overexpressed (HER2). This can be used for the detection of antibodies against both MUC1 and HER2 in human ser.[14]

Nowadays, pancreatic cancer is the fourteenth most common cancer globally and the seventh leading cause of cancer-related deaths.[15] In Spain, pancreatic tumor was the third leading cause of deaths provoked by cancer in 2019.[16] Lethality of this disease is a consequence of the late detection of this tumor at a very advanced stage.[17] Currently, surgery, chemotherapy and radiation are the most used therapies employed to treat pancreatic cancer, even though the life expectancy of patients diagnosed with this cancer is between 4 and 6 months.[18] Thus, screening and early diagnosis are crucial for improving its prognosis. However, to date, biomarkers have not the sufficient sensitivity and specificity to be used in clinical assays.

The only serum biomarker approved by the U.S. Food and Drug Administration (FDA), carbohydrate antigen 19-9 (sialyl-Lewis^A, CA19-9) has limited specificity to be used in the early screening of pancreatic cancer.[19] Therefore, some studies have focused on the discovering of novel serum biomarkers to be used in combination with CA19-9 to reduce the proportion of false negative.[20]

Of note, in pancreatic cancer overexpression of MUC1 is correlated with high metastases and poor prognosis.[21] In this regard, circulating anti-MUC1 antibodies could play an essential role in the prognosis for patients with pancreatic cancer in first stages of the disease.[22] Most of the examples found in the bibliography based on the detection of anti-MUC1 antibodies do not exhibit any alteration of the natural antigens. Further modifications on the antigen structure could be a key point for the enhancement of the interaction between antigen and antibody and therefore improving the selectivity and sensitivity of a possible detection system.[23,24] For this purpose, it is essential to know the molecular interactions by which these antibodies recognize their targets. Regarding this issue, our research group has reported the X-ray structure of MUC1-based glycopeptide in a complex with SM3 antibody (Figure 5.1A), which provides important information concerning how this antibody interacts with the antigen and therefore paves the way to the design of non-natural antigens with improved characteristics for recognition.[25]

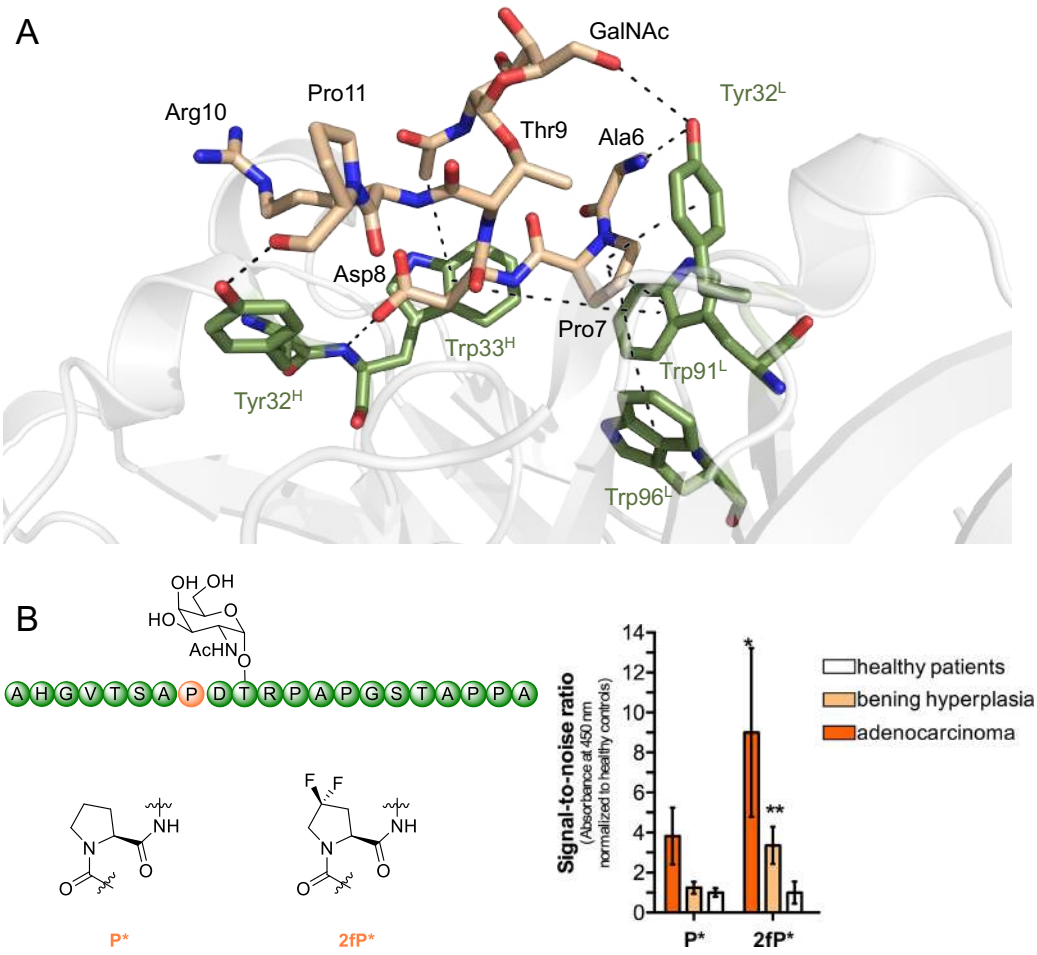


Figure 5.1. (A) Key binding interactions of glycopeptide APDT*RP in complex with antibody SM3. The antibody fragment is shown as a light grey cartoon and the glycopeptide fragment of the antigen is shown as brown carbon atoms. Interactions between the antibody and the antigen are shown as dashed black lines. The PDB entry for this complex is 5A2K. (B) Detection of circulating anti-MUC1 antibodies in sera of patients with prostate cancer using a natural glycopeptide or a synthetic variant (**2fP***).

In line with this research, our group has recently described the synthesis of a non-natural glycopeptide featuring a fluoroproline residue. This new antigen was used for the detection of prostate cancer achieving best results than those obtained for the natural antigen (Figure 5.1B).[26]

On the other hand, as it was described in the previous chapter, we have recently achieved the X-ray structure of a short peptide bearing APGST*AP in complex with 5E5 antibody (Figure 5.2), which also provides crucial information about the binding process of the antigen with this antibody and could help in the

development of structure-guided potent antigens. That is, antigens with improved affinity for anti-MUC1 antibodies related to naturally occurring MUC1-like antigens.

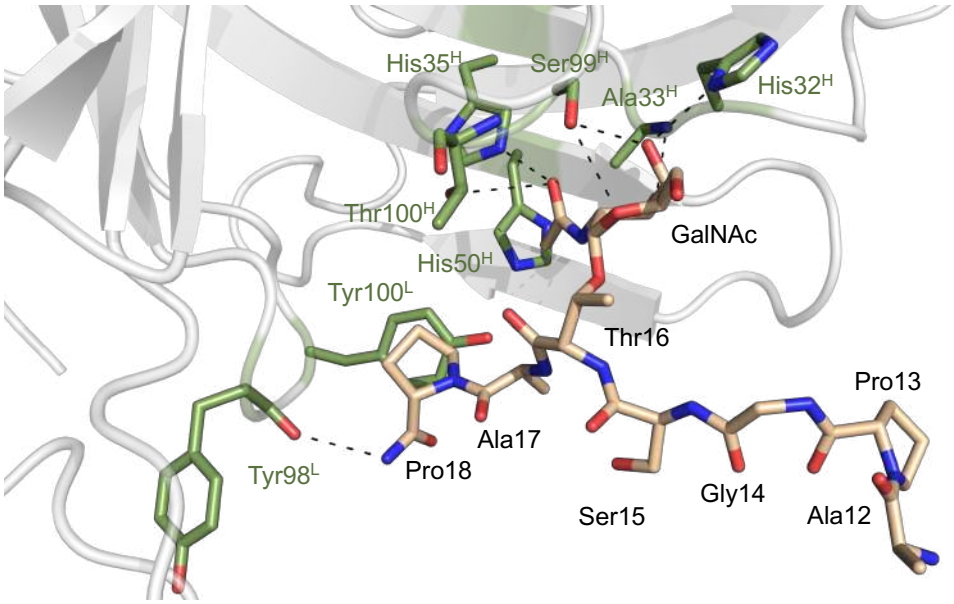


Figure 5.2. Key binding interactions of glycopeptide APGST*AP in complex with antibody 5E5. The antibody fragment is shown as a light grey cartoon and the glycopeptide fragment of the antigen is shown as brown carbon atoms. Interactions between the antibody and the antigen are shown as dashed black lines. The PDB entry for this complex is 6TNP.

5.2. Objectives

In this chapter, we examine how minor changes in key amino acids of the MUC1 sequence can alter the recognition of these antigens by two different anti-MUC1 antibodies. With this purpose, different non-natural amino acids will be synthesized and incorporated in different epitopes of MUC1 to determine their binding affinity for the SM3 and 5E5 anti-MUC1 antibodies by SPR experiments. Then, gold nanoparticles (AuNPs) will be synthesized and conjugated to glycopeptides featuring modifications in their sequence. This nano-system will be used also to carry out the binding affinity studies between the antibodies and the unnatural antigens by densitometric analysis. Lastly, we will perform a serological assay using the AuNPs conjugated to the new glycopeptides to detect anti-MUC1 autoantibodies in sera from pancreatic cancer patients.

5.3. Results

5.3.1. Structure-based design of new antigen mimetics

Herein, we have exploited the X-ray structures of complexes scFv-SM3/glycopeptide **4**[25] and scFv-5E5/glycopeptide **2** reported in the previous chapter in order to design new antigen mimetics with higher affinity than naturally occurring ones.

In the case of scFv-SM3 antibody, crystallographic analysis reveals that $A_6P_7DT_9RP_{11}A_{12}$ fragment is the most recognized part, where numerous hydrogen bonds and stacking interactions stabilize the complex. Although the GalNAc moiety is not essential, it also established, as we have described, a CH-π interaction and a hydrogen bond with residues of the antibody. On this basis, we decided to design short glycopeptides containing the epitope of MUC1 and introduce different modifications in the structure to attempt the improvement of the affinity with this antibody.

Interestingly, non-terminal proline (Pro7) plays a central role in the stabilization through CH-π interactions with the aromatic rings of the $Trp91^L$, $Trp96^L$ and $Tyr32^L$ residues of the antibody (Figure 5.3A). In this context, recent studies have shown that CH-π interactions can be significantly improved by increasing the polarization of the CH bonds involved in them.[27,28] This objective could be achieved by replacing hydrogen atoms with fluorine atoms in specific positions of the proline following the same strategy as in a previous work of our group.[26] A detailed analysis of the X-ray structure suggested the use of the commercially available (4*S*)-4-fluoro-L-proline for this purpose.

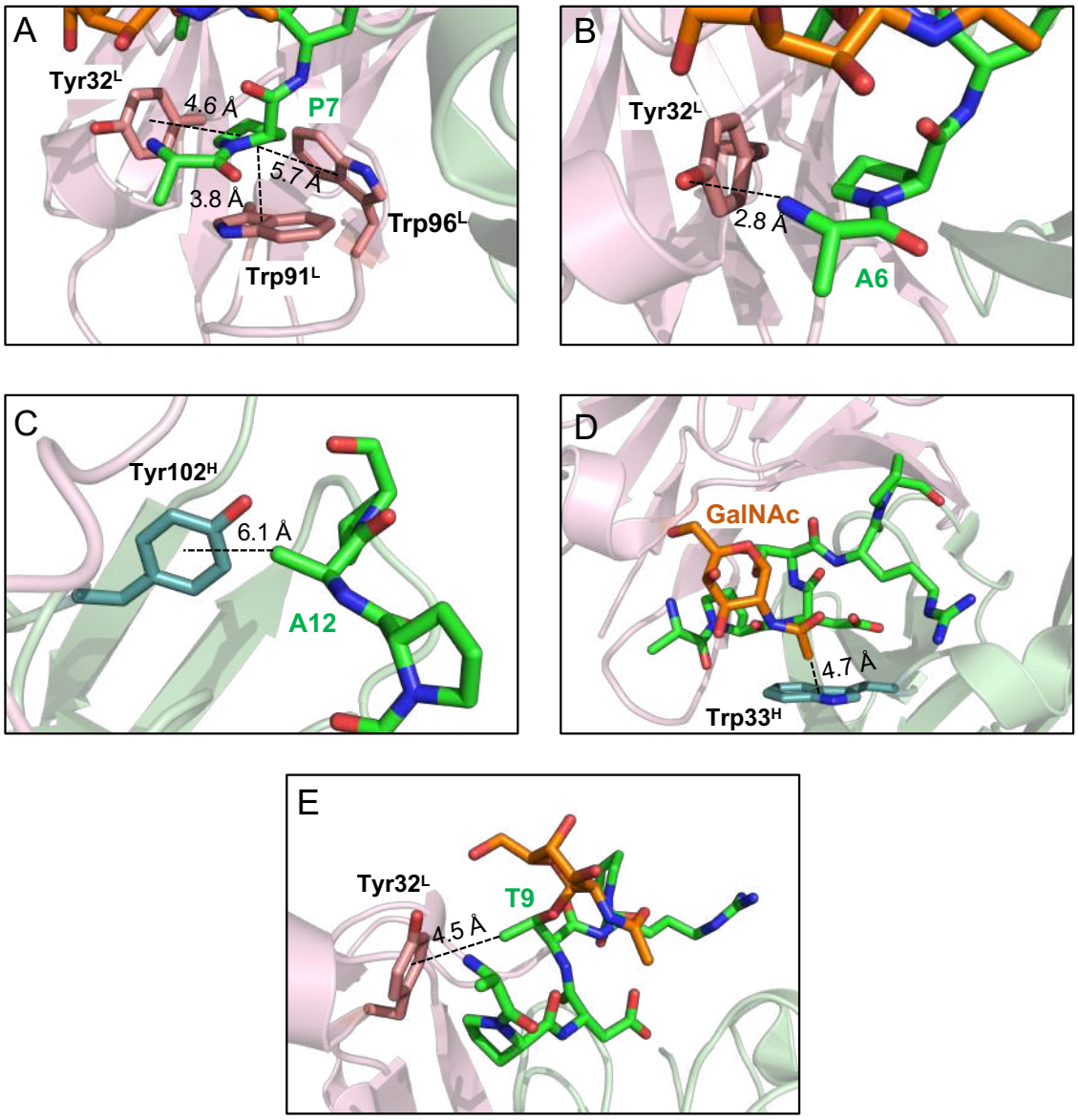
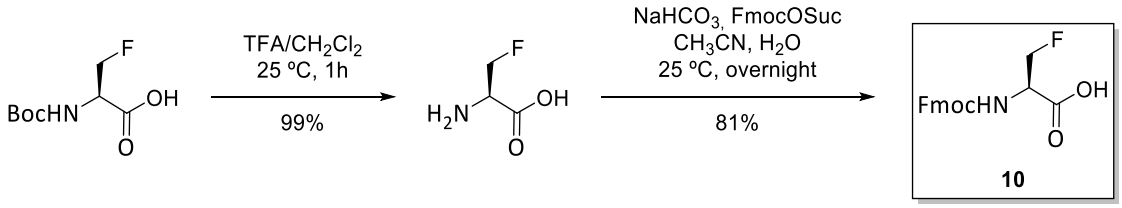


Figure 5.3. Key binding interactions of the antigen with SM3 antibody, as observed in the X-ray crystal structures. The antibody fragment is shown as a light pink (light chain) and light green (heavy chain) cartoon. The peptide fragment of the antigen and the GalNAc moiety are shown as green and orange carbon atoms, respectively. Interactions between the antibody and Pro7 (A), Ala6 (B), A12 (C), GalNac (D) and Thr9 (E) are shown as dashed black lines. The PDB entries for the complexes are 5A2K and 1SM3.

It is also clear from the X-ray structure of the complex scFv-SM3/**4** that a hydrogen bond is established between the NH group of Ala6 and the oxygen of OH of Tyr32^{L} (Figure 5.3B). This interaction could be enhanced by replacing hydrogen atoms in the methyl group of the alanine with fluorine atoms, and in

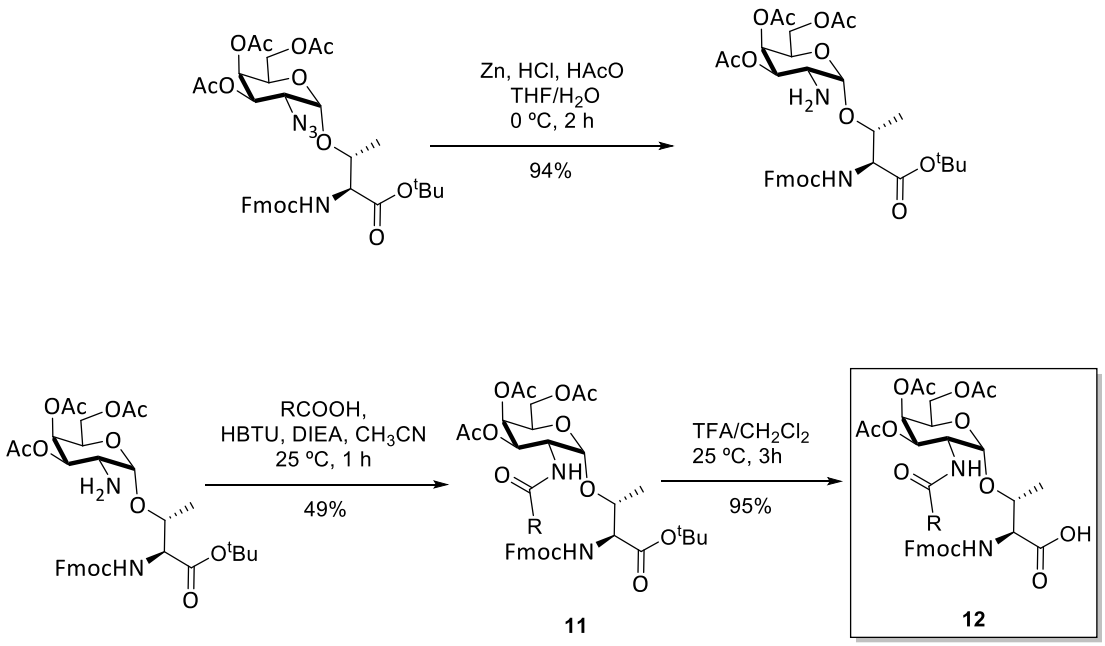
this way increasing the hydrogen donor character of N-H. The following synthetic route was used for the synthesis of this unnatural fluorinated amino acid. Firstly, commercially available Boc-F-Ala-OH was treated with TFA to obtain the naked amino acid, which was then protected with Fmoc to obtain compound **10** ready for use in SPPS.



Scheme 5.1. Synthetic route followed to obtain Fmoc-F-Ala-OH (compound **10**)

Similarly, the X-ray structure reported in the literature for a complex between scFv-SM3 and a longer peptide[29] reveals a CH-π interaction between Ala12 of $SA_6P_7DT_9RPA_{12}P$ peptide and Tyr102^H of SM3 antibody. (Figure 5.3C). The improvement of this binding could be achieved using the F-Alanine previously described, which could increase the polarization of the CH bonds involved in the stabilizing interaction.

On the other hand, the X-ray structure of the complex scFv-SM3/**4** revealed that *N*-acetyl group of the sugar stacks with the aromatic ring of Trp33^H (Figure 5.3D), thus providing the selectivity of SM3 for GalNAc-containing antigens. In theory, this interaction could be also improved by the addition of a fluorine atom in the acetyl group of the carbohydrate. Therefore, we carried out the synthesis of Fmoc-Thr[GalNFAc$(Ac)_3$-α-D]-OH following the methodology described in the literature.[30] Alternatively, we proposed the substitution of the *N*-acetyl group by another group to enhance the interaction with Trp33^H. Thus, we synthesized a Thr-Tn analog which is under patent process, Fmoc-Thr[GalNCOR$(Ac)_3$-α-D]-OH, being R an alkyl or aryl group (compound **12**, Scheme 5.2). Firstly, we synthesized Fmoc-Thr[Gal$NH_2(Ac)_3$-α-D]-O^{t}Bu, following the procedure described[31] and this compound was then reacted with an acid to afford compound **11**. After purification by column chromatography, compound **11** was treated with TFA/CH_2Cl_2 to obtain compound **12** ready-to-use in SPPS without further purification.



Scheme 5.2. Synthetic route followed to obtain Fmoc-Thr[GalNCOR(Ac)$_3$-α-D]-OH (compound **12**)

Lastly, from the crystal analysis, it is possible to observe an interaction between the methyl group of Thr9 with Tyr32^L (Figure 5.3E). The distance between these two residues could be diminished by the elongation of the β-methyl group. Thus, the use of Tn-glycosylated hydroxynorvaline could be a strategy to achieve this objective. This unnatural amino acid was previously used in our group obtaining as result a similar conformation behavior to the natural Thr compound and comparable K_D values of both Thr and Hnv derived glycopeptides.[32] On this basis, Fmoc-Hnv[GalNAc(Ac)$_3$-α-D]-OH was synthesized following the methodology described in the Thesis defended by Iris A. Bermejo (Universidad de La Rioja, 2018).[32]

On the other hand, crystallographic analysis of the complex between scFv-5E5 and a small glycopeptide containing the GST*AP epitope reveals that the most recognized fragment of the glycopeptide is the carbohydrate.[33–35] Thus, the GalNAc unit is involved in several hydrogen bonds with the antibody and stablishes a CH-π interaction between the methyl group and His50^H (Figure 5.4). Interestingly, a hydrogen bond between the –CONH$_2$ group in the *C*-terminal region of the peptide and the carbonyl group of Tyr98^L aromatic moiety. Therefore, we have proposed the design of short glycopeptides containing the

most recognized epitope of MUC1 featuring small modifications in the structure to study the changes of the binding with this antibody.

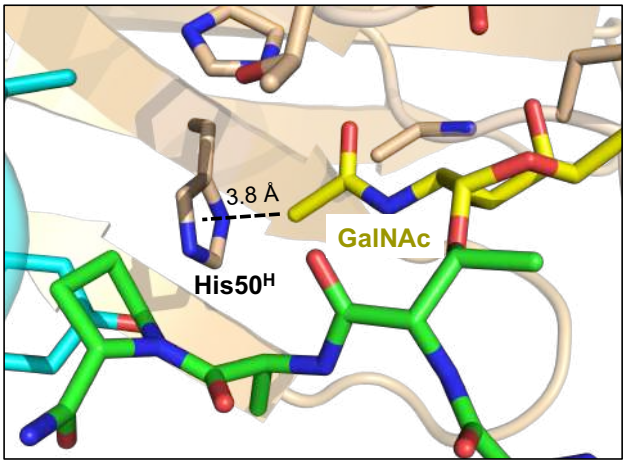


Figure 5.4. Key binding interactions of glycopeptide **2** with 5E5 antibody, as observed in the X-ray crystal structure. The antibody fragment is shown as a light blue (light chain) and light brown (heavy chain) cartoon. The peptide fragment of the antigen and the GalNAc moiety are shown as green and yellow carbon atoms, respectively. Interactions between the antibody and GalNAc are shown as dashed black lines. The PDB entry for this complex is 6TNP.

On this basis, we anticipated two strategies to enhance the binding towards scFv-5E5. The first strategy was the hydrogen-to-fluorine substitution of the methyl group of the *N*-acetyl group, which could induce an enhancement of the corresponding CH-π interaction.[26–28,36–38] Moreover, the substitution of the acetyl group of GalNAc by an alkyl/aryl group could also improve the binding, owing to the reduction of the distance between the antigen and the antibody. For this purpose, compound **12** was also used for the synthesis of a new antigen mimetic.

5.3.2. Synthesis of glycopeptides for affinity assays

Unnatural amino acids Fmoc-Thr[GalNAc(Ac)$_3$-α-D]-OH (Tn-Thr), Fmoc-FAla-OH, Fmoc-Thr[GalNFAc(Ac)$_3$-α-D]-OH, Fmoc-Thr[GalNCOR(Ac)$_3$-α-D]-OH and Fmoc-Hnv[GalNAc(Ac)$_3$-α-D]-OH were synthesized following the procedures described in the literature (see Experimental section 5.5.1).[30,31,39]

The different glycopeptides were then prepared by stepwise microwave assisted solid-phase synthesis (MW-SPPS) on a Liberty Blue synthesizer with Rink Amide MBHA resin using the Fmoc strategy. The building blocks with the unnatural residues were manually coupled as described in Experimental section 5.5.2. The number of equivalents was reduced and the reaction time was extended to increase the yield of this coupling step.[40] The *O*-acetyl groups of GalNAc moiety were then removed in a mixture of NH_2NH_2/MeOH (7:3). Afterward, glycopeptides were released from the resin, and all acid sensitive protecting groups were simultaneously removed using a cleavage cocktail consisting of a mixture of TFA 95%, TIS (triisopropylsilane) 2.5% and H_2O 2.5%, followed by precipitation with cold diethyl ether. Finally, crude products were purified by HPLC, lyophilized and characterized.

In Figure 5.5, the library of glycopeptides synthesized in this work for the affinity assays is summarized. These glycopeptides are divided in two groups: glycopeptides containing SM3 epitope (PDTRP) and glycopeptides containing 5E5 epitope (GSTAP).

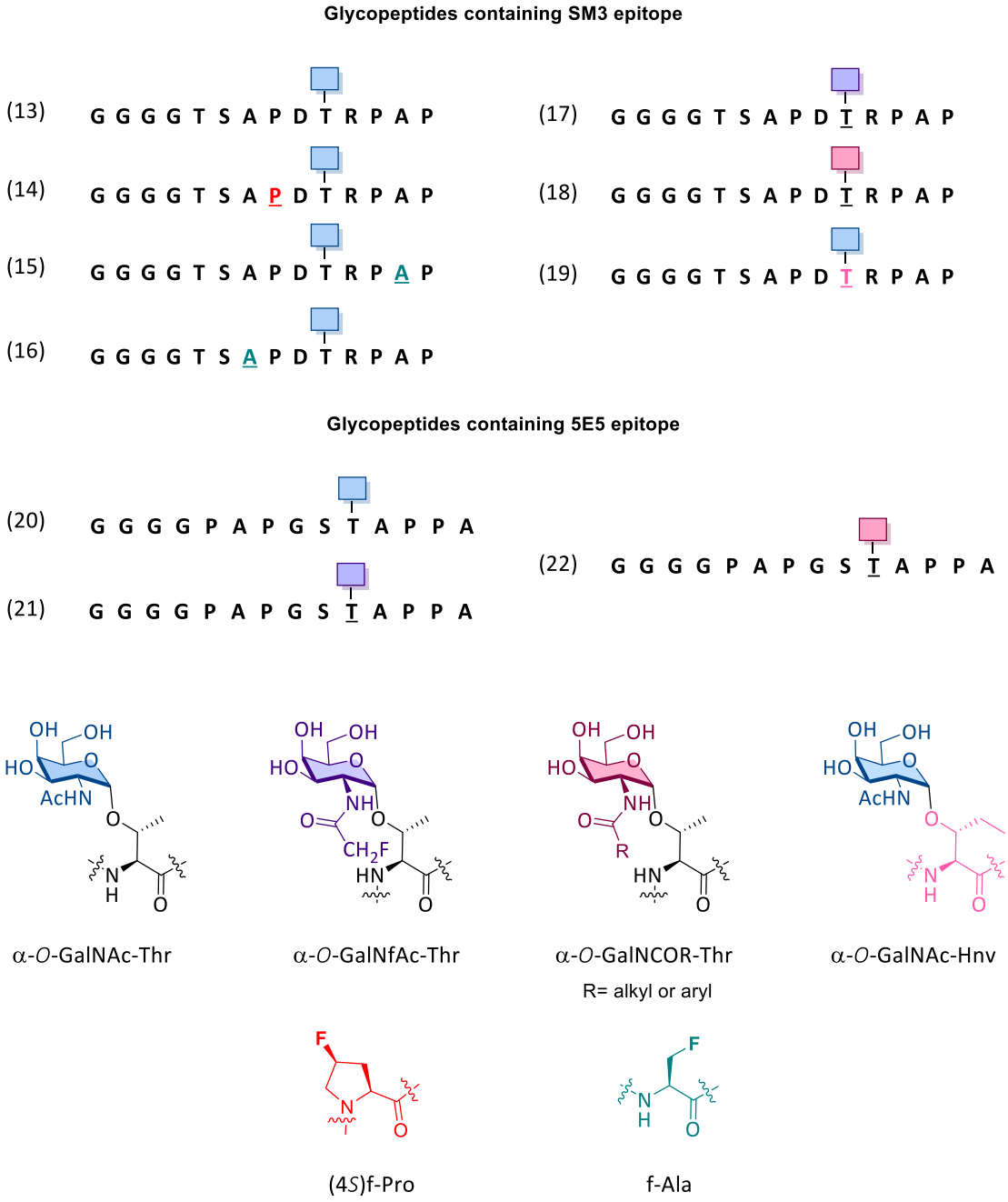


Figure 5.5. Glycopeptides synthesized and studied in this work.

5.3.3. Affinity assays

Surface plasmon resonance (SPR) assays were used to evaluate the binding of glycopeptides shown in Figure 5.5 with 5E5 and SM3 antibodies.

K_D values obtained by SPR for the glycopeptides containing SM3 epitope are summarized in Figure 5.6 (see the green bars). The highest affinity was observed for glycopeptide **14**. In fact, significant improved affinity (~7-fold) was obtained

for this derivative in comparison to the glycosylated epitope **13**. This glycopeptide contains a (4*S*)-4-fluoro-L-proline in the main epitope and its affinity was like that previously reported in our group.[26] This result confirms the importance of the proline residue for the recognition by the SM3 antibody and how it is possible to modulate CH-π interactions by replacing hydrogen atoms with fluorine atoms in specific positions of the proline. A similar affinity value regarding to glycosylated epitope **13** was obtained for derivative **16**, which contains a F-Alanine. On the contrary, glycopeptides **15**, **17**-**19** exhibit similar affinity among them, with K_D values around 2-3-fold lower than glycopeptide **13**.

In the case of 5E5 antibody, the K_D values determined by SPR for different glycopeptides are summarized in Figure 5.6 (see the blue bars). The highest affinity was observed for glycopeptide **22** which contains a Tn antigen mimetic. In this case, an improved affinity (~7-fold) was obtained in comparison to the glycosylated epitope **20**. Interestingly, glycopeptide **21,** which contains a fluorine atom in the GalNAc moiety, has similar K_D value than the natural glycopeptide **20**.

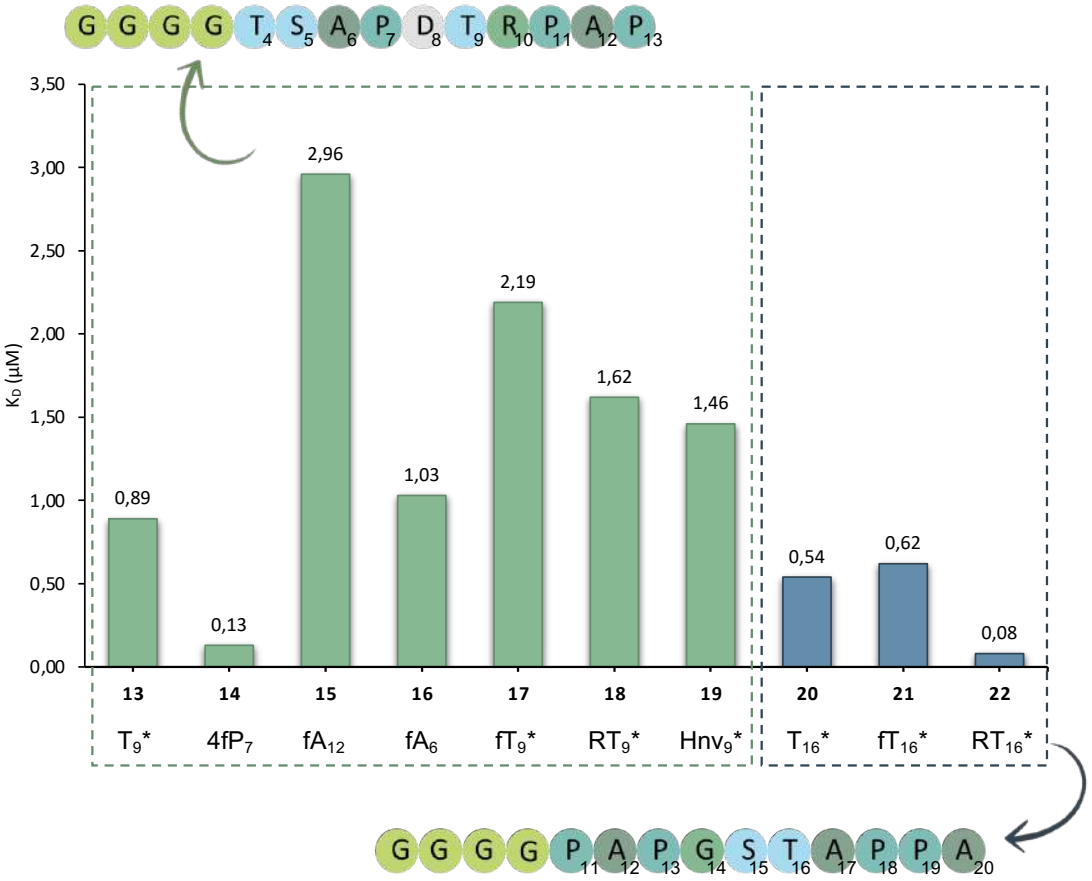


Figure 5.6. K_D constants derived from SPR experiments for the studied glycopeptides. Green and blue bars represent the affinity of glycopeptides for SM3 antibody and 5E5 antibody, respectively.

5.3.4. Synthesis of glycopeptides for nanoparticle-based assay

For the nanoparticles-based assay it was necessary to synthesize the glycopeptides introducing a cysteine at the *N*-terminus to allow the conjugation to AuNPs. Glycopeptides were synthesized by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer with Rink Amide MBHA resin using the Fmoc strategy as previously described. Finally, crude products were purified by HPLC, lyophilized and characterized.

Figure 5.7 shows the library of glycopeptides synthesized in this work for the nanoparticle dot-blot assay. Glycopeptides were divided in three groups: glycopeptides containing SM3 epitope, glycopeptides containing 5E5 epitope and glycopeptides containing the complete MUC1 tandem repeat sequence.

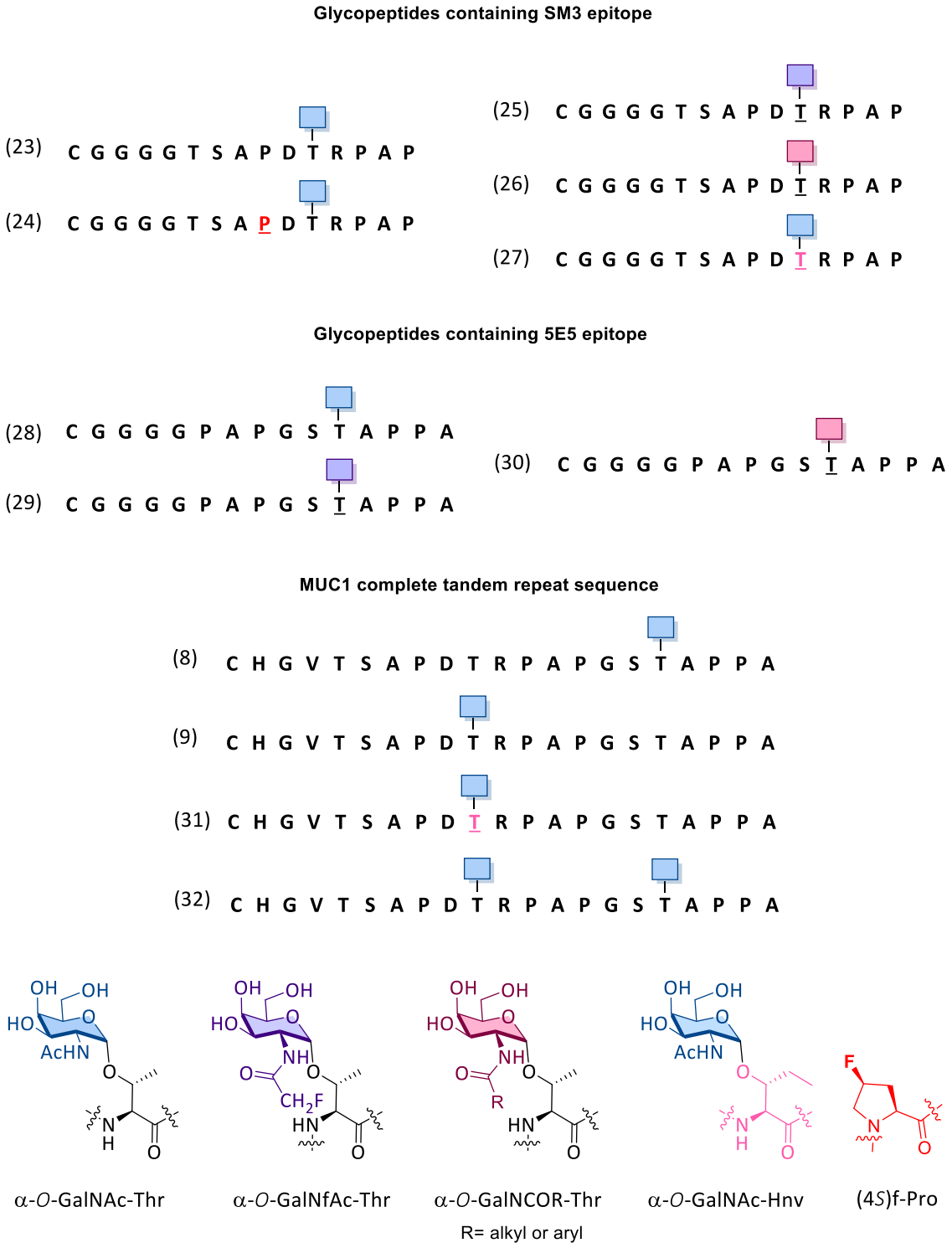


Figure 5.7. Glycopeptides synthesized and studied in this work for nanoparticle-based assay. Note that glycopeptides described in this figure contain a *N*-acetyl cysteine as *N*-terminus.

5.3.5. Synthesis of gold nanoparticles conjugated to glycopeptides

During a short-term stay in the group of Dr. Fiammengo (University of Verona), we performed the synthesis of uniform AuNPs (see Figure 5.8). Following the procedure described in the literature, we could obtain citrate coated AuNPs with a diameter of 13.0 ± 1.0 nm.[41] Then, AuNPs were passivated by formation of self-assembled monolayers of alkyl-PEG600 thiols on their surface[42] by treatment with a mixture of carboxy- and amino-terminated PEG600-thiols (I and II) with the molar fraction of amino terminated derivative x_{NH2} = 0.06 (Figure 5.8). This low molar fraction limits the number of glycopeptides that could be attached to the nanoparticles reducing the possible negative consequences of a broad biofunctionalization. Then, a heterobifunctional linker SM(PEG)$_2$, featuring a maleimide group on one side and an *O*-succinimide reactive ester on the other was coupled through an amide formation reaction with amino-terminated PEG600-thiols.

The last step was the coupling of the glycopeptides to the gold nanoparticles. To this purpose, and before the conjugation reaction, glycopeptides were treated with TCEP in order to prevent the existence of disulfide bridges. The TCEP-treated peptides were added to a freshly prepared cold solution of linker-AuNPs and shaken overnight at 4 °C.

The AuNP-glycopeptides were further characterized by measuring the hydrodynamic radius and the ξ-potential using dynamic light scattering (DLS). These measurements confirmed that the conjugation of the glycopeptides increases the hydrodynamic radius.

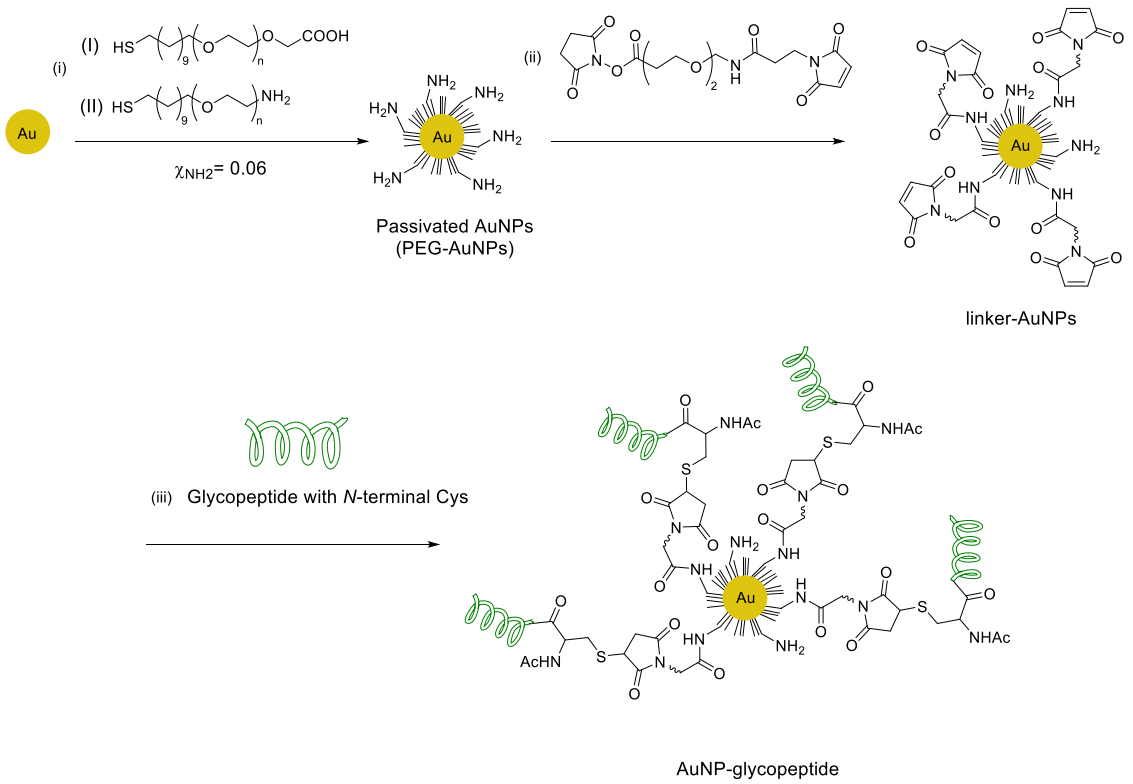


Figure 5.8. Synthetic route for the AuNP-glycopeptides: (i) passivation of AuNPs with alkyl-PEG600 thiols, (ii) coupling of $SM(PEG)_2$ linker, (iii) coupling of the glycopeptides.

5.3.6. Nanoparticles-based dot-blot assay with anti-MUC1 antibodies

The peptide-functionalized nanoparticles were tested in a dot blot assay to detect two different anti-MUC1 antibodies, SM3 and 5E5. The assay was divided into two experiments, the first one was performed using the AuNPs containing glycopeptides with the SM3 epitope (AuNP-glycopeptides **23**-**27**, **9**, **31**) and, the second one was focus on using AuNPs containing the 5E5 epitope (AuNP-glycopeptides **28-30**, **8**). Each group of AuNPs was spotted onto a nitrocellulose membrane with a capillary (16 nM) and allowed to dry at room temperature. Later, each membrane was blocked with 5% non-fat milk in tris-buffered saline (TBS) for 1 h to block nonspecific interactions. Then, membranes were washed in TBS with 0.1% Tween 20 (3×10 min) and then incubated overnight at room temperature with commercially available anti-MUC1 mouse monoclonal antibodies 5E5 or SM3 using different concentrations (ranging from 1 µg/mL to to 0.0001 g/mL in 5% non-fat milk in TBS). After washing with TBS containing 0.1% Tween 20 (3×10 min), the membranes were incubated with goat anti-

mouse IgG (H+L) cross-adsorbed Dylight 800 secondary antibody, diluted 1:2500 in milk for 1 h at room temperature. This secondary antibody allows the detection of primary antibodies bound to the glycopeptides immobilized on the AuNPs. Finally, membranes were washed in TBS with 0.1% Tween 20 (3×10 min) and with TBS (2x10 min) and visualized using an Oddysey Infrared Imaging System from LI-COR Biosciences (Figure 5.9).

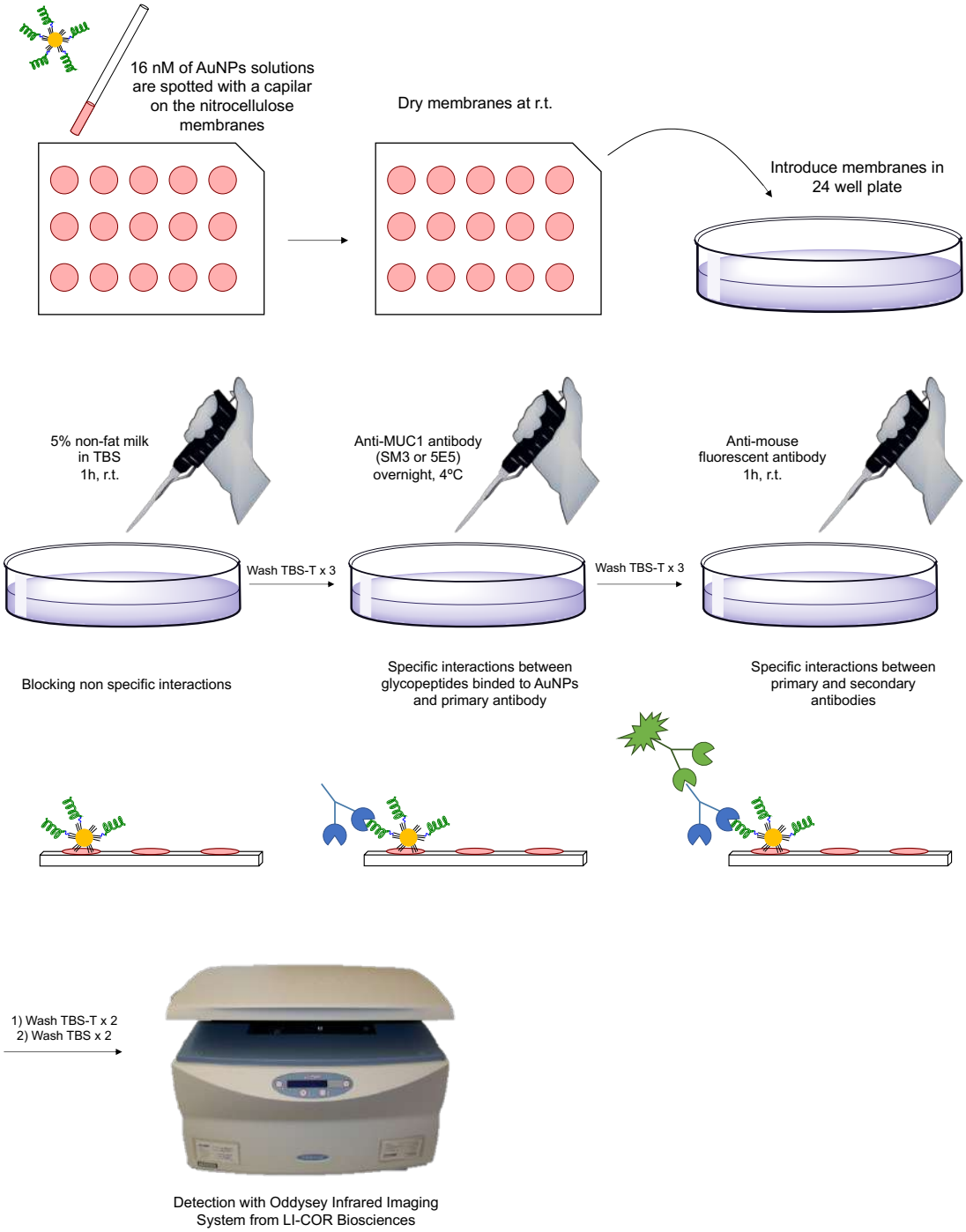


Figure 5.9. Scheme of dot-blot assay with anti-MUC1 antibodies.

Quantification of the intensity of the spots of the membranes was done via densitometry. Previously background subtracted membranes were analyzed to obtain an integrated density of every spot. Results were reflected in a concentration of antibody vs intensity graph. Nonlinear regression was used to fit affinity curves. Concerning SM3 antibody, binding curves for AuNP-glycopeptides **9**, **23**-**27** and **31** are represented in Figure 5.10. The K_D values obtained by densitometry for the glycopeptides containing SM3 epitope are summarized in Figure 5.11 (see the green bars).

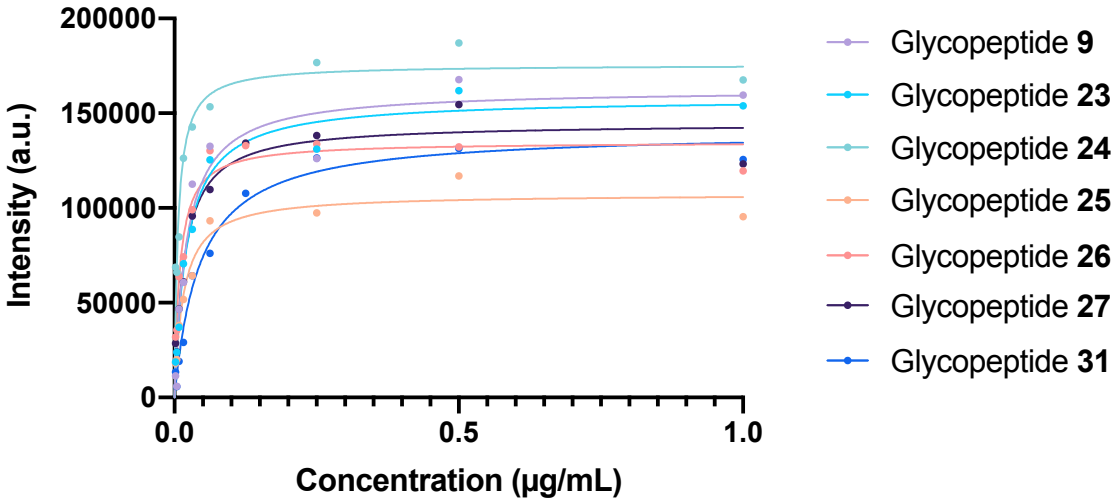


Figure 5.10. Binding curves for AuNP-glycopeptides **9**, **23-27** and **31** with the commercial SM3 antibody using a one-site binding mode.

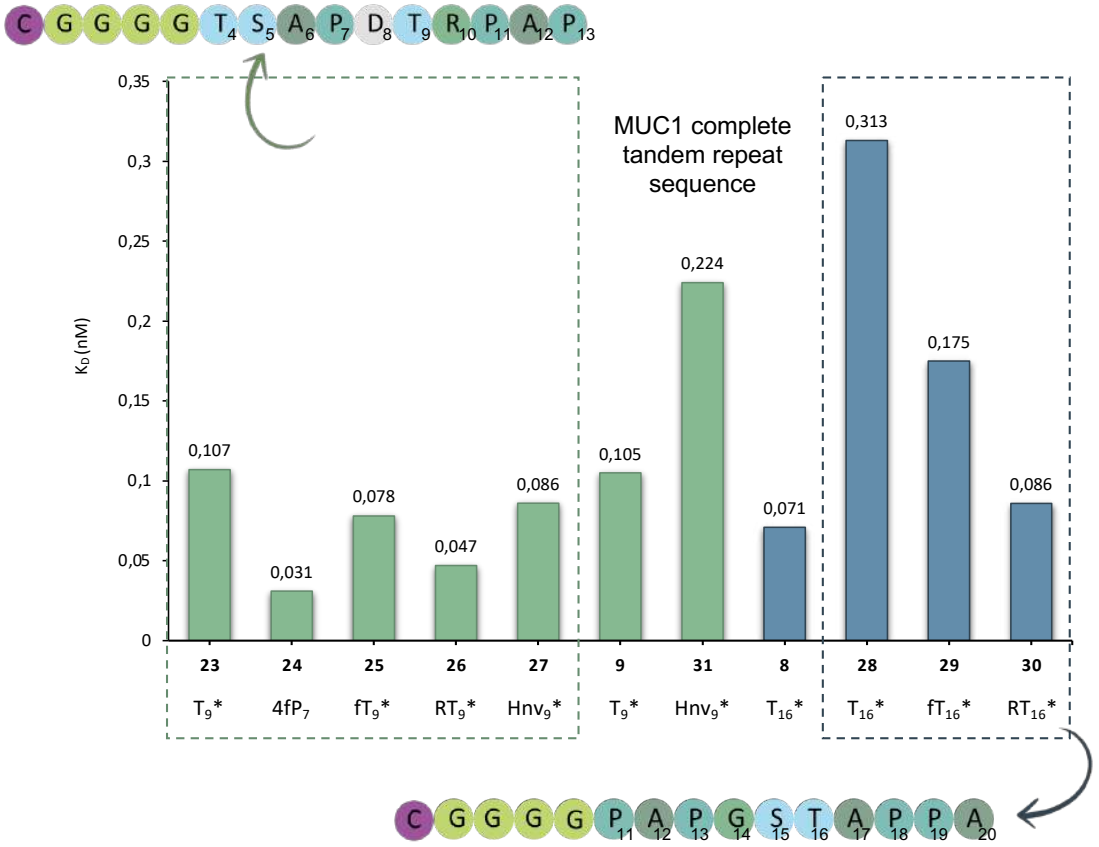


Figure 5.11. K_D constants derived from experiments for the studied AuNP-glycopeptides conjugates. Green and blue bars represent the affinity of glycopeptides for SM3 antibody and 5E5 antibody, respectively.

The highest affinity was observed for glycopeptide **24** which contains a (4*S*)f-Pro in the main epitope. This result is in agreement with those obtained with SPR assays previously reported. A similar affinity value was observed for glycopeptide **26**, which contains a mimetic Tn antigen with a modification of the NHAc chain of the carbohydrate. Glycopeptides **23, 25** and **27** have similar affinity values that are ~3.5 to 2.5 times lower than those of glycopeptide **24**. These results show the same trend than those determined by SPR assays.

Concerning the complete MUC1 tandem repeat peptides (glycopeptide **9** and **31**) the one containing Tn-Thr antigen exhibits 2-fold higher affinity than the Tn-Hnv derivative.

On the other hand, regarding 5E5 antibody, binding curves for AuNP-glycopeptides **8** and **28-30** are represented in Figure 5.12. The K_D values obtained by densitometry for the glycopeptides containing 5E5 epitope are summarized in in Figure 5.11 (see the blue bars).

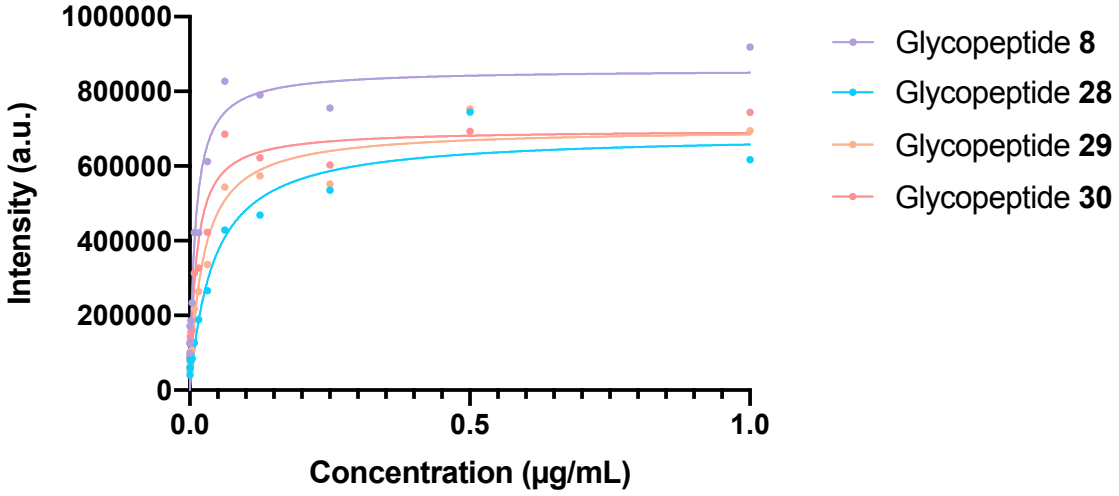


Figure 5.12. Binding curves for AuNP-glycopeptides **8** and **28-30** with the 5E5 antibody using a one-site binding mode.

The highest affinity was observed for glycopeptide **30** which contains a Tn antigen mimetic with an alkyl/aryl group in the *N*-acetyl position. This result is in agreement with the affinity values obtained with SPR. Glycopeptide **28**, which consists in the short 5E5 sequence with Tn antigen in the GSTA epitope, and glycopeptide **29** featuring a fluorine atom at the NHAc group of the sugar present lower affinity values respecting the best candidate **30**. These results also concur with those obtained by SPR assays.

Glycopeptide **8** has one of the best affinity values, similar to glycopeptide **30.** This fact can be due to other interactions of this longer antigen with the antibody through the additional peptide backbone.

5.3.7. Nanoparticles-based dot-blot assay with sera of cancer patients.

The peptide-functionalized nanoparticles were subsequently used in a dot-blot assay to detect possible anti-MUC1 antibodies in human sera. The study was carried out using healthy patients' sera (n= 20) and sera of patients with pancreatic cancer (n= 20) provided by Hospital San Pedro (Logroño) (see Experimental section 5.5.5)

Each group of AuNPs was spotted onto a nitrocellulose membrane with a capillary (16 nM) and allowed to dry at room temperature. Then, each membrane was blocked with 5% BSA in tris-buffered saline (TBS) for 1 h to prevent nonspecific interactions. Finally, a procedure similar to that previously described for the commercially available antibodies was used (Figure 5.13).

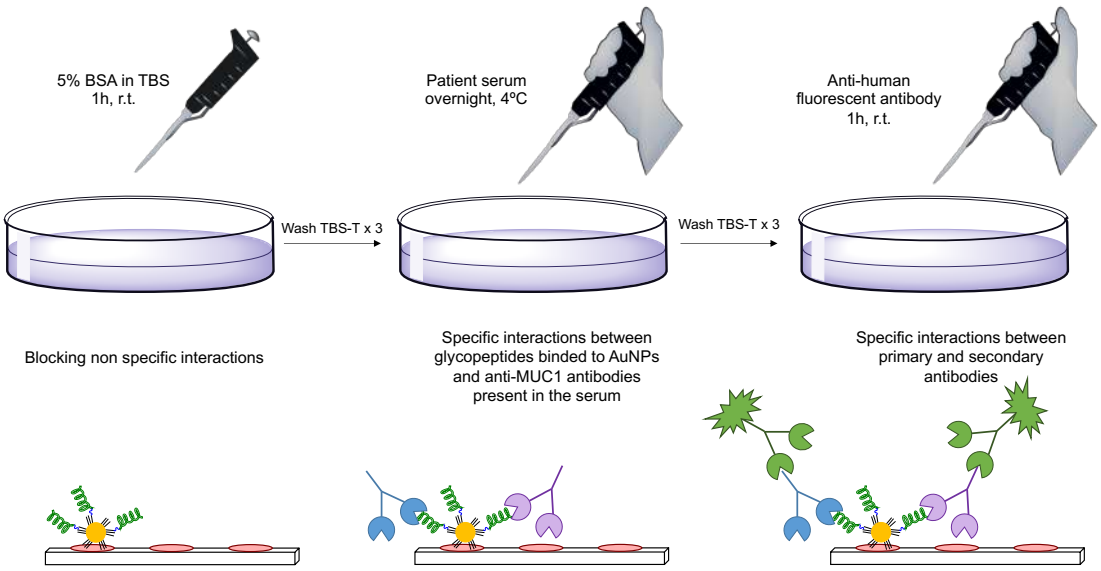


Figure 5.13. Scheme of dot-blot assay with patient sera.

Statistic study comparing the diverse antigens

Dot-blot signal intensities were compared using Student's t-test, and the results were represented as heat map and boxplot graphs (see Experimental section 5.5.5). Significant differences, in terms of intensity, between the sera of patients and controls were found using the glycopeptides containing the 5E5 epitope

(Figure 5.14). This result may suggest that pancreatic cancer patients generate 5E5-like antibodies that can be efficiently detected with our dot-blot assay.

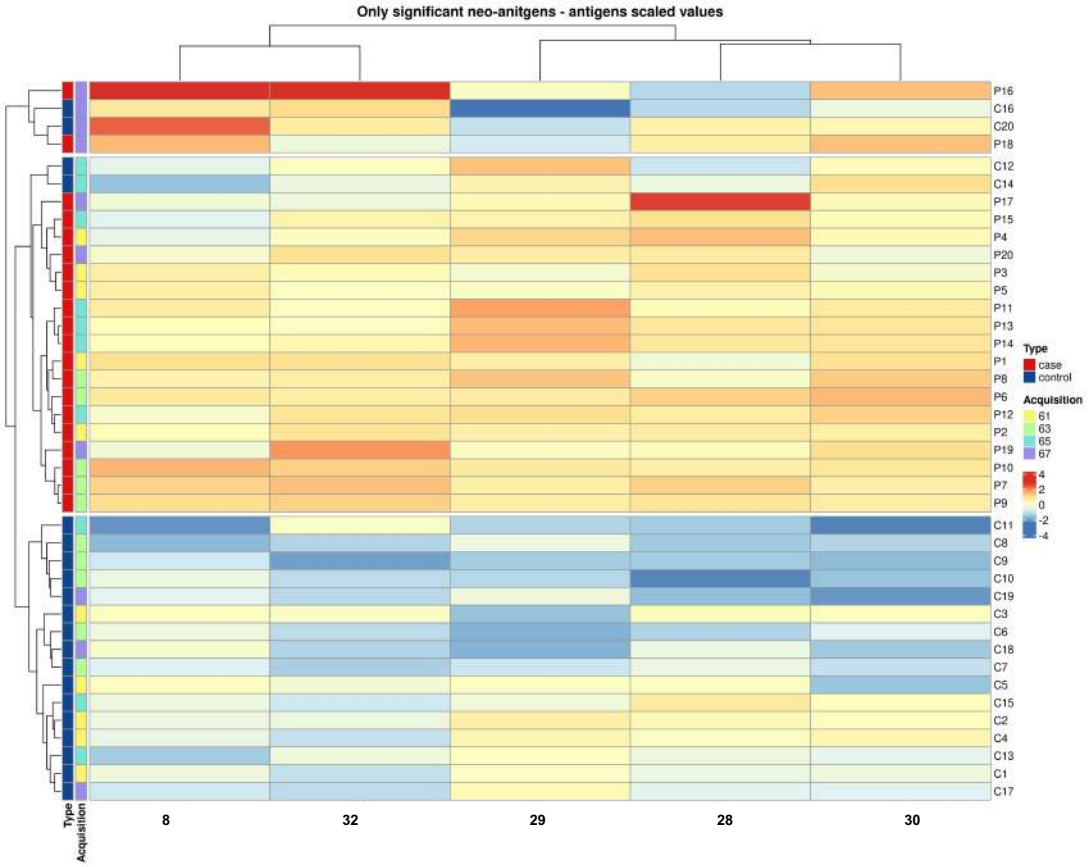


Figure 5.14. Heat map of the signal intensities obtained with the dot-blot assay for 40 sera samples from healthy patients (C1-C20, control) and patients with pancreatic cancer (P1-P20, cases) screened with the 5 most reactive antigens (glycopeptides **8**, **32**, **29**, **28**, **30**). Each row represents a patient and each column an antigen. The degree of expression is indicated by different colors, with reactivity increasing between blue (low) and red (high).

The levels of MUC1 antibodies found in cancer patients (11.76 ± 0.14) was statistically significantly higher ($p<0.0001$) than in healthy volunteers (11.38 ± 0.26) when it was used glycopeptide **30**. (Fig. 5.15A).

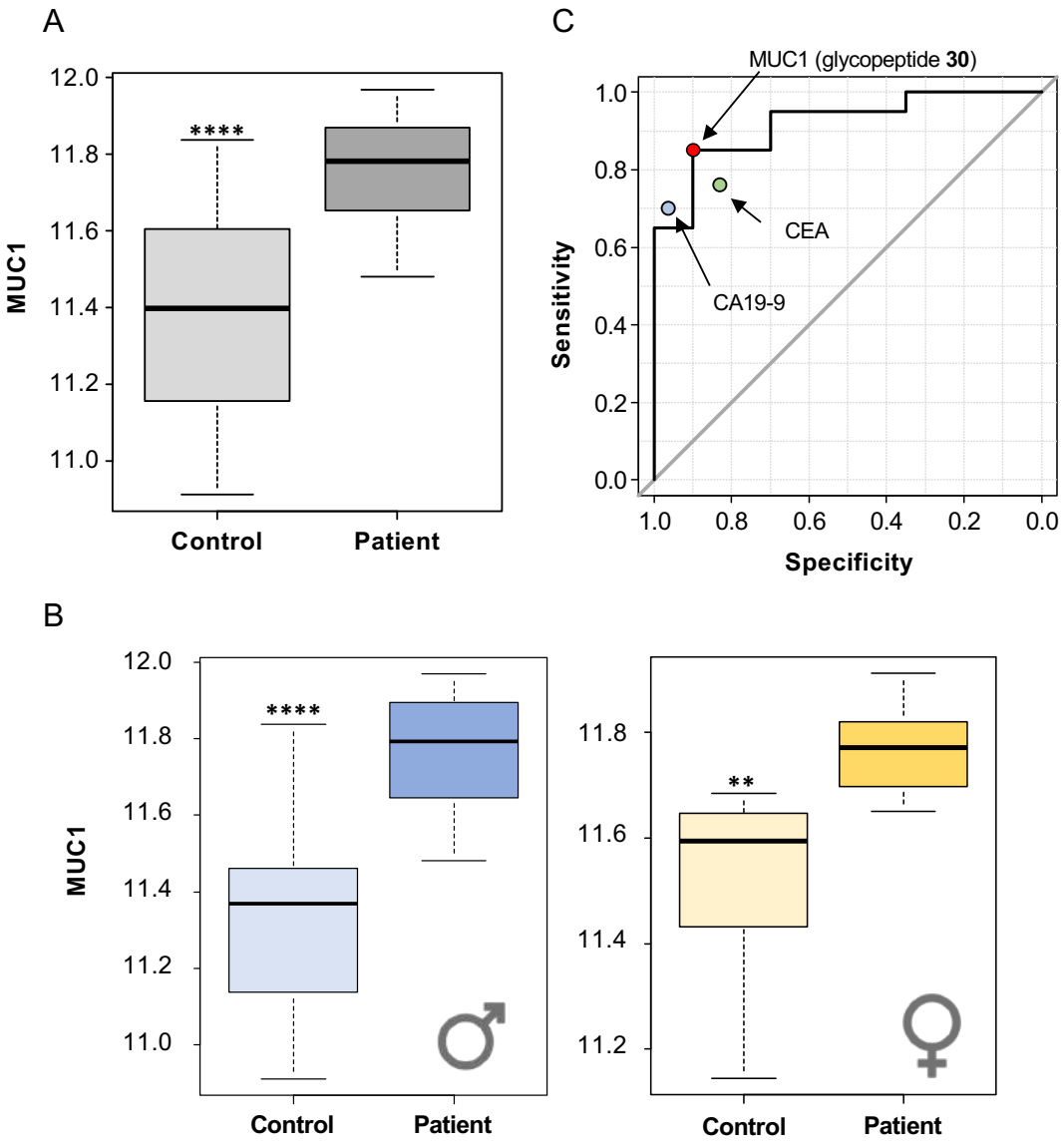


Figure 5.15. (A) Levels of MUC1 antibodies in healthy volunteers (control) and in pancreatic cancer patients using glycopeptide **30**. Box plots represent the interquartile range with the median as a horizontal line. Whiskers encompass the maximum and minimum values of the population. ****: $p<0.0001$. (B) Levels of antibodies in healthy volunteers (control) and in pancreatic cancer patients divided by sex using glycopeptide **30**. Box plots represent the interquartile range with the median as a horizontal line. Whiskers encompass the maximum and minimum values of the population. **: $p<0.01$; ****: $p<0.0001$. (C) Receiver operating curve (ROC) representing the sensitivity and specificity of MUC1 (glycopeptide **30**) values, with an area under the curve (AUC) of 0.918 (95% CI: 0.832-1.000), 85% sensitivity and 90% specificity, with a threshold at 11.63. Blue and green dots represent sensitivity and specificity of two currently used markers, CA19-9[43] and CEA,[44] respectively.

There was no correlation of MUC1 antibodies levels with either age (p=0.40) or sex (p=0.37), but when analyzing the differences between patients and controls in men or women, the differences remained significant irrespective of the patient´s sex (Fig. 5.15.B). Among the male population, the statistical significance remained high ($p<0.0001$) whereas among females the significance was a little lower ($p=0.007$).

To better understand the potential immunoreactivity of glycopeptide **30** as a diagnostic tool, a receiver operating curve (ROC) was built.[45] An area under the curve (AUC) of 0.918 (95% CI: 0.832-1.000) was obtained. Optimal threshold was calculated at 11.63 for the maximum sum of sensitivity (85%) plus specificity (90%) (Fig. 5.15.C). These values were compared with two tumor markers currently in use for the detection of pancreatic cancer, CA19-9,[43] a tetrasaccharide also known as sialyl-Lewis^A and a carcinoembryonic antigen, CEA.[44] It was observed an enhancement of the sensitivity when using our assay and an excellent value of sensitivity, which suggest that an assay based on this glycopeptide can be used for the detection of pancreatic cancer.

5.3.8. Molecular Dynamics simulations of glycopeptide 30 in complex to 5E5 antibody

Considering the great results obtained for glycopeptide **30** both in the affinity assays and in the serological study we decided to get insight into the molecular recognition of this glycopeptide in solution. To this purpose, we run 0.5 μs Molecular Dynamics (MD) simulations of on this derivative in complex with 5E5 antibody (Figure 5.16). Our calculations show that the complex is stable throughout the trajectory. Interestingly, all interactions present in the X-ray structure of the natural glycopeptide are also preserved in the new complex. The replacement of the *N*-acetyl by an alkyl or aryl group of the carbohydrate enhances the CH/π interactions with His50^H, which can result in a stronger interaction of the antigen with the 5E5 antibody.

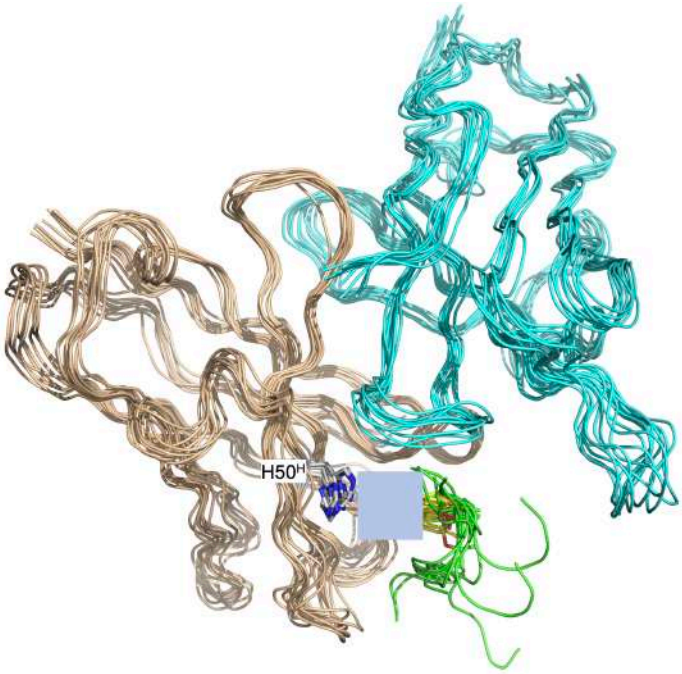


Figure 5.16. Structural ensembles derived from 0.5 μs MD simulation performed on glycopeptide **30** in complex to 5E5 antibody. The antibody is shown as light blue (light chain) and brown (heavy chain) cartoon. The peptide backbone of the antigen is shown in green, and the carbon atoms of the sugar moiety are shown in yellow. The complete structure of the glycopeptide is not shown in the figure because this compound is under patent process.

5.4. Conclusions

Cancer is the main health problem worldwide, with pancreatic cancer being one of the deadliest tumors. In this regard, circulating anti-MUC1 antibodies may play an essential role in the prognosis of patients with pancreatic cancer in the early stages of the disease.[22] In this chapter, a comprehensive study of the molecular basis on which two different anti-MUC1 antibodies (SM3 and 5E5) recognize their targets was performed. On this basis, subtle chemical modifications to the antigen structure have been proposed to enhance the antigen-antibody interactions. Subsequently, using the results of affinity assays, we have demonstrated how the rational structure-based design of new antigens can lead to more efficient antigens in terms of affinity to anti-MUC1 antibodies.

In addition, a serological study based on a nanoparticle dot-blot assay was performed, showing significant differences between healthy patients and

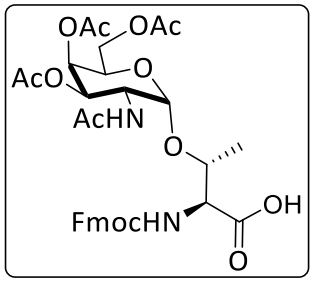
patients with pancreatic cancer when using glycopeptides containing the 5E5 epitope.

Finally, our studies confirm the efficiency of the APGST*AP epitope in detecting pancreatic cancer in serum. This result suggests that patients with pancreatic cancer produce 5E5-like antibodies that can be efficiently detected by our dot-blot assay.

5.5. Experimental section

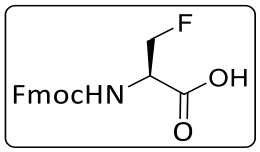
5.5.1. Synthesis of non-natural amino acids

Synthesis of Fmoc-Thr[GalNAc(Ac)$_3$-α-D]-OH



Synthesis of this building block was carried out following the literature.[31]

Synthesis of Fmoc-F-Ala-OH (compound 10)



Boc-F-Ala-OH commercially available (200 mg, 0.97 mmol) was dissolved in a mixture of TFA/CH_2Cl_2 (12mL, 1:1). After stirring for 1 h at room temperature, the solvent was removed and free amino acid H_2N-F-Ala-OH was obtained as a white solid without further purification (103 mg, 0.96mmol, 99% yield).

^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.92 (dddd, 2H, *J* = 47.6, 29.9, 10.7, 3.3 Hz, H_β) 4.37 (dt, 1H, *J* = 29.9, 3.2 Hz, H_α).

^{19}F NMR (282 MHz, $CDCl_3$) δ (ppm)**:** -232.69-(-232.05) (m, 1F, CH_2F^{Ala}).

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_3H_7FNO_2$: 108.0455; Found: 108.0457.

Then, this amino acid (103 mg, 0.96mmol) was dissolved in H_2O (2 mL) and it was added $NaHCO_3$ (243 mg, 2.89 mmol) and the solution was diluted with acetonitrile (4 mL), followed by the addition of Fmoc-OSu (561 mg, 1.66 mmol). The solution was stirred at room temperature overnight. After this time, acetonitrile was removed under reduced pressure and the aqueous solution was washed with Et_2O, followed by acidification with HCl 6M solution and subsequent extraction with a mixture of CH_3Cl/iPrOH (3:1). The organic layer

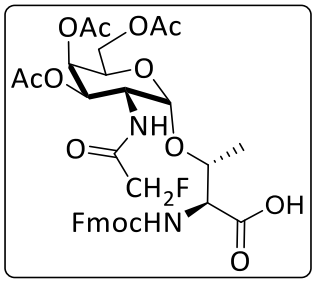
was concentrated and compound **10** was obtained without further purification (256 mg, 0.78 mmol, 81%).[46]

^{1}H NMR (400 MHz, D_2O) δ (ppm): 7.74 – 7.83 (m, 2H, arom.), 7.59 – 7.68 (m, 2H, arom.), 7.30 – 7.48 (m, 4H, arom.), 5.56 (d, 1H, *J* = 7.9 Hz, NHFmoc), 4.55 – 5.03 (m, 3H, H_β, H_α), 4.37 – 4.52 (m, 2H, CH_{2Fmoc}), 4.25 (t, 1H, *J* = 6.9, CH_{Fmoc}).

Decoupled ^{19}F NMR (282 MHz, $CDCl_3$) δ (ppm)**:** -230.4 (s, 1F, CH_2F^{Ala}).

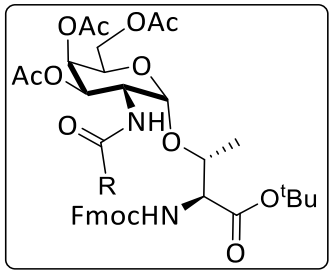
HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{18}H_{17}FNO_4$: 330.1136; Found: 330.1130.

Synthesis of Fmoc-Thr[GalNFAc(Ac)$_3$- α-D]-OH



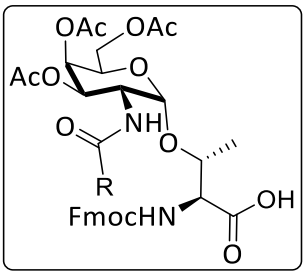
Synthesis of this building block was carried out following the literature.[30]

Synthesis of Fmoc-Thr[GalNCOR(Ac)$_3$-α-D]-O^{t}Bu (compound 11)



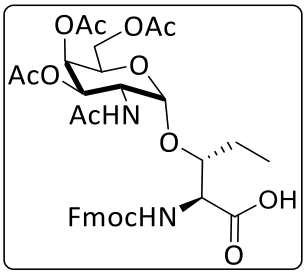
Synthesis of this compound is not described because it is under patent process.

Synthesis of Fmoc-Thr[GalNCOR(Ac)$_3$-α-D]-OH (compound 12)



Synthesis of this compound is not described because it is under patent process.

Synthesis of Fmoc-Hnv[GalNAc(Ac)$_3$- α-D]-OH



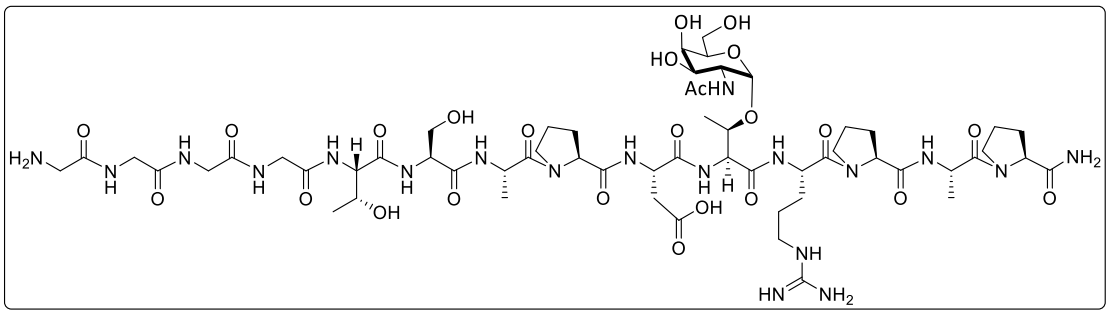
Synthesis of this building block was carried out following the literature.[32]

5.5.2. General procedure for solid-phase peptide synthesis (SPPS)

All glycopeptides were synthesized by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). Non-natural amino acids were synthesized as described before with the exception of Fmoc-(4*S*)-4-fluoro-L-proline commercially available, and manually coupled using HBTU [(2-(1H-benzotriazol-1-yl)- 1,1,3,3-tetramethyluronium hexafluorophosphate], 0.9 equiv., and 0.25 mL of DIPEA (2.0 M in NMP) dissolved in 1 mL of DMF, while all other Fmoc amino acids (5.0 equiv) were automatically coupled using oxyma pure/DIC (N,N'-diisopropylcarbodiimide). The *O*-acetyl groups of GalNAc moiety were removed in a mixture of NH_2NH_2/MeOH (7:3) for the Tn antigen. Glycopeptides were then released from the resin, and all acid sensitive protecting groups were simultaneously removed using TFA 95%, TIS (triisopropylsilane) 2.5% and H_2O 2.5%, followed by precipitation with cold diethyl ether. The crude products were

purified by HPLC on a Phenomenex Luna C18(2) column (10 µm, 250 mm × 21.2 mm) and a dual absorbance detector, with a flow rate of 20 mL/min.

Synthesis of glycopeptide 13 (GGGGT$_4$S$_5$A$_6$P$_7$D$_8$T*$_9$R$_{10}$P$_{11}$A$_{12}$P$_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **13** was obtained and purified by semi-preparative HPLC.

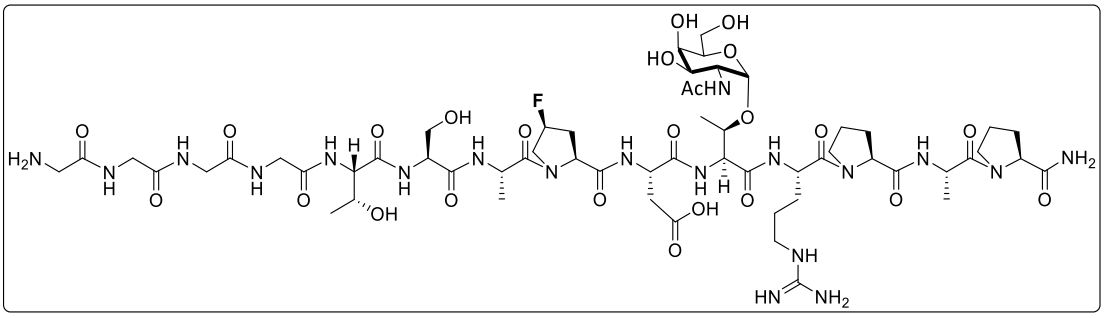
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.80 – 4.86 (m, 1H, H_{1S}), 4.20 – 4.64 (m, 11H, $H\beta_{Thr4}$, $H\beta_{Thr9}$, $H\alpha_{Pro7}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Ala6}$, $H\alpha_{Ala12}$, $H\alpha_{Thr4}$, $H\alpha_{Thr9}$, $H\alpha_{Ser}$, $H\alpha_{Asp}$), 3.53 – 4.10 (m, 23H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\alpha_{Arg}$), 3.16 – 3.25 (m, 2H, $2H\delta_{Arg}$), 2.89 (ddd, 2H, J = 50.5, 17.0, 7.0 Hz, $2H\beta_{Asp}$), 2.21 – 2.33 (m, 3H, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$), 1.80 – 2.08 (m, 14H, $2H\beta_{Arg}$, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $2H\gamma_{Pro7}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$, $NHCOCH_3$), 1.64 – 1.71 (m, 2H, $2H\gamma_{Arg}$) 1.30 – 1.39 (m, 6H, CH_{3Ala6}, CH_{3Ala12}), 1.15 – 1.25 (m, 6H, CH_{3Thr4}, CH_{3Thr9}).

Semi-preparative HPLC: Rt = 13.2 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (12.5:87.5), 15 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 8.3 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{58}H_{96}N_{19}O_{24}$: 1442.6870; Found: 1442.6853.

Synthesis of glycopeptide 14 (GGGGT$_4$S$_5$A$_6$(4*S*)fP$_7$D$_8$T*$_9$R$_{10}$P$_{11}$A$_{12}$P$_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **14** was obtained and purified by semi-preparative HPLC.

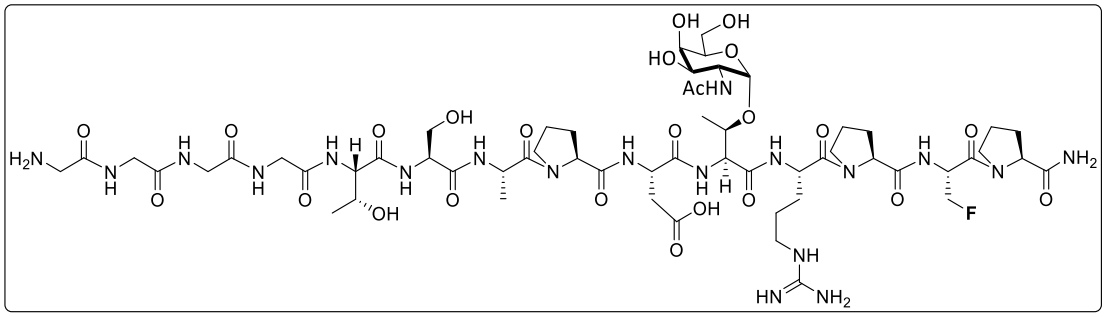
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.84 – 4.90 (m, 1H, H_{1S}), 4.22 – 4.60 (m, 11H, $H\beta_{Thr4}$, $H\beta_{Thr9}$, $H\alpha_{Pro7}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Ala6}$, $H\alpha_{Ala12}$, $H\alpha_{Thr4}$, $H\alpha_{Thr9}$, $H\alpha_{Ser}$, $H\alpha_{Asp}$), 3.59 – 4.08 (m, 23H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\alpha_{Arg}$), 3.18 – 3.24 (m, 2H, $2H\delta_{Arg}$), 2.90 (ddd, 2H, *J* = 48.5, 17.0, 6.7 Hz, $H\beta_{Asp}$), 1.76 – 2.70 (m, 16H, $2H\beta_{Arg}$, $2H\beta_{Pro7}$, $2H\beta_{Pro11}$, $2H\beta_{Pro13}$, $H\gamma_{Pro7}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$, $NHCOCH_3$), 1.65 – 1.71 (m, 2H, $2H\gamma_{Arg}$) 1.34 – 1.42 (m, 6H, CH_{3Ala6}, CH_{3Ala12}), 1.17 – 1.26 (m, 6H, CH_{3Thr4}, CH_{3Thr9}).

Semi-preparative HPLC: Rt = 13.0 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (12.5:87.5), 15 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 11.4 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{58}H_{95}FN_{19}O_{24}$: 1460.6776; Found: 1460.6705.

Synthesis of glycopeptide 15 (GGGGT$_4$S$_5$A$_6$P$_7$D$_8$T*$_9$R$_{10}$P$_{11}$fA$_{12}$P$_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **15** was obtained and purified by semi-preparative HPLC.

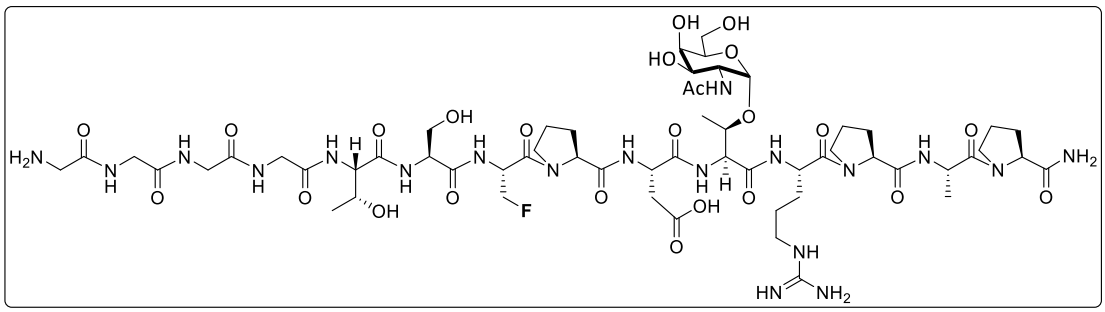
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.84 – 4.90 (m, 1H, H_{1S}), 4.22 – 4.60 (m, 13H, $H\beta_{Thr4}$, $H\beta_{Thr9}$, $2H\beta_{Ala12}$, $H\alpha_{Pro7}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Ala6}$, $H\alpha_{Ala12}$, $H\alpha_{Thr4}$, $H\alpha_{Thr9}$, $H\alpha_{Ser}$, $H\alpha_{Asp}$), 3.54 – 4.13 (m, 23H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\alpha_{Arg}$), 3.14 – 3.25 (m, 2H, $2H\delta_{Arg}$), 2.80 (ddd, 2H, J = 23.7, 17.0, 7.3 Hz, $H\beta_{Asp}$), 2.23 – 2.35 (m, 3H, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$), 1.78 – 2.08 (m, 14H, $2H\beta_{Arg}$, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $2H\gamma_{Pro7}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$, $NHCOCH_3$), 1.65 – 1.73 (m, 2H, $2H\gamma_{Arg}$) 1.35 (d, 3H, J = 7.0, CH_{3Ala6}), 1.16 – 1.26 (m, 6H, CH_{3Thr4}, CH_{3Thr9}).

Semi-preparative HPLC: Rt = 14.2 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 15.1 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{58}H_{96}FN_{19}O_{24}$: 730.8424; Found 730.8421.

Synthesis of glycopeptide 16 ($GGGGT_4S_5fA_6P_7D_8T^*_9R_{10}P_{11}A_{12}P_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **16** was obtained and purified by semi-preparative HPLC.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.92 – 5.05 (m, 1H, H_{1S}), 4.15 – 4.73 (m, 13H, $H\beta_{Thr4}$, $H\beta_{Thr9}$, $2H\beta_{Ala6}$, $H\alpha_{Pro7}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Ala6}$, $H\alpha_{Ala12}$, $H\alpha_{Thr4}$, $H\alpha_{Thr9}$, $H\alpha_{Ser}$, $H\alpha_{Asp}$), 3.45 – 4.16 (m, 23H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\alpha_{Arg}$), 3.14 – 3.27 (m, 2H, $2H\delta_{Arg}$), 2.87 (ddd, 2H, J = 23.6, 17.0, 7.0 Hz, $H\beta_{Asp}$), 2.21 – 2.33 (m, 3H, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$), 1.78 – 2.08 (m, 14H, $2H\beta_{Arg}$,

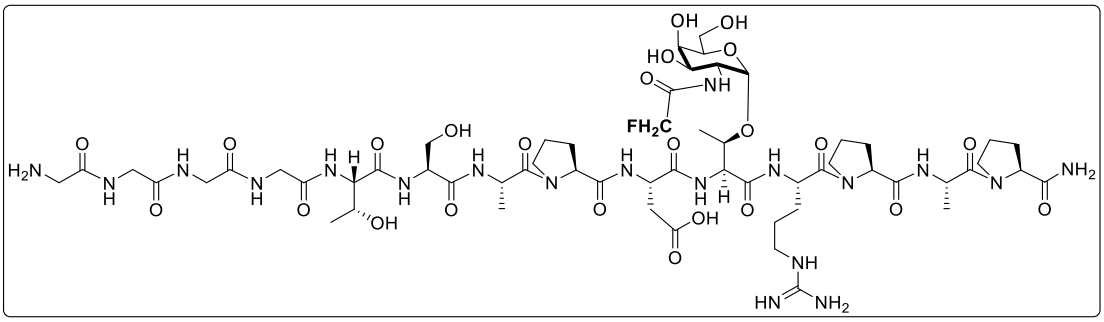
$H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $2H\gamma_{Pro7}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$, $NHCOC\underline{H}_3$), 1.65 – 1.70 (m, 2H, $2H\gamma_{Arg}$) 1.36 (d, 3H, J = 7.0, CH_{3Ala12}), 1.15 – 1.26 (m, 6H, CH_{3Thr4}, CH_{3Thr9}).

Semi-preparative HPLC: Rt = 13.3 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 12.8 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{58}H_{96}FN_{19}O_{24}$: 730.8424; Found 730.8421.

Synthesis of glycopeptide 17 ($GGGGT_4S_5A_6P_7D_8T(GalNFAc)_9R_{10}P_{11}A_{12}P_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **17** was obtained and purified by semi-preparative HPLC.

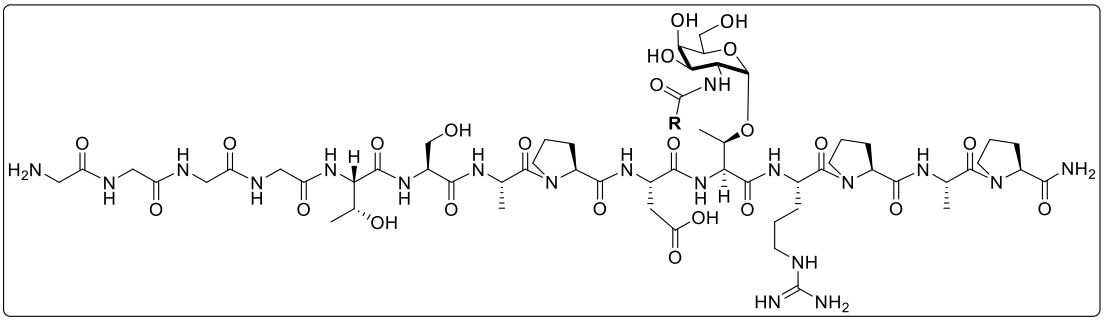
^1H NMR (400 MHz, D_2O) δ (ppm): 4.84 – 5.03 (m, 3H, H_{1S}, $NHCOC\underline{H}_2F$), 4.20 – 4.64 (m, 11H, $H\beta_{Thr4}$, $H\beta_{Thr9}$, $H\alpha_{Pro7}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Ala6}$, $H\alpha_{Ala12}$, $H\alpha_{Thr4}$, $H\alpha_{Thr9}$, $H\alpha_{Ser}$, $H\alpha_{Asp}$), 3.53 – 4.10 (m, 23H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\alpha_{Arg}$), 3.16 – 3.25 (m, 2H, $2H\delta_{Arg}$), 2.89 (ddd, 2H, J = 50.5, 17.0, 7.0 Hz, $2H\beta_{Asp}$), 2.22 – 2.33 (m, 3H, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$), 1.80 – 2.07 (m, 11H, $2H\beta_{Arg}$, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $2H\gamma_{Pro7}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$), 1.63 – 1.71 (m, 2H, $2H\gamma_{Arg}$) 1.32 – 1.39 (m, 6H, CH_{3Ala6}, CH_{3Ala12}), 1.16 – 1.26 (m, 6H, CH_{3Thr4}, CH_{3Thr9}).

Semi-preparative HPLC: Rt = 13.5 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (12.5:87.5), 15 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 14.1 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{58}H_{96}FN_{19}O_{24}$: 730.8424; Found: 730.8427.

Synthesis of glycopeptide 18 (GGGGTSAPDT(GalNCOR)RPAP)

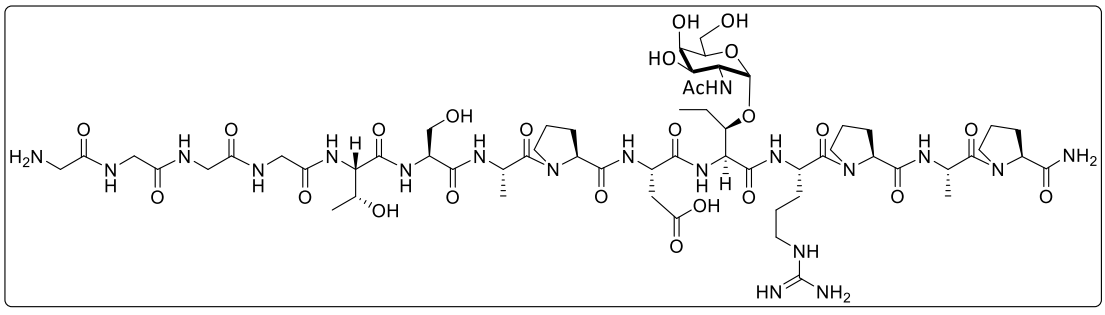


Following SPPS methodology with the adequately protected amino acids, glycopeptide **18** was obtained and purified by semi-preparative HPLC.

Semi-preparative HPLC: Rt = 15.2 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 15.6 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 1 mL/min, λ= 212 nm)

Synthesis of glycopeptide 19 ($GGGGT_4S_5A_6P_7D_8Hnv(GalNAc)_9R_{10}P_{11}A_{12}P_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **19** was obtained and purified by semi-preparative HPLC.

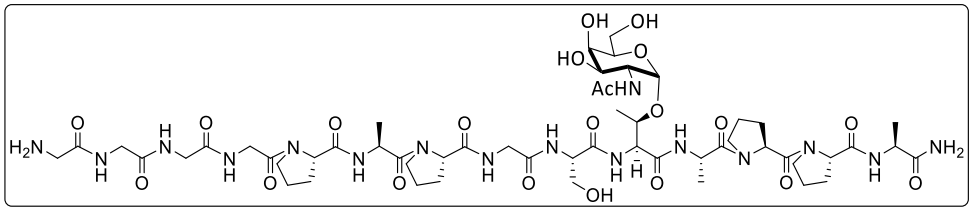
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.83 – 4.87 (m, 1H, H_{1S}), 3.53 – 4.65 (m, 34H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\alpha_{Arg}$, $H\beta_{Thr4}$, $H\beta_{Hnv}$, $H\alpha_{Pro7}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Ala6}$, $H\alpha_{Ala12}$, $H\alpha_{Thr4}$, $H\alpha_{Hnv}$, $H\alpha_{Ser}$, $H\alpha_{Asp}$), 3.15 – 3.24 (m, 2H, $2H\delta_{Arg}$), 2.91 (ddd, 2H, *J* = 23.7, 17.1, 7.1 Hz, $2H\beta_{Asp}$), 2.22 – 2.34 (m, 3H, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$), 1.80 – 2.08 (m, 18H, $2H\beta_{Arg}$, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $2H\gamma_{Pro7}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$, $2H\gamma_{Hnv}$, $2H\gamma_{Arg}$, $NHCOCH_3$), 1.33 – 1.39 (m, 6H, CH_{3Ala6}, CH_{3Ala12}), 1.19 (d, 3H, *J* = 6.4 Hz, CH_{3Thr4}), 0.83 (t, 3H, *J* = 7.4 Hz, CH_{3Hnv}).

Semi-preparative HPLC: Rt = 15.9 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 16.8 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{61}H_{100}N_{19}O_{25}$: 1498.7132; Found 1498.7102.

Synthesis of glycopeptide 20 ($GGGGP_{11}A_{12}P_{13}G_{14}S_{15}T^*_{16}A_{17}P_{18}P_{19}A_{20}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **20** was obtained and purified by semi-preparative HPLC.

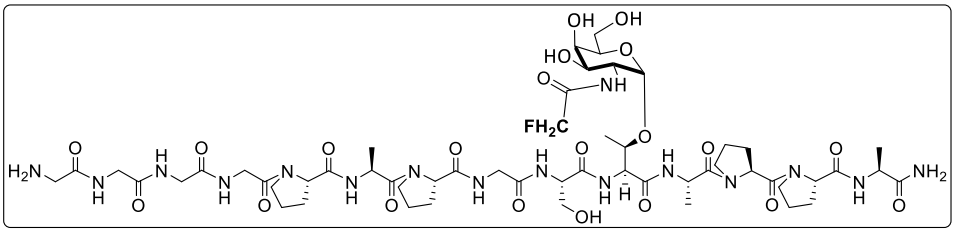
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.91 (d, 1H, *J* = 4.9 Hz, H_{1S}), 3.56 – 4.66 (m, 36H, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H\delta_{Pro18}$, $2H\delta_{Pro19}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\beta_{Thr16}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Pro18}$, $H\alpha_{Pro19}$, $H\alpha_{Ala12}$, $H\alpha_{Ala17}$, $H\alpha_{Ala20}$, $H\alpha_{Thr16}$, $H\alpha_{Ser}$), 2.18 – 2.40 (m, 4H, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $H\beta_{Pro18}$, $H\beta_{Pro19}$), 1.84 – 2.08 (m, 15H, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $H\beta_{Pro18}$, $H\beta_{Pro19}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$, $2H\gamma_{Pro18}$, $2H\gamma_{Pro19}$, $NHCOCH_3$), 1.30 – 1.40 (m, 9H, CH_{3Ala12}, CH_{3Ala17}, CH_{3Ala20}), 1.24 (d, 3H, *J* = 6.4 Hz, CH_{3Thr16}).

Semi-preparative HPLC: Rt = 12.7 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (12.5:87.5), 15 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 15.8 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{54}H_{87}N_{16}O_{21}$: 1295.6226; Found 1295.6184.

Synthesis of glycopeptide 21 ($GGGGP_{11}A_{12}P_{13}G_{14}S_{15}T(GalNFAc)_{16}A_{17}P_{18}P_{19}A_{20}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **21** was obtained and purified by semi-preparative HPLC.

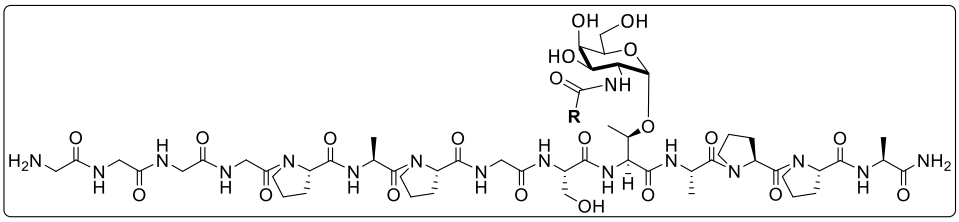
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.82 – 4.98 (m, 3H, H_{1S}, NHCOC$\underline{H}_2$F), 3.52 – 4.69 (m, 36H, 2Hδ_{Pro11}, 2Hδ_{Pro13}, 2Hδ_{Pro18}, 2Hδ_{Pro19}, 2H$_{6S}$, H$_{2S}$, H$_{3S}$, H$_{4S}$, H$_{5S}$, 2Hα_{Gly}, 2Hβ_{Ser}, Hβ_{Thr16}, Hα_{Pro11}, Hα_{Pro13}, Hα_{Pro18}, Hα_{Pro19}, Hα_{Ala12}, Hα_{Ala17}, Hα_{Ala20}, Hα_{Thr16}, Hα_{Ser}), 2.16 – 2.44 (m, 4H, Hβ_{Pro11}, Hβ_{Pro13}, Hβ_{Pro18}, Hβ_{Pro19}), 1.80 – 2.15 (m, 12H, Hβ_{Pro11}, Hβ_{Pro13}, Hβ_{Pro18}, Hβ_{Pro19}, 2Hγ_{Pro11}, 2Hγ_{Pro13}, 2Hγ_{Pro18}, 2Hγ_{Pro19}), 1.29 – 1.41 (m, 9H, CH_{3Ala12}, CH_{3Ala17}, CH_{3Ala20}), 1.24 (d, 3H, *J* = 6.4 Hz, CH_{3Thr16}).

Semi-preparative HPLC: Rt = 13.5 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (12.5:87.5), 15 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 13.7 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{54}H_{87}N_{16}O_{21}$: 657.3102; Found 657.3101.

Synthesis of glycopeptide 22 (GGGGPAPGST(GalNCOR)APPA)

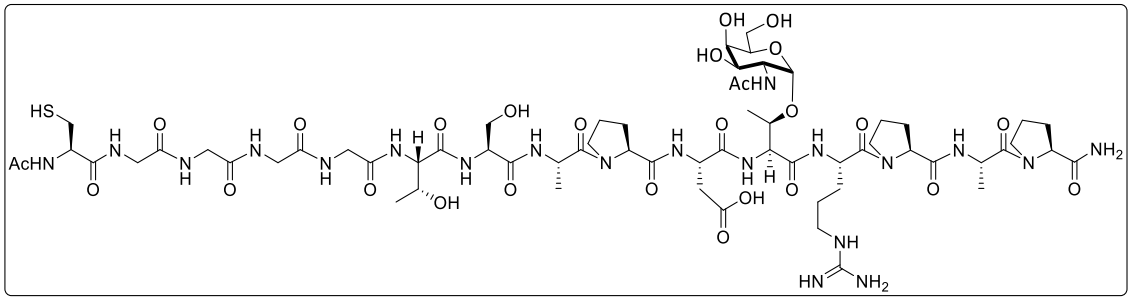


Following SPPS methodology with the adequately protected amino acids, glycopeptide **22** was obtained and purified by semi-preparative HPLC.

Semi-preparative HPLC: Rt = 14.9 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 15.8 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

Synthesis of glycopeptide 23 ($CGGGGT_4S_5A_6P_7D_8T^*_9R_{10}P_{11}A_{12}P_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **23** was obtained and purified by semi-preparative HPLC.

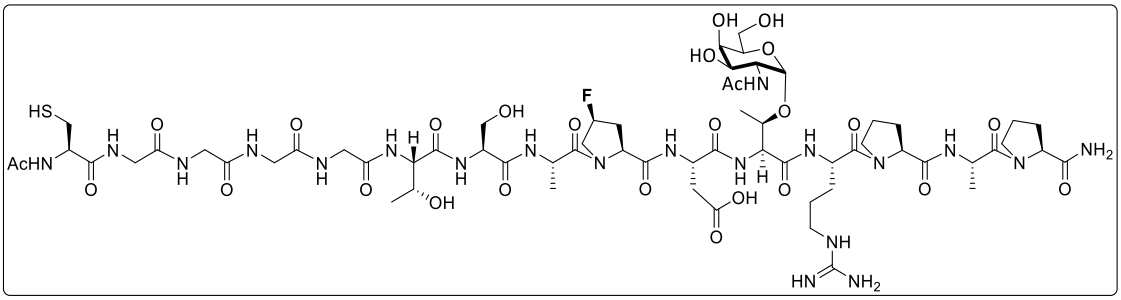
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.82 – 4.85 (m, 1H, H_{1S}), 4.19 – 4.66 (m, 12H, $H\beta_{Thr4}$, $H\beta_{Thr9}$, $H\alpha_{Pro7}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Ala6}$, $H\alpha_{Cys}$, $H\alpha_{Ala12}$, $H\alpha_{Thr4}$, $H\alpha_{Thr9}$, $H\alpha_{Ser}$, $H\alpha_{Asp}$), 3.53 – 4.10 (m, 23H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\alpha_{Arg}$), 3.13 – 3.23 (m, 2H, $2H\delta_{Arg}$), 2.79 – 2.99 (m, 4H, $2H\beta_{Asp}$, $2H\beta_{Cys}$), 2.23 – 2.33 (m, 3H, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$), 1.83 – 2.08 (m, 17H, $2H\beta_{Arg}$, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $2H\gamma_{Pro7}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$, $NHCOCH_3$, $NHCOCH_{3term}$), 1.65 – 1.70 (m, 2H, $2H\gamma_{Arg}$) 1.33 – 1.36 (m, 6H, CH_{3Ala6}, CH_{3Ala12}), 1.16 – 1.23 (m, 6H, CH_{3Thr4}, CH_{3Thr9}).

Semi-preparative HPLC: Rt = 17.2 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 17.9 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{63}H_{104}N_{20}O_{26}S$: 794.3517; Found: 794.3555.

Synthesis of glycopeptide 24 (CGGGGT$_4$S$_5$A$_6$(4*S*)fP$_7$D$_8$T*$_9$R$_{10}$P$_{11}$A$_{12}$P$_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **24** was obtained and purified by semi-preparative HPLC.

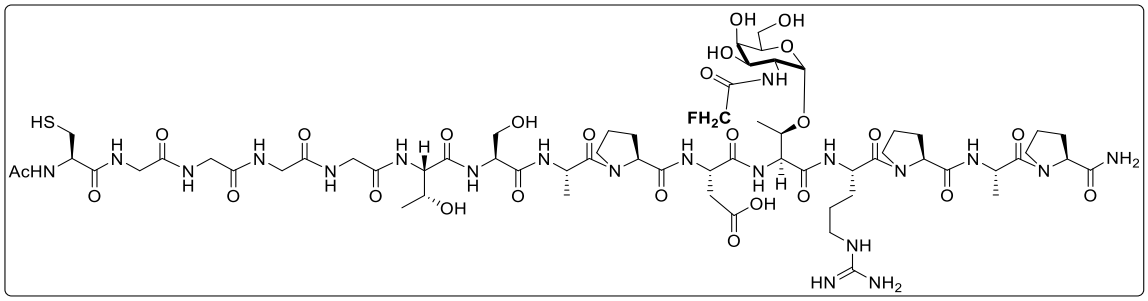
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.84 – 4.89 (m, 1H, H_{1S}), 4.17 – 4.74 (m, 12H, $H\beta_{Thr4}$, $H\beta_{Thr9}$, $H\alpha_{Pro7}$, $H\alpha_{Pro11}$, $H\alpha_{Pro13}$, $H\alpha_{Ala6}$, $H\alpha_{Cys}$, $H\alpha_{Ala12}$, $H\alpha_{Thr4}$, $H\alpha_{Thr9}$, $H\alpha_{Ser}$, $H\alpha_{Asp}$), 3.59 – 4.07 (m, 23H, $2H\delta_{Pro7}$, $2H\delta_{Pro11}$, $2H\delta_{Pro13}$, $2H_{6S}$, H_{2S}, H_{3S}, H_{4S}, H_{5S}, $2H\alpha_{Gly}$, $2H\beta_{Ser}$, $H\alpha_{Arg}$), 3.15 – 3.23 (m, 2H, $2H\delta_{Arg}$), 2.80 – 3.00 (m, 4H, $2H\beta_{Asp}$, $2H\beta_{Cys}$), 2.25 – 2.57 (m, 3H, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$), 1.80 – 2.08 (m, 18H, $2H\beta_{Arg}$, $H\beta_{Pro7}$, $H\beta_{Pro11}$, $H\beta_{Pro13}$, $H\gamma_{Pro7}$, $2H\gamma_{Pro11}$, $2H\gamma_{Pro13}$, NHCOC$\underline{H}_3$, NHCOC$\underline{H}_{3term}$, $2H\gamma_{Arg}$) 1.34 – 1.43 (m, 6H, CH_{3Ala6}, CH_{3Ala12}), 1.17 – 1.26 (m, 6H, CH_{3Thr4}, CH_{3Thr9}).

Semi-preparative HPLC: Rt = 17.1 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 17.7 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{63}H_{103}FN_{20}O_{26}S$: 803.3523; Found: 803.3517.

Synthesis of glycopeptide 25 (CGGGGT$_4$S$_5$A$_6$P$_7$D$_8$T(GalNFAc)$_9$R$_{10}$P$_{11}$A$_{12}$P$_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **25** was obtained and purified by semi-preparative HPLC.

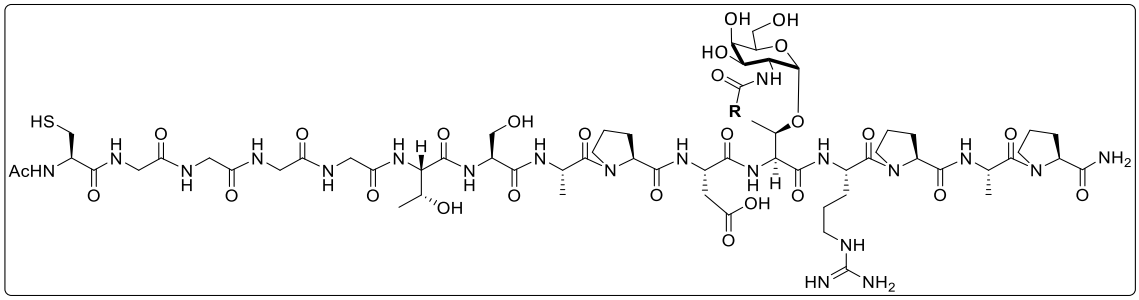
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.88 – 5.04 (m, 3H, H_{1S}, NHCOC$\underline{H}_2$F), 4.15 – 4.59 (m, 12H, Hβ$_{Thr4}$, Hβ$_{Thr9}$, Hα$_{Pro7}$, Hα$_{Pro11}$, Hα$_{Pro13}$, Hα$_{Ala6}$, Hα$_{Cys}$, Hα$_{Ala12}$, Hα$_{Thr4}$, Hα$_{Thr9}$, Hα$_{Ser}$, Hα$_{Asp}$), 3.51 – 4.05 (m, 23H, 2Hδ$_{Pro7}$, 2Hδ$_{Pro11}$, 2Hδ$_{Pro13}$, 2H$_{6S}$, H$_{2S}$, H$_{3S}$, H$_{4S}$, H$_{5S}$, 2Hα$_{Gly}$, 2Hβ$_{Ser}$, Hα$_{Arg}$), 3.16 – 3.23 (m, 2H, 2Hδ$_{Arg}$), 2.73 – 2.98 (m, 4H, 2Hβ$_{Asp}$, 2Hβ$_{Cys}$), 2.23 – 2.33 (m, 3H, Hβ$_{Pro7}$, Hβ$_{Pro11}$, Hβ$_{Pro13}$), 1.80 – 2.07 (m, 14H, 2Hβ$_{Arg}$, Hβ$_{Pro7}$, Hβ$_{Pro11}$, Hβ$_{Pro13}$, 2Hγ$_{Pro7}$, 2Hγ$_{Pro11}$, 2Hγ$_{Pro13}$, NHCOC$\underline{H}_{3term}$), 1.65 – 1.71 (m, 2H, 2Hγ$_{Arg}$) 1.33 – 1.40 (m, 6H, CH_{3Ala6}, CH_{3Ala12}), 1.16 – 1.27 (m, 6H, CH_{3Thr4}, CH_{3Thr9}).

Semi-preparative HPLC: Rt = 17.6 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 17.9 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{63}H_{102}FN_{20}O_{26}S$: 1605.6973; Found: 1605.6922.

Synthesis of glycopeptide 26 (CGGGGTSAPDT(GalNCOR)RPAP)

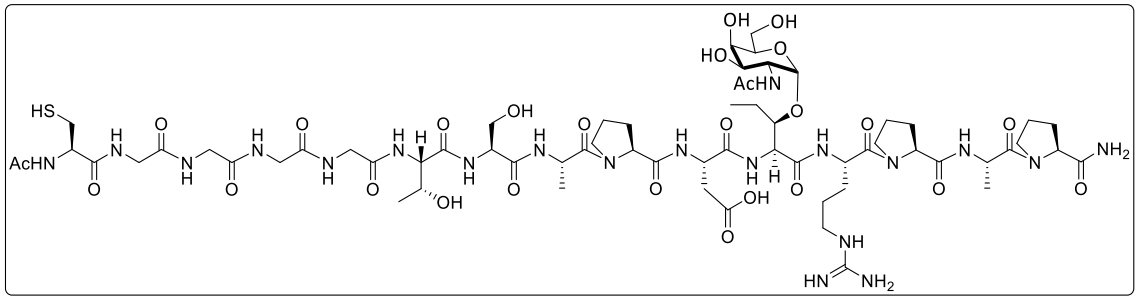


Following SPPS methodology with the adequately protected amino acids, glycopeptide **26** was obtained and purified by semi-preparative HPLC.

Semi-preparative HPLC: Rt = 19.0 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (16:84), 22 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 19.6 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 1 mL/min, λ= 212 nm)

Synthesis of glycopeptide 27 (CGGGGT$_4$S$_5$A$_6$P$_7$D$_8$Hnv*$_9$R$_{10}$P$_{11}$A$_{12}$P$_{13}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **27** was obtained and purified by semi-preparative HPLC.

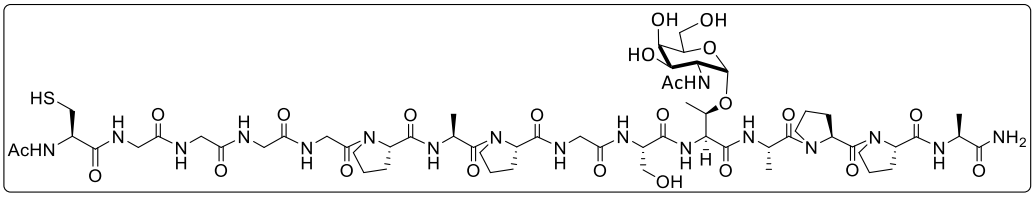
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.85 – 4.87 (m, 1H, H_{1S}), 3.60 – 4.41 (m, 35H, 2Hδ$_{Pro7}$, 2Hδ$_{Pro11}$, 2Hδ$_{Pro13}$, 2H$_{6S}$, H$_{2S}$, H$_{3S}$, H$_{4S}$, H$_{5S}$, 2Hα$_{Gly}$, 2Hβ$_{Ser}$, Hα$_{Arg}$, Hβ$_{Thr4}$, Hβ$_{Hnv}$, Hα$_{Pro7}$, Hα$_{Pro11}$, Hα$_{Pro13}$, Hα$_{Ala6}$, Hα$_{Cys}$, Hα$_{Ala12}$, Hα$_{Thr4}$, Hα$_{Hnv}$, Hα$_{Ser}$, Hα$_{Asp}$), 3.17 – 3.24 (m, 2H, 2Hδ$_{Arg}$), 2.77 – 3.02 (m, 4H, 2Hβ$_{Asp}$, 2Hβ$_{Cys}$), 2.24 – 2.32 (m, 3H, Hβ$_{Pro7}$, Hβ$_{Pro11}$, Hβ$_{Pro13}$), 1.69 – 2.07 (m, 21H, 2Hβ$_{Arg}$, Hβ$_{Pro7}$, Hβ$_{Pro11}$, Hβ$_{Pro13}$, 2Hγ$_{Pro7}$, 2Hγ$_{Pro11}$, 2Hγ$_{Pro13}$, 2Hγ$_{Hnv}$, 2Hγ$_{Arg}$, NHCOC$\underline{H}_3$, NHCOC$\underline{H}_{3term}$), 1.32 – 1.38 (m, 6H, CH$_{3Ala6}$, CH$_{3Ala12}$), 1.19 (d, 3H, *J* = 6.4 Hz, CH$_{3Thr4}$), 0.83 (t, 3H, *J* = 7.4 Hz, CH$_{3Hnv}$).

Semi-preparative HPLC: Rt = 18.5 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 19.0 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{64}H_{106}N_{20}O_{26}S$: 801.3648; Found 801.3656.

Synthesis of glycopeptide 28 (CGGGGP$_{11}$A$_{12}$P$_{13}$G$_{14}$S$_{15}$T*$_{16}$A$_{17}$P$_{18}$P$_{19}$A$_{20}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **28** was obtained and purified by semi-preparative HPLC.

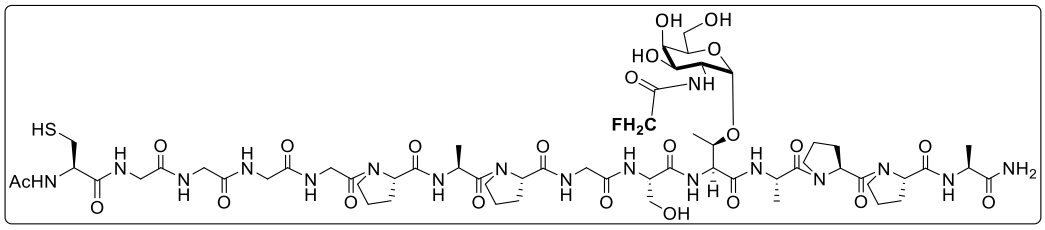
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.91 (d, 1H, *J* = 3.8 Hz, H_{1S}), 3.54 – 4.66 (m, 37H, 2Hδ$_{Pro11}$, 2Hδ$_{Pro13}$, 2Hδ$_{Pro18}$, 2Hδ$_{Pro19}$, 2H$_{6S}$, H$_{2S}$, H$_{3S}$, H$_{4S}$, H$_{5S}$, 2Hα$_{Gly}$, 2Hβ$_{Ser}$, Hβ$_{Thr16}$, Hα$_{Pro11}$, Hα$_{Pro13}$, Hα$_{Pro18}$, Hα$_{Pro19}$, Hα$_{Ala12}$, Hα$_{Ala17}$, Hα$_{Ala20}$, Hα$_{Cys}$,Hα$_{Thr16}$, Hα$_{Ser}$), 2.91 (d, 2H, *J* = 6.1 Hz, 2Hβ$_{Cys}$), 2.14 – 2.39 (m, 4H, Hβ$_{Pro11}$, Hβ$_{Pro13}$, Hβ$_{Pro18}$, Hβ$_{Pro19}$), 1.83 – 2.09 (m, 18H, Hβ$_{Pro11}$, Hβ$_{Pro13}$, Hβ$_{Pro18}$, Hβ$_{Pro19}$, 2Hγ$_{Pro11}$, 2Hγ$_{Pro13}$, 2Hγ$_{Pro18}$, 2Hγ$_{Pro19}$, NHCOC<u>H</u>$_3$, NHCOC<u>H</u>$_{3term}$), 1.29 – 1.40 (m, 9H, CH$_{3Ala12}$, CH$_{3Ala17}$, CH$_{3Ala20}$), 1.24 (d, 3H, *J* = 6.4 Hz, CH$_{3Thr16}$).

Semi-preparative HPLC: Rt = 17.4 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 18.0 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{59}H_{94}N_{17}O_{23}S$: 1440.6424; Found 1440.6399.

Synthesis of glycopeptide 29 ($CGGGGP_{11}A_{12}P_{13}G_{14}S_{15}T(GalNFAc)_{16}A_{17}P_{18}P_{19}A_{20}$)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **29** was obtained and purified by semi-preparative HPLC.

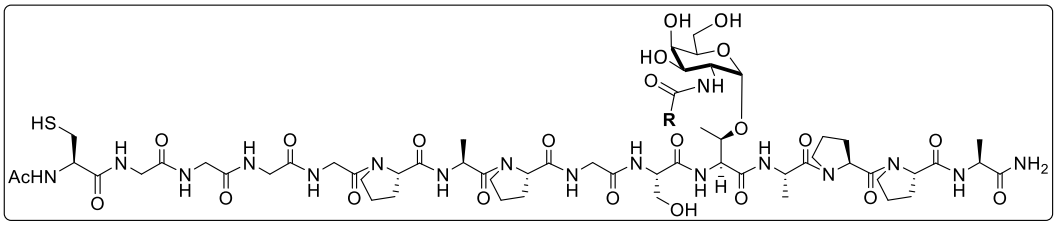
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.83 – 4.98 (m, 3H, H_{1S}, NHCOC$\underline{H}_2$F), 3.56 – 4.70 (m, 37H, 2Hδ_{Pro11}, 2Hδ_{Pro13}, 2Hδ_{Pro18}, 2Hδ_{Pro19}, 2H_{6S}, H_{2S}, H_{3S}, H_{4S}, H_{5S}, 2Hα_{Gly}, 2Hβ_{Ser}, Hβ_{Thr16}, Hα_{Pro11}, Hα_{Pro13}, Hα_{Pro18}, Hα_{Pro19}, Hα_{Ala12}, Hα_{Ala17}, Hα_{Ala20}, Hα_{Cys}, Hα_{Thr16}, Hα_{Ser}), 2.92 (d, 2H, *J* = 6.1 Hz, 2Hβ_{Cys}), 2.16 – 2.38 (m, 4H, Hβ_{Pro11}, Hβ_{Pro13}, Hβ_{Pro18}, Hβ_{Pro19}), 1.82 – 2.09 (m, 15H, Hβ_{Pro11}, Hβ_{Pro13}, Hβ_{Pro18}, Hβ_{Pro19}, 2Hγ_{Pro11}, 2Hγ_{Pro13}, 2Hγ_{Pro18}, 2Hγ_{Pro19}, NHCOC$\underline{H}_{3term}$), 1.27 – 1.39 (m, 9H, CH_{3Ala12}, CH_{3Ala17}, CH_{3Ala20}), 1.24 (d, 3H, *J* = 6.4 Hz, CH_{3Thr16}).

Semi-preparative HPLC: Rt = 17.8 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (15:85), 20 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 18.5 min (Phenomenex Luna C18 (2), 5 µm 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{59}H_{93}FN_{17}O_{23}S$: 1458.6329; Found 1458.6280.

Synthesis of glycopeptide 30 (CGGGGPAPGST(GalNCOR)APPA)

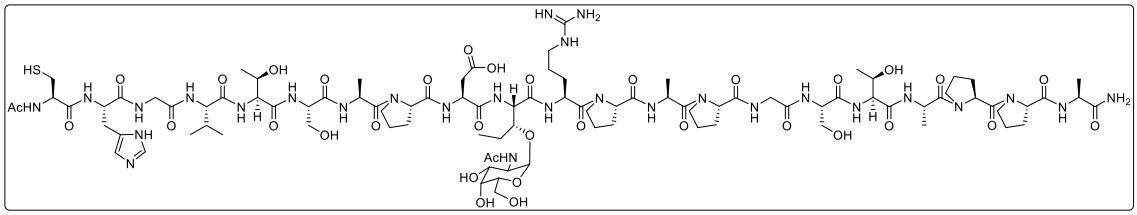


Following SPPS methodology with the adequately protected amino acids, glycopeptide **30** was obtained and purified by semi-preparative HPLC.

Semi-preparative HPLC: Rt = 19.1 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (16:84), 22 min, 20 mL/min, λ= 212 nm)

Analytical HPLC: Rt = 19.9 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 1 mL/min, λ= 212 nm)

Synthesis of glycopeptide 31 (CHGVTSAPDHnv(GalNAc)RPAPGSTAPPA)



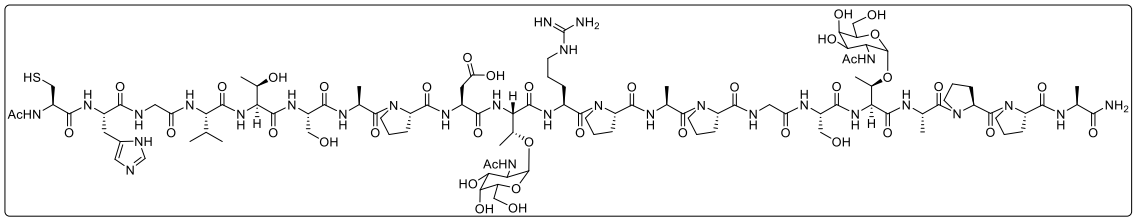
Following SPPS methodology with the adequately protected amino acids, glycopeptide **31** was obtained and purified by semi-preparative HPLC.

Semi-preparative HPLC: Rt = 13.6 min (Phenomenex Luna C18 (2), 10 µ, 21.2x250mm, Grad: acetonitrile/water+0.1% TFA (8:92) → (25:75), 20 min, 10 mL/min, λ= 212 nm).

Analytical RP-HPLC: Rt = 11.2 min (Phenomenex Luna C18 (2), 5µm, 4.6x250mm, Grad: acetonitrile/water+0.1% TFA (8:92) → (25:75), 15 min, 1 mL/min, λ = 212 nm).

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{94}H_{151}N_{28}O_{34}S$: 2248.0663; Found 2247.7265.

Synthesis of glycopeptide 32 (CHGVTSAPDT*RPAPGST*APPA)



Following SPPS methodology with the adequately protected amino acids, glycopeptide **32** was obtained and purified by semi-preparative HPLC.

Semi-preparative HPLC: Rt = 21.5 min (Phenomenex Luna C18 (2), 10 μ, 21.2x250mm, Grad: acetonitrile/water+0.1% TFA (5:95) → (20:80), 30 min, 10 mL/min, λ= 212 nm).

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{101}H_{163}N_{29}O_{39}S$: 1219.0686; Found 1219.0653.

5.5.3. Binding studies by Surface Plasmon Resonance (SPR)

SPR experiments were performed in the Instituto de Química Médica (Madrid) with a Biacore X-100 apparatus (Biacore, GE) in HBS-EP buffer at pH 7.5 (Hepes 10 mM, NaCl 150 mM, EDTA 3 mM, with 2% DMSO and 0.05% Tween X100 as the running buffer at 25 ºC. The antibodies were immobilized on a CM5 sensor chip (Biacore, GE) following standard amine coupling method.[47] Briefly, the carboxymethyl dextran surface of the flow cell 2 was activated with a 7-min injection of a 1:1 ratio of aqueous 0.4 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 0.1 M sulfo-*N*-hydroxysuccinimide. Then, the antibodies were coupled to the surface during a 7-min injection using several dilutions in 10 mM sodium acetate, pH 4.0. The unreacted active esters on the surface were quenched by a 7-min injection of aqueous 0.1 M ethanolamine-HCl (pH 8.0). The levels of immobilization were around 5500 resonance units (RU). Flow cell 1 treated as a flow cell 2 (amine coupling procedure) without protein was used as a reference. Prior to use, 50 mM stock solutions of peptide ligand were diluted to the final concentration in the running buffer. Typically, a series of different compounds were injected onto the sensor chip a flow rate of 30 μl/min for a period of 1 min followed by a dissociation period of 1 min. No regeneration was needed. Sensogram data were double-referenced using the Biaevaluation X100 software (Biacore, GE). The experimental data of affinity measurements were fitted to a two site-specific model binding using Prism software (Figures 5.17-5.26)

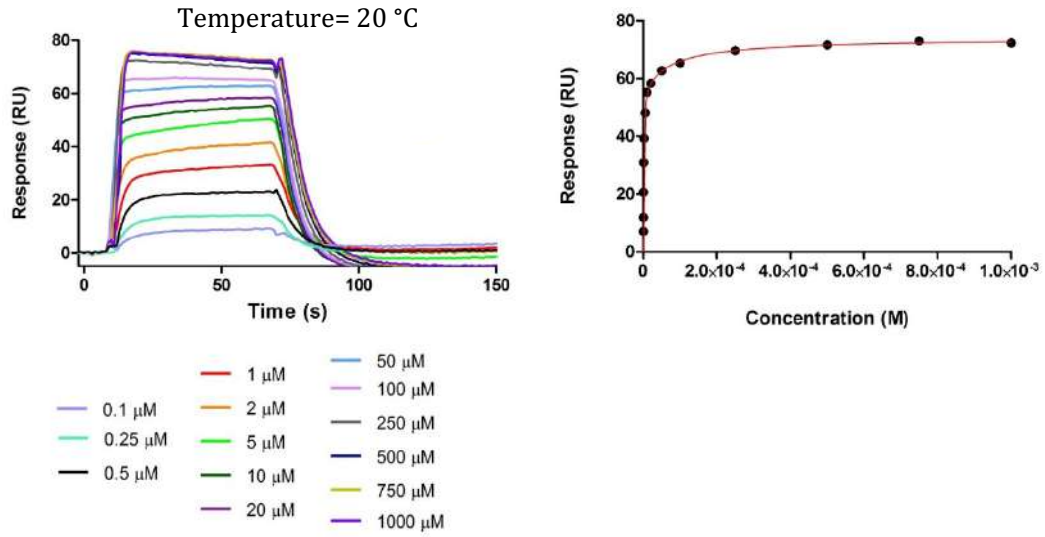


Figure 5.17. SPR curves and response-concentration fit obtained for the binding of glycopeptide **13** towards SM3 antibody at 20 °C.

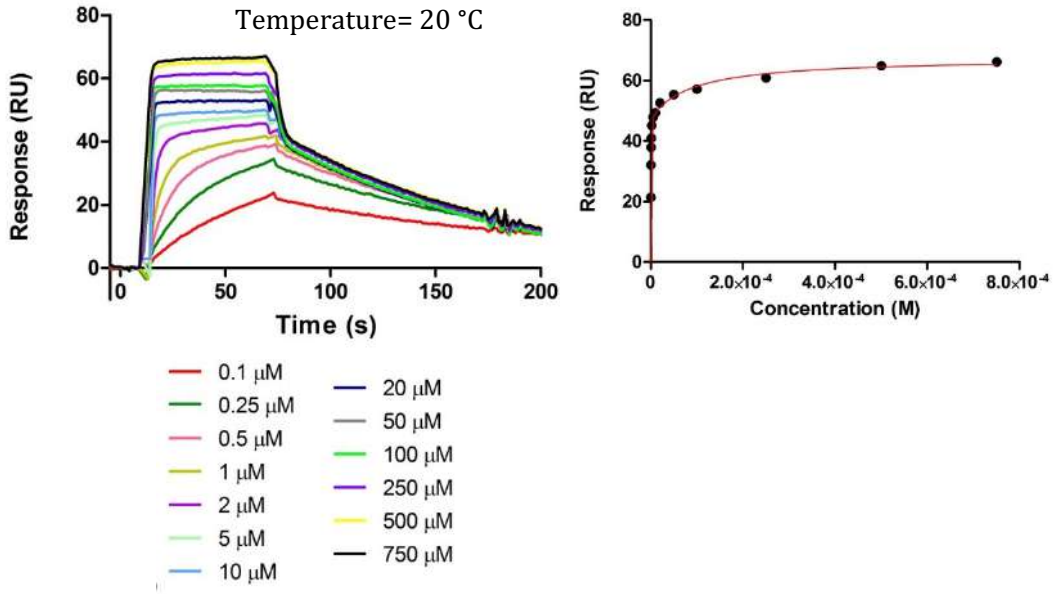


Figure 5.18. SPR curves and response-concentration fit obtained for the binding of glycopeptide **14** towards SM3 antibody at 20 °C.

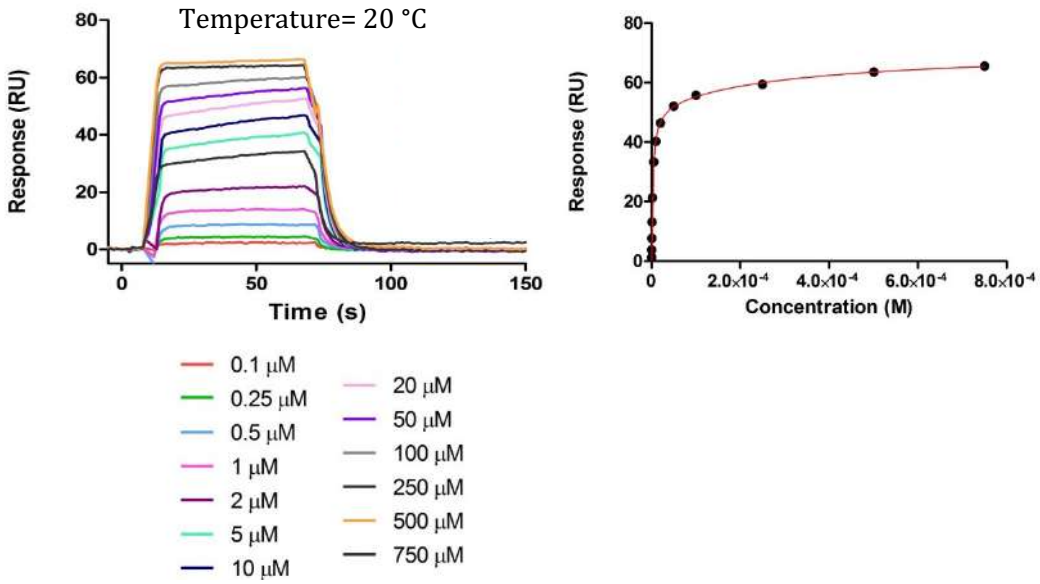


Figure 5.19. SPR curves and response-concentration fit obtained for the binding of glycopeptide **15** towards SM3 antibody at 20 °C.

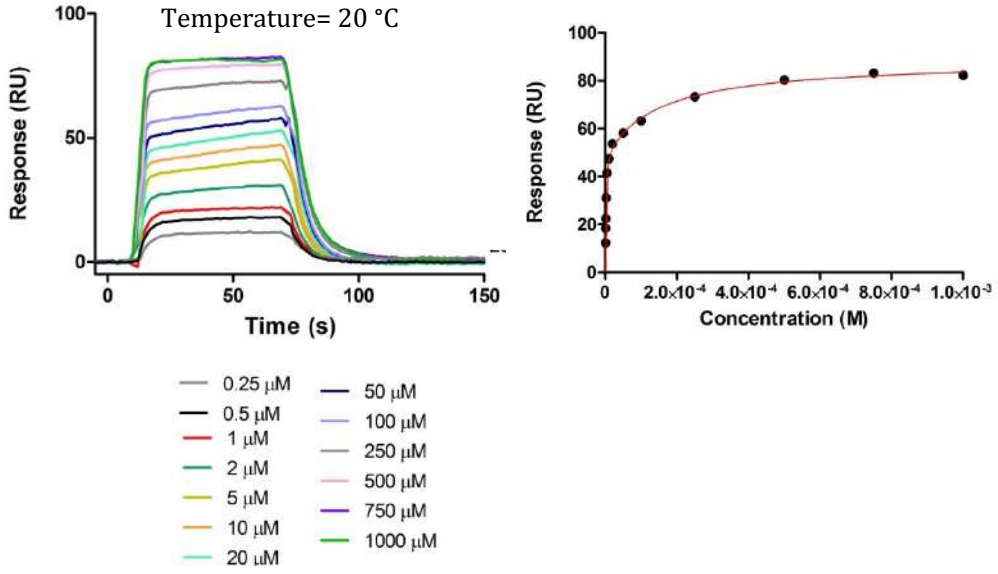


Figure 5.20. SPR curves and response-concentration fit obtained for the binding of glycopeptide **16** towards SM3 antibody at 20 °C.

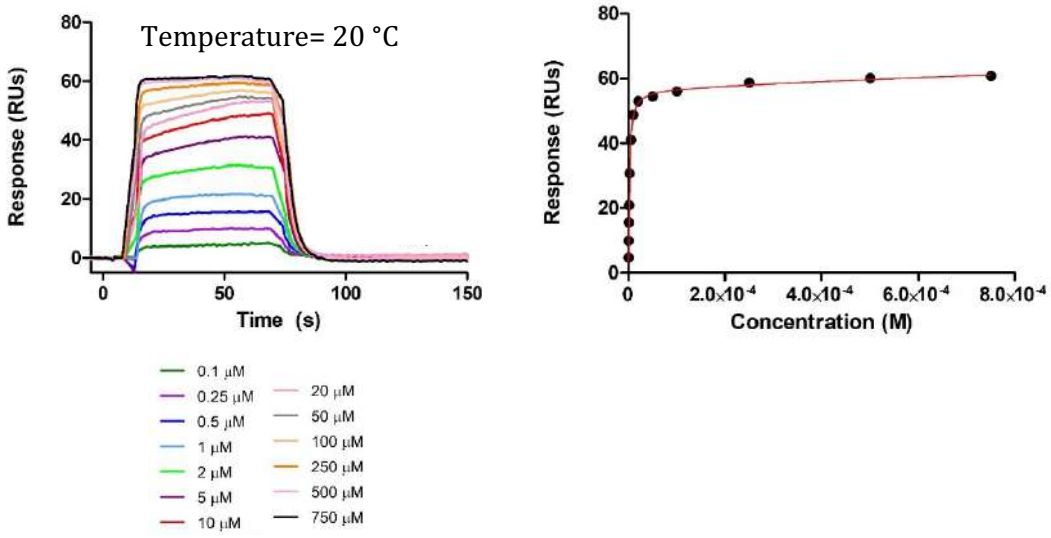


Figure 5.21. SPR curves and response-concentration fit obtained for the binding of glycopeptide **17** towards SM3 antibody at 20 °C.

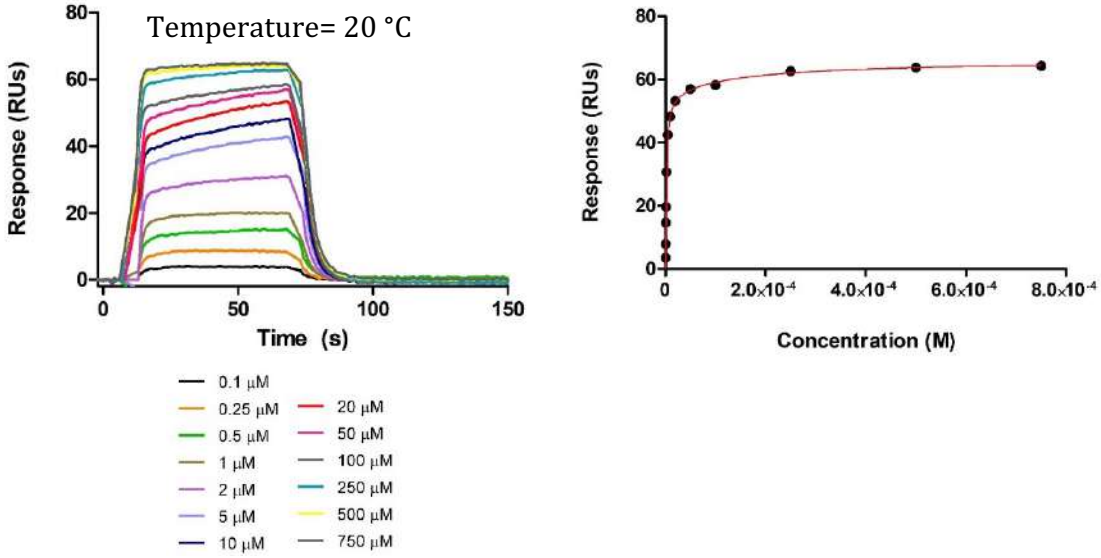


Figure 5.22. SPR curves and response-concentration fit obtained for the binding of glycopeptide **18** towards SM3 antibody at 20 °C.

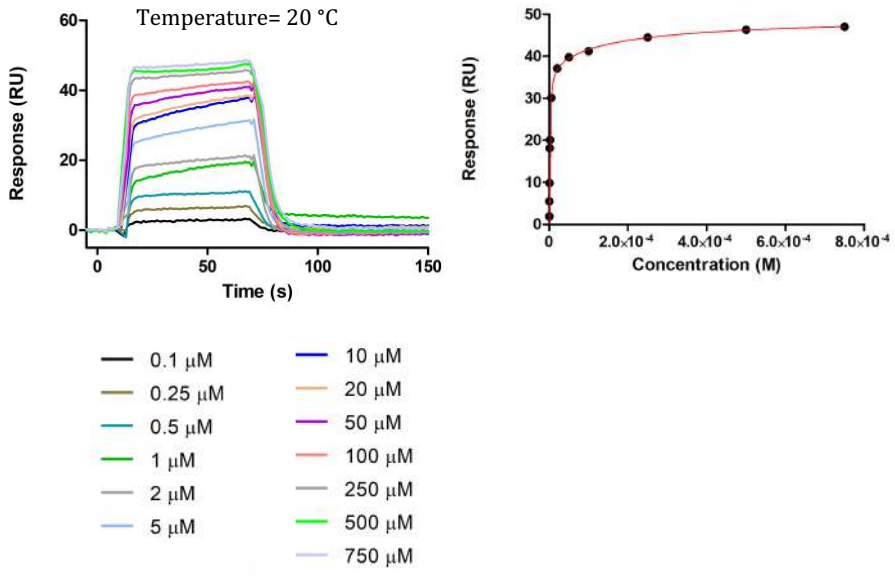


Figure 5.23. SPR curves and response-concentration fit obtained for the binding of glycopeptide **19** towards SM3 antibody at 20 °C.

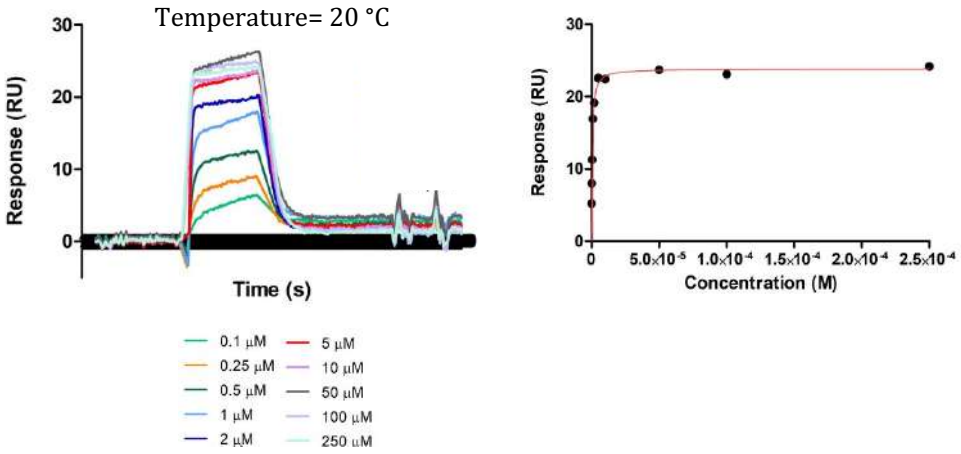


Figure 5.24. SPR curves and response-concentration fit obtained for the binding of glycopeptide **20** towards 5E5 antibody at 20 °C.

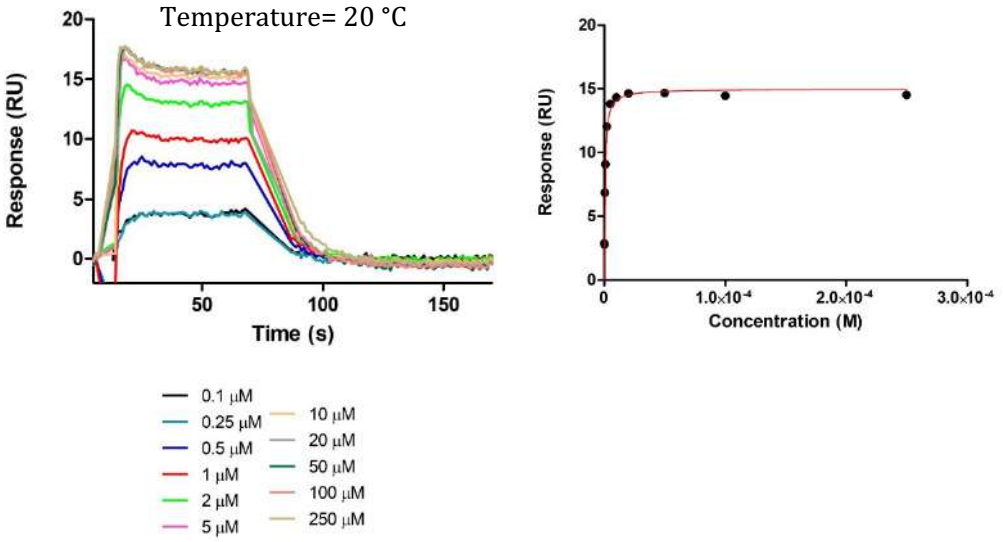


Figure 5.25. SPR curves and response-concentration fit obtained for the binding of glycopeptide **21** towards 5E5 antibody at 20 °C.

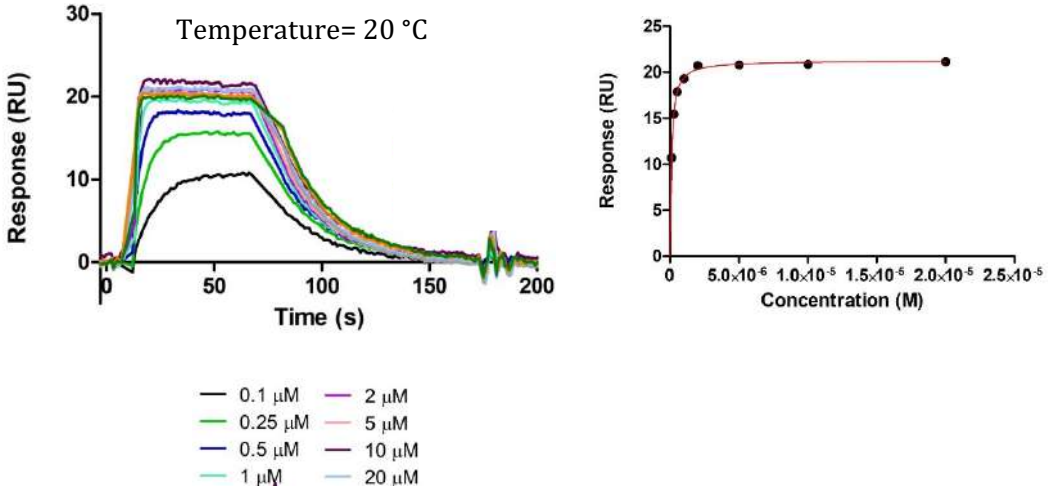


Figure 5.26. SPR curves and response-concentration fit obtained for the binding of glycopeptide **22** towards 5E5 antibody at 20 °C.

Table 5.1. Thermodynamic binding parameters obtained by SPR for SM3 with several glycopeptides at 25 °C and pH 7.5.

Compound	Peptide	K_{D1} (μM)	$Rmax_1$ (RU)	K_{D2} (μM)	$Rmax_2$ (RU)	Chi^2 (RU^2)
13	GGGGTSAPDT*RPAP	80.8	17	0.892	57	0.63
14	GGGGTSA(4*S*)fPDT*RPAP	90.8	19	0.128	48	1.07
15	GGGGTSAPDT*RPfAP	326	19	2.96	52	0.14
16	GGGGTSfAPDT*RPAP	153	39	1.03	49	3.40
17	GGGGTSAPDT(GalNFAc)RPAP	2.19	55	–	-	2
18	GGGGTSAPDT(GalNCOR)RPAP	142	11	1.62	55	0.8
19	GGGGTSAPDHnv(GalNAc)RPAP	175	11	1.46	38	-

Table 5.2. Thermodynamic binding parameters obtained by SPR for 5E5 with several glycopeptides at 25 °C and pH 7.5.

Compound	Peptide	K_D (μM)	Rmax (RU)	Chi^2 (RU^2)
20	GGGGPAPGST*APPA	0.54	23	0.5
21	GGGGPAPGST(GalNFAc)APPA	0.62	15	0.49
22	GGGGPAPGST(GalNCOR)APPA	0.083	23	0.04

5.5.4. Nanoparticle-based dot-blot assay

Synthesis of nanoparticles and dot-blot assays were performed during the secondment at Dr. Fiammengo's group in the University of Verona (Italy).

All glassware employed for nanoparticle preparation was cleaned with aqua regia (HCl (37%)/HNO_3 (65%) 3:1). Ultrapure deionized water (Millipore Elix® 35 water purification system, 18.2 MΩ cm) was used for the preparation of all

aqueous solutions. All solutions used for nanoparticle preparation were filtered through a 0.2 µm membrane filter (cellulose acetate, Whatman®). AuNPs were purified by ultrafiltration (Amicon Ultra-4 or Amicon Ultra-15 Centrifugal Filter Units, regenerated cellulose – 100 kDa from Millipore) and gel filtration (Illustra NAP-10 columns from GE Healthcare). AuNPs were characterized by Dynamic Light Scattering (DLS) (Zetasizer, Malvern Instruments). AuNP core diameter was measured via TEM (JEOL JEM-1011 transmission electron microscope operating at an accelerating voltage of 100 kV). UV/Vis measurements were carried out using a TECAN Infinite M200 Pro plate reader. The concentration of gold was determined via inductively coupled plasma-optical emission spectrometry (Agilent 720 ICP-OES)

Citrate coated AuNPs with a diameter of 13.0 ± 1.0 nm were synthesized according to a literature procedure.[41] AuNPs were passivated by formation of self-assembled monolayers of alkyl-PEG600 thiols on their surface.[42] A mixture of carboxy- and amino-terminated PEG600-thiols (I and II) was used with the molar fraction of amino terminated derivative being x_{NH2} = 0.06. Stock solutions of thiols in EtOH (5 mM) were freshly prepared and directly used. Final concentrations in the passivation reaction were: 80 – 90 nM AuNPs, 25 mM $NaHCO_3$, 1 mM thiols (total) and therefore the reaction contained 20% v/v of EtOH. Stirring at room temperature was continued for 96 h after which PEG-AuNPs were purified by ultrafiltration in Amicon Ultra-15 centrifugal filters (1×15 mL 25 mM $NaHCO_3$, 1×15 mL 2:8 v/v EtOH/50 mM $NaHCO_3$ and 2×15 mL 50 mM $NaHCO_3$ buffer). Purified, PEG-AuNPs were taken up in H_2O at a final concentration of 332 nM.

Peptides previously described were coupled to PEG-AuNPs accordingly to a previously reported procedure.[48] Shortly, $SM(PEG)_2$ linker (2.1 mg, final conc. 5 mM) was added to 350 µL of an approx. 332 nM aqueous solution of PEG-AuNPs in 30 mM phosphate buffer at pH 8.2. The reaction mixture was shaken 4 h at 1 °C. Linker-AuNPs were then purified from excess linker via ultrafiltration (2 × 4 mL 2:8 EtOH/20 mM phosphate buffer pH 8.2 and 2 × 4 mL H_2O) performing all steps on ice, in a 4 °C-cooled centrifuge, and with ice-cold wash solutions. Purified linker-AuNPs were taken up in water at a concentration of approx. 398 nM. In parallel, 200 µL of 1 mM solution in water of each peptide were incubated with 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) for 2 h at 0 °C to fully reduce any possible symmetric disulfide. For the coupling reaction, the

necessary amount of freshly TCEP-treated peptide was added to a freshly prepared cold solution of linker-AuNPs and shaken overnight at 4 °C, 500 rpm. The typical reaction volume was 500 µL and the final concentrations in the reaction mixture were: linker-AuNPs ~100 nM, 15 mM phosphate buffer pH 7.0, 50 mM of peptide. Excess uncoupled peptide was removed by centrifugation (20800 ×g, 45 min, 18 °C), removal of the supernatant, and resuspension in 1 mL 25 mM $NaHCO_3$ twice with 1 mL 10 mM $NaHCO_3$. Peptide-coupled AuNP were taken up in 350 µL 10 mM $NaHCO_3$ to reach a final concentration between 100 and 137 nM, which was stored protected from light at +4° C

The assay exploits gold nanoparticles (AuNPs) functionalized with synthetic glycopeptides **23-32**, **8** and **9**. The peptides were coupled to PEG-AuNPs accordingly to a previously reported procedure.[48] AuNPs were spotted in triplicate with a capillary onto nitrocellulose membranes (16 nM AuNPs) and allowed to dry at room temperature.

For the affinity assays, each membrane was blocked with 5% non-fat milk in tris-buffered saline (TBS) during 1 h, washed in TBS with 0.1% Tween 20 (3×10 min and incubated overnight at room temperature with commercially available anti-MUC1 mouse monoclonal antibody SM3 (sc-53381, aurogene), or with anti-MUC1 mouse monoclonal antibody 5E5 provided by Dr. Clausen (University of Copenhagen). (1 µg/mL to 0.000122 µg/mL in 5% non-fat milk in TBS). After washing with TBS containing 0.1% Tween 20 (3×10 min), the membranes were incubated with goat anti-mouse IgG (H+L) cross-adsorbed Dylight 800 secondary antibody (039610-145-121, tebu-bio), diluted 1:2500 in milk for 1 h at room temperature.

For the sera analysis, each membrane was blocked with 5% BSA in TBS for 1 h, washed in TBS with 0.1% Tween 20 (3×10 min) and incubated overnight at room temperature with patient sera (1:100 dilution in TBS with 1% BSA). After washing with TBS containing 0.1% Tween 20 (3×10 min), the membranes were incubated with rabbit anti-human IgG (H+L) cross-adsorbed Dylight 800 secondary antibody (039609-445-002, tebu-bio), diluted 1:5000 in TBS for 1 h at room temperature.

Finally, membranes were washed in TBS with 0.1% Tween 20 (3×10 min) and with TBS (2x10 min) and visualized using an Oddysey Infrared Imaging System

from LI-COR Biosciences. The intensity of the spots was quantified via densitometry with ImageJ, subtracting the background.

Dynamic light scattering (DLS) and ζ-Potential measurements.

The hydrodynamic radius of AuNPs was determined by dynamic light scattering. AuNPs were suspended in 10 mM $NaHCO_3$. ζ-potentials were measured in the same buffer. The values are reported in Table 5.3.

Table 5.3. Physicochemical characterization of AuNPs used in this work.

	DLS[a]		ζ-Potential[a]
AuNP	Ø(nm)[b]	PDI	(mV)
AuNP-glycopeptide **23**	42 ± 1	0.494	- 35 ± 1
AuNP-glycopeptide **24**	40 ± 1	0.409	- 29 ± 3
AuNP-glycopeptide **25**	44 ± 2	0.528	- 32 ± 1
AuNP-glycopeptide **26**	40 ± 1	0.461	- 28 ± 4
AuNP-glycopeptide **27**	41 ± 1	0.495	- 28 ± 7
AuNP-glycopeptide **28**	39 ± 1	0.365	- 36 ± 5
AuNP-glycopeptide **29**	39 ± 1	0.386	- 31 ± 5
AuNP-glycopeptide **30**	37 ± 1	0.477	- 34 ± 6
AuNP-glycopeptide **8**	39 ± 1	0.462	- 24 ± 4
AuNP-glycopeptide **9**	37 ± 1	0.246	- 18 ± 2
AuNP-glycopeptide **31**	35 ± 1	0.241	- 14 ± 1
AuNP-glycopeptide **32**	21± 1	0.541	- 27± 1

[a] Measurements in 10 mM $NaHCO_3$. [b] Derived from intensity distributions.

Gel electrophoresis

Gel electrophoresis was carried out on 0.6% agarose gel using sodium boric acid (SB) buffer pH 8.5, for approx. 2 h at 70 V.[49] Nanoparticle samples were diluted with loading buffer (1:1 SB buffer/glycerol) to 25 nM AuNPs before loading on the gel (~10 µL/lane).

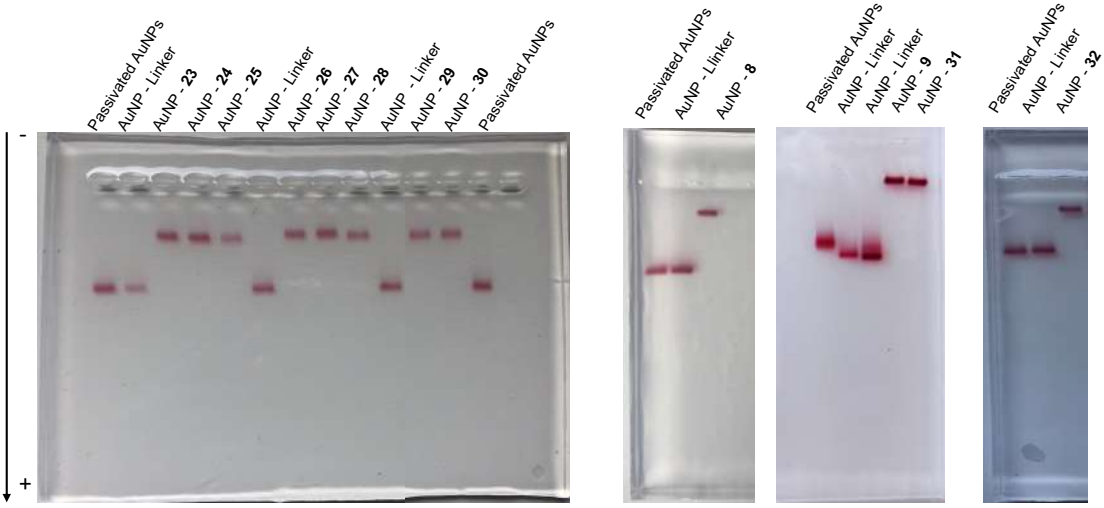


Figure 5.27. Agarose (0.6 %) gel electrophoresis of peptide functionalized AuNPs. Peptide functionalized AuNPs are negatively charged and migrate towards the positive electrode. Successful functionalization is demonstrated by the slower migration of peptide functionalized AuNPs in comparison to control AuNPs.

Membrane dot blot assay with anti-MUC1 antibodies

AuNP-glycopeptides **9**, **23**, **24**, **25**, **26**, **27** and **31** were tested in a dot blot assay to detect the commercially available SM3 antibody. Each membrane was incubated with a decreasing concentration of anti-MUC1 antibody: 1 µg/mL, 0.5 µg/mL, 0.25 µg/mL 0.125 µg/mL, 0.0625 µg/mL, 0.03125 µg/mL, 0.01563 µg/mL, 0.007813 µg/mL, 0.003906 µg/mL, 0.001953 µg/mL, 0.000977 µg/mL, 0.000489 µg/mL, 0.000244 µg/mL, 0.000122 µg/mL.

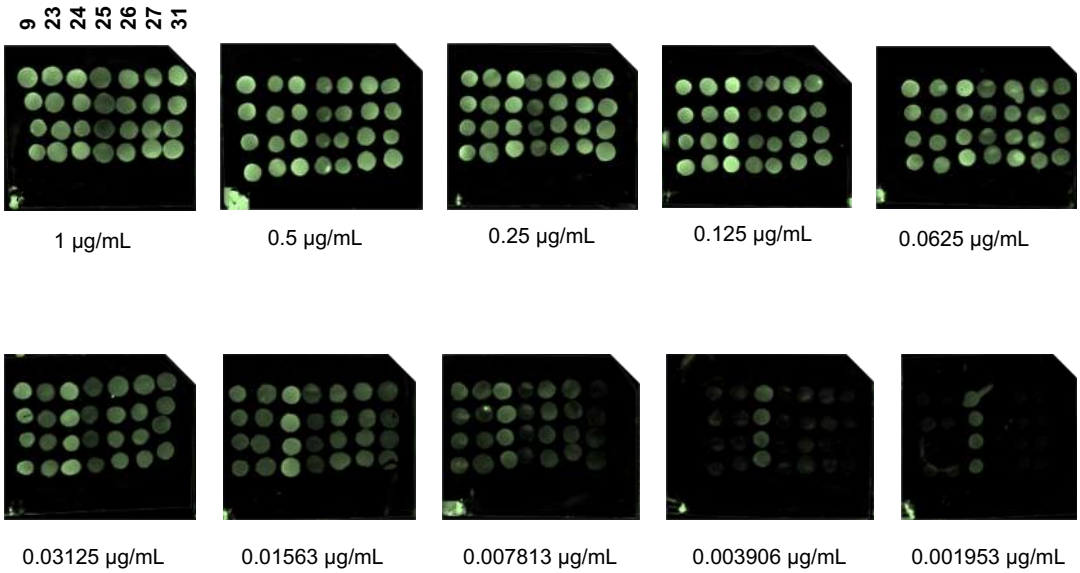


Figure 5.28. Images of the scanned membranes of the assay of AuNP-glycopeptides **9**, **23**, **24**, **25**, **26**, **27** and **31** with a decreased concentration of SM3 antibody.

After preliminary optimization of the procedure, it was observed a fluorescent signal for the membranes incubated with anti-MUC1 antibody concentrations in the range between 1µg/mL and 0.001953 µg/mL. It was not possible to observe any signal for lower concentrations.

Table 5.4. Signal average values after background subtraction with different concentration of SM3 antibody with AuNP-glycopeptides **9**, **23**, **24**, **25**, **26**, **27** and **31.**

Conc.	Glycopeptides						
µg/mL	**9**	**23**	**24**	**25**	**26**	**27**	**31**
1	159.543	153.954	167.595	95.486	119.598	123.216	125.503
0.5	167.683	161.864	187.010	116.866	132.221	154.609	131.439
0.25	126.031	131.074	176.672	97.473	133.740	138.315	126.430
0.125	164.292	175.158	222.639	113.422	132.871	134.312	107.699
0.0625	132.610	125.437	153.473	93.224	130.279	109.695	76.099
0.03125	112.632	88.799	142.694	64.250	99.165	95.685	64.262
0.01563	60.710	70.656	126.279	51.828	74.230	61.079	29.111
0.007813	46.579	37.164	84.729	38.020	63.710	46.716	19.048
0.003906	5.997	23.958	66.062	20.022	35.581	24.003	5.677
0.001953	11.384	18.770	68.724	18.089	31.927	28.621	13.423

Glycopeptides **8**, **28**, **29** and **30** were tested in a dot blot assay to detect 5E5 antibody. Each membrane was incubated with a decreasing concentration of anti-MUC1 antibody: 1 µg/mL, 0.5 µg/mL, 0.25 µg/mL 0.125 µg/mL, 0.0625 µg/mL, 0.03125 µg/mL, 0.01563 µg/mL, 0.007813 µg/mL, 0.003906 µg/mL, 0.001953 µg/mL, 0.000977 µg/mL, 0.000489 µg/mL, 0.000244 µg/mL, 0.000122 µg/mL.

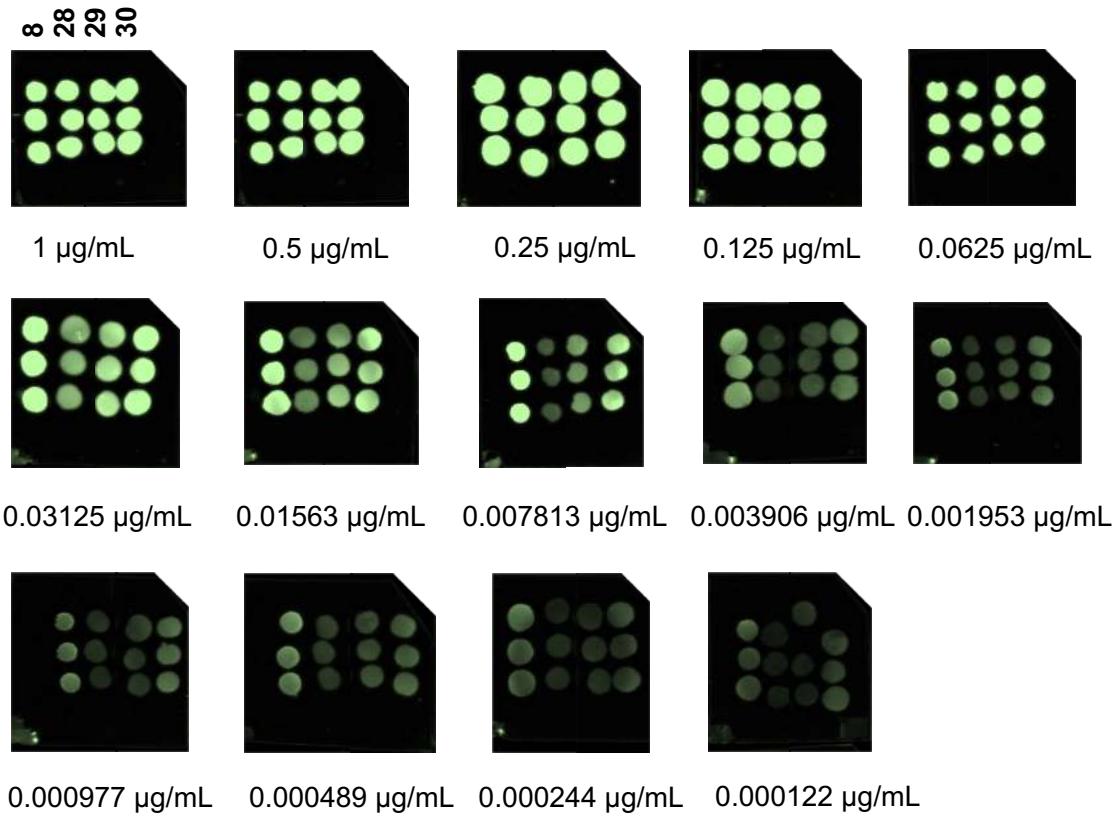


Figure 5.29. Images of the scanned membranes of the assay of glycopeptides **8**, **28**, **29** and **30** with a decreased concentration of 5E5 antibody.

After preliminary optimization of the procedure, it was observed a fluorescent signal for all membranes incubated with anti-MUC1 antibody concentrations in the range between 1 µg/mL and 0.000122 µg/mL.

Table 5.5. Signal average values after background subtraction with different concentration of 5E5 antibody with AuNPs-glycopeptides **8**, **28**, **29** and **30**.

Conc.	**Glycopeptides**			
µg/mL	**8**	**28**	**29**	**30**
1	918334	617267	694504.667	743574.333
0.5	1241627.67	744664.333	752543.667	693057
0.25	755618.667	535193.667	551807	602610
0.125	790563.667	468843.333	574158	621792.333
0.0625	827129	429122	543839.333	685180.333
0.03125	612491.333	266330	336053.667	422979.667
0.01563	421620	188534.667	264073.333	326776
0.007813	421936.333	126404.667	217923.333	312998.333
0.003906	234138.333	84171	107141.333	161246
0.001953	186592.333	77538.3333	105865.667	155945
0.0009766	129740	58990	77369	144126
0.00048883	171168.667	83376.6667	100541	124197.333
0.000244	124824.333	60454.6667	62363.6667	78650.3333
0.0001221	98796	40337	50311	89766.3333

The K_D values obtained by densitometry for the glycopeptides containing SM3 epitope and 5E5 epitope are summarized in in Table 5.6 and Table 5.7, respectively.

Table 5.6. K_D values for AuNP-glycopeptides **9**, **23-27** and **31** with the commercial SM3 antibody using a one-site binding mode.

Comp.	AuNP-glycopeptide	K_D (μg/mL)	K_D(nM)
9	CHGVTSAPDT*RPAPGSTAPPA	0.02097	0.10485
23	CGGGGTSAPDT*RPAP	0.02136	0.10680
24	CGGGGTSA(4*S*)fPDT*RPAP	0.00611	0.03056
25	CGGGGTSAPDT(GalNFAc)RPAP	0.01567	0.07835
26	CGGGGTSAPDT(GalNCOR)RPAP	0.00945	0.04724
27	CGGGGTSAPDHnv(GalNAc)RPAP	0.01710	0.08550
31	CHGVTSAPDHnv(GalNAc)RPAPGSTAPPA	0.04472	0.22360

Table 5.7. K_D values for AuNP-glycopeptides **8** and **28-30** with the 5E5 antibody using a onesite binding mode.

Compound	AuNP-glycopeptide	K_D (μg /mL)	K_D (nM)
8	CHGVTSAPDTRPAPGST*APPA	0.00964	0.07141
28	CGGGGPAPGST*APPA	0.04221	0.31267
29	CGGGGPAPGS(GalNFAc)APPA	0.02363	0.17504
30	CGGGGPAPGST(GalNCOR)APPA	0.01167	0.08644

5.5.5. Serological study with pancreatic cancer patients and healthy volunteers

Patients diagnosed with pancreatic adenocarcinoma (n=20) at the Medical Oncology Service of the Hospital San Pedro (Logroño, Spain) were recruited for the study and signed the informed consent form. In addition, age- and sex-matched healthy volunteers (n=20) were recruited among the blood donors of

the Blood Bank at the same hospital. The study was designed as a retrospective, observational, and longitudinal clinical study. All data were anonymized. Personal and clinical data collected for the study are in line with the Spanish Data Protection Act (Ley Orgánica 3/2018 de 5 de diciembre de Protección de Datos Personales). The study adheres to all tenets of the Declaration of Helsinki, and was approved by the local review board (Comité Ético de Investigación Clínica de La Rioja, Ref. CEICLAR P.I. 260). Blood was collected from all subjects and serum was extracted and subjected to MUC1 immunoreactivity assay.

Statistics analysis

The statistical analysis was performed by the group of G. Malerba (University of Verona) using the t-test with a discriminative parameter of sample type (case/control). The main R library used was Limma and the support R libraries were readr, data.table, dplyr, sqldf, stringr, reshape2, tidyr, ggplot2, ComplexHeatmap.

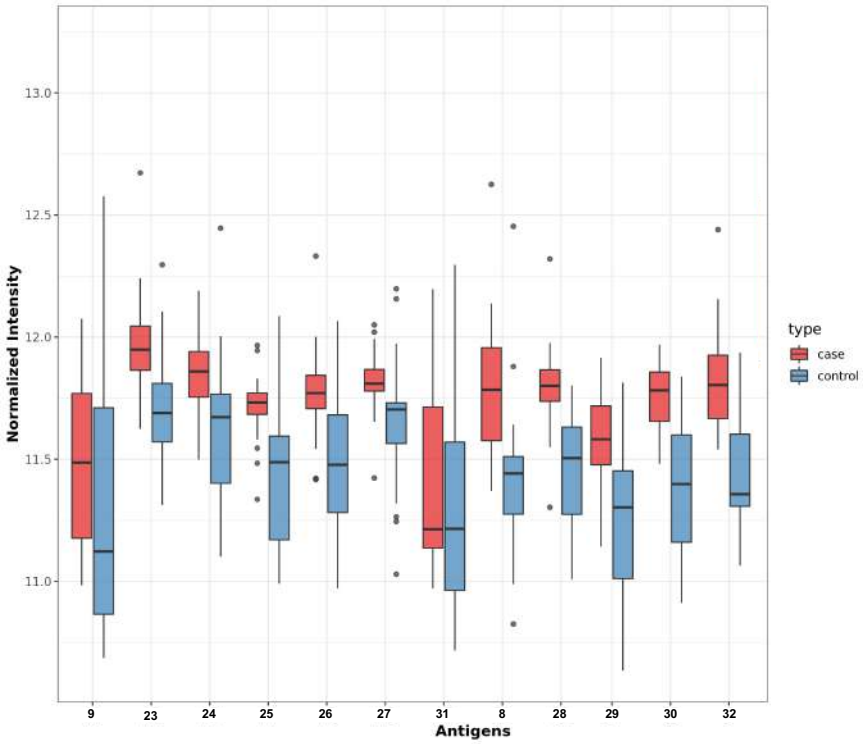


Figure 5.30. Boxplot of the signal intensities obtained with the dot-blot assay divided by healthy patients in blue (control) and patients with pancreatic cancer in red (cases) screened with all the glycopeptides.

In order to verify the significance showed by the t-test and by the graphs (boxplot and heatmap), 1000 permutations of the analyses were carried out with randomisation of the case/control variable. The results showed that only glycopeptides **8**, **28**, **29**, **30** and **32** antigens are not dependent on chance.

To note, results obtained by these 5 glycopeptides are correlated. Therefore, it is possible to screen a serum sample with one single glycopeptide or with two glycopeptides to obtain a more reliable result.

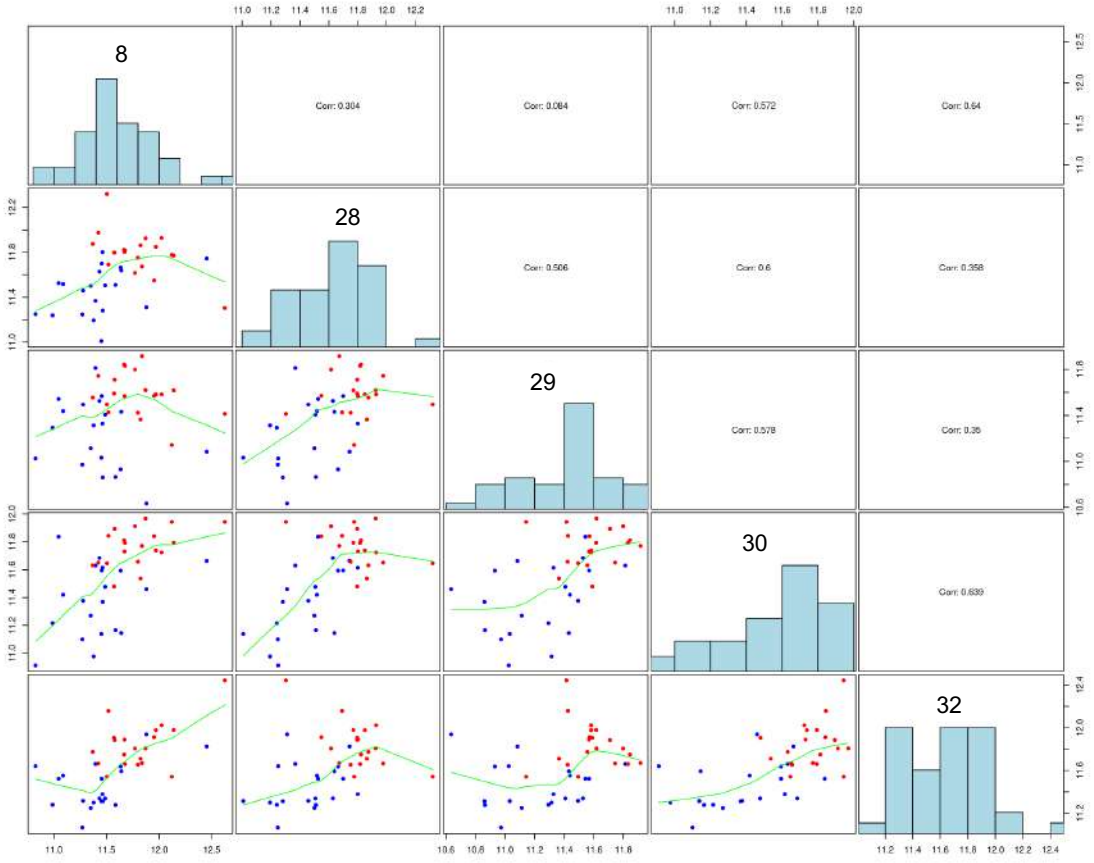


Figure 5.31. Correlation plot of the most significant glycopeptides analyzed in this serological assay

In parallel, another statistical analysis was performed by A. Martinez (CIBIR). R statistical software was used for statistics and Excel for drawing some plots. For comparisons, data distribution was checked for normality (Shapiro-Wilk test) and homoscedasticity (Levene test). When datasets were normally distributed with equal variances, *t*-test was used. Welch test was applied in case of normal

data and heteroscedasticity. Non-parametric Mann Whitney U test was used for not-normal data. The significance level was set at $p < 0.05$.

A group of 20 pancreatic cancer patients (65% male, 65.2 ± 9.0 years of age) and 20 age- and sex- matched healthy volunteer (65% male, 63.4 ± 7.5 years) participated in the study. No significant differences were found between the groups regarding either age or sex ratios (Table 5.8).

Table 5.8. Age and sex distribution of cancer patients and healthy volunteers in the study. Statistical tests employed: Chi square (*) or *t*-test (†).

	Healthy volunteers (n=20)	**Cancer patients (n=20)**	*p* value
Sex (males, %)	13 (65%)	13 (65%)	1.000*
Age (years)	63.4 ± 7.5 years	65.2 ± 9.0	0.507†

In table 5.9 are summarized the values of sensitivity and specificity of two tumor markers currently in use for the detection of pancreatic cancer and the results obtained with our dot-blot assay using glycopeptide **30**.

Table 5.9. Sensitivity and specificity values of two currently used tumor markers for pancreatic cancer and results obtained for our dot-blot assay (glycopeptide **30**)

	Sensitivity	**Specificity**	**AUC**
CEA[44]	77.0	83.2	~0.8 (not published)
CA19-9[43]	69.5	98.2	0.875
MUC1 (glycopeptide 30)	85.0	90.0	0.918

5.6. References

(1) Kufe, D. W. *Nat. Rev. Cancer* **2009**, *9*, 874–885.

(2) Levitin, F.; Stern, O.; Weiss, M.; Gil-Henn, C.; Ziv, R.; Prokocimer, Z.; Smorodinsky, N. I.; Rubinstein, D. B.; Wreschner, D. H. *J. Biol. Chem.* **2005**, *280*, 33374–33386.

(3) Boshell, M.; Lalani, E.; Pemberton, L.; Burchell, J.; Gendler, S.; Taylor-Papadimitriou, J. *Biochem. Biophys. Res. Commun.* **1992**, *185*, 1–8.

(4) Smorodinsky, N.; Weiss, M.; Hartmann, M.; Baruch, A.; Harness, E.; Yaakobovitz, M.; Keydar, I.; Wreschner, D. *Biochem. Biophys. Res. Commun.* **1996**, *228*, 115–121.

(5) Yousefi, M.; Dehghani, S.; Nosrati, R.; Zare, H.; Evazalipour, M.; Mosafer, J.; Tehrani, B. S.; Pasdar, A.; Mokhtarzadeh, A.; Ramezani, M. *Biosens. Bioelectron.* **2019**, *130*, 1–19.

(6) Moreno, M.; Bontkes, H. J.; Scheper, R. J.; Kenemans, P.; Verheijen, R. H. M.; von Mensdorff-Pouilly, S. *Cancer Lett.* **2007**, *257*, 47–55.

(7) Frenette, P. S.; Thirlwell, M. P.; Trudeau, M.; Thomson, D. M. P.; Joseph, L.; Shuster, J. S. *Tumor Biol.* **1994**, *15*, 247–254.

(8) Rughetti, A.; Fama, A.; von Mensdorff-Pouilly, S.; Taurino, F.; Rahimi, H.; Ribersani, M.; Natalino, F.; D'Elia, G. M.; Bizzoni, L.; Latagliata, R.; Breccia, M.; Foà, R.; Alimena, G.; Girelli, G.; Frati, L.; Nuti, M.; Tafuri, A. *Blood* **2008**, *112*, 5237–5237.

(9) Tothill, I. *Semin. Cell Dev. Biol.* **2009**, *20*, 55–62.

(10) Meyer, T.; Rustin, G. *Br. J. Cancer* **2000**, *82*, 1535–1538.

(11) Basil, C.; Zhao, Y.; Zavaglia, K.; Jin, P.; Panelli, M.; Voiculescu, S.; Mandruzzato, S.; Lee, H.; Seliger, B.; Freedman, R.; Taylor, P. R.; Hu, N.; Zanovello, P.; Marincola, F. M.; Wang, E. *Cancer Res.* **2006**, *66*, 2953–2961.

(12) Bohunicky, B.; Mousa, S. *Nanotechnol. Sci. Appl.* **2010**, *4*, 1–10.

(13) Tang, Y.; Wang, L.; Zhang, P.; Wei, H.; Gao, R.; Liu, X.; Yu, Y.; Wang, L. *Clin. Vaccine Immunol.* **2010**, *17*, 1903–1908.

(14) Gheybi, E.; Amani, J.; Salmanian, A. H.; Mashayekhi, F.; Khodi, S. *Tumor Biol.* **2014**, *35*, 11489–11497.

(15) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. *CA. Cancer J. Clin.* **2021**, *71*, 209–249.

(16) Spain: number of deaths from cancer by type 2019 | Statista https://www.statista.com/statistics/777097/number-of-deaths-from-cancer-in-spain-by-type-of-cancer/ (accessed Nov 3, 2021).

(17) Bose, M.; Mukherjee, P. *Vaccines* **2020**, *8*, 1–21.

(18) Curry, J. M.; Thompson, K. J.; Rao, S. G.; Besmer, D. M.; Murphy, A. M.; Grdzelishvili, V. Z.; Ahrens, W. A.; McKillop, I. H.; Sindram, D.; Iannitti, D. A.; Martinie, J. B.; Mukherjee, P. *J. Surg. Oncol.* **2013**, *107*, 713–722.

(19) Yang, J.; Xu, R.; Wang, C.; Qiu, J.; Ren, B.; You, L. *Cancer Commun.* **2021**, *41*, 1257–1274.

(20) Martínez-Bosch, N.; Cristóbal, H.; Iglesias, M.; Gironella, M.; Barranco, L.; Visa, L.; Calafato, D.; Jiménez-Parrado, S.; Earl, J.; Carrato, A.; Manero-Rupérez, N.; Moreno, M.; Morales, A.; Guerra, C.; Navarro, P.; García de Frutos, P. *eBioMedicine* **2022**, *75*, 103797.

(21) Tinder, T.; Subramani, D.; Basu, G.; Bradley, J.; Schettini, J.; Million, A.; Skaar, T.; Mukherjee, P. *J. Immunol.* **2008**, *181*, 3116–3125.

(22) Hamanakai, Y.; Suehiro, Y.; Fukui, M.; Shikichi, K.; Imai, K.; Hinoda, Y. *Int. J. Cancer* **2003**, *103*, 97–100.

(23) Martínez-Sáez, N.; Peregrina, J. M.; Corzana, F. *Chem. Soc. Rev.* **2017**, *46*, 7154–7175.

(24) Compañón, I.; Guerreiro, A.; Mangini, V.; Castro-López, J.; Escudero-Casao, M.; Avenoza, A.; Busto, J. H.; Castillo, S.; Asensio, J. L.; Jiménez-Osés, G.; Boutureira, O.; Peregrina, J. M.; Hurtado-Guerrero, R.; Fiammengo, R.; Bernardes, G. J. L.; Corzana, F. *J. Am. Chem. Soc.* **2019**, *141*, 4063–4072.

(25) Martínez-Sáez, N.; Castro-López, J.; Valero-González, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem. Int. Ed.* **2015**, *54*, 9830–9834.

(26) Somovilla, V. J.; Bermejo, I. A.; Albuquerque, I. S.; Martínez-Sáez, N.; Castro-López, J.; García-Martín, F.; Compañón, I.; Hinou, H.; Nishimura, S. I.; Jiménez-Barbero, J.; Asensio, J. L.; Avenoza, A.; Busto, J. H.; Hurtado-Guerrero, R.; Peregrina, J. M.; Bernardes, G. J. L.; Corzana, F. *J. Am. Chem. Soc.* **2017**, *139*, 18255–18261.

(27) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. *Acc. Chem. Res.* **2013**, *46*, 946–954.

(28) Hsu, C. H.; Park, S.; Mortenson, D. E.; Foley, B. L.; Wang, X.; Woods, R. J.; Case, D. A.; Powers, E. T.; Wong, C. H.; Dyson, H. J.; Kelly, W. *J. Am. Chem. Soc.* **2016**, *138*, 7636–7648.

(29) Dokurno, P.; Bates, P. A.; Band, H. A.; Stewart, L. M. D.; Lally, J. M.; Burchell, J. M.; Taylor-Papadimitriou, J.; Snary, D.; Sternberg, M. J. E.; Freemont, P. S. *J. Mol. Biol.* **1998**, *284*, 713–728.

(30) Bermejo, I. A.; Usabiaga, I.; Compañón, I.; Castro-López, J.; Insausti, A.; Fernández, J. A.; Avenoza, A.; Busto, J. H.; Jiménez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Jiménez-Osés, G.; Hurtado-Guerrero, R.; Cocinero, E. J.; Corzana, F. *J. Am. Chem. Soc.* **2018**, *140*, 9952–9960.

(31) Plattner, C.; Höfener, M.; Sewald, N. *Org. Lett.* **2011**, *13*, 545–547.

(32) Bermejo, I. A. 'Structure-Based Design of Glycopeptides Featuring Unnatural Tn Antigens and Their Applications to Cancer Vaccination and Diagnosis'. Doctoral dissertation (Universidad de La Rioja 2018).

(33) Posey, A. D.; Schwab, R. D.; Boesteanu, A. C.; Steentoft, C.; Mandel, U.; Engels, B.; Stone, J. D.; Madsen, T. D.; Schreiber, K.; Haines, K. M.; Cogdill, A. P.; Chen, T. J.; Song, D.; Scholler, J.; Kranz, D. M.; Feldman, M. D.; Young, R.; Keith, B.; Schreiber, H.; Clausen, H.; Johnson, L. A.; June, C. H. *Immunity* **2016**, *44*, 1444–1454.

(34) He, Y.; Schreiber, K.; Wolf, S. P.; Wen, F.; Steentoft, C.; Zerweck, J.; Steiner, M.; Sharma, P.; Michael Shepard, H.; Posey, A.; June, C. H.; Mandel, U.; Clausen, H.; Leisegang, M.; Meredith, S. C.; Kranz, D. M.; Schreiber, H. *JCI Insight* **2019**, *4*, e130416.

(35) Sørensen, A. L.; Reis, C. A.; Tarp, M. A.; Mandel, U.; Ramachandran, K.; Sankaranarayanan, V.; Schwientek, T.; Graham, R.; Taylor-Papadimitriou, J.; Hollingsworth, M. A.; Burchell, J.; Clausen, H. *Glycobiology* **2006**, *16*, 96–107.

(36) Jiménez-Moreno, E.; Jiménez-Osés, G.; Gómez, A. M.; Santana, A. G.; Corzana, F.; Bastida, A.; Jiménez-Barbero, J.; Asensio, J. L. *Chem. Sci.* **2015**, *6*, 6076–6085.

(37) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. *J. Am. Chem. Soc.* **2015**, *137*, 15152–15160.

(38) Unione, D. L.; Alcalá, M.; Echeverria, B.; Serna, D. S.; Ardá, D. A.; Franconetti, D. A.; Cañada, P. D. F. J.; Diercks, D. T.; Reichardt, D. N.; Jiménez-Barbero, P. D. J. *Chemistry* **2017**, *23*, 3957–3965.

(39) Liebe, B.; Kunz, H. *Tetrahedron Lett.* **1994**, *35*, 8777–8778.

(40) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595–598.

(41) Grabar, K. C.; Freeman, R. G.; Hommer, M. B.; Natan, M. J. *Anal. Chem.* **2002**, *67*, 735–743.

(42) Maus, L.; Dick, O.; Bading, H.; Spatz, J. P.; Fiammengo, R. *ACS Nano* **2010**, *4*, 6617–6628.

(43) Murakami, M.; Nagai, Y.; Tenjin, A.; Tanaka, Y. *Endocr. J.* **2018**, *65*, 639–643.

(44) Pasanen, P. A.; Eskelinen, M.; Partanen, K.; Pikkarainen, P.; Penttila, I.; Alhava, E. *Br. J. Cancer* **1993**, *67*, 852–855.

(45) Zou, K. H.; O'Malley, A. J.; Mauri, L. *Circulation* **2007**, *115*, 654–657.

(46) Carpentier, C.; Godbout, R.; Otis, F.; Voyer, N. *Tetrahedron Lett.* **2015**, *56*, 1244–1246.

(47) Johnsson, B.; Löfås, S.; Lindquist, G. *Anal. Biochem.* **1991**, *198*, 268–277.

(48) Cai, H.; Degliangeli, F.; Palitzsch, B.; Gerlitzki, B.; Kunz, H.; Schmitt, E.; Fiammengo, R.; Westerlind, U. *Bioorg. Med. Chem.* **2016**, *24*, 1132–1135.

(49) Brody, J.; Kern, S. *Biotechniques* **2004**, *36*, 214–216.

6 Molecular basis for the activity of antifreeze glycoproteins: reconciling the opposite proposed models

6.1. Introduction

Antifreeze proteins (AFP) and antifreeze glycoproteins (AFGP) are a class of biomolecules indispensable for life to organisms that inhabit polar waters to survive at subzero temperatures.[1,2,11,12,3–10] These proteins prevent freezing of organism fluids by inhibiting recrystallization of ice and preventing cell damage that crystals could cause. They are attracting much interest in recent years due to their wide range of applications in fields such as medicine and nutrition. Thus, for instance, these proteins are used to preserve cells, tissues and organs at low temperature.[13] They are also used in cryosurgery,[14] aquaculture[15] and for improve the texture of frozen foods.[16]

AFGPs are generally repeats of the threonine-alanine-alanine tripeptide in which the hydroxyl group of Thr is glycosylated with β-D-galactosyl-(1-3)-α-*N*-acetyl-D-galactosamine bound to each threonine residue (Figure 6.1A). These are classified by size, from AFGP-1 (50 repeating units, 33 kDa) to AFGP-8 (4 repeating units, 2.6 kDa).[6,10]

AFGPs inhibit ice growth by binding to the surface of the primary prism plane (Figure 6.1B) of incipient ice crystals which causes curvatures on ice surface surrounded by the AFGP molecules. This curvature hinders water molecules to bind to ice and therefore lowers the local freezing point through the Kelvin effect (Figure 6.1C).[17] This binding results in a lowering of the freezing point without an appreciable change in the melting point. The difference between melting and freezing temperatures, known as thermal hysteresis (TH), is used to detect and quantify the antifreeze activity of these proteins (Figure 6.1D). These compounds alter also the morphology of the ice crystal, giving rise to crystals with a hexagonal bipyramid morphology (Figure 6.1E).[18]

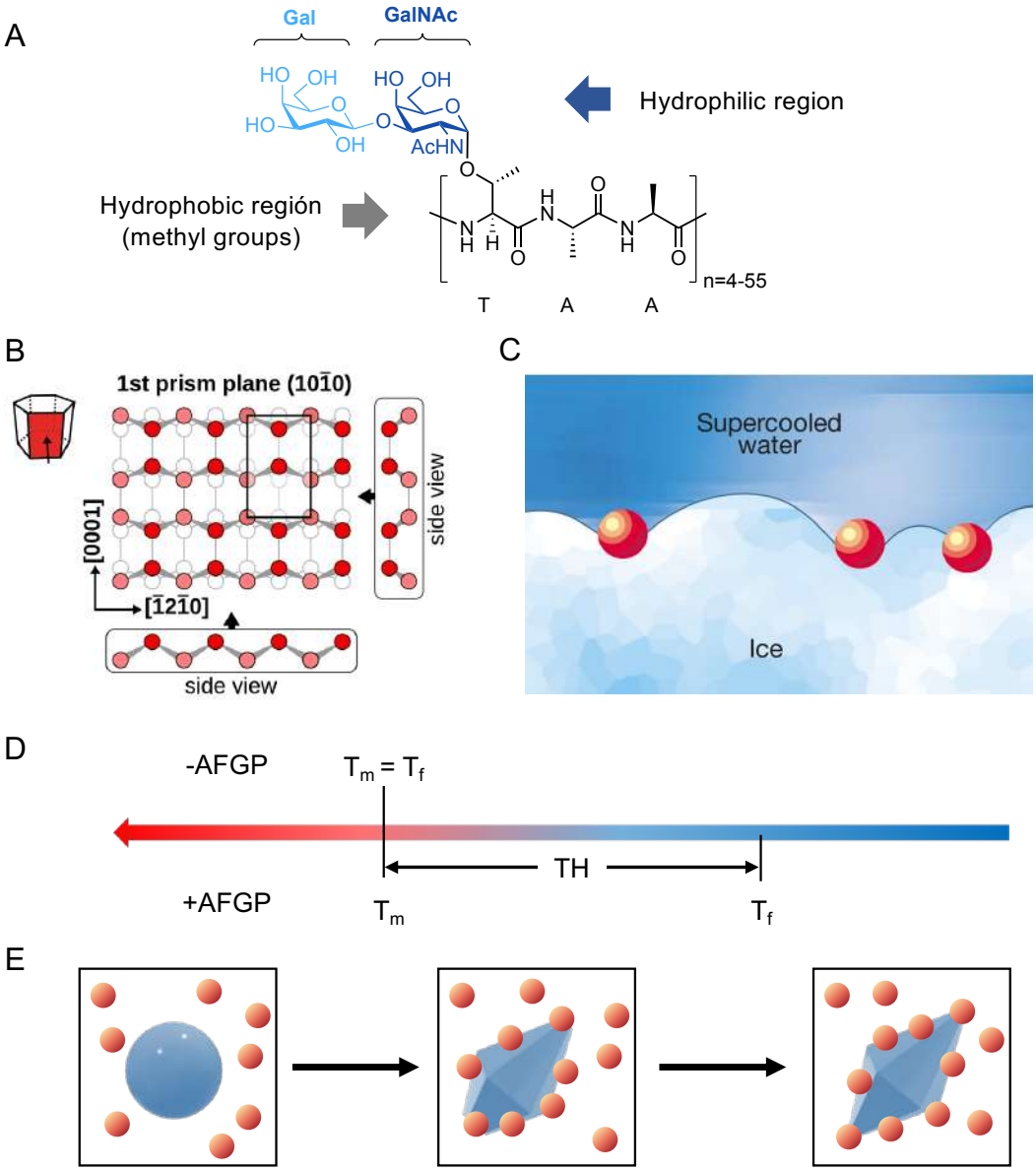


Figure 6.1. Summary of antifreeze glycoproteins properties. (A) Structure of AFGPs. (B) Primary prism face of ice.[19] (C) Model of ice growth inhibition by antifreeze glycoproteins.[17] (D) Schematic representation of thermal hysteresis. (E) Growth of ice crystals and their morphology.

AFGPs were the first proteins with antifreeze activity to be described. However, very little is known about their mechanism of action.[20] This result is principally due to the difficulty in isolating sufficient amounts of pure material from the plasma of organisms in which they are present. Furthermore, attempts to synthesize pure AFGP using cell lines have not yet been successful. Consequently, chemical synthesis of antifreeze glycopeptides[7,18,21–25] as well as unnatural

derivatives, has become a suitable alternative to produce pure AFGP for detailed structure-function studies.

Thus, the rational design of non-natural antifreezing derivatives has generated significant research attention (Figure 6.2). In this sense, Ben and co-workers[26] have synthesized *C*-linked analogues of AFGPs with a variable length between the peptide and the carbohydrate moieties observing different behaviors as inhibitors of ice recrystallization. Conversely, proline oligomers could be used as AFGP mimics,[27] due to the helical type-II conformation adopted by these derivatives which mimics the native AFGP form. Therefore, Sumii *et al.* proposed the synthesis of non-natural AFGPs combining proline oligomers (hydrophobic part) and carbohydrates (hydrophilic part). In addition, fluorinated proline oligomers were investigated, given that these derivatives adopt a more rigid conformation. Interestingly, it was observed a higher antifreeze activity of fluorinated galactose-proline oligomers, probably caused by the different form adopted by this derivative for the fluorine gauche effect.[28]

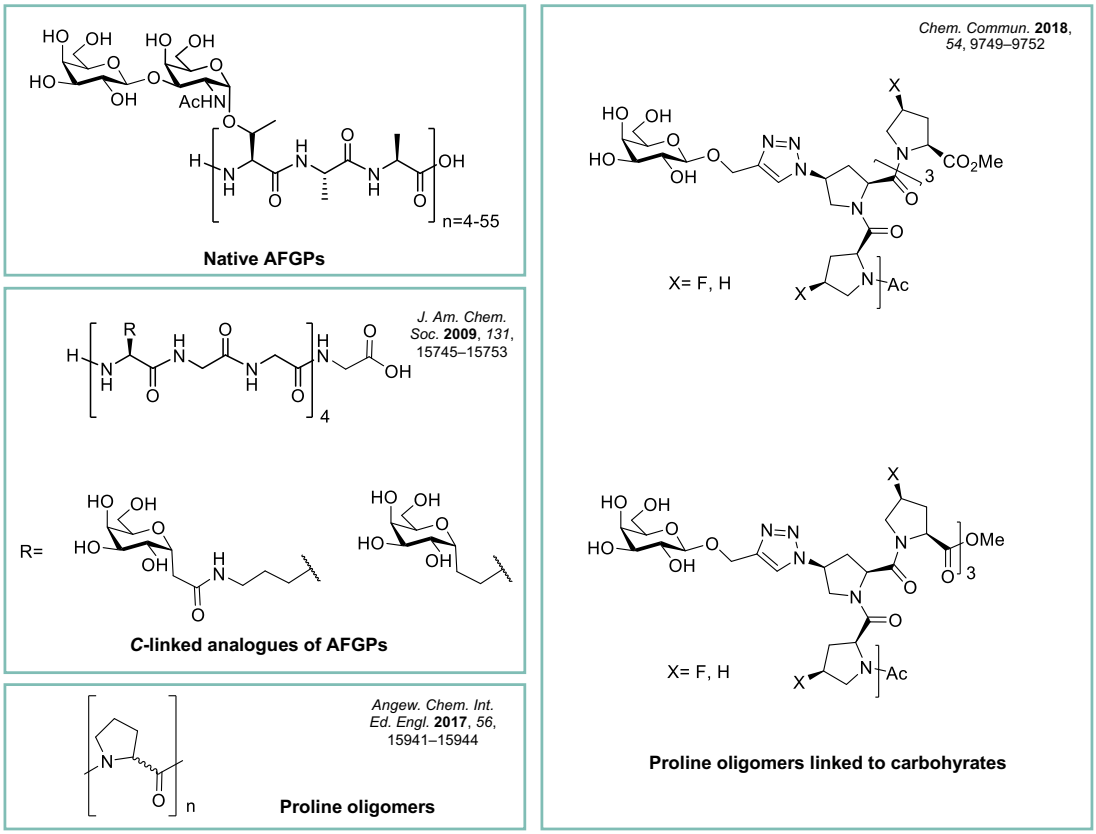


Figure 6.2. Structures of natural AF(G)Ps and previously reported antifreeze analogues.

In this context, Nishimura's group reported the most comprehensive study of AFGPs to date. Mimics modifying the peptide part, the nature of the sugar and the configuration of the glycosidic linkage were synthesized which allowed to stablish the relationship between the structure of these glycoproteins and their activity as inhibitors of ice crystal growth (see Table 2.1 in Chapter 2).[18]

According to these studies, it can be concluded that three key motifs are essential for the antifreeze activity, such as *N*-acetyl group at C2 position of the first carbohydrate attached to the peptide, the α-configuration of the *O*-glycosidic linkages between sugars and the peptide chain, and the methyl group of the threonyl residue.

The conformational analysis of these derivatives in aqueous solution, based on circular dichroism and supported by molecular dynamics (MD) calculations, indicated that the peptide chain of derivatives with antifreeze properties adopts a polyproline-like conformation (PPII).[29] For the glycopeptide with four units, a conformational analysis was accomplished based on MD calculations and NMR experiments (mainly distances derived from NOEs experiments).[18] The main populated conformer shows a PPII helix for the peptide chain, with the carbohydrates oriented towards the same face of the peptide, thus forming a hydrophilic face. Likewise, the methyl groups of the alanine residues and the acetyl group of the carbohydrate are oriented towards the opposite face, giving rise to the hydrophobic zone (Figure 6.3).

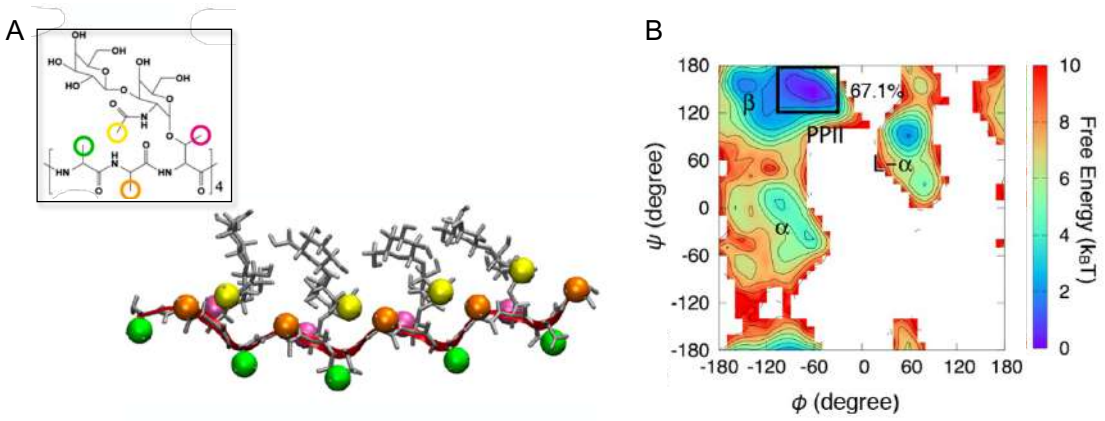


Figure 6.3. Structure of AFGP8 in solution derived from NMR data and MD simulations.[18] (A) Three-dimensional structure with the peptide backbone in PPII helix conformation. The spheres represent the different methyl groups present in the structure. The disaccharide is shown in gray and the peptide chain in red. (B) Ramachandran plot of AFGP8 at 268 K highlighting the population of PPII-helix. The regions for PPII-helix (PPII), β-sheet (β), α-helix (α) and left-handed α-helix (L-α) secondary structures are labeled in the graph.[30]

Despite these experimental data, it is important to note that the molecular mechanism by which AFGPs inhibit ice growth is not currently known in detail.[8,9,37,38,10,30–36] In fact, two opposing models have been proposed to explain the mechanism of inhibition of ice growth by these molecules.[30,34] The "classical" model suggests that the hydroxyl groups of carbohydrates are those that irreversibly bind to the surface of ice, thus inhibiting its growth.[33] (Figure 6.4A). Here, it is proposed that the hydroxyl groups of the carbohydrate units are directly included into the ice crystal surface, thus, forming three hydrogen bonds with water (Figure 6.4B).[7] In favor of this mechanism, Meister and co-workers tested the irreversibility of AFGP binding to ice by microfluidic solution exchange experiment. Fluorescent-labeled AFGP was introduced into the microfluidic channel which provokes the formation of crystal with hexagonal bipyramid structure. After the replacement of this solution with water, it was found that AFGP were irreversibly bound to ice (Figure 6.4C).[34]

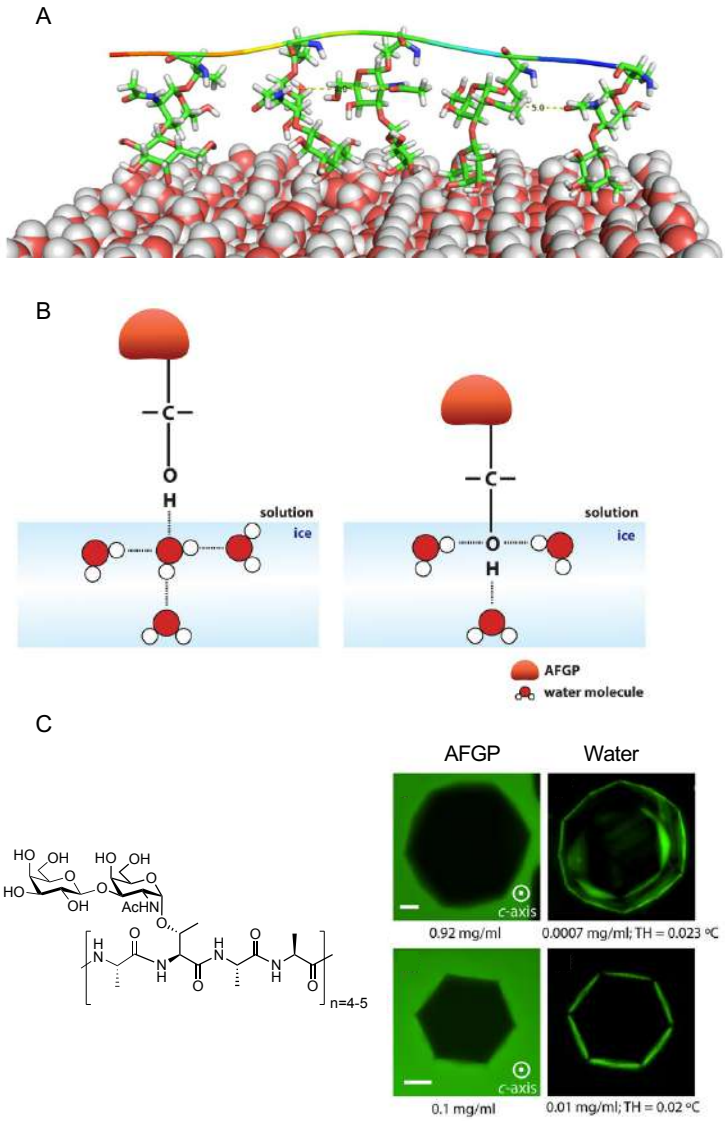


Figure 6.4. Proposed mechanism for the inhibition of ice growth by antifreeze glycopeptides. (A) In this model, hydroxyl groups of the galactose residue irreversible interact with the ice surface. (B) An alternative model in which the hydroxyl groups of the sugar are part of the structure of the ice crystal lattice. (C) Irreversible binding of AFGP7–8 isoforms to ice. Images of ice crystals before and after the AFGP solution was exchanged with pure water.

Recently, a theoretical model has been proposed, based on molecular dynamics calculations,[30] in which the methyl groups of the peptide chain and the acetyl groups of the carbohydrates of antifreeze glycoproteins are the responsible to bind *reversibly* to ice (Figure 6.5A). This mechanism has also been proposed for antifreeze proteins.[39] This model states that constrained water molecules on the ice binding site might be released into the bulk solvent during the proteins

adsorption to ice which cause a gain in entropy, driving in this way the binding process. Similarly, constrained molecules of water around hydrophobic groups (mainly methyl groups of the glycoprotein) are released into the solvent, entropically favoring the bind of the hydrophobic groups (Figure 6.5B).[40] Alternatively, diverse modeling studies[35,41–43] proposed that water in the ice bind site is organized into a constrained ice-like pattern which can merges and freezes with the 'quasi-liquid layer'[38,44] of the ice lattice (Figure 6.5C).

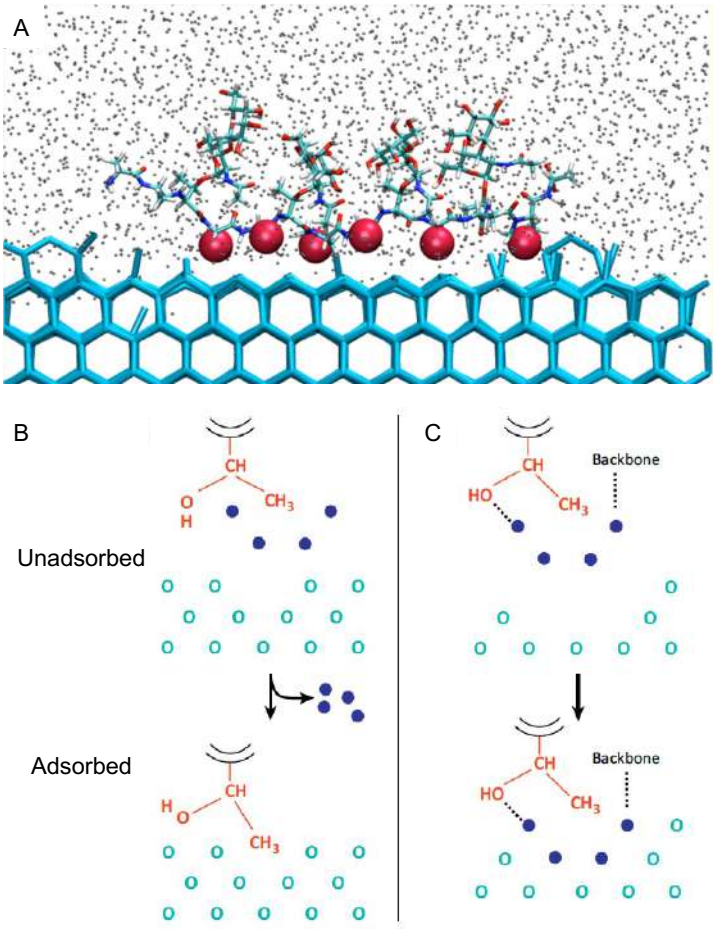


Figure 6.5. Proposed mechanism for the inhibition of ice growth by antifreeze glycopeptides. (A) In this model, the methyl groups of the peptide chain and of the carbohydrate residues interact reversibly with the ice surface. In this case, the methyl groups of the alanine residues have been represented as red spheres. (B) Constrained molecules of water around hydrophobic groups are released into the solvent (top) entropically favoring the bind of the hydrophobic groups (bottom) (C) Anchored clathrate hypothesis. Hydrogen bonds between the ice-like clathrate waters around hydrophobic groups and the protein backbone and side chain groups (top). These clathrate molecules of water merge with the quasi-liquid layer waters becoming ice (bottom).

6.2. Objectives

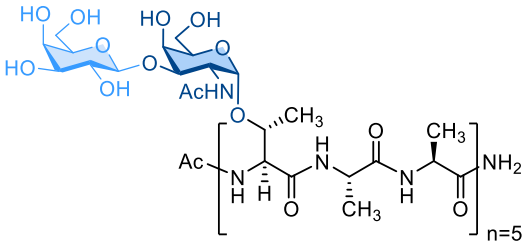
Firstly, in order to determine the structural motifs essential for the antifreeze activity of AFGPs, we decide to develop small glycopeptides containing several changes in the chemical structure of AFGPs. In a second step, we will analyze the conformational properties of these new derivatives and their antifreeze activity. With this information in hand, we will propose an alternative mechanism that attempts to reconcile the currently proposed models.

6.3. Results

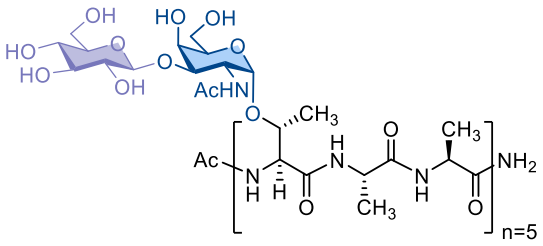
6.3.1. Synthesis of glycopeptides

In this context, our first approach was to synthesize glycopeptides with only two tandem repeating units. We did not observe any antifreeze activity in these derivatives, which contrasts with the data reported by Nishimura and co-workers.[18] Therefore, we decided to prepare glycopeptides with more tandem repeating units, especially with *n*=5 (Figure 6.6).

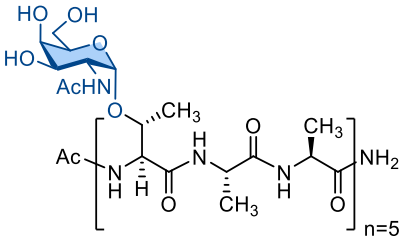
Glycopeptides were synthesized by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer with Rink Amide MBHA resin using the methodology described in previous chapters. The crude products were purified by HPLC, lyophilized and characterized.



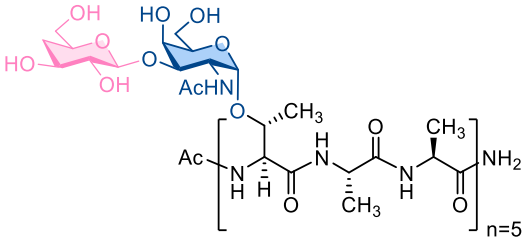
(33) (GalGalNAc-TAA)$_5$



(34) (GlcGalNAc-TAA)$_5$



(35) (GalNAc-TAA)$_5$



(36) (4-deoxy-GalGalNAc-TAA)$_5$

Figure 6.6. Glycopeptides synthesized and studied in this work.

6.3.2. Thermal hysteresis

We first measured the thermal hysteresis (TH) of these derivatives to evaluate their antifreeze activity. Freezing and melting points were determined through a microscopic observation method, as described in the experimental section 6.5.2. The temperature at which ice crystal growth starts is considered the freezing temperature, whereas the melting temperature is established when the seed crystal melts. The difference between both temperatures is the thermal hysteresis which is a direct measure of antifreeze activity. Figure 6.7 shows a comparison of thermal hysteresis curves of glycopeptides **33**-**36**.

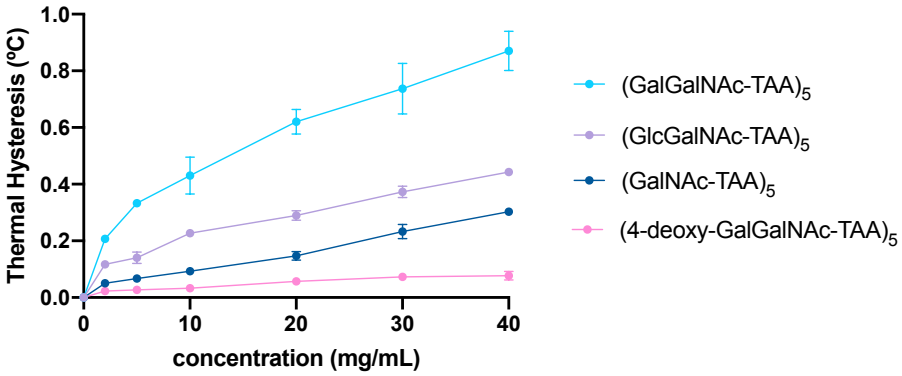


Figure 6.7. Thermal hysteresis activity of glycopeptides

Based on these results, we found that the highest value of thermal hysteresis (around 1°C) was obtained for the natural glycopeptide which contains the disaccharide Galβ(1-3)GalNAc. A significant decrease in TH was observed when the carbohydrate attached to the GalNAc unit was substituted for a D(+)-glucose. When the carbohydrate attached to the threonine was the monosaccharide GalNAc, a lower TH was obtained. This result is in agreement with that reported by Nishimura[18] confirming that the terminal Gal, although not required, slightly improves the activity. It is noteworthy that the incorporation of a 4-deoxy-GalGalNAc leads to an almost complete loss of freezing activity, indicating the importance of the OH-4 of the galactose unit. Indeed, the addition of 4-deoxy-Gal leads to a greater decrease in activity than the removal of this unit.

6.3.3. Ice crystal morphology

Furthermore, antifreeze glycoproteins alter the morphology of the ice crystal, giving rise to crystals with a hexagonal bipyramid morphology. Figure 6.8 shows a scheme of the diverse morphology of ice crystal correlated to the antifreeze activity of these glycoproteins.

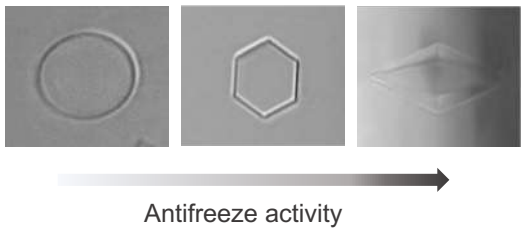


Figure 6.8. The ice-crystal morphologies depending on the antifreeze activity of the antifreeze glycoproteins.

Hence, it was analyzed the ice-crystal morphology of water in the presence of the synthesized glycopeptides at different concentrations (Figure 6.9). Interestingly, all glycopeptides except for the (4-deoxyGalGalNAc-TAA)$_5$ derivative can shape ice crystals into hexagonal bipyramids, being the natural glycopeptide **33** the derivative with best antifreeze activity. These results are consistent with those obtained with the thermal hysteresis experiments.

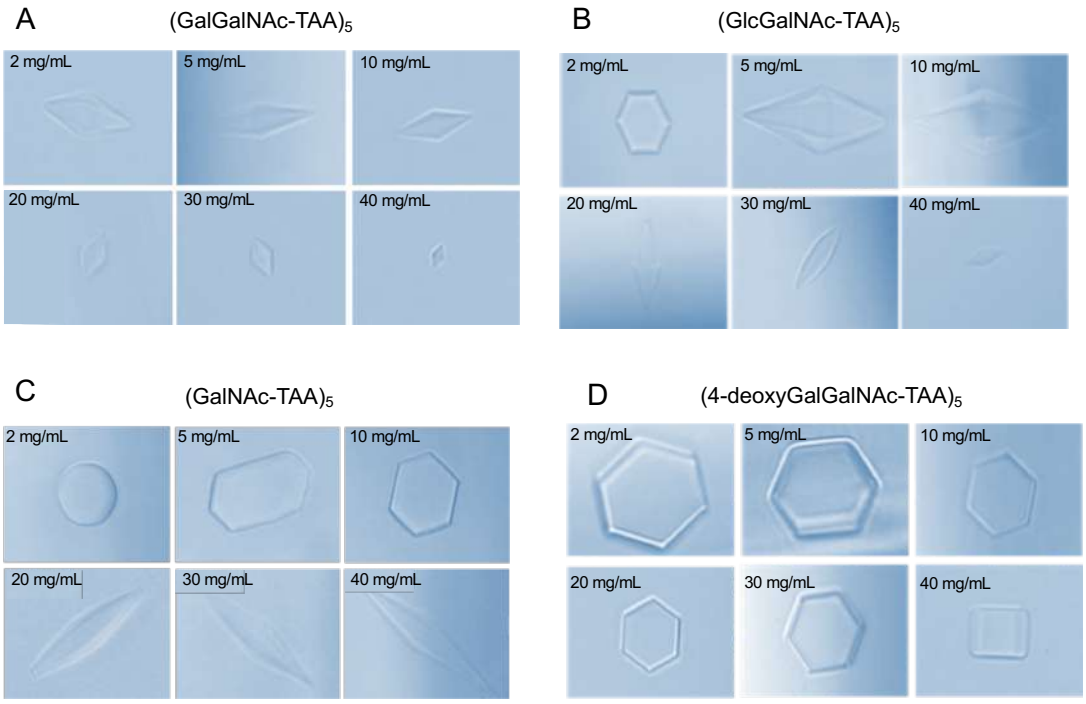


Figure 6.9. The ice-crystal morphologies in the presence of diverse concentration of (A) glycopeptide **33**, (B) glycopeptide **34**, (C) glycopeptide **35**, (D) glycopeptide **36** in water.

6.3.4. Conformation analysis in solution

As a next step, we decided to study the structure of these glycopeptides in order to evaluate if the difference of their antifreeze activity can be attributed to a different conformational behavior in water solution.

For this purpose, the conformation of the glycopeptides **33**-**36** was first analyzed using circular dichroism experiments. From the CD curves (Figure 6.10), it is possible to observe a maximum at $\lambda \approx 216$-217 nm and a minimum at $\lambda \approx 194$-195 nm for all the derivates, which fits with a polyproline-like conformation (PPII).[45,46]

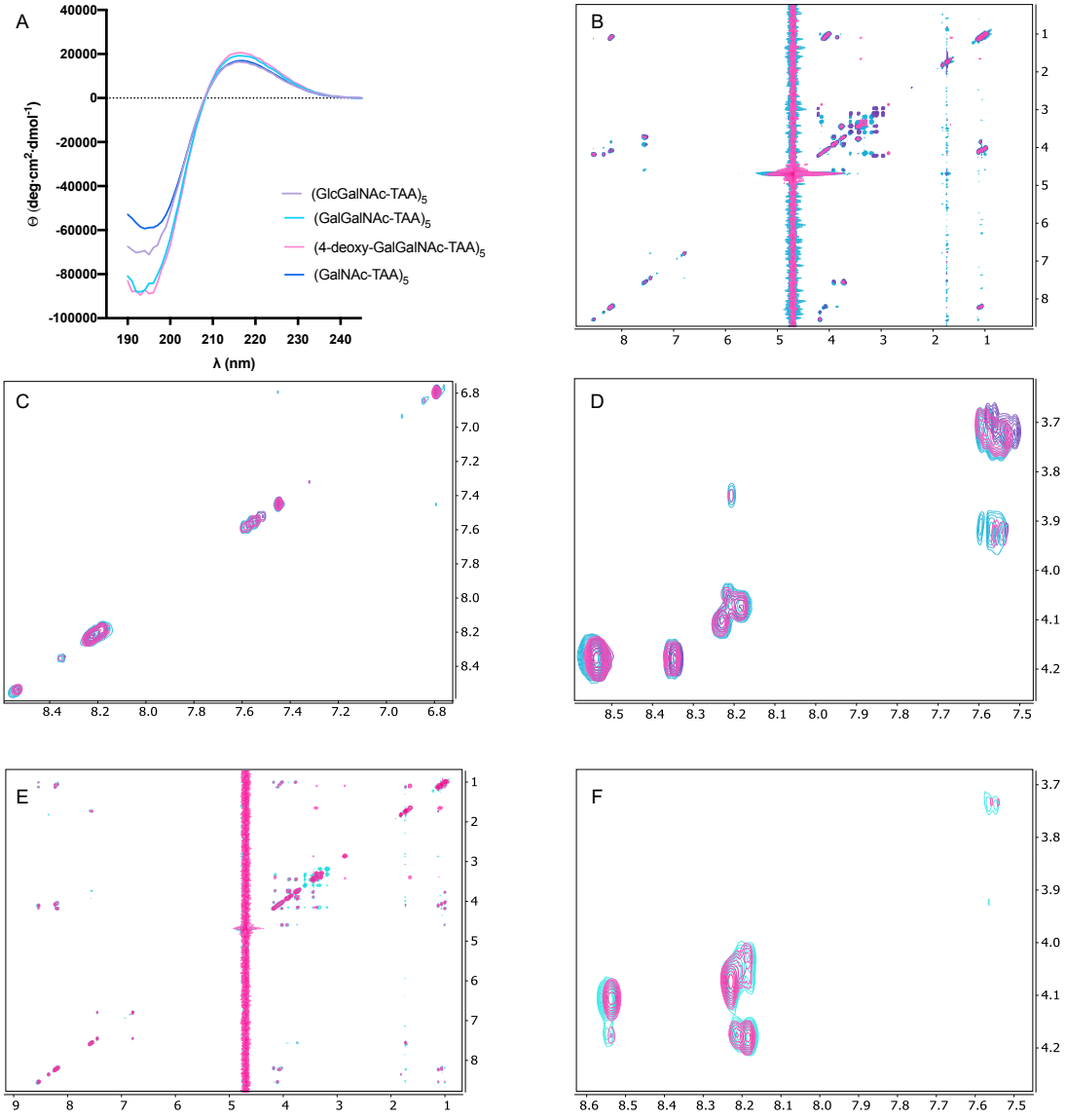


Figure 6.10. Conformational analysis in solution of synthetic glycopeptides. (A) CD spectra of compounds **33**-**36** (0.5 mg·mL^{-1}) in water at 5 °C. (B) Superposition of TOCSY spectra of compounds **33**, **34** and **36** in H_2O/D_2O (95:5) at 5.0°C recorded in a 600 MHz NMR spectrometer. (C) Zoom on the amide region of the TOCSY spectra (D) Zoom on the NH-Hα cross-peaks region of the TOCSY spectra. (E) Superposition of NOESY spectra of compounds **33** and **36** in H_2O/D_2O (95:5) at 5.0°C recorded in a 600 MHz NMR spectrometer. (F) Zoom on the NH-Hα cross-peaks region of NOESY spectra.

As a next step, we carried out the NMR analysis on glycopeptides **33**, **34** and **36**. The analysis of the TOCSY and NOESY spectra indicate that these derivatives display the NMR same pattern (Figure 6.10), which also confirms that all of them

exhibit the same conformation in solution. Furthermore, no difference are observed in chemical shift for the different tandem repeat units which suggest a similar environment for the 5 TAA units. In addition, NOE contacts between consecutive Hα-NH groups were no observed, which confirms an extended conformation for the backbone of these peptides.[47]

Therefore, once confirmed that all glycopeptides exhibit the same conformation, yet their antifreeze activity is very different. This result indicates that the carbohydrate moiety is essential for the activity of these peptides which is difficult to reconcile with the model proposed by Molinero and co-coworkers,[30] in which the methyl groups of the peptide and GalNAc are those that bind to ice and the carbohydrates plays a purely structural effect .

6.3.5. MD simulations

In addition to these experimental data, we performed 1 μs Molecular Dynamic (MD) simulations on these derivatives in explicit water to analyze the structure adopted by these peptides in solution. The analysis of the ϕ/ψ distribution of all residues in all compounds shows that they adopt mainly an extended-like conformation in solutions (data not shown), which is compatible with the PPII conformation. However, it is important to note that derivate (4-deoxy-GalGalNAcThrAlaAla)$_5$, also populates helix-like conformations for some of the residues. These aspects can be also inferred from the superposition of several frames evenly spaced along the MD trajectories (Figure 6.11).

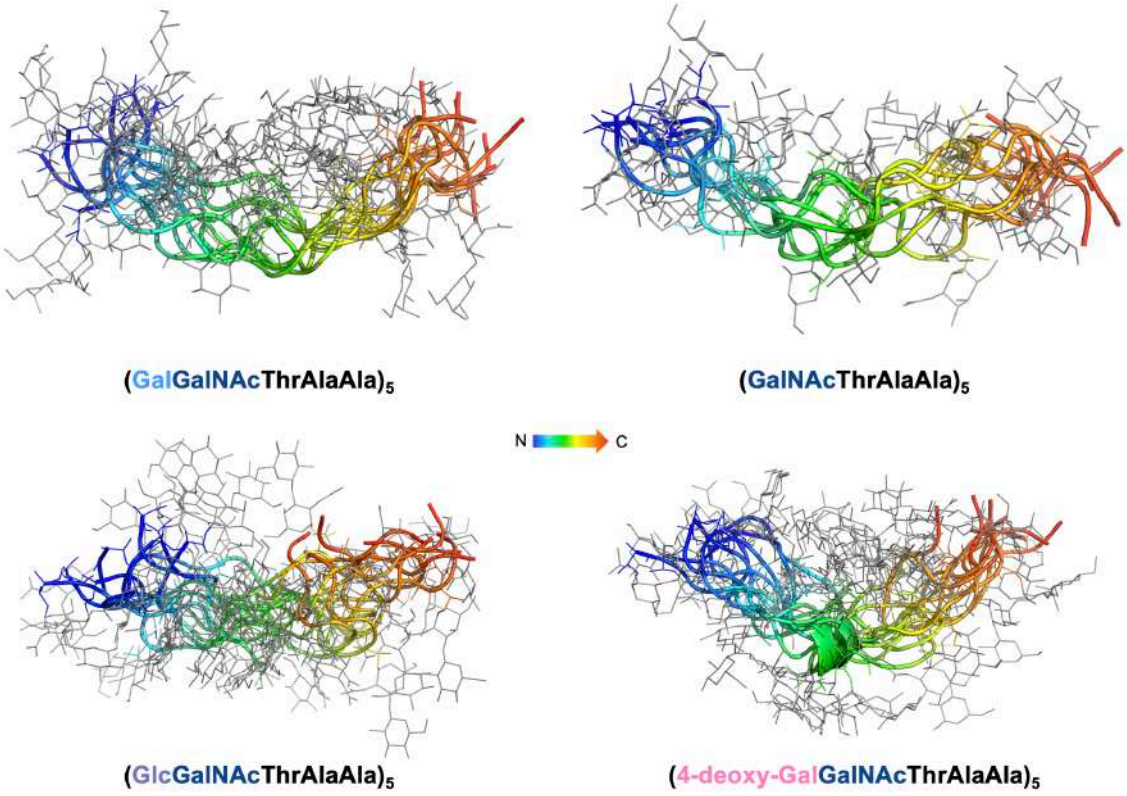


Figure 6.11. Overlay of 10 frames evenly spaced along the 1 μs MD trajectory for glycopeptides **33-36**.

For the natural-like glycopeptide **33**, we performed a cluster analysis of the MD trajectory. Figure 6.12A shows the three most populated conformations of glycopeptide **33** in solution, being the one shown at the right the main conformer proposed in the literature.[18] In this conformer, the peptide adopts a PPII helix with the carbohydrates oriented towards the same face of the peptide (hydrophilic face), whereas the methyl groups are oriented towards the opposite face (hydrophobic zone) as illustrated in Figure 6.12B.

Furthermore, the distance distribution between the OH-4 of the galactose units of adjacent tandem repeats was calculated throughout the MD simulation, obtaining in all cases a peak at 4.5 Å (Figure 6.12D). Notably, this distance matches with the distance between oxygen atoms in the primary prism plane of ice (Figure 6.12C).

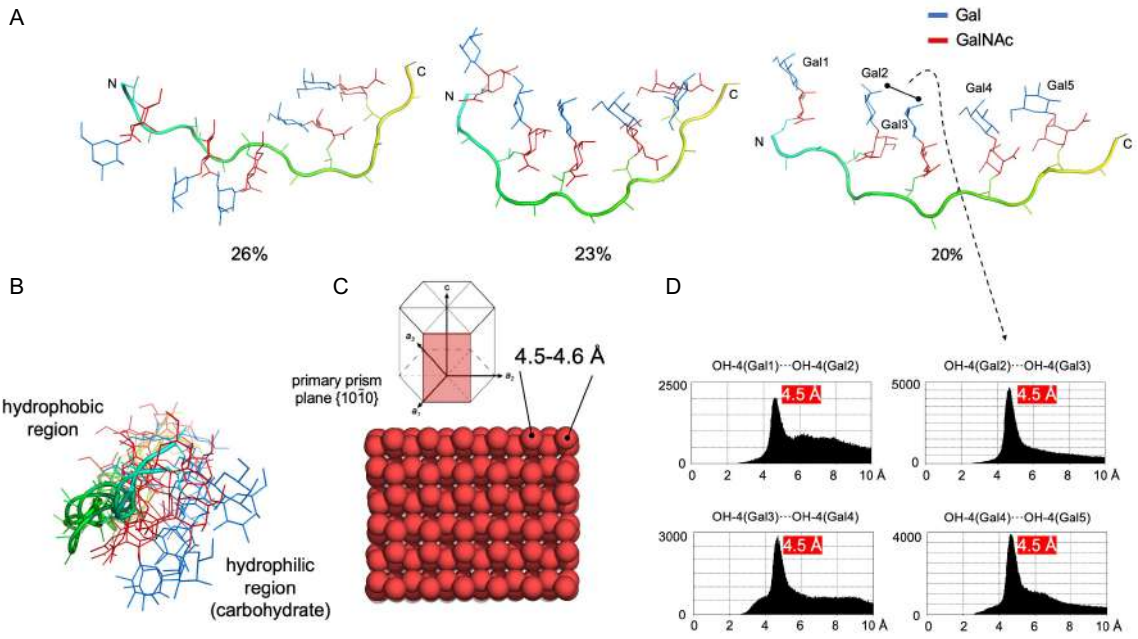


Figure 6.12. Conformational study in solution of glycopeptide **33** by MD simulations. (A) Main conformers obtained in solution for glycopeptide **33** by MD simulations. (B) Different regions observed for conformer shown at the right in panel A. (C) Distance between oxygen atoms of the primary prism plane of ice. (D) Distance distribution between the adjacent OH-4 of galactose units derived from the MD simulations.

6.3.6. Synthesis of deuterated glycopeptides

From these experimental data, it is clear the importance of the carbohydrate for the antifreeze activity, which is in accordance with the model proposed by Nishimura and others.[33,34,48–50]

At a next step, we decided to test if the methyl groups of threonine and alanine residues, along with the acetamide group of GalNAc have also a key role in the antifreeze activity. For this reason, we proposed the replacement of hydrogen atoms with deuterium atoms in specific positions of the glycopeptide. It is important to note that this 'subtle' substitution do not should provoke a change in the conformational behavior of the compounds. Our hypothesis implies that if the -CD_3 groups are directly involved in hydrophobic interactions with the ice surface, then the antifreeze activity of these compounds should be attenuated related to those bearing the -CH_3 groups.[51]

Therefore, we carried out the synthesis of diverse deuterated glycopeptides by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer with Rink Amide MBHA resin using the Fmoc strategy above mentioned. The deuterated residues used in this work are commercially available (Figure 6.13).

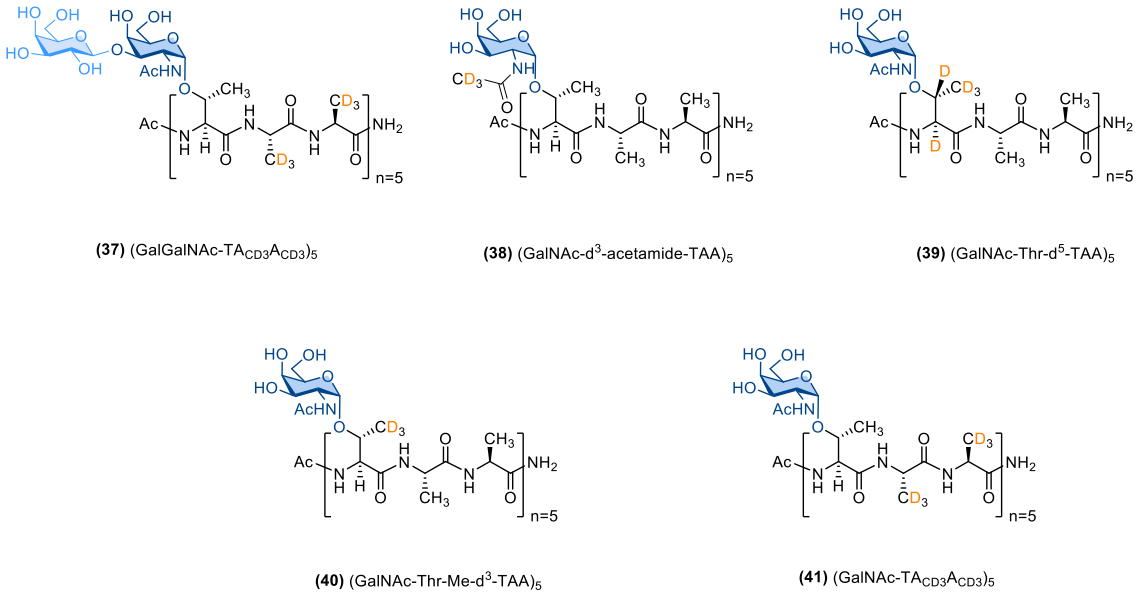


Figure 6.13. Deuterated glycopeptides synthesized and studied in this work.

Interestingly, when we performed the HPLC purification of a mixture of glycopeptides **33** and **37**, we observed that the deuterated derivative had a smaller retention time than the natural derivative (Figure 6.14B), which in a way confirms that the first eluted has less hydrophobic interactions with the alkyl chains of the C18 column (Figure 6.14A).

As a next step, in order to investigate the relevance of isotopic effect (H by D substitution) in hydrophobic interactions, we carried out a computational study using small molecules such as CH_4 and CD_4 with H_2O-ice and D_2O-ice by QM-MD calculations. These calculations were performed in collaboration with Dra. Rovira (University of Barcelona). According to these simulations, the CH_4 molecule interacts more tightly with the ice surface than the deuterated derivative, which is reflected in a more compact occupancy function for CH_4, with a clear preference for a specific position of ice structure (Figure 6.14C). In contrast, CD_4 interacts with different parts of the ice surface through the QM-MD trajectory due likely to the lower hydrophobic interactions of the deuterium

atoms (Figure 6.14D). Furthermore, when D_2O ice was used instead of H_2O, this effect was enhanced, as shown in Figure 6.14E.

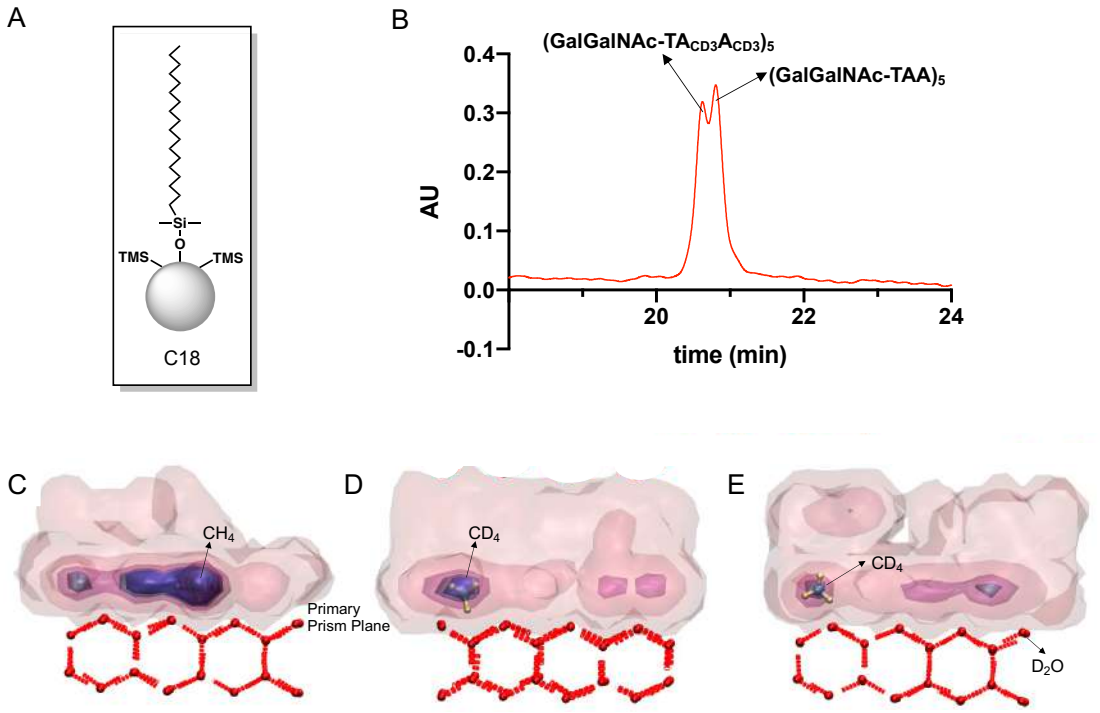


Figure 6.14. Study of the isotopic effect through HPLC assays and QM-MD simulations. (A) Chemical structure of the C18 stationary phase in the porous silica support of the HPLC column used in this work. (B) HPLC chromatogram of a mixture 1:1 of glycopeptides **33** and **37** obtained using a Phenomenex Luna 5 µm C18(2) column, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) →(11:89), 30 min, 1 mL/min, λ= 212 nm. Occupancy function of (C) CH_4/ice (D) CD_4/ice and (E) CD_4/D_2O-ice along the whole QM-MD trajectory (100 ps).

Then, we determined the thermal hysteresis and the ice crystal morphology of the deuterated glycopeptides shown in Figure 6.15. Of note, although not differences were observed in the ice crystal morphology between glycopeptides **33** and **37**, the TH values were significantly lower (p= 0.0014 (**)) for the deuterated glycopeptide **37**, which is bearing Ala residues with deuterated methyl groups.

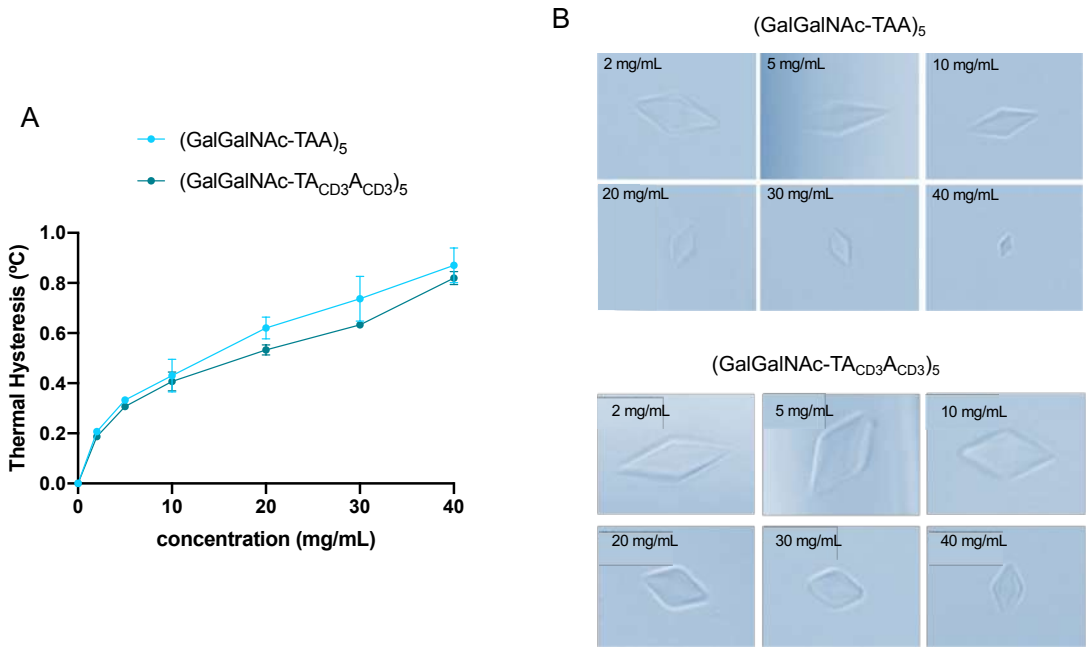


Figure 6.15. Comparison of natural $(GalGalNAc\text{-}TAA)_5$ (glycopeptide **33**) and the deuterated mimetic $(GalGalNAc\text{-}TA_{CD3}A_{CD3})_5$ (glycopeptide **37**). (A) Thermal hysteresis of both glycopeptides. (B) Ice crystal morphology at diverse concentration of both glycopeptides.

Regarding the glycopeptides **38-41**, which contain exclusively the monosaccharide GalNAc and hydrogen-by-deuterium substitution positions, their antifreeze activity was also tested (Figure 6.16A). A slightly descent of their thermal hysteresis was observed related to the non-deuterated glycopeptide **35**. Of note, the lowest thermal hysteresis was obtained for glycopeptide **41**, bearing deuterium atoms in the alanine residues ($-CD_3$). This result could entail that the methyl groups of the alanines have a greater interaction with the ice surface than the other hydrophobic groups of the molecule. In general, these results are supported by the ice-crystal images (Figure 6.16B). All these glycopeptides produce hexagonal bipyramidal ice crystals. However, the non-deuterated glycopeptide **35** present the capacity to shape ice crystals into hexagonal bipyramids at lower concentration than the rest of the mimetics.

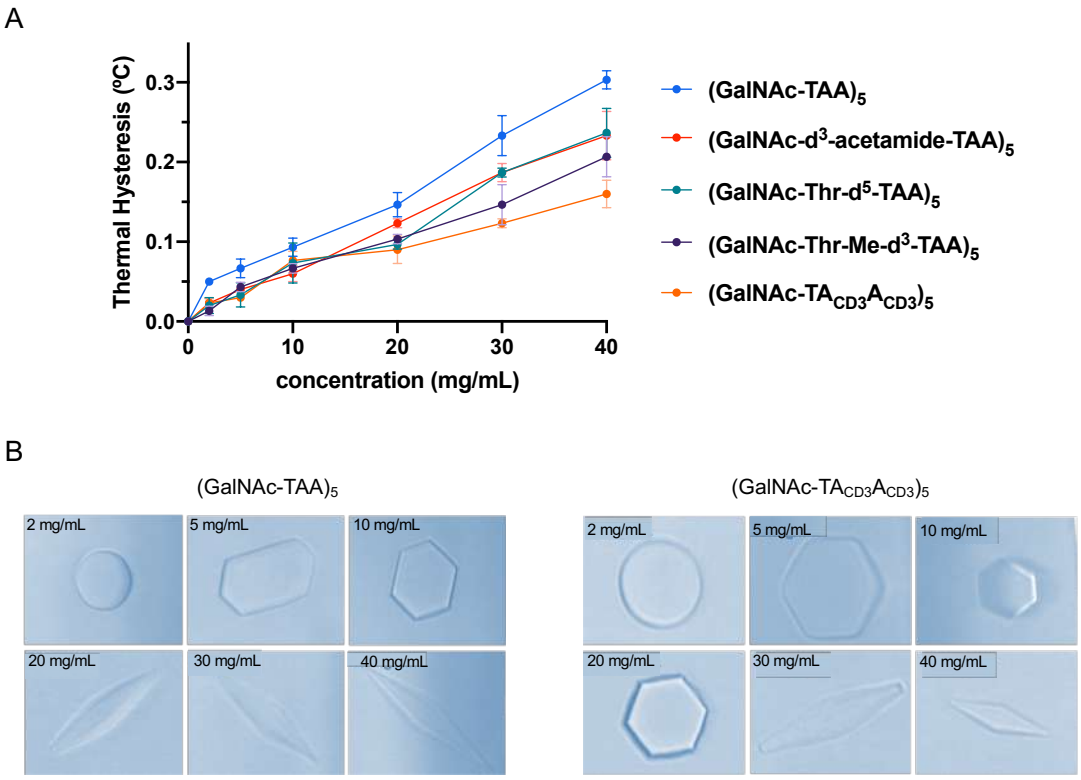


Figure 6.16. Comparison of deuterated glycopeptides **38**-**41** and glycopeptide **35**. (A) Thermal hysteresis of all glycopeptides. (B) Ice crystal morphology at diverse concentration of $(GalNAc\text{-}TAA)_5$ (glycopeptide **35**) and $(GalNAc\text{-}TA_{CD3}A_{CD3})_5$ (glycopeptide **41**)

It is important to note that for these glycopeptides bearing a monosaccharide, the isotopic effect is more significant than for the glycopeptides **33** and **37**, which present a disaccharide and higher antifreeze activity.

Considering the results disclosed in this chapter, they seem to indicate that both the methyl groups and the sugar moieties play a key role in the antifreeze effect. As we have previously mentioned, there is a layer of semi-ordered ice, known as quasi-liquid layer (QLL) between the interface of the ice lattice and bulk water.[52] In this regard, some studies suggest that interactions between the carbohydrate and this layer can be responsible of the inhibition of ice crystallization,[38,53] being the stereochemistry of the carbohydrate essential for their antifreeze activity due to their different hydration indexes. For instance, galactose has the larger hydration index.[44] This finding is consistent with our MD simulations on the free glycopeptides, which show that the disaccharide Galβ(1-3)GalNAc has the highest number of water molecules compared to the disaccharide with glucose

or 4-deoxy-galactose at the non-reducing end (Figure 6.17A). The MD simulations show that the residence time of the hydrogen bonds between the hydroxyl groups of the sugar and water molecules is generally higher for the galactose residue than for glucose or 4-deoxy-galactose as well (Figure 6.17B). Thus, these two features that are unique for galactose cause a greater disruption of the QLL and consequently this sugar may be a better ice recrystallization inhibitor.

While the first carbohydrate unit (GalNAc) is essential to maintain the proper conformation of the backbone, the sugar attached to this first unit (second unit) plays an essential role in the antifreeze activity. In this sense, the higher antifreeze activity of the derivative featuring a Galβ(1-3)GalNAc in comparison with the Glcβ(1-3)GalNAc derivative could be explain by the ability of galactose to hold the large number of water molecules tightly to its OH groups which consequently increases the disorder of the water molecules and therefore disfavors ice formation.

With all these data in hand, we propose a new model that try to reconcile the previous ones: a) the model where the methyl groups of the glycopeptide interact with ice through hydrophobic interactions, and b) the model where the carbohydrate residues interact with the quasi-liquid layer (QLL).

On this way, we propose a hybrid model, in which the main features of previous models could coexist. Methyl groups of alanine and threonine residues interact with ice through hydrophobic interactions, and the non-reducing end of the carbohydrates interact with the quasi-liquid layer (QLL), deforming its structure (Figure 6.17C).

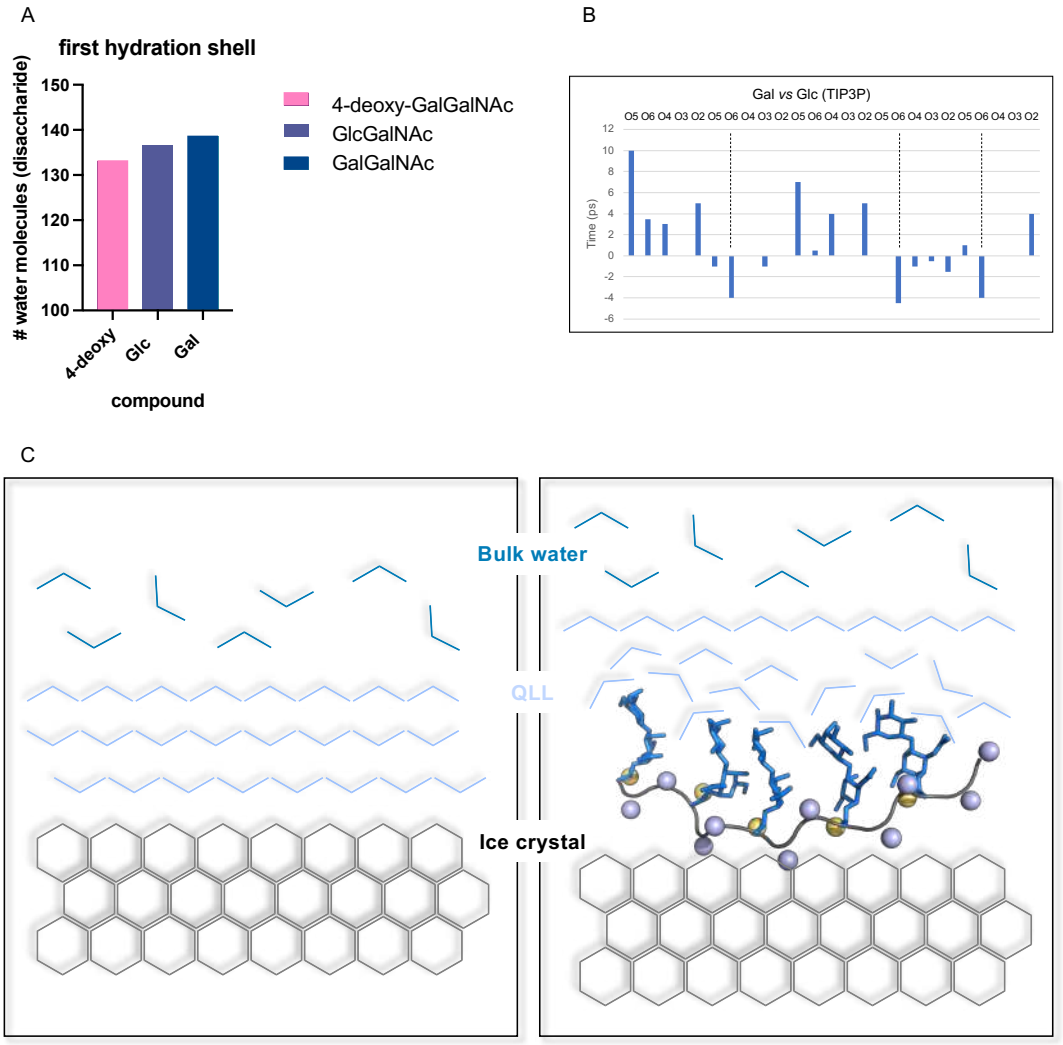


Figure 6.17. (A) Comparison of water molecules observed in the first hydration shell of disaccharides 4-deoxy-GalGalNAc, GlcGalNAc and GalGalNAc. (B) Residence time of the hydrogen bonds between the hydroxyl groups of the sugar and water molecules obtained by 1 μs MD simulations using TIP3P water model (hydrogen bonding **Gal**β(1-3)GalNAc vs **Glc**β(1-3)GalNAc hydroxyl groups). (C) Proposed model for explaining the antifreeze activity of glycoproteins. Methyl groups of alanine and threonine residues (violet and yellow spheres, respectively) interact with ice, and the non-reducing end of the carbohydrates (blue sticks) interact with the quasi-liquid layer (QLL).

6.4. Conclusions

Antifreeze glycoproteins (AFGP) have attracted much interest in recent years due to their wide range of applications in medicine, for instance, preservation of cells, tissues and organs at low temperature or cryosurgery and nutrition for improve the texture of frozen foods. Nevertheless, the molecular mechanism by which these glycoproteins act as inhibitors of ice formation is currently unclear. In this chapter, subtle changes to the chemical structure of AFGPs have been proposed in order to disclose the molecular basis of this family of glycoproteins. By the conformational analysis and the study of the antifreeze activity of these derivatives we propose an alternative mechanism that attempts to reconcile the currently proposed models. Results obtained in this chapter reveal the importance of both the methyl groups and the carbohydrate in the interaction of these glycoproteins with ice.

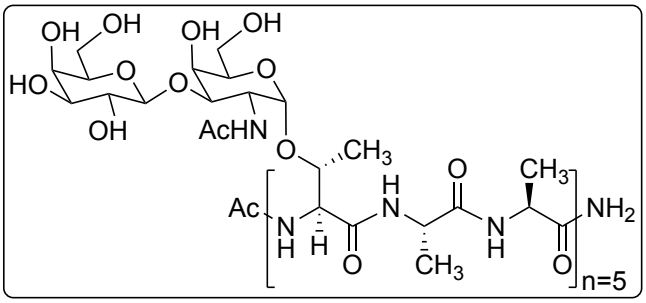
6.5. Experimental section

6.5.1. General procedure for solid-phase peptide synthesis (SPPS)

All glycopeptides were synthesized by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). Fmoc-Thr[GalNAc(Ac)$_3$-α-D]-OH was synthesized as described in the literature,[54] Fmoc-Thr[Gal(Ac)$_4$-β-(1-3)-GalNAc(Ac)$_3$-α-D]-OH and Fmoc-Ala(CD$_3$)-OH were commercially available. Dr. Asensio's group in Instituto de Química Orgánica General in CSIQ (Madrid, Spain), Dr. Ester Jimenez and Dr. Ismael Compañón members of our group have also participated in this synthesis. Non-natural amino acids were manually coupled using HBTU [(2-(1H-benzotriazol-1-yl)- 1,1,3,3-tetramethyluronium hexafluorophosphate], 0.9 equiv., and 0.25 mL of DIPEA (2.0 M in NMP) dissolved in 1 mL of DMF, while all other Fmoc amino acids (5.0 equiv) were automatically coupled using oxyma pure/DIC (N,N'-diisopropylcarbodiimide). The *O*-acetyl groups of GalNAc moiety were removed in a mixture of NH_2NH_2/MeOH (7:3) for the Tn antigen. (Glyco)peptides were then released from the resin, and all acid sensitive protecting groups were simultaneously removed using TFA 95%, TIS

(triisopropylsilane) 2.5% and H_2O 2.5%, followed by precipitation with cold diethyl ether. The crude products were purified by HPLC on a Phenomenex Luna C18(2) column (10 µm, 250 mm × 21.2 mm) using a dual absorbance detector, with a flow rate of 20 mL/min.

Synthesis of glycopeptide 33 (GalGalNAc-TAA)$_5$



Following SPPS methodology with the adequately protected amino acids, glycopeptide **33** was obtained and purified by semi-preparative HPLC.

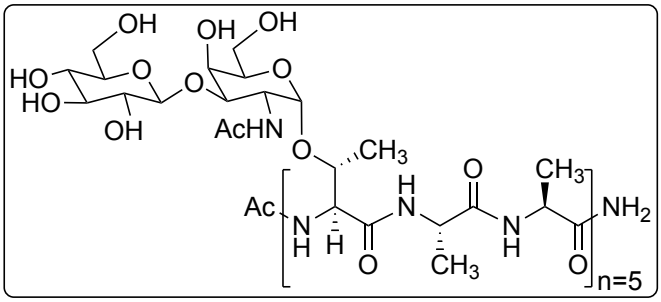
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.87 – 4.92 (m, 5H, H_{1S}), 4.30 – 4.49 (m, 25H, $H\alpha_{Thr}$, $H\alpha_{Ala}$, $H\beta_{Thr}$, H_{1S}'), 4.15 – 4.24 (m, 10H, H_{4S}, H_{2S}), 3.95– 4.11 (m, 10H, H_{5S}, H_{3S}), 3.87 – 3.91 (m, 5H, H_{6S}'), 3.68 – 3.78 (m, 20H, H_{6S}', H_{3S}', $2H_{6S}$), 3.59 – 3.67 (m, 10H, H_{4S}', H_{5S}'), 3.45 – 3.50 (m, 5H, H_2'), 2.11 (s, 3H, NHCOC$\underline{H}_3$), 1.99 – 2.05 (m, 15H, NHCOC$\underline{H}_3$), 1.34 – 1.44 (m, 30H, CH_{3Ala}), 1.26-1.33 (d, 15H, J = 6.4 Hz, CH_{3Thr}).

Semi-preparative HPLC: Rt = 20.7 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (11:89), 30 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 16.7 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (15:85), 30 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+3H]^{3+}$ Calcd for $C_{122}H_{208}N_{21}O_{71}$: 1034.4432; Found: 1034.4423.

Synthesis of glycopeptide 34 (GlcGalNAc-TAA)$_5$



Following SPPS methodology with the adequately protected amino acids, glycopeptide **34** was obtained and purified by semi-preparative HPLC.

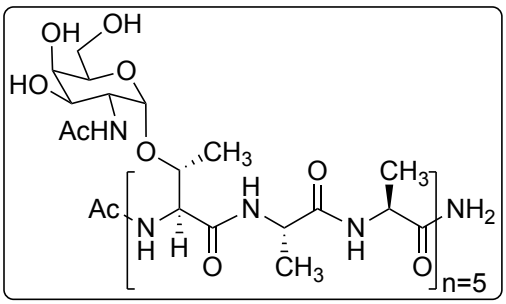
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.85 – 4.91 (m, 5H, H_{1S}), 4.29 – 4.51 (m, 25H, $H\alpha_{Thr}$, $H\alpha_{Ala}$, $H\beta_{Thr}$, H_{1S}'), 4.15 – 4.24 (m, 10H, H_{2S}, H_{4S}), 3.94– 4.03 (m, 10H, H_{5S}, H_{3S}), 3.83 – 3.88 (m, 5H, H_{6S}'), 3.66 – 3.75 (m, 15H, H_{6S}', $2H_{6S}$), 3.34– 3.44 (m, 15H, H_{3S}', H_{4S}', H_{5S}'), 3.21 – 3.29 (m, 5H, H_2'), 2.10 (s, 3H, NHCOC$\underline{H}_3$), 1.97 – 2.03 (m, 15H, NHCOC$\underline{H}_3$), 1.32 – 1.40 (m, 30H, CH_{3Ala}), 1.24-1.30 (d, 15H, J = 6.4 Hz, CH_{3Thr}).

Semi-preparative HPLC: Rt = 16.7 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (11:89), 30 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 20.4 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (10:90), 30min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{122}H_{207}N_{21}O_{71}$: 1551.1611; Found: 1551.1551.

Synthesis of glycopeptide 35 (GalNAc-TAA)$_5$



Following SPPS methodology with the adequately protected amino acids, glycopeptide **35** was obtained and purified by semi-preparative HPLC.

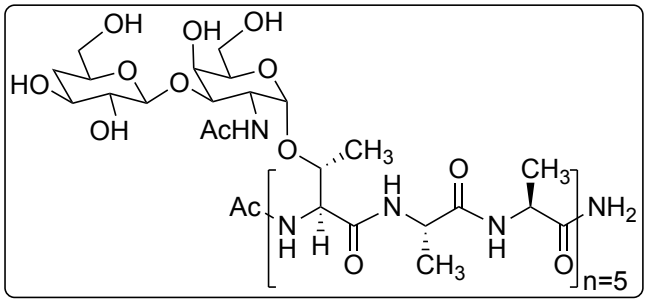
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.90 – 4.95 (m, 5H, H_{1S}), 4.48 – 4.53 (m, 5H, $H\alpha_{Thr}$), 4.29 – 4.45 (m, 15H, $H\alpha_{Ala}$, $H\beta_{Thr}$), 4.08 – 4.15 (m, 5H, H_{2S}), 3.87 – 4.06 (m, 15H, H_{3S}, H_{4S}, H_{5S}), 3.73 – 3.80 (m, 10H, $2H_{6S}$), 2.13 (s, 3H, NHCOC$\underline{H}_3$), 2.05 – 2.09 (m, 15H, , NHCOC$\underline{H}_3$), 1.36 – 1.47 (m, 30H, CH_{3Ala}), 1.30 (d, 15H, J = 6.4 Hz, CH_{3Thr}).

Semi-preparative HPLC: Rt = 21.1 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (0.5:99.5) → (15:85), 29 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 20.6 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (21:79), 40 min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{92}H_{157}N_{21}O_{46}$: 1146.0290; Found: 1146.0262.

Synthesis of glycopeptide 36 (4-deoxy-GalGalNAc-TAA)$_5$



Following SPPS methodology with the adequately protected amino acids, glycopeptide **36** was obtained and purified by semi-preparative HPLC.

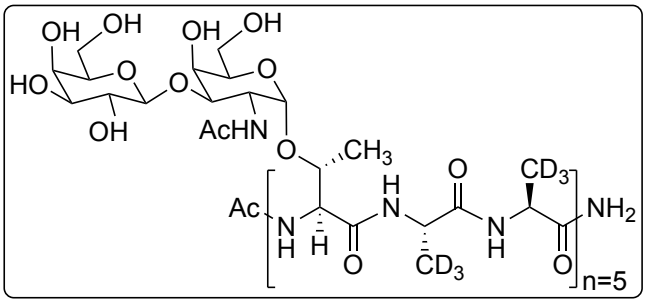
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.91 – 4.95 (m, 5H, H_{1S}), 4.24 – 4.54 (m, 35H, $H\alpha_{Thr}$, $H\alpha_{Ala}$, $H\beta_{Thr}$, H_{1S}', H_{2S}, H_{4S}), 4.01 – 4.12 (m, 10H, H_{5S}, H_{3S}), 3.59 – 3.77 (m, 30H, $2H_{6S}$, $2H_{6S}$', H_{3S}', H_{5S}'), 3.15 – 3.24 (m, 5H, H_2'), 2.14 (s, 3H, NHCOC$\underline{H}_3$), 2.03 – 2.07 (m, 15H, NHCOC$\underline{H}_3$), 1.94 – 2.01 (m, 5H, H_{4S}'), 1.38 – 1.47 (m, 35H, CH_{3Ala}, H_{4S}'), 1.29-1.35 (d, 15H, J = 6.4 Hz, CH_{3Thr}).

Semi-preparative HPLC: Rt = 19.8 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (11:89), 30 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 19.6 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (15:85), 30min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+3H]^{3+}$ Calcd for $C_{122}H_{208}N_{21}O_{66}$: 1007.7849; Found: 1007.7854.

Synthesis of glycopeptide 37 (GalGalNAc-TA$_{CD3}$A$_{CD3}$)$_5$



Following SPPS methodology with the adequately protected amino acids, glycopeptide **37** was obtained and purified by semi-preparative HPLC.

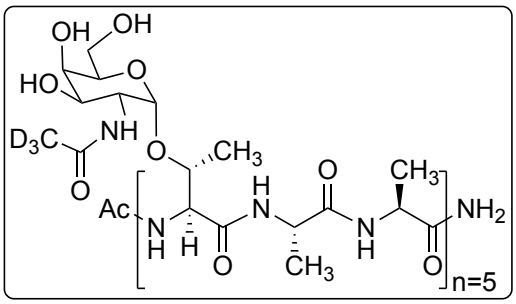
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.90 – 4.95 (m, 5H, H_{1S}), 4.34 – 4.54 (m, 25H, $H\alpha_{Thr}$, $H\alpha_{Ala}$, $H\beta_{Thr}$, H_{1S}'), 4.20 – 4.30 (m, 10H, H_{2S}, H_{4S}), 4.01– 4.11 (m, 10H, H_{5S}, H_{3S}), 3.90 – 3.94 (m, 5H, H_{6S}'), 3.73 – 3.80 (m, 20H, H_{6S}', H_{3S}', $2H_{6S}$), 3.62– 3.70 (m, 10H, H_{4S}', H_{5S}'), 3.48 – 3.54 (m, 5H, H_2'), 2.14 (s, 3H, NHCOC$\underline{H}_3$), 2.02 – 2.08 (m, 15H, NHCOC$\underline{H}_3$), 1.30-1.35 (d, 15H, *J* = 6.4 Hz, CH_{3Thr}).

Semi-preparative HPLC: Rt = 20.5 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (11:89), 30 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 16.9 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (15:85), 30min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{122}H_{177}D_{30}N_{21}O_{71}$: 1566.2552; Found: 1566.2490.

Synthesis of glycopeptide 38 (GalNAc-d³-acetamide-TAA)₅



Following SPPS methodology with the adequately protected amino acids, glycopeptide **38** was obtained and purified by semi-preparative HPLC.

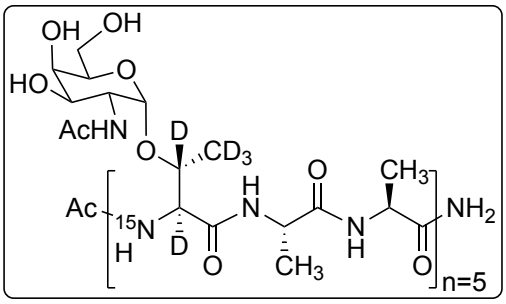
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.87 – 4.92 (m, 5H, H_{1S}), 4.44 – 4.50 (m, 5H, $H\alpha_{Thr}$), 4.20 – 4.41 (m, 15H, $H\alpha_{Ala}$, $H\beta_{Thr}$), 3.80 – 4.12 (m, 20H, H_{2S} H_{3S}, H_{4S}, H_{5S}), 3.70 – 3.77 (m, 10H, $2H_{6S}$), 2.10 (s, 3H, NHCOC$\underline{H}_3$), 1.34 – 1.46 (m, 30H, CH_{3Ala}), 1.27 (d, 15H, J = 6.3 Hz, CH_{3Thr}).

Semi-preparative HPLC: Rt = 20.2 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (0.5:99.5) → (15:85), 29 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 20.5 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (0.5:99.5) → (13:87), 25 min, 1 mL/min, λ= 212 nm).

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{92}H_{142}D_{15}N_{21}O_{46}$: 1153.5761; Found: 1153.5745.

Synthesis of glycopeptide 39 (GalNAc-Thr-d⁵-TAA)₅



Following SPPS methodology with the adequately protected amino acids, glycopeptide **39** was obtained and purified by semi-preparative HPLC.

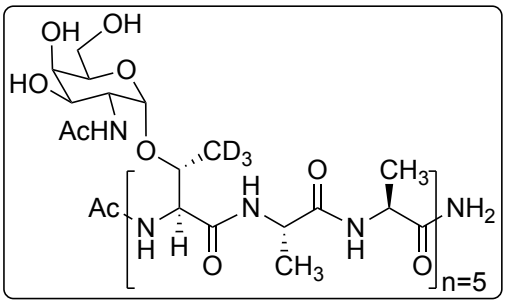
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.86 – 4.92 (m, 5H, H_{1S}), 4.30 – 4.41 (m, 10H, $H\alpha_{Ala}$), 3.80 – 4.17 (m, 20H, H_{2S}, H_{3S}, H_{4S}, H_{5S}), 3.72 (d, 10H, *J* = 5.9 Hz, $2H_{6S}$), 2.10 (s, 3H, NHCOC$\underline{H}_3$), 2.01 – 2.05 (m, 15H, NHCOC$\underline{H}_3$), 1.32 – 1.43 (m, 30H, CH_{3Ala}).

Semi-preparative HPLC: Rt = 20.3 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (0.5:99.5) → (11:89), 21 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 20.4 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (16:84), 30 min, 1 mL/min, λ= 212 nm).

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{92}H_{132}D_{25}N_{16}{}^{15}N_5O_{46}$: 1161.1001; Found: 1161.0984.

Synthesis of glycopeptide 40 (GalNAc-Thr-Me-d^3-TAA)$_5$



Following SPPS methodology with the adequately protected amino acids, glycopeptide **40** was obtained and purified by semi-preparative HPLC.

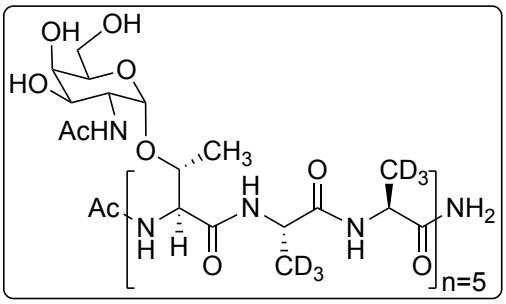
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.87 – 4.92 (m, 5H, H_{1S}), 4.44 – 4.48 (m, 5H, $H\alpha_{Thr}$), 4.24 – 4.41 (m, 15H, $H\alpha_{Ala}$, $H\beta_{Thr}$), 3.81 – 4.15 (m, 20H, H_{2S}, H_{3S}, H_{4S}, H_{5S}), 3.74 (d, 10H, *J* = 5.6 Hz, $2H_{6S}$), 2.10 (s, 3H, NHCOC$\underline{H}_3$), 2.02 – 2.05 (m, 15H, NHCOC$\underline{H}_3$), 1.33 – 1.43 (m, 30H, CH_{3Ala}).

Semi-preparative HPLC: Rt = 20.4 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (0.5:99.5) → (11:89), 21 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 20.5 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (16:84), 30 min, 1 mL/min, λ= 212 nm).

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{92}H_{142}D_{15}N_{21}O_{46}$: 1153.5761; Found: 1153.5758.

Synthesis of glycopeptide 41 (GalNAc-TA$_{CD3}$A$_{CD3}$)$_5$



Following SPPS methodology with the adequately protected amino acids, glycopeptide **41** was obtained and purified by semi-preparative HPLC.

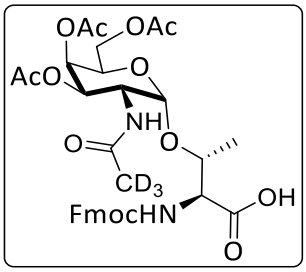
^{1}H NMR (400 MHz, D_2O) δ (ppm): 4.91 – 4.95 (m, 5H, H_{1S}), 4.48 – 4.52 (m, 5H, $H\alpha_{Thr}$), 4.29 – 4.41 (m, 15H, $H\alpha_{Ala}$, $H\beta_{Thr}$), 3.84 – 4.16 (m, 20H, H_{2S} H_{3S}, H_{4S}, H_{5S}), 3.77 (d, 10H, *J* = 5.9 Hz, $2H_{6S}$), 2.13 (s, 3H, NHCOC$\underline{H}_3$), 2.04 – 2.09 (m, 15H, NHCOC$\underline{H}_3$), 1.30 (d, 15H, *J* = 6.3 Hz, CH_{3Thr}).

Semi-preparative HPLC: Rt = 20.3 min (Phenomenex Luna C18 (2), 10 µm, 21.2×250mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (11:89), 30 min, 20 mL/min, λ = 212 nm).

Analytical HPLC: Rt = 20.7 min (Phenomenex Luna C18 (2), 5 µm, 4.6×250 mm, Grad: acetonitrile/water+0.1% TFA (1:99) → (21:79), 40min, 1 mL/min, λ= 212 nm)

HRMS (ESI) m/z: $[M+2H]^{2+}$ Calcd for $C_{122}H_{177}D_{30}N_{21}O_{71}$: 1566.2552; Found: 1566.2490.

Synthesis of Fmoc-Thr[GalNAc-d^3(Ac)$_3$-α-D]-OH (compound 42)



Fmoc-Thr [GalNAc-d3(Ac)$_3$-α-D]-O^{t}Bu (300 mg, 0.41 mmol) was dissolved in a mixture of TFA/CH_2Cl_2 (1:2, 18 mL). After stirring for 3 h at room temperature, the solvent was removed and compound **42** was obtained as a white solid without further purification (269 mg, 0.40 mmol, 97% yield).

^{1}H NMR (400 MHz, D_2O) δ (ppm): 7.78 – 7.85 (m, 2H, arom.), 7.65 – 7.73 (m, 2H, arom.), 7.28 – 7.44 (m, 4H, arom.), 5.35 – 5.43 (m, 1H, H_{4S}), 5.07 (dd, 1H, *J* = 11.5, 3.1 Hz, H_{3S}), 4.92 – 4.94 (m, 1H, H_{1S}), 4.43 – 4.62 (m, 1H, CH_{2Fmoc}), 4.34 – 4.42 (m, 2H, H_{β}, H_{2S}), 4.21 – 4.32 (m, 3H, CH_{Fmoc}, H_{5S}, H_{α}), 4.06 – 4.16 (m, 2H, H_{6S}), 2.14 (s, 3H, $COCH_3$), 2.03 (s, 3H, $COCH_3$), 1.94 (s, 3H, $COCH_3$), 1.24 (d, 3H, *J* = 6.4 Hz, CH_{3Thr}).

^{13}C NMR (100 MHz, D_2O) δ (ppm): 173.7, 173.3, 172.1, 172.1, 171.9(CON), 159.2, 145.4, 145.1, 142.7, 142.7, 128.8, 128.2, 126.2, 126.0, 121.0, 121.0 (arom.), 100.8 (C_{1S}), 77.6 (C_{β}), 69.7 (C_{3S}), 68.8 (C_{4S}), 68.2 (C_{5S}), 67.7 (CH_{2Fmoc}), 63.3(C_{6S}), 59.8 (C_{α}), 48.6 (C_{2S}, CH_{Fmoc}), 20.6, 20.6, 20.5 (CO), 19.2 (CH_{3Thr}).

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{33}H_{36}D_3N_2O_{13}$: 674.2635; Found: 674.2637.

6.5.2. Antifreeze activity evaluation

Antifreeze activity was measured by Nishimura's group in Hokkaido University (Sapporo, Japan) as thermal hysteresis using a Clifton Nanolitre Osmomer (Clifton Technical Physics, Hartford, NY), as previously described.[55] Glycopeptides were dissolved in 0.01M ammonium bicarbonate solution. The ice crystal morphology of synthetic compounds were observed using a Leica DMLB 100 photomicroscope equipped with a Linkam LK 600 temperature controller as described previously.[56] The glycopeptides were dissolved in water (2 mg/mL to 40 mg/mL), momentarily frozen (approximately – 22 °C), and warmed to 0 °C on the sample stage of the optical microscope and create several ice nucleolus in

the solution. This solution was then cooled at a rate of 0.07 °C per min, and the crystal morphologies were monitored. The photos were taken at – 0.1 °C in the case the compound have ice crystal growth inhibition activity, or at 0.0 °C for those did not have the activity.

6.5.3. MD simulations

The simulations were carried out by with AMBER 20 package[57] implemented with ff14SB[58] and GLYCAM06j[59] force fields. The glycopeptides were prepared using the GYCAM-Web utilities (https://glycam.org). Each compound was immersed in a water box with a 10 Å buffer of TIP3P water molecules.[60] A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then gently heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 30 kcal·mol^{-1} were applied to the solute, and the Andersen temperature-coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method.[61] An 8 Å cut-off was applied to Lennard-Jones interactions. Each system was equilibrated for 2 ns with a 2-fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for additional 1 μs under the same simulation conditions.

6.6. References

(1) DeVries, A. L. *Science* **1971**, *172*, 1152–1155.

(2) Yeh, Y.; Feeney, R. E. *Chem. Rev.* **1996**, *96*, 601–618.

(3) Harding, M. M.; Anderberg, P. I.; Haymet, A. D. J. *Eur. J. Biochem.* **2003**, *270*, 1381–1392.

(4) Bouvet, V.; Ben, R. N. *Cell Biochem. Biophys.* **2003**, *39*, 133–144.

(5) Tachibana, Y.; Monde, K.; Nishimura, S. I. *Macromolecules* **2004**, *37*, 6771–6779.

(6) Haridas, V.; Naik, S. *RSC Adv.* **2013**, *3*, 14199–14218.

(7) Urbańczyk, M.; Góra, J.; Latajka, R.; Sewald, N. *Amino Acids* **2017**, *49*, 209–222.

(8) Biggs, C. I.; Bailey, T. L.; Ben Graham; Stubbs, C.; Fayter, A.; Gibson, M. I. *Nat. Commun.* **2017**, *8*, 1–12.

(9) Voets, I. K. *Soft Matter* **2017**, *13*, 4808–4823.

(10) Pandey, P.; Mallajosyula, S. S. *Phys. Chem. Chem. Phys.* **2019**, *21*, 3903–3917.

(11) Ramløv, H.; Friis, D. S. *Antifreeze Proteins Volume 1: Environment, Systematics and Evolution*; Springer International Publishing, 2020.

(12) Ramløv, H.; Friis, D. S. *Antifreeze Proteins Volume 2: Biochemistry, Molecular Biology and Applications*; Springer International Publishing, 2020.

(13) Heisig, M.; Mattessich, S.; Rembisz, A.; Acar, A.; Shapiro, M.; Booth, C. J.; Neelakanta, G.; Fikrig, E. *PLoS One* **2015**, *10*, e0116562.

(14) Venketesh, S.; Dayananda, C. *Crit. Rev. Biotechnol.* **2008**, *28*, 57–82.

(15) Zbikowska, H. M. *Transgenic Res.* **2003**, *12*, 379–389.

(16) Li, B.; Sun, D. W. *J. Food Eng.* **2002**, *54*, 175–182.

(17) Knight, C. A. *Nature* **2000**, *406*, 249–251.

(18) Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S. I. *Angew. Chem. Int. Ed.* **2004**, *43*, 856–862.

(19) Sánchez, M. A.; Kling, T.; Ishiyama, T.; Van Zadel, M. J.; Bisson, P. J.; Mezger, M.; Jochum, M. N.; Cyran, J. D.; Smit, W. J.; Bakker, H. J.; Shultz, M. J.; Morita, A.; Donadio, D.; Nagata, Y.; Bonn, M.; Backus, E. H. G. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 227–232.

(20) Harding, M. M.; Ward, L. G.; Haymet, A. D. J. *Eur. J. Biochem.* **1999**, *264*, 653–665.

(21) Peltier, R.; Brimble, M. A.; Wojnar, J. M.; Williams, D. E.; Evans, C. W.; Devries, A. L. *Chem. Sci.* **2010**, *1*, 538–551.

(22) Wilkinson, B. L.; Stone, R. S.; Capicciotti, C. J.; Thaysen-Andersen, M.; Matthews, J. M.; Packer, N. H.; Ben, R. N.; Payne, R. J. *Angew. Chem. Int. Ed.* **2012**, *51*, 3606–3610.

(23) Izumi, R.; Matsushita, T.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Tsuda, S.; Hinou, H.; Nishimura, S. I. *Chem. Eur. J.* **2013**, *19*, 3913–3920.

(24) Orii, R.; Sakamoto, N.; Fukami, D.; Tsuda, S.; Izumi, M.; Kajihara, Y.; Okamoto, R. *Chem. Eur. J.* **2017**, *23*, 9253–9257.

(25) Liu, X.; Liu, J.; Wu, Z.; Chen, L.; Wang, S.; Wang, P. *Chem. Sci.* **2019**, *10*, 8694–8700.

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

(27) Graham, B.; Bailey, T. L.; Healey, J. R. J.; Marcellini, M.; Deville, S.; Gibson, M. I. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 15941–15944.

(28) Sumii, Y.; Hibino, H.; Saidalimu, I.; Kawahara, H.; Shibata, N. *Chem. Commun.* **2018**, *54*, 9749–9752.

(29) Adzhubei, A. A.; Sternberg, M. J. E.; Makarov, A. A. *J. Mol. Biol.* **2013**, *425*, 2100–2132.

(30) Mochizuki, K.; Molinero, V. *J. Am. Chem. Soc.* **2018**, *140*, 4803–4811.

(31) Mallajosyula, S. S.; Vanommeslaeghe, K.; Mackerell, A. D. *J. Phys. Chem. B* **2014**, *118*, 11696–11706.

(32) Groot, C. C. M.; Meister, K.; Devries, A. L.; Bakker, H. J. *J. Phys. Chem. Lett.* **2016**, *7*, 4836–4840.

(33) Knight, C. A.; Driggers, E.; DeVries, A. L. *Biophys. J.* **1993**, *64*, 252–259.

(34) Meister, K.; Devries, A. L.; Bakker, H. J.; Drori, R. *J. Am. Chem. Soc.* **2018**, *140*, 9365–9368.

(35) Nutt, D. R.; Smith, J. C. *J. Am. Chem. Soc.* **2008**, *130*, 13066–13073.

(36) Ben, R. N. *ChemBioChem* **2001**, *2*, 161–166.

(37) Ebbinghaus, S.; Meister, K.; Born, B.; Devries, A. L.; Gruebele, M.; Havenith, M. *J. Am. Chem. Soc.* **2010**, *132*, 12210–12211.

(38) Balcerzak, A. K.; Capicciotti, C. J.; Briard, J. G.; Ben, R. N. *RSC Adv.* **2014**, *4*, 42682–42696.

(39) Sönnichsen, F. D.; DeLuca, C. I.; Davies, P. L.; Sykes, B. D. *Structure* **1996**, *4*, 1325–1337.

(40) Davies, P. L. *Trends Biochem. Sci.* **2014**, *39*, 548–555.

(41) Yang, C.; Sharp, K. A. *Biophys. Chem.* **2004**, *109*, 137–148.

(42) Smolin, N.; Daggett, V. *J. Phys. Chem. B* **2008**, *112*, 6193–6202.

(43) Garnham, C. P.; Campbell, R. L.; Davies, P. L. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 7363–7367.

(44) Tam, R. Y.; Ferreira, S. S.; Czechura, P.; Ben, R. N.; Chaytor, J. L. *J. Am. Chem. Soc.* **2008**, *130*, 17494–17501.

(45) Drake, A. F.; Siligardi, G.; Gibbons, W. A. *Biophys. Chem.* **1988**, *31*, 143–146.

(46) Makowska, J.; Rodziewicz-Motowidło, S.; Bagińska, K.; Vila, J. A.; Liwo, A.; Chmurzyński, L.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1744–1749.

(47) Corzana, F.; Busto, J. H.; De Luis, M. G.; Fernández-Tejada, A.; Rodríguez, F.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. *Eur. J. Org. Chem.* **2010**, *2010*, 3525–3532.

(48) Celik, Y.; Drori, R.; Pertaya-Braun, N.; Altan, A.; Barton, T.; Bar-Dolev, M.; Groisman, A.; Davies, P. L.; Braslavsky, I. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1309–1314.

(49) Drori, R.; Celik, Y.; Davies, P. L.; Braslavsky, I. *J. R. Soc. Interface* **2014**, *11*.

(50) Drori, R.; Davies, P. L.; Braslavsky, I. *Langmuir* **2015**, *31*, 5805–5811.

(51) Turowski, M.; Yamakawa, N.; Meller, J.; Kimata, K.; Ikegami, T.; Hosoya, K.; Tanaka, N.; Thornton, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 13836–13849.

(52) Karim, O. A.; Haymet, A. D. J. *Chem. Phys. Lett.* **1987**, *138*, 531–534.

(53) Czechura, P.; Tam, R. Y.; Dimitrijevic, E.; Murphy, A. V; Ben, R. N. *J. Am. Chem. Soc.* **2008**, *130*, 2928–2929.

(54) Plattner, C.; Höfener, M.; Sewald, N. *Org. Lett.* **2011**, *13*, 545–547.

(55) Kao, M. H.; Fletcher, G. L.; Wang, N. C.; Hew, C. L. *Can. J. Zool.* **1986**, *64*, 578–582.

(56) Miura, K.; Ohgiya, S.; Hoshino, T.; Nemoto, N.; Suetake, T.; Miura, A.; Spyracopoulos, L.; Kondo, H.; Tsuda, S. *J. Biol. Chem.* **2001**, *276*, 1304–1310.

(57) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S.A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York D. M.; Kollman, P. A. (2020), AMBER 2020, University of California, San Francisco.

(58) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

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

(60) Kiyohara, K.; Gubbins, K. E.; Panagiotopoulos, A. Z. *Mol. Phys.* **1998**, *94*, 803–808.

(61) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

7 | Conclusions

7.1. Conclusions

7.2. Conclusiones

7.3. Scientific publication derived from this dissertation

7.4. Other scientific publications

7.5. Contribution to congresses

7.1. Conclusions

The following conclusions can be drawn from the results obtained along this dissertation:

- Diverse antigens, such as, Tn-Thr (GalNAc-α-1-*O*-Thr), Tn-Ser (GalNAc-α-1-*O*-Ser) or STn-Thr (Neu5Ac-α-(2-6)-GalNAc-α-1-*O*-Thr) have been incorporated in different peptidic fragments of MUC1 to disclose the molecular aspects of how 5E5 mAb recognizes MUC1-Tn epitopes. By conducting a multidisciplinary approach it was uncovered that 5E5 mAb is able to bind preferably to -T*-X-P-. Our studies indicate that the recognition of the GalNAc moiety is the main driving force of 5E5 mAb binding and the neighboring downstream proline residue also plays a key role. This implies that 5E5 antibody recognizes selectively the Tn antigen and can be useful as a diagnostic tool to detect Tn- and STn- expressing cancers and as a therapeutic and specific anticancer drug. Finally, using an approach based on gold nanoparticles, a serologic assay with prostate cancer patients have been performed achieving the detection of tumor-associated autoantibodies when using the epitope APGST*AP.

- An exhaustive study of the molecular basis by which two different anti-MUC1 antibodies (SM3 and 5E5) recognize their target has been performed, which has offered clues of how to modify glycopeptides in order to enhance antigen-antibody interactions. On this basis, subtle chemical modifications on the antigen structure have been accomplished and these new antigens have been incorporated into different epitopes of MUC1. Results obtained with affinity assays have demonstrated how rational structure-based design of new antigens can lead to more efficient antigens in terms of affinity for anti-MUC1 antibodies.

Conclusions

- A serologic assay using gold nanoparticles conjugated to diverse glycopeptides has been performed to detect anti-MUC1 autoantibodies in sera from patients with pancreatic cancer. Notably, significant differences between healthy patients and patients with pancreatic cancer were observed when using glycopeptides containing the 5E5 epitope. These results demonstrate the presence of 5E5 antibodies in patients with pancreatic cancer and the efficiency of epitope APGST*AP to detect them.

- By the conformational analysis and the study of the antifreeze activity of diverse synthetic AFGPs derivatives we have disclosed an alternative molecular mechanism by which these glycoproteins act as inhibitors of ice formation. Our model proposes that both, the methyl groups and the carbohydrates play a key role in the interaction with ice, and therefore in the antifreeze activity of these glycoproteins.

7.2. Conclusiones

De los resultados obtenidos a lo largo de esta tesis se pueden extraer las siguientes conclusiones:

- Se han incorporado diversos antígenos, como, Tn-Thr (GalNAc-α-1-*O*-Thr), Tn-Ser (GalNAc-α-1-*O*-Ser) o STn-Thr (Neu5Ac-α-(2-6)-GalNAc-α-1-*O*-Thr) en diferentes fragmentos peptídicos de MUC1 para descubrir los aspectos moleculares de cómo el anticuerpo 5E5 reconoce los epítopos de MUC1-Tn. Al llevar a cabo un enfoque multidisciplinar se descubrió que el anticuerpo 5E5 es capaz de unirse preferentemente a -T*-X-P-. Nuestros estudios indican que el reconocimiento de la fracción GalNAc es la principal fuerza motriz de la unión del anticuerpo 5E5 y que el residuo de prolina adyacente desempeña también un papel clave. Esto implica que el anticuerpo 5E5 reconoce selectivamente el antígeno Tn y puede ser útil como herramienta de diagnóstico para detectar cánceres que expresen Tn- y STn- y como fármaco terapéutico y específico contra el cáncer. Finalmente, utilizando una aproximación basada en nanopartículas de oro, se ha realizado un ensayo serológico con pacientes de cáncer de próstata logrando la detección de auto-anticuerpos asociados al tumor cuando se utiliza el epítopo APGST*AP.

- Se ha realizado un estudio exhaustivo de las bases moleculares por las que dos anticuerpos anti-MUC1 (SM3 y 5E5) reconocen su diana, lo que ha ofrecido pistas de cómo modificar los glicopéptidos para mejorar las interacciones antígeno-anticuerpo. Sobre esta base, se han logrado sutiles modificaciones químicas en la estructura del antígeno y estos nuevos antígenos se han incorporado a diferentes epítopos de MUC1. Los resultados obtenidos con los ensayos de afinidad han demostrado cómo el diseño racional basado en la estructura de los nuevos antígenos puede conducir a antígenos más eficientes en términos de afinidad por los anticuerpos anti-MUC1.

- Se ha realizado un ensayo serológico utilizando nanopartículas de oro conjugadas con diversos glicopéptidos para detectar auto-anticuerpos anti-MUC1 en el suero de pacientes con cáncer de páncreas. En particular, se observaron diferencias significativas entre los pacientes sanos y los pacientes con cáncer de páncreas cuando se utilizaron glicopéptidos que contenían el epítopo 5E5. Estos resultados demuestran la presencia de anticuerpos 5E5 en pacientes con cáncer de páncreas y la eficacia del epítopo APGST*AP para detectarlos.

- Mediante el análisis conformacional y el estudio de la actividad anticongelante de diversos derivados AFGPs sintéticos hemos propuesto un mecanismo molecular alternativo por el que estas glicoproteínas actúan como inhibidores de la formación de hielo. Nuestro modelo expone que tanto los grupos metilo como los carbohidratos juegan un papel clave en la interacción con el hielo, y por tanto en la actividad anticongelante de estas glicoproteínas.

7.3. Scientific publications derived from this dissertation

- Derived from Chapter 4:

  'Structural characterization of an unprecedented lectin-like antitumoral anti-MUC1 antibody'
  Macías-León, J.; Bermejo, I. A.; Asín, A.; García-García, A.; Compañón, I.; Jiménez-Moreno, E.; Coelho, H.; Mangini, V.; Albuquerque, I. S.; Marcelo, F.; Asensio, J. L.; Bernardes, G. J. L.; Joshi, H. J.; Fiammengo, R.; Blixt, O.; Hurtado-Guerrero, R.; Corzana, F.
  Chemical Communication **2020**, *56*, 15137–15140.

7.4. Other scientific publications

- 'Structure-based design of anti-cancer vaccines: the significance of antigen presentation to boost the immune response'
  Asín, A.; García-Martín, F.; Busto, J. H.; Avenoza, A.; Peregrina, J. M.; Corzana, F.
  Current Medicinal Chemistry **2022**, *29*, 1258–1270.

- 'Cell-penetrating peptides containing fluorescent D-cysteines'
  Navo, C. D.; Asín, A.; Gómez-Orte, E.; Gutiérrez-Jiménez, M. I.; Compañón, I.; Ezcurra, B.; Avenoza, A.; Busto, J. H.; Corzana, F.; Zurbano, M. M.; Jiménez-Osés, G.; Cabello, J.; Peregrina, J. M.
  Chemistry - A European Journal **2018**, *24*, 7991–8000.

7.5. Contribution to congresses

- Poster presentation: 'Stereoselective *S*-Michael additions of sulphur nucleophiles to a chiral dehydroalanine. Synthesis of fluorescent peptides' Asín, A.; Gutiérrez-Jiménez, M. I.; Gómez-Orte, E.; Navo, C. D.; Compañón, I.; Ezcurra, B.; Avenoza, A.; Busto, J. H.; Corzana, F.; Zurbano, M. M.; Jiménez-Osés, G.; Cabello, J.; Peregrina, J. M.
  16th Iberian peptide meeting 4th ChemBio Group Meeting. Barcelona (Spain). February 5-7th of 2018.

- Poster presentation: 'Structure-based design of new MUC1 derivatives for detection of antibodies in patients with cancer' Asín, A.; Jiménez-Moreno, E.; Peregrina, J. M.; Corzana, F.
  6-ECBS/LS-EuCheMS. Madrid (Spain). April 3-5th of 2019.

- Poster presentation: 'Structure-based design of new MUC1 derivatives for detection of antibodies in patients with cancer' Asín, A.; Jiménez-Moreno, E.; Peregrina, J. M.; Corzana, F.
  XXXVII Bienal RSEQ. San Sebastián (Spain). May 26-30th of 2019.

- Poster presentation: 'Diseño de nuevos derivados de MUC1 para la detección de anticuerpos en pacientes con cáncer.' Asín, A.; Jiménez-Moreno, E.; Peregrina, J. M.; Corzana, F.
  X Jornadas de Química CISQ. Logroño (Spain). June 20-21st of 2019.

- Poster presentation: 'Exploring the use of unnatural MUC1 derivatives for detection of antibodies in cancer patients.' Asín, A.; Compañón, I.; Fiammengo, R.; Avenoza, A.; Busto, J. H.; Peregrina, J. M.; Hurtado-Guerrero, R; Corzana, F.
  V GEQB ChemBio Group Meeting. Granada (Spain). February 19-21st of 2020.

- Oral communication: 'Caracterización estructural de anticuerpos anti-MUC1 y su aplicación para el diagnóstico de cáncer' Asín, A.
  XI Jornadas de Química CISQ. Logroño (Spain). December 2-3rd of 2021.

8 Supplementary material

8.1. Reagents and general procedures

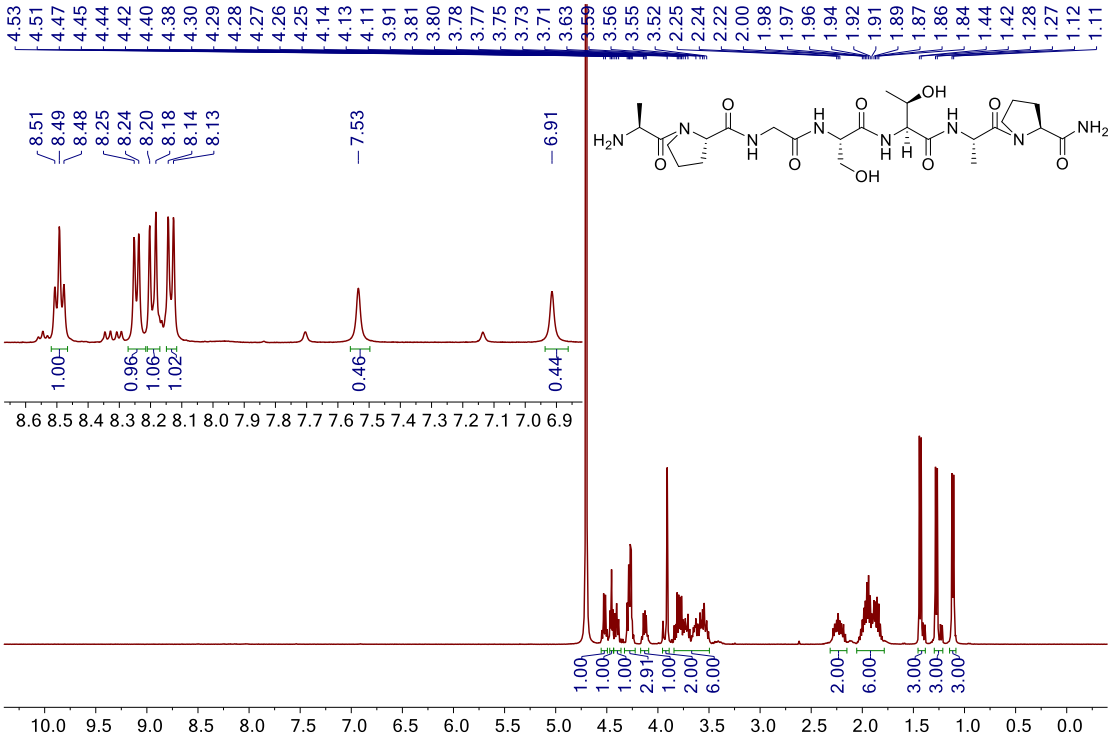
Commercial reagents were used without further purification. Analytical thin layer chromatography (TLC) was performed on Macherey-Nagel precoated aluminum sheets with a 0.20 mm thickness of silica gel 60 with fluorescent indicator UV254. TLC plates were visualized with UV light and by staining with phosphomolybdic acid (PMA) solution (5 g of PMA in 100 mL of absolute ethanol) or sulfuric acid-ethanol solution (1:20). Column chromatography was performed on silica gel (230–400 mesh). ^{1}H, ^{13}C and ^{19}F NMR spectra were measured with a 400 MHz spectrometer with TMS as the internal standard. Multiplicities are quoted as singlet (s), broad singlet (br s), doublet (d), doublet of doublets (dd), triplet (t), quadruplet (q) or multiplet (m). Signals were assigned using COSY and HSQC experiments. NMR chemical shifts (δ) were reported in ppm and coupling constants (*J*) in Hz. 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.

8.2. NMR experiments

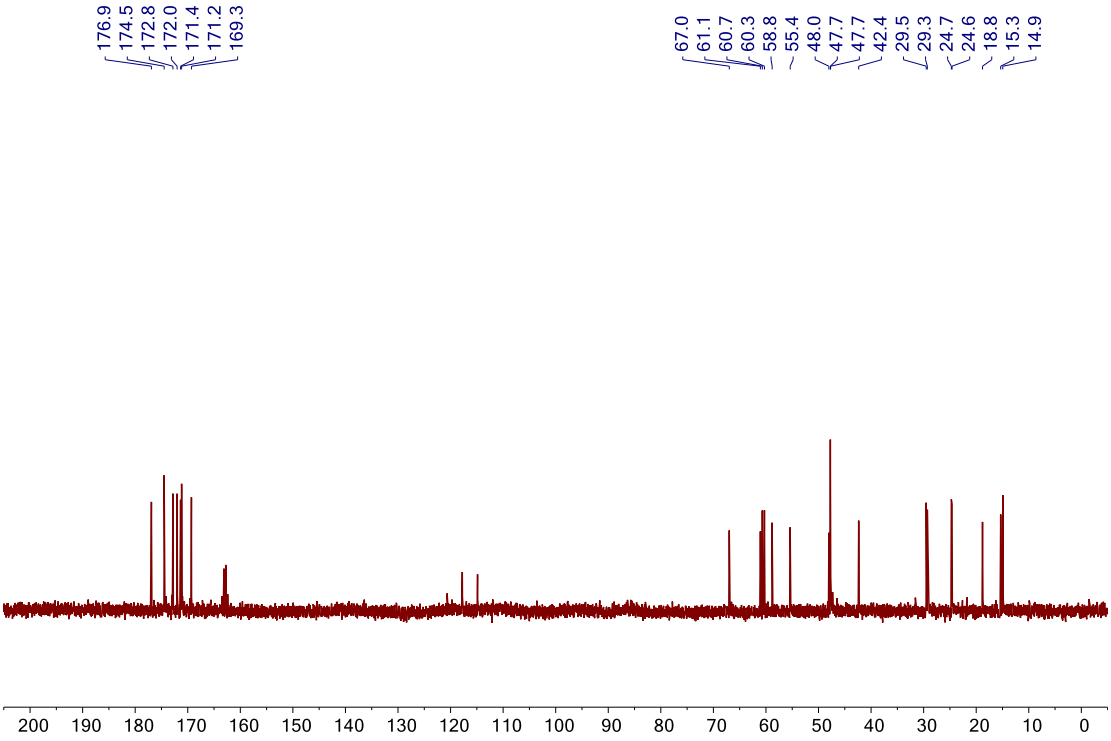
NMR experiments were performed on a 400 MHz spectrometer at 298K. Magnitude-mode ge-2D COSY spectra were acquired with gradient by using the *cosygpqf* pulse program with a pulse width of 90°. Phase-sensitive ge-2D HSQC spectra were acquired by using z-filter and selection before t1 removing the decoupling during acquisition by use of the *invigpndph* pulse program with CNST2 (*J*HC) = 145. Chemical shifts are given in ppm (δ) and coupling constants (*J*) in Hz. The results were processed using MestReNova software.

8.3. Supplementary material of Chapter 4

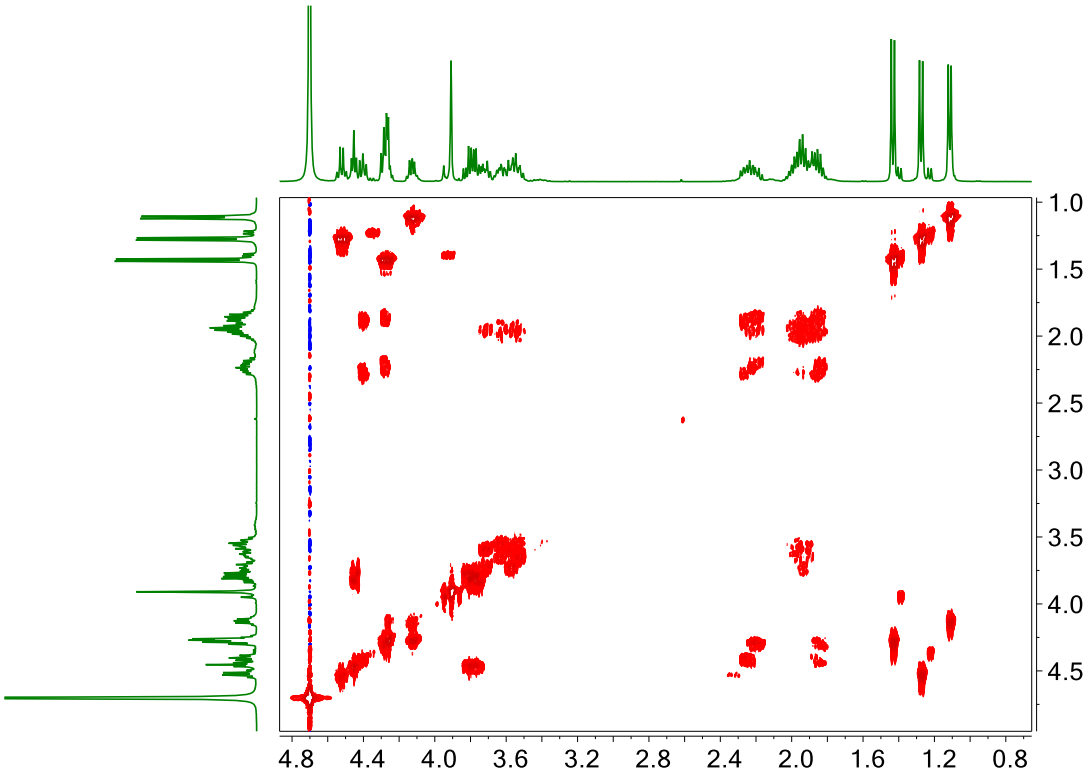
^{1}H NMR 400 MHz in D_2O registered at 298 K of **peptide 1**



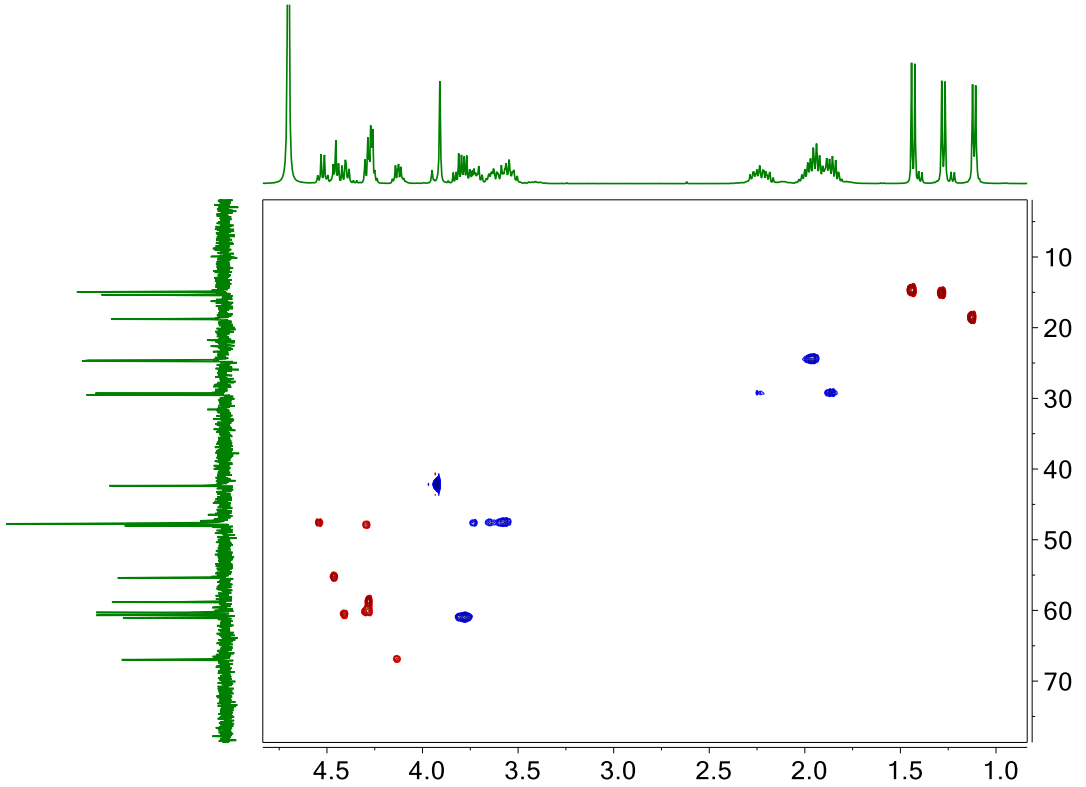
^{13}C NMR 100 MHz in D_2O registered at 298 K of **peptide 1**



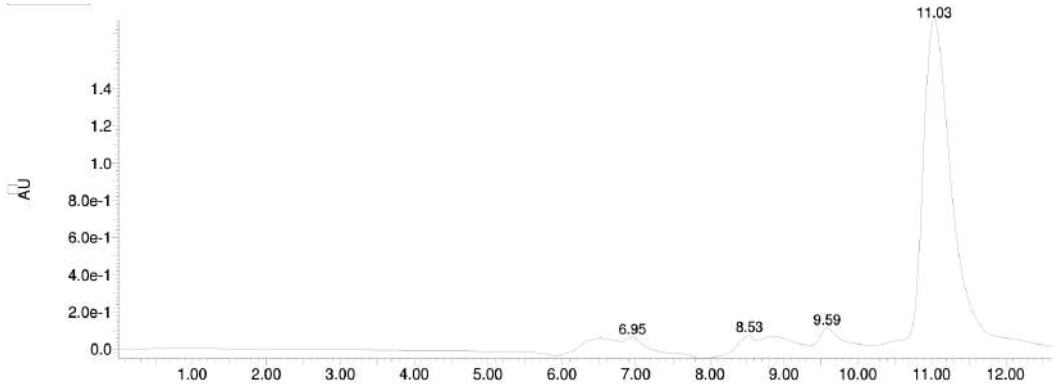
COSY in D_2O registered at 298 K of **peptide 1**



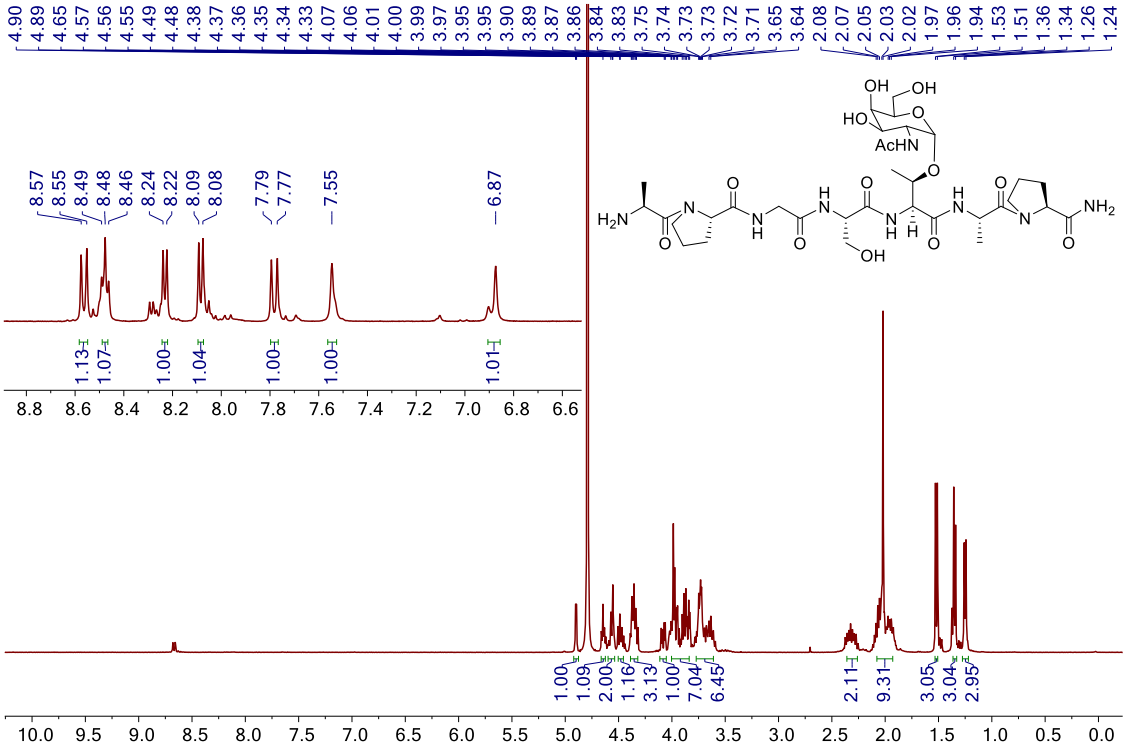
HSQC in D_2O registered at 298 K of **peptide 1**



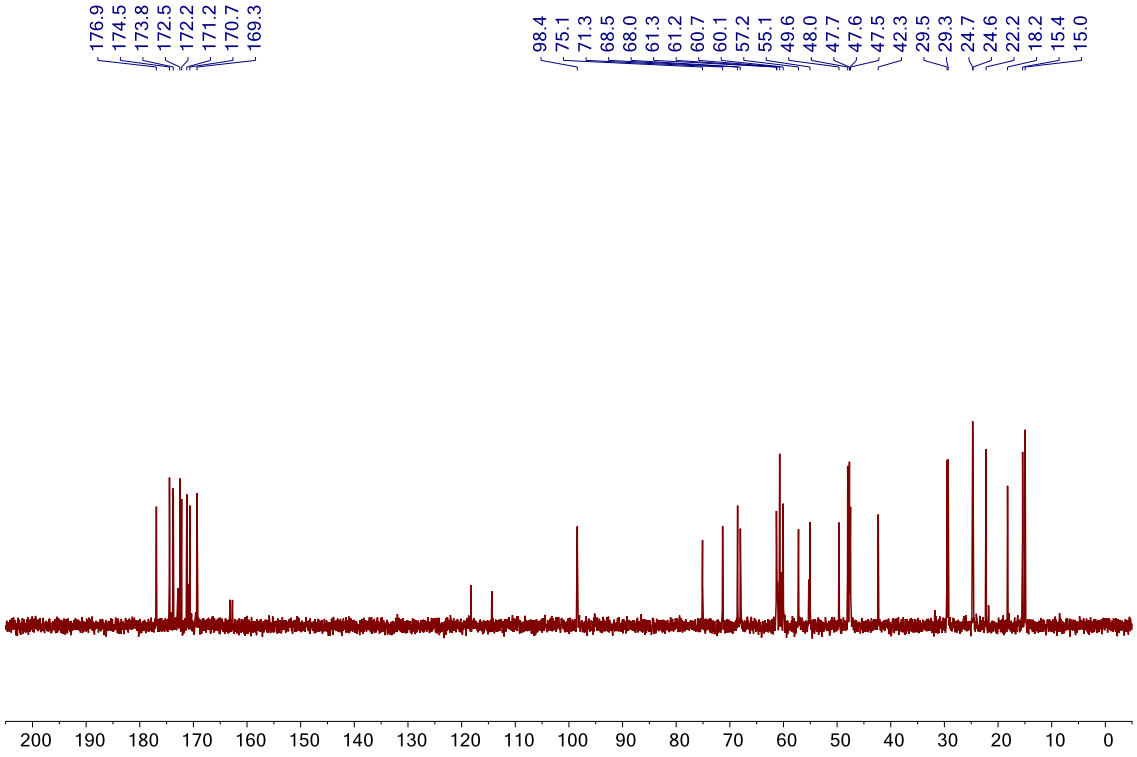
Semi-preparative HPLC chromatogram of **peptide 1**



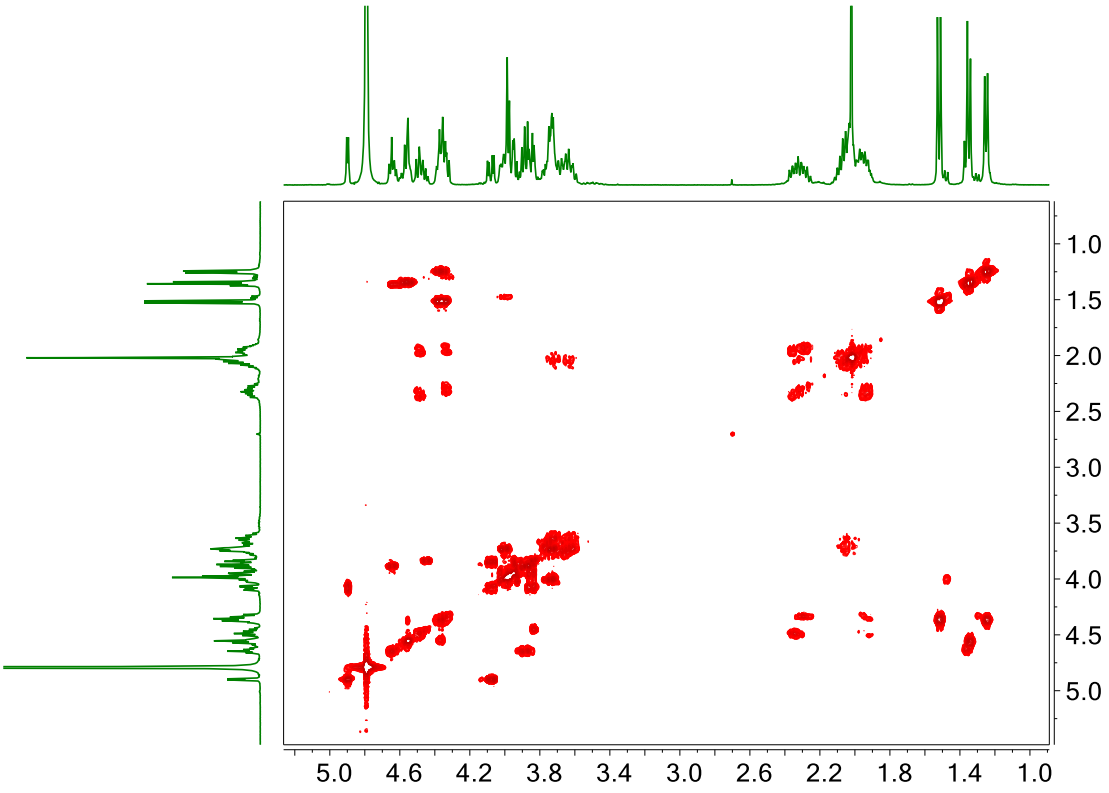
^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 2**



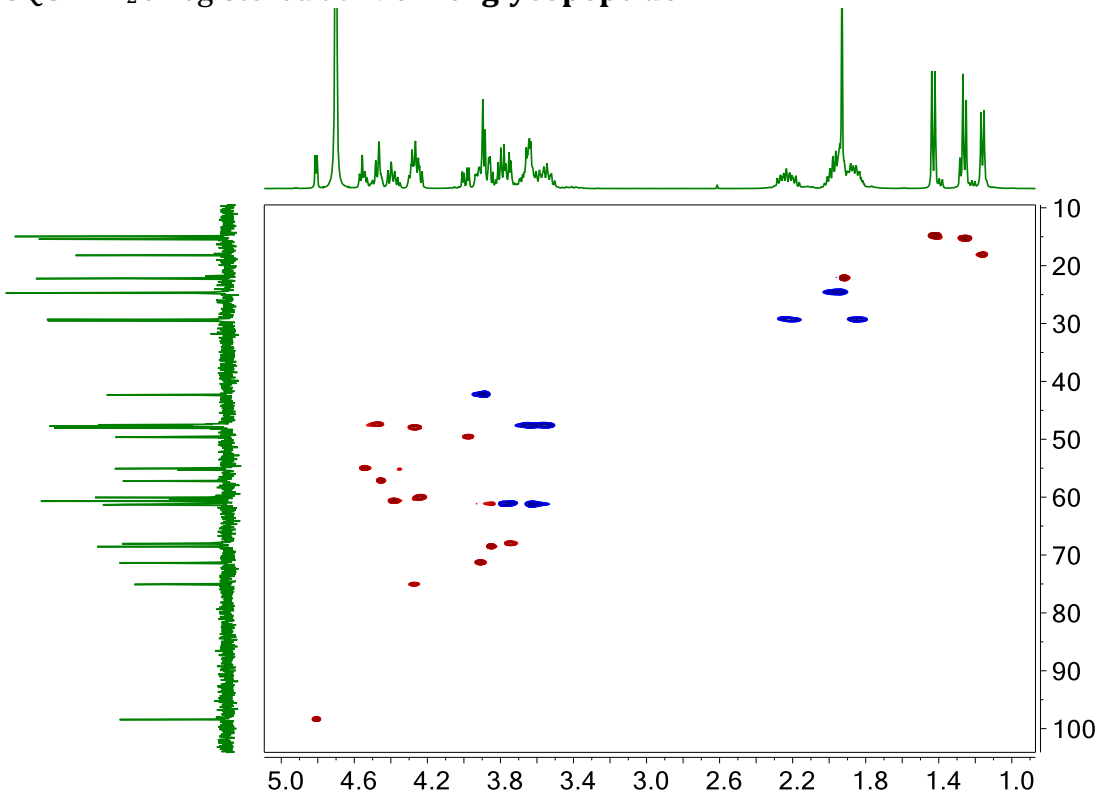
^{13}C NMR 100 MHz in D_2O registered at 298 K of **glycopeptide 2**



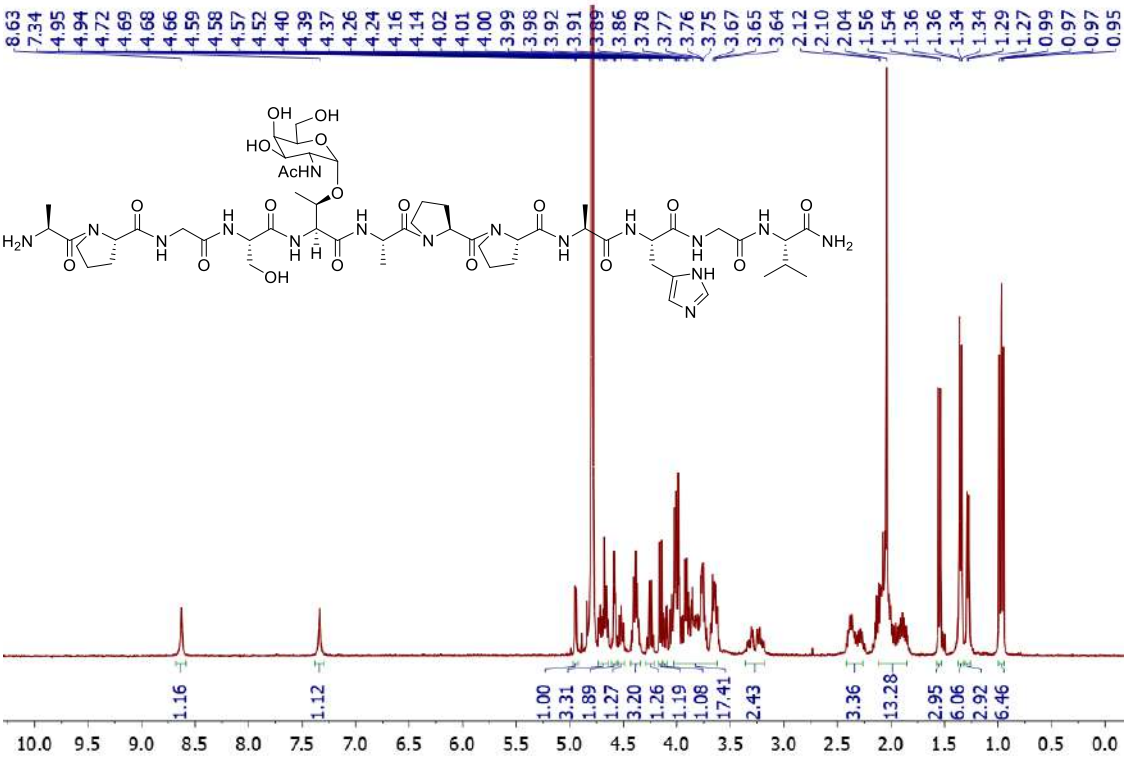
COSY in D_2O registered at 298 K of **glycopeptide 2**



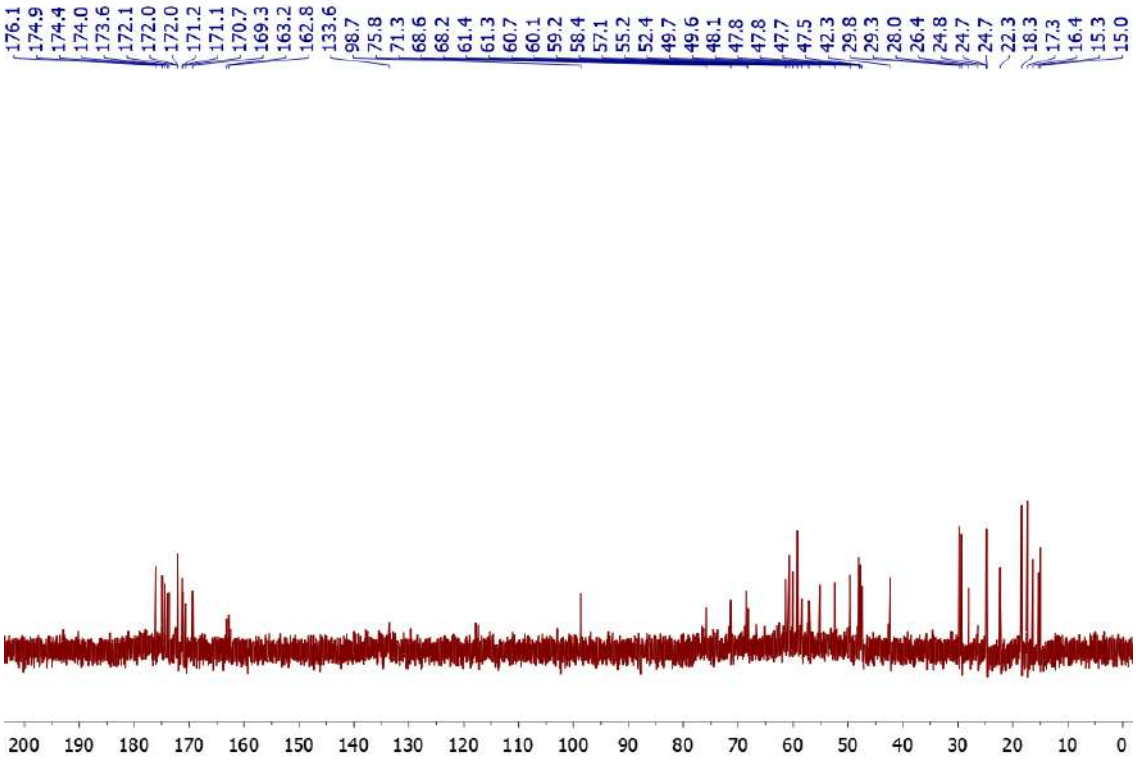
HSQC in D_2O registered at 298 K of **glycopeptide 2**



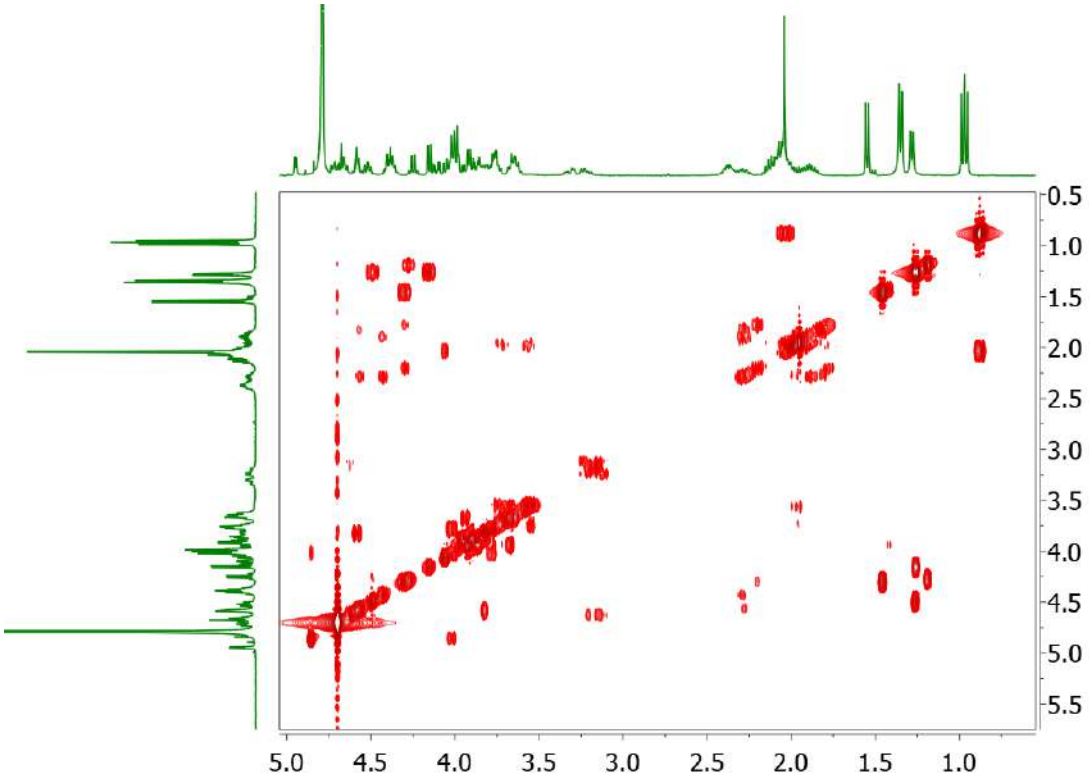
^1H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 3**



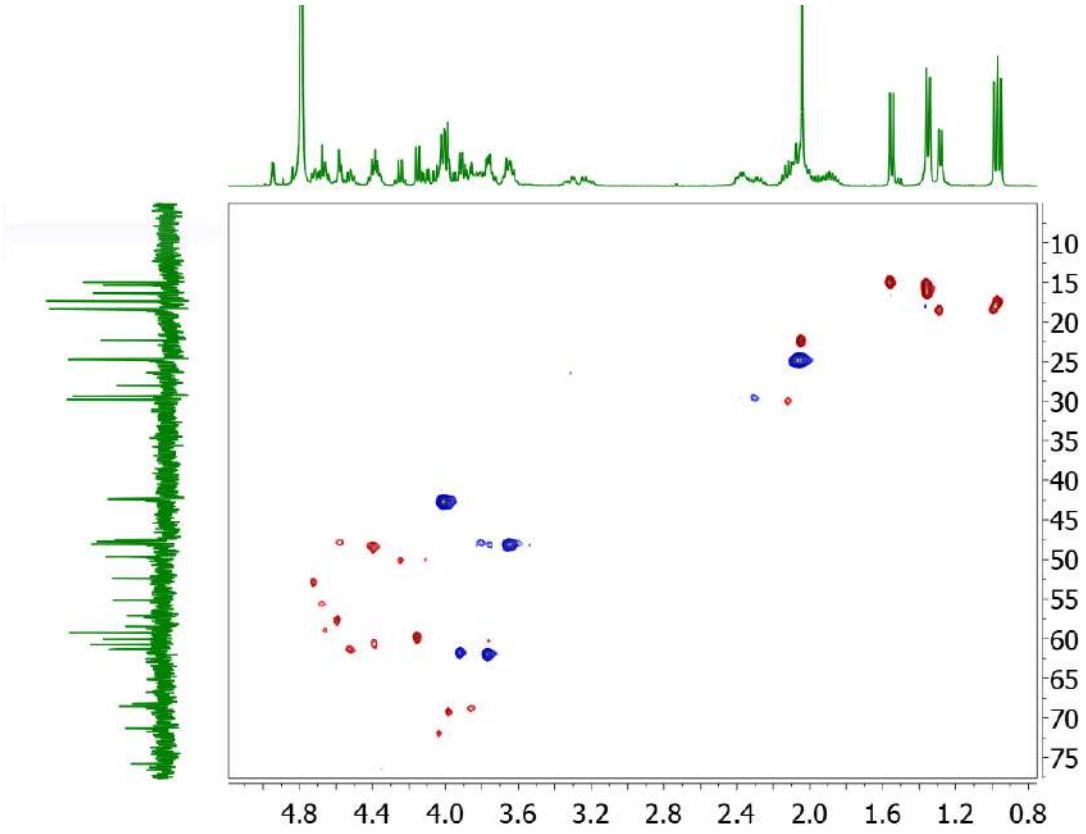
^{13}C NMR 100 MHz in D_2O registered at 298 K of **glycopeptide 3**



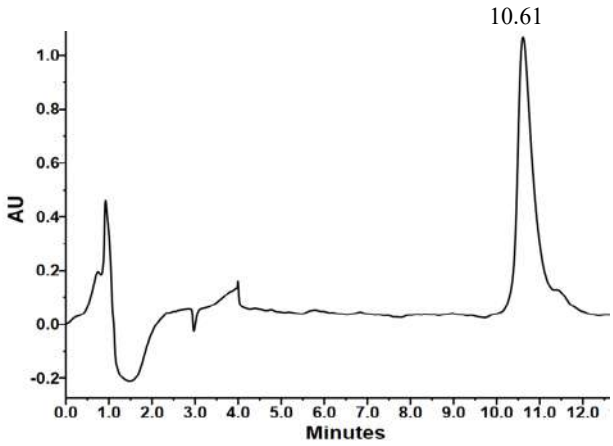
COSY in D_2O registered at 298 K of **glycopeptide 3**



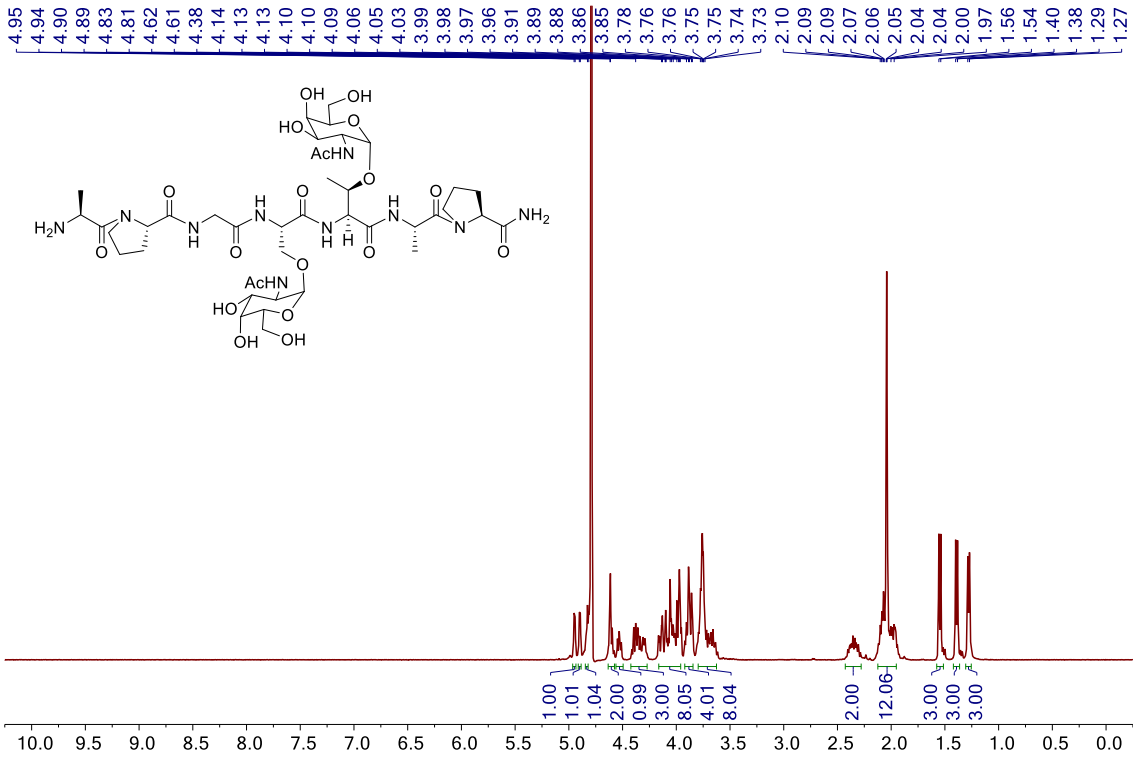
HSQC in D_2O registered at 298 K of **glycopeptide 3**



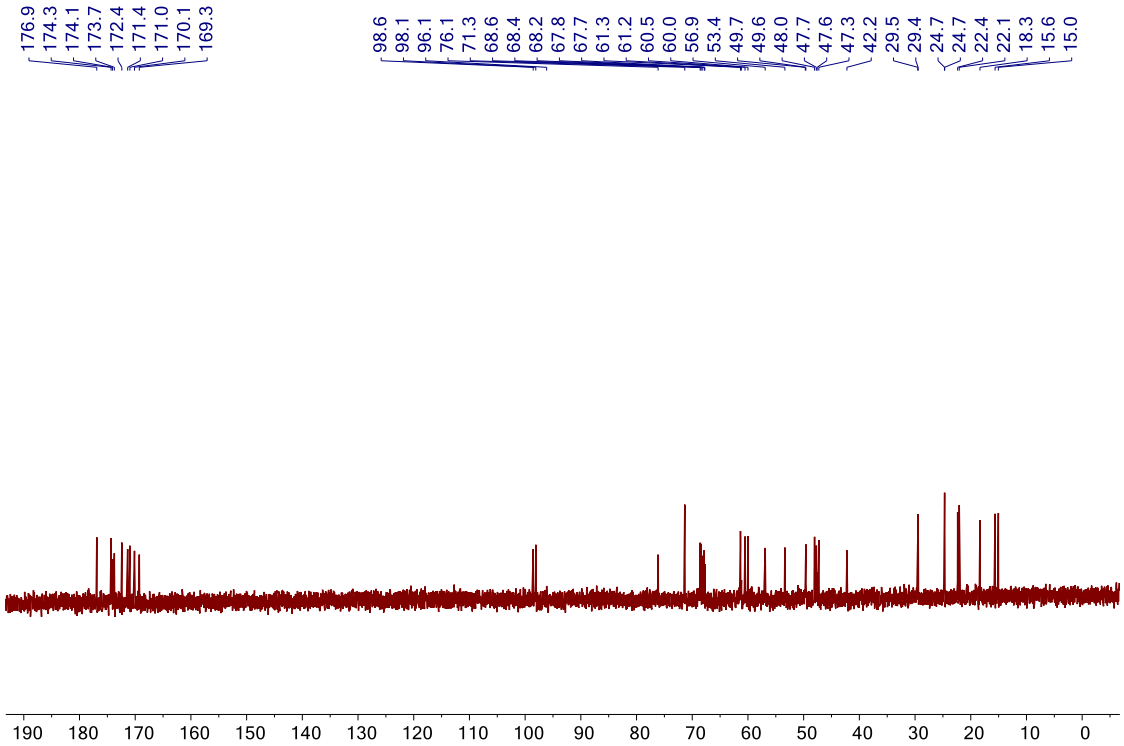
Analytical HPLC chromatogram of **glycopeptide 3**



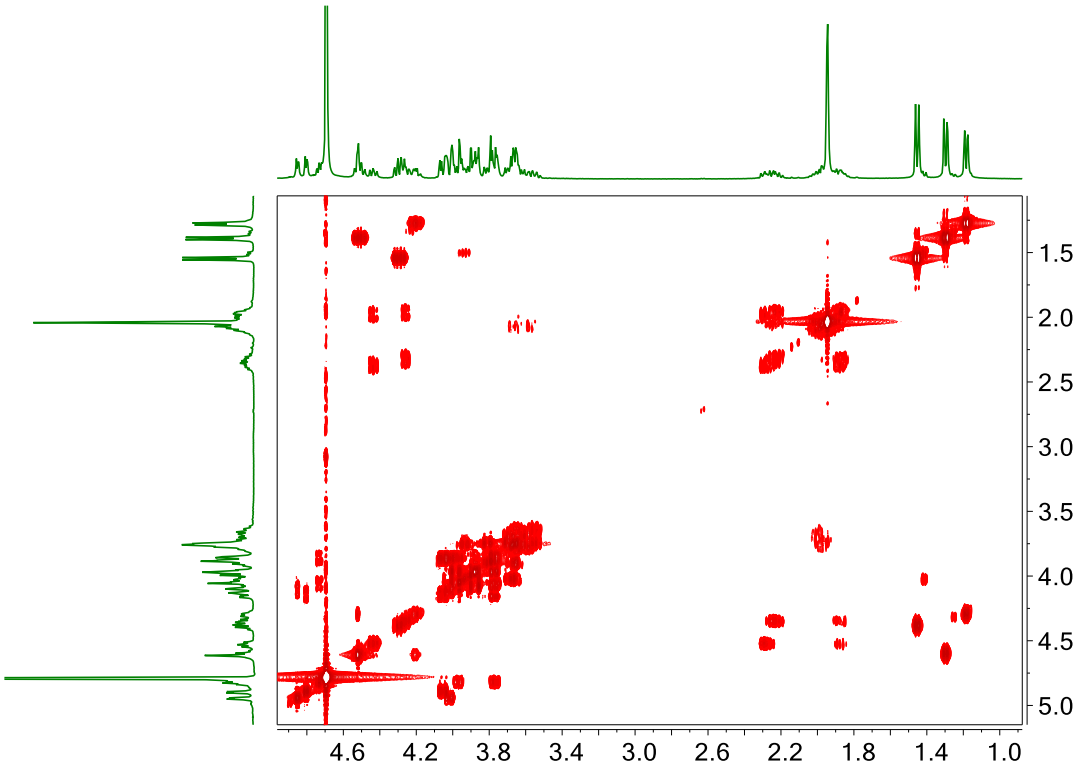
^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 5**



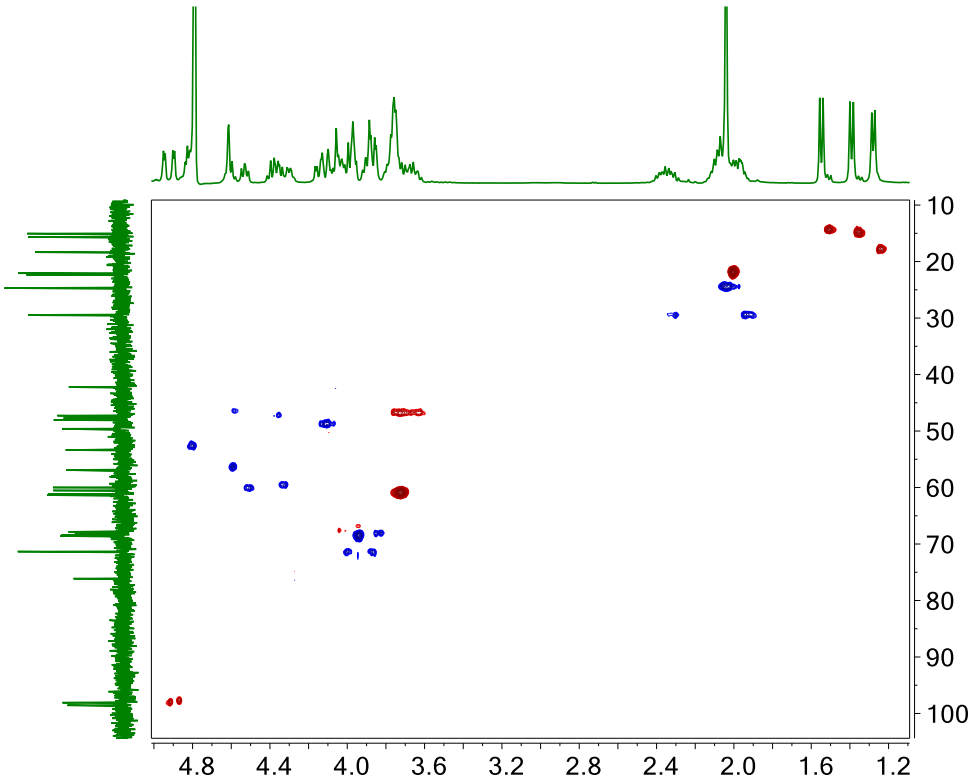
^{13}C NMR 100 MHz in D_2O registered at 298 K of **glycopeptide 5**



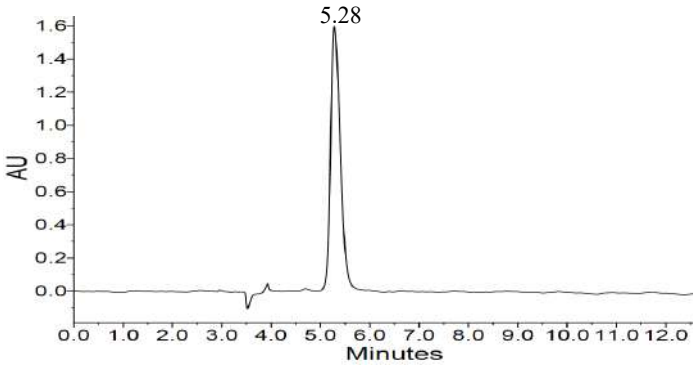
COSY in D_2O registered at 298 K of **glycopeptide 5**



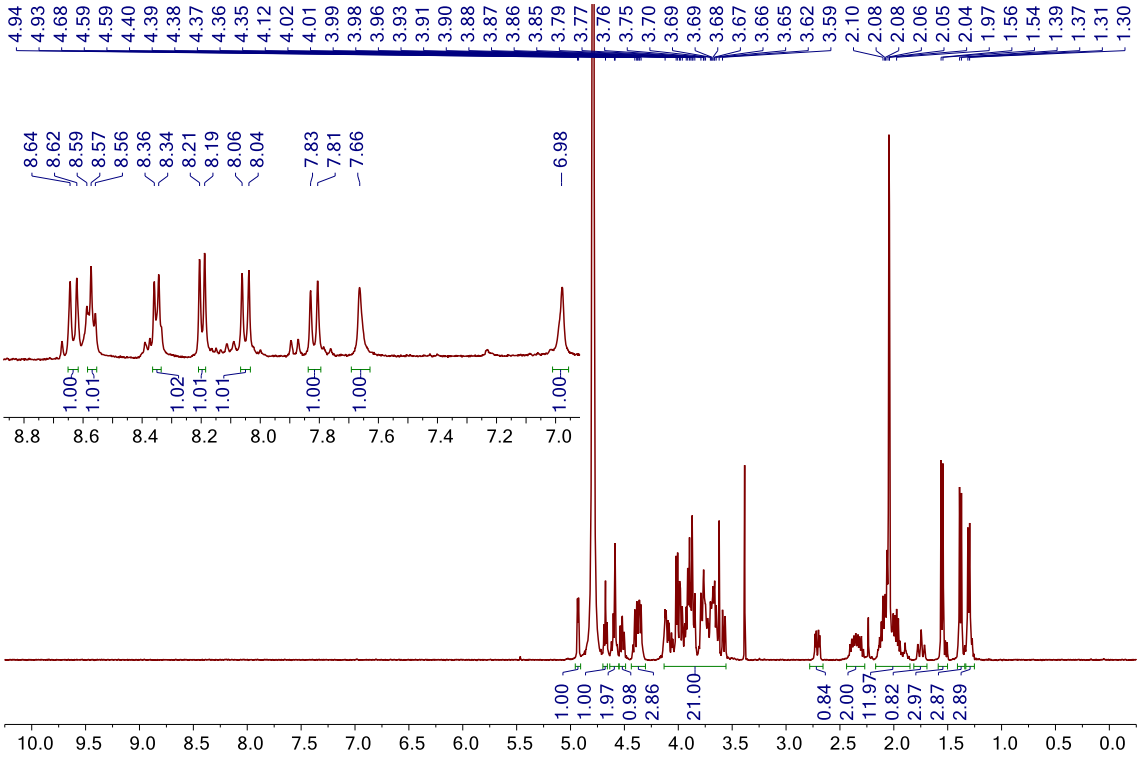
HSQC in D_2O registered at 298 K of **glycopeptide 5**



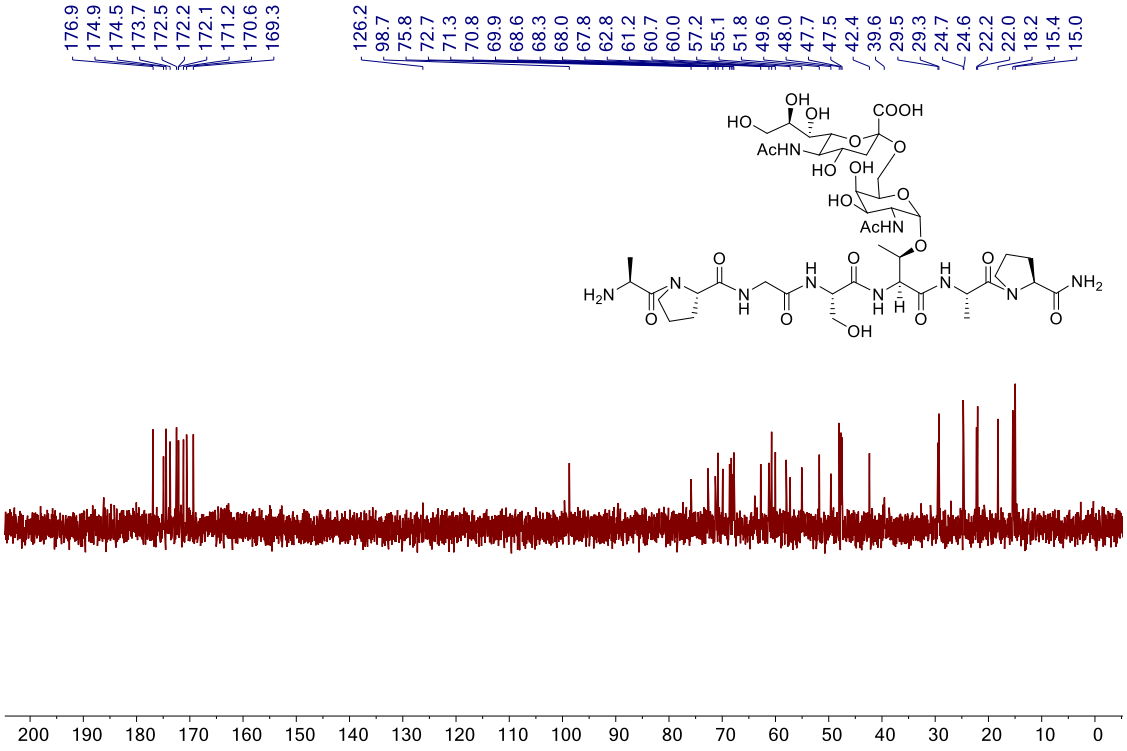
Analytical HPLC chromatogram of **glycopeptide 5**



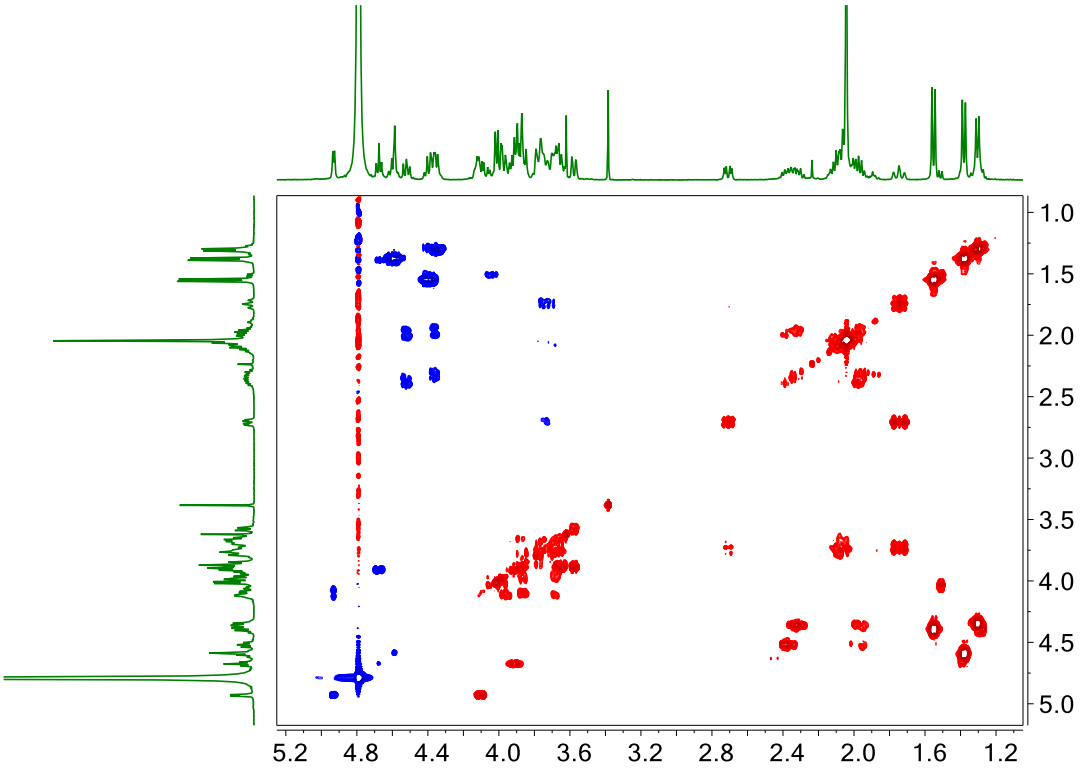
^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 6**



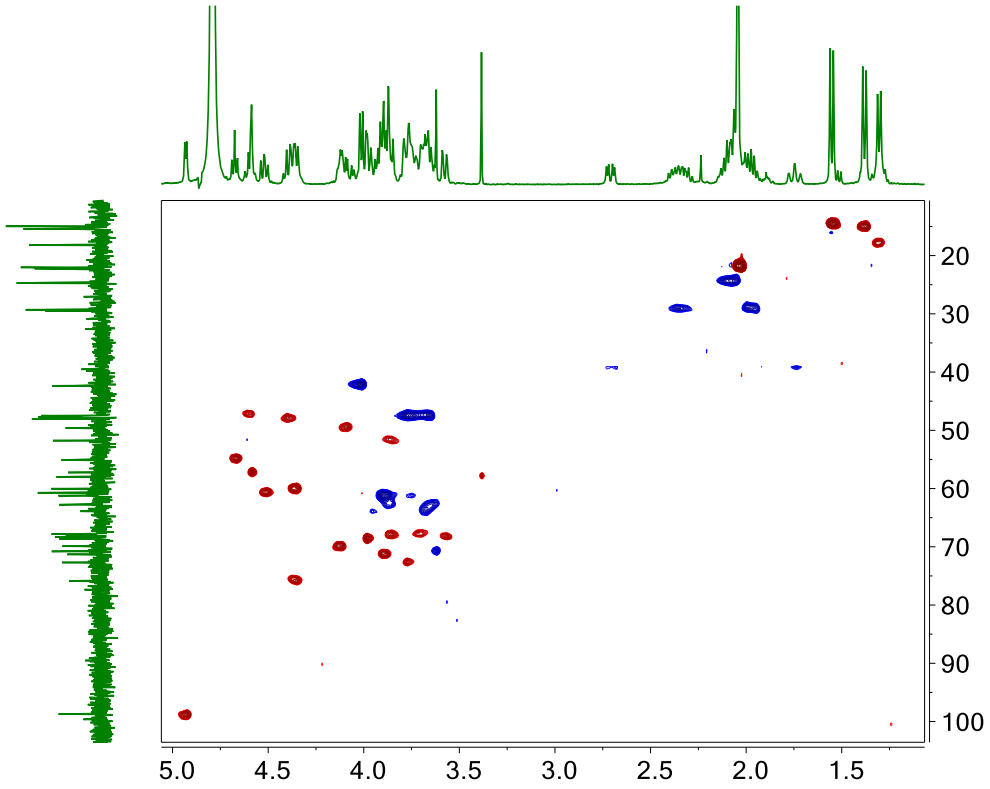
^{13}C NMR 100 MHz in D_2O registered at 298 K of **glycopeptide 6**



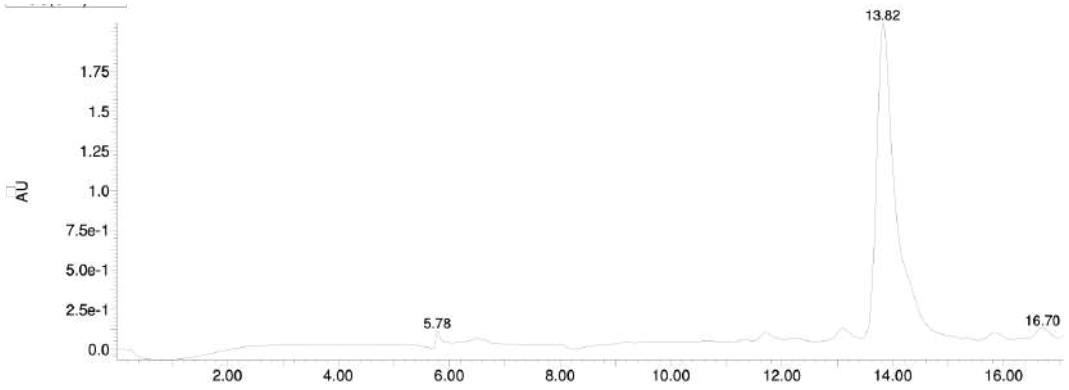
COSY in D_2O registered at 298 K of **glycopeptide 6**



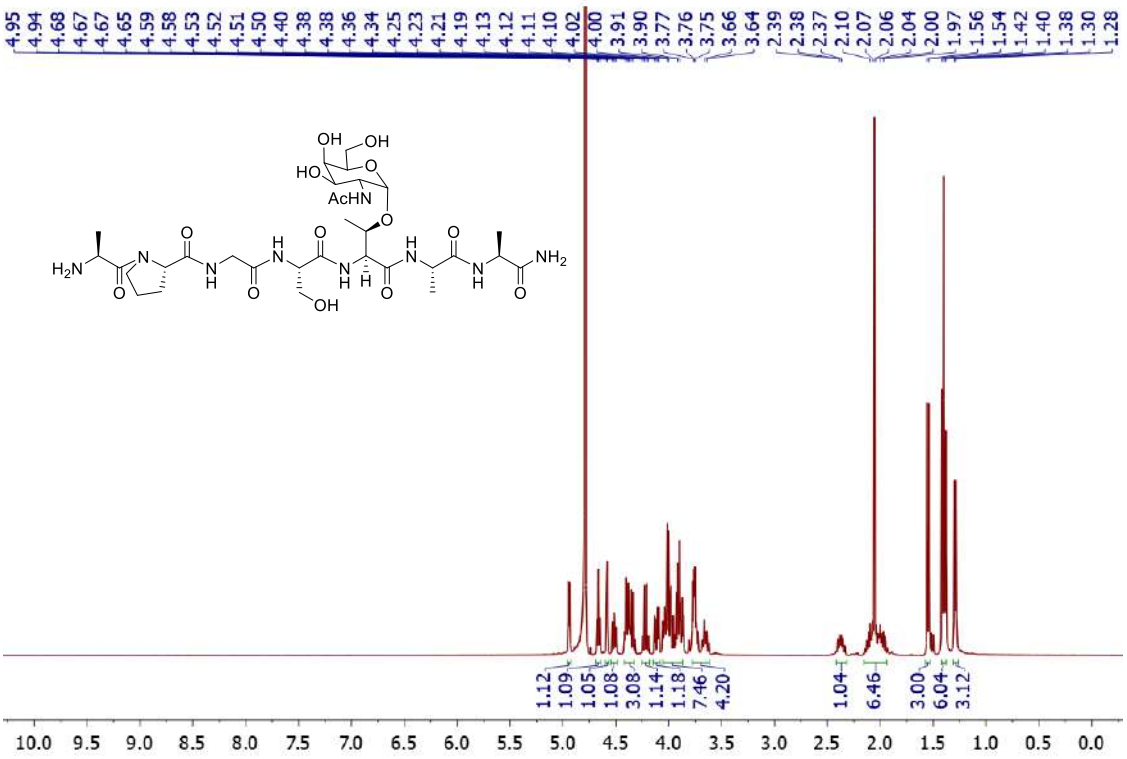
HSQC in D_2O registered at 298 K of **glycopeptide 6**



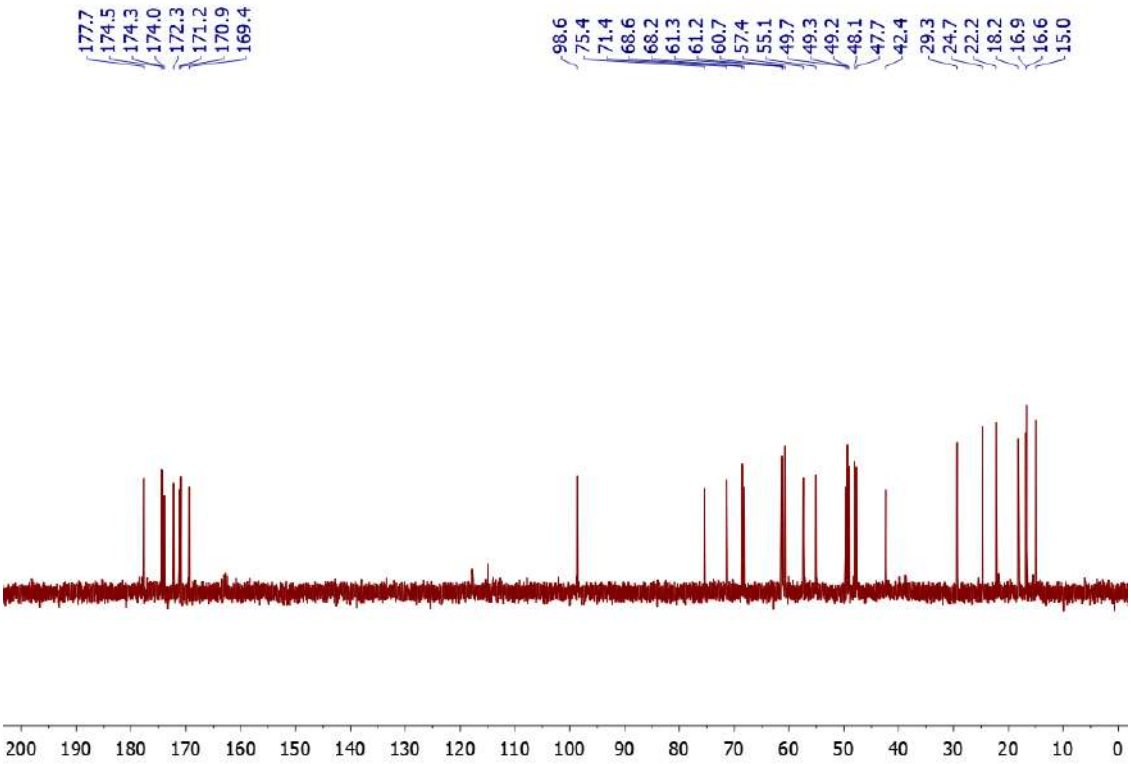
Semi-preparative HPLC chromatogram of **glycopeptide 6**



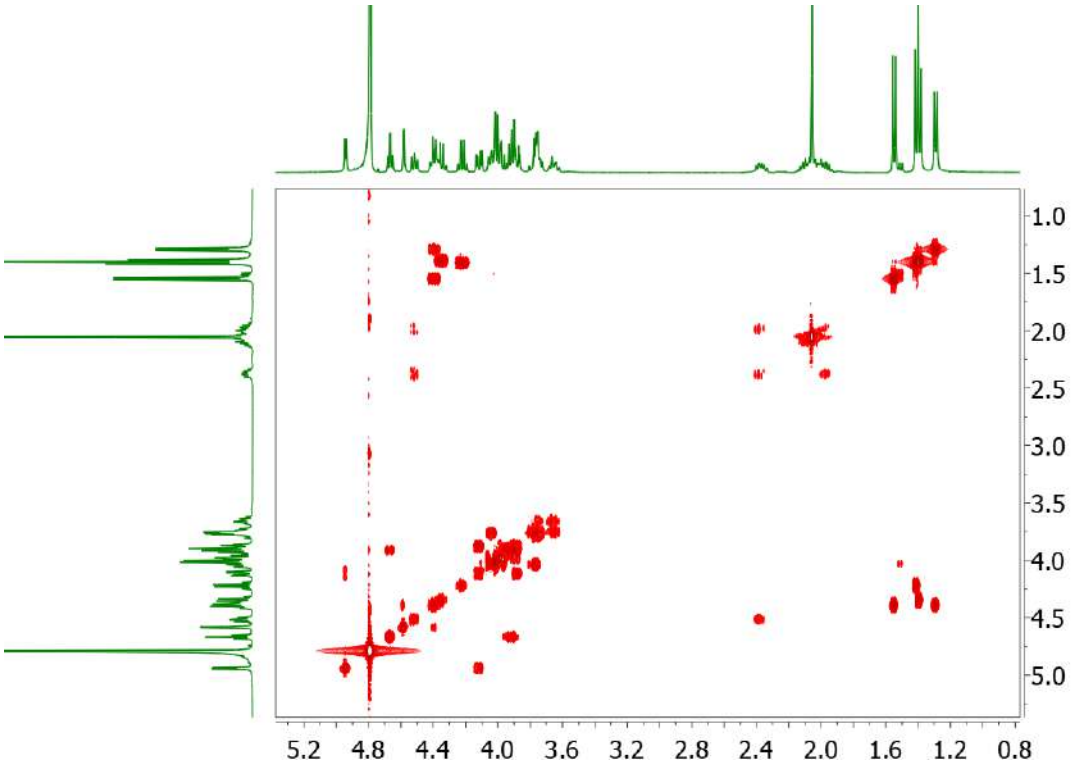
^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 7**



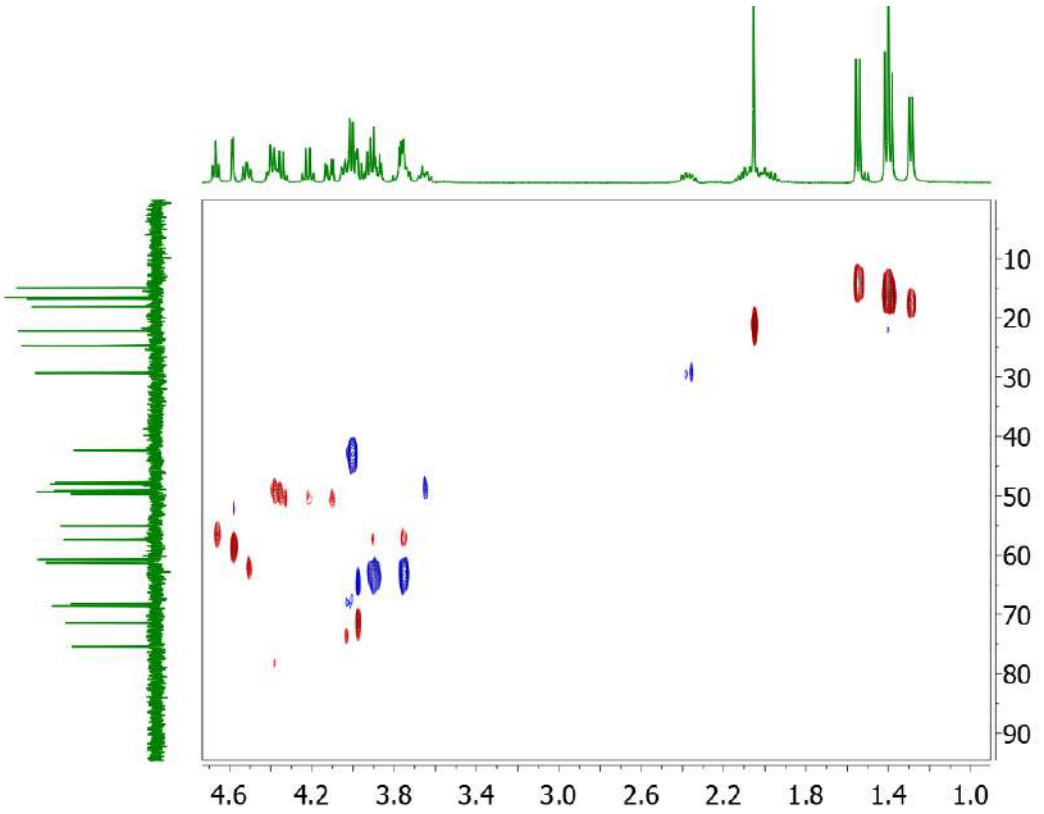
^{13}C NMR 100 MHz in D_2O registered at 298 K of **glycopeptide 7**



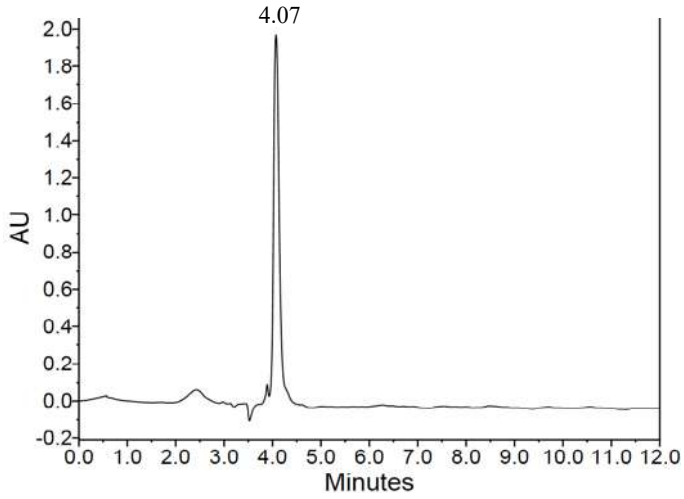
COSY in D_2O registered at 298 K of **glycopeptide 7**



HSQC in D_2O registered at 298 K of **glycopeptide 7**

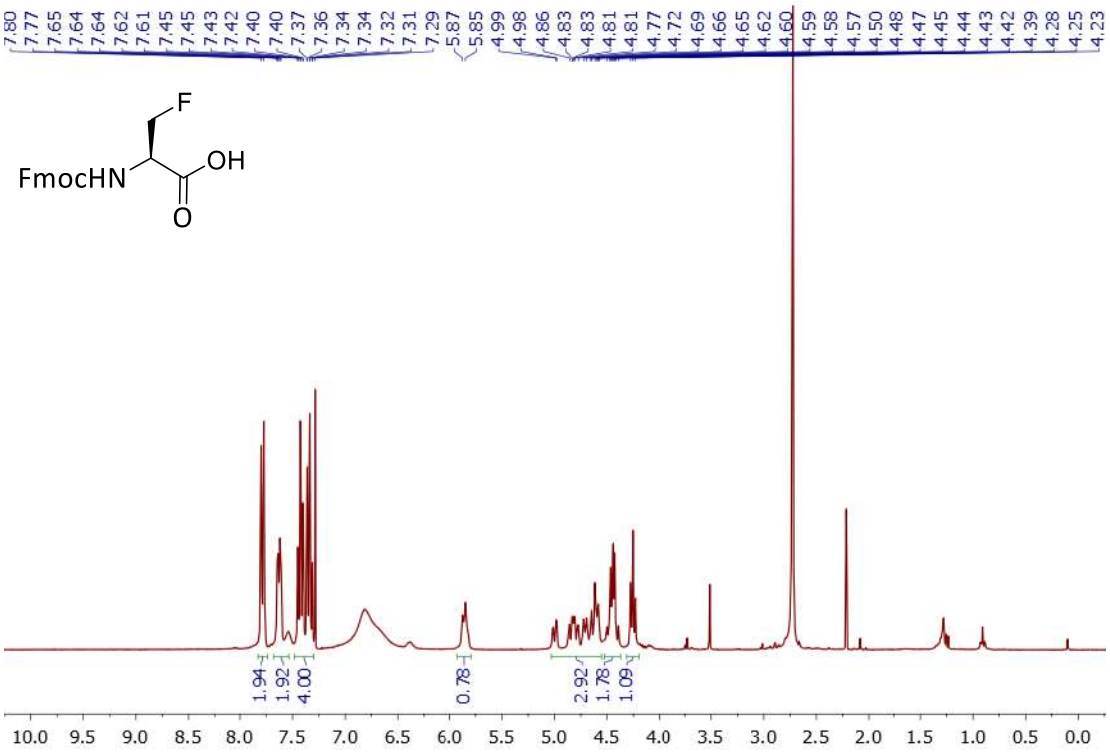


Analytical HPLC chromatogram of **glycopeptide 7**

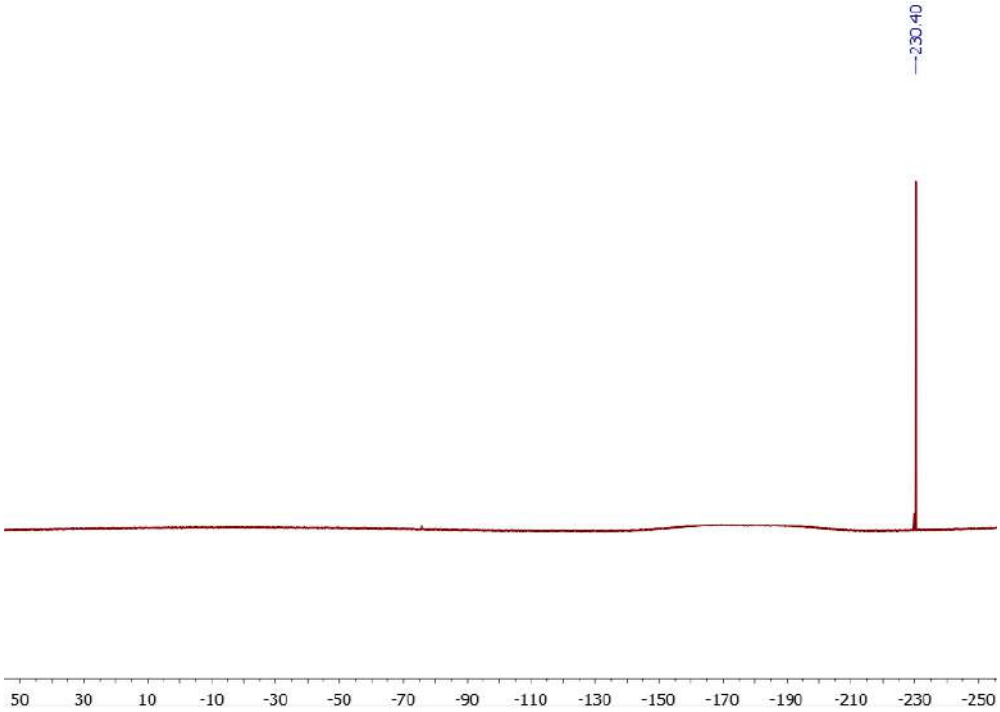


8.4. Supplementary material of Chapter 5

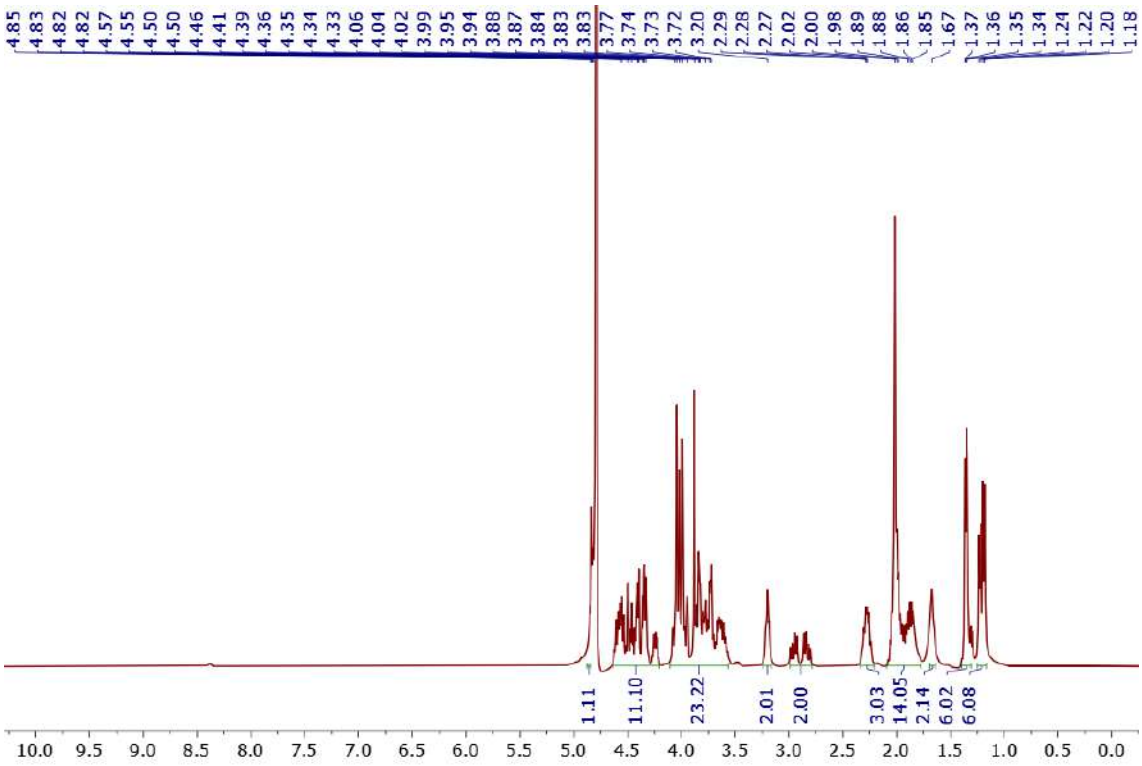
^{1}H NMR 400 MHz in $CDCl_3$ registered at 298 K of **compound 10**



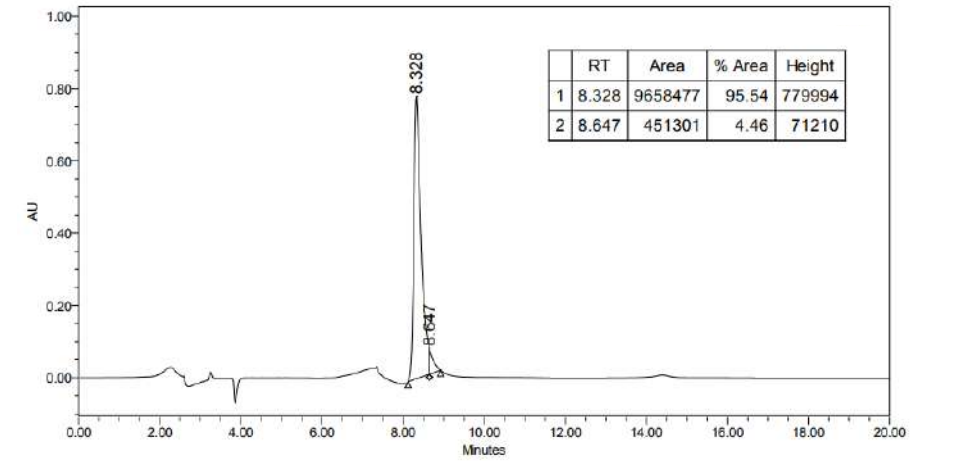
Decoupled ^{19}F NMR 282 MHz in $CDCl_3$ registered at 298 K of **compound 10**

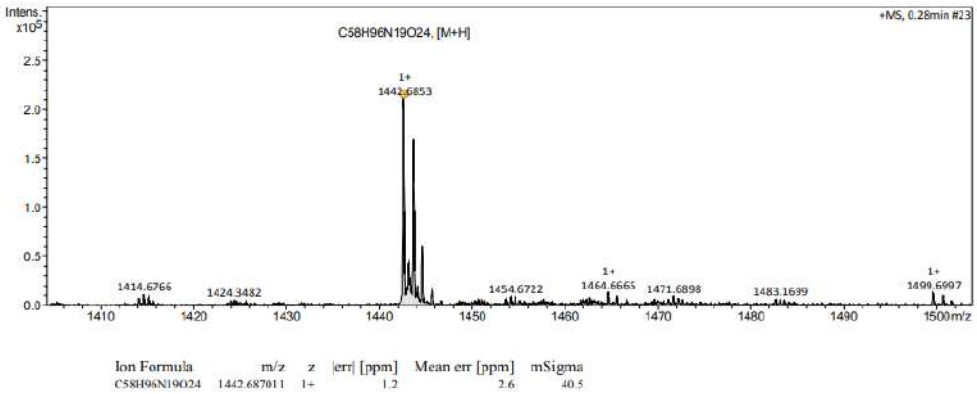


^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 13**



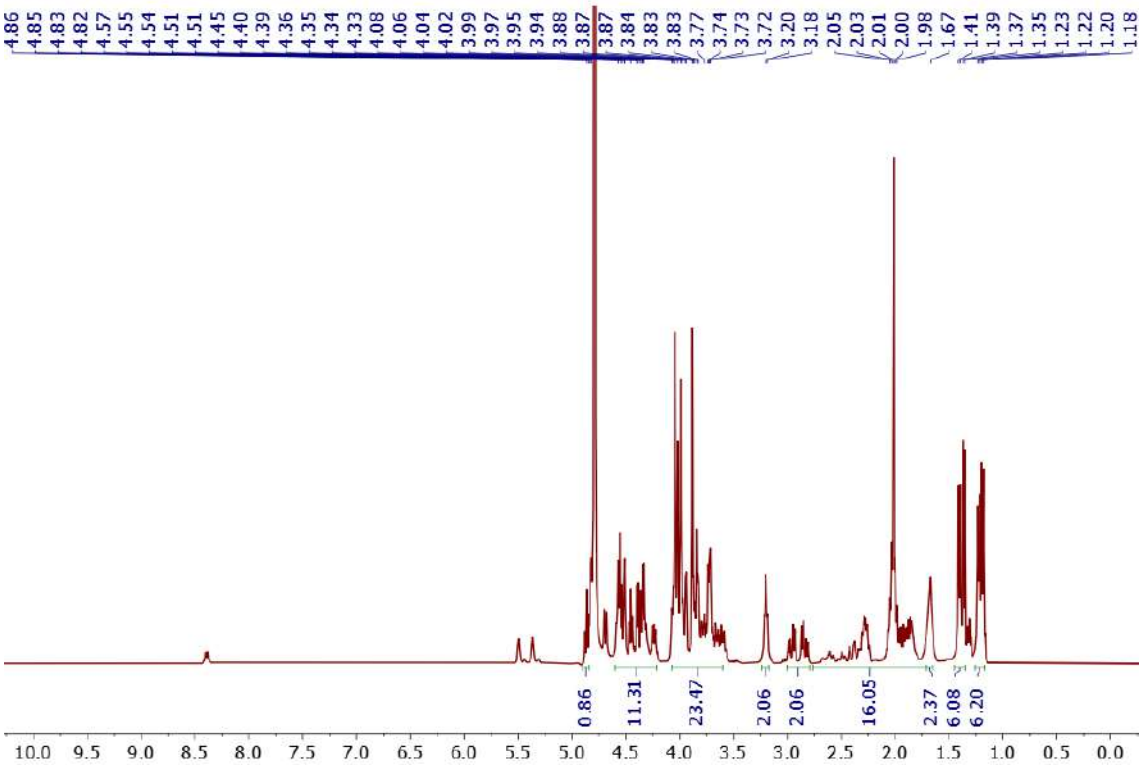
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 13**



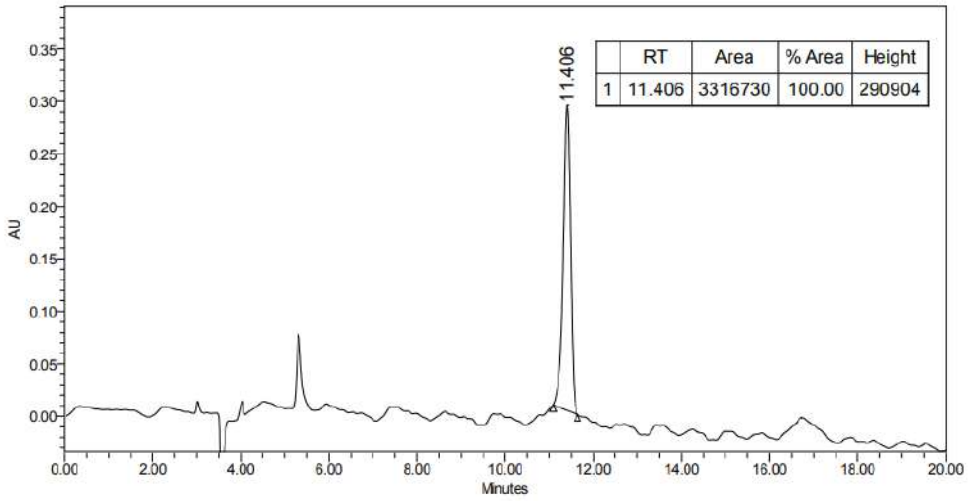


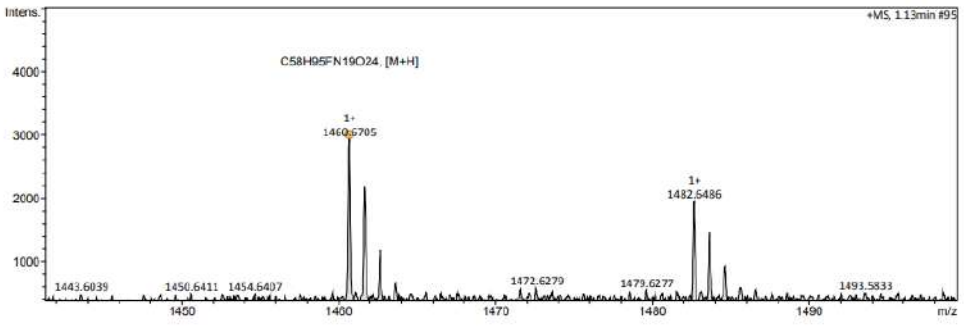
Ion Formula	m/z	z	err [ppm]	Mean err [ppm]	mSigma
C58H96N19O24	1442.687011	1+	1.2	2.6	40.5

^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 14**



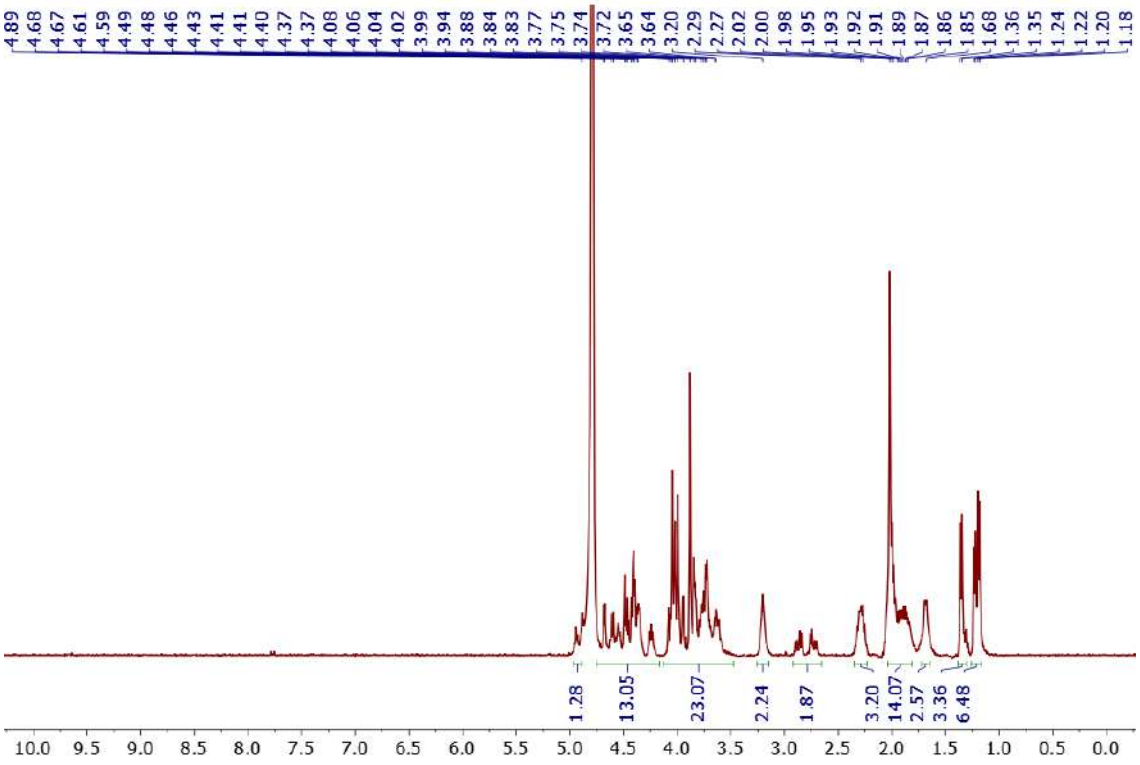
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 14**



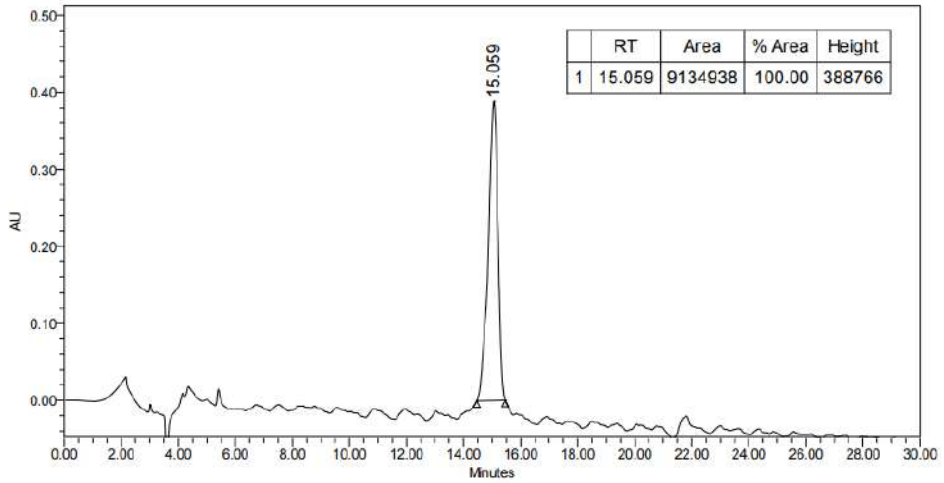


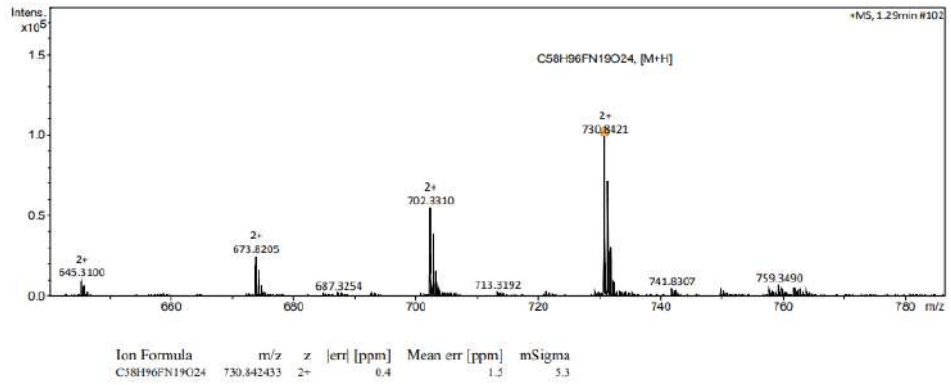
Ion Formula	m/z	z	err [ppm]	Mean err [ppm]	mSigma
C58H95FN19O24	1460.677590	1+	4.8	347.7	75.7

^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 15**



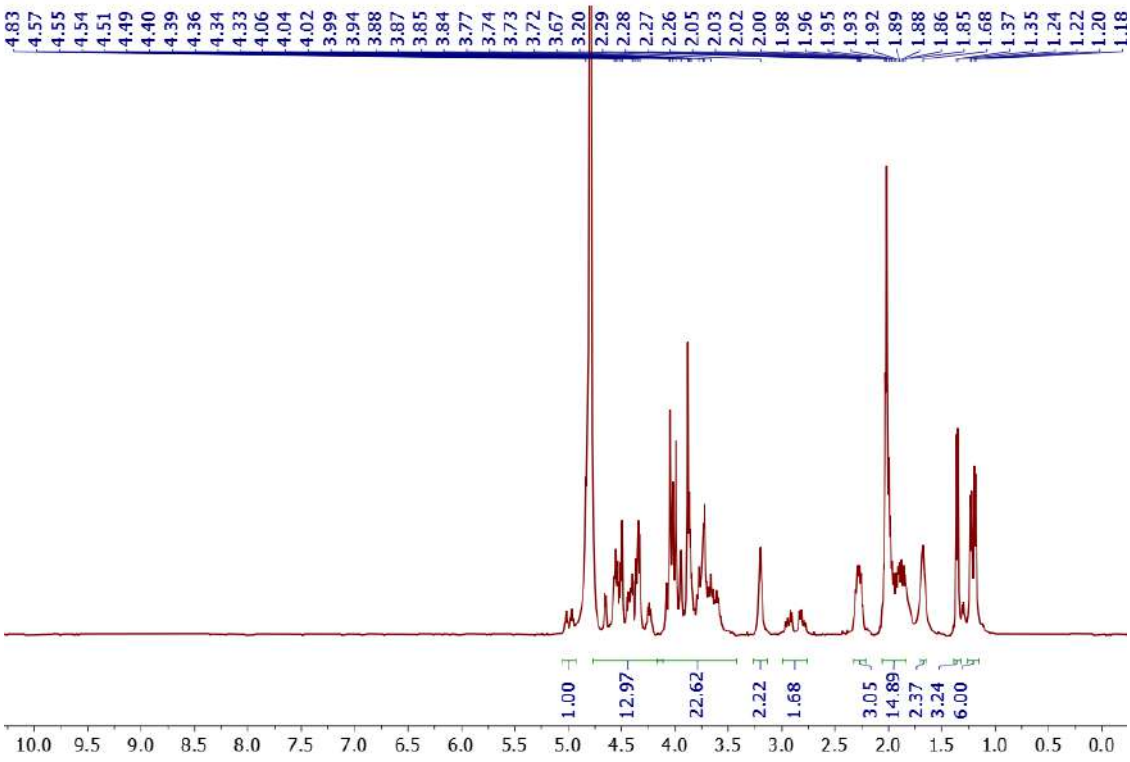
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 15**



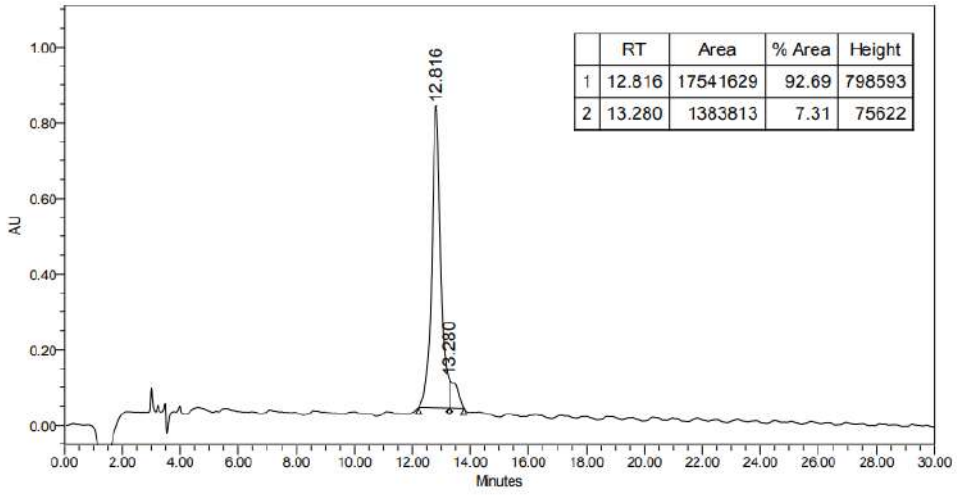


Ion Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C58H96FN19O24	730.842433	2+	0.4	1.5	5.3

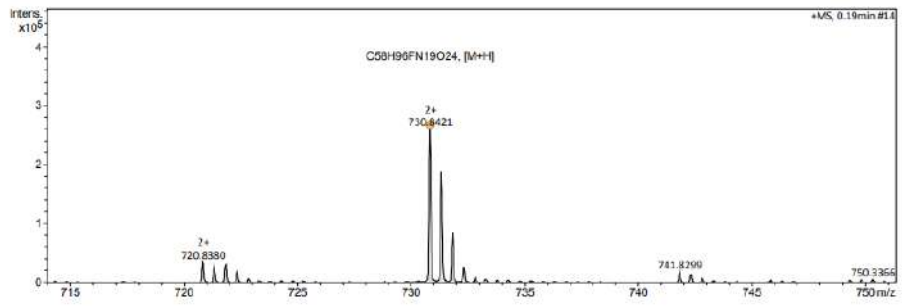
^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 16**



Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 16**

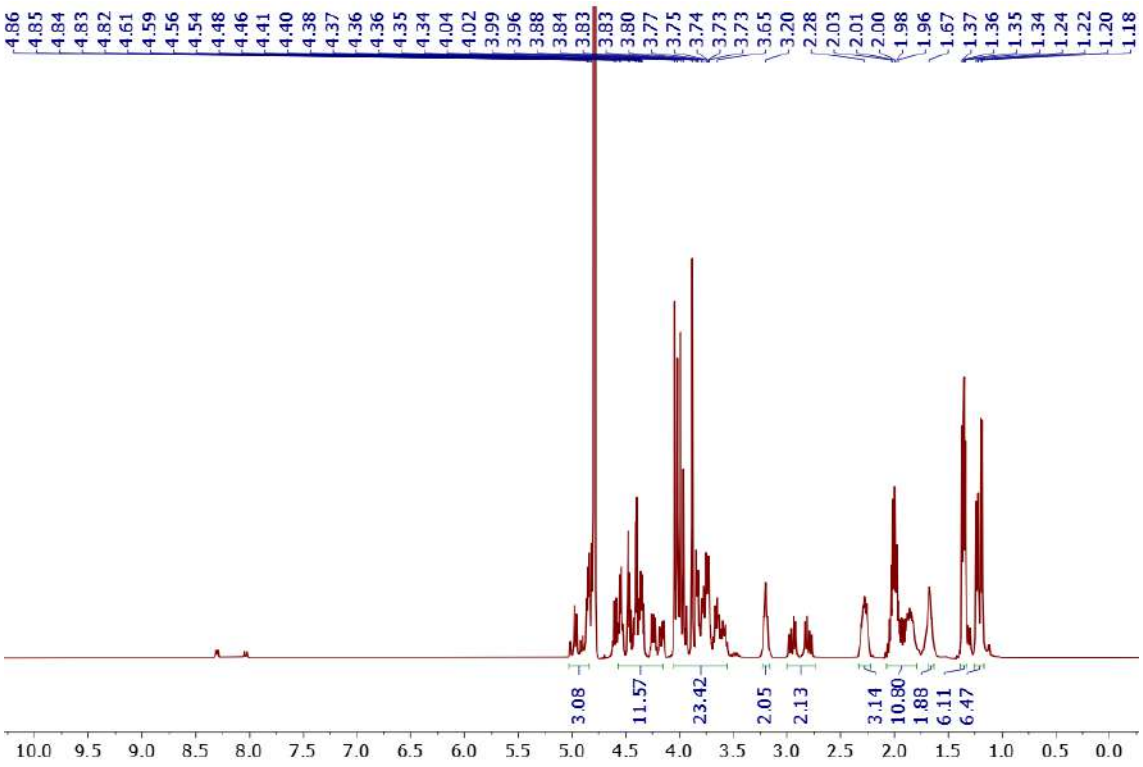


	RT	Area	% Area	Height
1	12.816	17541629	92.69	798593
2	13.280	1383813	7.31	75622

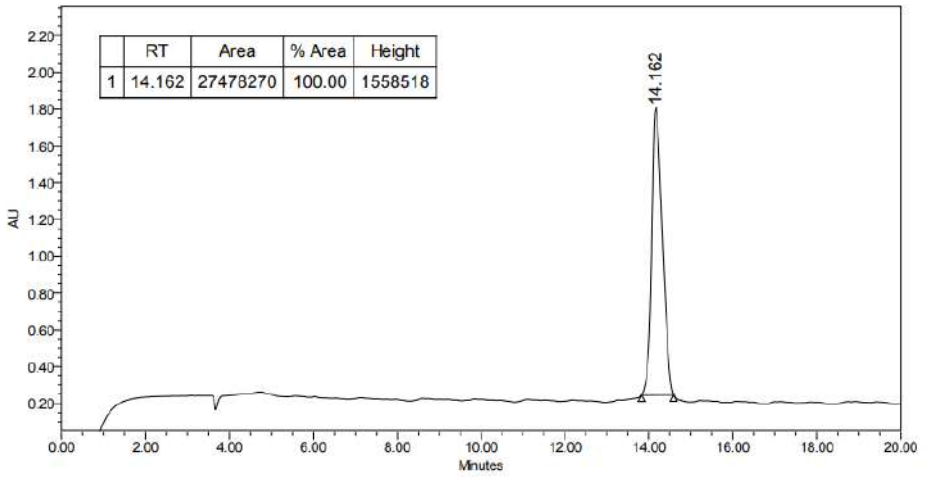


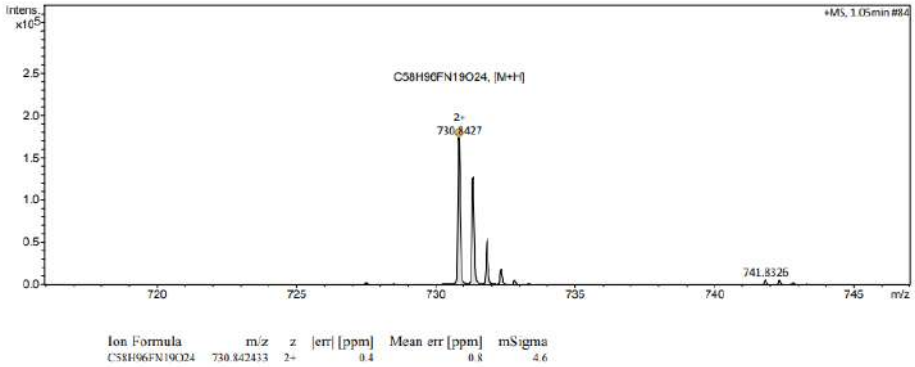
Ion Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C58H96FN19O24	730.842433	2+	0.4	2.9	10.5

^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 17**



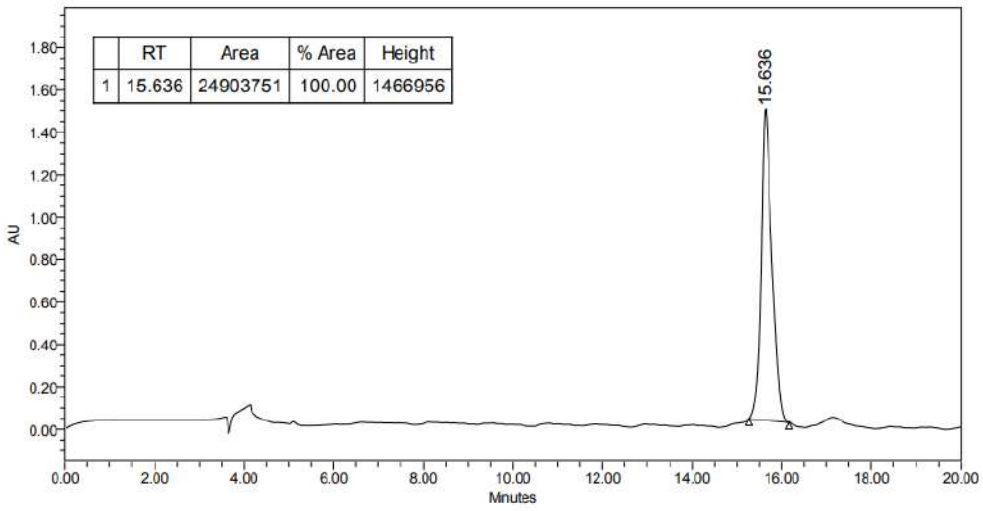
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 17**



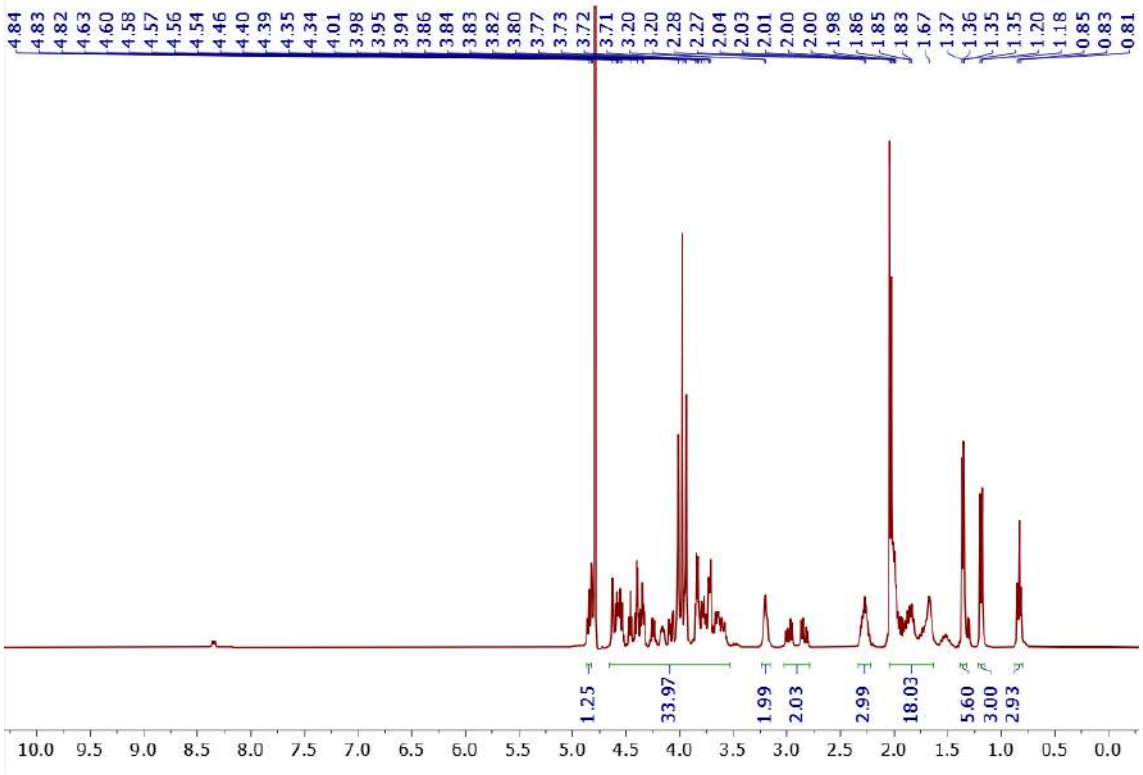


Ion Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C58H96FN19O24	730.842433	2+	0.4	0.8	4.6

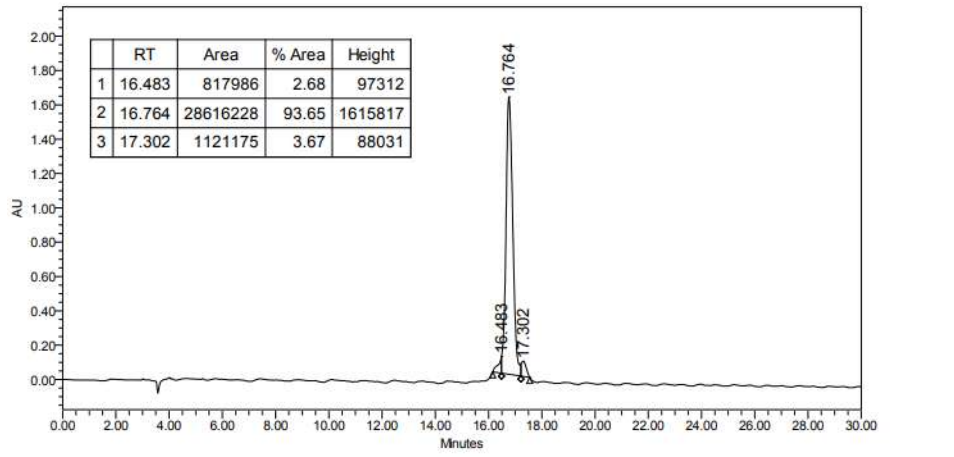
Analytical HPLC chromatogram of **glycopeptide 18**

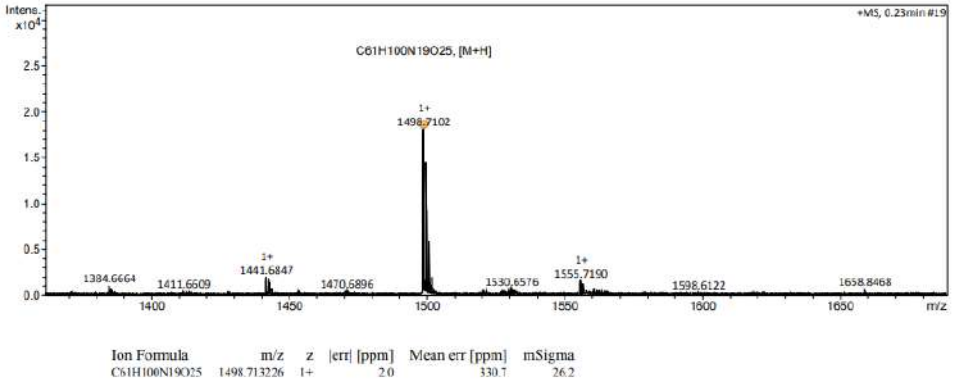


^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 19**



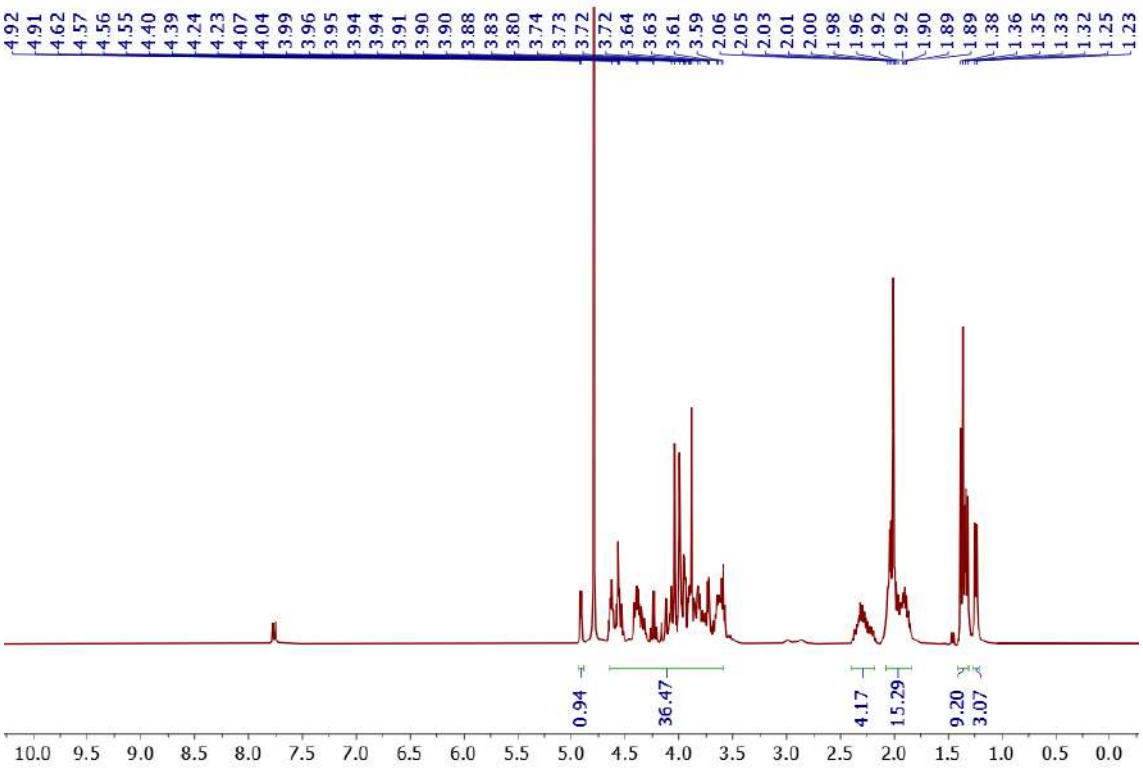
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 19**



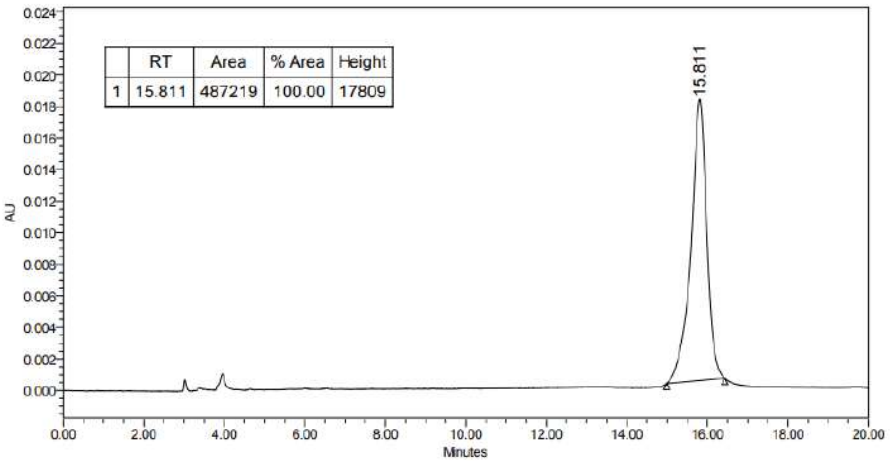


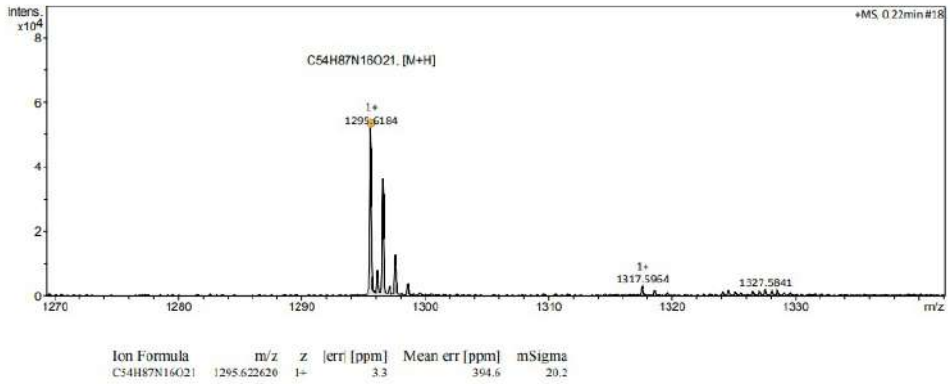
Ion Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C61H100N19O25	1498.713226	1+	2.0	330.7	26.2

^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 20**



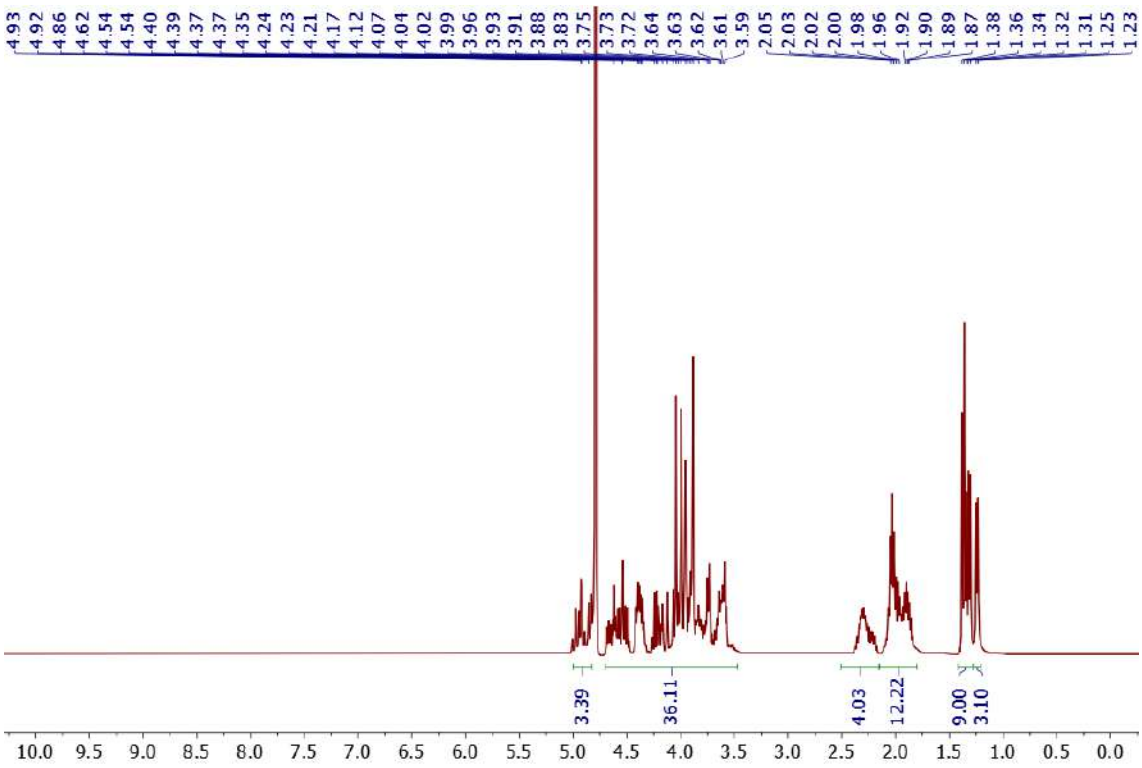
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 20**



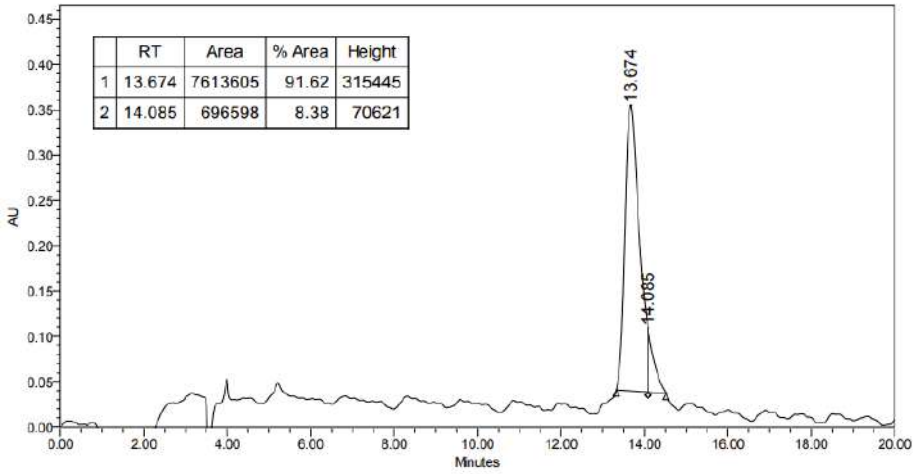


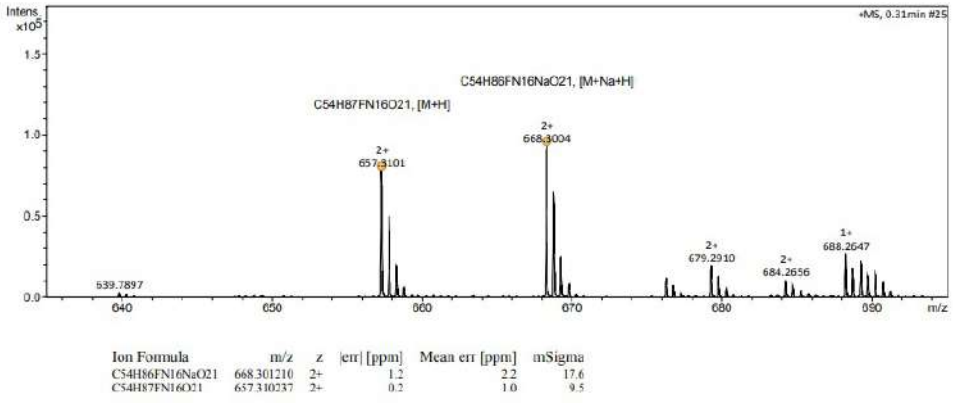
Ion Formula	m/z	z	err [ppm]	Mean err [ppm]	mSigma
C54H87N16O21	1295.622620	1+	3.3	394.6	20.2

^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 21**



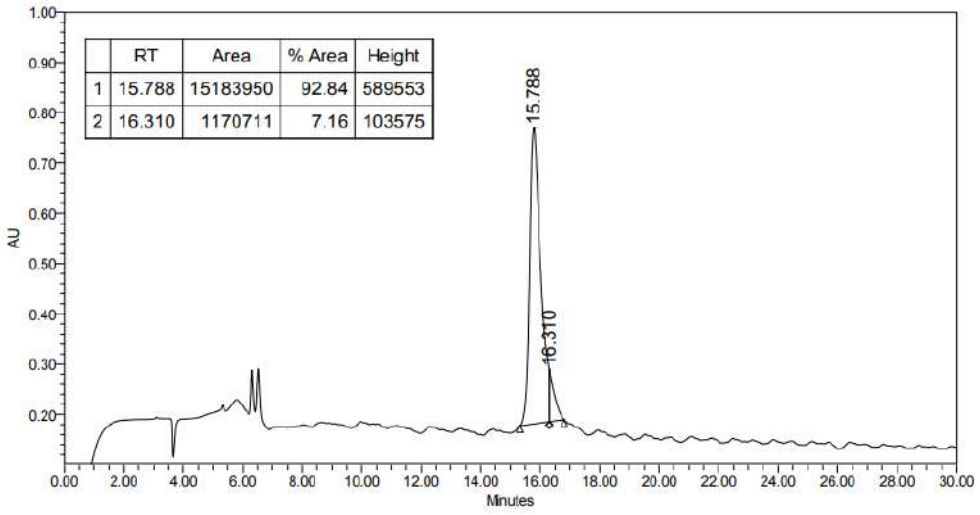
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 21**



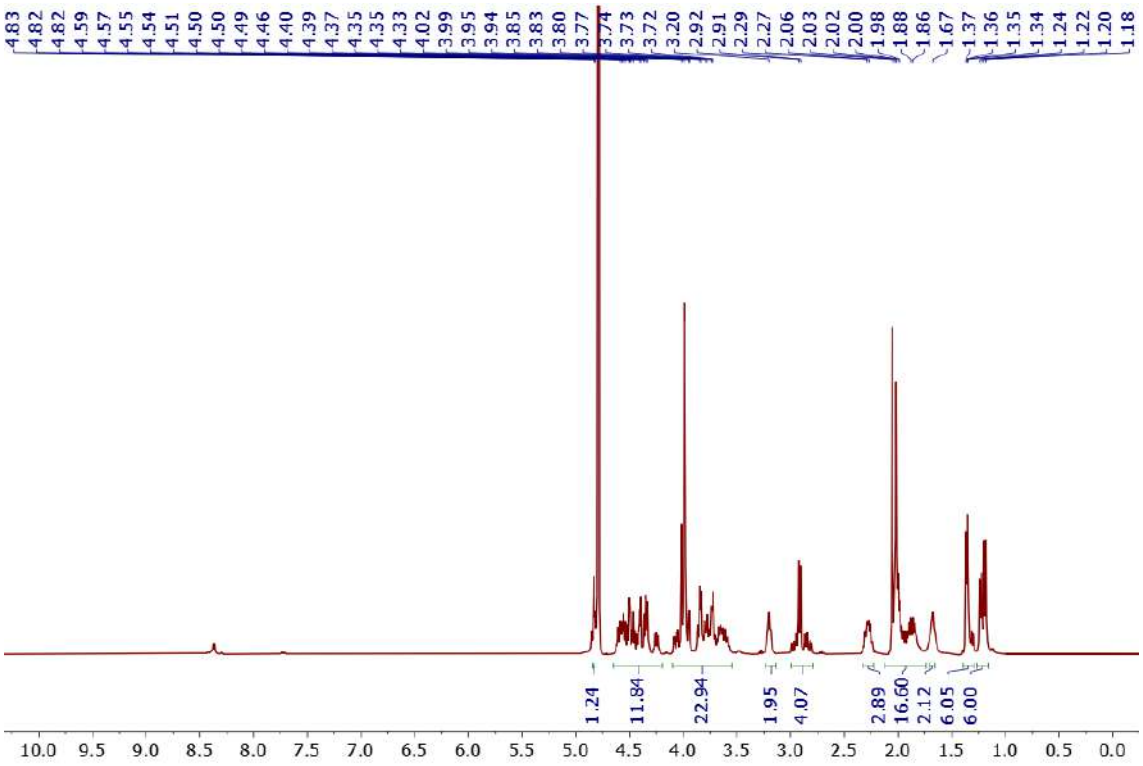


Ion Formula	m/z	z	err [ppm]	Mean err [ppm]	mSigma
C54H86FN16NaO21	668.301210	2+	1.2	2.2	17.6
C54H87FN16O21	657.310237	2+	0.2	1.0	9.5

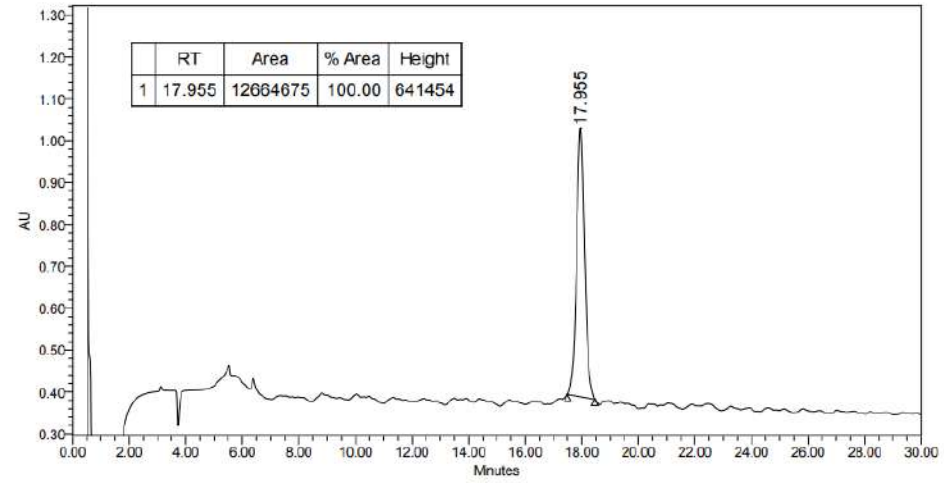
Analytical HPLC chromatogram of **glycopeptide 22**

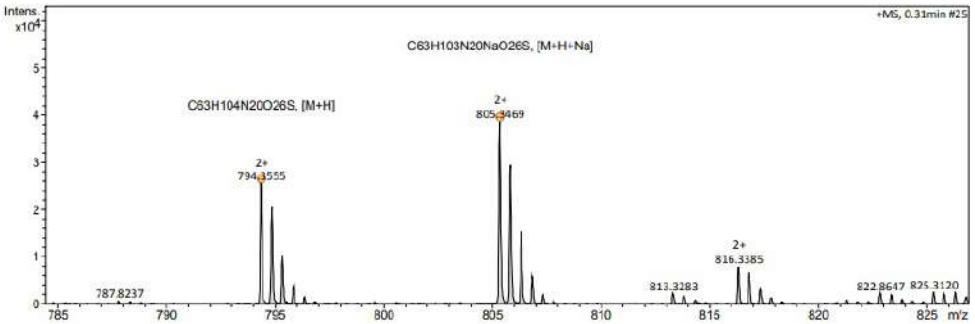


^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 23**



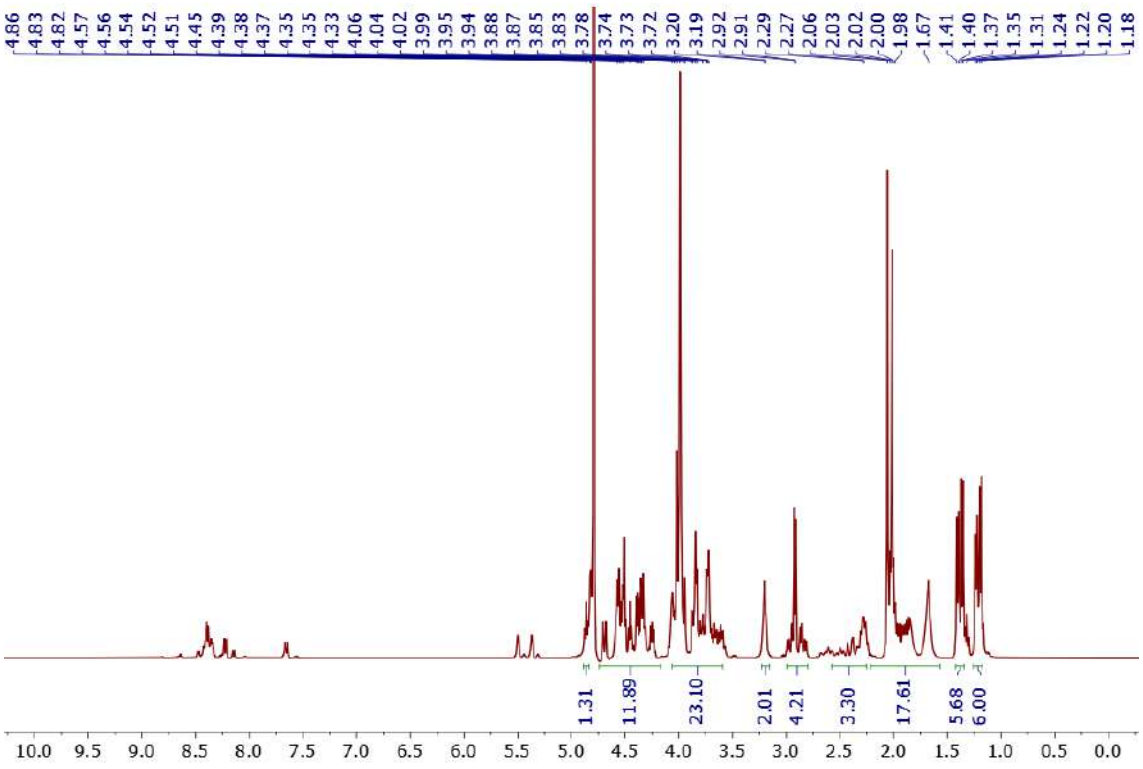
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 23**



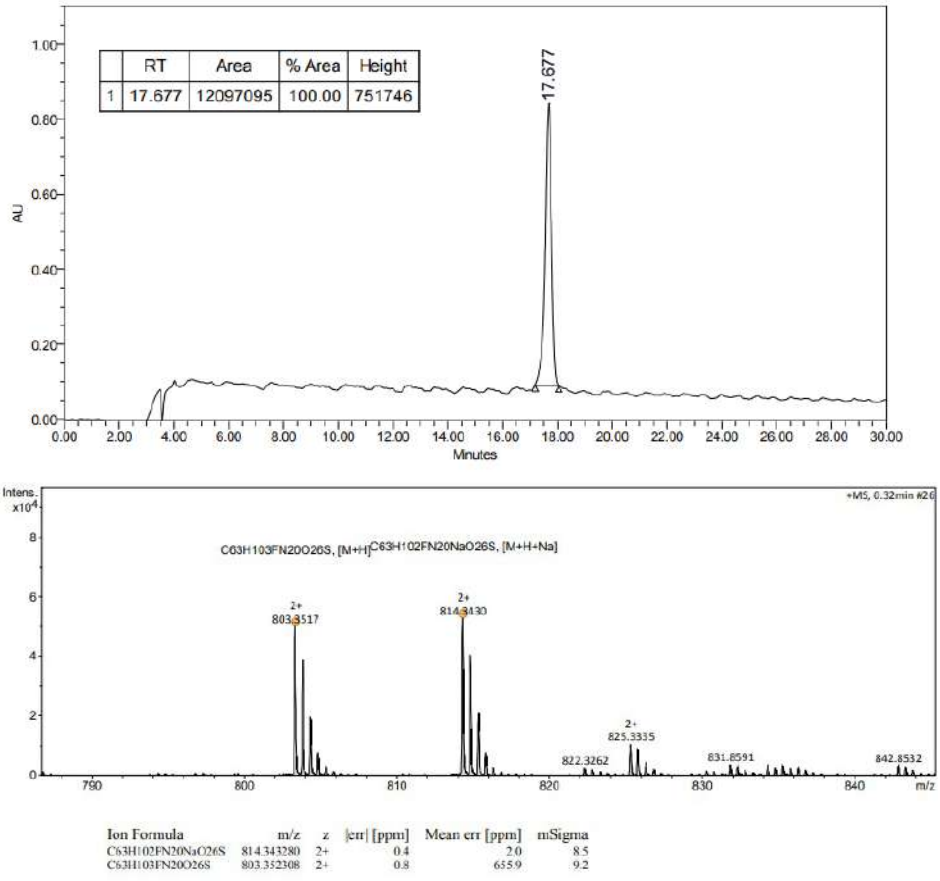


Ion Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C63H103N20NaO26S	805.347991	2+	1.4	3.1	8.2
C63H104N20O26S	794.357019	2+	2.0	664.6	14.4

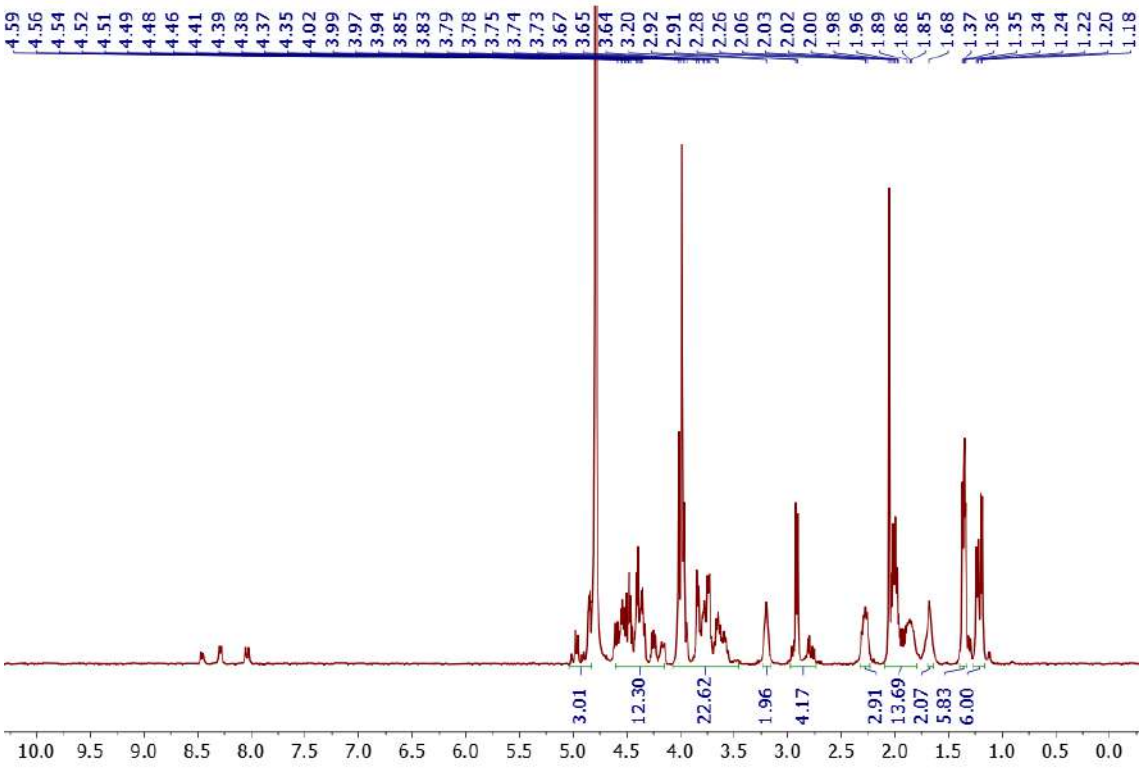
^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 24**



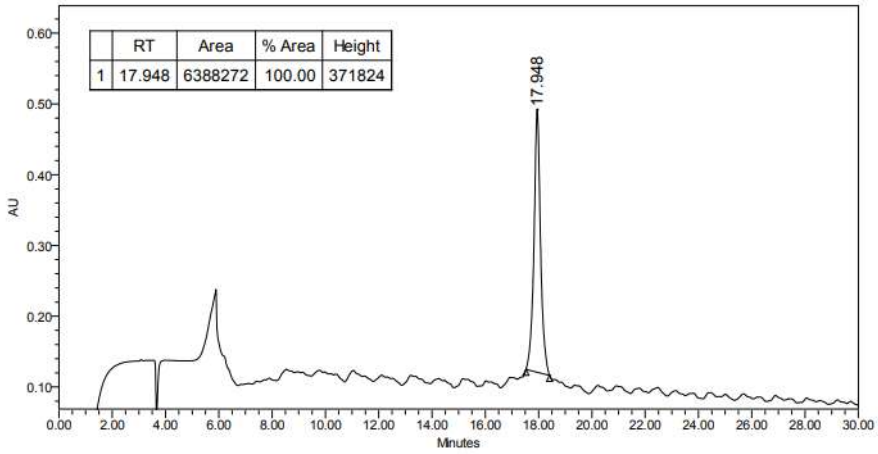
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 24**

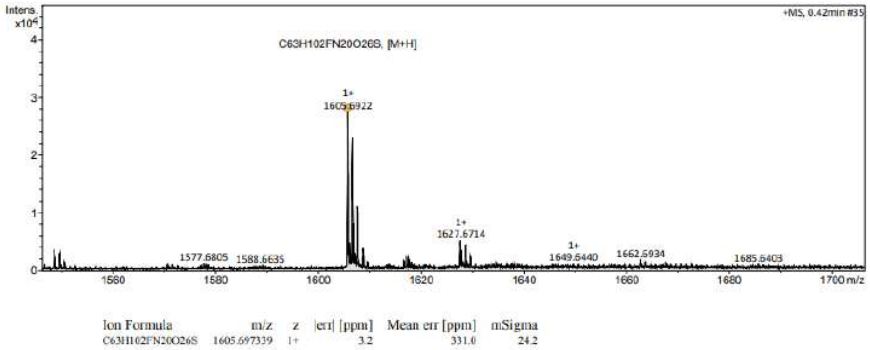


^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 25**



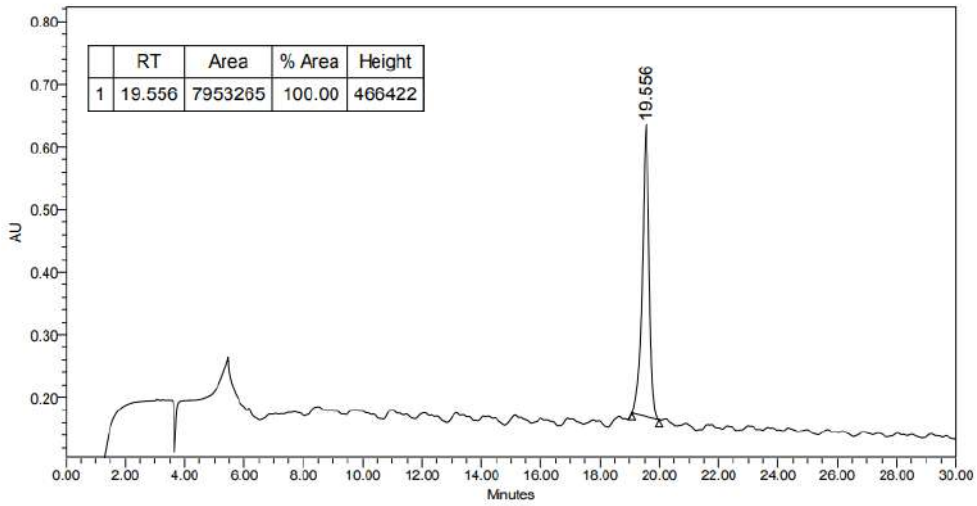
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 25**



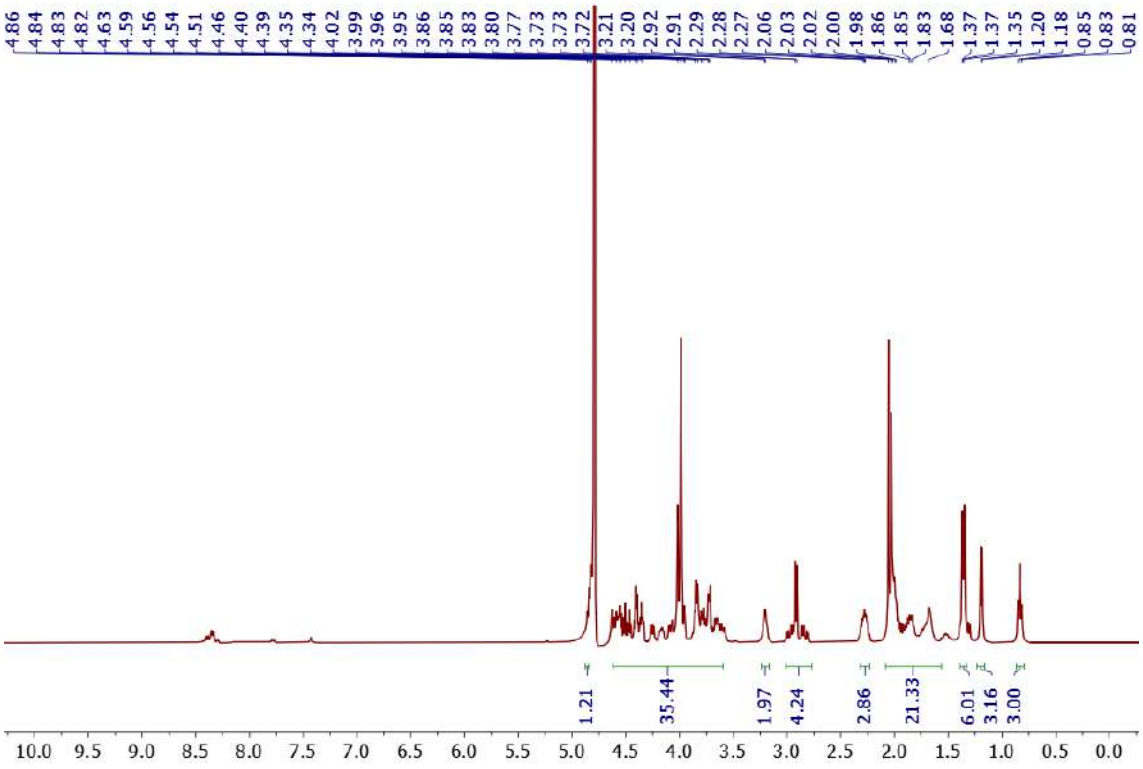


Ion Formula	m/z	z	err [ppm]	Mean err [ppm]	mSigma
C63H102FN20O26S	1605.697339	1+	3.2	331.0	24.2

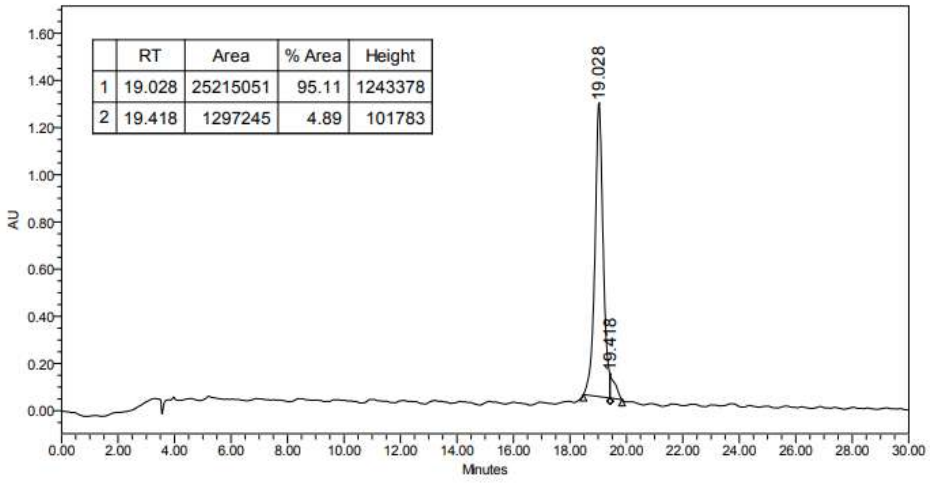
Analytical HPLC chromatogram of **glycopeptide 26**

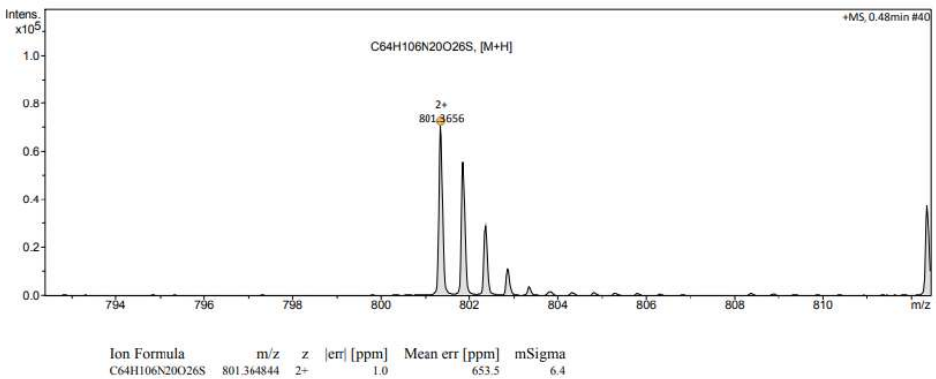


^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 27**



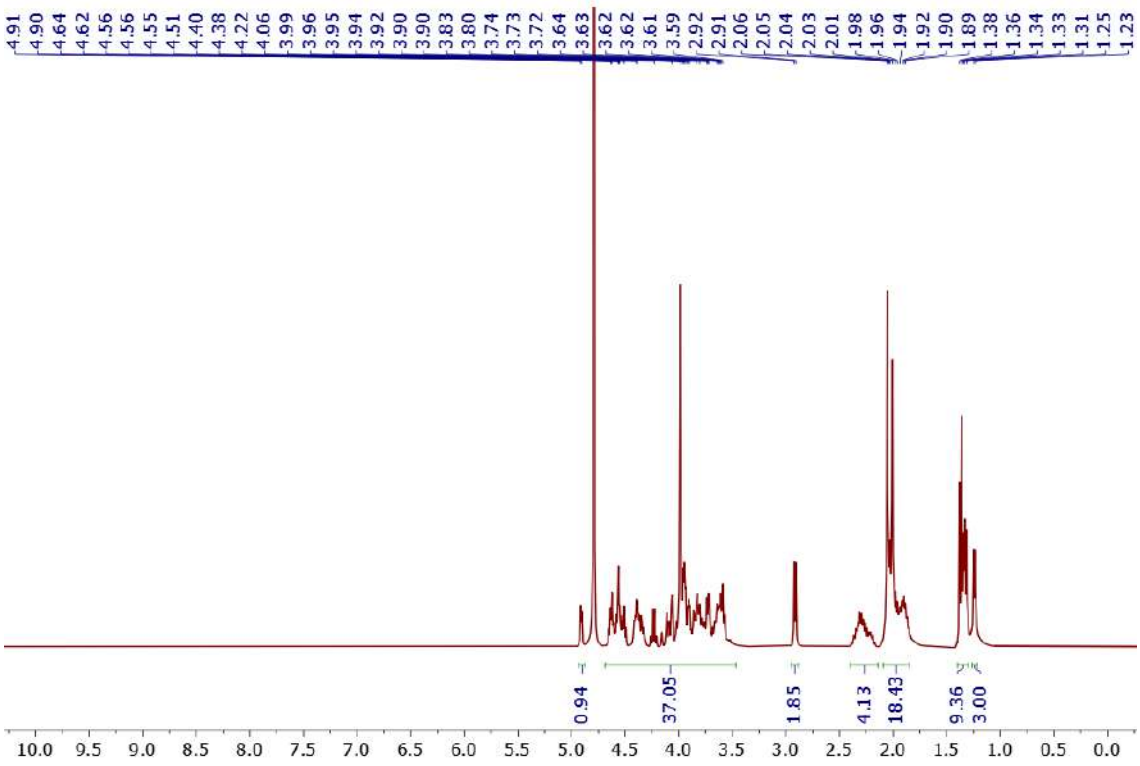
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 27**



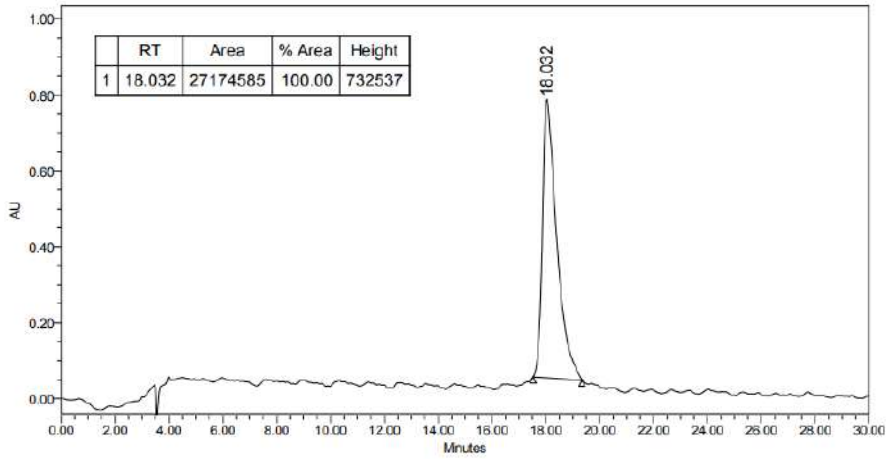


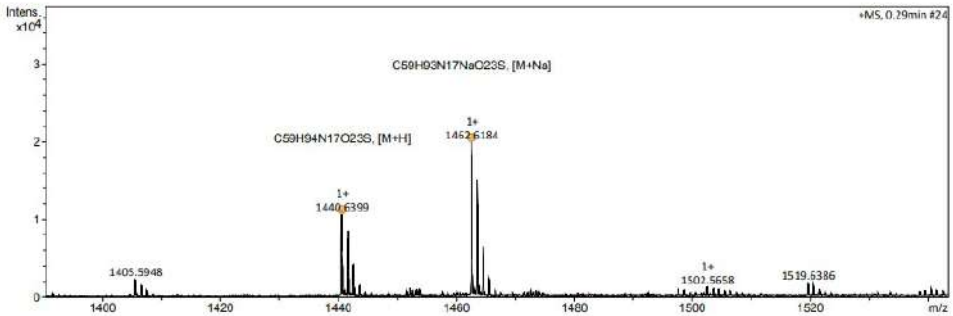
Ion Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
$C_{64}H_{106}N_{20}O_{26}S$	801.364844	2+	1.0	653.5	6.4

^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 28**



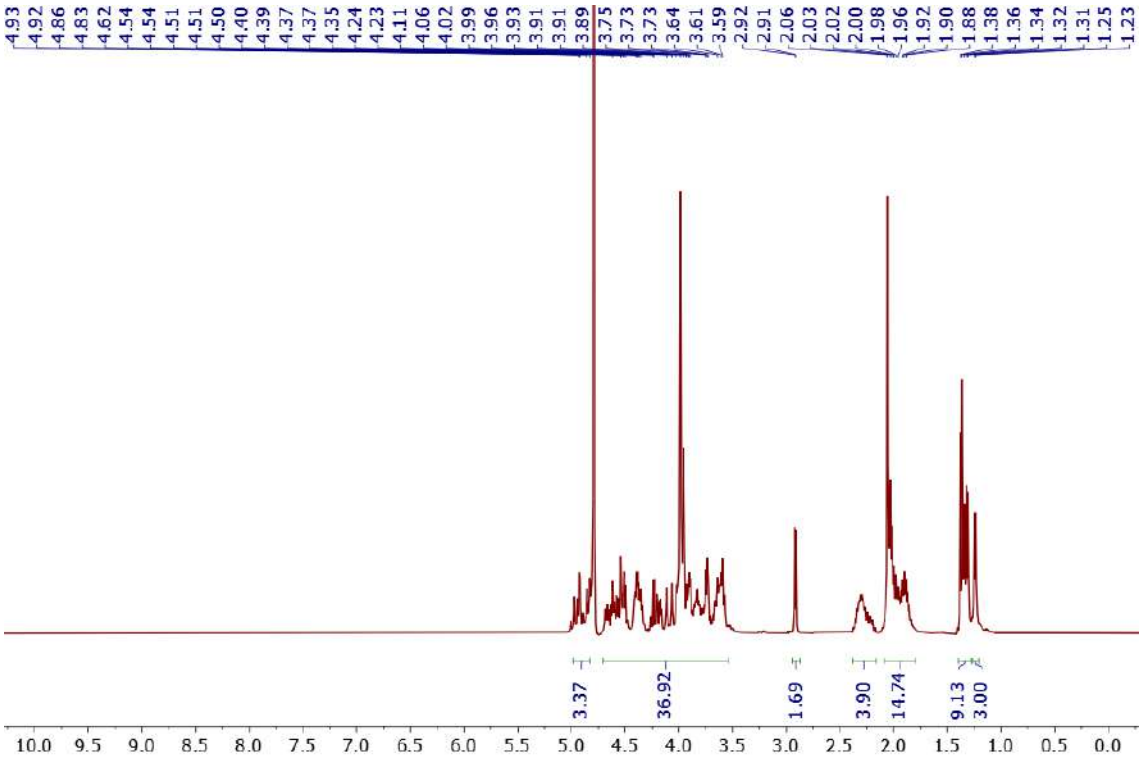
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 28**



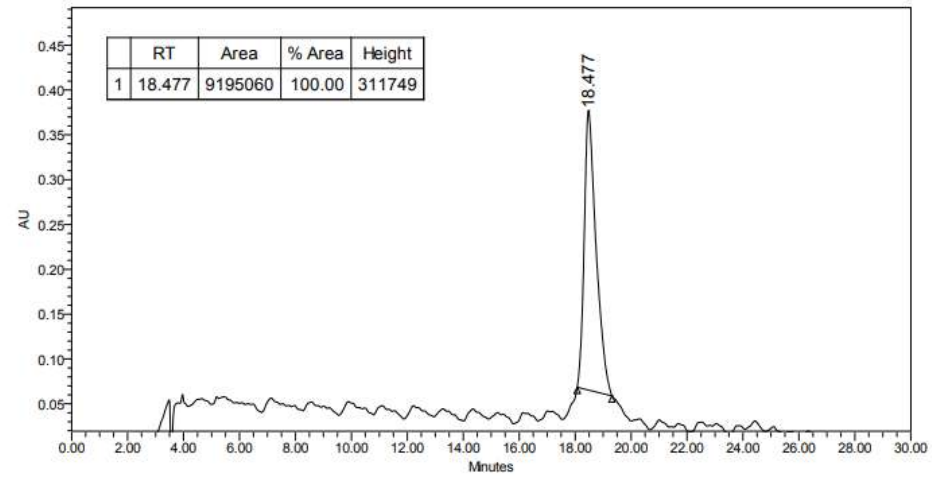


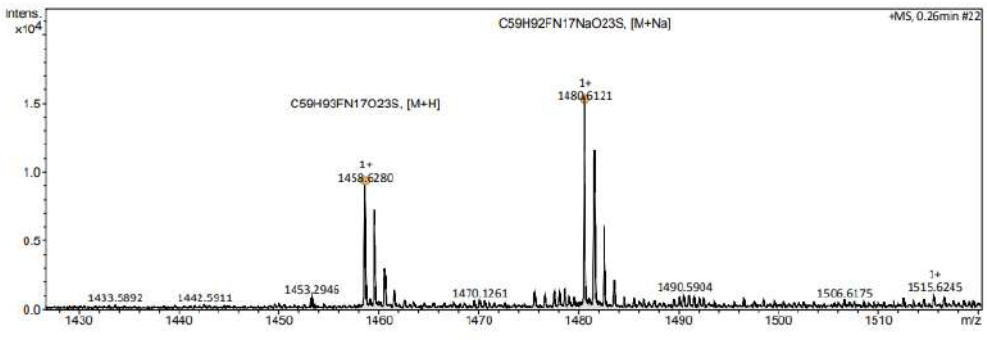
Ion Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C59H93N17NaO23S	1462.624314	1+	4.0	4.9	17.7
C59H94N17O23S	1440.642370	1+	1.7	682.7	41.7

^{1}H NMR 400 MHz in D_2O registered at 298 K of **glycopeptide 29**



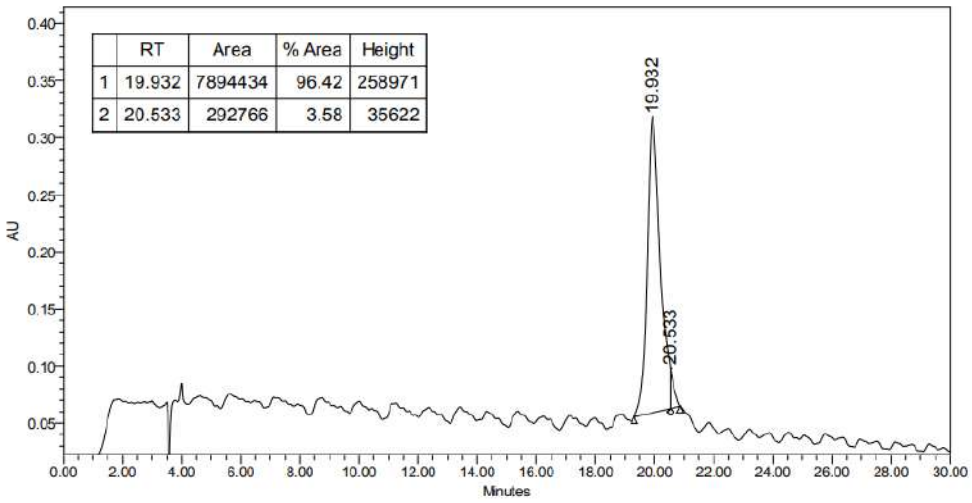
Analytical HPLC chromatogram and HRMS-ESI spectrum of **glycopeptide 29**





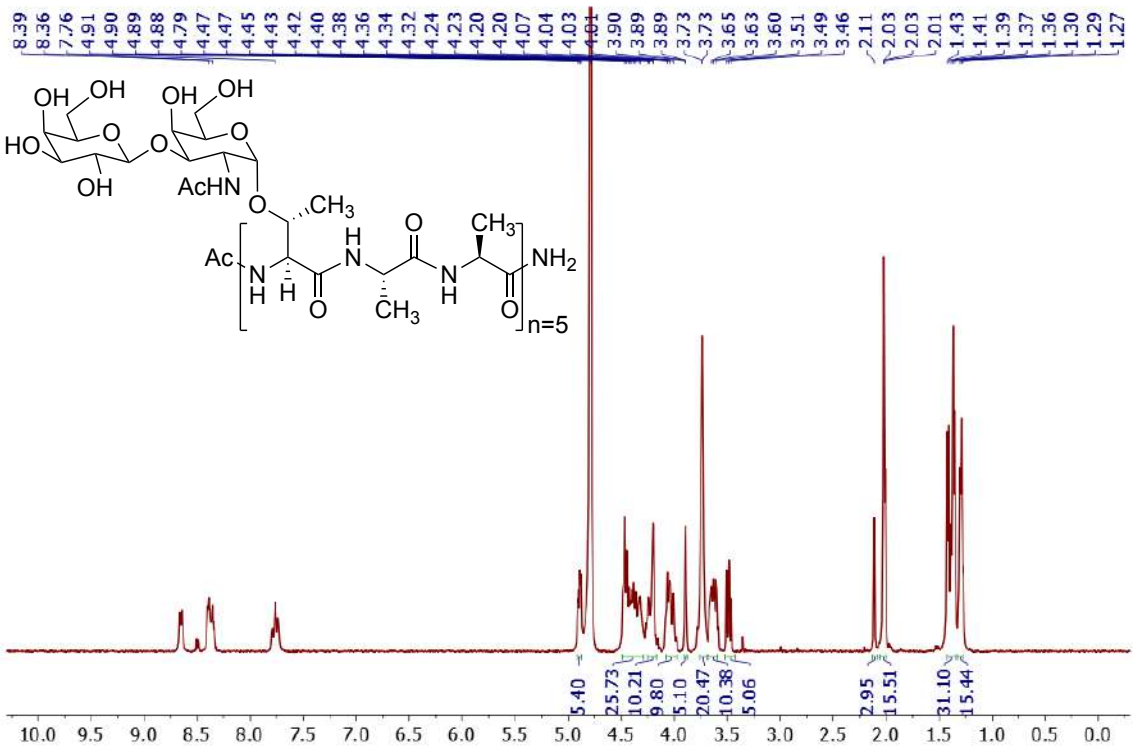
Ion Formula	m/z	z	|err| [ppm]	Mean err [ppm]	mSigma
C59H92FN17NaO23S	1480.614892	1+	1.9	365.3	31.6
C59H93FN17O23S	1458.632948	1+	3.4	675.2	39.8

Analytical HPLC chromatogram of **glycopeptide 30**

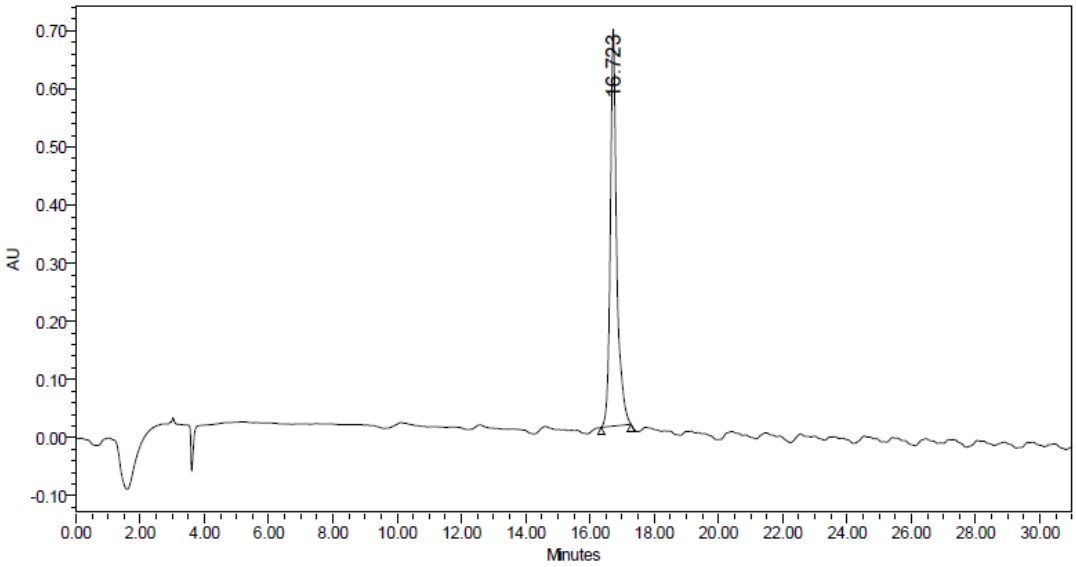


8.5. Supplementary material of Chapter 6

^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 33**

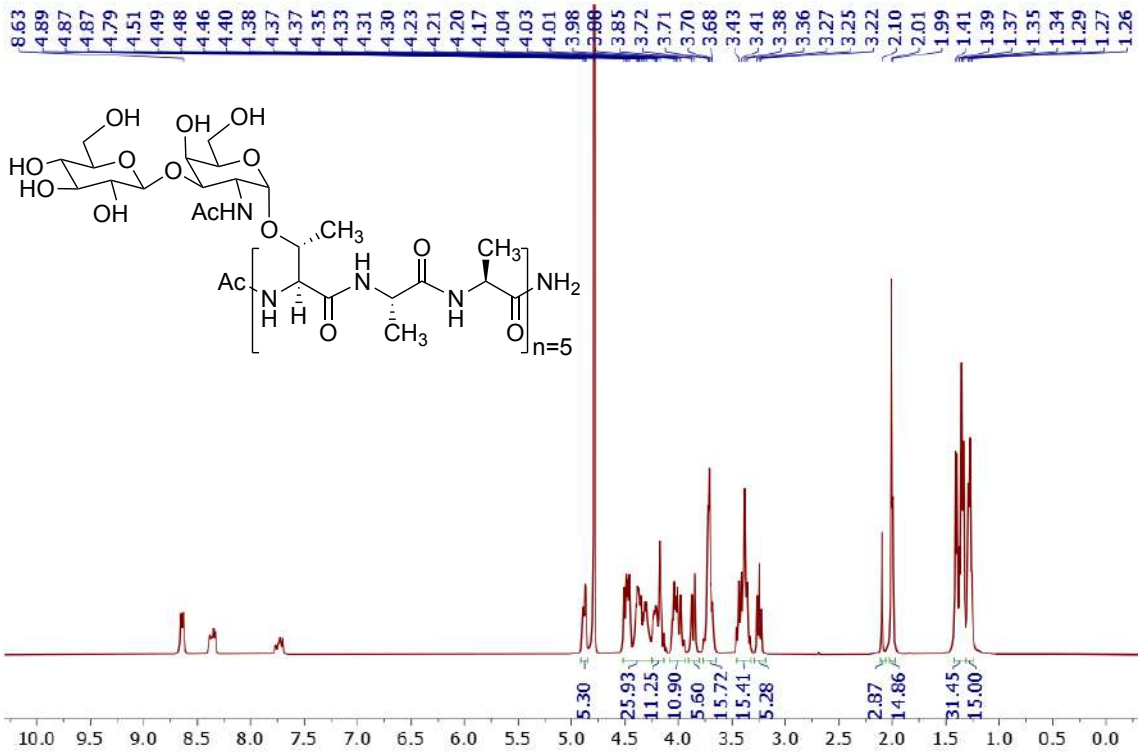


Analytical HPLC chromatogram of **glycopeptide 33**

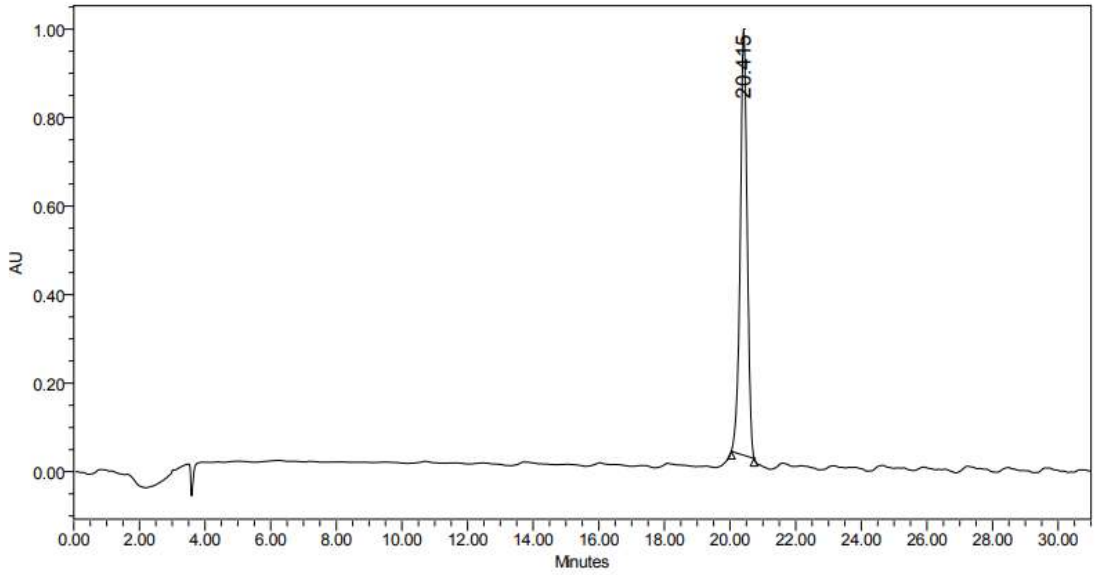


	RT	Area	% Area	Height
1	16.723	9456261	100.00	682897

^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 34**

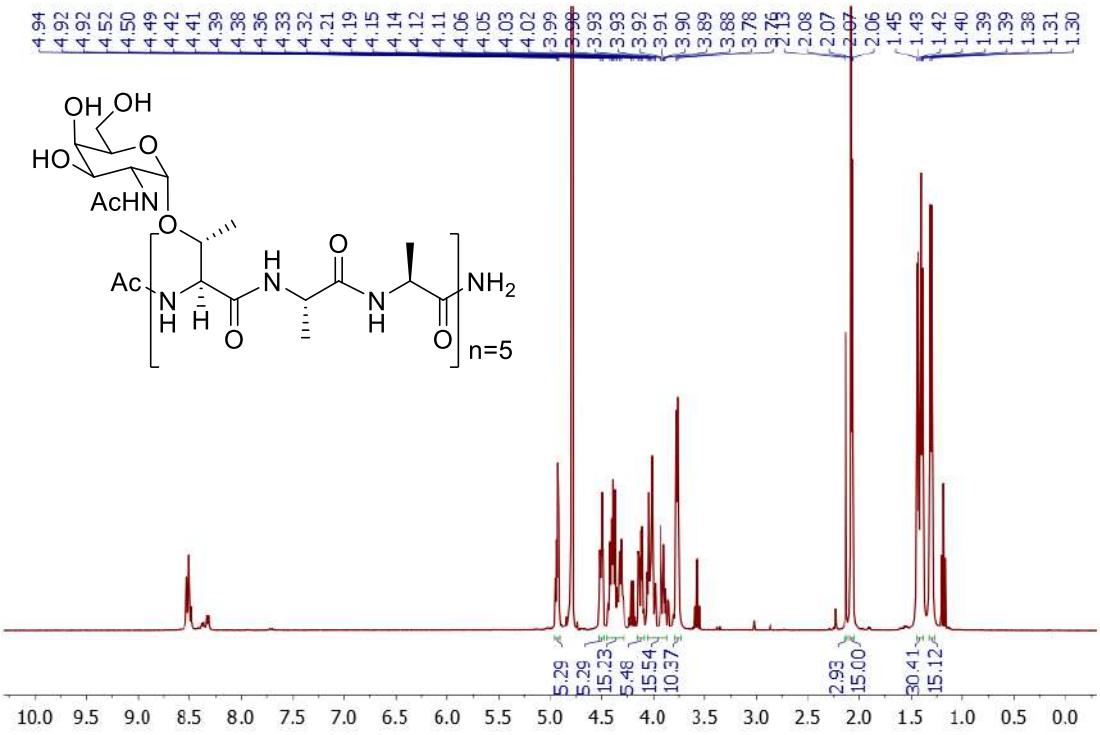


Analytical HPLC chromatogram of **glycopeptide 34**

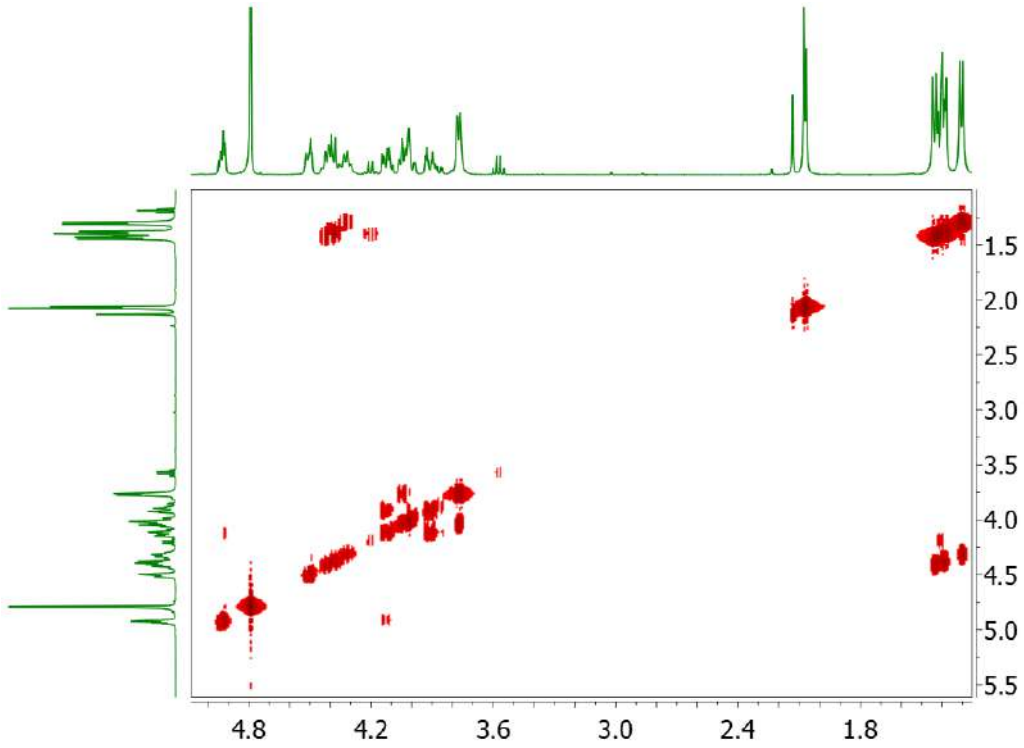


	RT	Area	% Area	Height
1	20.415	14105515	100.00	963125

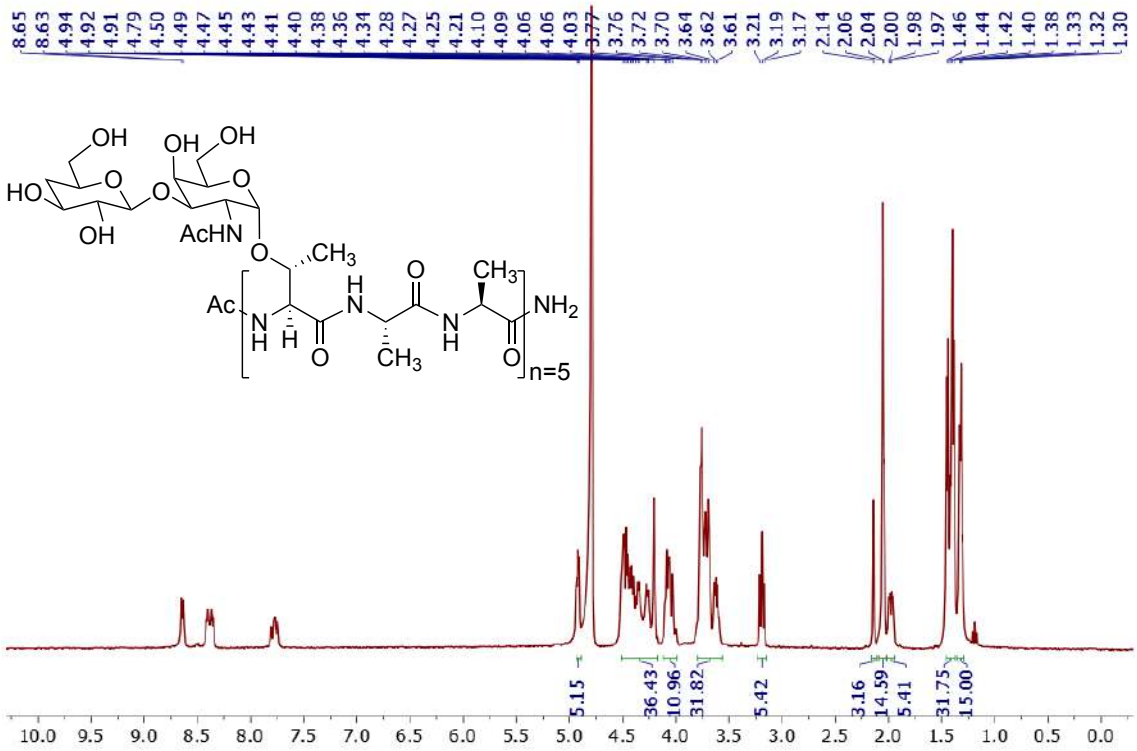
^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 35**



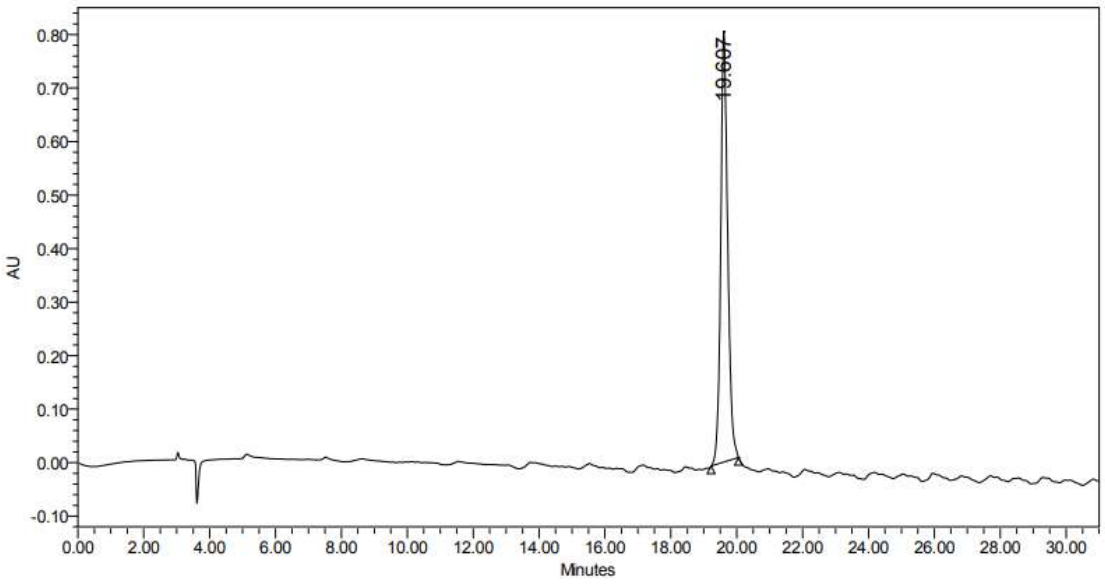
COSY in D_2O registered at 298 K of **compound 35**



^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 36**

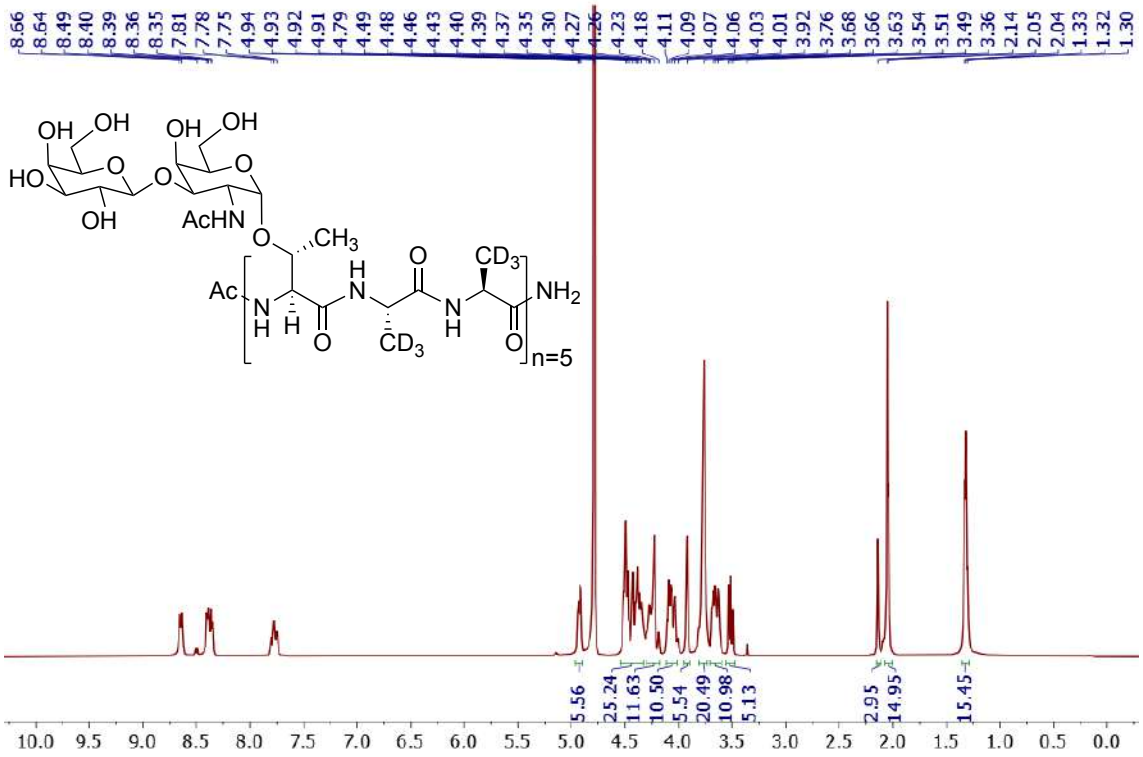


Analytical HPLC chromatogram of **glycopeptide 36**

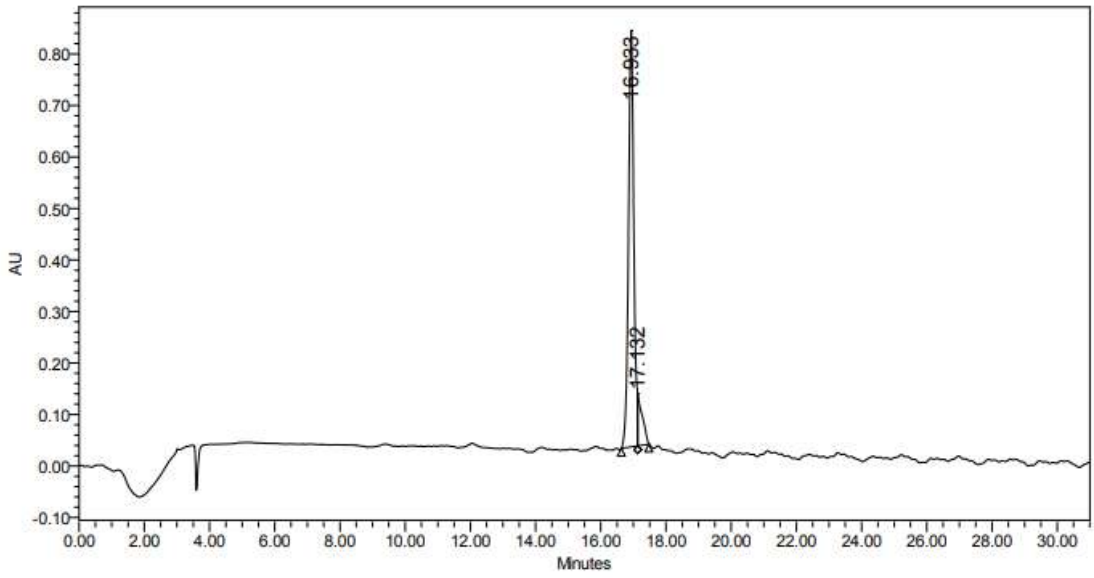


	RT	Area	% Area	Height
1	19.607	11855379	100.00	805249

^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 37**

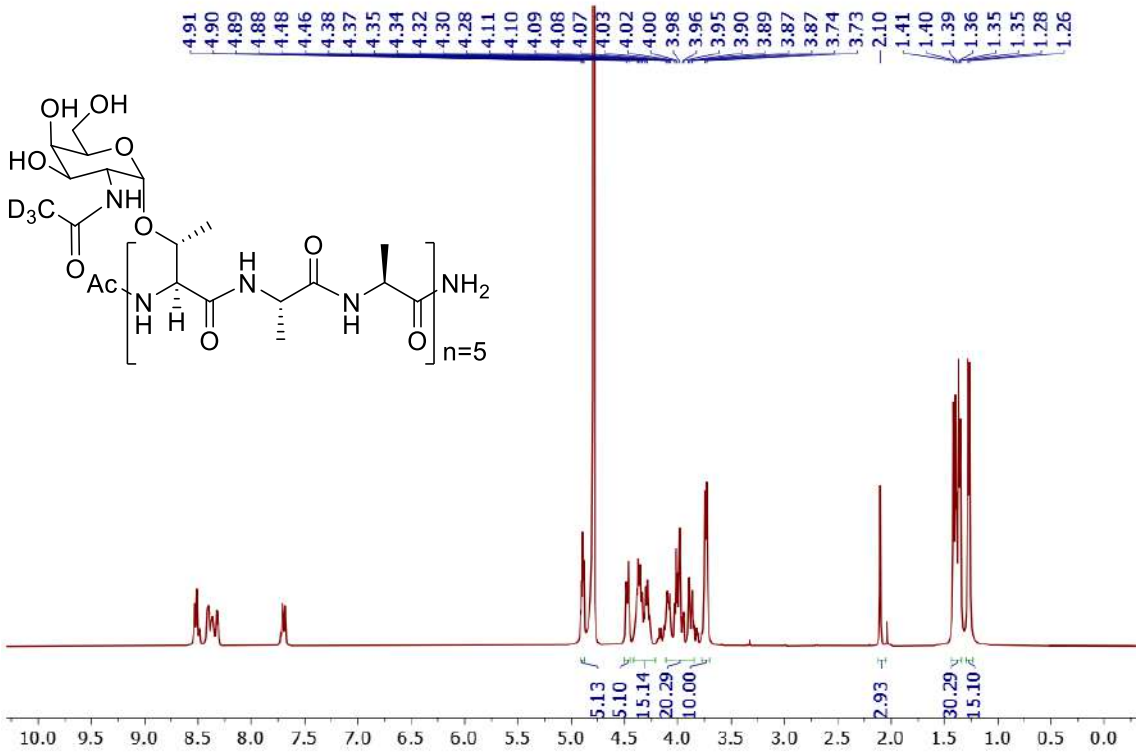


Analytical HPLC chromatogram of **glycopeptide 37**

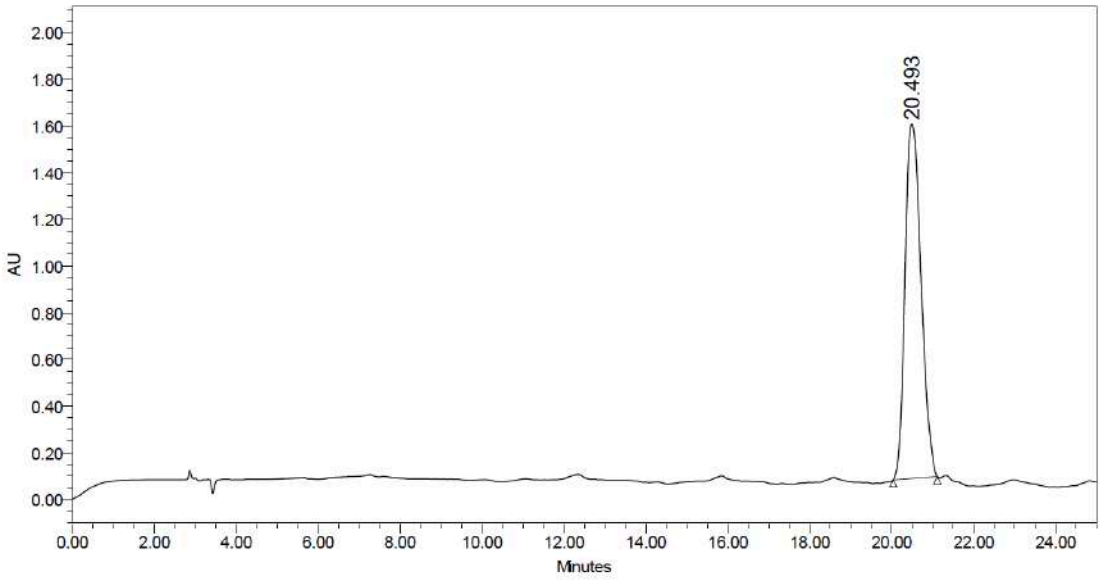


	RT	Area	% Area	Height
1	16.933	9197228	90.37	808744
2	17.132	980417	9.63	101015

^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 38**

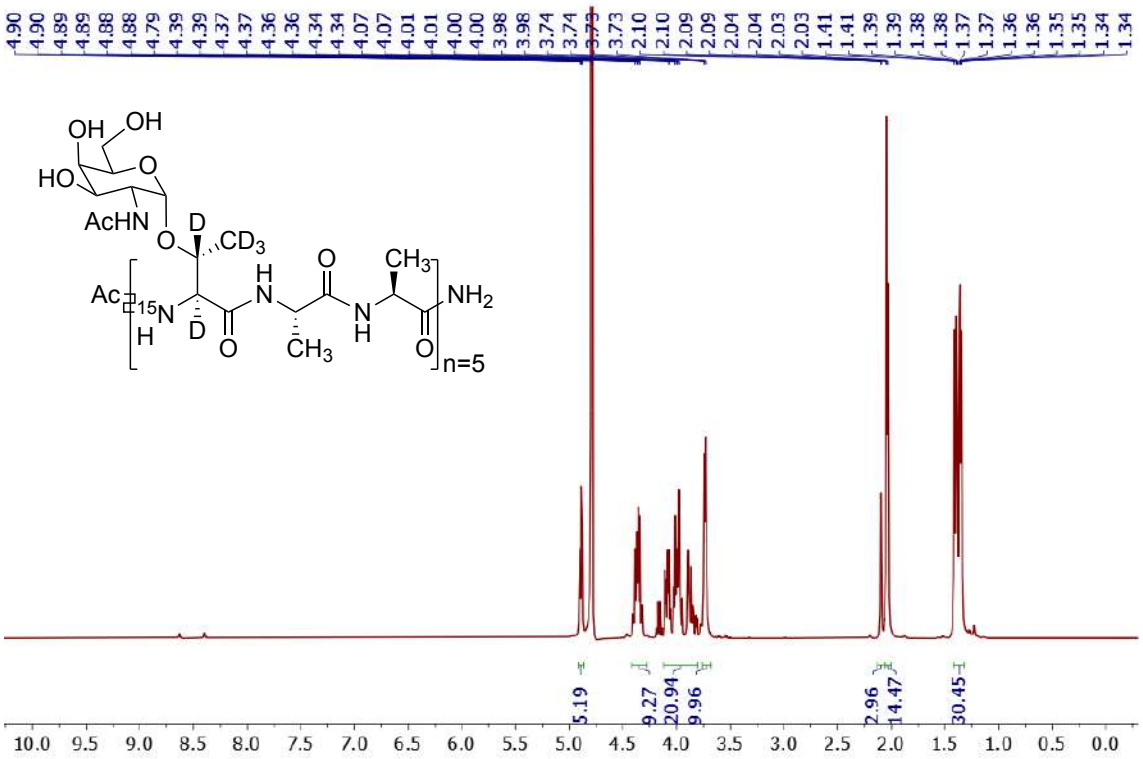


Analytical HPLC chromatogram of **glycopeptide 38**

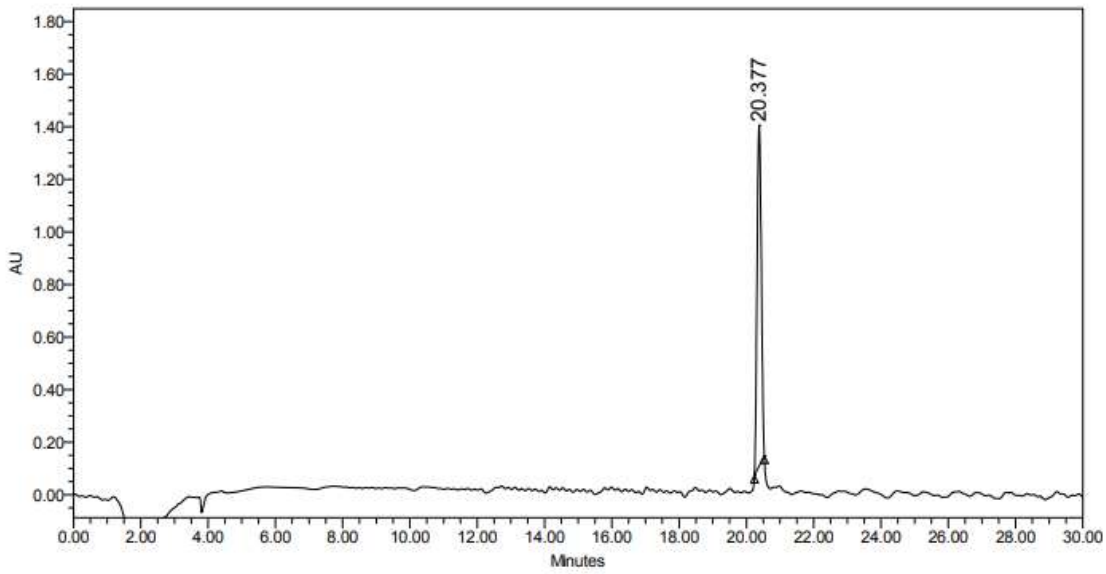


	RT	Area	% Area	Height
1	20.493	41336777	100.00	1517316

^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 39**

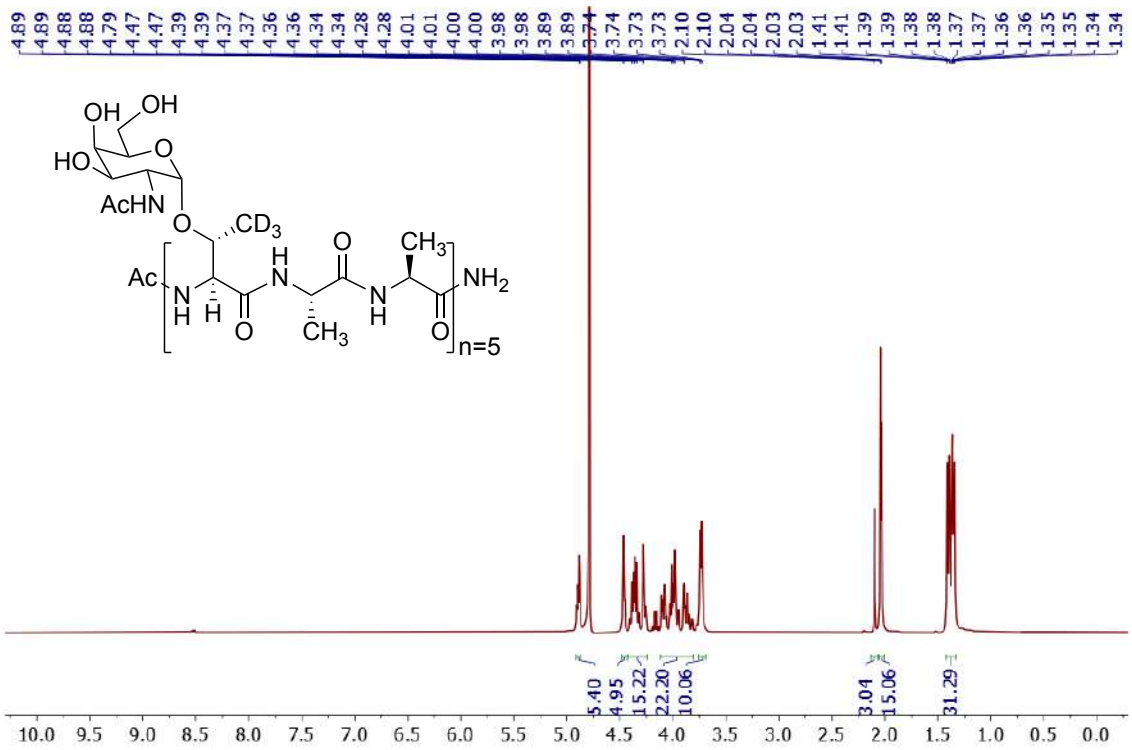


Analytical HPLC chromatogram of **glycopeptide 39**

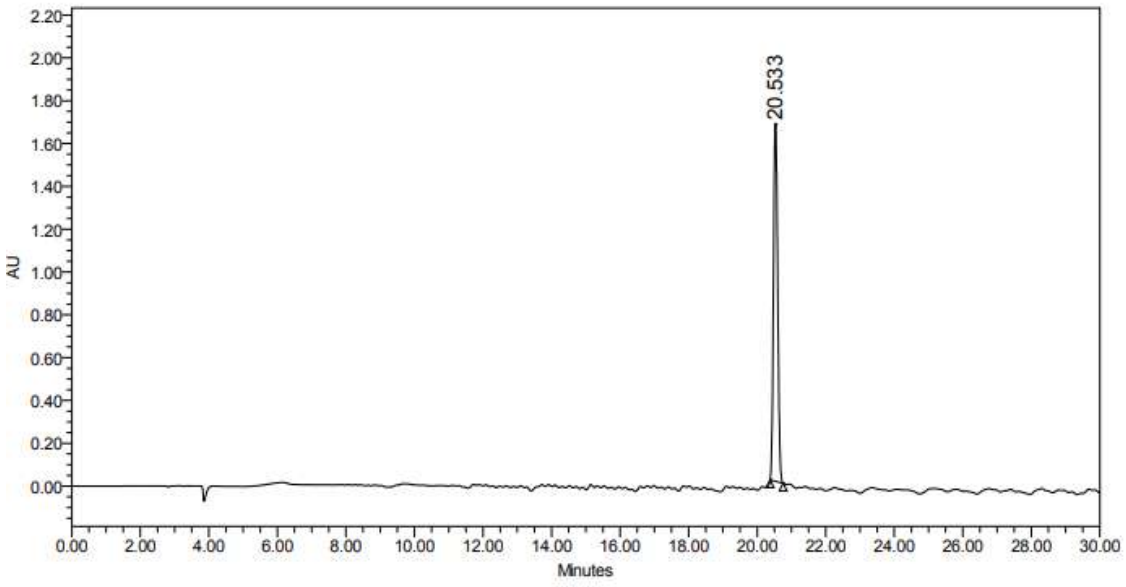


	RT	Area	% Area	Height
1	20.377	11300504	100.00	1297244

^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 40**

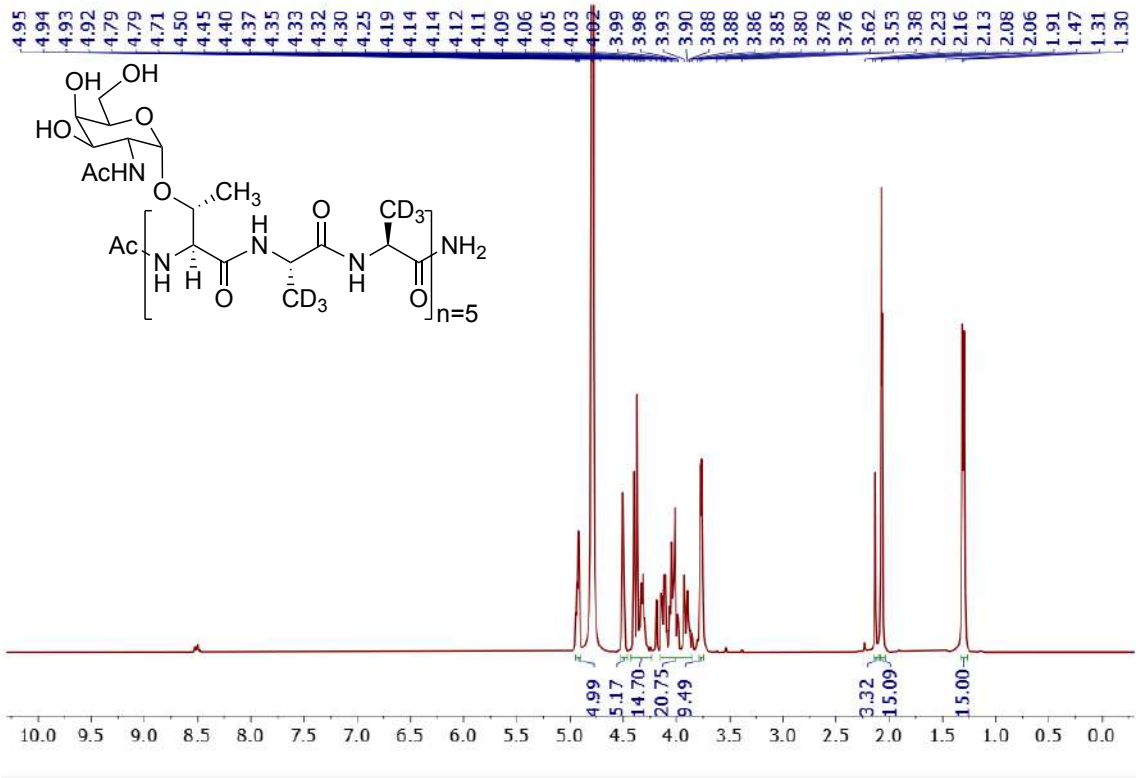


Analytical HPLC chromatogram of **glycopeptide 40**

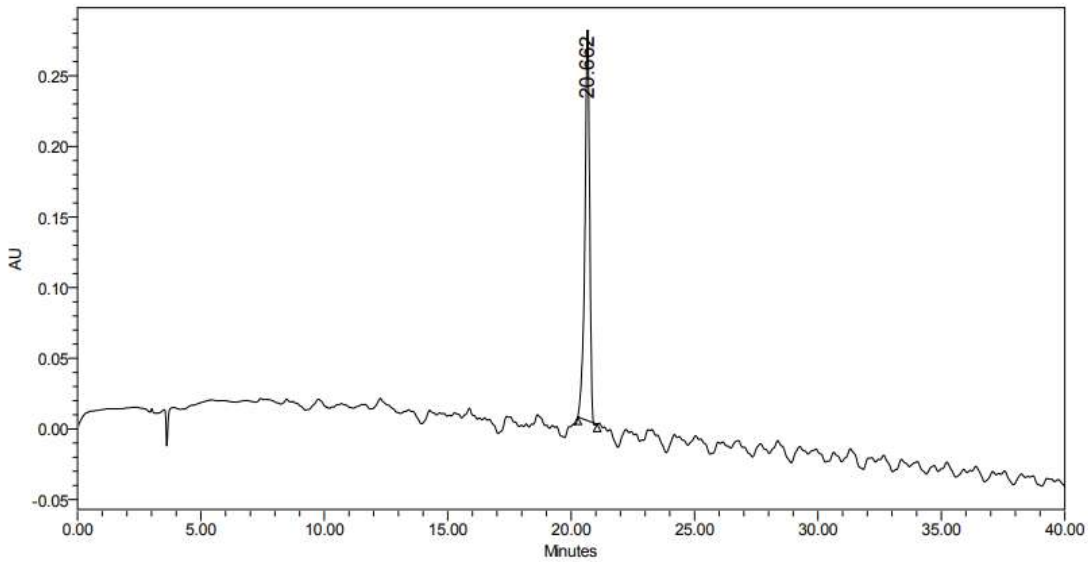


	RT	Area	% Area	Height
1	20.533	14265461	100.00	1669793

^{1}H NMR 400 MHz in D_2O registered at 298 K of **compound 41**

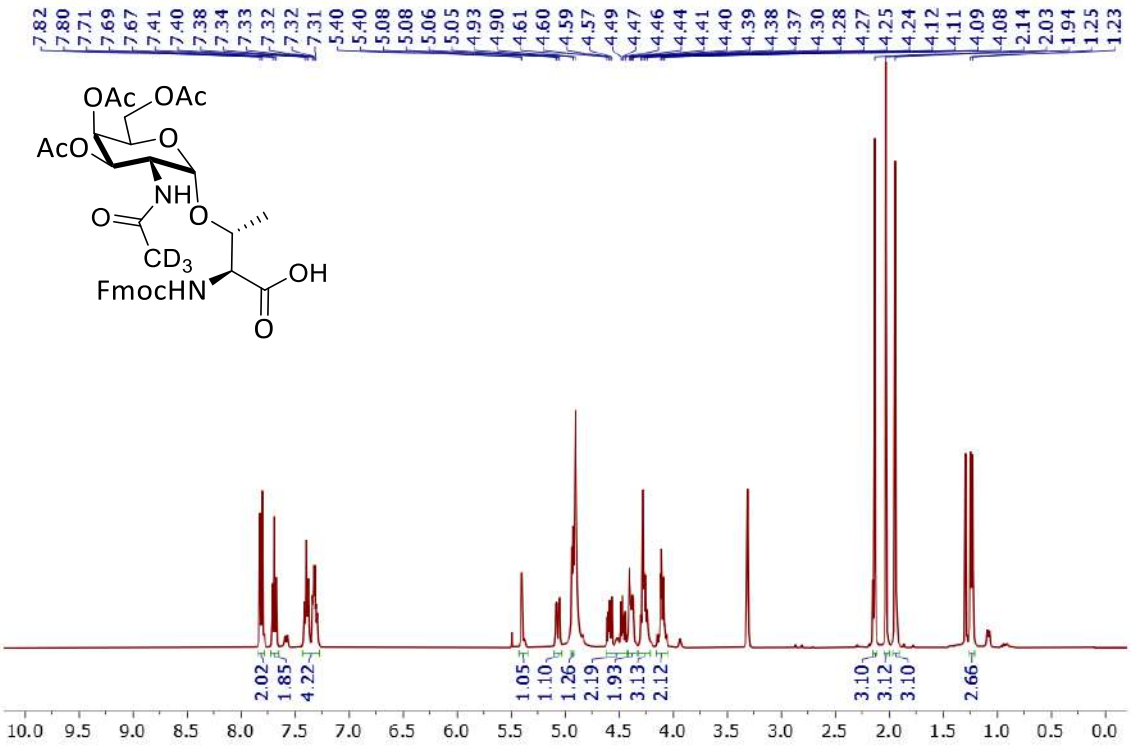


Analytical HPLC chromatogram of **glycopeptide 41**

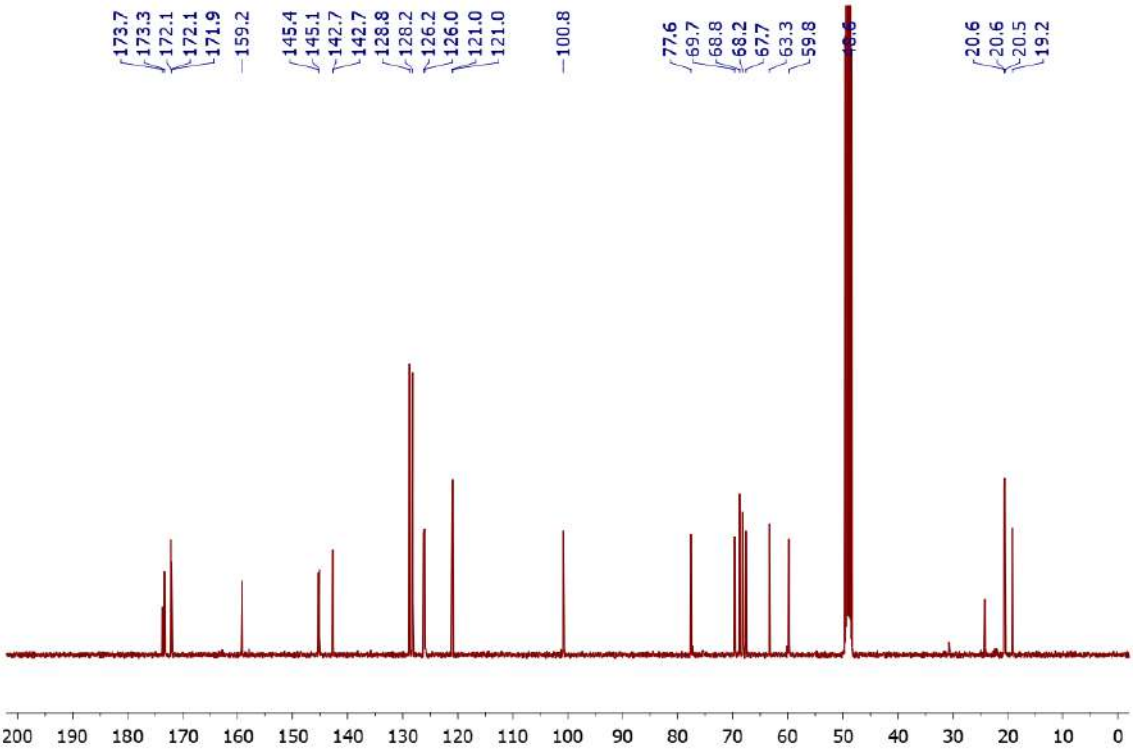


	RT	Area	% Area	Height
1	20.662	3725294	100.00	276141

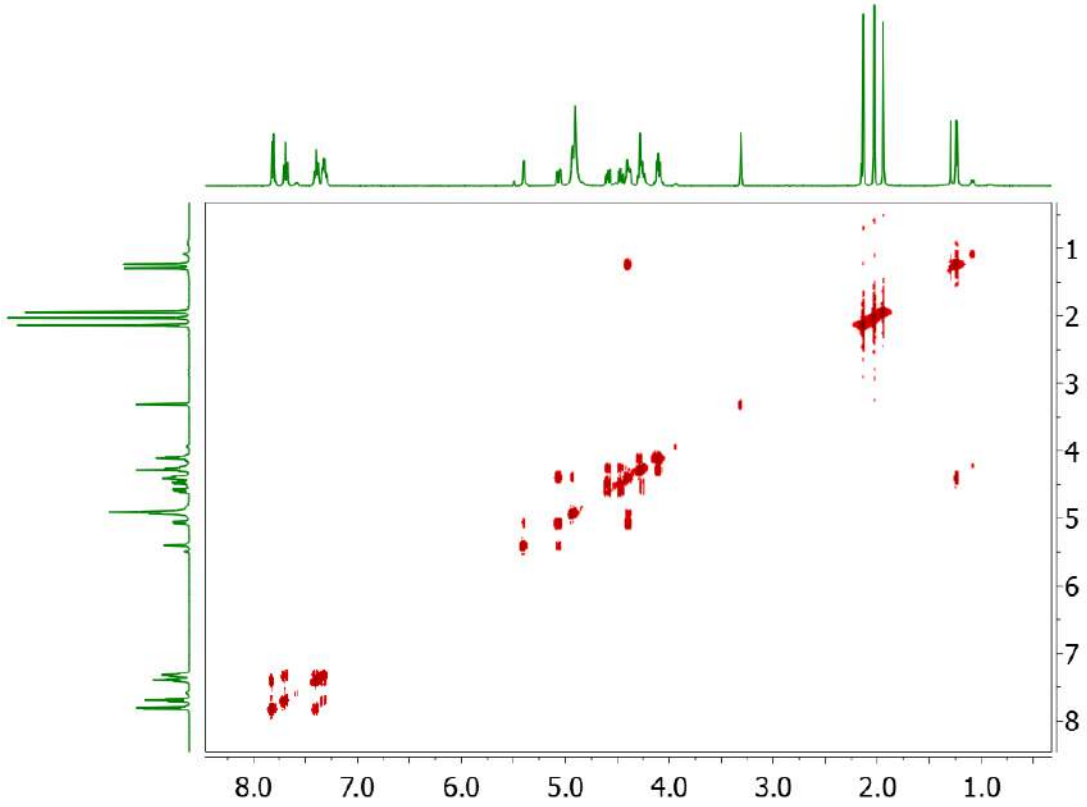
^{1}H NMR 400 MHz in MeOH registered at 298 K of **compound 42**



^{13}C NMR 100 MHz in MeOH registered at 298 K of **compound 42**



COSY in MeOH registered at 298 K of **compound 42**



HSQC in MeOH registered at 298 K of **compound 42**