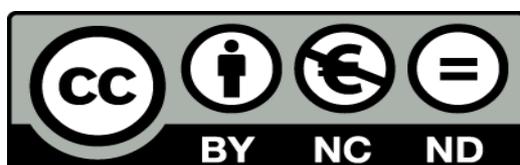# UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL

Título
Structure-based design of glycopeptides featuring unnatural Tn antigens and their applications to cancer vaccination and diagnosis

Autor/es
Iris A. Bermejo Ruiz

Director/es
Francisco Corzana López y Jesús Manuel Peregrina García

Facultad
Facultad de Ciencia y Tecnología

Titulación

Departamento
Química

Curso Académico

FACULTAD DE CIENCIAS, ESTUDIOS
AGROALIMENTARIOS E INFORMÁTICA

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

UNIVERSIDAD
DE LA RIOJA

TESIS DOCTORAL

Structure-based design of glycopeptides featuring unnatural Tn antigens and their applications to cancer vaccination and diagnosis.

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

Iris A. Bermejo Ruiz

Septiembre 2018

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

Dr. JESÚS MANUEL PEREGRINA GARCÍA, Catedrático de Química Orgánica del Departamento de Química de la Universidad de La Rioja

HACEN CONSTAR:

Que la memoria "STRUCTURE-BASED DESIGN OF GLYCOPEPTIDES FEATURING UNNATURAL Tn ANTIGENS AND THEIR APPLICATIONS TO CANCER VACCINATION AND DIAGNOSIS", realizada por Iris Alicia Bermejo Ruiz 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, septiembre de 2018

Los directores

Francisco Corzana   Jesús M. Peregrina

Dime qué precio has pagado por tu libertad
es la distancia una cárcel, es noche y olvido.
Es aburrido pensar en que no hay que pensar
qué triste acordarse del triste final
de los que cayeron durante el camino.
Dime que te has acordado de guardar
nuestras hazañas entre los cajones.

Nos volveremos a ver, La Raíz

Agradecimientos

Bueno, pues parecía que este momento nunca iba a llegar, pero está aquí finalmente. Con el tiempo una se da cuenta de que "hacer la tesis" (en química, al menos) no se reduce al mero hecho de subir al laboratorio a poner reacciones, si no que se trata de una formación integral, tanto profesional como personal, en la que el fin último consiste en abordar diferentes situaciones o escenarios problemáticos y ser capaz de resolverlos. Para ello, además de armarse de la mayor cantidad de paciencia posible, es conveniente rodearse del mejor equipo de gente que uno encuentre. Agradezco, por tanto, a todas las personas referidas a continuación que, de un modo u otro, han contribuido durante estos años de extraño devenir temporal.

Para empezar, comienzo agradeciendo a mis directores Paco y Pere. A **Pere** por permitirme formar parte del grupo de investigación y por el interés mostrado hacia mi trabajo. A **Paco** por su disponibilidad, literalmente, en cualquier momento, por sus consejos y por ser un supervisor que siempre te va ayudar con una sonrisa y, sobre todo, mucha paciencia! Gracias también por "traernos" a Xhenti. Esta apreciación por supuesto se extiende al resto de jefes del grupo de investigación en Química Biológica, a **Alberto**; por ser un pozo de conocimiento y por algunas sugerencias muy acertadas a lo largo de estos años; a **Marimar**, por su simpatía y por encargarse de las compras del grupo con tanta diligencia; a **Héctor**, por "captarme" hace años ya y meterme en esta historia de las mucinas y a **Gonzalo**, por organizar los *group meetings*, intentar poner orden en el laboratorio y, en general, preocuparte de las condiciones de los predoc e intentar mejorarlas; suerte en el futuro.

Al final, las personas con las que más tiempo se pasa son los compañeros de laboratorio, creo que he aprendido algo de cada uno de vosotros y espero haberos ayudado yo también. Por ello, agradezco a **Ismael** por sus incontables y cruciales contribuciones a esta tesis y por la lección de que hay que dar siempre el 120% en todo lo que uno haga. A **Nuria** y **Claudio**, por los cientos de horas compartidas en el laboratorio, por ser tan buenos compañeros, mucha suerte en la *hermosa* carrera investigadora. Nuria, después de la carrera, el máster y la tesis juntas ir a un laboratorio y que no estés va a ser muy raro. Claudio, muchas gracias por las decenas de RMNs que me has lanzado y faseado y por pillar siempre mis referencias a los Simpsons. Joróbate Flanders. **Xhenti**, eres TOP, hasta que llegaste tú el laboratorio y La Rioja estaban incompletos, cuando te marches lo volverán a estar. Ha sido genial ver la transformación de una chica fina florentina a una "amante" riojana de pueblo tan maja jajaja en fin, ya sabes que en Logroño tienes tu casa y en mí a una amiga. **Pablo**, tu mera presencia en el lab hace que tenga ganas de ir a trabajar, aunque luego concentrarse sea imposible entre bromas y chistes, ánimo con la tesis! **Guille**, aquella primera noche en la terraza del Odeón me conquistaste para siempre; eres imprescindible en comidas, cenas y noches de copas, y me da la sensación de que hemos compartido menos tiempo del que debíamos, hay que arreglar esto. **Emilio**, muchas gracias por todos esos consejos de hermano mayor (eres demasiado joven para ser mi padre), por aportar el punto de vista racional el nuestras discusiones absurdas de becarios y por tu *particular* sentido del humor. A las nuevas, **Alicia** y **Paula**, mucha suerte con la tesis, y mucha paciencia! Alicia, en tus manos quedan las mucinas, dales duro. Paula, me quedo tranquila porque contigo en el lab no van a faltar buenas ideas a la hora de hacer planes, gymkanas y pruebas en general. Por supuesto, gracias a los ya doctores, **Víctor R**, cuando te fuiste pensaba que el laboratorio iba a implosionar, sorprendentemente no ha sido así. **Iván** y **Mada**, me lo he pasado genial con vosotros en la uni, pero mucho mejor cuando hemos quedado fuera jajaja **Martisa** suerte con lo que se te viene encima! Muchas gracias a todos por los buenos momentos compartidos.

No puedo olvidarme de los vecinos fotoquímicos con los que hemos celebrado tantas cenas, cumpleaños, *lunches*, etc. **David**, compañero oficial de sitio en la mesa en cenas de grupo,

gracias por tu sentido del humor y los ratos cerveciles. **Raúl**, *persona importante* dentro de la Universidad de la Rioja, mil gracias por tantos buenos ratos, en el laboratorio, de cena, en el monte… por tu disposición inmejorable y por tener siempre un momento para echar una mano. **Elena**, confío en ti para que inculques los valores *adecuados* en las generaciones venideras. **Edu**, el fotoquímico de los péptidos, suerte con el sintetizador! y los jefes, **Pedro**, **Diego** y, especialmente, **Miguel Ángel**. Esta apreciación queda extendida a los compañeros inorgánicos, **Jesús**, **Nora**, **Mattia**, **Mónica**, **Chuchi**, **Quintana**… con los siempre hay un hueco para tomar *algo* en la cafetería o en la Laurel.

Agradezco además a la gente del PAS por su buena disposición. Gracias al servicio de laboratorios, por quitarnos trabajo y por su simpatía, a **Jorge** por soportarnos a la hora de comer. A **Ernesto**, por su ayuda con los problemas técnicos de laboratorio y a **Fer**, por los RMNs; pero sobretodo a **Nines**, a la que voy a echar terriblemente de menos cuando tenga que hacer un masas en el futuro. Además de ser el eje neurálgico del CCT e imprescindible en el desarrollo de la tesis de cualquiera.

Regarding to my short stay in Dortmund, I am really grateful to **Ulrika**, for allowing me be part of your lab crew and make feel like one of them. **Vanessa**, thank you so much for your advice, lunches and good moments together. My time in Germany would not have been the same without you. **Sandra**, thanks for all your help in the lab, all the best with your PhD! **Adele**, you are one of the strongest and most determined persons I have ever met, I am sure a bright future awaits you, thanks for inspiration. **Manuel**, scientist and singer, never lose your sense of humor, I will never look at a tea bag the same way! Hahaha **Jin**, thanks a lot for your help with the peptide synthesizer, peptides work-up and purification.

La vida no solo gira en torno al CCT, así que muchas gracias a los amigos "no químicos", por vuestro cariño y por distraerme de la tesis: **María** (Dr. Virumbrales, *scientific advisor*), **Cristina**, **Elena**, **Ana**, **Jorge**, **Josu**, **Laura**, **Bárbara**… A mi familia, en especial a mis padres, **Hortensia** y **Paco**, gracias por vuestro apoyo y por escuchar mis quejas recurrentes sobre la tesis y la vida en general. Agradezco también a la familia de Lorenzo, por su cariño, sobretodo a **Isabel** ("la Sisa" jaja) y **Vera**, por cuidarnos a todos de manera incansable.

No sé si a alguien le quedaba alguna duda: **Sombra**, eres la mejor perra del mundo y un recibimiento inigualable al llegar a casa. Haré todo lo posible por que en el próximo sitio en el que vivamos tengas tu propio jardín y gatos a los que perseguir sin éxito.

En último lugar, pero no por ello menos importante, (San) **Lorenzo**, muchas gracias por tu infinita paciencia, especialmente en estos últimos meses de escritura. Te dedico a ti esta tesis que espero que sirva para darnos un futuro mejor a los dos, en el que no haga falta preocuparse tanto por el dinero, sino por la próxima montaña que ascender o la próxima ciudad que visitar. As I said before: best is yet to come!!

Abstract

Within this dissertation the Tn antigen, alone and engaged in the peptide scaffold of the MUC1 mucin, has been conformationally studied. Knowing the structure of this kind of glycopeptides is key to improve their interactions with biomolecules such as lectins, enzymes or, as in this case, antibodies.

The Tn antigen is classically referred to as *N*-acetylgalactosamine (GalNAc) α-*O*-linked to either a serine (Ser) or a threonine (Thr) residue, without any distinction. Nonetheless, both derivatives present different carbohydrate orientation in solution; whereas GalNAc moiety appears perpendicularly oriented with respect to the backbone in the Thr derivative, in the case of the Ser, the sugar moiety is arranged in a parallel way over the underlying amino acid. Strikingly, an inspiring work conducted, for the first time, in the gas phase, conclude that their geometry is indistinguishable from one another, which proves the importance of the first hydration shell surrounding Tn antigen in sculpting their unique behavior in solution.

Mucin-like glycopeptides and glycoproteins, particularly MUC1, are strongly related in the development of some kinds of cancer, representing an ideal candidate to be used as part of a cancer vaccine or as an early detection diagnostic tool. Since MUC1 mucin is an endogenous glycoprotein, it is perceived as a self-antigen by the immune system, leading into immunotolerance and immunosuppression. To overcome this issue, unnatural mucins have been synthesized and their binding affinity to antibodies has been tested. Once the original structure of the MUC1 is unveiled, modifications over the carbohydrate or the underlying amino acid on the Tn antigen have been carried out, demonstrating the strong influence that subtle structural modifications may produce in their overall biological recognition and vaccine effectiveness. Within this context, a vaccine based on unnatural glycopeptide bearing an iminosugar has been tested in mice. The results show that the vaccine can boost the immune system, electing antibodies able to recognize tumor-associated MUC1 presented on the surface of cancer cell lines.

Besides, improving MUC1-antibody recognition might has implications in the use of mucins as tumor biosensors or monitoring therapy. Again, knowing the natural structure of the MUC1 mucin allows to choose a particular residue to be modified to enhance binding affinity towards antibodies, hence, increasing sensibility of this kind of tests to be applied as early detection tools. As a proof-of-concept, a unique antigen featuring a fluorinated glycopeptide has been successfully tested to find circulating antibodies in patients with prostate cancer.

Resumen

El objetivo principal de la presente tesis doctoral es el estudio conformacional del antígeno Tn, ya sea solo o incluido en el esqueleto peptídico de la mucina MUC1. Conocer la estructura de este tipo de glicoproteínas es clave para poder mejorar su interacción con moléculas de interés biológico, tales como lectinas, enzimas o, el caso que nos ocupa, anticuerpos.

El antígeno Tn se define comúnmente como una unidad de *N*-acetilgalactosamina (GalNAc) conectada mediante un enlace α-*O*-glicosídico a un aminoácido, ya sea una serina (Ser) o una treonina (Thr), sin ningún tipo de distinción entre las mismas. Sin embargo, ambos derivados presentan diferencias en la orientación del carbohidrato en disolución; mientras que la unidad de GalNAc se presenta de manera perpendicular sobre el esqueleto peptídico del residuo Thr, en el caso de la Ser ésta se dispone de manera paralela sobre el aminoácido subyacente. Sorprendentemente, y observado por primera vez en fase gas, esta diferencia desaparece, siendo indistinguible la geometría de ambos aminoácidos glicosilados. Este hecho pone de manifiesto la importancia que presentan las moléculas de agua que rodean a este tipo de moléculas y su crucial contribución estructural.

Glicopéptidos y glicoproteínas del tipo mucina, en particular la mucina MUC1, están relacionados de manera intrínseca con el desarrollo de ciertos tipos de cáncer, y se presentan, por tanto, como candidatos prometedores a la hora de diseñar una vacuna o como método para el diagnóstico temprano. Dado que la mucina MUC1 es una glicoproteína endógena, el sistema inmune la percibe como propia durante el desarrollo de la enfermedad, produciéndose circunstancias en las que el sistema inmune tolera su aparición y evita cualquier tipo de respuesta inmune contra ella. Para solucionar esta cuestión, durante esta tesis, se han sintetizado mucinas no naturales y se ha probado su afinidad con diferentes anticuerpos. Una vez se conoce la estructura original de la MUC1, se han llevado a cabo modificaciones sobre el carbohidrato o el aminoácido del antígeno Tn, demostrando la influencia que sutiles modificaciones estructurales pueden producir en el reconocimiento biológico global de este tipo de compuestos o su efectividad como vacunas. En este contexto, una vacuna basada en un glicopéptido no natural conteniendo un iminoazúcar ha sido probada en ratones. Los resultados muestran que esta vacuna es reconocida por el sistema inmune, siendo capaz de generar anticuerpos que reconocen MUC1 natural que aparece en la superficie de células cancerosas.

Además, la mejora del reconocimiento MUC1-anticuerpo puede presentar interesantes implicaciones en el uso de las mucinas como sensores tumorales o en el seguimiento del tratamiento. De nuevo, conocer la estructura de la mucina MUC1 natural permite poder elegir un residuo determinado para ser modificado. Así, la afinidad hacia los anticuerpos será mayor y se podrá aumentar la sensibilidad de este tipo de pruebas, que pueden ser aplicadas en el diagnóstico precoz del cáncer. Como prueba de concepto, un glicopéptido que incorpora una prolina fluorada se ha probado de manera exitosa como sensor para determinar anticuerpos anti-MUC1 en el suero de pacientes con cáncer de próstata.

Index

Abbreviations

δ	chemical shift
ΔG	relative Gibbs free energy
ΔH	relative enthalpy
ΔS	relative entropy
ºC	celsius degree
^{1}H NMR	proton nuclear magnetic resonance
^{13}C NMR	carbon nuclear magnetic resonance
Å	angstrom
AAb	autoantibody
Ab	antibody
Ac	acetyl
Ac$_2$O	acetic anhydride
AcOEt	ethyl acetate
AcOH	acetic acid
APC	antigen presenting cell
Arom.	aromatic
BCR	B cell receptor
BLI	bio-layer interferometry
BSA	Bovine Serum Albumin
Bn	benzyl
ca.	*circa* (around)
calcd.	calculated
COSY	1H-1H correlation NMR spectroscopy
CRC	colorectal cancer
d	doublet
DIC	*N,N'*-diisopropylcarbodiimide
DIPEA	*N,N*-diisopropylethylamine
DC	dendritic cell
DCM	dichloromethane
DMF	dimethylformamide
EDC	ethyl -3(3-dimethylamino)propylcarbodiimide
EDT	1,2-ethanodithiol
ELISA	enzyme linked immunosorbent assay
ESI-MS	electrospray-mass spectrometry
Et$_2$O	diethyl ether
Fmoc	9-fluorenylmethyloxycarbonyl
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GlnNAc	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
HRP	horseradish peroxidase
HSQC	heteronuclear single quantum correlation
i.e.	for example
^{i}PrOH	isopropanol
IR	infrared
IRID	infrared ion-dip
IUD	intrauterine contraceptive device
J	coupling constant
K$_D$	dissociation constant
KLH	Keyhole Limpet Hemocyanin
m	multiplet
mAb	monoclonal antibody
MALDI	matrix-assisted laser desorption ionization
MBHA	4-methylbenzhydrylamine
MD	molecular dynamics
MD-tar	molecular dynamics with weighted time-averaged restraints
Me	methyl
MBS	m-maleidobenzoyl-*N*-hydroxysuccinimide
MeOH	methanol
MeONa	sodium methoxyde
MeSer	α-methylserine
MHC	major histocompatibility complex
MS	mass spectrometry
MW-SPPS	microwave-assisted solid-phase peptide synthesis
Neu5Ac	*N*-acetylneuraminic acid or sialyc acid
NHBn	benzylic amide
NHS	*N*-hydroxysuccinimide
NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
OVA	Ovoalbumin
PADRE	Pan DR
Pam$_3$Cys	dipalmitoyloxypropyl-*N*-palmitoylcysteine
PBS	phosphate buffered saline
PDB	protein data bank
PEG	polyethylene glycol
PG	protecting group
Ph	phenyl
ppm	parts per million

PSA	prostate specific antigen
QM	quantum mechanical
R	universal gas constant or substituent
RAFT	regioselectively addresable functionalised template
RFU	relative fluorescence intensity
ROE	rotating frame nuclear Overhauser effect
ROESY	rotating frame nuclear Overhauser effect spectroscopy
Rt	retention time
RU	relative unit
s	singlet
scFv	single-chain variable fragment
SPPS	solid-phase peptide synthesis
SPR	surface plasmon resonance
STD	saturation transfer difference
STF antigen	Sialyl Thomsen-Friedenreich antigen
STn antigen	Sialyl Tn antigen
Su	succinimide
t	triplet
TAA	tumor-associated antigens
TACA	tumor-associated carbohydrate antigens
^{t}Bu	*tert*-butyl
TCR	T cell receptor
TF antigen	Thomsen-Friedenreich antigen
TFA	trifluoroacetyc acid
THF	tetrahydrofuran
TIS	triisopropylsilane
TLC	thin layer chromatography
TMB	3,3',5,5'-tetrametylbenzidine
TMS	tetramethylsilane
TSA	tumor-specific antigens
Ttox	Tetanus Toxoid
UDP	uridine diphosphate
UV	ultraviolet
V	potential energy
z	charge

A Ala — Alanine

R Arg — Arginine

N Asn — Asparagine

D Asp — Aspartate

C Cys — Cysteine

Q Gln — Glutamine

E Glu — Glutamate

G Gly — Glycine

H His — Histidine

I Ile — Isoleucine

L Leu — Leucine

K Lys — Lysine

M Met — Metionine

F Phe — Phenilalanine

P Pro — Proline

S Ser — Serine

T Thr — Threonine

W Trp — Tryptophan

Y Tyr — Tyrosine

V Val — Valine

Legend:
- Negatively charged
- Positively charged
- Polar uncharged
- Nonpolar aliphatic
- Aromatic

Introduction

1

1.1. Cancer and the immune system

Human cells grow and divide in an orderly way as the human body requires them. When this process grows out of control, due to abnormalities in the genetic material, the balance proliferation/death in cells is altered. The ensemble of all the diseases where this disruption appears is commonly designated as cancer.

Some causes of cancer might be genetic factors and/or environmental factors, for instance physical or chemical agents, such as UV and ionizing radiation, tobacco, alcohol, asbestos or air pollution. An unhealthy life style, lack of physical activity and unbalanced diet habits are also potentially related to the incidence of cancer.

Cancer is one of the most relevant public health problems all around the world due to its incidence, prevalence and mortality rates. It constitutes the second leading cause of death in developed countries, only after cardiovascular diseases.[1] For example, in 2017, 1.688.780 new cancer cases and 600.920 cancer deaths were predicted to occur in the United States.[2]

The traditional treatments for cancer include surgery, radiation and chemotherapy. In general, radiation and chemotherapy come with many side effects. Unfortunately, they attack not only the tumor, but also the surrounding tissues and the rapidly dividing normal cells. Therefore, undesired effects, like immunosuppression, ulcers, hair loss and fatigue could appear. Despite the efforts put on their improvement, these conventional procedures are clearly not sufficient to cure cancer. For this reason, new approaches should be envisioned. One of the most promising alternatives is immunotherapy. That is, the stimulation of the host's immune system to identify and destroy cancer cells.

The immune system plays a key role in cancer development, trying to maintain an equilibrium between immune recognition and tumor growth. The concept of immunoediting[3,4] refers to a dynamic process whereby the immune system not only protects against cancer development, but also shapes the character of emerging tumors.

Cancer immunoediting is composed of three phases: elimination, equilibrium and escape.[5–7] During the first stage, the immune system is able to recognize and destroy most cancer cells, because they present tumor antigens. Even so, constant tumor cell division may elude immune elimination, resulting in a balanced state between elimination and production of new cancer cells, designated as equilibrium, which gives the appearance of cancer dormancy. The final escape phase occurs when tumor progression overcomes immunosurveillance, resulting in clinical cancer. Therefore, the most important goals of immunotherapy consist in provide the immunological machinery with adequate danger signals when tumor cells develop, to overcome immunological tolerance[8] and focus just on cancer cells, to avoid secondary effects.

In pursuit of an effective cancer treatment, numerous approaches have been envisaged, always considering that the basis for it are targeting the appropriate immunizing antigen and priming the suitable costimulatory signals.

1.2. Cancer immunotherapy and cancer vaccines

William Bradley Coley, a surgeon in New York, is considered the father of cancer immunotherapy.[9] At the very end of the 19th century,[10] he started treating inoperable sarcoma patients with *Streptococcus Pyrogenes* and *Serratia Marcescens*, inoculating these bacteria directly into the tumor. In that way, he provoked an immune response – an acute infection known as *Erysipelas*. He reported numerous cases,[11,12] in many of which the patient successfully recovered from cancer. This approach led to the use of the "Coley's toxins"[13] as an early immunological treatment of cancer, which was rapidly substituted by an emerging radiotherapy, due to the lack of reproducibility and the aggressiveness of Coley's approach.[14]

Due to the insufficient understanding of immune mechanisms, the development of these therapies was stalled for nearly a century. The idea of using immunotherapy in cancer was reevaluated when the theory of cancer immunosurveillance was proposed in 1957 by Burnet and Thomas.[15] They postulated that the immune system can recognize and destroy somatic cells transformed by spontaneous mutations.[3,16] It was later extended into the more acute concept of cancer immunoediting, mentioned above, which nowadays is the most accepted approach.

From the 1970s onwards, scientific and methodological innovations included the development of the hybridoma technology to produce monoclonal antibodies[17] as tools to engage immune mechanisms and to increase the understanding of pathways and targets of the immune system.

During the past 20 years, multiple different immunotherapeutic treatments have been developed, mainly divided into two main categories: passive and active therapies.

Passive immunotherapy involves the administration of exogenously produced components to mediate an immune response.[18] Serve as an example therapeutic antibodies, such as trastuzumab, widely used in clinical practice.[19–21]

On the other hand, in the active therapies the body is exposed to an antigen that boost an immune response and also immunological memory.[22] It can also be mediated by antibodies, known as immunostimulatory antibodies,[23] like ipilimumab, that has been used in ovarian cancer clinical studies. [24]

Another popular trend in active therapies are anticancer vaccines, which are designed to either increase immune recognition of tumor cells or to enhance the anti-tumor immune response through lymphocyte activation.[25] These come into two variations, either prophylactic form, aiming at delaying the relapse of the disease, which have been

administered in cancers of viral origin, such as human papillomavirus, HPV;[26,27] or therapeutic form, which could trigger immune response to attack existing cancer cells.[28]

The major end in the development of therapeutic vaccines is to maximize the immunogenicity without compromising safety and tolerability. These vaccines must be safe, but also able to prime a potent and long-lasting immune response. Key characteristics on their synthesis are size, shape and surface molecule organization.[29]

One of the first attempts to develop a cancer vaccine consisted in the removal of the patient's immune cells to be inactivated and re-injected then into the patient.[30] Unfortunately, this methodology is too expensive in labor and cost of such personalized medicine. Another route to obtain vaccines involves a biological source, such as engineered bacteria like *Salmonella*[31,32] and *Listeria monocytogenes*,[33] where the epitope is isolated. The main drawback of this approach is the fact that immunogenic epitopes exist as heterogeneous mixtures of molecules and for the interaction with specific antibodies is required a precise structure, high purity and sufficient amount.

Therefore, significant efforts are currently focused on the development of chemical and enzymatic methods for the synthesis of structurally well-defined and highly pure vaccines in relatively large amounts. In this respect, synthetic cancer vaccination encompasses a diverse range of strategies, one of the most important approaches is the construction of the so-called subunit vaccines. Their formulation involves an ensemble composed by an antigen, an adjuvant and a carrier.

Antigens are preferably molecules only present or overexpressed on cancer cells – usually a protein or a synthetic peptide,[34,35] oligosaccharide or DNA fragment.[36] They can be classified as Tumor-Specific Antigens – TSA, only displayed in tumor cells – or Tumor-Associated Antigens – TAA, presented in both tumor cells and normal cells. Antigens are capable of triggering a selective immune response in the host organism, which makes them essential within the scaffold of any vaccine.

The antigen is combined with an **adjuvant** or adjuvant system, to enhance antigen immunogenicity, elicit stronger immune response and to reverse the tumor-mediated immunosuppressive mechanisms. They should also be able to create an antigen depot,[29] which results in prolonged exposure of the antigen to the immune system. Some classical examples are water-in-oil emulsions, such as Freund's adjuvant, which has been used widely in cancer vaccines, but failed to prove efficacy in clinical trials[18] and aluminum salts.[37,38] More recent strategies involve the use of a family of transmembrane proteins known as Toll-like receptors[39,40] that recognize conserved molecular patterns of microbial origin, priming an immune response. They are able to induce strong antitumor activity[41] and T cell-mediated responses.[42] Toll-like receptors are usually linked to the antigen, overcoming side effects while immunogenicity persists, since only the cells that interact with the antigen result activated by the adjuvant.

Vaccines also incorporate a delivering system, named **carrier**. Classical approaches involve the conjugation of the antigen to a carrier protein, for example keyhole limpet hemocyanin – KLH –, bovine serum albumin – BSA –, ovalbumin – OVA – or tetanus

toxoid – TTox.[43] Other strategies include recombinant viral vectors,[44] liposomal microspheres[45,46] or polymeric[47,48] and inorganic nanoparticles.[49]

The carriers are necessary for the vaccine to be uptaken by the antigen presenting cells (APC), they enhance the presentation of the antigen and induce the activation of T helper cells.[50] Nevertheless, the major drawback of employing carrier proteins is that they are inherently immunogenic, eliciting B cell responses, hence leading, in some cases, to epitope suppression.

The subunit vaccine strategy presents many advantages, since they can be produced and administrated in an easy and inexpensive manner. On the contrary, its major challenge is the selection of the appropriate antigen, adjuvant and carrier to boost the optimal immune response and ensure the delivery of the vaccine into APC.[51]

The scope of the present PhD dissertation lies in the design and study of novel antigens based on unnatural glycopeptides to be used in cancer treatment. Hence, the rest of this chapter will be focused on introducing the glycopeptides used in this work.

1.3. Tumor-Associated Carbohydrate Antigens

Glycans are essential for a vast range of biological functions in the organism. For example, glycans are crucial for protein stability and secretion, since they regulate the folding process for freshly synthesized proteins, in controlling the proper functioning of the endoplasmic reticulum and for protein targeting in the secretory pathway.[52] When glycans are presented on the cell surface, they participate in a large number of processes like cell communication, adhesion and migration, signal transduction, host-pathogen interactions and immune surveillance.[53,54] Besides, it has been demonstrated the implication of glycoproteins in tumor pathology,[55–57] since the majority of human cancers are characterized by aberrant glycosylation.[58–61]

The surface of healthy cells is characterized by the presence of long and highly-branched carbohydrate chains.[62] Nevertheless, as a result of an incorrect or ineffective action of the glycosyltransferases, the glycan structures of cancer cells present overexpression of the truncated versions of the aforementioned oligosaccharides. The carbohydrates and peptide backbones that are normally shielded, in cancer tissues are exposed, becoming more accessible to the immune system and acting as tumor antigens or, since they involve carbohydrates, Tumor-associated Carbohydrate Antigens, TACA.[63–65] (Figure 1.1)
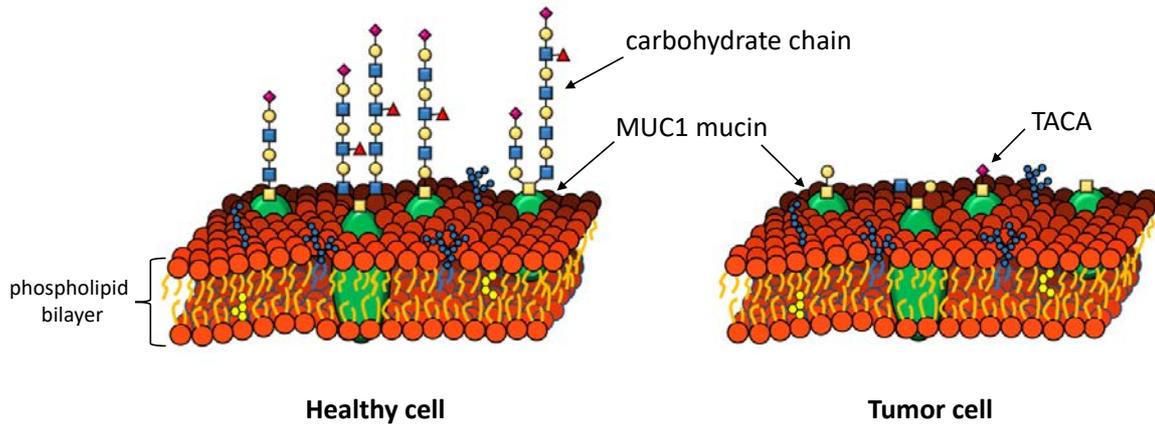
Figure 1.1 Different glycosylation patterns in healthy (left) and tumor (right) cells.

TACA are found in the glycocalyx of tumor cells, acting as biomarkers that can be used to differentiate cancer cells from normal cells. There are two main divisions of TACA:[55,66] glycolipid antigens, which are carbohydrates linked to lipids, anchored to the lipid bilayer on the cell surface, or glycoprotein antigens, where the sugars are attached to a protein backbone.

Glycolipid antigen category includes gangliosides, such GD2, GD3, GM2 and GM3; globo class, for example Globo-H, and blood group determinants, like Lewis^x, Lewis^y and their sialylated derivatives.[63] In glycoproteins, glycans can be bound to proteins involving amide linkages to asparagine residues (*N*-glycosylation) and glycosidic linkages to serine (Ser) and threonine (Thr) side chains (*O*-glycosylation).[67–69]

Within the wide variety of O-glycans, such as *O*-galactose, *O*-glucose, *O*-fucose, *O*-mannose and *O*-GlcNAc; *O*-GalNAc represents the most abundant, occurring on at least one Ser/Thr residue in more than 85% of secreted proteins.[70] In this context, the most significant cancer antigens resulting from incomplete *O*-glycosylation are the Thomsen-Friedenreich (TF)-related antigens, constituted by short *O*-GalNAc glycans: TF or T antigen, T antigen *nouvelle* (Tn antigen) and their sialylated derivatives: STF and STn antigens. (Figure 1.2)

The Tn antigen features the simplest structure, it comprises the monosaccharide GalNAc α-*O*-linked to a Ser/Thr residue within the peptide backbone, while the T antigen consists of the addition of Gal to the GalNAc moiety (Gal-β–1,3-GalNAc-α1-*O*-Ser/Thr). Their sialylated counterparts involve the binding of terminal sialic acid to carbon 6 of GalNAc (leading to the STn antigen) and to carbon 3 on Gal (ST antigen), preventing further elongation of the glycan.
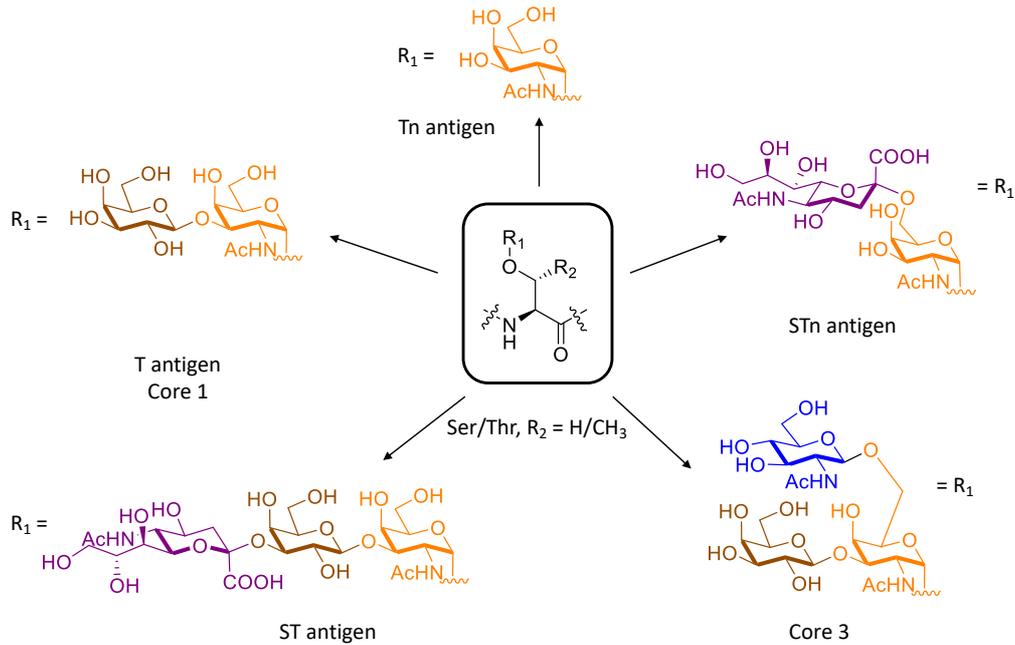
Figure 1.2 Structures of some of the simplest TACA.

The aforementioned antigens are rarely exposed in normal tissue, but are overexpressed on the majority of human tumor cells – such as breast,[71,72] stomach,[73–75] colorectum,[76] pancreas,[77,78] bladder,[79] ovary[80] and breast.[71] Thus, they represent an excellent target for cancer vaccine development.

TACA as described before appear on the surface of diverse glycoproteins, for instance, in the family of mucins, described in the next section.

1.4. Mucins

Epithelial cells in the human body use a mucus lining to protect themselves from pathogens, lubricate the cell surface and lower the mechanical stress on the epithelium. The protein component of this mucus is constituted mainly by the mucin family. These glycoproteins can be widely found within the human body, for example, in the airways of the respiratory system, the stomach, the intestinal tract and in secretory epithelial surfaces of liver, pancreas, bladder and kidney, ocular surface, nasal cavity and salivary and lacrimal glands.[81,82] They are used as cell-surface receptors and sensors, coordinating responses like proliferation, differentiation, apoptosis and secretion.[83] More importantly, mucins are strongly related in the development of cancer.

Mucins are commonly divided into two different categories:[84,85] secreted – entirely extracellular – and transmembrane mucins, which are anchored to the cell membrane, and present both a cytoplasmic tail and extracellular regions. The former type is represented by MUC2, MUC5AC, MUC5B, MUC6-8 and MUC19, whereas the transmembrane kind is comprised by MUC1, MUC3, MUC4, MUC12-17, MUC20 and MUC21 mucins. They present several unique functional domains, specific to each group.

For the purpose of this thesis, we focused on MUC1 mucin, which is aberrantly glycosylated and overexpressed on the surface of several types of carcinoma cells. It localizes to the apical membranes of normal secretory epithelial cells,[86] mainly in the breast, pancreas, eyes membranes, respiratory system gastrointestinal tract and internal organs of the female reproductive system.[82]

As a membrane-tethered mucin, it presents a C-terminal transmembrane subunit – MUC1-C – anchored to the cell surface, which is non-covalently complexed with the N-terminal subunit – MUC1-N. The latter constitutes a physical protecting barrier around the cell surface that can shed from it to the circulation,[87,88] leaving the former as receptor for transformation and tumor progression pathways.[89] (Figure 1.3)
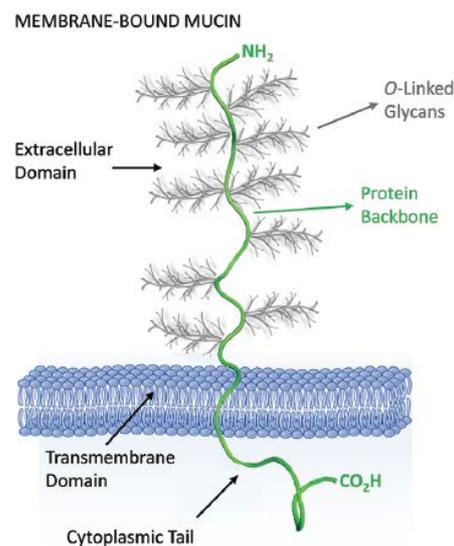
Figure 1.3 MUC1 mucin transmembrane structure.[90]

The N-terminal subunit, the extracellular domain, is comprised by highly conserved tandem repeats of 20 amino acids each (HGVTSAPDTRPAPGSTAPPA),[84,91,92] containing five potential primary *O*-glycosylation sites – three Thr and two Ser residues. Both acting as linkage amino acids through GalNAc, constitute the first anchoring site for the rest of the glycan structure. The goal of the present research is focused on the N-terminal subunit, which from this point forward will be designated just as MUC1 mucin.

As mentioned in the previous section, when the glycosyltransferases fail to elongate the saccharide chain, the truncated sugars appear exposed to the immune system, behaving as TACA.

In MUC1 mucin, *O*-glycosylation begins with GalNAc-Ser/Thr (Tn antigen) in the Golgi apparatus, the subsequent sugars are then individual and sequentially added.[93,94] First, galactose is attached building the Core 1 or T antigen, acting in turn as a substrate for the GlcNAc residue, creating Core 2 from where the chain elongation commences (Figure 1.4). The oligosaccharide is terminated by sialic acid or fucose.[95]
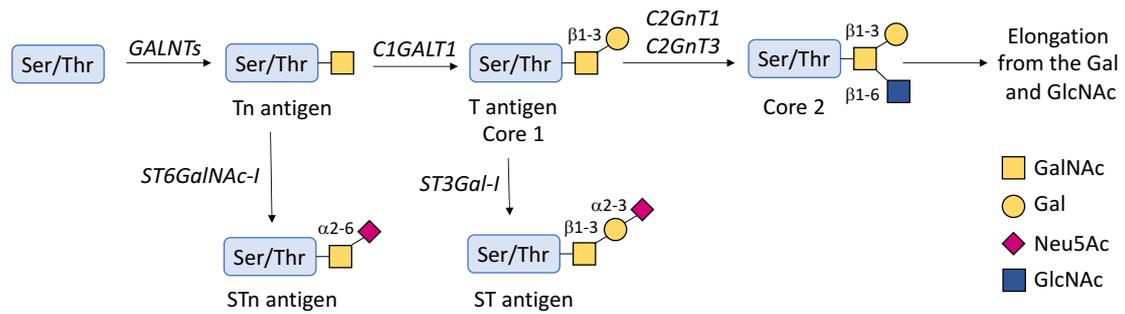
Figure 1.4 The O-linked glycosylation biosynthetic pathway.

Nevertheless, as a result of the down-regulation of glycosyltransferases, in cancer cells T and Tn antigens are expressed in MUC1, since transformation of Core 1 into Core 2 is severely reduced.[90] Besides, sialic acid is prematurely added to Core 1, terminating chain extension and yielding into the formation of ST and STn antigens.[93]

In addition to the presence of the aforementioned TACA, MUC1 expression appears exclusively on the apical surface in normal epithelial cells, bordering a lumen. In tumor cells, however, MUC1 mucin is highly expressed all over the surface, being found across the plasma membrane and in the cytoplasm.[84,93]

1.5. Glycopeptide-based cancer vaccines

Based on these findings, MUC1 mucin represents an ideal candidate to be used as part of a cancer vaccine or as an early detection diagnostic tool. Actually, the glycopeptide-based vaccines[34,96] are one of the best-known modalities of active cancer vaccination. They elicit endogenous tumor-specific response, representing an attractive cancer treatment, since their ease of use and low toxicity in clinical trials.[35,97]

Nevertheless, TACA-based antitumor vaccines development remains a challenge due to their inherently T cell independent nature.[63,98] This means that they do not bind to major histocompatibility complexes (MHC), hence they cannot activate T cells by themselves. On the contrary, TACA are prompted to trigger B cell responses, by cross-linking the B cell receptor (BCR) without any MHC-II interaction,[96] producing only low-affinity IgM antibodies (Figure 1.5 upper panel).
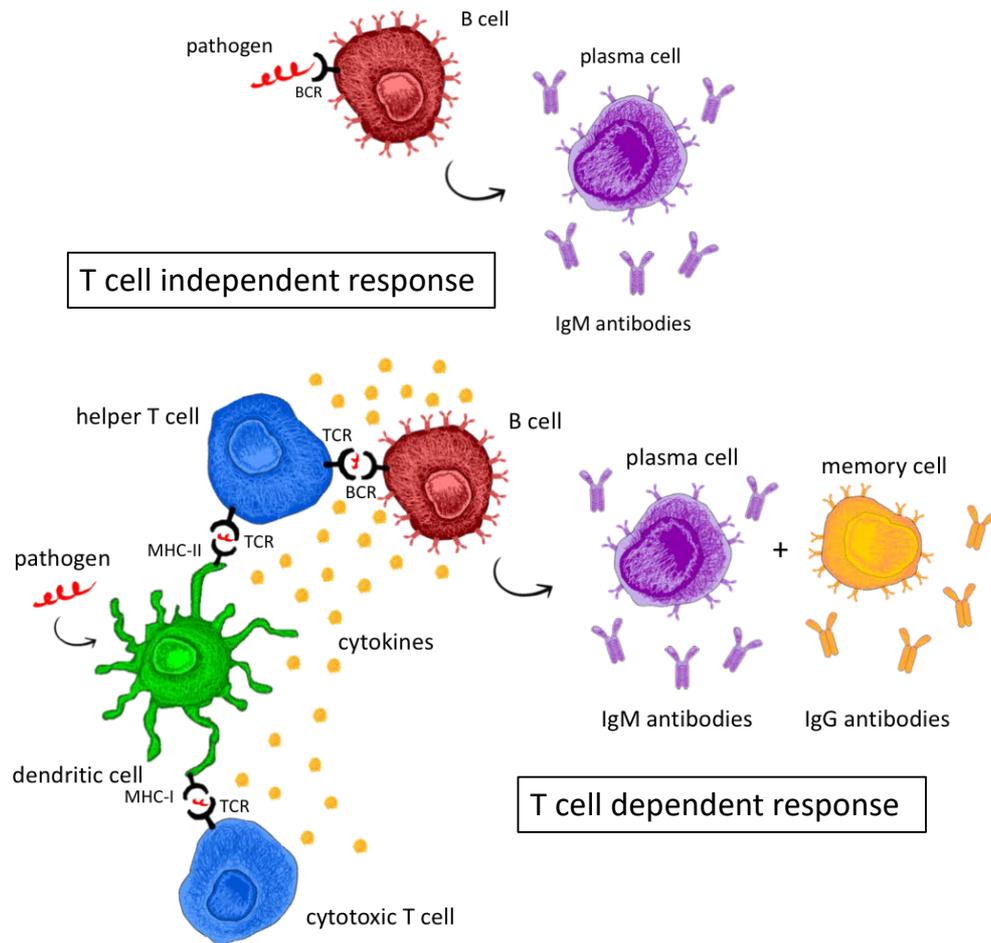
Figure 1.5 Simplification of the process of the immune response.

To solve this issue, carbohydrate-based vaccines need to be bound to peptides that can be captured and internalized by APCs, for instance, dendritic cells, DCs (Figure 1.5 lower panel). This kind of cells cleave the vaccine into short peptide epitopes to be presented onto the cell surface complexed to the MHC-II, activating helper T cells. These interaction triggers the release of chemical substances – cytokines – to attract macrophages to the site of infection, which activate both cytotoxic T cells, that control the infection, and B cells. When B cells are stimulated this way, antibody class-switching occurs, producing high-affinity IgG antibodies and memory cells, which are key to induce a robust and long-lasting immune response.

However, even if TACA are overexpressed in malignant cell surfaces, they are still endogenous host cells and appear presented in small amounts on normal cells. This fact leads into immunotolerance and immunosuppression, since they are perceived as self-antigens by the immune system. To overcome immunotolerance there have been used immunostimulant components, such as carriers, or chemically modified TACA.[99–101]

Hence, the main goal of the present PhD dissertation is to study the structure of both MUC1 mucin and Tn antigen and analyze their interactions with anti-MUC1 antibodies. Understanding molecular recognition processes would allow us to modify selectively the antigen structure to enhance binding affinity.

1.6. References

1.   Bray, F., Ren, J. S., Masuyer, E. & Ferlay, J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int. J. Cancer* **132,** 1133–1145 (2013).
2.   Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2017. *CA. Cancer J. Clin.* **67,** 7–30 (2017).
3.   Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat. Immunol.* **3,** 991–998 (2002).
4.   Shankaran, V., Ikeda, H., Bruce, a T., White, J. M., Swanson, P. E., Old, L. J. & Schreiber, R. D. IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410,** 1107–1111 (2001).
5.   Mittal, D., Gubin, M. M., Schreiber, R. D. & Smyth, M. J. New insights into cancer immunoediting and its three component phases-elimination, equilibrium and escape. *Curr. Opin. Immunol.* **27,** 16–25 (2014).
6.   Dunn, G. P., Old, L. J. & Schreiber, R. D. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* **21,** 137–148 (2004).
7.   Dunn, G. P., Old, L. J. & Schreiber, R. D. The three Es of cancer immunoediting. *Annu. Rev. Immunol.* **22,** 329–360 (2004).
8.   Pardoll, D. M. Cancer vaccines. *Nat. Med. Vaccine Suppl.* **4,** 525–531 (1998).
9.   McCarthy, E. F. The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. *Iowa Orthop. J.* **26,** 154–8 (2006).
10. Coley, W. B. II. Contribution to the Knowledge of Sarcoma. *Ann. Surg.* **14,** 199–220 (1891).
11. Coley, W. B. The treatment of malignant tumors by repeated inoculations of erysipelas: with a report of ten original cases. *Am. J. Med. Sci.* **105,** 487–511 (1893).
12. Coley, W. B. The Treatment of Inoperable Sarcoma by Bacterial Toxins (the Mixed Toxins of the Streptococcus erysipelas and the Bacillus prodigiosus). *Proc. R. Soc. Med.* **3,** 1–48 (1910).
13. Zacharski, L. R. & Sukhatme, V. P. Coley's toxin revisited: Immunotherapy or plasminogen activator therapy of cancer? *J. Thromb. Haemost.* **3,** 424–427 (2005).
14. Hoption Cann, S. A., van Netten, J. P. & van Netten, C. Dr William Coley and tumour regression: a place in history or in the future. *Postgrad. Med. J.* **79,** 672–680 (2003).
15. Burnet, F. M. Cancer; a biological approach. I. The processes of control. *Br. Med. J.* **1,** 779–786 (1957).
16. Burnet, F. M. The Concept of Immunological Surveillance. *Prog. Exp. Tumor Res.* **13,** 1–27 (1970).
17. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256,** 495–497 (1975).
18. Melero, I., Gaudernack, G., Gerritsen, W., Huber, C., Parmiani, G., Scholl, S., Thatcher, N., Wagstaff, J., Zielinski, C., Faulkner, I. & Mellstedt, H. Therapeutic vaccines for cancer: an overview of clinical trials. *Nat. Rev. Clin. Oncol.* **11,** 509–524 (2014).
19. Strome, S. E., Sausville, E. A. & Mann, D. A Mechanistic Perspective of Monoclonal Antibodies in Cancer Therapy Beyond Target-Related Effects. *Oncologist* **12,** 1084–1095 (2007).
20. Castillo, J., Winer, E. & Quesenberry, P. Newer monoclonal antibodies for hematological malignancies. *Exp. Hematol.* **36,** 755–768 (2008).
21. Boyiadzis, M. & Foon, K. A. Approved monoclonal antibodies for cancer therapy. *Expert Opin. Biol. Ther.* **8,** 1151–1158 (2008).
22. Baxter, D. Active and passive immunity, vaccine types, excipients and licensing. *Occup. Med. (Chic. Ill).* **57,** 552–556 (2007).
23. Melero, I., Hervas-Stubbs, S., Glennie, M., Pardoll, D. M. & Chen, L. Immunostimulatory monoclonal antibodies for cancer therapy. *Nat. Rev. Cancer* **7,** 95–106 (2007).

24. Hodi, F. S., Mihm, M. C., Soiffer, R. J., Haluska, F. G., Butler, M., Seiden, M. V., Davis, T., Henry-Spires, R., MacRae, S., Willman, A., Padera, R., Jaklitsch, M. T., Shankar, S., Chen, T. C., Korman, A., Allison, J. P. & Dranoff, G. Biologic activity of cytotoxic T lymphocyte-associated antigen 4 antibody blockade in previously vaccinated metastatic melanoma and ovarian carcinoma patients. *Proc. Natl. Acad. Sci.* **100,** 4712–4717 (2003).

25. Baxevanis, C. N., Perez, S. A. & Papamichail, M. Cancer immunotherapy. *Crit. Rev. Clin. Lab. Sci.* **46,** 167–189 (2009).

26. Bosch, F. X., Broker, T. R., Forman, D., Moscicki, A., Gillison, M. L., Doorbar, J., Stern, P. L., Stanley, M., Arbyn, M., Poljak, M., Cuzick, J., Castle, P. E., Schiller, J. T., Markowitz, L. E., Fisher, W. A., Canfell, K., Denny, L. A., … Albero, G. Comprehensive control of human papillomavirus infections and related diseases. *Vaccine* **31,** F1–F31 (2013).

27. Burg, S. H. Van Der & Melief, C. J. M. Therapeutic vaccination against human papilloma virus induced malignancies. *Curr. Opin. Immunol.* **23,** 252–257 (2011).

28. Mellman, I., Coukos, G. & Dranoff, G. Cancer immunotherapy comes of age. *Nature* **480,** 480–489 (2014).

29. Bachmann, M. F. & Jennings, G. T. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat. Rev. Immunol.* **10,** 787–96 (2010).

30. Ward, S., Casey, D., Labarthe, M. C., Whelan, M., Dalgleish, A., Pandha, H. & Todryk, S. Immunotherapeutic potential of whole tumour cells. *Cancer Immunol. Immunother.* **51,** 351–357 (2002).

31. Hoiseth, S. K. & Stocker, B. A. D. Aromatic-dependent Salmonella typhimurium anr non-virulent and effective as live vaccines. *Nature* **291,** 238–239 (1981).

32. Sadoff, J. C., Ballou, W. R., Baron, L. S., Majarian, W. R., Brey, R. N., Hockmeyer, W. T., Young, J. F., Cryz, S. J., Ou, J., Lowell, G. H. & Chulay, J. D. Oral Salmonella typhimurium vaccine expressing circumsporozoite protein protects against malaria. *Science.* **240,** 336–338 (1988).

33. Shen, H. A. O., Slifka, M. K., Matloubian, M., Jensen, E. R., Ahmed, R. & Miller, J. F. Recombinant Listeria monocytogenes as a live vaccine vehicle for the induction of protective anti-viral cell-mediated immunity. *Proc. Natl. Acad. Sci. U S A.* **92,** 3987–3991 (1995).

34. Parmiani, G., Castelli, C., Dalerba, P., Mortarini, R., Rivoltini, L., Marincola, F. M. & Anichini, A. Cancer immunotherapy with peptide-based vaccines: What have we achieved? Where are we going? *J. Natl. Cancer Inst.* **94,** 805–818 (2002).

35. Parmiani, G., Russo, V., Maccalli, C., Parolini, D., Rizzo, N. & Maio, M. Peptide-based vaccines for cancer therapy. *Hum. Vaccin. Immunother.* **10,** 3175–3178 (2014).

36. Rice, J., Ottensmeier, C. H. & Stevenson, F. K. DNA vaccines: Precision tools for activating effective immunity against cancer. *Nat. Rev. Cancer* **8,** 108–120 (2008).

37. Marrack, P., McKee, A. S. & Munks, M. W. Towards an understanding of the adjuvant action of aluminium. *Nat. Rev. Immunol.* **9,** 287–293 (2009).

38. Bomford, R. Aluminium Salts: Perspectives in their use as Adjuvants. in *Immunological Adjuvants and Vaccines* (eds. Gregoriadis, G., Allison, A. C. & Poste, G.) 35–41 (Springer US, 1989).

39. Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annu. Rev. Immunol.* **21,** 335–376 (2003).

40. Li, Q. & Guo, Z. Recent advances in Toll-like receptor-targeting glycoconjugate vaccines. *Molecules* **23,** 1583–1607 (2018).

41. Shi, M., Chen, X., Ye, K., Yao, Y. & Li, Y. Application potential of Toll-like receptors in cancer immunotherapy: Systematic review. *Medicine (Baltimore).* **95,** 1–7 (2016).

42. Rakoff-Nahoum, S. & Medzhitov, R. Toll-Like Receptors and cancer. *Nat. Rev. Cancer* **9,** 57–63 (2009).

43. Cipolla, L., Peri, F. & Airoldi, C. Glycoconjugates in Cancer Therapy. *Anticancer. Agents Med. Chem.* **8,** 92–121 (2008).

44.   Bruno, B. J., Miller, G. D. & Lim, C. S. Basics and recent advances in peptide and protein drug delivery. *Ther. Deliv.* **4,** 1443–1467 (2014).

45.   Schwendener, R. A. Liposomes as vaccine delivery systems: a review of the recent advances. *Ther. Adv. Vaccines* **2,** 159–182 (2014).

46.   Schmidt, S. T., Foged, C., Korsholm, K. S., Rades, T. & Christensen, D. Liposome-based adjuvants for subunit vaccines: Formulation strategies for subunit antigens and immunostimulators. *Pharmaceutics* **8,** 1–22 (2016).

47.   Leleux, J. & Roy, K. Micro and Nanoparticle-Based Delivery Systems for Vaccine Immunotherapy: An Immunological and Materials Perspective. *Adv. Healthc. Mater.* **2,** 72–94 (2013).

48.   Hartmann, S., Nuhn, L., Palitzsch, B., Glaffig, M., Stergiou, N., Gerlitzki, B., Schmitt, E., Kunz, H., Zentel, R., Glaffi, M., Stergiou, N., Gerlitzki, B., Schmitt, E., Kunz, H. & Zentel, R. CpG-Loaded Multifunctional Cationic Nanohydrogel Particles as Self-Adjuvanting Glycopeptide Antitumor Vaccines. *Adv. Healthc. Mater.* **4,** 2192–2659 (2014).

49.   Wang, F., Li, C., Cheng, J. & Yuan, Z. Recent advances on inorganic nanoparticle-based cancer therapeutic agents. *Int. J. Environ. Res. Public Health* **13,** 1–15 (2016).

50.   Hagan, D. T. O. & Valiante, N. M. Recent advances in the discovery and delivery of vaccine adjuvants. *Nat. Rev. Drug Discov.* **2,** 727–735 (2003).

51.   Mire, C. E. & Geisbert, T. W. Viral-vectored vaccines to control pathogenic filoviruses. in *Novel Technologies for Vaccine Development* (eds. Lukashevich, I. S. & Shirwan, H.) 33–61 (Springer US, 2014).

52.   Nagata, K. Mechanism and components of endoplasmic reticulum- associated degradation. *J. Biochem.* **147,** 19–25 (2010).

53.   Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126,** 855–867 (2006).

54.   Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3,** 97–130 (1993).

55.   Fuster, M. M. & Esko, J. D. The sweet and sour of cancer: Glycans an novel therapeutic targets. *Nat. Rev. Cancer* **5,** 526–542 (2005).

56.   Reis, C. A., Osorio, H., Silva, L., Gomes, C. & David, L. Alterations in glycosylation as biomarkers for cancer detection. *J. Clin. Pathol.* **63,** 322–329 (2010).

57.   Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15,** 540–555 (2015).

58.   Springer, G. F. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *J. Mol. Med.* **75,** 594–602 (1997).

59.   Kim, Y. J. & Varki, A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconj. J.* **14,** 569–576 (1997).

60.   Hakomori, S.-I. Cancer-associated glycosphingolipid antigens: Their structure, organization, and function. *Acta Anat. (Basel).* **161,** 79–90 (1998).

61.   Brooks, S. A., Carter, T. M., Royle, L., Harvey, D. J., Fry, S. A., Kinch, C., Dwek, R. A. & Rudd, P. M. Altered glycosylation of proteins in cancer: What is the potential for new anti-tumour strategies. *Anticancer. Agents Med. Chem.* **44,** 2–21 (2008).

62.   Shimizu, M. & Yamauchi, K. Isolation and characterization of Mucin-like glycoprotein in human milk fat globule membrane. *J. Biochem.* **91,** 515–524 (1982).

63.   Feng, D., Shaikh, A. S. & Wang, F. Recent advance in Tumor-Associated Carbohydrate Antigens (TACAs)-based antitumor vaccines. *ACS Chem. Biol.* **11,** 850–863 (2016).

64.   McDonald, D. M., Byrne, S. N. & Payne, R. J. Synthetic self-adjuvanting glycopeptide cancer vaccines. *Front. Chem.* **3,** 1–8 (2015).

65.   Koganty, R. R., Yalamati, D. & Jiang, Z.-H. Glycopeptide-based cancer vaccines: The role of synthesis and structural definition. in *Carbohydrate-Based Vaccines* (ed. Roy, R.) 311–334 (American Chemical Society, 2008).

66.   Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity,

synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13,** 448–462 (2012).

67. Schachter, H. The joys of HexNAc. The synthesis and function of N- and O- glycan branches. *Glycoconj. J.* **17,** 465–483 (2000).

68. Clausen, H. & Bennett, E. P. A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology* **6,** 635–646 (1996).

69. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A. & Fritz, T. A. Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22,** 736–756 (2012).

70. Steentoft, C., Vakhrushev, Y., Joshi, H. J., Kong, Y., Vester-christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B. & Clausen, H. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32,** 1478–1488 (2013).

71. Dalziel, M., Whitehouse, C., Mcfarlane, I., Brockhausen, I., Gschmeissner, S., Schwientek, T., Clausen, H., Burchell, J. M. & Taylor-papadimitriou, J. The relative activities of the C2GnT1 and ST3Gal-I Glycosyltransferases determine O-glycan structure and expression of a tumor-associated epitope on MUC1. *J. Biol. Chem.* **276,** 11007–11015 (2001).

72. Sewell, R., Ba, M., Dalziel, M., Gschmeissner, S., Karlsson, H., Noll, T., Clausen, H., Hansson, G. C., Burchell, J. & Taylor-papadimitriou, J. The ST6GalNAc-I sialyltransferase localizes throughout the Golgi and is responsible for the synthesis of the tumor-associated Sialyl-Tn O-glycan in human breast cancer. *J. Biol. Chem.* **281,** 3586–3594 (2006).

73. Duarte, H. O., Freitas, D., Gomes, C., Gomes, J., Magalhães, A. & Reis, C. A. Mucin-type O-glycosylation in gastric carcinogenesis. *Biomolecules* **6,** 1–19 (2016).

74. Pinho, S., Marcos, N. T., Ferreira, B., Carvalho, A. S., Oliveira, M. J., Santos-silva, F. & Harduin-lepers, A. Biological significance of cancer-associated sialyl-Tn antigen: Modulation of malignant phenotype in gastric carcinoma cells. *Cancer Lett.* **249,** 157–170 (2007).

75. Miles, D. W., Linehan, J., Smith, P. & Filipe, I. Expression of sialyl-Tn in gastric cancer: Correlation with known prognostic factors. *Br. J. Cancer* **71,** 1074–1076 (1995).

76. Dall'Olio, F., Malagolini, N., Trinchera, M. & Chiricolo, M. Mechanisms of cancer-associated glycosylation changes. *Front. Biosci.* **17,** 670–699 (2012).

77. Itzkowitz, S., Kjeldsen, T., Friera, A., Hakomori, S.-I., Yang, U.-S. & Kim, Y. S. Expression of Tn, Sialosyl Tn, and T antigens in human pancreas. *Gastroenterology* **100,** 1691–1700 (1991).

78. Radhakrishnan, P., Dabelsteen, S., Brus, F., Francavilla, C., Kopp, K. L., Yin, G., Chen, L., Song, H., Bak, M., Hlady, R. A., Peters, S. L. & Opavsky, R. Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc Natl Acad Sci U S A.* **111,** E4066–E4075 (2014).

79. Ferreira, J. A., Videira, P. A., Lima, L., Pereira, S., Silva, M., Amado, F. & Santos, L. L. Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours Jos. *Mol. Oncol.* **7,** 719–731 (2013).

80. Ricardo, S., Marcos-silva, L., Pereira, D., Pinto, R., Mandel, U., Clausen, H., Almeida, R., Ola, S., Felix, A., Lunet, N. & David, L. Detection of glyco-mucin profiles improves specificity of MUC16 and MUC1 biomarkers in ovarian serous tumours. *Mol. Oncol.* **9,** 503–512 (2015).

81. Forstner, J. F. Intestinal mucins in health and disease. *Digestion* **17,** 234–263 (1978).

82. Corfield, A. P. Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochim. Biophys. Acta* **1850,** 236–252 (2015).

83. Hollingsworth, M. A. & Swanson, B. J. Mucins in cancer: Protection and control of the cell surface. *Nat. Rev. Cancer* **4,** 45–60 (2004).

84. Hattrup, C. L. & Gendler, S. J. Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol.* **70,** 431–457 (2008).

85. Senapati, S., Das, S. & Batra, S. K. Mucin-interacting proteins: from function to therapeutics. *Trends Biochem. Sci.* **35,** 236–245 (2009).

86. Inghirami, G., Abe, M., Justi-Wheeler, H. & Schlom, J. Reactivity of a novel monoclonal antibody (DF3) with human malignant versus benign breast tumors differential. *Hybridoma* **3,** 223–232 (1984).

87. Hayes, D. F., Sekine, H., Ohno, T., Abe, M., Keefe, K. & Kufe, D. W. Use of a Murine monoclonal antibody for detection of circulating plasma DF3 antigen levels in breast cancer patients. *J. Clin. Invest.* **75,** 1671–1678 (1985).

88. Abe, M. & Kufe, D. Structural analysis of the DF3 human breast carcinoma-associated protein. *Cancer Res.* **49,** 2834–2839 (1989).

89. Kufe, D. W. Targeting the human MUC1 oncoprotein : A tale of two proteins Targeting the human MUC1 oncoprotein. *Cancer Biol. Ther.* **7,** 81–84 (2008).

90. Martínez-Sáez, N., Peregrina, J. M. & Corzana, F. Principles of mucin structure: implications for the rational design of cancer vaccines derived from MUC1-glycopeptides. *Chem. Soc. Rev.* **46,** 7154–7175 (2017).

91. Hang, H. C. & Bertozzi, C. R. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg. Med. Chem.* **13,** 5021–5034 (2005).

92. Tarp, M. A. & Clausen, H. Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochim. Biophys. Acta* **1780,** 546–563 (2008).

93. Taylor-Papadimitriou, J., Burchell, J., Miles, D. W. & Dalziel, M. MUC1 and cancer. *Biochim. Biophys. Acta* **1455,** 301–313 (1999).

94. Röttger, S., White, J., Wandall, H. H., Olivo, J., Stark, A., Bennett, E. P., Whitehouse, C., Berger, E. G., Clausen, H. & Nilsson, T. Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus. *J. Cell Sci.* **111,** 45–60 (1998).

95. Nath, S. & Mukherjee, P. MUC1: A multifaceted oncoprotein with a key role in cancer progression. *Trends Mol. Med.* **20,** 332–342 (2014).

96. Adamo, R., Nilo, A., Castagner, B., Boutureira, O., Berti, F. & De, G. J. L. B. Synthetically defined glycoprotein vaccines: current status and future directions. *Chem. Sci.* **4,** 2995–3008 (2013).

97. Pol, J., Bloy, N., Buqué, A., Eggermont, A., Cremer, I., Sautès-Fridman, C., Galon, J., Tartour, E., Zitvogel, L., Kroemer, G. & Galluzzi, L. Trial Watch: Peptide-based anticancer vaccines. *Oncoimmunology* **4,** e974411-1-e974411-12 (2015).

98. Buskas, T., Thompson, P. & Boons, G.-J. Immunotherapy for cancer: Synthetic carbohydrate-based vaccines. *Chem. Commun.* 5335–5349 (2009).

99. Astronomo, R. D. & Burton, D. R. Carbohydrate vaccines: Developing sweet solutions to sticky situations? *Nat. Rev. Drug Discov.* **9,** 308–324 (2010).

100. Berti, F. & Adamo, R. Recent mechanistic insights on glycoconjugate vaccines and future perspectives. *ACS Chem. Biol.* **8,** 1653–1663 (2013).

101. Wilson, R. M. & Danishefsky, S. J. A vision for vaccines built from fully synthetic tumor-associated antigens: From the laboratory to the clinic. *J. Am. Chem. Soc.* **135,** 14462–14472 (2013).

Objectives

2

Considering the relevance of MUC1 mucin and the Tn antigens in cancer, the primary goal of the present dissertation is to study in deep both their structure in solution and in the bound state with some anti-MUC1 antibodies. This research will help us to rationally adjust the glycopeptide backbone, as well as the sugar moiety to improve the binding affinity against these antibodies and to synthesize subrogates more resistant to enzymatic degradation. These two features are desirable for the construction of either cancer vaccines or diagnostic tools, such as cancer biomarkers.

Within this context, the study of the conformational preferences of the Tn antigens (α-*O*-GalNAc-Ser and α-*O*-GalNAc-Thr) will be performed, updating the studies conducted by our group more than 10 years ago. To confirm the active role of water molecules in the conformational preferences of these important molecules unambiguously, an unprecedented experimental and theoretically study of these derivatives in the gas phase will be conducted.

The role in molecular recognition antigen-antibody of the β-methyl group of threonine residue located in the most immunogenic fragment of MUC1 will be analyzed. To this purpose, novel glycopeptides incorporating either the epimer of Thr at the Cβ, *allo*-threonine (*allo*-Thr), or (*2S,3R*)-hydroxynorvaline (Hnv), which displays an ethyl-by-methyl substitution at the position mentioned above, will be exanimated in the free state and in complex to two anti-MUC1 antibodies. Additionally, the relevance of the hydroxymethyl group of the GalNAc unit for the binding will be investigated by replacing this sugar by the corresponding iminosugar. A vaccine based on an unnatural glycopeptide with this surrogate will be designed and tested in mice.

The strength of a proline residue at the most immunogenic fragment of MUC1 will be tuned by introducing fluorine atoms to this amino acid. These innovative derivatives will be tested as biosensors to detect antibodies in true samples of patients with prostate cancer.

The synthesis of the STn antigens, both with serine and threonine, in a gram scale and ready to use in solid-phase peptide synthesis will be carried out. To this purpose an international short-stay in the Dr. Westerlind's lab will be done.

Different shape of Tn antigen: GalNAc-Ser vs. GalNAc-Thr

3

3.1. Introduction

3.1.1. The Tn antigen

As mentioned in the Introduction section, Tumor-Associated Cancer Antigens, TACA allow the differentiation among normal and cancer cells. One of the most studied of these antigens is the so-called Tn antigen[1–3] (α-*O*-GalNAc-Ser/Thr) shown in Figure 3.1. It consists of a *N*-acetylgalactosamine (GalNAc) unit linked to a serine (Ser) or threonine (Thr) residue through a α-*O*-linkage.
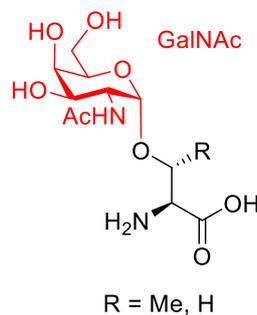

R = Me, H

Figure 3.1 Tn antigen structure.

The discovery of the Tn antigen is not associated to cancer disease, but to a hematological disorder: Tn polyagglutination syndrome.[4] This illness was first reported more than fifty years ago[5,6] and it is characterized by the tendency of red blood cells to autoagglutinate due to the exposure of Tn antigen on the surface of a portion of blood cells.[7] Since a similar phenomenon was observed previously with T antigen,[8] this new molecule received the name of T *nouvelle* – new – antigen, Tn antigen. In the cluster differentiation nomenclature is designated CD175.[9]

The Tn antigen structure was described in 1975 using coelution in gas-liquid chromatography to identify the carbohydrate as GalNAc. The presence of GalNAc-Ser/Thr was demonstrated on the carbohydrate moiety of tryptic erythrocyte glycopeptides from people with Tn syndrome.[10]

At the same time, Springer *et al* observed for the first time the presence of unmasked Tn antigen in the cell membranes of breast cancer, but not in healthy tissues or benign tumors.[11] They postulated later the presence of this antigen on about 90% of the most common human malignant tumors, including colon, breast, lung and pancreas; and correlated the density of Tn receptors with carcinoma aggressiveness and tumor invasiveness.[12] Tn antigen was later related to bladder, cervix, ovary, prostate and stomach cancer.[13,14]

Regarding to GalNAc-Ser/Thr synthesis, many approaches appear in bibliography concerning *O*-glycosylation reactions. In some examples mercuric salts are employed as coupling agent, including Helferich[15,16] or Kaifu-Osawa[17] reactions (Scheme 3.1).
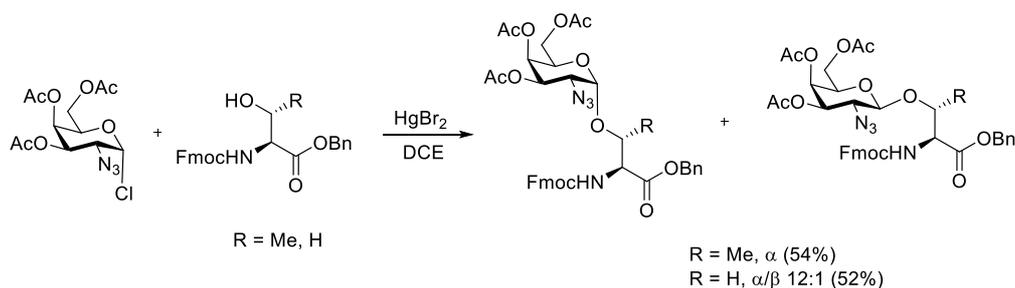
Scheme 3.1 Helferich methodology for O-glycosylation and stereoselectivity rates.[18]

Additionally, the Koenigs-Knorr reaction has been widely employed as a glycosylation methodology (Scheme 3.2 upper panel).[19] It consists in the use of carbohydrate halides, commonly bromine[20] or chlorine[20,21], as glycosyl donors, and the hydroxyl group of the Ser or Thr as glycosyl acceptor. Silver carbonate and silver perchlorate serve as promoters to form the α anomer stereoselectively.[22]
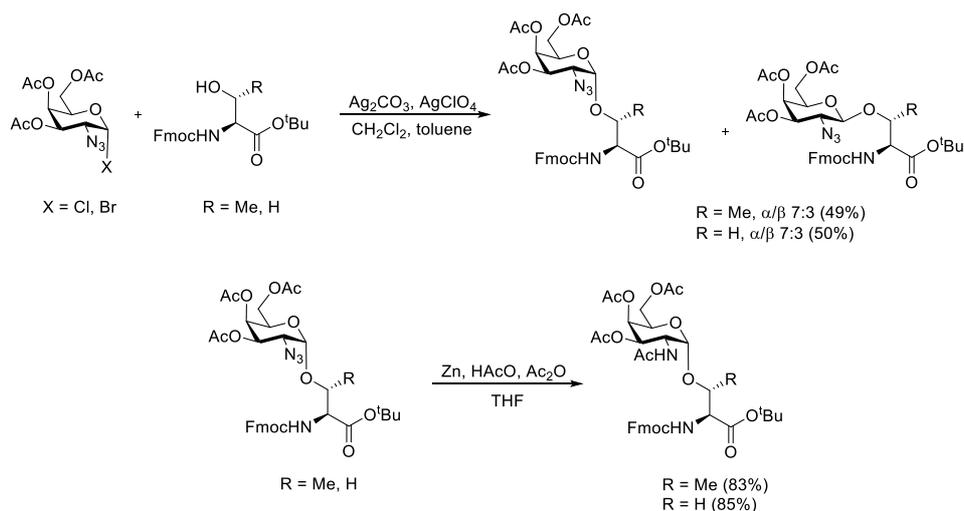
Scheme 3.2 Koenigs-Knorr reaction and stereoselectivity rates observed by our group,[23,24] upper panel, and subsequent reduction of the azido group, bottom panel.

The presence of an azido group at C2 of the glycosyl donor is also crucial, since it promotes a selective α-glycosylation reaction. This group is reduced to primary amine and acetylated in a subsequent one-pot reaction, yielding the GalNAc moiety. (Scheme 3.2 bottom panel). Variations to this methodology also appear in bibliography.[25–27]

Different leaving groups on the azidosugar donor have been reported, such as trichloroacetimidate[28,29] or thiophenyl,[30] and many different chemical glycosylation strategies have been conducted to date.[25,26,28–36] Koenigs-Knorr methodology has been applied in this thesis to obtain Tn antigen in a similar fashion, as previously reported by our group.[23,24]

One relatively recent strategy, developed by Schmidt and co-workers, is based on a Michael addition reaction.[37] (Scheme 3.3)
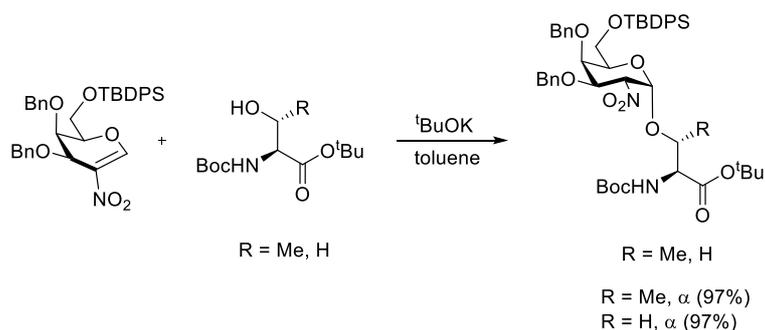
Scheme 3.3 Schmidt methodology for O-glycosylation and stereoselectivity rates.[38]

Traditionally, the Tn antigen is referred indistinctively to GalNAc α-*O*-linked to either serine or threonine, without any specification to the amino acid. Nevertheless, biological targets have shown differences on the recognition of Tn antigen. For example, in glycopeptides carrying GalNAc-Ser or GalNAc-Thr,[39] their recognition varies when interact with cancer antibodies,[40–43] lectins[44,45] or in antimicrobial peptides.[46] This distinction is also related to some illnesses, for instance Kanzaki disease, associated to an enzyme that hydrolyzes differently the α-*O*-GalNAc glycosidic linkage when bound either to serine or threonine,[47,48] and structural changes of protein tau in Alzheimer's disease.[49] In the context of antifreeze glycoproteins, whose tandem repeat is (Thr(Galβ(1,3)-GalNAc)-Ala-Ala)$_n$, the replacement of threonine by serine involves the loss of the antifreeze activity.[50]

Considering the aforementioned examples, serine and threonine amino acids within Tn antigen appear to be nonequivalent. To shed some light on this issue, α-*O*-GalNAc-Ser/Thr conformation, previously studied in our group,[51–53] has been revised and updated with new techniques in the present chapter.

Prior to a detailed description of Tn antigen structure, some explanations regarding to conformational analysis in solution must be provided.

3.1.2. Introduction to conformational analysis in solution

In general, two main techniques are combined to obtain the conformational preferences of a given molecule in solution: nuclear magnetic resonance (NMR), particularly nuclear Overhauser effect spectroscopy (NOESY), coupled with modeling, and particularly molecular dynamics (MD) simulations.

NOESY is a 2D NMR methodology that correlates protons that are spatially close to each other (< 5 Å). The magnitude of the NOE depends on the distance separation of the interacting spins (r) and its motional properties (τ_C):[54]

$$NOE \sim r^{-6} \cdot f(\tau_C)$$

The study of the NOESY cross-peaks in a molecule is very useful to experimentally determine those H-H distances that are relevant for the conformational analysis. Within this context, ROESY (Rotating frame nuclear Overhauser effect spectroscopy) experiments are employed in the present dissertation alternatively to NOESY, since the rotational correlation time of molecules analyzed herein appears in a range where NOE signals are close to zero due to their molecular weight. ROESY experiments have a different dependence between the correlation time and cross-relaxation rate constant. Indeed, whereas in NOESY these properties go from positive to negative values, in ROESY the cross-relaxation rate constant is always positive.

Vicinal coupling constants are also relevant for the conformational analysis, as they provide valuable stereochemical information via Karplus equation and related equations.[55] In particular, ^3J coupling constants are widely used in the conformational analysis of (glyco)peptides.

Concerning the dihedral angles required to define the conformational behavior of a peptide, the peptide backbone is described, as shown in Figure 3.2, by ψ_P and ϕ_P torsional angles, which define the rotation around $C\alpha$-C_1 and $C\alpha$-N_2 bonds, respectively. Additionally, the rotation of the side chain is described by χ^1.
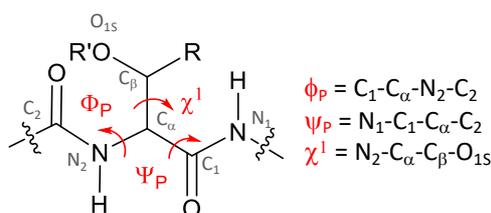

$$\phi_P = C_1\text{-}C_\alpha\text{-}N_2\text{-}C_2$$
$$\psi_P = N_1\text{-}C_1\text{-}C_\alpha\text{-}C_2$$
$$\chi^1 = N_2\text{-}C_\alpha\text{-}C_\beta\text{-}O_{1S}$$

Figure 3.2 Dihedral angles in structural analysis of proteins.

When working with glycopeptides, two additional angles must be considered: ψ_S and ϕ_S, which define the rotation around $C\beta$–O_{1S} and O_{1S}–C_{1S}, respectively (Figure 3.3).[56]

$$\phi_S = O_{5S}\text{-}O_{1S}\text{-}O_{1S}\text{-}C_\beta$$
$$\psi_S = C_{1S}\text{-}O_{1S}\text{-}C_\beta\text{-}C_\alpha$$

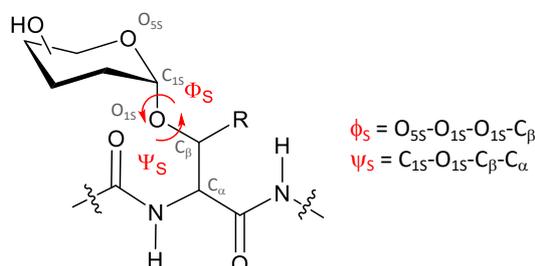*Figure 3.3 Significant angles in glycosidic linkage conformational analysis.*

Although many combinations of values could be ascribed to ψ and ϕ angles, only some of them are energetically stable secondary structures, for instance, α-helix and β-sheet. Both dihedral angles, ψ_P and ϕ_P, of each residue may also be represented graphically, in

the so-called Ramachandran plots,[57] where energetically allowed regions of amino acids in protein structure are shown.

On the other hand, MD simulations study the dynamic properties of molecules in a theoretical way, providing structural, dynamical and thermodynamical information at atomic level. In this type of calculations, atoms are considered as rigid spheres with a fixed partial charge and chemical bonds as springs, so bond order is associated to an elastic constant.

The force field chosen to perform the simulations is a crucial step. It is composed by mathematical equations that describe the potential energy (V) of the given molecule based on atomic coordinates and a set of parameters (Figure 3.4), which are obtained either from experimental data (X-ray, NMR, IR…) or Quantum Mechanical (QM) calculations. From the different force fields available on literature, in this thesis AMBER and GLYCAM force fields are employed. The former is needed to perform the conformational analysis of peptides and proteins and the latter is required to accurately reproduce the dynamics and conformational behavior of carbohydrates or glycoproteins. In addition, since MD experiments are conducted for molecules in solution (water), interactions with solvent must be also considered. There have been developed two models to represent the solvent. The simplest one, which considers the solvent as just a dielectric constant (implicit model), or a more sophisticated one that includes real molecules of solvent (explicit model). In this latter case, several models to describe water molecule have been developed. One of the most studied and popular is TIP3P water model,[58,59] which will be used along this thesis.

$$\text{Potential energy (V)} = E_{\text{bonding terms}} + E_{\text{non-bonding terms}}$$

$$\text{Potential energy (V)} = \sum_{bonds} K_{bond}\,(r - r_0)^2 \qquad V_{\text{bonds}}$$

$$+ \sum_{angles} K_{angle}\,(\theta - \theta_0)^2 \qquad V_{\text{angles}}$$

$$+ \sum_{dihedral} \frac{E'}{2}\,(1 + \cos n\varphi - \gamma) \qquad V_{\text{dihedral}}$$

$$+ \sum_{m,n} \frac{q_m\,q_n}{\varepsilon\,(r_{mn})r_{mn}} \qquad V_{\text{electrostatic}}$$

$$+ \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \qquad V_{\text{Van der Waals}}$$

Figure 3.4 Force field general equations.

In general, to obtain more accurate MD results, particularly when unnatural residues are studied, NMR experimental data are introduced as restraints. Thus, proton-proton distances relevant for the conformational analysis, obtained from the NOE experiments, together with $^3J_{\text{H-H}}$ coupling constants, are used as experimental restraints. It is

important to notice that these experimental data are averaged over all the possible conformations present in solution for a given molecule. Consequently, the use of experiment-guided MD simulations,[60] that is, MD simulations with time-averaged restraints are mandatory to study flexible molecules in solution employing NMR data. This type of calculations renders an experimental ensemble of low energy conformers (at a given temperature) able to reproduce all experimental data.

3.1.3. Previous conformational studies of Tn antigen in solution

Conformational analyses on the Tn antigen in solution performed previously by our group[51] using MD-tar calculations showed remarkable differences between α-*O*-GalNAc-Ser and α-*O*-GalNAc-Thr entities (Figure 3.5).
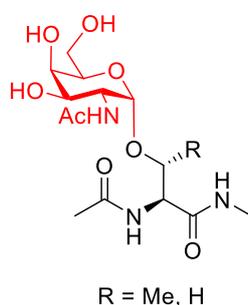

Figure 3.5 Ser and Thr diamide derivatives synthetized.

Regarding to the distribution of the glycosidic linkage, ϕ_S adopts a value around 80° for the two derivatives, which is quite rigid due to the exo-anomeric effect.[61,62] On the contrary, ψ_S torsional angles are noticeably different for Tn antigen derivatives. In fact, while in α-*O*-GalNAc-Thr, ψ_S is rather rigid, presenting a value of 120°, where Cβ-O$_{1S}$ bond presents an eclipsed conformation, in GalNAc-Ser this torsional angle is more flexible, with values close to 180°. In this case, the typical staggered conformation for the glycosidic linkage is observed. This conclusion concurs with the finding that in Ser derivative $^3J_{H\alpha,H\beta}$ value is larger than in glycopeptide derived from Thr, suggesting that rotation of the side-chain in the latter, defined by χ^1, is restricted to a value close to 60°. In Tn-Ser, this torsional angle takes the three possible staggered values, being *gg* conformer the main populated.

As a result, the orientation of the sugar moiety is considerably distinct for both molecules. While in Ser compound the carbohydrate adopts a parallel disposition; in the Thr molecule the GalNAc unit is almost perpendicular to the peptide backbone, allowing the GalNAc group to be located closer to the peptide backbone (Figure 3.6).
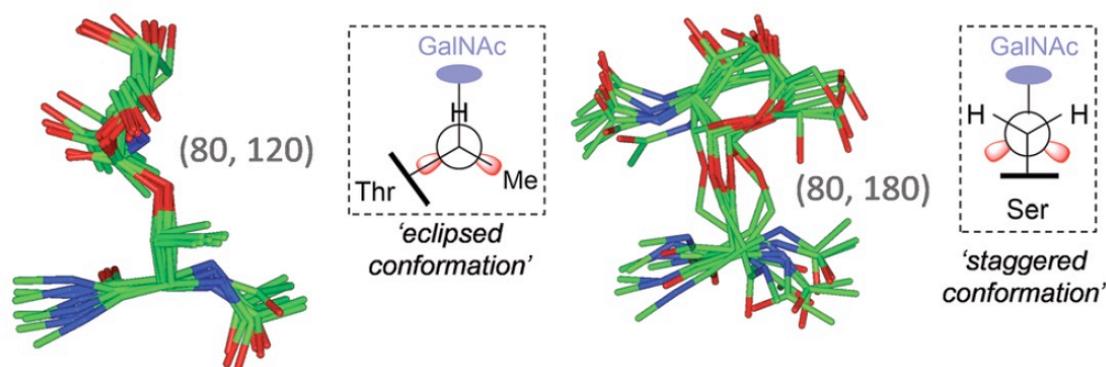
Figure 3.6 Main conformation in solution of Tn-Thr (left) and Tn-Ser (right), as well as a Newman projection for the C_β-O_{1S} bond for both derivatives.

For a better understanding, these computational results appear summarized on Table 3.1.

Table 3.1 Most relevant results derived from MD-tar simulations on Tn antigen.

	χ^1	Φ_S	Ψ_S	Conformation C_β-O_{1S}	Carbohydrate orientation
α-*O*-GalNAc-Ser	60°	~80°	180°	Staggered	Parallel
α-*O*-GalNAc-Thr	~60°	80°	120°	Eclipsed	Perpendicular

In general, α-*O*-glycosylation with GalNAc forces an extended conformation of the peptide backbone. This feature has been attributed to steric effects that GalNAc unit exerts on the peptide backbone[35,63] or to specific hydrogen bonds between the carbohydrate and the underlying amino acids, which fix the sugar orientation over the peptide chain,[64] forcing it into an extended conformation. Within this context, our group has proposed the presence of 'water pockets' connecting the sugar and peptide moieties in the Tn antigens.[51,52,65] In Tn-Ser, a water bridge has been proposed connecting the NH of GalNAc to the carbonyl group of the glycosylated amino acid. In contrast, in the Thr derivative, the water-bridging molecule appears between the NH groups of the sugar and the amino acid residue (Figure 3.7). These observations deduced from MD-tar simulations are, at least partially, experimentally supported by the presence of a medium-size NOE cross-peak between NH group of Thr and NH proton of the GalNAc moiety, corroborating also the eclipsed conformation above mentioned. This NOE contact does not appear in the Ser derivative, showing the different behavior of the two Tn antigens in solution.
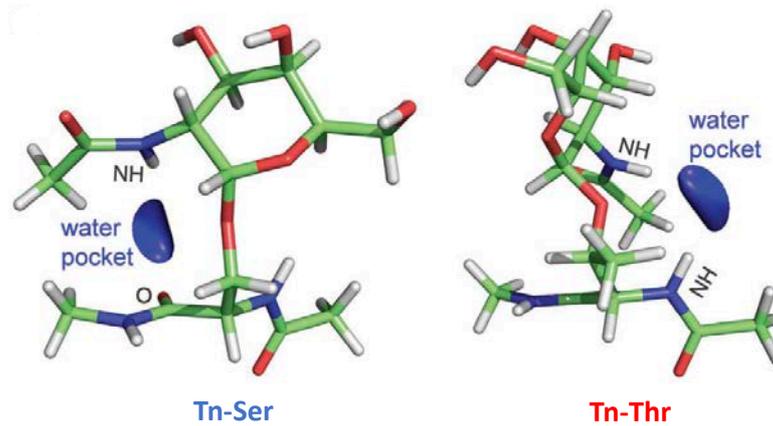
Figure 3.7 Different accommodation of the water pocket in Tn antigen derivatives.

A powerful tool to experimentally demonstrate the presence of the aforementioned water pocket in GalNAc-Thr glycopeptides would be X-ray crystallography. Actually, crystal structures of MUC1-like glycopeptides bound to anti-MUC1 antibodies present this hydrophilic pocket,[41,66] although the water molecule has never been captured.

Interestingly, several X-ray structures evidence that the water bridging molecule might be substituted by an oxygen atom, either within the backbone structure of the ligand or the receptor, binding NH groups of sugar and Thr residue. For instance, the antibody 237 in complex with its glycopeptide epitope (pdb ID: 3IET)[66] presents this kind of linkage with Thr92 of the antibody. Also PDT(α-*O*-GalNAc)R epitope (pdb ID: 4D69)[67] when bound to Soybean agglutinin lectin is folded so that carbonyl oxygen of Asp within the own ligand bridges GalNAc and Thr NH groups. The same kind of linkage has been observed in GalNAc-Transferase 2 in complex with MUC5AC-13 (pdb ID: 5AJP) and MUC5AC-3-13 (pdb ID: 5AJO). The former involves a Ser residue from the ligand as the bridge and the later forms this interaction with Glu334[68] (Figure 3.8).
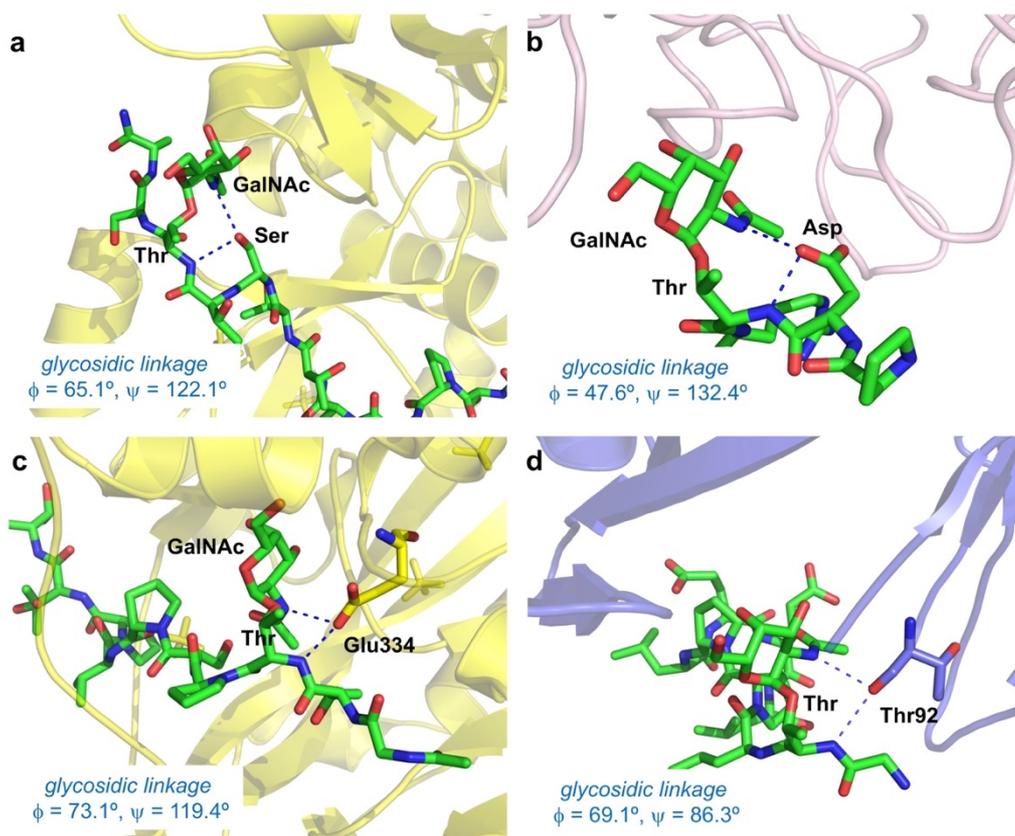
Figure 3.8 Zoom-in of the crystal structures of GalNAc-Transferase 2 in complex with UDP and MUC5AC-13 (a) and MUC5AC-3-13 (c), Soybean agglutinin with the glycopeptide PDT(α-O-GalNAc)R (b) and antibody 237 with its glycopeptide epitope (d).

The presence of a bridging oxygen atom in a specific region of the binding pocket leads to the idea of the importance of this particular non-bonding interaction, where the water molecule, if not an oxygen-assistant molecule, apparently stabilizes the bound state and helps the antigen to display the bioactive conformation, hence facilitating molecular recognition.

Despite our structural observations regarding Tn antigen structure have been endorsed by later investigations,[65,69] a direct experimental demonstration of the aforementioned water pocket has not been provided yet.

3.2. Objectives

α-*O*-GalNAc-Ser and α-*O*-GalNAc-Thr moieties are classically designated as Tn antigen, without any distinction. Nevertheless, as previously reported by our group,[51] they both present different conformation in solution, proved by MD-tar simulations and NMR experiments. Our goal is to provide experimental support for this observation, as well as recheck our theoretical findings and contribute with new ones.

We firstly synthesized Tn antigen-like derivatives incorporating a benzylamide group, required for the conformational analysis in the gas phase (see below).

These compounds were then analyzed in solution by NMR experiments and MD-tar simulations, observing that they behave similarly to corresponding methylamide variants.

They were then studied in the gas phase, by means of laser spectroscopy and quantum chemical calculations.

Additionally, fluorinated Tn antigen derivatives were synthesized and incorporated to the MUC1 APDTRP epitope. Fluorinated glycopeptides were crystallographically studied, complexed to SM3 antibody, trapping, for the first time, the proposed bridging water molecules in Tn-Thr analogs.

3.3. Tn antigen benzylamide derivatives

For the purpose of studying Tn antigen derivatives in the gas phase, avoiding any solvent interference, compounds described in the following section were prepared. These were synthesized as benzylamide derivatives to facilitate their detection through mass-selected ultraviolet photoionization.

3.3.1. Synthesis

Tn antigen derivatives, denoted as **Tn-Thr'** and **Tn-Ser'**, were synthesized as described in Scheme 3.4, using Fmoc-(α-O-GalNAc)-Ser/Thr-OH building-blocks, whose syntheses have been previously described in our group.[23,24] See also *Experimental section 8.9.1*.
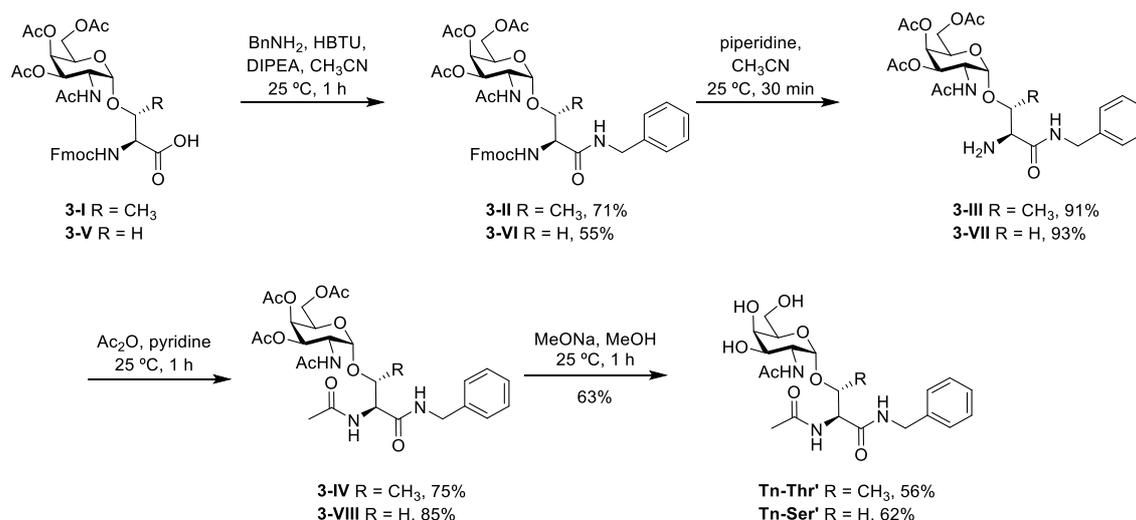
*Scheme 3.4 Synthetic route followed to obtain **Tn-Ser'** and **Tn-Thr'** derivatives.*

Compounds **3-II** and **3-VI** were synthesized by treatment of derivatives **3-I** and **3-V** with benzylamine in presence of the base DIPEA and the coupling agent HBTU. After

purification by column chromatography, Fmoc was removed by adding a solution of piperidine in CH_3CN to give products **3-III** and **3-VII** in good yield. The amine group was then acetylated using a solution of Ac_2O in pyridine to obtain compounds **3-IV** and **3-VIII** in relatively good yields after purification. Finally, acetyl groups of the carbohydrate were removed using MeOH/MeONa and adjusting pH value at 9.5. The treatment with a Dowex resin allowed to obtain the pure compounds **Tn-Ser'** and **Tn-Thr'** without further purification.

3.3.2. Conformational analysis in solution

3.3.2.1. NMR studies

Tn-Thr' and **Tn-Ser'** structures were studied in solution by NMR. Apart from the usual ^{1}H, ^{13}C, COSY and HSQC experiments (See *Supplementary information 9.1.*), 2D-ROESY studies were conducted. Amide region of these compounds is shown in Figure 3.9 and Figure 3.10, respectively. As can be observed, both compounds present low-size cross-peaks between benzylic amide (NHBn) and the NH of Ser or Thr. A medium size cross-peak was detected between NHBn and Hα of the amino acids. Also, between Ser or Thr amide H and Hα. Finally, within the carbohydrate, ROE contacts have been observed among NH of GalNAc and H1 and H3.
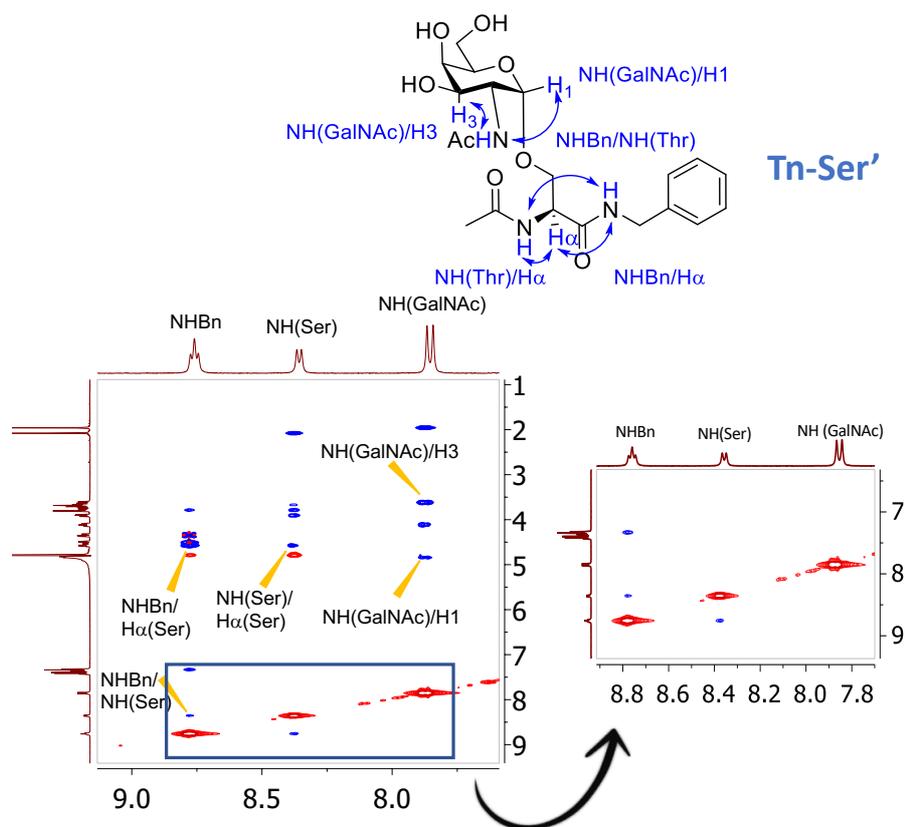

*Figure 3.9 ROESY spectrum of **Tn-Ser'** together with its structure and ROE interactions.*

In the case of **Tn-Thr'** derivative, an additional ROE signal appears between the NH of the threonine and the NH of GalNAc (highlighted in Figure 3.10). This signal is representative of the eclipsed conformation of the glycosidic linkage in Tn-Thr antigen.
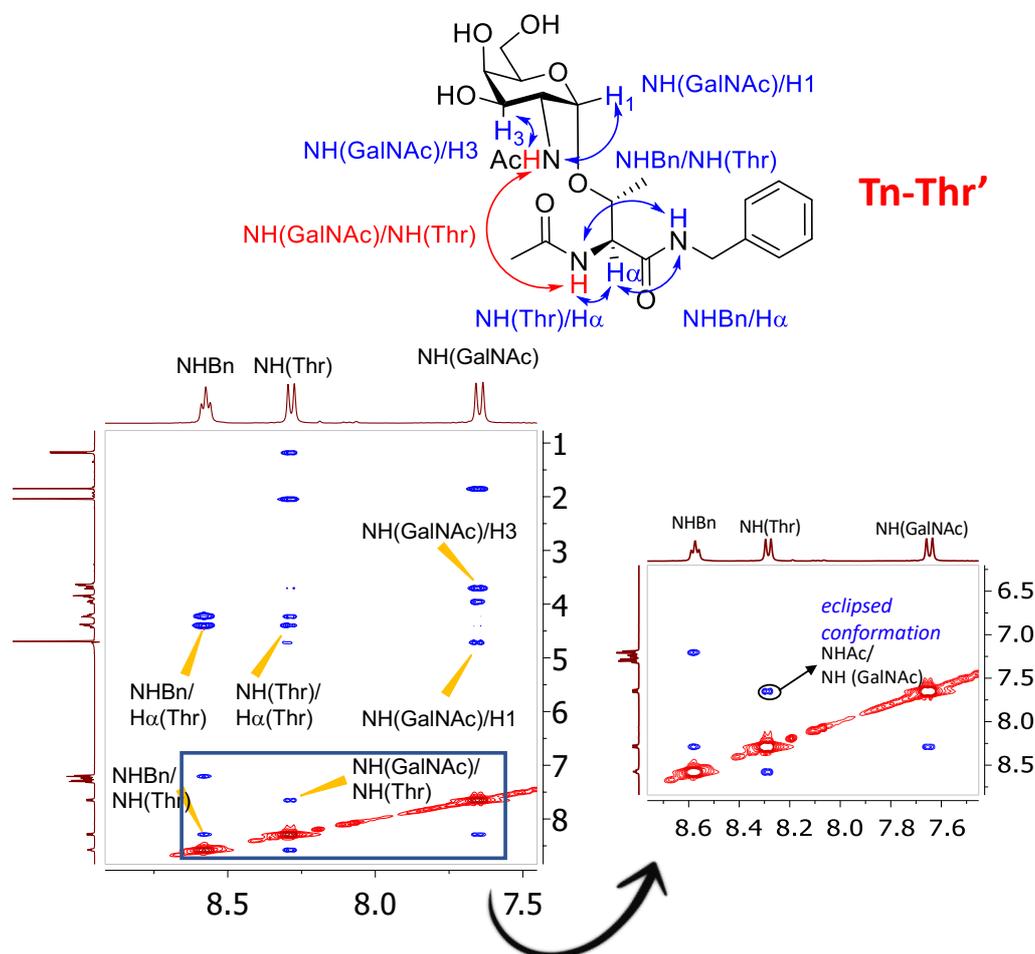

*Figure 3.10 ROESY spectrum of **Tn-Thr'** together with its structure and ROE interactions.*

It also should be noted that same cross-peaks were previously observed in the Tn-derivatives with methylamide.[51]

From these ROESY experiments it is possible to obtain H-H distances that are relevant for the conformational analysis. These data were used as restraints in MD-tar analysis described in the next section.

3.3.2.2.   MD-tar simulations

As presented in Table 3.2, there is a good agreement between the experimental and theoretically derived distances, validating the MD-tar calculations.

*Table 3.2 Comparison of experimental and simulated distances for **Tn-Ser'** and **Tn-Thr'** antigens.*

	Distances	Experimental values (Å)	Experiment-guided MD simulations (Å)
Tn-Ser'	NH(Ser)-Hα(Ser)	3.1	3.0
	NHBn-Hα(Ser)	2.4	2.3
	NH(Ser)-NHBn	> 3.55	4.3
	NH(Ser)-NH(GalNAc)	> 3.5	3.8
Tn-Thr'	NH(Thr)-Hα(Thr)	2.9	3.0
	NHBn-Hα(Thr)	2.3	2.3
	NH(Thr)-NHBn	3.1	3.1
	NH(Thr)-NH(GalNAc)	> 3.5	4.4

The relevant ϕ_S, ψ_S and χ^1 torsional angles were analyzed (Figure 3.11).
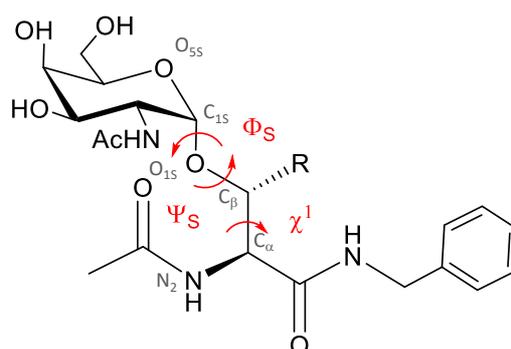
Tn-Thr' R = CH$_3$
Tn-Ser' R = H

*Figure 3.11 Representation of ϕ_S and ψ_S torsional angles in **Tn-Ser'** and **Tn-Thr'** derivatives.*

The ϕ_S/ψ_S distribution of the glycosidic linkage in solution for both compounds was studied. As in previous investigations,[51] ϕ_S adopts a value around 60º-80º for the two derivatives; whereas ψ_S value for the Thr derivative is 120º and for the Ser compound presents a value *ca.* 180º. We can conclude that **Tn-Thr'** presents eclipsed conformation, while **Tn-Ser'** shows a staggered conformation for the glycosidic linkage (Figure 3.12).
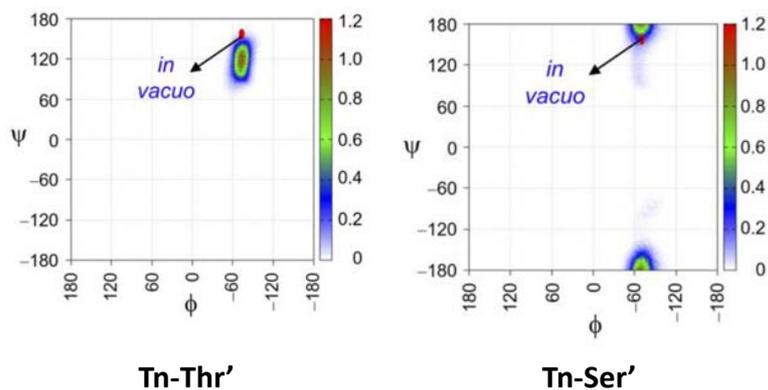
Tn-Thr'   **Tn-Ser'**

*Figure 3.12 ϕ_s and ψ_s torsional angles for the glycosidic linkage in **Tn-Ser'** and **Tn-Thr'** derivatives. The main conformer populated observed in gas phase is emphasized as a red dot*

Regarding to χ^1 torsional angle, the side-chain distribution for the Thr derivative presents only a value of 60º. Conversely, in the case of the more flexible **Tn-Ser'** antigen, χ^1 may adopt three different values, 60º, -60º and, the less populated, 180º (Figure 3.13).
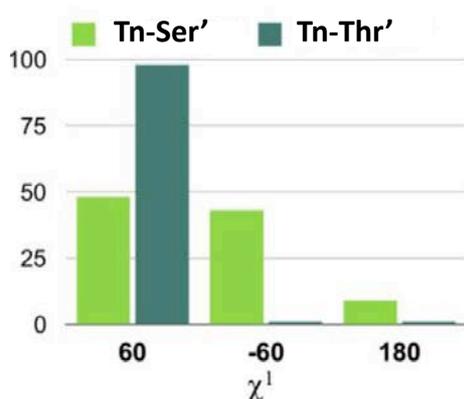
*Figure 3.13 χ^1 torsional angle values in **Tn-Ser'** and **Tn-Thr'** derivatives.*

Structural ensembles derived from MD-tar simulations of both studied derivatives are depicted in Figure 3.14. As expected, the orientation of the carbohydrate in the **Tn-Thr'** antigen is nearly perpendicular to the peptide backbone, while in the Ser derivative the GalNAc moiety presents a parallel disposition. This observation also supports previous results.[51] Hence, we can conclude that benzylamide moiety does not interfere with glycosyl amino acid conformation.

*Figure 3.14 Structural ensembles of **Tn-Thr'** and **Tn-Ser'** antigens derived from MD-tar calculations.*
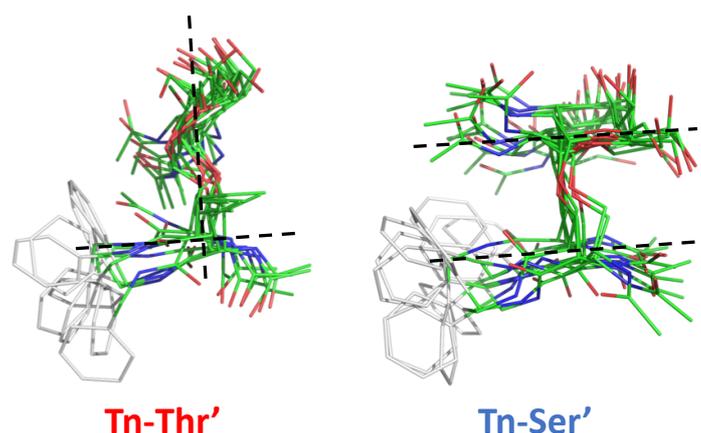
In conclusion, conformational analysis in solution demonstrates that Tn antigen structure is closely connected with the underlying amino acid moiety, either serine or threonine. To clarify whether this structural variation resides in intrinsic steric interactions related to the structure itself or the effect of water, conformational studies in the gas phase were conducted to observe molecular preferences free of the interference of solvent.

3.3.3. Conformational analysis in the gas phase

The conformation of **Tn-Ser'** and **Tn-Thr'** derivatives was evaluated in the gas phase in order to study the structure of both molecules in the absence of solvent. In this case, mass- and conformer-selected infrared laser spectroscopy, complemented with QM calculations, were employed. (Both described in *Experimental section 8.9.2.* and *8.9.3.*). The use of the benzylamide derivatives facilitates the detection through mass-selected ultraviolet (UV) photoionization. This study was conducted in collaboration with researchers from the Biofisika Institute (University of the Basque Country, Spain).

Once the samples were vaporized by laser desorption, the infrared ion-dip (IRID) spectra in the gas phase were recorded. The region in the range 2800-3800 cm^{-1}, corresponding to N-H and O-H stretching vibrational modes, is the most relevant for the conformational analysis, since this region is highly sensitive to the presence of hydrogen bonds. As can be seen in Figure 3.17 (see below), the recorded spectra a virtually identical for both molecules, which indicates that both antigens displayed the same pattern of hydrogen bonds in the gas phase.

In parallel, a conformational search was conducted for both derivatives following a Monte Carlo algorithm[70] and using several force fields (MMFFs,[71] AMBER[72] and OPLS2005[73]). Only those structures with the lowest energy (<15 kJ·mol^{-1}) were re-optimized through density functional theory calculations. The most populated conformers of **Tn-Thr'** and **Tn-Ser'** are shown in Figure 3.15 and Figure 3.16, respectively. In addition, their associated harmonic OH and NH vibrational spectra were predicted and compared to the experimental spectrum (Figure 3.15 and Figure 3.16 left panel).

The best fitting between experiment and theory corresponds to the molecules designated as 01* in both Figure 3.15 and Figure 3.16 (red spectrum on the left panel) and is further discussed below.
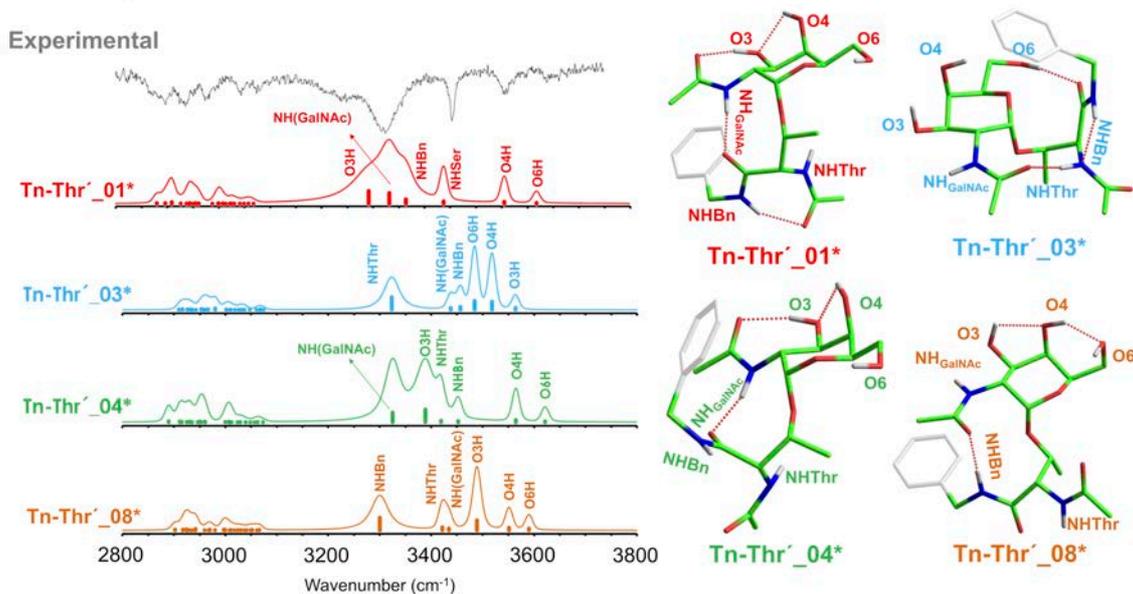

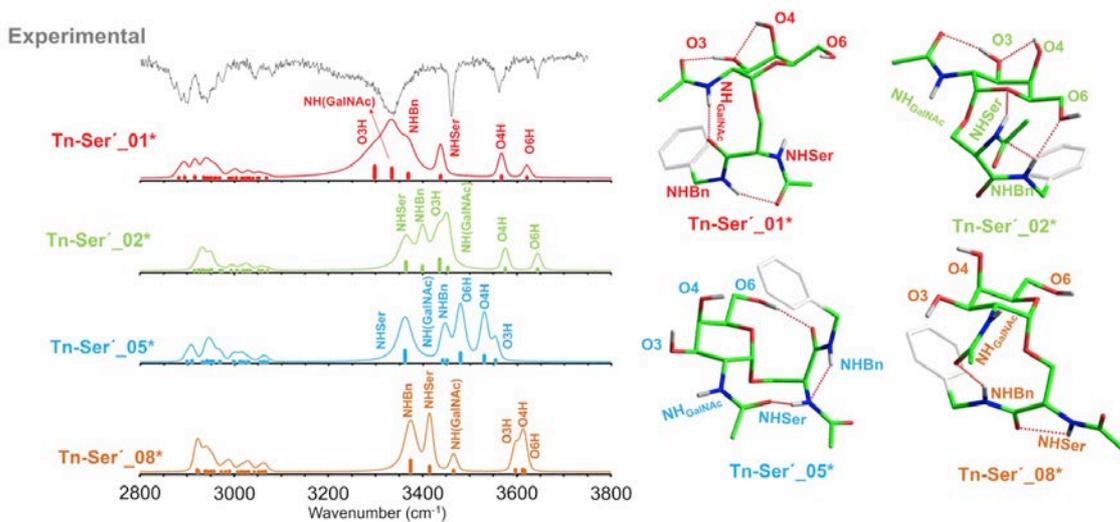*Figure 3.15 Analysis of the main conformers of **Tn-Thr'** antigen in the gas phase.*

*Figure 3.16 Analysis of the main conformers of **Tn-Ser'** antigen in the gas phase.*

The lowest free energy conformers calculated for both derivatives using different QM methods are very similar and show an excellent agreement between experimental and calculated spectra (Figure 3.17).
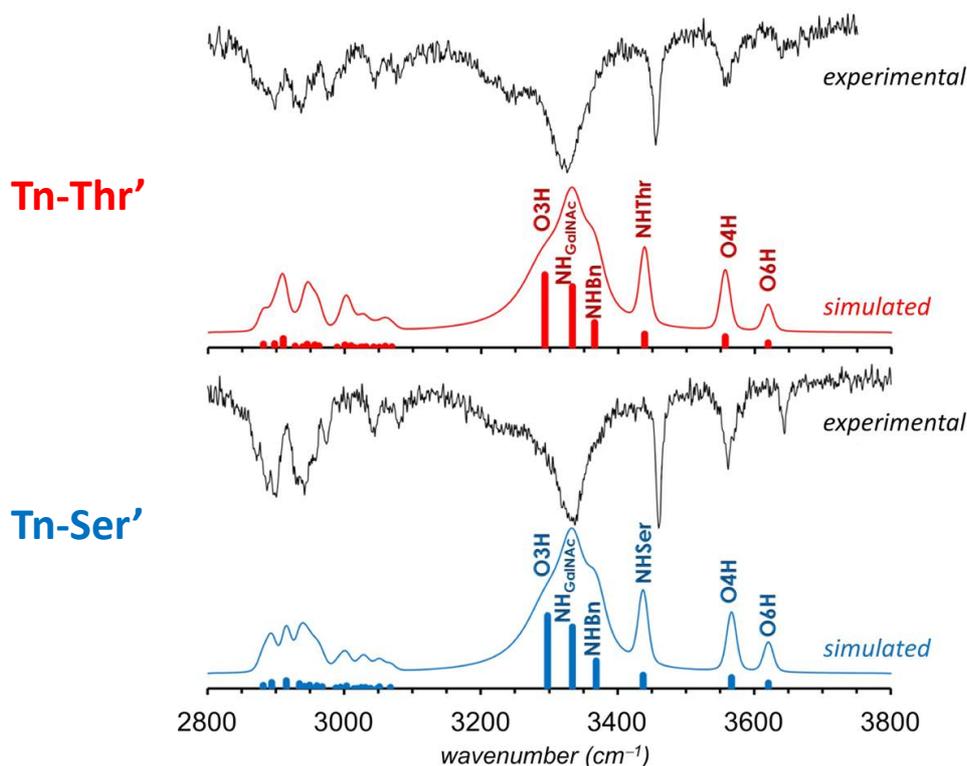
*Figure 3.17 Experimental and simulated IRID spectra of **Tn-Thr'** and **Tn-Ser'** compounds.*
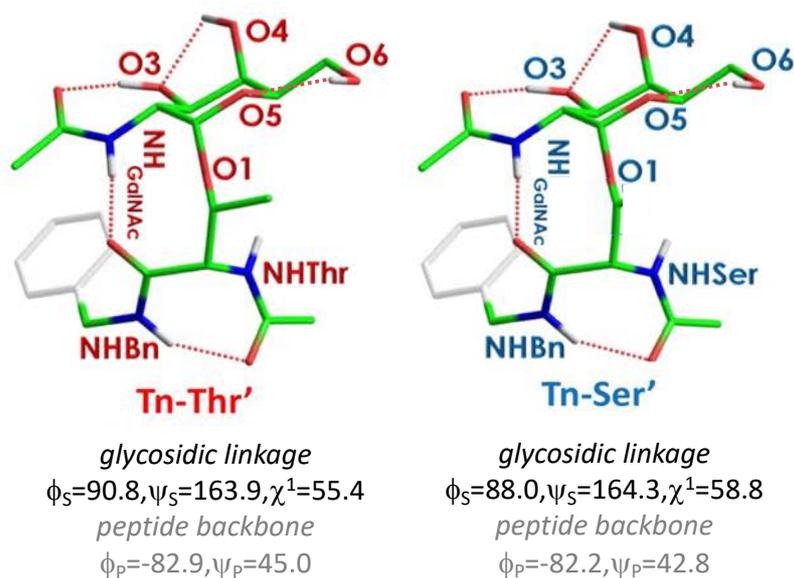
*Figure 3.18 Lowest energy conformers calculated for **Tn-Thr'** and **Tn-Ser'** at the M06-2X/6-31G(d,p) level and intramolecular hydrogen bonds as red dashed lines.*

In these preferred conformations, the amino acid backbone adopts an inverse γ-turn stabilized by a strong hydrogen bond between the amino acid *C*-terminal amide and the *N*-terminal acetamide carbonyl group (band at 134 ~3370 cm^{-1} for NH) (Figure 3.18). Synergistically with the amino acid conformation, the *N*-acetyl group of the

carbohydrate −essential for biological activity− is engaged in two strong hydrogen bonds that constitute the main driving force for the special architecture of the glycosidic bond in the Tn antigens in the gas phase: one connecting the carbonyl and the hydroxyl O3H groups of the GalNAc moiety and the other involving the NH of GalNAc to the carbonyl group of each amino acid. The occurrence of these common stabilizing interactions is in good agreement with the bands observed within the $\sim$3200–3350 cm^{-1} region for the O3H and the NH of the carbohydrate. Additional weaker hydrogen bonds of the carbohydrate appear engaging OH4 with O3 and also O6H is connected to the endocyclic oxygen O5. Strikingly, in the gas phase, **Tn-Ser'** and **Tn-Thr'** present a staggered conformation for the glycosidic linkage. They both present similar values for the torsional angles (Figure 3.18); ψ_S is close to 164 °, ϕ_S to 90º and the χ^1 torsional angle is around 60º. These outcomes point out the crucial role of surrounding water molecules in the conformational preferences of these entities in solution. Therefore, first hydration shell will be studied in the next section.

3.3.4. First hydration shell

As aforementioned, both molecules (**Tn-Ser'** and **Tn-Thr'**) display the same conformation in the gas phase. The values of the most relevant torsional angles are shown in Figure 3.19.
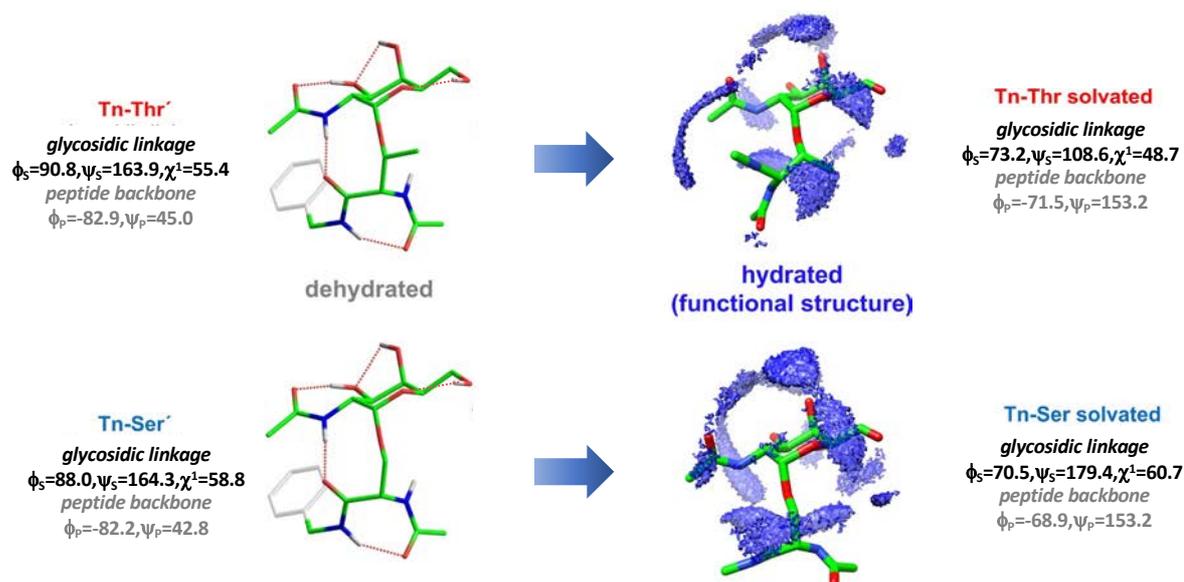

Tn-Thr'
glycosidic linkage
ϕ_S=90.8,ψ_S=163.9,χ^1=55.4
peptide backbone
ϕ_P=-82.9,ψ_P=45.0

dehydrated

hydrated (functional structure)

Tn-Thr solvated
glycosidic linkage
ϕ_S=73.2,ψ_S=108.6,χ^1=48.7
peptide backbone
ϕ_P=-71.5,ψ_P=153.2

Tn-Ser'
glycosidic linkage
ϕ_S=88.0,ψ_S=164.3,χ^1=58.8
peptide backbone
ϕ_P=-82.2,ψ_P=42.8

Tn-Ser solvated
glycosidic linkage
ϕ_S=70.5,ψ_S=179.4,χ^1=60.7
peptide backbone
ϕ_P=-68.9,ψ_P=153.2

Figure 3.19 Lowest energy conformers calculated at the M06-2X/6-31+G level for **Tn-Thr'** (upper panel) and **Tn-Ser'** (lower panel) combined with the first hydration shell.*

When these Tn antigen derived conformers are combined with the first hydration shell derived from the 200 ns MD-tar simulations (Figure 3.19 right side), their structure varies. In Tn-Ser derivative the first hydration shell promotes the extended arrangement of the peptide backbone and staggered conformation observed in vacuum is retained. A water pocket connects the NH of GalNAc to the carbonyl group of the glycosylated amino acid (Figure 3.20-A). In contrast, Tn-Thr molecule is forced to turn over O1-Cβ bond to accommodate the complete solvation of the antigen, otherwise it would be

interrupted by the β-methyl group of Thr. ψ_s torsion angle rotates 60°, from staggered conformation in the gas phase (ψ_s= 180°) to the eclipsed form (ψ_s= 120°) observed in solution. In this case, the water pocket appears between the NH groups of the sugar and the amino acid (Figure 3.20-B)
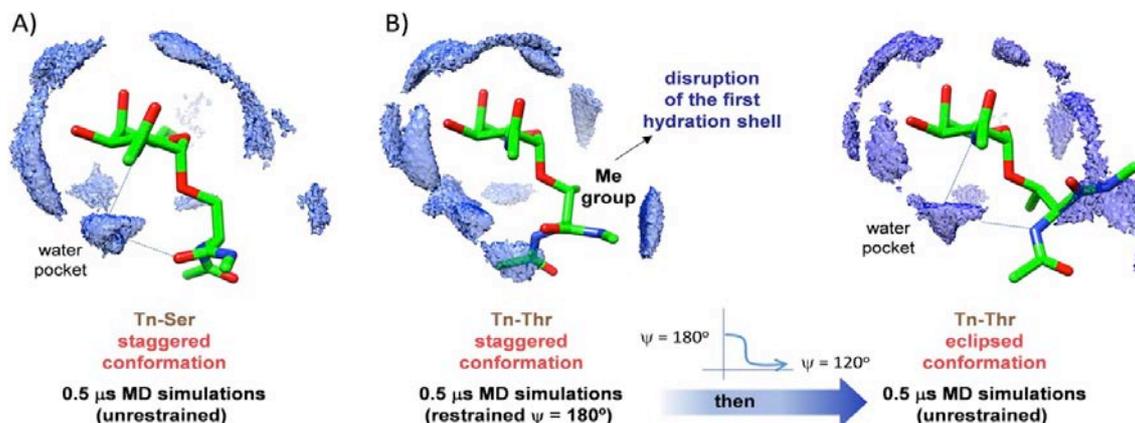

Figure 3.20 Tn-Ser (A) and Tn-Thr (B) lowest energy conformers together with the first hydration shell derived from MD-tar simulations.

Therefore, Tn antigen conformation relies upon both its first hydration shell and the presence of a methyl group on Cβ. In vacuum, Tn-Thr and Tn-Ser depict a staggered form; whereas when they are studied surrounded by water molecules the β-methyl group of Thr disrupts the first hydration shell forcing this antigen into an eclipsed conformation.

3.4. Fluorinated glycopeptides

The presence of a bridging water molecule, connecting the NH groups of GalNAc and Thr, has not been experimentally demonstrated yet, probably due to the low residence time of the water molecule. We hypothesized that a more hydrophilic water pocket would capture the water molecule as a crystallization water, so it could be observed by X-ray. Thus, the hydrogen-bonding donor character of the sugar *N*-acetyl fragment could be enhanced by replacing its constituent methyl group by fluoromethyl groups.

In this respect, previously in our group, mono- and difluorinated Tn antigen derivatives, compounds **Tn-fThr*'** and **Tn-2fThr*'**, respectively, were synthesized and subsequently engaged into MUC1-related glycopeptides,[74] **F-Thr*'** and **2F-Thr*'**, to be studied in the solid state (Figure 3.21).
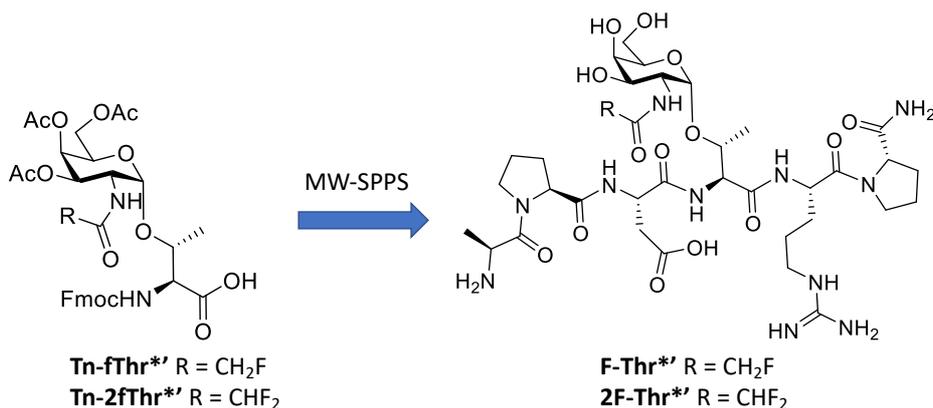
Figure 3.21 Fluorinated glycopeptide synthetized.

3.4.1. Conformational analysis in the solid state

Quantum mechanical calculations were performed on a reduced model of **2F-Thr*'** (Figure 3.22, left panel). These evidenced that the water molecule is stabilized through an O-H···F contact, in addition to the bridging contact between NH groups of sugar and Thr residue. Besides, water oxygen distances to NH-GalNAc (1.97 Å) and NH-Thr (1.91 Å) are in agreement with those previously observed in MD-tar simulations for its natural counterpart (2.03 Å and 1.95 Å, respectively).[51]

Furthermore, 200 ns MD simulations performed on glycopeptide **2F-Thr*'** bound to scFv-SM3 fragment showed a high-water density placed where the water pocket is commonly observed in Tn-Thr antigen (Figure 3.22, right panel). This finding suggests that the fluorinated GalNAc unit configures a more hydrophilic environment, thus water molecule would be more easily captured in the crystal structure.
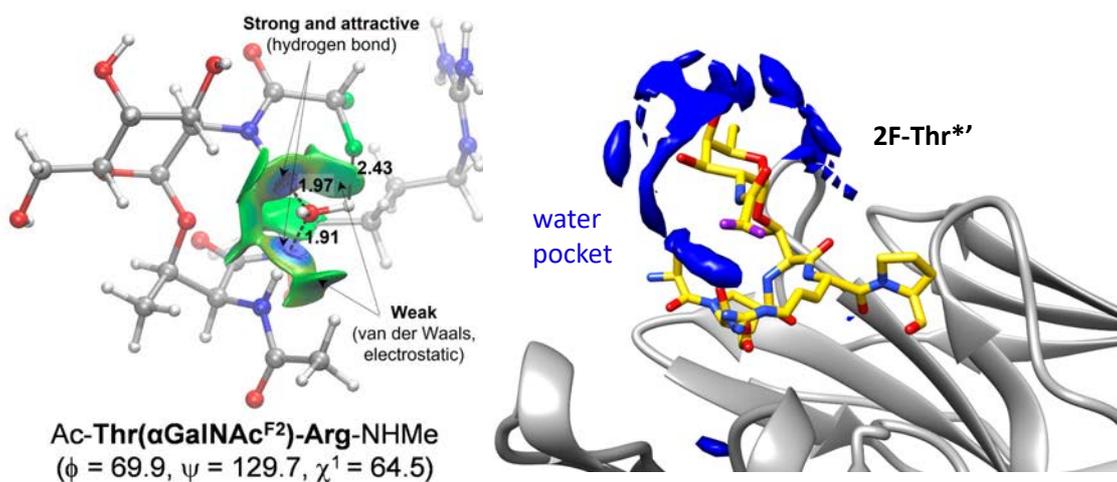

*Figure 3.22 QM calculations (M06-2X/6-311G(d,p)) performed on a reduced model of **2F-Thr*'** (left) and first hydration shell over **2F-Thr*'** complexed with scFv-SM3 derived from MD simulations (right).*

X-Ray crystallography studies were conducted as previously described.[41] (See also *Experimental section 8.8.).* Fluorinated glycopeptides **F-Thr*'** and **2F-Thr*'** complexed to scFv-SM3 antibody were solved at high resolution (<2.0 Å). As shown in Figure 3.23, interfacial water molecules were observed for the first time. As was anticipated in solution,[51,75] water molecules appear bridging the NH groups of GalNAc and the Thr moiety. The surface groove nicely fits all the peptide residues in both cases, irrespective of the presence of the fluorine atom, retaining the overall conformation of the natural glycopeptide. The stabilizing contacts are mostly hydrogen bonds, some of them mediated by water molecules, and also stacking interactions, again similar to that found in the natural variant. Finally, the glycosidic linkage of fluorinated compounds retains the "eclipsed" conformation ($\psi \approx 120^\circ$) previously reported for its natural counterpart in solid phase.[41]

glycosidic linkage
$\phi_S = 66.0^\circ$, $\psi_S = 125.9^\circ$

glycosidic linkage
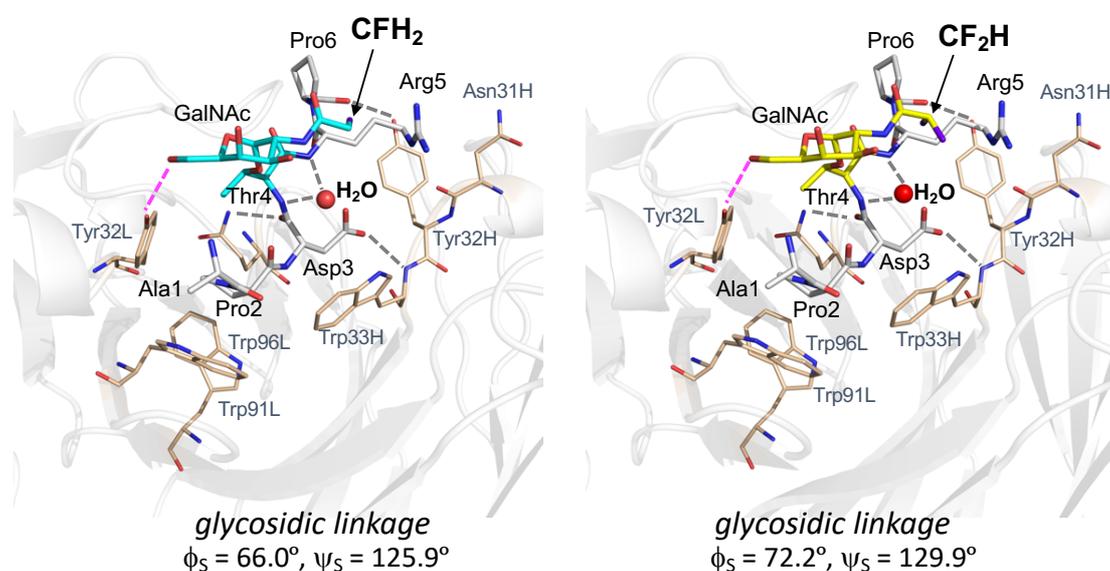$\phi_S = 72.2^\circ$, $\psi_S = 129.9^\circ$

*Figure 3.23 Crystal structure of glycopeptides **F-Thr*'** (pdb ID: 6FZR, left picture) and **2F-Thr*'** (pdb ID: 6FZQ, right picture) bound to scFv-SM3 antibody.*

3.5. Conclusions

The work compiled in this chapter provides new insight in Tn antigen conformation. Herein we have demonstrated the nonequivalence of Tn-Ser and Tn-Thr, both experimental and theoretically.

Initially, Tn antigen benzylamide derivatives, **Tn-Thr'** and **Tn-Ser'**, were synthesized. Their structure was then studied in solution by NMR experiments and MD-tar simulations. Supporting our previous observations, both compounds present different carbohydrate orientation. Thus, whereas GalNAc moiety appears perpendicularly oriented over the peptide backbone in the Thr derivative; in the case of the Ser moiety it is arranged in a parallel way over the underlying peptide.

Tn-Thr' and **Tn-Ser'** conformational behavior was also analyzed in the gas phase through IRID spectroscopy combined with QM calculations. Strikingly, the geometry that both derivatives present is indistinguishable from one another. This finding states the importance of the water molecules surrounding Tn antigen in sculpting their unique behavior in solution.

Consequently, first hydration shell was studied by extensive MD calculations, revealing a ψ_s torsion angle rotation of 60º for the Tn-Thr derivative. This turn serves to accommodate the complete solvation of the antigen, otherwise it would be interrupted by the β-methyl group of Thr. In this way, this molecule turns from staggered conformation in the gas phase (ψ_s = 180º) to the eclipsed form (ψ_s = 120º) observed in solution. Conversely, Tn-Ser derivative retains the staggered conformation shown in the gas phase.

Finally, the structure of the Tn-Thr antigen in the solid state was investigated. To that end mono- and difluorinated Tn-Thr antigen moieties were synthesized and incorporated in the APDTRP epitope. In this way, a more hydrophilic water pocket could be stablished, as supported by QM and MD calculations. Certainly, the crystal structures reveal a structural water molecule located between the amino group of the GalNAc and the NH group of the glycosylated Thr residue, as we have proposed earlier for the natural Tn-Thr antigen in solution.[51] To the best of our knowledge, this is the first confirmation of a commonly accepted hypothesis in the field of *O*-glycopeptides.

In conclusion, Tn-Ser and Tn-Thr structures have been proved nonequivalent. Different arrangement of the glycosidic linkage is due not only to the β-methyl group of Thr, but also to the key role of water in the modulation of the conformational preferences of both antigens.

This work has been published as an article entitled *Water Sculpts the Distinctive Shapes and Dynamics of the Tn Antigens: Implications for their Molecular Recognition* in the *Journal of American Chemical Society*, DOI: 10.1021/jacs.8b04801.[74]

3.6. References

1. Ju, T., Otto, V. I. & Cummings, R. D. The Tn antigena-structural simplicity and biological complexity. *Angew. Chem. Int. Ed.* **50,** 1770–1791 (2011).
2. Freire, T. & Osinaga, E. Immunological and biomedical relevance of the Tn antigen. *Inmunología* **22,** 27–38 (2003).
3. Sadraei, S. I., Reynolds, M. R. & Trant, J. F. The synthesis and biological characterization of acetal-free mimics of the tumor-associated carbohydrate antigens. in *Advances in Carbohydrate Chemistry and Biochemistry* (ed. Baker, D.) **74,** 137–237 (Elsevier Inc., 2017).
4. Berger, E. G. Tn-syndrome. *Biochim. Biophys. Acta* **1455,** 255–268 (1999).
5. Dausset, J., Moullec, J. & Bernard, J. Acquired hemolytic anemia with polyagglutinability of red blood cells due to a new factor present in normal human serum (anti-Tn). *Blood* **XIV,** 1079–1093 (1959).
6. Moreau, R., Dausset, J., Bernard, J. & Moullec, J. Anémie hémolytique acquise avec polyagglutinabilité des hématies par un nouveau facteur présent dans le sérum humain normal (anti-Tn). *Bull. Mem. Soc. Med. Hop. Paris* **73,** 569–587 (1957).
7. Beck, M. L. Red blood cell polyagglutination: Clinical aspects. *Semin. Hemathology* **37,** 186–196 (2000).
8. Friedenreich, V. *The Thomsen Hemagglutination Phenomenon: Production of a Specific Receptor Quality in Red Corpuscles by Bacterial Activity.* (Levin & Munksgaard, 1930).
9. Gabius, H., Kaltner, H., Kopitz, J. & Andre, S. The glycobiology of the CD system : a dictionary for translating marker designations into glycan/lectin structure and function. *Trends Biochem. Sci.* **40,** 360–376 (2015).
10. Dahr, W., Uhlenbruck, G., Gunson, H. H. & Van der Hart, M. Molecular Basis of Tn-Polyagglutinability. *Vox Sang.* **29,** 36–50 (1975).
11. Springer, G. F., Desai, P. R. & Banatwala, I. Blood group MN specific substances and precursors in normal and malignant human breast tissues. *Naturwissenschaften* **61,** 457–458 (1974).
12. Springer, G. T and Tn, general carcinoma autoantigens. *Science (80-.).* **224,** 1198–1206 (1984).
13. Springer, G. F. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *J. Mol. Med.* **75,** 594–602 (1997).
14. Desai, P. R. Immunoreactive T and Tn antigens in malignancy: role in carcinoma diagnosis, prognosis, and immunotherapy. *Transfus. Med. Rev.* **14,** 312–25 (2000).
15. Helferich, B. & Weis, K. Zur synthese von glucosiden und von nicht-reduxierenden disscchariden. *Chem. Ber.* **89,** 314–321 (1956).
16. Wang, Z. Helferich Glycosylation. in *Comprehensive Organic Name Reactions and Reagents* (ed. Wang, Z.) 1367–1370 (Wiley, 2010).
17. Kaifu, R. & Osawa, T. Synthesis of O-(2-acetamido-2-deoxy-alpha-D-galactopyranosyl)-N-tosyl-L-serine. *Carbohydr. Res.* **58,** 235–239 (1977).
18. Martins-Teixeira, M. B., Campo, V. L., Biondo, M., Sesti-Costa, R., Carneiro, Z. A., Silva, J. S. & Carvalho, I. α-selective glycosylation affords mucin-related GalNAc amino acids and diketopiperazines active on Trypanosoma cruzi. *Bioorganic Med. Chem.* **21,** 1978–1987 (2013).
19. Koenigs, W. & Knorr, E. Ueber einige derivate des traubenzuckers und der galactose. *Chem. Ber.* **34,** 957–981 (1901).
20. Liu, M., Young, V. G., Lohani, S., Live, D. & Barany, G. Syntheses of Tn building blocks N-alpha-(9-fluorenylmethoxycarbonyl)-O-serine/L-threonine pentafluorophenyl esters: comparison of protocols and elucidation of side reactions. *Carbohydr. Res.* **340,** 1273–1285 (2005).

21.   Plattner, C., Michael, H. & Sewald, N. One-pot azidochlorination of glycals. *Org. Lett.* **13,** 545–547 (2011).

22.   Paulsen, H. & Hölck, J.-P. Synthese der glycopeptide O-beta-D-galactopyranosyl(1-3)-O-(2-acetamido-2-desoxy-alpha-D-galactopyranosyl)-(1-3)-L-serin   und   -L-threonin. *Carbohydr. Res.* **109,** 89–107 (1982).

23.   Madariaga, D. New insights into molecular recognition of epitopes related to MUC1 human mucin by lectins. Doctoral dissertation. (Universidad de La Rioja, 2015).

24.   Martínez-Sáez, N. Diseño racional de derivados de mucina como candidatos a vacunas contra el cáncer. Doctoral dissertation. (Universidad de La Rioja, 2013).

25.   Kunz, H. & Birnbach, S. Synthesis of O-glycopeptides of the tumor-associated Tn- and T-antigen type and their binding to bovine serum albumin. *Angew. Chem. Int. Ed.* **25,** 360–362 (1986).

26.   Liebe, B. & Kunz, H. Solid-phase synthesis of a tumor-associated sialyl-Tn antigen glycopeptide with a partial sequence of the 'tandem repeat' of the MUC-1 mucin. *Angew. Chem. Int. Ed.* **36,** 618–621 (1997).

27.   Satyanarayana, J., Gururaja, T. L., Naganagowda, G. A., Ramasubbu, N. & Levine, M. J. A concise methodology for the stereoselective synthesis of O-glycosylated amino acid building blocks: complete 1H NMR assignments and their application in solid-phase glycopeptide synthesis. *J. Pept. Res.* **52,** 165–179 (1998).

28.   Payne, R. J., Ficht, S., Tang, S., Brik, A., Yang, Y., Case, D. A., Wong, C. & Wong, C. Extended sugar-assisted glycopeptide ligations: Development, scope, and applications. *J. Am. Chem. Soc.* **129,** 13537–13536 (2007).

29.   Toyokuni, T., Dean, B. & Hakomori, S.-I. Synthetic vaccines: I. Synthesis of multivalent Tn antigen cluster-lysyllysine conjugates. *Tetrahedron Lett.* **31,** 2673–2676 (1990).

30.   Cato, D., Buskas, T. & Boons, G.-J. Highly efficient stereospecific preparation of Tn and TF building blocks using thioglycosyl donors and the Ph2SO/Tf2O promotor system. *J. Carbohydr. Chem.* **24,** 503–516 (2005).

31.   McDonald, D. M., Byrne, S. N. & Payne, R. J. Synthetic self-adjuvanting glycopeptide cancer vaccines. *Front. Chem.* **3,** 1–8 (2015).

32.   Buskas, T., Ingale, S. & Boons, G. Glycopeptides as versatile tools for glycobiology. *Glycobiology* **16,** 113R–136R (2006).

33.   Kunz, H., Birnbach, S. & Wernig, P. Synthesis of glycopeptides with the T, and T antigen struc- tures, and their coupling to bovine serum albumin. *Carbohydr. Res.* **202,** 207–223 (1990).

34.   Grogan, M. J., Pratt, M. R., Marcaurelle, L. A. & Bertozzi, C. R. Homogeneous glycopeptides and glycoproteins for biological investigation. *Annu. Rev. Biochem.* **71,** 593–634 (2002).

35.   Pratt, M. R. & Bertozzi, C. R. Synthetic glycopeptides and glycoproteins as tools for biology. *Chem. Soc. Rev.* **34,** 58–68 (2005).

36.   Kuduk, S. D., Schwarz, J. B., Chen, X., Glunz, P. W., Sames, D., Ragupathi, G., Livingston, P. O. & Danishefsky, S. J. Synthetic and immunological studies on clustered modes of mucin-related Tn and TF O-linked antigens: the preparation of a glycopeptide-based vaccine for clinical trials against prostate. *J. Am. Chem. Soc.* **120,** 12474–12485 (1998).

37.   Das, J. & Schmidt, R. R. Convenient glycoside synthesis of amino sugars: Michael-type addition to 2-Nitro-D-galactal. *Eur. J. Org. Chem.* 1609–1613 (1998).

38.   García de Luis, M. Glicopéptidos modelo que incorporan el antígeno Tn. Doctoral dissertation. (Universidad de La Rioja, 2008).

39.   Naganagowda, G. A., Gururaja, T. L., Satyanarayana, J. & Levine, M. J. NMR analysis of human salivary mucin (MUC7) derived O-linked model glycopeptides comparison of structural features and carbohydrate-peptide interactions. *J. Pept. Res.* **54,** 290–310 (1999).

40.   Mazal, D., Lo-Man, R., Bay, S., Pritsch, O., Dériaud, E., Ganneau, C., Medeiros, A., Ubillos,

L., Obal, G., Berois, N., Bollati-Fogolin, M., Leclerc, C. & Osinaga, E. Monoclonal antibodies toward different Tn-amino acid backbones display distinct recognition patterns on human cancer cells. Implications for effective immuno-targeting of cancer. *Cancer Immunol. Immunother.* **62,** 1107–1122 (2013).

41. 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. Deciphering the non-equivalence of serine and threonine O-glycosylation points: implications for molecular recognition of the Tn antigen by an anti-MUC1 antibody. *Angew. Chem. Int. Ed. Engl.* **127,** 9968–9972 (2015).

42. 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. The quest for anticancer vaccines: deciphering the fine-epitope specificity of cancer-related monoclonal antibodies by combining microarray screening and saturation transfer difference NMR. *J. Am. Chem. Soc.* **137,** 12438–12441 (2015).

43. Grinstead, J. S., Koganty, R. R., Krantz, M. J., Longenecker, B. M. & Campbell, A. P. Effect of glycosylation on MUC1 humoral immune recognition: NMR studies of MUC1 glycopeptide-antibody interactions. *Biochemistry* **41,** 9946–9961 (2002).

44. Zhang, Y., Li, Q., Rodriguez, L. G. & Gildersleeve, J. C. An array-based method to identify multivalent inhibitors. *J. Am. Chem. Soc.* **132,** 9653–9662 (2010).

45. Madariaga, D., Martínez-Sáez, N., Somovilla, V. J., García-García, L., Berbis, M. Á., Valero-Gónzalez, J., Martín-Santamaría, S., Hurtado-Guerrero, R., Asensio, J. L., Jiménez-Barbero, J., Avenoza, A., Busto, J. H., Corzana, F. & Peregrina, J. M. Serine versus threonine glycosylation with alpha-O-GalNAc: unexpected selectivity in their molecular recognition with lectins. *Chem. Eur. J.* **20,** 12616–12627 (2014).

46. Lele, D. S., Talat, S., Kumari, S., Srivastava, N. & Kaur, K. J. Understanding the importance of glycosylated threonine and stereospecific action of Drosocin, a proline rich antimicrobial peptide. *Eur. J. Med. Chem.* **92,** 637–647 (2015).

47. Kanzaki, T., Mizuno, N., Yokota, M., Matsumoto, Y. & Hirabayashi, Y. Novel lysosomal glycoaminoacid storage disease with angiokeratoma corporis diffusum. *Lancet* **333,** 875–877 (1989).

48. Kanekura, T., Sakuraba, H., Matsuzawa, F., Aikawa, S., Doi, H., Hirabayashi, Y., Yoshii, N., Fukushige, T. & Kanzaki, T. Three dimensional structural studies of alpha-N-acetylgalactosaminidase (alpha-NAGA) in alpha-NAGA deficiency (Kanzaki disease): different gene mutations cause peculiar structural changes in alpha-NAGAs resulting in different substrate specificities and c. *J. Dermatol. Sci.* **37,** 15–20 (2005).

49. Brister, M. A., Pandey, A. K., Bielska, A. A. & Zondlo, N. J. OGlcNAcylation and phosphorylation have opposing structural effects in tau: Phosphothreonine induces particular conformational order. *J. Am. Chem. Soc.* **136,** 3803–3816 (2014).

50. Tachibana, Y., Fletcher, G. L., Fujitani, N., Tsuda, S., Monde, K. & Nishimura, S.-I. Antifreeze glycoproteins: elucidation of the structural motifs that are essential for antifreeze activity. *Angew. Chem. Int. Ed.* **116,** 874–880 (2004).

51. Corzana, F., Busto, J. H., Jiménez-Osés, G., De Luis, M. G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M. & Avenoza, A. Serine versus threonine glycosylation: The methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *J. Am. Chem. Soc.* **129,** 9458–9467 (2007).

52. Corzana, F., Busto, J. H., Jiménez-Osés, G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M. & Avenoza, A. New insights into alpha-GalNAc-Ser motif: Influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc.* **128,** 14640–14648 (2006).

53. Corzana, F., Busto, J. H., García de Luis, M., Jiménez-Barbero, J., Avenoza, A. & Peregrina, J. M. The nature and sequence of the amino acid aglycone strongly modulates the

conformation and dynamics effects of Tn antigen's clusters. *Chem. Eur. J.* **15,** 3863–3874 (2009).

54. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. Investigation of exchange processes by twodimensional NMR spectroscopy. *J. Chem. Phys.* **71,** 4546–4553 (1979).

55. Karplus, M. Vicinal proton coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **85,** 2870–2871 (1963).

56. Tvaroska, I. The conformational properties of the glycosidic linkage. *Carbohydr. Res.* **90,** 173–185 (1981).

57. Ramachandran, G. N. & Sasisekharan, V. Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23,** 283–438 (1968).

58. Kiyohara, K., Gubbins, K. & Panagiotopoulos, A. Phase coexistence properties of polarizable water models. *Mol. Phys.* **94,** 803–808 (1998).

59. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).

60. Corzana, F., Busto, J. H., Engelsen, S. B., Jimenez-Barbero, J., Asensio, J. L., Peregrina, J. M. & Avenoza, A. Effect of beta-O-glucosylation on L-Ser and L-Thr Diamides: A bias toward alpha-helical conformations. *Chem. Eur. J.* **12,** 7864–7871 (2006).

61. Wiberg, K. B., Bailey, W. F., Lambert, K. M. & Stempel, Z. D. The anomeric effect : It's complicated. *J. Org. Chem.* **83,** 5242–5255 (2018).

62. Filloux, C. M. The problem of origins and origins of the problem: Influence of language on studies concerning the anomeric effect. *Angew. Chem. Int. Ed.* **54,** 8880–8894 (2015).

63. Movahedin, M., Brooks, T. M., Supekar, N. T., Gokanapudi, N., Boons, G. & Brooks, C. L. Glycosylation of MUC1 influences the binding of a therapeutic antibody by altering the conformational equilibrium of the antigen. *Glycobiology* **27,** 677–687 (2017).

64. Coltart, D. M., Royyuru, A. K., Williams, L. J., Glunz, P. W., Sames, D., Kuduk, S. D., Schwarz, J. B., Chen, X., Danishefsky, S. J. & Live, D. H. Principles of Mucin architecture: Structural studies on synthetic glycopeptides bearing clustered mono-, di-, tri-, and hexasaccharide glycodomains. *J. Am. Chem. Soc.* **124,** 9833–9844 (2002).

65. Mallajosyula, S. S. & Mackerell, A. D. Influence of solvent and intramolecular hydrogen bonding on the conformational properties of O-linked glycopeptides. *J. Phys. Chem.* **115,** 11215–11229 (2011).

66. 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. Antibody recognition of a unique tumor-specific glycopeptide antigen. *Proc. Natl. Acad. Sci. U S A.* **107,** 10056–10061 (2010).

67. Madariaga, D., Martínez-Sáez, N., Somovilla, V. J., Coelho, H., Valero-González, J., Castro-Lopez, J., Asensio, J. L., Jiménez-Barbero, J., Busto, J. H., Avenoza, A., Marcelo, F., Hurtado-Guerrero, R., Corzana, F. & Peregrina, J. M. Detection of tumor-associated glycopeptides by Lectins: The peptide context modulates carbohydrate recognition. *ACS Chem. Biol.* **10,** 747–756 (2015).

68. Lira-Navarrete, E., de las Rivas, M., Compañón, I., Pallarés, M. C., Kong, Y., Iglesias-Fernández, J., Bernardes, G. J. L., Peregrina, J. M., Rovira, C., Bernadó, P., Bruscolini, P., Clausen, H., Lostao, A., Corzana, F. & Hurtado-Guerrero, R. Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates protein O-glycosylation. *Nat. Commun.* **6,** 6937–6949 (2015).

69. Mezei, P. D. & Csonka, G. I. Unified picture for the conformation and stabilization of the O-glycosidic linkage in glycopeptide model structures. *Struct. Chem.* **26,** 1367–1376 (2015).

70. Chang, G., Guida, W. C. & Still, C. W. An internal coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **111,** 4379–4386 (1989).

71. Halgren, T. A. Merck Molecular Force Field I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **17,** 490–519 (1996).

72. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc. A* **117,** 5179–5197 (1995).

73. Shivakumar, D., Harder, E., Damm, W., Friesner, R. A. & Sherman, W. Improving the prediction of absolute solvation free energies using the next generation OPLS Force Field. *J. Chem. Theory Comput.* **8,** 2553–2558 (2012).

74. 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. Water sculpts the distinctive shapes and dynamics of the Tn antigens: Implications for their molecular recognition. *J. Am. Chem. Soc.* **140,** 9952–9960 (2018).

75. Corzana, F., Busto, J. H., García de Luis, M., Fernández-tejada, A., Rodríguez, F., Jiménez-Barbero, J., Avenoza, A. & Peregrina, J. M. Dynamics and hydration properties of small antifreeze-like glycopeptides containing non-natural amino acids. *Eur. J. Org. Chem.* 3525–3532 (2010).

Replacement of Thr9 by unnatural moieties: amino acid and sugar modifications

4

4.1. Introduction and background

As previously mentioned, MUC1 represent an attractive target in the field of cancer therapy because while it features long carbohydrate chains when presented on the surface of healthy epithelial cells, in cancer cells sugar chains appear truncated and overexpressed. In the latter case, carbohydrate antigens are exposed to the immune system and can boost an immune response. In order to take advantage of this particular hallmark; MUC1 mucin, in combination with diverse TACA, has been extensively employed as a cancer vaccine.[1–4] Nevertheless, glycoproteins are poorly immunogenic, hence unable to prime a potent and long-lasting immune response. Although numerous examples have been reported in the literature where synthetic MUC1-based glycopeptide vaccines are capable of eliciting an antibody response, this is in general only humoral immune response, lacking an immunological memory.[4–6] As a consequence, they usually end up failing in clinical trials due to the tolerance of the immune system towards endogenous structures. To overcome this problem, a variety of different approaches have been employed to elicit an improved immune response. Proposed alternatives are, for instance, multivalent presentation of TACA, as dendrimeric structures[7–14] or RAFT (Regioselectively Addresable Functionalised Template) cyclopeptide scaffolds.[15–20]

Also, to enhance immunogenicity, MUC1 glycopeptides have been conjugated with T-cell epitope peptides, carrier proteins or lipopeptide ligands, used as immunostimulants. Resulting from this approach, we can find synthetic glycopeptide vaccines combining two or three of these constituents. A popular trend has been the so-called two-component vaccines (Figure 4.1), consisting of a MUC1 glycopeptide, as B cell stimulant, together with a T cell epitope, included to induce an immunoglobulin class-switching from IgM to IgG antibodies, hence immunological memory.
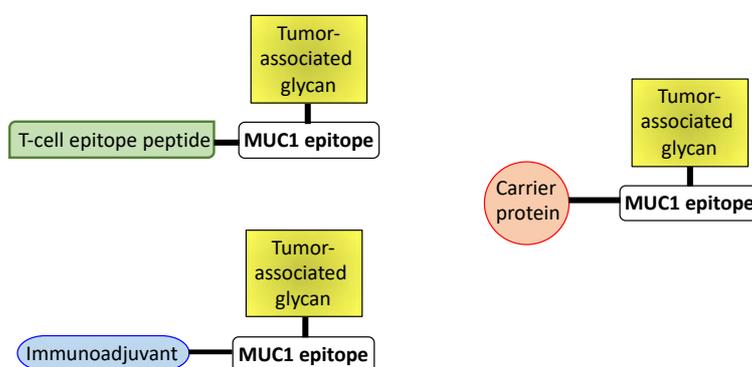

Figure 4.1 Design of two-components synthetic vaccines.

T-cell epitope peptides derived from Tetanus toxoid (TTox),[21] ovoalbumin (OVA)[22,23] or Pan DR (PADRE)[24,25] have been used in this approach. Most common carrier proteins are

bovine serum albumin (BSA),[6,26–28] Keyhole limpet hemocyanin (KLH)[29–32] and TTox protein.[33–36] Finally, immunoadjuvants such as the lipopeptides dipalmitoyloxypropyl-*N*-palmitoylcysteine (Pam$_3$Cys)[37–41] or MALP2[42] have been coupled to MUC1 mucin B cell epitopes to enhance their immunogenicity.

Considering that carrier proteins boost an antibody response themselves, hindering the effect of the saccharide antigen, fully synthetic vaccines arose as an alternative. The predictable next step was the development of three-component vaccines, consisting of B cell - the MUC1 (glyco)peptide - and T cell helper epitopes as well as the immunoadjuvant Pam$_3$Cys.

The first tripartite vaccine was reported more than ten years ago and is presented in Figure 4.2Figure 4.2,[43] it is composed of a fragment of the MUC1, a T cell epitope peptide derived from Poliovirus and the Pam$_3$Cys moiety. This scaffold showed an efficient humoral and cellular immune response, as well as the installation of an immunological memory. Derived from this strategy, similar vaccines have been reported, either from the same group,[44–46] or others.[41,47,48]
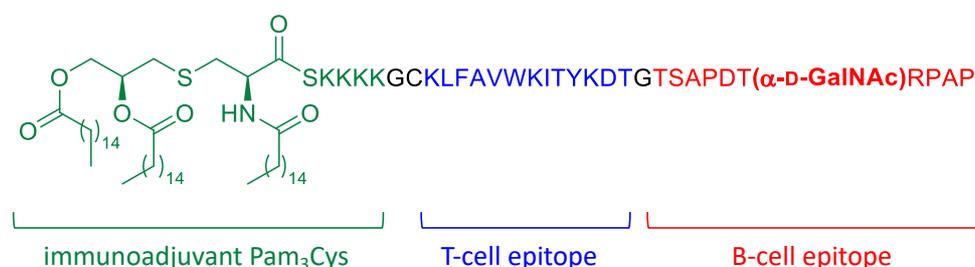

Figure 4.2 First three-component vaccine reported.

Another different strategy to elicit a robust immune response involves the introduction of unnatural variants within vaccine construction. In this way, *in vivo* biostability may be improved, since unnatural epitopes are more resistant to chemical and enzymatic degradation. Besides, the exogenous nature of these derivatives would overcome the tolerance that their natural counterparts present.

Owing to this thesis is focused on Tn antigen, just unnatural variants of this moiety will be detailed herein. Despite that, modifications over other cancer antigens, such as STn and T antigens, have also been reported.[49–54]

Possibly, the most common modification has been the so-called sulfa-Tn antigen, where the oxygen in the *O*-glycosidic bond is substituted by a sulfur atom. Three main strategies have been employed to obtain this building block. The first synthetic route uses cysteine conveniently protected as nucleophile and an azidosugar as glycosyl donor.[55]

The other two approaches utilize a nucleophilic substitution where C1-thio-GalNAc[56] acts as a nucleophile over either β-bromoalanine[57,58] or α,β-dehydroalanine as a Michael acceptor.[59–62] (Scheme 4.1)
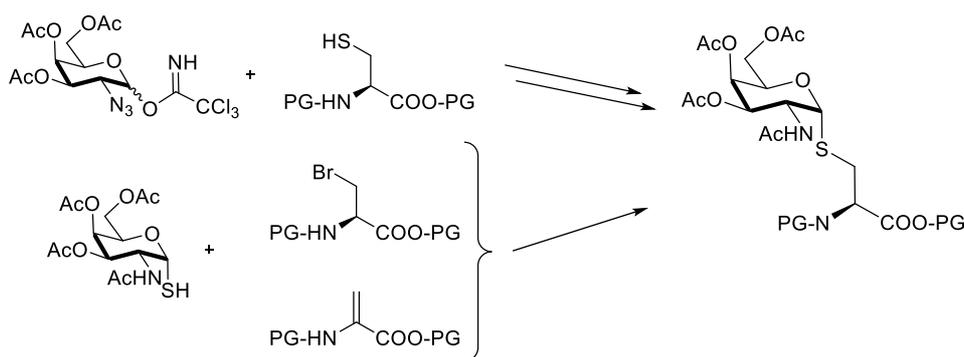
Scheme 4.1 Synthetic routes employed to synthetize sulfa-Tn antigen.

Regretfully, all the aforementioned compounds are only Tn-Ser antigen mimetics, since α-S-GalNac-Thr has not been described to date.

On the other hand, *C*-glycoside analogs of Tn-Ser antigen have also been reported. One of the strategies proposes the condensation of anomeric organosamarium of GalNAc with a convenient aldehyde precursor derived from aspartic acid, which is followed by a deoxygenation step.[63] The other suggests a Wittig-Horner reaction of a C-aldehyde glycoside with a phosphorylglycine moiety.[64] (Scheme 4.2)
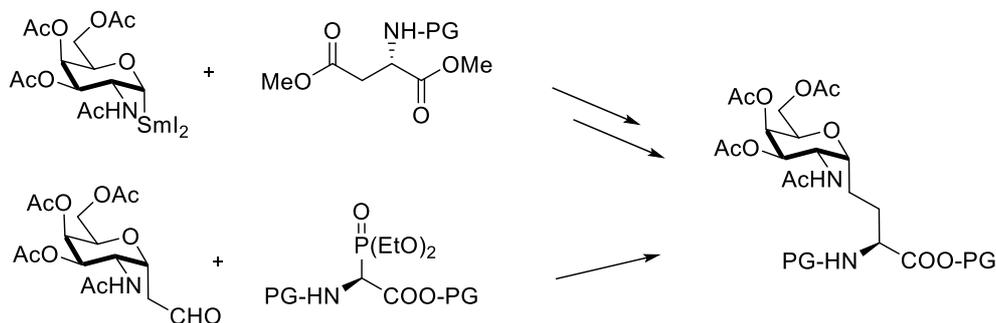
Scheme 4.2 Synthetic routes employed to synthetize C-Tn antigen.

A different modification consists in the elongation of the side chain of the amino acid, for example, replacement of serine by hydroxynorleucine[65–67] or homoserine.[68] A similar approach was conducted through hydrothiolation reaction of a glycosyl-thiol and a double bond included in the side chain of Ser, Thr or Cys derivatives.[58] (Figure 4.3)
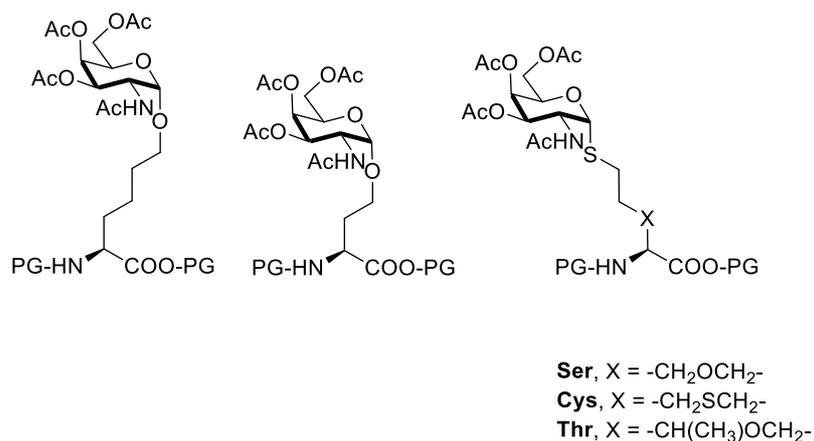
Figure 4.3 Diverse non-natural mimetics of hydroxynorleucine (left), homoserine (center) and glycosyl-thiol derivatives (right).

Otherwise, an oxyme moiety between an α-*O*-GalNAc residue and the amino acid within the peptide backbone has been stablished, preparing this linker via reductive ozonolysis of dehydroleucine selectively condensed with aminooxy sugars[69] (Scheme 4.3). Furthermore, an oxyme linker has been employed to bond a *C*-glycosyl Tn antigen to a T cell epitope to construct a vaccine.[70,71]
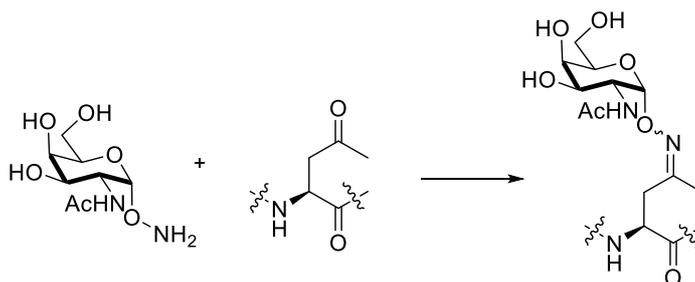
Scheme 4.3 Oxime-linked analogues of the Tn antigen.

Besides, triazol moiety resulting from Cu(I)-catalyzed Huisgen azide-alkyne cycloaddition has been employed to couple carbohydrate and amino acid derivatives. The synthesis of this linker has been reported, either using peptides containing one or more propargylglycines and azido-sugars[72,73] or, on the contrary, azido alanine and propargylated GalNAc[74] (Scheme 4.4). Click cycloaddition strategy has also been use in the synthesis of a multiantigenic vaccine construction.[75]
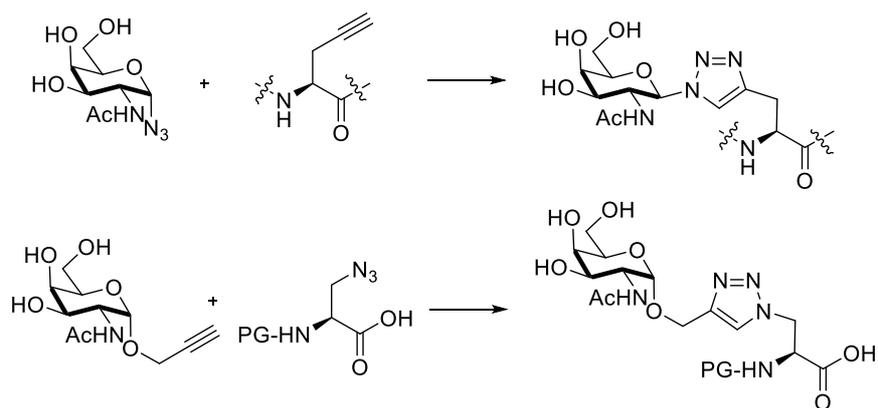
Scheme 4.4 Synthetic routes to obtain triazol-linked Tn antigen derivatives.

The substitution of the underlying amino acid of the Tn antigen by a unnatural residue has also been explored. One representative example corresponds to the quaternary amino acid α-methylserine (MeSer)[76,77] and another to β-amino acid isoserine, whose *O*-glycosylation is accomplished by a ring-opening reaction of isoserine sulfamidate with the carbohydrate.[78] (Figure 4.4)
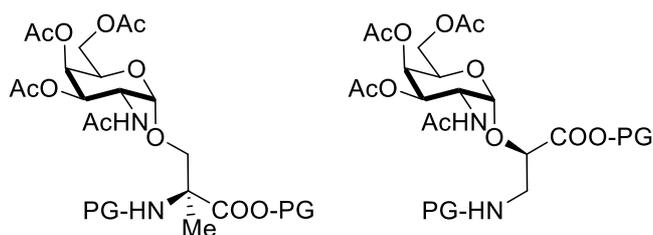

Figure 4.4 α-methyl-serine (left) and isoserine (right) derivatives.

There are other variants using β-amino acids. In one of the cases, the Tn antigen analogue bearing a β-amino acid is accomplished through the modification of an aspartic acid and its subsequently use as glycosyl acceptor,[79,80] and in the other the Arndt-Eistert homologation is employed as a key step.[81] (Scheme 4.5)
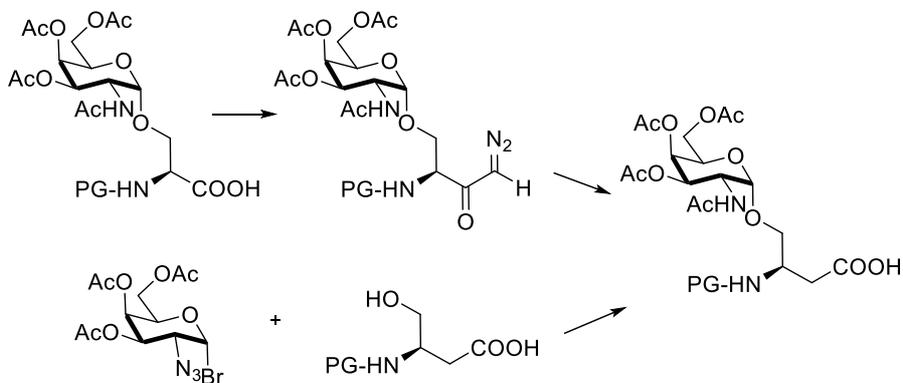
Scheme 4.5 Synthetic routes to obtain β-amino acid analogues.

Regarding to carbohydrate modifications; iminosugar-, thiosugar- or carbasugar-containing vaccines have not yet been reported, since they suffer from low stability or high synthetic complexity.[82,83]

The most remarkable sugar modification over the Tn antigen is represented by a sp^2-iminosugar glycomimetic replacing the α-O-GalNAc moiety. In this derivative, the hydroxymethyl group of the sugar is part of a cyclic carbamate, which limits the conformational flexibility of the system and promotes complete α-stereoselectivity in the glycosylation reaction.[82] (Figure 4.5)
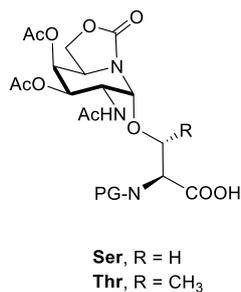

Ser, R = H
Thr, R = CH$_3$

Figure 4.5 sp²-iminosugar mimetic of the Tn antigen.

Also, a fluorine atom have been used to substitute the hydroxyl group at position C6 on the Tn-Thr antigen.[84] It has been reported the synthesis of *N*-mono- and difluoroacetyl galactosamide derivatives, obtained through reduction of the azide group of the sugar moiety after glycosylation reaction and subsequent coupling of *N*-acetyl modified groups.[49] (Figure 4.6)
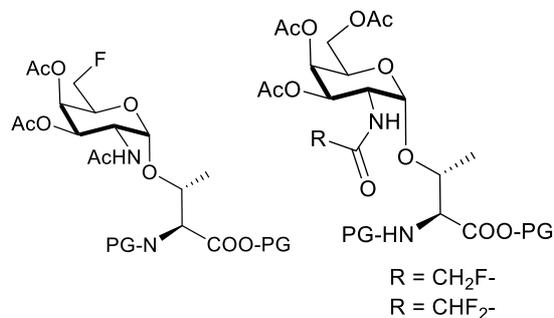

R = CH$_2$F-
R = CHF$_2$-

*Figure 4.6 Fluorinated derivatives of Tn antigen, either replacing OH6 (left)
or over N-acetyl group (right).*

Finally, the synthesis of the mimetic 2-deoxy-2-thio-α-O-galactoside that retains the $^4C_1$ chair conformation of the native Tn antigen, shown in Figure 4.7, has also been described.[15,85]
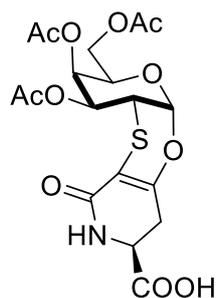
Figure 4.7 Conformationally restricted Tn antigen mimetic.

Some of these unnatural Tn-antigen mimetics have been incorporated and tested *in vivo* within vaccine constructions, leading to results comparable to those obtained from their natural counterparts.

Earliest attempts correspond to multiantigenic glycopeptide vaccines.[29,30,86,87] This kind of scaffolds, designated as cassettes, was comprised by multiple TACA displayed on a single peptide backbone and bound to it though a hydroxynorleucine amino acid (Figure 4.8). These constructions, subsequently bound to KLH protein, elicited IgG and IgM antibody titers.
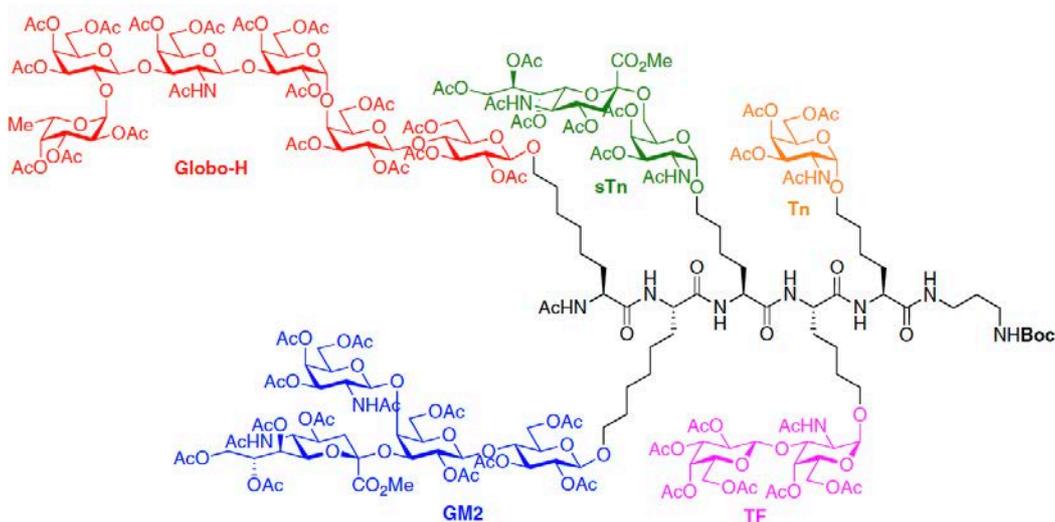

Figure 4.8 N-terminal protected B-cell epitope of glycosylamino acids cassette approach.[86]

In another example, a fully synthetic vaccine composed of the Tn antigen mimetic conjugated to a RAFT scaffold and a OvaPADRE as T-cell epitope (Figure 4.9),[15] induced significant levels of IgG/IgM antibodies, which were still present in the serum 240 days after the last immunization. Moreover, tumor size and mouse survival were monitored, showing a notable reduction in tumor volume and enhanced survival rate compared to controls.
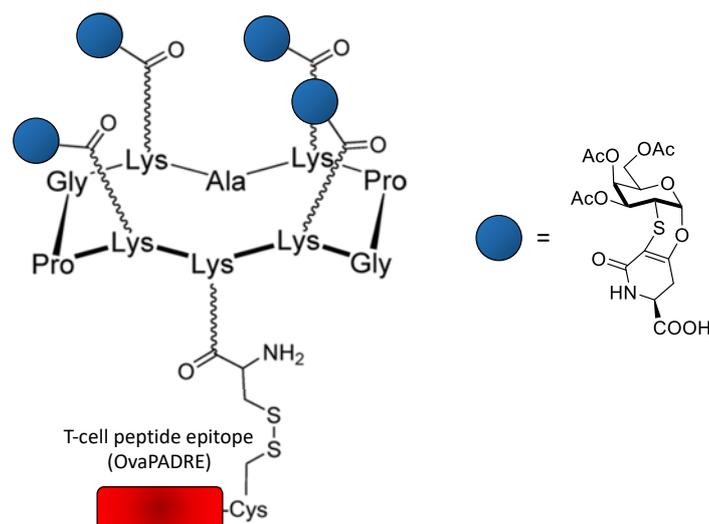
Figure 4.9 Fully synthetic vaccine composed of the Tn antigen mimetic conjugated to a RAFT scaffold and a OvaPADRE as T-cell epitope.

Diverse *N*-acyl fluorinated Tn saccharide analogs were used as vaccines in mice. Modified derivatives shown in Figure 4.10 were conjugated to the carrier protein CRM197[88–90] and their immunogenicity was tested.[91] *N*-monofluoroacetyl compound elicited higher IgG titers than the natural moiety, and also the other conjugates. In addition, antibodies generated by the aforementioned modified vaccine are able to recognize the tumor cells that express the natural Tn antigen on their surface.

R = CH$_3$CO-
R = CH$_3$CH$_2$CO-
R = CF$_3$CO-
R = CHF$_2$CO-
R = CH$_2$FCO-

Figure 4.10 Different N-acyl fluorinated Tn saccharide analogs.

Finally, reported by our group, a fragment of MUC1 tandem repeat containing the synthetic amino acid α-GalNAc-MeSer, together with a Pam$_3$Cys lipopeptide and a T-helper cell epitope, comprised the first unnatural three-component vaccine ever tested[77] (Figure 4.11). This construction elicited humoral and cellular immune response. Moreover, antisera obtained from immunizations with this vaccine displayed recognition of MUC1 expressed in tumor cells. Yet, it did not improve the effectiveness of its natural counterpart,[44] which includes a Thr residue instead of the unnatural amino acid, MeSer-containing vaccine presents better stability in serum.
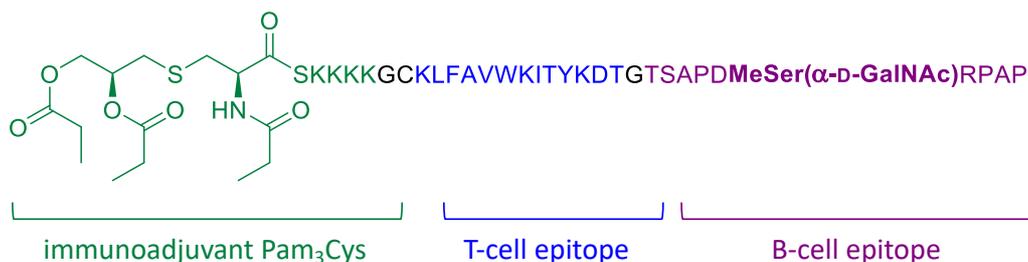
Figure 4.11 Three-component non-natural vaccine.

Aiming to provide with a rational explanation of the immune response elicited by the later vaccine, we decided to study the conformational behavior of its B cell epitope, that is the MUC1 glycopeptide, and how the unnatural moiety influences the presentation of the glycopeptide.

Thus, the shorter fragment of this mucin mimicry, PDMeSer*R, was synthesized and its conformational behavior studied by combining NMR experiments and MD-tar simulations.

According to these experiments, side chain of the MeSer is rather flexible and the glycosidic linkage adopts a staggered conformation, presenting a value around 80º for the ϕ torsional angle and close to 180º for the ψ dihedral angle. This outcome indicates that the conformational preferences of GalNAc-MeSer are similar to those reported for the Tn-Ser derivative.[92]

In summary, the MeSer derivative displays a higher flexibility at the sugar moiety, distorting the native conformation of the glycopeptide, which probably leads to a decrease of the recognition by the immune system.

In addition, from our previous research,[93] we can infer the key binding interactions in the bound state between the MUC1 epitopes APDT*RP (Figure 4.12) or APDS*RP and the anti-MUC1 antibody SM3 from the analysis of the X-ray structures.
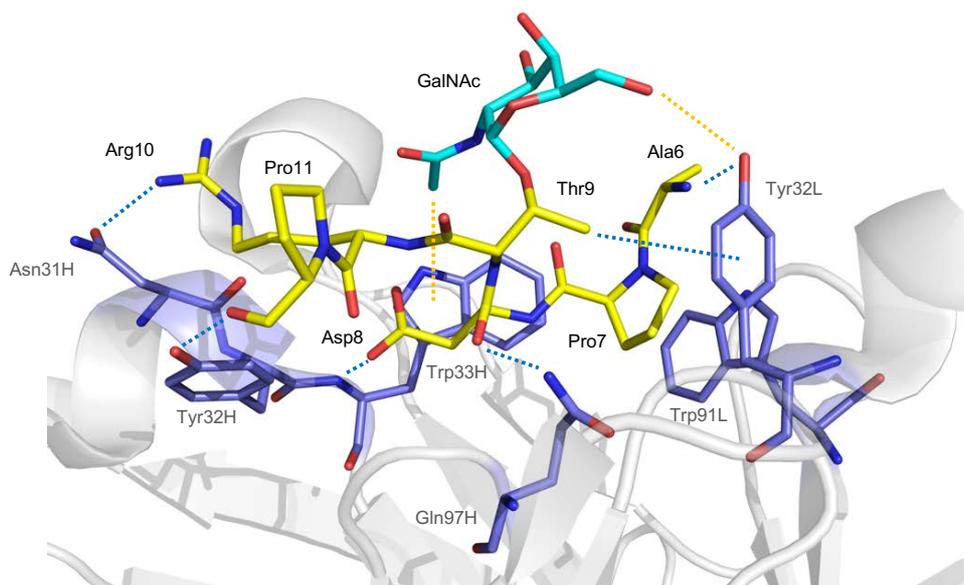
Figure 4.12 Key binding interactions of APDT*RP with SM3 antibody (PDB id: 5a2k). Dashed orange lines represent GalNAc-SM3 contacts and dashed blue lines peptide backbone-SM3 hydrogen bonds.

This analysis reveals that the glycosidic linkage on threonine glycopeptide adopts an 'eclipsed' conformation, allowing the formation of an intermolecular hydrogen bond between the O6 group of the sugar moiety and the hydroxyl group of Tyr32L of the antibody. Besides, we detected two CH/π interactions, between the methyl group of the GalNAc unit and the β-methyl group of Thr with the aromatic ring of Trp 33H and Tyr32L, respectively (Figure 4.12 orange dashed lines).

Conversely, the GalNAc unit of APDS*RP adopts a different conformation, characterized by ψ close to -90 degrees. This conformation, rarely populated in water, impedes the occurrence of stabilizing contacts between the sugar moiety and the antibody. This unusual conformer cannot be displayed by the threonine residue due to the steric hindrance that its β-methyl group prompts with the sugar unit in this conformation.

Considering the divergence between serine and threonine, both in the geometry of the glycosidic linkage[93,94] and in the affinity towards SM3 antibody,[93] we decided to focus on the influence that the β-methyl group of the Tn-Thr antigen brings to the overall glycopeptide, and the effect that it might provide regarding to antibody recognition. These outcomes will render useful information regarding the appropriate design of coming anti-cancer vaccines.

4.2. Objectives

In this chapter, glycosylated Thr9 within the MUC1 20-amino acids tandem repeat would be substituted by diverse unnatural moieties.

Firstly, we have studied the structural differences caused by the substitution of the β-methyl group at Thr, maintaining sugar moiety. To this purpose, the unnatural amino acids *allo*-threonine (*allo*-Thr), which is the epimer of Thr at the Cβ, and (2S,3R)-hydroxy-L-norvaline (Hnv), which bears an ethyl-by-methyl substitution at the aforementioned position, have been employed.

Both adequately protected *allo*-Thr and Hnv were glycosylated and incorporated within the peptide backbone of MUC1-like (glyco)peptides following SPPS methodology. NMR and MD-tar structural studies of the unnatural variants were then accomplished in solution. Furthermore, binding affinity studies were performed with two anti-MUC1 antibodies, SM3 and VU3C6, employing microarray and biolayer interferometry techniques. Also, dissociation constants were measured at diverse temperatures by surface plasmon resonance (SPR), allowing to carry out thermodynamic parameters estimations. Finally, crystal structure of the glycosylated Hnv derivative was studied complexed to the SM3 antibody to observe the structure in the associated state.

Secondly, maintaining the Thr residue, GalNAc moiety is substituted by a sp^2-iminosugar derivative. This derivative is engaged within MUC1 mucin backbone, coupled to the KLH carrier protein and tested as a cancer vaccine in mice.

4.3. Modifications in the amino acid

4.3.1. Synthesis of unnatural glycosyl amino acids

The strategy used to obtain **Tn-Hnv** (Scheme 4.6) and **Tn-*allo*Thr** (Scheme 4.7) building blocks involved the Koenigs-Knorr methodology described in Chapter 3. (See also experimental details in *Experimental section 8.10.1.).*
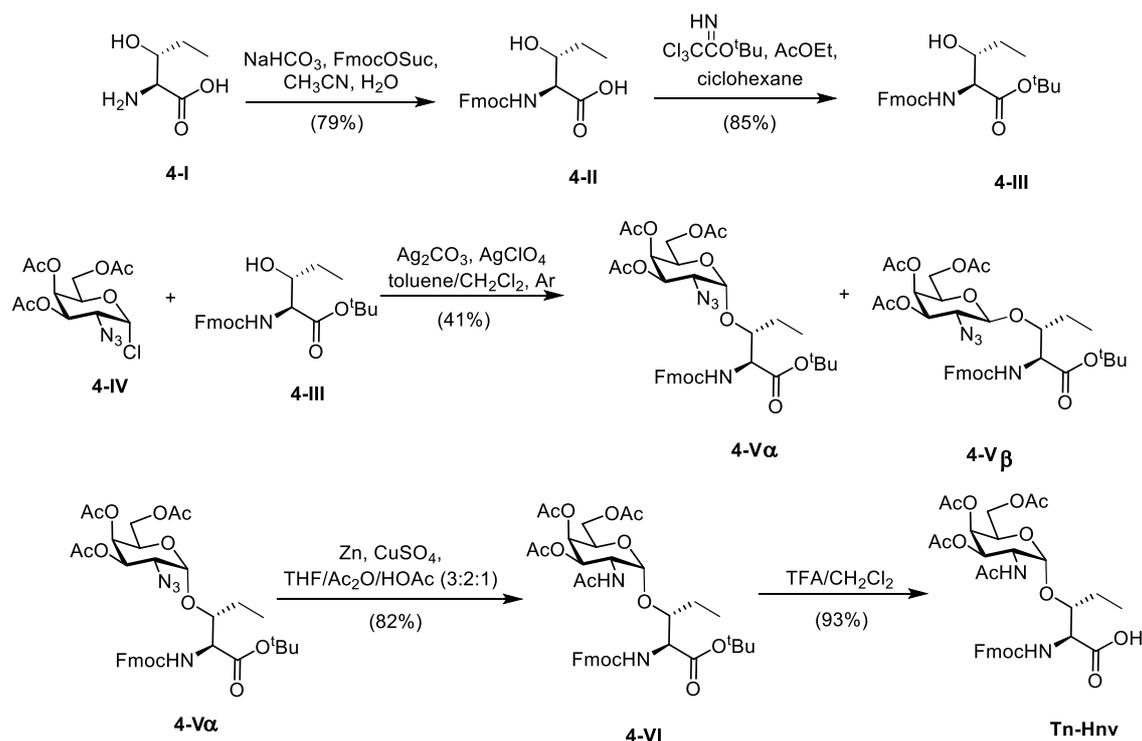
*Scheme 4.6 Synthetic route followed to obtain the **Tn-Hnv** derivative.*

Firstly, naked amino acid **4-I** was conveniently protected. Thus, the amino group was protected with Fmoc to give derivative **4-II** with good yield, which was then reacted with *tert*-butyl trichloroacetimidate in AcOEt/cyclohexane to yield derivative **4-III** as a white foam after purification by column chromatography. As mentioned before, Koenigs-Knorr reaction was employed, leading to a mixture of α and β anomers, **4-Vα** and **4-Vβ**, respectively in a 8:2 ratio. Once purified by column chromatography, compound **4-Vα** was transformed into **4-VI** following a one-pot reduction and subsequent acetylation of the resulting amino group of the carbohydrate moiety, giving the desired compound with an 82% overall yield in this step. Finally, the removal of the *tert*-butyl group in acid media yielded **Tn-Hnv**, which was ready-to-use in SPPS without further purification.

In parallel, **Tn-*allo*Thr** derivative was synthesized following the synthetic route shown in Scheme 4.7:
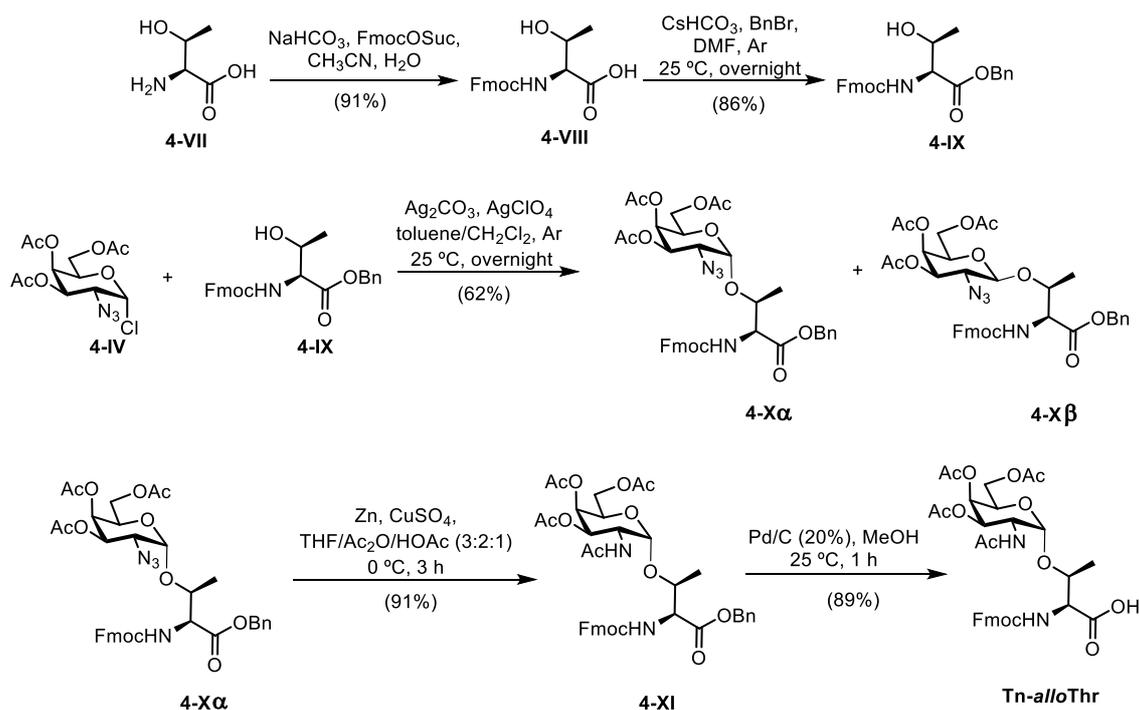
*Scheme 4.7 Synthetic route followed to obtain the **Tn-alloThr** derivative.*

Thus, the amino group was protected with Fmoc to give **4-VIII** with a 90% yield. This compound was protected as a benzyl ester obtaining the convenient protected amino acid **4-IX** as a white foam. Koenigs-Knorr glycosylation was then used, giving a mixture of both anomers (with a **4-Xα/4-Xβ** ratio = 7:3). After separation by column chromatography, the azido group in **4-Xα** was reduced and then acetylated to give compound **4-XI** in a good yield. Finally, the carboxylic acid was deprotected through a hydrogenation reaction with Pd/C to afford the desired **Tn-alloThr** derivative.

4.3.2. Synthesis of unnatural glycopeptides

The synthesis of (glyco)peptides shown in Figure 4.13 was carried out using the MW-SPPS protocol with a Rink amide MBHA resin, Fmoc methodology and commercially available side chain protected natural amino acids (See *Experimental section 8.10.2.*).

Figure 4.13 MUC1-like peptides and glycopeptides synthesized and studied.

As in previous cases, the building blocks with the unnatural residues, compounds **Tn-Hnv** and **Tn-*allo*Thr**, were manually coupled to reduce the number of equivalents employed and the reaction time was extended to three hours to increase the yield of this step. The reaction progress was checked by the classical Kaiser test.[95] Once the peptide sequence was synthesized, the *O*-acetyl groups of the carbohydrate were removed in a mixture of NH_2NH_2/MeOH (7:3). Afterwards, glycopeptides were detached from the resin, simultaneously with all the acid-labile side-chain protecting groups, using a cleavage cocktail consisting of a mixture of TFA/TIS/H_2O (95:2.5:2.5). Lastly, all glycopeptides were purified by semi-preparative HPLC, lyophilized and characterized.

4.3.3. NMR studies

The conformational analysis of short glycopeptides (**Thr*'**, ***allo*Thr*'** and **Hnv*'**) were studied by NMR spectroscopy. To this purpose, the full assignment of the protons in these compounds were carried out first by using COSY and HSQC experiments (*Supplementary information 9.2.*). Then, 2D-ROESY experiments in H_2O/D_2O (9:1) and at 20 °C were carried out.

4.3.3.1. ROESY experiments

Figure 4.14 shows amide regions of **Hnv*'** and ***allo*Thr*'** glycopeptides. The most relevant cross-peaks of the unnatural amino acid with the sugar moiety have been

signaled. As can be observed in Figure 4.14, both spectra present ROE contacts between GalNAc amide proton and H2 and H3 of the sugar moiety, in agreement with a $^4C_1$ chair conformer for this sugar. In addition, medium size cross-peaks were observed between the NH of both Hnv and *allo*-Thr and their respective Hα and Hβ protons, which provides information related to the glycosidic geometry.
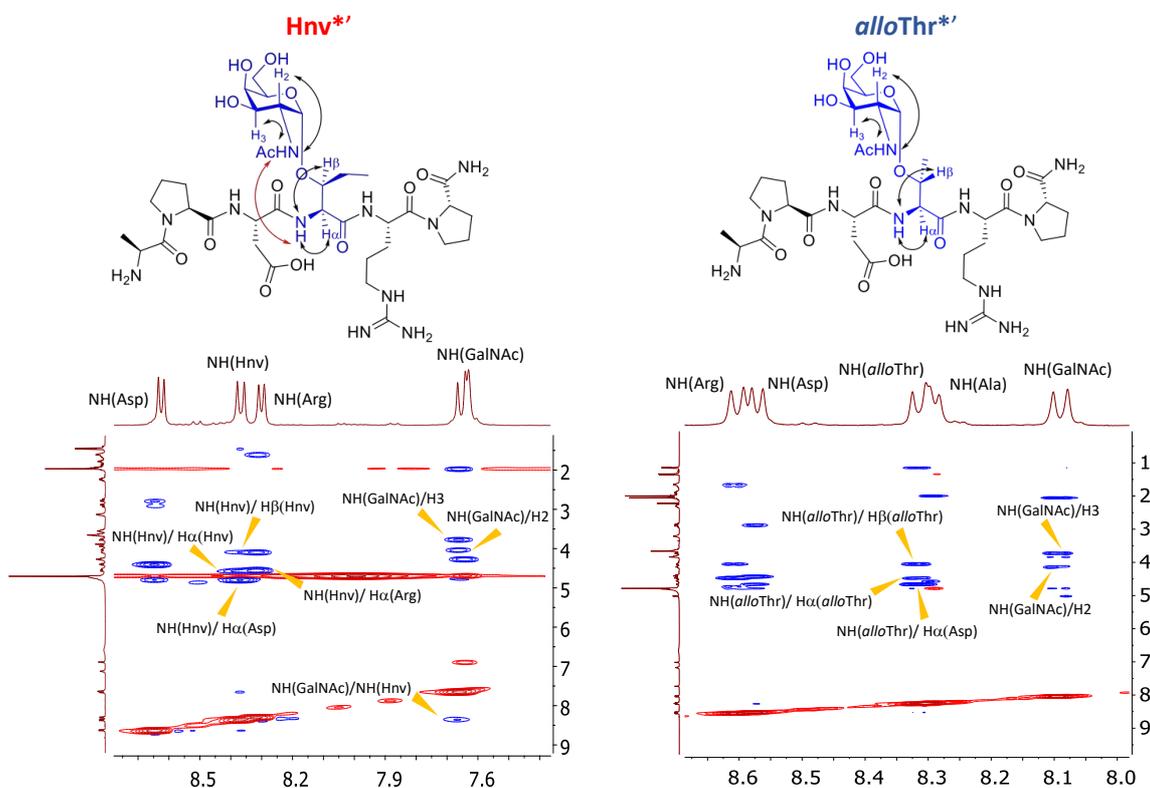

*Figure 4.14 ROESY spectrum of **Hnv*'** and **alloThr*'** glycopeptides.*

Interestingly, in glycopeptide **Hnv*'**, an extra medium-size ROE contact, highlighted in red in the structure, has been detected between the Hnv amide proton and GalNAc amide proton. As in natural variants, this cross-peak contact is representative of the 'eclipsed' conformation of the glycosidic linkage.[94,96] Conversely, this cross-peak has not been observed in the ***allo*Thr*'** analogue, where the Cβ presents a (*S*)-configuration. This observation points towards the idea that the glycosidic bond of GalNAc-*allo*-Thr may display a staggered geometry, as observed in GalNAc-Ser derivatives.

4.3.4. MD simulations with time-averaged restraints (MD-tar) or experiment-guided MD simulations

Those H-H distances relevant for the conformational analysis and obtained from the ROESY spectra, together with $^3J_{H-H'}$ distances, were used as experimental restraints in experiment-guided MD simulations. We use time-averaged restraints to avoid the production of virtual conformations and to obtain a theoretical ensemble in the free-state of these glycopeptides able to reproduce the experimental data.
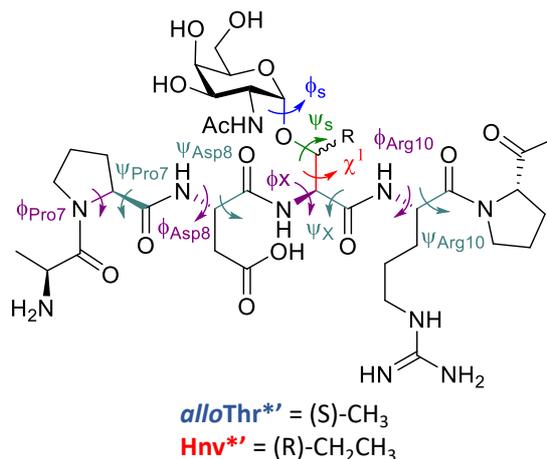
alloThr*' = (S)-CH$_3$
Hnv*' = (R)-CH$_2$CH$_3$

*Figure 4.15 Representation of torsional angles in **alloThr*'** and **Hnv*'** derivatives.*

The distribution of the backbone (represented by dihedral ϕ and ψ in Figure 4.15), for both glycopeptides is shown in Figure 4.16.
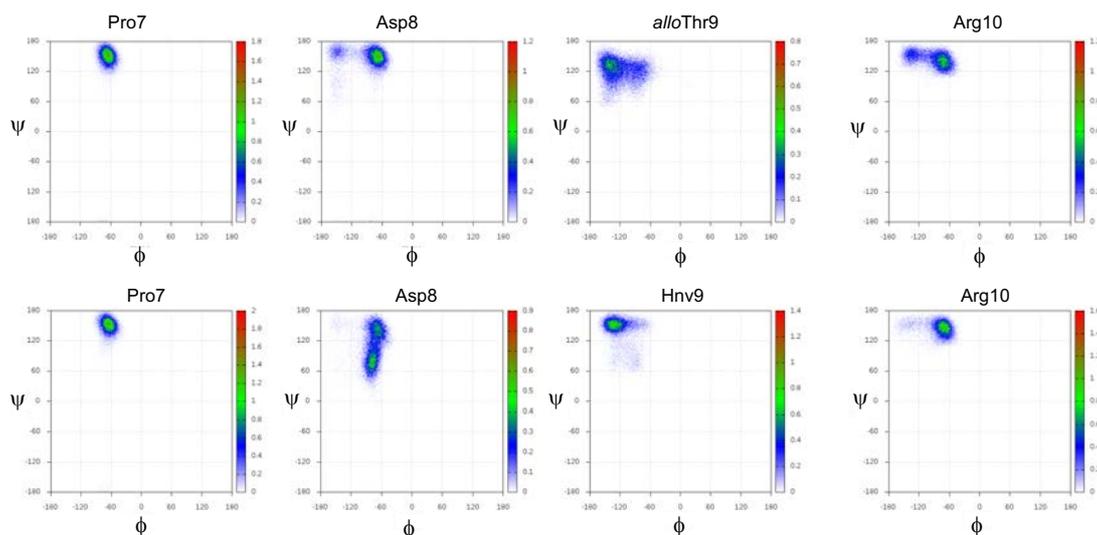

Figure 4.16 ϕ and ψ torsional angles for the unnatural glycopeptides
allo*Thr*' (upper panel) and **Hnv*'** (lower panel).*

While an extended conformation is the most populated for glycopeptide **alloThr*'**, the glycopeptide containing Hnv displays also a β-turn conformation for the Asp residue, as it occurs in the Thr-containing variant.[93]
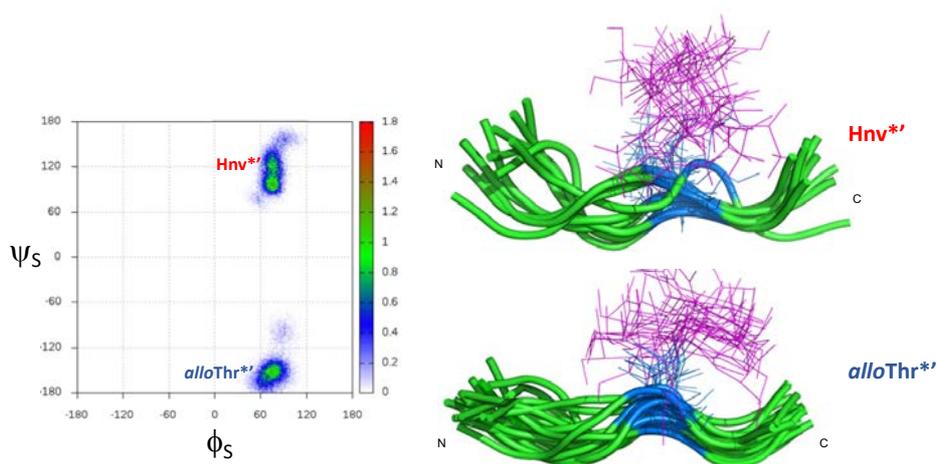
*Figure 4.17 ϕs and ψs torsional angles for the glycosidic linkage (left panel) and MD-tar derived conformational ensembles (right panel) of **allo**Thr' and **Hnv'** derivatives.*

Concerning the glycosidic linkage (Figure 4.17), in both cases ϕ_S adopts a value around 60°, in agreement with the exo-anomeric effect.[97,98] The value for the ψs torsional angle in the **Hnv*'** compound (around 120°) is comparable to that observed in its natural counterpart; hence the unnatural moiety presents the same structural behavior, a eclipsed conformation for the glycosidic linkage. Conversely, the ***allo*Thr*'** glycopeptide displays a value close to −180°, with the sugar adopting a parallel orientation of the carbohydrate with respect to the peptide backbone. This 3D disposition resembles the one displayed by the GalNAc-Ser glycosidic linkages. The value of the ψ_S is negative due to the change on the configuration at Cβ of the glycosylated amino acid.

The conformational ensembles derived from the simulations of both glycopeptides are shown in Figure 4.17 right panel, with the peptide backbone as a grey ribbon, except for the unnatural amino acid that is highlighted in blue, and the GalNAc moiety in purple.

Additionally, the presence of a high-water density placed between NH-GalNAc and NH-Hnv has been detected (Figure 4.18). This water pocket, equivalent to that observed in the GalNAc-Thr derivative, reinforces the idea that peptide backbone and glycosidic linkage of **Thr*'** and **Hnv*'** are comparable.
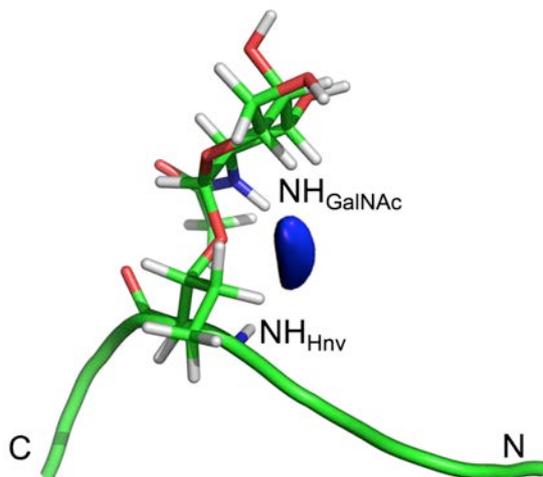
*Figure 4.18 MD-tar derived conformation of **Hnv*'** showing a water pocket between NH-GalNAc and NH-Hnv.*

In conclusion, as previously observed in NMR spectra, **Hnv*'** short epitope displays a conformation in solution similar to that observed in the **Thr*'** variant. On the contrary, ***allo*Thr*'** derivative presents a glycosidic bond that behaves similarly to the natural GalNAc-Ser counterpart.

Next, the conformation of the **Hnv*'** derivative was also studied bound to the anti-MUC1 antibody scFv-SM3 fragment by X-ray crystallography as well as STD experiments.

4.3.5. Crystallographic studies

Crystals of the complex scFv-SM3:**Hnv*'** were grown by sitting drop diffusion at 18 °C (see *Experimental section 8.8* and *8.10.3.*), as previously described,[93] allowing the acquisition of a structure at high resolution (<2.0 Å). Crystal structure of **Hnv*'** fragment in complex with scFv-SM3 is shown in Figure 4.19.
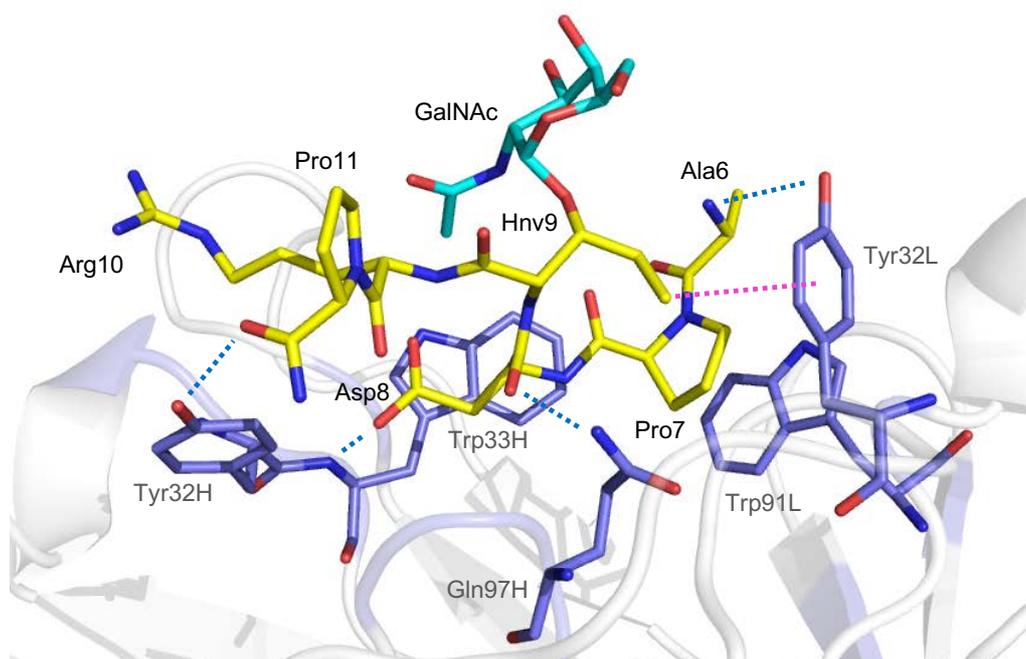
*Figure 4.19 Key binding interactions of **Hnv*'** with scFv-SM3 fragment. Dashed line in pink indicates the CH/π interaction between the β-ethyl group of the Hnv residue and the aromatic ring of Tyr32L of SM3 antibody and dashed blue lines indicate antigen-antibody hydrogen bonds. Peptide backbone carbon atoms appear in yellow, GalNAc carbon atoms are shown in cyan and carbon atoms of key residues of the antibody are highlighted in purple and the rest of the antibody is show as grey ribbons.*

The stabilizing contacts found in this complex are mainly intermolecular hydrogen bonds, many of them water-mediated (see *Experimental section 8.10.3.*), as well as CH/π interactions. Contrary to previous observations on its Thr counterpart,[93] GalNAc moiety, which is also exposed to the solvent, does not stablished a hydrogen bond with Tyr32L. Its methyl group is, however, engaged in a CH-π interaction as in the threonine variant.

Regarding to the peptide backbone, **Hnv*'** derivative shows resemblance to the threonine derived glycopeptide. The unnatural peptide is involved in stacking interactions of Pro7 with indole ring of Trp91L and the side chains of Asp8 and Arg10 are also hydrophobically engaged with Trp33H and Tyr32H, respectively. In addition, several hydrogen bonds between the peptide fragment and the antibody are found (dashed blue lines in Figure 4.19). For instance, the carbonyl groups of Hnv9 and Pro11 form hydrogen bonds with Gln97H and Tyr32H, respectively, and the NH group of Ala6 with Tyr32L.

Most important is the CH-π interaction established between the β-ethyl group of the Hnv residue and the aromatic ring of Tyr32L on the SM3 antibody. As expected, elongation of the alkyl group at Cβ with respect to Thr results in a shortening of the hydrophobic contact in the bound state, from 4.5Å to 4.2Å. However, as previously mentioned, this substitution can be the responsible for the disappearance of the hydrogen bond between the hydroxymethyl group of GalNAc and the side chain of Tyr32L, observed in the natural derivative.

Glycosidic bond angles in the bound state are similar. Thus, ψ_S torsional angle is 81° in the **Hnv*'** derivative, whereas the same angle for the glycosidic bond on threonine variant is 91°. Values of ϕ_S and χ_1 in the unnatural glycopeptide are 65° and 50°, respectively, also close to those of its natural counterpart.

In conclusion, changing the β-methyl group on the threonine for a slightly longer aliphatic chain, such as a β-ethyl group, involves the decrease of some intermolecular interactions. This loss would apparently be counteracted by the enhancement of the CH/π interaction between the β-ethyl group of Hnv residue and the aromatic ring of Tyr32L of SM3 antibody.

4.3.6. STD experiments

Saturation transfer difference (STD) NMR is a technique widely employed to characterize and identify binding contacts between a ligand and its receptor in solution. It is very sensitive for weak to medium binders, as it occurs in our antigen-antibody complexes, and highly accurate for detecting the hydrogen atoms of a ligand that are in close contact with those of the receptor.

Actually, previous examples of STD experiments have been reported regarding to the glycosylated MUC1 epitope PDT(α-*O*-GalNAc)RP bound to either SM3[99] or VU3C6[100] anti-MUC1 antibodies. Briefly, the results from both works point out to different recognition patterns over the peptide backbone of this epitope. The DTRP fragment presents medium STD signals, with an enhanced signal for Pro7, when bound to SM3. Conversely, in complex with VU3C6, the higher STD signals correspond to the TRP motif.

In relation to antibody-carbohydrate contacts, in the case of the SM3 antibody, the signals of the GalNAc ring protons are of lower intensity than each of the amino acids and only the *N*-acetyl methyl group presents a high STD effect. The interaction between GalNAc and VU3C6 is verify through a strong STD signal in the H2 proton and a medium intensity signal in the *N*-acetyl methyl group, while the rest of the sugar protons present low STD signals.

Although, we could not obtain a good STD spectrum for **Hnv*'** bound to SM3, we were able to study by STD experiments the binding of this unnatural glycopeptide with VU3C6. We compared the epitope binding of the glycopeptide with that reported for the natural variant.[100]
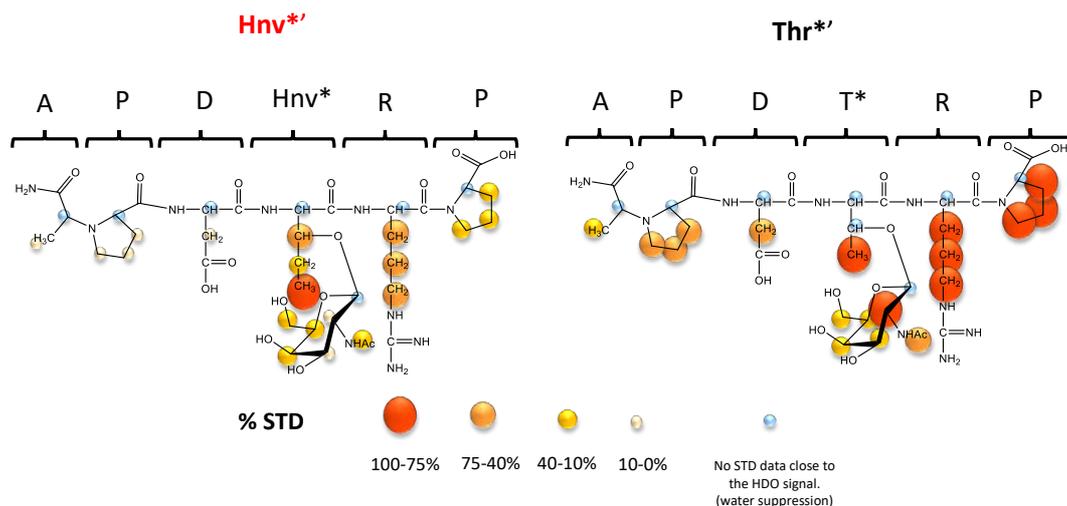
Figure 4.20 STD-derived epitope mapping of both glycopeptides with VU3C6 antibody.

Epitope mapping of the **Thr*'** and **Hnv*'** derivatives, presented in Figure 4.20, showed a strong STD effect for the RP amino acids and the β-methyl group of GalNAc-Thr moiety, denoted as T*. The same contacts, but of lower intensity, were observed in the unnatural epitope, except for the β-ethyl group of the Hnv residue, which maintains a high STD signal. STD signals over the GalNAc moiety of the **Hnv*'** glycopeptide show a similar binding profile compared to that of the natural epitope, presenting all protons of them a medium intensity. The most remarkable difference appears on H2: whereas STD effect is quite high in the **Thr*'** glycopeptide, it disappears in the non-natural derivative.

In conclusion, STD contacts of **Hnv*'** with the anti-MUC1 antibody VU3C6 present a similar binding patter than the **Thr*'** epitope, but of lower intensity; excluding the terminal methyl of the ethyl group, whose STD signal is comparable to that observed in the natural epitope. Therefore, peptide conformation appears to be quite like that observed in the natural derivative. Besides, the interaction of the β-ethyl group with the antibody seems to be equivalent to the β-methyl group in the **Thr*'** derivative.

4.3.7. Affinity assays

We tested the impact of the chemical nature of the unnatural residues in the binding affinity to anti-MUC1 antibodies. To this end we employed biolayer interferometry and microarray techniques with peptides and glycopeptides containing the tandem repeat sequence of MUC1 and antibodies SM3 and VU3C6. This study was completed with a Surface Plasmon Resonance (SPR) assay conducted with glycopeptide **Hnv*'** and the antibody SM3.

4.3.7.1. *Microarray assays*

Microarray assays were performed to qualitatively estimate the affinity of derivatives shown in Figure 4.21 against the commercially available antibodies SM3 and VU3C6.

Thr   AHGVTSAPDTRPAPGSTAPPA

Thr*   AHGVTSAPDTRPAPGSTAPPA

*allo*Thr   AHGVTSAPD◇RPAPGSTAPPA

*allo*Thr* AHGVTSAPD◇RPAPGSTAPPA

Hnv   AHGVTSAPD◆RPAPGSTAPPA

Hnv*   AHGVTSAPD◆RPAPGSTAPPA

Figure 4.21 (Glyco)peptides subjected to affinity assays.

These (glyco)peptides were spotted on microarray slides at different concentrations, from 500 µM to 15.6 µM. (See *Experimental section 8.6.).* The fluorescent image scans of both anti-MUC1 antibodies, SM3 and VU3C6, at 50 µg/mL, are shown in Figure 4.22.

Figure 4.22 Fluorescent image scan of SM3 and VU3C6 (50 mg/mL both).

According to these assays, antibody SM3 recognizes only (glyco)peptides where Thr has been substituted by Hnv. In these cases the fluorescence intensity is a bit higher for the glycosylated derivatives, which is in agreement with the binding studies previously reported by our group.[93] Conversely, compounds carrying a *allo*-Thr residue are not recognized by SM3 at detectable levels. Conversely, VU3C6 antibody shows binding with all the unnatural derivatives studied in this work. Fluorescence intensity is slightly higher for the glycosylated peptides, which qualitatively present an antibody recognition comparable to the **Thr*** derivative. This finding suggests a different recognition profile

for both anti-MUC1 antibodies, as previously inferred from the STD experiments (see Section 4.3.6).

In order to obtain a quantitative affinity value, these (glyco)peptides were subjected to biolayer interferometry experiments.

4.3.7.2. Biolayer interferometry (BLI) assays

To study binding affinity of the MUC1 antigens shown previously in Figure 4.21 with the scFv-SM3 antibody, the latter was immobilized on amine-reactive biosensors to perform BLI experiments. (*Experimental section 8.7* and *8.10.5.*). Dissociation constants (K_D) for each glyco(peptide) are shown in Figure 4.23.
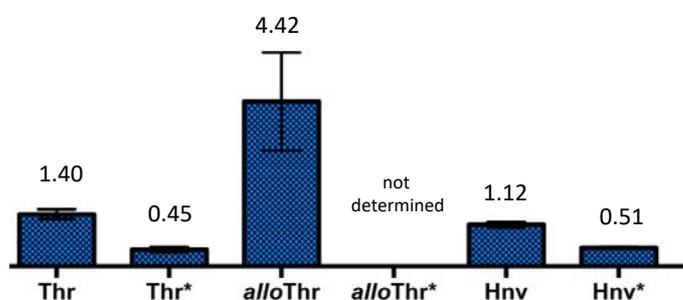
Figure 4.23 K_D constants (μM) observed for the diverse (glyco)peptides obtained by BLI experiments.

The BLI results highlight that both glycosylated and naked peptide derived from the unnatural Hnv amino acid show lower K_D values (that is, higher affinity) than their natural counterparts, **Thr** and **Thr***. Of note, derivative **Hnv*** displayed *ca.* 3-fold enhancement with respect to the naked Thr-variant.

In contrast, the binding of the *allo*-Thr derived peptide is *ca.* 3-fold lower than that presented by the natural derivative. Moreover, the affinity of *allo***Thr*** derivative could not be determined, likely due to its low affinity. This outcome is line with the result derived from antibody SM3 microarray experiments presented in previous section.

4.3.7.3. Surface Plasmon Resonance (SPR)

Herein, a thermodynamic analysis of the antigen-antibody binding was performed using SPR technique.[101–103] It was applied to **Thr*'** and **Hnv*'** derivatives; additionally, the natural non-glycosylated peptide APDTRP, denoted as **Thr'**, was also compared to observe the effect of the glycosidic moiety. The anti-MUC1 antibody chosen for this experiment was the commercially available SM3 antibody.

In the context of ligand-receptor binding, the affinity of the interaction can be represented as the binding free energy difference, ΔG, between the associated and

unassociated states of the interacting molecules. It is related to the dissociation constant, K_D through Equation 4.1.

$$\Delta G = RT \ln K_D$$

Equation 4.1

From the aforementioned expression, it follows that high-affinity interactions (low K_D values) would yield into a negative free binding difference. The forces driving the interaction are comprised by two terms, enthalpy change (ΔH) and entropy change (ΔS), Equation 4.2:

$$\Delta G = \Delta H - T\Delta S$$

Equation 4.2

On the one hand, the change in enthalpy refers to the heat emitted or absorbed due to the formation or disruption of hydrogen bonds network surrounding both reactants, as well as intramolecular hydrophobic interactions, such as van der Waals interactions and salt bridges, and how they vary from free to bound state. Hence, a negative change in enthalpy would favor the binding. On the other hand, entropy is defined as the "disorder" of the system or, in other words, as the measure of the change in the degree of freedom of mobility, rotation or vibration of the interacting molecules and the surrounding solvent. Roughly, it can be represented as the flexibility of the (glyco)peptides; a higher ligand mobility is expected in the free state, which is reduced when the derivative bounds to the receptor. Therefore, a flexible (glyco)peptide would yield into a higher entropy penalty than a more rigid variant than present the bioactive conformation in solution. However, other players, such as water molecules must be considered. Indeed, solvent reorganization modifies ΔS when it is displaced from the binding groove to the bulky solvent or trapped between two interacting surfaces. For instance, the displacement of water molecules from a hydrophobic surface would favor entropy.

To estimate the thermodynamic parameters associated to the binding of the studied glycopeptides to antibody SM3, the dissociation constant K_D was obtained at different temperatures for each complex (see *Experimental Section 8.10.6.*). Afterwards, thermodynamic parameters, such as enthalpy and entropy changes, were determined by a linear regression combining Equation 4.1 and Equation 4.2 in the following expression, denoted as the linear form of the van't Hoff equation:

$$R \ln K_D = \Delta H \cdot \frac{1}{T} - \Delta S$$

Equation 4.3

where R is the universal gas constant and T the temperature of each measurement. By plotting R·ln K_D against 1/T, ΔH and ΔS may be calculated, respectively, as the slope and Y-axis intercept of the resulting straight line.

Thermodynamic parameters for each antigen at physiological conditions (T = 310.15 K) are shown in Figure 4.24 as a bar chart. ΔG has been deduced as the sum of enthalpy and entropy parameters using Equation 4.2.
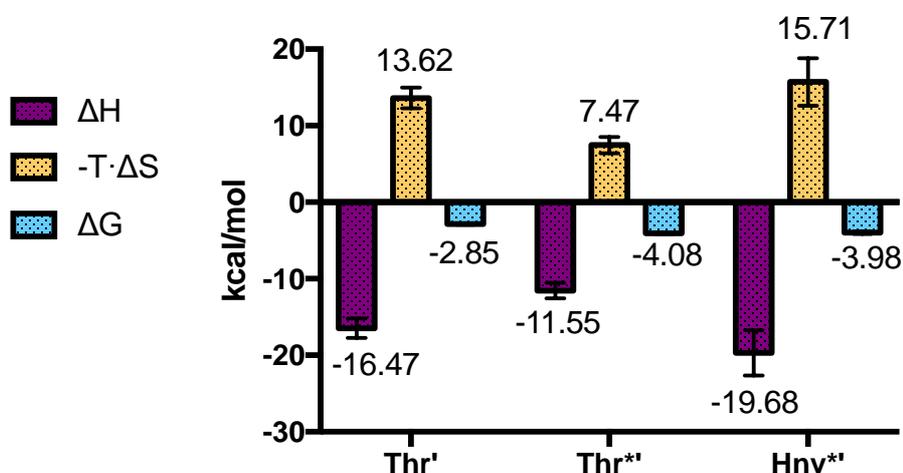

Figure 4.24 Thermodynamic parameters deduced from SPR for each MUC1 epitope at 310.15 K (all figures in Kcal/mol).

Most evident conclusion drawn from the thermodynamic results shown in Figure 4.24 is that bindings for all peptides are favored, meaning that free energy difference is negative, as a result of an unfavorable entropy contribution and a favorable enthalpy contribution. A closer look, however, would reveal clear differences.

Firstly, enthalpy contribution is quite in high in the three cases, being more significant for the unnatural glycopeptide **Hnv*'**. On the other hand, unfavorable entropy change may arise from the flexibility of each (glyco)peptide. Merely speculative, the most flexible molecule in the unbound state corresponds to **Hnv*'** derivative. As a result of the decrease in its high mobility, in particular the ethyl group, the most unfavorable entropic change arises from its binding. Even though, enthalpy and entropy contributions yield into a favorable binding affinity for **Hnv*'**.

Worth noticing also the influence of the carbohydrate on the natural threonine derivatives. Enthalpy change is more favored in the non-glycosylated derivative, suggesting an increase in the number of interactions in the bound state. On the contrary, entropy change is smaller for the glycopeptide, probably due to an enhanced stiffness caused by the carbohydrate moiety.

4.4. Modifications in the GalNAc moiety

In a previously reported work,[82] a battery of diverse Tn mimics incorporating sp^2-iminosugars were incorporated into the APDTRP epitope of MUC1 and their binding affinity to SM3 anti-MUC1 antibody was tested.
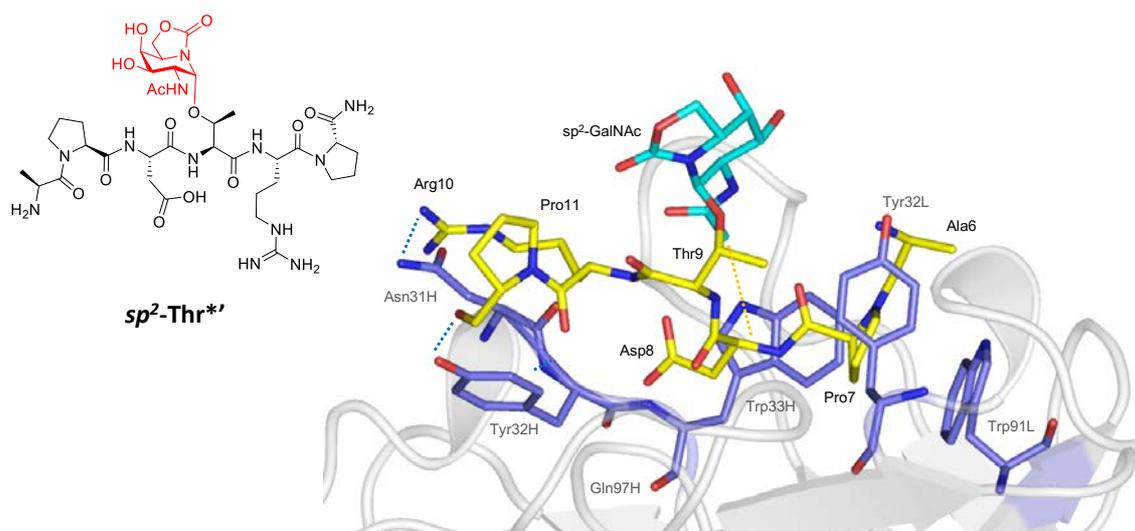
Figure 4.25 Structure of sp²-Thr' (left panel), together with sp²-Thr*':SM3 complex crystal structure (right panel).*

Attending to the binding affinity studies, the most promising candidate for a MUC1-based vaccine scaffold would be glycopeptide **sp²-Thr*'**, shown in Figure 4.25. Hence, to test its potential application as a cancer drug, the synthesis of a *sp²*-iminosugar containing vaccine was performed.[104]

Within this context, we synthesized the 20-amino acids tandem repeat of MUC1 mucin incorporating the *sp²*-iminosugar, as a B cell epitope; which was subsequently coupled to the KLH carrier protein and tested in mice.

4.5. Design of an unnatural cancer vaccine

4.5.1. Glycopeptide synthesis

The synthesis of unnatural glycopeptide **Cys-sp²-Thr***, together with the natural variants **Cys-Thr*** and **Cys-Thr** (Figure 4.26), was accomplished through MW-SPPS protocol with a Rink amide MBHA resin, Fmoc methodology and commercially available side chain protected amino acids (See *Experimental section 8.10.2.).* The Cys residue at the *N*-terminal position, instead of the typical Ala amino acid, permits further conjugation through a maleimide scaffold previously installed on the surface of the KLH carrier protein.

Cys-Thr   CHGVTSAPDTRPAPGSTAPPA

Cys-Thr*   CHGVTSAPDTRPAPGSTAPPA

Cys-sp²-Thr*   CHGVTSAPDTRPAPGSTAPPA

-*O*-GalNAc-Thr   α-*O*-sp²-GalNAc-Thr

Figure 4.26 MUC1-like peptide and glycopeptides synthesized and studied.

Following our own methodology, glycosylated building blocks were manually coupled to reduce the number of equivalents employed and the reaction time was extended compared to the in-machine synthesis and the end of the coupling was checked by Kaiser test.[95]

Once synthesized, we proceeded as previously described removing the *O*-acetyl groups of the carbohydrate. Glycopeptides were then detached from the resin, using a cleavage cocktail consisting TFA/TIS/H₂O/EDT (92.5:2.5:2.5:2.5), and finally purified by semi-preparative HPLC, lyophilized and characterized.

4.5.2. Protein carrier conjugation

Having successfully synthesized the aforementioned (glyco)peptides, those were conjugated to commercially available KLH protein and carrier through a Michael reaction.[105]

Maleimide heterobifunctional linker was coupled to the amino groups of Lys residues of each protein and is then reacted with the thiol component of Cys residue at the N-terminal position of MUC1 glycopeptides. (Figure 4.27, see *Experimental Section 8.10.7.*).

KLH conjugates, **KLH-Cys-Thr*** and **KLH-Cys-sp²-Thr***, were subsequently used to immunize the mice.

Figure 4.27 General procedure for (glyco)peptide conjugation to KLH.

We proved that the conjugation of glycopeptides **Cys-Thr*** and **Cys-sp²-Thr*** to KLH was efficient by immunoassay experiments using commercially available anti-MUC1 antibodies.

Once vaccine scaffolds were synthesized and confirmed, mice immunization was carried out.

4.5.3. Immunization experiments

Pairs of Balb/c mice were immunized four times at biweekly intervals with both vaccines, **KLH-Cys-sp2-Thr*** and **KLH-Cys-Thr***, as a control experiment with comparative purposes. These compounds were administered with complete or incomplete Freund's adjuvant depending on the immunization session.

One week after the last immunization, the mice were sacrificed, and serum harvested. The induction of antibodies was checked by analyzing this serum by ELISA assay, coating the plates with **Cys-Thr*** derivative (See *Experimental section 8.10.9.*).
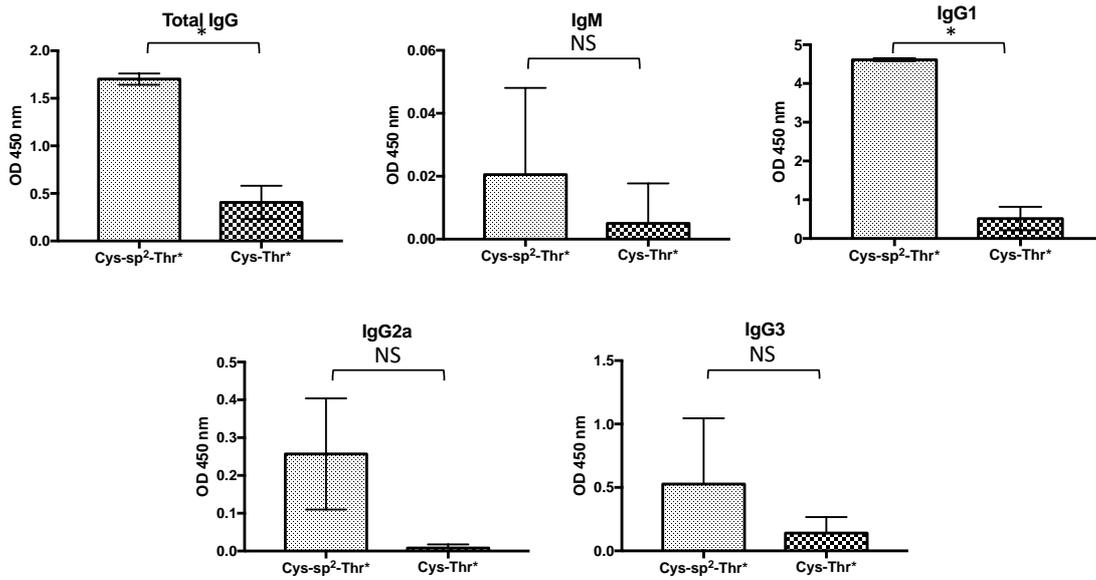
*Figure 4.28 Antibody response observed after immunization with vaccines **KLH-Cys-sp²-Thr*** and **KLH-Cys-Thr*** (n=2 both). Asterisk indicate statistically significant difference (* P < 0.05) and NS indicates no significant difference.*

As shown in Figure 4.28, humoral response was higher for those mice immunized with the unnatural vaccine **KLH-Cys-sp²-Thr*** in all cases. Of note, IgG1 sub-type and total IgG antibody cases show significantly different responses from those mice treated with the natural vaccine **KLH-Cys-Thr***. Additionally, IgM antibody values were low and not significantly different each other, indicating a T cell-mediated class-switching.

In parallel, a similar study was repeated coating ELISA plates with **Cys-Thr**, instead of the natural glycopeptide. It can be inferred from Figure 4.29 that the antibodies produced by the unnatural vaccine, in particular IgG1, could recognize glycosylated and unglycosylated MUC1 mucins in a similar way, indicating that antibodies recognize mainly the peptide and not the glycan moiety.
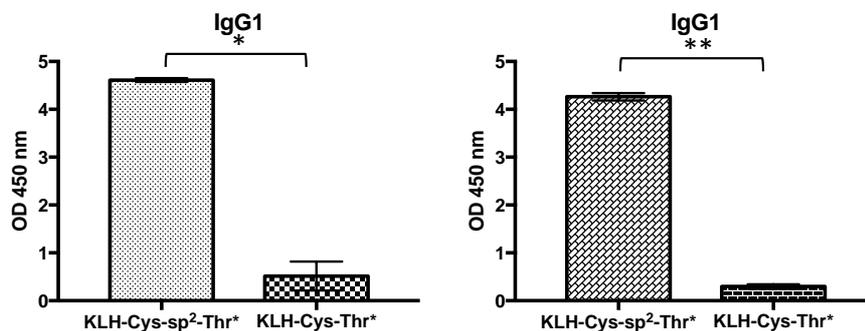
*Figure 4.29 IgG1 antibody response observed after immunization with vaccines **KLH-Cys-sp2-Thr*** and **KLH-Cys-Thr*** (n=2 both). ELISA plates were coated with **Cys-Thr*** (left) and **Cys-Thr** (right) derivatives. Asterisks indicate statistically significant difference (* P < 0.05, ** P < 0.01).*

4.6. Conclusions

We have studied the influence that the β-methyl group of Thr9 glycopeptide asserts on both, the conformation in the unbound and bound states of mucin-like glycopeptides and on the recognition by antibodies. To this purpose, Thr9 residue was replaced by its epimer at the Cβ, alloThr, and the Hnv amino acid, which displays a β-ethyl group instead of the β-methyl one.

Firstly, both unnatural amino acids were glycosylated following Koenigs-Knorr approach and incorporated within the peptide backbone. To check the effect of the sugar moiety, the naked unnatural variants were also prepared. Additionally, for comparative purposes, Thr derived peptides were obtained. Complete MUC1 tandem repeat, as well as shorter 6-amino acids variants, were synthesized incorporating these compounds.

The structure of the aforementioned short epitopes was studied in solution, combining NMR experiments, such as NOESY and STD, and experiment-guided MD simulations. The study suggests that Hnv derived glycopeptide presents a closer conformational behavior to the natural Thr compound than the *allo*-Thr glycosylated variant. In fact, the latter ones exhibited a behavior similar to that previously described for the Ser analog.

Binding affinity of the whole MUC1 mucin tandem repeat was then obtained from micro arrays and BLI experiments, showing a comparable K_D values for both Thr and Hnv derived (glyco)peptides, in contrast with their *allo*-Thr counterparts, which present low affinity for the SM3 anti-MUC1 antibody.

STD experiments showed a similar conformation for both epitopes. Besides, the recognition of the β-ethyl group of Hnv seems to be equivalent to the β-methyl group of the Thr derivative.

Thermodynamic parameters of short epitopes were studied by SPR technique, measuring the dissociation constant at different temperatures. The obtained results are in accordance to those derived from the microarrays and BLI experiments.

Finally, bound state of the Hnv derived glycopeptide complexed with scFv-SM3 antibody fragment was studied by X-ray crystallography. Hydrogen bonds, as well as hydrophobic contacts, were observed displaying a similar pattern of interactions than in the Thr variant. Most remarkable differences were associated to the ethyl-by-methyl substitution, which provokes a small deviation on the glycosidic linkage.

Herein, we can conclude that subtle modifications on MUC1 mucin structure may yield into great differences in antibody recognition. We have demonstrated that a change in the stereochemistry of Cβ on Thr leads to a decrease in binding affinity, whereas an ethyl-by-methyl substitution maintains conformational behavior, which is translated into a similar binding affinity.

On the other hand, an unnatural sp^2-iminosugar derived vaccine has been synthesized and tested in mice. The main conclusion drawn from the immunological study derived from this construction points out that a unnatural MUC1-based cancer vaccine that

emulates conformational features of its natural counterpart is able to slightly improve humoral response of the later.

Altogether, the results derived from the (glyco)peptides investigated herein set up the basis for a rational design of synthesized variants with enhanced immunological activity.

4.7. References

1. Westerlind, U. Synthetic glycopeptides and glycoproteins with applications in biological research. *Beilstein J. Org. Chem.* **8,** 804–818 (2012).
2. Westerlind, U. & Kunz, H. Synthetic vaccines from tumor-associated glycopeptide antigens. *Chimia (Aarau).* **65,** 30–34 (2011).
3. Gaidzik, N., Westerlind, U. & Kunz, H. The development of synthetic antitumour vaccines from mucin glycopeptide antigens. *Chem. Soc. Rev.* **42,** 4421–42 (2013).
4. Wilson, R. M. & Danishefsky, S. J. A vision for vaccines built from fully synthetic tumor-associated antigens: From the laboratory to the clinic. *J. Am. Chem. Soc.* **135,** 14462–14472 (2013).
5. Rabinovich, G. A., Kooyk, Y. Van & Cobb, B. A. Glycobiology of immune responses. *Ann. N. Y. Acad. Sci.* **1253,** 1–15 (2012).
6. Sarkar, S., Lombardo, S. A., Herner, D. N., Talan, R. S., Wall, K. A. & Sucheck, S. J. Synthesis of a single-molecule L-Rhamnose-containing three-component vaccine and evaluation of antigenicity in the presence of anti-L-rhamnose antibodies. *J. Am. Chem. Soc.* **132,** 17236–17246 (2010).
7. Glaffig, M., Palitzsch, B., Stergiou, N., Schüll, C., Strasburger, D., Schmitt, E., Frey, H. & Kunz, H. Enhanced immunogenicity of multivalent MUC1 glycopeptide antitumour vaccines based on hyperbranched polymers. *Org. Biomol. Chem.* **13,** 10150–10154 (2015).
8. Tam, J. P. Synthetic peptide vaccine design: synthesis and properties of a high-density multiple antigenic peptide system. *Proc Natl Acad Sci U S A.* **85,** 5409–5413 (1988).
9. Glaffig, M., Palitzsch, B., Hartmann, S., Schüll, C., Nuhn, L., Gerlitzki, B., Schmitt, E., Frey, H. & Kunz, H. A fully synthetic glycopeptide antitumor vaccine based on multiple antigen presentation on a hyperbranched polymer. *Chem. Eur. J.* **20,** 4232–4236 (2014).
10. Shiao, T. C. & Roy, R. Glycodendrimers as functional antigens and antitumor vaccines. *New J. Chem.* **36,** 324–339 (2012).
11. Chen, P.-G., Huang, Z.-H., Sun, Z.-Y., Li, Q.-Q., Chen, Y.-X., Zhao, Y.-F. & Li, Y.-M. Synthesis of an MUC1 glycopeptide dendrimer based on β-cyclo-dextrin by click chemistry. *Synlett* **28,** 1961–1965 (2017).
12. Huang, Z., Shi, L., Sun, Z., Cai, H., Chen, Y., Zhao, Y. & Li, Y. A totally synthetic, self-assembling, adjuvant-free MUC1 glycopeptide vaccine for cancer therapy. *J. Am. Chem. Soc.* **134,** 8730–8733 (2012).
13. Keil, S., Kaiser, A., Syed, F. & Kunz, H. Dendrimers of vaccines consisting of tumor-associated glycopeptide antigens and T-cell epitope peptides. *Synthesis (Stuttg).* **8,** 1355–1369 (2009).
14. Chun, C. K. Y. & Payne, R. J. Synthesis of MUC1 peptide and glycopeptide dendrimers. *Aust. J. Chem.* **62,** 1339–1343 (2009).
15. Richichi, B., Thomas, B., Fiore, M., Bosco, R., Qureshi, H., Nativi, C., Renaudet, O. & BenMohamed, L. A cancer therapeutic vaccine based on clustered Tn-antigen mimetics induces strong antibody-mediated protective immunity. *Angew. Chem. Int. Ed.* **53,** 11917–11920 (2014).
16. Fallarini, S., Brittoli, A., Fiore, M., Lombardi, G., Renaudet, O., Richichi, B. & Nativi, C. Immunological characterization of a rigid α-Tn mimetic on murine iNKT and human NK cells. *Glycoconj. J.* **34,** 553–562 (2017).
17. Thomas, B., Pifferi, C., Daskhan, G. C., Fiore, M., Berthet, N. & Renaudet, O. Divergent and convergent synthesis of GalNAc-conjugated dendrimers using dual orthogonal ligations. *Org. Biomol. Chem.* **13,** 11529–11538 (2015).
18. Pifferi, C., Berthet, N. & Renaudet, O. Cyclopeptide scaffolds in carbohydrate-based synthetic vaccines. *Biomater. Sci.* **5,** 953–965 (2017).

19. Jeon, I., Lee, D., Krauss, I. J. & Danishefsky, S. J. A new model for the presentation of tumor-associated antigens and the quest for an anticancer vaccine: A solution to the synthesis challenge via ring-closing metathesis. *J. Am. Chem. Soc.* **131,** 14337–14344 (2009).

20. Renaudet, O., Benmohamed, L., Dasgupta, G., Bettahi, I. & Dumy, P. Towards a self-adjuvanting multivalent B and T cell epitope containing synthetic glycolipopeptide cancer vaccine. *Chem. Med. Chem.* **3,** 737–741 (2008).

21. Keil, S., Claus, C., Dippold, W. & Kunz, H. Towards the development of antitumor vaccines: A synthetic conjugate of a tumor-associated MUC1 glycopeptide antigen and a tetanus toxin epitope. *Angew. Chem. Int. Ed.* **40,** 366–369 (2001).

22. Dziadek, S., Hobel, A., Schmitt, E. & Kunz, H. A fully synthetic vaccine consisting of a tumor- associated glycopeptide antigen and a T-cell epitope for the induction of a highly specific humoral immune response. *Angew. Chem. Int. Ed.* **44,** 7630–7635 (2005).

23. Westerlind, U., Hobel, A., Gaidzik, N., Schmitt, E. & Kunz, H. Synthetic vaccines consisting of tumor-associated MUC1 glycopeptide antigens and a T-cell epitope for the induction of a highly specific humoral immune response. *Angew. Chem. Int. Ed.* **47,** 7551–7556 (2008).

24. Cremer, G.-A., Bureaud, N., Piller, V., Kunz, H., Piller, F. & Delmas, A. F. Synthesis and biological evaluation of a multiantigenic Tn/TF-containing glycopeptide mimic of the tumor-related MUC1 glycoprotein. *Chem Med Chem* **1,** 965–968 (2006).

25. Alexander, J., Sidney, J., Southwood, S., Ruppert, J., Oseroff, C., Maewal, A., Snoke, K., Serra, H. M., Kubo, R. T., Sette, A. & Grey, H. M. Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. *Immunity* **1,** 751–761 (1994).

26. Wittrock, S., Becker, T. & Kunz, H. Synthetic vaccines of tumor-associated glycopeptide antigens by immune-compatible thioether linkage to bovine serum albumin. *Angew. Chem. Int. Ed.* **46,** 5226–5230 (2007).

27. Kunz, H. & Birnbach, S. Synthesis of O-glycopeptides of the tumor-associated Tn- and T-antigen type and their binding to bovine serum albumin. *Angew. Chem. Int. Ed.* **25,** 360–362 (1986).

28. Cai, H., Chen, M., Sun, Z., Zhao, Y., Kunz, H. & Li, Y. Self-adjuvanting synthetic antitumor vaccines from MUC1 glycopeptides conjugated to T-cell epitopes from Tetanus Toxoid. *Angew. Chem. Int. Ed.* **52,** 6106–6110 (2013).

29. Ragupathi, G., Koide, F., Livingston, P. O., Cho, Y. S., Endo, A., Wan, Q., Spassova, M. K., Keding, S. J., Allen, J., Ouerfelli, O., Wilson, R. M. & Danishefsky, S. J. Preparation and evaluation of unimolecular pentavalent and hexavalent antigenic constructs targeting prostate and breast cancer: A synthetic route to anticancer vaccine candidates. *J. Am. Chem. Soc.* **128,** 2715–2725 (2006).

30. Zhu, J., Wan, Q., Lee, D., Yang, G., Spassova, M. K., Ouerfelli, O., Ragupathi, G., Damani, P., Livingston, P. O. & Danishefsky, S. J. From synthesis to biologics: preclinical data on a chemistry derived anticancer vaccine. *J. Am. Chem. Soc.* **131,** 9298–9303 (2009).

31. Slovin, S. F., Ragupathi, G., Musselli, C., Olkiewicz, K., Verbel, D., Kuduk, S. D., Schwarz, J. B., Sames, D., Danishefsky, S., Livingston, P. O. & Scher, H. I. Fully synthetic carbohydrate-based vaccines in biochemically relapsed prostate cancer: clinical trial results with alpha-N-acetylgalactosamine-O-Serine/Threonine conjugate vaccine. *J Clin Oncol* **21,** 4292–4298 (2003).

32. Kagan, E., Ragupathi, G., Yi, S. S., Reis, C. A., Gildersleeve, J., Kahne, D., Clausen, H., Danishefsky, S. J. & Livingston, P. O. Comparison of antigen constructs and carrier molecules for augmenting the immunogenicity of the monosaccharide epithelial cancer antigen Tn. *Cancer Immunol. Immunother.* **54,** 424–430 (2005).

33. Kaiser, A., Gaidzik, N., Westerlind, U., Kowalczyk, D., Hobel, A., Schmitt, E. & Kunz, H.

Antitumor vaccines a synthetic vaccine consisting of a tumor-associated Sialyl-Tn-MUC1 tandem-repeat glycopeptide and tetanus toxoid: induction of a strong and highly selective immune response. *Angew. Chem. Int. Ed.* **48,** 7551–7555 (2009).

34. Gaidzik, N., Kaiser, A., Kowalczyk, D., Westerlind, U., Gerlitzki, B., Sinn, H. P., Schmitt, E. & Kunz, H. Synthetic antitumor vaccines containing MUC1 glycopeptides with two immunodominant domains — induction of a strong immune response against breast tumor tissues. *Angew. Chem. Int. Ed.* **50,** 9977–9981 (2011).

35. Stergiou, N., Glaffig, M., Jonuleit, H., Schmitt, E. & Kunz, H. Immunization with a synthetic human MUC1 glycopeptide vaccine against tumor-associated MUC1 breaks tolerance in human MUC1 transgenic mice. *Chem. Med. Chem.* **12,** 1424–1428 (2017).

36. Palitzsch, B., Gaidzik, N., Stergiou, N., Stahn, S., Hartmann, S., Gerlitzki, B., Teusch, N., Flemming, P., Schmitt, E. & Kunz, H. A synthetic glycopeptide vaccine for the induction of a monoclonal antibody that differentiates between normal and tumor mammary cells and enables the diagnosis of human pancreatic cancer. *Angew. Chem. Int. Ed.* **55,** 2894–2898 (2016).

37. Kaiser, A., Gaidzik, N., Becker, T., Menge, C., Groh, K., Cai, H., Li, Y., Gerlitzki, B., Schmitt, E. & Kunz, H. Fully synthetic vaccines consisting of tumor-associated MUC1 glycopeptides and a lipopeptide ligand of the Toll-like receptor 2. *Angew. Chem. Int. Ed.* **49,** 3688–3692 (2010).

38. Cai, H., Huang, Z.-H., Shi, L., Zhao, Y.-F., Kunz, H. & Li, Y.-M. Towards a fully synthetic MUC1-based anticancer vaccine: Efficient conjugation of glycopeptides with mono-, di-, and tetravalent lipopeptides using click chemistry. *Chem. Eur. J.* **17,** 6396–6406 (2011).

39. Shi, L., Cai, H., Huang, Z.-H., Sun, Z.-Y., Chen, Y.-X., Zhao, Y.-F., Kunz, H. & Li, Y.-M. Synthetic MUC1 antitumor vaccine candidates with varied glycosylation pattern bearing R/S-configured. *Chem. Bio. Chem.* **17,** 1412–1415 (2016).

40. Cai, H., Sun, Z. Y., Chen, M. S., Zhao, Y. F., Kunz, H. & Li, Y. M. Synthetic multivalent glycopeptide-lipopeptide antitumor vaccines: Impact of the cluster effect on the killing of tumor cells. *Angew. Chem. Int. Ed.* **53,** 1699–1703 (2014).

41. Cai, H., Sun, Z., Huang, Z., Shi, L., Zhao, Y., Kunz, H. & Li, Y.-M. Fully synthetic self-adjuvanting thioether-conjugated glycopeptide lipopeptide antitumor vaccines for the induction of complement-dependent cytotoxicity against tumor cells. *Chem. Eur. J.* **19,** 1962–1970 (2013).

42. Mcdonald, D. M., Wilkinson, B. L., Corcilius, L., Thaysen-andersen, M., Byrne, S. N. & Payne, R. J. Synthesis and immunological evaluation of self-adjuvanting MUC1-macrophage activating lipopeptide 2 conjugate vaccine candidates. *Chem. Commun.* **50,** 10273–10276 (2014).

43. Ingale, S., Wolfert, M. A., Gaekwad, J., Buskas, T. & Boons, G. Robust immune responses elicited by a fully synthetic three-component vaccine. *Nat. Chem. Biol.* **3,** 663–667 (2007).

44. 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. Immune recognition of tumor-associated mucin MUC1 is achieved by a fully synthetic aberrantly glycosylated MUC1 tripartite vaccine. *Proc. Natl. Acad. Sci. U S A.* **109,** 261–266 (2012).

45. Abdel-Aal, A.-B. M., Lakshminarayanan, V., Thompson, P., Supekar, N., Bradley, J. M., Wolfert, M. A., Cohen, P. A., Gendler, S. J. & Boons, G.-J. Immune and anticancer responses elicited by fully synthetic aberrantly glycosylated MUC1 tripartite vaccines modified by a TLR2 or TLR9 agonist. *Chem. Bio. Chem.* **15,** 1508–1513 (2014).

46. Supekar, N. T., Lakshminarayanan, V., Capicciotti, C. J., Sirohiwal, A., Madsen, C. S., Wolfert, M. A., Cohen, P. A., Gendler, S. J. & Boons, G.-J. Synthesis and immunological evaluation of a multicomponent cancer vaccine candidate containing a long MUC1 glycopeptide. *Chem. Bio. Chem.* **19,** 121–125 (2018).

47. Wilkinson, B. L., Day, S., Chapman, R., Perrier, S., Apostolopoulos, V. & Payne, R. J. Synthesis and immunological evaluation of self-assembling and self-adjuvanting

tricomponent glycopeptide cancer-vaccine candidates. *Chem. Eur. J.* **18,** 16540–16548 (2012).

48. Wilkinson, B. L., Malins, L. R., Chun, C. K. Y. & Payne, R. J. Synthesis of MUC1 – lipopeptide chimeras. *Chem. Commun.* **46,** 6249–6251 (2010).

49. Xiao, A., Zheng, X.-J., Song, C., Gui, Y., Huo, C.-X. & Ye, X.-S. Synthesis and immunological evaluation of MUC1 glycopeptide conjugates bearing N-acetyl modified STn derivatives as anticancer vaccines. *Org. Biomol. Chem.* **14,** 7226–7237 (2016).

50. Oberbillig, T., Mersch, C., Wagner, S. & Hoffmann-Röder, A. Antibody recognition of fluorinated MUC1 glycopeptide antigens. *Chem. Commun.* **48,** 1487–1489 (2012).

51. Yang, F., Zheng, X., Huo, C., Wang, Y., Zhang, Y. & Ye, X. Enhancement of the immunogenicity of synthetic carbohydrate vaccines by chemical modifications of STn antigen. *ACS Chem. Biol.* **6,** 252–259 (2011).

52. Hoffmann-Röder, A., Kaiser, A., Wagner, S., Gaidzik, N., Kowalczyk, D., Westerlind, U., Gerlitzki, B., Schmitt, E. & Kunz, H. Synthetic antitumor vaccines from Tetanus Toxoid conjugates of MUC1 glycopeptides with the Thomsen – Friedenreich antigen and a fluorine-substituted analogue. *Angew. Chem. Int. Ed.* **49,** 8498–8503 (2010).

53. Awad, L., Madani, R., Gillig, A., Kolympadi, M., Philgren, M., Muhs, A., Gérard, C. & Vogel, P. A C-linked disaccharide analogue of Thomsen – Friedreich epitope induces a strong immune response in mice. *Chem. Eur. J.* **18,** 8578–8582 (2012).

54. Kuberan, B., Sikkander, S. A., Tomiyama, H. & Linhardt, R. J. Synthesis of a C-Glycoside analogue of STn: An HIV- and tumor-associated antigen. *Angew. Chem. Int. Ed.* **42,** 2073–2075 (2003).

55. Bousquet, E., Spadaro, A., Pappalardo, M. S., Bernardini, R., Romeo, R., Panza, L. & Ronsisvalle, G. Synthesis and immunostimulating activity of a thioglycolipopeptide glycomimetic as a potential anticancer vaccine derived from Tn antigen. *J. Carbohydr. Chem.* **19,** 537–541 (2000).

56. Knapp, S. & Myers, D. S. Synthesis of alpha-GalNAc thioconjugates from an alpha-GalNAc mercaptan. *J. Org. Chem.* **67,** 2995–2999 (2002).

57. Thayer, D. A., Yu, H. N., Galan, M. C. & Wong, C. A general strategy toward S-linked glycopeptides. *Angew. Chem. Int. Ed.* **44,** 4596–4599 (2005).

58. Rojas-Ocáriz, V., Compañón, I., Aydillo, C., Castro-López, J., Jiménez-Barbero, J., Hurtado-Guerrero, R., Avenoza, A., Zurbano, M. M., Peregrina, J. M., Busto, J. H. & Corzana, F. Design of α-S-Neoglycopeptides derived from MUC1 with a flexible and solvent-exposed sugar moiety. *J. Org. Chem.* **81,** 5929–5941 (2016).

59. Galonic, D. P., Van der Donk, W. A. & Gin, D. Y. Oligosaccharide-peptide ligation of glycosyl thiolates with glycopeptide conjugates. *Chem. Eur. J.* **9,** 5997–6006 (2003).

60. Zhu, Y. & Van der Donk, W. A. Convergent synthesis of peptide conjugates using dehydroalanines for chemoselective ligations. *Org. Lett.* **3,** 3603–3606 (2001).

61. Aydillo, C., Compañón, I., Avenoza, A., Busto, J. H., Corzana, F., Peregrina, J. M. & Zurbano, M. M. S-Michael additions to chiral dehydroalanines as an entry to glycosylated cysteines and a sulfa-Tn antigen mimic. *J. Am. Chem. Soc.* **136,** 789–800 (2014).

62. Gutiérrez-Jiménez, M. I., Aydillo, C., Navo, C. D., Avenoza, A., Corzana, F., Jiménez-Osés, G., Zurbano, M. M., Busto, J. H. & Peregrina, J. M. Bifunctional chiral dehydroalanines for peptide coupling and stereoselective S-Michael Addition. *Org. Lett.* **18,** 2796–2799 (2016).

63. Urban, D., Skrydstrup, T. & Beau, J.-M. First synthesis of a C-glycoside anologue of a tumor-associated carbohydrate antigen employing samarium diiodide promoted C - glycosylation. *Chem. Commun.* 955–956 (1998).

64. Röhrig, C. H., Takhi, M. & Schmidt, R. R. Synthesis of a C-glycoside analog of the tumor-associated Tn antigen. *Synlett* **7,** 1170–1172 (2001).

65. Allen, J. R., Harris, C. R. & Danishefsky, S. J. Pursuit of optimal carbohydrate-based

anticancer vaccines: Preparation of a multiantigenic unimolecular glycopeptide containing the Tn, MBr1, and Lewis antigens. *J. Am. Chem. Soc.* **123,** 1890–1897 (2001).

66. Keding, S. J., Endo, A. & Danishefsky, S. J. Synthesis of non-natural glycosylamino acids containing tumor-associated carbohydrate antigens. *Tetrahedron* **59,** 7023–7031 (2003).

67. Keding, S. J., Atsushi, E., Biswas, K., Zatorski, A., Coltart, D. M. & Danishefsky, S. J. Hydroxynorleucine as a glycosyl acceptor is an efficient means for introducing amino acid functionality into complex carbohydrates. *Tetrahedron Lett.* **44,** 3413–3416 (2003).

68. Vichier-Guerre, S., Lo-man, R., Huteau, V., Dériaud, E., Leclerc, C. & Bay, S. Synthesis and immunological evaluation of an antitumor neoglycopeptide vaccine bearing a novel homoserine Tn antigen. *Bioorg. Med. Chem. Lett.* **14,** 3567–3570 (2004).

69. Marcaurelle, L. A., Shin, Y., Goon, S. & Bertozzi, C. R. Synthesis of oxime-linked mucin mimics containing the tumor-related Tn and Sialyl Tn antigens. *Org. Lett.* **3,** 3691–3694 (2001).

70. Peri, F., Cipolla, L., Rescigno, M., La Ferla, B. & Nicotra, F. Synthesis and biological evaluation of an anticancer vaccine containing the C-glycoside analogue of the Tn epitope. *Bioconjug. Chem.* **12,** 325–328 (2001).

71. Cipolla, L., Rescigno, M., Leone, A., Peri, F., La Ferla, B. & Nicotra, F. Novel Tn antigen-containing neoglycopeptides: synthesis and evaluation as anti tumor vaccines. *Bioorg. Med. Chem.* **10,** 1639–1646 (2002).

72. Lee, D. J., Mandal, K., Harris, P. W. R., Brimble, M. A. & Kent, S. B. H. A one-pot approach to neoglycopeptides using orthogonal native chemical ligation and click chemistry. *Org. Lett.* **11,** 5270–5273 (2009).

73. Lee, D. J., Harris, P. W. R. & Brimble, M. A. Synthesis of MUC1 Neoglycopeptides using efficient microwave-enhanced chaotrope-assisted click chemistry. *Org. Biomol. Chem.* **9,** 1621–1626 (2011).

74. Miller, N., Williams, G. M. & Brimble, M. A. Synthesis of fish antifreeze neoglycopeptides using microwave-assisted "Click Chemistry". *Org. Lett.* **11,** 2409–2412 (2009).

75. Wan, Q., Chen, J., Chen, G. & Danishefsky, S. J. A potentially valuable advance in the synthesis of carbohydrate-based anticancer vaccines through extended cycloaddition chemistry. *J. Org. Chem.* **71,** 8244–8249 (2006).

76. Corzana, F., Busto, J. H., Marcelo, F., García de Luis, M., Asensio, J. L., Martín-Santamaría, S., Sáenz, Y., Torres, C., Jiménez-Barbero, J., Avenoza, A., Peregrina, J. M. & Torres, C. Rational design of a Tn antigen mimic. *Chem. Commun.* **47,** 5319–5321 (2011).

77. 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. Mucin architecture behind the immune response: design , evaluation and conformational analysis of an antitumor vaccine derived from an unnatural MUC1 fragment. *Chem. Sci.* **7,** 2294–2301 (2016).

78. Tovillas, P., García, I., Oroz, P., Mazo, N., Avenoza, A., Corzana, F., Jiménez-Osés, G., Busto, J. H. & Peregrina, J. M. Tn antigen mimics by ring-opening of chiral cyclic sulfamidates with carbohydrate C1-S- and C1-O-nucleophiles. *J. Org. Chem.* **83,** 4973–4980 (2018).

79. Norgren, A. S., Norberg, T. & Arvidsson, P. I. Glycosylated foldamers: synthesis of carbohydrate-modified β3hSer and incorporation into β-peptides. *J. Pept. Sci.* **13,** 717–727 (2007).

80. Caputo, R., Cassano, E., Longobardo, L. & Palumbo, G. Chiral N-Protected beta-iodoamines from alpha-aminoacids: A general synthesis. *Tetrahedron Lett.* **36,** 167–168 (1995).

81. Karch, F. & Hoffmann-Röder, A. Synthesis of glycosylated β3-homo-threonine conjugates for mucin-like glycopeptide antigen analogues. *Beilstein J. Org. Chem.* **6,** 1–8 (2010).

82. Fernández, E. M. S., Navo, C. D., Martínez-Sáez, N., Gonçalves-Pereira, R., Somovilla, V. J., Avenoza, A., Busto, J. H., Bernardes, G. J. L., Jiménez-Osés, G., Corzana, F., Fernández,

J. M. G., Mellet, C. O. & Peregrina, J. M. Tn Antigen Mimics Based on sp2-Iminosugars with Affinity for an anti-MUC1 Antibody. *Org. Lett.* **18,** 3890–3893 (2016).

83. Sadraei, S. I., Reynolds, M. R. & Trant, J. F. The synthesis and biological characterization of acetal-free mimics of the tumor-associated carbohydrate antigens. in *Advances in Carbohydrate Chemistry and Biochemistry* (ed. Baker, D.) **74,** 137–237 (Elsevier Inc., 2017).

84. Wagner, S., Mersch, C. & Hoffmann-Röder, A. Fluorinated glycosyl amino acids for mucin-like glycopeptide antigen analogues. *Chem. Eur. J.* **16,** 7319–7330 (2010).

85. Jiménez-Barbero, J., Dragoni, E., Venturis, C., Nannucci, F., Ardá, A., Fontanella, M., André, S., Cañada, F. J., Gabius, H. J. & Nativi, C. α-O-linked glycopeptide mimetics: Synthesis, conformation analysis, and interactions with viscumin, a galactoside-binding model lectin. *Chem. Eur. J.* **15,** 10423–10431 (2009).

86. Lee, D. & Danishefsky, S. J. 'Biologic' level structures through chemistry: a total synthesis of a unimolecular pentavalent MUCI glycopeptide construct. *Tetrahedron Lett.* **50,** 2167–2170 (2009).

87. Zhu, J., Warren, J. D. & Danishefsky, S. J. Synthetic carbohydrate-based anticancer vaccines: the Memorial Sloan-Kettering experience. *Expert Rev. Vaccines* **8,** 1399–1413 (2009).

88. Uk, G. Menveo Group A, C, W135 and Y conjugate vaccine - Summary of Product Characteristics (SPC) - (eMC). *https://www.medicines.org.uk/emc/medicine/27347#companyDetails* (2016).

89. Buzzi, S., Rubboli, D., Buzzi, G., Buzzi, A. M., Morisi, C. & Pironi, F. CRM197 (nontoxic diphtheria toxin): effects on advanced cancer patients. *Cancer Immunol. Immunother.* **53,** 1041–1048 (2004).

90. Bröker, M., Costantino, P., DeTora, L., Mcintosh, E. D. & Rappuoli, R. Biochemical and biological characteristics of cross-reacting material 197 (CRM197), a non-toxic mutant of diphtheria toxin: Use as a conjugation protein in vaccines and other potential clinical applications. *Biologicals* **39,** 195–204 (2011).

91. Song, C., Sun, S., Huo, C., Li, Q., Zheng, X., Tai, G., Zhou, Y. & Ye, X.-S. Synthesis and immunological evaluation of N-acyl modified Tn analogues as anticancer vaccine candidates. *Bioorg. Med. Chem.* **24,** 915–920 (2016).

92. Corzana, F., Busto, J. H., Jiménez-Osés, G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M. & Avenoza, A. New insights into alpha-GalNAc-Ser motif: Influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc.* **128,** 14640–14648 (2006).

93. 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. Deciphering the non-equivalence of serine and threonine O-glycosylation points: implications for molecular recognition of the Tn antigen by an anti-MUC1 antibody. *Angew. Chem. Int. Ed.* **127,** 9968–9972 (2015).

94. Corzana, F., Busto, J. H., Jiménez-Osés, G., De Luis, M. G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M. & Avenoza, A. Serine versus threonine glycosylation: The methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *J. Am. Chem. Soc.* **129,** 9458–9467 (2007).

95. Kaiser, E., Colescott, R. L., Bossinger, C. D. & Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* **34,** 595–598 (1970).

96. 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. Water sculpts the distinctive

shapes and dynamics of the Tn antigens: Implications for their molecular recognition. *J. Am. Chem. Soc.* **140,** 9952–9960 (2018).

97. Wiberg, K. B., Bailey, W. F., Lambert, K. M. & Stempel, Z. D. The anomeric effect : It's complicated. *J. Org. Chem.* **83,** 5242–5255 (2018).

98. Filloux, C. M. The problem of origins and origins of the problem: Influence of language on studies concerning the anomeric effect. *Angew. Chem. Int. Ed.* **54,** 8880–8894 (2015).

99. Möller, H., Serttas, N., Paulsen, H., Burchell, J. M., Taylor-Papadimitriou, J. & Meyer, B. NMR-based determination of the binding epitope and conformational analysis of MUC-1 glycopeptides and peptides bound to the breast cancer-selective monoclonal antibody SM3. *Eur. J. Biochem.* **269,** 1444–1455 (2002).

100. 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. The quest for anticancer vaccines: deciphering the fine-epitope specificity of cancer-related monoclonal antibodies by combining microarray screening and saturation transfer difference NMR. *J. Am. Chem. Soc.* **137,** 12438–12441 (2015).

101. Brooks, J. M. *Transition state thermodynamic analysis using Biacore T100*. (2007).

102. Van der Merwe, P. A. Surface Plasmon Resonance: General principles of Biacore experiments. in *Protein-Ligand Interactions: hydrodynamics and calorimetry.* (eds. Harding, S. & Chowdhry, P. Z.) (Oxford University Press, 2001).

103. Willcox, B. E., Gao, G. F., Wyer, J. R., Ladbury, J. E., Bell, J. I., Jakobsen, B. K. & Van der Merwe, P. A. TCR binding to peptide-MHC stabilizes a flexible recognition interface. *Immunity* **10,** 357–365 (1999).

104. Navo, C. D. Conformationally restricted Tn antigen mimics: synthesis, structural analysis and biological properties. Doctoral dissertation. (Universidad de La Rioja, 2018).

105. Sun, Y., Liu, H., Cheng, L., Zhu, S., Cai, C., Yang, T. & Ding, P. Thiol Michael addition reaction: a facile tool for introducing peptides into polymer-based gene delivery systems. *Polym. Int.* **67,** 25–31 (2018).

Hydrogen-by-fluorine substitution at Pro7: modified MUC1 mucin as a biosensor

5

5.1. Introduction and background

As previously commented, within the glycocalyx of epithelial cells we can find the MUC1 mucin, comprised by two subunits:[1–3] MUC1-C and MUC1-N. The former is a transmembrane protein anchored to the cell membrane, whereas the later appears exposed to the cytoplasm. As a result of a loss of apical-basal polarity in the outer of carcinoma cells,[4,5] MUC1-N can be shed form the surface and, therefore, detected in the circulation of cancer patients.[6,7]

Within this context, the release of substances from cancer cells into blood and tissues are of considerable clinical importance since their presence can be used as an early diagnosis tool, for determining prognosis and monitoring disease course.[8,9] This approach presents many advantages over traditional techniques, such as mammography and biopsy, because of its minimal invasiveness and cost. These substances are termed tumor biomarkers and are produced either by the tumor or by the host in response to the tumor.[10] Their detection should be both specific and sensitive to allow early diagnosis. Biomarkers can come from diverse molecular origins, i.e. DNA, RNA or proteins (hormones, antibodies, tumor suppressors and enzymes).[11,12]

Many examples of peptide antigens used as cancer biomarkers appear in the bibliography.[13–16] For example, Cancer Antigen 15.3 (CA15.3), whose structure corresponds to MUC1-N mucin,[17] has been widely employed mainly in breast cancer monitoring therapy.[18–21] Nevertheless, most of antigenic biomarkers have not yet demonstrated sufficient sensitivity and specificity for treatment monitoring or clinical use. In addition to this, due to the abundance of non-specific serum proteins their detection is still challenging.[22]

Conversely, tumor-associated autoantibodies (AAbs)[23] represent an appealing alternative to the detection of cancer antigens, since they are more sensitively detected than the latter. AAbs are defined as antibodies produced against host substances that can be an indicative of autoimmune disease or cancer.[24]

AAbs present many useful features as biomarkers. Firstly, they develop during early stages in tumorigenesis, being measurable several years before the development of symptoms[25–28] and remain long time in the circulation. Secondly, they are produced in relatively high concentrations and can be detected at high titers in patients with early stages of cancer.[29] Finally, antibodies are stable in serum samples, compared to other polypeptides, and not subject to proteolysis.[30]

Regarding to the use of MUC1 as tool to detect and measure AAbs in human serum samples, many studies can be found in the bibliography, several of them referring to breast cancer. In all the examples quoted below, ELISA tests are employed to detect the concentration of these AAbs and MUC1 appears coating the plate in different manners. In most cases, the peptide backbone appears naked, but in some examples a

glycosylated MUC1 mucin is employed.[31–33] In many of the investigations the isotype of the Ab is also studied, being IgG and IgM the most frequent. Alternative methodologies used to detect AAbs are protein microarrays[32–34] and it has been recently reported a nanoplasmonic-based biosensor.[35]

In order to provide a better understanding, examples have been classified according to the type of cancer they analyze and the similarities they may present on the proposed work methodology or results.

Regarding to ovarian cancer, the presence of AAbs has been observed in the serum of patients suffering from this disease as well as healthy[36,37] and pregnant women.[37] In fact, there has been reported notably higher concentration of anti-MUC1 IgG AAbs in ovarian cancer patients compared to healthy and pregnant women, but similar IgM AAbs levels in the three scenarios.[37] Other authors noted that IgM[36] and IgG[38] levels were significantly superior in early than in late stage of the disease, which might play a role in controlling disease spread and early diagnosis. It has also been observed the decrease of AAb levels with age, called immunosenescence.[36,39,40] The explanation to this phenomenon is attributed to the formation of circulating immune complexes between the AAbs and the MUC1, so in middle-aged women (over 40 years) high MUC1 and low free AAb levels are due to an augmented release of MUC1 into circulation.[36] More interestingly, events related to the formation of anti-MUC1 AAbs are inversely associated with ovarian cancer risk and the amount of ovulation cycles has been directly connected with a higher incidence for this illness. For instance, women who developed ovarian cancer were more likely to be nulliparous, have a family history of this disease and a relatively shorter durations of oral contraceptive use.[39]

It is argued that events related to stoppage on the ovulation cycles, such as pregnancy or breast-feeding; contraceptive methods, like tubal ligation[39] or IUD use; or even infection episodes, for example bone fracture, osteoporosis or mastitis;[41] are linked to the triggering of an immune response,[39,40,42] since the shed of MUC1 into the circulation provokes the development of protective anti-MUC1 AAbs.

Finally, MUC1-specific AAbs have been pointed as prognostic immune biomarkers in platinum-resistant ovarian cancer, due to their elevated concentration in the platinum-resistant and platinum-refractory tumors studied.[43]

Significant efforts have been made on the development of an efficient methodology to detect anti-MUC1 AAbs in the serum of breast cancer patients, since their apparition is usually related to a better prognosis.[44–47] As in ovarian cancer studies, differences in anti-MUC1 AAbs levels between early-stage breast cancer patients and healthy controls have been observed. Various authors have reported that rates of circulating anti-MUC1 AAbs in patients with benign breast tumors and breast cancer were higher than those in healthy women.[45,48–50] Conversely, other authors claim that these antibodies are present in both healthy controls and breast cancer patients.[34,51–54]

The presence of anti-MUC1 AAbs has also been studied in other body fluids. For instance, in nipple aspirate fluids no significant difference between benign breast lesions

and healthy women was observed.[52] On the contrary, in saliva there has been reported a higher expression of salivary AAbs when compared to healthy controls.[49]

In addition, the progression on the concentration of IgG AAbs in serum with cancer phases has been studied. Slightly higher level of this isotype has been found in patients with stage II compared to stage I, it decreases in stage III and rises again in stage IV.[46,48,55] Nevertheless, other studies only found IgG AAbs in stages I and II, but not in late stage patients.[32] Conversely, different authors point that the presence of these antibodies is characteristic in later stages of breast cancer.[56] Regarding to IgM AAbs, their levels in stages II, III, IV are slightly higher compared to phase I.[48]

Due to the presence of breast cancer antibodies in healthy controls, again, as in ovarian cancer, their apparition was related to reproductive events in women. For example, an increased IgM and IgG MUC1 circulating AAbs levels were demonstrated in non-pregnant versus pregnant individuals and IgG in lactating respect to non-lactating women. Besides, no differences were found on both antibody isotypes levels among the trimesters of pregnancy.[57]

Actually, detection of anti-MUC1 IgG is frequently related to a previous event that stimulates the immune system, such as pregnancy or lactation.[48,54,57] This happens because of the access of MUC1 antigens to the circulation during pregnancy and lactation, and also in benign breast diseases.[53]

The formation of circulating immune complexes between the AAbs and the MUC1 has also been observed in breast cancer. Their presence has been confirmed not only in healthy patients,[57] but also in breast cancer patients.[53,55] According to bibliography, MUC1 can circulate complexed with both IgG and IgM AAbs.[55,57]

Eventually, there is a recent study stating that even if AAbs to MUC1 may be useful for determining prognosis in women with early-stage breast cancer, this kind of experiment is unlikely to be useful as a screening test for cancer within the general population. They extend this assertion to ovarian, lung and pancreatic cancer.[34]

Concerning to colorectal cancer (CRC), there have not been addressed differences between the incidence of anti-MUC1 AAbs in early (I and II) and later clinical stages (III and IV), suggesting that this kind of study might be used for the screening of CRC in the early stages of the disease.[58–60]

The relevance of the carbohydrate within the peptide backbone was reported in CRC, showing the relevance of this scaffold in the detection of MUC1-specifc IgG AAbs using glycopeptide microarray libraries.[61]

Most recent studies demonstrate an increased sensitivity when the assay combines MUC1 with another cancer antigen, like p53 protein,[62] or when a multi-marker combination is employed.[63] These authors propose a multi-marker blood test for early detection of colorectal cancer.

The rest of cancer types have attracted less attention. For example, in lung cancer, the highest level of AAb sensitivity was found to MUC1 antigen among a cohort of 7 different ags.[64] As in other kinds of cancer, the presence of IgG AAbs in healthy individuals has been reported.[65]

It has been demonstrated that circulating anti-MUC1 AAbs could be a favorable prognosis factor in pancreatic cancer cases.[66] In gastric cancer patients IgG immune response to MUC1 is significantly enhanced in patients than in healthy controls, which suggests a better outcome.[44] Also in gastric cancer, it has also been found that in *Helicobacter pylori* infected individuals the IgG immune response to MUC1 is up regulated in gastric cancer patients.[67]

Lastly, in multiple myeloma, yet IgM and IgG AAbs were detected in patients, the mean levels of both circulating AAbs were lower than those detected in healthy donors.[68] It has also been observed an anti-MUC1 AAbs response in cervical carcinogenesis[69] and prostate cancer.[33]

The main conclusion that can be drawn from all the mentioned data is that, regardless of the vast research done, there is not yet a clear protocol to be applied on the determination of anti-MUC1 AAbs for early detection of cancer or treatment evolution.

Nowadays, prostate cancer has become one of the major cancer types diagnosed in men, presenting an elevated mortality rate. Besides, at early stages it is generally asymptomatic and classic diagnostic procedures are highly invasive.[70] A common prostate cancer biomarker found in serum is Prostate Specific Antigen, PSA,[71–73] a glycoprotein enzyme secreted by the epithelial cells of the prostate gland. Nevertheless, PSA detection lacks of sensitivity and specificity, added to the observation of elevated levels of PSA in nonmalignant conditions, results in a high false positive rate.[74–76] For these reasons, alternative tests based on the detection of prostate cancer AAbs have been recently reported,[77–79] as well as our own research presented herein.

We hypothesized that the election of the best antigen to be bound to the ELISA plate would be crucial to improve the sensibility of AAb detection, which is directly related to the antigen-antibody affinity.

To enhance antibody recognition in serum of patients suffering from cancer, and consequently, detect small amounts of AAbs, in this investigation we propose to modify the structure of the antigen (MUC1). By means of the thorough study of the antigen-antibody complex binding, most relevant interactions can be found and they may also be improved by substituting key atoms within antigen peptide backbone.

Regarding this issue, within our research group, X-ray structures of diverse small (glyco)peptides in complex to the SM3 antibody have been recently published.[80] In particular, the complexes between this antibody and the most immunogenic epitope of MUC1,[81,82] the MUC1 fragment APDTRP,[83] and its glycosylated counterpart, APDT(α-*O*-GalNAc)RP, were examined.

Crystallographic analysis revealed that the surface groove of the antibody fits all of the peptide residues in each of the three substrates, whereas the presence of the GalNAc is not critical in the conformation of the peptide backbone complexed to SM3. Moreover, the stabilizing contacts observed include numerous hydrogen bonds, some of them mediated by water molecules, together with various staking interactions. The most significant interactions involving the peptide backbone (Figure 5.1) correspond to Pro7, staking against aromatic units of Trp91L, Trp96L and Tyr32L; and Asp8 and Arg10, whose side-chains are engaged in hydrophobic contacts with Trp33H and Tyr32H, respectively. Hydrogen bonds have been also observed in both glycopeptides between the NH group of Ala6 and the carbonyl group of Thr9/Ser9 with Tyr32L and Gln97H, respectively.
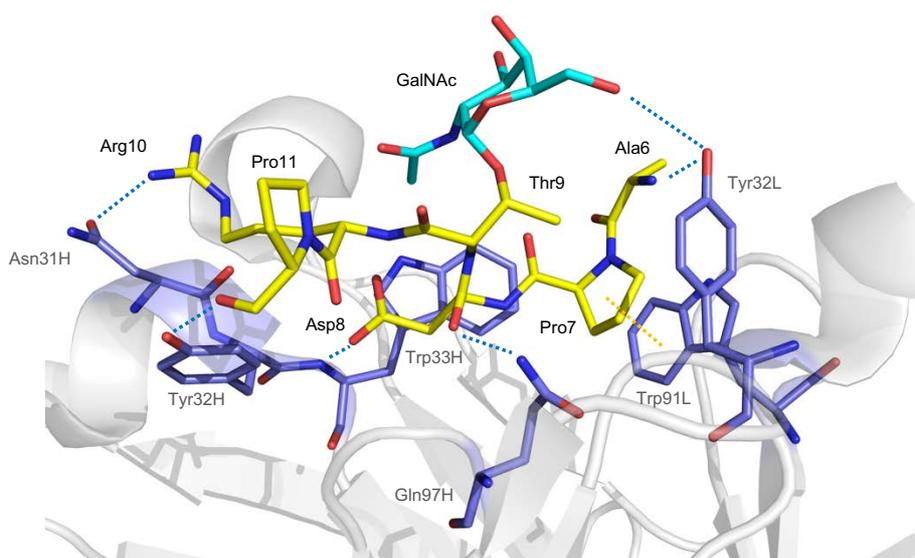
Figure 5.1 Key binding interactions of glycopeptide APDT*RP in complex with antibody SM3 (pdb ID: 5a2k). Dashed orange line represent Pro7-Trp91L stacking interaction and dashed blue lines glycopeptide-SM3 hydrogen bonds.

As a first approach to improve antigen-antibody affinity, hence, sensibility on AAb recognition, we proposed to modify the natural peptide structure of the MUC1 mucin. We decided to modify one particular residue to evaluate stabilizing interactions on the unnatural peptide.

In this particular case, we have chosen Pro7 to be modified. Since this amino acid presents, as aforementioned, several staking interactions with the aromatic rings of diverse residues of the SM3 antibody, we proposed that increasing the polarization of bonds in the interacting C-H moieties , will be translated into a significantly enhanced of the corresponding CH/π interaction, as shown previously in blibliography.[84–88] Consequently, we speculated that by means of the substitution of one or two particular hydrogens of the Pro7 by a highly electronegative atom, for instance fluorine, the CH/π interaction would be reinforced, since the fluorine atom withdraws electronic density making hydrogens more electropositive (Figure 5.2). As unnatural residues, we chose (4*S*)-4-fluoro-L-proline and 4,4-difluoro-L-proline.
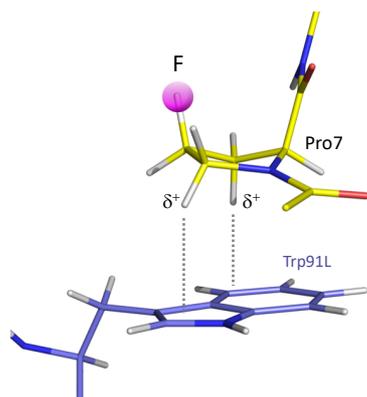
Figure 5.2 Utilization of a fluorine atom to enhance the CH/π interaction between Pro7 and Trp91L.

To provide theoretical support for this hypothesis, we previously calculated the electrostatic potential surface of a proline residue as a diamide and its fluorinated derivatives (Figure 5.3). Their interaction energies with an indole ring were also evaluated at the M06-2X/6-31+G(d,p) level of theory.[89] The interacting Pro face displays a significantly larger positive electrostatic potential in both fluorinated derivatives.

L-proline   (*4S*)-4-fluoro-L-proline   4,4-difluoro-L-proline

$-6.16 \cdot 10^{-2}$ $4.16 \cdot 10^{-2}$
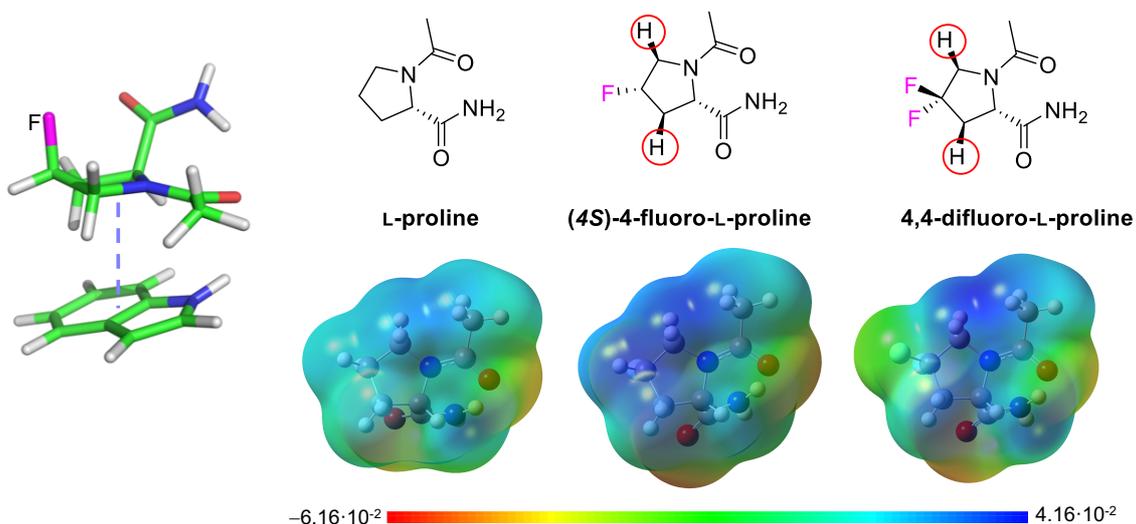
Figure 5.3 Electrostatic potential surfaces (in au) calculated at the M0-2X/ 6-31+G(d,p) level in vacuum, showing re faces of the Pro derivatives. Blue/red indicates positive/negative potentials.

In the same way, this polarization effect is reflected in the theoretical interaction energies, decreasing binding energy in more than 1 kcal/ mol, meaning that the complex stabilization is larger than 1 kcal/mol, and also reducing theoretical distance Pro-indole in 1 Å. (Table 5.1).

	Binding Energy (kcal·mol^{-1})	Distance Pro-indole (Å)
L-proline	-14.7	3.7
(4*S*)-4-fluoro-L-proline	-16.0	3.6
4,4-difluoro-L-proline	-15.9	3.6

5.2. Objectives

Herein, our aim is to study how hydrogen-by-fluorine substitution at Pro7 within the MUC1 mucin backbone influences antigen-antibody binding affinity. Besides, we would like to analyze (glyco)peptide structures bound to the SM3 anti-MUC1 antibody, to provide with a rational explanation for the differences observed in recognition. Finally, we also desire to test the sensibility of our derivatives in serum of prostate cancer patients, in order to validate their use as tumor biomarkers.

With this purpose, both mono- and difluorinated derivatives of the MUC1 mucin and their glycosylated counterparts were synthesized. To determine their binding affinity to two different anti-MUC1 antibodies, SM3 and VU3C6, microarray and BLI experiments were performed.

Detailed structural studies of the fluorinated antigens were carried out on a shorter version of the MUC1 mucin. This fragment was studied crystallographically in complex with a single-chain variable fragment of the SM3 anti-MUC1 antibody (scFv-SM3) and also by NMR and MD simulations.

In the end, as a proof-of-concept, we performed a serologic assay to study antigen-antibody affinity by indirect ELISA; using the unnatural mucins synthesized by us as antigens and circulating anti-MUC1 AAbs in serum samples of patients presenting benign and malignant prostate tumors.

5.3. Synthesis

The synthesis of all the peptides and glycopeptides shown in Figure 5.4, as well as their simplified version, **fP*'**, **P*'** and **2fP*'**, was performed using the MW-SPPS protocol with a Rink amide MBHA resin, Fmoc methodology and commercially available side chain protected amino acids. (See *Experimental section 8.11.1).*

P   AHGVTSAPDTRPAPGSTAPPA

P*   AHGVTSAPDTRPAPGSTAPPA

P*'   GVTSAPDTRPAP

fP   AHGVTSA⬠DTRPAPGSTAPPA

fP*   AHGVTSA⬠DTRPAPGSTAPPA

fP*'   GVTSA⬠DTRPAP

2fP   AHGVTSA⬠DTRPAPGSTAPPA

2fP*'   GVTSA⬠DTRPAP

2fP*   AHGVTSA⬠DTRPAPGSTAPPA

1   7 9   20

Pro   **F-Pro**   **2F-Pro**   **α-*O*-GalNAc-Thr**

P   ⬠   ⬠   🟨
                                            T
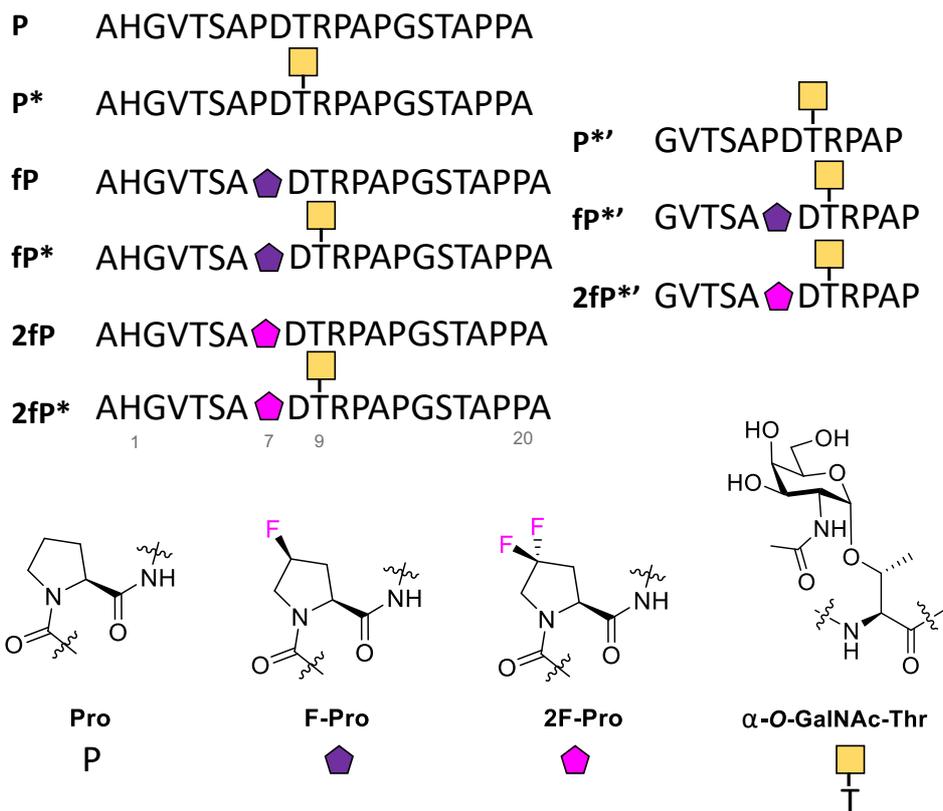
Figure 5.4 MUC1-like peptides and glycopeptides synthesized and studied.

The protected glycosyl amino acid (α-*O*-GalNAc-Thr) and fluorinated prolines ((4*S*)-4-fluoro-L-proline and 4,4-difluoro-L-proline) were coupled manually as described in *Experimental section 8.2.*, in order to reduce the number of equivalents required. The reaction time was extended compared to the in-machine synthesis and the end of the coupling was checked by Kaiser test.[90] The *O*-acetyl groups of the carbohydrate were removed in a mixture of NH_2NH_2/MeOH (7:3). Afterwards, peptides were released from the resin, simultaneously with all the acid-labile side-chain protecting groups, using a cocktail cleavage consisting of TFA/TIS/H_2O (95:2.5:2.5). Finally, the (glyco)peptides were purified by semi-preparative HPLC, lyophilized and characterized.

5.4. Affinity assays

With these compounds in hand, we performed diverse affinity studies to evaluate the impact of the hydrogen-by-fluorine substitution on the antigen in association with the anti-MUC1 antibodies SM3 and VU3C6.

5.4.1. Microarray assays

Microarray assays were performed to determine the affinity of MUC1 mucin antigens with commercially available antibodies SM3 and VU3C6.[91]

(Glyco)peptides were spotted on microarrays slides in quadruplicate at six different concentrations, from 500 μM to 15.6 μM. (See *Experimental Section 8.6.*). The

fluorescent image scans of both anti-MUC1 antibodies, SM3 and VU3C6, at 50 μg/mL, are shown in Figure 5.5.
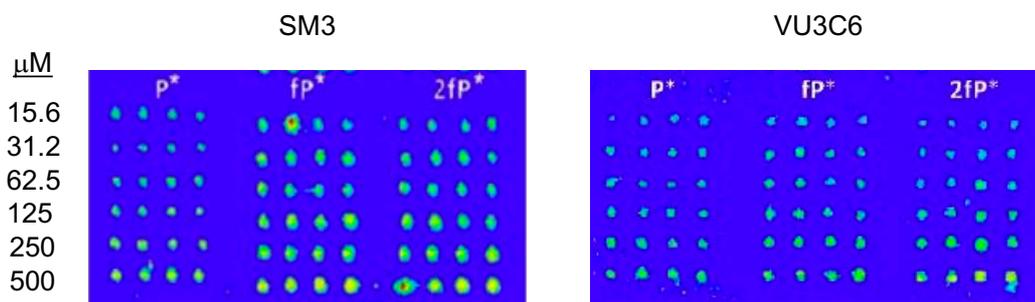
Figure 5.5 Fluorescent image scan of SM3 and VU3C6 (50 mg/mL both).

In general, microarray experiments show an enhancement of antigen-antibody affinity for the fluorinated peptides. Both SM3 and VU3C6 antibodies recognize the mucins in a similar manner, yet affinity to SM3 is slightly higher for the unnatural glycosylated derivatives, compared to VU3C6. For instance, the affinity of both **fP*** and **2fP*** (500 μM) for SM3 is approximately two-fold compared to the natural mucin, whereas the same scenario is less than 1.5-fold for VU3C6. To confirm these results the same (glyco)peptides were subjected to biolayer interferometry experiments from which quantitative binding data can be obtained.

5.4.2. Biolayer interferometry (BLI)

To determine affinity of the MUC1 antigens **fP**, **fP***, **2fP** and **2fP*** with the scFv-SM3 antibody, the later were immobilized on amine-reactive biosensors to perform BLI experiments. (*Experimental section 8.11.2.*). Dissociation constants (K_D) for each glyco(peptide) are shown in Figure 5.6.

As previously reported by our group,[80] the glycosylated counterpart, **P***showed better affinity, more than 3-fold, than the naked peptide, **P**. The same observation is applicable to the fluorinated derivatives. In fact, MUC1 mucins containing (4S)-4-fluoro-L-proline (**fP** and **fP***) and 4,4-difluoro-L-proline (**2fP** and **2fP***) displayed enhanced affinity compared to their natural counterparts (**P** and **P***). Of note, in the case of the glycosylated antigens, both **fP*** and **2fP*** present 3-fold enhancement with respect to the natural **P***.
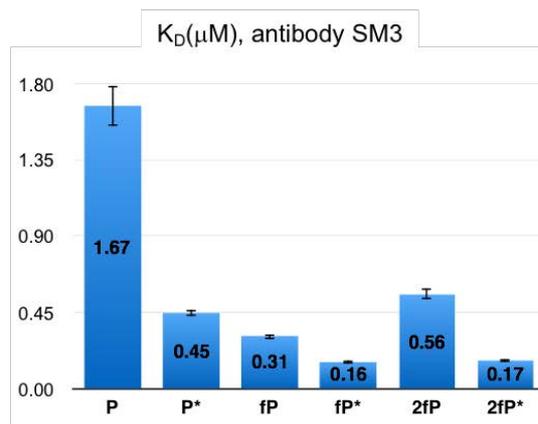
Figure 5.6 k_D constants obtained from BLI experiments for studied peptides.

5.5. Crystallographic studies

To study the conformation of fluorinated peptides in complex with anti-MUC1 antibodies, and consequently validate the aforementioned binding affinity experiments, crystallographic studies on the MUC1 mucins in complex with scFv-SM3 antibody were also carried out.

To facilitate the crystallization, a shorter model of the MUC1 antigen was synthesized, comprising the sequence GVTSAfPDT*RPAP, denoted as **fP*'**. This fragment was complexed to scFv-SM3 antibody, as previously described.[80] Crystals were grown by sitting drop diffusion at 18 °C. (See *Experimental section 8.11.3.*), enabling the acquisition of a structure at high resolution (<2.0 Å).
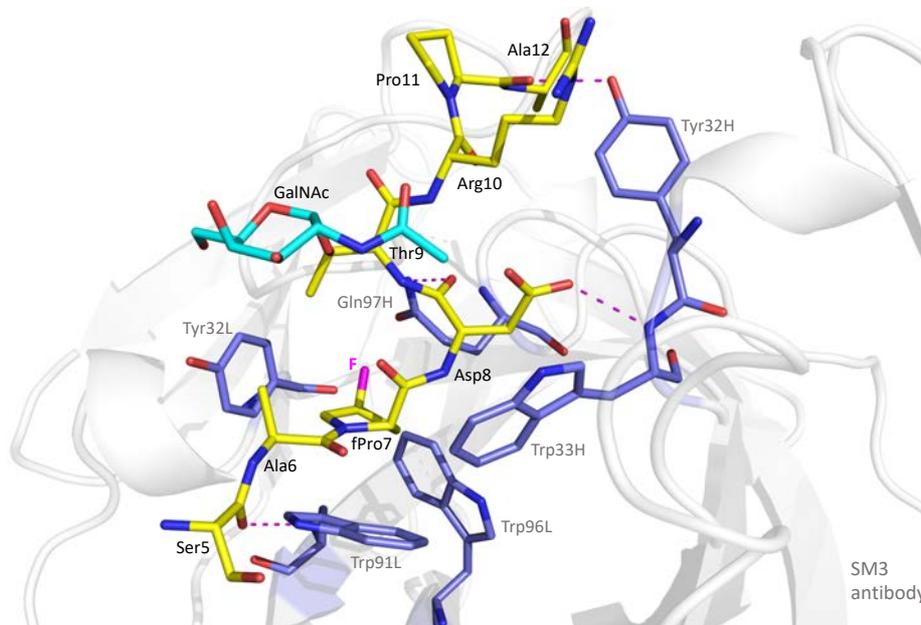
*Figure 5.7 Crystal structure of **fP*'** bound to scFv-SM3 (pdb ID: 5OWP). Peptide backbone carbon atoms appear in yellow, GalNAc carbon atoms are shown in cyan and carbon atoms of key residues of SM3 are highlighted in purple. The rest of the antibody is show as grey ribbons. Fluorine atom appears in pink. Dashed lines indicating hydrogen bonds between peptide backbone and SM3 antibody.*

Crystal structure of **fP*'** fragment complexed with scFv-SM3 is shown in Figure 5.7 (pdb ID: 5OWP). Only SAPDTRP residues from the whole **fP*'** peptide could be resolved, probably due to the higher flexibility of the rest of the amino acids.

In general, the interactions between the unnatural glycopeptide and scFv-SM3 antibody are the same as those found for the natural glycopeptide described above (Figure 5.1). Key hydrogen bonds within the peptide backbone and the antibody have been highlighted as dashed lines in Figure 5.7. In addition, there is an intermolecular hydrogen bond between the hydroxymethyl group of GalNAc and the side chain of Tyr32L of the SM3. Besides, the *N*-acetyl group of the sugar stacks with the aromatic ring of Trp33H, which explains the enhanced affinity of anti-MUC1 antibodies for glycosylated antigens. As expected, a stacking interaction is also observed between the hydrogens 3 and 5 from the pyrrolidine (4*S*)-4-fluoro-L-proline ring and Trp91L. Both hydrogens appear perpendicularly presented over the indole ring of Trp91L.

In Figure 5.8, experimental distances between the center of the Pro7 ring and Trp91 of diverse antigen-antibody complexes previously reported by us[80] and others[83] have been represented. The effect that fluorine plays on the reduction of the aforementioned distance is noticeable, since it has been shortened 0.2 Å compared with its natural counterpart, this is, APDT(α-*O*-GalNAc)RP, denoted as 5a2k. The same feature has been observed for the non-glycosylated peptide and the distance is 0.3 Å longer when fluorinated glycopeptide is compared with APDS(α-*S*-GalNAc)RP, denoted as 5a2j. Therefore, the hydrogen-by-fluorine substitution, in position 4 reinforces CH/π interaction with the tryptophan residue.
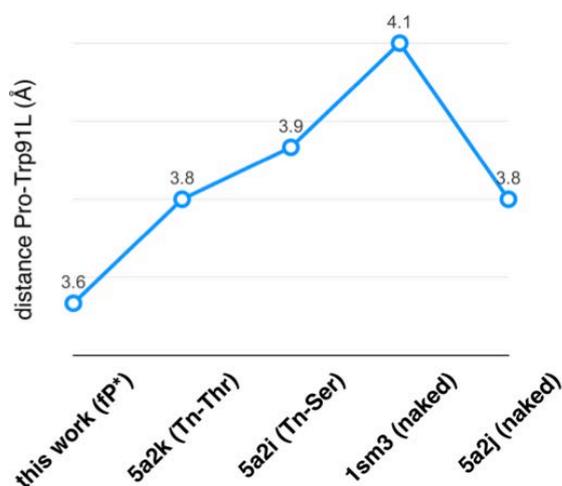
Figure 5.8 Experimental distances Pro–Trp91L obtained from X-ray structures of various MUC1-like derivatives in complex with antibody SM3.

Additionally, to verify the perfect fitting of **fP*'** in the surface groove of the scFv-SM3 antibody, its structure has been superposed to that of its natural counterparts. As can be observed in Figure 5.9, the superposition of the peptide backbone of glycopeptides **fP*'** and APDT(α-*O*-GalNAc)RP,[80] together with SAPDTRPAP peptide[83] in complex with

SM3, remains unchanged. This fact suggests that the incorporation of a fluorine atom does not significantly modify the structure of the peptide in the bound state.

rmsd (*backbone*) = 0.2 Å
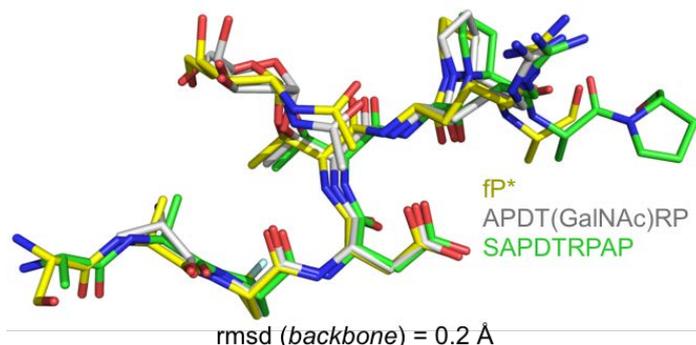
Figure 5.9 Superposition of X-Ray structure of fP*', APDT(α-O-GalNAc)RP and SAPDTRPAP (glyco)peptides.

This implies that the antigen-antibody hydrogen-bonding network is identical for the unnatural glycopeptide presented herein, **fP*'**, and the previously reported complexes.[80,83]

5.6. Molecular Dynamics (MD) simulations

As we couldn't obtain de crystal structure of the **2fP*'** in the bound state with antibody SM3, we envisioned a theoretical study based on MD simulations. These calculations were also performed on **fP*'**/SM3 and **P*'**/SM3 complexes for comparative purposes. The 200 ns MD simulations in explicit solvent indicated that three complexes were stable through the whole simulation time.

fP*', R = H
2fP*', R = F
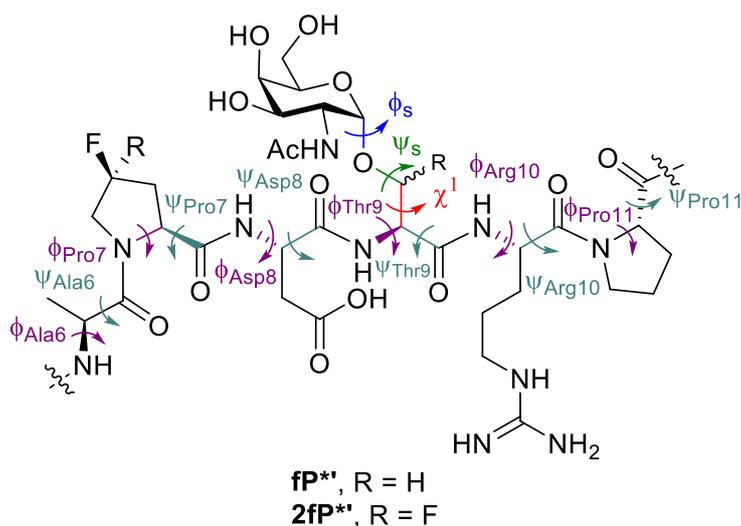
Figure 5.10 Representation of ϕ and ψ torsional angles for Af/2fPD(α-O-GalNA)RP fragment of fP*' and 2fP*' glycopeptides.

First, we analyzed ϕ_s, ψ_s and χ^1 torsional angles for the glycosidic linkage (GalNAc-Thr, shown in Figure 5.10) in glycopeptides **fP*'** and **2fP*'** in complex with SM3 antibody. As

depicted in Figure 5.11, both present a similar conformational behavior, with ψ_s ranging from 90 to 180º.[92] It is important to note that the glycosidic linkage in the natural counterpart bound to SM3 displays a value of ψ_s around 90º, likely to favor a CH/π interaction between the *N*-acetyl group of GalNAc and Trp33H and a hydrogen bond between the hydroxymethyl group of the sugar and Tyr32L.[80]

For the glycopeptide **fP*'** in complex with SM3 antibody, the main conformer populated observed in the crystal structure (pdb ID: 5OWP) is highlighted as a red dot in Figure 5.11, which is different from that observed in solution by MD.
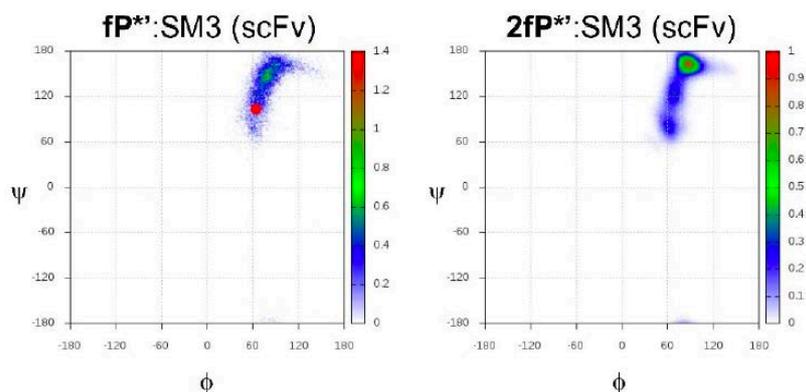
*Figure 5.11 ϕ_s and ψ_s torsional angles for the glycosidic linkage (GalNAc-Thr) in glycopeptides **fP*'** and **2fP*'** bound to SM3 antibody.*

In the same way, ϕ and ψ torsional angles (Figure 5.10) for the backbone of fluorinated glycopeptides, **fP*'** and **2fP*'**, were monitored (Figure 5.12 and Figure 5.13, respectively). For both derivatives, Ala6, Asp8, Arg10 and both Pro7 and Pro11 exhibit an extended conformation, whereas GalNAc-Thr9 displays a folded conformation.

As in the previous case, for the glycopeptide **fP*'** bound to SM3, the main conformer observed in the X-Ray structure is depicted as a red dot (Figure 5.12). This is in fully agreement with that calculated in solution for all the residues, except for Pro11, which is slightly diverted. This results concerning to the peptide conformation are also concordant with previous studies performed by our group over the natural peptide APDT(α-*O*-GalNAc)RP in complex with SM3.[80]

fP*':SM3 (scFv)
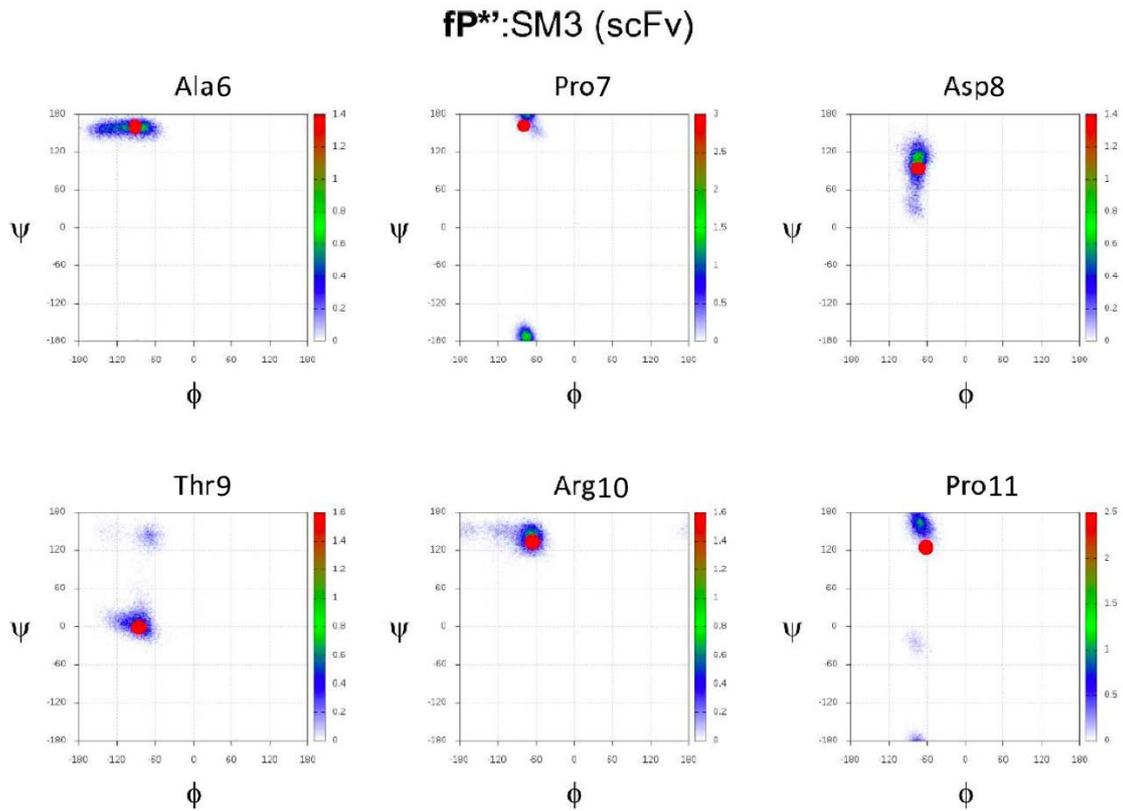
*Figure 5.12 ϕ and ψ torsional angles for the $AfPDT(\alpha-O-GalNAc)RP$ fragment in glycopeptide **fP*'** bound to SM3 antibody.*

2fP*':SM3 (scFv)
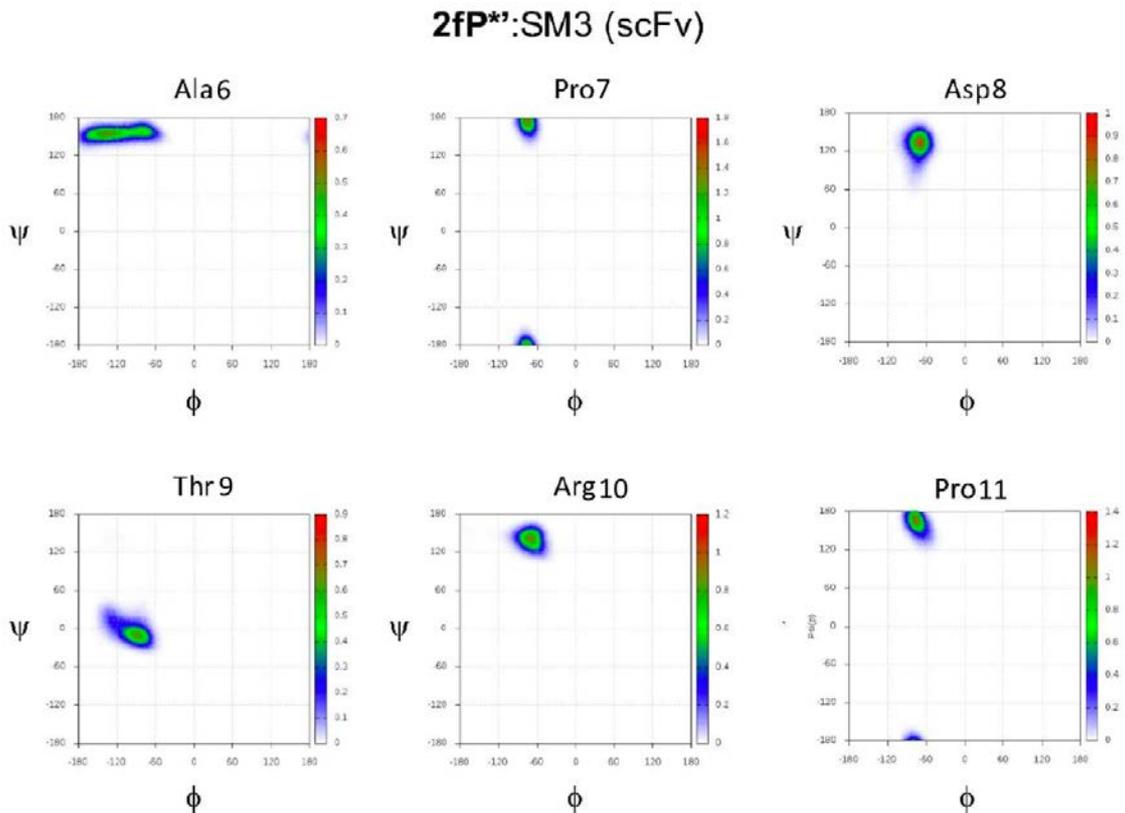
*Figure 5.13 ϕ and ψ torsional angles for the $A2fPDT(\alpha-O-GalNAc)RP$ fragment in glycopeptide **2fP*'** bound to SM3 antibody.*

We monitored then the distances between Pro and Trp91L through the MD simulations. Interestingly, this distance was shorter in the case of the fluorinated glycopeptides: 4.0 ± 0.4 Å was found in the case of the derivative **fP*'**, which is in agreement with X-ray structure previously described, and 4.1 ± 0.6 Å distance was observed for **2fP*'**; whilst for the natural glycopeptide **P*'** the distance Pro-Trp91L was 4.7 ± 0.4 Å. (Figure 5.14)
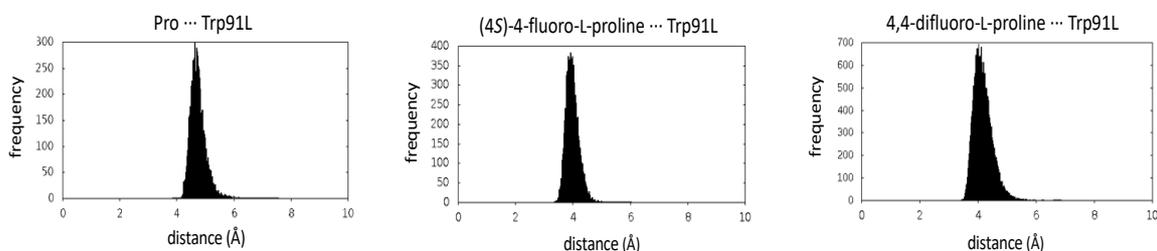
Figure 5.14 Distance distribution of Pro-Trp91L obtained for glycopeptide P' (left panel), fP*'(middle panel) and 2fP*' (right panel) bound to the antibody SM3.*

We can infer from these results that both mono- and difluorinated peptide conformations remain as in the natural derivative, therefore binding affinity would not be conformationally influenced by the hydrogen-by-fluorine substitution.

Conversely, the enhancement in the fluorinated proline/tryptophan staking interaction, shortens the distance between these residues, providing with a theoretic explanation for the improvement in the antigen-antibody binding observed in the microarray and BLI experiments.

5.7. Serologic assay

In order to check if the fluorinated mucins could detect anti-MUC1 AAbs in a more sensible way than their natural counterparts, we established an indirect Enzyme Linked Immunosorbent Assay (ELISA) protocol using human serum samples obtained from Biobanco-iMM, Lisbon Academic Medical Center (Lisbon, Portugal). (Table 5.2).

Table 5.2 Age distribution of sera donors used for human anti-MUC1 AAbs detection.

Diagnostic	Age distribution (years)	Number of samples
Prostatic adenocarcinoma	65.3 ± 6.0 (p=0.0947, n.s.)	9
Prostatic benign hyperplasia	71.6 ± 8.1 (p=0.0092, **)	11
Male controls	60.2 ± 2.5	5

Therefore, in this indirect ELISA experiment, both natural MUC1 mucin, denoted as **P***, and the difluorinated moiety **2fP*** were checked on different serum samples from prostatic adenocarcinoma and prostatic benign hyperplasia, as well as healthy controls.

We chose to limit this study only to glycosylated peptides due to their enhanced binding affinity compared to their naked counterparts, as previously demonstrated by both BLI and microarray experiments.

The experiment proceeded as described in *Experimental Section 8.11.4*.
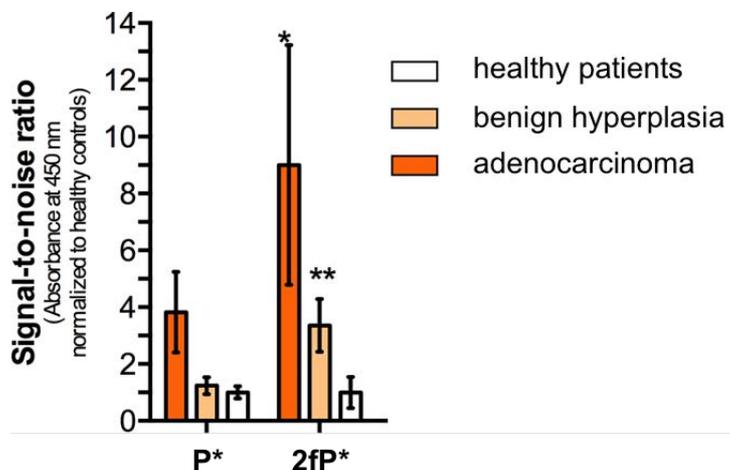
*Figure 5.15 Serologic assay results of **P*** and ***2fP*** MUC1 antigens in human serum of prostate cancer patients at different stages. *p < 0.05; **p < 0.02.*

In Figure 5.15, ELISA experiment results for the natural MUC1 mucin, **P***, and the fluorinated derivative, **2fP***, are shown. The values obtained have been normalized to healthy controls, disclosing a 6-fold increased binding affinity for **2fP*** compared to **P*** in benign hyperplasia cases and more than 2-fold enhancement in adenocarcinoma serum.

ELISA experiments have demonstrated that hydrogen-by-fluorine substitution positively alters antigen-antibody binding, enhancing sensibility of the recognition. Therefore, previous affinity assays and structural studies have been proved trustworthy, since they are supported by serologic assay.

5.8. Conclusions

In this chapter, our starting hypothesis has been that the replacement of one specific residue within the most immunogenic epitope of MUC1 mucin, APDTRP, would contribute to stabilize the interactions with anti-MUC1 antibodies, improving their recognition.

In this case, we decided to substitute Pro in position 7 by two fluorinated derivatives, either (4*S*)-4-fluoro-L-proline or 4,4-difluoro-L-proline. We postulated that this replacement would reinforce staking interactions that Pro7 presents with the aromatic ring of various residues of the SM3 antibody. Due to an increase in the polarization of the interacting CH moieties, the CH/π bonds would be significantly enhanced.

Binding assays, such as BLI and microarray experiments, showed that effectively antigen-antibody interaction is higher when fluorine modified peptides are tested. We

compared mono- and difluorinated (glyco)peptides with their natural derivatives using two different anti-MUC1 antibodies, SM3 and VU3C6, obtaining in all cases better response in the two former cases. We also observed that K_D constants were 3-fold lower in case of fluorinated glycopeptides.

These results have been supported by the crystal structure of the APDTRP epitope of MUC1 bound to scFv-SM3 antibody. We could experimentally observe the staking interactions among hydrogens 3 and 5 from fluoroproline ring and indole ring of Trp91L, as we previously hypothesized. This influence reduces the Pro-Trp distance, as we have measured, but does not modify the peptide conformation, in comparison with natural peptides.

Besides, X-ray structure of the fluorinated epitope in the bound state was superposed to that of natural peptides, showing no significant differences on their disposition. Consequently, the antigen–antibody hydrogen-bonding network is identical to that found in the previously reported complexes

Backbone structure in complex with SM3 antibody was also studied by MD simulations. In line with crystal structure, peptide conformation remains unchanged for the fluorinated peptides. The most relevant difference was the shortening in the distance distribution of Pro7-Trp91L due to the influence of hydrogen-by-fluorine substitution.

Finally, to validate aforementioned experiments, serologic assays, were conducted experiments on serum samples from prostatic adenocarcinoma, prostatic benign hyperplasia and healthy controls. ELISA experiment confirmed that the fluorinated glycopeptide **2fP*** presents enhanced sensibility for AAbs than natural MUC1 mucin **P***.

In light of these results, we state that hydrogen-by-fluorine substitution in Pro7 within MUC1 mucin backbone successfully enhances antigen-antibody binding affinity. Therefore, this kind of derivatives might be applied on the determination of anti-MUC1 AAbs for early detection or monitoring of cancer disease.

This work has been published as an article entitled *The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer* in the *Journal of American Chemical Society*, DOI: 10.1021/jacs.7b09447.[93]

5.9. References

1. Siddiqui, J., Abe, M., Hayes, D., Shani, E., Yunis, E. & Kufe, D. Isolation and sequencing of a cDNA coding for the human DF3 breast carcinoma-associated antigen. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2320–2323 (1988).

2. Levitin, F., Stern, O., Weiss, M., Gil-Henn, C., Ziv, R., Prokocimer, Z., Smorodinsky, N. I., Rubinstein, D. B. & Wreschner, D. H. The MUC1 SEA module is a self-cleaving domain. *J. Biol. Chem.* **280,** 33374–33386 (2005).

3. Ligtenberg, M. J. L., Kruijshaar, L., Buijs, F., Van Meijer, M., Litvinov, S. V. & Hilkens, J. Cell-associated episialin is a complex containing two proteins derived from a common precursor. *J. Biol. Chem.* **267,** 6171–6177 (1992).

4. Lee, M. & Vasioukhin, V. Cell polarity and cancer - cell and tissue polarity as a non-canonical tumor suppressor. *J. Cell Sci.* **121,** 1141–1150 (2008).

5. Royer, C. & Lu, X. Epithelial cell polarity: A major gatekeeper against cancer. *Cell Death Differ.* **18,** 1470–1477 (2011).

6. Boshell, M., Lalanl, E., Pemberton, L., Burchell, J., Gendler, S. & Taylor-papadimitriou, J. The product of the human MUC1 gene when secreted by mouse cells transfected with the full-lenght cDNA lacks the cytoplasmic tail. *Biochem. Biophys. Res. Commun.* **185,** 1–8 (1992).

7. Smorodinsky, N., Weiss, M., Hartmann, M. L., Baruch, A., Harness, E., Yaakobovitz, M., Keydar, I. & Wreschner, D. H. Detection of a secreted MUC1/SEC protein by MUC1 isoform specific monoclonal antibodies. *Biochem. Biophys. Res. Commun.* **228,** 115–121 (1996).

8. Somerfield, M. R. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. *J. Clin. Oncol.* **14,** 2843–2877 (1996).

9. 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. Common cancer biomarkers. *Cancer Res.* **66,** 2953–2961 (2006).

10. Kilpatrick, E. S. & Lind, M. J. Appropriate requesting of serum tumour markers. *BMJ* **339,** 859–862 (2009).

11. Kabel, A. M. Tumor markers of breast cancer: New prospectives. *J. Oncol. Sci.* **3,** 5–11 (2017).

12. Sell, S. *Human Cancer Markers*. (Springer US, 1982).

13. Meyer, T. & Rustin, G. J. Role of tumour markers in monitoring epithelial ovarian cancer. *Br. J. Cancer* **82,** 1535–1538 (2000).

14. Smith, D. S., Humphrey, P. A. & Catalona, W. J. The early detection of prostate carcinoma with prostate specific antigen. *Cancer* **80,** 1852–1856 (1997).

15. Tothill, I. E. Biosensors for cancer markers diagnosis. *Semin. Cell Dev. Biol.* **20,** 55–62 (2009).

16. Bohunicky, B. & Mousa, S. A. Biosensors: The new wave in cancer diagnosis. *Nanotechnol. Sci. Appl.* **4,** 1–10 (2011).

17. Duffy, M. CA 15-3 and related mucins as markers in breast cancer: a critical review. *Ann. Clin. Biochem.* 579–586. (2000).

18. Duffy, M. J. Serum tumor markers in breast cancer: Are they of clinical value? *Clin. Chem.* **52,** 345–351 (2006).

19. Fu, Y. & Li, H. Assessing Clinical Significance of Serum CA15-3 and Carcinoembryonic Antigen (CEA) Levels in Breast Cancer Patients: A Meta-Analysis. *Med. Sci. Monit.* **22,** 3154–3162 (2016).

20. Klee, G. G. & Schreiber, W. E. MUC1 gene-derived glycoprotein assays for monitoring breast cancer (CA 15-3, CA 27.29, BR): Are they measuring the same antigen? *Arch. Pathol. Lab. Med.* **128,** 1131–1135 (2004).

21. Lumachi, F., Basso, S. M. M., Brandes, A. A., Pagano, D. & Ermani, M. Relationship between hormone receptors, MIB-1 index and serum tumour markers CEA and CA 15–3 in patients with pT1–2 breast cancer. *Anticancer Res.* **24,** 3221–3224 (2004).

22. Lu, H. L., Goodell, V. & Disis, M. L. Humoral Immunity Directed against Tumor-Associated Antigens As Potential Biomarkers for the Early Diagnosis of Cancer. *J. Proteome Res.* **7,** 1388–1394 (2008).

23. Wu, J., Li, X., Song, W., Fang, Y., Yu, L. & Liu, S. The roles and applications of autoantibodies in progression, diagnosis, treatment and prognosis of human malignant tumours. *Autoimmun. Rev.* **16,** 1270–1281 (2017).

24. MacDonald, I. K., Parsy-Kowalska, C. B. & Chapman, C. J. Autoantibodies: Opportunities for Early Cancer Detection. *Trends in Cancer* **3,** 198–213 (2017).

25. Turnbull, A. R., Turner, D. T. L., Fraser, J. D., Lloyd, R. S., Lang, C. J. & Wright, R. Autoantibodies in early breast cancer: A stage-ralated phenomenon? *Br. J. Cancer* **38,** 461–463 (1978).

26. 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. Immunobiomarkers in small cell lung cancer: Potential early cancer signals. *Clin. Cancer Res.* **17,** 1474–1480 (2011).

27. Zhong, L., Coe, S. P., Stromberg, A. J., Khattar, N. H., Jett, J. R. & Hirschowitz, E. A. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J. Thorac. Oncol.* **1,** 513–519 (2006).

28. Trivers, G. E., De Benedetti, V. M., Cawley, H. L., Caron, G., Harrington, a M., Bennett, W. P., Jett, J. R., Colby, T. V, Tazelaar, H., Pairolero, P., Miller, R. D. & Harris, C. C. Anti-p53 antibodies in sera from patients with chronic obstructive pulmonary disease can predate a diagnosis of cancer. *Clin. Cancer Res.* **2,** 1767–1775 (1996).

29. Disis, B. M. L., Pupa, S. M., Gralow, J. R., Dittadi, R., Menard, S. & Cheever, M. a. High-Titer HER-2 / neu Protein-Specific Antibody Can Be Detected in Patients With Early-Stage Breast Cancer. *J. Clin. Oncol.* **15,** 3363–3367 (1997).

30. Anderson, K. S. & Labaer, J. The Sentinel Within : Exploiting the Immune System for Cancer. *J. Proteome Res.* **4,** 1123–1133 (2005).

31. Kurtenkov, O., Klaamas, K., Vahter, L., Sergejev, B., Miljukhina, L. & Shljapnikova, L. IgG immune response to tumor-associated carbohydrate antigens (TF, Tn, alpha-Gal) in patients with breast cancer: impact of neoadjuvant chemotheraphy and relation to the survival. *Exp. Oncol.* **27,** 136–140 (2005).

32. Blixt, O., Bueti, D., Burford, B., Allen, D., Julien, S., Hollingsworth, M., Gammerman, A., Fentiman, I., Taylor-papadimitriou, J. & Burchell, J. M. Autoantibodies to aberrantly glycosylated MUC1 in early stage breast cancer are associated with a better prognosis. *Breast Cancer Res.* **13,** R25 (2011).

33. Wandall, H. H., Blixt, O., Tarp, M. A., Pedersen, J. W., Bennett, E. P., Mandel, U., Ragupathi, G., Livingston, P., Hollingsworth, M. A., Taylor-papadimitriou, J., Burchell, J. & Clausen, H. Auto-antibody signatures to aberrant O-glycopeptide epitopes serve as undiscovered biomarkers of cancer. *Cancer Res.* **70,** 1306–1313 (2010).

34. Burford, B., Graham, R., Allen, D., Pedersen, J. W., Nudelman, A. S., Blixt, O., Fourkala, E. O., Bueti, D., Dawnay, A., Ford, J., Desai, R., David, L., Trinder, P., Acres, B. & Schwientek, T. Autoantibodies to MUC1 glycopeptides cannot be used as a screening assay for early detection of breast, ovarian, lung or pancreatic cancer. *Br. J. Cancer* **108,** 2045–2055 (2013).

35. Soler, M., Estevez, M., Villar-vazquez, R., Casal, J. I. & Lechuga, L. M. Label-free nanoplasmonic sensing of tumor-associate autoantibodies for early diagnosis of colorectal cancer. *Anal. Chim. Acta* **930,** 31–38 (2016).

36. Richards, E. R., Devine, P. L., Quin, R. J., Fontenot, J. D., Ward, B. G. & Mcguckin, M. A. Antibodies reactive with the protein core of MUC1 mucin are present in ovarian cancer patients and healthy women. *Cancer Immunol. Immunother.* **46,** 245–252 (1998).

37. Snijdewint, F. G. M. G., von Mensdorff-Pouilly, S., Karuntu-Wanamarta, A. H. H., Verstraeten, A. A. A., van Zanten-Przybysz, I., Hummel, P., Nijman, H. W. W., Kenemans, P. & Hilgers, J. Cellular and humoral immune responses to MUC1 mucin and tandem-repeat peptides in ovarian cancer patients and controls. *Cancer Immunol. Immunother.* **48,** 47–55 (1999).

38. Oei, A. L. M., Moreno, M., Verheijen, R. H. M., Sweep, F. C. G. J., Thomas, C. M. G., Massuger, L. F. A. G. & Mensdorff-Pouilly, S. Induction of IgG antibodies to MUC1 and survival in patients with epithelial ovarian cancer. *Int. J. Cancer* **123,** 1848–1853 (2008).

39. Pinheiro, S. P., Hankinson, S. E., Tworoger, S. S., Rosner, B. A., Mckolanis, J. R., Finn, O. J. & Cramer, D. W. Anti-MUC1 antibodies and ovarian cancer risk: Prospective data from the nurses' health studies. *Cancer Epidemiol. Biomarkers Prev.* **19,** 1595–1602 (2010).

40. Terry, K. L., Titus-Ernstoff, L., Mckolanis, J. R., Welch, W. R., Finn, O. J. & Cramer, D. W. Incessant ovulation, Mucin 1 immunity, and risk for ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* **16,** 30–36 (2007).

41. Cramer, D. W., Williams, K., Vitonis, A. F., Titus, L. & Fichorova, R. N. Puerperal mastitis: a reproductive event of importance affecting anti-mucin antibody levels and ovarian cancer risk. *Cancer Causes Control* **24,** 1911–1923 (2013).

42. Cramer, D. W., Titus-Ernstoff, L., Mckolanis, J. R., Welch, W. R., Vitonis, A. F., Berkowitz, R. S. & Finn, O. J. Conditions associated with antibodies against the tumor-associated antigen MUC1 and their relationship to risk for ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* **14,** 1125–1131 (2005).

43. Budiu, R. A., Mantia-Smaldone, G., Elishaev, E., Chu, T., Thaller, J., McCabe, K., Lenzner, D., Edwards, R. P. & Vlad, A. M. Soluble MUC1 and serum MUC1-specific antibodies are potential prognostic biomarkers for platinum-resistant ovarian cancer. *Cancer Immunol. Immunother.* **60,** 975–984 (2011).

44. Kurtenkov, O., Klaamas, K., Miljukhina, L., Shljapnikova, L., Chužmarov, V., Klaamas, K., Miljukhina, L. & Shljapnikova, L. Humoral immune response to MUC1 and to the Thomsen-Friedenreich (TF) glycotope in patients with gastric cancer: Relation to survival. *Acta Oncol. (Madr).* **46,** 316–323 (2009).

45. Fremd, C., Stefanovic, S., Beckhove, P., Pritsch, M., Wallwiener, M., Heil, J., Golatta, M., Rom, J., Schneeweiss, A., Schuetz, F., Domschke, C., Fremd, C., Stefanovic, S., Beckhove, P., Pritsch, M., Wallwiener, M., Heil, J., … Heil, J. Mucin 1-specific B cell immune responses and their impact on overall survival in breast cancer patients. *Oncoimmunology* **5,** e1057387-1–7 (2016).

46. von Mensdorff-Pouilly, S., Verstraeten, A. A., Kenemans, P., Snijdewint, F. G. M., Kok, A., Van Kamp, G. J., Paul, M. A., Van Diest, P. J., Meijer, S. & Hilgers, J. Survival in early breast cancer patients is favorably influenced by a natural humoral immune response to polymorphic epithelial mucin. *J. Clin. Oncol.* **18,** 574 (2000).

47. Storr, S. J., Chakrabarti, J., Barnes, A., Murray, A., Chapman, C. J. & Robertson, J. F. R. Use of autoantibodies in breast cancer screening and diagnosis. *Expert Rev. Anticancer Ther.* **6,** 1215–1223 (2006).

48. Tang, Y., Wang, L., Zhang, P., Wei, H., Gao, R., Liu, X., Yu, Y., Wang, L. & Mmunol, C. L. I. N. V. A. I. Detection of circulating anti-Mucin 1 (MUC1) antibodies in breast tumor patients by indirect enzyme-linked immunosorbent assay using a recombinant MUC1 protein containing six tandem repeats and expressed in Escherichia coli. *Clin. Vaccine Immunol.* **17,** 1903–1908 (2010).

49. Laidi, F., Bouziane, A., Errachid, A. & Zaoui, F. Usefulness of salivary and serum auto-antibodies against tumor biomarkers HER2 and MUC1 in breast cancer screening. *Asian Pacific J. Cancer Prev.* **17,** 335–339 (2016).

50. Robertson, J. F. R., Chapman, C., Cheung, K., Murray, A. & Pinder, S. E. Autoantibodies in early breast cancer. *J. Clin. Oncol.* **23,** 2005:549 (2005).

51. Tang, Y., Cui, X., Xiao, H., Qi, S., Hu, X., Yu, Q., Shi, G., Zhang, X., Gu, J., Yu, Y., Wang, L. &

Li, Y. Binding of circulating anti-MUC1 antibody and serum MUC1 antigen in stage IV breast cancer. *Mol. Med. Rep.* **15,** 2659–2664 (2017).

52.   Menekse, E., Mckolanis, J., Finn, O. J., Mcauliffe, P. F., Johnson, R. & Soran, A. Anti-MUC1 antibody in nipple aspirate fluids correlates with tumor aggressiveness in breast cancer : A feasibility study. *Dis. Markers* **2015,** 1–6 (2015).

53.   von Mensdorff-Pouilly, S., Gourevitch, M. M., Kenemans, P., Verstraeten, A. A., Litvinov, S. V, Kamp, G. J. Van, Meijer, S., Vermorken, J. & Hilgers, J. Humoral immune response to polymorphic epithelial mucin (MUC-1) in patients with benign and malignant breast tumours. *Eur. J. Cancer* **32A,** 1325–1331 (1996).

54.   Mensdorff-pouilly, S. Von, Kok, A., Paul, M. A. & Hilgers, J. An Enzyme-Linked Immunosorbent Assay for the measurement of polymorphic epithelial mucin (MUC-1). *Tumor Biol.* **19,** 186–195 (1998).

55.   Croce, V., Isla-larrain, M. T., Demichelis, S. O., Gori, J. R., Price, R. & Segal-eiras, A. Tissue and serum MUC1 mucin detection in breast cancer patients. *Breast Cancer Res. Treat.* **81,** 195–207 (2003).

56.   Gheybi, E., Amani, J. & Salmanian, A. H. Designing a recombinant chimeric construct contain MUC1 and HER2 extracellular domain for prediagnostic breast cancer. *Tumor Biol.* **35,** 11489–11497 (2014).

57.   Croce, M. V, Isla-larrain, M. T., Capafons, A. & Price, M. R. Humoral immune response induced by the protein core of MUC1 mucin in pregnant and healthy women. *Breast Cancer Res. Treat.* **69,** 1–11 (2001).

58.   Nakamura, H., Hinoda, Y., Nakagawa, N., Makiguchi, Y., Itoh, F. & Endo, T. Detection of circulating anti-MUC1 mucin core protein antibodies in patients with colorectal cancer. *J. Gastroenterol.* **33,** 354–361 (1998).

59.   Willman, A., Schoen, R. E., Potter, D. M. & Finn, O. J. Humoral immune response to abnormal MUC1 in subjects with colorectal adenoma and cancer. *Mol. Immunol.* **47,** 52–56 (2009).

60.   Wang, Y., Zhang, H., Liu, C., Xia, Q. & Wu, H. Correlation between auto-antibodies to survivin and MUC1 variable number tandem repeats in colorectal cancer. *Asian Pacific J. Cancer Prev.* **13,** 5557–5562 (2012).

61.   Pedersen, J. W., Blixt, O., Bennett, E. P., Tarp, M. A., Dar, I., Mandel, U., Poulsen, S. S., Pedersen, A. E., Rasmussen, S., Jess, P., Clausen, H. & Wandall, H. H. Seromic profiling of colorectal cancer patients with novel glycopeptide microarray. *Int. J. Cancer* **128,** 1860–1871 (2011).

62.   Pedersen, J. W., Gentry-Maharaj, A., Alexander, N., Fourkala, E., Dawnay, A., Burnell, M., Zaikin, A., Burchell, J., Papadimitriou, J. T., Clausen, H., Jacobs, I., Menon, U. & Wandall, H. H. Cancer-associated autoantibodies to MUC1 and MUC4—A blinded case–control study of colorectal cancer in UK Collaborative Trial of Ovarian Cancer Screening. *Int. J. Cancer* **134,** 2180–2188 (2014).

63.   Chen, H., Werner, S., Butt, J., Zörnig, I., Knebel, P., Eichmüller, S. B., Jäger, D., Waterboer, T., Pawlita, M. & Brenner, H. Prospective evaluation of 64 serum autoantibodies as biomarkers for early detection of colorectal cancer in a true screening setting. *Oncotarget* **7,** 16420–16432 (2016).

64.   Chapman, C. J., Murray, A., Mcelveen, J. E., Sahin, U. & Luxemburger, U. Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. *Thorax* **63,** 228–233 (2008).

65.   Hirasawa, Y., Kohno, N., Yokoyama, A., Kondo, K., Hiwada, K. & Miyake, M. Natural autoantibody to MUC1 is a prognostic indicator for non-small cell lung cancer. *Am. J. Respir. Crit. Care Med.* **161,** 589–594 (2000).

66.   Hamanaka, Y., Suehiro, Y., Fukui, M., Shikichi, K., Imai, K. & Hinoda, Y. Circulating anti-MUC1 IgG antibodies as a favorable prognostic factor for pancreatic cancer. *Int. J. Cancer* **103,** 97–100 (2003).

67. Klaamas, K., Kurtenkov, O., Shljapnikova, L., Miljukhina, L., Brjalin, V. & Lipping, A. Impact of Helicobacter pylori infection on the humoral immune response to MUC1 peptide in patients with chronic gastric diseases. *Immunol. Invest.* **36,** 371–386 (2007).

68. Treon, S. P., Maimonis, P., Bua, D., Young, G., Raje, N., Mollick, J., Chauhan, D., Tai, Y., Hideshima, T., Shima, Y., Hilgers, J., Von, S., Belch, A. R., Pilarski, L. M., Anderson, K. C., Treon, S. P., Maimonis, P., … Tai, Y. Elevated soluble MUC1 levels and decreased anti-MUC1 antibody levels in patients with multiple myeloma Elevated soluble MUC1 levels and decreased anti-MUC1 antibody levels in patients with multiple myeloma. *Blood J.* **96,** 3147–3153 (2000).

69. Jin, Y., Kim, S. C., Kim, H. J., Ju, W., Kim, Y. H. & Kim, H. Use of autoantibodies against tumor-associated antigens as serum biomarkers for primary screening of cervical cancer. *Oncotarget* **8,** 105425–105439 (2017).

70. O'Reilly, J.-A. & O'Kennedy, R. J. Prostate cancer detection: complexities and strategies. *J. Cancer Treat. Diagnosis* **2,** 18–25 (2017).

71. Catalona, W. J., Smith, D. S., Ratliff, T. L., Dodds, K. M., Coplen, D. E., Yuan, J. J. J., Petros, J. A. & Andriole, G. L. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* **324,** 1156–1161 (1991).

72. Loeb, S. & Catalona, W. J. Prostate-specific antigen in clinical practice. *Cancer Lett.* **249,** 30–39 (2007).

73. Peracaula, R., Tabarés, G., Royle, L., Harvey, D. J., Dwek, R. A., Rudd, P. M. & de Llorens, R. Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology* **13,** 457–470 (2003).

74. Ulmert, D., O'Brien, F. M., Bjartell, A. S. & Lilja, H. Prostate kallikrein markers in diagnosis, risk stratification and prognosis. *Nat. Rev. Urol.* **6,** 384–391 (2009).

75. Makarov, D. V, Loeb, S., Getzenberg, R. H. & Partin, A. W. Biomarkers for prostate cancer. *Annu. Rev. Med.* **60,** 139–151 (2009).

76. Stephan, C., Rittenhouse, H., Hu, X., Cammann, H. & Jung, K. Prostate-specific antigen (PSA) screening and new biomarkers for prostate cancer (PCa). *eJIFCC* **25,** 55–78 (2014).

77. Llop, E., Ferrer-Batallé, M., Barrabés, S., Guerrero, P. E., Ramírez, M., Saldova, R., Rudd, P. M., Aleixandre, R. N., Comet, J., de Llorens, R. & Peracaula, R. Improvement of prostate cancer diagnosis by detecting PSA glycosylation-specific changes. *Theranostics* **6,** 1190–1204 (2016).

78. Schipper, M., Wang, G., Giles, N. & Ohrnberger, J. Novel prostate cancer biomarkers derived from autoantibody signatures. *Transl. Oncol.* **8,** 106–111 (2015).

79. Massoner, P., Lueking, A., Goehler, H., Höpfner, A., Kowald, A., Kugler, K. G., Amersdorfer, P., Horninger, W., Bartsch, G., Schulz-Knappe, P., Klocker, H., Schulz-knappe, P., Klocker, H. & Ag, P. Serum-autoantibodies for discovery of prostate cancer specific biomarkers. *Prostate* **72,** 427–436 (2012).

80. 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. Deciphering the non-equivalence of serine and threonine O-glycosylation points: implications for molecular recognition of the Tn antigen by an anti-MUC1 antibody. *Angew. Chem. Int. Ed.* **127,** 9968–9972 (2015).

81. Burchell, J., Taylor-Papadimitriou, J., Boshell, M., Gendler, S. & Duhig, T. A short sequence, within the amino acid tandem repeat of a cancer-associated mucin, contains immunodominant epitopes. *Int. J. Cancer* **44,** 691–696 (1989).

82. Price, M. R., Hudecz, F., Sullivan, C. O., Baldwin, R. W., Edwards, P. M. & Tendler, S. J. B. Immunological and structural features of the protein core of human polymorphic epithelial mucin. *Mol. Immunol.* **27,** 795–802 (1990).

83. 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. Crystal structure at 1.95

Å resolution of the breast tumour-specific antibody SM3 complexed with its peptide epitope reveals novel hypervariable loop recognition. *J. Mol. Biol.* **284,** 713–728 (1998).

84.   Asensio, J. L., Ardá, A., Cañada, F. J. & Jiménez-Barbero, J. Carbohydrate - Aromatic Interactions. *Acc. Chem. Res.* **46,** 946–954 (2013).

85.   Jiménez-Moreno, E., Jiménez-Osés, G., Gómez, A. M., Santana, A. G., Corzana, F., Bastida, A., Jiménez-Barbero, J. & Asensio, J. L. A thorough experimental study of CH/π interactions in water: quantitative structure–stability relationships for carbohydrate/aromatic complexes. *Chem. Sci.* **6,** 6076–6085 (2015).

86.   Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L. & Woolfson, D. N. Carbohydrate–aromatic Interactions in proteins. *J. Am. Chem. Soc.* **137,** 15152–15160 (2015).

87.   Hsu, C., Park, S., Mortenson, D. E., Foley, B. L., Wang, X., Woods, R. J., Case, D. A., Powers, E. T., Wong, C., Dyson, H. J. & Kelly, W. The dependence of carbohydrate–aromatic interaction strengths on the structure of the carbohydrate. *J. Am. Chem. Soc.* **138,** 7636–7648 (2016).

88.   Unione, L., Echeverria, B., Serna, S., Franconetti, A., Caçada, F. J., Diercks, T. & Reichardt, N. Fluoroacetamide moieties as NMR spectroscopy probes for the molecular recognition of GlcNAc-containing sugars: modulation of the CH – pi stacking interactions by different fluorination patterns. *Chem. Eur. J.* **23,** 3957–3965 (2017).

89.   Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* **120,** 215–241 (2008).

90.   Kaiser, E., Colescott, R. L., Bossinger, C. D. & Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* **34,** 595–598 (1970).

91.   Karsten, U., Serttas, N., Paulsen, H., Danielczyk, A. & Goletz, S. Binding patterns of DTR-specific antibodies reveal a glycosylation-conditioned tumor-specific epitope of the epithelial mucin (MUC1). *Glycobiology* **14,** 681–692 (2004).

92.   Corzana, F., Busto, J. H., Jiménez-Osés, G., De Luis, M. G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M. & Avenoza, A. Serine versus threonine glycosylation: The methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *J. Am. Chem. Soc.* **129,** 9458–9467 (2007).

93.   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. The use of fluoroproline in MUC1 antigen enables efficient detection of antibodies in patients with prostate cancer. *J. Am. Chem. Soc.* **139,** 18255–18261 (2017).

International short-term research stay

6

6.1. Synthesis of the STn antigen

To expand the scope of this dissertation beyond the Tn antigen, we proposed as the main goal for a short-term stay the synthesis of the STn antigen and its further incorporation into a MUC1-like peptide epitope. This work was conducted in the lab of Dr. Westerlind (Leibniz-Institut für Analytische Wissenschaften, ISAS) in Dortmund, Germany.

As mentioned in Introduction section, the STn antigen (α−Neu5Ac-2,6-GalNAc-α-*O*-Ser/Thr, Figure 6.1) is a truncated *O*-glycan containing a sialic acid α-2,6 linked to the Tn antigen (α-*O*-GalNAc-Ser/Thr). In healthy cells sialic acid terminates the carbohydrate chain, hence paying a fundamental role in cellular recognition and cell adhesion and signaling;[1,2] whereas the STn antigen appears exposed to the immune system on the surface of tumor epithelial cells, in a similar fashion as the Tn antigen. For this reason, the STn antigen has also been extensively applied in cancer therapy.[3–9] Of note, STn is often co-expressed with the Tn antigen in cancer tissues, but in contrast to this one, the STn antigen is not a normal synthetic bioprecursor, meaning that its expression is necessarily pathologic.[10,11]
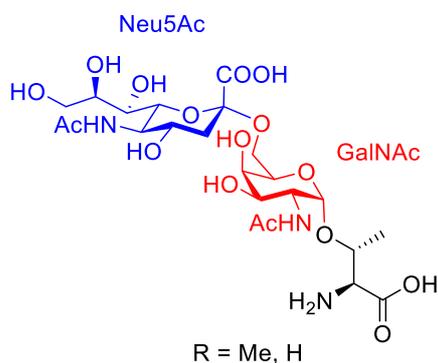
Figure 6.1 STn antigen structure.

The synthetic route followed to obtain both **STn-Ser** and **STn-Thr** proceeds as follows: first the Tn antigen skeleton is synthesized and then sialic acid is incorporated. To learn a different procedure from that of our own laboratory, we decided to start the synthesis from the beginning. Therefore, the azidosugar donor was obtained as described in Scheme 6.1.[12,13] Carbohydrate halide **6-I** was reduced to compound **6-II** with zinc and copper sulfate, which was then treated with ammonium cerium (IV) nitrate (CAN) and sodium azide (NaN$_3$) in acetonitrile under argon atmosphere. After the purification of compound **6-III** by column chromatography, it was treated with lithium bromide (LiBr) to yield azidosugar donor **6-IV**.
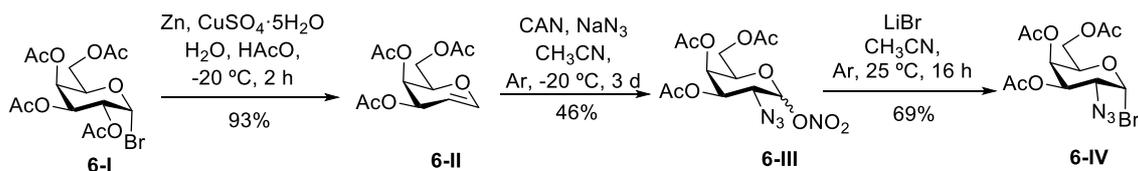
Scheme 6.1 Synthesis of the azidosugar moiety.

This sugar moiety was then reacted with either Ser (**6-VIII**) or Thr (**6-VII**) amino acids, conveniently protected, through the Koenigs-Knorr methodology to give derivatives **6-IX** and **6-X**, respectively. Azide group was subsequently transformed into the acetamide group of the GalNAc unit after treatment with thioacetic acid in pyridine, allowing to obtain compounds **Tn-Thr''** and **Tn-Ser''** after purification by column chromatography. (Scheme 6.2).
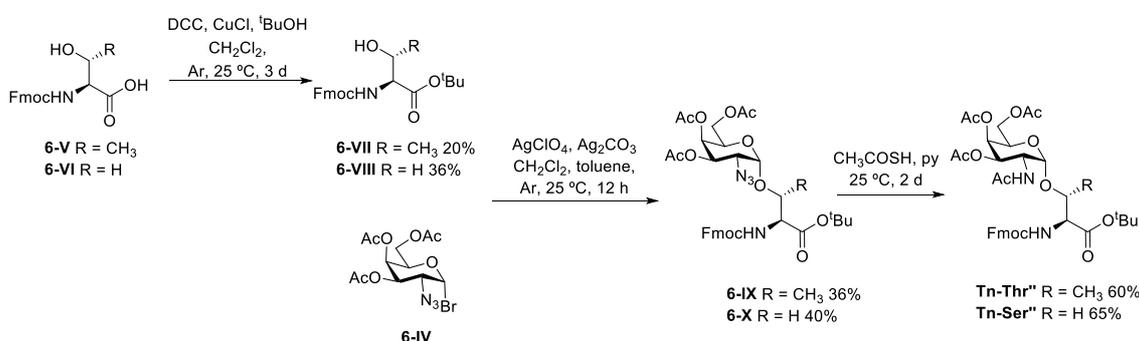
Scheme 6.2 Synthetic route followed to obtain **Tn-Ser''** *and* **Tn-Thr''** *derivatives.*

In parallel, commercially available sialic acid was protected as described in Scheme 6.3 to yield glycosyl donor **6-XV**, which was used to form the disaccharide building block. Firstly, hydroxyl groups were protected as acetyl groups (**6-XII**) and carboxylic acid as benzyl ester to obtain compound **6-XIII**, which was subsequently transformed into sialyl chloride **6-XIV**. Compound **6-XIV** was allowed to react with *O*-ethyl *S*-potassium dithiocarbonate in ethanol to obtain α-sialyl xanthate **6-XV**.[14]
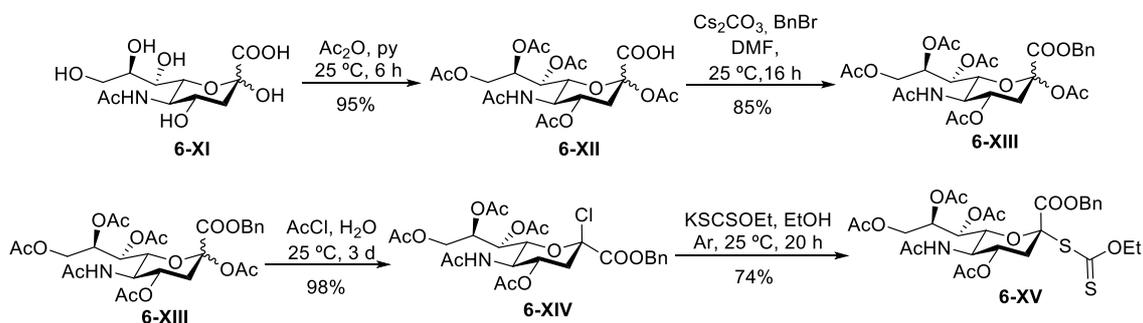
*Scheme 6.3 Synthetic route followed to obtain α-sialyl xanthate (**XV**).*

Prior to glycosylation, acetyl groups of **Tn-Thr''/Ser''** were removed with MeONa/MeOH to give derivatives **6-XVI** and **6-XVII**, respectively. Next, glycosyl donor **6-XV** was coupled

to them in the presence of silver triflate and methanesulfenyl bromide to obtain respective derivatives **6-XVIII** and **6-XIX,** in an anomeric ratio of α/β 4:1 for both compounds.[15] Hydroxyl groups of the GalNAc moiety were acetylated again to give compounds **6-XX** and **6-XXI**. Finally, the removal of the *tert*-butyl group in acid media gave **STn-Thr** and **STn-Ser**, which were ready-to-use in SPPS without further purification. (Scheme 6.4).
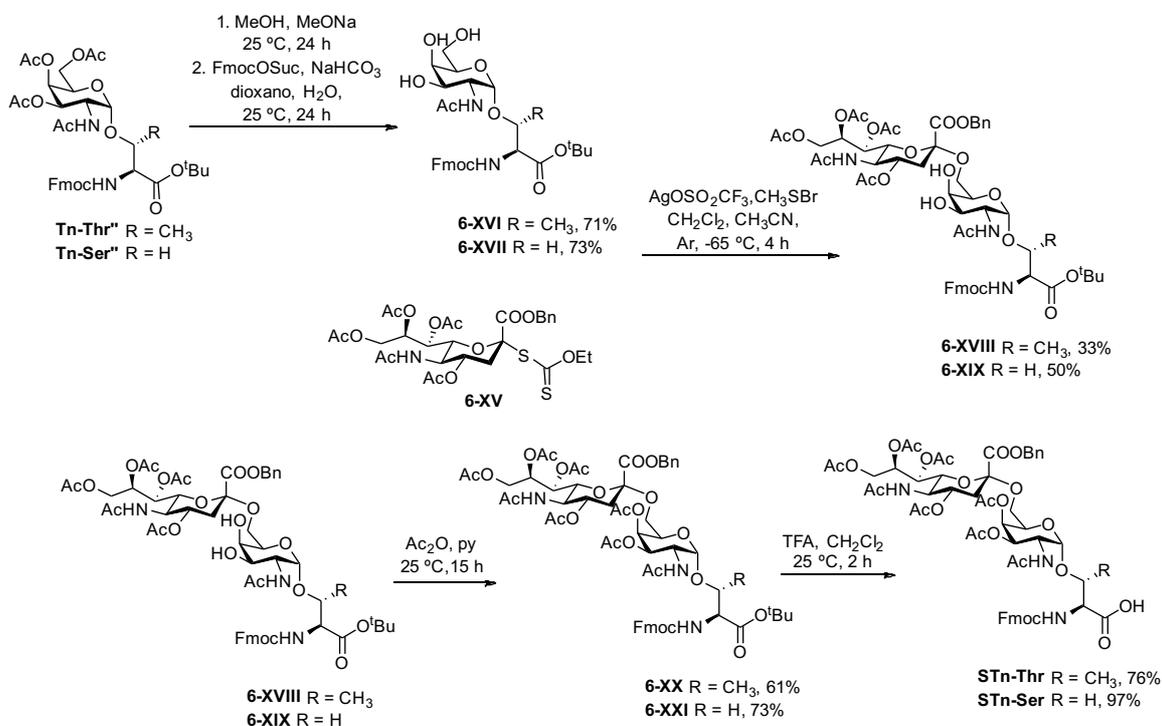
Scheme 6.4 Synthetic route followed to obtain **STn-Ser** *and* **STn-Thr** *derivatives.*

STn-Thr and **STn-Ser** were subsequently incorporated into the APDT/SRP epitope through an in-machine coupling, but with an extended reaction time compared to the naked amino acids. Once the peptide synthesis was completed, derivatives were cleaved from the resin, using a mixture of TFA/TIS/H_2O (15:0.9:0.9). The carboxylic acid was then deprotected through a hydrogenolysis reaction with Pd/C in MeOH, followed by deacetylation with MeONa/MeOH. Glycopeptides **STn-Thr*'** and **STn-Ser*'** (Figure 6.2) were finally purified by semi-preparative HPLC and lyophilized.
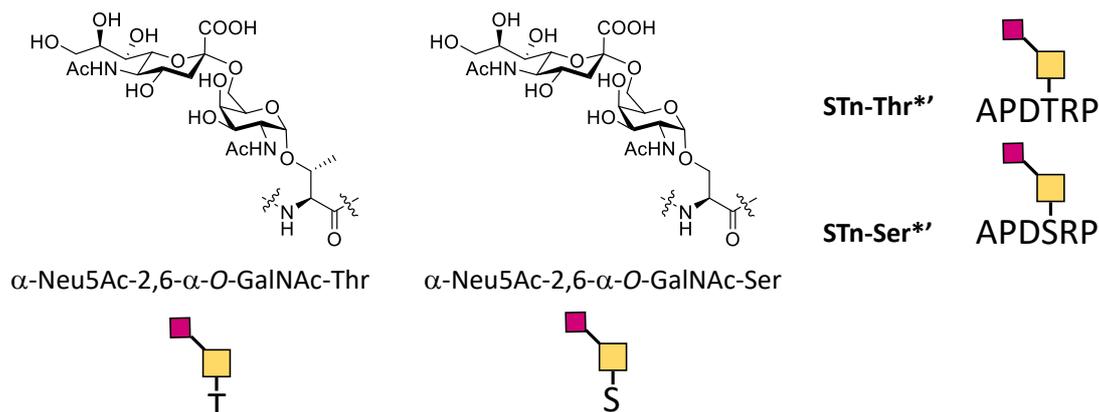
Figure 6.2 MUC1 epitopes containing the STn antigen synthesized.

Out of the scope of this Thesis, the next step will be the thorough study of the conformational preferences of both glycopeptides, **STn-Thr*'** and **STn-Ser*'**, in solution, using NMR and MD simulations, as well as in the SM3-bound state, applying X-ray crystallography. This study will result in a deeper comprehension of the STn antigen conformational preferences, thus, allowing us to understand the molecular basis of antigen-antibody recognition process as we previously showed with the Tn antigen. We also envision the synthesis of unnatural STn derivatives, conveniently modified in specific positions, with an enhanced binding affinity for antibodies. The ultimate goal of this research will be a rational design of refined tools in cancer treatment or diagnosis including modified STn moieties within their structure.

6.2. References

1.   Fu, C., Zhao, H., Wang, Y., Cai, H., Xiao, Y., Zeng, Y. & Chen, H. Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen. *HLA* **88,** 275–286 (2016).

2.   Munkley, J. The role of sialyl-Tn in cancer. *Int. J. Mol. Sci.* **17,** 275–284 (2016).

3.   Zhu, J., Wan, Q., Lee, D., Yang, G., Spassova, M. K., Ouerfelli, O., Ragupathi, G., Damani, P., Livingston, P. O. & Danishefsky, S. J. From synthesis to biologics: preclinical data on a chemistry derived anticancer vaccine. *J. Am. Chem. Soc.* **131,** 9298–9303 (2009).

4.   Stergiou, N., Glaffig, M., Jonuleit, H., Schmitt, E. & Kunz, H. Immunization with a synthetic human MUC1 glycopeptide vaccine against tumor-associated MUC1 breaks tolerance in human MUC1 transgenic mice. *Chem. Med. Chem.* **12,** 1424–1428 (2017).

5.   Palitzsch, B., Gaidzik, N., Stergiou, N., Stahn, S., Hartmann, S., Gerlitzki, B., Teusch, N., Flemming, P., Schmitt, E. & Kunz, H. A synthetic glycopeptide vaccine for the induction of a monoclonal antibody that differentiates between normal and tumor mammary cells and enables the diagnosis of human pancreatic cancer. *Angew. Chem. Int. Ed.* **55,** 2894–2898 (2016).

6.   Westerlind, U. & Kunz, H. Synthetic vaccines from tumor-associated glycopeptide antigens. *Chimia* **65,** 30–34 (2011).

7.   Keil, S., Claus, C., Dippold, W. & Kunz, H. Towards the development of antitumor vaccines: A synthetic conjugate of a tumor-associated MUC1 glycopeptide antigen and a tetanus toxin epitope. *Angew. Chem. Int. Ed.* **40,** 366–369 (2001).

8.   Wittrock, S., Becker, T. & Kunz, H. Synthetic vaccines of tumor-associated glycopeptide antigens by immune-compatible thioether linkage to bovine serum albumin. *Angew. Chem. Int. Ed.* **46,** 5226–5230 (2007).

9.   Wu, J. & Guo, Z. Improving the antigenicity of sTn antigen by modification of its sialic acid residue for development of glycoconjugate cancer vaccines. *Bioconjug. Chem.* **17,** 1537–1544 (2006).

10. Kudelka, M. R., Ju, T., Heimburg-Molinaro, J. & Cummings, R. D. *Simple sugars to complex disease-mucin-type O-glycans in cancer*. *Advances in Cancer Research* **126,** (Elsevier Inc., 2015).

11. Loureiro, L. R., Carrascal, M. A., Barbas, A., Ramalho, J. S., Novo, C., Delannoy, P. & Videira, P. A. Challenges in antibody development against Tn and sialyl-Tn antigens. *Biomolecules* **5,** 1783–1809 (2015).

12. Lemieux, R. U. & Ratcliffe, R. M. The azidonitration of tri-O-acetyl-D-galactal. *Can. J. Chem.* **57,** 1244–1251 (1979).

13. Heggemann, C., Budke, C., Schomburg, B., Majer, Z., Wibrock, M., Koop, T. & Sewald, N. Antifreeze glycopeptide analogues: microwave-enhanced synthesis and functional studies. *Amino Acids* **38,** 213–222 (2010).

14. Martichonok, V. & Whitesides, G. M. Stereoselective α-sialylation with sialyl xanthate and phenylsulfenyl triflate as a promotor. *J. Org. Chem.* **61,** 1702–1706 (1996).

15. Liebe, B. & Kunz, H. Synthesis of sialyl-Tn antigen. Regioselective sialylation of a galactosamine threonine conjugate unblocked in the carbohydrate portion. *Tetrahedron Lett.* **35,** 8777–8778 (1994).

Conclusions

7

7.1. Conclusions

The following conclusions can be drawn from the work described along this dissertation:

❖ Benzylamide derivatives of both Tn antigens (α-*O*-GalNAc-Ser and α-*O*-GalNAc-Thr) have been synthesized and their conformational preferences studied in solution. As it occurs with the methylated variant, whereas the sugar moiety appears perpendicularly oriented with respect to the backbone in the Thr derivative, in the case of the Ser analogue, the sugar moiety adopts a parallel arrangement over the backbone of the underlying peptide. This work has been completed with the study of these compounds in the gas phase by analyzing their infrared ion-dip spectra. The data revealed that the two Tn antigen variants present identical geometry free of any interference of solvent. This fact highlights the strong influence that the water molecules within the first hydration shell plays in sculpting the 3D structure of these determinants in solution. Thus, water molecules prompt the rotation around the glycosidic linkage in the threonine derivative, forcing the eclipsed conformation of this linkage to shield the hydrophobic methyl group and allow an optimal solvation of the polar region of the antigen. The unusual arrangement of α-*O*-GalNAc-Thr features a water molecule bound into a "pocket" between the sugar and the threonine. This mechanism is supported by trapping such localized water in the crystal structures of the SM3 antibody bound to two glycopeptides that comprise fluorinated Tn antigens in their structure

❖ The impact that a change in the β-methyl group of Thr9 causes on the 20-mer MUC1 mucin conformation and its recognition by antibodies has been studied. To this purpose, Thr9 residue was replaced by its epimer at the Cβ, *allo*Thr, and (2S,3R)-hydroxy-L-norvaline (Hnv), which displays a β-ethyl group instead of the β-methyl one. The binding studies conducted on several synthesized MUC1-like glycopeptides point out the significance that the the β-methyl group of Thr has in shape complementary with the antibodies. In fact , the *allo*-Thr analogue, in which the methyl group is located in a different region, the CH/π interaction is not possible, causing a significant drop in the affinity. Notably, a comparable binding affinity was observed for the glycosylated Hnv derivative and its natural counterpart. The crystal structure of this unnatural epitope in complex to scFV-SM3 antibody revealed a similar pattern of interactions than in the natural Thr variant, with a CH/π interaction between the ethyl group and Tyr32L of the SM3. Altogether, this study indicates that a MUC1-like peptide containing the glycosylated Hnv moiety may be a promising candidate for the designing an efficacious anti-cancer vaccine.

❖ An unnatural Tn glycomimetic incorporating a sp^2-iminosugar was used to synthesize an anti-cancer vaccine. The *in vivo* experiments performed with the

novel vaccine showed that this scaffold was more immunogenic than its natural Thr-derived counterpart. The study also demonstrated that the antibodies elicited by the unnatural vaccine were capable of recognizing natural MUC1 glycopeptides, which represent an essential feature in vaccine design.

❖ The stacking of Pro7 in the antigen against aromatic residues of the SM3 antibody is a crucial feature of the antibody–antigen complex. The substitution of Pro7 within MUC1 mucin backbone by two fluorinated derivatives, either (4*S*)-4-fluoro-L-proline or 4,4-difluoro-L-proline allowed to enhance the strength of this interaction. Besides, crystallographic studies proved that fluorinated glycopeptides conformation remains unchanged compared to the natural Thr epitope. A shortening in the in the distance distribution of Pro7-Trp91L, due to the influence of hydrogen-by-fluorine substitution, was observed. The fluorinated derivatives also presented improved sensibility for antibodies on serum samples from prostatic adenocarcinoma, compared to natural MUC1 mucin.

❖ In a short-time research stay in the lab of Dr. Westerlind (Leibniz-Institut für Analytische Wissenschaften, ISAS, Germany) the STn antigens (α-Neu5-2,6-α-*O*-GalNAc-Thr and α-Neu5-2,6-α-*O*-GalNAc-Ser) were synthesized and incorporated into the short epitope of MUC1. Out of the scope of this dissertation, the conformation of these glycopeptides will be studied in solution and complexed to the SM3 anti-MUC1 antibody by X-ray diffraction.

❖ The research outcomes herein presented were deduced by a multidisciplinary approach that involves the use of diverse techniques, such as Nuclear Magnetic Resonance (NMR), Molecular Dynamics simulations (MD), Quantum Mechanical calculations (QM), X-ray diffraction, Bio-Layer Interferometry (BLI), Surface Plasmon Resonance (SPR), micro array and Enzyme-Linked Immunosorbent Assays (ELISA).

❖ The results compiled in this dissertation demonstrate, firstly, the non-equivalence of Ser and Thr residues when are part of the Tn antigen and the key role of water in the modulation of their conformational preferences. Secondly, based on a rational design derived from structural data, the incorporation of unnatural residues, either within Tn antigen moiety or substituting a MUC1 mucin amino acid, has proved to enhance binding affinity towards antibodies, resulting in an improved vaccination outcome or more sensible tumor biomarkers.

7.2. Conclusiones

Las conclusiones derivadas del trabajo recogido en la presente Tesis Doctoral son las siguientes:

❖ Han sido sintetizados derivados del antígeno Tn, α-*O*-GalNAc-Ser y α-*O*-GalNAc-Thr, cuyo extremo amino se encuentra protegido con benzilamida. Se ha analizado su conformación en disolución, mostrando que el carbohidrato se presenta de manera diferente para ambos. Mientras que el GalNAc se dispone de manera perpendicular con respecto a la cadena peptídica en el derivado de Thr, en el caso de la Ser éste adopta una disposición paralela sobre el esqueleto peptídico. Este trabajo se ha completado con el estudio de estos compuestos en fase gas, analizándolos mediante espectroscopia infrarroja, en la cual presentan la misma geometría. Este hecho pone de manifiesto la fuerte influencia que la solvatación del agua ejerce sobre el antígeno Tn, conformando la estructura tridimensional de estos compuestos en disolución. Así, las moléculas de agua provocan un giro alrededor del enlace glicosídico en el caso del derivado de treonina, forzando una conformación eclipsada que oculta el grupo metilo hidrofóbico y permite una solvatación óptima de la región polar del antígeno. Esta disposición inusual de la α-*O*-GalNAc-Thr permite la aparición de una molécula de agua puente entre el azúcar y la treonina, cuya presencia ha sido demostrada mediante la captura de ésta en las estructuras cristalinas del anticuerpo SM3 unido a dos glicopéptidos que incluyen derivados fluorados del antígeno Tn en su estructura.

❖ Se ha estudiado la influencia que un cambio en el grupo metilo en posición β de la treonina 9 causa en la conformación de la mucina MUC1 de 20 residuos y su reconocimiento molecular mediante anticuerpos. El residuo de Thr9 fue reemplazado por su epímero en el carbono β, *allo*Thr, y por el aminoácido (2S,3R)-hidroxy-L-norvalina (Hnv), el cual presenta un grupo etilo en β en vez de un metilo. Estudios de afinidad llevados a cabo en diversos glicopéptidos de tipo MUC1 señalan la importancia que el grupo β-metilo de la Thr tiene en el reconocimiento mediante anticuerpos. De hecho, el análogo de *allo*-Thr, en el que el grupo metilo se localiza en una región diferente, carece de una interacción CH/π fundamental, causando una notable disminución de la afinidad. Sin embargo, en el derivado glicosilado de Hnv se observa una afinidad comparable a la del derivado natural. La estructura cristalina de este epítopo no natural acomplejado con el anticuerpo scFV-SM3 presenta un patrón de interacciones muy similar al del derivado de treonina natural, observándose una interacción CH/π entre el grupo etilo y la Tyr32L del SM3. Por estos motivos, un péptido derivado de MUC1 que contenga un aminoácido de Hnv glicosilada representa un candidato prometedor en el diseño de una vacuna contra el cáncer.

❖ Un glicomimético del antígeno Tn derivado de iminoazúcares de tipo sp^2 ha sido incorporado en una vacuna derivada de MUC1. Ensayos in vivo han demostrado que este derivado es más inmunogénico que el derivado natural de Thr con la

misma estructura. Además, los anticuerpos que fueron generados tras el tratamiento con la vacuna no natural han sido capaces de reconocer glicopéptidos de tipo MUC1 naturales, siendo ésta una característica esencial para el diseño de vacunas de este tipo.

❖ La interacción de la Pro7 del antígeno sobre residuos aromáticos del anticuerpo SM3 es fundamental en el complejo antígeno-anticuerpo. La sustitución de la Pro7 dentro del esqueleto peptídico de la mucina MUC1 por dos derivados fluorados, bien sea (4*S*)-4-fluoro-L-prolina o 4,4-difluoro-L-prolina, permite aumentar su interacción en este tipo de sistemas. Además, posteriores estudios cristalográficos han probado que los glicopéptidos fluorados mantienen la misma conformación que su equivalente natural. También se ha observado una disminución en la distancia de la interacción Pro7-Trp91L, debido a la influencia de la sustitución de hidrógeno por flúor. Así mismo, los derivados fluorados han presentado una mejora en la sensibilidad a la hora de reconocer anticuerpos obtenidos en muestras de suero de pacientes de cáncer de próstata, en comparación con las mucinas naturales.

❖ Durante una estancia breve de investigación en el laboratorio de la Dr. Westerlind (ISAS, Alemania) se llevó a cabo la síntesis de los antígenos STn (α-Neu5-2,6-α-*O*-GalNAc-Thr y α-Neu5-2,6-α-*O*-GalNAc-Ser), los cuales fueron incorporados en el epítopo de la MUC1. Fuera de las competencias de esta tesis, la conformación de estos glicopéptidos será estudiada en disolución o formando un complejo con el anticuerpo anti-MUC1 SM3 mediante difracción de rayos X.

❖ Los resultados de esta investigación han sido obtenidos mediante una aproximación multidisciplinar que comprende el empleo de diversas técnicas como por ejemplo, Resonancia Magnética Nuclear, cálculos de Dinámica Molecular y Mecánica Cuántica, difracción de rayos X y experimentos para conocer afinidad entre sistemas como Bio-Layer Interferometry (BLI), Surface Plasmon Resonance (SPR), micro arrays y Enzyme-Linked Immunosorbent Assays (ELISA).

❖ Los resultados derivados de esta Tesis demuestran, en primer lugar, que los residuos de serina y treonina, componentes del antígeno Tn, no son equivalentes y el agua circundante juega un papel crucial en la modulación de su conformación. En segundo lugar, que basándonos en un diseño racional a partir de datos estructurales, es posible incorporar aminoácidos no naturales, ya sea modificando la estructura del antígeno Tn o sustituyendo otro aminoácido de la mucina MUC1; de manera que estos cambios permitan mejorar la afinidad hacia anticuerpos, dando como resultado una mejora en su uso como vacunas o marcadores tumorales más sensibles.

7.3. Scientific publications derived from this dissertation

❖ <u>Derived from chapter 3</u>*: Water Sculpts the Distinctive Shapes and Dynamics of the Tn Antigens: Implications for their Molecular Recognition.*
<u>I. A. Bermejo,</u> I. Usabiaga, I. Compañón, J. Castro-López, A. Insausti, J. A. Fernández, A. Avenoza, J. H. Busto, J. Jiménez-barbero, J. L. Asensio, J. M. Peregrina, G. Jiménez-Osés, R. Hurtado-Guerrero, E. J. Cocinero, F. Corzana.
Journal of American Chemical Society, **2018**, *140*, 9952-9960.

Featured in JACS Spotlights: *J. Am. Chem. Soc.* **2018**, *140*, 10383–10384.

❖ <u>Derived from chapter 5</u>*: The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer.*
V. J. Somovilla, <u>I. A. Bermejo</u>, I. S. Albuquerque, N. Martínez-Sáez, J. Castro-López, F. García-Martín, I. Compañón, H. Hinou, S.-I. Nishimura, J. Jiménez-Barbero, J. L. Asensio, A. Avenoza, J. H. Busto, R. Hurtado-Guerrero, J. M. Peregrina, G.J. L. Bernardes, F. Corzana.
Journal of American Chemical Society, **2017**, *139*, 18255-18261.

Featured in JACS Spotlights: *J. Am. Chem. Soc.* **2018**, *140*, 1–1.

❖ <u>Background of chapter 4</u>*: Mucin architecture behind the immune response: design, evaluation and conformational analysis of an antitumor vaccine derived from an unnatural MUC1 fragment.*
N. Martínez-Sáez, N. T. Supekar, M. A. Wolfert, <u>I. A. Bermejo</u>, R. Hurtado-Guerrero, J. L. Asensio, J. Jiménez-Barbero, J. H. Busto, A. Avenoza, G.-J. Boons, J. M. Peregrina, F. Corzana.
Chemical Science, **2016**, *7*, 2294-2301.

7.4. Other scientific publications

❖ *Nucleoside azide–alkyne cycloaddition reactions under solvothermal conditions or using copper vials in a ball mill.*
A. J. Cummings, F. Ravalico, K. I. S. McColgan-Bannon, O. Eguaogie, P. A. Elliott, M. R. Shannon, <u>I. A. Bermejo</u>, A. Dwyer, A. B. Maginty, J. Mack, J. S. Vyle.
Nucleosides, Nucleotides and Nucleic Acids, **2015**, *34*, 361-370.

7.5. Contribution to congresses

❖ **Flash communication:** *Diseño de nuevas vacunas contra el cáncer basadas en glicopéptidos.* <u>I. A. Bermejo</u>, N. Martínez-Sáez, J. Castro-López, R. Hurtado-Guerrero, A. Avenoza, J. H. Busto, J. M. Peregrina, F. Corzana. XIII Simposio de

Investigadores Jóvenes RSEQ - Sigma Aldrich. Logroño (Spain). November 8-11th of 2016.

❖ **Poster:** *Synthesis of a novel Tn antigen to experimentally determine its first hydration shell*. I. A. Bermejo, A. Avenoza, J. H. Busto, E. J. Cocinero, F. Corzana, J. M. Peregrina. II Reunión Bienal del GEQB/ XIV Encuentro Peptídico Ibérico. Bilbao (Spain). September 4-6th of 2014.

❖ **Poster:** *Designing non-natural cancer vaccines by replacing the methyl group of threonine-10.* I. A. Bermejo, N. Martínez-Sáez, V. J. Somovilla, L. García-García, V. Lázaro, F. Corzana, A. Avenoza. J. H. Busto, J. M. Peregrina, G.-J. Boons. XI Carbohydrate Symposium. Logroño (Spain). May 28-30th of 2014.

❖ **Poster:** *Reversible photo-control of an antigenic peptide conformation.* I. A. Bermejo, D. Martínez-López, A. Avenoza, J. H. Busto, D. Sampedro, F. Corzana, J. M. Peregrina. XI Simposio de Investigadores Jóvenes RSEQ - Sigma Aldrich. Bilbao (Spain). November 4-7th of 2014.

❖ **Poster:** *A comparison between the conformation of threonine and* allo-*threonine in MUC1 antigens.* I. A. Bermejo, N. Martínez-Sáez, V.J. Somovilla, J.L. Asensio, J. Jiménez-Barbero, G.J. L. Bernardes, R. Hurtado-Guerrero, A. Avenoza, J. H. Busto, F. Corzana, Jesús M. Peregrina. XXXV Biennial Meeting of the RSEQ. A Coruña (Spain). July 19th-23rd of 2015.

❖ **Poster:** *New generation of cancer vaccines based on IONPs.* I. A. Bermejo, A. Ruiz de Angulo, N. Martínez-Sáez, J. C. Mareque-Rivas, J. H. Busto, A. Avenoza, J. M. Peregrina, F. Corzana. III Biennial Meeting. Madrid (Spain). March 14-16th of 2016.

❖ **Poster:** *Inmunoterapia para tratar el cáncer: mucinas modificadas.* I. A. Bermejo, N. Martínez-Sáez, A. Avenoza, J. H. Busto, F. Corzana, J. M. Peregrina. III Jornadas Doctorales. Jaca (Spain). July 4-5th of 2016.

❖ **Poster:** *Inmunoterapia para tratar el cáncer: mucinas modificadas.* I. A. Bermejo, N. Martínez-Sáez, A. Avenoza, J. H. Busto, F. Corzana, J. M. Peregrina. V Jornadas Doctorales del Grupo 9 de Universidades (G-9). Palma de Mallorca (Spain). 8-10th February 2017.

❖ **Poster:** *Untangling Tn antigen structure and its first hydration shell.* I. A. Bermejo, I. Usabiaga, A. Avenoza, J. H. Busto, E. J. Cocinero, J. M. Peregrina, F. Corzana. EUROCARB 2017. Barcelona (Spain). 2nd-6th July of 2017.

❖ **Poster:** *Untangling Tn antigen structure and its first hydration shell.* I. A. Bermejo, I. Usabiaga, A. Avenoza, J. H. Busto, E. J. Cocinero, J. M. Peregrina, F. Corzana. 28th Joint Glycobiology Meeting 2017. Aachen (Germany). 17-19th September 2017.

7.6. Poster awards

❖ *Inmunoterapia para tratar el cáncer: mucinas modificadas.*
V Jornadas Doctorales del Grupo 9 de Universidades (G-9). Palma de Mallorca, February 2017.

❖ *New generation of cancer vaccines based on IONPs.*
III Biennial Meeting of the Chemical Biology Group/ XII Carbohydrate Symposium. Madrid, March 2016.

❖ *Unnatural MUC1 derivatives based on allo-threonine as a suitable tool to get further insights into the molecular recognition of MUC1 derivatives by antibodies.*
XXXV Bienal RSEQ. A Coruña, July 2015.

Experimental section

8

8.1. Reagents and general procedures

Commercial reagents were used without further purification. Analytical thin layer chromatography (TLC) was performed on glass plates precoated with a 0.25 mm thickness of silica gel. TLC plates were visualized with UV light and by staining with phosphomolibdic acid (4.5 g) in ethanol (100 mL) or 10% H_2SO_4 in ethanol solutions. Column chromatography was performed on silicagel (230-400 mesh). Mass spectrometry analyses were performed on HP 59987B, using electron impact ionization. Electrospray-mass spectrometry (ESI-MS) were performed on the same equipment with HP 59987A interface and were registered in positive or negative ion mode. A Bruker Microtof-Q spectrometer was also used.

8.2. Solid Phase Peptide Synthesis (SPPS)

All (glyco)peptides were synthesized by a micro-wave assisted solid-phase peptide synthesis on a Liberty Blue synthesizer, using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol) and oxyma Pure and DIC as coupling agents.

Unnatural and glycosylated amino acids were coupled manually as follows: 1.5 equiv. of the corresponding amino acid together with HBTU (0.9 equiv.) and 0.25 mL of DIPEA (2.0 M in NMP) were dissolved in 1 mL of DMF. Reaction mixture was added to the resin and vortex shacked for usually 3 h, until the coupling was completed as deduced by Kaiser test.[1] The resin was then placed in the reactor vessel of the peptide synthesizer again to complete the sequence. Once the synthesis was completed, the resin was removed from the peptide synthesizer. In glycopeptides, acetate groups of the carbohydrate were deprotected using 5 mL of a hydrazine/MeOH (7:3) solution. Afterwards, resin is rinsed with MeOH and DCM.

(Glyco)peptides were released from the resin, as well as acid sensitive side-chain protecting groups, using 3 mL of TFA/TIS/H_2O (95:2.5:2.5), followed by precipitation with cold diethyl ether (10 mL). The resin was placed in the peptide synthesizer again and microwaved for 40 min at 38 °C. In Cys-containing (glyco)peptides, a solution of TFA/TIS/H_2O/EDT (92.5:2.5:2.5:2.5) was used for 3 h at 25 °C without microwave irradiation. (Glyco)peptides were centrifuged for 6 min at 6500 rpm and the supernatant solution was discarded, this process was repeated three times. Finally, (glyco)peptides were dried and redissolved in water to be purified by reverse phase HPLC on a Phenomenex Luna C18(2) column (10 μm, 250 mm x 21.2 mm) with a flow rate of 10 or 20 mL/min. UV detection was done at 212 nm.

8.3. NMR studies

^{1}H NMR experiments were performed at 300 or 400 MHz, ^{13}C NMR experiments were conducted at 75.5 or 100.6 MHz and decoupled ^{19}F{^{1}H} NMR experiments at 282 MHz.

Magnitude-mode ge-2D COSY spectra were acquired with gradients by using the *cosygpqf* pulse program with a pulse width of 90°. Phase-sensitive ge-2D HSQC spectra were acquired by using z-filter and selection before t1 removing the decoupling during acquisition by using of the *invigpndph* pulse program with CNST2 (JHC)=145.

ROESY experiments were recorded on a Bruker Avance 400 spectrometer at 298 K and pH 6.5 in H_2O/D_2O (9:1). The experiments were conducted by using phase-sensitive ge-2D ROESY with the WATERGATE method. ROESY intensities were normalized with respect to the diagonal peak at zero mixing time. The number of scans used was 32 and the mixing time was 500 ms. Distances involving NH protons were semi-quantitatively determined by integrating the volume of the corresponding cross-peaks.

Chemical shifts are given in ppm (δ) and coupling constants (*J*) in Hz. Chloroform, with TMS as internal reference, methanol and water were used as deuterated solvents. The results were processed using MestReNova software.

8.4. Unrestrained molecular dynamics (MD) simulations

All molecular dynamics (MD) simulations were carried out on in-house GPU clusters. The starting geometries for the complexes were generated from the available data deposited in the Protein Data Bank (pdb IDs: 5A2K –scFv-SM3–) and modified accordingly. Each model complex was immersed in a 10 Å-sided cube with pre-equilibrated TIP3P water molecules.[2,3]

Firstly, only the water molecules are minimized, and then heated to 300 K. The water box, together with ions, was then minimized, followed by a short MD simulation. 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 modelled using the particle-mesh-Ewald method.[4] An 8 Å cut-off was applied to Lennard-Jones interactions. MD simulations were performed with the Sander module of AMBER 12 (parm99 force field),[5] which was implemented with GAFF parameters,[6] and AMBER 16 package,[7] implemented with ff14SB,[8] GAFF[6] and GLYCAM06j[9] force fields. The parameters and charges for the unnatural amino acids were generated with the antechamber module of AMBER, using GAFF force field and AM1-BCC method for charges.[10]

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 (usually) additional 200 ns under the same simulation conditions.

8.5. MD simulations with time-averaged restraints or Experiment-guided MD simulations

The MD simulations with time averaged restraints were performed with AMBER 16 package,[7] implemented with ff14SB,[8] GAFF[6] and GLYCAM06j[9] force fields. The specific simulation conditions were also similar to those described previously by our group.[11]

The ROESY-derived distances were imposed as time-averaged constraint, applying an r^{-6} averaging. The equilibrium distance range was set to: $r_{exp} - 0.2$ Å $\leq r_{exp} \leq r_{exp} + 0.2$ Å. Trajectories were run at 298 K, with usually a decay constant of 100 ns and a time step of 1 fs. The force constants rk_2 and rk_3 used in each case were 10 kcal·mol^{-1}·Å^{-2}. The overall simulation length was 1 μs. The coordinates were saved each 1 ps, obtaining MD trajectories of 1000000 frames each. A convergence within the equilibrium distance range was obtained in the simulations.

8.6. Microarray binding assay

Microarray binding experiments were carried out by Dr. Fayna Garcia-Martin in the Hokkaido University (Hokkaido, Japan). Plastic microarrays slides (75 mm long, 25 mm wide and 1 mm thick) were purchased from Sumitomo Bakelite Co., Ltd. (Tokyo, Japan), as well as hybridization covers (60 × 25 × 0.7 mm). (Glyco)peptides were spotted by MicroSys 5100 (Cartesian Technologies, CA, USA). Anti-MUC1 mouse mAbs clon SM3 (0.2 mg/mL) was purchased from Santa Cruz Biotechnology (TX, USA) and VU-3C6 (0.86 mg/mL) from Exalpha (MA, USA). FluoroLinkTM CyTM3-labeled goat anti-mouse IgG was from Amersham Biosciences (Buckinghamshire, UK).

For the binding assay the next buffers and solutions were employed: buffer for the solution of mAb: 50 mM Tris-HCl, 100 mM NaCl, 1 mM $CaCl_2$, $MnCl_2$, $MgCl_2$, 0.05% Tween-20, 0.1% BSA, pH 7.4. Washing buffer: 50 mM Tris-HCl, 100 mM NaCl, 1 mM $CaCl_2$, $MnCl_2$, $MgCl_2$, 0.05% Triton X-100, pH 7.4.

8.6.1. Microarray printing

(Glyco)peptides were covalently immobilized to their *N*-terminal amino group with a 0.6-mm pitch using a Filgen solid spin (200 μm pin diameter). Each compound was printed in quadruplicate with 0.3 mm distance between spots of same compound and 0.6 mm gap among different compounds (Figure 8.1, left panel). Each (glyco)peptide was printed at six different concentrations from 15.6 μM to 500 μM (Figure 8.1, right panel). Cy3 labeled BSA protein (25 μg/mL) was used as grid.
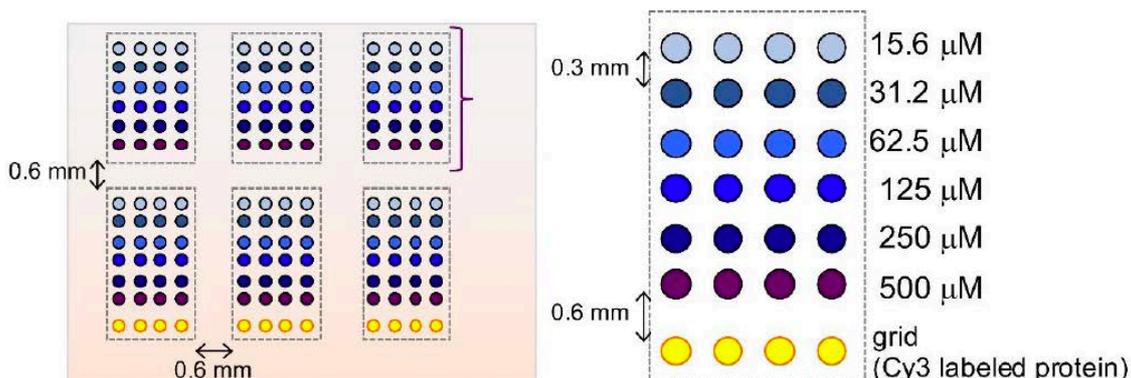
Figure 8.1 Microarray slides, schematic microarray printing on chamber slide (left panel) and printing pattern of each compound group (right panel).

Spotting conditions were 23 °C and 60% of humidity. After printing, slides were incubated overnight on dry conditions. Non-reacted groups were inactivated by blocking buffer at 37 °C for 1 h under slow agitation. Finally, slides were rinsed by washing buffer (3 x 5 min) and dried by centrifugation and then used for further binding assay of mAb.

8.6.2. Microarray mAb binding assay

For the mAb incubation, 20 µL of mAb solution in buffer (50.0 µg/mL) was carefully added onto each chamber of slides and they were kept at room temperature for 2 h on humid conditions. Afterwards, 3 x 2 min washing buffer washing steps and centrifugation drying were followed by incubation with secondary Ab (Cy3-labeled Ab), diluted to 1 µg/mL in buffer. It was infused between hybridization covers and slides for 1 h at room temperature in the absence of light.

Slides were then washed by different methods (stated for each case): (a) washing buffer (3 x 2 min) and dried up by centrifugation; (b) previous method (a) followed by washing buffer (2 x 2 min), water washing (2 x 2 min) and dried up by centrifugation. To storage the slides, they were degassed under vacuum and kept at 4 °C.

Slides were subjected to fluorescent image scanning on a Tryphoon Trio Plus instrument (GE Healthcare). Array Vision software was used to quantify the fluorescence of each spot. The average value of relative fluorescence intensity (RFU) was used; spot intensities were determined by subtracting the average pixel intensity from the median pixel intensity of the local background within the spots.

8.7. Bio-layer Interferometry

BLI experiments were conducted by Dr. Gonçalo Bernardes in Cambridge University (Cambridge, United Kingdom).

Binding assays were performed on an Octet Red Instrument (fortéBIO). Ligand immobilization, binding reactions, regeneration and washes were conducted in wells of black polypropylene 96-well microplates.

(Glyco)peptides (10 mg/mL) were immobilized on amine-reactive biosensors (AR2G biosensors) in 10 mM sodium acetate pH 5.5 buffer, using 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide and *N*-hydroxysuccinimide as coupling agent for 10 min at 1000 rpm at 25 °C. The excess of reactive esters was then blocked with a solution of ethanolamine hydrochloride (1M, pH 8.5), followed by regeneration (glycine pH 2.0 buffer) and washing.

Binding analyses were carried out at 25 °C, 1000 rpm in 10 mM sodium phosphate buffer (pH 7.4) containing 150 mM NaCl, using different concentrations of the recombinant SM3 antibody to obtain the association curves, with a 120 s of association followed by a 180 s of dissociation. The surface was thoroughly washed with the running buffer without regeneration solution.

Data were analyzed using Data Analysis (fortéBIO), with Savitzky-Golay filtering. Binding was fitted to a 2:1 heterogeneous ligand model. Steady state analysis was performed to obtain the binding kinetics constants (K_D).

8.8. Crystallization

Crystal structures were successfully obtained by Dr. Ramón Hurtado-Guerrero in the Instituto de Biocomputación y Física de Sistemas Complejos and the University of Zaragoza (Zaragoza, Spain).

Expression and purification of scFv-SM3 has been described previously.[12]

Crystals were grown by sitting drop diffusion at 18 °C. The drops were prepared by mixing 0.5 µL of protein solution, containing 15 mg/mL of scFv-SM3 and 10 mM of respective glycopeptides with 0.5 µL of the mother liquor.

Crystals of scFv-SM3 with the aforementioned glycopeptide were grown in 20% PEG 3350, 0.2 M disodium hydrogen phosphate. The crystals were cryoprotected in mother liquor containing 15% ethylene glycol and frozen in a nitrogen gas stream cooled to 100 K.

The data was processed and scaled using XDS package[13] and CCP4 software.[14] The crystal structures were solved by molecular replacement with Phaser[15] and using PDB entry 1SM3 as the template. Initial phases ware further improved by cycles of manual model building in Coot63 and refinement with REF-MAC5.[16] The final models were validated with PROCHECK.[17]

8.9. Experimental section chapter 3

8.9.1. Synthesis

Synthesis of compound 3-II

3-II
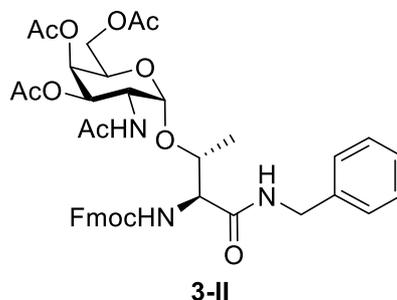
To a solution of derivative **3-I**,[31] (600 mg, 0.89 mmol) in CH_3CN (15 mL), DIPEA (311 µL, 1.79 mmol) and HBTU (339 mg, 0.89 mmol) were added. After stirring the reaction mixture for approximately 10 min, NH_2Bn (98 µL, 0.89 mmol) was added. After 1 h, EtAcO (25 mL) was added and the subsequent solution was washed with a saturated solution of K_2CO_3 (25 mL) and 0.5 M HCl (25 mL), filtered and concentrated in vacuum. The residue was purified by silica gel column chromatography (CH_2Cl_2/MeOH, 95:5) to give compound **3-II** (485 g, 71%) as a white solid.

HRMS (m/z): $[M]^+$ calcd. for $C_{40}H_{46}N_3O_{12,}$ 760.3076; found, 760.3072.

^{1}H NMR (400 MHz, CDCl₃) δ (ppm): 1.26 (d, 3H, J = 4.3 Hz, Me$_{Thr}$), 1.91 (s, 3H, NHCOC$\underline{H}_3$), 2.00 (s, 3H, COCH₃), 2.02 (s, 3H, COCH₃), 2.16 (s, 3H, COCH₃), 4.03 − 4.09 (m, 2H, 2H$_{6S}$), 4.16 − 4.31 (m, 4H, CHFmoc, H$_{5S}$, Hα, Hβ), 4.33 − 4.61 (m, 5H, H$_{2S}$, CH₂Fmoc, C$\underline{H}_2$Ph), 4.91 (s, 1H, H$_{1S}$), 5.03 (d, 1H, J = 11.0, H$_{3S}$), 5.36 (s, 1H, H$_{4S}$), 7.18 − 7.24 (m, 2H, Fmoc), 7.28 − 7.43 (m, 7H, Fmoc, Ph), 7.56 − 7.64 (m, 2H, Fmoc), 7.73 − 7.78 (m, 2H, Fmoc).

^{13}C NMR (100 MHz, CDCl₃) δ (ppm): 17.5 (Me$_{Thr}$), 20.7 ($\underline{C}H_3CO$), 20.8 ($\underline{C}H_3CO$), 20.9 ($\underline{C}H_3CO$), 23.0 (NHCO$\underline{C}H_3$), 43.8 (CH₂Ph), 47.2 (CHFmoc), 47.7 (C$_{2S}$), 58.8 (Cα), 62.0 (C$_{6S}$), 67.3 (CH₂Fmoc), 67.5 (C$_{4S}$), 67.7 (C$_{5S}$), 68.8 (C$_{3S}$), 77.4 (Cβ), 100.1 (C$_{1S}$), 120.1, 125.0, 127.2, 127.9 (Fmoc), 127.6, 127.6, 127.9, 129.0, 129.0, 137.2 (Ph), 141.4, 143.5, 143.6 (Fmoc), 156.8 (NCOO), 169.9, 170.3, 170.4, 170.7, 170.9 (CO).

Synthesis of compound 3-III

3-III
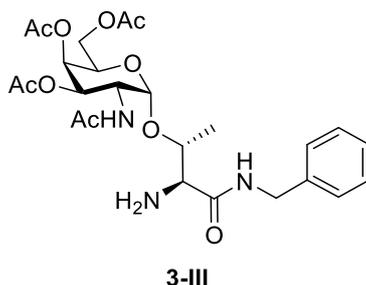
Compound **3-II** (356 mg, 0.47 mmol) was dissolved in CH_3CN/piperidine (4 mL, 3:1). After stirring the mixture for 30 min at room temperature, the solvent was removed, and the crude product was purified by silica gel column chromatography (hexane/ethyl acetate, 1:1, then CH_2Cl_2/MeOH, 9:1) to give building block **3-III** (229 mg, 91%) as a white solid.

HRMS (m/z): $[M]^+$ calcd. for $C_{25}H_{36}N_3O_{10}$, 538.2395; found, 538.2396.

^{1}H NMR (400 MHz, CDCl$_3$) δ (ppm): 1.31 (d, 3H, J = 6.4 Hz, Me$_{Thr}$), 1.93 (s, 3H, NHCOC$\underline{H}_3$), 1.99 (s, 3H, COCH$_3$), 2.04 (s, 3H, COCH$_3$), 2.16 (s, 3H, COCH$_3$), 3.34 (d, 1H, J = 2.7 Hz, Hα), 4.08 (d, 2H, J = 6.6 Hz, 2H$_{6S}$), 4.27 (td, 1H, J = 6.6, 1.3 Hz, H$_{5S}$), 4.39 − 4.47 (m, 2H, Hβ, C$\underline{H}_2$Ph), 4.53 − 4.61 (m, 2H, C$\underline{H}_2$Ph, H$_{2S}$), 4.95 (d, 1H, J = 3.5 Hz, H$_{1S}$), 5.02 (dd, 1H, J = 11.3, 3.2 Hz, H$_{3S}$), 5.37 (dd, 1H, J = 3.1, 1.3 Hz, H$_{4S}$), 7.25 − 7.31 (m, 3H, Bn), 7.32 − 7.37 (m, 2H, Bn).

^{13}C NMR (100 MHz, CDCl$_3$) δ (ppm): 16.9 (Me$_{Thr}$), 20.7 ($\underline{C}$H$_3$CO), 20.8 ($\underline{C}$H$_3$CO), 20.9 ($\underline{C}$H$_3$CO), 22.9 (NHCO$\underline{C}$H$_3$), 43.4 ($\underline{C}$H$_2$Ph), 47.8 (C$_{2S}$), 59.0 (Cα), 62.0 (C$_{6S}$), 67.4 (C$_{4S}$), 67.5 (C$_{5S}$), 68.9 (C$_{3S}$), 77.4 (Cβ), 98.9 (C$_{1S}$), 127.6, 127.6, 128.8, 128.8, 138.0 (Ph), 170.5, 170.8, 170.9, 172.8 (CO).

Synthesis of compound 3-IV

3-IV
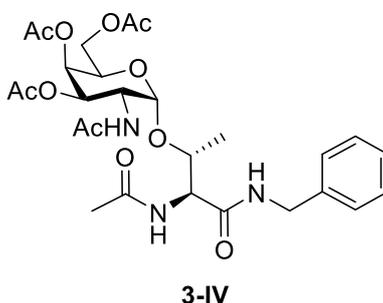
Compound **3-III** (229 mg, 0.43 mmol) was dissolved in Ac$_2$O/pyridine (4.5 mL, 2:1). After stirring the mixture for 1 h at room temperature, the solvent was removed, and the crude product was purified by silica gel column chromatography (CH$_2$Cl$_2$/MeOH, 9:1) to give building block **3-IV** (185 mg, 75%) as a white solid.

HRMS (m/z): [M]$^+$ calcd. for $C_{27}H_{38}N_3O_{11}$, 580.2501; found, 580.2501.

^{1}H NMR (400 MHz, CDCl$_3$) δ (ppm): 1.27 (d, 3H, *J* = 6.5 Hz, Me$_{Thr}$), 1.96 (s, 3H, NHCOC$\underline{H}_3$), 1.99 (s, 3H, COCH$_3$), 2.02 (s, 3H, COCH$_3$), 2.03 (s, 3H, NHCOC$\underline{H}_{3\ Thr}$), 2.16 (s, 3H, COCH$_3$), 4.02 – 4.11 (m, 2H, 2H$_{6S}$), 4.15 (m, 1H, Hβ), 4.20 (m, 1H, H$_{5S}$), 4.34 – 4.45 (m, 2H, C$\underline{H}_2$Ph), 4.57 (ddd, *J* = 11.1, 9.4, 3.6 Hz, 1H, H$_{2S}$), 4.66 (dd, 1H, *J* = 9.0, 2.4 Hz, Hα), 4.88 (d, 1H, *J* = 3.7 Hz, H$_{1S}$), 5.06 (dd, 1H, *J* = 11.3, 3.2 Hz, H$_{3S}$), 5.35 (d, 1H, *J* = 3.1 Hz, H$_{4S}$), 7.18 – 7.24 (m, 2H, Bn), 7.27 – 7.37 (m, 3H, Bn).

^{13}C NMR (100 MHz, CDCl$_3$) δ (ppm): 18.2 (Me$_{Thr}$), 20.6 ($\underline{C}$H$_3$CO), 20.7 ($\underline{C}$H$_3$CO), 20.8 ($\underline{C}$H$_3$CO), 23.0 (NHCO$\underline{C}$H$_3$ s), 23.2 (NHCO$\underline{C}$H$_3$ Thr), 43.7 ($\underline{C}$H$_2$Ph), 47.5 (C$_{2S}$), 56.6 (Cα), 62.2 (C$_{6S}$), 67.3 (C$_{4S}$), 67.4 (C$_{5S}$), 69.0 (C$_{3S}$), 78.2 (Cβ), 100.7 (C$_{1S}$), 127.6, 127.6, 129.9, 127.9, 137.2 (Ph), 170.1, 170.4, 170.5, 170.8, 171.0, 171.3 (CO).

Synthesis of Tn-Thr'

Tn-Thr'
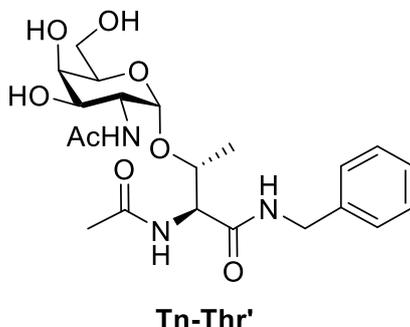
Compound **3-IV** (189 mg) was dissolved in MeONa/MeOH (6 mL, 1:2), adjusting pH value at 9.5. After stirring for 1 h at room temperature, Dowex resin is added until neutral pH is obtained. The reaction mixture is then filtered and concentrated, yielding final product **Tn-Thr'** as a pure white solid (86 mg, 58%).

HRMS (m/z): [M]$^+$ calcd. for $C_{21}H_{32}N_3O_8$, 454.2184; found, 454.2173.

^{1}H NMR (400 MHz, D$_2$O) δ (ppm): 1.15 (d, 3H, *J* = 6.3 Hz, Me$_{Thr}$), 1.81 (s, 3H, NHCOC$\underline{H}_3$), 2.01 (s, 3H, NHCOC$\underline{H}_{3\ Thr}$), 3.61 (m, 2H, 2H$_{6S}$), 3.67 (dd, 1H, *J* = 11.0, 2.6 Hz, H$_{3S}$), 3.80 – 3.86 (m, 2H, H$_{4S}$, H$_{5S}$), 3.93 (dd, 1H, *J* = 11.0, 3.6 Hz, H$_{2S}$), 4.15 – 4.23 (m, 2H, Hβ, C$\underline{H}_2$Ph), 4.41 – 4.35 (m, 2H, Hα, C$\underline{H}_2$Ph), 4.70 (d, 1H, *J* = 3.6 Hz, H$_{1S}$), 7.16 – 7.30 (m, 5H, Ph).

^{13}C NMR (100 MHz, D$_2$O) δ (ppm): 17.9 (Me$_{Thr}$), 21.7 (NHCO$\underline{C}$H$_3$), 22.1 (NHCO$\underline{C}$H$_3$), 43.0 ($\underline{C}$H$_2$Ph), 49.8 (C$_{2S}$), 58.0 (Cα), 61.2 (C$_{6S}$), 68.0 (C$_{3S}$), 68.5 (C$_{4S}$), 71.3 (C$_{5S}$), 75.5 (Cβ), 98.7 (C$_{1S}$), 127.1, 127.1, 127.5, 128.8, 128.8, 137.8 (Ph), 171.5, 174.2, 174.8 (CO).

Synthesis of compound 3-VI

3-VI
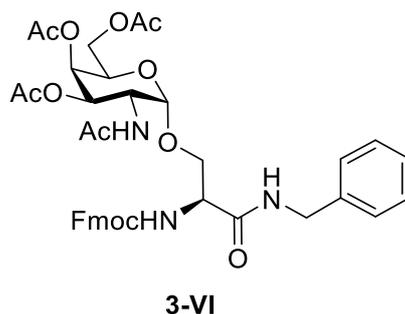
Following the same methodology as described for **3-II** but starting from derivative **3-V** (900 mg, 1.37 mmol), compound **3-VI** was obtained as white solid in a 55% yield (560 mg).

HRMS (m/z): [M]$^+$ calcd. for $C_{39}H_{44}N_3O_{12}$, 746.2920; found, 746.2917.

^{1}H NMR (400 MHz, CDCl$_3$) δ (ppm): 1.93 (s, 3H, NHCOC$\underline{H}_3$), 1.97 (s, 3H, COCH$_3$), 2.04 (s, 3H, COCH$_3$), 2.20 (s, 3H, COCH$_3$), 3.88 (m, 1H, Hβ), 3.96 – 4.05 (m, 4H, 2H$_{6S}$, Hβ, H$_{5S}$), 4.24 (t, 1H, J = 6.5 Hz, CHFmoc), 4.42 – 4.55 (m, 5H, CH$_2$ moc, C$\underline{H}_2$Ph, Hα), 4.56 – 4.63 (m, 1H, H$_{2S}$) 4.86 (d, 1H, d, J = 3.7 Hz, H$_{1S}$), 5.04 (dd, 1H, J = 11.4, 3.2 Hz, H$_{3S}$), 5.26 (d, 1H, J = 3.2 Hz, H$_{4S}$), 7.26 – 7.30 (m, 2H, Fmoc), 7.31 – 7.38 (m, 5H, Ph), 7.42 – 7.47 (m, 2H, Fmoc), 7.58 – 7.63 (m, 2H, Fmoc), 7.77 – 7.81 (m, 2H, Fmoc).

^{13}C NMR (100 MHz, CDCl$_3$) δ (ppm): 20.6, 20.7, 20.8 (C$\underline{H}_3$CO), 23.1 (NHCOC$\underline{H}_3$), 43.7 (CH$_2$ Ph), 47.0 (C$\underline{H}$Fmoc), 47.6 (C$_{2S}$), 54.7 (Cα), 62.0 (C$_{6S}$), 67.1 (C$\underline{H}_2$Fmoc), 67.2 (C$_{4S}$), 67.3 (C$_{5S}$), 68.3 (C$_{3S}$), 77.2 (Cβ), 99.5 (C$_{1S}$), 120.1 (Fmoc), 124.9 (Fmoc), 127.1, 127.1 (Ph), 127.7 (Fmoc), 127.9 (Fmoc), 128.9, 128.9 (Ph), 137.7 (Ph), 141.4 (Fmoc), 143.6 (Fmoc), 143.5 (Fmoc), 156.2 (NCOO), 169.1, 170.3, 170.4, 170.5, 170.9 (CO).
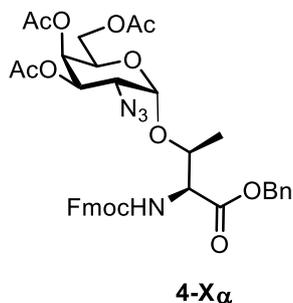
Synthesis of compound 3-VII
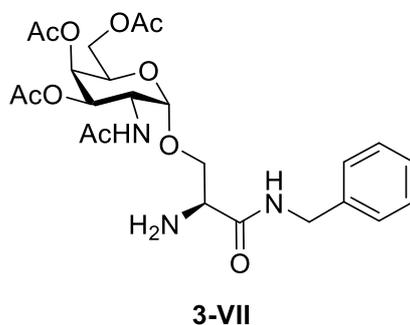
3-VII

Following the same methodology as described for **3-III** but starting from derivative **3-VI** (500 mg, 0.67 mmol), compound **3-VII** was obtained as white solid in a 93 % yield (325 mg).

HRMS (m/z): [M]$^+$ calcd. for $C_{24}H_{34}N_3O_{10}$, 524.2239; found, 524.2239.

^{1}H NMR (400 MHz, CDCl$_3$) δ (ppm): 1.95 (s, 3H, NHCOC$\underline{H}_3$), 2.05 (s, 3H, COCH$_3$), 2.08 (s, 3H, COCH$_3$), 2.21 (s, 3H, COCH$_3$), 3.68 (t, 1H, J = 4.3 Hz, Hα), 3.94 (dd, 1H, J = 9.8 Hz, 3.6 Hz, Hβ), 4.03 (dd, 1H, J = 9.8, 4.9 Hz, Hβ), 4.12 (dd, 2H, J = 6.5, 1.9 Hz, 2H$_{6S}$), 4.23 (td, 1H, J = 6.4, 1.1 Hz, H$_{5S}$), 4.52 (d, 2H, J = 6.0 Hz, C$\underline{H}_2$Ph), 4.63 (ddd, 1H, J = 11.3, 9.6, 3.6 Hz, H$_{2S}$), 4.94 (d, 1H, J = 3.6 Hz, H$_{1S}$), 5.12 (dd, 1H, J = 11.4, 3.2 Hz, H$_{3S}$), 5.39 (dd, 1H, J = 3.2, 1.2 Hz, H$_{4S}$), 7.31 − 7.35 (m, 3H, Ph), 7.37 − 7.42 (m, 2H, Ph).

^{13}C NMR (100 MHz, CDCl$_3$) δ (ppm): 20.6 ($\underline{C}H_3CO$), 20.7 ($\underline{C}H_3CO$), 20.8 ($\underline{C}H_3CO$), 23.1 (NHCO$\underline{C}H_3$), 43.3 ($\underline{C}H_2Ph$), 47.7 (C$_{2S}$), 54.9 (Cα), 61.9 (C$_{6S}$), 67.2 (C$_{4S}$), 67.3 (C$_{5S}$), 68.5 (C$_{3S}$), 72.3 (Cβ), 99.1 (C$_{1S}$), 127.6, 127.6, 127.6, 128.8, 128.8, 138.1 (Ph), 170.4, 170.5, 170.6, 170.9, 171.9 (CO).

Synthesis of compound 3-VIII
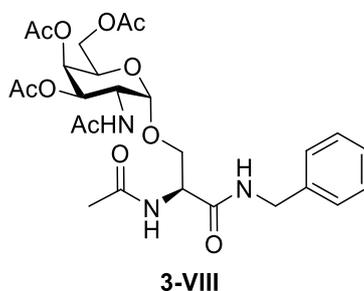
3-VIII

Following the same methodology as described for **3-IV** but starting from derivative **3-VII** (360 mg, 0.69 mmol), building block **3-VIII** was obtained as white solid in an 85 % yield (330 mg).

HRMS (m/z): [M]$^+$ calcd. for $C_{26}H_{36}N_3O_{11}$, 566.2344; found, 566.2336.

^{1}H NMR (400 MHz, CDCl$_3$) δ (ppm): 1.93 (s, 3H, NHCOC$\underline{H}_3$), 2.00 (s, 3H, NHCOC$\underline{H}_3$), 2.02 (s, 6H, COCH$_3$, COCH$_3$), 2.16 (s, 3H, COCH$_3$), 3.78 − 3.90 (m, 2H, 2Hβ), 3.99 − 4.09 (m, 3H, H$_{5S}$, 2H$_{6S}$), 4.43 (d, 2H J = 5.7 Hz, C$\underline{H}_2$Ph), 4.56 (ddd, 1H, J = 11.2, 9.5, 3.7 Hz, H$_{2S}$), 4.73 (m, 1H, Hα), 4.84 (d, 1H, J = 3.6 Hz, H$_{1S}$), 5.02 (dd, 1H, J = 11.3, 3.3 Hz, H$_{3S}$), 5.25 (d, 1H, J = 3.3 Hz, H$_{4S}$), 7.24 − 7.36 (m, 5H, Ph).

^{13}C NMR (100 MHz, CDCl$_3$) δ (ppm): 20.6 ($\underline{C}H_3CO$), 20.7 ($\underline{C}H_3CO$), 20.8 ($\underline{C}H_3CO$), 23.1 (NHCO$\underline{C}H_3$), 23.2 (NHCO$\underline{C}H_3$), 43.7 ($\underline{C}H_2Ph$), 47.5 (C$_{2S}$), 52.7 (Cα), 62.0 (C$_{6S}$), 67.2 (C$_{4S}$), 67.3 (C$_{5S}$), 68.5 (C$_{3S}$), 69.53 (Cβ), 99.3 (C$_{1S}$), 127.8, 127.8, 127.9, 128.9, 128.9, 137.5 (Ph), 169.3, 170.3, 170.5, 170.6, 170.7, 171.0 (CO).

Synthesis of Tn-Ser'
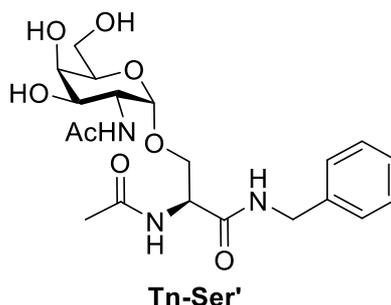
Tn-Ser'

Following the same methodology as described for **Tn-Thr'** but starting from derivative **3-VIII** (330 mg, 0.58 mmol), compound **Tn-Ser'** was obtained as white solid in a 62 % yield (159 mg).

HRMS (m/z): [M]$^+$ calcd. for $C_{20}H_{30}N_3O_8$, 440.2027; found, 440.2028.

^{1}H NMR (400 MHz, MeOD) δ (ppm): 1.98 (s, 3H, NHCOC$\underline{H}_3$), 2.05 (s, 3H, NCOC$\underline{H}_3$), 3.61 – 3.67 (m, 2H, H$_{3S}$, H$_{6S}$), 3.68 – 3.79 (m, 3H, H$_{6S}$, H$_{4S}$, H$_{5S}$), 3.79 – 3.94 (m, 2H, 2Hβ), 4.27 (dd, 1H, *J* = 10.9, 3.7 Hz, H$_{2S}$), 4.35 – 4.50 (m, 2H, C$\underline{H}_2$Ph), 4.63 (t, 1H, *J* = 5.1 Hz, Hα), 4.81 (d, 1H, *J* = 3.8 Hz, H$_{1S}$), 7.22 – 7.39 (m, 5H, Ph).

^{13}C NMR (100 MHz, MeOD) δ (ppm): 21.2 (NHCO$\underline{C}$H$_3$), 21.4 (NHCO$\underline{C}$H$_3$), 42.8 ($\underline{C}$H$_2$Ph), 50.0 (C$_{2S}$), 53.6 (Cα), 61.5 (C$_{6S}$), 68.4 (C$_{3S}$), 68.5 (Cβ), 68.9 (C$_{4S}$), 71.5 (C$_{5S}$), 98.8 (C$_{1S}$), 127.0, 127.3, 127.3, 128.2, 128.2, 138.4 (Ph), 170.6, 172.0, 172.5 (CO).

8.9.2. IRID spectra

Experimental system has been previously described.[18–21] Briefly, **Tn-Ser'** and **Tn-Thr'** antigens were mixed with carbon nanotubes (Multi-Walled Carbon Nanotubes, purity >90%, 10-30 mm diameter, Sun NanoTech Co Ltd) and were deposited in a cylindrical sample holder (4.5 mm diameter, 15 mm long) by pressing the cylinder against the desired mixture. They were then vaporized by a laser desorption system (1064 nm, Quantel Ultra) coupled to a pulsed valve into a supersonic jet of argon that passed through a 4 mm skimmer to create a collimated molecular beam. This was crossed by two tunable laser beams, UV ionization and IR depopulation lasers, in the extraction region of a linear time-of-flight mass spectrometer (Jordan Inc.).

One color mass-selected photoionization spectra, recorded using a frequency-doubled pulsed Nd:YAG-pumped dye laser operating at 10 Hz, were followed by conformer-specific spectroscopy in the IR using IR-UV ion dip (IRID) double resonance spectroscopy. The tunable IR radiation was provided by an OPO/OPA laser system (LaserVision). The delay between the pump and the probe laser pulses was ~150 ns in the IRID double resonance experiments.

8.9.3. Conformational search and QM calculations on *Tn-Thr'* and *Tn-Ser'* derivatives

The conformational search followed an iterative approach. The spectroscopic calculations began with completely unrestricted and exhaustive surveys of the conformational landscapes of each of the **Tn-Thr'** and **Tn-Ser'** antigens, and their hydrated complexes, using a fast molecular mechanics method (with several force fields MMFFs,[22] AMBER[23] and OPLS2005[24]) until no additional new structures were obtained. Their conformers were identified with the advanced hybrid[25] Large-Scale Low-Mode[26]/Monte Carlo algorithm[27] (LM/MC) implemented in MacroModel (v.9.2), Schrödinger, LLC21.[28] These surveys, generated 2000-5000 structures depending on the system (<20 kJ·mol^{-1}). The initial sets of structures were grouped into families.

The ~400 lowest-lying energy conformers (<15 kJ·mol^{-1}) and a representative member of each group that might have a significant population in the cooled adiabatic expansion (typically ~150 structures), were re-optimized through density functional theory calculations [M06-2X/6-31+G(d)] using the Gaussian 09 program package[29] to provide more accurate structures and energies. In addition, their associated harmonic vibrational spectra were predicted. The superposition error was also corrected by BSSE calculations for hydrated antigens.

The optimized geometries were ordered by relative Gibbs free energy at 298K (Equation 1). The energy at 0 K (Equation 2) is also reported:

$$\textit{Equation 1: } \Delta G_{298K} = \text{Elec.} + \text{ZPE} + \text{BSSE} + \Delta H_{298K} - T \cdot \Delta S_{298K}$$
$$\textit{Equation 2: } \Delta G_{0K} = \text{Elec.} + \text{ZPE} + \text{BSSE}$$

Conformational assignments were based primarily on the level of correspondence between the experimental and computed OH and NH vibrational spectra, scaled by the 'anharmonicity' factor, 0.95. The best agreement between experiment and theory corresponded with the calculated minimum free energy structures.

8.10.   Experimental section chapter 4

8.10.1. Synthesis

Synthesis of compound 4-III
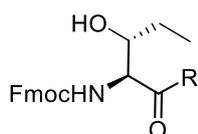
4-II, R = OH
4-III. R = O^{t}Bu

Commercially available compound **4-I** (200 mg, 1.50 mmol) was dissolved in H_2O (7 mL) and $NaHCO_3$ (252 mg, 3 mmol) was then added. The resulting mixture was stirred at room temperature until complete dissolution was achieved. The reaction mixture was diluted with acetonitrile (13 mL), followed by the addition of Fmoc–OSu (761 mg, 2.25 mmol). The white suspension was stirred vigorously at room temperature overnight. Acetonitrile was then removed under reduced pressure and the aqueous solution was extracted with Et_2O, followed by acidification and subsequent extraction with a mixture of $CHCl_3$/^iPrOH (3:1). The organic layer was concentrated, and the corresponding amino acid **4-II** (420 mg, 79%) was obtained without further purification.

Tert-Butyl 2,2,2-trichloroacetimidate (423 μL, 2.36 mmol) dissolved in cyclohexane (2.50 mL) was added drop-by-drop to a solution of derivative **4-II** (420 mg, 1.18 mmol) in AcOEt (6.30 mL) and stirred vigorously overnight. Reaction mixture was then washed twice with a saturated solution of $NaHCO_3$ (10 mL). Byproducts were recrystallized over CH_2Cl_2 and supernatant was concentrated and purified by column chromatography (hexane/AcOEt, 6:4) to give amino acid **4-III** fully protected as white foam (413 mg, 85%).

HRMS (ESI+) m/z: calcd. for $C_{24}H_{29}NNaO_5$ [M+Na]$^+$: 434.1938; found: 434.1920.

^1H NMR (300 MHz, $CDCl_3$) δ (ppm): 0.99 (t, 3H, J = 7.3 Hz, CH_3 Hnv), 1.48 − 1.52 (m, 11H, CH_2 Hnv, C(CH$_3$)$_3$), 4.00 (t, 1H, J = 5.9 Hz, Hβ), 4.21 − 4.26 (m, 1H, CHFmoc), 4.30 (d, 1H, J = 9.3 Hz, Hα), 4.41 (m, 2H, CH_2Fmoc), 7.29 − 7.34 (m, 2H, Fmoc), 7.38 − 7.43 (m, 2H, Fmoc), 7.59 − 7.63 (m, 2H, Fmoc), 7.75 − 7.78 (m, 2H, Fmoc).

^{13}C NMR (75 MHz, $CDCl_3$) δ (ppm): 10.1 (CH$_3$ Hnv), 26.9 (CH$_2$ Hnv), 28.0 (C($\underline{C}$H$_3$)), 47.2 (CH-Fmoc), 58.0 (Cα), 67.1 (CH$_2$-Fmoc), 73.8 (Cβ), 82.6 ($\underline{C}$(CH$_3$)), 119. 9, 120.0, 125.1, 125.1, 127.1, 127.1, 127.7, 127.7, 143.3, 143.3, 143.8, 143.9 (Fmoc), 156.7 (NCOO), 170.6 (COO).

Synthesis of compound 4-Vα

4-Vα
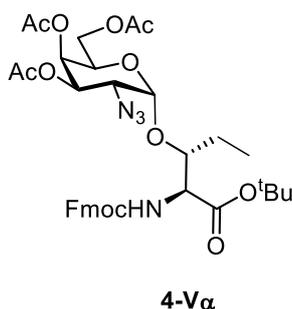
Compound **4-III** (900 mg, 2.26 mmol) was dissolved in a mixture of dry toluene/CH_2Cl_2 (18 mL, 1:1) under argon atmosphere with molecular sieves (4 Å). After stirring for 1 h, silver salts, Ag_2CO_3 (936 mg, 3.39 mmol) and $AgClO_4$ (25 mg, 0.12 mmol), were added at 0 °C. Reaction mixture is kept under vigorous stirring for 30 min and glycosidic donor **4-IV** (1.38 g, 3.94 mmol)[30] was finally added in a mixture of toluene/CH_2Cl_2 (11 mL, 4:7). Reaction then stirred overnight at room temperature in the absence of light. Once it had finished, reaction mixture was filtered through Celite, concentrated and purified by

chromatography column toluene/acetone (9:1) yielding a mixture of both α and β anomers (887 mg, 56%) as a white foam that was used in the following step without further purification.

HRMS (ESI+) m/z: calcd. for $C_{36}H_{45}N_4NaO_{12}$ [M+H]$^+$: 725.3028; found: 725.3020.

Synthesis of compound 4-VI

4-VI
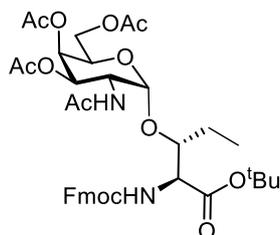
Compounds **4-Vα** and **4-Vβ** (230 mg, 0.317 mmol) were dissolved in a mixture of THF/AcOH/Ac$_2$O (8 mL, 3:2:1). Keeping the mixture at 0 °C, activated Zn (210 mg, 3.21 mmol) and 0.27 mL of a saturated solution of CuSO$_4$ were added and reaction was stirred for 2 h. After that time, it was filtered through Celite, then washed with a saturated solution of NaHCO$_3$ (2x5 mL), dried and evaporated. After column chromatography (hexane/AcOEt, 7:3), building block **4-VI** was finally obtained as a white foam (179 mg, 76%).

HRMS (ESI+) m/z: calcd. for $C_{38}H_{49}N_2O_{13}$ [M+H]$^+$: 741.3229; found: 741.3230.

^{1}H NMR (400 MHz, CDCl$_3$) δ (ppm): 0.98 (t, 3H, J = 7.3 Hz, CH$_3$ Hnv), 1.47 (s, 9H, C(CH$_3$)$_3$), 1.61 − 1.70 (m, 2H, CH$_2$ Hnv), 2.00 (s, 3H, NHCOCH$_3$), 2.01, 2.05, 2.17 (s, 9H, 3COCH$_3$), 3.86 − 3.94 (m, 1H, Hβ), 4.10 (m, 2H, CH$_2$Fmoc), 4.24 − 4.30 (m, 2H, H$_{5S}$, CHFmoc), 4.40 (d, 1H, J = 8.7 Hz, Hα), 4.44 − 4.51 (m, 2H, H$_{6S}$), 4.61 − 4.68 (m, 1H, H$_{2S}$), 4.92 (d, 1H, J = 3.4 Hz, H$_{1S}$), 5.10 (dd, 1H, J = 11.3, 2.9 Hz, H$_{3S}$), 5.38 − 5.41 (s, 1H, H$_{4S}$), 7.31 − 7.36 (m, 2H, Fmoc), 7.39 − 7.43 (m, 2H, Fmoc), 7.61 − 7.67 (m, 2H, Fmoc), 7.76 − 7.80 (m, 2H, Fmoc).

^{13}C NMR (101 MHz, CDCl$_3$) δ (ppm): 10.0 (CH$_3$ Hnv), 20.6, 20.7, 20.8 (CH$_3$CO), 23.3 (NHCOCH$_3$), 24.8 (CH$_2$ Hnv), 28.1 (C(CH$_3$)), 47.2 (CH-Fmoc), 47.3 (C$_{2S}$), 55.6 (Cα), 62.2 (CH$_2$-Fmoc), 67.3 (C$_{6S}$), 67.4 (C$_{4S}$), 67.5 (C$_{5S}$), 68.8 (C$_{3S}$), 82.5 (Cβ), 83.5 (C(CH$_3$)), 99.8 (C$_{1S}$), 120.0, 120.1, 125.0, 125.1, 127.0, 127.1, 127.7, 127.8 (Fmoc), 141.3, 141.4, 143.7, 143.9 (Fmoc), 156.3 (NCOO), 170.2, 170.3, 170.4, 170.5, 171.0 (COO).

Synthesis of compound Tn-Hnv
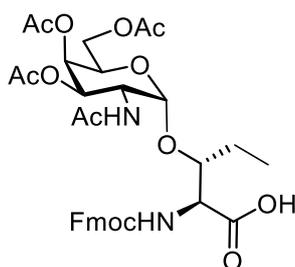
Tn-Hnv

A solution of derivative **4-VI** (179 mg, 0.242 mmol) in a mixture of CH_2Cl_2/TFA (10 mL, 1:1) was stirred for 2 h. It was then concentrated until TFA was removed. Final compound **Tn-Hnv** was obtained as a white foam (152 mg, 92%), ready to be coupled into a peptide.

HRMS (ESI) (m/z): cald. for $C_{34}H_{41}N_2O_{13}^+$[M+H]$^+$: 685.2603; found: 685.2599.

Synthesis of compound 4-IX
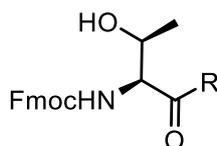
4-IX

4-VIII, R = OH
4-IX. R = OBn

Following the same methodology as described for **4-II** but starting from derivative **4-VII** (1.00 g, 8.39 mmol), compound **4-VIII** was obtained as white solid in a 91 % yield (2.60 g). Compound **4-VIII** (2.60 g, 7.62 mmol) was subsequently dissolved in DMF (15 mL) under argon atmosphere. Cs_2CO_3 (3.72 g, 11.42 mmol) was added at room temperature and stirred for 1 h, followed by the addition of benzyl bromide (1.09 mL, 9.14 mmol) and stirred at room temperature overnight. The reaction mixture was poured onto a saturated solution of LiBr (100 mL), extracted with ethyl acetate (50 mL), and washed with water (50 mL) and brine (25 mL). The organic layer was dried (Na_2SO_4), filtered and the filtrate was concentrated in vacuum. The residue was purified by silica gel column chromatography (hexane/ethyl acetate, 1:1) to give compound **4-IX** (3.12 g, 86%) as a white solid.

HRMS (ESI+) m/z: calcd. for $C_{26}H_{25}NO_5$ [M+H]$^+$: 432.1805 found: 432.1810.

^{1}H NMR (CDCl$_3$) δ (ppm): 1.17 (d, 3H, *J* = 6.1 Hz, CH$_3$ *allo*Thr), 2.79 (s, 1H, OH), 4.10 – 4.26 (m, 2H, Hβ alloThr, CH-Fmoc), 3.36 – 4.46 (m, 2H, CH$_2$-Fmoc), 4.49 (d, 1H, *J* = 3.8 Hz, Hα *allo*Thr), 5.08 – 5.31 (m, 2H, CH$_2$-Bn), 7.19 – 7.47 (m, 9H, Arom.), 7.51 – 7.67 (m, 2H, Arom.), 7.70 – 7.81 (m, 2H, Arom.).

^{13}**C NMR (CDCl₃) δ (ppm):** 18.9 (CH₃ *allo*Thr), 47.2 (CH-Fmoc), 59.5 (Cα *allo*Thr), 67.3 (CH₂-Fmoc), 67.6(CH₂-Bn), 69.0(Cβ *allo*Thr), 120.0, 120.1, 125.1, 127.1, 127.8, 128.2, 128.3, 128.4, 128.5, 128.6, 128.7, 135.0, 141.4, 143.7, 143.7 (Arom.), 156.65, 170.17 (COO).

Synthesis of compound 4-Xα

4-Xα

Following the same methodology as described for **4-III** but starting from derivative **4-IX** (869 mg, 2.01 mmol), compound **4-Xα** was obtained as white foam (928 mg, 62%).

HRMS (ESI+) m/z: calcd. for $C_{38}H_{40}N_4O_{12}^+$ [M+H]$^+$: 745.2715 found: 745.2746.

^{1}**H NMR (CDCl₃) δ (ppm):** 1.40 (d, 3H, *J* = 6.5 Hz, CH₃ *allo*Thr), 1.95 (s, 3H, COCH₃), 2.09 (s, 3H, COCH₃), 2.18 (s, 3H, COCH₃), 3.62 (dd, 1H, *J* = 11.2, 3.3 Hz, H₂ₛ), 4.00 – 4.10 (m, 2H, 2H₆ₛ), 4.10 – 4.21 (m, 1H, Hβ *allo*Thr), 4.25 (dd, 1H, *J* = 15.4, 8.5 Hz, CH-Fmoc), 4.32 – 4.47 (m, 2H, CH₂-Fmoc, H₅ₛ), 4.49 – 4.66 (m, 2H, CH₂ Fmoc, Hα *allo*Thr), 4.99 (d, 1H, *J* = 3.3 Hz, H₁ₛ), 5.21 – 5.37 (m, 3H, CH₂-Bn, H₃ₛ), 5.40 – 5.48 (m, 1H, H₄ₛ), 7.18 – 7.50 (m, 9H, Arom.), 7.63 – 7.72 (m, 2H, Arom.), 7.73 – 7.85 (m, 2H, Arom.).

^{13}**C NMR (CDCl₃) δ (ppm):** 16.9 (CH₃ *allo*Thr), 20.5, 20.6, 20.7 (3 CO<u>C</u>H₃), 47.2 (CH Fmoc), 57.4 (C₂ₛ), 58.6(Cα *allo*Thr), 61.9(C₆ₛ), 67.1 (CH₂ Fmoc), 67.4(C₅ₛ,), 67.7, 67.7, 67.8 (CH₂-Bn, C₄ₛ, C₃ₛ), 78.2(Cβ *allo*Thr), 98.5 (C₁ₛ), 120.0, 125.1, 125.2, 127.1, 127.7, 128.3, 128.7, 128.7, 128.7, 129.1, 134.9, 141.4, 143.8, 143.9 (Arom.), 155.9, 168.9, 169.6, 170.0, 170.4 (COO).

Synthesis of compound 4-XI

4-XI
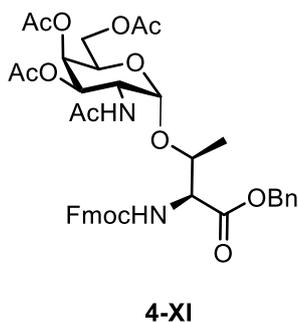
Compound **4-Xα** (620 mg, 0.832 mmol) was dissolved in a mixture of THF/AcOH/Ac$_2$O (12 mL, 3:2:1). Keeping the mixture at 0 °C, activated Zn (707 mg, 10.82 mmol) and 1.25 mL of a saturated solution of CuSO$_4$ were added and reaction is stirred for 2 h. After that time, it was filtered through Celite, then washed with a saturated solution of NaHCO$_3$ (2x15 mL), dried and evaporated. After column chromatography (hexane/AcOEt, 7:3), building block **4-XI** was finally obtained as a white foam (580 mg, 91%).

HRMS (ESI+) m/z: calcd. for C$_{40}$H$_{44}$N$_2$O$_{13}$ [M+H]$^+$: 761.2916 found: 761.2914.

^{1}H NMR (CDCl$_3$) δ (ppm): 1.27 (d, 3H, *J* = 6.0 Hz , CH$_3$ *allo*Thr), 1.92 (s, 3H, COCH$_3$), 1.97 (s, 3H, COCH$_3$), 2.05 (s, 3H, COCH$_3$), 2.20 (s, 3H, COCH$_3$), 3.97 – 4.12 (m, 3H, 2CH$_2$ $_{6S}$, Hβ *allo*Thr), 4.18 – 4.30 (m, 2H, CH-Fmoc, H$_{5S}$), 4.41 – 4.61 (m, 4H, CH2-Fmoc, Hα *allo*Thr, H$_{2S}$), 4.92 (s, 1H, H$_{1S}$), 4.95 – 5.05 (m, 1H, H$_{3S}$), 5.14 – 5.36 (m, 3H, CH$_2$-Bn, H$_{4S}$) , 5.45 (d, 1H, *J* = 8.7 Hz, NHAc), 7.27 – 7.49 (m, 9H, Arom.), 7.55 – 7.68 (m, 2H, Arom.), 7.75 – 7.84 (m, 2H, Arom.).

^{13}C NMR (CDCl$_3$) δ (ppm): 17.1 (CH$_3$ *allo*Thr), 20.6, 20.7, 20.8, 23.2 (4 CO<u>C</u>H$_3$), 47.2 (CH-Fmoc), 47.9 (C$_{2S}$), 58.7 (Cα *allo*Thr), 62.1 (C$_{6S}$), 67.4 (CH$_2$-Fmoc), 67.6, 68.1 (CH$_2$-Bn, C$_{3S}$), 97.4 (C$_{1S}$), 120.0, 125.0, 127.1, 127.7, 128.8, 128.9, 135.0, 141.4, 143.7 (Arom.), 169.9, 170.5, 170.9 (COO).

Synthesis of compound Tn-*allo*Thr
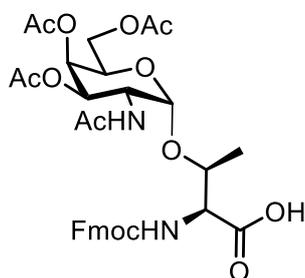
Tn-*allo*Thr

A solution of derivative **4-XI** (580 mg, 0.762 mmol) in 20 mL of MeOH, together with 3-4 drops of HCl 37.5%, was treated with 10% Pd/C (116 mg) and shaken under H$_2$ atmosphere for 1 h at 25°C. Catalyst was then removed through Celite filtration and the solution was concentrated prior to column chromatography in CH$_2$Cl$_2$/MeOH (9:1, 0.1% AcOH), yielding compound **Tn-*allo*Thr** as a white foam (460 mg, 89%) ready to be coupled in a peptide backbone.

HRMS (ESI+) m/z: calcd. for C$_{33}$H$_{38}$N$_2$O$_{13}$$^+$ [M+H]$^+$: 671.2447 found: 671.2443.

8.10.2. (Glyco)peptide synthesis

Synthesis of peptide *allo*Thr

Peptide ***allo*Thr** was synthesized as previously described.[31]
Semipreparative HPLC gradient R_t = 25.1 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	5	95
35	10	23	77

HRMS (ESI+) m/z: calcd. for $C_{83}H_{133}N_{27}O_{28}$: $[M+2H]^{2+}$: 978.9906 found: 978.9912. Calcd. for $[M+3H]^{3+}$: 652.9938 found: 652.9946.

Synthesis of glycopeptide *allo*Thr*

Natural amino acids were coupled following SPPS-MW methodology, whereas glycosylated moiety was coupled manually (100.5 mg, 0.15 mmol). After deprotection of acetyl groups and cleavage, glycopeptide ***allo*Thr*** was purified by reverse-phase HPLC and lyophilized.

Semipreparative HPLC gradient R_t = 16.7 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	10	90
30	10	18	82

HRMS (ESI+) m/z: calcd. for $C_{91}H_{146}N_{28}O_{33}$: $[M+3H]^{3+}$: 720.6869 found: 720.6875.

Synthesis of peptide Hnv

Peptide **Hnv** was synthesized as previously described.[31]

Semipreparative HPLC gradient R_t = 27.2 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	5	95
35	10	23	77

HRMS (ESI+) m/z: calcd. for $C_{84}H_{138}N_{27}O_{28}$: $[M+3H]^{3+}$: 658.0011 found: 658.0016.

Synthesis of glycopeptide Hnv*

Natural amino acids were coupled following MW-SPPS methodology using 0.05 mmol of resin, whereas glycosylated moiety was added manually through a double coupling strategy (41.1 mg, 0.06 mmol, then 33 mg, 0.05 mmol). After deprotection of acetyl groups and cleavage, glycopeptide **Hnv*** was purified by reverse-phase HPLC and lyophilized.

Semipreparative HPLC gradient R_t = 19.2 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	10	90
20	10	14	86

HRMS (ESI+) m/z: calcd. for $C_{92}H_{151}N_{28}O_{33}$: $[M+3H]^{3+}$: 725.3658 found: 725.3666.

Synthesis of glycopeptide *allo*Thr*'

Natural amino acids were coupled following SPPS-MW methodology, whereas glycosylated moiety was coupled manually (100.5 mg, 0.15 mmol). After deprotection of acetyl groups and cleavage, glycopeptide ***allo*Thr*'** was purified by reverse-phase HPLC and lyophilized.

Semipreparative HPLC gradient R_t = 11.0 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	10	90
14	10	12	88

HRMS (m/z): $[M]^+$ calcd. for $C_{35}H_{60}N_{11}O_{14}$, 858.4316; found 858.4289.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 1.04 (d, 3H, J = 6.1 Hz, Me$_{alloThr}$), 1.23 (d, 3H, J = 7.1 Hz, Me$_{Ala}$), 1.51 − 1.57 (m, 2H, 2Hγ$_{Arg}$), 1.72 − 1.81 (m, 2H, 2Hβ$_{Arg}$), 1.81− 2.02 (m, 12H, NHCOC**H**$_{3\ alloThr}$, NHCOC**H**$_{3\ Ala}$, 2Hγ$_{Pro7}$, 2Hγ$_{Pro11}$, Hβ$_{Pro7}$, Hβ$_{Pro11}$), 2.14 − 2.23 (m, 2H, Hβ$_{Pro7}$, Hβ$_{Pro11}$), 2.71 − 2.84 (m, 2H, 2Hβ$_{Asp}$), 3.08 − 3.12 (m, 2H, 2Hδ$_{Arg}$), 3.49 − 3.57 (m, 5H, Hδ$_{Pro7}$, Hδ$_{Pro11}$, H$_{5S}$, 2H$_{6S}$), 3.62 (m, 1H, J = 7.1, 4.0 Hz, H$_{3S}$), 3.67 − 3.74 (m, 3H, Hδ$_{Pro7}$, Hδ$_{Pro11}$, H$_{4S}$), 3.90 − 3.97 (m, 1H, Hβ$_{Thr}$), 4.02 (dd, 1H, J = 11.1, 3.8 Hz, H$_{2S}$), 4.26 − 4.33 (m, 2H, Hα$_{Pro11}$, Hα$_{Pro7}$), 4.35 (d, 1H, J = 8.4 Hz, Hα$_{alloThr}$), 4.41 − 4.47 (m, 1H, Hα$_{Ala}$), 4.55 (t, 1H, J = 6.7 Hz, Hα$_{Asp}$), 4.63 − 4.64 (m, 1H, Hα$_{Arg}$), 4.91 (d, 1H, J = 3.9 Hz, H$_{1S}$).

^{1}H NMR (400 MHz, D_2O/H_2O, 1:9, amide region) δ (ppm): 7.02 (s, 1H, NH$_{2\ Pro11}$), 7.16 − 7.19 (m, 1H, NH $_{Arg}$), 7.74 (s, 1H, NH$_{2\ Pro11}$), 8.04 (d, 1H, J = 9.1 Hz, N**H**COCH$_3$), 8.24 − 8.28 (m, 2H, NH $_{Ala}$, NH $_{Thr}$), 8.52 − 8.57 (m, 2H, NH $_{Asp}$, NH $_{Arg}$).

^{13}C NMR (101 MHz, D_2O) δ (ppm): 14.8 (Me$_{alloThr}$), 15.3 (Me$_{Ala}$), 21.4 (NHCOC**H**$_{3\ Ala}$), 21.9 (NHCOC**H**$_{3\ alloThr}$), 24.0 (Cγ$_{Arg}$), 24.6 (Cγ$_{Pro11}$), 24.7 (Cγ$_{Pro7}$), 28.0 (Cβ$_{Arg}$), 29.3 (Cβ$_{Pro11}$), 29.6 (Cβ$_{Pro7}$), 35.2 (Cβ$_{Asp}$), 40.5 (Cδ$_{Arg}$), 47.7, 47.8, 47.9 (Cδ$_{Pro7}$, Cδ$_{Pro11}$, Cα$_{Ala}$), 49.8 (C$_{2S}$), 50.1 (Cα$_{Asp}$), 50.8 (Cα$_{Arg}$), 57.9 (Cα$_{alloThr}$), 60.2 (Cα$_{Pro11}$), 60.4 (Cα$_{Pro7}$), 60.5 (C$_{6S}$), 67.5 (C$_{3S}$), 68.3 (C$_{4S}$),71.1 (C$_{5S}$), 71.2 (Cβ$_{alloThr}$), 94.0 (C$_{1S}$), 156.8 (Cε$_{Arg}$), 170.8, 171.1, 172.0, 173.4, 173.7, 173.8, 174.0, 174.6, 176.7 (COO).

Synthesis of glycopeptide Hnv*'

Natural amino acids were coupled following MW-SPPS methodology, whereas glycosylated moiety was added manually through a double coupling strategy (82.2 mg, 0.12 mmol, then 68.5 mg, 0.1 mmol). After deprotection of acetyl groups and cleavage, glycopeptide **Hnv*'** was purified by reverse-phase HPLC and lyophilized.

Semipreparative HPLC gradient R$_t$ = 12.7 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	8	92
14	10	14	86

HRMS (m/z): $[M]^+$ calcd. for $C_{36}H_{62}N_{11}O_{14}$, 872.4472; found 872.4446.

^{1}H NMR (400 MHz, D_2O) δ (ppm): 0.87 (t, 3H, J = 7.4 Hz, Me$_{Hnv}$), 1.47 − 1.64 (m, 5H, Me$_{Ala}$), 1.69 − 2.07 (m, 12H, NHCOC$\underline{H}_3$, Hβ$_{Pro7}$, Hβ$_{Pro11}$, 2Hγ$_{Pro7}$, 2Hγ$_{Pro11}$, 2Hβ$_{Arg}$, 2Hγ$_{Arg}$), 2.28 − 2.39 (m, 2H, Hβ$_{Pro7}$, Hβ$_{Pro11}$), 2.84 − 3.06 (m, 2H, 2Hβ$_{Asp}$), 3.21 − 3.26 (m, 2H, 2Hδ$_{Arg}$), 3.62 − 3.68 (m, 2H, Hδ$_{Pro7}$, Hδ$_{Pro11}$), 3.71 − 3.78 (m, 4H, Hδ$_{Pro7}$, Hδ$_{Pro11}$, 2H$_{6S}$), 3.86 (dd, 1H, J = 11.0, 3.0 Hz, H$_{3S}$), 3.98 (d, 1H, J = 2.6 Hz, H$_{4S}$), 4.05 (t, 1H, J = 6.1 Hz, H$_{5S}$), 4.13 (dd, 1H, J = 11.0, 3.8 Hz, H$_{2S}$), 4.16 − 4.22 (m, 1H, Hβ$_{Hnv}$), 4.34 − 4.41 (m, 2H, Hα$_{Pro11}$, Hα$_{Ala}$), 4.48 − 4.52 (m, 1H, Hα$_{Pro7}$), 4.61 − 4.68 (m, 2H, Hα$_{Arg}$, Hα$_{Hnv}$), 4.85 − 4.92 (m, 2H, H$_{1S}$, Hα$_{Asp}$),

^{1}H NMR (400 MHz, D_2O/H_2O, 1:9, amide region) δ (ppm): 6.99 (s, 1H, NH$_{2\ Pro11}$), 7.21 − 7.23 (m, 1H, NH $_{Arg}$), 7.73 − 7.76 (m, 2H, N$\underline{H}$COCH$_3$, NH$_{2\ Pro11}$), 8.40 (d, 1H, J = 7.2 Hz, NH $_{Arg}$), 8.46 (d, 1H, J = 8.9 Hz, NH $_{Hnv}$), 8.72 (d, 1H, J = 7.2 Hz, NH $_{Asp}$).

^{13}C NMR (75 MHz, D_2O) δ (ppm): 9.0 (Me$_{Hnv}$), 15.04 (Me$_{Ala}$), 22.30 (NHCOC$\underline{H}_3$), 25.3 (CH$_{2\ Hnv}$), 24.2 (Cγ$_{Arg}$), 24.6 (Cγ$_{Pro7}$), 24.7 (Cγ$_{Pro11}$), 27.4 (Cβ$_{Arg}$), 29.4 (Cβ$_{Pro6}$), 29.5 (Cβ$_{Pro2}$), 35.1 (Cβ$_{Asp}$), 40.5 (Cδ$_{Arg}$), 47.6 (Cδ$_{Pro7}$), 47.7 (Cδ$_{Pro11}$), 48.0 (Cα$_{Ala}$), 49.5 (Cα$_{Asp}$), 49.7 (C$_{2S}$), 51.2 (Cα$_{Arg}$), 53.9 (Cα $_{Hnv}$), 60.1 (Cα$_{Pro11}$), 60.2 (Cα$_{Pro7}$), 61.3 (C$_{6S}$), 68.2 (C$_{3S}$), 68.5 (C$_{4S}$), 71.5 (C$_{5S}$),79.7 (Cβ $_{Hnv}$), 98.3 (C$_{1S}$),156.7 (Cε$_{Arg}$), 169.2, 170.0, 171.5, 172.7, 173.3, 173.8, 174.1, 176.8 (CO).

Synthesis of peptide Cys-Thr

Amino acids were coupled following MW-SPPS methodology. After cleavage, peptide **Cys-Thr** was purified by reverse-phase HPLC and lyophilized.

Semipreparative HPLC gradient R_t = 15.4 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	10	90
17	10	17	83

HRMS (ESI+) m/z: calcd. for $C_{83}H_{135}N_{27}O_{28}S$: $[M+2H]^{2+}$: 994.9840 found: 994.9840.

Synthesis of glycopeptide Cys-Thr*

Natural amino acids were coupled following MW-SPPS methodology, whereas glycosylated moiety was coupled manually (100.5 mg, 0.15 mmol). After deprotection of acetyl groups and cleavage, glycopeptide **Cys-Thr*** was purified by reverse-phase HPLC and lyophilized.
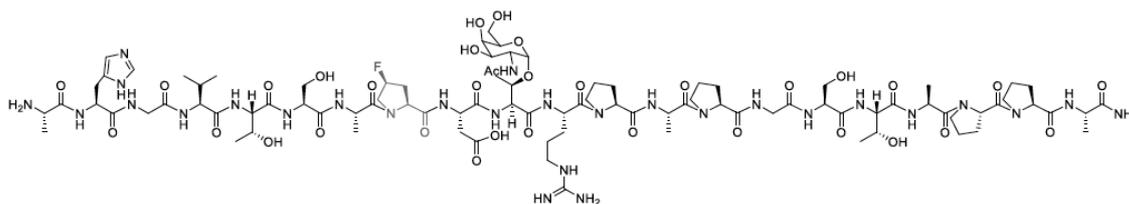
Semipreparative HPLC gradient R_t = 15.6 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	10	90
20	10	17	83

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

Synthesis of glycopeptide Cys-*sp²*-Thr*

Natural amino acids were coupled following MW-SPPS methodology, whereas glycosylated moiety was coupled manually (95 mg, 0.15 mmol). After deprotection of acetyl groups and cleavage, glycopeptide **Cys-*sp²*-Thr*** was purified by reverse-phase HPLC and lyophilized.
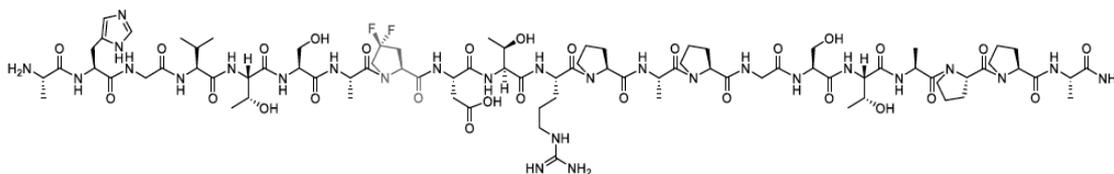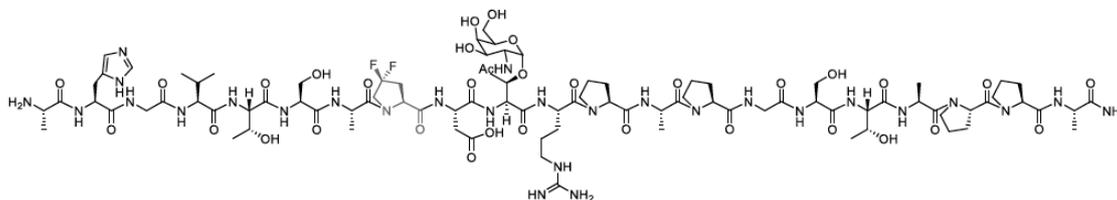
Semipreparative HPLC gradient R_t = 8.5 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	10	90
17	10	17	83

HRMS (ESI+) m/z: calcd. for $C_{92}H_{147}N_{29}O_{33}S$: $[M+2H]^{2+}$: 1109.0213 found: 1109.0232.

8.10.3. Crystal structure

X-ray structure of **Hnv*'** was obtained as previously described in *Experimental section 8.8*. Electron density map of glycopeptide **Hnv*'** in complex with scFV-SM3 antibody was F_O–F_C syntheses (blue) contoured at 2.0 σ (Figure 8.2).

*Figure 8.2 Electron density map of glycopeptide **Hnv*'** in complex with scFV-SM3 antibody.*

Hydrogen bonds found in the X-ray structures between **Hnv*'** epitope and scFv-SM3 anti-MUC1 antibody were summarized in

Table 8.1.

*Table 8.1 Hydrogen bonds found in the X-ray structures between **Hnv*'** epitope and scFv-SM3 fragment.*

Hnv*'	scFv-SM3	Hydrogen bond	distances (Å)
Ala6	Tyr[32H]	N-Oη	3.5
	Arg[52H]	O-water-NH2	2.8, 2.8
Asp8	Gln[97H]	O-Nε	3.0
	Trp[33H]	Oδ2-N	2.9
	Gly[96H]	O-water-N	3.3, 2.8
	Gly[96H]	O-water-N	3.5, 2.8
	Asn[31H]	Oδ1-water-O	2.7, 2.6
	Asn[31H]	Oδ2-water-O	3.3, 2.6
	Val[95H]	Oδ2-water-N	2.7, 2.8
	Trp[33H]	Oδ2-water-O	2.7, 2.7
Arg10	Asn[31H]	Nε-O	3.6
	Asn[31H]	NH1-Nδ	3.1
	Asn[31H]	NH2-Nδ	3.9
	Gly[96H]	O-water-N	2.8, 2.8
Pro11	Tyr[32H]	O-Oη	2.8
	Tyr[32H]	N-Oη	3.7

8.10.4. STD experiments

STD experiments were carried out by Dr. Filipa Macerlo in the Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (Lisbon, Portugal).

STD experiments were recorded on a Bruker Avance 600 MHz spectrometer as previously reported.[32] The samples were prepared in 20 mM PBS (pH = 7.1) in D_2O and the experiments were performed in a 40:1 glycopeptide:mAb molar ratio at 310 K with 8 μM of mAb. Commercial VU3C6 mAb was purchased from Abcam. A series of Gaussian-shaped pulses of 50 ms each one was employed with a total saturation time for the protein envelope of 2 s and a maximum B1 field strength of 60 Hz. An excitation sculpting module with gradients was employed to suppress the water proton signals. Selective saturation of the protein resonances (on-resonance spectrum) was performed by irradiating at δ = −0.5 ppm and for the reference spectrum (off-resonance) the

samples were irradiated at δ = 100 ppm. A total number of 4864 scans were acquired. The STD intensities were then normalized to that with the highest response. The signal of the anomeric proton of as well as, the Hα protons of the amino acids could not be analyzed due to their close distance to the H_2O signal.

STD spectrum (blue) and the off-resonance spectrum (red) for the STD-NMR experiment of the **Hnv*'** glycopeptide in presence of mAb VU-3C6 is shown in Figure 8.3. The interaction of the natural Thr*' derivative was previously reported.[32]

*Figure 8.3 STD spectrum (blue) and the off-resonance spectrum (red) for the STD-NMR experiment of the **Hnv*'** derivative in presence of mAb VU-3C6.*

8.10.5. BLI experiments

(Glyco)peptides at different concentrations (0 to 10 µM), were immobilized on amine-reactive biosensors following the protocol described in the *Experimental section 8.7* Figure 8.4 shows the typical BLI curves and the corresponding fit.
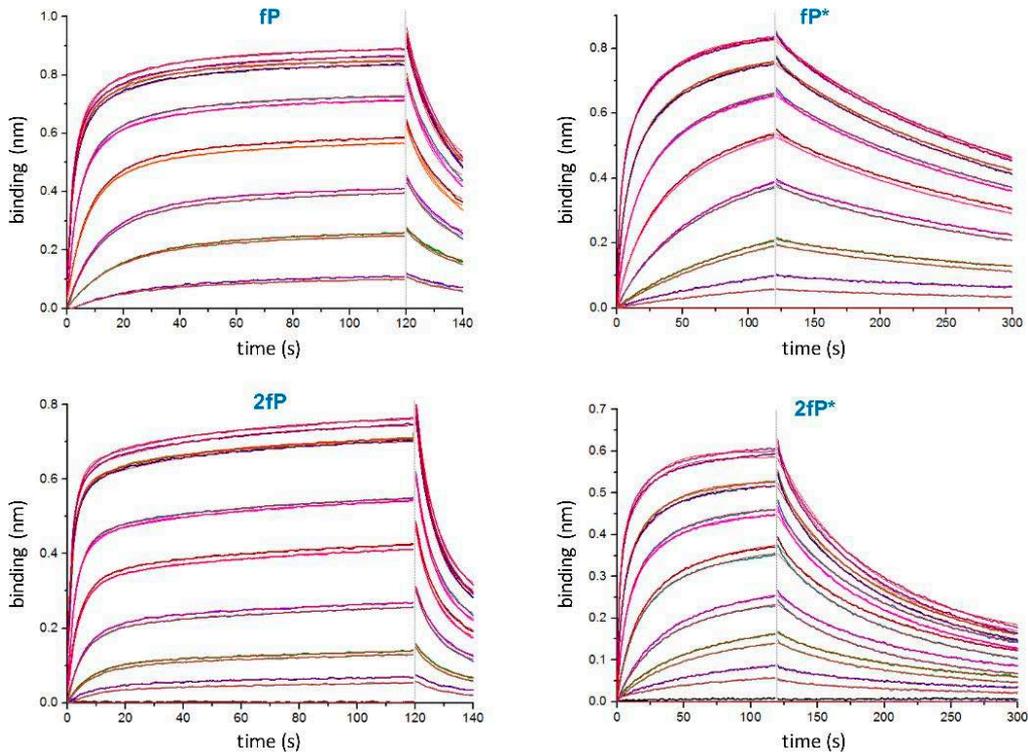
Figure 8.4 BLI curves and fit obtained for the (glyco)peptides.

Steady state analyses were then performed using each maximal binding response *versus* each range of concentrations to obtain the dissociation curves for the different (glyco)peptides. Dissociation constants (K_D) were consequently deduced from each chart. (Figure 8.5).

Figure 8.5 Steady state analyses and K_D constants obtained for the studied (glyco)peptides.

8.10.6. Surface Plasmon Resonance (SPR)

SPR experiments were performed with a Biacore X-100 apparatus (Biacore GE) using 25 mM PBS buffer with 0.005% tween as running buffer at a range of temperatures of 7-35

°C. Flow cells (CM5 sensor chip; Biacore) were activated for 7 min by injecting 140 μL of 50 mM *N*-hydroxysuccinimide (NHS): 200 mM ethyl-3(3-dimethylamino)propylcarbodiimide (EDC). mAb SM3 (ABCAM) was immobilized on the activated gold chip in flow cell 2 by injection of a 100 μg/mL protein solution diluted with 10 mM sodium acetate buffer with a flow rate of 10 μL/min for 7 min followed by an injection of 130 μL ethanolamine to block any remaining activated groups on the surface. The level of immobilization reached was about 3000 RUs. Flow cell 1, used as reference, was blocked with ethanolamine at the same conditions of flow cell 2 without immobilization of protein. Affinity experiments were made using a series of different concentrations of each epitope the range of 0.025–5 mM with a flow rate of 30 μL/min during 60 s. Each injection was followed by a 100 s injection of running buffer (dissociation phase). No regeneration steps were performed between injections. Response data were collected at real time and analyzed with the Biacore® X-100. Evaluation software and plotted as response shift versus analyte concentration.

SPR curves obtained and respective steady states analyses for each derivative, **Thr'**, **Thr*'** and **Hnv*'**, at different temperatures are gathered in *Supplementary information 9.3.*

For a better understanding, K_D values of analysed (glyco)peptides at each temperature are collated in Table 8.2; together with the reverse temperature and the product of R and ln K_D , which are used to obtain thermodynamic values by using Van't Hoff equation:[33]

$$\mathrm{R} \ln K_D = \Delta H \cdot \frac{1}{T} - \Delta S$$

Equation 8.1

Table 8.2 K_D values of analysed (glyco)peptides determined by SPR at different temperatures.

T (ºC)	Thr'		Thr*'		Hnv*'	
	K_D (M)	R ln (K_D)	K_D (M)	R ln (K_D)	K_D (M)	R ln (K_D)
37	-	-	0.001377	-54.77447823	0.001989	-51.776306
25	0.003480	-47.065923	0.0006189	-61.42371772	0.0003036	-67.422731
15	0.001180	-56.058173	0.0002721	-68.25633515	0.0001264	-74.716716
10	-	-	-	-	0.00009427	-77.065355
7	0.0005872	-61.860878	0.0001967	-70.95429751	-	-

By plotting the product of R and ln K_D against reverse temperature, ΔH and ΔS may be calculated, respectively, as the slope and Y-axis intercept of the resulting straight line. Both thermodynamic values for **Thr'**, **Thr*'** and **Hnv*'** are respectively shown in Figure 8.6, Figure 8.7 and Figure 8.8.

Thr'

$\Delta H = -(16.47 \pm 1.28)$ kcal/mol

$\Delta S = -(0.044 \pm 0.004)$ kcal/mol·K

$\Delta H = -(6.88982 \pm 0.5352)\cdot10^4$ J/mol

$-\Delta S = 183.7 \pm 18.5$ J/mol·K

Equation	y = a + b*x
Plot	R*ln(KD)
Weight	No Weighting
Intercept	183.71632 ± 18.5486
Slope	-6.88982 ± 0.5352
Residual Sum of Square	0.66653
Pearson's r	-0.997
R-Square(COD)	0.994
Adj. R-Square	0.988

*Figure 8.6 Thermodynamic values obtained for **Thr'**.*

Thr*'

$\Delta H = -(11.56 \pm 1.01)$ kcal/mol

$\Delta S = -(0.024 \pm 0.003)$ kcal/mol·K

$\Delta H = -(4.83519 \pm 0.42273)\cdot10^4$ J/mol

$-\Delta S = 100.8 \pm 14.4$ J/mol·K

Equation	y = a + b*x
Plot	R*ln(KD)
Weight	No Weighting
Intercept	100.76784 ± 14.40
Slope	-4.83519 ± 0.42273
Residual Sum of Squar	2.38086
Pearson's r	-0.99244
R-Square(COD)	0.98494
Adj. R-Square	0.97741

*Figure 8.7 Thermodynamic values obtained for **Thr*'**.*

Hnv*'

ΔH = -(19.68 $\pm$ 2.96) kcal/mol

ΔS = -(0.051 $\pm$ 0.010) kcal/mol·K

ΔH = -(8.23575 $\pm$ 1.23691)·10^4 J/mol

$-\Delta S$ = 211.9 $\pm$ 42.0 J/mol·K

*Figure 8.8 Thermodynamic values obtained for **Hnv*'**.*

Finally, thermodynamic parameters for each antigen at physiological conditions (T = 310.15 K) are shown in Table 8.3. ΔG has been deduced as the sum of enthalpy and entropy parameters using Equation 8.2:

$$\Delta G = \Delta H - T\Delta S$$
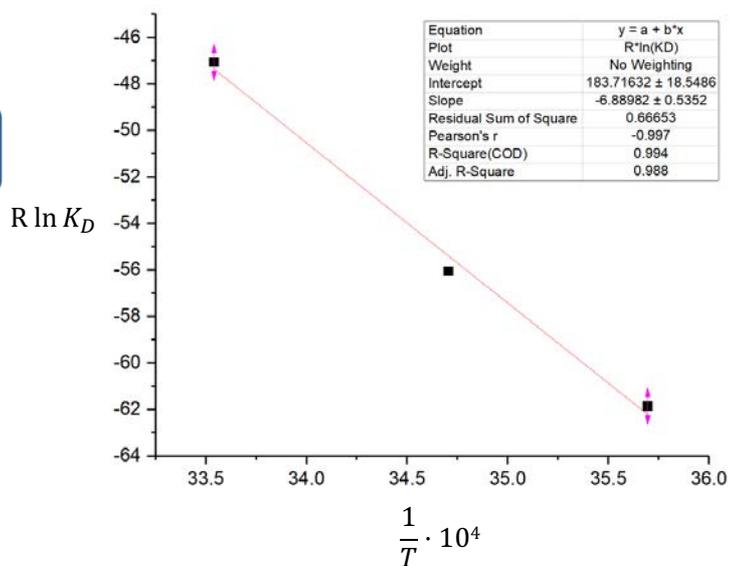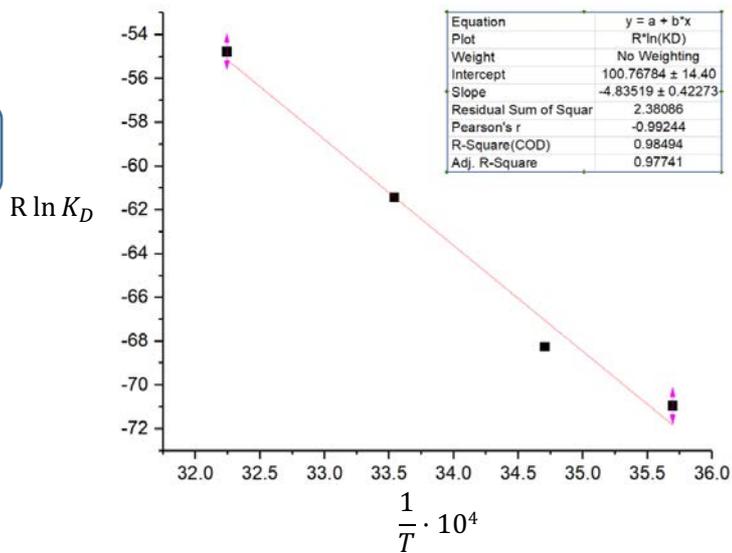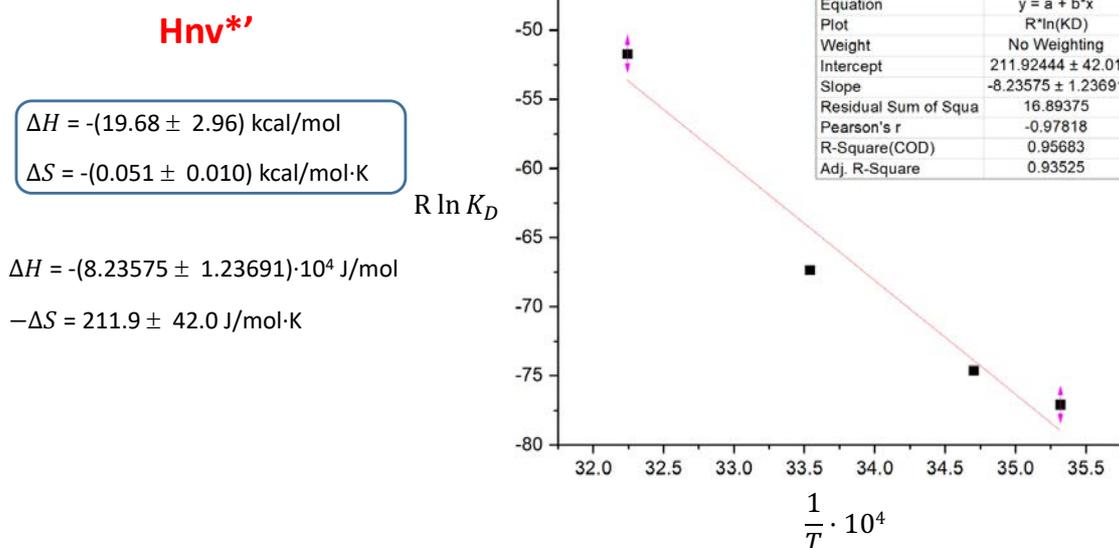
Equation 8.2

Table 8.3 Thermodynamic parameters for each antigen at physiological conditions, T = 310.15 K (all figures in Kcal/mol).

	Thr'	**Thr*'**	**Hnv*'**
ΔH	-(16.47±1.28)	-(11.55±1.01)	-(19.68±2.96)
$-310.15 \cdot \Delta S$	13.62±1.37	7.47±1.07	15.71±3.11
ΔG ($T = 310.15\ K$)	-(2.85±0.09)	-(4.08±0.06)	-(3.98±0.16)

8.10.7. Conjugation

Conjugation of MUC1 derivatives was conducted by INYCOM Biotech (Zaragoza, Spain) following a standard procedure as follows. First, m-maleimidobenzoyl-*N*-hydroxysuccinimide (3 mg) ester (MBS) was dissolved in 200 μL DMF. 70 μL of that mixture were then added to a KLH solution (5 mg KLH in 0.5 ml 10 mM phosphate buffer, pH 7.0). After stirring gently for 30 min at room temperature, the non-coupled (glyco)peptides were removed using a Sephadex G25 size exclusion column, which was equilibrated in 50 mM phosphate buffer (pH 6.0). The reaction mixture was then loaded and eluted using 50 mM phosphate buffer (pH 6.0). Purified KLH/MBS was obtained (3.5 mL) and 0.5 mL of water were then added. (Glyco)peptides (5 mg) were respectively dissolved in 100 μL of DMF and purified KLH/MBS (1 mL) was added. Reaction mixture

was shacked and NaOH 2 M (11 µL) was added (pH = 7.5). Stirring is kept for 3 hours at room temperature or overnight at 4°C. Finally, ammonium bicarbonate 0.1 M (3 mL) was added prior to lyophilization. The conjugation of glycopeptides to KLH was proved to be efficient by immunoassay experiments.

8.10.8. Immunizations

Immunizations were carried out by INYCOM Biotech (Zaragoza, Spain) in the following way. Eight to 12-week-old MUC1.Tg mice (Balb/c) that express human MUC1 at physiological level were immunized four-times at biweekly intervals at the base of the tail intradermally with **KLH-Cys-Thr*** or **KLH-Cys-sp^2-Thr*** vaccines (25 µg), together with complete or incomplete Freund's adjuvant. Endpoint was one week after 4th immunization.

8.10.9. Serologic assays

Anti-MUC1 IgG, IgG1, IgG2a, IgG2b, IgG3 and IgM antibody titers were determined by enzyme-linked immunosorbent assay, which were conducted by INYCOM Biotech (Zaragoza, Spain) as follows. ELISA plates (JetBioFil, China) were coated with 75 µL of a 5 µg/mL solution of either **Cys-Thr*** or **Cys-Thr** antigens in PBS and incubated for 1 h at 37 °C. Unreacted sites were blocked with 225 µL of BSA 1% in washing buffer (0.05% Tween-20 in PBS 1x) for 1 h at 37 °C. Afterwards, wells were washed twice with washing buffer. 75 µL of sera samples (blood taken from each mouse's tail diluted in 800 µL of BSA 1% solution) were allowed to bind to immobilized antigens for 30 min at 37 °C. Wells were then washed twice with washing buffer, prior to incubation with 75 µL of biotinylated anti-mouse antibodies (1:1000 dilution in conjugated solvent, INYCOM BIOTECH) for 30 min at 37 °C. After washing three times with washing buffer, incubation with 75 µL streptavidin-HRP (Sigma), 1:500 dilution in conjugation solvent (INYCOM BIOTECH), was carried out for 30 min at 37 °C. Wells were washed five times with washing buffer. 75 µL of TMB (INYCOM BIOTECH) were then incubated for 5 min at room temperature. For the final color-developing step, 75 µL of 0.2 M HCl were added to stop the reaction. Absorbance at 450 nm was read, using a iMark™ BIO RAD Microplate Reader.

Background absorbance values were subtracted from each measurement, i.e. absorbance obtained for wells coated with a different antigen to MUC1 mucin. **KLH-Cys-sp^2-Thr*** groups were compared to equivalent **KLH-Cys-Thr*** for each antibody isotype using a paired T-test with two-tailed correction; * $P < 0.05$, ** $P < 0.01$.

Raw data obtained for detection of circulating human antibodies against MUC1 vaccines, where plates were coated with **Cys-Thr*** glycopeptides, are shown in Table 8.4. Corresponding charts are depicted in Figure 4.28 chapter 4, except for IgG2b, whose titers are not recognized at detectable levels.

Table 8.4 Data obtained for detection of circulating human antibodies against KLH-Cys-Thr* and KLH-Cys-sp²-Thr*.

	Total IgG		IgG1		IgG2a		IgG2b		IgG3		IgM	
KLH-Cys-Thr*	0.529	0.282	0.729	0.304	0.015	0.001	-0.003	-0.007	0.230	0.051	0.014	-0.004
KLH-Cys-sp²-Thr*	1.744	1.660	4.641	4.583	0.153	0.361	0.005	-0.004	0.894	0.160	0.040	0.001

On the other hand, IgG1 antibody response observed after immunization with MUC1 vaccines, where plates where coated either with Cys-Thr* or Cys-Thr are shown in Table 8.5.

Table 8.5 Data obtained for detection of circulating human antibodies against KLH-Cys-Thr* and KLH-Cys-sp²-Thr*with different antigens coating the plates.

	Cys-Thr*		Cys-Thr	
KLH-Cys-Thr*	0.729	0.304	0.331	0.263
KLH-Cys-sp²-Thr*	4.641	4.583	4.317	4.208

8.11.   Experimental section chapter 5

8.11.1. (Glyco)peptides synthesis

Synthesis of peptide fP
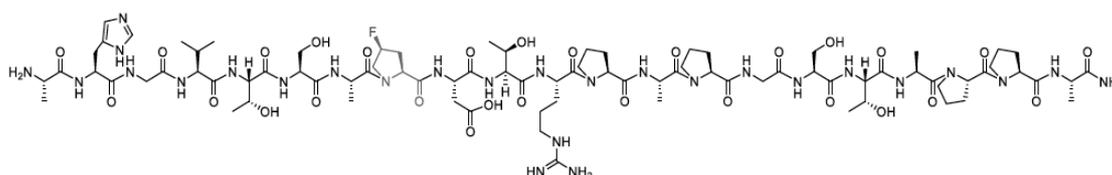
Semipreparative HPLC gradient R_t = 15.5 min

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	10	10	90
18	10	20	80

HRMS (ESI+) m/z: calcd. for $C_{83}H_{132}FN_{27}O_{28}$: [M+3H]$^{3+}$: 658.9979 found: 658.9991.

Synthesis of glycopeptide fP*

Semipreparative HPLC gradient R$_t$: 7.9 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H$_2$O+0.1% TFA (%)
0	20	11	89
9	20	15	85

HRMS (ESI+) m/z: calcd. for C$_{91}$H$_{145}$FN$_{28}$O$_{33}$: [M+3H]$^{3+}$: 726.6838 found: 726.6871.

Synthesis of peptide 2fP

Semipreparative HPLC gradient R$_t$: 6.6 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H$_2$O+0.1% TFA (%)
0	20	13	87
8	20	19	81

HRMS (ESI+) m/z: calcd. for C$_{83}$H$_{131}$F$_2$N$_{27}$O$_{28}$: [M+3H]$^{3+}$: 664.9948 found: 664.9973.

Synthesis of glycopeptide 2fP*

Semipreparative HPLC gradient R$_t$: 7.9 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H$_2$O+0.1% TFA (%)
0	20	12	88
9	20	17	83

HRMS (ESI+) m/z: calcd. for $C_{91}H_{144}F_2N_{28}O_{33}$: $[M+3H]^{3+}$: 732.6806 found: 732.6880.

Synthesis of glycopeptide P*'
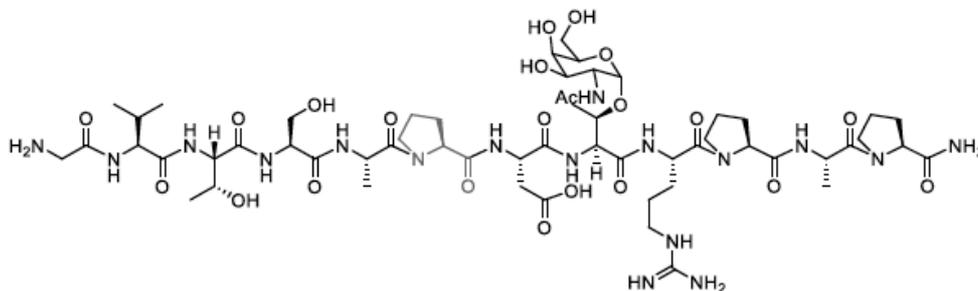
Semipreparative HPLC gradient R_t: 15.6 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H$_2$O+0.1% TFA (%)
0	20	5	95
20	20	15	85

HRMS (ESI+) m/z: calcd. for $C_{57}H_{95}N_{17}O_{22}$: $[M+H]^+$: 1370.6838 found: 1370.6885.

¹H NMR (400 MHz, D2O) δ (ppm): 0.95-0.97 (m, 6H, Val3γ), 1.21-1.26 (m, 6H, Thr4γ, Thr9γ), 1.31-1.41 (m, 6H, Ala6β, Ala12β), 1.69-2.34 (m, 20H, Ac, Pro7β, Pro11β, Pro13β, Pro7γ, Pro11γ, Pro13γ, Arg10β, Arg10γ, Val3β), 2.83-3.04 (m, 2H, Asp8β), 3.17- 3.27 (m, 2H, Gly2α).

Synthesis of glycopeptide fP*'
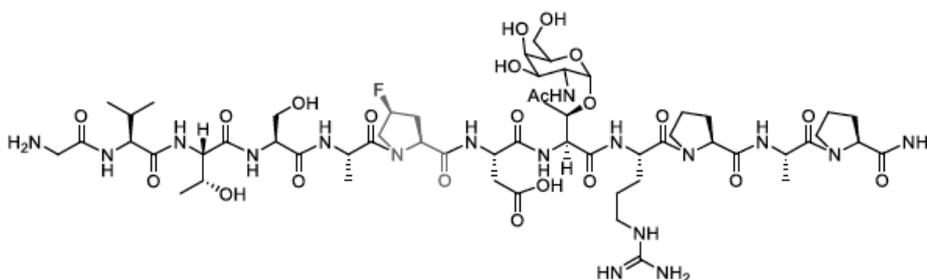
Semipreparative HPLC gradient R_t: 15.2 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H$_2$O+0.1% TFA (%)
0	20	5	95
20	20	15	85

HRMS (ESI+) m/z: calcd. for $C_{57}H_{94}FN_{17}O_{22}$: $[M+2H]^{2+}$: 694.8444 found: 694.8455.

¹H NMR (400 MHz, D2O) δ (ppm): 0.95-0.97 (m, 6H, Val3γ), 1.21-1.25 (m, 6H, Thr4γ, Thr9γ), 1.31-1.43 (m, 6H, Ala6β, Ala12β), 1.69-2.69 (m, 18H, Ac, Pro7β, Pro11β, Pro13β, Pro11γ, Pro13γ, Arg10β, Arg10γ, Val3β), 2.84-3.07 (m, 2H, Asp8β), 3.17-3.27 (m, 2H, Gly2α), 5.33-5.52 (m, 1H, Pro7γ).

Synthesis of glycopeptide 2fP*'
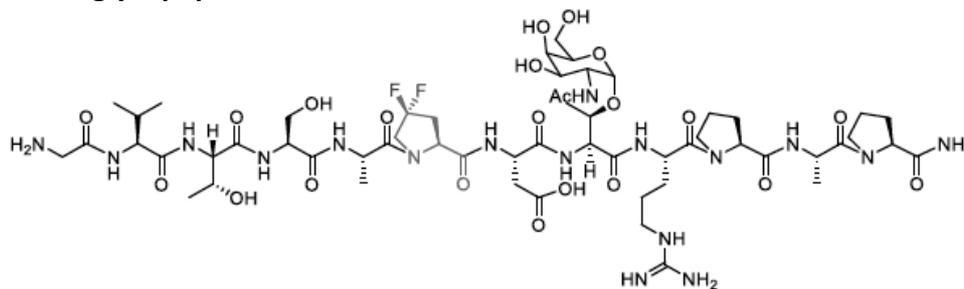
Semipreparative HPLC gradient R_t: 17.8 min.

Time (min)	Flow (mL/min)	Acetonitrile (%)	H_2O+0.1% TFA (%)
0	20	5	95
20	20	15	85

HRMS (ESI+) m/z: calcd. for $C_{57}H_{94}F_2N_{17}O_{22}$: $[M+H]^+$: 1406.6449 found: 1406.6703.

¹H NMR (400 MHz, D2O) δ (ppm): 0.95-0.97 (m, 6H, Val3γ), 1.19-1.28 (m, 6H, Thr4γ, Thr9γ), 1.31-1.42 (m, 6H, Ala6β, Ala12β), 1.69-3.05 (m, 20H, Ac, Pro7β, Pro11β, Pro13β, Pro11γ, Pro13γ, Arg10β, Arg10γ, Val3β, Asp8β), 3.17-3.27 (m, 2H, Gly2α).

8.11.2. BLI experiments

(Glyco)peptides at different concentrations (0 to 10 µM), were immobilized on amine-reactive biosensors (AR2G biosensors). Reactive sites were subsequently blocked. Following the protocol described in the *Experimental section 8.7*, antibody binding step was then performed. Figure 8.9 shows the typical BLI curves and the corresponding fit.

Figure 8.9 BLI curves and fit obtained for the (glyco)peptides.

Steady state analyses were then performed using each maximal binding response *versus* each range of concentrations to obtain the dissociation curves for the different (glyco)peptides. Dissociation constants (K_D) were consequently deduced from each chart (Figure 8.10).
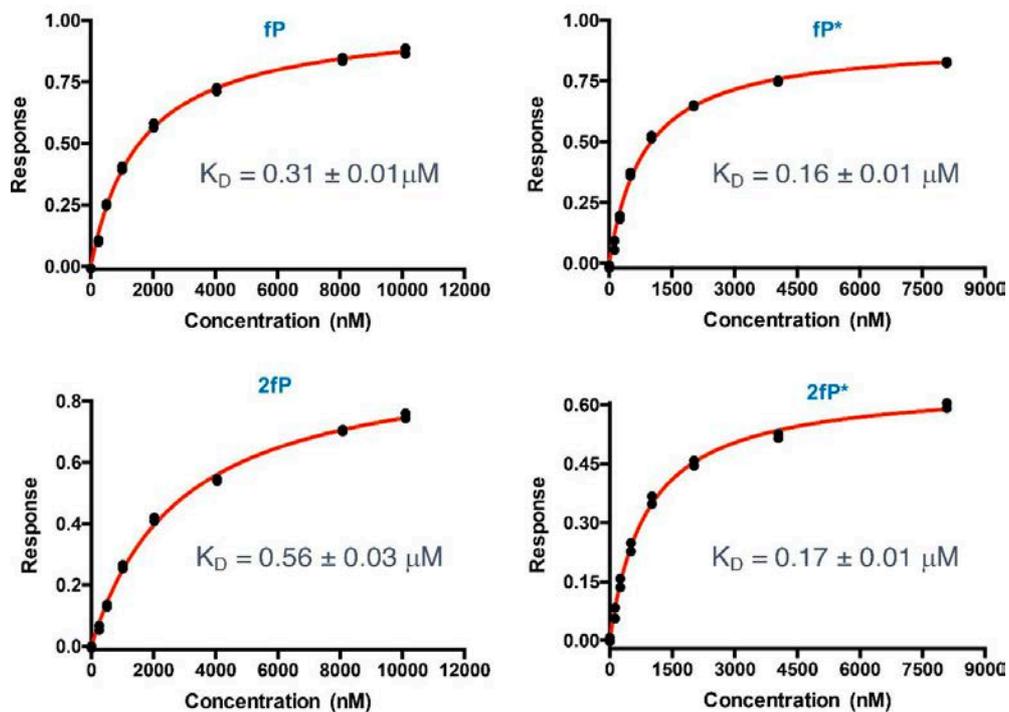
Figure 8.10 Steady state analyses and K_D constants obtained for the studied (glyco)peptides.

171

8.11.3. Crystal structure

X-ray structure of **fP*'** (pdb ID: 5OWP) was obtained as previously described in *Experimental section 8.8*. Electron density maps are F_O–F_C syntheses (blue) contoured at 2.2 σ for glycopeptide **fP*'** (Figure 8.11). The amino acid residues and the GalNAc moiety are colored in blue and green, respectively. The fluorine atom is in magenta. It is important to note that Pro at the C-terminal region could not be resolved.
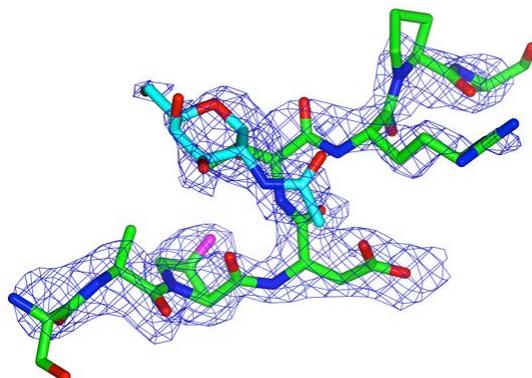
*Figure 8.11 Electron density maps for glycopeptide **fP*'** (pdb ID: 5OWP).*

8.11.4. Serologic Assay

High-binding ELISA plates (JetBioFil, China) were coated with 100 µL of a 20 µM solution (in PBS buffer, pH 7) of each glycopeptide and incubated for 2 h at 37 °C. Wells were washed three times with washing buffer (0.05% Tween-20 in PBS 1x) and blocked for 1 h at 37 °C with 100 µL of ELISA buffer (1% BSA in washing buffer).

The wells were again washed 3 times with washing buffer, before incubation with sera samples (dilution 1:50 in ELISA buffer), performed at room temperature, for 90 min. After incubation, three washing buffer washing steps followed before incubation with donkey anti human IgG H&L antibody conjugated to HRP (Abcam, UK) for 1 h at room temperature (final concentration 300 ng/mL).

After washing, incubation with 90 µL of 3,3',5,5'-tetramethylbenzidine (TMB) 1x solution (eBiosciences, ThermoFisher Scientific, USA) was carried out for 10 min at room temperature. For the final color-developing step 50 µL of 2N H_2SO_4 were added to stop the reaction. Absorbance at 450 nm was read within the next 10 min, using a Tecan Infinite M200 plate-reader.

Background absorbance values were subtracted, *i.e.* absorbance obtained for wells coated with the same MUC1 mucin variant but incubated only with the secondary antibody. Values were normalized to those obtained for healthy controls.

All groups were compared to **P*** using Wilcoxon matched-pairs signed rank test; Experimental groups were compared to equivalent healthy controls using an unpaired T-test with Welch's correction (one-tailed); * p<0.05; ** p<0.02.

Raw data obtained for detection of circulating human antibodies against MUC1 glycopeptides are shown in Table 8.6. Data are shown as absorbance at 450 nm to which the background absorbance value was subtracted. "Signal-to-noise" corresponds to the average of absorbance values of each group normalized to the values obtained for "healthy controls".

Table 8.6 Data obtained for detection of circulating human antibodies against P* and 2fP*.

Diagnostic	Age	Absorbance at 450 nm (corrected for "blank")		"Signal-to-noise" ratio	
		P*	2fP*	P*	2fP*
Prostatic Adenocarcinoma	60	0.3171	0.1736	1.4806	2.4892
	61	0.1407	0.0431	0.6573	0.6178
	61	0.0385	-0.0026	0.1800	-0.0380
	62	0.2853	0.1971	1.3321	2.8254
	63	0.1354	0.0381	0.6323	0.5469
	64	2.2600	1.9428	10.5546	27.8491
	68	0.9091	0.5183	4.2456	7.4298
	71	2.3987	2.3477	11.2021	33.6532
	78	0.8835	0.3914	4.1260	5.6107
Average		**0.8187 *** (p=0.0424 vs Healthy)	**0.6277 *** (p=0.0484 vs Healthy)	**3.8234**	**8.9982 *** (p=0.0273 vs P*)
Prostatic Benign Hyperplasia	64	0.2351	0.1880	1.0982	2.6950
	65	0.2384	0.3711	1.1136	5.3197
	65	0.1864	0.2030	0.8707	2.9100
	65	0.1055	0.2215	0.4927	3.1759
	67	0.0781	0.1462	0.3650	2.0958
	68	0.3628	0.6662	1.6945	9.5492
	72	0.0511	-0.0236	0.2389	-0.3390
	72	0.5257	0.1613	2.4548	2.3115
	78	0.3893	0.2420	1.8178	3.4698
	84	0.0408	-0.0878	0.1908	-1.2593
	87	0.7009	0.4896	3.2735	7.0183
Average		**0.2649** N.S. (p=0.2646 vs Healthy)	**0.2343 *** (p=0.00234 vs Healthy)	**1.2373**	**3.3588 **** (p=0.0068 vs P*)
Healthy Male Controls	56	0.2068	0.0015	0.9658	0.0215
	60	0.2738	0.1211	1.2787	1.7360
	61	0.2408	0.1148	1.1248	1.6456
	62	0.3064	0.1542	1.4307	2.2097
	62	0.0428	-0.0427	0.2001	-0.6128
Average		**0.2141**	**0.0698**	**1.0000**	**1.0000**

8.12.　References

1. Kaiser, E., Colescott, R. L., Bossinger, C. D. & Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* **34,** 595–598 (1970).

2. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).

3. Kiyohara, K., Gubbins, K. & Panagiotopoulos, A. Phase coexistence properties of polarizable water models. *Mol. Phys.* **94,** 803–808 (1998).

4. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

5. Wang, J., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules ? *J. Comput. Chem.* **21,** 1049–1074 (2000).

6. Wang, J., Wolf, R. M., Caldwell, J., Kollman, P. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25,** 1157–1174 (2004).

7. Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., … Kollman, P. A. *Amber 2016 Reference Manual.* (2016). doi:10.1021/ct200909j

8. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. & Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11,** 3696–3713 (2015).

9. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Lachele, F. B. & Woods, R. J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29,** 622–655 (2007).

10. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23,** 1623–1641 (2002).

11. Corzana, F., Busto, J. H., Engelsen, S. B., Jimenez-Barbero, J., Asensio, J. L., Peregrina, J. M. & Avenoza, A. Effect of beta-O-glucosylation on L-Ser and L-Thr Diamides: A bias toward alpha-helical conformations. *Chem. Eur. J.* **12,** 7864–7871 (2006).

12. 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. Deciphering the non-equivalence of serine and threonine O-glycosylation points: implications for molecular recognition of the Tn antigen by an anti-MUC1 antibody. *Angew. Chem. Int. Ed.* **127,** 9968–9972 (2015).

13. Kabsch, W. XDS. *Acta Crystallogr.* **D66,** 125–132 (2010).

14. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., Mccoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr.* **D67,** 235–242 (2011).

15. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Biol. Crystallogr.* **D60,** 2126–2132 (2004).

16. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr.* **D67,** 355–367 (2011).

17. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 283–291 (1993).

doi:10.1107/S0021889892009944

18. Cocinero, E. J., Stanca-Kaposta, E. C., Gamblin, D. P., Davis, B. G. & Simons, J. P. Peptide secondary structures in the gas phase: consensus motif of N-linked glycoproteins. *J. Am. Chem. Soc.* **131,** 1282–1287 (2009).

19. Usabiaga, I., González, J., León, I., Arnáiz, P. F., Cocinero, E. J. & Fernández, J. A. Influence of the anomeric conformation intermolecular interactions of glucose in the intermolecular interactions of glucose. *J. Phys. Chem. Lett.* **8,** 1147–1151 (2017).

20. Robertson, E. G. & Simons, J. P. Getting into shape: Conformational and supramolecular landscapes in small biomolecules and their hydrated clusters. *Phys. Chem. Chem. Phys.* **3,** 1–18 (2001).

21. Çarçabal, P., Patsias, T., Hünig, I., Liu, B., Kaposta, C., Snoek, L. C., Gamblin, D. P., Davis, B. G. & Simons, J. P. Spectral signatures and structural motifs in isolated and hydrated monosaccharides: phenyl alpha- and beta-L-fucopyranoside. *Phys. Chem. Chem. Phys.* **8,** 129–136 (2006).

22. Halgren, T. A. Merck Molecular Force Field I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **17,** 490–519 (1996).

23. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc. A* **117,** 5179–5197 (1995).

24. Shivakumar, D., Harder, E., Damm, W., Friesner, R. A. & Sherman, W. Improving the prediction of absolute solvation free energies using the next generation OPLS Force Field. *J. Chem. Theory Comput.* **8,** 2553–2558 (2012).

25. Parish, C. A., Lombardi, R., Smith, E. & Goldberg, A. A comparison of the Low Mode and Monte Carlo conformational search methods. *J. Mol. Graph. Model.* **21,** 129–150 (2002).

26. Kolossváry, I. & Guida, W. C. Low mode search. An sfficient, automated computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. *J. Am. Chem. Soc.* **118,** 5011–5019 (1996).

27. Chang, G., Guida, W. C. & Still, C. W. An internal coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **111,** 4379–4386 (1989).

28. Fariborz, M., J., R. N. G., C., G. W., Rob, L., Mark, L., Craig, C., George, C., Thomas, H. & Clark, S. W. Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **11,** 440–467 (1990).

29. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., … Fox, D. J. Gaussian 09, Revision E.01. *Gaussian Inc., Wallingford CT* (2009).

30. Plattner, C., Michael, H. & Sewald, N. One-pot azidochlorination of glycals. *Org. Lett.* **13,** 545–547 (2011).

31. Martínez-Sáez, N. Diseño racional de derivados de mucina como candidatos a vacunas contra el cáncer. Doctoral dissertation. (Universidad de La Rioja, 2013).

32. 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. The quest for anticancer vaccines: deciphering the fine-epitope specificity of cancer-related monoclonal antibodies by combining microarray screening and saturation transfer difference NMR. *J. Am. Chem. Soc.* **137,** 12438–12441 (2015).

33. Atkins, P. & De Paula, J. *Physical Chemistry*. (Oxford, 2006).

Supplementary information

9

9.1. **NMR experiments chapter 3**

Compound 3-II

^{1}H NMR 400 MHz in CDCl$_3$ registered at 298K
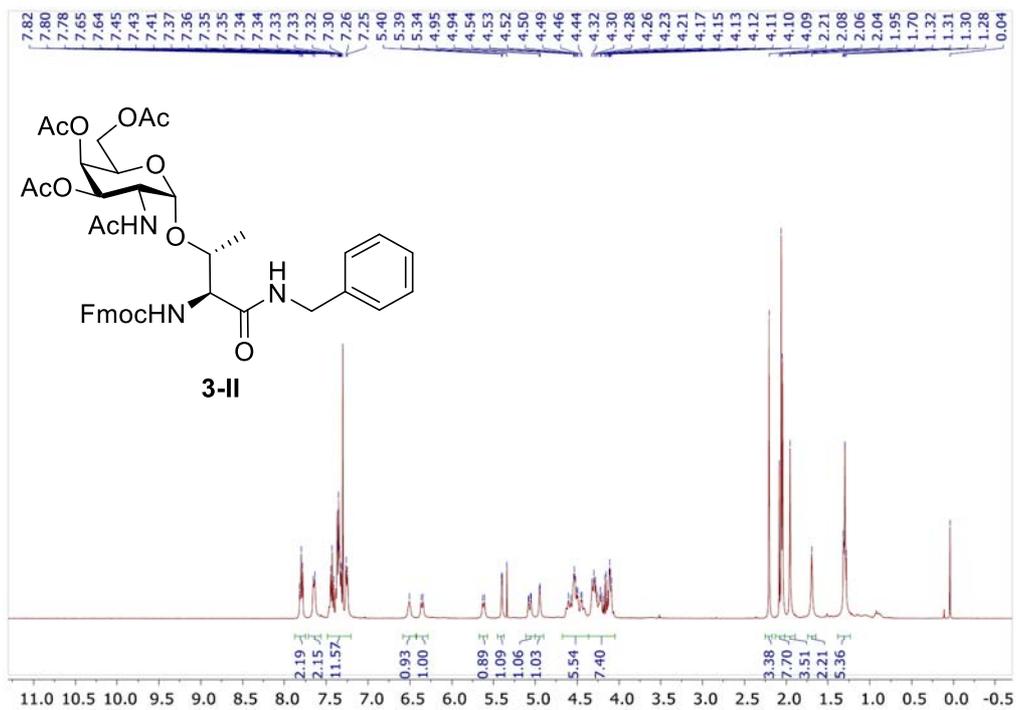
^{13}C NMR 100 MHz in CDCl$_3$ registered at 298K
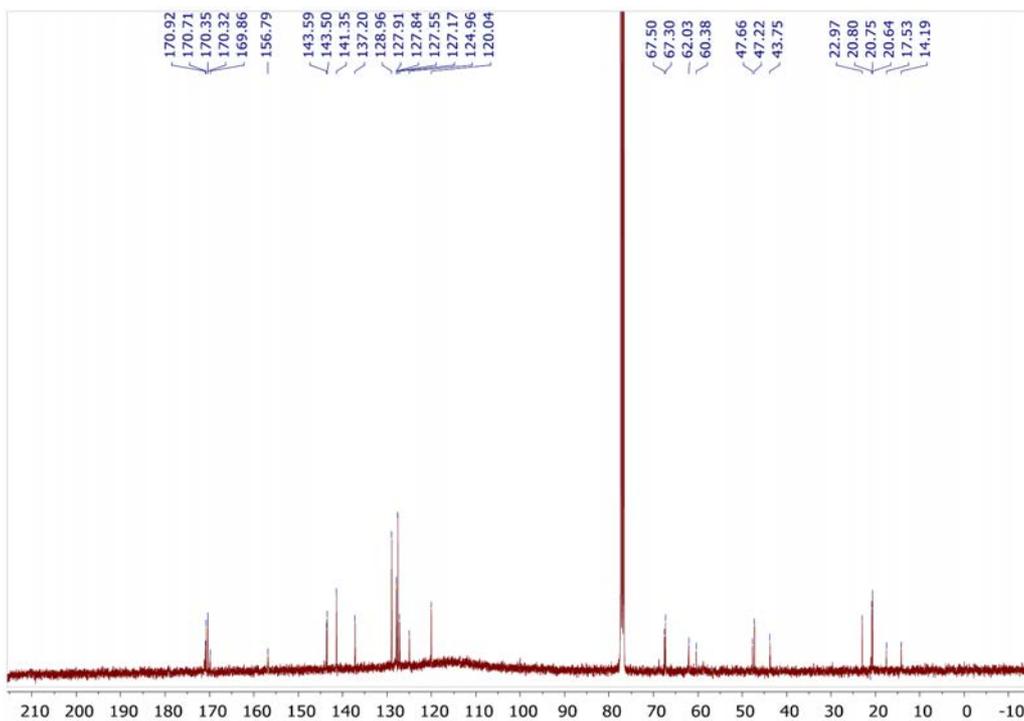
COSY in CDCl₃ registered at 298K
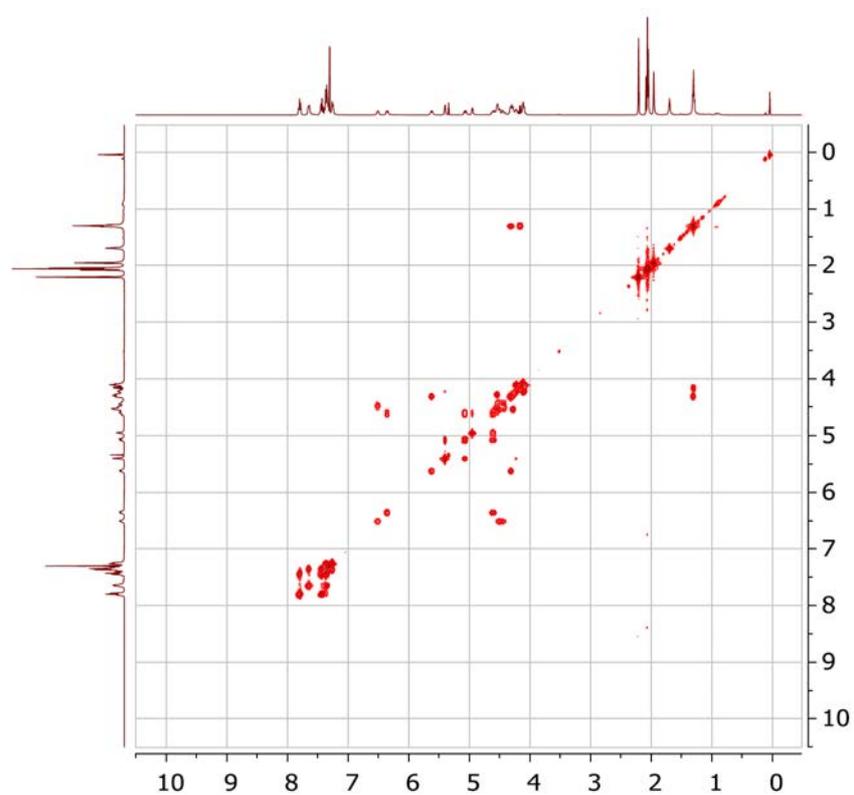
HSQC in CDCl₃ registered at 298K
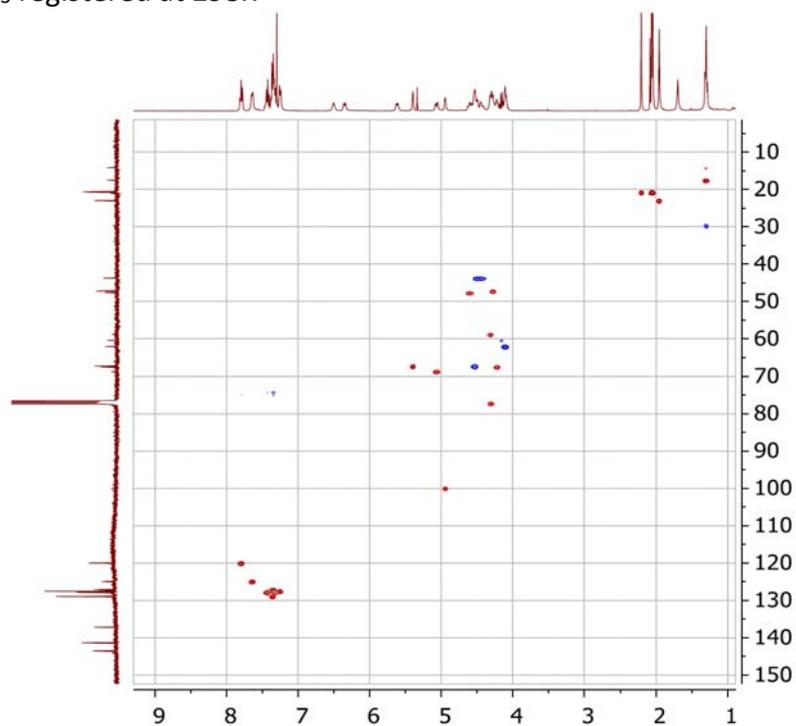
Compound 3-III
¹H NMR 400 MHz in CDCl₃ registered at 298K
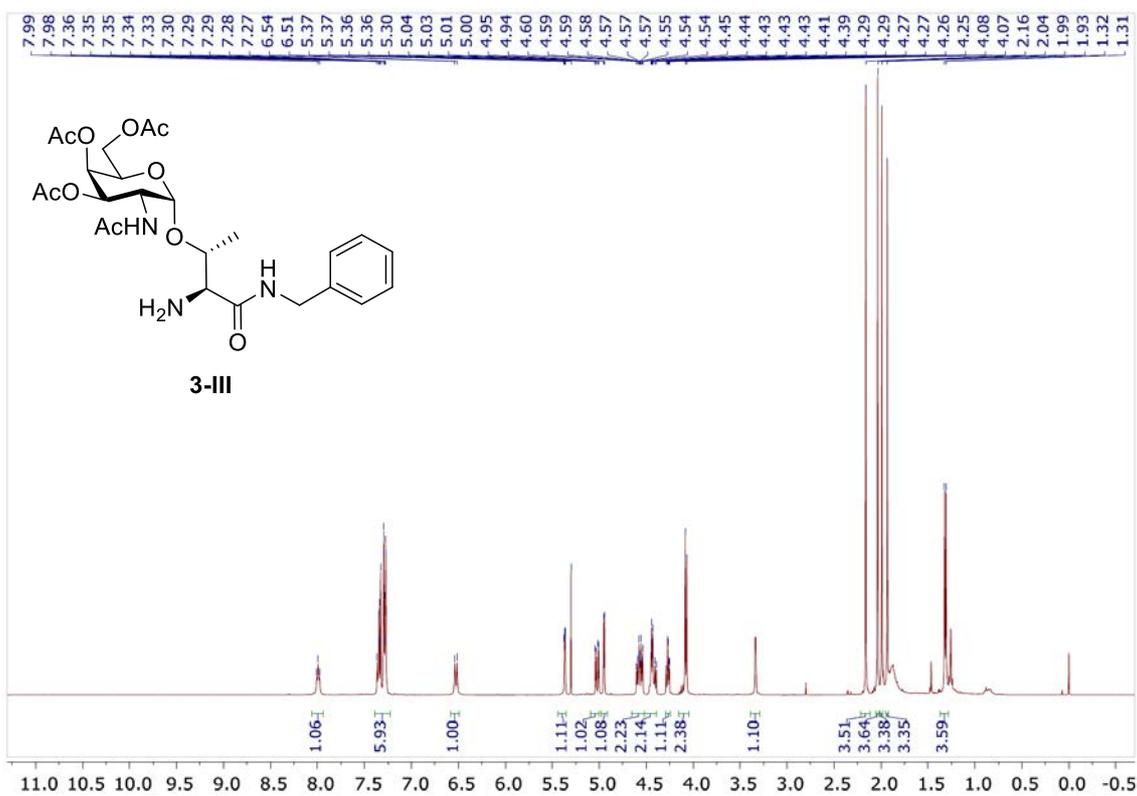
¹³C NMR 100 MHz in CDCl₃ registered at 298K
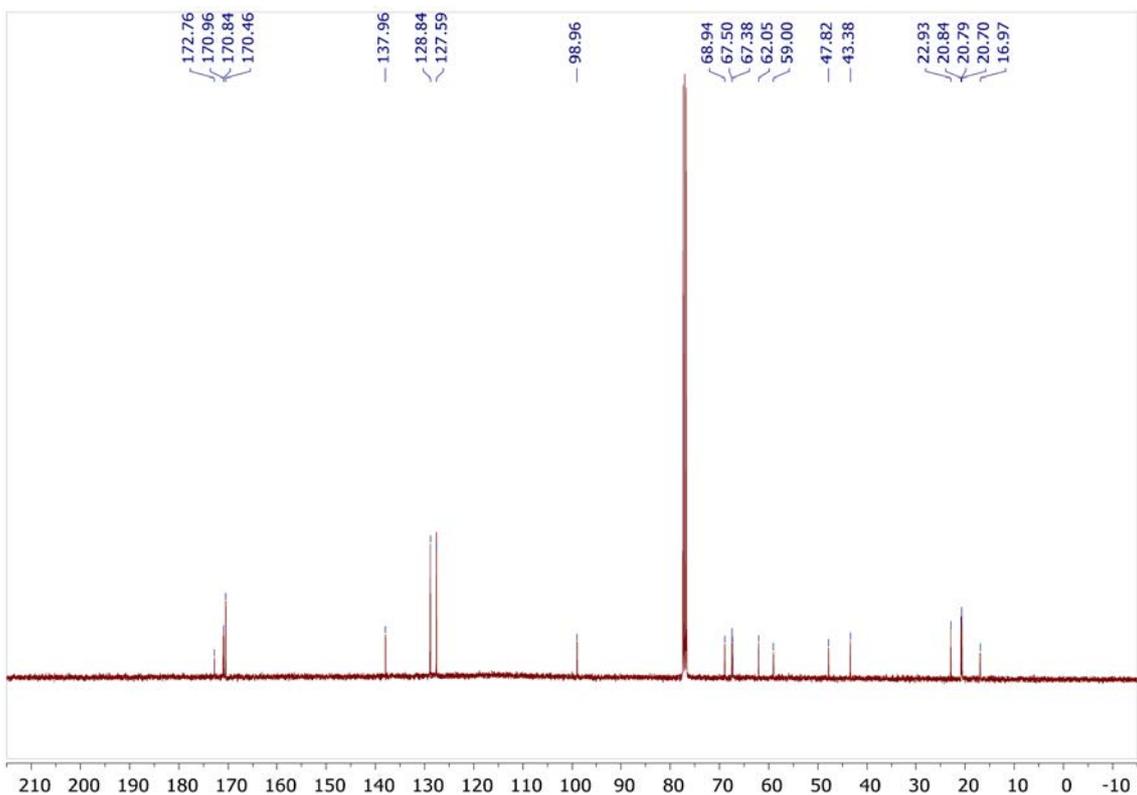
COSY in CDCl₃ registered at 298K
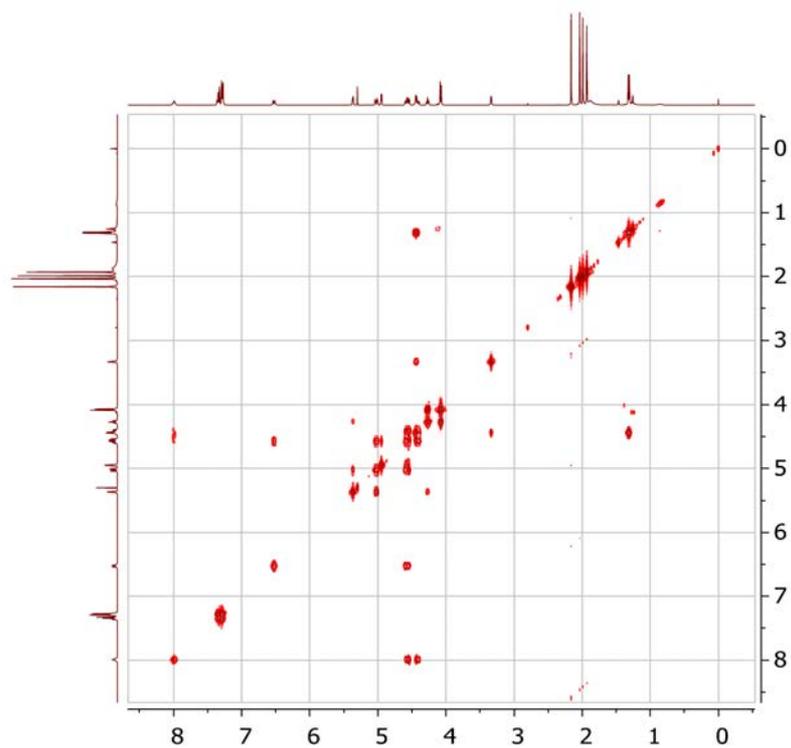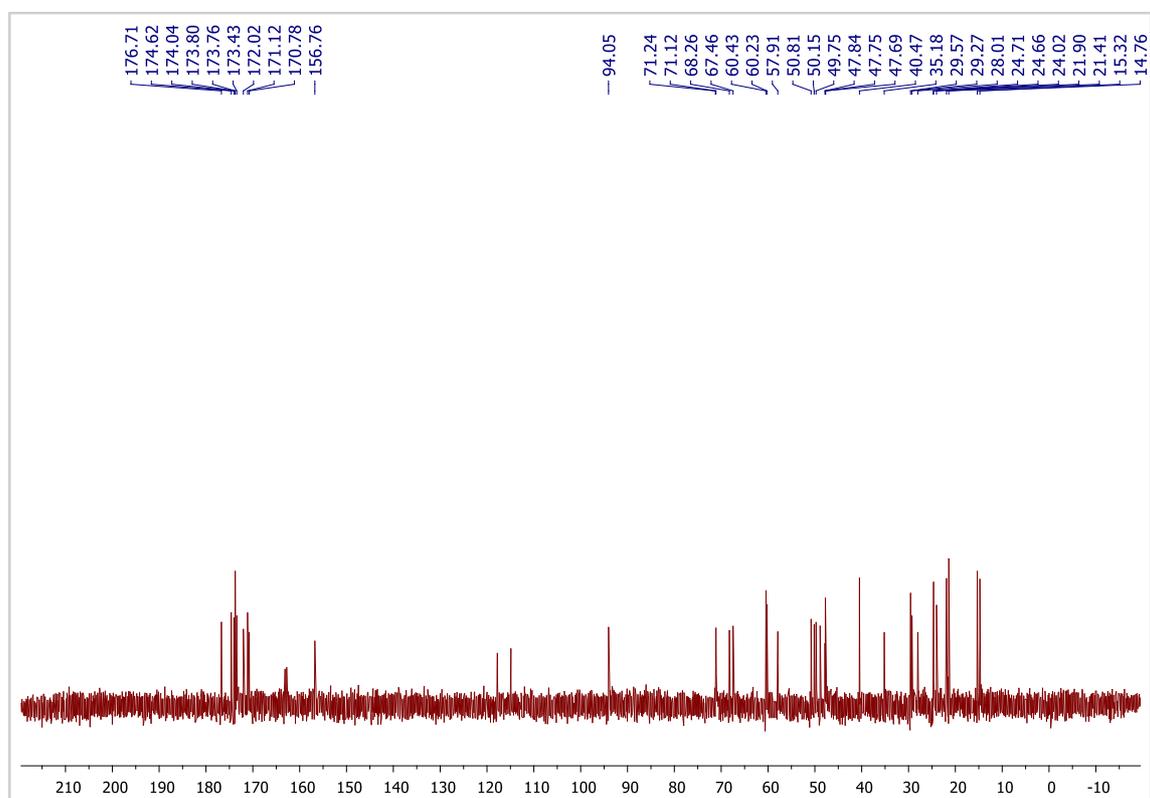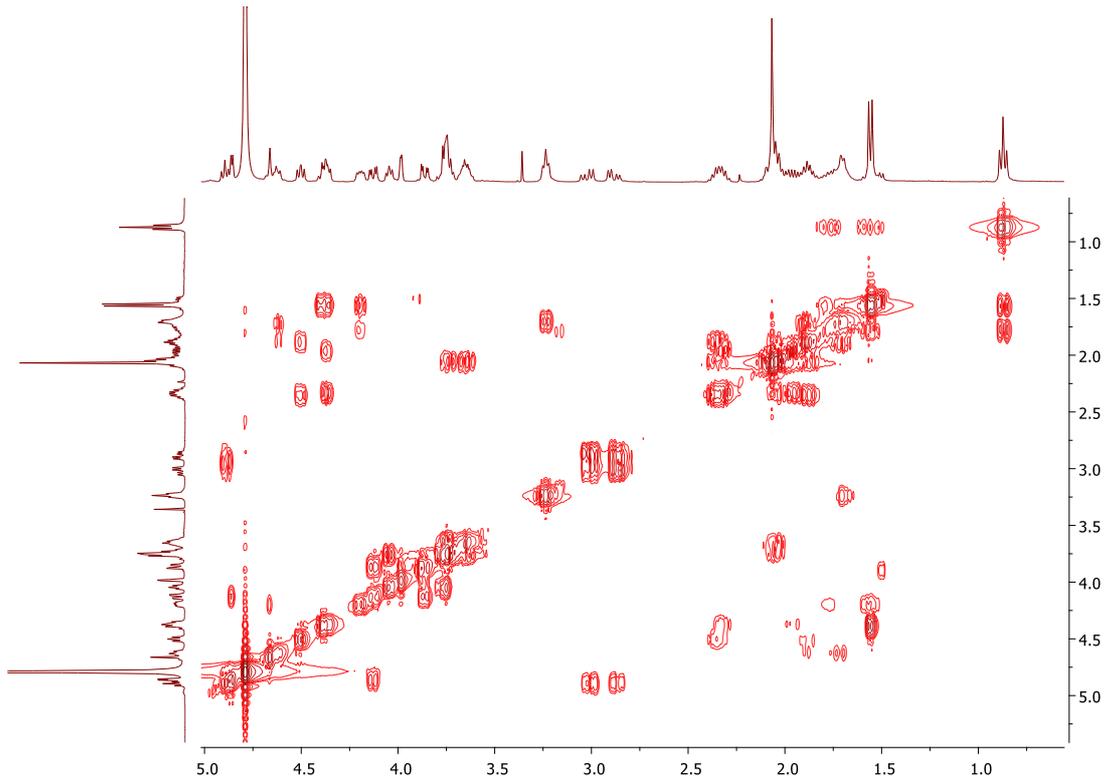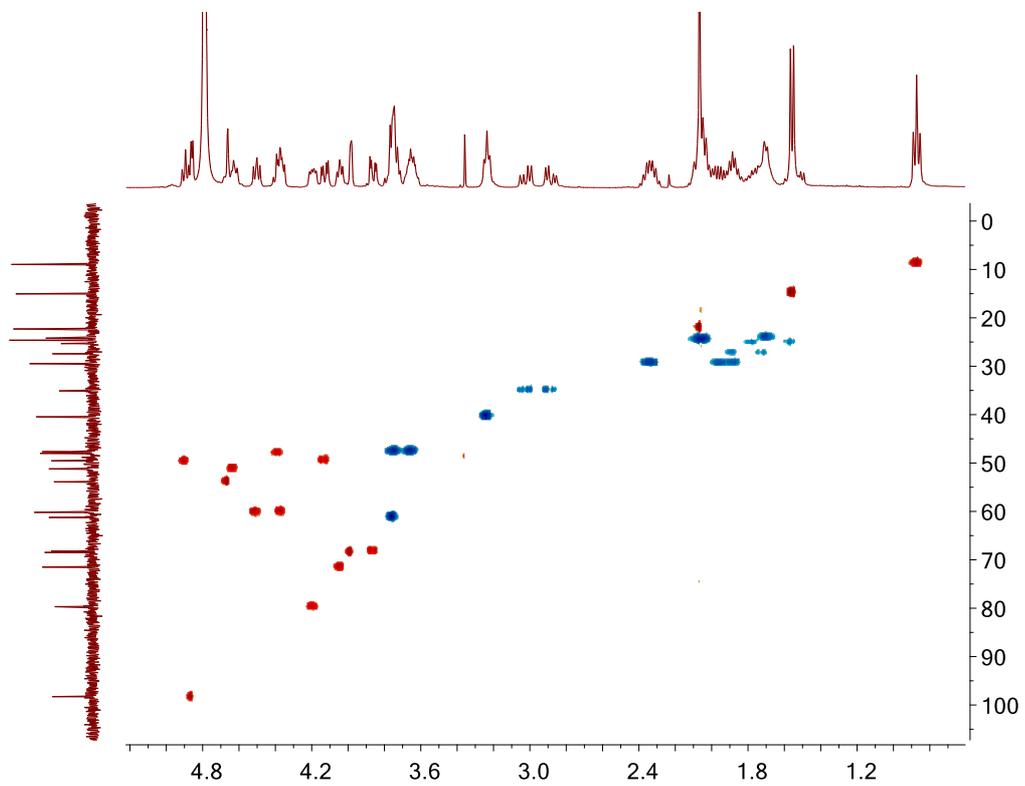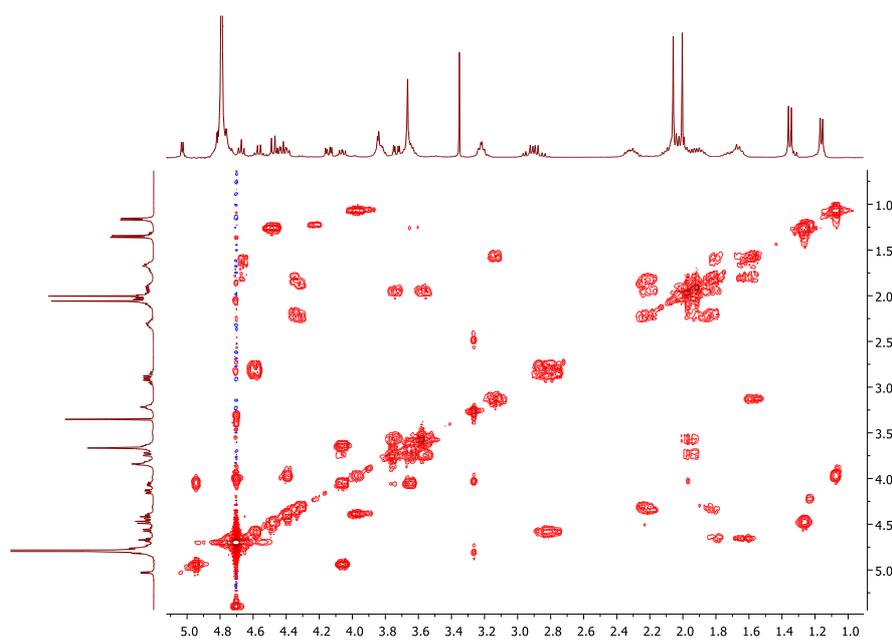

HSQC in CDCl₃ registered at 298K
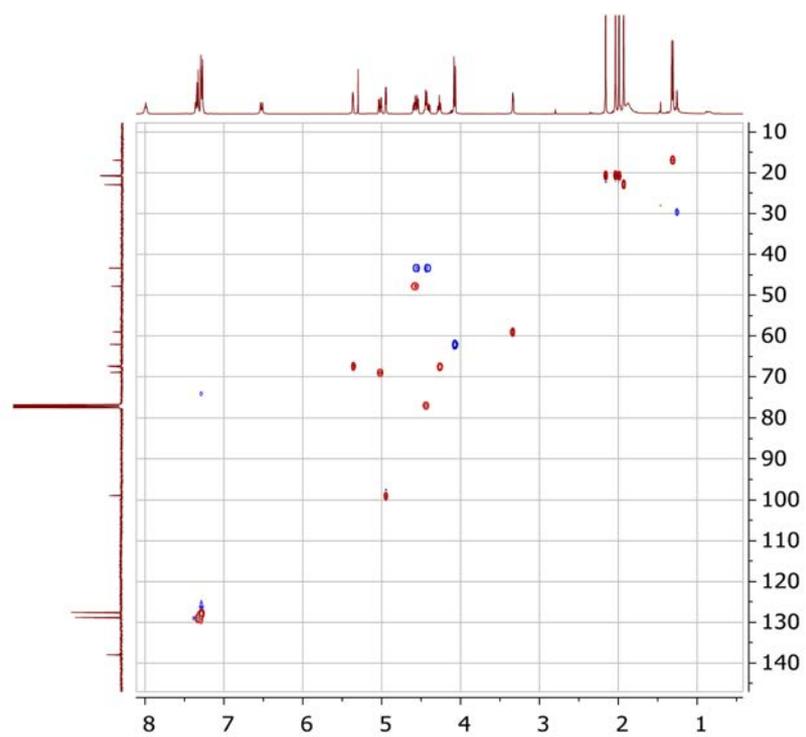
Compound 3-IV
^{1}H NMR 400 MHz in CDCl$_3$ registered at 298K
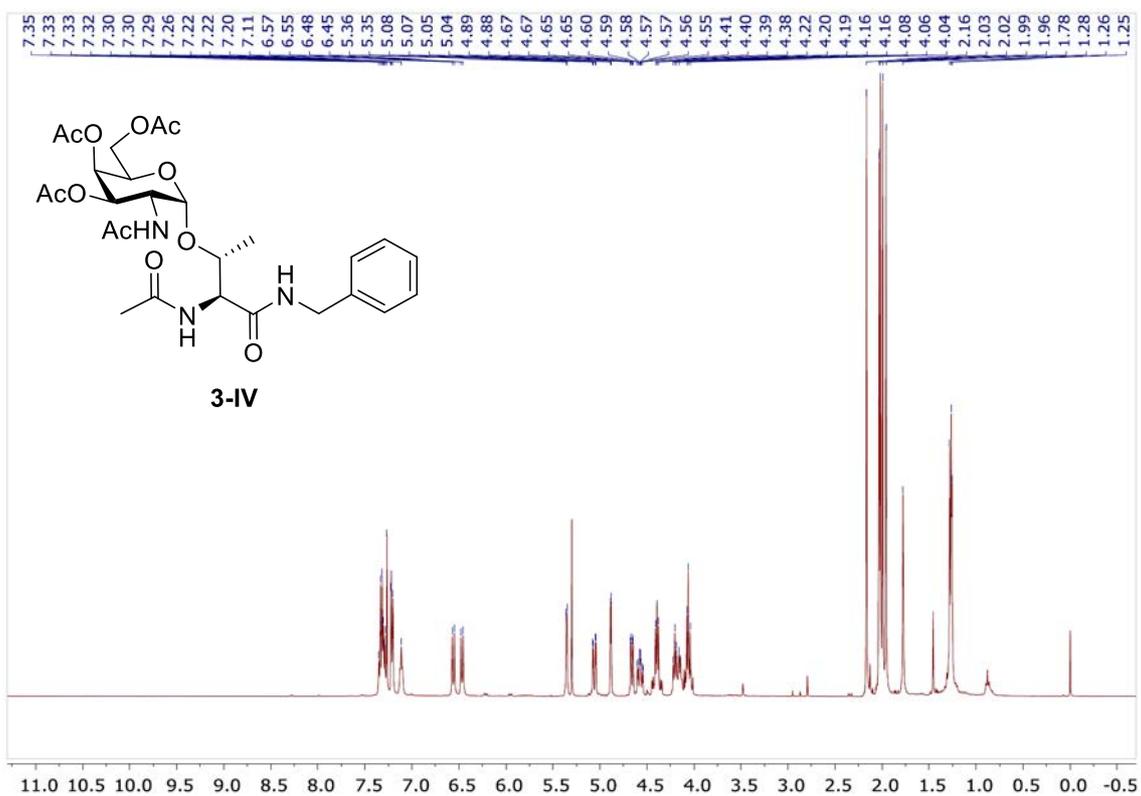
^{13}C NMR 100 MHz in CDCl$_3$ registered at 298K
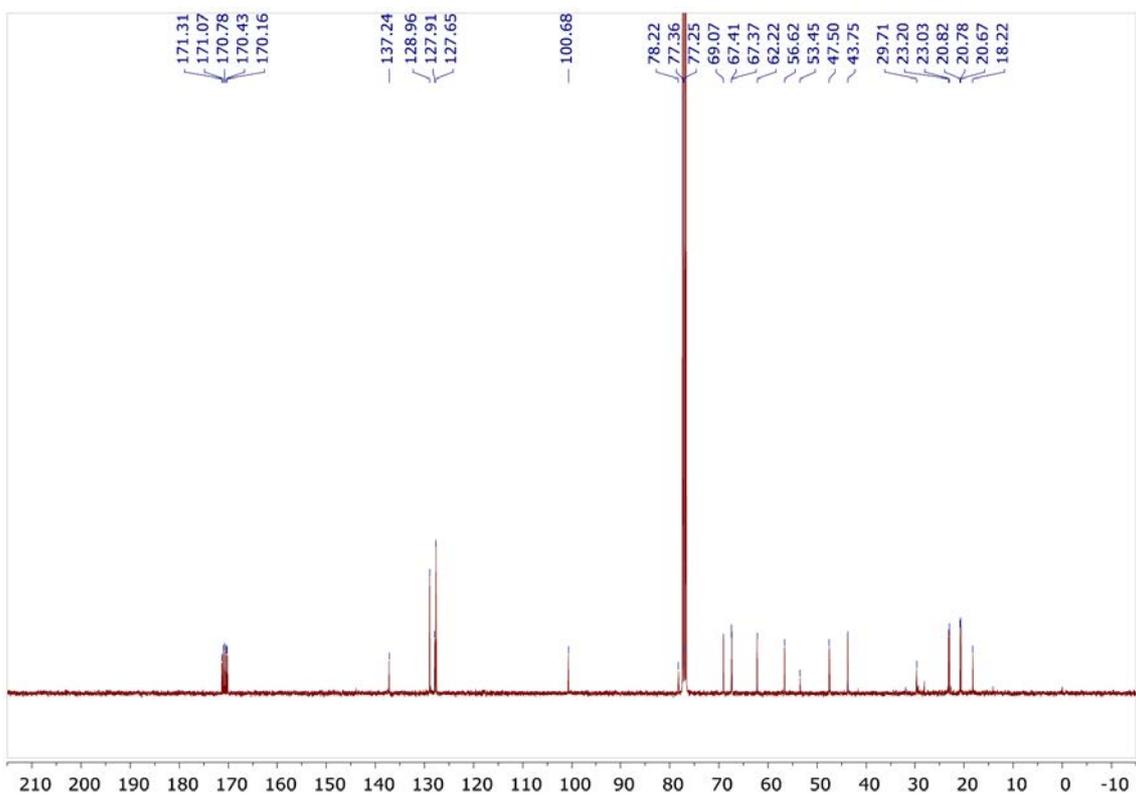
COSY in CDCl₃ registered at 298K
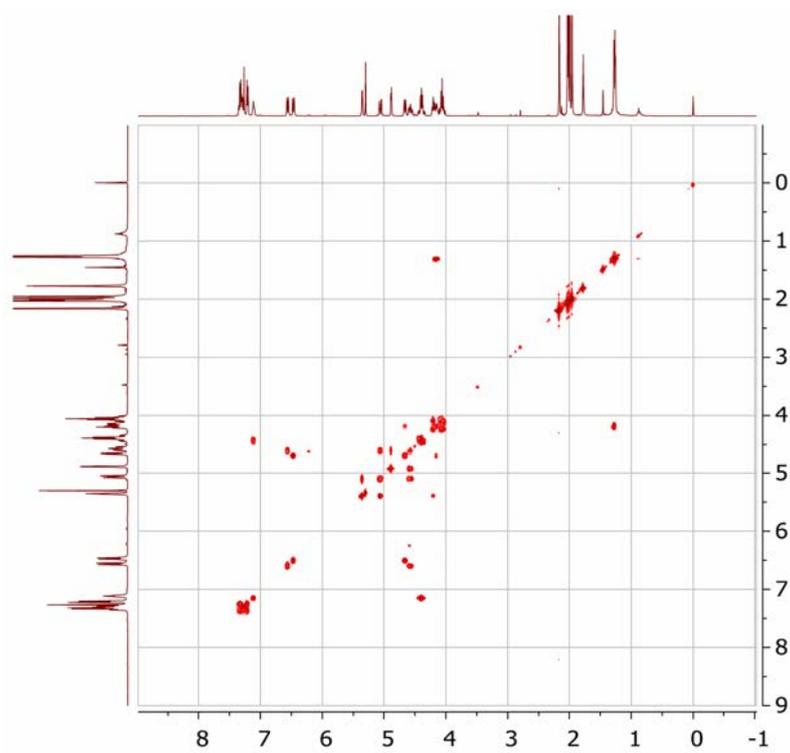
HSQC in CDCl₃ registered at 298K
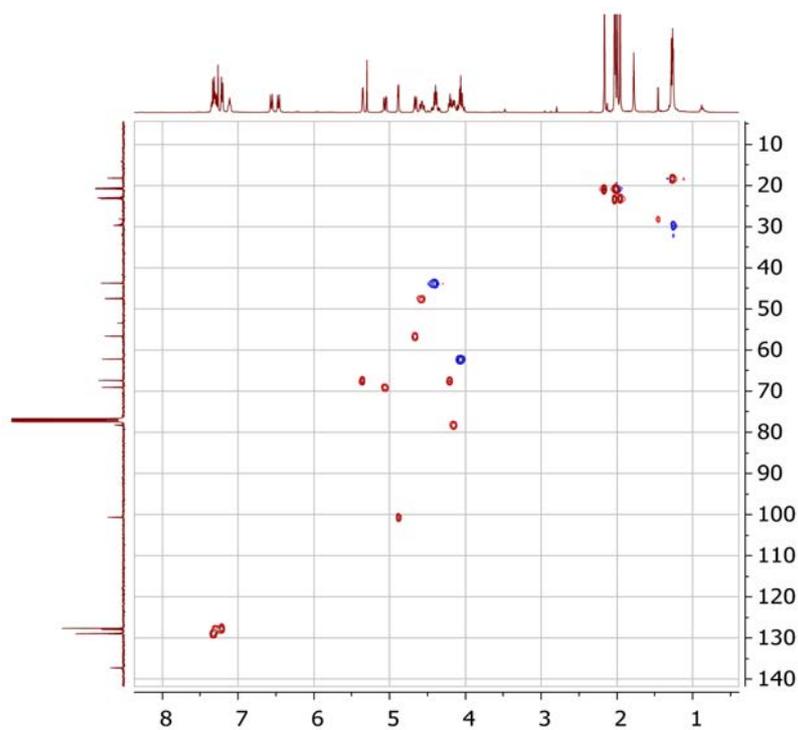
Tn-Thr'

^{1}H NMR 400 MHz in D$_2$O registered at 298K
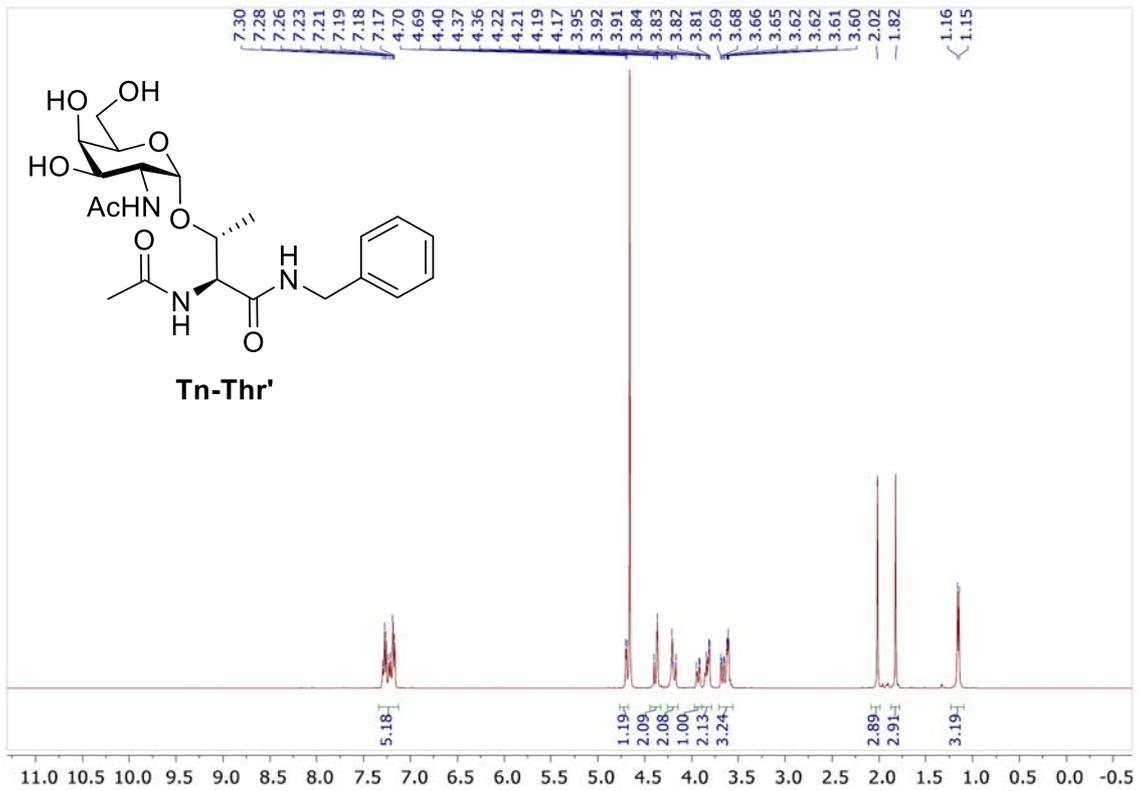
^{13}C NMR 100 MHz in D$_2$O registered at 298K
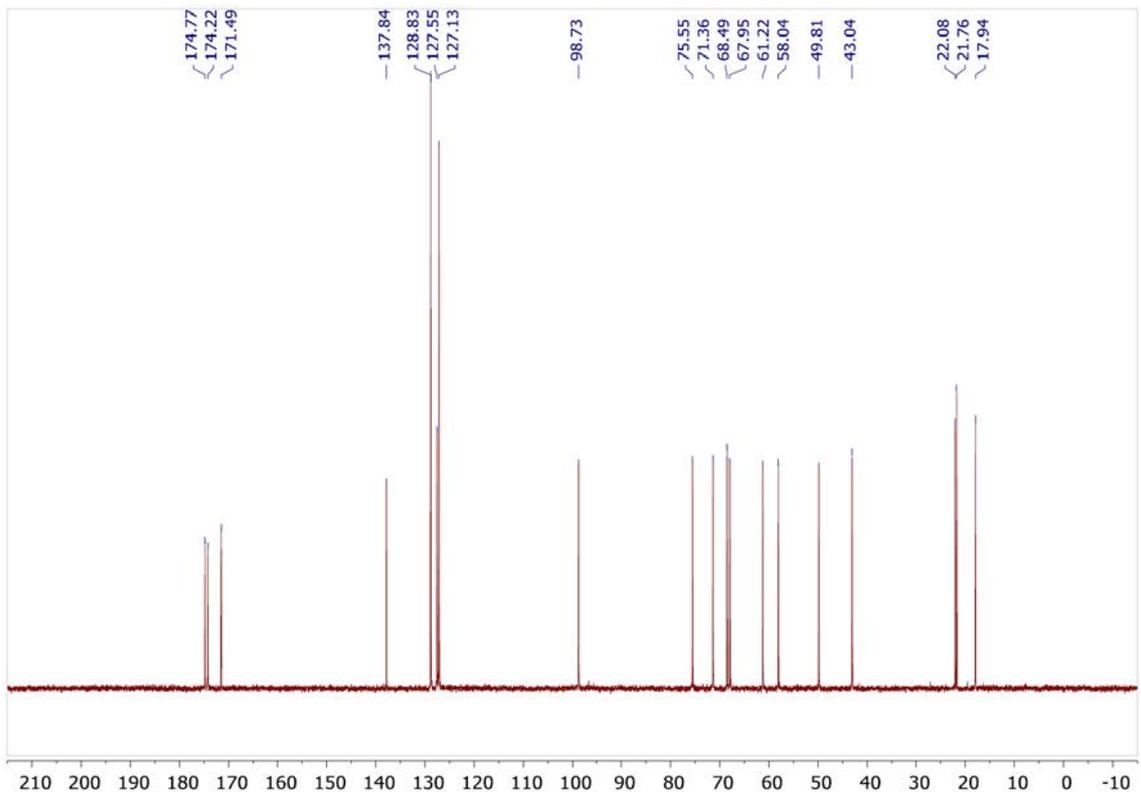
COSY in D₂O registered at 298K
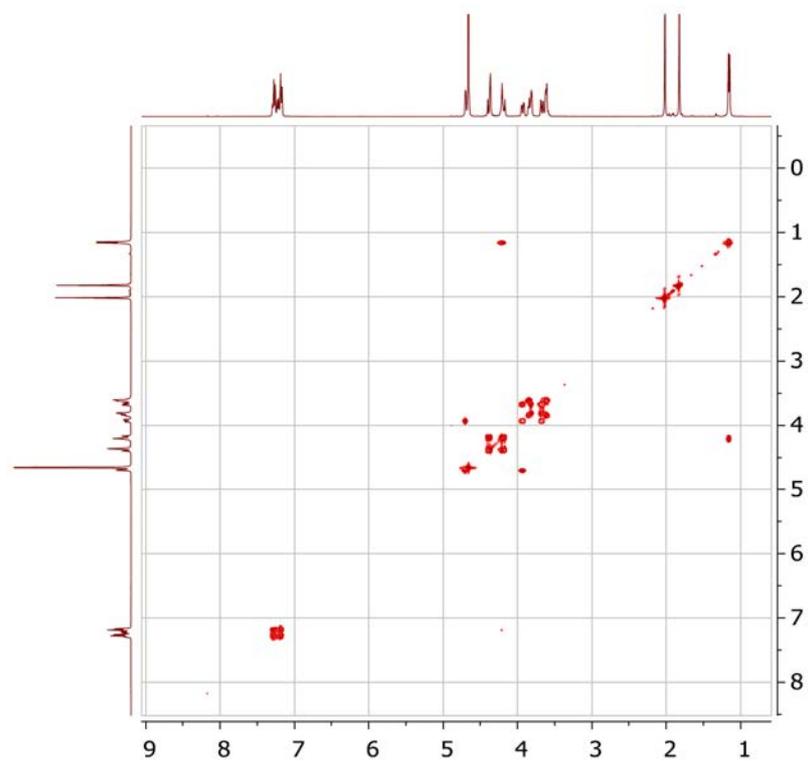
HSQC in D₂O registered at 298K
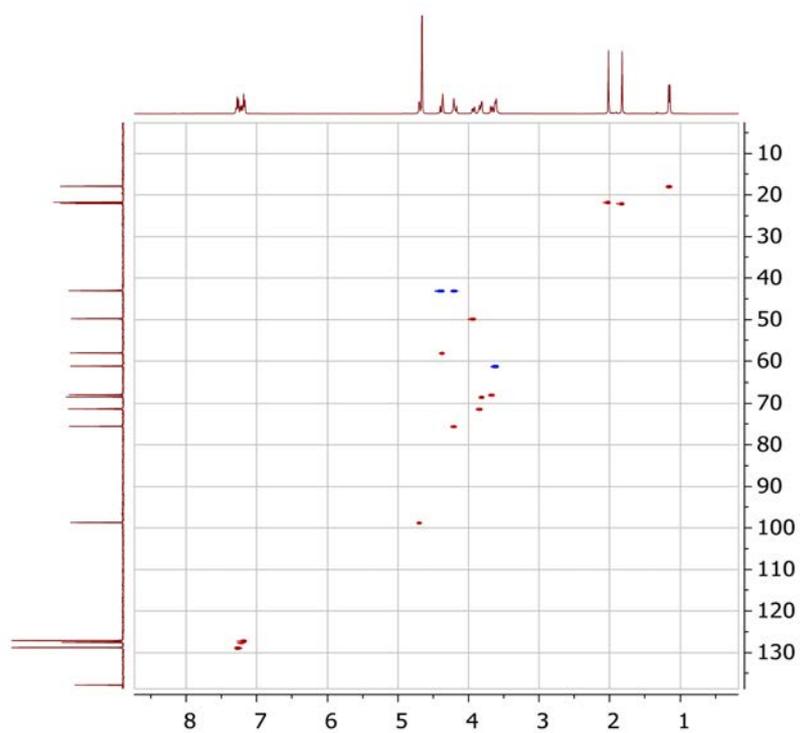
Compound 9

^{1}H NMR 400 MHz in CDCl$_3$ registered at 298K
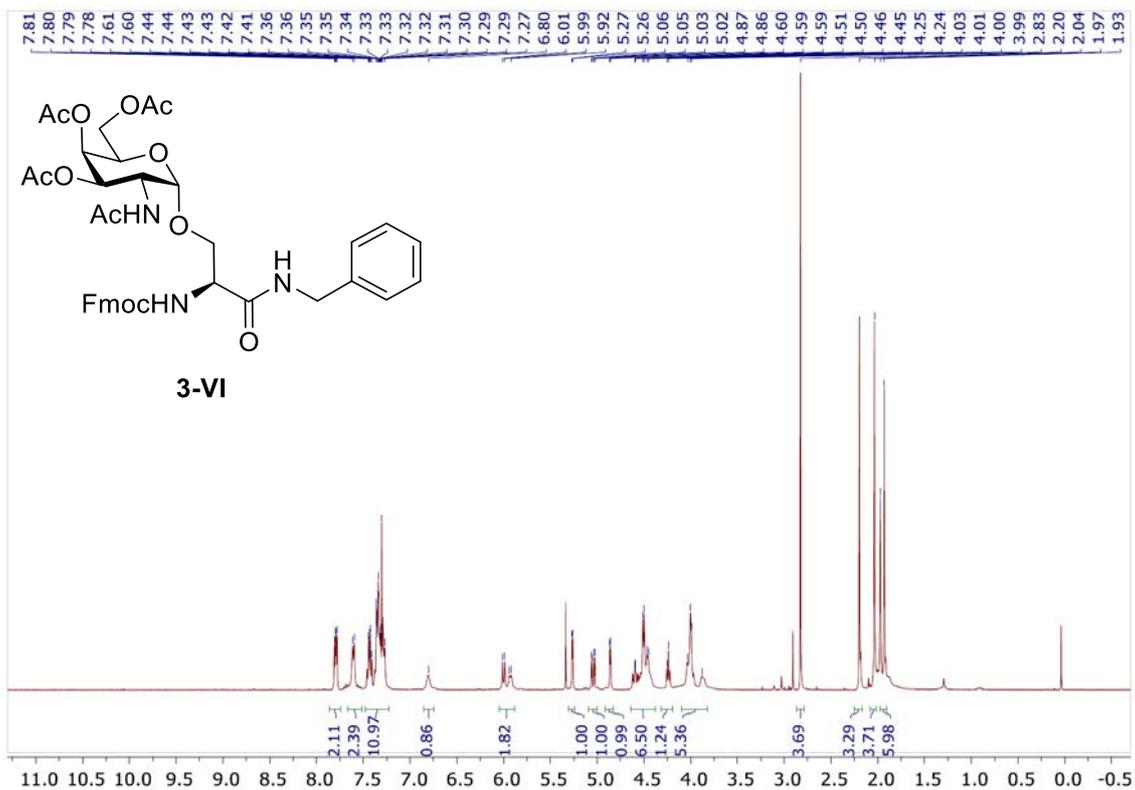
^{13}C NMR 100 MHz in CDCl$_3$ registered at 298K
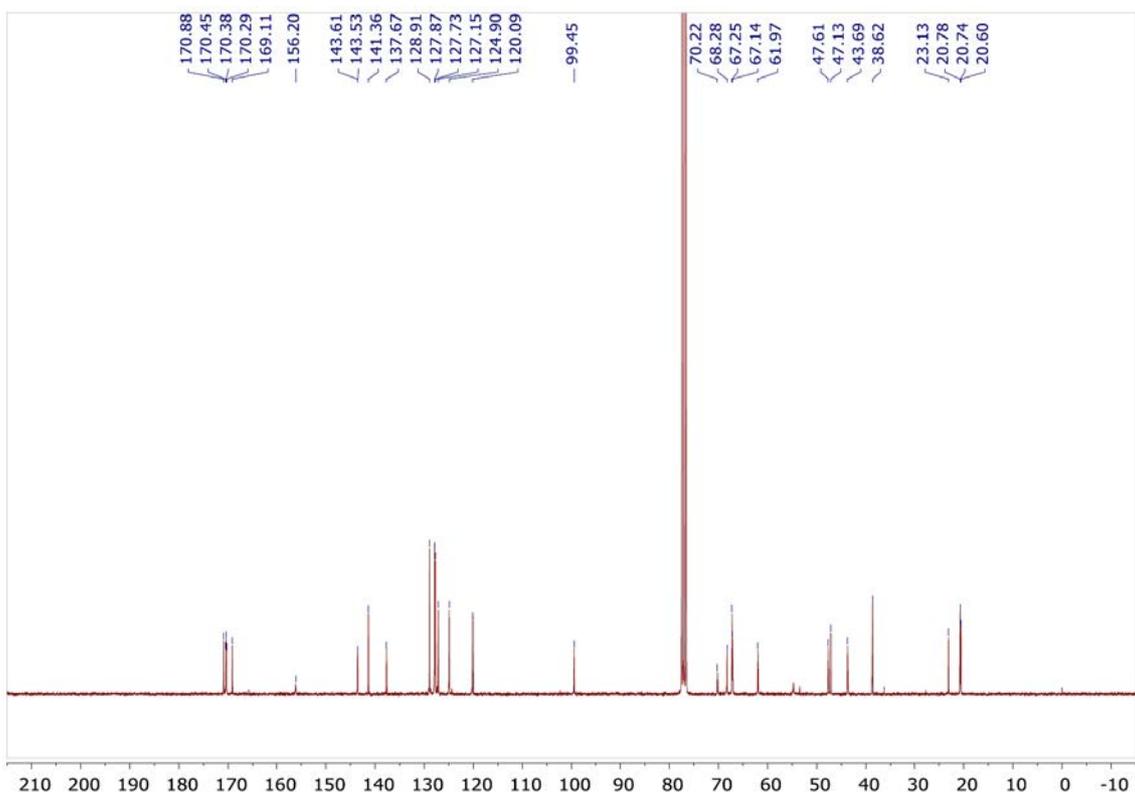
COSY in CDCl₃ registered at 298K
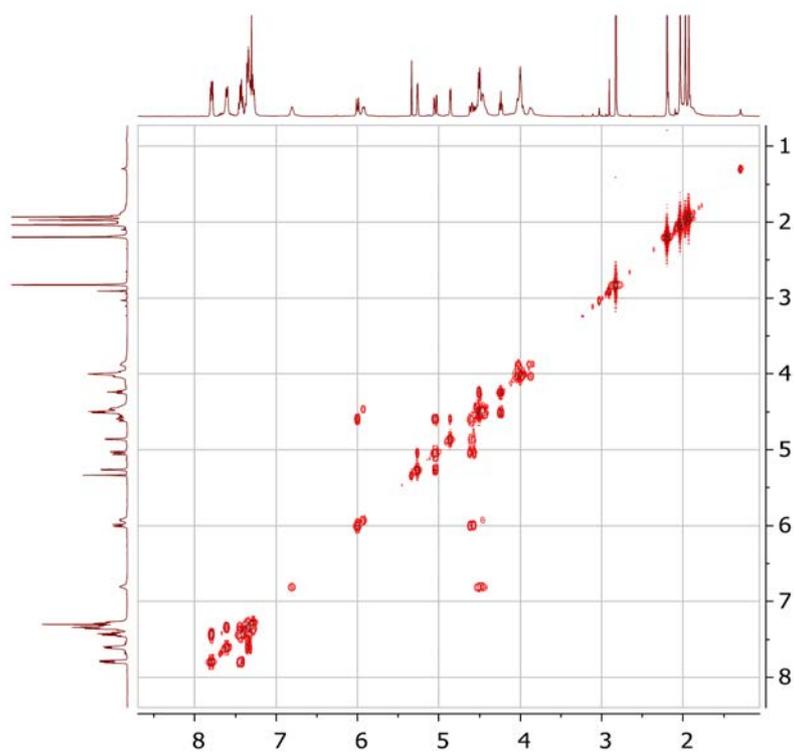
HSQC in CDCl₃ registered at 298K
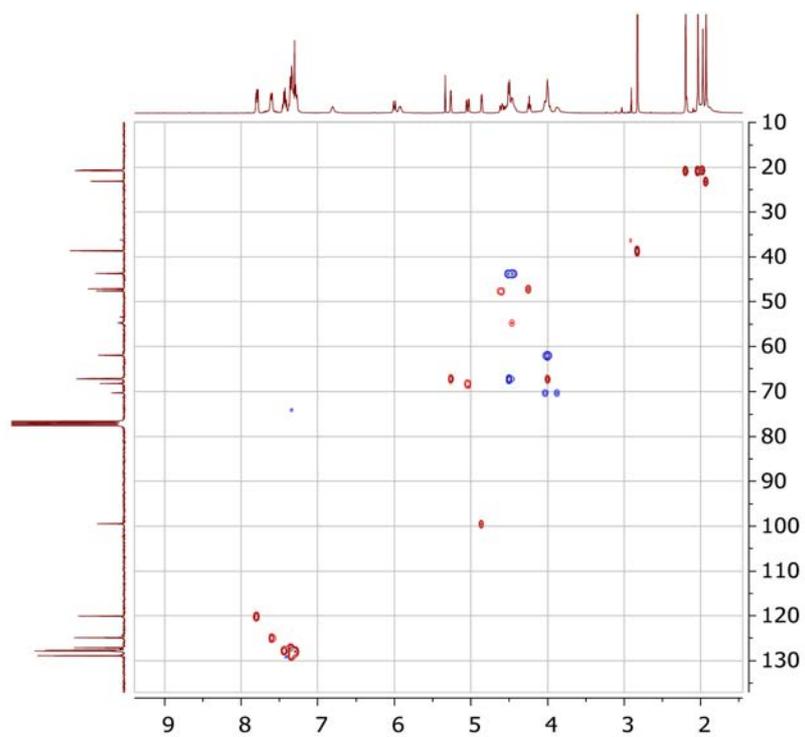
Compound 3-VII

^{1}H NMR 400 MHz in CDCl$_3$ registered at 298K
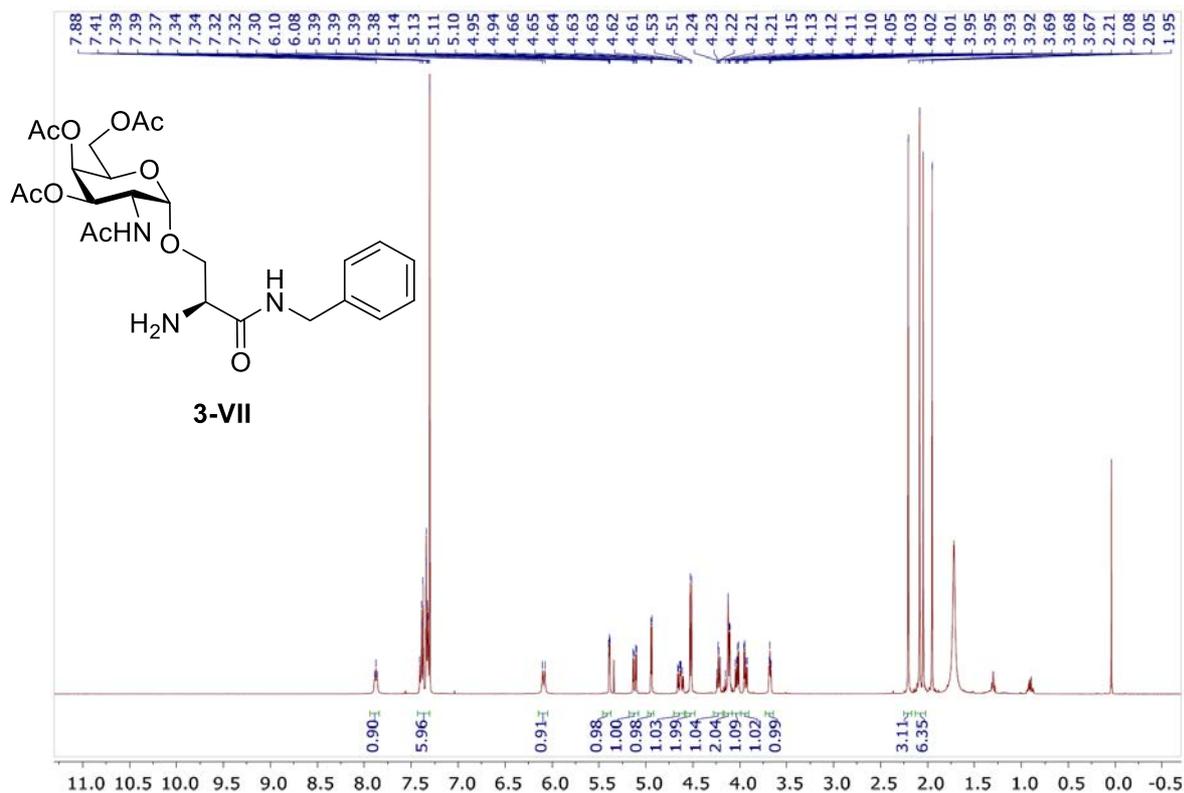
^{13}C NMR 100 MHz in CDCl$_3$ registered at 298K
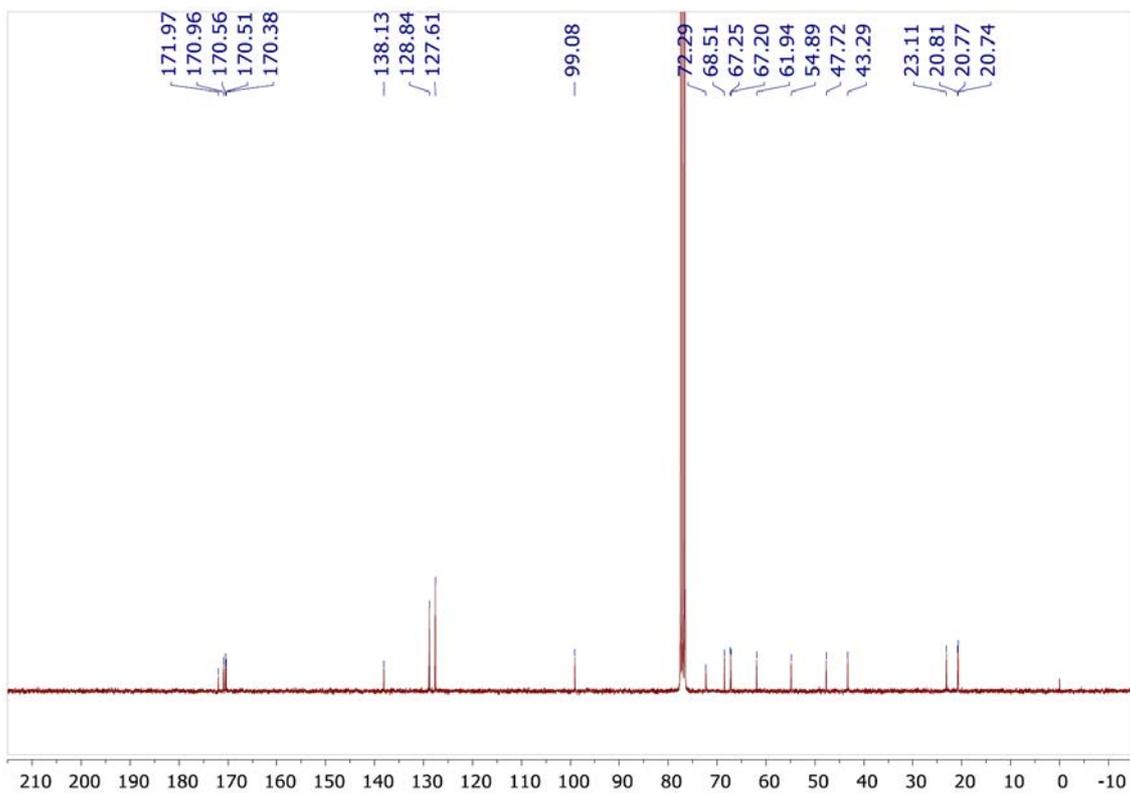
COSY in CDCl₃ registered at 298K
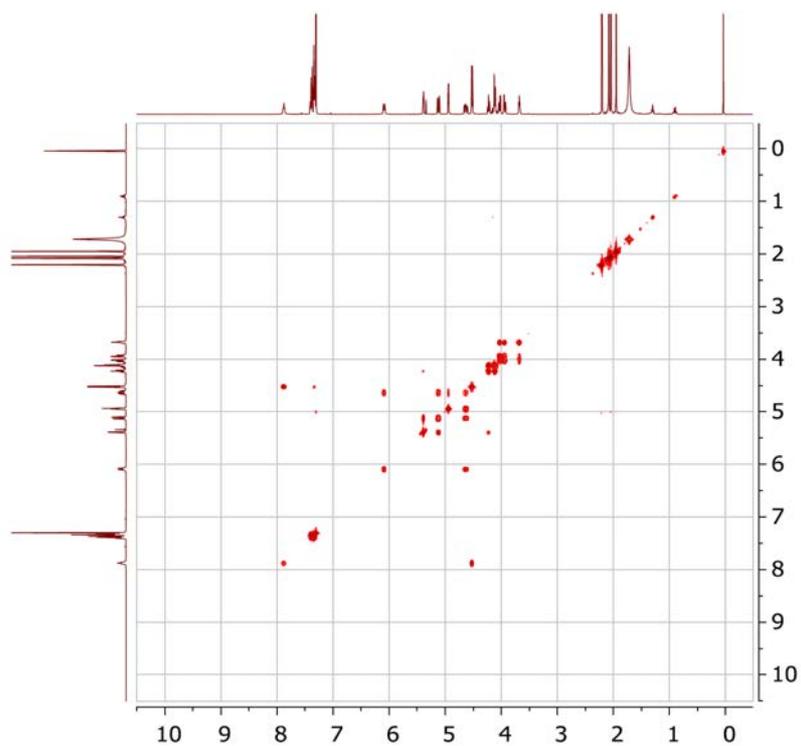
HSQC in CDCl₃ registered at 298K
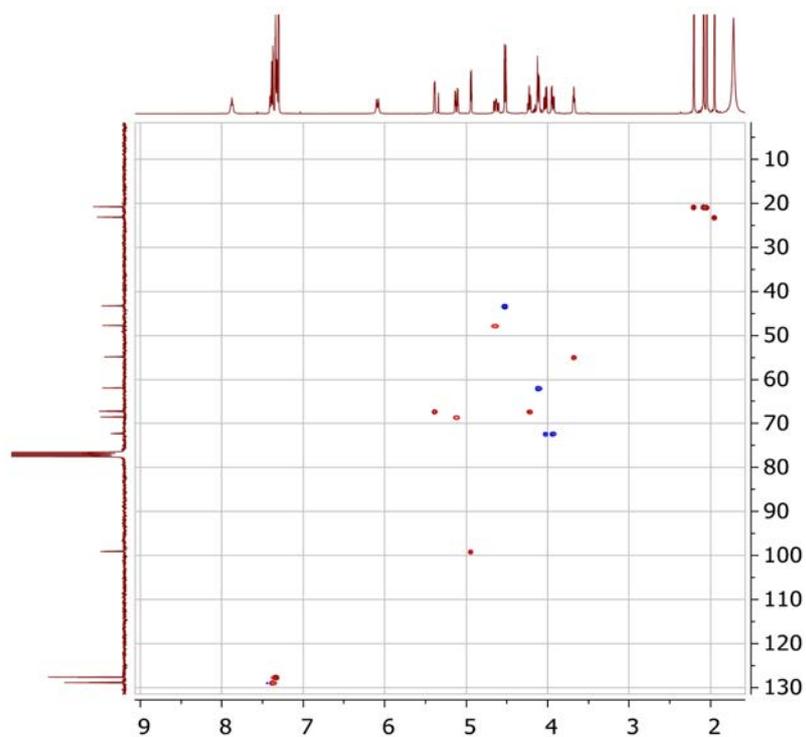
Compound V-III
¹H NMR 400 MHz in CDCl₃ registered at 298K
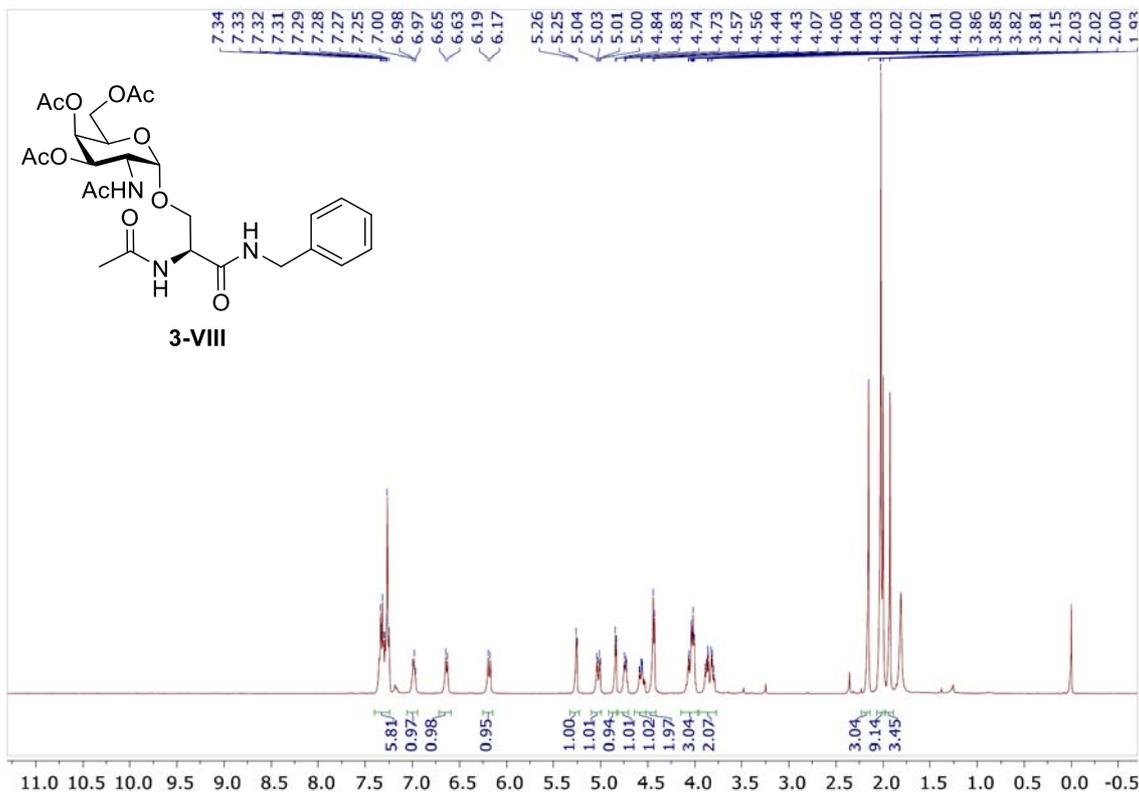
¹³C NMR 100 MHz in CDCl₃ registered at 298K
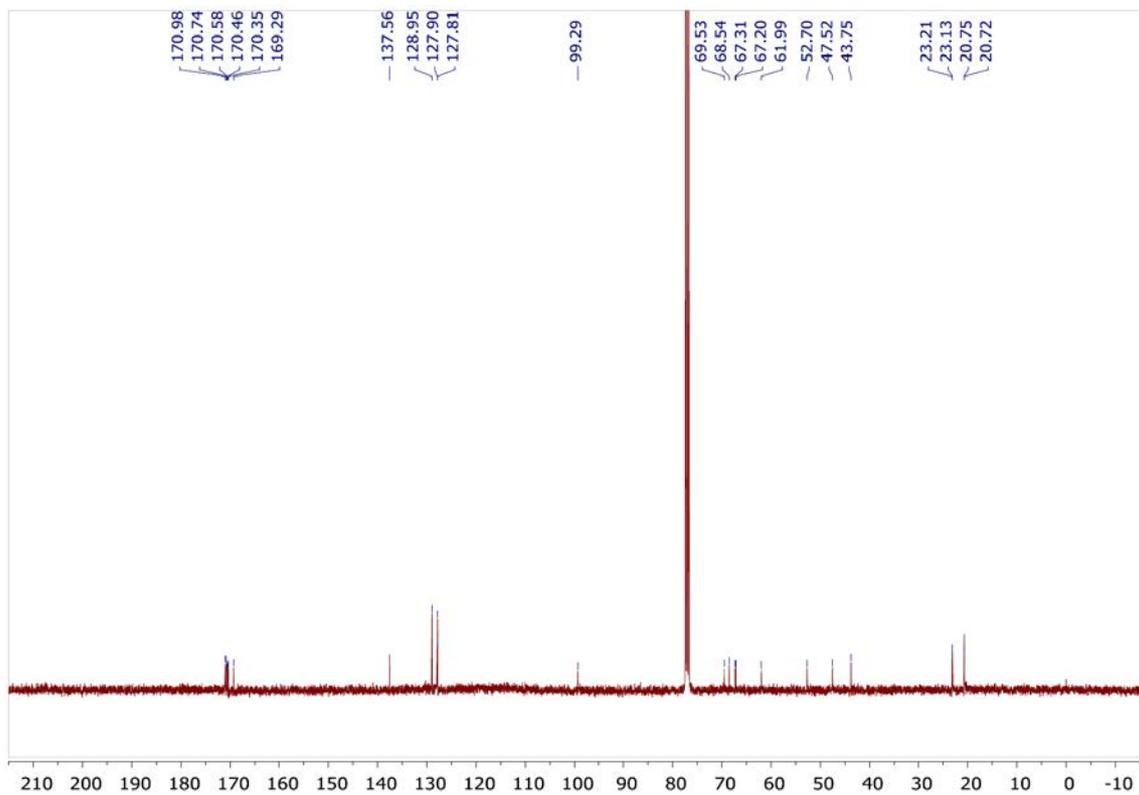
COSY in CDCl₃ registered at 298K
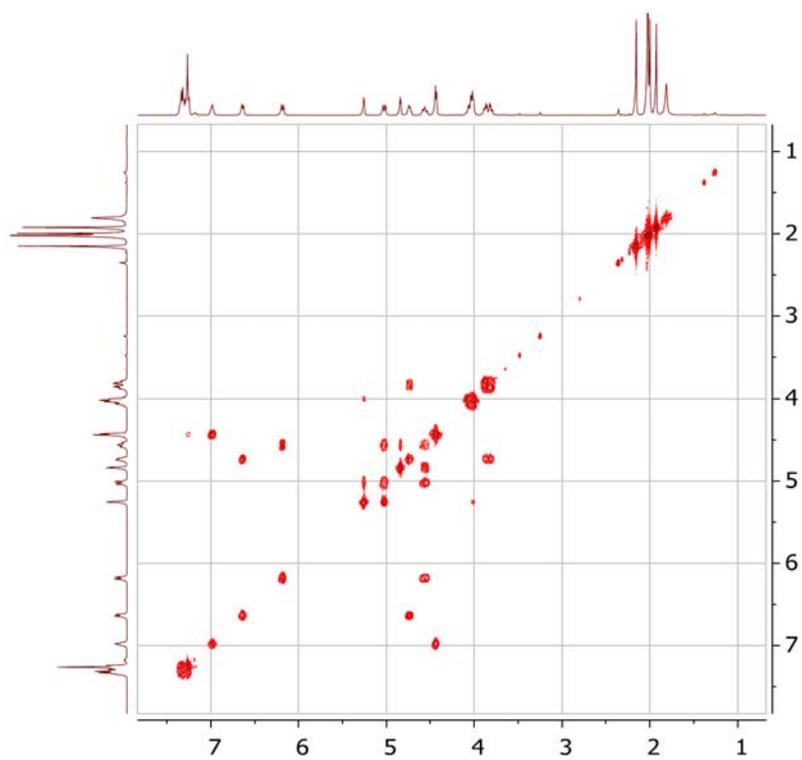
HSQC in CDCl₃ registered at 298K
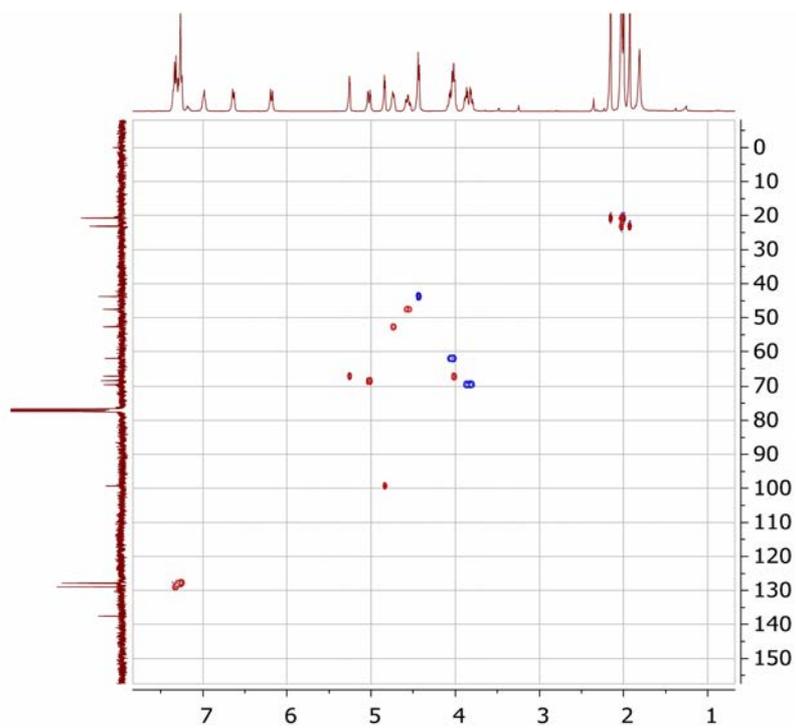
Tn-Ser'

^{1}H NMR 400 MHz in MeOD$_3$ registered at 298K
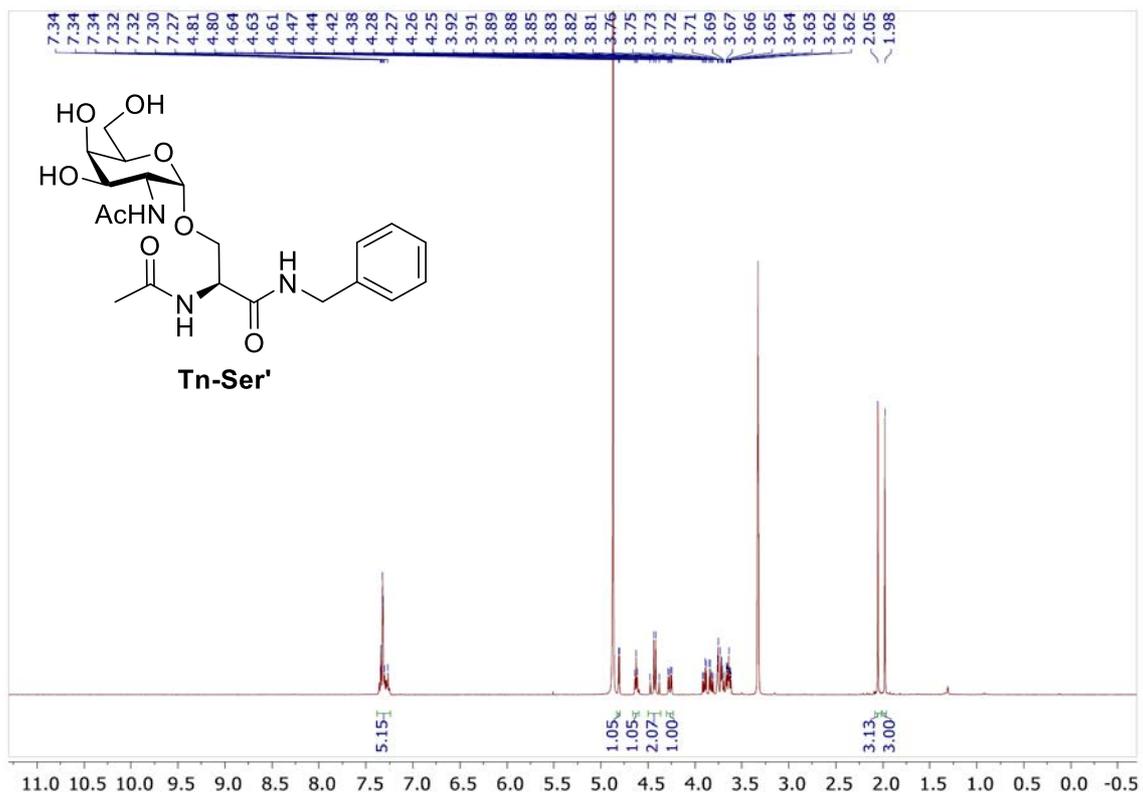
^{13}C NMR 100 MHz in MeOD$_3$ registered at 298K
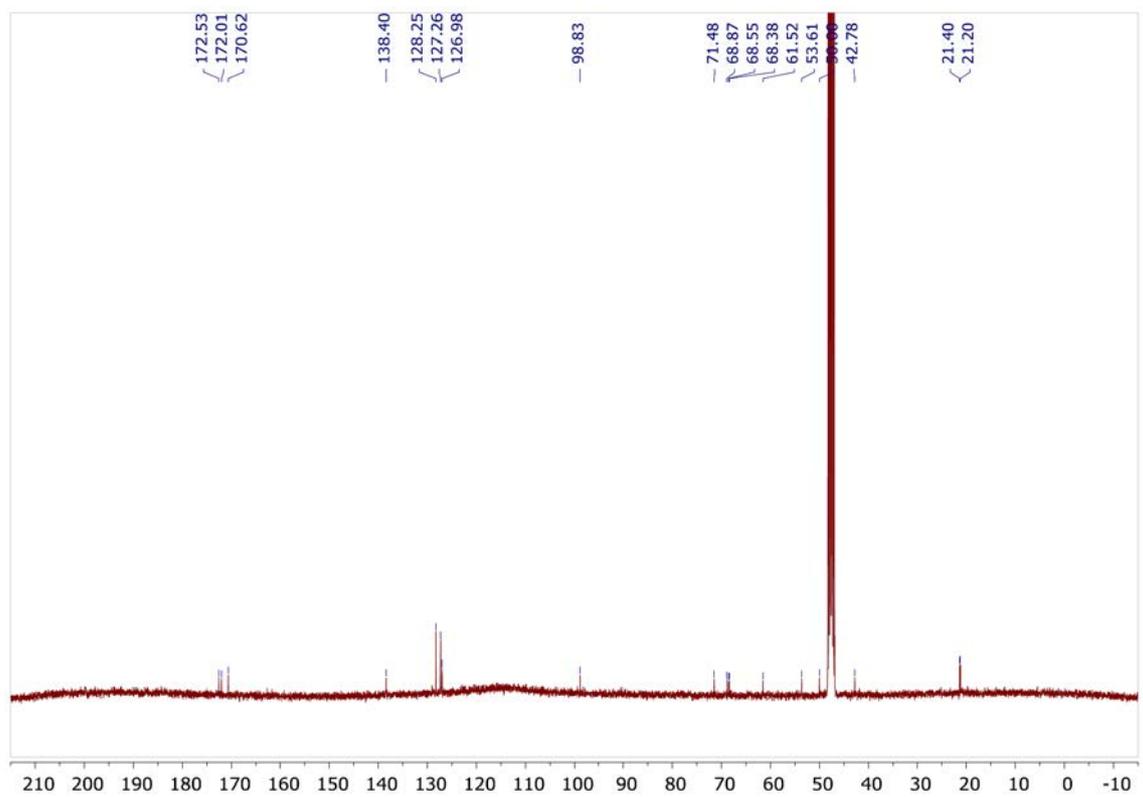
COSY in MeOD₃ registered at 298K
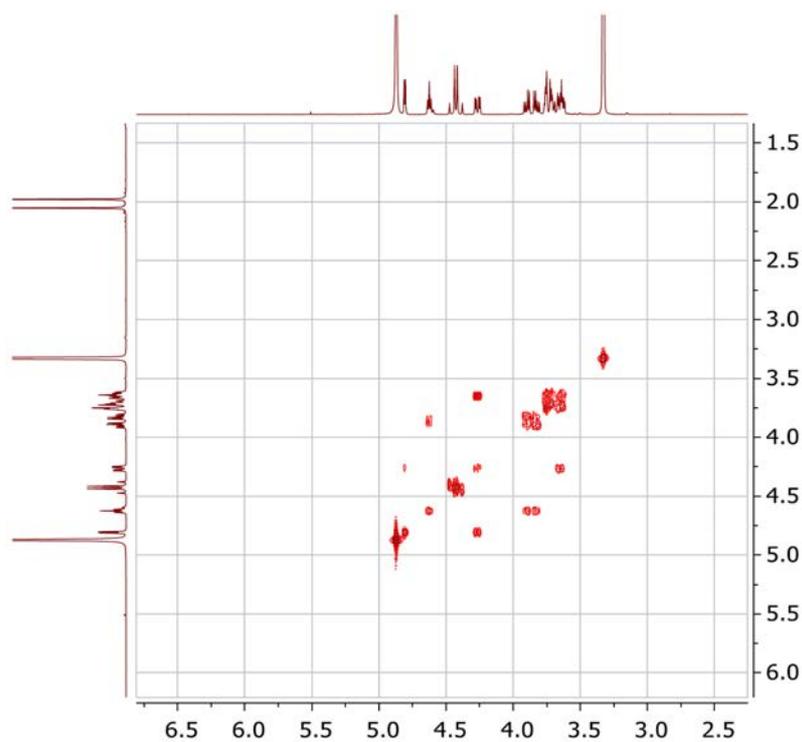
HSQC in MeOD₃ registered at 298K
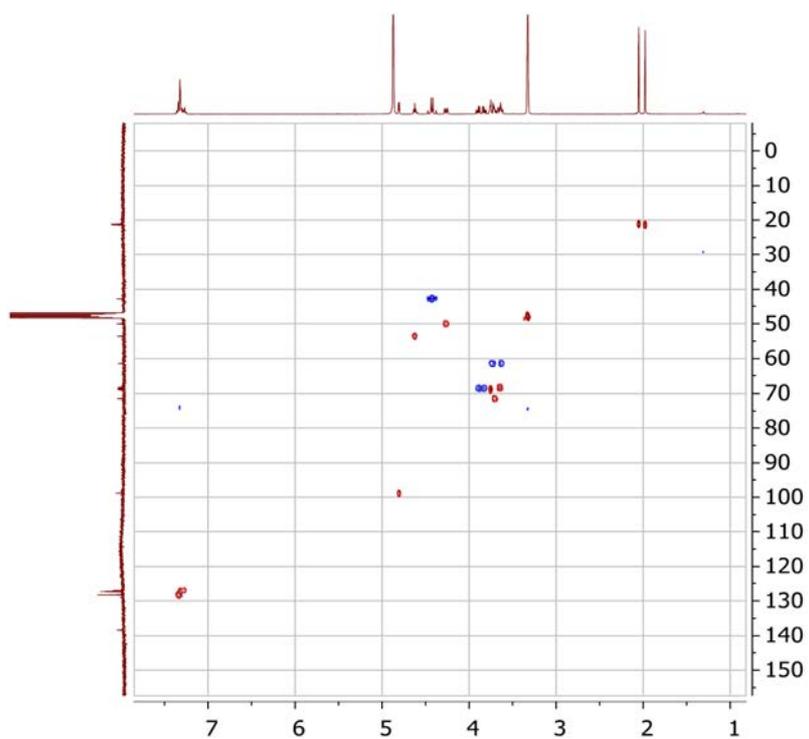
9.2. **NMR experiments and chromatograms chapter 4**

Compound 4-III

^{1}H NMR 400 MHz in CDCl$_3$ registered at 298K
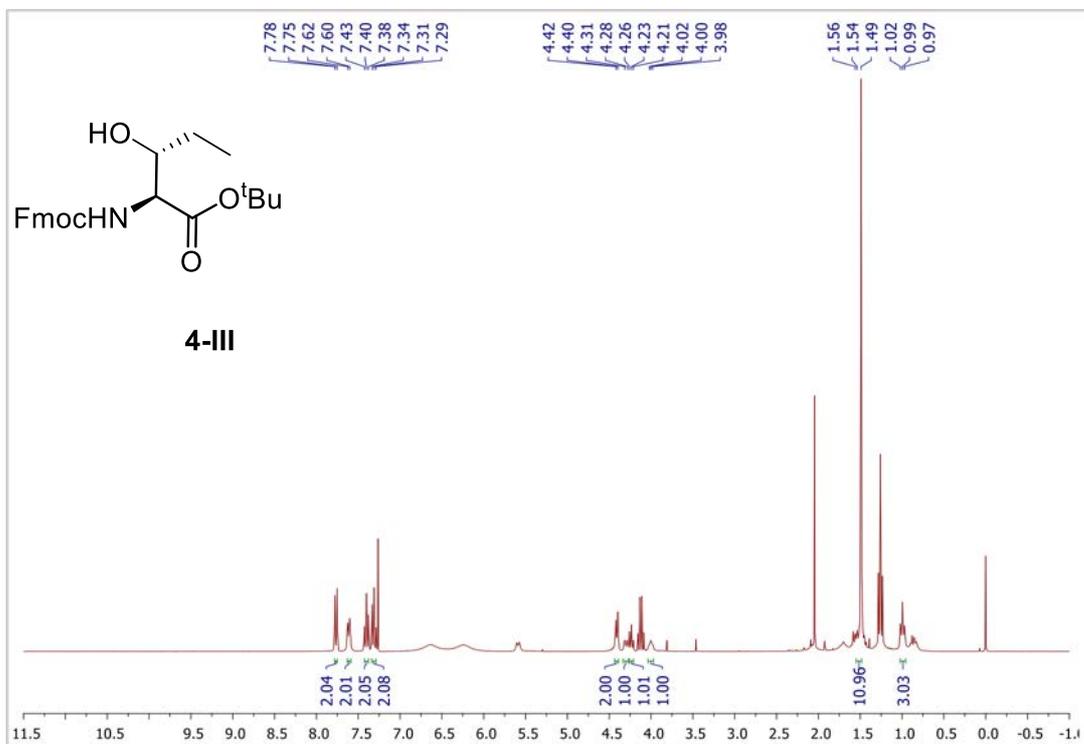
^{13}C NMR 100 MHz in CDCl$_3$ registered at 298K
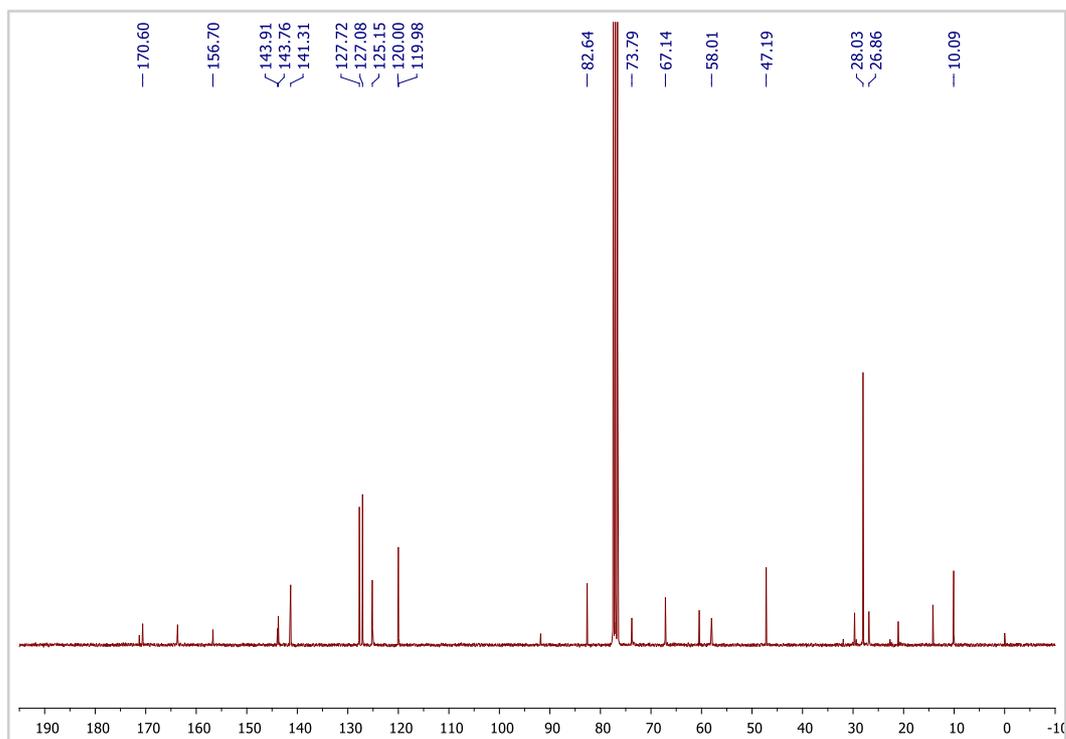
COSY in CDCl₃ registered at 298K
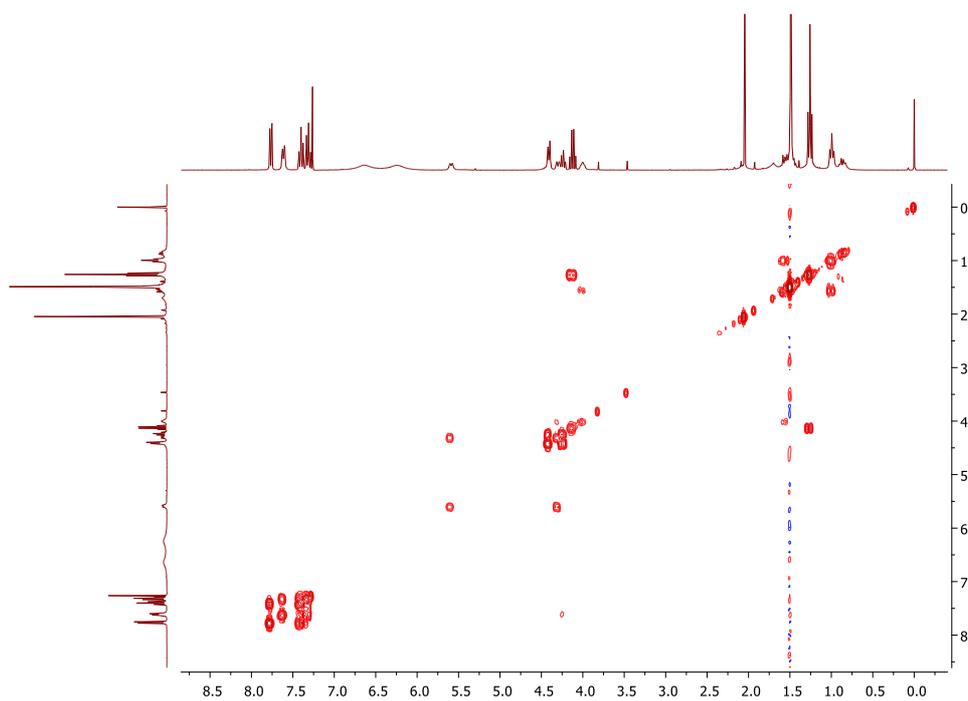
HSQC in CDCl₃ registered at 298K
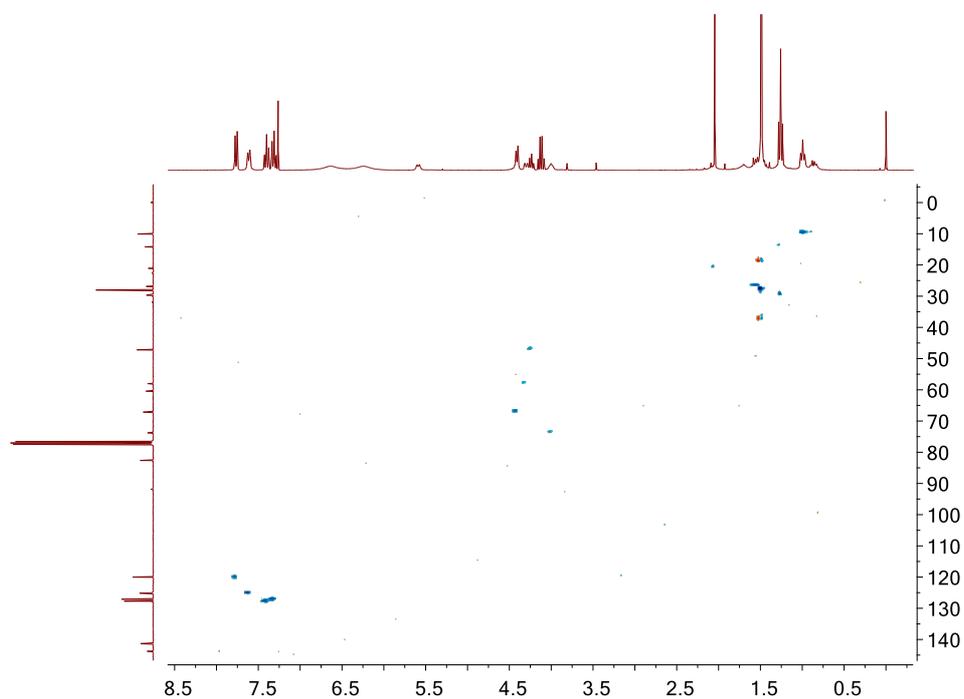
Compound 4-VI

^{1}H NMR 400 MHz in CDCl$_3$ registered at 298K
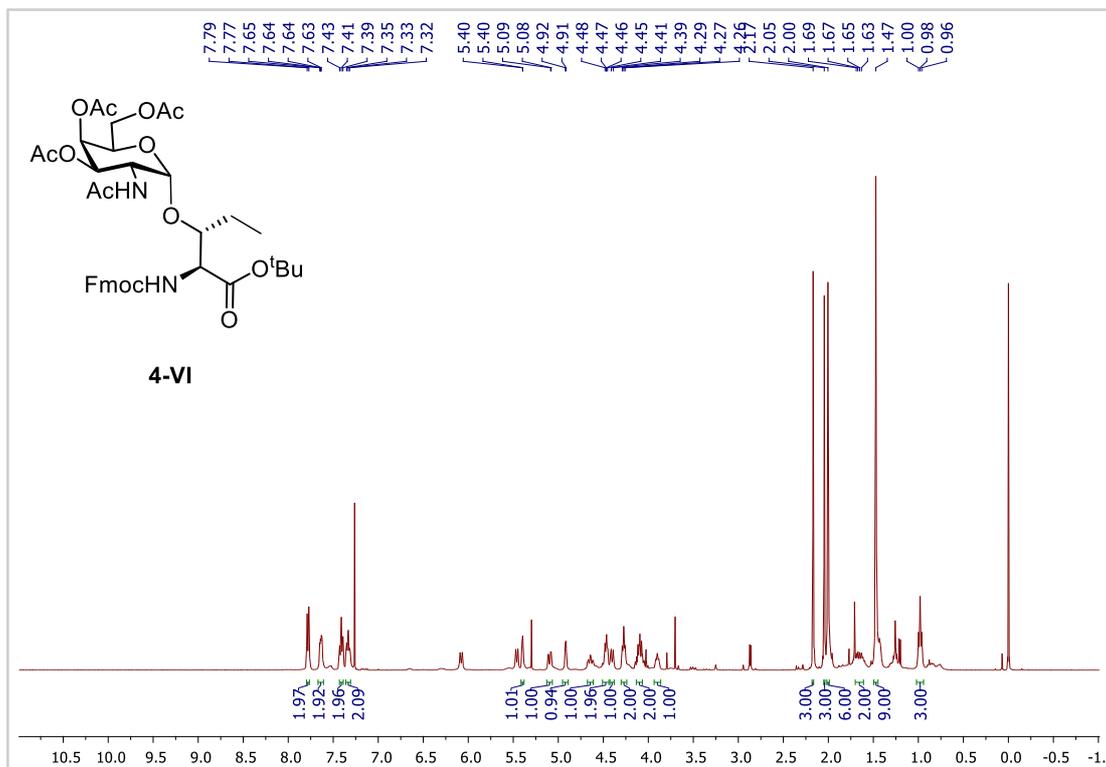
^{13}C NMR 100 MHz in CDCl$_3$ registered at 298K
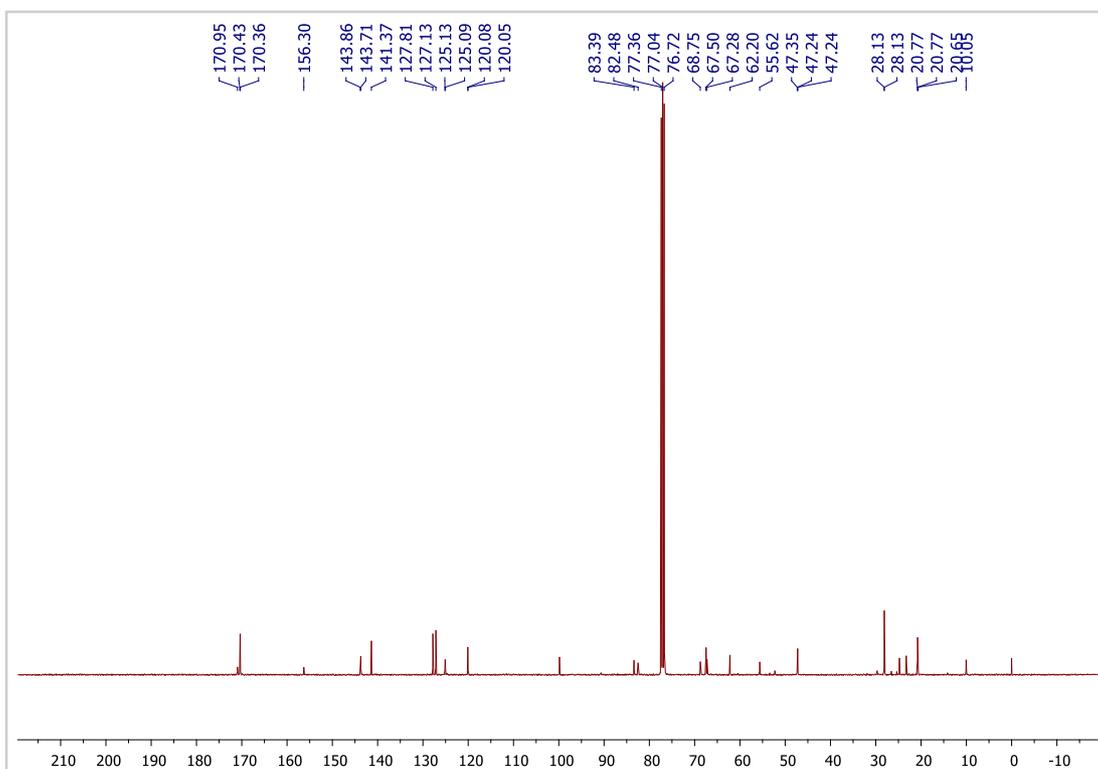
COSY in CDCl₃ registered at 298K
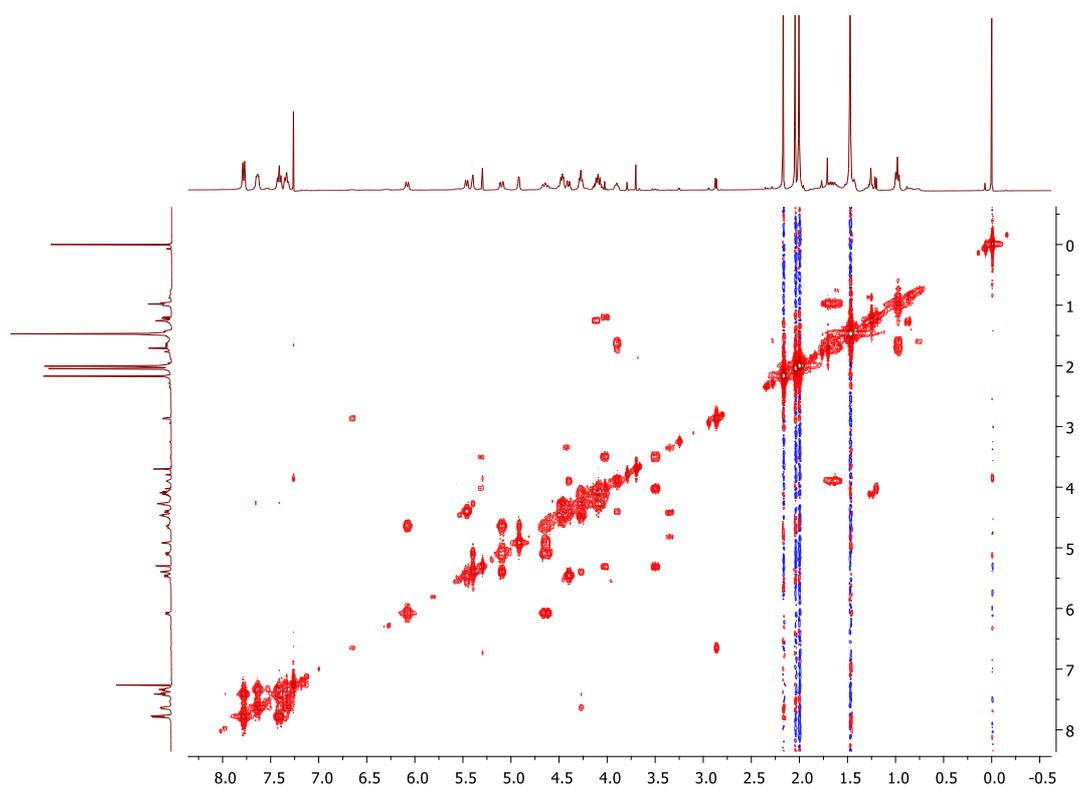
HSQC in CDCl₃ registered at 298K
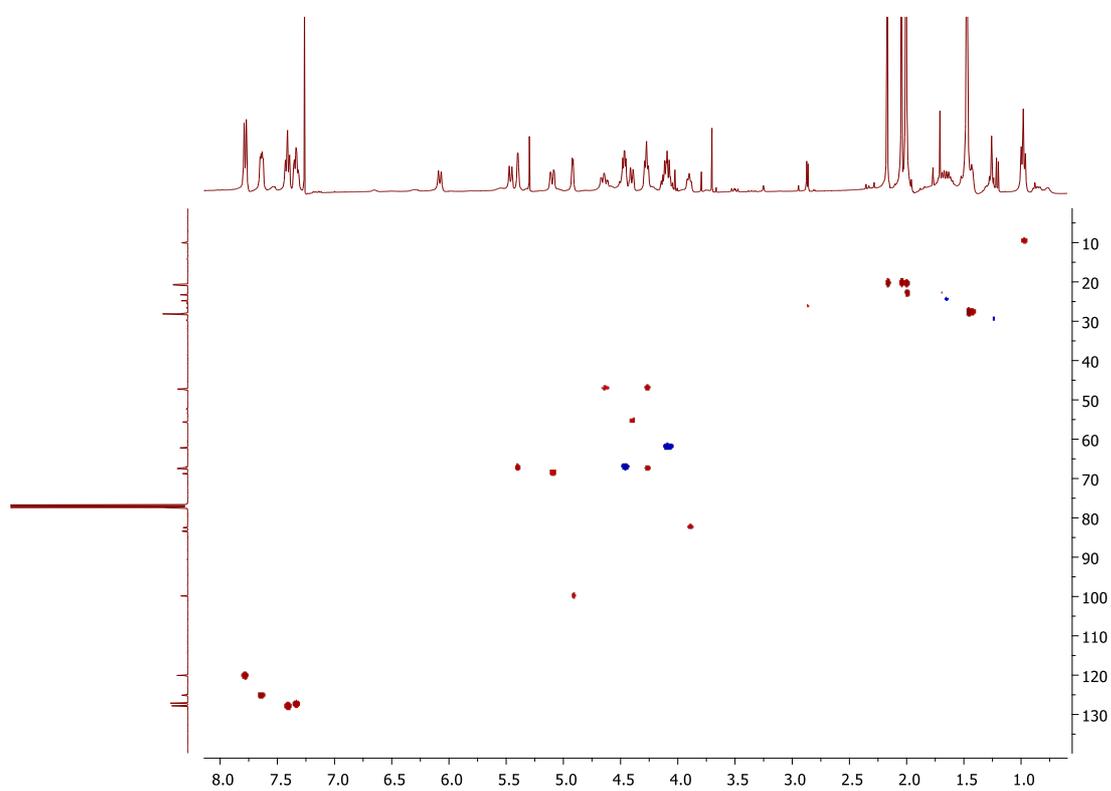
Compound 4-IX

^{1}H NMR 400 MHz in CDCl$_3$ registered at 298K
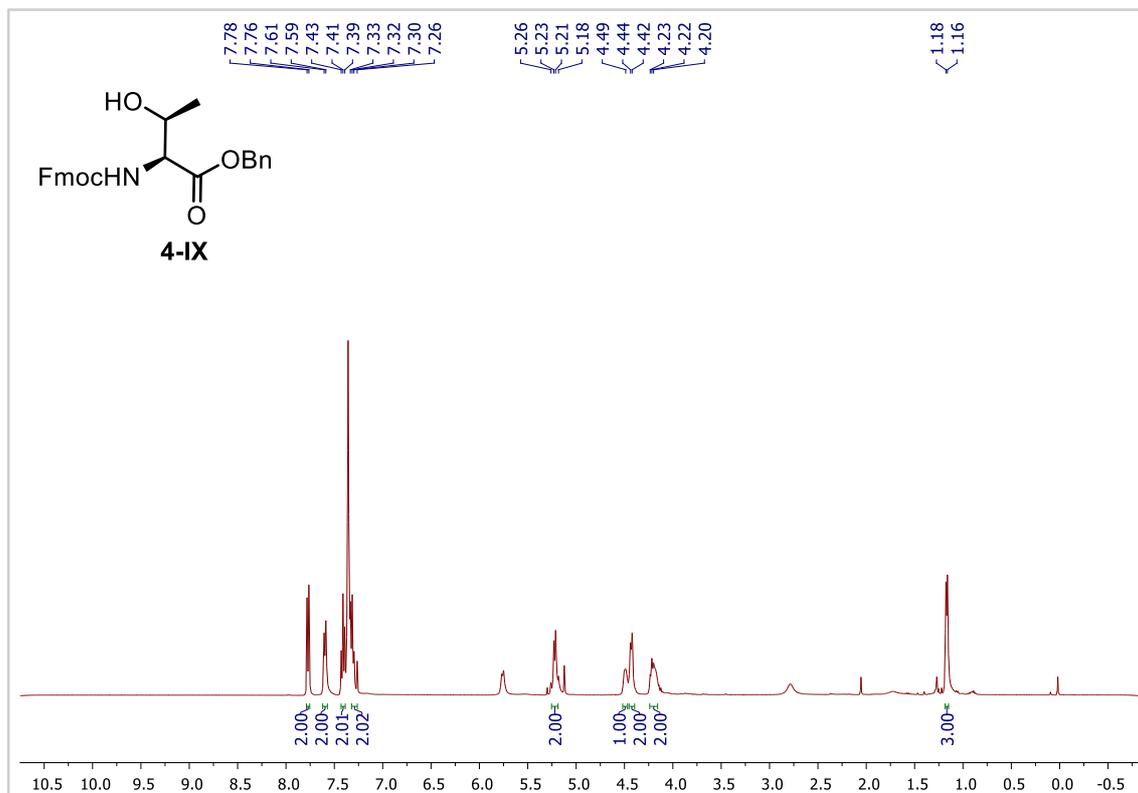
^{13}C NMR 100 MHz in CDCl$_3$ registered at 298K
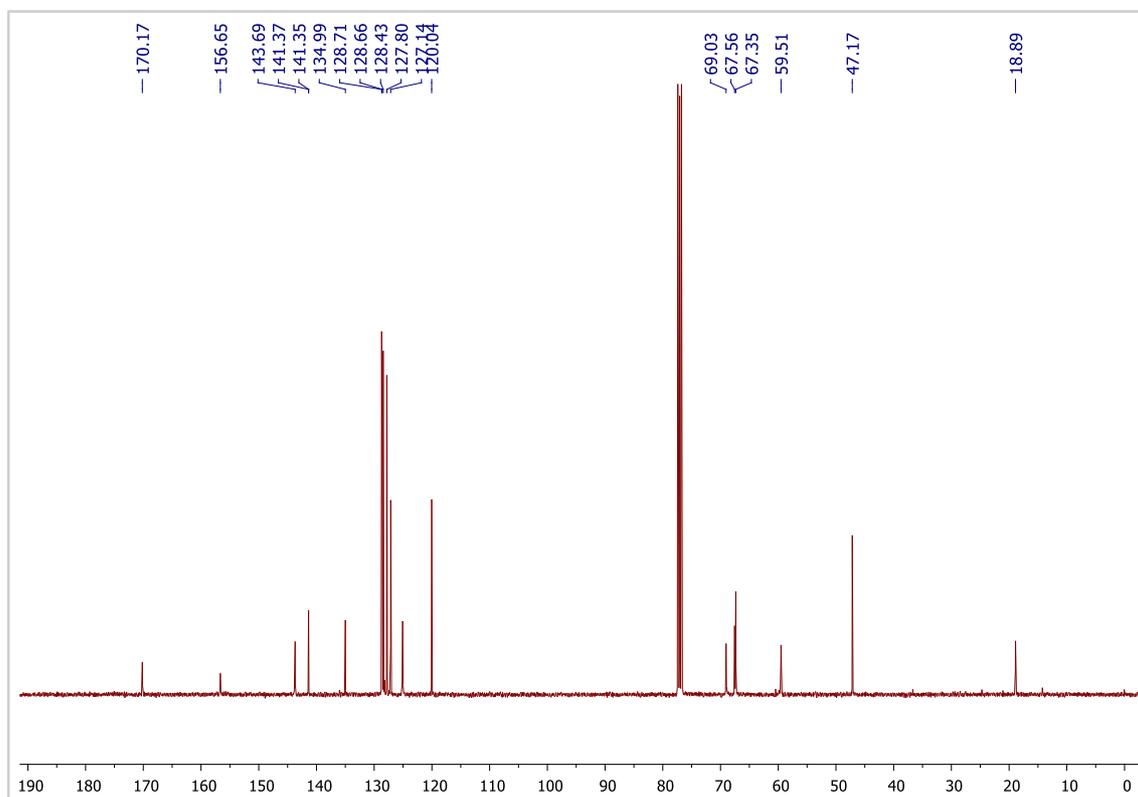
COSY in CDCl₃ registered at 298K
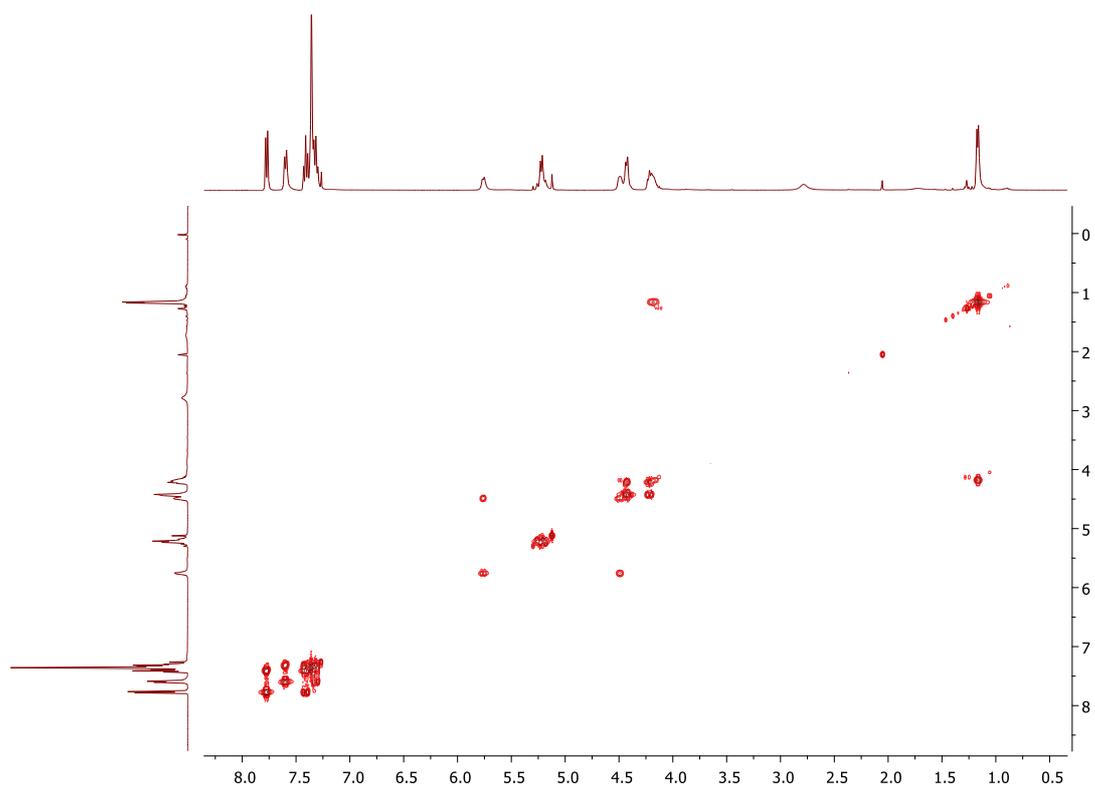
HSQC in CDCl₃ registered at 298K
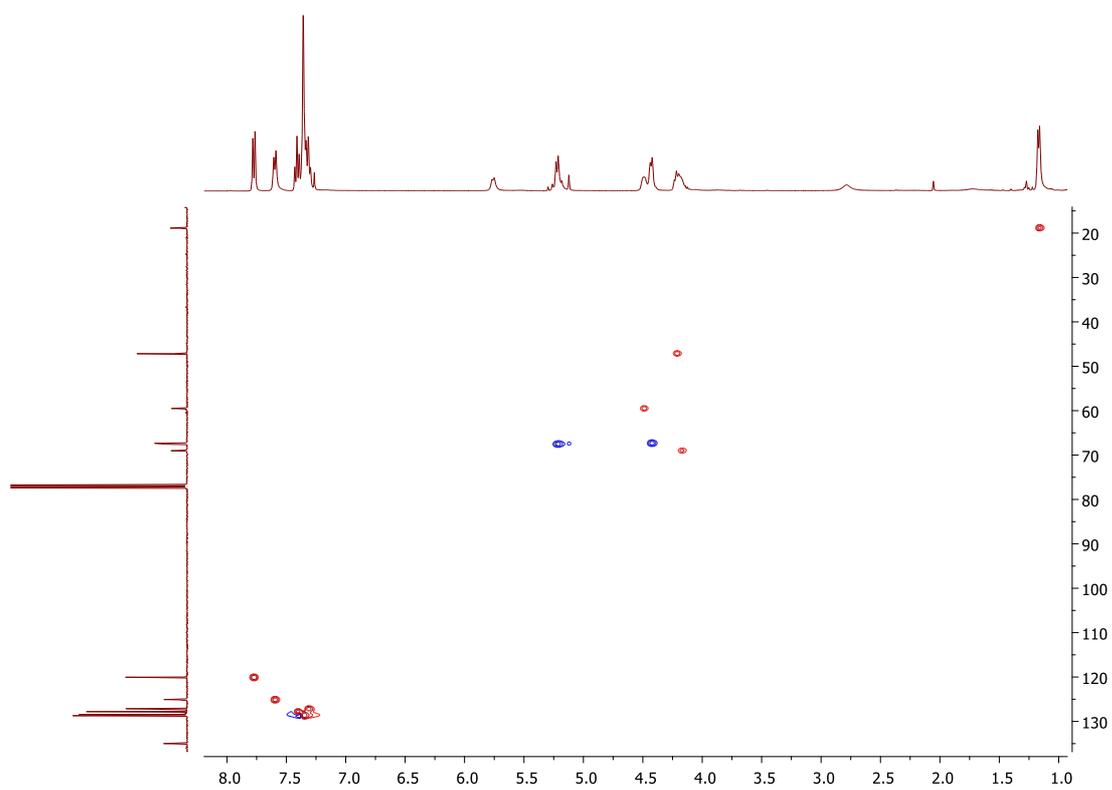
Compound 4-XI

^{1}H NMR 400 MHz in CDCl$_3$ registered at 298K
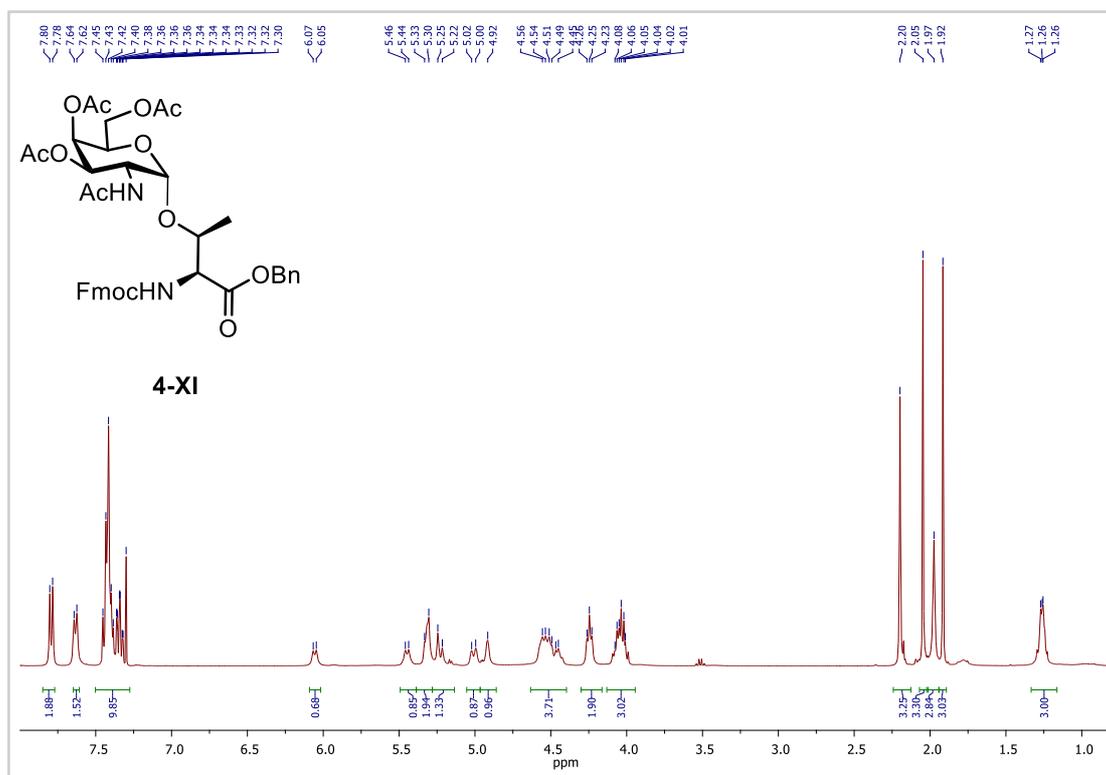
^{13}C NMR 100 MHz in CDCl$_3$ registered at 298K
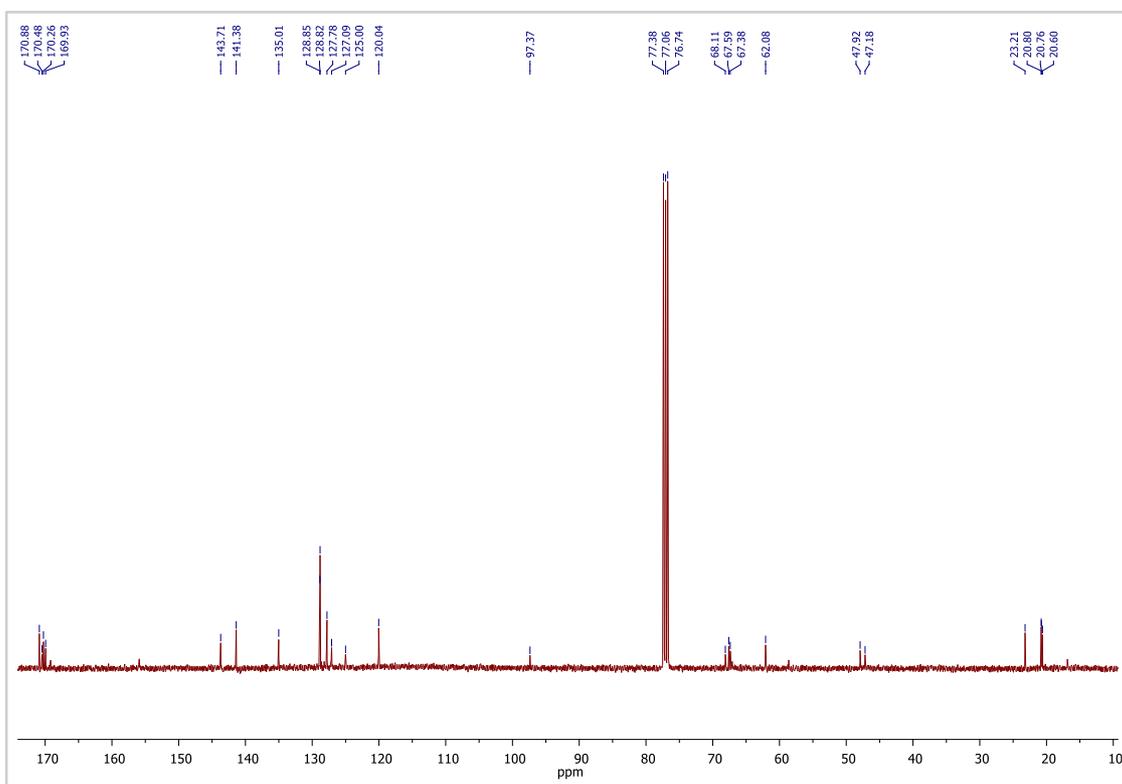
COSY in CDCl₃ registered at 298K
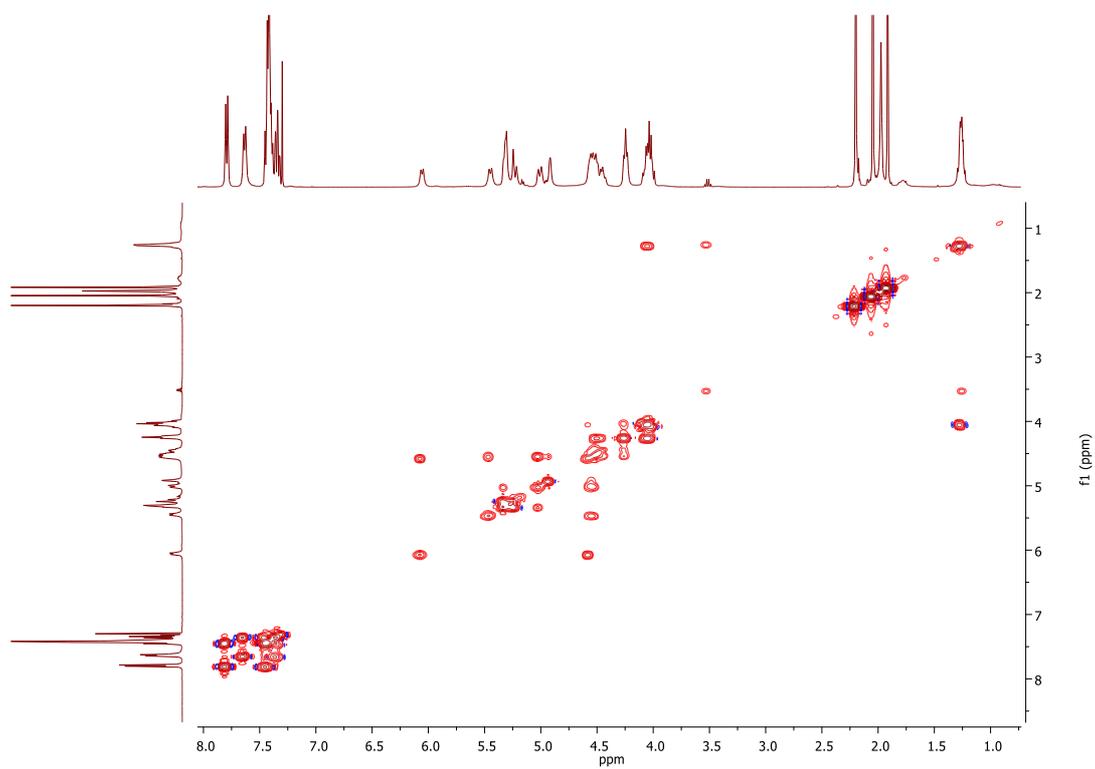
HSQC in CDCl₃ registered at 298K
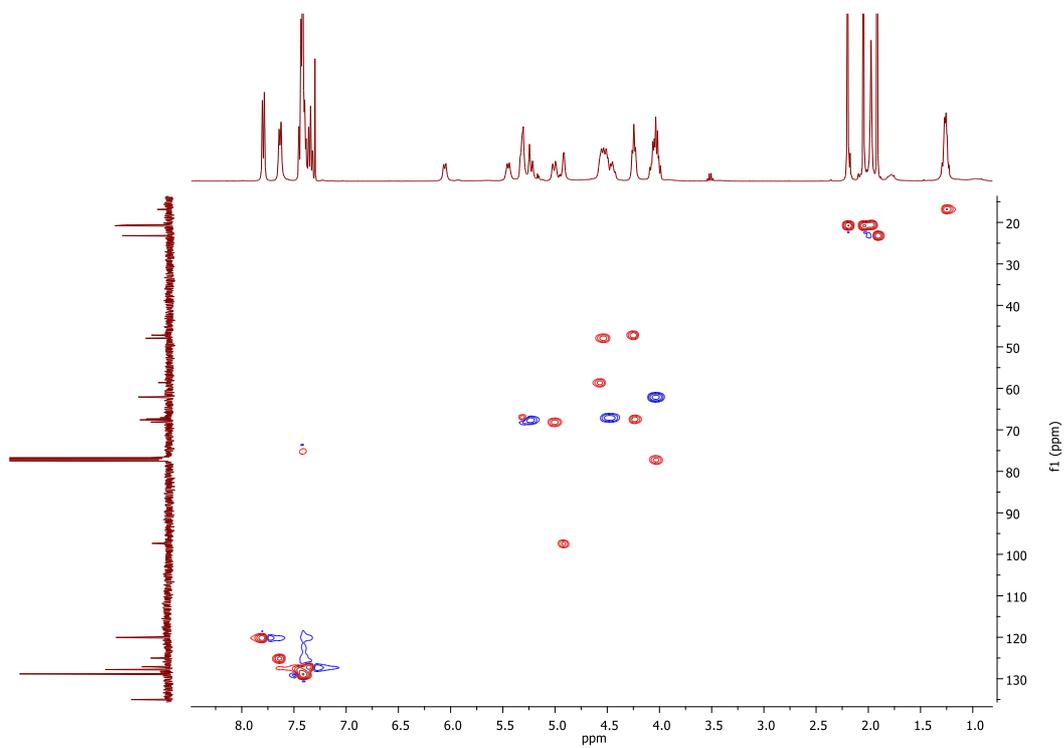
Glycopeptide Hnv*'
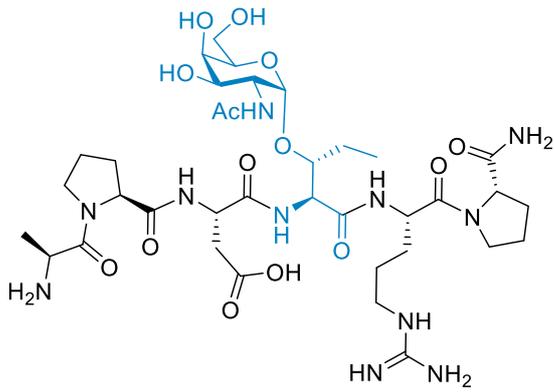
HPLC chromatogram
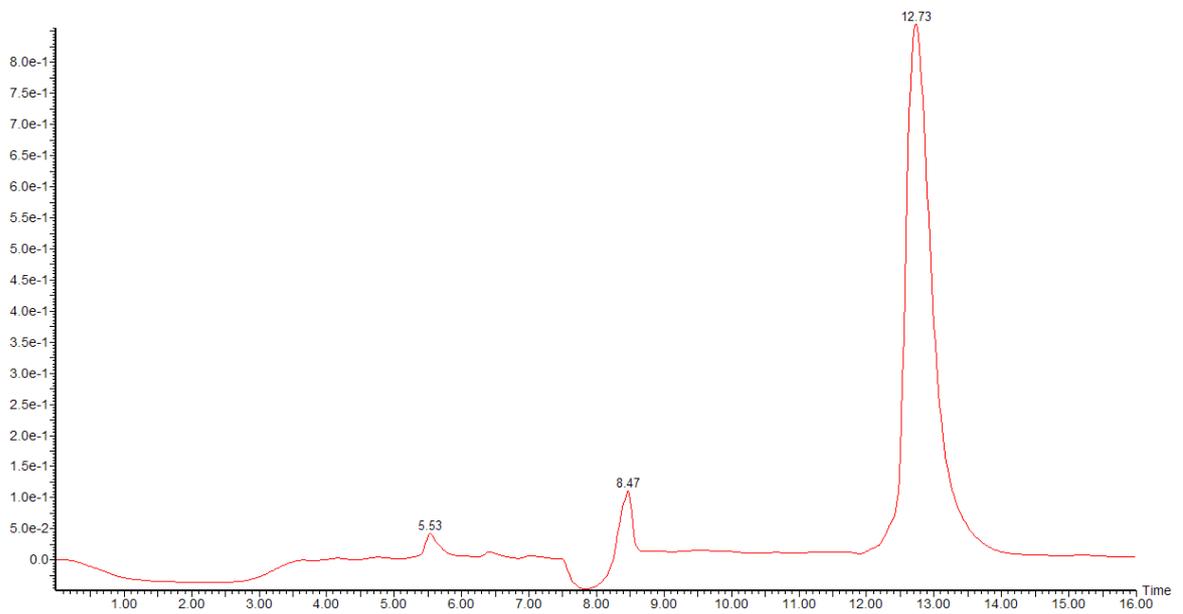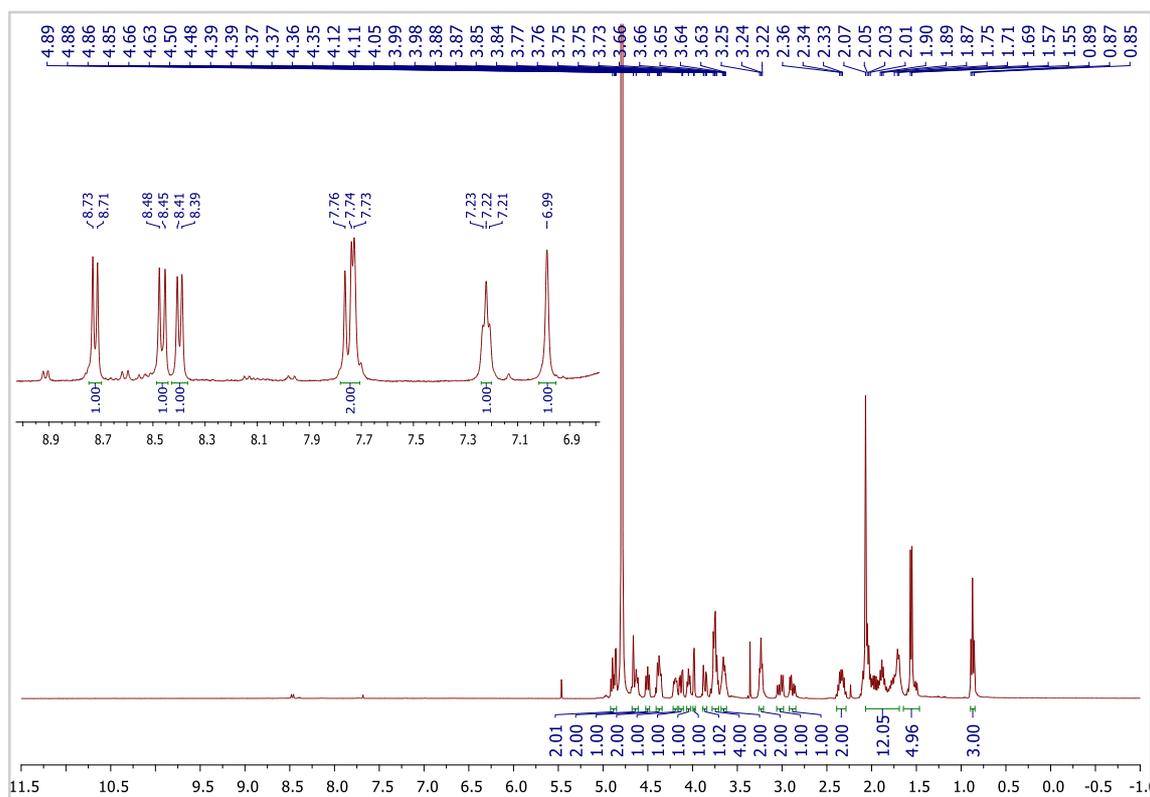
^{1}H NMR 400 MHz in D$_2$O registered at 298K

^{13}C NMR 75 MHz in D$_2$O registered at 298K

COSY in D₂O registered at 298K

HSQC in D₂O registered at 298K

ROESY in D₂O registered at 298K
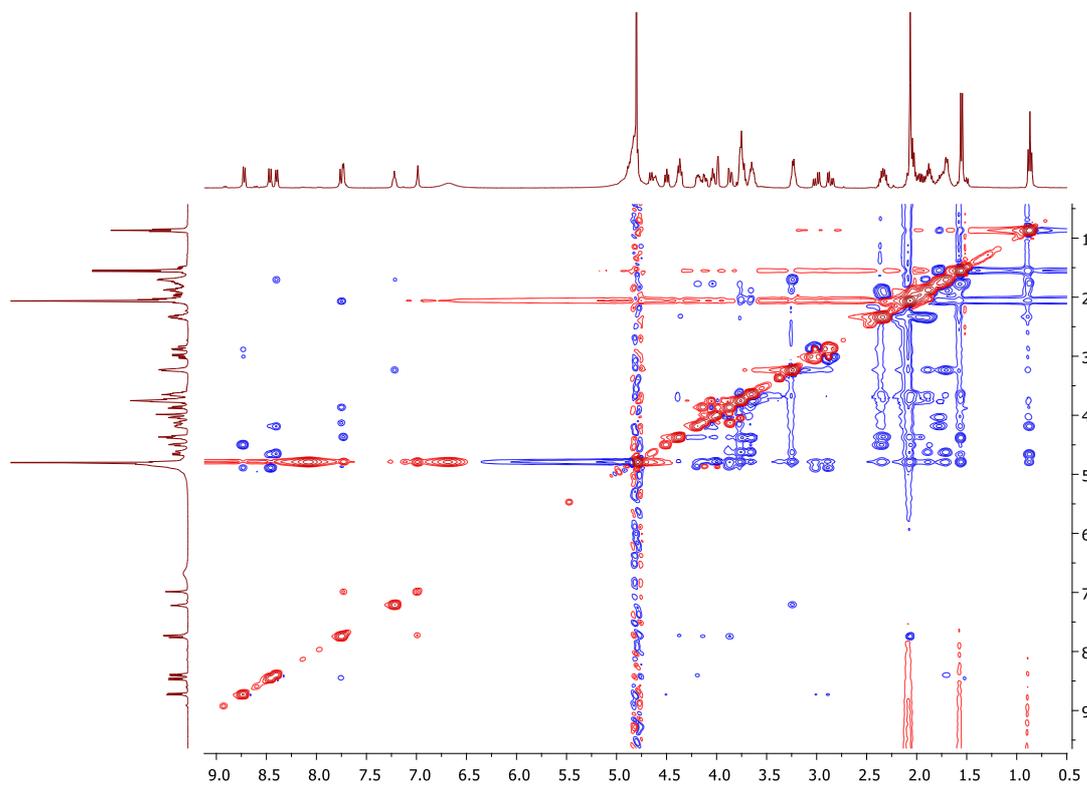
Glycopeptide *allo*Thr*'
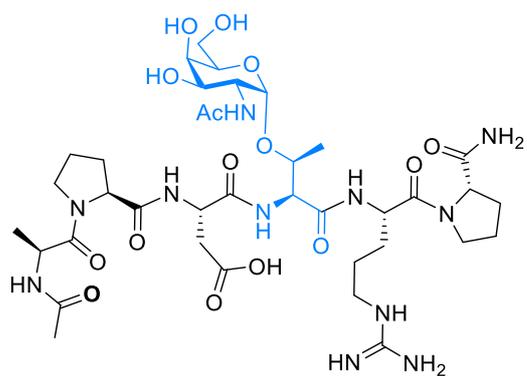
HPLC chromatogram
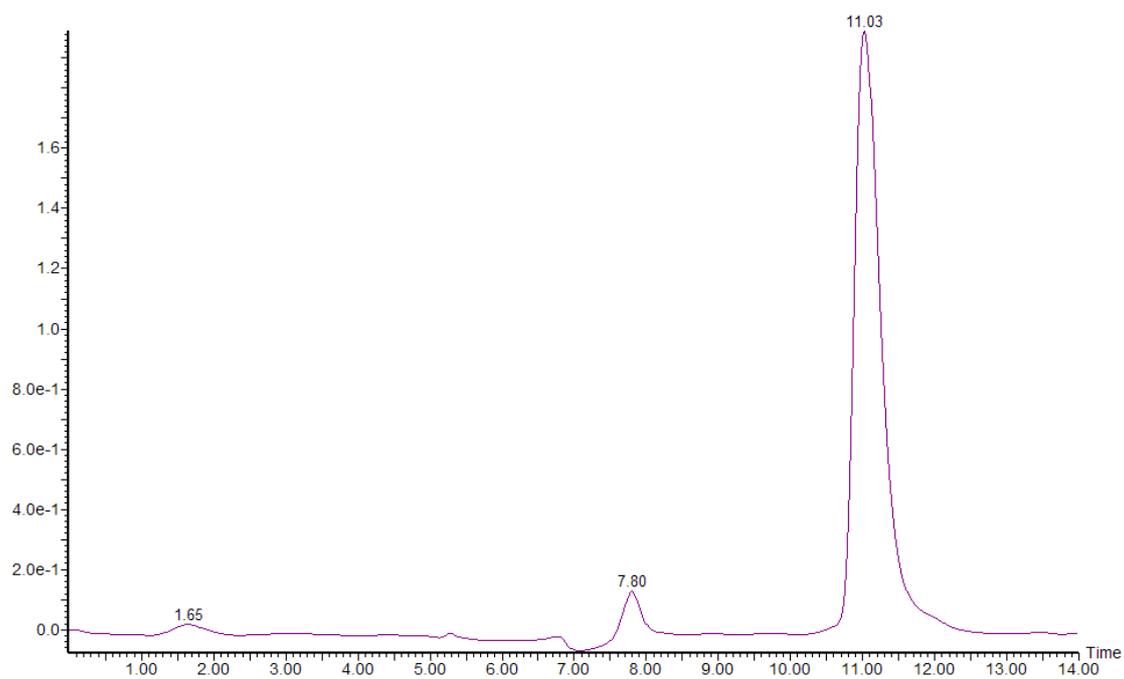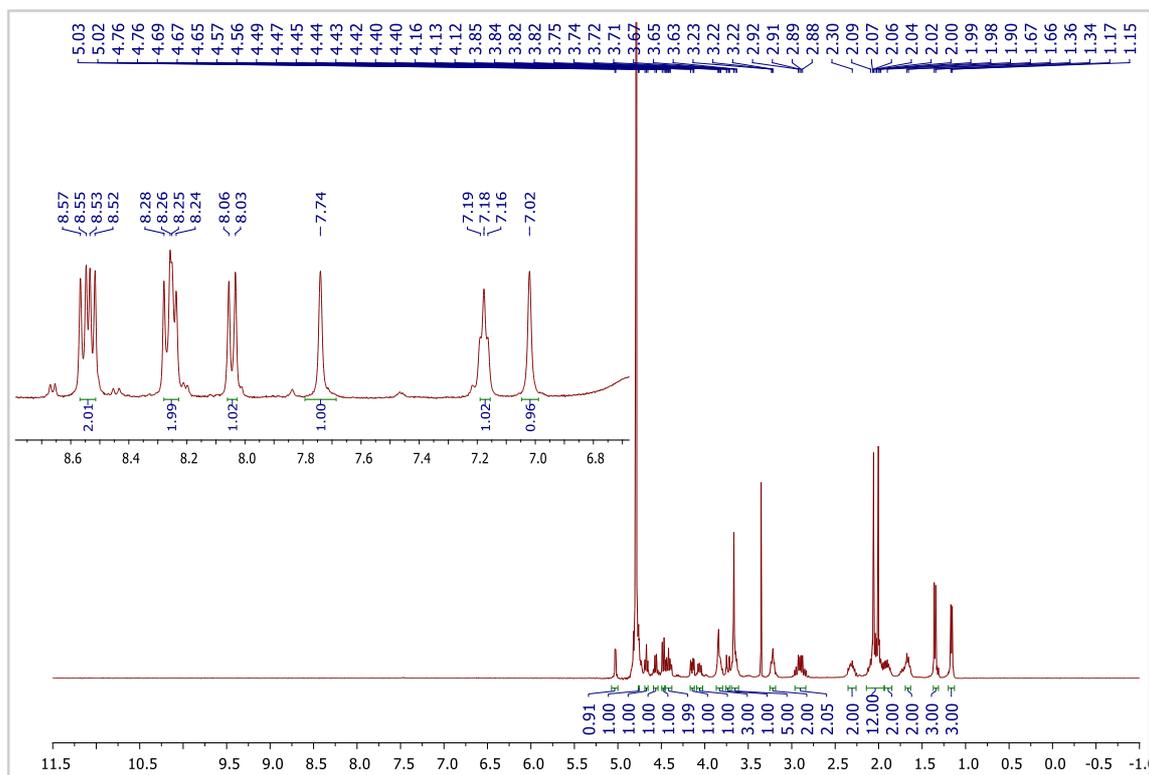
^{1}H NMR 400 MHz in D$_2$O registered at 298K

^{13}C NMR 75 MHz in D$_2$O registered at 298K
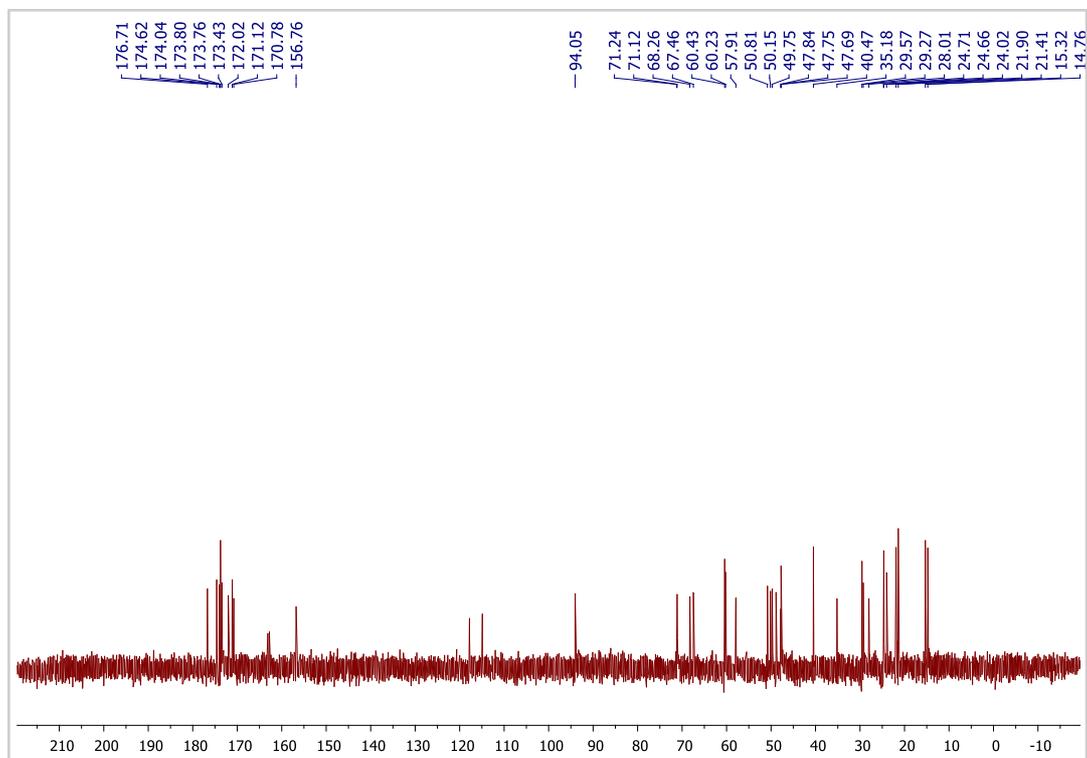
COSY in D₂O registered at 298K

HSQC in D₂O registered at 298K
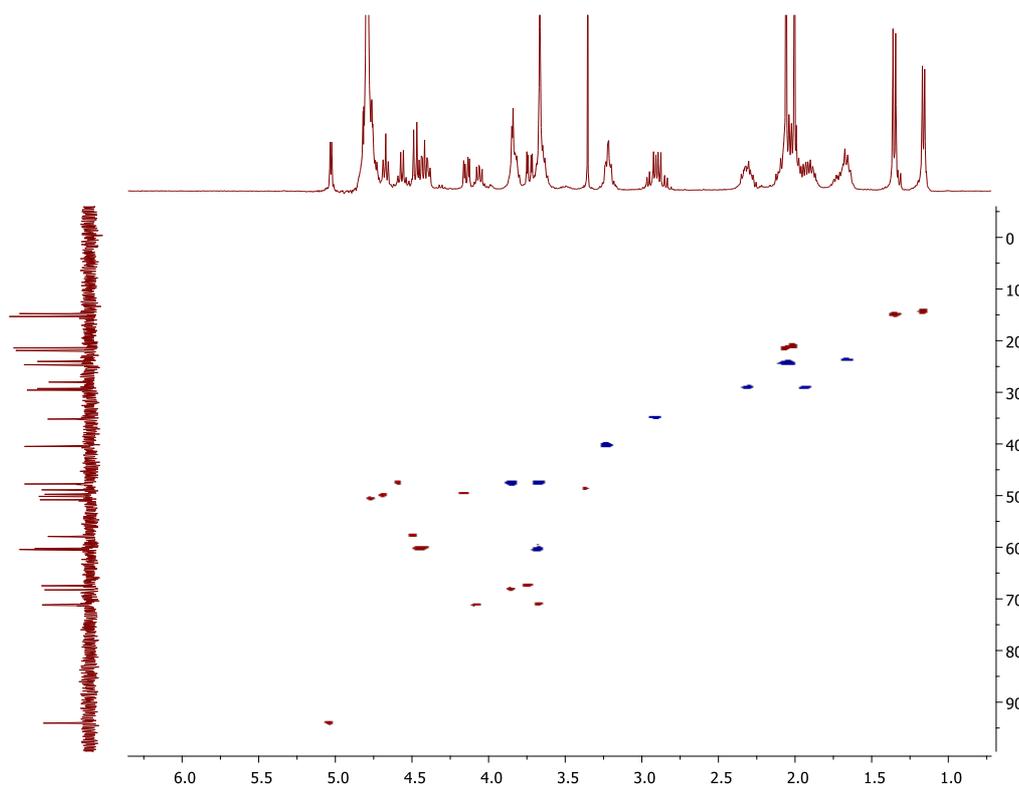
9.3. SPR curves and steady states analyses of compounds Thr', Thr*' and Hnv*'

9.3.1. Thr' analysis.

SPR curves obtained at 7 °C
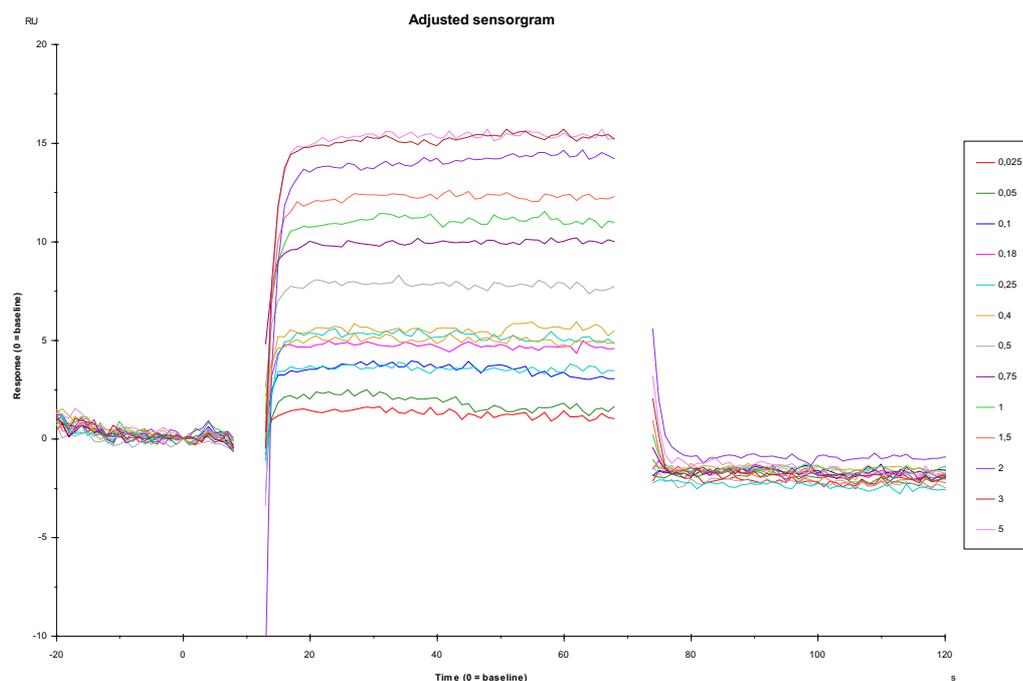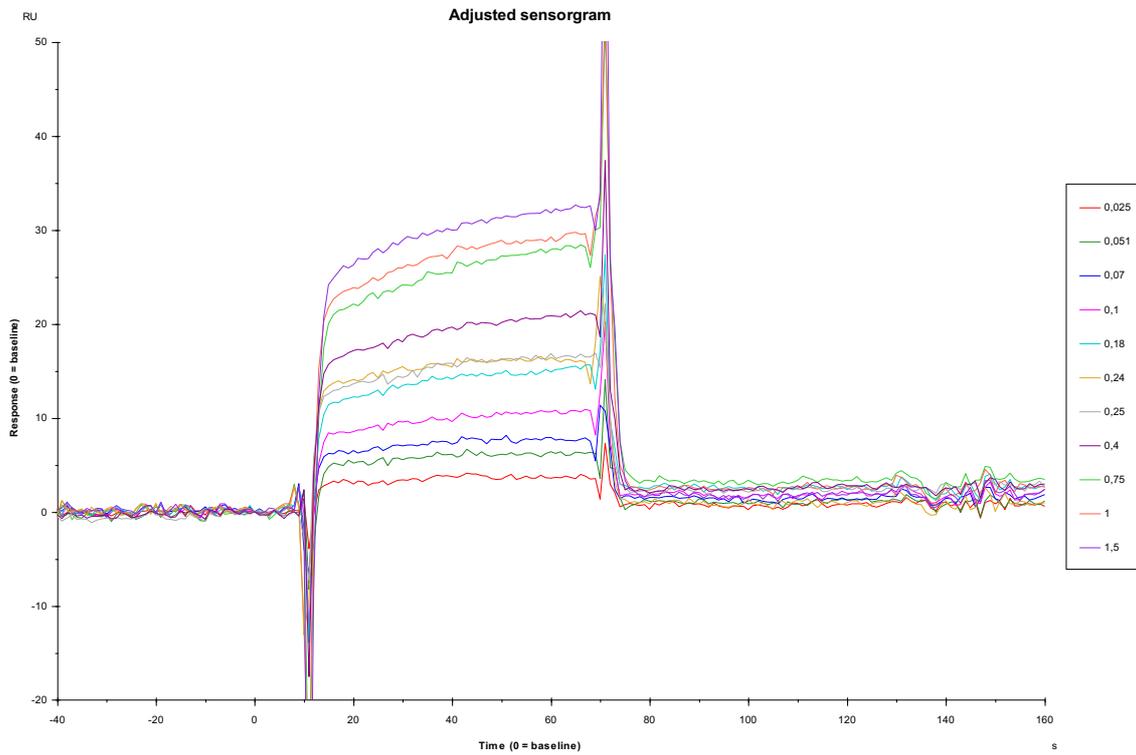

Steady state analysis obtained at 7 °C
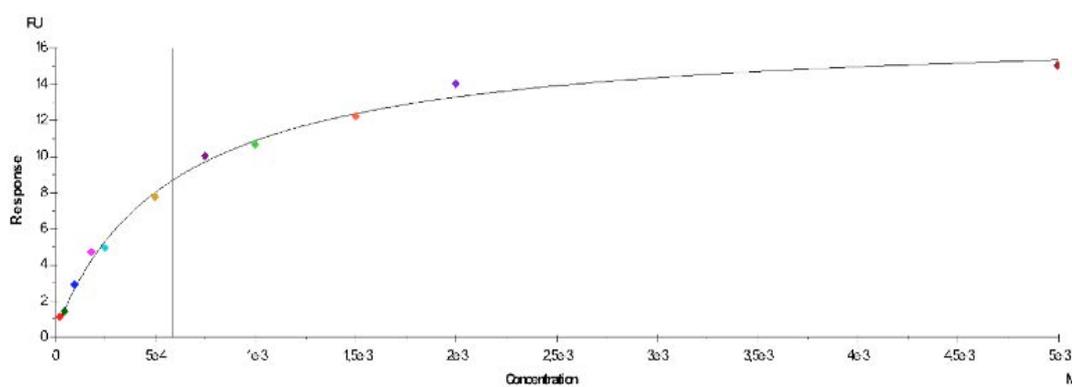
K_D (M)	R_{max} (RU)	Chi² (RU²)
5,872E-4	16,85	0,164

SPR curves obtained at 15 ºC
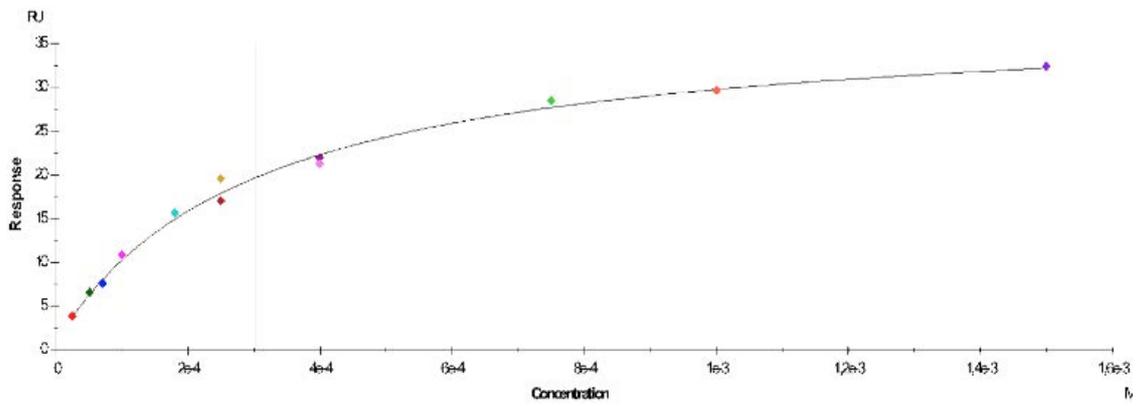

Steady state analysis obtained at 15 ºC

| K_D (M)	R_{max} (RU)	Chi² (RU²)
0,001180	20,44	0,418

SPR curves obtained at 25 ℃

Steady state analysis obtained at 25 ℃

K_D (M)	R_{max} (RU)	Chi² (RU²)
0,003480	32,29	0,0930

9.3.2. Thr*' analysis.

SPR curves obtained at 7 °C
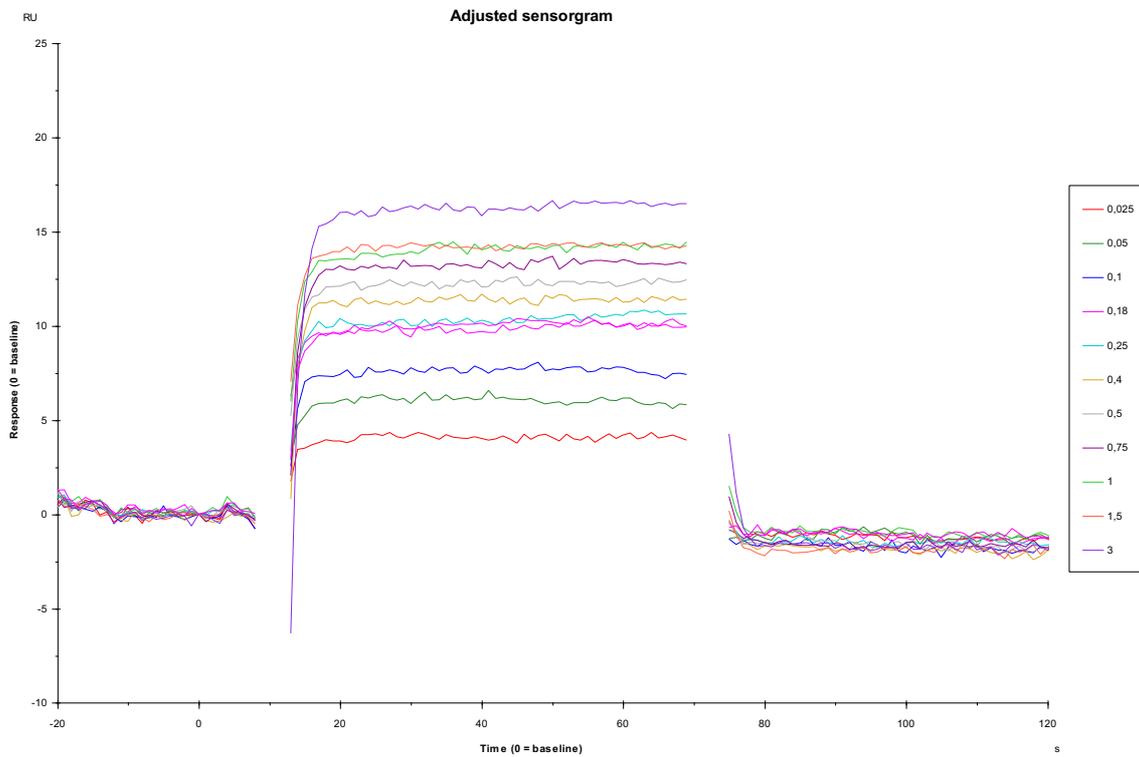
Steady state analysis obtained at 7 °C
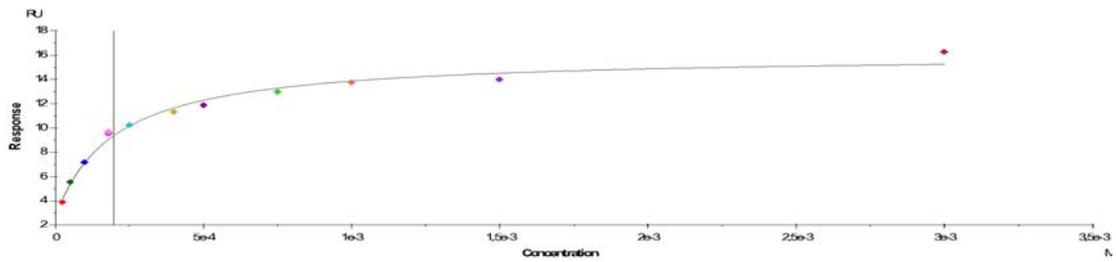
K_D (M)	R_{max} (RU)	Chi² (RU²)
1,967E-4	13,39	0,266

SPR curves obtained at 15 °C
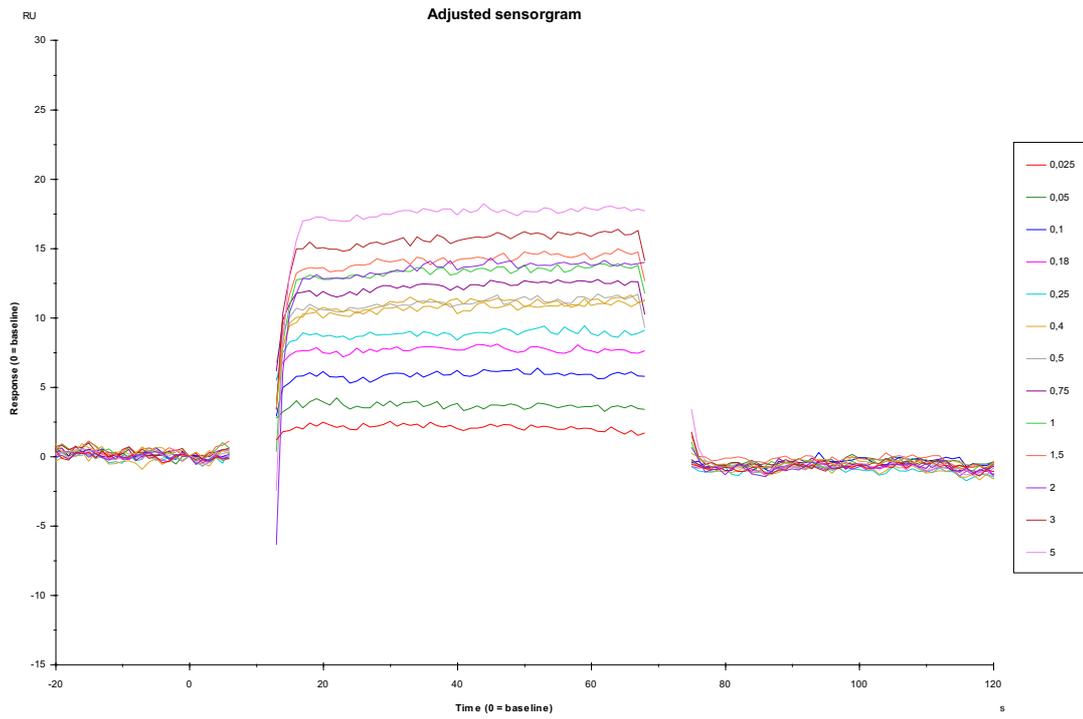
Steady state analysis obtained at 15 °C
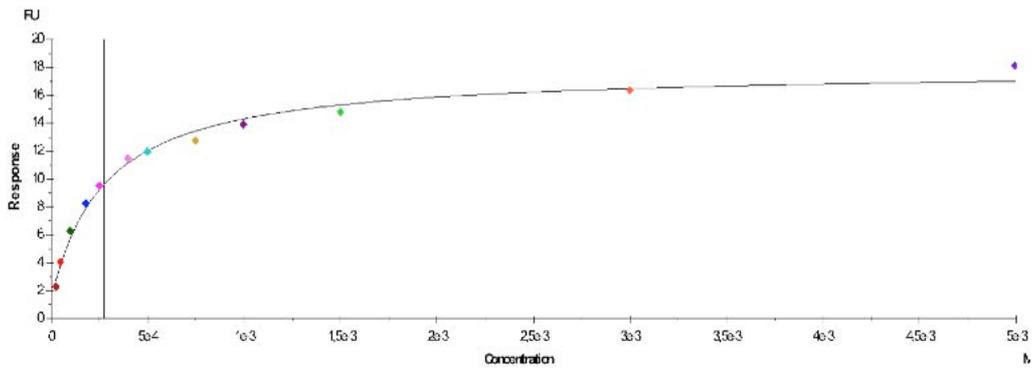
K_D (M)	R_{max} (RU)	Chi² (RU²)
2,721E-4	16,54	0,331

SPR curves obtained at 25 °C
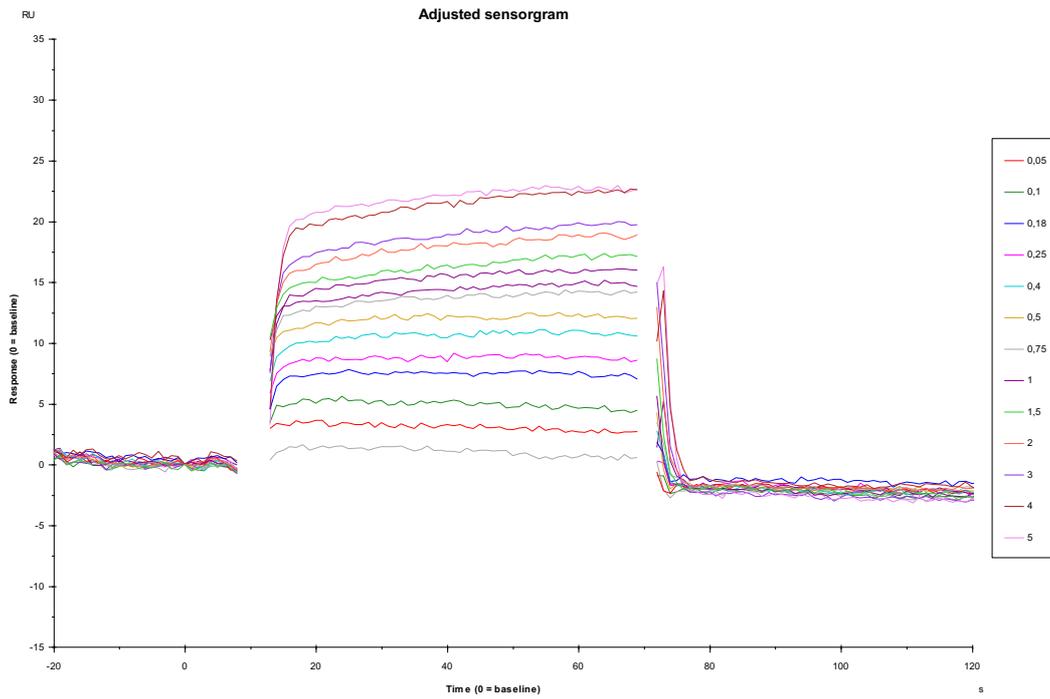

Steady state analysis obtained at 25 °C
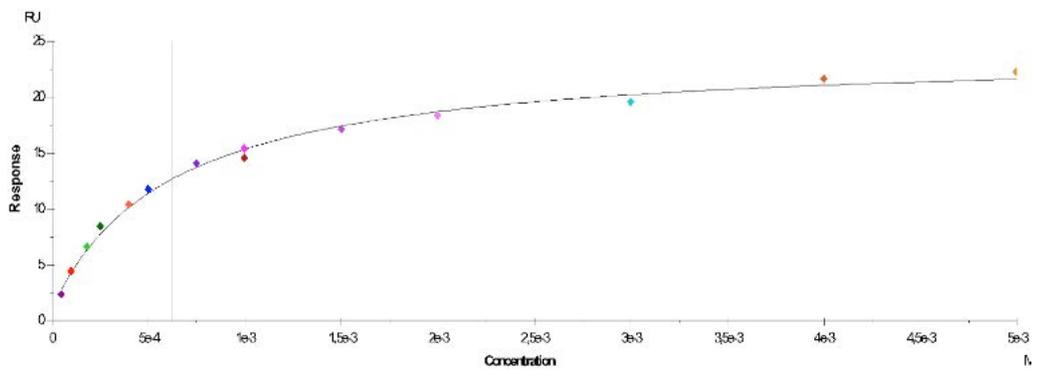
K_D (M)	R_{max} (RU)	Chi² (RU²)
6,189E-4	23,05	0,271

SPR curves obtained at 37 °C
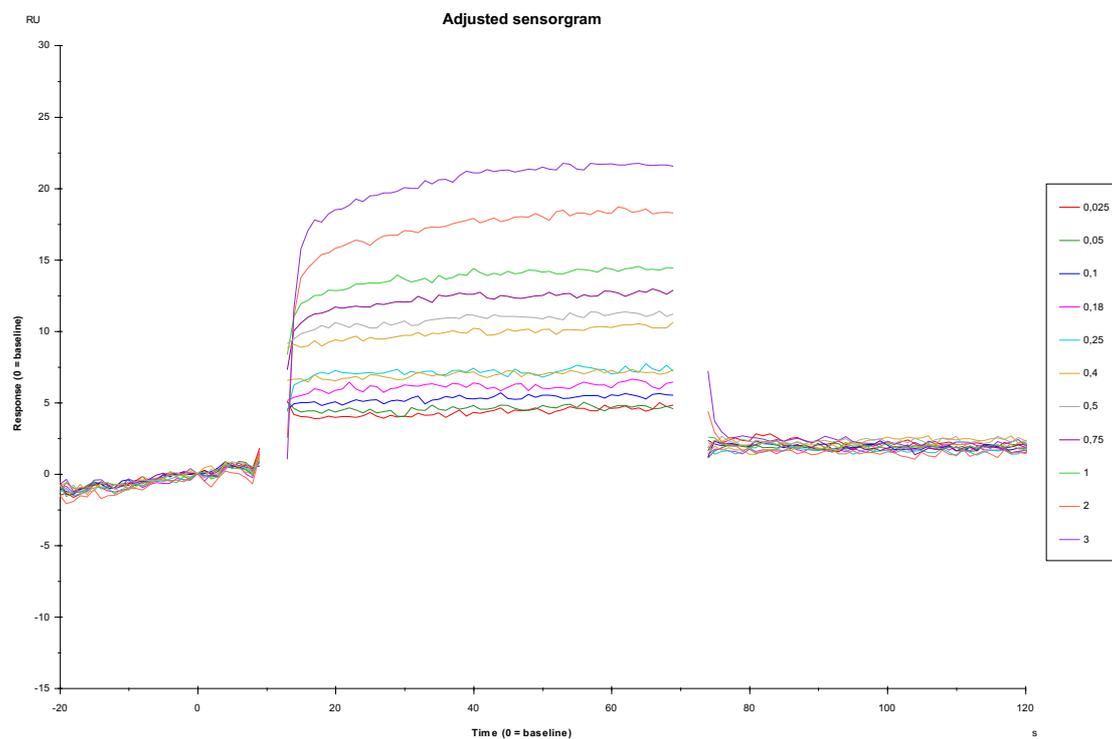

Steady state analysis obtained at 37 °C
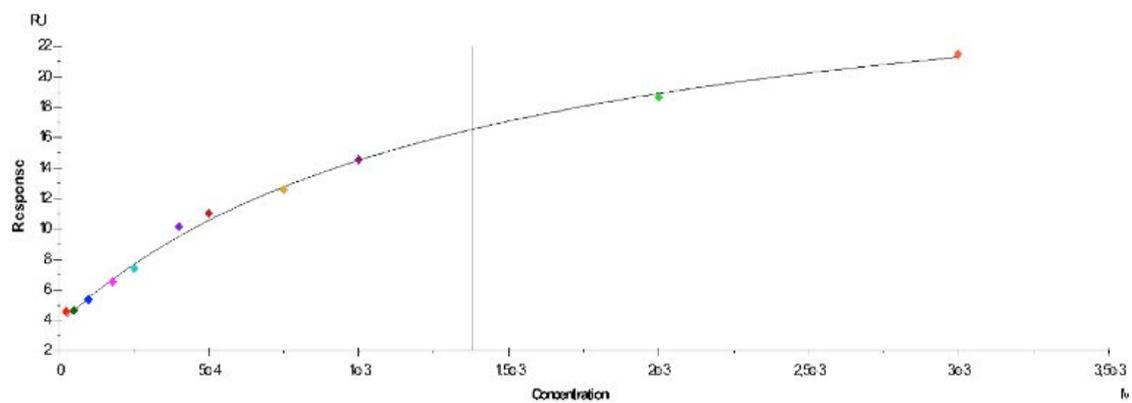
K$_D$ (M)	R$_{max}$ (RU)	Chi² (RU²)
0,001377	25,58	0,112

9.1.1. Hnv*' analysis.

SPR curves obtained at 10 °C
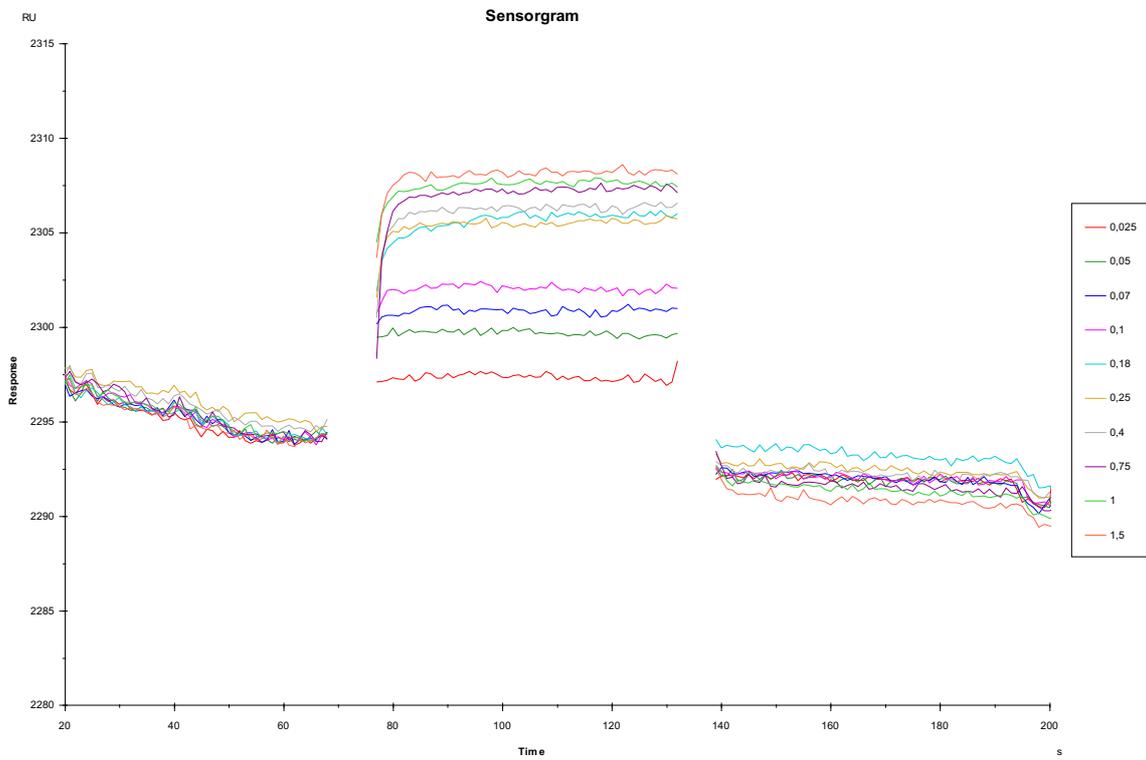
Steady state analysis obtained at 10 °C
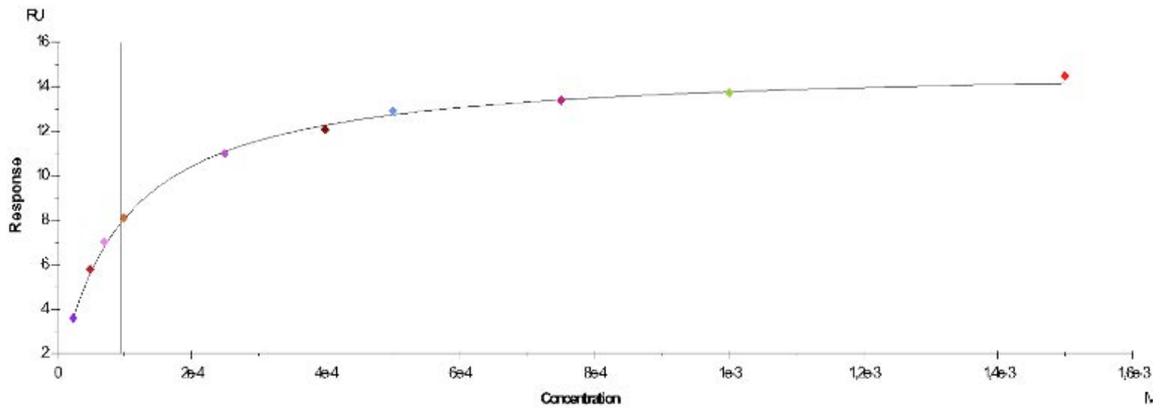
K_D (M)	R_{max} (RU)	Chi² (RU²)
9,427E-5	14,32	0,0381

SPR curves obtained at 15 ºC

Steady state analysis obtained at 15 ºC

K_D (M)	R_{max} (RU)	Chi² (RU²)
1,264E-4	27,81	0,634

SPR curves obtained at 25 ºC

Steady state analysis obtained at 25 ºC

K_D (M)	R_{max} (RU)	Chi² (RU²)
3,036E-4	37,55	0,700

SPR curves obtained at 37 °C
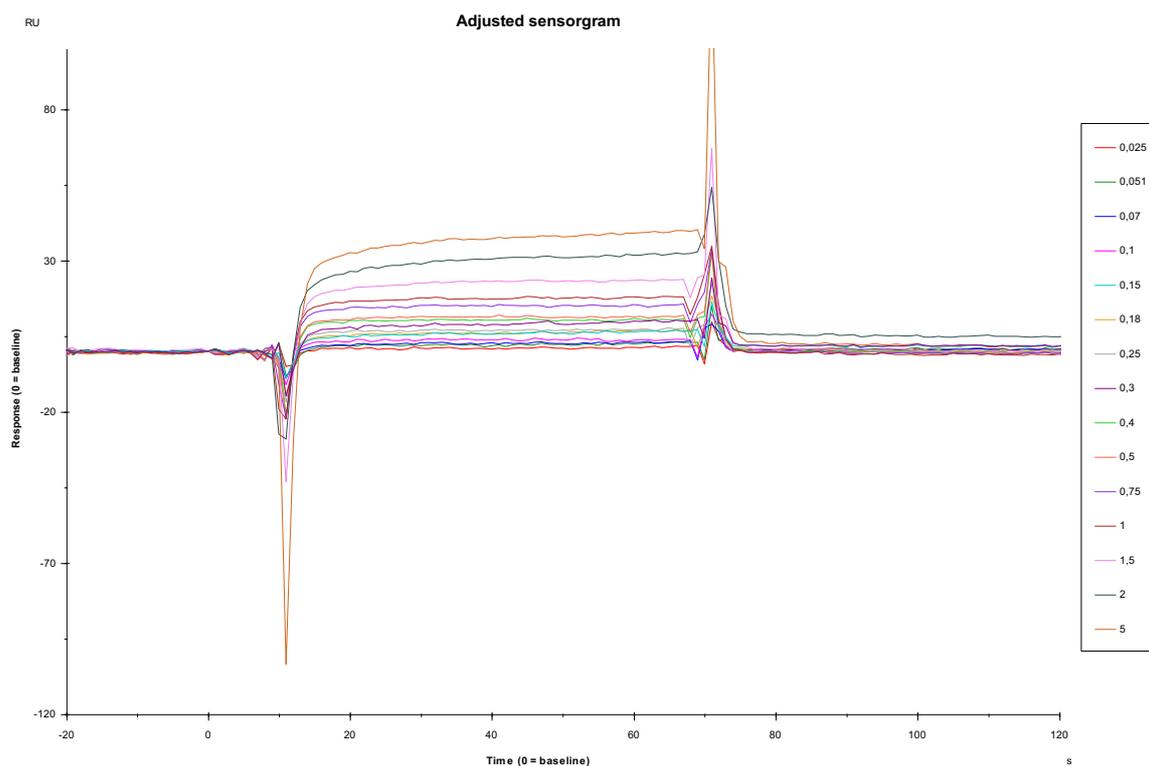
Steady state analysis obtained at 37 °C
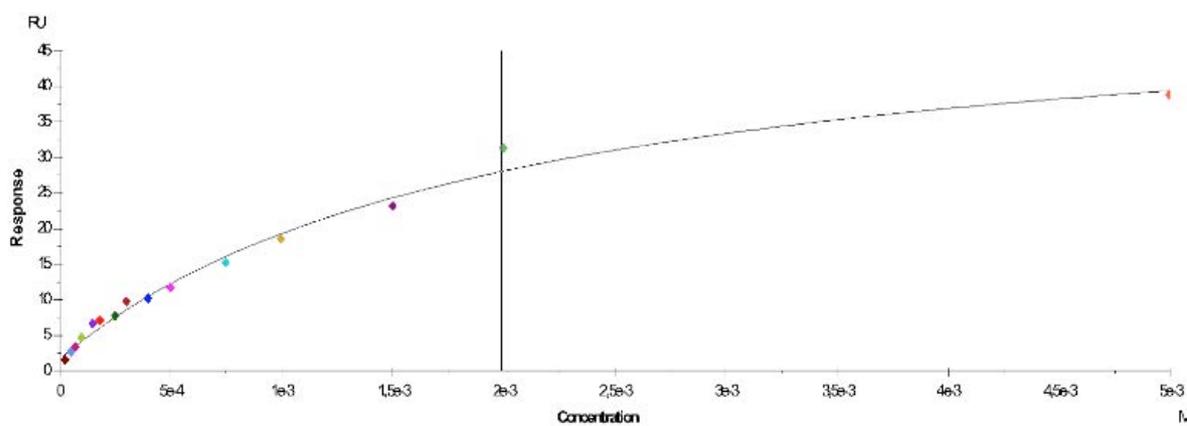
K_D (M)	R_{max} (RU)	Chi² (RU²)
0,001989	52,69	1,687

9.4. NMR spectra and chromatograms chapter 5
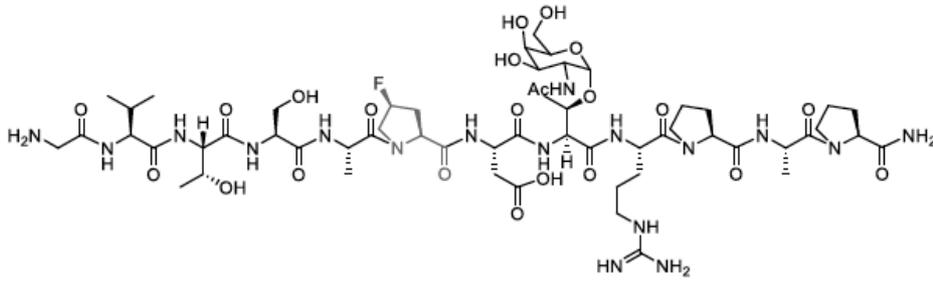
Glycopeptide P*'
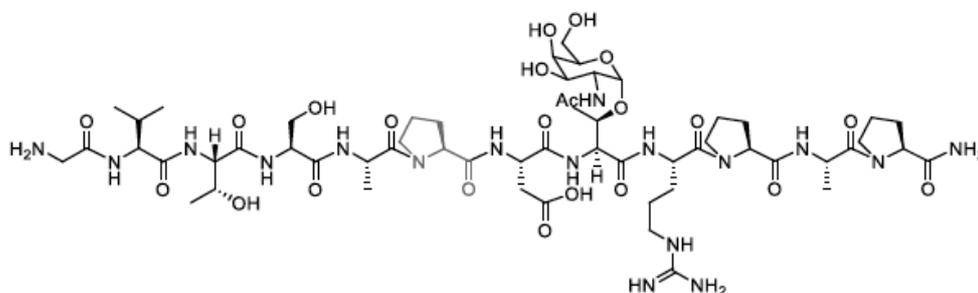
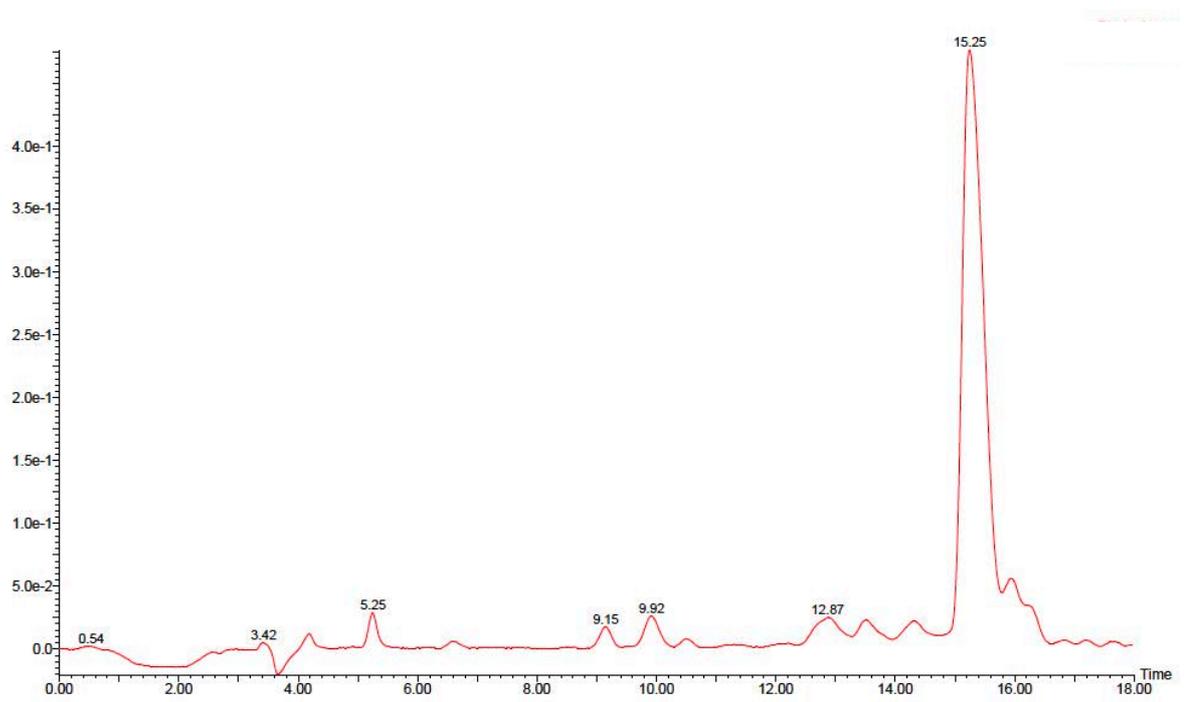HPLC chromatogram
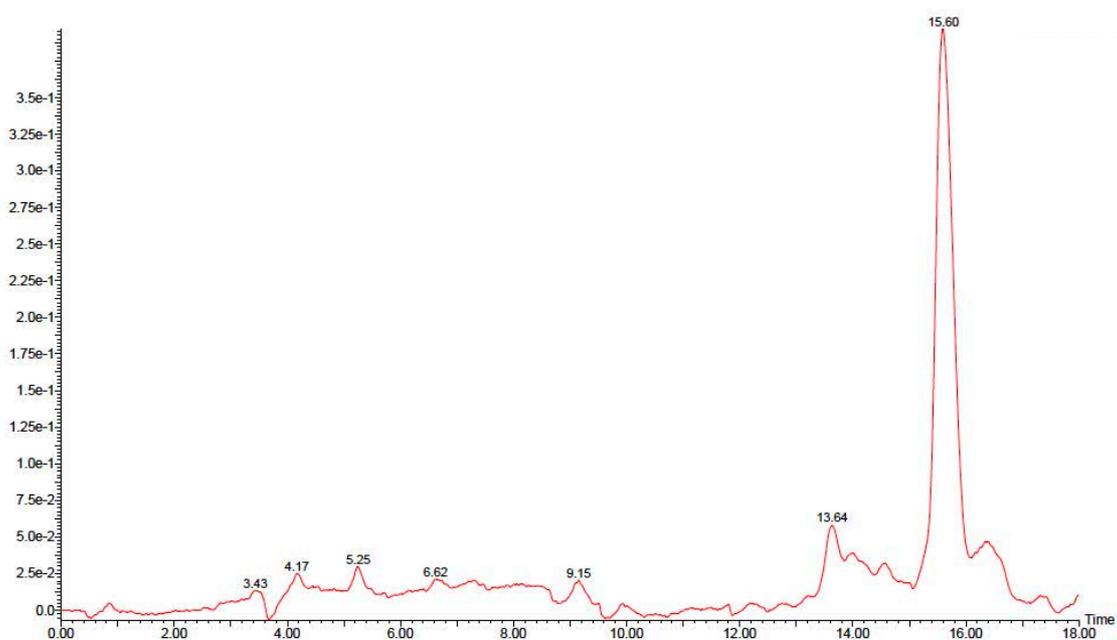
¹H NMR 400 MHz in D₂O registered at 298K
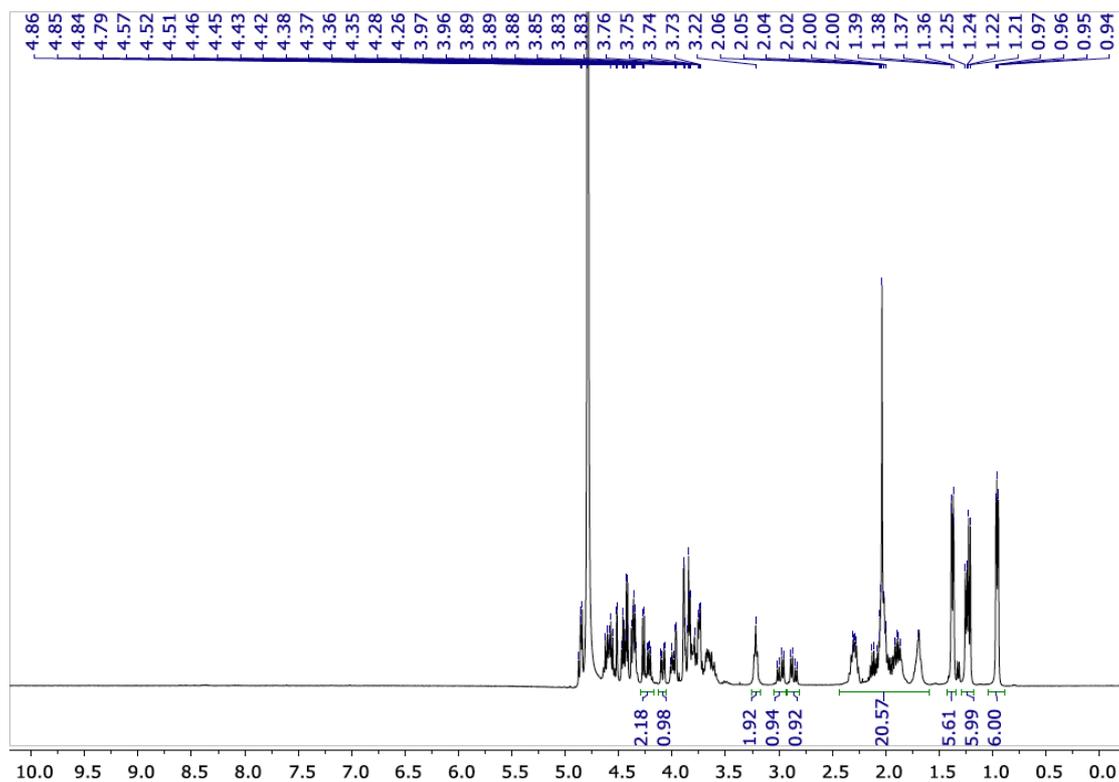
¹³C NMR 75 MHz in D₂O registered at 298K
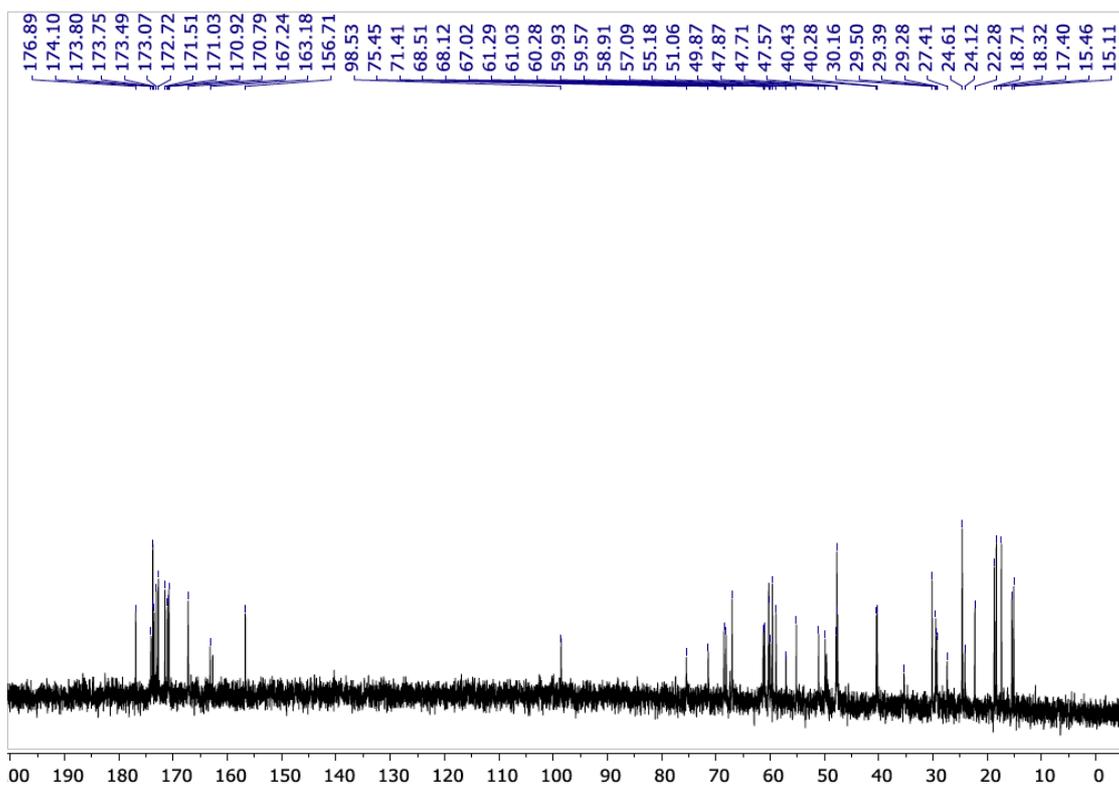
Glycopeptide fP*'

HPLC chromatogram
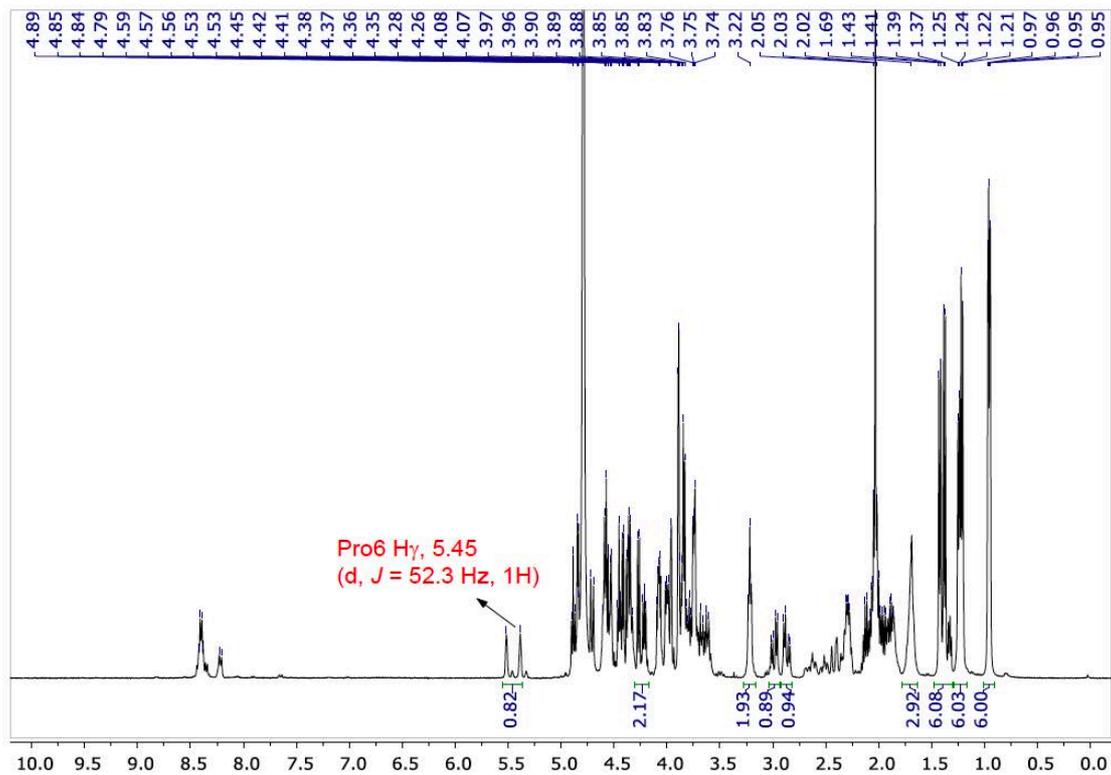
^{1}H NMR 400 MHz in D$_2$O registered at 298K

Peak labels (top): 4.89, 4.85, 4.84, 4.79, 4.59, 4.57, 4.56, 4.53, 4.53, 4.45, 4.42, 4.41, 4.38, 4.37, 4.36, 4.35, 4.28, 4.26, 4.08, 4.07, 3.97, 3.96, 3.90, 3.89, 3.88, 3.85, 3.85, 3.83, 3.76, 3.75, 3.74, 3.22, 2.05, 2.03, 2.02, 1.69, 1.43, 1.41, 1.39, 1.37, 1.25, 1.24, 1.22, 1.21, 0.97, 0.96, 0.95, 0.95

Pro6 Hγ, 5.45
(d, J = 52.3 Hz, 1H)
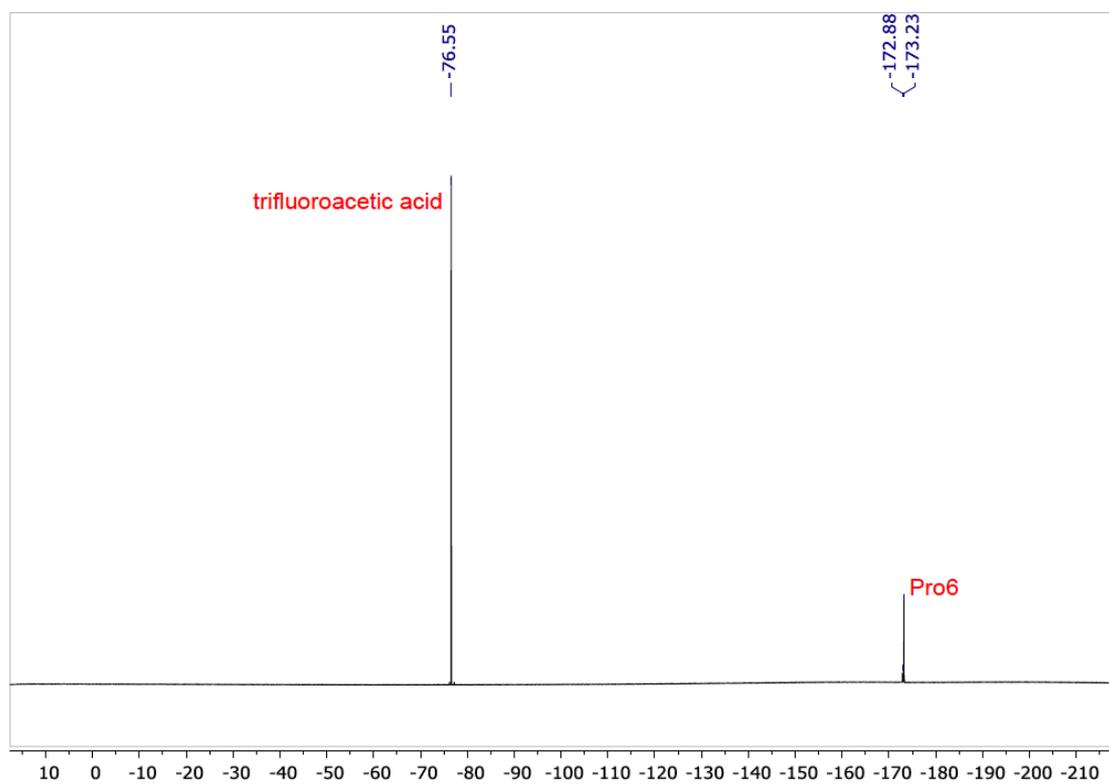
Integrals: 0.82, 2.17, 1.93, 0.89, 0.94, 2.92, 6.08, 6.03, 6.00

A second set of signals (in a small percentage) is observed. They correspond to the cis disposition of the amide bond of proline residues.

^{13}C NMR 75 MHz in D$_2$O registered at 298K
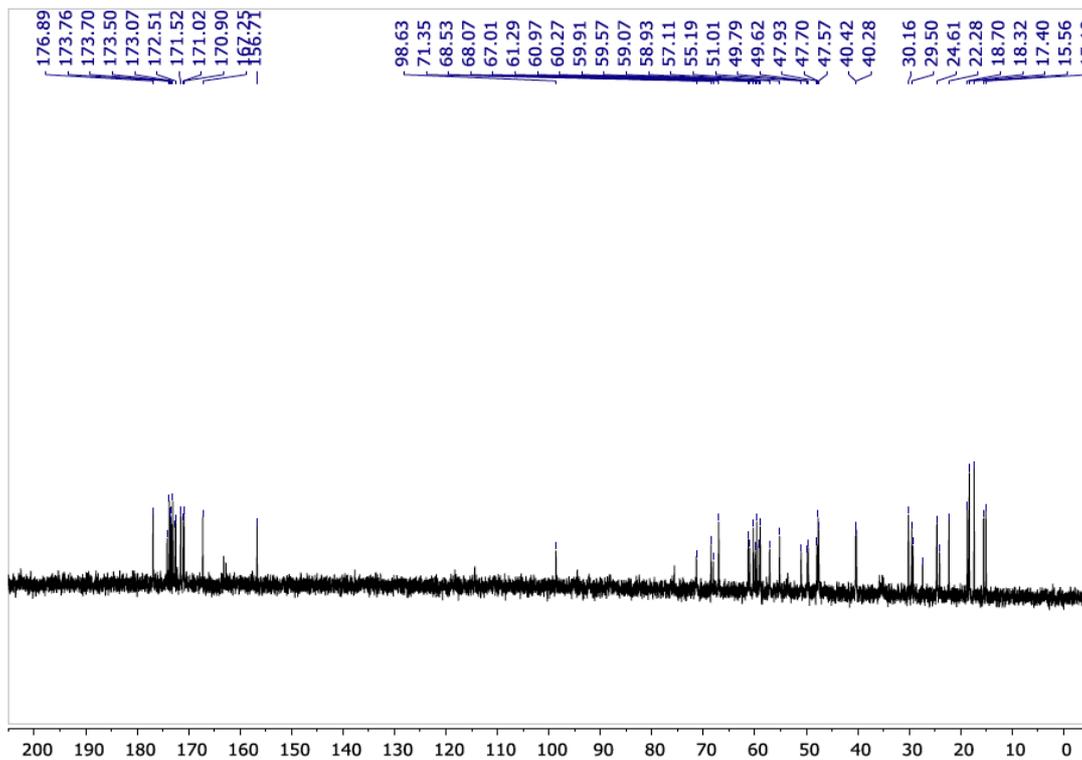
Peak labels (top): 176.89, 173.76, 173.70, 173.50, 173.07, 172.51, 171.52, 171.02, 170.90, 166.75, 98.63, 71.35, 68.53, 68.07, 67.01, 61.29, 60.97, 60.27, 59.91, 59.57, 59.07, 58.93, 57.11, 55.19, 51.01, 49.79, 49.62, 47.93, 47.70, 47.57, 40.42, 40.28, 30.16, 29.50, 24.61, 22.28, 18.70, 18.32, 17.40, 15.56, 15.10

224
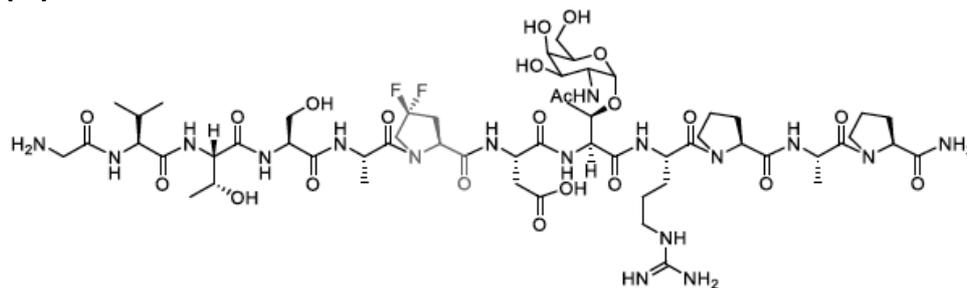
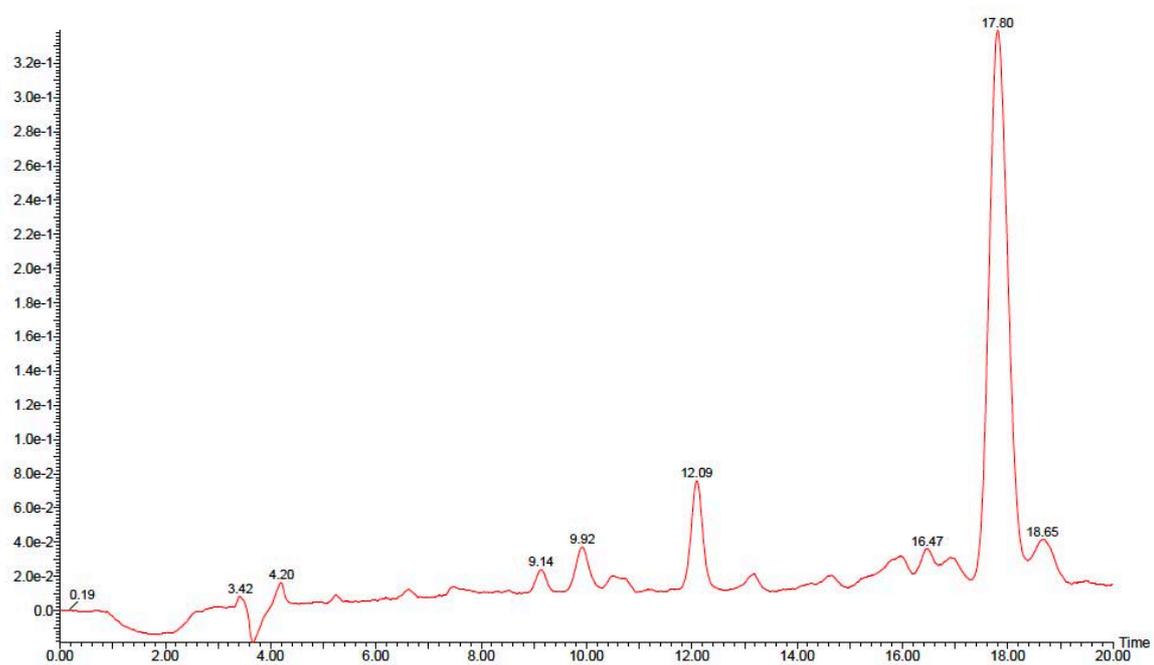Decoupled $^{19}F\{^{1}H\}$ NMR 282 MHZ in D_2O registered at 298K

A second set of signals (in a small percentage) is observed. They correspond to the cis disposition of the amide bond of proline residues.
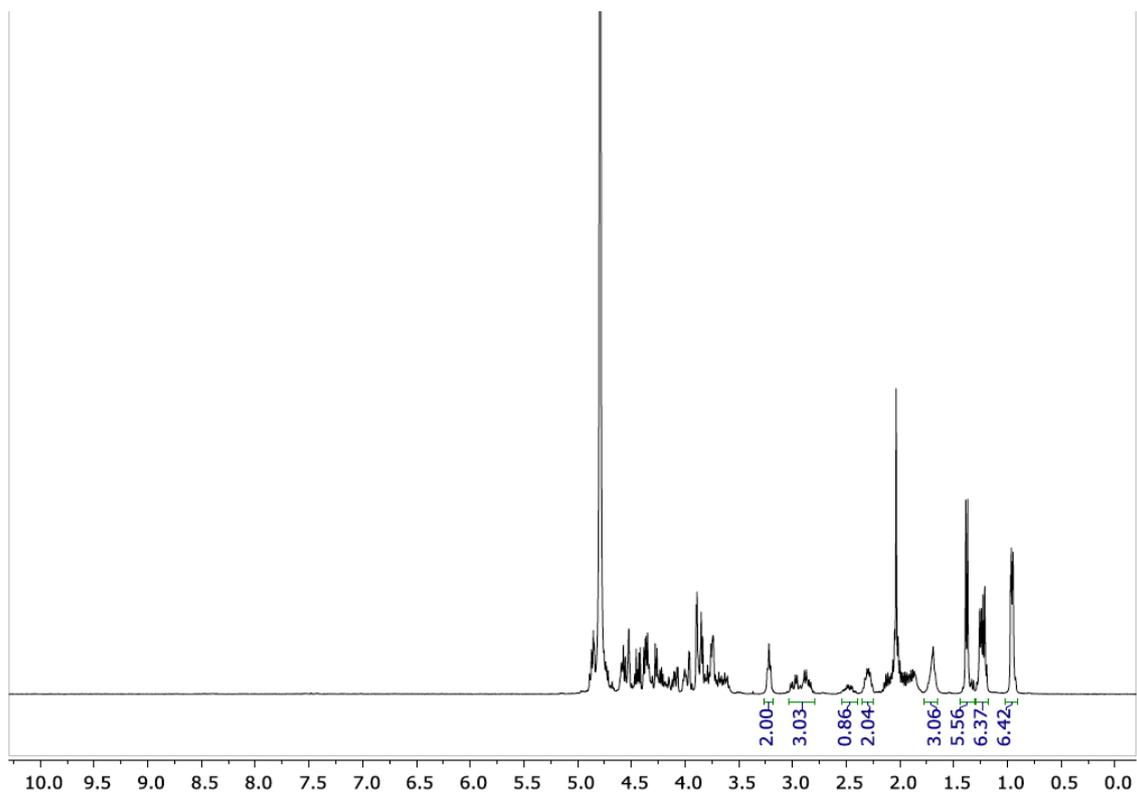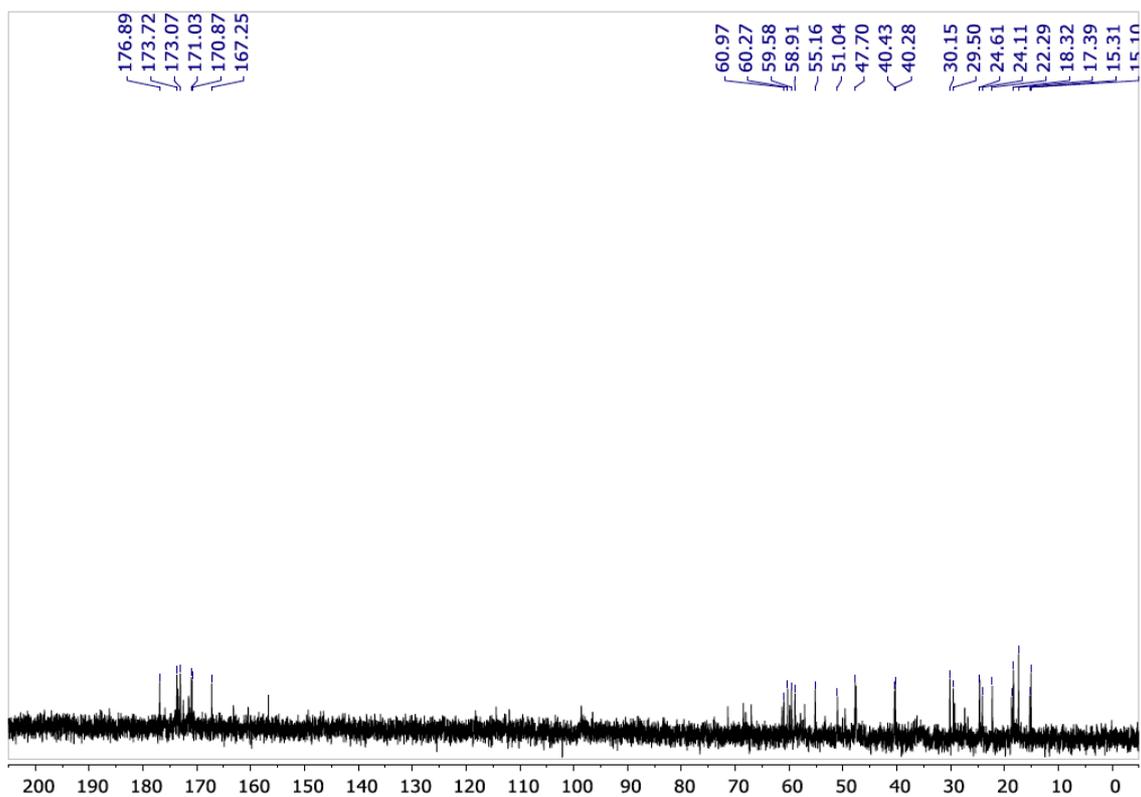
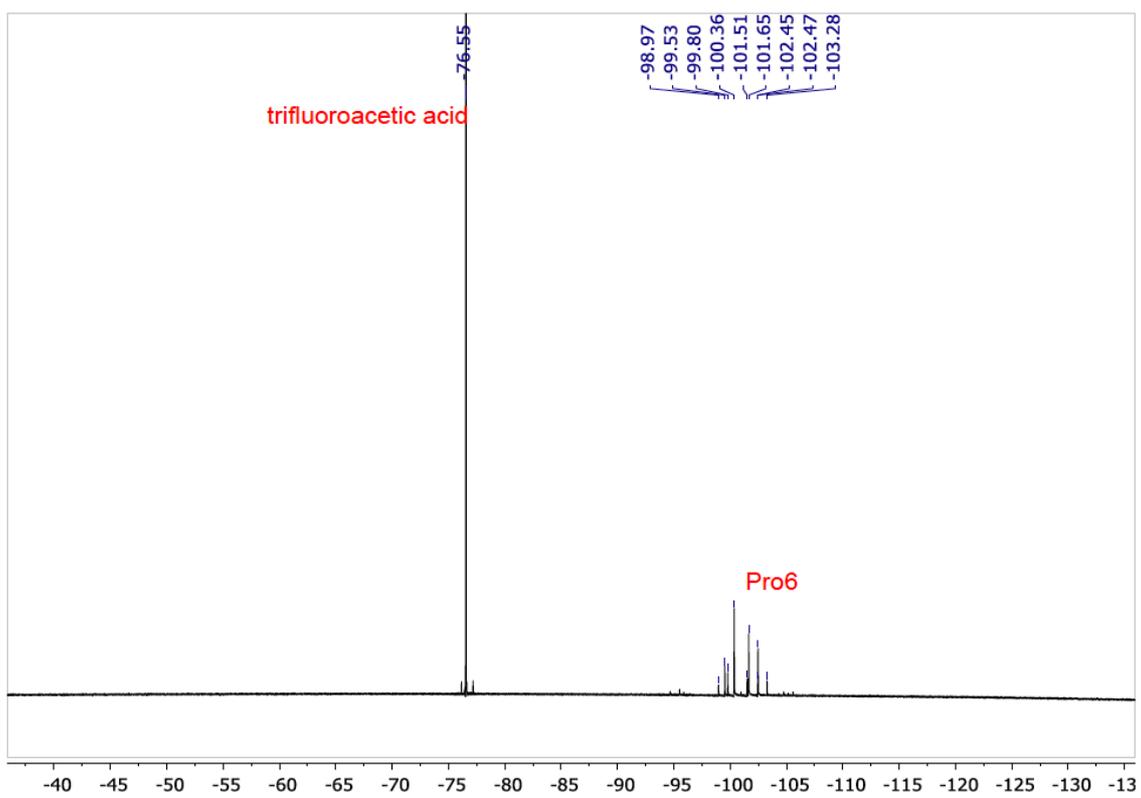**Glycopeptide 2fP*'**

HPLC chromatogram

¹H NMR 400 MHz in D₂O registered at 298K

2.00
3.03
0.86
2.04
3.06
5.56
6.37
6.42

¹³C NMR 75 MHz in D₂O registered at 298K

176.89
173.72
173.07
171.03
170.87
167.25
60.97
60.27
59.58
58.91
55.16
51.04
47.70
40.43
40.28
30.15
29.50
24.61
24.11
22.29
18.32
17.39
15.31
15.10

Decoupled ^{19}F{^{1}H} NMR 282 MHZ in D$_2$O registered at 298K

A second set of signals (in a small percentage) is observed. They correspond to the cis disposition of the amide bond of proline residues.