UNIVERSIDAD DE LA RIOJA

# TESIS DOCTORAL

| Título |
| --- |
| **Reconocimiento Molecular de Análogos de Antígenos Tn como Base para el Diseño de Vacunas Terapéuticas contra el Cáncer** |
| Autor/es |
| **Ismael Compañón Pérez** |
| Director/es |
| Francisco Corzana López y Alberto Avenoza Aznar |
| Facultad |
| Facultad de Ciencia y Tecnología |
| Titulación |
| |
| Departamento |
| Química |
| Curso Académico |
| |

FACULTAD DE CIENCIA Y TECNOLOGÍA

UNIVERSIDAD DE LA RIOJA

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

TESIS DOCTORAL

# Reconocimiento Molecular de Análogos de Antígenos Tn como Base para el Diseño de Vacunas Terapéuticas contra el Cáncer

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

**Ismael Compañón Pérez**

September 2019

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

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

HACEN CONSTAR:

*Que la memoria ''* Reconocimiento Molecular de Análogos de Antígenos Tn como Base para el Diseño de Vacunas Terapéuticas contra el Cáncer*''* *realizada por* ***Ismael Compañón Pérez****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, a 13 de Septiembre de 2019

Los directores

Francisco Corzana López

Alberto Avenoza Aznar

**Resumen**

Las mucinas son una familia de glicoproteínas que se encuentran altamente expresadas en las células de los tejidos epiteliales. La característica principal de las mucinas es la presencia en su estructura de dominios ampliamente glicosilados constituidos por glicanos muy complejos. Estos se unen a la cadena peptídica a través de enlaces α-*O*-glicosídicos con los residuos de treonina y serina, siendo el primer carbohidrato la N-acetilgalactosamina (GalNAc). En las células tumorales se producen cambios importantes en los patrones de glicosilación de la MUC1, lo cual va asociado a la exposición de determinados antígenos que se encuentran enmascarados en las moléculas presentes en las células sanas. Esta característica abre la posibilidad de su utilización como diana terapéutica ya que, mediante su incorporación en vacunas multicomponente, se consigue que el sistema inmune genere anticuerpos anti-MUC1.

Por ello, el objetivo principal de esta tesis es el diseño de antígenos que presenten modificaciones capaces de mejorar la afinidad mostrada por los anticuerpos anti-MUC1 y que además poseen una mayor inmunogenicidad y resistencia a la hidrólisis enzimática.

En este contexto, se ha desarrollado una metodología que permite la síntesis estereoselectiva de una amplia variedad de derivados de *S*-glicosilcisteína.Por otro lado, se han estudiado en detalle los elementos claves para el reconocimiento molecular de antígenos tipo MUC1 por el anticuerpo scFv-SM3, logrando explicar las diferencias encontrada en el reconocimiento de los antígenos Tn de Ser y Thr. Dichos estudios han motivado el diseño de rutas sintéticas que permiten la obtención de los *building blocks* de *S*-(α-D-GalNAc)-*tiotreonina* y *Se*-(α-D-GalNAc)-*selenotreonina*, los cuales son aptos para la incorporación en secuencias peptídicas mediante el empleo de la técnica de síntesis de péptidos en fase

solida (SPPS). Además, se ha demostrado la capacidad como antígeno del péptido no natural que incorpora el aminoácido *S*-(α-D-GalNAc)-*tiotreonina* mediante la formulación de una vacuna terapéutica.

Asimismo, se han efectuado una serie de colaboraciones con diversos grupos de investigación, a nivel nacional e internacional, que han tenido como objetivo la síntesis, purificación y caracterización de péptidos y glicopéptidos.

## Abstract

Mucins are a family of glycoproteins that are highly expressed in epithelial tissue cells. The main characteristic of mucins is the presence in their structure of widely glycosylated domains featuring very complex glycans. These carbohydrates bind to the peptide backbone through α-*O*-glycosidic bonds with threonine and serine residues, being the first carbohydrate N-acetylgalactosamine (GalNAc). Notably, significant changes in the glycosylation patterns of MUC1 occur in tumor cells, which is associated with the exposure of antigens that are masked in the molecules present in healthy tissues. This characteristic opens the possibility of using tumor-associated MUC1 as a therapeutic target. Its incorporation in multicomponent vaccines can result in a robust immune response able to generate anti-MUC1 antibodies.

Therefore, the main objective of this thesis is the design of antigens that present structural modifications capable of improving the affinity shown by the anti-MUC1 antibodies and with high immunogenicity and resistance to enzymatic hydrolysis. In this context, a methodology has been developed that allows stereoselective synthesis of a wide variety of S-glycosylcysteine derivatives.

On the other hand, the critical elements for the molecular recognition of MUC1 type antigens by scFv-SM3 antibody have been studied in detail, explaining the differences found in recognition of Ser and Thr Tn antigens. These studies have motivated the design of synthetic routes that allows obtaining the building blocks of S-(α-D-GalNAc)-thiothreonine and Se-(α-D-GalNAc)- selenothreonine, which are ready-to-use in solid-phase peptide synthesis technique (SPPS). Moreover, it has been possible to demonstrate the capacity as the antigen of the unnatural peptide that incorporates the

amino acid S-(α-D-GalNAc)-thiothreonine by formulating a therapeutic vaccine.

Besides, a series of collaborations have been made with various research groups, nationally and internationally, to perform synthesis, purification and characterization of several peptides and glycopeptides.

## Abreviaciones

| | |
|---|---|
| **δ** | desplazamiento químico |
| **ΔG** | incremento de energía libre de Gibbs |
| **ΔH** | incremento de entalpía |
| **ΔS** | incremento de entropía |
| **ºC** | grados Celsius |
| **$^{1}$H NMR** | *proton nuclear magnetic resonance* |
| **$^{13}$C NMR** | *carbon nuclear magnetic resonance* |
| **Å** | angstrom |
| **AAb** | *autoantibody* |
| **Ab** | *antibody* |
| **Ac** | acetil |
| **$Ac_2O$** | anhidrido acético |
| **AcOEt** | acetato de etilo |
| **Allyl** | alilo |
| **AcCl** | cloruro de acetilo |
| **BLI** | *bio-layer interferometry* |
| **Boc** | *terc*-butoxicarbonilo |
| **Bn** | bencilo |
| **Cbz** | benciloxicarbonilo |
| **COSY** | *correlated spectroscopy* |
| **DBU** | 1,8-diazabiciclo[5.4.0]undec-7-eno |
| **DFT** | *density functional theory* |
| **Dha** | deshidroalanina |
| **Dhb** | deshidrobutirina |
| **DIEA** | diisopropiletilamina |
| **DMF** | dimetilformamida |
| **Dr** | *diasteromeric ratio* |

| | |
|---|---|
| **ELISA** | *enzyme-linked imminosorbent assay* |
| **ELLA** | *enzyme-linked lectin assay* |
| **ESI** | *electrospray ionization* |
| **Fmoc** | 9-fluoroenilmetoxicarbonilo |
| **Fmoc-*O*Su** | 9-fluoroenilmetoxicarbonato de *O*-succinimida |
| **Gal** | galactosa |
| **GalNAc** | *N*-acetilgalactosamina |
| **Glc** | glucosa |
| **GlcNAc** | *N*-acetilglucosamina |
| **Grad** | gradiente |
| **HBTU** | hexafluorofosfato de O-(benzotriazol-1-il)-N,N,N',N'-tetrametiluronio |
| **HEK293T** | *human embryonic kidney* 293T |
| **HPLC** | *high performance liquid cromatography* |
| **HSQC** | *heteronuclear single quantum correlation* |
| **Hz** | hercio |
| **IgG** | inmunoglobulina G |
| **iPrOH** | isopropanol |
| **IRID** | *infrared ion-dip* |
| **J** | constante de acoplamiento |
| **$K_D$** | constante de disociación |
| **KHDMS** | bis(trimetillsilil)amiduro de potasio |
| **KLH** | *keyhole limpet haemocyanin* |
| **Lan** | lantionina |
| **mAB** | *monoclonal antibody* |
| **MD** | *molecular dinamics* |
| **MD-tar** | *molecular dynamics whit weighted time-averaged restraints* |
| **Me** | metil |

| | |
|---|---|
| **MeLan** | metillantionina |
| **MHBA** | 4-metilbencidrilamina |
| **MW** | *microwave* |
| **NPs** | nanoparticulas |
| **NOE** | *nuclear overhauser effect* |
| **NOESY** | *nuclear overhauser effect spectroscopy* |
| **OVA** | ovoalbumina |
| **PBS** | *phosphate buffered saline* |
| **PEG** | polietilenglicol |
| **PPII** | poliprolina II |
| **ppm** | partes por millón |
| **Py** | piridina |
| **QM** | *quantum mechanics* |
| **Rf** | *rate factor* |
| **ROESY** | *rotating-frame nuclear overhauser effect correlation spectroscopy* |
| **$R_t$** | *retention time* |
| **SBA** | *soybean agglutinin* |
| **ScFv** | *single-chain variable fragment* |
| **SPE** | *solid phase extraction* |
| **SPPS** | *solid phase peptide synthesis* |
| **SPR** | *surface plasmon resonance* |
| **STD** | *saturation transfer difference* |
| **Su** | succinimida |
| **Syn** | *synplanar* |
| **$^t$Bu** | *terc*-butil |
| **TACA** | *tumor-associated carbohydrate antigen* |
| **TBTU** | tetrafluoroborato de O-(benzotriazol-1-il)-N,N,N’,N’-tetrametiluronio |

| | |
|---|---|
| **TFA** | ácido trifluoroacetico |
| **THF** | tetrahidrofurano |
| **TIS** | triisopropilsilano |
| **TLC** | *thin layer chromatography* |
| **TLRs** | *toll-like receptors* |
| **TMB** | tetrametilbencidina, terametoxibutano |
| **TOF** | *time of flight* |
| **tr-NOE** | *transferred nuclear overhauser effect* |
| **UDP** | *uridine diphosphate* |
| **UV** | ultravioleta |
| **V** | potencial |

# Introducción

1

Las mucinas son una familia de glicoproteínas que se encuentran altamente expresadas en tejidos mucosos.[1] Aunque presentan características estructurales comunes, existen varias clases de mucinas con funciones diferentes (Tabla 1).[2] Las hay que se encuentran unidas a la membrana celular de las células de los tejidos epiteliales y otras que son secretadas en la superficie de las mucosas. Estas últimas forman una capa protectora que cubren todo el tejido mucoso.[3–5] Por el contrario, las mucinas asociadas a la membrana son las encargadas, entre otras funciones, de generar señales para la regulación del crecimiento celular y la apoptosis.

| MUCINA | Subtipo | Distribución en tejido sano |
|---|---|---|
| MUC1 | Asociada a membrana | Pulmón, estomago, mama, endometrio, cavidad oral, cérvix, riñón, intestino y vejiga |
| MUC2 | Secretada | Intestino, cérvix y pulmón |
| MUC3A | Asociada a membrana | intestino |
| MUC3B | Asociada a membrana | Intestino |
| MUC4 | Asociada a membrana | Pulmón, cavidad oral e intestino |
| MUC5AC | Secretada | Pulmón, estomago, cérvix e intestino |
| MUC5B | Secretada | Pulmón, cavidad oral, cérvix e intestino |
| MUC6 | Secretada | Estomago e intestino |
| MUC7 | Secretada | Cavidad oral y pulmón |
| MUC12 | Asociada a membrana | Intestino |
| MUC13 | Asociada a membrana | Intestino y pulmón |
| MUC15 | Asociada a membrana | Intestino y mama |
| MUC16 | Asociada a membrana | Pulmón, cavidad oral, estomago intestino, tracto reproductivo y peritoneo |
| MUC17 | Asociada a membrana | Intestino |
| MUC19 | Secretada | Cavidad oral |
| MUC20 | Asociada a membrana | Riñón |
| MUC21 | Asociada a membrana | Pulmón, intestino grueso, timo y testículos |

**Tabla 1.** Mucinas presentes en distintos tejidos humanos.

La expresión anómala de estas mucinas puede desencadenar procesos inflamatorios, infecciones y el desarrollo de cáncer en estos tejidos

mucosos.[6,7] Debido a esto último, las mucinas son usadas como marcadores biológicos de la presencia de tumores y como agentes terapéuticos contra el cáncer.[2] La característica principal de todos los tipos de mucinas es la presencia de dominios ampliamente glicosilados constituidos por glicanos muy complejos, los cuales están unidos a residuos de treonina y serina mediante enlaces *O*-glicosídicos. Esto último hace que las mucinas sean unas de las glicoproteínas más complejas expresadas en los mamíferos.

Los glicanos presentes en estas mucinas pueden llegar a aportar hasta el 80% de la masa de estas glicoproteínas, atribuyéndoles la mayoría de sus propiedades. Estos glicanos se ensamblan paso a paso en el aparato de Golgi por acción de las glicosiltransferasas, siendo GalNAc el primer carbohidrato unido a la secuencia peptídica la *N*-acetilgalactosamina.

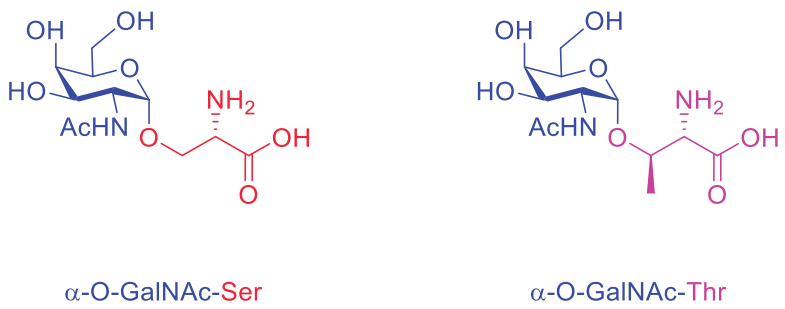



**Figura 1.** *Unión del carbohidrato GalNAc a una serina o treonina mediante un enlace α-O-glicosídico.*

Mediante la regulación de estas glicosiltransferasas, las células epiteliales pueden presentar una gran variedad de glicanos dentro de un mismo tejido. Por ejemplo, la MUC2, presente en el intestino, puede contener hasta 200 *O*-glicanos diferentes.[8] Estos glicanos son fundamentales para regular las interacciones célula-célula y célula-matriz, así como el reconocimiento celular y la señalización.

Generalmente, en las células tumorales se producen cambios importantes en los patrones de glicosilación, lo cual va asociado a la exposición de

determinados antígenos que se encuentran enmascarados en las moléculas presentes en las células sanas. Esta característica abre la posibilidad de utilizar las mucinas como marcadores de la presencia de tumores.

Una de las mucinas más estudiadas es la denominada MUC1.[9–11] Esta glicoproteína está formada por la repetición de la secuencia de 20 aminoácidos AHGVTSAPDTRPAPGSTAPP.[12]

La MUC1 está sobrexpresada en las células cancerosas, pudiendo encontrase en más del 90% de los tumores.[13] Esta glicoproteína, además de estar sobrexpresada, presenta fallos en su estructura debido, fundamentalmente, al mal funcionamiento de las glicosiltransferasas. Como consecuencia de esto, la MUC1 asociada a tumores presenta carbohidratos sencillos y que en muchos casos se reduce a uno o dos monosacáridos (Figuras 2 y 3).

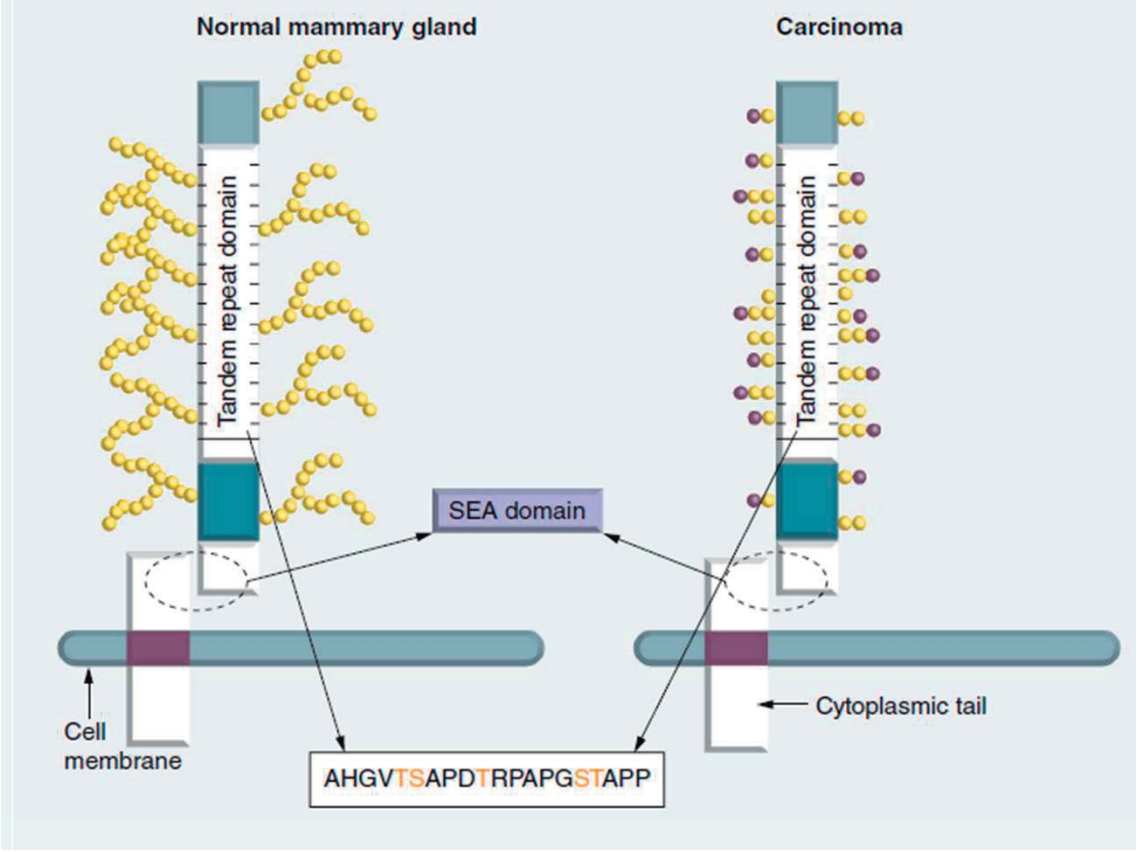



**Figura 2.** *MUC1 expresada por una glándula mamaria normal u por un carcinoma de mama. La secuencia de repetición de 20 aminoácidos se muestra en la parte inferior de la figura y contiene cinco puntos de glicosilación mostrados en color naranja.*

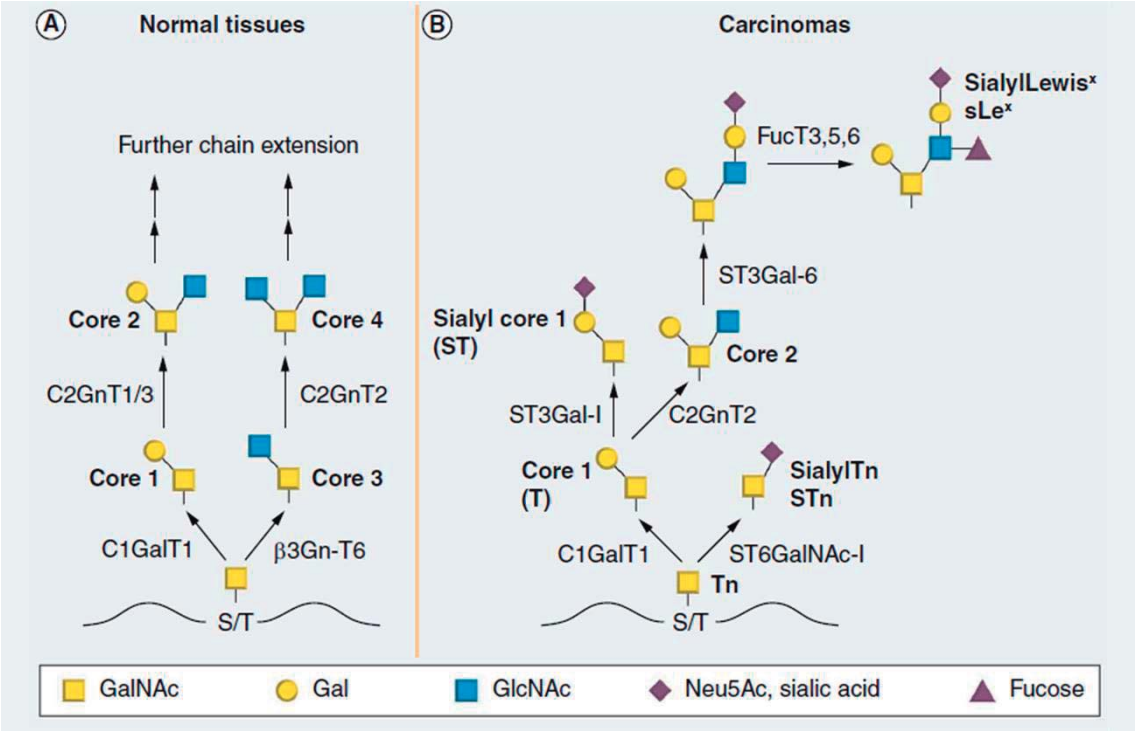



**Figura 3.** *Diferencias en la glicosidación de las mucinas presentes en un tejido sano (A) y un tejido canceroso (B).*

Así, antígenos como el Tn (GalNAcα1Ser/Thr), sTn (NeuAcα2,6GalNAcα1Ser/Thr) T (Galβ1,3GalNAcα1Ser/Thr), y sT (Neu5Ac2,3Galβ1,3GalNAcα1Ser/Thr), conocidos como TACAs (*tumor-associated carbohydrate antigens*), o el fragmento peptídico APDTRP, que en células sanas se encuentran enmascarados, quedan expuestos al sistema inmune.[14–18]

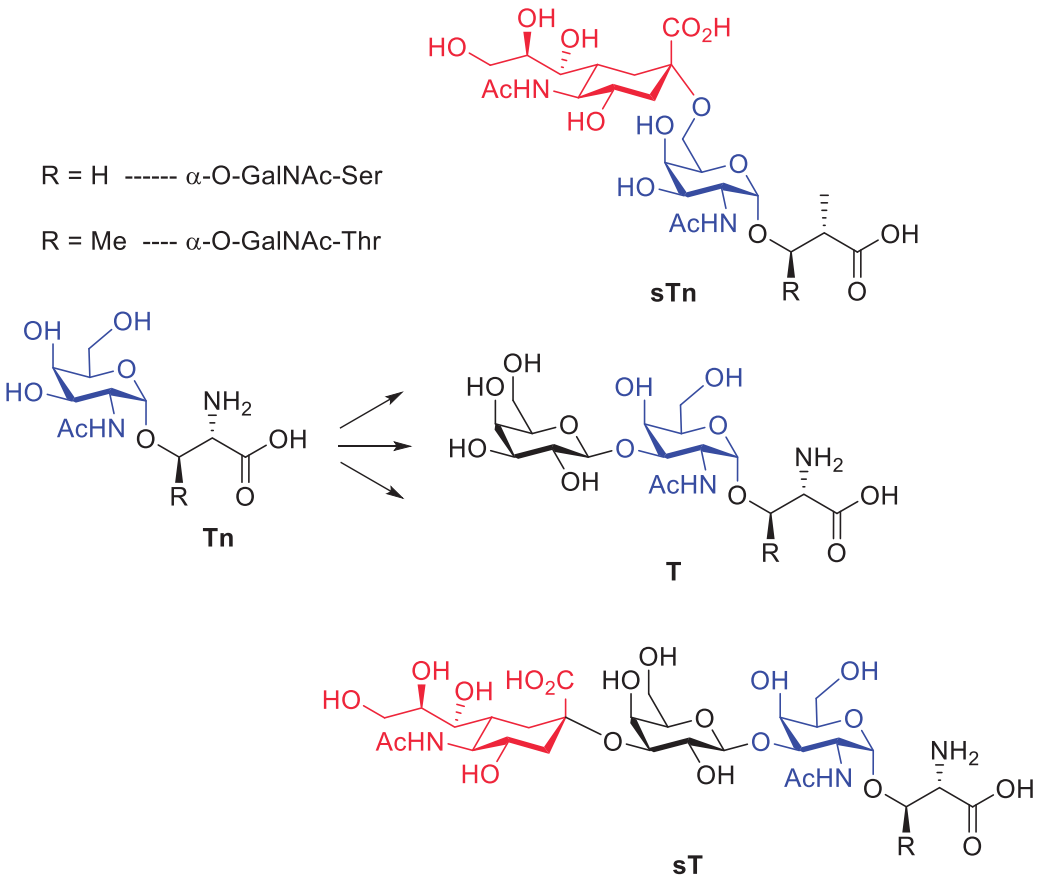



**Figura 3.** *Antígenos asociados a tumores formados por glicosilaminoácidos (TACAs) presentes en la MUC1.*

Los anticuerpos han jugado un papel muy importante en el estudio de la estructura y la biología de las mucinas y en la detección y caracterización de las glicoformas alteradas que se encuentran en el cáncer. Un avance importante en la elucidación de la estructura de las mucinas se produjo al clonar los genes que codifican sus proteínas centrales, pudiéndose definir de esta manera la estructura de las proteínas. La estrategia consistía en el desarrollo de anticuerpos específicos para dichas proteínas, los cuales se utilizaron para realizar una criba de las diferentes proteínas expresadas. Estos anticuerpos dieron la primera pista de que en los tejidos cancerosos existían diferencias en la glicosilación de las mucinas. La primera secuencia caracterizada siguiendo esta estrategia fue la de la mucina de membrana MUC1, para la cual se utilizó el anticuerpo SM3. La tinción

inmunohistoquímica con SM3 mostró que las células de tejido mamario canceroso se teñían fuertemente, al contrario que las células de tejido sano que lo hacían débilmente. Esto se explica por el hecho de que el acceso a los epítopos de la proteína que proviene de un tejido sano está impedido por la presencia de glicanos complejos. Los anticuerpos contra *O*-glicanos cortos han estado disponibles durante muchos años, siendo utilizados como biomarcadores de tumores.

Esta deficiencia en la glicosilación afecta al metabolismo celular. Debido a que los glicanos juegan un papel fundamental en el tráfico celular, los cambios en su estructura pueden ser importantes mediadores en el comportamiento metastásico. Uno de los glicanos que pueden encontrarse en las mucinas es el sialil-Lewis$^X$ (sLe$^X$) el cual es reconocido por selectinas. La expresión de este tipo de moléculas está asociada con la metástasis hematógena de algunos canceres como el de colon, de mama y de huesos.[19]

Los cambios en la glicosidación de las mucinas también pueden permitir interacciones con galectinas, las cuales están involucradas en la metástasis del cáncer de mama y colon.[20]

Además del potencial que posee la MUC1 en la diagnosis, también es posible su utilización como diana terapéutica. En las últimas décadas, el tratamiento inmunitario pasivo con anticuerpos monoclonales antitumorales (mAbs) se ha utilizado con éxito en ensayos clínicos.[21,22] Estos avances han arrojado luz sobre el desarrollo de vacunas terapéuticas contra el cáncer.

A pesar del potencial que tienen las vacunas antitumorales basadas en los TACAs, estos poseen algunos problemas. Al contrario que los péptidos y proteínas, al no unirse a la mayoría de los complejos de histocompatibilidad (MHCs) directamente, no son capaces de activar las células T por sí mismos,[23] lo cual es crucial para la inmunoterapia contra el cáncer. Se sabe que los péptidos son los principales agentes capaces de estimular a las células

T mediante su unión a los MHCs. Debido a esto, una estrategia muy común es conjugar los TACAs a péptidos o proteínas transportadoras que contengan epítopos capaces de estimular dichas células.[24–28]

A través de los años, diferentes tipos de proteínas como la albumina del suero bovino (BSA), la albumina del suero humano (HSA), la ovoalbúmina (OVA) y la toxina del tétano (TTox) se han utilizado como moléculas transportadoras en la síntesis de vacunas antitumorales. Kagan y colaboradores mostraron como la *keyhole limpet hemocyanin* (KLH) es uno de los transportadores más eficientes a la hora sintetizar vacunas terapéuticas contra el cáncer basadas en los TACAs.[29] Recientemente, la proteína CRM197 se ha convertido en otro candidato prometedor.[30] Además, también se ha demostrado que la utilización de distintos conectores tiene diferentes efectos en las propiedades inmunológicas de las moléculas conjugadas resultantes.

Un ejemplo de vacuna de tres componentes es la diseñada por el grupo del Prof. Boons, la cual ha sido capaz de generar respuesta inmune en ratones e incluso reducir el tamaño en los tumores mediante la incorporación de un fragmento agonista de los receptores tipo Toll (TLRs), un epítopo reconocido por las células T y un glicopéptido tipo MUC1 (Figura 4).[31]

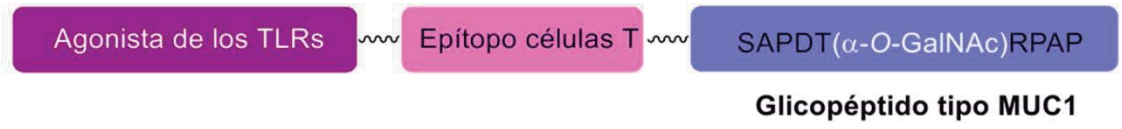



**Figura 4.** *Vacuna de tres componentes basada en MUC1.*

Otra estrategia posible es el empleo de nanopartículas como moléculas transportadoras. En este contexto, el grupo del profesor Huang describió la utilización de la cápside del virus del mosaico del caupí (CPMV) como

molécula transportadora inmunogénica de un derivado del antígeno Tn de Ser (Figura 5).[32]

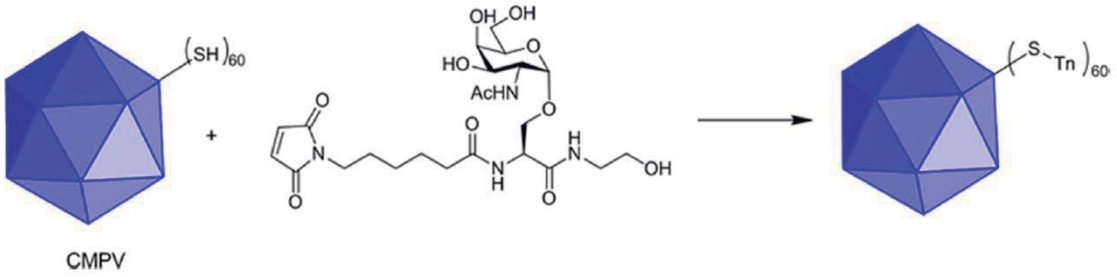



**Figura 5.** *Funcionalización del conjugado Tn-CMPV.*

El interés en las nanopartículas de oro (AuNP) ha crecido de manera notable, especialmente con respecto a sus aplicaciones en el tratamiento del cáncer y la inmunización mediante nanovacunas. El interés por este tipo de nanopartículas se debe a la capacidad que tienen de penetrar los vasos sanguíneos y las barreras de los tejidos, pudiendo dirigirse a una célula específica. Muy recientemente, Biswas y colaboradores sintetizaron conjugados de antígeno T unidos a AuNP a través de un conector de PEG. Los resultados obtenidos respaldan el uso de AuNP como agentes terapéuticos antitumorales, cuando son dirigidos contra líneas celulares que expresan lectinas específicas que interactúan con el antígeno T.[11]

El otro problema existente es la baja inmunogenicidad de los TACAs. A diferencia de agentes como virus y bacterias, las células cancerosas son células huésped endógenas. Aunque los TACAs están sobrexpresados en la superficie de las células tumorales, también están presentes en menor cantidad en las células sanas, por lo que son percibidos por el sistema inmunológico como autoantígenos. Por ello, debido a sus propiedades endógenas, la inmunotolerancia y la inmunosupresión se inducen más fácilmente. Una de las claves en la investigación de vacunas basadas en carbohidratos es evitar esa inmunotolerancia.

Este problema se puede solventar mediante la incorporación de antígenos de tipo TACAs no naturales. Además de poseer una mayor inmunogenicidad y biodisponibilidad, es posible el diseño de antígenos que presenten modificaciones capaces de mejorar la respuesta que tiene el sistema inmunológico por los derivados naturales.

En este contexto, una de las opciones posibles es la introducción de átomos de flúor en la estructura del carbohidrato. Esta resulta una buena idea debido la ausencia de enlaces C-F en la mayoría de los organismos, siendo además un enlace muy robusto. Siguiendo esta línea de investigación, el grupo de la profesora Hoffmann-Röder sustituyo varios grupos hidroxilos del antígeno T por átomos de flúor, consiguiendo de esta manera buenos resultados en la inmunización de ratones con vacunas que poseían dichos antígenos no naturales.[33,34]

Otra posibilidad es la de incorporar aminoácidos no naturales conformacionalmente restringidos. En este sentido, en nuestro grupo de investigación, se desarrolló una vacuna de tres componentes la cual posee en su secuencia el mimético del antígeno Tn (*O*-(α-D-GalNAc)-α-metilserina).[35]

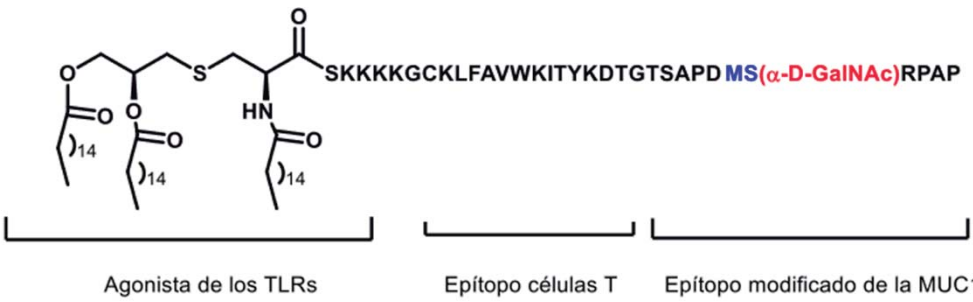



**Figura 6.** *Vacuna de tres componentes que incorpora en su estructura un fragmento de MUC1 no natural. MS= α-metilserina.*

Un último ejemplo de la utilización de estos miméticos no naturales en el desarrollo de vacunas terapéuticas es el llevado a cabo por el grupo de

investigación de la profesora Nativi.[36] En este trabajo la vacuna propuesta se compone de un *cluster* de cuatro moléculas miméticas del antígeno Tn y el epítopo peptídico inmunogénico OvaPADRE, conjugados todos ellos a la molécula transportadora ciclopeptídica RAFT (Figura 7).

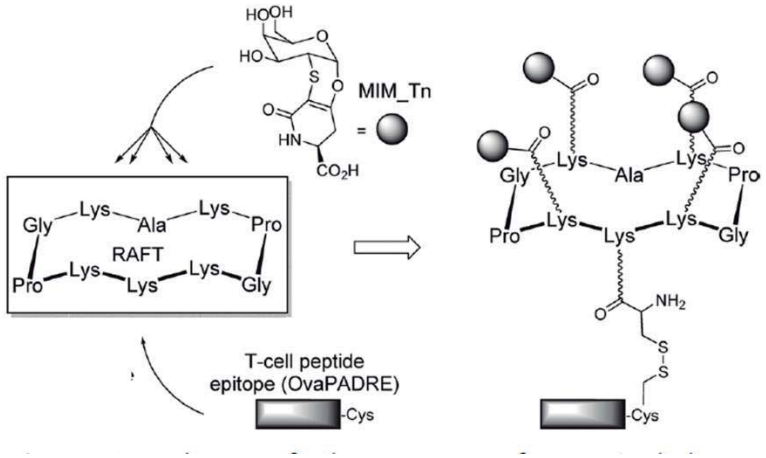



**Figura 7**. *Vacuna diseñada por el grupo de la profesora Nativi.*

Actualmente, ninguna de las vacunas terapéuticas basadas en MUC1 ha pasado la fase clínica III. A pesar de esto, el interés en la MUC1 con un fin terapéutico ha crecido significativamente en los últimos años, como lo demuestra el aumento en el número de ensayos clínicos basados en MUC1 iniciados en 2017 (Figura 8).[37]

Por otro lado, la eficacia de la vacuna debe estar ligada a la presentación de los antígenos. El análisis conformacional de estos derivados, tanto en el estado asociado como libre puede ayudar al desarrollo de vacunas más eficaces.

Estas dos líneas de investigación (derivados no naturales) y análisis conformacional de los antígenos tipo Tn serán los pilares básicos sobre los que se sustenta el presente trabajo.

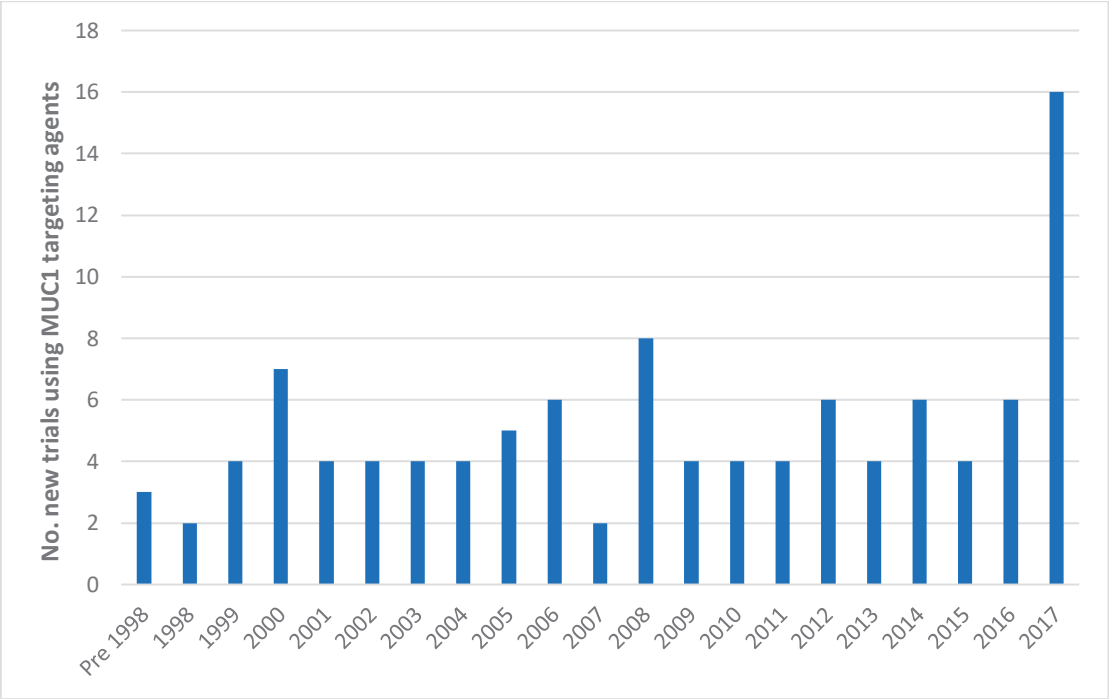



**Figura 8.** *Número de ensayos basados en la MUC1 en cada año*

## Bibliografía

(1) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.

(2) Taylor-Papadimitriou, J.; Burchell, J. M.; *Mucins and Cancer*, Future Medicine, **2013**.

(3) Hasnain, S. Z.; Wang, H.; Ghia, J.; Haq, N.; Deng, Y.; Velcich, A.; Grencis, R. K.; Thornton, D. J.; Khan, W. I. *Gastroenterology* **2010**, *138*, 1763-1771.

(4) Bergstrom, K. S. B.; Kissoon-Singh, V.; Gibson, D. L.; Ma, C.; Montero, M.; Sham, H. P.; Ryz, N.; Huang, T.; Velcich, A.; Finlay, B. B.; Chadee, K.; Vallance, B.A. *PLoS Pathog.* **2010**, *6*, e1000902.

(5) Hasnain, S. Z.; Evans, C. M.; Roy, M.; Gallagher, A. L.; Kindrachuk, K. N.; Barron, L.; Dickey, B. F.; Wilson, M. S.; Wynn, T. A.; Grencis, R. K.; Thornton, D. J. *J. Exp. Med.* **2011**, *208*, 893–900.

(6) Van der Sluis, M.; De Koning, B. A. E.; De Bruijn, A. C. J. M.; Velcich, A.; Meijerink, J. P. P.; Van Goudoever, J. B.; Büller, H. A.; Dekker, J.; Van Seuningen, I.; Renes, I. B.; Einerhand, A.W.C. *Gastroenterology* **2006**, *131*, 117–129.

(7) Velcich, A.; Yang, W.; Heyer, J.; Fragale, A.; Nicholas, C.; Viani, S.; Kucherlapati, R.; Lipkin, M.; Yang, K.; Augenlicht, L. *Science* **2002**, *295*, 1726–1729.

(8) Holmén Larsson, J. M.; Karlsson, H.; Sjövall, H.; Hansson, G. C. *Glycobiology* **2009**, *19*, 756–766.

(9) Taylor-Papadimitriou, J.; Burchell, J. M.; Miles, D.; Dalziel, M. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1999**, *1455*, 301–313.

(10) Taylor-Papadimitriou, J.; Burchell, J. M.; Plunkett, T.; Graham, R.; Correa, I.; Miles, D.; Smith, M. *J. Mammary Gland Biol. Neoplasia* **2002**, *7*, 209–221.

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

(12) Müller, S.; Goletz, S.; Packer, N.; Gooley, A.; Lawson, A. M.; Hanisch, F. G. *J. Biol. Chem.* **1997**, *272*, 24780–24793.

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

(14) Sell, S. *Hum. Pathol.* **1990**, *21*, 1003–1019.

(15) Hakomori, S.-I.; Zhang, Y. *Chem. Biol.* **1997**, *4*, 97–104.

(16) Taylor-Papadimitriou, J.; Epenetos, A. A. *Trends Biotechnol.* **1994**, *12*, 227–233.

(17) Gabius, H.-J. *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 1267–1276.

(18) Schuman, J.; Campbell, A. P.; Koganty, R. R.; Longenecker, B. M. *J.*

*Pept. Res.* **2003**, *61*, 91–108.

(19) Julien, S.; Ivetic, A.; Grigoriadis, A.; QiZe, D.; Burford, B.; Sproviero, D.; Picco, G.; Gillett, C.; Papp, S. L.; Schaffer, L.; et al. *Cancer Res.* **2011**, *71*, 7683–7693.

(20) Barrow, H.; Guo, X.; Wandall, H. H.; Pedersen, J. W.; Fu, B.; Zhao, Q.; Chen, C.; Rhodes, J. M.; Yu, L.-G. *Clin. Cancer Res.* **2011**, *17*, 7035–7046.

(21) Dougan, M.; Dranoff, G. *Annu. Rev. Immunol.* **2009**, *27*, 83–117.

(22) Weiner, L. M.; Surana, R.; Wang, S. *Nat. Rev. Immunol.* **2010**, *10*, 317–327.

(23) Danishefsky, S. J.; Allen, J. R. *Angew. Chemie Int. Ed.* **2000**, *39*, 836–863.

(24) Haurum, J. S.; Arsequell, G.; Lellouch, A. C.; Wong, S. Y.; Dwek, R. A.; McMichael, A. J.; Elliott, T. *J. Exp. Med.* **1994**, *180*, 739–744.

(25) Abdel-Motal, U. M.; Berg, L.; Rosén, A.; Bengtsson, M.; Thorpe, C. J.; Kihlberg, J.; Dahmén, J.; Magnusson, G.; Karlsson, K.-A.; Jondal, M. *Eur. J. Immunol.* **1996**, *26*, 544–551.

(26) Speir, J. A.; Abdel-Motal, U. M.; Jondal, M.; Wilson, I. A. *Immunity* **1999**, *10*, 51–61.

(27) Dudler, T.; Altmann, F.; Carballido, J. M.; Blaser, K. *Eur. J. Immunol.* **1995**, *25*, 538–542.

(28) Dengjel, J.; Rammensee, H.-G.; Stevanovic, S. *J. Mass Spectrom.* **2005**, *40*, 100–104.

(29) Kagan, E.; Ragupathi, G.; Yi, S. S.; Reis, C. A.; Gildersleeve, J.; Kahne, D.; Clausen, H.; Danishefsky, S. J.; Livingston, P. O. *Cancer*

*Immunol. Immunother.* **2005**, *54*, 424–430.

(30) Shinefield, H. R. *Vaccine* **2010**, *28*, 4335–4339.

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

(32) Miermont, A.; Barnhill, H.; Strable, E.; Lu, X.; Wall, K. A.; Wang, Q.; Finn, M. G.; Huang, X. *Chem. - A Eur. J.* **2008**, *14*, 4939–4947.

(33) Hoffmann-Röder, A.; Johannes, M. *Chem. Commun.* **2011**, *47*, 9903–9905.

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

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

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

(37) Taylor-Papadimitriou, J.; Burchell, J. M.; Graham, R.; Beatson, R. *Biochem. Soc. Trans.* **2018**, *46*, 659–668.

# Antecedentes



2

Aunque el diseño racional de las vacunas de tipo glicopeptídico se basa en estudios de RMN, rayos X y estudios computacionales realizados en los últimos 15 a 20 años, aún queda mucho por aprender sobre los factores que rigen la presentación de los antígenos y cuáles son los elementos estructurales necesarios para lograr interacciones óptimas antígeno-anticuerpo. Para avanzar en el diseño de vacunas, es necesario racionalizar los efectos estructurales que los carbohidratos tienen en el fragmento inmunogénico y en regiones distantes que puedan afectar a la disposición del antígeno. Es crucial determinar si el primer residuo GalNAc unido al esqueleto peptídico es responsable de la modificación de la estructura del péptido, o si los posteriores azúcares enlazados también podrían tener un impacto en la conformación de este y, en consecuencia, en la presentación del antígeno. También es importante poder determinar qué posiciones en MUC1 deben ser glicosiladas para aumentar la inmunidad específica.

En este apartado se van a detallar los estudios estructurales más destacados que se han realizado hasta la fecha sobre MUC1. Para ello, inicialmente, se presentarán los conceptos más relevantes relacionados con el análisis conformacional de péptidos y glicopéptidos. Seguidamente, se expondrán las diferencias más notables, encontradas hasta la fecha, en las preferencias conformacionales de los antígenos Tn de serina y treonina. A continuación, se mostrarán algunos de los miméticos del antígeno Tn sintetizados con anterioridad a la elaboración de esta Tesis, así como la implicación que tiene la sustitución del antígeno Tn natural por los mismos en la manera de presentar el antígeno al sistema inmune. Por último, se evaluará la sustitución de otros aminoácidos presentes en el epítopo de reconocimiento APDTRP por aminoácidos no naturales y cómo dicha modificación puede mejorar la afinidad presentada por la MUC1 frente al anticuerpo anti-MUC1 SM3.

## 2.1. Análisis conformacional de (glico)péptidos en disolución

El estudio conformacional de péptidos y glicopéptidos en disolución se suele llevar a cabo combinando experimentos de RMN con cálculos de dinámica molecular (DM).

En concreto, dentro de los experimentos de RMN, la técnica NOESY[1] (del inglés *nuclear Overhauser effect spectroscopy*), nos permite obtener información sobre distancias protón-protón promediadas para todas las posibles conformaciones que una determinada molécula puede adoptar en el tiempo de la adquisición del experimento.

Esto se debe a que la intensidad del pico de cruce NOE depende de la proximidad en el espacio de los núcleos estudiados; por ejemplo, dos hidrógenos, (r), así como el tiempo de correlación de la molécula ($\tau_c$) (Ecuación 1).[1]

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

**Ecuación 1.** *Dependencia de la intensidad de la señal NOE con la distancia H-H y el tiempo de correlación de la molécula.*

En este contexto, la técnica ROESY (del inglés *Rotating frame nuclear Overhauser effect correlation spectroscopy*) se puede utilizar de manera alternativa, ya que $\tau_c$, que a su vez depende de la masa molecular de la molécula, puede aparecer en un rango donde las señales NOE sean nulas o muy débiles. Los experimentos ROESY[2] tienen una dependencia diferente entre el tiempo de correlación y la constante de velocidad de relajación. De hecho, mientras que en los experimentos NOESY estas propiedades abarcan desde valores negativos hasta valores positivos, en la técnica ROESY, los valores son siempre positivos.

Además de estas técnicas, también se puede extraer información complementaria para el estudio conformacional a través de las constantes de acoplamiento protón-protón (J). Dichas constantes proporcionan información sobre ángulos diedros a través de la ecuación de Karplus y otras ecuaciones relacionadas.[3] En concreto, el análisis de las constantes de acoplamiento $^{3}J_{NH\text{-}NH}$ y $^{3}J_{NH,Ha}$ dan información acerca la conformación del esqueleto peptídico.

Tanto las distancias H-H como las constantes de acoplamiento experimentales se interpretan, con la ayuda de la MD.[4] Esta técnica computacional permite conocer las propiedades dinámicas de una molécula (o conjunto de moléculas) desde un punto de vista teórico. En este tipo de cálculos, los átomos son considerados esferas rígidas con una carga parcial fija, mientras que los enlaces son representados mediante muelles, asociando su orden de enlace a la constante elástica de dichos muelles.

Para llevar a cabo dichas simulaciones es necesario el empleo de un campo de fuerzas. Los campos de fuerzas son un conjunto de ecuaciones y de parámetros que describen la energía potencial del sistema objeto de estudio (Ecuación 2). Los campos de fuerzas empleados para el desarrollo de esta tesis han sido AMBER[5] (en concreto, ff14SB, parametrizado para péptidos y proteínas) y GLYCAM[6] (parametrizado para el análisis conformacional de carbohidratos y glicopéptidos).

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

$$\begin{aligned}
\text{Potential energy (V)} = & \sum_{bonds} K_{bond}\,(r - r_0)^2 && V_{\text{bonds}} \\
& + \sum_{angles} K_{angle}\,(\theta - \theta_0)^2 && V_{\text{angles}} \\
& + \sum_{dihedral} \frac{E'}{2}\,(1 + \cos n\varphi - \gamma) && V_{\text{dihedral}} \\
& + \sum_{m,n} \frac{q_m\, q_n}{\varepsilon\,(r_{mn}) r_{mn}} && V_{\text{electrostatic}} \\
& + \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) && V_{\text{Van der Waals}}
\end{aligned}$$

**Ecuación 2.** *Ecuación típica de un campo de fuerzas en DM en la que aparecen desglosados los distintos términos de energía potencial.*

Además, estos cálculos se llevan a cabo en presencia de moléculas de agua explícita. Si bien existen varios modelos descritos para emular las moléculas de agua, el modelo TIP3P[7,8] es uno de los más utilizados. Dicho modelo reproduce bastante bien las propiedades del agua, siendo el tiempo de computación empleado en su cálculo no excesivamente elevado.

Aunque en se pueden llevar a cabo cálculo de MD sin restricciones. En general, estos cálculos emplean los datos experimentales como restricciones promediadas en el tiempo (dando lugar a las llamadas MD-tar, del inglés *molecular dynamics with time averaged restraints*).[9] De esta forma, se reduce, en gran medida, la presencia de conformaciones "virtuales" (de alta energía), que pueden aparecer cuando se aplican constricciones a sistemas flexibles, como son los pequeños glicopéptidos. Además, el empleo de datos experimentales guía también al cálculo de MD, evitando así las desviaciones provocadas por los campos de fuerza, sobre todo cuando se estudian derivados no naturales.

Como resultado de las MD-tar se obtiene un conjunto de conformaciones de baja energía (llamado *ensemble*) que satisface los datos experimentales.

Otros parámetros importantes en el estudio conformacional de péptidos y glicopéptidos son los ángulos diedros. Con respecto a la cadena peptídica se definen el ángulo diedro Ψ, el cual describe la rotación en torno al enlace $C_\alpha$-$C_1$ y el ángulo diedro ϕ que tiene en cuenta la rotación respecto al enlace $C_\alpha$-$N_2$. En cuanto a la cadena lateral de los aminoácidos, el ángulo diedro $\chi^1$ describe la rotación entorno al enlace $C_\alpha$-$C_\beta$ (Figura 1).

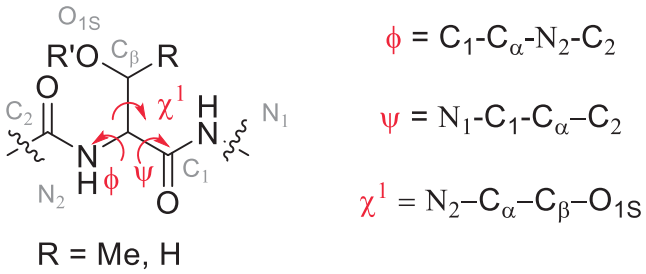



**Figura 1.** *Ángulos diedros relevantes para el análisis conformacional de péptidos y proteínas.*

Cuando se trabaja con glicopéptidos es necesaria la utilización de otros dos ángulos diedros: $\Psi_S$ y $\phi_S$ los cuales describen la rotación respecto a los enlaces $O_{1S}$-$C_\beta$ y $C_{1S}$-$O_{1S}$, respectivamente (Figura 2).[10]

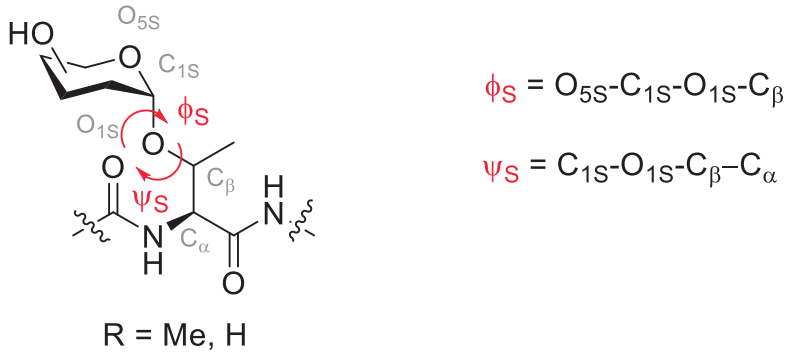



**Figura 2.** *Ángulos diedros relevantes para el análisis conformacional de glicopéptidos.*

En el caso de los péptidos, aunque pueden existir numerosas combinaciones de los ángulos diedros ϕ y Ψ, solo unas pocas de ellas presentan conformaciones energéticamente estables, dando lugar a estructuras secundarias como, por ejemplo, la α-hélice y la lamina β. Todos estos

ángulos se pueden representar mediante los diagramas de Ramachandran,[11] los cuales permiten ver de manera sencilla las disposición tridimensionales de menor energía adoptadas por los distintos aminoácidos de un péptido o proteína.

### 2.2. Estudio conformacional de los antígenos Tn en disolución

Nuestro grupo de investigación ha realizado un exhaustivo estudio conformacional en disolución acuosa del antígeno Tn, tanto de serina como de treonina.[12] Dicho estudio conformacional muestra que existen importantes diferencias estructurales entre α-*O*-GalNAc-Ser y α-*O*-GalNAc-Thr (Figura 3).

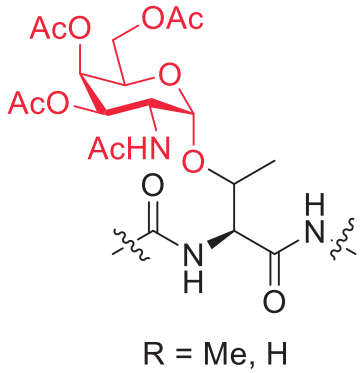



**Figura 3.** *Estructura de los antígenos Tn de Ser y Thr en forma de diamida para emular una cadena peptídica.*

En concreto, mientras que ambos derivados muestran un valor de $\phi_S$ próximo a 80°, lo que está de acuerdo con el efecto exo-anomérico[13,14] para el ángulo diedro $\psi_S$ existen notables diferencias. En el caso de α-*O*-GalNAc-Thr, $\psi_S$ adopta valores próximos a 120°, presentando un cierto grado de rigidez. Por el contrario, en el caso de α-*O*-GalNAc-Ser, $\psi_S$ posee más flexibilidad y se sitúa en torno a los 180°.

Como consecuencia, la disposición del residuo GalNAc en ambos derivados es muy diferente. Mientras que en el caso del antígeno Tn de serina el carbohidrato se orienta de manera paralela con respecto al aminoácido, en el

antígeno Tn de treonina lo hace de manera perpendicular. De esta forma, el GalNAc en la treonina queda próximo a la cadena peptídica, pudiendo interaccionar de una manera más eficiente con ella. Esta conformación está experimentalmente avalada por la presencia de un NOE entre los grupos NH del acetamido del GalNAc y del aminoácido Thr.

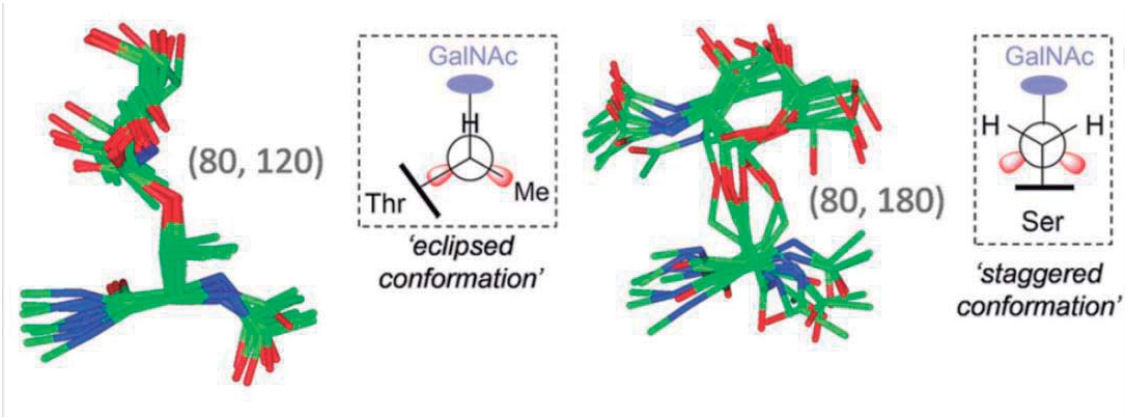



**Figura 4.** *Conformaciones principales en disolución encontradas para los antígenos Tn de Thr (derecha) y Ser (izquierda) en disolución acuosa, junto con las proyecciones de Newman de los enlaces glicosídicos.*[12,15,16]

Por otro lado, como puede apreciarse en la figura 4, la cadena lateral de los aminoácidos (representada por $\chi_1$) es más flexible para el derivado de serina, lo que está de acuerdo con el mayor valor observado experimentalmente para la constante de acoplamiento $J_{H\alpha,H\beta}$ ( 5.5 Hz para Tn-Ser y 2.5 Hz para Tn-Thr).

En la tabla 1 se recogen los diferentes valores para los ángulos diedros $\psi_S$, $\phi_S$ y $\chi_1$ calculados mediante MD-tar.

| | $\chi_1$ | $\psi_S$ | $\phi_S$ | Conformación $C_\beta$-$O_{1S}$ | Orientación del carbohidrato |
|---|---|---|---|---|---|
| α-*O*-GalNAc-Ser | 60° | 180° | ~80° | Alternada | Paralela |
| α-*O*-GalNAc-Thr | ~60° | 120° | 80° | Eclipsada | Perpendicular |

**Tabla 1.** Valores de los *ángulos diedros relevantes obtenidos por MD-tar para las conformaciones más relevantes en disolución acuosa para los antígenos Tn.*

Con respecto a la disposición de la cadena peptídica, en general, la *O*-glicosilación con GalNAc favorece conformaciones extendidas.[17] Esta característica es atribuida al impedimento estérico que el GalNAc ejerce sobre la cadena peptídica o a las interacciones especificas con el carbohidrato,[18,19] las cuales fijan la orientación del azúcar sobre la cadena peptídica impidiendo que esta se pliegue.[20] En este contexto, aunque varios grupos han propuesto la existencia de enlaces de hidrógeno entre el carbohidrato y el péptido,[17] nuestro grupo de investigación propuso la presencia de bolsillos hidrofóbicos en los antígenos Tn los cuales conectan ambas partes.[12,15,21] En el caso de Tn-Ser, dicho bolsillo se sitúa entre el NH del GalNAc y el carbonilo del aminoácido, mientras que en el caso de Tn-Thr, este se encuentra entre el NH del carbohidrato y el NH del aminoácido (Figura 5).

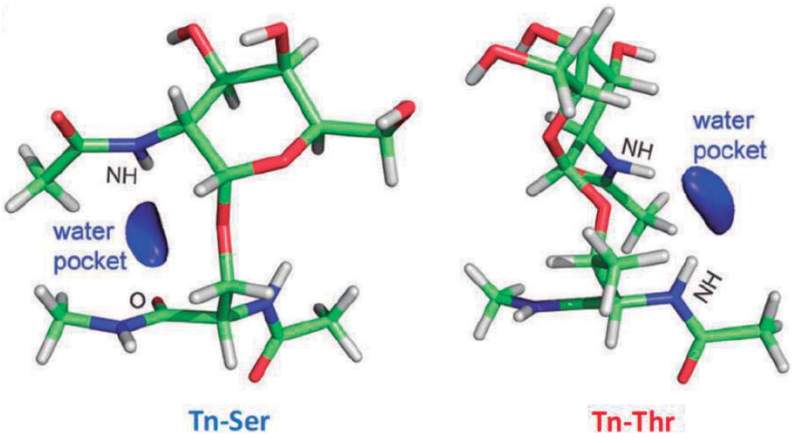



**Figura 5.** *Diferentes situaciones de los bolsillos hidrofílicos en los antígenos Tn.*

Es importante señalar que para el derivado Tn-Thr, nuestro grupo ha podido capturar estas moléculas de agua puente en dos estructuras de rayos X del complejo formado por dos antígenos Tn no naturales, presentes en la secuencia APDTRP, y el anticuerpo scFv-SM3. En concreto, estos derivados presentan un grupo fluroacetilo o difluoroacetilo en el GalNAc que es responsable de aumentar el carácter dador del NH del GalNAc, lo cual ha podido contribuir de forma decisiva a la captura de estas moléculas de agua puente (Figura 6).[16]

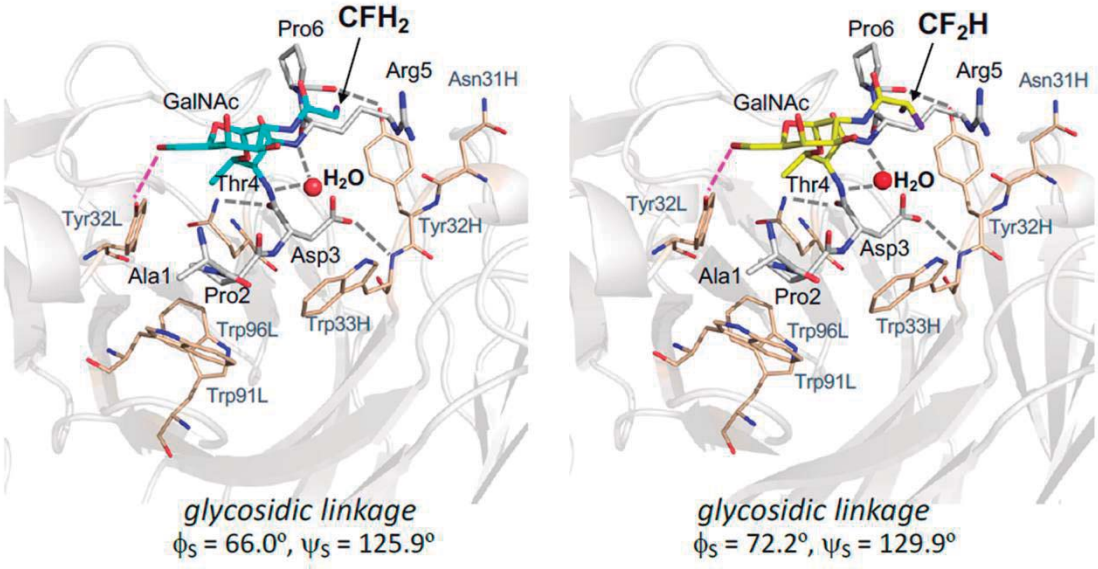



**Figura 6.** *Estructuras de rayos X de dos antígenos fluorados derivados de MUC1 y el anticuerpo scvF-SM3 en las que se pueden observar las moléculas de agua puente entre la parte peptídica y el GalNAc.*

## 2.3. Estudio conformacional de los antígenos Tn en fase gas

En el año 2018 el grupo de investigación llevo a cabo un estudio experimental de las conformaciones de los antígenos Tn-Ser y Tn-Thr en fase gas[16] y, por tanto, sin interferencia del disolvente, que se muestran en la Figura 7. Pare ello se empleó la espectroscopia de IR de alta resolución [*infrared ion-dip* (IRID)] en colaboración con el Dr. Cocinero (UPV).

Mediante ablación láser se consiguió pasar una pequeña fracción de estos dos compuestos a fase gas. La utilización de un láser fue fundamental para evitar los problemas de descomposición térmica. Además, el uso de los derivados con bencilamida facilita su posterior detección al incorporar un fluoróforo.

Tras el análisis de los espectros de infrarrojo, se obtuvieron, con ayuda de cálculos DFT, las conformaciones para ambos derivados capaces de satisfacer los espectros IR experimentales. Como puede observarse en la Figura 7, en ausencia de agua, ambos derivados adoptan la misma conformación para el enlace glicosídico. Conviene señalar que las conformaciones que explican los datos experimentales se corresponden con las de menor energía.

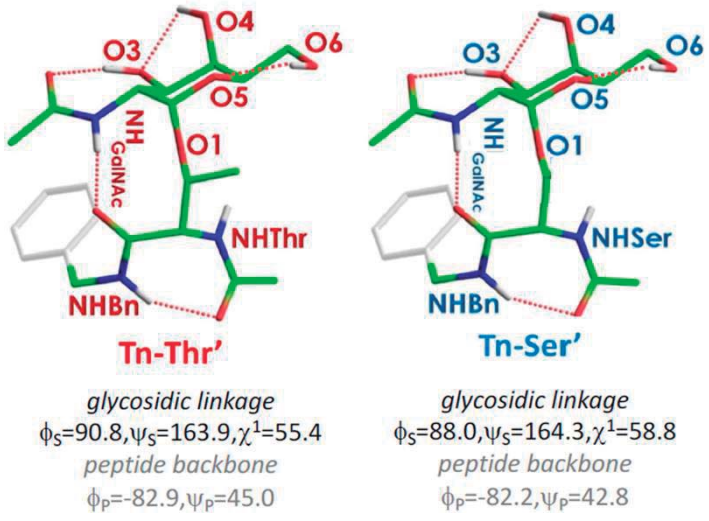



**Figura 7.** *Confórmeros de menor energía calculados para los derivados de Tn-Thr y Tn-Ser en fase gas junto con los enlaces de hidrogeno que las estabilizan.*

En concreto, la cadena peptídica adopta una disposición *γ-turn* estabilizada por la presencia de un fuerte enlace de hidrogeno entre el NH de la amida *C*-terminal y el carbonilo del acetamido *N*-terminal. A su vez, el grupo acetamido del GalNAc establece dos enlaces de hidrógeno: uno con el grupo O3H a través del carbonilo y el otro involucra al grupo NH y al carbonilo *C*-

terminal del aminoácido. Sorprendentemente, en fase gas, ambos derivados adoptan una conformación alternada para el enlace glicosídico, con un valor para el ángulo diedro $\psi_S$ en torno a 164°, lo que hace señalar la importancia de la presencia de moléculas de agua en la conformación que estos derivados adoptan en disolución. De hecho, cuando se realizan simulaciones MD de 200 ns introduciendo la primera esfera de hidratación, el derivado de serina retiene la conformación establecida en la fase gaseosa. Por el contrario, el derivado de treonina se ve forzado a rotar el enlace glicosídico para evitar la perturbación de la esfera de solvatación por el grupo metilo situado en la posición beta. En este caso el ángulo diedro $\psi_S$ adopta un valor próximo a 120°, estableciéndose, como ya se ha comentado, un bolsillo de agua entre el carbohidrato y el aminoácido (Figura 8).

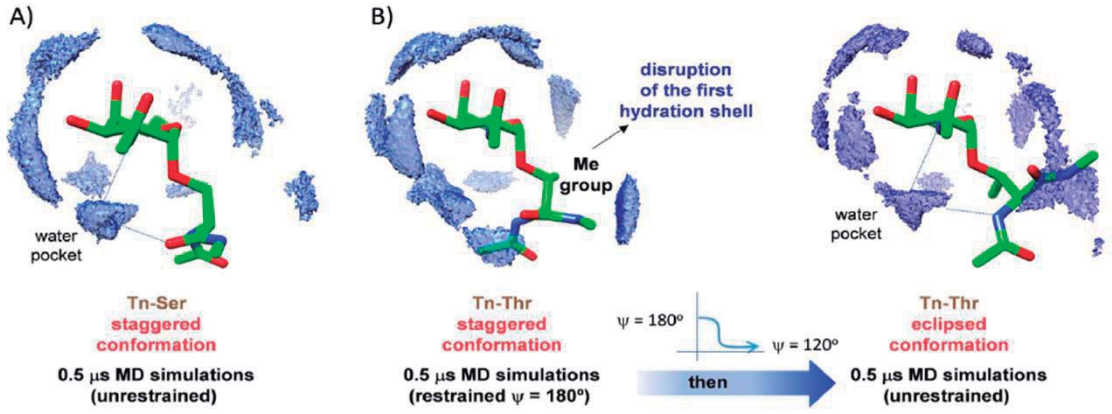



**Figura 8.** *Confórmeros mayoritarios en disolución para el derivado de Tn-Ser (A) y Tn-Thr (B) mediante MD-tar, junto con la primera esfera de hidratación. Como puede observase, cuando se elimina la restricción conformacional en el derivado de Thr, el ángulo diedro* $\psi_S$*, adopta rápidamente la conformación eclipsada.*

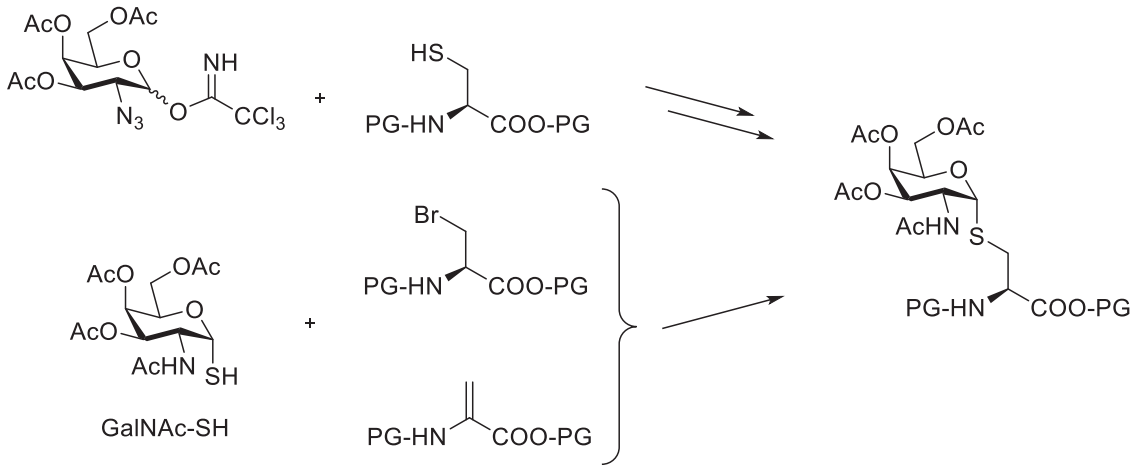
Todo esto indica que la presencia del grupo metilo en el derivado de treonina, es fundamental a la hora de explicar las diferencias conformacionales presentadas por los dos antígenos en disolución.

## 2.4. Modificaciones del antígeno Tn

Con motivo de ahondar en la importancia de la presencia del grupo metilo en el antígeno Tn-Thr, así como el interés que suscita la obtención de miméticos de éste para el desarrollo de nuevas vacunas más potentes; a continuación, se describen algunos de los derivados no naturales más relevantes sintetizados hasta la fecha.

### 2.4.1. Modificaciones en el aminoácido

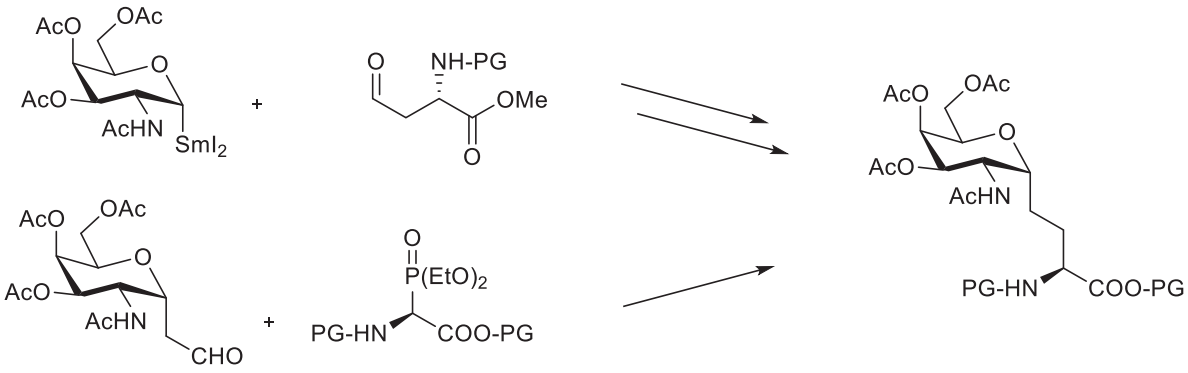
Una de las modificaciones más interesantes a la hora de la obtención de derivados no naturales del antígeno Tn es la sustitución del enlace O-glicosídico por un enlace *S*-glicosídico. En este contexto, existen tres estrategias principales que permiten la obtención del denominado antígeno sulfa-Tn. La primera ruta sintética parte de la cisteína convenientemente protegida y un carbohidrato activado con cloroacetimidato como dador glicosídico.[22] En las otras dos estrategias, el GalNAc-SH actúa como nucleófilo[23] sobre la β-bromoalanina[24,25] o la α,β-deshidroalanina[26–29] que actúan como electrófilos (Esquema 1).




**Esquema 1.** *Rutas empleadas para la síntesis del antígeno sulfa-Tn.*

Sin embargo, todos los derivados de este tipo son solo miméticos del antígeno Tn de serina, ya que el α-*S*-GalNAc-Thr no ha sido descrito hasta la fecha.

Por otro lado, también han sido descritos análogos de Tn-Ser *C*-glicosídicos. Una de las metodologías propuestas es la condensación de un complejo organometálico de samario con GalNAc con un precursor de aldehído derivado del aminoácido aspártico.[30] La otra ruta sintética propone como etapa clave una reacción Wittig-Horner entre un aldehído C-glicósido con un residuo de fosfoglicina[31] (Esquema 2).




**Esquema 2.** *Rutas sintéticas empleadas para la síntesis del antígeno Tn con enlaces C-glicosídicos*.

Otra modificación relevante consiste en la elongación de la cadena lateral del aminoácido. Los nuevos derivados pueden obtenerse mediante la sustitución de la serina por una hidroxinorleucina[32–34] u homoserina.[35] Con el mismo objetivo, puede emplearse también reacciones tipo tio-eno entre un tiocarbohidrato y un doble enlace introducido en la cadena lateral de Ser, Thr o Cys (Figura 9).[25]

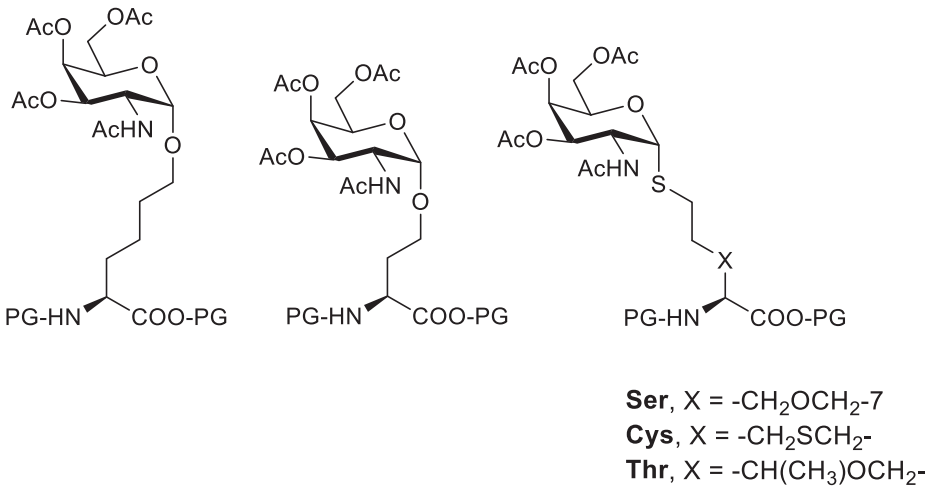
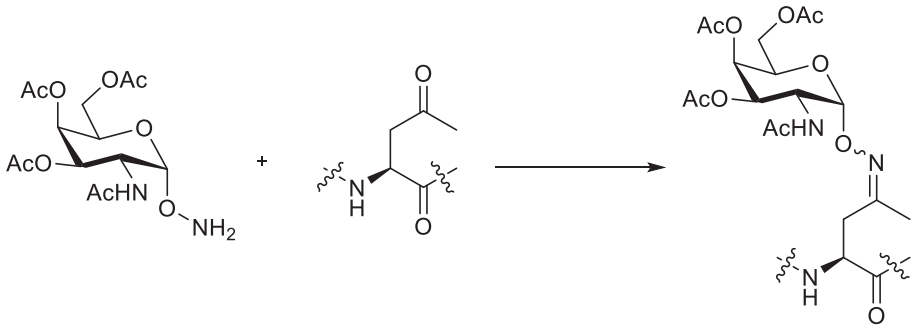


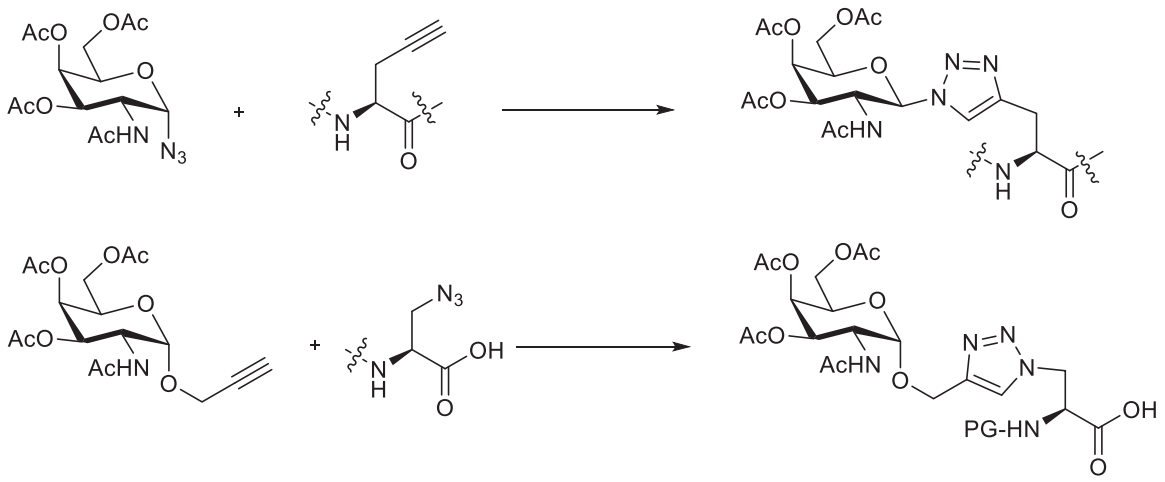

**Figura 9.** *Derivados no naturales del antígeno Tn que presentan cadenas laterales carbonadas con más de un átomo de carbono .*

La formación de derivados de oxima también se ha empleado para la unión del carbohidrato con el esqueleto peptídico, dando lugar a compuestos como los que se muestran en el esquema 3.[36]



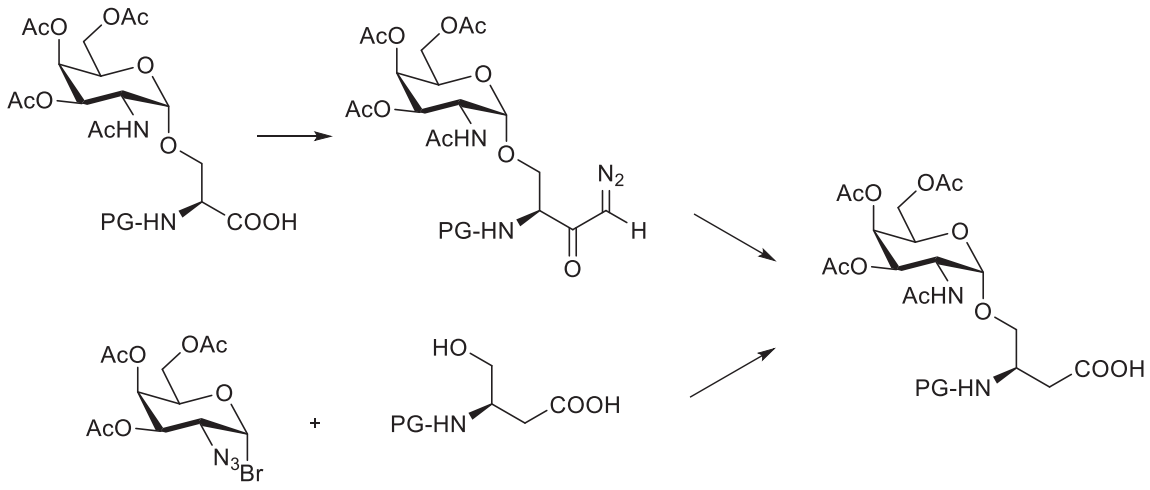

**Esquema 3.** *Análogo del antígeno Tn con un enlace glicosídico tipo oxima.*

Otra estrategia es la utilización del triazol como conector entre la parte peptídica y el carbohidrato. Esta modificación se puede realizar mediante la cicloadición 1,3-dipolar entre un alquino y un grupo azida, los cuales pueden estar presentes tanto en el aminoácido[37] como en el carbohidrato[38,39] (Esquema 4).



**Esquema 4.** *Rutas sintéticas para la obtención de análogos de Tn que incorporan triazol como conector entre el aminoácido y el carbohidrato.*

La incorporación de β-aminoácidos también ha dado lugar a derivados no naturales del antígeno Tn. A partir de a-aminoácidos, y empleando como etapa clave una reacción de Arndt-Eistert[40] se puede obtener el correspondiente β-aminoácido. Con el mismo objetivo, también es posible la utilización de derivados del ácido aspártico como aceptores glicosídicos[41,42] (Esquema 5).



**Esquema 5.** *Rutas sintéticas empleadas para la preparación de antígeno Tn que presentan en su estructura β-aminoácidos.*

### 2.4.2. Modificaciones en el carbohidrato GalNAc

En este contexto, una de las modificaciones más importantes es la sustitución del carbohidrato GalNAc por un mimético de tipo *sp*$^2$-iminocarbohidrato como el que se muestra en la figura 10. Es importante señalar que en la preparación de estos derivados, el iminocarbohidrato promueve una completa estereoselectividad hacia el anómero α en la reacción de glicosilación (Figura 10).[43]

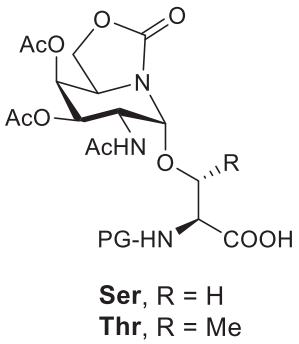



**Figura 10.** *Sp2-iminocarbohidrato mimético del antígeno Tn.*

Otra de las modificaciones posibles en el carbohidrato es la introducción de átomos de flúor. Con esta idea, se ha realizado la sustitución del grupo hidroxilo en posición 6 por un átomo flúor en el antígeno Tn de treonina.[44] Además, también se han realizado modificaciones en las cuales se ha sustituido el acetamido del carbohidrato por los grupos *N*-monofluoroacetil y *N*-difluoroacetil (Figura 11).[45]

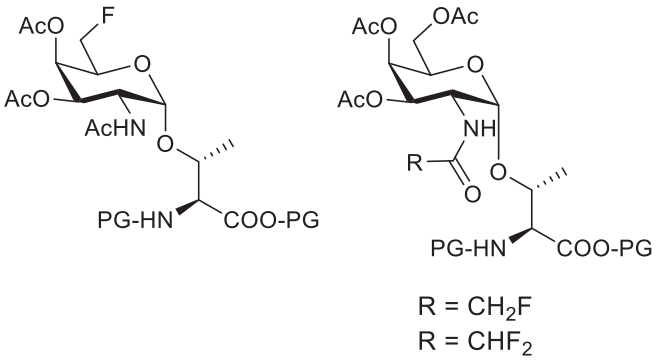



**Figura 11.** *Derivados fluorados del antígeno Tn de treonina.*

Por último, también se ha descrito la síntesis de un mimético del antígeno Tn, el cual es capaz de retener la conformación $^4C_1$ del carbohidrato mediante la utilización de la estructura tricíclica representada en la figura 12.[46,47]

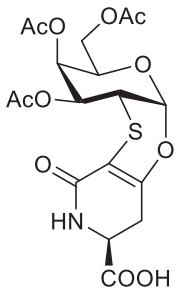



**Figura 12.** *Mimético del antígeno Tn conformacionalmente restringido.*

### 2.4.3. Modificaciones en la prolina del epítopo APD(α-*O*-GalNAc)TRP

Con motivo de mejorar la afinidad que este epítopo de reconocimiento posee frente a anticuerpos anti-MUC1, en nuestro grupo de investigación, se han llevado a cabo modificaciones en la prolina 2. Dicho residuo está implicado en interacciones tipo CH/π con el anticuerpo scFV-SM3. En este sentido, la polarización de los enlaces C-H directamente implicados en este tipo de interacciones favorecerán la formación del complejo antígeno-anticuerpo. Esto puede conseguirse sustituyendo determinados átomos de hidrógeno por flúor. Con esta idea, se escogieron la (4*S*)-4-fluoro-L-prolina (F-Pro) y la 4,4-difluoro-L-prolina (2F-Pro) (Figura 13).

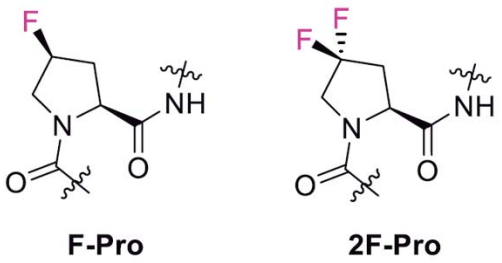



**Figura 13.** *Prolinas fluoradas escogidas para el estudio.*

Estos residuos no naturales de prolina se incorporaron a la secuencia de 20 aminoácidos de la MUC1, tanto glicosiladas con Tn (fP* y 2fP*), como sin glicosilar (fP y 2fP). Además, y con un fin comparativo, se sintetizaron los correspondientes derivados naturales y se midió la afinidad de estos compuestos frente al scvF-SM3 por experimentos de BLI (bio-layer interferometry) (Figura 14).

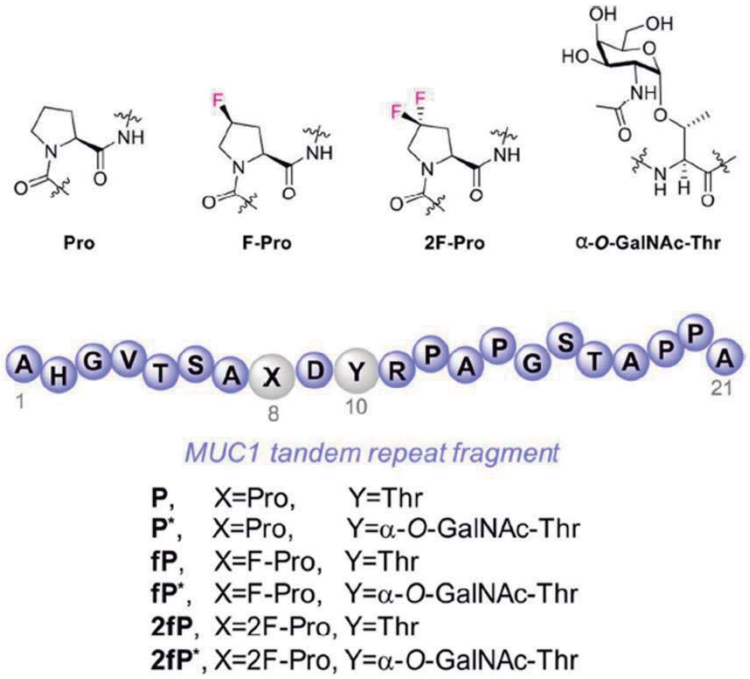



**Figura 14.** *Péptidos y glicopéptidos tipo mucina sintetizados y estudiados por BLI.*

Como se puede apreciar en la figura 15, la afinidad presentada por los derivados fluorados es mejor que la encontrada para los análogos naturales en todos los casos, tanto glicosilados como no glicosilados. Esta mejora se hace muy evidente en el caso de fP*, cuya constante de disociación $K_D$ es tres veces superior a la presentada por el derivado natural P*.[48]

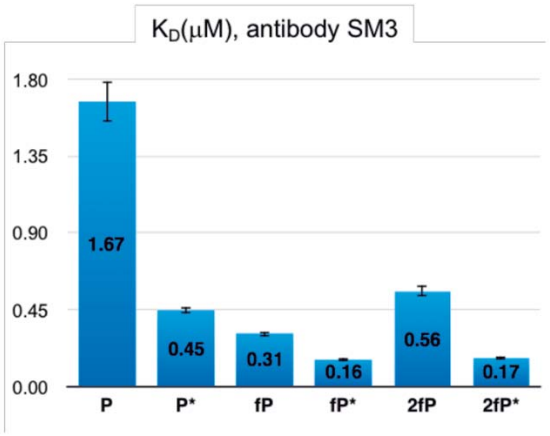



**Figura 15.** *Constantes de disociación ($K_D$) obtenidas mediante experimentos BLI. fP y 2fP hacen referencia al péptido de MUC1 con un residuo de (4S)-4-fluoro-L-prolina y de 4,4-difluoro-L-prolina, respectivamente.*

Los derivados fluorados *fP** y *2fP** demostraron ser excelentes candidatos a la hora de detectar anticuerpos anti-MUC1 en pacientes con cáncer de próstata. Estos derivados pueden, por tanto, contribuir significativamente al desarrollo de nuevas herramientas de diagnóstico del cáncer.

## 2.5. Bibliografia

(1) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **2008**, *71*, 4546–4553.

(2) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *63*, 207–213.

(3) Karplus, M. *J. Am. Chem. Soc*. **1963,** *85*, 2870-2871.

(4) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459–466.

(5) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

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

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

(8) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(9) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457–479.

(10) Mazeau, K.; Tvaroska, I. *Carbohydr. Res.* **1992**, *225*, 27–41.

(11) Anfinsen, C. B.; Anson, M. L.; Edsall, J. T.;Richards, F. M. *Advances in Protein Chemistry*, Elsevier, **1968**.

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

(13) Wiberg, K. B.; Bailey, W. F.; Lambert, K. M.; Stempel, Z. D. *J. Org. Chem.* **2018**, *83*, 5242–5255.

(14) Filloux, C. M. *Angew. Chem. Int. Ed.* **2015**, *54*, 8880–8894.

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

(16) 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-Oses, G.; Hurtado-Guerrero, R.; Cocinero, E. J.; Corzana, F. *J. Am. Chem. Soc.* **2018**, *140*, 9952–9960.

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

(18) Pratt, M. R.; Bertozzi, C. R. *Chem. Soc. Rev.* **2005**, *34*, 58–68.

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

(20) Coltart, D. M.; Royyuru, A. K.; Williams, L. J.; Glunz, P. W.; Sames, D.; Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Danishefsky, S. J.; Live, D. H. *J. Am. Chem. Soc.* **2002**, *124*, 9833–9844.

(21) Mallajosyula, S. S.; MacKerell, A. D. *J. Phys. Chem. B* **2011**, *115*, 11215–11229.

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

(23) Knapp, S.; Myers, D. S. *J. Org. Chem.* **2002**, *67*, 2995–2999.

(24) Thayer, D. A.; Yu, H. N.; Galan, M. C.; Wong, C.-H. *Angew. Chem., Int. Ed.* **2005**, *44*, 4596–4599.

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

(26) Galonic, D. P.; van der Donk, W. A.; Gin, D. Y. *Chem. Eur. J.* **2003**, *9*, 5997–6006.

(27) Zhu, Y.; van der Donk, W. A. *Org. Lett.* **2001**, *3*, 1189–1192.

(28) Aydillo, C.; Compañón, I.; Avenoza, A.; Busto, J. H.; Corzana, F.; Peregrina, J. M.; Zurbano, M. M. *J. Am. Chem. Soc.* **2014**, *136*, 789–800.

(29) 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. *Org. Lett.* **2016**, *18*, 2796–2799.

(30) Urban, D.; Skrydstrup, T.; Beau, J.-M. *Chem. Commun.* **1998**, 955–956.

(31) Röhrig, C. H.; Takhi, M.; Schmidt, R. R. *Synlett* **2001**, 1170–1172.

(32) Allen, J. R.; Harris, C. R.; Danishefsky, S. J. *J. Am. Chem. Soc*. **2001,** *123*, 1890-1897

(33) Keding, S. J.; Endo, A.; Danishefsky, S. J. *Tetrahedron* **2003**, *59*, 7023–7031.

(34) Keding, S. J.; Atsushi, E.; Biswas, K.; Zatorski, A.; Coltart, D. M.; Danishefsky, S. J. *Tetrahedron Lett.* **2003**, *44* (16), 3413–3416.

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

(36) Marcaurelle, L. A.; Shin, Y.; Goon, S.; Bertozzi, C. R. *Org. Lett.*

**2001**, *3*, (23), 3691-3694

(37) Miller, N.; Williams, G. M.; Brimble, M. A. *Org. Lett.* **2009**, *11*, 2409–2412.

(38) Lee, D. J.; Mandal, K.; Harris, P. W. R.; Brimble, M. A.; Kent, S. B. H. *Org. Lett.* **2009**, *11*, 5270–5273.

(39) Lee, D. J.; Harris, P. W. R.; Brimble, M. A. *Org. Biomol. Chem.* **2011**, *9*, 1621–1626.

(40) Karch, F.; Hoffmann-Röder, A. *Beilstein J. Org. Chem.* **2010**, *6*, 1–8.

(41) Norgren, A. S.; Norberg, T.; Arvidsson, P. I. *J. Pept. Sci.* **2007**, *13*, 717–727.

(42) Caputo, R.; Cassano, E.; Longobardo, L.; Palumbo, G. *Tetrahedron Lett.* **1995**, *36*, 167–168.

(43) 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.; García-Fernádez, J. M.; Ortiz-Mellet, C.; Peregrina, J. M. *Org. Lett.* **2016**, *18*, 3890–3893.

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

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

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

(47) Nativi, C.; Papi, F.; Roelens, S. *Chem. Commun.* **2019**, *55*, 7729–7736.

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

# Objetivos

3

## Objetivos

Debido a la importancia que tiene el acceso a nuevos miméticos del antígeno Tn que presenten un enlace *S*-glicosídico, se pretende desarrollar una metodología novedosa que permita la síntesis estereoselectiva de una amplia variedad de derivados de *S*-glicosilcisteína mediante el empleo de la adición sulfa-Michael doblemente estereoselectiva de tiocarbohidratos convenientemente protegidos sobre deshidroalaninas quirales. Algunos de estos compuestos serán derivatizados a su correspondientes *building blocks* con el fin de poder emplearlos en síntesis de péptidos en fase sólida.

Además, se realizará un estudio conformacional del mimético azufrado del antígeno Tn con serina (sulfa-Tn-Ser) mediante experimentos de RMN y cálculos de dinámica molecular.

Se estudiarán en detalle los elementos claves para el reconocimiento molecular de antígenos tipo MUC1 por el anticuerpo scFv-SM3. Para ello se intentarán conseguir diferentes estructuras cristalinas de los distintos complejos antígeno-anticuerpo con el fin de llevar a cabo un estudio mediante difracción de rayos X

Se pretende que dichos estudios puedan explicar las diferencias encontradas entre los antígenos Tn de Ser y Thr a la hora de ser reconocidos por anticuerpos anti-MUC1, pudiendo tener importantes implicaciones a la hora del diseño de nuevos anticuerpos y biosensores.

Con el objetivo de tener acceso a los *building blocks* de *S*-(α-D-GalNAc)-*S*Thr y *Se*-(α-D-GalNAc)-*Se*Thr, se pretende diseñar una ruta sintética que permita su obtención, con el fin de incorporarlos en secuencias peptídicas mediante el empleo de la técnica de síntesis de péptidos en fase solida.

Por último, se llevará a cabo la síntesis de una vacuna terapéutica que incorpore el aminoácido *S*-(α-D-GalNAc)-*S*Thr con el fin de realizar estudios inmunológicos *in vivo*.

# Análogos S-glicosidados del antígeno Tn



4

## 4.1. Introducción

Un gran número de glicoproteínas se encuentran en la superficie celular, en la cual, los oligosacáridos que forman parte de su estructura participan en un gran número de sistemas de reconocimiento molecular.[1–3] Para poder realizar el estudio de dichas interacciones a nivel atómico, es necesario, en primer lugar, llevar a cabo la síntesis de dichas glicoproteínas o pequeños glicopéptidos derivados de ellas.[4–9] En este contexto, la síntesis de proteínas modificadas, y en particular de *O*-glicoproteínas, es un área de rápida expansión en la química biologica.[10–12] En los glicopéptidos naturales, el residuo carbohidrato y la secuencia peptídica están unidos mediante un enlace *O*-glicosídico. Sin embargo, estos productos naturales presentan, en general, una baja estabilidad química y, en consecuencia, en muchos casos la utilización como agentes terapéuticos se ve obstaculizada. Debido a ello, varios grupos de investigación se han centrado en la síntesis de miméticos de glicopéptidos que incorporan un enlace glicosídico estable.[13,14] En concreto, los glicopéptidos unidos mediante enlaces *S*-glicosídicos, pueden evitar este problema debido a su mayor estabilidad química y resistencia enzimática.[15–19]

En la naturaleza podemos encontrar *S*-glicopeptidos provenientes de la *S*-glicosilación de la cisteína[20,21] en las bacteriocinas glicocina F[22,23] y sublancina,[24–26] las cuales, contienen una β-*N*-acetilglucosamina (β-GlcNAc) y una β-glucosa (β-Glc), respectivamente. Esta glicosilación es esencial para la actividad antimicrobiana que poseen estos compuestos.

Dada la relevancia actual de los *S*-glicopéptidos, se han descrito una gran variedad de metodologías de glicosilación,[20,21,27] en las que, generalmente, se utilizan carbohidratos convenientemente protegidos, que actúan como electrófilos, y derivados de cisteína como nucleófilos. Por otro lado, se pueden utilizar como nucleófilos derivados de tipo tiocarbohidrato[28] en la

adición conjugada 1,4 sobre la deshidroalanina (Dha) y deshidrobutirina (Dhb). Sin embargo, estas últimas estrategias suelen conducir a malos resultados en la diastereoselectividad.[29,30] La deshidroalanina es un residuo insaturado de gran interés biológico y sintético.[31] Ésta se puede encontrar en las etapas iniciales de la biosíntesis de los péptidos antibióticos que contienen lantionina, conocidos como lantibióticos.[32–34] Además, la deshidroalanina también es un precursor útil para llevar a cabo un gran número de modificaciones postraduccionales mediante la adición conjugada de tioles.[35–39] En este contexto, nuestro grupo de investigación ha descrito la síntesis de miméticos de lantionina llevando a cabo la adición sulfa-Michael asimétrica de derivados de la D-cisteína sobre deshidroalaninas quirales[40] (Figura 1).

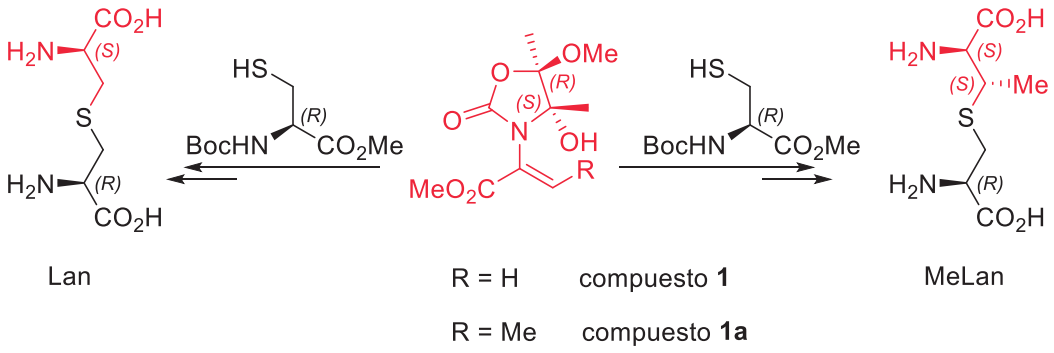



**Figura 1.** *Síntesis esteroselectiva de lantioninas a partir de un deshidroaminoácido quiral.*

## 4.2. Objetivos

Con el fin de tener acceso a los *building blocks* de glicosilaminoácidos que incorporan un enlace *S*-glicosídico (*S*-glicosilcisteínas) y poder incorporar estos motivos en secuencias peptídicas, se propuso el diseño de una metodología sintética basada en la reacción de adición tipo sulfa-Michael[41] asimétrica de distintos tiocarbohidratos, adecuadamente protegidos, como

los que se muestran en la figura 2, sobre la deshidroalanina quiral **1,** anteriormente comentada.

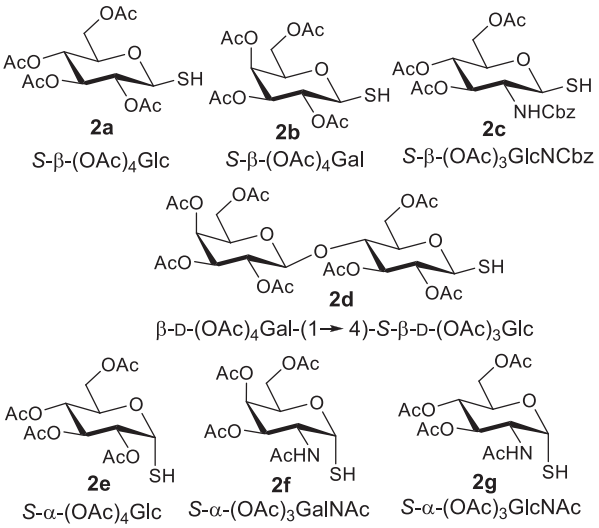



**Figura 2.** *Tioazúcares usados como nucleófilos en las reacciones de adición de Michael.*

Además, se pretende llevar a cabo el estudio conformacional en disolución del derivado **3f** (Figura 3), el cual es un análogo azufrado de Tn-Ser, y compararlo con el antígeno Tn-Ser natural. Este estudio se realizará mediante el empleo de dinámica molecular con restricciones promediadas en el tiempo (MD-tar).

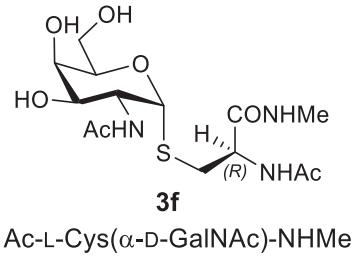


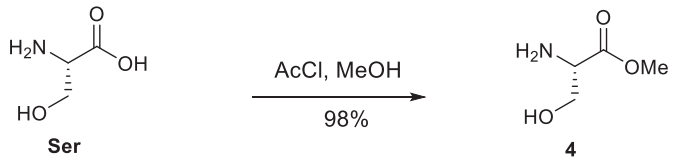

**Figura 3.** *Mimético azufrado del antígeno Tn- Ser.*

Por otro lado, y con el fin de conocer el epítopo de reconocimiento del derivado **3f** por el anticuerpo scFv-SM3, se realizarán medidas de RMN basadas en STD (*Saturation-Transfer Difference*).

## 4.3. Discusión de resultados

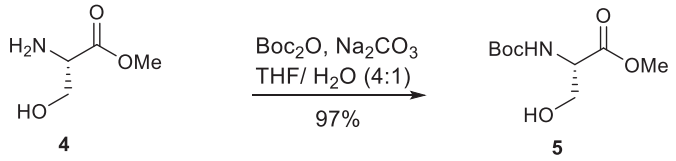
### 4.3.1. Síntesis mediante adiciones S-Michael estereoselectivas

Inicialmente, se procedió con la síntesis de la deshidroalanina quiral **1**. Para ello, se utilizó la estrategia diseñada por el grupo de investigación con anterioridad.[40,42,43] Así, partiendo del aminoácido comercial L-serina se formó el correspondiente éster metílico mediante la esterificación de *Fischer*, la cual consiste en la utilización de cloruro de acetilo y metanol, generando *in situ* cloruro de hidrógeno. De esta manera se consigue catalizar la reacción del metanol con el ácido carboxilo del aminoácido y así poder obtener de manera cuantitativa el éster metílico **4** (Esquema 1).



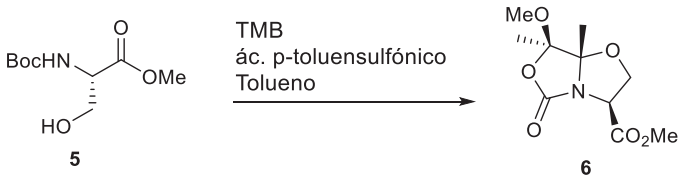

**Esquema 1.** *Esterificación de Fischer del aminoácido L-serina.*

Posteriormente, se trató el compuesto **4** con anhídrido de Boc ($Boc_2O$) en presencia de $Na_2CO_3 \cdot H_2O$ en una mezcla THF/$H_2O$ 4:1, obteniéndose tras su purificación por cromatografía de gel de sílice, el compuesto **5** (Esquema 2).



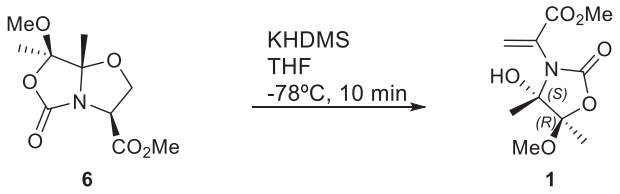

**Esquema 2.** *Síntesis del derivado de serina **5** convenientemente protegido.*

Una vez se tuvo el aminoácido convenientemente protegido, se le hizo reaccionar con 2,2,3,3-tetrametoxibutano (TMB) en presencia de ácido *p*-toluensulfónico como catalizador en reflujo de tolueno durante tres horas, dando el *N,O*-acetal bicíclico **6** con un rendimiento moderado del 75%, tras su purificación en columna cromatográfica (Esquema 3).




**Esquema 3.** *Síntesis del N,O-acetal bicíclico* ***6****.*

El tratamiento del compuesto **6** con bis(trimetilsilil)amiduro de potasio (KHDMS) en THF como disolvente a una temperatura de -78°C genera la deshidroalanina quiral objetivo (**1**) (Esquema 4).




**Esquema 4.** *Obtención de la deshidroalanina quiral* ***1****.*

Análogamente, se preparó la deshidroalanina quiral enantiómera **ent-1** partiendo en este caso de D-Ser (Figura 4). La síntesis se completó con la preparación de los correspondientes derivados con ésteres *terc*-butilicos (**1'** y **ent-1')** y del éster bencílico **1''** que aparecen en la figura 4.

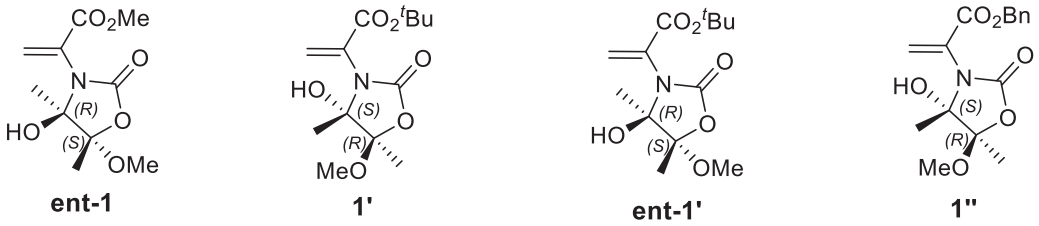
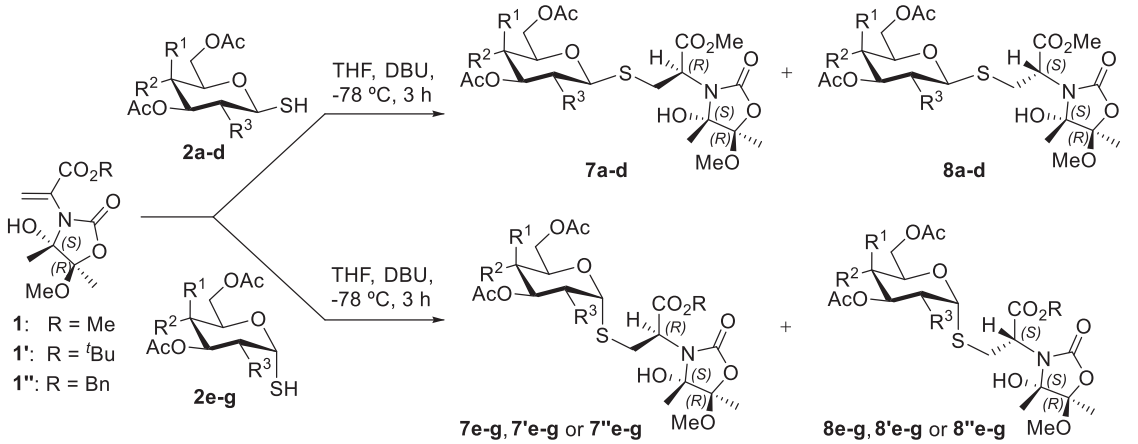



**Figura 4.** *Deshidroalaninatos quirales análogos.*

Una vez que se obtuvieron las deshidroalaninas quirales, de forma enantiopura, se llevó a cabo la adición de tipo *S*-Michael con los tiocarbohidratos anteriormente mencionados y mostrados en la figura 2, α- y β-tioazucares, tanto monosacáridos como disacáridos, junto con 2-amino-2-desoxi-D-tiocarbohidratos *N*-sustituidos, como los compuestos **2f** y **2g**.

Todos estos tiocarbohidratos fueron preparados de acuerdo a los procedimientos descritos en la literatura.[44–54]

El método general para llevar a cabo estas adiciones nucleófilas, consistió en la utilización de una base orgánica impedida, como es la diazabiciclo[5.4.0]-undec-8-eno (DBU), utilizando THF como disolvente a una temperatura de reacción de -78°C (condiciones cinéticas). Cabe destacar, que bajo estas condiciones no se observa mutarrotación de los tiocarbohidratos.[55]



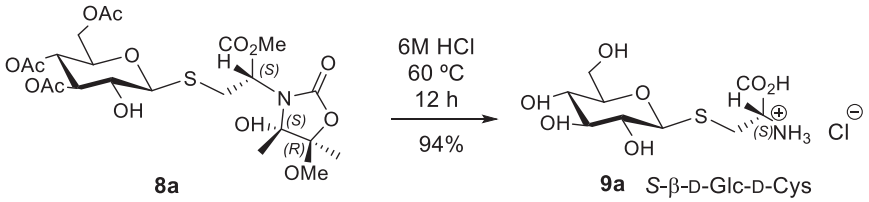

| entrada | Aceptor Michael | R | Tioazúcar | $R^1$ | $R^2$ | $R^3$ | Rdto (%) | Aducto Michael | (dr) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Me | 2a | H | OAc | OAc | 84 | 7a/8a | >5:95 |
| 2 | 1 | Me | 2b | OAc | H | OAc | 82 | 7b/8b | >5:95 |
| 3 | 1 | Me | 2c | H | OAc | NHCbz | 56 | 7c/8c | >5:95 |
| 4 | 1 | Me | 2d | H | β-Gal | OAc | 78 | 7d/8d | >5:95 |
| 5 | 1 | Me | 2e | H | OAc | OAc | 79 | 7e/8e | 23/77 |
| 6 | 1 | Me | 2f | OAc | H | NHAc | 67 | 7f/8f | >5:95 |
| 7 | 1 | Me | 2g | H | OAc | NHAc | 80 | 7g/8g | >5:95 |
| 8 | 1' | $^t$Bu | 2f | OAc | H | NHAc | 86 | 7'f/8'f | >5:95 |
| 9 | 1' | $^t$Bu | 2g | H | OAc | NHAc | 74 | 7'g/8'g | >5:95 |
| 10 | 1'' | Bn | 2g | H | OAc | NHAc | 71 | 7''g/8''g | 15:85 |

**Tabla 1.** *Reacciones S-Michael ensayadas con los tioazúcares **2a-2g** y la deshidroalanina **1**.*

Es importante destacar que, en general, la mayoría de los tiocarbohidratos utilizados dieron buenos rendimientos y que la diastereoselectividad

obtenida fue excelente en casi todos los casos, con una relación diastereomérica, evaluada mediante $^{1}$H RMN, mayor de un 95:5, con la excepción de los derivados **2e**[56] (entrada 5 en la Tabla 1) y **2g** (entrada 10 en la Tabla 1) en la que las relaciones diastereoméricas fueron 23/77 y 15/85, respectivamente. Para el resto de los compuestos, la purificación por columna cromatográfica de los crudos de reacción proporcionó un único estereoisómero.

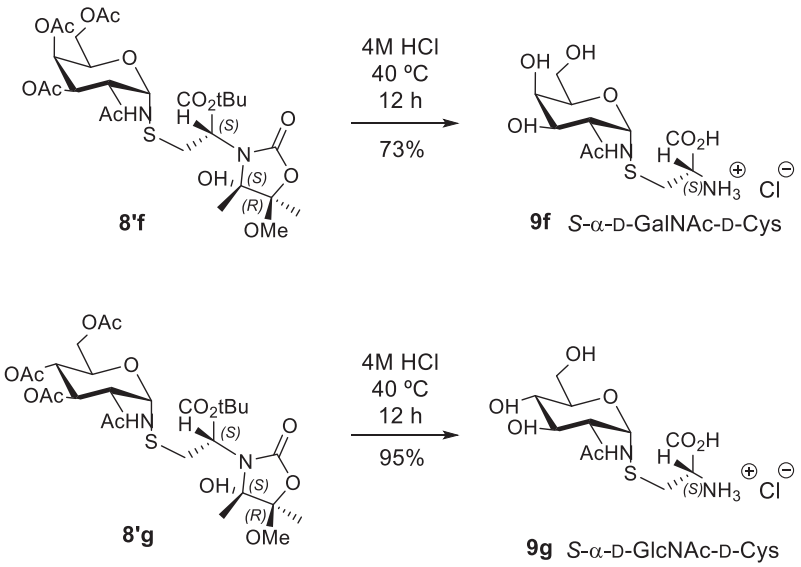
En el caso del producto **8a**, el cual es un ejemplo representativo de los derivados β-glicosídicos, la configuración absoluta del nuevo estereocentro formado fue determinada mediante la derivatización de este compuesto al clorhidrato correspondiente del tioglicosilaminoácido *S*-β-D-Glc-D-Cys. Esta transformación se realizó mediante hidrólisis ácida, utilizando una disolución de HCl 6M a 60 °C (Esquema 5).




**Esquema 5.** *Síntesis del glicosilaminoácido **9a**.*

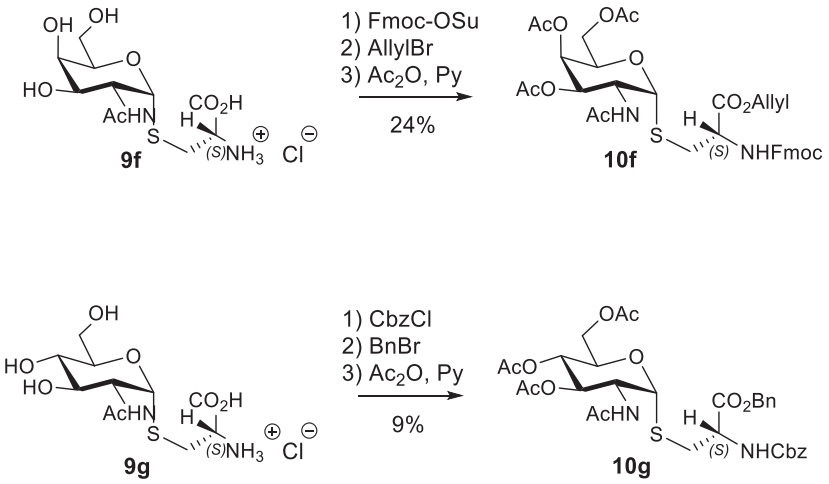
De esta manera, se pudo determinar el valor de rotación óptica especifica de dicho clorhidrato, siendo +43, y que tras compararse con el encontrado en la literatura para el compuesto S-β-D-Glc-L-Cys ($[\alpha]^{20}_{D}$ = –40), permitió asignar como *S* la configuración absoluta del nuevo estereocentro creado tras la reacción de Michael. De manera similar, se llevó a cabo la hidrólisis ácida los derivados **8f** y **8g**. Sin embargo, en estas condiciones los grupos acetamido de dichos compuestos también se hidrolizan parcialmente, lo que dificulta de manera considerable la purificación de los compuestos de interés. Este problema se pudo solventar mediante la utilización de unas condiciones de hidrólisis ácida más suaves sobre los derivados *terc*-butilicos

**8f'** y **8g'**, disminuyendo casi en su totalidad la hidrólisis de los grupos acetamido. Los compuestos **9a**, **9f** y **9g**[57] no se habían sintetizado hasta la realización de este trabajo (Esquema 6).



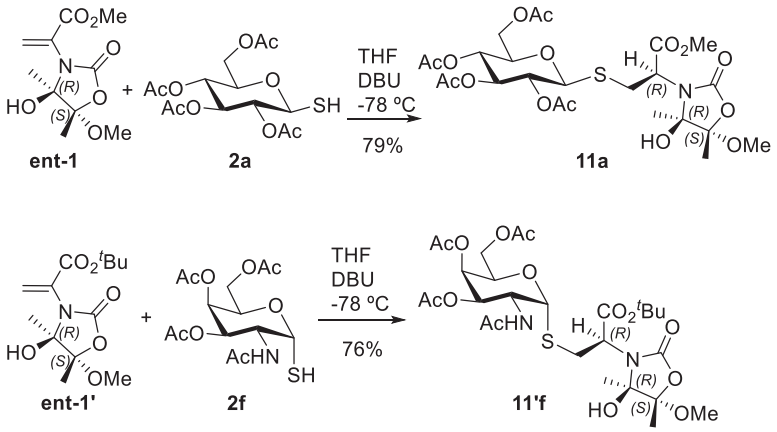

**Esquema 6.** *Síntesis de los glicosilaminoácidos **9f** y **9g**.*

De hecho, es importante destacar que esta es la primera vez que se describen este tipo de glicoconjugados en los que tanto el carbohidrato como la cisteína pertenecen ambos a la serie D. Teniendo en cuenta que tampoco están descritos los compuestos *S*-α-D-GalNAc-D-Cys **9f** y *S*-α-D-GlcNAc-D-Cys **9g** y con el fin de determinar la configuración absoluta de los aductos de Michael, se procedió a la derivatización de dichos compuestos a sus análogos protegidos **10f** y **10g**, respectivamente (Esquema 7). Estas transformaciones involucran tres pasos: protección del grupo amino con Fmoc, para el compuesto **10f**, y de Cbz, en el caso **10g**; derivatización de los grupos ácido al éster bencílico en **10g** y éster alílico en **10f** y la acetilación de los grupos hidroxilo en ambos casos. La comparación de los datos espectroscópicos de estos compuestos con los encontrados en la bibliografía[16,58,59] 5-α-D-$(OAc)_3$GalNAc-Fmoc-L-Cys-OAlil y 5-α-D-$(OAc)_3$GalNAc-Cbz-L-Cys-OBn permitieron confirmar que los nuevos estereocentros generados tras la adición de Michael también poseen una configuración absoluta *S*.



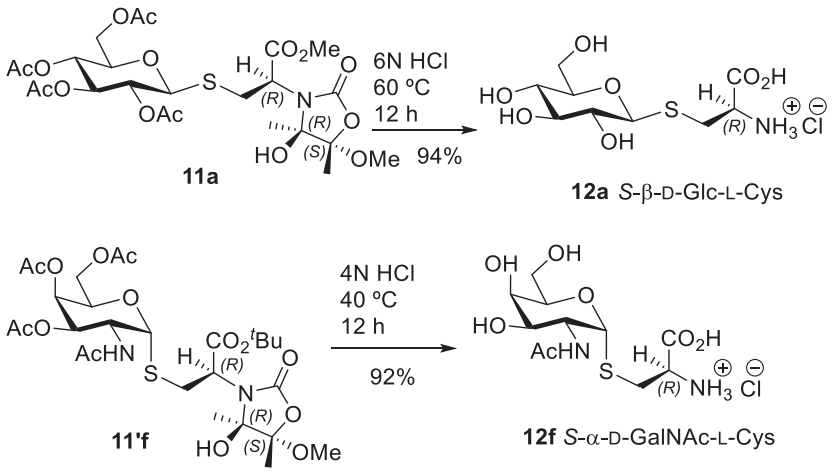

**Esquema 7.** *Derivatización de los aductos de Michael **9f** y **9g** a los esteres **10f** y **10g**, respectivamente.*

Con el fin de tener acceso a glicopéptidos que presenten el aminoácido de la serie L, se llevó a cabo la adición conjugada de Michael con los tiocarbohidratos **2a** y **2f** y los deshidroalaninatos **ent1** y **ent1',** respectivamente, utilizando las mismas condiciones anteriormente descritas (Esquema 8). Los mejores resultados proporcionaron los compuestos **11a** y **11'f** con total diasteroselectividad y unos rendimientos tras columna cromatográfica del 79% y 76%, respectivamente.



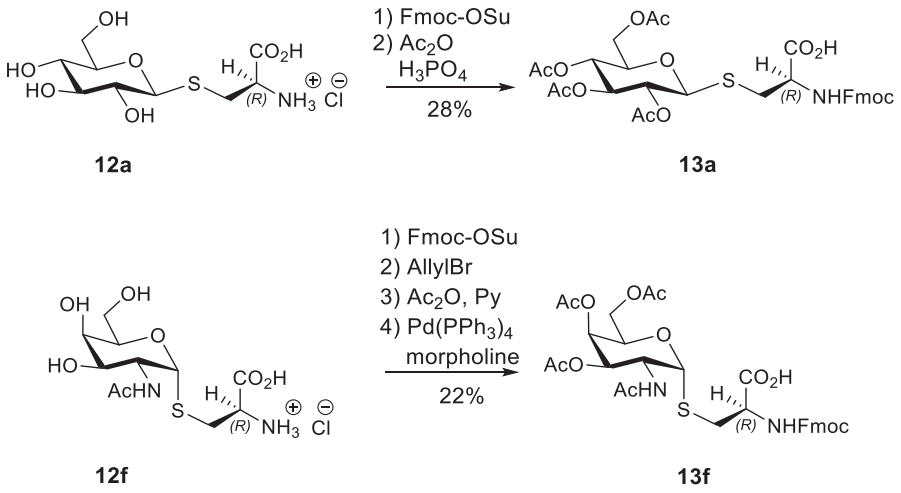

**Esquema 8.** *Síntesis de los aductos de Michael **11a** y **11'f**.*

Estos derivados fueron hidrolizados a los correspondientes clorhidratos *S*-β-D-Glc-L-Cys **12a**[57] y *S*-α-D-GalNAc-L-Cys **12f** (Esquema 9).




**Esquema 9.** *Síntesis de S-β-D-Glc-L-Cys y S-α-D-GalNAc-L-Cys.*

Con el fin de poder incorporar estos glicosilaminoácidos en secuencias peptídicas mediante la técnica de síntesis de péptidos en fase sólida (SPPS), se procedió a la derivatización de los mismos a los correspondientes *"building blocks"* **13a** y **13f**. Para ello, primero se protegió el grupo amino del compuesto **12a** con Fmoc−OSu en medio ligeramente básico. Posteriormente, se llevó a cabo la acetilación de los grupos hidroxilos del carbohidrato mediante la utilización de $Ac_2O$ en presencia de ácido fosfórico, obteniéndose de esta manera el *"building block"* **13a**.[26,60] Para la obtención del compuesto **13f** fue necesario el empleo de cuatro etapas: (a) protección del grupo amino del aminoácido con Fmoc−OSu, (b) la transformación del ácido carboxílico en el correspondiente éster alílico, (c) la acetilación[59] de los grupos hidroxilos del carbohidrato mediante el empleo de $Ac_2O$/Py y (d) la desprotección del éster alílico anteriormente formado mediante el empleo de morfolina y $Pd(PPh_3)_4$ como catalizador (Esquema 10).



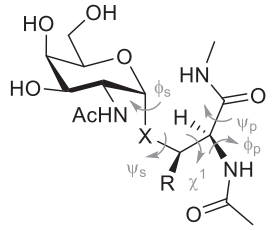

**Esquema 10.** *Síntesis de los "building blocks" **13a** y **13f** para su utilización en SPPS.*

### 4.3.2. Análisis conformacional en disolución acuosa del compuesto **3f**

Como se ha comentado anteriormente, en nuestro grupo de investigación, se había realizado un estudio conformacional de los análogos peptídicos Ac-L-Ser(α-D-GalNAc)-NHMe y Ac-L-Thr(α-D-GalNAc)-NHMe en agua utilizando experimentos de RMN y cálculos de MD. En dicho estudio se observó que existen moléculas de agua puente entre el residuo carbohidrato y el aminoácido glicosilado que contribuyen a la estabilización de conformaciones extendidas de este último y que dichas conformaciones pueden estar relacionadas con importantes implicaciones biológicas. Además, los derivados Ac-L-Thr/Ser(α-D-GalNAc)-NHMe son particularmente importantes debido a que incorporan la subestructura del antígeno Tn, el cual se encuentra sobrexpresado en células tumorales (Figura 3) y, en consecuencia, podrían ser utilizados en diseño de vacunas terapéuticas contra el cáncer .[61,62]

Para poder entender la actividad biológica de dichas subestructuras es muy importante conocer sus preferencias conformacionales en disolución acuosa. El modo en el que el carbohidrato es presentado resulta crucial a la hora de

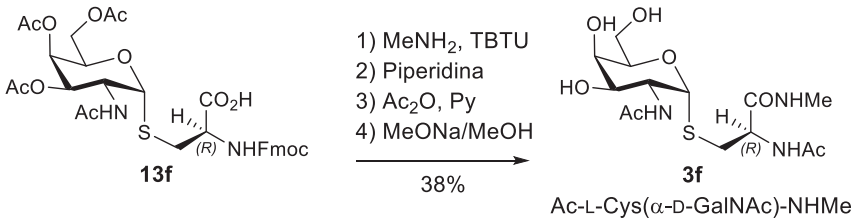
llevarse a cabo el reconocimiento molecular de dichos antígenos. En este contexto, se llevó a cabo un estudio comparativo de las preferencias conformacionales en agua de las moléculas Ac-L-Ser(α-D-GalNAc)-NHMe, Ac-L-Thr(α-D-GalNAc)-NHMe y Ac-L-Cys(α-D-GalNAc)-NHMe (**3f**) con motivo de determinar las posibles diferencias inducidas por la presencia del átomo de azufre en el enlace glicosídico (Figura 5).




X = O, R = Me: Ac-L-Thr(α-D-GalNAc)-NHMe
(*Antígeno Tn de Thr*)
X = O, R = H: Ac-L-Ser(α-D-GalNAc)-NHMe
(*Antígeno Tn de Ser*)
X = S, R = H: Ac-L-Cys(α-D-GalNAc)-NHMe
(compuesto **3f**)

**Figura 5.** Estructura de los *antígenos Tn de Ser, Thr y del compuesto **3f***.

Debido a ello, se sintetizó el compuesto objetivo **3f** a partir del *"building block"* **13f** (Esquema 11). Primeramente, se transformó el ácido carboxílico en la correspondiente metilamida mediante el empleo de tetrafluoroborato de *O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetrametilamonio (TBTU) como agente de acoplamiento y metilamina. Seguidamente, se llevó a cabo la desprotección del grupo amino, por tratamiento con piperidina, y la consiguiente acetilación de la amina generada con mezcla de $Ac_2O$/Py. Por último, con el empleo de una disolución de NaOMe en metanol, se consiguió la desacetilación de los grupos hidroxilo del carbohidrato (Esquema 11).




**Esquema 11.** *Síntesis de Ac-L-Cys(α-D-GalNAc)-NHMe **3f**.*

Una vez obtenida la molécula objeto de estudio **3f**, se realizó el estudio conformacional de la misma en agua. A partir de los experimentos NOESY (Figura 6) se obtuvieron distancias protón-protón relevantes para el análisis conformacional. Dichas distancias, que se recogen en la Tabla 2, se utilizaron como restricciones promediadas en el tiempo en las simulaciones de MD (MD-tar, del inglés *molecular dynamics with time-averaged restraints*). [63,64]

| Ac-L-Cys(α-D-GalNAc)-NHMe **3f** | | |
|---|---|---|
| **Distancias (Å)** | **Valores experimentales** | **Valores obtenidos por MD-tar** |
| NH1-Hα | 2.5 | 2.7 |
| NH2- Hα | 2.9 | 2.9 |
| NH1-$H1_{(GalNAc)}$ | 4.0 | 3.9 |
| NH2-$H1_{(GalNAc)}$ | 3.3 | 3.3 |
| NH1-NH2 | 3.1 | 2.8 |

**Tabla 2.** *Comparación de las distancias H-H obtenidas de manera teórica y experimental.*

De esta forma se obtuvo una distribución de conformaciones de baja energía de acuerdo con los datos experimentales obtenidos por RMN.

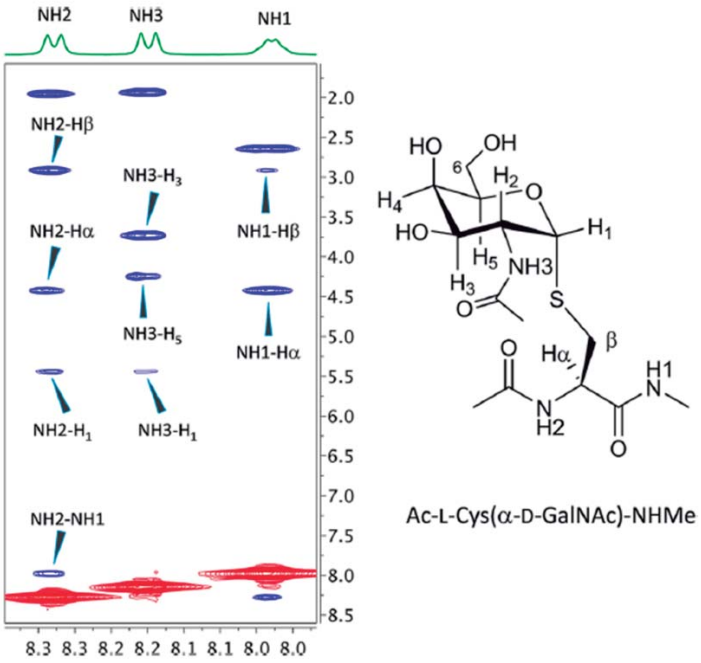



**Figura 6.** *2D-NOESY de Ac-L-Cys(α-D-GalNAc)-NHMe* ***3f*** *en $H_2O/D_2O$ (9:1) a pH = 5.5 y 298 K.*

La figura 7 muestra la distribución de confórmeros obtenida por MD-tar del compuesto **3f**, tanto para la cadena peptídica como para la disposición del enlace *S*-glicosídico. A su vez, se muestra la comparación con los antígenos Tn naturales de serina y treonina.[61,62]

En el compuesto **3f**, la cadena peptídica adopta, principalmente, una conformación plegada tipo α-hélice, corroborada por la señal intensa de los picos de cruce NOE NH2−NH1, NH1−Hα y NH2−Hα (Figura 6). Por el contrario, tal y como nuestro grupo de investigación describió previamente, ambos antígenos Tn naturales adoptan conformaciones extendidas de la cadena peptídica. En lo relativo al enlace glicosídico, el ángulo diedro $\phi_s$ (definido como O5−C1−S−C$\beta$) es bastante rígido, en concordancia con el efecto exo-anomérico.[65] Sin embargo, se observan diferencias significativas para el ángulo diedro $\psi_s$ (C1−S−Cβ−Cα) cuando se comparan con los derivados naturales. En el antígeno Tn de serina, este ángulo es cercano a los 180º, mientras que en el derivado de treonina el valor es aproximadamente

120º (conformación eclipsada). Sin embargo, para el derivado **3f** que contiene un enlace *S*-glicosídico, este valor se sitúa próximo a 60º. Es importante tener en cuenta que este confórmero se encuentra en uno de los mínimos locales calculados para la 1-*O*-metil-4-tio-α-maltosa.[66]

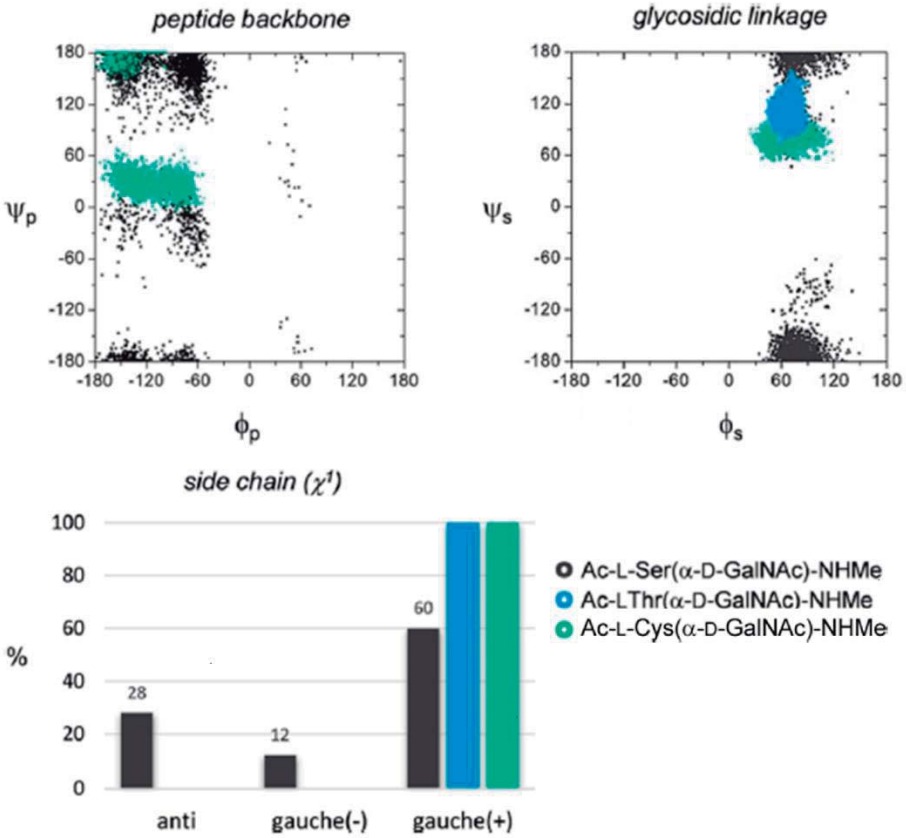



**Figura 7.** *Distribuciones obtenidas por MD-tar para la cadena peptídica, enlace glicosídico y cadena lateral de los derivados Ac-L-Ser(α-D-GalNAc)-NHMe, Ac-L-Thr(α-D-GalNAc)-NHMe y Ac-L-Cys(α-D-GalNAc)-NHMe (**3f**).*

Teniendo en cuenta estos resultados, es importante señalar que se puede modular el valor del ángulo diedro $\psi_s$ desde 60º a 180º mediante la incorporación de cisteína, serina o treonina. Con respecto a la cadena lateral, definida por el ángulo diedro $\chi^1$ (X−Cβ−Cα−N), las simulaciones indican que en el derivado **3f** ésta es bastante rígida, encontrándose solo el confórmero *gauche* (+), al igual que en el derivado Ac-L-Thr(α-D-GalNAc)-NHMe. A la vista de todas estas observaciones se puede concluir que el

comportamiento del derivado **3f** está estrechamente relacionado con el que muestra el antígeno Tn de treonina. De hecho, como se puede observar en la figura 8, mientras que en el derivado de serina el residuo carbohidrato adopta una disposición paralela con respecto a la cadena peptídica, en el caso de treonina y cisteína éste se dispone de manera perpendicular (Figura 8).

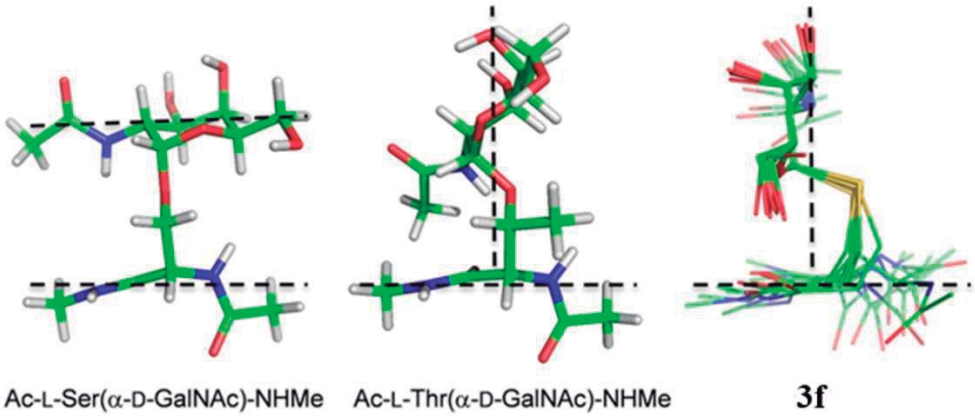



**Figura 8.** *Conformaciones representativas para los antígenos Tn naturales junto con la superposición de las conformaciones mayoritarias para* ***3f*** *calculadas mediante simulaciones de MD-tar.*

La conformación mayoritaria para el derivado **3f** permite la formación de un bolsillo hidrofílico entre el grupo NH del acetamido del carbohidrato (denominado N3 en la figura 9a) y el grupo amino del aminoácido (N2 en la figura 9a). Este resultado está de acuerdo con la función de distribución radial 2D que se muestra en la figura 9b. Dicha función calcula, en este caso, la probabilidad de encontrar una molécula de agua a una distancia < 3.5 Å de los nitrógenos N2 y N3, con respecto a la distribución de agua pura. Además, esta disposición espacial también está corroborada experimentalmente por la existencia de un pico de cruce NOE entre el protón del N2 y el protón $H_1$ del carbohidrato (NH2−H1, Figura 6). La presencia de estas moléculas de agua puente entre el carbohidrato y el aminoácido contribuyen a aumentar la rigidez del mimético de Tn. Este resultado es

similar al previamente descrito[62] por nuestro grupo de investigación para el caso del derivado Ac-L-Thr(α-D-GalNAc)-NHMe.

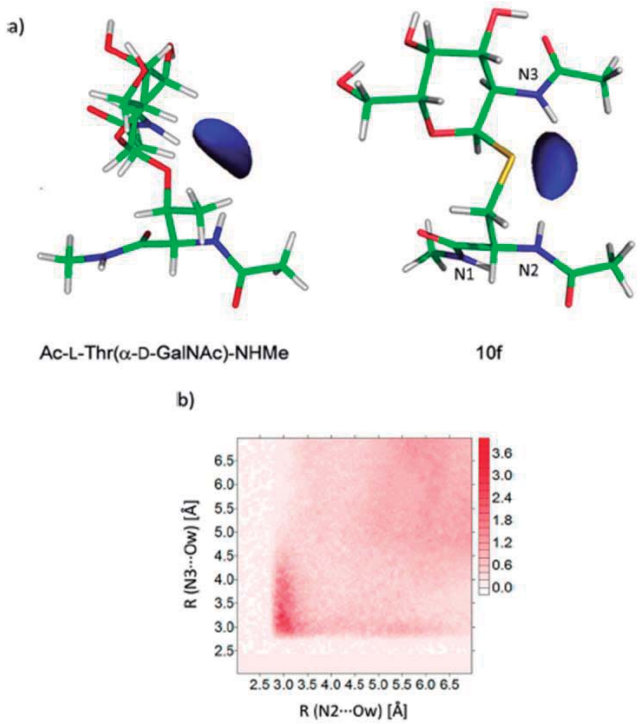



**Figura 9.** *a) Comparación de los bolsillos de agua para los derivados Ac-L-Thr(α-D-GalNAc)-NHMe y Ac-L-Cys(α-D-GalNAc)-NHMe (**3f**) obtenidos por cálculos de MD-tar. b) Función de distribución radial 2D para los átomos de nitrógeno N2 y N3 del derivado **3f** calculada por MD-tar.*

### 4.3.3. Estudios de afinidad

Estudios recientes sobre el diseño de vacunas contra el cáncer basadas en derivados de MUC1 han planteado la incorporación de análogos no naturales con el fin de aumentar la inmunogenicidad de las potenciales vacunas.[67–70] Hasta la fecha, se han publicado diferentes estrategias para la preparación de análogos de antígenos glicopeptídicos de tipo mucina, tales como la incorporación de átomos de fluor,[71–74] la síntesis de derivados *C*-glicósidos[75–82] y *S*-glicósidos[83–87] e incluso la utilización de homoserina y $\beta^3$-homotreonina en la cadena peptídica.[88,89] En este contexto, nuestro grupo de

investigación ha contribuido recientemente con la síntesis y el estudio de análogos del antígeno Tn conformacionalmente restringidos.[82] Sin embargo, apenas se encuentran en la literatura ejemplos de la utilización de *S*-glicósidos como análogos del antígeno Tn en la preparación de vacunas contra el cáncer.[84,90] Debido a ello, como primer paso para el desarrollo de una potencial vacuna, nos planteamos llevar a cabo estudios de afinidad de nuestro análogo azufrado con una molécula biológicamente relevante como lo es la lectina *Soybean agglutinin* (SBA). Los motivos de esta elección fueron la fácil disponibilidad y gran estabilidad de la lectina SBA, la alta afinidad que muestra con el carbohidrato GalNAc, antígeno Tn y otros miméticos de éste[82,91] y la disponibilidad de la estructura de rayos X del complejo de dicha lectina con galactosa.[92,93]

La afinidad del derivado **3f** fue evaluada mediante estudios de competición ELLA (del inglés, *enzyme-linked lectin assay)*, comparándose la afinidad de **3f** frente al derivado natural Ac-L-Ser(α-D-GalNAc)-NHMe. Para ello, se une de manera covalente a los pocillos de la placa ELLA el glicopéptido Ala-Pro-Asp-Thr(α-D-GalNAc)-Arg (100 mmol por pocillo) mediante la reacción con maleimida del amino terminal. Posteriormente, se añade una cantidad fija de lectina y cantidades crecientes del derivado **3f** o del derivado de serina. Los resultados obtenidos de los ensayos se representaron mediante el tanto por ciento de inhibición (Tabla 3). Esta medida se calcula como el cociente entre el valor de la densidad óptica observada a 450 nm en presencia de cada derivado de Tn y el valor de densidad óptica en ausencia de este (Tabla 3).

| nmol | Ac-L-Ser(α-D-GalNAc)-NHMe | 3f |
|---|---|---|
| 50 | 72.2 | 54.4 |
| 100 | 91.5 | 83.9 |
| 150 | 95.3 | 91.3 |

**Tabla 3.** *Porcentaje de inhibición (%), a distintas concentraciones, de los compuestos Ac-L-Ser( α-D-GalNAc)-NHMe y Ac-L-Cys( α-D-GalNAc)-NHMe (**3f**) para la unión SBA/Ala-Pro-Asp-Thr( α-D-GalNAc)-Arg.*

Como se puede ver en la tabla 3, aunque los valores son ligeramente superiores para el derivado de serina, ambos presentan unos valores de inhibición similares. Estos resultados están de acuerdo con estudios recientes en los que se demuestra una afinidad similar de *S*- y *O*-glicoconjugados frente a diferentes lectinas.[94]

Con el fin explicar estos resultados de afinidad, se llevaron a cabo estudios de MD con el complejo SBA/**3f** (Figura 10a).

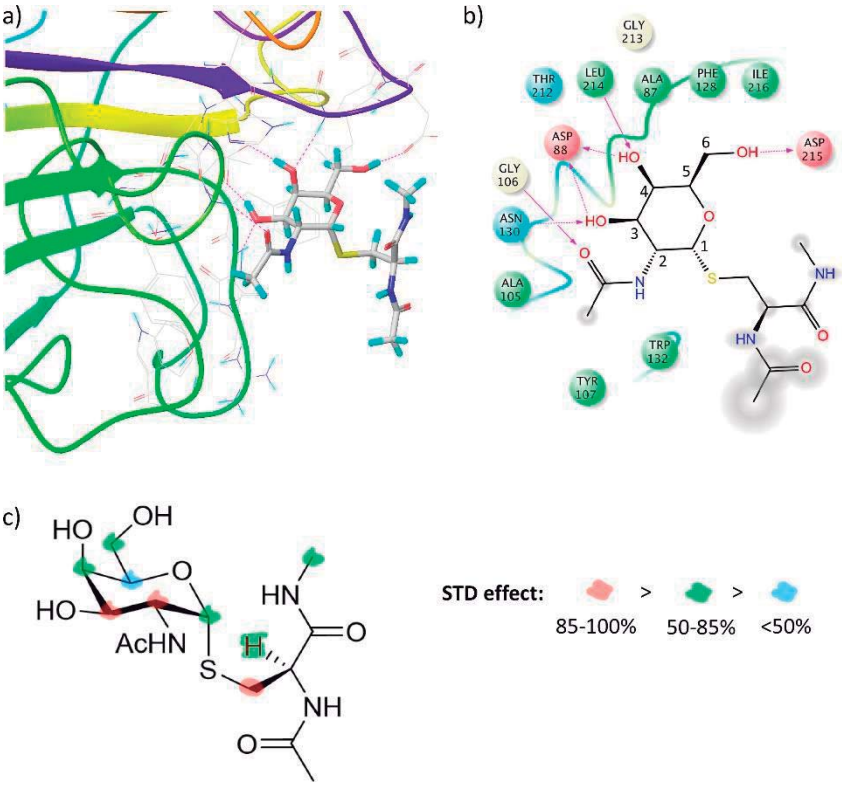



**Figura 10.** *a) Imagen tomada de la simulación de MD de 25 ns del complejo SBA/**3f** SBA. b) Diagrama de interacción de ligando para el complejo SBA/**3f** obtenido utilizando el programa Maestro 9.3.5 para un "frame" representativo de la MD. c) Epítopo de reconocimiento del derivado **3f** por SBA obtenido mediante experimentos de RMN de tipo STD.*

Los resultados de estos cálculos muestran que el compuesto **3f** interactúa con la lectina principalmente por la parte del carbohidrato. Se observan enlaces de hidrógeno entre el OH-3 y OH-4 del GalNAc y la cadena lateral del Asp88, el OH-6 y la cadena lateral del Asp215 y el OH-3 y la cadena lateral de Asn130. Además, los hidrógenos de la cara β del GalNAc, fundamentalmente H3 y H5, establecen interacciones CH/π con la Phe128. (Figura 10 b). Estas interacciones son las mismas que se pueden encontrar en la estructura de rayos X para el complejo SBA/ galactosa (pdb ID: 1SBF).[92,93]

Con el fin de corroborar de manera experimental este modelo teórico, se decidió llevar a cabo estudios de RMN del tipo STD (*saturation transfer difference*) entre la lectina SBA y el derivado **3f**.[95,96]

Como puede verse en la figura 10c, las intensidades STD más relevantes se sitúan en el carbohidrato, fundamentalmente en los hidrógenos H2 y H3, lo que, a grandes rasgos, está de acuerdo con el modelo teórico propuesto.

### 4.4. Conclusiones

Se ha desarrollado una metodología novedosa que permite la síntesis estereoselectiva de una amplia variedad de derivados de *S*-glicosilcisteína mediante el empleo de la adición sulfa-Michael doblemente estereoselectiva de tiocarbohidratos convenientemente protegidos sobre deshidroalaninas quirales. Estos compuestos son fácilmente derivatizables a los correspondientes *building blocks*, pudiéndose de esta manera emplear en la síntesis de péptidos en fase sólida.

Además, se ha realizado un estudio conformacional del mimético azufrado del antígeno Tn mediante MD y experimentos de RMN. En dicho estudio se han comparado tanto las preferencias conformacionales como la primera esfera de hidratación del derivado azufrado con los *O*-glicosilaminoácidos naturales, pudiéndose determinar que la disposición espacial que adoptan en disolución estos derivados azufrados es muy similar a la encontrada en el antígeno Tn de treonina. De hecho, en ambas moléculas el carbohidrato se dispone de manera perpendicular con respecto a la cadena peptídica, generando de esta manera el mismo bolsillo de agua entre la parte carbohidrato y la cadena peptídica.

Por último, se ha llevado a cabo un estudio comparativo de afinidad entre el derivado azufrado del antígeno Tn de serina y la molécula natural frente a la lectina *Soybean agglutinin* (SBA) empleando ensayos ELLA de

competición. Estos estudios demuestran que ambas moléculas poseen afinidades similares por la lectina SBA.

Con el fin de entender estos datos experimentales a nivel atómico se han realizado cálculos de MD del complejo SBA/**3f** que señalan que el reconocimiento tiene lugar por el GalNAc. Dichos estudios están avalados por experimentos de RMN del tipo STD.

Finalmente, esta metodología de síntesis se puede utilizar para la obtención de péptidos de elevado interés biológico.

## 4.5. Parte experimental

**General Procedure for the Sulfa-Michael Additions.** The corresponding α,β-dehydroamino acid (1 equiv.) was dissolved in dry THF (concentration of 10 mg/mL), and the resulting solution was introduced in a Schlenk tube under an argon atmosphere. This

solution was stirred and cooled to −78 °C. The corresponding thiocarbohydrate (**2a−g**, 1.2 equiv.) dissolved in THF was introduced into the Schlenk tube by a syringe, and DBU (1.2 equiv.) was then added slowly by another syringe. After the mixture was stirred at this temperature for 3 h, a saturated $NH_4Cl$ solution (the same volume as THF) was added. This mixture was stirred vigorously while it was allowed to warm to room temperature. After that, the reaction crude was diluted with ethyl ether, and the aqueous phase was extracted with ethyl acetate. The combined organic phases were washed with brine and dried with anhydrous $Na_2SO_4$. The solvent was filtered and evaporated, and the residue was purified by silica gel columna chromatography to give the corresponding sulfa-Michael adduct.

**Methyl (2S)-2-((4S,5R)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin- 3-yl)-3-(tetra-O-acetyl-β-D-glucosylthio)propanoate (8a).**

Compound **8a** was obtained from α,β-dehydroamino acid **1** (550 mg, 2.24 mmol) and tetra-O-acetyl-1-thio-β-D-glucose **2a** (980 mg, 2.69 mmol) following the procedure described for sulfa-Michael additions, using DBU (403 μL, 2.69 mmol) as a base. After columna chromatography (ethyl acetate/hexane, 1:1), compound **8**a was obtained as a white foam (1.147 g, 1.88 mmol, 84%).

$[\alpha]_D^{25}$ = −34.5 (c 1.03, $CHCl_3$).

HRMS ESI+ (m/z): calcd for $C_{24}H_{35}NO_{15}SNa^+$ [M + Na]+ 632.1620, found 632.1626.

$^1$H NMR (400 MHz, $CDCl_3$): δ (ppm) 1.40 (s, 3H; $CH_3$), 1.57 (s, 3H; $CH_3$), 1.99 (s, 3H; Ac), 2.02 (s, 3H; Ac), 2.04 (s, 3H; Ac), 2.05 (s, 3H; Ac), 3.31 (dd, J = 5.3, 14.3, Hz, 1H; $CHCH_2$), 3.43 (s, 3H; $OCH_3$), 3.49−3.62 (m, 1H; $CHCH_2$), 3.76 (s, 3H; CO2$CH_3$), 3.70−3.78 (m, 1H; H-5), 4.08−4.19 (m, 2H; H-6a; H-6b), 4.19−4.29 (m, 1H; $CHCH_2$), 4.64 (d, J = 10.1 Hz, 1H; H-1), 4.95− 5.09 (m, 2H; H-2; H-4), 5.21 (t, J = 9.3 Hz, 1H; H-3).

$^{13}$C NMR (100 MHz, $CDCl_3$): δ (ppm) 15.6 ($CH_3$), 19.8 (CH3), 20.4, 20.5, 20.6, 20.7 (Ac), 30.7 ($CHCH_2$), 50.8 ($OCH_3$), 52.9 ($CO_2CH_3$), 55.8 ($CHCH_2$), 62.0 (C-6), 68.0 (C-4), 70.0 (C-2), 73.4 (C-3), 76.2 (C-5), 84.9 (C-1), 90.0 ($CNCH_3OH$), 108.4 ($CCH_3OCH_3$), 154.5 ($NCO_2$), 169.3, 169.4, 169.5, 170.0, 170.4 (4×Ac; $CO_2CH_3$).

**Methyl (2S)-2-((4S,5R)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin- 3-yl)-3 (tetra-O-acetyl-β-D-galactosylthio)propanoate (8b).**

Compound **8b** was obtained from α,β-dehydroamino acid **1** (200 mg, 0.82 mmol) and tetra-O-acetyl-1-thio-β-D-galactose **2b** (327 mg, 0.82 mmol)

following the procedure for sulfa-Michael additions, using DBU (136 μL, 0.89 mmol) as a base. After column chromatography (ethyl acetate/hexane, 1:1), **8b** was obtained as a white amorphous solid (406 mg, 0.67 mmol, 82%). $[\alpha]_D^{25}$ = −24.6 (c 1.00, $CHCl_3$).

Mp: 64−66 °C.

HRMS ESI+ (m/z): calcd for $C_{24}H_{35}NO_{15}SNa^+$ $[M + Na]^+$ 632.1620, found 632.1606.

$^1$H NMR (400 MHz, $CDCl_3$): δ (ppm) 1.43 (s, 3H; $CH_3$), 1.60 (s, 3H; $CH_3$), 1.99 (s, 3H; Ac), 2.04 (s, 3H; Ac), 2.08 (s, 3H; Ac), 2.19 (s, 3H; Ac), 3.36 (dd, J = 5.0, 14.8 Hz, 1H; $CHCH_2$), 3.44 (s, 3H; $OCH_3$), 3.53−3.59 (m, 2H; $CHCH_2$; OH), 3.77 (s, 3H; $CO_2CH_3$), 3.98 (t, J = 6.31 Hz, 1H; H-5), 4.07−4.17 (m, 2H; H-6a; H-6b), 4.21− 4.25 (m, 1H; $CHCH_2$), 4.59 (d, J = 9.9 Hz, 1H; H-4), 5.03−5.06 (m, 1H; H-2), 5.24 (t, J = 9.9 Hz, 1H; H-3), 5.43 (d, J = 3.2 Hz, 1H; H-1).

$^{13}$C NMR (100 MHz, $CDCl_3$): δ (ppm) 15.5 ($CH_3$), 19.8 ($CH_3$), 20.6, 20.7, 20.8, 20.9 (Ac), 30.2 ($CHCH_2$), 50.8 ($OCH_3$), 53.0 ($CO_2CH_3$), 55.9 ($CHCH_2$), 61.8 (C-6), 67.0 (C-3), 67.2 (C-1), 71.6 (C-2), 75.1 (C-5), 85.0 (C-4), 90.1 ($CNCH_3OH$), 108.5 ($CCH_3OCH_3$), 154.5 ($NCO_2$), 169.4, 169.7, 170.0, 170.3, 170.4 (4×Ac; $CO_2CH_3$).

**Methyl (2S)-2-((4S,5R)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)-3-(tri-O-acetyl-2-benzyloxycarbonylamino-2-deoxy-β-D-glucosylthio) propanoate (8c).**

Compound **8c** was obtained from α,β-dehydroamino acid **1** (22 mg, 0.09 mmol) and tri-Oacetyl-2-benzyloxycarbonyl-2-deoxy-1-thio-β-D-glucose **2c** (45 mg, 0.10 mmol) following the procedure for sulfa-Michael additions, using DBU (15 μL, 0.10 mmol) as a base. After columna chromatography (ethyl acetate/hexane, 1:1), **8c** was obtained as a white amorphous solid (36 mg, 0.05 mmol, 56%).

$[\alpha]_D^{25}$ = −27.4 (c 1.00, $CHCl_3$).

Mp: 68−70 °C.

HRMS ESI+ (m/z): calcd for $C_{30}H_{40}N_2O_{15}SNa+$ $[M + Na]^+$ 723.2042, found 723.2037.

$^1$H NMR (400 MHz, $CDCl_3$): δ (ppm) 1.40 (s, 3H; $CH_3$), 1.59 (s, 3H; $CH_3$), 1.93 (s, 3H; Ac), 2.02 (s, 3H; Ac), 2.06 (s, 3H; Ac), 3.31 (dd, J = 4.5, 14.9 Hz, 1H; $CHCH_2$), 3.44 (s, 3H; $OCH_3$), 3.63−3.72 (m, 3H; $CHCH_2$; H-5; H-4), 3.76 (s, 3H; $CO_2CH_3$), 4.10−4.23 (m, 2H; H-6a; H-6b), 4.28 (d, J = 6.1 Hz, 1H; $CHCH_2$), 4.84 (d, J = 10.0 Hz, 1H; H-1), 4.95 (d, J = 8.8 Hz, 1H; $NH_{Cbz}$), 5.01−5.13 (m, 3H; H-2; $OCH_2Ph$), 5.20 (d, J = 9.2 Hz, 1H; H-3), 7.29−7.43 (m, 5H; Ph).

$^{13}$C NMR (100 MHz, $CDCl_3$): δ (ppm) 15.7 ($CH_3$), 19.9 ($CH_3$), 20.6, 20.7, 20.7 (Ac), 30.8 ($CHCH_2$), 51.0 ($OCH_3$), 53.1 ($CO2CH_3$), 55.6 ($CHCH_2$), 56.0 (C-4), 62.3 (C-6), 67.2 ($CH_2Ph$), 68.5 (C-2), 73.2 (C-3), 76.3 (C-5), 86.0 (C-1), 90.3 ($CNCH_3OH$), 108.6 ($CCH_3OCH_3$), 128.2, 128.4, 128.7 (Ph), 154.8 ($NCO_2$), 155.9 ($NCO_2Bn$), 169.6, 169.6, 170.7, 170.8 (3×Ac; $CO_2CH_3$).

**Methyl (2S)-2-((4*S*,5*R*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)-3 (hepta-*O*-acetyl-*β*-D-lactosylthio)propanoate (8d).** Compound **8d** was obtained from *α*,*β*-dehydroamino acid **1** (200 mg, 0.82 mmol) and hepta-*O*-acetyl-1-thio-*β*-D-lactose **2d** (535 mg, 0.82 mmol) following the procedure for sulfa-Michael additions, using DBU (136 μL, 0.89 mmol) as a base. After columna chromatography (ethyl acetate/hexane, 1:1), **8d** was obtained as a white amorphous solid (574 mg, 0.64 mmol, 78%).

$[\alpha]_D^{25}$ = −21.0 (c 1.00, $CHCl_3$).

Mp: 97−99 °C.

HRMS ESI+ (m/z): calcd for $C_{36}H_{51}NO_{23}SNa^+$ $[M + Na]^+$ 920.2465, found 920.2470.

$^{1}$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.36 (s, 3H; $CH_3$), 1.53 (s, 3H; $CH_3$), 1.92 (s, 3H; Ac), 2.00 (s, 3H; Ac), 2.01 (s, 3H; Ac), 2.02 (s, 3H; Ac), 2.03 (s, 3H; Ac), 2.05 (s, 3H; Ac), 2.11 (s, 3H; Ac), 3.24 (dd, J = 4.6, 14.9 Hz, 1H; $CHCH_2$), 3.40 (s, 3H; $OCH_3$), 3.55 (dd, J = 10.3, 14.9 Hz, 1H; $CHCH_2$), 3.60−3.68 (m, 2H; HGlc-5; OH), 3.69−3.78 (m, 4H; $CO_2CH_3$; HGlc-4), 3.85 (t, J = 6.8 Hz, 1H; HGal-5),3.98 (dd, J = 5.9, 12.1 Hz, 1H; HGlc-6a), 4.04−4.11 (m, 2H; HGal-6a; HGal-6b), 4.19 (dd, J = 4.5, 10.3 Hz, 1H; $CHCH_2$), 4.44 (d, J = 7.9 Hz, 1H; HGal-1), 4.51 (dd, J = 1.5, 12.0 Hz, 1H; HGlc-6b), 4.59 (d, J = 10.0 Hz, 1H; HGlc-1), 4.85−4.95 (m, 2H; HGlc-2; HGal-3), 5.05 (dd, J = 7.9, 10.4 Hz, 1H; HGal-2), 5.16 (t, J = 9.1 Hz, 1H; HGlc-3), 5.30 (d, J = 3.0 Hz, 1H; HGal-4).

$^{13}$C NMR (100 MHz, $CDCl_3$): δ (ppm) 14.2 ($CH_3$), 15.6 ($CH_3$), 19.8, 20.5, 20.6, 20.6, 20.7, 20.7, 21.0 (Ac), 30.6 ($CHCH_2$), 50.9 ($OCH_3$), 52.9 ($CO_2CH_3$), 55.8 ($CHCH_2$), 60.9 (CGal-6), 62.1 (CGlc-6), 66.7 (CGal-4), 69.1 (CGal-2), 70.3 (CGlc-2), 70.8 (CGal-5), 71.0 (CGal-3), 73.3 (CGlc-3), 75.9 (CGlc-4), 77.1 (CGlc-5), 84.8 (CGlc-1), 90.0 ($CNCH_3OH$), 101.0 (CGal-1), 108.5 ($CCH_3OCH_3$), 154.5 ($NCO_2$), 169.1, 169.3, 169.6, 169.6, 170.1, 170.1, 170.3, 170.4 (7×Ac; $CO_2CH_3$).

**Methyl 2-((4*S*,5*R*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)-3-(tetra-*O*-acetyl-*α*-D-glucosylthio)propanoate, Mixture (7e/8e).**

Diastereomeric mixture **7e**/**8e** was obtained from *α*,*β*- dehydroamino acid **1** (101 mg, 0.41 mmol) and tetra-*O*-acetyl-1-thio-*α*-D-glucose **2e** (151 mg, 0.41 mmol) following the procedure for sulfa-Michael additions, using DBU (67 μL, 0.45 mmol) as a base. After column chromatography (ethyl acetate/hexane, 1:1), diastereomeric mixture **7e**/**8e** was obtained as a colorless syrup (197 mg, 0.32 mmol, 79%) in a diastereomeric ratio 23/77, as determined by $^{1}$H NMR.

HRMS ESI+ (m/z): calcd for $C_{24}H_{35}NO_{15}SNa^+$ $[M + Na]^+$ 632.1620, found 632.1617.

$^1$H NMR data of major compound (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.46 (s, 3H), 1.59 (s, 3H), 2.00 (s, 3H), 2.06 (s, 3H), 2.12 (s, 3H), 2.16 (s, 3H), 3.27−3.41 (m, 3H), 3.44 (s 3H), 3.78 (s, 3H), 4.08−4.24 (m, 2H), 4.26−4.37 (m, 2H), 5.19−5.36 (m, 4H).

$^{13}$C NMR (100 MHz, $CDCl_3$): $\delta$ (ppm) 15.7, 19.8, 20.6, 20.7, 20.9, 29.9, 50.9, 53.1, 53.4, 62.2, 66.1, 69.2, 69.6, 70.8, 81.3, 90.6, 108.3, 154.4, 169.5, 169.7, 169.8, 170.1, 170.7.

**Methyl (2*S*)-2-((4*S*,5*R*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)-3-tri-*O*-acetyl-2-acetamido-2-deoxy-α-D-galactosylthio) propanoate (8f).**

Compound **8f** was obtained from *α*,*β*-dehydroamino acid **1** (424 mg, 1.73 mmol) and tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-*α*-D-galactose **2f** (690 mg, 1.90 mmol) following the procedure for sulfa-Michael additions, using DBU (310 μL, 2.07 mmol) as a base. After column chromatography ($CHCl_3$/MeOH, 95:5), **8f** was obtained as a white amorphous solid (704 mg, 0.64 mmol, 67%).

$[\alpha]_D^{25}$ = +76.6 (c 1.10, $CHCl_3$).

Mp: 179−181 °C.

HRMS ESI+ (m/z): calcd for $C_{24}H_{36}N_2O_{14}SNa^+$ $[M + Na]^+$ 631.1785, found 631.1788.

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.41 (s, 3H; CH3), 1.57 (s, 3H; $CH_3$), 1.97 (s, 3H; Ac), 2.00 (s, 3H; Ac), 2.07 (s, 3H; Ac), 2.14 (s, 3H; Ac), 3.31−3.35 (m, 2H; $CHCH_2$), 3.43 (s, 3H; $OCH_3$), 3.77 (s, 3H; $CO_2CH_3$), 4.01−4.22 (m, 4H; $CHCH_2$; H-6a; H-6b; OH), 4.52 (t, J = 6.3 Hz, 1H; H-5), 4.61−4.71 (m, 1H; H-2), 5.06 (dd, J = 3.2, 11.7 Hz, 1H; H-3), 5.40 (d, J = 2.8 Hz, 1H; H-4), 5.78 (d, J = 5.4 Hz, 1H; H-1), 6.16 (d, J = 7.2 Hz, 1H; NH).

$^{13}$C NMR (100 MHz, $CDCl_3$): $\delta$ (ppm) 15.8 ($CH_3$), 19.9 ($CH_3$), 20.7, 20.7, 20.8, 23.1 (Ac), 28.8 ($CHCH_2$), 48.7 (C-2), 50.9 ($OCH_3$), 53.0 ($CO2CH_3$), 53.1 ($CHCH_2$), 62.0 (C-6), 67.2 (C-4), 67.6 (C-5), 68.4 (C-3), 82.5 (C-1), 90.6 ($CNCH_3OH$), 108.3 ($CCH_3OCH_3$), 154.6 ($NCO_2$), 169.6, 170.2, 170.6, 171.1, 171.3 (4×Ac; $CO_2CH_3$).

**Methyl (2*S*)-2-((4*S*,5*R*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)-3-(tri-*O*-acetyl-2-acetamido-2-deoxy-*α*-D-glucosylthio) propanoate (8g).**

Compound **8g** was obtained from *α*,*β*-dehydroamino acid **1** (143 mg, 0.58 mmol) and tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-*α*-D-glucose **2g** (234 mg, 0.64 mmol) following the procedure for sulfa-Michael additions, using DBU (105 μL, 0.70 mmol) as a base. After column chromatography ($CHCl_3$/MeOH, 95:5), **8g** was obtained as a white foam (284 mg, 0.46 mmol, 80%).

$[\alpha]_D^{25}$ = +54.7 (c 1.14, $CHCl_3$).

HRMS ESI+ (m/z): calcd for $C_{24}H_{36}N_2O_{14}SNa^+$ $[M + Na]^+$ 631.1785, found 631.1780.

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.38 (s, 3H; $CH_3$), 1.53 (s, 3H; $CH_3$), 1.91 (s, 3H; Ac), 1.98 (s, 3H; Ac), 1.99 (s, 3H; Ac), 2.06 (s, 3H; Ac), 3.23−3.35 (m, 2H; $CHCH_2$), 3.39 (s, 3H; $OCH_3$), 3.72 (s, 3H; $CO_2CH_3$), 4.05−4.28 (m, 3H; H-6a; H6-b; $CHCH_2$), 4.28−4.38 (m, 1H; H-5), 4.38−4.47 (m, 1H; H-2); 4.98−5.09 (m, 2H; H-3; H-4), 5.57 (d, J = 5.4 Hz, 1H; H-1), 6.19 (d, J = 8.0 Hz, 1H; NHAc).

$^{13}$C NMR (100 MHz, $CDCl_3$): $\delta$ (ppm) 15.6 ($CH_3$), 19.8 ($CH_3$), 20.6, 20.7, 20.7, 23.0 (Ac), 29.8 ($CHCH_2$), 50.9 ($OCH_3$), 52.5 (C-2), 53.0 ($CO_2CH_3$), 53.3 ($CHCH_2$), 62.0 (C-6), 68.3 (C-4), 68.7 (C-5), 71.1 (C-3), 83.2 (C-1), 90.7 ($CNCH_3OH$), 108.3 ($CCH_3OCH_3$), 154.7 ($NCO_2$), 169.4, 169.5, 170.7, 171.0, 171.6 (4xAc; $CO_2CH_3$).

**tert-Butyl (2*S*)-2-((4*S*,5*R*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)3-(tri-*O*-acetyl-2-acetamido-2-deoxy-*α*-D-galactosylthio) propanoate (8'f)**

Compound **8'f** was obtained from *α*,*β*-dehydroamino acid **1′** (720 mg, 2.51 mmol) and tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-*α*-D-galactose **2f** (1.00 g, 2.73 mmol) following the procedure for sulfa-Michael additions, using DBU (310 μL, 2.07 mmol) as a base. After column chromatography ($CHCl_3$/MeOH, 95:5), **8'f** was obtained as a white amorphous solid (1.40 g, 2.15 mmol, 86%).

$[\alpha]_D^{25}$ = +89.3 (c 1.00, $CHCl_3$).

Mp: 184−186 °C.

HRMS ESI+ (m/z): calcd for $C_{27}H_{42}N_2O_{14}SNa^+$ $[M + Na]^+$ 673.2249, found 673.2243.

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.41 (s, 3H; $CH_3$), 1.44 (s, 9H; tBu), 1.56 (s, 3H; $CH_3$), 1.96 (s, 3H; Ac), 1.99 (s, 3H; Ac), 2.07 (s, 3H; Ac), 2.14 (s, 3H; Ac), 3.29 (d, J = 7.6 Hz, 2H; $CHCH_2$), 3.41 (s, 3H; $OCH_3$), 4.00 (t, J = 7.6 Hz, 1H; $CHCH_2$), 4.06−4.19 (m, 3H; H-6a; H6-b; OH), 4.53 (t, J = 6.3 Hz, 1H; H-5), 4.62−4.72 (m, 1H; H-2), 5.06 (dd, J = 3.2, 11.7 Hz, 1H; H-3), 5.39 (d, J = 2.6 Hz, 1H; H-4), 5.72 (d, J = 5.4 Hz, 1H; H-1), 6.14 (d, J = 7.5 Hz, 1H; NH).

$^{13}$C NMR (100 MHz, $CDCl_3$): $^{\delta}$ (ppm) 15.4 ($CH_3$), 19.8 ($CH_3$), 20.6, 20.7, 20.7, 23.1 (Ac), 27.8 ($C(CH3)_3$), 28.7 ($CHCH_2$), 48.6 (C-2), 50.7 ($OCH_3$), 54.1 ($CHCH_2$), 61.9 (C-6), 67.2 (C-4), 67.7 (C-5), 68.4 (C-3), 82.6 ($C(CH_3)_3$), 83.1 (C-1), 90.3 ($CNCH_3OH$), 108.2 ($CCH_3OCH_3$), 154.6 ($NCO_2$), 168.0, 170.2, 170.7, 171.1, 171.2 (4xAc; $CO_2$ tBu).

**tert-Butyl (2*S*)-2-((4*S*,5*R*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)3-(tri-*O*-acetyl-2-acetamido-2-deoxy-*α*-D-glucosylthio) propanoate Ester (8'g).**

Compound **8'g** was obtained from *α*,*β*-dehydroamino acid **1′** (504 mg, 1.75 mmol) and tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-*α*-D-glucose **2g** (765 mg, 2.10 mmol) following the procedure for sulfa-Michael additions, using DBU (288 μL, 1.92 mmol) as a base. After column chromatography ($CHCl_3$/MeOH, 19:1), **8'g** was obtained as a white amorphous solid (845 mg, 1.30 mmol, 74%).

$[\alpha]_D^{25}$ = +59.5 (c 0.99, $CHCl_3$).

Mp: 71−73 °C.

HRMS ESI+ (m/z): calcd for $C_{27}H_{42}N_2O_{14}SNa^+$ $[M + Na]^+$ 673.2249, found 673.2258.

$^1$H NMR (400 MHz, $CDCl_3$): *δ* (ppm) 1.39 (s, 3H; $CH_3$), 1.42 (s, 9H; tBu), 1.53 (s, 3H; $CH_3$), 1.92 (s, 3H; Ac), 1.99 (s, 3H; Ac), 2.00 (s, 3H; Ac), 2.07 (s, 3H; Ac), 3.19−3.33 (m, 2H; $CHCH_2$), 3.39 (s, 3H; $OCH_3$), 4.00 (dd, 1H; J = 5.9, 9.0 Hz, $CHCH_2$), 4.08 (d, 1H; J = 12.2 Hz; H-6a), 4.20−4.30 (m, 2H; H-6b; OH), 4.30−4.37 (m, 1H; H-5), 4.44 (dd, 1H; J = 2.6, 5.3 Hz; H-2), 5.13−4.97 (m, 2H; H-3; H-4), 5.56 (d, J = 5.4 Hz, 1H; H-1), 6.15 (d, J = 7.8 Hz, 1H; NHAc).

$^{13}$C NMR (100 MHz, $CDCl_3$): *δ* (ppm) 15.4 ($CH_3$), 19.7 ($CH_3$), 20.6, 20.7, 20.7, 23.1 (Ac), 27.8 ($C(CH_3)_3$), 29.4 ($CHCH_2$), 50.7 ($OCH_3$), 52.6 (C-2), 54.2 ($CHCH_2$), 61.9 (C-6), 68.2 (C-4), 68.6 (C-5), 71.2 (C-3), 82.8 ($C(CH_3)_3$), 83.1 (C-1), 90.4 ($CNCH_3OH$), 108.2 ($CCH_3OCH_3$), 154.7 ($NCO_2$), 168.0, 169.3, 170.8, 170.9, 171.7 (4×Ac; $CO_2$tBu).

**Benzyl (2*S*)-2-((4*S*,5*R*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)-3-(tri-*O*-acetyl-2-acetamido-2-deoxy-*α*-D-glucosylthio) propanoate (8''g).**

Compound **8''g** was obtained from *α,β*-dehydroamino acid **1''** (271 mg, 0.85 mmol) and tri-*O*-acetyl-2- acetamido-2-deoxy-1-thio-α-D-glucose **2g** (337 mg, 0.92 mmol) following the procedure for sulfa-Michael additions, using DBU (161 μL, 1.00 mmol) as a base. After columna chromatography ($CHCl_3$/MeOH, 19:1), **8''g** was obtained as a white amorphous solid (411 mg, 0.60 mmol, 71%) with the presence of minor compound **7''g** in a ratio **7''g**/**8''g** of 15:85.

$[\alpha]_D^{25}$ = +60.4 (c 1.03, $CHCl_3$).

Mp: 46−48 °C.

HRMS ESI+ (m/z): calcd for $C_{30}H_{40}N_2O_{14}SNa^+$ $[M + Na]^+$ 707.2092, found 707.2090.

$^1$H NMR (400 MHz, $CDCl_3$): *δ* (ppm) 1.36 (s, 3H; $CH_3$), 1.52 (s, 3H; $CH_3$), 1.91 (s, 3H; Ac), 1.99 (s, 3H; Ac), 2.00 (s, 3H; Ac), 2.00 (s, 3H; Ac), 3.26−3.48 (m, 5H; $OCH_3$; $CHCH_2$), 4.01−4.12 (m, 1H; H-6a), 4.14−4.23 (m, 2H; H-6b; $CHCH_2$), 4.28−4.36 (m, 1H; H-5), 4.38−4.47 (m, 1H; H-2), 5.01−5.09 (m, 2H; H-3; H-4), 5.15 (s, 2H; $CH_2Ph$), 5.59 (d, J = 5.4 Hz, 1H; H-1), 6.23 (d, J = 7.8 Hz, 1H; NHAc), 7.30 (m, 5H; Ph).

$^{13}$C NMR (100 MHz, $CDCl_3$): *δ* (ppm) 15.5 ($CH_3$), 19.7 ($CH_3$), 20.6, 20.6, 20.7, 23.0 (Ac), 29.7 ($CHCH_2$), 50.7 ($OCH_3$), 52.6 (C-2), 54.5 ($CHCH_2$), 62.0 (C-6), 67.8 ($CH_2Ph$), 68.3 (C-4), 68.7 (C-5), 71.2 (C-3), 83.2 (C-1), 90.6 ($CNCH_3OH$), 108.4 ($CCH_3OCH_3$), 128.1, 128.4, 128.6, 135.0 (Ph), 154.7 ($NCO_2$), 168.9, 169.3, 170.7, 171.0, 171.6 (4×Ac; $CO_2Bn$).

**_S_-(_β_-D-Glucosyl)-D-cysteine Hydrochloride (9a).**

Compound **8a** (220 mg, 0.36 mmol) was introduced in a flask with 6 N HCl (5 mL). The mixture was stirred overnight at 60 °C. The solvent was removed in vacuo, and the crude was dissolved in water (3 mL) and extracted with ethyl acetate (3 mL). Aqueous layer was evaporated to give **9a** (*S*-*β*-D-Glc-D-Cys·HCl) as a colorless syrup (108 mg, 0.34 mmol, 94%).

$[\alpha]_D^{25}$ = +43.0 (c 0.56, $H_2O$).

HRMS ESI + (m/z): calcd for $C_9H_{18}NO_7S^+$ $[M]^+$ 284.0798, found 284.0799.

$^1$H NMR (400 MHz, $D_2O$): $\delta$ (ppm) 3.21−3.55 (m, 6H; $CHCH_2$; H-2; H-4; H-3; H-5), 3.71 (dd, J = 6.3, 12.2 Hz, 1H; H-6a), 3.80−3.97 (m, 1H; H-6b), 4.31−4.40 (m, 1H; $CHCH_2$), 4.62 (d, J = 9.7 Hz, 1H; H-1).

$^{13}$C NMR (100 MHz, $D_2O$): $\delta$ (ppm) 31.1 ($CHCH_2$), 53.2 ($CHCH_2$), 60.9 (C-6), 69.4 (C-4), 72.0 (C-2), 76.9 (C-3), 79.9 (C-5), 85.8 (C-1), 174.0 ($CO_2H$).

**_S_-(2-Acetamido-2-deoxy-*α*-D-galactosyl)-D-cysteine Hydrochloride (9f).** Compound **8'f** (1.40 g, 2.15 mmol) was introduced in a flask with 4 N HCl (25 mL). The mixture was stirred overnight at 40 °C. The solvent was removed in vacuo, and the crude was disolved in water (20 mL) and extracted with ethyl acetate (20 mL). Aqueous layer was evaporated, and the crude was purified with a LC-18 SPE tube to give **9f** (*S*-α-D-GalNAc-D-Cys·HCl) as a colorless syrup (567 mg, 1.57 mmol, 73%).

$[\alpha]_D^{25}$ = +141.6 (c 1.00, $H_2O$).

HRMS ESI+ (m/z): calcd for $C_{11}H_{21}N_2O_7S^+$ $[M]^+$ 325.1064, found 325.1056.

1H NMR (400 MHz, D2O): δ (ppm) 2.02 (s, 3H; Ac), 3.10 (dd, J = 7.8, 14.9 Hz, 1H; $CHCH_2$), 3.34 (dd, J = 4.6, 14.9 Hz, 1H; CHCH2), 3.69−3.81 (m, 2H; H-6a; H-6b), 3.84 (dd, J = 3.1, 11.4 Hz, 1H; H-3), 3.99 (d, J = 3.1 Hz, 1H; H-4), 4.23 (dd, J = 4.8, 7.2 Hz, 1H; H-5), 4.29 (dd, J = 4.6, 7.6 Hz, 1H; $CHCH_2$), 4.36 (dd, J = 5.5, 11.4 Hz, 1H; H-2), 5.59 (d, J = 5.5 Hz, 1H; H-1).

$^{13}$C NMR (100 MHz, $D_2O$): δ (ppm) 21.9 (Ac), 30.5 ($CHCH_2$), 50.0 (C-2), 52.9 ($CHCH_2$), 61.1 (C-6), 67.4 (C-3), 68.3 (C-4), 72.1 (C-5), 84.3 (C-1), 170.2 (Ac), 174.8 ($CO_2H$).

**_S_-(2-Acetamido-2-deoxy-*α*-D-glucosyl)-D-cysteine Hydrochloride (9g).**

Compound **8'g** (845 mg, 1.30 mmol) was introduced in a flask with 4 N HCl (20 mL). The mixture was stirred overnight at 40 °C. The solvent was removed in vacuo, and the crude was purified by a LC-18 SPE tube to give 5g (*S*-*α*-D-GlcNAc-D-Cys·HCl) as a colorless syrup (447 mg, 95%). This compound was obtained along with a small impurity (83/17 ratio by NMR) corresponding to the hydrolysis of the N-acetyl group of carbohydrate moiety, which was not possible to separate.

$^{1}$H NMR (400 MHz, $D_2O$): *δ* (ppm) 1.96 (s, 3H; Ac), 3.09 (dd, J = 6.9, 14.9 Hz, 1H; $CHCH_2$), 3.27 (dd, J = 4.7, 14.9 Hz, 1H; $CHCH_2$), 3.42 (t, J = 9.4 Hz, 1H; H-4), 3.60 (dd, J = 8.8, 10.9 Hz H-3), 3.68−3.82 (m, 2H; H-6a; H-6b), 3.91 (ddd, J = 2.3, 5.1, 10.1 Hz 1H; H-5), 4.04 (dd, J = 5.3, 11.0 Hz, 1H; H-2), 4.28 (dd, J = 4.8, 6.9 Hz, 1H; $CHCH_2$), 5.45 (d, J = 5.3 Hz, 1H; H-1).

$^{13}$C NMR (100 MHz, $D_2O$): *δ* (ppm) 21.9 (Ac), 30.5 ($CHCH_2$), 52.6 ($CHCH_2$), 53.7 (C-2), 60.3 (C-6), 70.0 (C-4), 70.6 (C-3), 72.9 (C-5), 84.2 (C-1), 169.9 (Ac), 174.5 ($CO_2H$).

HRMS ESI+ (m/z): calcd for $C_{11}H_{21}N_2O_7S^+$ $[M]^+$ 325.1064, found 325.1059.

**_S_-(Tri-_O_-acetyl-2-acetamido-2-deoxy-_α_-D-galactosyl)-_N_-Fmoc-D-cysteine Allyl Ester (10f).**

(1) Compound **9f** (567 mg, 1.57 mmol) was dissolved in water (10 mL), and $NaHCO_3$ (263 mg, 3.14 mmol) was added with stirring. Fmoc-OSu (793 mg, 2.35 mmol) was then added as a solution in acetonitrile (20 mL). The resulting mixture was stirred overnight at room temperature. Water was then added, and the mixture was acidified to a pH of 1 with HCl 2 N. Next, acetonitrile was evaporated, and the mixture was filtered. Fmoc-protected derivative was collected and dried as a white solid (608 mg, 1.11 mmol, 70%).

(2) Fmoc-D-Cys-(*α*-D-GalNAc)-OH (518 mg, 0.94 mmol) was introduced as a solution in dry DMF (10 mL), under argon atmosphere, in a Schlenk

previously charged with 3 Å molecular sieves and $Cs_2CO_3$ (370 mg, 1.13 mmol). Allyl bromide (131 μL, 1.51 mmol) was then added by a syringe, and the mixture was stirred overnight at room temperature. After that, the solvent was evaporated, and the crude was dissolved in AcOEt and filtered. Evaporation of the solvent yielded the compound Fmoc-D-Cys(*α*-D-GalNAc)-OAllyl (217 mg, 0.37 mmol, 37%).

(3) Treatment with $Ac_2O$ in pyridine (1:2, 5 mL) at room temperature for 1 h produced compound 6f (249 mg, 0.35 mmol, 94%), which was purified by a silica gel columna chromatography ($CHCl_3$/AcOEt 1:1) affording **10f** as a colorless syrup.

$[\alpha]_D^{25}$ = +70.8 (c 1.00, $CHCl_3$).

HRMS ESI+ (m/z): calcd for $C_{35}H_{40}N_2O_{12}SNa^+$ $[M + Na]^+$ 735.2194, found 735.2202.

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.94 (s, 3H; Ac), 1.99 (s, 3H; Ac), 2.05 (s, 3H; Ac), 2.18 (s, 3H; Ac), 3.07 (dd, J = 5.8, 14.0 Hz, 1H; $CHCH_2$), 3.27 (dd, J = 4.5, 13.8 Hz, 1H; $CHCH_2$), 3.95−4.17 (m, 2H; H-6a; H-6b), 4.27 (t, J = 6.2 Hz, 1H; $CH_{Fmoc}$), 4.36 (t, J = 6.1 Hz, 1H; H-5), 4.40−4.56 (m, 1H; $CH_{2Fmoc}$), 4.56−4.62 (m, 1H; $CH_{2Fmoc}$), 4.63−4.72 (m, 3H; $CHCH_2$; $CH_{2Allyl}$), 4.78 (ddd, J = 5.4, 8.5, 11.9 Hz, 1H; H-2), 5.00 (dd, J = 3.1, 11.7 Hz, 1H; H-3), 5.26−5.46 (m, 3H; H-4; $CH_2CH_2$), 5.58 (d, J = 5.2 Hz, 1H; H-1), 5.66 (d, J = 8.6 Hz, 1H; NHAc), 5.80 (d, J = 7.8 Hz, 1H; $N_{HFmoc}$), 5.92 (ddd, J = 6.0, 11.2, 16.5 Hz, 1H; $CH_2CH_2$), 7.37 (dd, J = 6.5, 13.5 Hz, 2H; Fmoc), 7.40−7.50 (m, 2H; Fmoc), 7.60−7.64 (m, 2H; Fmoc), 7.80 (d, J = 7.4 Hz, 2H; Fmoc).

$^{13}$C NMR (75 MHz, $CDCl_3$): $\delta$ (ppm) 20.4, 20.7, 20.7, 23.2 (4xAc), 33.5 (CHCH2), 47.0 ($CH_{Fmoc}$), 48.1 (C-2), 53.7 ($CHCH_2$), 62.0 (C-6), 66.6 ($CH_{2Allyl}$), 66.6 ($CH_{2Fmoc}$), 67.1 (C-4), 67.5 (C-5), 68.2 (C-3), 85.6 (C-1), 119.6 ($CH_2CH_2$), 120.0, 124.8, 124.8, 127.1, 127.1, 127.7 (Fmoc), 131.1

($CH_2CH_2$), 141.3, 141.3, 143.6, 143.7 (Fmoc), 155.7 ($NCO_2$), 169.7, 170.2, 170.3, 170.5, 170.9 (4×Ac; $CO_{2allyl}$).

**_S_-((Tri-_O_-acetyl)-2-acetamido-2-deoxy-$\alpha$-D-glucosyl)-_N_-Cbz-Dcysteine Benzyl Ester (10g).**

(1) Compound **9g** (244 mg, 0.68 mmol) was dissolved in water (2.5 mL), and $NaHCO_3$ (114 mg, 1.36 mmol) was added with stirring. The resulting solution was cooled at 0 °C, and Cbz-Cl (145 μL, 1.02 mmol) was added as a solution in 1,4-dioxane (2.5 mL). The resulting mixture was stirred at 0 °C for 1 h and allowed to warm to room temperature overnight. Water was then added, and the mixture was acidified to a pH of 1 with HCl 2 N. The aqueous layer was extracted with AcOEt (3 × 10 mL) and $CHCl_3$/iPrOH (3:1) (4 × 10 mL). The organic layers were combined and evaporated, and the crude was purified by LC-18 SPE tube to give the N-Cbz-protected derivative (91 mg, 0.20 mmol, 29%).

(2) Cbz-D-Cys ($\alpha$-D-GlcNAc)-OH was introduced as a solution in dry DMF (2 mL), under argon atmosphere, in a Schlenk tube previously charged with 3 Å molecular sieves and $Cs_2CO_3$ (77 mg, 0.24 mmol). Benzyl bromide (38 μL, 0.32 mmol) was then added by a syringe, and the mixture was stirred overnight at room temperature. Next, the solvent was evaporated, and the crude was dissolved in AcOEt and filtered. Evaporation of the solvent gave the benzyl ester derivative Cbz-D-Cys($\alpha$-D-GlcNAc)-OBn (60 mg, 0.11 mmol, 55%).

(3) Treatment with $Ac_2O$ and pyridine (1:2, 1.5 mL) at room temperature for 1 h yielded compound 6g (42 mg, 0.063 mmol, 57%), which was purified by a silica gel column chromatography (hexane/AcOEt 3:7) affording 6g as a colorless syrup.

$[\alpha]_D^{25}$ = +71.0 (c 1.01, CHCl3).

HRMS ESI+ (m/z): calcd for $C_{32}H_{38}N_2O_{12}SNa^+$ $[M + Na]^+$ 697.2038, found 697.2035.

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.94 (s, 3H; Ac), 2.03 (s, 3H; Ac), 2.03 (s, 3H; Ac), 2.05 (s, 3H; Ac), 3.01 (dd, J = 5.6, 13.9 Hz, 1H; $CHCH_2$), 3.21 (dd, J = 4.5, 13.9 Hz, 1H; CHCH2), 4.05 (d, J = 10.5 Hz, 1H; H-6a), 4.15−4.28 (m, 2H; H-6b; H-5), 4.47 (ddd, J = 11.0, 8.7, 5.4 Hz, 1H; H-2), 4.61−4.71 (m, 1H; CHCH2), 4.98 (dd, J = 11.1, 9.4 Hz, 1H; H-3), 5.03−5.21 (m, 5H; H-4; 2×CH2Ph), 5.44 (d, J = 5.3 Hz, 1H; H-1), 5.64 (d, J = 7.7 Hz, 1H; NHCbz), 5.76 (d, J = 8.4 Hz, 1H; NHAc), 7.36 (m, 10H; 2xPh).

$^{13}$C NMR (100 MHz, $CDCl_3$): $\delta$ (ppm) 20.6, 20.7, 20.7, 23.2 (4×Ac), 34.6 ($CHCH_2$), 52.5 (C-2), 54.0 ($CHCH_2$), 61.8 (C-6), 67.3 ($CH_2Ph$), 67.9 ($CH_2Ph$), 68.0 (C-4), 68.9 (C-5), 71.2 (C-3), 85.8 (C-1), 128.1, 128.3, 128.6, 128.6, 128.8, 128.8, 134.8, 136.0 (2×Ph), 155.7 ($NCO_2$), 169.2, 169.9, 170.1, 170.7, 171.6 (4×Ac; $CO_2Bn$).

**Methyl (2*R*)-2-((4*R*,5*S*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)-3-(tetra-*O*-acetyl-*β*-D-glucosylthio)propanoate (11a).**

Compound **11a** was obtained from *α*,*β*-dehydroamino acid **ent-1** (2.20 g, 8.97 mmol) and tetra-*O*-acetyl-1-thio-*β*-D-glucose **2a** (3.57 g, 9.85 mmol) following the procedure for sulfa-Michael additions, using DBU (1.60 mL, 10.50 mmol) as a base. After column chromatography (ethyl acetate/hexane, 1:1), **11a** was obtained as a white amorphous solid (4.32 g, 7.09 mmol, 79%). $[\alpha]_D^{25}$ = +11.3 (c 1.00, $CHCl_3$).

Mp: 54−56 °C.

HRMS ESI+ (m/z): calcd for $C_{24}H_{35}NO_{15}SNa^+$ $[M + Na]^+$ 632.1620, found 632.1630.

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.42 (s, 3H; $CH_3$), 1.57 (s, 3H; $CH_3$), 2.02 (s, 3H; Ac), 2.03 (s, 3H; Ac), 2.07 (s, 3H; Ac), 2.09 (s, 3H; Ac), 3.37

(dd, J = 10.0, 14.9, Hz, 1H; $CHCH_2$), 3.44 (s, 3H; $OCH_3$), 3.54−3.65 (m, 1H; $CHCH_2$), 3.71−3.83 (m, 1H; H-5), 3.79 (s, 3H; $CO_2CH_3$), 4.09−4.26 (m, 2H; H-6a; $CHCH_2$), 4.26−4.35 (m, 1H; H-6b), 4.56 (d, J = 10.0 Hz, 1H; H-1), 5.03−5.15 (m, 2H; H-2; H-4), 5.22 (t, J = 9.5 Hz, 1H; H-3).

$^{13}$C NMR (100 MHz, $CDCl_3$): $\delta$ (ppm) 15.4 ($CH_3$), 19.6 ($CH_3$), 20.6, 20.6, 20.7, 20.7 (Ac), 30.1 ($CHCH_2$), 50.8 ($OCH_3$), 53.0 ($CO_2CH_3$), 55.7 ($CHCH_2$), 61.4 (C-6), 67.7 (C-4), 69.3 (C-2), 73.6 (C-3), 76.3 (C-5), 84.7 (C-1), 90.1 ($CNCH_3OH$), 108.4 ($CCH_3OCH_3$), 154.7 ($NCO_2$), 169.2, 169.4, 169.4, 170.0, 170.6 (4×Ac; $CO_2CH_3$).

**tert-Butyl (2*R*)-2-((4*R*,5*S*)-4,5-Dimethyl-4-hydroxy-5-methoxy-2-oxooxazolidin-3-yl)-3-(tri-O-acetyl-2-acetamido-2-deoxy-α-D-galactosylthio) propanoate (11'f).**

Compound **11'f** was obtained from α,β-dehydroamino acid ent-1′ (764 mg, 2.66 mmol) and tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-*α*-D-galactose **2f** (1.06 g, 2.93 mmol) following the procedure for sulfa-Michael additions, using DBU (529 μL, 3.19 mmol) as a base. After column chromatography ($CHCl_3$/MeOH, 19:1), **11'f** was obtained as a white amorphous solid (1.31 g, 2.02 mmol, 76%).

$[\alpha]_D^{25}$ = +141.8 (c 1.00, $CHCl_3$).

Mp: 140−142 °C.

HRMS ESI+ (m/z): calcd for $C_{27}H_{42}N_2O_{14}SNa^+$ $[M + Na]^+$673.2249, found 673.2257.

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 1.45 (s, 12H; $C(CH_3)_3$; $CH_3$), 1.57 (s, 3H; $CH_3$), 1.96 (s, 3H; Ac), 1.99 (s, 3H; Ac), 2.03 (s, 3H; Ac), 2.14 (s, 3H; Ac), 3.33 (dd, J = 7.3, 10.5 Hz, 2H; $CHCH_2$), 3.41 (s, 3H; $OCH_3$), 4.05−4.18 (m, 3H; H-6a; H-6b; $CHCH_2$), 4.52 (t, J = 6.4 Hz, 1H; H-5), 4.60−4.77 (m, 1H; H-2), 5.02 (dd, J = 3.0, 11.6 Hz, 1H; H-3), 5.37 (d, J = 2.4 Hz, 1H; H-4), 5.64 (d, J = 5.3 Hz, 1H; H-1), 5.95 (d, J = 8.1 Hz, 1H; NHAc).

$^{13}$C NMR (100 MHz, $CDCl_3$): $\delta$ (ppm) 15.3 ($CH_3$), 19.9 ($CH_3$), 20.7, 20.7, 20.8, 23.2 (Ac), 27.9 ($C(CH_3)_3$), 30.8 ($CHCH_2$), 48.6 (C-2), 50.7 ($OCH_3$), 56.4 ($CHCH_2$), 61.8 (C-6), 67.2 (C-4), 67.7 (C-5), 68.4 (C-3), 83.3 ($C(CH_3)_3$), 85.5 (C-1), 90.5 ($CNCH_3OH$), 108.1 ($CCH_3OCH_3$), 154.8 ($NCO_2$), 168.3, 170.2, 170.7, 170.7, 171.1 (4×Ac; $CO_2$tBu).

**_S_-(_β_-D-Glucosyl)-L-cysteine Hydrochloride (12a).**

Compound **11a** (4.32 g, 7.09 mmol) was introduced in a flask with 6 N HCl (100 mL). The mixture was stirred overnight at 60 °C. The solvent was removed in vacuo, and the crude was dissolved in water (25 mL) and extracted with ethyl acetate (25 mL). Aqueous layer was evaporated to give **8a** (*S*-*β*-D-Glc-L-Cys·HCl) as a colorless syrup (2.13 g, 6.66 mmol, 94%).

$[\alpha]_D^{25} = -37.4$ (c 0.63, $H_2O$).

HRMS ESI+ (m/z): calcd for $C_9H_{18}NO_7S^+$ $[M + H]^+$ 284.0798, found 284.0787.

$^1$H NMR (400 MHz, $D_2O$): $\delta$ (ppm) 3.14 (dd, J = 8.0, 15.3 Hz, 1H; $CHCH_2$), 3.36−3.47 (m, 3H; $CHCH_2$; H-2; H-4), 3.47−3.56 (m, 2H; H-3; H-5), 3.72 (dd, J = 5.8, 12.4 Hz, 1H; H-6a), 3.88−3.99 (m, 1H; H-6b), 4.02 (dd, J = 3.7, 7.8 Hz, 1H; $CHCH_2$), 4.56 (d, J = 9.7 Hz, 1H; H-1).

$^{13}$C NMR (100 MHz, $D_2O$): $\delta$ (ppm) 30.4 ($CHCH_2$), 54.5 ($CHCH_2$), 60.9 (C-6), 69.4 (C-4), 71.7 (C-2), 77.0 (C-3), 80.1 (C-5), 84.7 (C-1), 172.2 ($CO_2H$). A small percentage of mutarotation was observed in the 1H NMR spectrum.[55]

**_S_-(2-Acetamido-2-deoxy-_α_-D-galactosyl)-L-cysteine Hydrochloride (12f).**

Compound **11′f** (1.31 g, 2.02 mmol) was introduced in a flask with 4 N HCl (25 mL). The mixture was stirred overnight at 40 °C. The solvent was removed in vacuo, and the crude was disolved in water (10 mL) and extracted

with ethyl acetate (10 mL). Aqueous layer was evaporated to give **12f** (*S*-*α*-D-GalNAc-L-Cys·HCl) as a colorless syrup (670 mg, 1.85 mmol, 92%).

$[\alpha]_D^{25}$ = +144.9 (c 1.00, $H_2O$).

HRMS ESI+(m/z): calcd for $C_{11}H_{21}N_2O_7S^+$ $[M + H]^+$325.1064, found 325.1061.

$^1$H NMR (400 MHz, $D_2O$): $\delta$ (ppm) 2.05 (s, 3H; Ac), 3.09−3.19 (m, 1H; $CHCH_2$), 3.34 (dd, J = 6.2, 15.0 Hz, 1H; $CHCH_2$), 3.74−3.87 (m, 3H; H-6a; H-6b; H-3), 4.00 (s, 1H; H-3), 4.04−4.11 (m, 1H; $CHCH_2$), 4.26−4.36 (m, 1H; H-5), 4.40 (dd, J = 5.4, 11.4 Hz, 1H; H-2), 5.54 (d, J = 5.3 Hz, 1H; H-1).

$^{13}$C NMR (100 MHz, $D_2O$): $\delta$ (ppm) 21.9 (Ac), 33.1 ($CHCH_2$), 49.9 (C-2), 54.4 ($CHCH_2$), 61.2 (C-6), 67.4 (C-3), 68.4 (C-4), 72.4 (C-5), 86.2 (C-1), 172.3 ($CO_2H$), 174.6 (Ac).

**_S_-(Tetra-_O_-acetyl-_β_-D-glucosyl)-_N_-Fmoc-L-cysteine (13a).**

Compound **12a** (975 mg, 3.05 mmol) was dissolved in water (25 mL), and $NaHCO_3$ (511 mg, 6.08 mmol) was added with stirring. Fmoc-OSu (1.54 g, 4.57 mmol) was then added as a solution in acetonitrile (50 mL). The resulting mixture was stirred at room temperature overnight. Water was then added, and the mixture was acidified to pH = 1 with 2 N HCl. Next, acetonitrile was evaporated, and the mixture was filtered. Fmoc-protected derivative, Fmoc-L-Cys(*β*-D-Glc)-OH, was collected and dried as a white amorphous solid (902 mg, 1.79 mmol, 59%). This compound was suspended in $Ac_2O$ (10 mL), and a few drops of $H_3PO_4$ (85 wt % in water) were added. The reaction was stirred for 2 h at room temperature and then quenched with saturated solution of $NaHCO_3$ (30 mL). The solution was acidified to pH = 1 with 2 N HCl and then extracted with ethyl acetate (3 × 15 mL). This crude compound was purified by flash silica gel chromatography ($CH_2Cl_2$/MeOH,

95:5) to give building block **13a** (566 mg, 0.84 mmol, 47%) as a colorless syrup.

$[\alpha]_D^{25}$ = −6.2 (c 1.00, $CHCl_3$).

HRMS ESI− (m/z): calcd for $C_{32}H_{34}NO_{13}S^-$ $[M + H]^-$ 672.1756, found 672.1778.

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ (ppm) 2.01 (s, 3H; Ac), 2.04 (s, 3H; Ac), 2.06 (s, 3H; Ac), 2.10 (s, 3H; Ac), 3.10 (dd, J = 4.3, 14.2 Hz, 1H; $CHCH_2$), 3.30 (d, J = 10.7 Hz, 1H; $CHCH_2$), 3.62−3.69 (m, 1H; H-5), 4.07−4.30 (m, 3H; $CHCH_{2Fmoc}$, H-6a; H-6b), 4.35−4.58 (m, 3H; $CHCH_{2Fmoc}$, H-1), 4.59−4.67 (m, 1H; $CHCH_2$), 4.99 (t, J = 9.7 Hz, 1H; H-2), 5.05 (m, 1H; H-4), 5.22 (t, J = 9.3 Hz, 1H; H-3), 5.94 (d, J = 7.1 Hz, 1H; NHFmoc), 7.32 (t, J = 7.4 Hz, 2H; Fmoc), 7.40 (t, J = 7.4 Hz, 2H; Fmoc), 7.55−7.64 (m, 2H; Fmoc), 7.77 (d, J = 7.5 Hz, 2H; Fmoc).

$^{13}$C NMR (100 MHz, CDCl3): $\delta$ (ppm) 20.6, 20.6, 20.7, 20.7 (Ac), 33.0 ($CHCH_2$), 47.2 ($CHCH_{2Fmoc}$), 53.8 ($CHCH_2$), 62.4 (C-6), 67.2 ($CHCH_{2Fmoc}$), 68.6 (C-4), 69.7 (C-2), 73.5 (C-3), 75.9 (C-5), 84.2 (C-1), 120.0, 125.1, 125.3, 127.2, 127.8 (Fmoc), 141.3, 143.7, 143.7 (Fmoc), 156.1 (NCO2), 169.4, 169.6, 170.2, 171.1, 176.7 (4×Ac; $CO_2H$).

**_S_-(Tri-O-acetyl-2-acetamido-2-deoxy-_α_-D-galactosyl)-N-Fmoc-L-cysteine (13f).**

Compound **12f** (670 mg, 1.85 mmol) was treated following the same methodology for the synthesis of Fmoc-D-Cys($\alpha$-D-$(OAc)_3$GalNAc)-OAllyl (**10f**).

(1) Fmoc-OSu, $NaHCO_3$, $H_2O$/ acetonitrile (1:2), room temperature, overnight (738 mg, 1.35 mmol, 73%).

(2) AllylBr, $Cs_2CO_3$, DMF, 3 Å molecular sieves, overnight (277 mg, 0.47 mmol, 35%).

(3) Ac2O, pyridine, 1 h (298 mg, 0.42 mmol, 89%). All physical and spectroscopic properties of Fmoc-L-Cys($\alpha$-D-$(OAc)_3$GalNAc)-OAllyl were identical to those reported in the literature.

(4) Fmoc-L-Cys($\alpha$-D-$(OAc)_3$GalNAc)-OAllyl (298 mg, 0.42 mmol) was dissolved in dry THF (10 mL) under argon atmosphere. $Pd(PPh_3)_4$ (5.2 mg, 4.5 ×10−3 mmol) and morpholine (182 μL, 2.1 mmol) were added, and the resulting solution was stirred at room temperature for 1 h. Next, the solvent was removed, and the crude was dissolved in $CHCl_3$/iPrOH (3:1, 10 mL). The solution was washed with 1 N HCl (2 × 5 mL), and the combined aqueous phases were extracted with more $CHCl_3$/iPrOH (2 × 5 mL). The organic layers were combined, dried with anhydrous $Na_2SO_4$, and solvent was removed. The crude product was purified by column chromatography (ethyl acetate/hexane 7:3 to $CH_2Cl_2$/MeOH 9:1) to give building block **13f** (268 mg, 0.40 mmol, 95%) as a yellow amorphous solid.

$[\alpha]_D^{25}$ = +120.3 (c 1.00, $CHCl_3$). Mp: 102−104 °C.

HRMS ESI− (m/z): calcd for $C_{32}H_{35}N_2O_{12}S^-$ $[M - H]^-$ 671.1916, found 671.1904.

$^1$H NMR (400 MHz, CDCl3): $\delta$ (ppm) 2.01 (s, 3H; Ac), 2.02 (s, 3H; Ac), 2.04 (s, 3H; Ac), 2.20 (s, 3H; Ac), 3.08 (d, J = 13.8 Hz, 1H; $CHCH_2$), 3.36 (d, J = 12.4 Hz, 1H; $CHCH_2$), 4.04−4.13 (m, 1H; H-6a), 4.20−4.30 (m, 2H; H-6b; $CHCH_{2Fmoc}$), 4.33−4.44 (m, 1H; $CHCH_{2Fmoc}$), 4.45−4.56 (m, 2H; $CHCH_{2Fmoc}$; H-5), 4.73−4.89 (m, 2H; H-2; $CHCH_2$), 4.94−5.03 (m, 1H; H-3), 5.40 (s, 1H; H-4), 5.58 (s, 1H; H-1), 6.42 (s, 1H; NHFmoc), 6.64 (s, 1H; NHAc), 7.31−7.46 (m, 4H; Fmoc), 7.60−7.68 (m, 2H; Fmoc), 7.79 (d, J = 7.4 Hz 2H; Fmoc).

$^{13}C$ NMR (100 MHz, $CDCl_3$): $\delta$ (ppm) 20.6, 20.7, 20.7, 23.1 (Ac), 36.4 ($CHCH_2$), 47.1 ($CHCH_{2Fmoc}$), 48.4 (C-2), 54.6 ($CHCH_2$), 61.6 (C-6), 67.0 ($CHCH_{2Fmoc}$), 67.2 (C-4), 68.0 (C-5), 68.3 (C-3), 87.4 (C-1), 120.0, 125.1, 125.1, 127.1, 127.8, 128.6, 128.7 (Fmoc), 141.3, 141.4, 143.7, 143.9 (Fmoc), 155.9 ($NCO_2$), 170.4, 170.6, 171.2, 171.4, 174.7 (4×Ac; $CO_2H$).

**_S_-(2-Acetamido-2-deoxy-$\alpha$-D-galactosyl)-_N_-(acetyl)-L-cysteine methylamide (3f).**

Fmoc-L-Cys($\alpha$-$(OAc)_3$-D-GalNAc)-OH (**13f**, 268 mg, 0.40 mmol) was dissolved in dry acetonitrile (10 mL) in the presence of 3 Å molecular sieves. TBTU (350 mg, 1.09 mmol) was added in one portion, followed by DIEA (732 μL, 4.2 mmol). After the mixture was stirred for 5 min, $MeNH_2 \cdot HCl$ (170 mg, 2.52 mmol) was added under argon atmosphere. The reaction was stirred overnight, the mixture was then filtered off over a Celite pad, and brine was added to the remaining solution. Aqueous layer was extracted with $CH_2Cl_2$ (2 × 5 mL), and organic layers were combined and washed with 1 N HCl (2 × 5 mL) and $NaHCO_3$ (2 × 5 mL) saturated solution. After evaporation of the solvents, the reaction crude was purified by column chromatography ($CH_2Cl_2$/MeOH, 95:5) yielding Fmoc-L-Cys($\alpha$-D-$(OAc)_3$GalNAc)-NHMe (158 mg, 0.23 mmol, 57%). This Fmoc-derivative was treated with piperidine (20% in acetonitrile, 10 mL) for 30 min at room temperature. The solvent was evaporated and the crude purified by flash column chromatography ($CH_2Cl_2$/MeOH, 90:10) yielding L-Cys($\alpha$-D-$(OAc)_3$GalNAc)-NHMe (102 mg, 0.22 mmol, 98%). This amine was reacted with $Ac_2O$ in the presence of pyridine (1:2, 5 mL) for 1 h at room temperature. The solvent was evaporated, and the reaction crude was purified by columna chromatography ($CH_2Cl_2$/MeOH, 95:5) yielding Ac-L-Cys($\alpha$-D-$(OAc)_3$GalNAc)-NHMe (96 mg, 0.19 mmol, 87%). This peracetylated

carbohydrate was treated with 0.5 M NaOMe solution in MeOH (2 mL) for 1.5 h. After neutralization with Dowex 50WX8, the solution was evaporated, and the crude was purified by a LC-18 SPE tube to give Ac-L-Cys($\alpha$-D-GalNAc)-NHMe **3f** as a white foam (58 mg, 0.15 mmol, 78% and 38% overall yield from **13f**).

$[\alpha]_D^{25}$ = +73.4 (c 0.50, $H_2O$).

HRMS ESI+(m/z): calcd for $C_{14}H_{26}N_3O_7S^+$ $[M + H]^+$ 380.1486, found 380.1471.

$^1$H NMR (400 MHz, $D_2O$): $\delta$ (ppm) 2.04 (s, 3H; AcN2), 2.06 (s, 3H; AcN3), 2.75 (s, 3H; NHMe), 2.97−3.10 (m, 2H; $CHCH_2$), 3.79 (d, J = 6.0 Hz, 2H; H-6a; H-6b), 3.84 (dd, J = 2.4, 11.2 Hz, 1H; H-3), 4.00 (s, 1H; H-4), 4.26 (t, J = 6.0 Hz, 1H; H-5), 4.36 (dd, J = 5.5, 11.4 Hz, 1H; H-2), 4.53 (t, J = 6.4 Hz, 1H; $CHCH_2$), 5.55 (d, J = 5.5 Hz, 1H; H-1). 1H NMR (400 MHz, $D_2O/H_2O$, 9:1): δ (ppm) 7.96−8.02 (m, 1H; NH1), 8.15 (d, J = 8.0 Hz, 1H; $NH_3$), 8.28 (d, J = 7.5 Hz, 1H; $NH_3$).

$^{13}$C NMR (100 MHz, $D_2O$): $\delta$ (ppm) 21.8 (Ac), 21.8 (Ac), 25.8 (NHMe), 31.7 ($CHCH_2$), 50.0 (C-2), 54.0 ($CHCH_2$), 61.0 (C-6), 67.4 (C-3), 68.4 (C-4), 71.6 (C-5), 84.6 (C-1), 172.5 (CONHMe), 174.3 (Ac), 174.5 (Ac).

## 4.6. Bibliografía


(1) Varki, A. *Glycobiology* **1993**, *3*, 97–130.

(2) Lis, H.; Sharon, N. *Eur. J. Biochem.* **1993**, *218*, 1–27.

(3) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.

(4) Jansson, A. M.; Hilaire, P. M. S.; Meldal, M.; Goodman, M.; Felix, A.; Moroder, L.; Toniolo, C. *Synthesis of Peptides and Peptidomimetics*; Thieme/Houben-Weyl Series, **2003**.

(5) Davis, B. G. *Chem. Rev.* **2002**, *102*, 579–602.

(6) Wittmann, V.; Fraser-Reid, B. O.; Tatsuta, K.; Thiem, J. *Glycoscience: Glycoproteins*; Springer, **2001**.

(7) Herzner, H.; Reipen, T.; Schultz, M.; Kunz, H. *Chem. Rev.* **2000**, *100*, 4495–4538.

(8) Arsequell, G.; Valencia, G. *Tetrahedron: Asymmetry* **1999**, *10*, 3045–3094.

(9) Taylor, C. M. *Tetrahedron* **1998**, *54*, 11317–11362.

(10) Carrico, I. S. *Chem. Soc. Rev.* **2008**, *37*, 1423–1431.

(11) Gamblin, D. P.; van Kasteren, S. I.; Chalker, J. M.; Davis, B. G. *FEBS J.* **2008**, *275*, 1949–1959.

(12) Qi, D.; Tann, C. M.; Haring, D.; Distefano, M. D. *Chem. Rev.* **2001**, *101*, 3081–3112.

(13) Marcaurelle, L. A.; Bertozzi, C. R. *Chem.—Eur. J.* **1999**, *5*, 1384–1390.

(14) Gamblin, D. P.; Scanlan, E. M.; Davis, B. G. *Chem. Rev.* **2009**, *109*, 131–163.

(15) Zhu, X.; Pachamuthu, K.; Schmidt, R. R. *J. Org. Chem.* **2003**, *68*, 5641–5651.

(16) Zhu, X.; Schmidt, R. R. *Tetrahedron Lett.* **2003**, *44*, 6063–6067.

(17) Eisele, T.; Schmidt, R. R. *Liebigs Ann.* **1997**, 865–887.

(18) Kiefel, M. J.; Thomson, R. J.; Radovanovic, M.; Itzstein, M. V. *J. Carbohydr. Chem.* **1999**, *18*, 937–959.

(19) Driguez, H. *ChemBioChem* **2001**, *2*, 311–318.

(20) Pachamuthu, K.; Schmidt, R. R. *Chem. Rev.* **2006**, *106*, 160–187.

(21) Novoa, A.; Barluenga, S.; Serba, C.; Winssinger, N. *Chem. Commun.* **2013**, *49*, 7608–7610.

(22) Stepper, J.; Shastri, S.; Loo, T. S.; Preston, J. C.; Novak, P.; Man, P.; Moore, C. H.; Havlicek, V.; Patchett, M. L.; Norris, G. E. *FEBS Lett.* **2011**, *585*, 645–650.

(23) Venugopal, H.; Edwards, P. J. B.; Schwalbe, M.; Claridge, J. K.; Libich, D. S.; Stepper, J.; Loo, T.; Patchett, M. L.; Norris, G. E.; Pascal, S. M. *Biochemistry* **2011**, *50*, 2748–2755.

(24) Oman, T. J.; Boettcher, J. M.; Wang, H.; Okalibe, X. N.; van der Donk, W. A. *Nat. Chem. Biol.* **2011**, *7*, 78–80.

(25) Wang, H.; van der Donk, W. A. *J. Am. Chem. Soc.* **2011**, *133*, 16394–16397.

(26) Katayama, H.; Asahina, Y.; Hojo, H. *J. Pept. Sci.* **2011**, *17*, 818–821.

(27) Hassan, H.; Reis, C. A.; Bennett, E. P.; Mirgorodskaya, E.; Roepstorff, P.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. *J. Biol. Chem.* **2000**, *275*, 38197–38205.

(28) Cohen, S. B.; Halcomb, R. L. *J. Am. Chem. Soc.* **2002**, *124*, 2534–2543.

(29) Zhu, Y.; van der Donk, W. A. *Org. Lett.* **2001**, *3*, 1189–1192.

(30) Galonic, D. P.; van der Donk, W. A.; Gin, D. Y. *Chem.—Eur. J.* **2003**, *9*, 5997–6006. b) Bogart, J. W.; Bowers, A. A. *Org. Biomol. Chem.* **2019**, *17*, 3653–3669

(31) a) Bonauer, C.; Walenzyk, T.; König, B. *Synthesis* **2006**, *1*, 1–20.

(32) Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007**, *61*, 477–501.

(33) Jack, R. W.; Jung, G. *Curr. Opin. Chem. Biol.* **2000**, *4*, 310–317.

(34) Jung, G. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1051–1068.

(35) Levengood, M. R.; van der Donk, W. A. *Nat. Protoc.* **2007**, *1*, 3001–3008.

(36) Wang, J.; Schiller, S. M.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2007**, *46*, 6849–6851.

(37) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 5052–5053.

(38) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2008**, *47*, 6399–6401.

(39) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. *Chem. Asian J.* **2009**, *4*, 630–640.

(40) Aydillo, C.; Avenoza, A.; Busto, J. H.; Jiménez-Osés, G.; Peregrina, J. M.; Zurbano, M. M. *Org. Lett.* **2012**, *14*, 334–337.

(41) Enders, D., Lüttgen, K.; Narine, A. *Synthesis.* **2007**, 959–980.

(42) Aydillo, C.; Avenoza, A.; Busto, J. H.; Jiménez-Osés, G.; Peregrina, J. M.; Zurbano, M. M. *Tetrahedron: Asymmetry* **2008**, *19*, 2829–2834.

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

(44) Caraballo, R.; Deng, L.; Amorim, L.; Brinck, T.; Ramström, O. *J. Org. Chem.* **2010**, *75*, 6115–6121.

(45) Černý, M.; Stanek, J.; Pacák, J. *Monatsh. Chem.* **1963**, *94*, 290–294.

(46) Frgala, J.; Černý, M.; Stanek, J. *Collect. Czech. Chem. Commun.* **1975**, *40*, 1411–1425.

(47) Dere, R. T.; Wang, X.; Zhu, X. *Org. Biol. Chem.* **2008**, *6*, 2061–2063.

(48) Horton, D.; Wander, J. D. In Carbohydrates: Chemistry and Biochemistry; Pigman, W. W.; Horton, D.; Academic Press: New York, **1990**

(49) Matta, K. L.; Girotra, R. N.; Barlow, J. L. *Carbohydr. Res.* **1975**, *43*,

101–109.

(50) Morais, L. L.; Bennis, K.; Ripoche, I.; Liao, L.; Auzanneau, F.-I.; Gelas, J. *Carbohydr. Res.* **2003**, *338*, 1369–1379.

(51) Knapp, S.; Myers, D. S. *J. Org. Chem.* **2001**, *66*, 3636–3638.

(52) Knapp, S.; Myers, D. S. *J. Org. Chem.* **2002**, *67*, 2995–2999.

(53) Jana, M.; Misra, A. K. *J. Org. Chem.* **2013**, *78*, 2680–2686.

(54) Zhu, X.; Dere, R. T.; Jiang, J.; Zhang, L.; Wang, X. *J. Org. Chem.* **2011**, *76*, 10187–10197.

(55) Dere, R. T.; Kumar, A.; Kumar, V.; Zhu, X.; Schmidt, R. R. *J. Org. Chem.* **2011**, *76*, 7539–7545.

(56) Käsbeck, L.; Kessler, H. *Liebigs Ann./Recl.* **1997**, 165–167.

(57) Cohen, S. B.; Halcomb, R. L. *Org. Lett.* **2001**, *3*, 405–407.

(58) Zhu, X.; Schmidt, R. R. *Chem.—Eur. J.* **2004**, *10*, 875–887.

(59) Thayer, D. A.; Yu, H. N.; Galan, M. C.; Wong, C.-H. *Angew. Chem., Int. Ed.* **2005**, *44*, 4596–4599.

(60) Hsieh, Y. S. Y.; Wilkinson, B. L.; O'Connell, M. R.; Mackay, J. P.; Matthews, J. M.; Payne, R. J. *Org. Lett.* **2012**, *14*, 1910–1913.

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

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

(63) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457–479.

(64) Pearlman, D. A. *J. Biomol. NMR* **1994**, *4*, 1–16.

(65) Navo, C. D.; Bermejo, I. A.; Oroz, P.; Tovillas, P.; Compañón, I.; Matías, C.; Avenoza, A.; Busto, J. H.; Zurbano, M. M.; Jiménez-Osés, G.; Corzana, F.; Peregrina, J. M. *ACS Omega* **2018,** *3*, 18142–18152.

(66) Mazeau, K.; Tvaroska, I. *Carbohydr. Res.* **1992**, *225*, 27–41.

(67) Guo, Z.; Wang, Q. *Curr. Opin. Chem. Biol.* **2009**, *13*, 608–617.

(68) Wang, Q.; Ekanayaka, S. A.; Wu, J.; Zhang, J.; Guo, Z. *Bioconjugate Chem.* **2008**, *19*, 2060–2067.

(69) Wang, Q.; Guo, Z. *ACS Med. Chem. Lett.* **2011**, *2*, 373–378.

(70) Liu, C.-C.; Ye, X.-S. *Glycoconjugate J.* **2012**, *29*, 259–271.

(71) Mersch, C.; Wagner, S.; Hoffmann-Röder, A. *Synlett* **2009**, 2167–2171.

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

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

(74) Yan, J.; Chen, X.; Wang, F.; Cao, H. *Org. Biomol. Chem.* **2013**, *11*, 842–848.

(75) Urban, D.; Skrydstrup, T.; Beau, J.-M. *Chem. Commun.* **1998**, 955–956.

(76) Röhrig, C. H.; Takhi, M.; Schmidt, R. R. *Synlett* **2001**, 1170–1172.

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

(78) Cipolla, L.; Rescigno, M.; Leone, A.; Peri, F.; La Ferla, B.; Nicotra, F. *Bioorg. Med. Chem.* **2002**, *10*, 1639–1646.

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

(80) Dondoni, A.; Marra, A. *Chem. Rev.* **2000**, *100*, 4395–4422.

(81) Peri, F.; Cipolla, L.; Rescigno, M.; La Ferla, B.; Nicotra, F. *Bioconjugate Chem.* **2001**, *12*, 325–328.

(82) Aydillo, C.; Navo, C. D.; Busto, J. H.; Corzana, F.; Zurbano, M. M.; Avenoza, A.; Peregrina, J. M. *J. Org. Chem.* **2013**, *78*, 10968–10977.

(83) Rich, J. J.; Bundle, D. R. *Org. Lett.* **2004**, *6*, 897–900.

(84) Bousquet, E.; Spadaro, A.; Pappalardo, M. S.; Bernardini, R.; Romeo,

R.; Panza, L.; Ronsisvalle, G. *J. Carbohydr. Chem.* **2000**, *19*, 527–541.

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

(86) Rich, J. J.; Wakarchuk, W. W.; Bundle, D. R. *Chem.—Eur. J.* **2006**, *12*, 845–858.

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

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

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

(90) Geraci, C.; Consoli, G. M.; Gelante, E.; Bousquet, E.; Pappalardo, M. S.; Spadaro, A. *Bioconjugate Chem.* **2008**, *19*, 751–758.

(91) Dam, T. K.; Gerken, T. A.; Cavada, B. S.; Nascimento, K. S.; Moura, T. R.; Brewer, C. F. *J. Biol. Chem.* **2007**, *282*, 28256–28263.

(92) Rao, V. S. R.; Lam, K.; Qasba, P. K. *J. Biomol. Struct. Dyn.* **1998**, *15*, 853–860.

(93) Elgavish, S.; Shaanan, B. *J. Mol. Biol.* **1998**, *277*, 917–932.

(94) Deng, L.; Wang, X.; Uppalapati, S.; Norberg, O.; Dong, H.; Joliton, A.; Yang, M.; Ramström, O. *Pure Appl. Chem.* **2013**, *85*, 1789–1801.

(95) Mayer, M.; Meyer, B. *Angew. Chem., Int. Ed.* **1999**, *38*, 1784–1788.

(96) Ardá, A.; Blasco, P.; Varón-Silva, D.; Schubert, V.; André, S.; Bruix, M.; Cañada, F. J.; Gabius, H.-J.; Unverzagt, C.; Jiménez-Barbero, J. *J. Am. Chem. Soc.* **2013**, *135*, 2667–2675.

# Influencia de la glicosidación en el Reconocimiento molecular de MUC1 por el anticuerpo SM3



5

## 5.1. Introducción

Como se ha mencionado en el capítulo de introducción, dentro del campo de las glicoproteínas podemos encontrar la familia de las mucinas. Dichas mucinas son glicoproteínas de alto peso molecular presentes en la membrana celular. Desde el punto de vista estructural, el primer residuo de carbohidrato unido a la cadena peptídica es generalmente la *N*-acetilgalactosamina, comúnmente conocido como GalNAc. Dicha unión tiene lugar a través de un enlace *O*-glicosídico con los grupos hidroxilo de los residuos de serina o treonina. A estos dos glicosilaminoácidos (α-*O*-GalNAc-Ser/Thr) se les conoce indistintamente, como ya hemos visto en el capítulo anterior, con el nombre de antígeno Tn.

El antígeno Tn está implicado en la infección por VIH[1] y es una de las estructuras asociadas a tumores humanos más específicas.[2] De hecho, se expresa en, aproximadamente, el 90% de los carcinomas, observándose una correlación directa entre la agresividad del carcinoma y la aparición de dicho antigeno.[3] Como consecuencia, el antígeno Tn ha encontrado un uso generalizado como biomarcador y como diana terapéutica con elevado potencial contra el cancer.[2,4–7] **Por ello, conocer las preferencias conformacionales de este estos dos compuesto (con serina y treonina) y dilucidar el mecanismo de reconocimiento por el sistema inmunológico es primordial para el desarrollo de potenciales vacunas terapéuticas contra el cáncer y de nuevos biosensores.**

Como se ha mencionado en el capítulo anterior, nuestro grupo de investigación ha encontrado diferencias conformacionales relevantes entre los dos antígenos (α-O-GalNAc-Ser/Thr), lo cual puede tener importantes implicaciones biológicas.[8–14]

Una de las biomoléculas en las que se puede encontrar el antígeno Tn es la glicoproteína de membrana MUC1. Dicha glicoproteína consiste en la repetición en tándem de la secuencia de 20 aminoácidos HGVTSAPDXRPAPGSTAPPA, la cual posee cinco posibles sitios de glicosilación, tres residuos de treonina y dos de serina (Figura 1).[15,16]

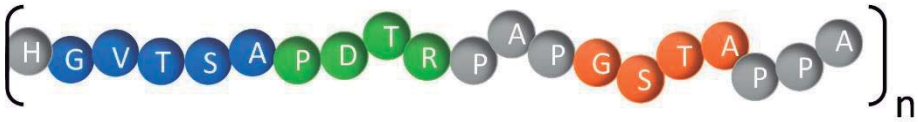



**Figura 1.** *Secuencia de repetición de la MUC1, donde n=20-150.*

Esta proteína esta sobrexpresada en células cancerosas, presentando, además, una glicosilación deficiente con carbohidratos estructuralmente más sencillos. Debido a ello, algunos fragmentos peptídicos que se encuentran enmascarados en células sanas, como el epítopo APDTRP y sus análogos glicosilados, son expuestos, pudiendo ser reconocidos por el sistema inmune (Figura 2).

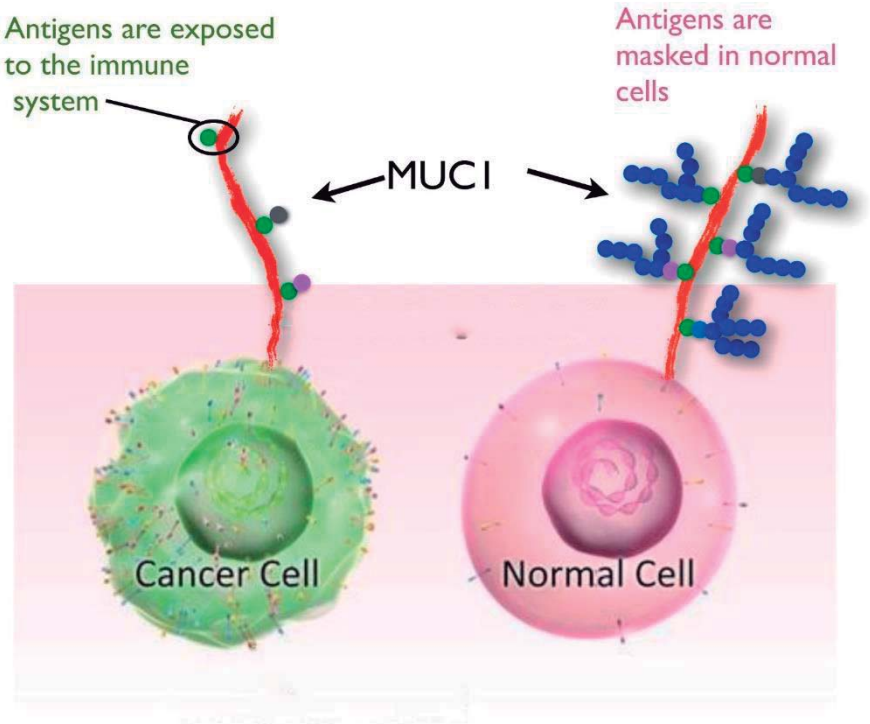


**Figura 2.** *Imagen de una célula sana y una célula cancerígena en la que se puede ver la diferente glicosilación de la MUC1. En una célula tumoral, el número de residuos glicosilados es menor y, además, los carbohidratos unidos al péptido presentan una estructura sencilla.*

En este contexto, la mejora de la afinidad de los epítopos glicosilados con respecto al antígeno "desnudo" ha sido atribuida a los cambios conformacionales inducidos por el glicano en el esqueleto peptídico;[17–20] si bien, hasta la fecha de la publicación del presente trabajo, no se conocía cómo el carbohidrato es reconocido por los anticuerpos anti-MUC1 .

De hecho, únicamente podemos encontrar en la bibliografía la estructura de rayos X del complejo formado por un anticuerpo y un glicopéptido que presenta el carbohidrato GalNAc, pero lamentablemente, dicho péptido no es de tipo mucina.[21] En este contexto, la única estructura de rayos X disponible es la de un péptido no glicosilado de tipo MUC1 y el anticuerpo SM3.[22] Afortunadamente, sí existe un estudio de RMN, basado en experimentos de STD y tr-NOE, del estado asociado de un péptido (PDTRP) y de su correspondiente glicopéptido con GalNAc (PDT*RP). Este estudio determina que el carbohidrato es el encargado de fijar la conformación bioactiva del fragmento peptídico además de interaccionar con la superficie del anticuerpo a través del grupo acetamido del carbohidrato.[23] Sin embargo, no se puede extraer, de este estudio, información precisa de las interacciones, a nivel atómico, existentes en el estado asociado.

## 5.2. Objetivos

El objetivo principal de este capítulo es conocer las preferencias conformacionales de los antígenos Tn de serina y de treonina en el estado asociado con un anticuerpo anti-MUC1 y determinar así el efecto de la

glicosidación en el reconocimiento molecular antígeno-anticuerpo. Para ello se llevarán a cabo los siguientes subobjetivos:

1) Síntesis del péptido **1**, así como de los glicopéptidos **1***, **2*** y **3***, los cuales son los epítopos de reconocimiento APDXRP pertenecientes a la secuencia de la MUC1, donde X=Ser/Thr (Figura 3).

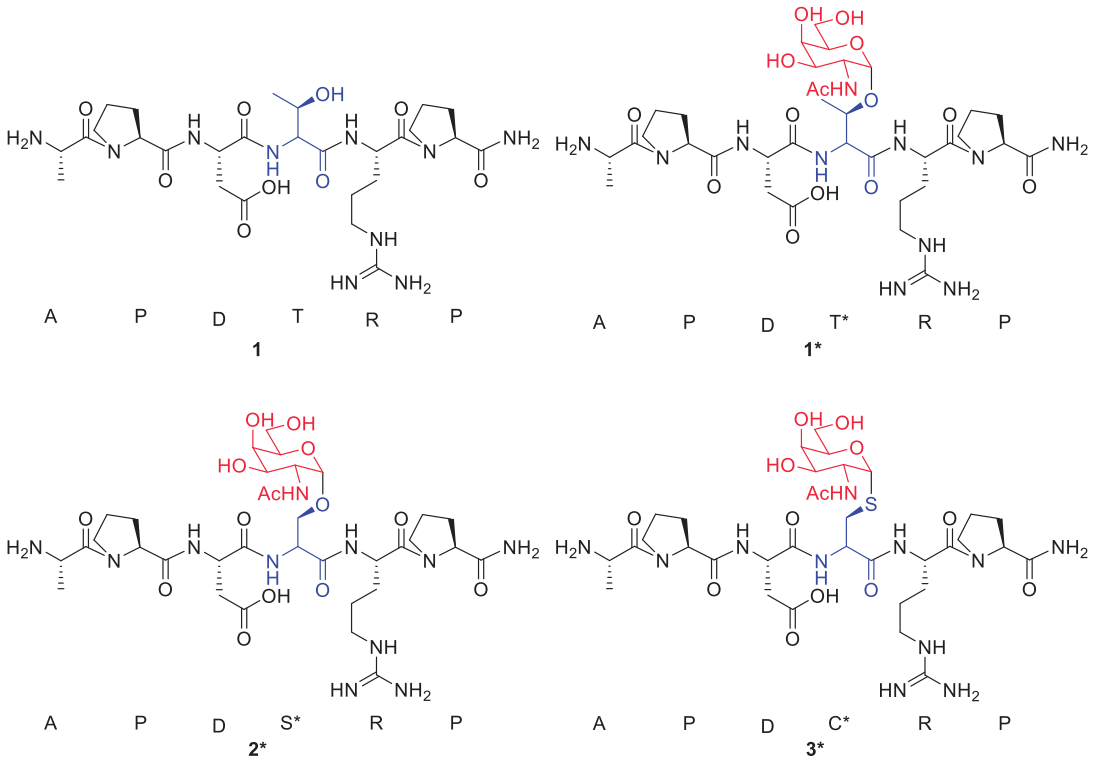



**Figura 3.** *Péptidos y glicopéptidos tipo MUC1 que constituyen el epítopo reconocimiento de anticuerpos anti-MUC1*.

2) Síntesis de los péptidos y glicopéptidos de 21 aminoácidos correspondientes al motivo de repetición en la MUC1 AHGVTSAPDXRPAPGSTAPPA, incorporando en la posición 10, representada con una X, los aminoácidos Thr (**m1**), Ser (**m2**) y los correspondientes glicosilaminoácidos: Thr(α-D-GalNAc) (**m1***) y Ser(α-D-GalNAc) (**m2***) (Figura 4).

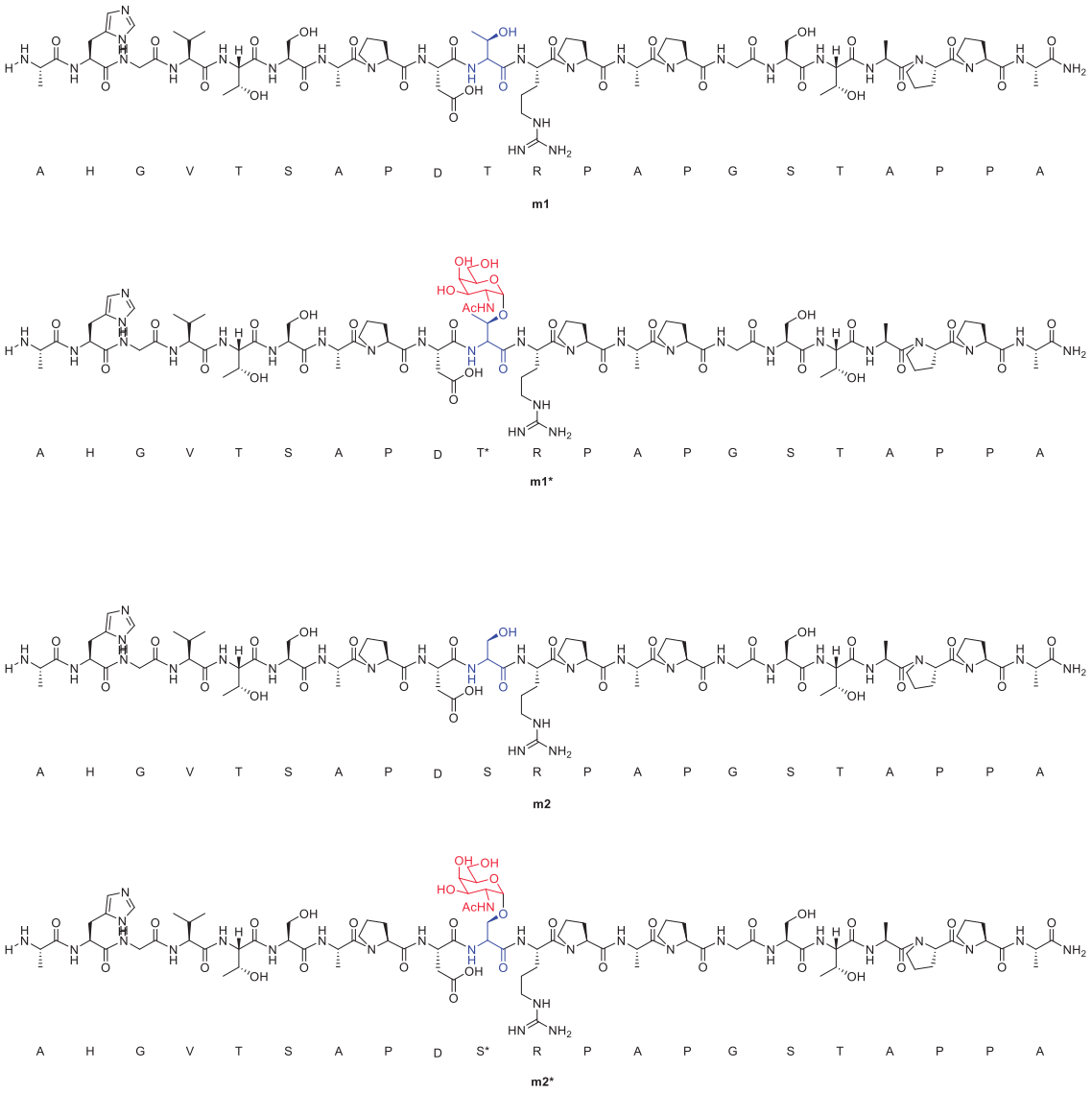



**Figura 4.** *Péptidos y glicopéptidos que incorporan en su estructura la secuencia de repetición de la MUC1.*

3) Estudios de afinidad entre las secuencias de 21 aminoácidos y el scFv-SM3, mediante la utilización de la técnica BLI (del inglés *Bio-layer interferometry*).
4) Estudio conformacional, tanto en el estado libre como asociado, de los epítopos de reconocimiento **1**, **1***, **2*** y **3***. Con motivo de evaluar a nivel atómico las interacciones presentes en el estado asociado al anticuerpo scFv-SM3, se planteó la obtención de las estructuras cristalinas de los complejos scFv-SM3:**1**, scFv-SM3:**1***, scFv-SM3:**2***, scFv-SM3:**3***.

## 5.3. Discusión de resultados

### 5.3.1. Síntesis

Primeramente, se realizó la síntesis de los "*building blocks"* **4** y **5** siguiendo la metodología descrita por nuestro grupo de investigación con anterioridad.[24] Así mismo, se realizó la síntesis del "*building block"* **6** mediante el empleo de la ruta sintética mostrada en el capítulo anterior.

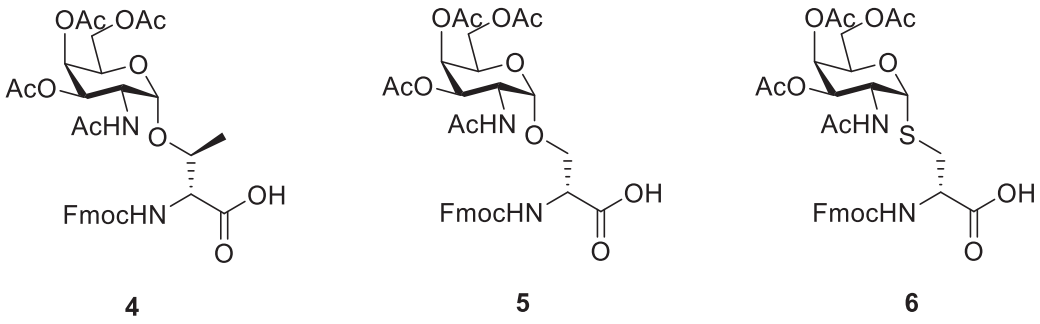



**Figura 5.** *"Building blocks" de los glicoaminoácidos necesarios para la preparación de los glicopéptidos tipo MUC1 por SPPS.*

Una vez sintetizados estos compuestos, se abordó la síntesis de los péptidos **1**, **m1** y **m2** y de los glicopéptidos **m1***, **m2***, **1***, **2*** y **3***.

Todos los péptidos y glicopéptidos se sintetizaron mediante el empleo de la técnica conocida como *"solid phase peptide synthesis"* (SPPS) utilizando la estrategia Fmoc. Para ello se empleó la resina polimérica Rink Amide MHBA (Figura 6) con el objetivo de obtener el extremo C-terminal como amida (-$CONH_2$).

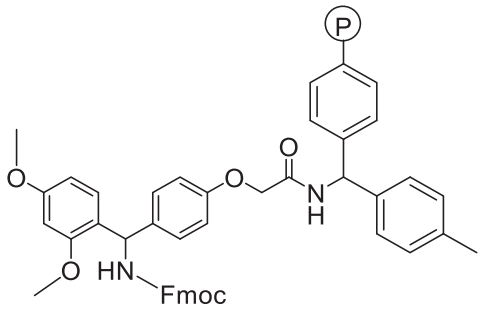



**Figura 6.** *Resina polimerica Rink Amide MHBA.*

Los aminoácidos con el grupo amino protegido con Fmoc y los grupos funcionales de las cadenas laterales con grupos protectores lábiles al medio ácido, se encuentran comercialmente disponibles y son aptos para la SPPS. Por lo tanto, se pudo llevar a cabo los acoplamientos de los péptidos y glicopéptidos de manera automática mediante la utilización de un sintetizador de péptidos Applied Biosystems 433A. Dichos acoplamientos fueron realizados utilizando diez equivalentes del aminoácido convenientemente protegido, hexafluorofosfato de N,N,N',N'-tetrametil-O-(1H-benzotriazolil)uronio (HBTU) como agente de acoplamiento y diisopropiletilamina (DIEA) como base en dimetilformamida (DMF) como disolvente. Llegado el momento de integrar al proceso automático de síntesis los *building blocks* de los glicoaminoácidos **4**, **5** y **6**, sintetizados en el laboratorio siguiendo rutas con varias etapas ,[25] se optó por su incorporación manual, pudiendo así aumentar el tiempo de reacción y reducir a dos el número de equivalentes utilizados. Para comprobar la finalización de dichos acoplamientos se utilizó el test de Kaiser.[26]

Una vez finalizada la síntesis de la cadena peptídica, se llevó a cabo la liberación del péptido de la resina a la vez que se procedió a la desprotección de los grupos de las cadenas laterales. Para ello se utilizó una mezcla de acidio trifluoroacético (TFA), triisopropilsilano (TIS) y agua. Cabe destacar que, en el caso de los glicopéptidos, es necesario realizar, en primer lugar, la

desprotección de los hidroxilos del carbohidrato, empleando una mezcla de hidrazina/MeOH 7:3.

Una vez liberados los compuestos de la resina y filtrada ésta, se añadió éter frío a la disolución, provocando así la precipitación de los (glico)péptidos, que seguidamente, fueron purificados por cromatografía líquida de alta resolución (HPLC).

### 5.3.2. Estudios de afinidad

A continuación, se procedió a evaluar la afinidad de los derivados **m1**, **m2**, **m1*** y **m2*** por el anticuerpo anti-MUC1 SM3. En concreto se utilizó su variante scFv-SM3, el cual es un anticuerpo monocatenario resultado de la unión, mediante un conector de glicinas, de las partes variables de las cadenas pesada y ligera de la inmunoglobulina original.

Dicha evaluación se llevó a cabo mediante la utilización de la técnica BLI. Esta técnica consiste en la inmovilización del anticuerpo en la punta de un biosensor por el cual se hace pasar una disolución de las moléculas objeto de estudio. El aumento del grosor de la capa, provocado por la interacción antígeno-anticuerpo, puede ser medido por el desplazamiento de la longitud de onda del haz de fotones que atraviesa dicha capa. De esta manera se obtuvieron los resultados que se muestran en la figura 7.

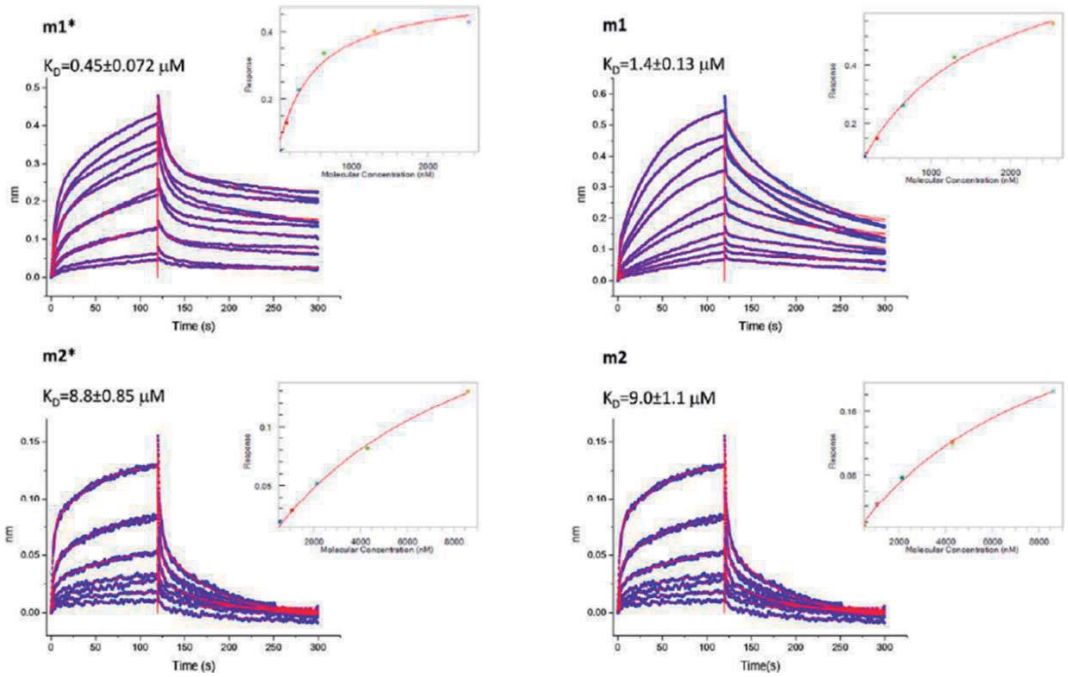



**Figura 7.** *Curvas BLI (en azul). Ajuste de las curvas respuesta vs concentración (en rojo) y $K_D$ obtenidas para **m1**, **m2**, **m1*** y **m2*** con scFv-SM3.*

Como se puede observar en la figura 7, la afinidad de los compuestos glicosilados es mayor ($K_D$ menor), siendo particularmente significativo el valor encontrado para **m1***, el cual presenta una afinidad tres veces mayor que su análogo no glicosilado **m1**.[27] Además, se observa una disminución del valor de afinidad en los derivados que presentan el aminoácido serina, demostrando la no equivalencia entre los dos antígenos Tn (Tn-Ser y Tn-Thr)

### 5.3.3. Estudio conformacional

Con el objetivo de encontrar una explicación a nivel atómico de los resultados obtenidos en los ensayos de afinidad, se expresó y purificó el anticuerpo scFv-SM3. De esta forma se pudo disponer de una cantidad suficiente de anticuerpo para preparar cristales de los distintos complejos. Como resultado, se obtuvieron cristales de buena calidad para los complejos SM3:**1**, SM3:**1***, SM3:**2*** (Figura 8), los cuales se pudieron resolver por difracción de rayos X con una resolución menor de 2 Å.

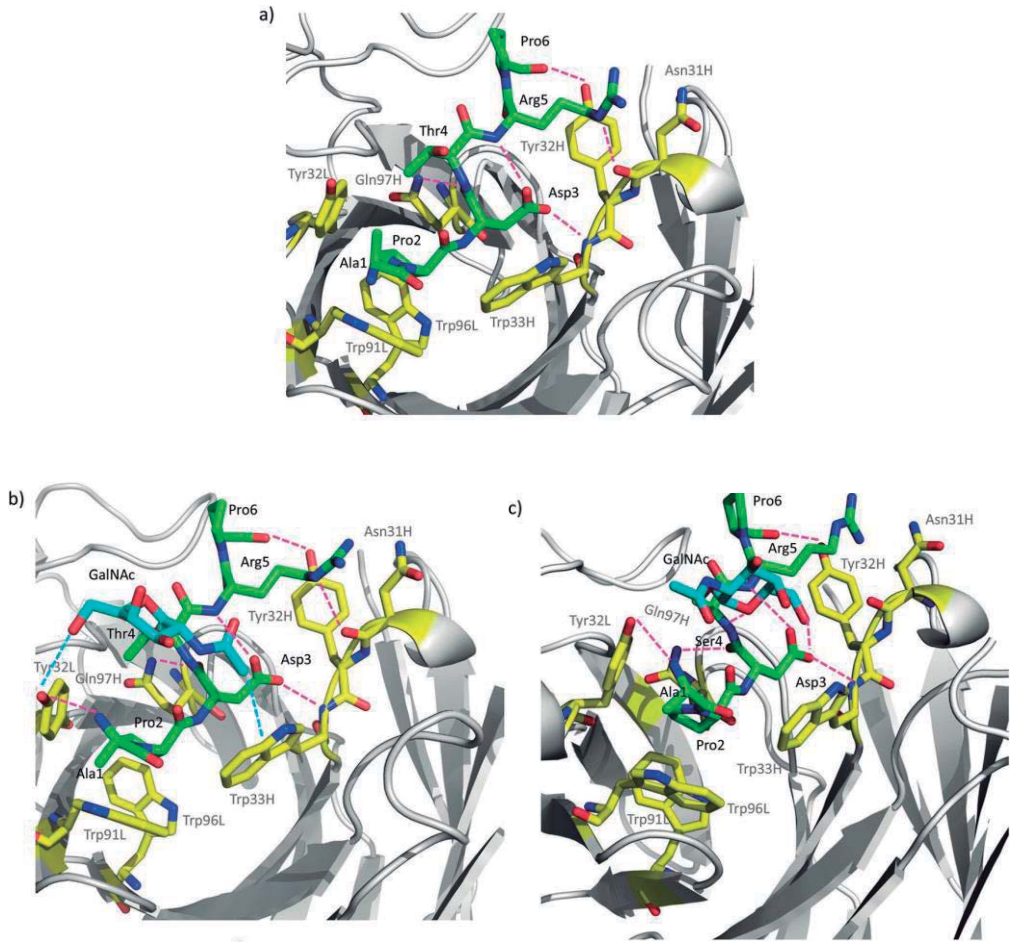



**Figura 8.** *Interacciones principales observadas por difracción de rayos X de los complejos formados por el péptido **1** (a), el glicopéptido **1*** (b) y el glicopéptido **2*** (c) con scFv-SM3. El anticuerpo se muestra en color gris. Las cadenas peptídicas de los antígenos se muestran en verde, los carbonos del GalNAc en azul y los residuos del anticuerpo que interaccionan directamente con el antígeno en amarillo. Las interacciones entre el GalNAc y el anticuerpo se representan mediante líneas azules punteadas y las interacciones de la cadena peptídica con el anticuerpo mediante líneas rosas punteadas. La etiqueta L y H en los distintos aminoácidos del anticuerpo hacen referencia a si estos están en la cadena ligera (light) o en la pesada (heavy), respectivamente.*

El análisis de rayos X revel,o que el sitio de unión del anticuerpo scFv-SM3 se adapta bien a todos los fragmentos peptídicos en los tres complejos estudiados (Figura 8), con independencia de la secuencia (Thr o Ser) y de la presencia o no del GalNAc. Además, la disposición presentada por los

fragmentos peptídicos de las tres variantes simplificadas de la MUC1 son prácticamente idénticas entre sí y a las estructuras descritas anteriormente para un péptido no glicosilado de mayor tamaño (Figura 9).[22]

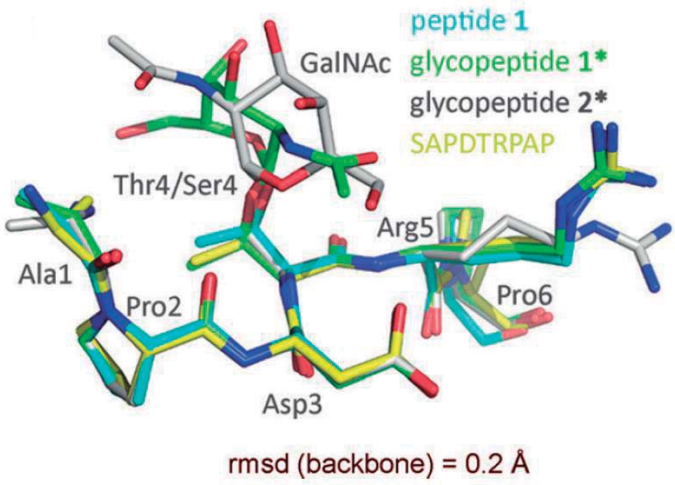



**Figura 9.** *Superposición de los esqueletos peptídicos de los compuestos **1**, **1***, **2*** y del péptido SAPDTRPAP[22] en los complejos con scFv-SM3.*

Por lo tanto, en este caso la presencia del carbohidrato GalNAc, independientemente del aminoácido al que se une, no afecta significativamente a la conformación del esqueleto peptídico en el estado asociado a SM3. De hecho, los pares anticuerpo-antígeno muestran el mismo patrón de interacciones con cada uno de los tres sustratos (Figura 8) e igual a los encontrados en la bibliografía.[22]

Las interacciones establecidas en estos complejos son tanto enlaces de hidrogeno, algunos mediados por moléculas de agua, así como interacciones de tipo CH/π (Figura 8). Por ejemplo, la Pro2 interacciona con los Trp91L y Trp96L y la Tyr32L a través de interacciones CH/π. Igualmente, las cadenas laterales de los residuos Asp3 y Arg5 están implicadas en este tipo de interacciones de *stacking* con el Trp33H y la Tyr32H, respectivamente. Adicionalmente, en los compuestos **1*** y **2***, el grupo NH de la Ala1 y el

grupo carbonilo de Thr4 ( o Ser4 en el compuesto **2***), están involucrados en enlaces de hidrogeno con los residuos Tyr32L y Gln97H, respectivamente. La disposición espacial de la cadena lateral de Arg5 en el glicopéptido **2*** (Figura 9) difiere significativamente de la encontrada para los otros análogos. Como resultado de ello, no se puede establecer el enlace de hidrogeno con el grupo carbonilo de Asn31H, el cual está presente en el resto de los derivados. Esta diferencia, junto con otros factores que se explicarán más adelante, pueden ser la causa de la baja afinidad encontrada por el scFv-SM3 para el glicopéptido **2***.

Es importante señalar que la principal diferencia entre los glicopéptidos **1*** y **2*** en el estado asociado reside en la geometría del enlace glicosídico. De hecho, en el complejo **1***:scFv-SM3, este enlace adopta la conformación exo-anomérica/*syn* esperada, con valores para los ángulos diedros $\phi$ y $\psi$ alrededor de 63° y 91°, respectivamente.[9] Esta conformación es similar a la encontrada en otros complejos que implican al Tn de treonina (Figura 10).[21,28]

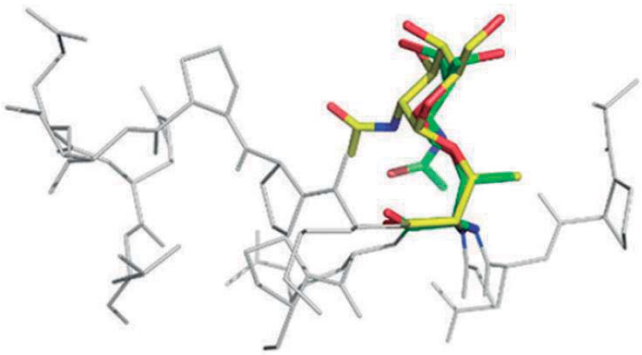


**Figura 10.** *Superposición del antígeno Tn de treonina en el estado asociado a SM3 (en verde) y en el estado asociado a 237-mAb*[21] *(en amarillo).*

Esta geometría del enlace glicosídico permite la formación de un enlace de hidrogeno entre el grupo hidroximetilo del carbohidrato y la Tyr32L del

anticuerpo. Por otra parte, el grupo acetamido del carbohidrato puede establecer una interacción CH/π con el anillo aromático del Trp33H, lo cual puede explicar la elevada afinidad que posee el SM3 por antígenos que incorporan el carbohidrato GalNAc. Cabe destacar que la disposición encontrada para el enlace glicosídico en el derivado de Thr es similar a la observada en otros glicopéptidos con α-*O*-GalNAc-Thr en disolución (Figura 11).[23]

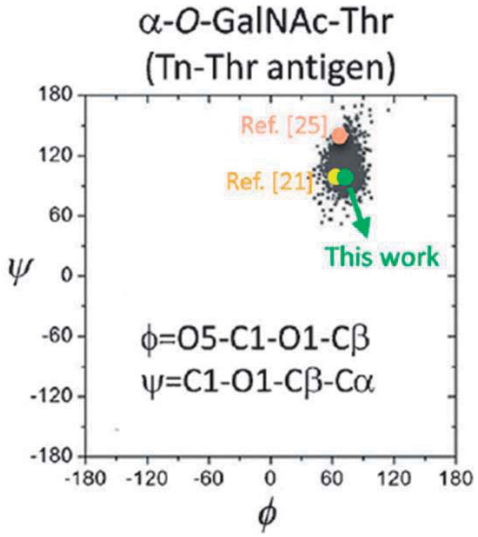



**Figura 11.** *Distribución del enlace glicosídico entre α-O-GalNAc y el residuo de Thr en el glicopéptido **1*** en agua, junto con valores encontrados para dicho enlace en el estado asociado de varios glicopéptidos con proteínas biológicamente relevantes.*[21,28–31]

Por el contrario, en el derivado de serina (compuesto **2***), el carbohidrato no forma puentes de hidrogeno significativos con el SM3. Además, en este derivado, el enlace glicosídico muestra una conformación poco poblada en el estado libre, con valores para el ángulo diedro ψ de -97°. De hecho, el análisis conformacional en el estado libre del glicopéptido **2***, llevado a cabo mediante experimentos de RMN y de MD-tar,[9] indica que el enlace glicosídico adopta esta geometría aproximadamente un 20% del total del tiempo de simulación (Figura 12).

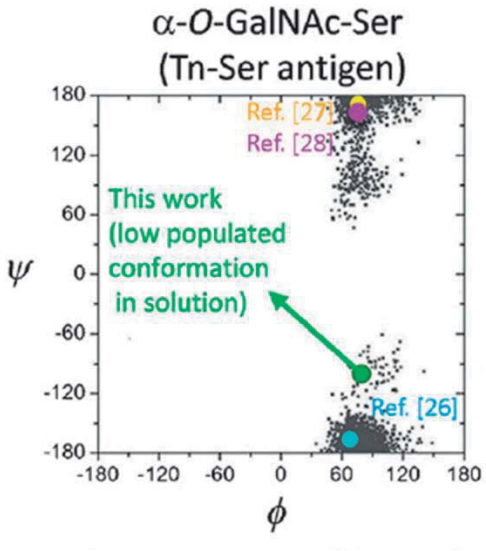



**Figura 12.** *Distribución del enlace glicosídico entre α-O-GalNAc y el residuo de Ser en agua para el glicopéptido* ***2****, junto con las geometrías encontradas para dicho enlace en el estado asociado con diferentes moléculas biológicamente relevantes.*[29–31]

Por lo tanto, en el derivado **1***, el enlace glicosídico adopta la típica conformación eclipsada previamente descrita por nuestro grupo, mientras que el derivado **2*** prefiere adoptar una conformación alternada.[8,9] De acuerdo con la estructura de rayos X, la conformación "inusual" que adopta el enlace glicosídico en el derivado  serina (compuesto **2***) en el estado asociado con scFv-SM3 está estabilizada por la presencia de dos enlaces de hidrogeno intramoleculares con la cadena peptídica. El oxígeno endocíclico del carbohidrato establece un enlace de hidrogeno con el NH del aminoácido de serina, mientras que el O6 interacciona con la cadena lateral del Asp3 (Figura 13).

Es importante remarcar que la calidad de la estructura cristalina del complejo **2***:scFv-SM3 es bastante más baja que la del resto de complejos y que además el factor-B para el glicopéptido **2*** es mucho más alto, pudiendo todo esto indicar que existe un mayor grado de flexibilidad para este complejo. Además, los estudios de MD llevados a cabo para este complejo corroboran dicha flexibilidad, ya que la conformación encontrada en la estructura

cristalina para el enlace glicosídico solo se mantiene durante los cuatro primeros nanosegundos de la trayectoria, adoptando después, la típica orientación alternada que presenta dicho enlace en **2*** el en el estado libre.[8] Con el fin de reforzar la importancia del grupo metilo en el $C_\beta$ de la treonina, a la hora de establecerse la conformación del enlace glicosídico y corroborar la excepcional disposición espacial de dicho enlace en glicopéptido **2*** asociado al anticuerpo SM3, se resolvió la estructura cristalina del complejo entre el análogo glicosilado de cisteína y el anticuerpo scFv-SM3 (Figura 13).

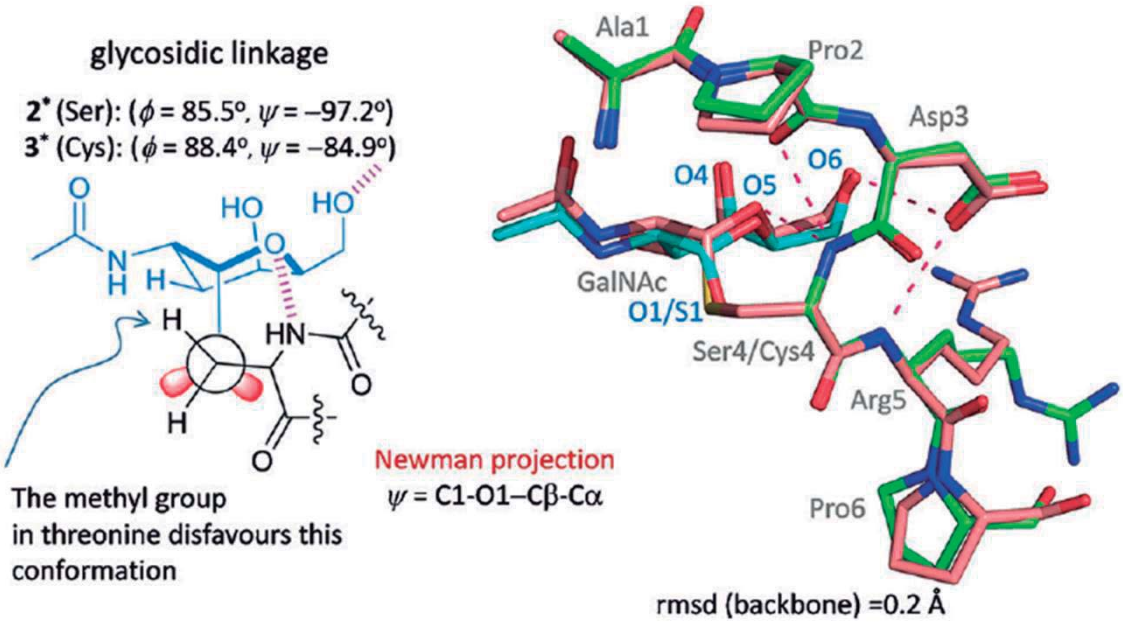



**Figura 13.** *Conformaciones de los glicopéptidos **2*** y **3*** en el complejo asociado a scFv-SM3, junto con la geometría del enlace glicosídico encontrada y las interacciones que la estabilizan.*

En dicha estructura se puede observar que el residuo carbohidrato adopta la misma disposición espacial que en el complejo **2***:scFv-SM3. Esta conformación no es posible en el complejo que contiene el glicopéptido de treonina **1***, ya que el metilo en el $C_\beta$, generaría un importante impedimento estérico con el carbohidrato, desestabilizando de forma significativa la

estructura. Como resultado, en el estado asociado, el carbohidrato adopta una disposición espacial completamente diferente cuando éste está unido al aminoácido treonina, exponiendo la mayoría de los grupos hidroxilos al disolvente.

## 5.4. Conclusiones

Se ha logrado sintetizar el péptido **1**, así como los glicopéptidos **1***, **2*** y **3***, los cuales son los epítopos de reconocimiento APDXRP pertenecientes a la secuencia de la mucina MUC1.

También se ha conseguido sintetizar los péptidos y glicopéptidos de 21 aminoácidos correspondientes al motivo de repetición en la MUC1 **m1**, **m2**, **m1*** y **m2***.

Se han conseguido las estructuras de rayos X de los complejos SM3:**1**, SM3:**1***, SM3:**2*** y SM3:**3***, las cuales han sido muy relevantes a la hora de explicar las diferencias encontradas en las afinidades por el anticuerpo SM3 para los diferentes derivados.

Se ha logrado explicar la diferencia existente en el reconocimiento molecular del péptido **1** y el glicopéptido **1*** por parte del anticuerpo anti-MUC SM3. Esto puede ser debido a las interacciones establecidas entre el carbohidrato y el anticuerpo (enlace de hidrogeno entre grupo hidroximetilo y Tyr32L e interacción CH/π entre el grupo acetamido y Trp32H).

Por otra parte, se han encontrado razones que explican la diferencia de afinidad que muestran los glicopéptidos **1*** y **2*** por el anticuerpo SM3. Mientras que en el caso del derivado **1*** la disposición del enlace glicosídico en disolución es similar a la encontrada en el estado asociado a SM3, en el derivado **2***, la disposición de dicho enlace glicosídico en el estado asociado se encuentra muy poco poblada en disolución.

Estas diferencias encontradas entre los derivados **1*** y **2*** podrían ser importantes a la hora del desarrollo racional de vacunas terapéuticas contra el cáncer.

### 5.5. Parte experimental

**Reagents and general procedures.** Commercial reagents were used without further purification. Solvents were dried and redistilled prior to use in the usual way. $^{1}H$ and $^{13}C$ NMR spectra were measured with a 400 MHz spectrometer with TMS as the internal standard. Multiplicities are quoted as singlet (s), broad singlet (br s), doublet (d), doublet of doublets (dd), triplet (t), or multiplet (m). Spectra were assigned using COSY and HSQC. All NMR chemical shifts (δ) were recorded in ppm and coupling constants (*J*) were reported in Hz. The results of these experiments were processed with MestreNova software. High-resolution electrospray mass (ESI) spectra were recorded on a microTOF spectrometer; accurate mass measurements were achieved by using sodium formate as an external reference.

**NMR experiments.** NMR experiments were performed on a 400 MHz spectrometer at 298 K. Magnitude-mode ge-2D COSY spectra were acquired with gradients by using the *cosygpqf* pulse program with a pulse width of 90º. Phase-sensitive ge-2D HSQC spectra were acquired by using zfilter and selection before t1 removing the decoupling during acquisition by use of the invigpndph pulse program with CNST2 (*J*HC)=145.

**2D ROESY experiments.** ROESY experiments were recorded on a 400 MHz spectrometer at 298 K and pH 6.5 in H2O/D2O (9:1). The experiments were conducted by using phase-sensitive ge-2D ROESY with WATERGATE for H2O/D2O (9:1) spectra. NOEs intensities were

normalized with respect to the diagonal peak at zero mixing time. Distances involving NH protons were semiquantitatively determined by integrating the volume of the corresponding cross-peaks. The number of scans used was 16 and the mixing time was 500 ms.

**Solid-phase (glyco)peptide synthesis.** All (glyco)peptides were synthesized by a stepwise solidphase peptide synthesis using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). The glycosylated amino acid building blocks (2.0 equiv.) were synthesized as described in the
literature and manually coupled using HBTU, while the other Fmoc amino acids (10.0 equiv.).
were automatically coupled on an Applied Biosystems 433A peptide synthesizer using HBTU. The *O*-acetyl groups of (AcO)3GalNAc moiety were removed in a mixture of NH2NH2/MeOH (7:3). The (glyco)peptides were then released from the resin, and all acid sensitive side-chain protecting groups were simultaneously removed using TFA 95%, TIS 2.5%, H2O 2.5%, followed by precipitation with cold diethyl ether. Finally, all (glyco)peptides were purified by HPLC using a Phenomenex Luna C18(2) column (10 μ, 250 mm × 21.2 mm) and a dual absorbance detector. Glycopeptides **m1*** and **m2*** were prepared as described previously by our group.

**Enzyme-linked immunosorbent assay (ELISA).** The ELISA plate (*Maleic anhydride 96-well plates*) was coated with 100 μL/well of a solution of MUC-1 derivatives (0-1000 nmol/well) in a phosphate buffer (0.2 M, pH 7.2) and incubated overnight at 25 ºC. The unbound sites were then blocked by adding 200 μL/well of blocking buffer (*Thermo Scientific SuperBlock Blocking Buffer*). After 1 h at 25 ºC, the blocking buffer was removed and the plate wells were washed 3 × 200 μL/well with PBST (Phosphate-buffered saline, 0.1 M sodium phosphate, 0.15 M sodium chloride, pH 7.2 containing

0.05% Tween-20 detergent). Next, wells were incubated with monoclonal MUC1 antibody SM3 (100 μL, diluted 1/150 in PBST buffer) for 2 h. After washing with PBST (3 × 200 μL/well, 2 min/well), the wells were incubated with *anti-mouse IgG, (H+L), horseradish peroxidase conjugated* (100 μL, diluted 1/3000 in PBST buffer) for 1 h at 25 ºC. All wells were again washed first with PBST (5 × 350 μL/well, 2 min/well) and then with 350 μL of water. 3,3′,5,5′-Tetramethylbenzidine (TMB) was added (90 μL/well) and after incubation for 10 min, the reaction was terminated with the addition of 50 mL/well of stop solution (1 M $H_2SO_4$). Absorbance detection of the wells was immediately performed at 450 nm using an ELISA plate reader. Average absorbance intensities of three replicates were plotted against mucins concentrations.

**Compound 1**

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

HRMS ESI+ (m/z) calcd. For $C_{27}H_{47}N_{10}O_9$ $[M+H]^+$ 655.3527, found 655.3546.

$^1$H NMR (400 MHz, $D_2O$) $\delta$ (ppm) 1.19 (d, 3H, $J$ = 6.4 Hz, $CH_3$ Thr4), 1.55 (d, 3H, $J$ = 7.0 Hz, CH3 Ala1), 1.62 – 2.13 (m, 10H, $2H_{\beta Arg5}$, $2H_{\gamma Arg5}$, $H_{\beta Pro2}$, $H_{\beta Pro6}$, $2H_{\gamma Pro2}$, $2H_{\gamma Pro6}$), 2.27 – 2.39 (m, 2H, $_{H\beta Pro2}$, $H_{\beta Pro6}$), 2.85 – 3.05 (m, 2H, $2H_{\beta Asp3}$), 3.22 (t, 2H, $J$ = 6.7 Hz, $2H_{\delta Arg5}$), 3.60 – 3.87 (m, 4H, $2H_{\delta Pro2}$, $2H_{\delta Pro6}$), 4.18 – 4.27 (m, 1H, $H_{\beta Thr4}$), 4.33 (d, 1H, $J$ = 4.6 Hz, $H_{\alpha Thr4}$), 4.34 – 4.42 (m, 2H, $H_{\alpha Ala1}$, $H_{\alpha Pro2}$), 4.46 – 4.52 (m, 1H, $H_{\alpha Pro6}$), 4.64 – 4.70 (m, 1H, $H_{\alpha Arg}5$), 4.73 – 4.80 (m, 1H, $H_{\alpha Asp3}$). 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 (100 MHz, $D_2O$) $\delta$ (ppm) 15.0 ($CH_3$ Ala1), 18.9 ($CH_3$ Thr4), 24.0, 24.6, 24.7 ($C_{\beta\ Arg5}$, $C_{\gamma\ Pro2}$, $C_{\gamma\ Pro6}$), 27.5 ($C_{\gamma\ Arg5}$), 29.4, 29.6 ($_{C\beta Pro2}$, $C_{\beta Pro6}$), 35.1

($C_{\beta Asp3}$), 40.6 ($C_{\delta Arg5}$), 47.7, 47.9 ($C_{\delta Pro2}$, $C_{\delta Pro6}$), 48.1 ($C_{\alpha Ala1}$), 50.1 ($C_{\alpha Asp3}$), 51.2 ($C_{\alpha Arg5}$), 59.0($C_{\alpha Thr4}$), 60.3, 60.5 ($C_{\alpha Pro2}$, $C_{\alpha Pro6}$), 67.0 ($C_{\beta Thr4}$), 156.8 ($CNH_{Arg5}$) 169.4, 171.4, 171.4, 172.5, 173.7, 173.9, 176.8 (7 CO).
Semi-preparative HPLC: Rt = 15.7 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/wáter+0.1% TFA (2:98) → (15:85), 40 min, λ = 212 nm).

**Compound 1***

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

HRMS ESI+ (m/z) calcd. for $C_{35}H_{60}N_{11}O_{14}$ $[M+H]^+$ 858.4321, found 858.4283.
$^1$H NMR (400 MHz, $D_2O$) $\delta$ (ppm) 1.27 (d, 3H, *J* = 6.3 Hz, $CH3_{Thr4}$), 1.54 (d, 3H, *J* = 7.0 Hz, $CH3_{Ala1}$), 1.61 – 1.79 (m, 3H, $H_{\beta Arg5}$, $2H\gamma_{Arg5}$), 1.82 – 1.97 (m, 3H, $H_{\beta Pro2}$, $H_{\beta Pro6}$, $H_{\beta Arg5}$), 1.99 – 2.15 (m, 7H, NHCOCH3, $2H_{\gamma Pro2}$ , $2H_{\gamma Pro6}$), 2.27 – 2.41 (m, 2H, $H_{\beta Pro2}$, $H_{\beta Pro6}$), 2.85 – 3.06 (m, 2H, $2H_{\beta Asp3}$), 3.24 (t, 2H, *J* = 6.8 Hz, $2H_{\delta Arg5}$), 3.60 – 3.68 (m, 2H, $H_{\delta Pro2}$, $H_{\delta Pro6}$), 3.68 – 3.78 (m, 4H, $H_{\delta Pro2}$, $H_{\delta Pro6}$, $2H_6$), 3.84 – 3.89 (dd, 1H, *J* =11.0, 3.1 Hz, $H_3$), 3.97 (d, 1H, *J* = 2.9 Hz, $H_4$), 4.02 (t, 1H, *J* = 6.4 Hz, $H_5$), 4.13 (dd, 1H, *J* = 11.0, 3.7 Hz, $H_2$), 4.26 – 4.43 (m, 3H, $H_{\alpha Ala1}$, $H_{\beta Thr4}$, $H_{\alpha Arg5}$), 4.48 – 4.58 (m, 3H, $H_{\alpha Thr4}$, $H_{\alpha Pro2}$, $H_{\alpha Pro6}$), 4.58 – 4.68 (m, 1H, $NH_{Arg5}$), 4.85 – 4.92 (m, 2H, $H_{\alpha Asp3}$, $H_1$).
$^1$H NMR (400 MHz, $H_2O/D_2O$, 9:1) δ (ppm) amide protons: 8.61 (t, J = 7.1 Hz, 1H, $NH_{Asp3}$), 8.35 (d, *J* = 8.9 Hz, 1H, $NH_{Thr4}$), 8.31 (d, *J* = 7.4 Hz, 1H, $NH_{Arg5}$), 7.59 (d, *J* = 9.7 Hz, 1H, $NH_{GalNAc}$). 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 (100 MHz, $D_2O$) $\delta$ (ppm) 15.1 ($CH_{3Ala1}$), 18.3 ($CH_{3Thr4}$), 22.3 ($NHCOCH_3$), 24.2, 24.7($C_{\gamma Arg5}$, $C_{\gamma Pro2}$, $C_{\gamma Pro6}$), 27.4 ($C_{\beta Arg5}$), 29.5, 29.5 ($C_{\beta Pro2}$, $C_{\beta Pro6}$), 35.2 ($C_{\beta Asp2}$), 40.5 ($C_{\delta Arg5}$), 47.7, 47.7 ($C_{\delta Pro}$), 48.0 ($C_{\alpha Ala1}$), 49.6, 49.9 ($C_2$, $C_{\alpha Asp3}$), 57.1 ($C_{\alpha Thr4}$), 60.1 ($C_{\alpha Pro}$), 60.2 ($C_{\alpha Arg5}$), 61.3 ($C_6$), 68.2, 68.5, 71.4 ($C_3$, $C_4$, $C_5$), 75.5 ($C_{\beta Thr4}$), 98.6 ($C_1$), 156.8 ($CNH_{Arg5}$), 169.2, 171.0, 171.1, 172.7, 173.5, 173.8, 174.0, 176.8 (8 CO).

Semi-preparative HPLC: Rt = 17.80 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/water+0.1% TFA (2:98) → (15:85), 40 min, $\lambda$ = 212 nm).

**Compound 2***

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

HRMS ESI+ (m/z) calcd. For $C_{34}H_{58}N_{11}O_{14}$ $[M+H]^+$ 844.4159, found 844.4182.

$^1$H NMR (400 MHz, $D_2O$) $\delta$ (ppm) 1.54 (d, 3H, *J* = 7.1 Hz, $CH_{3Ala1}$), 1.61 – 1.82 (m, 3H, $H_{\beta Arg5}$, $2H_{\gamma Arg5}$), 1.83 – 2.14 (m, 10H, $H_{\beta Pro2}$, $H_{\beta Pro6}$, $H_{\beta Arg5}$, $NHCOCH_3$, $2H_{\gamma Pro2}$, $2H_{\gamma Pro6}$), 2.26 – 2.40 (m, 2H, $H_{\beta Pro2}$, $H_{\beta Pro6}$), 2.83 – 3.03 (m, 2H, $2H_{\beta Asp3}$), 3.18-3.28 (m, 2H, $2H_{\delta Arg5}$), 3.58 – 4.00 (m, 11H, $2H_{\delta Pro2}$, $2H_{\delta Pro6}$, $2H_6$, $H_3$, $H_4$, $H_5$, $2H_{\beta Ser4}$), 4.18 (dd, 1H, *J* = 11.0, 3.7 Hz, H2), 4.34 – 4.42 (m, 2H, $H_{\alpha Ala1}$, $H_{\alpha Pro}$), 4.50 (dd, 1H, *J* = 8.3, 6.3 Hz, $H_{\alpha Pro}$), 4.58 – 4.65 (m, 1H, $H_{\alpha Ser4}$) 4.66 – 4.75 (m, 2H, $H_{\alpha Arg5}$, $H_{\alpha Asp3}$), 4.88 (d, 1H, *J* = 3.8 Hz, $H_1$).

$^1$H NMR (400 MHz, $H_2O/D_2O$, 9:1) $\delta$ (ppm) amide protons: 8.65 (t, J = 6.7 Hz, 1H, $NH_{Asp3}$), 8.49 (d, *J* = 7.3 Hz, 1H, $NH_{Arg5}$), 8.41 (d, *J* = 7.2 Hz, 1H, $NH_{Ser4}$), 7.94 (d, *J* = 9.4 Hz, 1H, $NH_{GalNAc}$). 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 (125 MHz, $D_2O$) $\delta$ (ppm) 15.0 ($CH_{3Ala1}$), 22.0 ($NHCOCH_3$), 24.0, 24.7($C_{\gamma Arg5}$, $C_{\gamma Pro2}$, $C_{\gamma Pro6}$), 27.5 ($C_{\beta Arg5}$), 29.3, 29.5 ($C_{\beta Pro2}$, $C_{\beta Pro6}$), 35.4 ($C_{\beta Asp2}$), 40.4 ($C_{\delta Arg5}$), 47.7, 47.8 ($C_{\delta Pro}$), 48.0 ($C_{\alpha Ala1}$), 49.6, ($C_2$), 50.2 ($C_{\alpha Asp3}$), 51.1 ($C_{\alpha Arg5}$) 53.2 ($C_{\alpha Ser4}$), 60.2, 60.2 ($2C_{\alpha Pro}$), 61.0 (C6), 66.7 ($C\beta_{Ser4}$), 67.7, 68.3, 71.3 ($C_3$, $C_4$, $C_5$), 97.7 ($C_1$), 156.7 ($CNH_{Arg5}$), 169.2, 170.8, 171.2, 172.3, 173.6, 174.4, 176.7 (CO).

HPLC Semipreparative HPLC: Rt = 17.30 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/wáter+0.1% TFA (2:98) → (15:85), 40 min, $\lambda$ = 212 nm).

**Compound 3***

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

HRMS ESI+ (m/z) calcd. for $C_{34}H_{58}N_{11}O_{13}S$ $[M+H]^+$ 860.3938, found 860.3936.

$^1$H NMR (400 MHz, $D_2O$) $\delta$ (ppm) 1.55 (d, *J* = 7.0 Hz, 3H, $3H_{\beta Ala1}$), 1.82–1.60 (m, 4H, $H_{\gamma Pro2}$, $H_{\gamma Pro6}$, $2H_{\gamma Arg5}$), 2.14–1.91 (m, 9H, $NHCOCH_3$, $2H_{\beta Arg5}$, $H_{\beta Pro2}$, $H_{\beta Pro6}$, $H_{\gamma Pro2}$, $H_{\gamma Pro6}$), 2.45–2.25 (m, 2H, $H_{\beta Pro2}$, $H_{\beta Pro6}$), 3.16–2.83 (m, 4H, $2H_{\beta Asp3}$, $2H_{\beta Cys4}$), 3.30–3.18 (m, 2H, $2H_{\delta Arg5}$), 3.85–3.59 (m, 7H, $2H_{\delta Pro2}$, $2H_{\delta Pro6}$, $2H_6$, $H_3$), 4.27– 4.17 (m, 1H, $H_5$), 4.00 (d, *J* = 2.9 Hz, 1H, $H_4$), 4.44–4.35 (m, 3H, $H_2$, $H_{\alpha Pro}$, $H_{\alpha Ala1}$), 4.54–4.46 (m, 1H, $H_{\alpha Pro}$), 4.65–4.57 (m, 1H, $H_{\alpha Cys4}$), 4.68 (d, *J* = 3.9 Hz, 1H, $H_{\alpha Arg5}$), 5.58 (d, *J* = 5.4 Hz, 1H, $H_1$), 4.73 (t, *J* = 6.7 Hz, 1H, $H_{\alpha Asp3}$).

$^1$H NMR (400 MHz, $H_2O/D_2O$, 9:1) $\delta$ (ppm) amide protons: 8.67 (t, J = 6.9 Hz, 1H, $NH_{Asp3}$), 8.40 (d, *J* = 7.5 Hz, 1H, $NH_{Cys4}$), 8.36 (d, *J* = 7.4 Hz, 1H, $NH_{Arg5}$), 8.20 (d, *J* = 8.2 Hz, 1H, $NH_{GalNAc}$). 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 (100 MHz, $D_2O$) $\delta$ (ppm) 15.1 ($C_{\beta Ala1}$), 21.9 ($NHCOCH_3$), 24.1 ($C_{\gamma Arg5}$), 24.6 ($C_{\beta Arg5}$), 24.7($C_{\gamma Pro}$), 27.7 ($C_{\gamma Pro}$), 29.3 ($C_{\beta Pro}$), 29.6 ($C_{\beta Pro}$), 32.6 ($C_{\beta Cys4}$), 35.2 ($C_{\beta Asp3}$), 40.6 ($C_{\delta Arg5}$), 47.7 ($C_{\delta Pro}$), 47.9 ($C_{\delta Pro}$), 48.0 ($C\alpha_{Ala1}$), 50.0 ($C_2$), 50.1 ($C_{\alpha Asp3}$), 51.1 ($C_{\alpha Arg5}$), 53.7 ($C_{\alpha Cys4}$), 60.3 ($C_{\alpha Pro}$ , $C_{\alpha Pro}$), 61.1 ($C_6$), 67.6 ($C_3$),68.4 ($C_4$), 71.9 ($C_5$), 85.4 ($C_1$), 156.8 ($CNH_{Arg5}$), 169.2 171.1, 171.4, 171.9, 173.6, 173.9, 174.6, 176.8 (8 CO).

HPLC Semi-preparative HPLC: Rt = 20.70 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/wáter+0.1% TFA (2:98) → (15:85), 40 min, λ = 212 nm).

**Compound m1**

HRMS (ESI+) m/z: calcd. for $C_{83}H_{135}N_{27}O_{28}$: $[M+2H]^{2+}$: 978.9906 found: 978.9917.

Semi-preparative HPLC:Rt = 19.15 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/wáter+0.1% TFA (5:95) → (23:77), 20 min, λ = 212 nm).

**Compound m2**

HRMS (ESI+) m/z: calcd. for $C_{82}H_{133}N_{27}O_{28}$: $[M+2H]^{2+}$: 971.9828 found: 971.9837.

Semi-preparative HPLC: Rt = 18.93 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/wáter+0.1% TFA (5:95) → (23:77), 20 min, λ = 212 nm).

**Compound m1***

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

Semi-preparative HPLC:Rt = 17.4 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/wáter+0.1% TFA (5:95) → (18:82), 30 min, λ = 212 nm).

**Compound m2***

HRMS (ESI+) m/z: calcd. for $C_{90}H_{147}N_{28}O_{33}$: $[M+3H]^{3+}$: 716.0223 found: 716.0234.

Semi-preparative HPLC:Rt = 18.9 min (Phenomenex Luna C18 (2), 21.20×250mm, Grad: acetonitrile/wáter+0.1% TFA (5:95) → (18:82), 30 min, λ = 212 nm).

## 5.6. Bibliografía

(1) Hansen, J. E.; Nielsen, C.; Arendrup, M.; Olofsson, S.; Mathiesen, L.; Nielsen, J. O.; Clausen, Broadly, H. *J. Virol.* **1991**, *65*, 6461–6467.

(2) Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem., Int. Ed.* **2011**, *50*, 1770–1791.

(3) Springer, G. F. *J. Mol. Med.* **1997**, *75*, 594–602.

(4) Buskas, T.; Thompson, P.; Boons, G.-J. *Chem. Commun.* **2009**, 5335–5349.

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

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

(7) Cai, H.; Chen, M.-S.; Sun, Z.-Y.; Zhao, Y.-F.; Kunz, H.; Li, Y.-M. *Angew. Chemie Int. Ed.* **2013**, *52*, 6106–6110.

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

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

(10) 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. *Chem. - A Eur. J.* **2014**, *20*, 12616–12627.

(11) 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. *Cancer Immunol. Immunother.* **2013**, *62*, 1107–1122.

(12) Zhang, Y.; Li, Q.; Rodriguez, L. G.; Gildersleeve, J. C. *J. Am. Chem. Soc.* **2010**, *132*, 9653–9662.

(13) Kanekura, T.; Sakuraba, H.; Matsuzawa, F.; Aikawa, S.; Doi, H.; Hirabayashi, Y.; Yoshii, N.; Fukushige, T.; Kanzaki, T. *J. Dermatol. Sci.* **2005**, *37*, 15–20.

(14) Brister, M. A.; Pandey, A. K.; Bielska, A. A.; Zondlo, N. J. *J. Am. Chem. Soc.* **2014**, *136*, 3803–3816.

(15) Taylor-Papadimitriou, J.; Burchell, J.; Miles, D. .; Dalziel, M. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1999**, *1455*, 301–313.

(16) Tarp, M. A.; Clausen, H. *Biochim. Biophys. Acta - Gen. Subj.* **2008**, *1780*, 546–563.

(17) Barchi, J. J. *Biopolymers* **2013**, *99*, 713–723.

(18) Schuman, J.; Campbell, A. P.; Koganty, R. R.; Longenecker, B. M. *J. Pept. Res.* **2008**, *61*, 91–108.

(19) Kinarsky, L.; Suryanarayanan, G.; Prakash, O.; Paulsen, H.; Clausen, H.; Hanisch, F.-G.; Hollingsworth, M. A.; Sherman, S. *Glycobiology* **2003**, *13*, 929–939.

(20) Matsushita, T.; Ohyabu, N.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Hinou, H.; Nishimura, S.-I. *Biochemistry* **2013**, *52*, 402–414.

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

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

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

(24) Liu, M.; Young, V. G.; Lohani, S.; Live, D.; Barany, G. *Carbohydr. Res.* **2005**, *340* (7), 1273–1285.

(25) Koenigs, W.; Knorr, E. *Berichte der Dtsch. Chem. Gesellschaft* **1901**, *34*, 957–981.

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

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

(28) Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; Coelho, H.; Valero-González, J.; Castro-López, J.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Corzana, F.; Peregrina, J. M. *ACS Chem. Biol.* **2015**, *10*, 747–756.

(29) Sousa, B. L.; Silva Filho, J. C.; Kumar, P.; Pereira, R. I.; Łyskowski, A.; Rocha, B. A. M.; Delatorre, P.; Bezerra, G. A.; Nagano, C. S.; Gruber, K.; Cavada, B. S. *Int. J. Biochem. Cell Biol.* **2015**, *59*, 103–110.

(30) Lescar, J.; Sanchez, J.-F.; Audfray, A.; Coll, J.-L.; Breton, C.; Mitchell, E. P.; Imberty, A. *Glycobiology* **2007**, *17*, 1077–1083.

(31) Babino, A.; Tello, D.; Rojas, A.; Bay, S.; Osinaga, E.; Alzari, P. M. *FEBS Lett.* **2003**, *536*, 106–110.

# Diseño racional de miméticos del epítopo de reconocimiento de la mucina MUC1



6

## 6.1. Introducción

Como se ha mencionado en capítulos anteriores, las mucinas son glicoproteínas que están sobrexpresadas en muchos tejidos cancerígenos.[1–4] Mientras que en las células sanas el esqueleto peptídico de estas moléculas se encuentra enmascarado por glicanos muy complejos, en las células tumorales estos glicanos son estructuralmente mucho más sencillos. En consecuencia, los antígenos asociados a tumores, como el Tn, están expuestos al sistema inmunológico,[5] pudiendo de esta manera ser identificados por anticuerpos anti-MUC1. El fragmento peptídico Pro-Asp-Thr-Arg-Pro[6] constituye una de las unidades mínimas que estos anticuerpos son capaces de reconocer.[7] Los derivados parcialmente glicosilados de estas mucinas han sido utilizados en la preparación de vacunas terapéuticas contra el cáncer.[8–12] Además, estos péptidos pueden actuar como biosensores en la detección de células cancerígenas, ya que se ha demostrado que pacientes con tumores presentan en sangre pequeñas cantidades de anticuerpos anti-MUC1.[13]

Una de las características estructurales de estos glicopéptidos es que la *O*-glicosilación con GalNAc fuerza a que la cadena peptídica adopte una conformación extendida en disolución. Esta conformación está estabilizada bien por la presencia de enlaces de hidrógeno[14] o por moléculas de agua que actúan de "puente" entre el GalNAc y la cadena peptídica.[15–17] Cabe destacar, sin embargo, que la estructura de rayos X del complejo formado por el dominio de reconocimiento del anticuerpo anti-MUC1 SM3[18] (scFv-SM3) y el glicopéptido APDT*RP (T*= αGalNAc-Thr) muestra una conformación tipo α-hélice para la treonina glicosilada (Figura 1).[19]

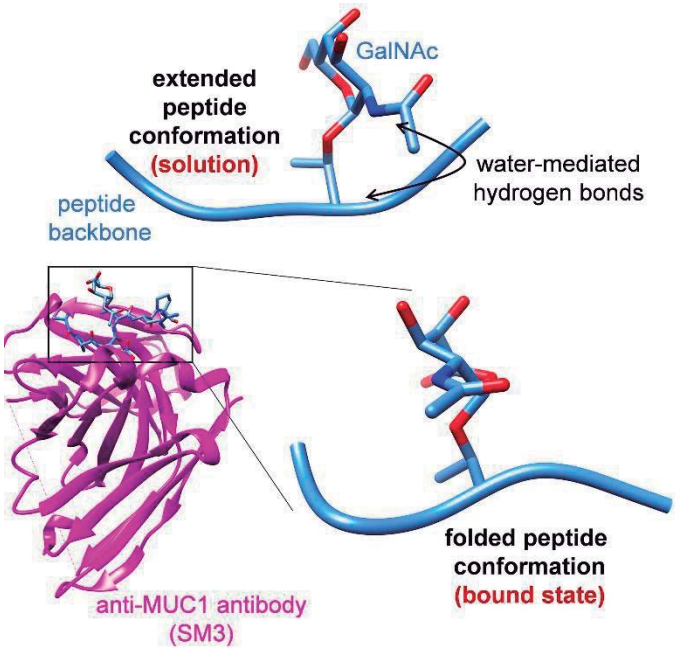



**Figura 1.** *Conformación de un glicopéptido tipo MUC1 en disolución (arriba) y en el estado asociado con el anticuerpo scFv-SM3.*

Como se puede ver en dicha figura, el carbohidrato fuerza a la cadena peptídica a adoptar una estructura en disolución diferente a la reconocida por el anticuerpo. La penalización entrópica, debida al cambio conformacional del péptido de la disolución al estado asociado es compensada por las contribuciones entálpicas[20,21] favorables entre el carbohidrato y el anticuerpo (un enlace de hidrógeno entre el grupo hidroximetilo y un residuo de tirosina y una interacción CH/π entre el metilo del GalNAc y un triptófano). Como resultado, la afinidad del anticuerpo resulta ser tres veces mayor para el glicopéptido en comparación con el correspondiente péptido.

Teniendo en cuenta esto, en este capítulo se propone la sustitución del oxígeno glicosídico por un átomo de azufre o selenio. Esta sustitución aumenta la distancia entre el carbohidrato y el péptido, a la vez que flexibiliza el enlace glicosídico, como consecuencia de la atenuación del efecto exo-anomérico.[22] Por tanto, se espera una menor interacción GalNAc-péptido, reduciéndose así el efecto de la glicosidación en la conformación

del esqueleto peptídico. A priori, es de esperar que los nuevos antígenos presenten conformaciones en el estado libre parecidas a la encontrada en el estado asociado, lo cual se traducirá en un menor coste entrópico asociado a la interacción antígeno-anticuerpo y en el consiguiente aumento de afinidad entre estas entidades (Figura 2).

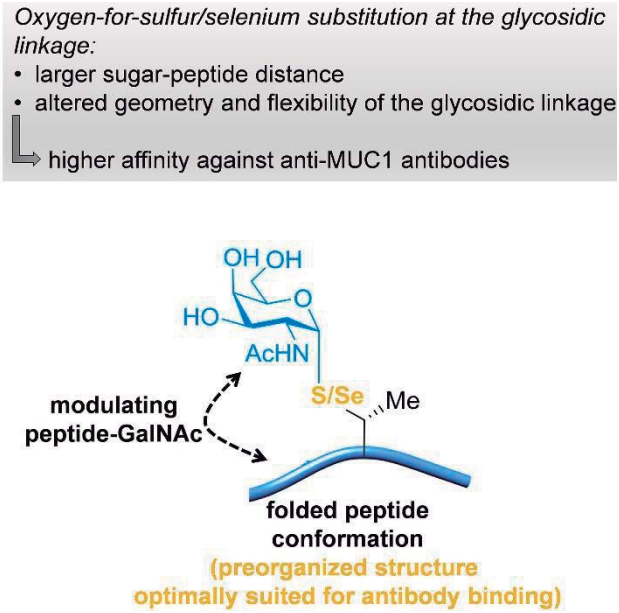



**Figura 2.** *Estrategia propuesta para inducir una preorganización del antígeno en disolución.*

## 6.2. Objetivos

En primer lugar, se planteó la síntesis de los glicopéptidos **1***, **2*** y **3*** y de los péptidos **1** y **2** (Figura 3) mediante la síntesis de péptidos en fase sólida (SPPS).

Con el fin de tener acceso a estos antígenos miméticos, se propuso el diseño de una metodología sintética que permitiese la obtención de los *building blocks* de *Se*-(α-D-GalNAc)-selenotreonina (*Se*Thr*, compuesto **4**) y *S*-(α-D-GalNAc)-tiotreonina (*S*Thr*, compuesto **5**).

A continuación, se determinarán las constantes de afinidad de los nuevos derivados por el scFv-SM3.

Posteriormente, se realizará un estudio conformacional exhaustivo de los mismos, tanto en el estado libre, mediante cálculos de dinámica molecular, como en el estado asociado al anticuerpo scFv-SM3 por difracción de rayos X.

Por último, se propondrá la síntesis de una vacuna terapéutica a partir de la secuencia de 21 aminoácidos que incorpora el glicosilaminoácido *S*Thr*, así como ensayos *in vivo* para caracterizar los anticuerpos producidos en ratones al ser inmunizados por dicha vacuna.

## 6.3. Discusión de resultados

### 6.3.1. Síntesis de los glicopéptidos no naturales

En primer lugar, se realizó la síntesis de los glicopéptidos **2*** y **3***, los cuales incorporan en la posición cuatro los glicosilaminoácidos *S*Thr* y *Se*Thr*, respectivamente. Además, también se realizó la síntesis del glicopéptido **1***, el cual incorpora el glicosilaminoácido natural *O*-(α-D-GalNAc)-Thr (Thr*), así como la síntesis de los péptidos **1** y **2** (Figura 3).

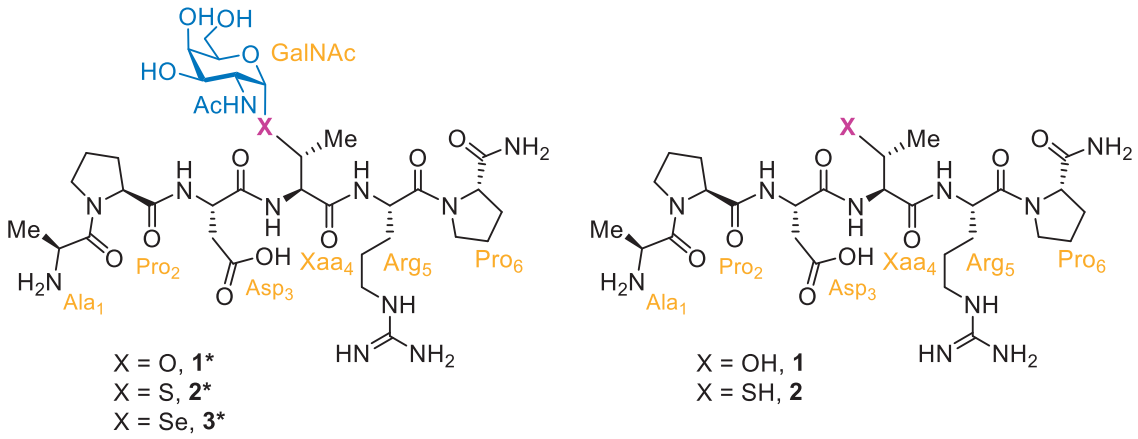
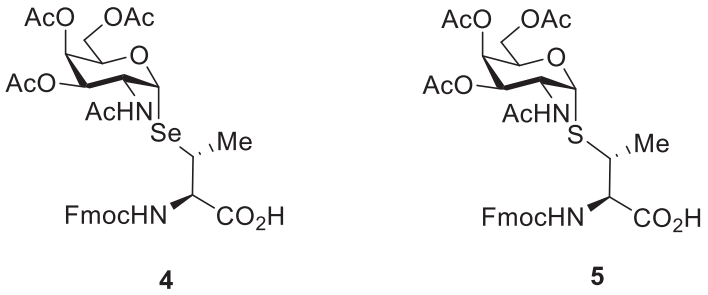



**Figura 3.** *Glicopéptidos y péptidos sintetizados en este capítulo.*

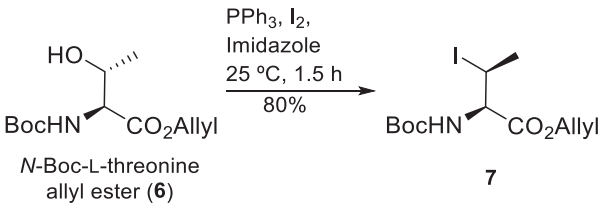
Aunque la síntesis del aminoácido (*S*Thr) ha sido descrita previamente,[23–25] no se conoce la preparación de los *building blocks* de los glicosilaminoácidos *Se*Thr* y *S*Thr*, que están protegidos de forma adecuada para su utilización en SPPS (compuestos **4** y **5**, respectivamente en la figura 4).



**4** **5**

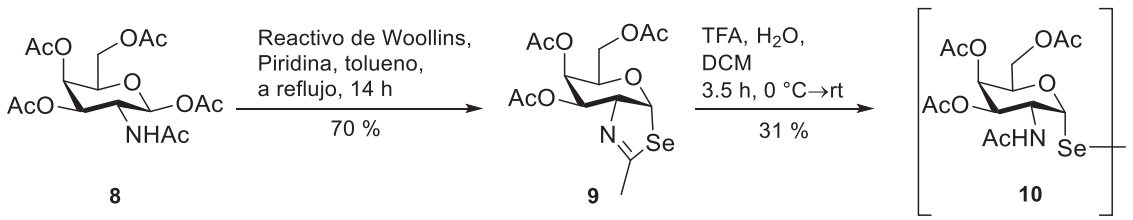
**Figura 4.** *Building blocks de los glicosilaminoacidos SeThr* y SThr*.*

En el esquema 1 se muestra la síntesis del derivado **7**, el cual ha sido adecuadamente protegido para poder ser utilizado en SPPS. Para ello, se hace reaccionar el derivado de treonina **6**, cuya síntesis ha sido descrita con anterioridad, con trifenilfosfina y iodo molecular en presencia de imidazol como base, generándose, de esta manera, el iodo-derivado **7** con inversión de la configuración en el carbono beta (Esquema 1).[26]



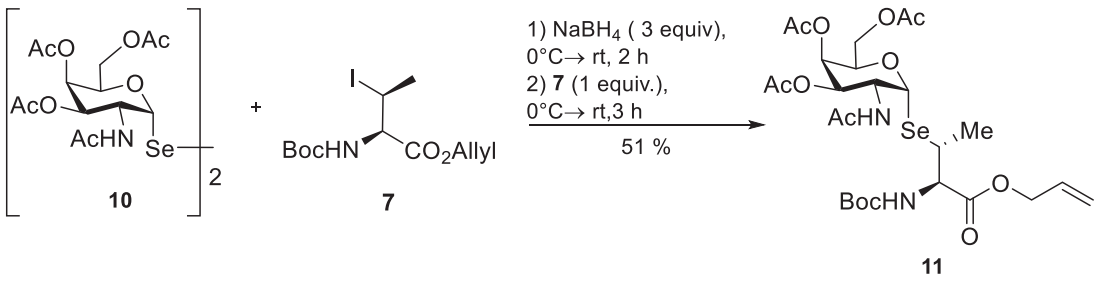

**Esquema 1.** *Síntesis del iodo-derivado* **7.**

De forma paralela, se preparó el selenocarbohidrato **10** en dos etapas de reacción, a partir de la galactosamina peracetilada **8** (Esquema 2). Inicialmente, se trata el compuesto **8** con el reactivo de Woollins en presencia de piridina, generándose así la selenazolina **9** con un rendimiento del 70%. La hidrólisis del compuesto **9** con ácido trifluoroacético (TFA) en agua proporciona con un rendimiento moderado el selenocarbohidrato **10** como dímero, debido a la formación del enlace diseleniuro (Esquema 2).



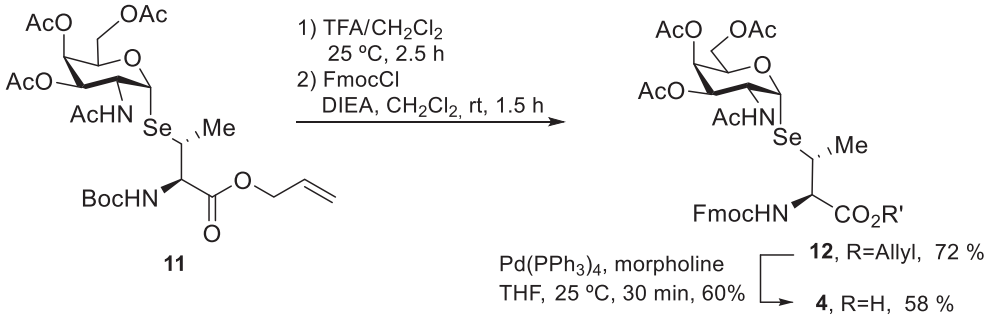

**Esquema 2.** *Síntesis del selenocarbohidrato* **10.**

La etapa clave en la preparación del derivado **11** es el ataque nucleófilo del compuesto **10**, previa reducción *in situ* con borohidruro de sodio, sobre el derivado **7**. Así, se obtiene **11** con un rendimiento del 51% y con inversión total de la configuración en el carbono beta, manteniendo la configuración alfa del carbono anomérico del carbohidrato (Esquema 3).



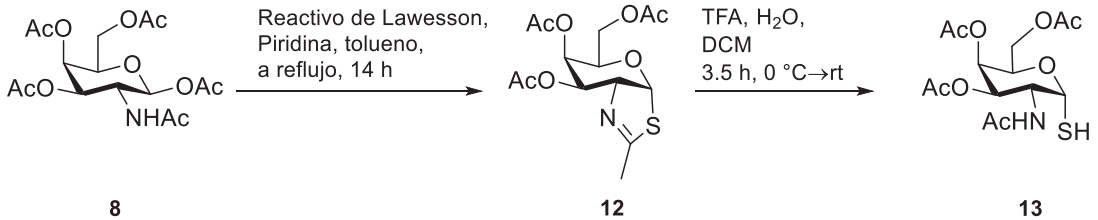

**Esquema 3.** *Preparación del derivado **11**.*

A continuación, sobre el compuesto **11** se llevaron a cabo las correspondientes protecciones y desprotecciones para la incorporación de los grupos protectores requeridos para su utilización en SPPS. El rendimiento global desde el compuesto **11** hasta el *building block* deseado **4** fue del 42% (Esquema 4).



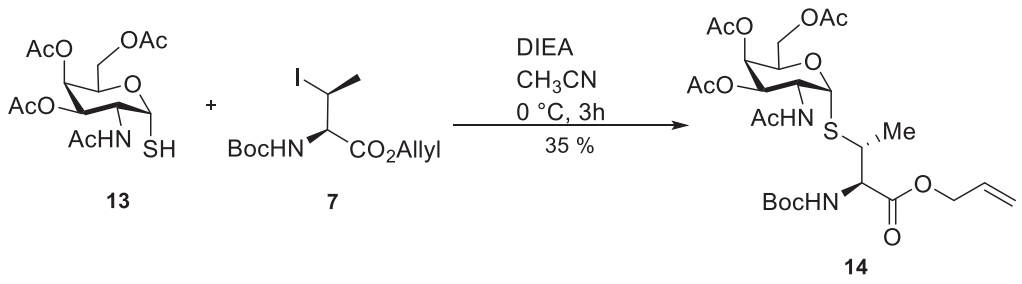

**Esquema 4.** *Derivatización para obtener el building block* **4.**

De manera similar, se preparó el *building block* de *S*Thr* **5**. Primeramente, se procedió con la síntesis del tiocarbohidrato utilizando en este caso el reactivo de Lawesson (Esquema 5).



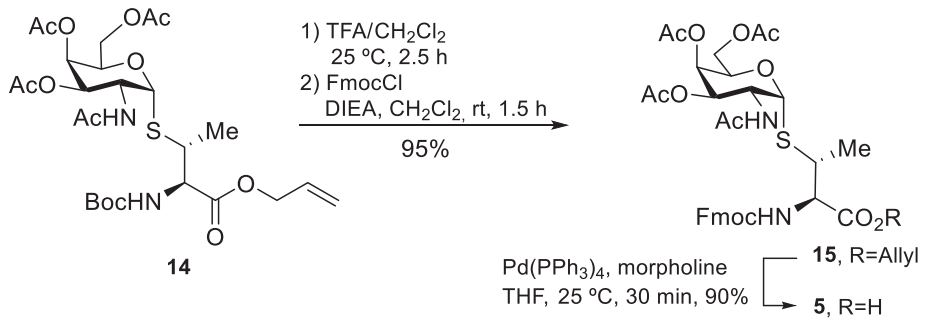

**Esquema 5.** *Síntesis del derivado* **13.**

A diferencia del derivado con selenio, el tiocarbohidrato generado es más estable, lo que permite que este se mantenga como tiol libre sin oxidarse al correspondiente dímero. En consecuencia, no es necesario realizar la reducción con borohidruro de sodio, siendo suficiente la adición de DIEA, como base, en la reacción de sustitución nucleófila sobre el iodo-derivado **7** (Esquema 6).




**Esquema 6.** *Síntesis del derivado **14**.*

Tras llevar a cabo la reacción de glicosilación se procedió a la derivatización del derivado **14** para obtener el *building block* **5** (Esquema 7).




**Esquema 7.** *Derivatización de **14** para obtener el building block **5.***

Una vez sintetizados los *building blocks* **4** y **5**, se procedió con la síntesis de los glicopéptidos. Dicha síntesis se realizó mediante la utilización del sintetizador de péptidos en fase solida asistido por microondas Liberty Blue siguiendo la metodología Fmoc.[13] Al igual que en el capítulo anterior, la incorporación de los glicosilaminoácidos se llevó a cabo fuera del

sintetizador para así poder utilizar únicamente 2 equivalentes de los derivados **4** y **5**.

### 6.3.2. Estudio conformacional

#### 6.3.2.1. Estudio conformacional en disolución

Una vez concluida la síntesis y purificación de los glicopéptidos, se realizó un estudio conformacional de los derivados no naturales **2*** y **3*** en disolución acuosa. Para ello se utilizaron datos de RMN, fundamentalmente distancias derivadas de los espectros ROESY (Figuras 5 y 6) y simulaciones de Dinámica Molecular.

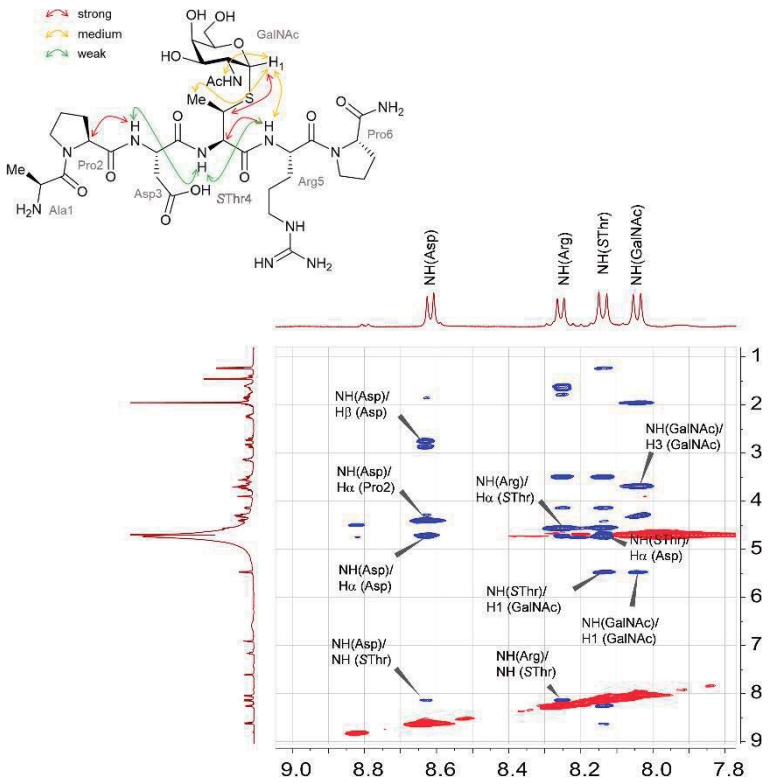



**Figura 5.** *Espectro ROESY en $H_2O/D_2O$ (9:1) a 298 K y pH= 5.6 para el glicopéptido **2**** mostrando la región de las amidas.*

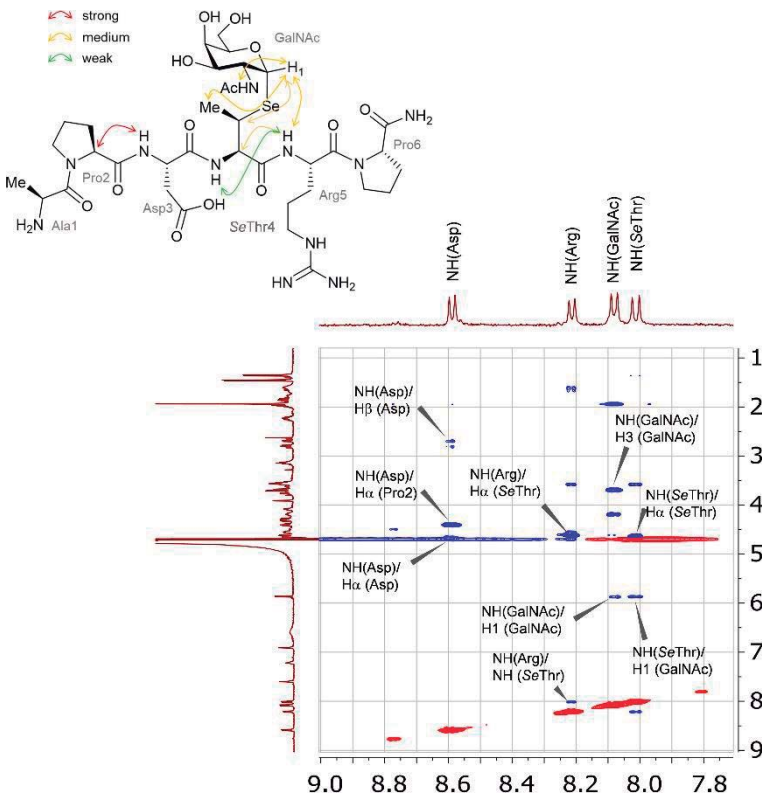



**Figura 6.** *Espectro ROESY en $H_2O/D_2O$ (9:1) a 298 K y pH= 5.6 para el glicopéptido **3*** mostrando la región de las amidas.*

La ausencia de pico de cruce ROESY entre el grupo NH de los aminoácidos *S*Thr y *Se*Thr y el grupo NH del carbohidrato (Figuras 5 y 6 ), el cual es característico de la conformación eclipsada del enlace glicosídico en GalNAcThr,[16,17] junto con la presencia de un pico de cruce entre el grupo NH del aminoácido no natural (*S*Thr4 y *Se*Thr4) y el protón anomérico H1 del carbohidrato, sugiere una conformación diferente para los glicopéptidos **2*** y **3*** respecto al natural **1***. Además, también se observan diferencias en la disposición de la cadena peptídica. En particular, los picos de cruce NH-NH para los residuos 4 y 5 de los glicopéptidos **2*** y **3** sugieren la existencia de una conformación plegada de los aminoácidos *S*Thr y *Se*Thr.[27]

A continuación, varias distancias H-H deducidas de los espectros ROESY (Tablas 1 y 2), se utilizaron como restricciones experimentales en cálculos de DM. En concreto, se emplearon MD con restricciones promediadas en el

tiempo (o también llamadas MD-tar).[28–31]Como puede observarse en estas tablas, la buena concordancia entre las distancias obtenidas experimental y teóricamente, validan los cálculos de MD.

| **Glycopéptido 2*** | | |
|---|---|---|
| **Distancias (Å)** | **Valores experimentales** | **Valores obtenidos por MD-tar** |
| NH(*S*Thr)-H1(GalNAc) | 2.9 | 3.1 |
| NH(*S*Thr)-NH(Arg) | 3.2 | 2.8 |
| NH(Asp)-Hα(Pro2) | 2.2 | 2.3 |
| NH(Asp)-Hα(Asp) | 2.5 | 2.7 |

**Tabla 1.** *Comparación de las distancias obtenidas por RMN y MD para el glicopéptido **2****.

| **Glycopéptido 3*** | | |
|---|---|---|
| **Distancias (Å)** | **Valores experimentales** | **Valores obtenidos por MD-tar** |
| NH(*Se*Thr)-H1(GalNAc) | 2.9 | 3.0 |
| NH(*Se*Thr )-NH(Arg) | 3.2 | 3.0 |
| NH(Asp)-Hα(Pro2) | 2.2 | 2.3 |
| NH(Asp)-Hα(Asp) | 2.5 | 2.8 |

**Tabla 2.** *Comparación de las distancias obtenidas por RMN y MD para el glicopéptido **3****.

Atendiendo a los cálculos de MD-tar, los enlaces *S*- y *Se*-glicosídicos de **2*** y **3*** muestran una única conformación, con valores para los ángulos diedros ϕ/ψ de aproximadamente 65°/70°. Los valores de ϕ están de acuerdo con el efecto exoanomérico,[22] y el ángulo diedro ψ se desvía de la disposición

eclipsada observada para el glicopéptido **1***, con ψ en torno a 120° [16] (Figura 7).

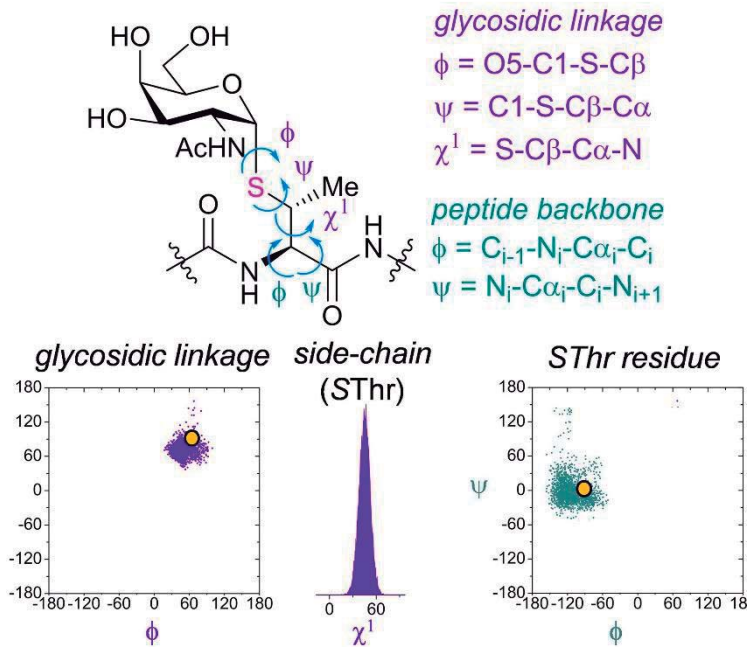



**Figura 7.** *Distribución de los ángulos* ϕ/ψ para el glicopéptido ***2*****,** *así como la geometría del enlace glicosídico. Los puntos amarillos corresponden a los valores encontrados en la estructura de rayos X del glicopéptido natural* ***1**** *asociado a scFv-SM3.*

Es importante remarcar que en el caso del derivado **2***, la conformación del enlace glicosídico determinada en disolución por MD se encuentran en un mínimo local calculado para la metil-4-tio-α-maltosa,[32] siendo, además, muy similar a la disposición encontrada para el antígeno no natural de cisteína descrito en el capítulo 4.[33] Las cadenas laterales de los residuos no naturales en **2*** y **3*** muestran gran rigidez estructural, con conformaciones caracterizadas por tener un ángulo diedro $\chi^1$ ~ 60°. La diferencia en la geometría de los enlaces *S*- y *Se*-glicosídicos con respecto al enlace *O*-glicosídico, junto con el mayor tamaño que poseen los átomos de azufre y selenio, dificultan las interacciones efectivas entre la cadena peptídica y el carbohidrato. De hecho, no se observan ni enlaces de hidrogeno ni bolsillos hidrofílicos entre ambas partes de las moléculas. Como consecuencia, los

compuestos **2*** y **3*** muestran una estructura plegada en torno a los aminoácidos no naturales *S*Thr4 y *Se*Thr4, respectivamente (Figuras 8 y 9).

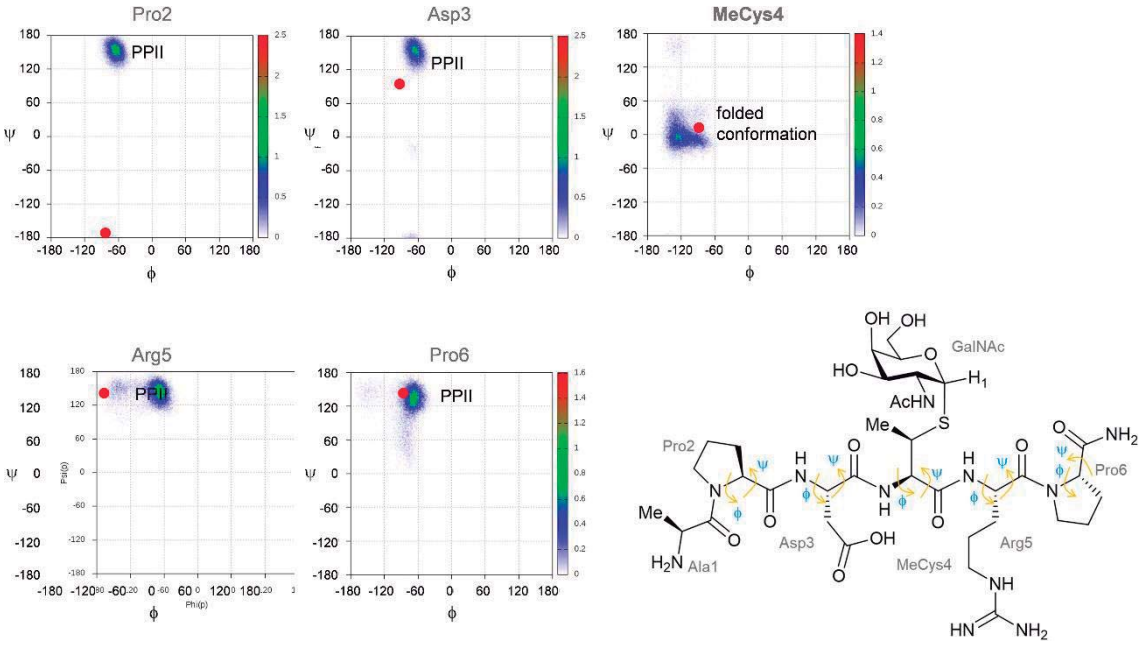



**Figura 8.** *Distribuciones de los ángulos ϕ/ψ del esqueleto peptídico del glicopéptido **2**** obtenidas de los cálculos de MD-tar de 20 ns en agua explicita. Los círculos rojos corresponden a la conformación encontrada por rayos X para los correspondientes aminoácidos del glicopéptido ***1**** en el estado asociado con scFv-SM3. PPII = conformación de poliprolina II.

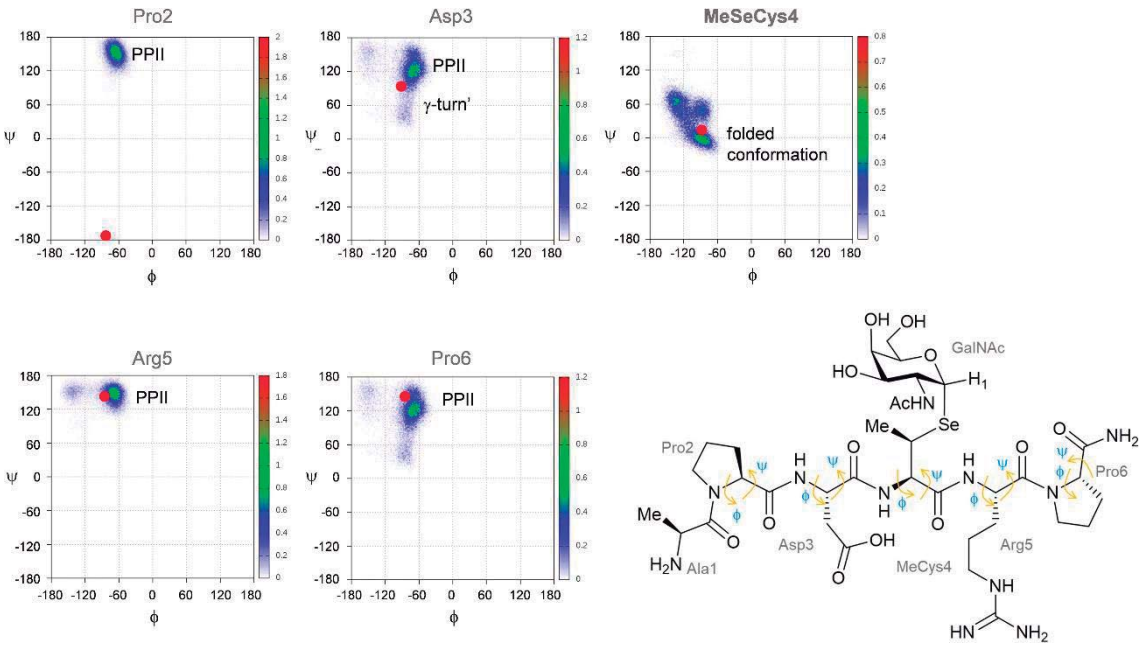


**Figura 9.** *Distribuciones de los ángulos ϕ/ψ del esqueleto peptídico del glicopéptido **3*** obtenidas de los cálculos de MD-tar de 20 ns en agua explicita. Los círculos rojos corresponden a la conformación encontrada por rayos X para los correspondientes aminoácidos del glicopéptido **1*** en el estado asociado con scFv-SM3. PPII = conformación de poliprolina II.*

Estas conformaciones son diferentes a la mostrada por el glicopéptido natural **1***, el cual adopta una conformación tipo PPII en disolución (Figura 11).[17,31]

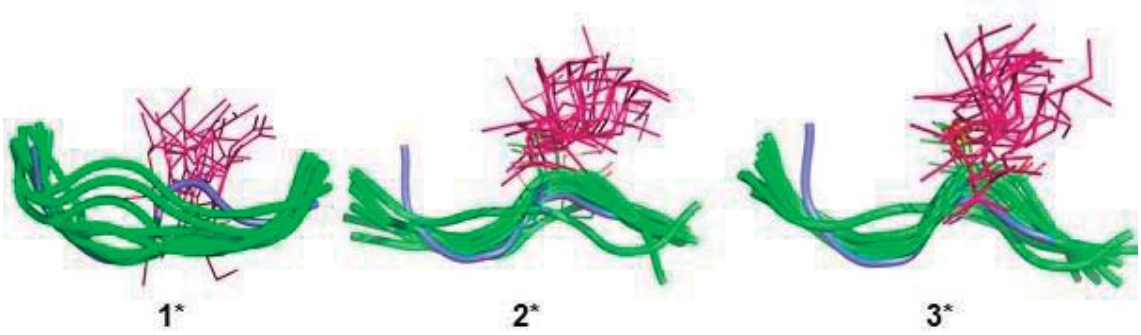



**Figura 10.** *Superposición de varias conformaciones calculadas por MD-tar de los glicopéptidos **1***, **2*** y **3*** en disolución. En verde se muestra la cadena peptídica y en fucsia el residuo de GaNAc. La conformación del esqueleto peptídico del derivado **1*** en la estructura de rayos X **1***:scFv-SM3 aparece en color púrpura.*

Estas diferencias en la disposición del esqueleto peptídico están avaladas por dicroísmo circular. De hecho, como puede verse en la figura 11, y a pesar de ser estructuras pequeñas, hay una clara diferencia entre los espectros registrados para los compuestos **1*** y **2***.

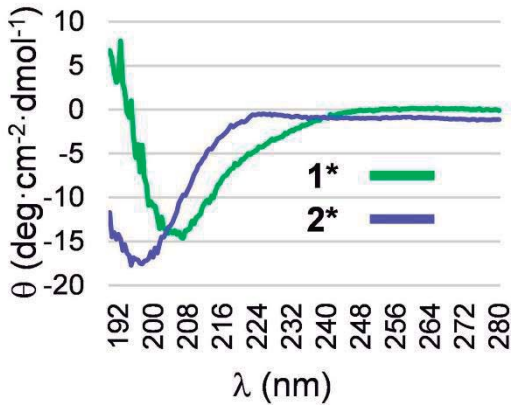


**Figura 11.** *Espectro de dicroísmo circular de los glicopéptidos **1*** y **2*** en buffer fosfato, (0.25 mM, pH = 7.5) a 20° C.*

Además, como muestra la MD sin restricciones efectuada para el péptido **2**, éste adopta una conformación desordenada (*random coil)* en disolución. Este resultado sugiere que el carbohidrato juega un papel en las preferencias conformacionales del fragmento péptido en **2*** y **3*,** a pesar de la ausencia de interacciones importantes entre el GAlNAc y el esqueleto peptídico. En este caso, se disminuyen las conformaciones que puede adoptar el péptido debido a los impedimentos estéricos derivados a la presencia del carbohidrato (Figura 12).

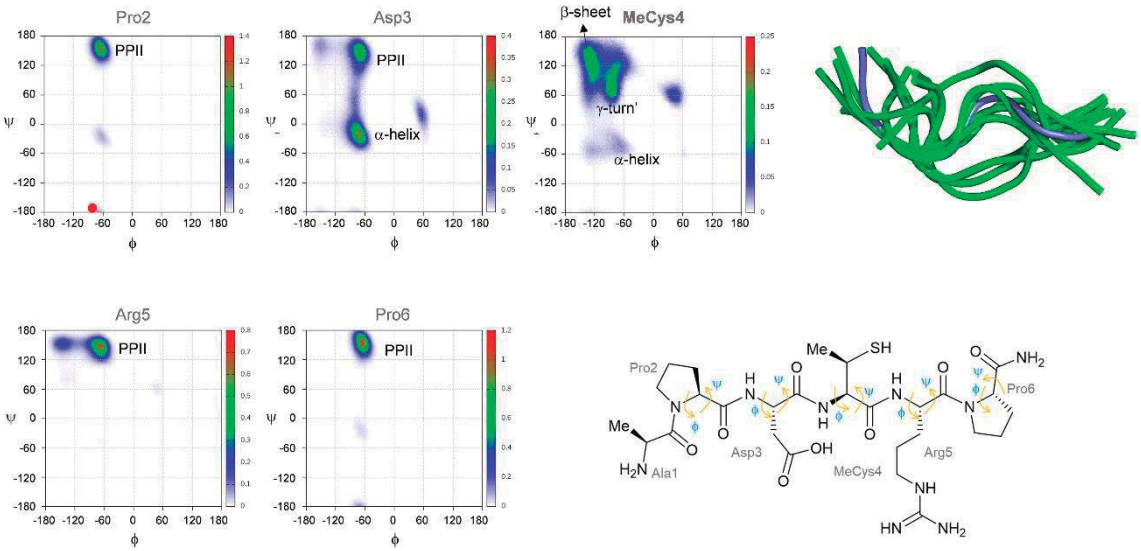



**Figura 12.** *Distribuciones de los ángulos ϕ/ψ del esqueleto peptídico del péptido **2** obtenidas de simulaciones de MD-tar de 1 μs en agua explicita, junto a un 'ensemble' estructural del esqueleto peptídico mostrado en color verde. La conformación del péptido encontrada en la estructura de rayos X del complejo **1:**scFv-SM3 aparece en color púrpura. PPII = conformación de poliprolina II.*

Con todos estos datos, se puede afirmar que, en general, los glicopéptidos **2*** y **3*** presentan una conformación en disolución diferente a la dispuesta por el antígeno natural **1***, simplemente por el cambio de un átomo (O→S/Se). En particular, las preferencias conformacionales del enlace glicosídico y el

esqueleto peptídico encontradas para los glicopéptidos **2*** y **3*** son muy próximas a las mostradas por el antígeno natural **1*** cuando este se encuentra asociado a anticuerpos anti-MUC1.[19] Es de esperar, por tanto, que el coste energético asociado al cambio conformacional desde el estado libre en disolución hasta el estado asociado será menor para los glicopéptidos **2*** y **3***, con la consiguiente mejora en la afinidad.

### 6.3.2.2. Análisis conformacional en el estado asociado

Con el fin de obtener la estructura en el estado asociado, se prepararon cristales de los complejos **2***:scFv-SM3 y **3***:scFv-SM3, algunos de los cuales fueron de una calidad adecuada para llevar a cabo difracción de rayos X (Figura 13).

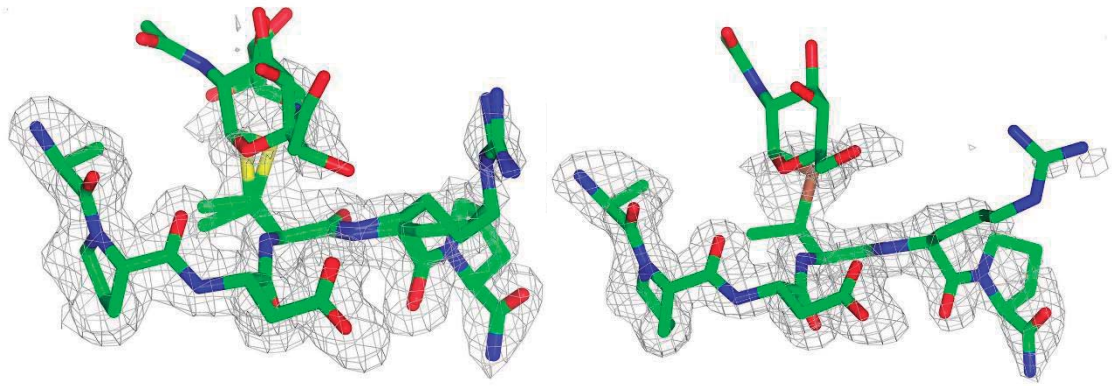


**Figura 13.** *Mapas de densidad electrónica de los glicopéptidos **2*** (izquierda) y **3*** (derecha) asociados a scFv-SM3.*

Las estructuras resueltas por difracción de rayos X de estos complejos revelan que la disposición adoptada por los glicopéptidos **2*** y **3*** unidos al anticuerpo es prácticamente idéntica a la adoptada por el glicopéptido **1*** en el complejo **1***:scFv-SM3 (Figura 14).

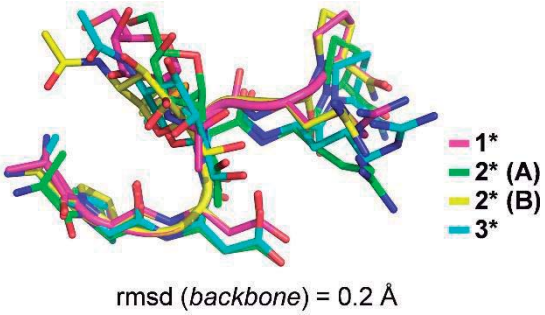



**Figura 14**. *Superposición de los glicopéptidos **1 ***, **2 *** y **3 *** en complejo con el anticuerpo scFv-SM3. Dichas estructuras se han obtenido por difracción de rayos X. A y B son los dos modos de unión encontrados para el derivado **2****

Estos resultados demuestran que el anticuerpo reconoce una conformación específica de los antígenos, independientemente de la naturaleza del aminoácido glicosilado presente en el glicopéptido. Al igual que para el glicopéptido **1***, las interacciones que estabilizan los complejos **2***:scFv-SM3 y **3***:scFv-SM3 son en su mayoría enlaces de hidrogeno entre el antígeno y el anticuerpo, algunos de ellos mediados por moléculas de agua, así como interacciones de tipo CH/π.

Cabe destacar que, en el caso del complejo **2***:scFvSM3, se pueden observar dos geometrías diferentes para el enlace glicosídico, apareciendo así dos modos de unión (llamados A y B en la Figura 15).

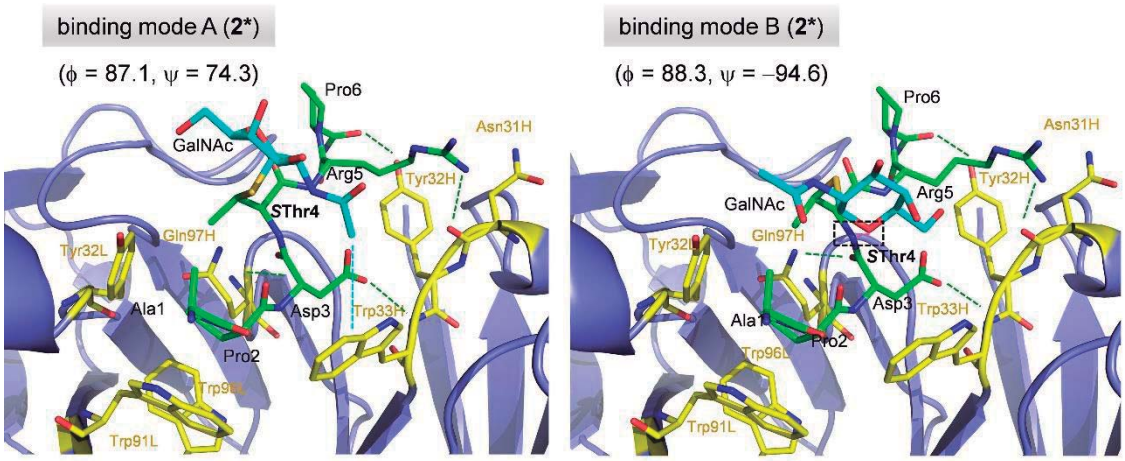


**Figura 15.** *Modos de unión encontrados en la estructura de rayos X del complejo 2*:scFv-SM3.*

El modo de unión A se caracteriza por poseer unos valores de ángulos diedros ϕ y ψ en torno a 87° y 74° respectivamente, mientras que en el B estos valores son próximos a 90° y -90°. Este segundo modo de unión es estabilizado por la presencia de un enlace de hidrogeno intramolecular establecido por el grupo NH del aminoácido *S*Thr4 y el oxígeno encíclico (O5) del carbohidrato (Figura 15). La disposición encontrada en B es igual a la que se observa en los complejos establecidos con las mucinas de serina y cisteína explicados en el capítulo anterior.[19] El modo de unión A permite que el grupo NHAc del carbohidrato establezca interacciones de tipo CH/π con el anillo aromático de un residuo de triptófano (Trp 33H) del scFv-SM3. Por el contrario, el modo B impide la existencia de contactos directos entre el carbohidrato y el anticuerpo.

Es importante señalar que la densidad electrónica del carbohidrato observada por difracción de rayos X en el complejo con el glicopéptido **3*** es bastante débil, lo que sugiere también la presencia de los dos modos de unión A y B de manera simultánea (Figura 13).

Por tanto, la presencia de estos dos modos de unión indica la flexibilidad del enlace glicosídico en el estado asociado, que también se observa en parte en disolución, como consecuencia de la disminución del efecto exo-anomérico al sustituir el oxígeno por azufre o selenio.[22]

### 6.3.3. Estudios de afinidad de los glicopéptidos **2*** y **3***

El estudio conformacional de los glicopéptidos **2*** y **3***, tanto en disolución como en el estado asociado a scFv-SM3, revela que estos antígenos no naturales presentan una preorganización en disolución de la conformación de la cadena peptídica reconocida por el anticuerpo que no existe en el derivado natural **1***. En consecuencia, cabría esperar que los glicopéptidos **2*** y **3*** posean una mayor afinidad frente al scFv-SM3 que la presentada por el antígeno **1***. Con el fin de confirmar dicha hipótesis, se realizaron ensayos de afinidad mediante la utilización de la técnica de resonancia de superficie de plasmón (SPR, del inglés *surface plasmon resonance*). Las medidas de SPR revelan que las afinidades presentadas por los glicopéptidos **2*** y **3*** son unas tres veces superiores a las mostradas por el antígeno **1***. Además, se puede observar que el valor de afinidad del glicopéptido **1*** es más sensible a la variación de temperatura, lo cual avala la preorganización existente en los glicopéptidos **2*** y **3*** predicha por el estudio conformacional (Figura 16).

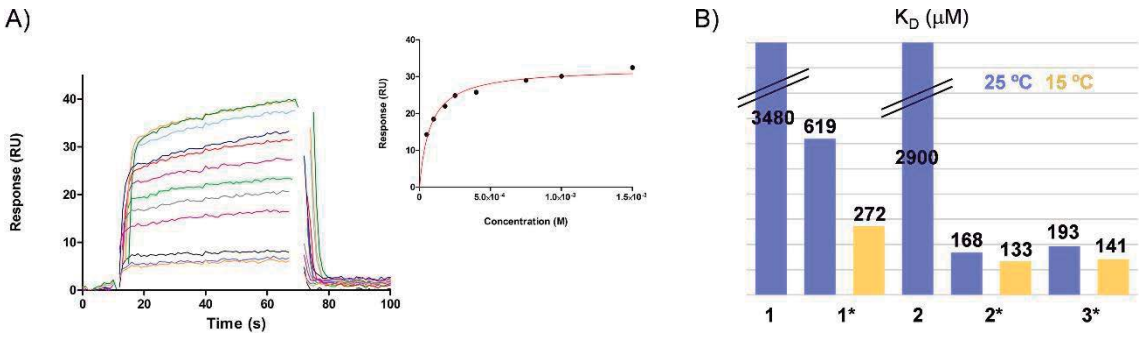



**Figura 17.** *(A) Curvas de SPR y ajuste respuesta-concentración obtenidos para determinar experimentalmente la afinidad de **2*** por scFv-SM3. (B) $K_D$ determinadas experimentalmente por SPR para los glicopéptidos **1*-3***, así como para los péptidos **1** y **2**.*

### 6.3.4. Preparación y ensayos inmunológicos *in vivo* de una vacuna contra el cáncer

Como se ha mencionado anteriormente, los derivados de MUC1 parcialmente glicosilados son buenos candidatos para el desarrollo de vacunas terapéuticas contra el cáncer.[34–37]

En este capítulo se ha demostrado que la sustitución de un solo átomo en el enlace glicosídico tiene un impacto notable en la estructura del glicopéptido en disolución, especialmente en torno al aminoácido glicosilado. Esto puede afectar de manera significativa a la presentación del péptido y del carbohidrato al sistema inmunológico y, por ello, también a la eficacia de la vacuna. Además, los glicopéptidos naturales pueden degradarse por la presencia de glicosidasas endógenas[38,39] pudiendo disminuir la eficacia como antígenos inmunizantes.[40] Por el contrario, los análogos *S*-glicosídicos poseen una mayor estabilidad frente a la degradación llevada a cabo por estas enzimas.[41,42] Debido a estas dos consideraciones, nos planteamos la posibilidad de utilizar los glicopéptidos no naturales como miméticos de otros antígenos naturales asociados a tumores.

Con ese motivo se realizó la síntesis del glicopéptido **16***, el cual posee la secuencia de repetición de la MUC1, en la cual se ha incorporado el residuo no natural *S*Thr* descrito anteriormente (Figura 17).

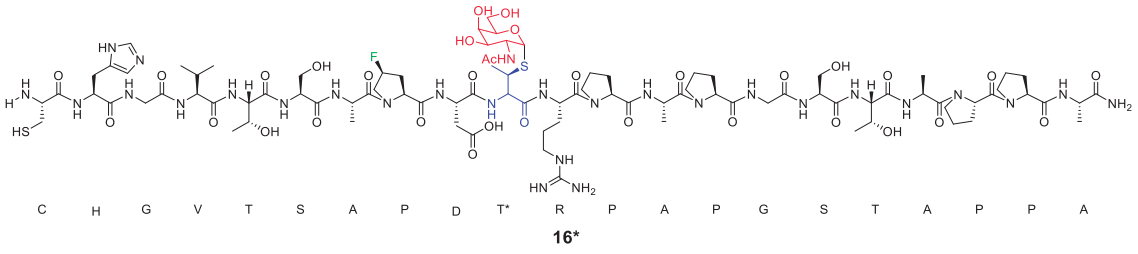



**Figura 17.** *Secuencia de la MUC1 correspondiente al glicopeptido **16**.*

Adicionalmente, este glicopéptido incorpora un residuo de (4*S*)-4-fluoro-L-prolina (fPro) el cual reemplaza a la primera prolina del epítopo de reconocimiento PDTRP. El motivo de este cambio fue el de combinar la mejora entrópica que aporta el residuo de *S*Thr* junto con la mejora de la contribución entálpica generada por el aminoácido fPro.[13]

Con el fin de mejorar las propiedades de la vacuna, se ha demostrado que las nanopartículas (NPs) de oro funcionalizadas con polietilenglicol (AuNPsPEG) pueden actuar como transportadores muy eficientes de antígenos de MUC1 en ratones y que los anticuerpos que estos generan reconocen los antígenos naturales presentados por las células tumorales de pacientes humanos con cáncer de mama.[43] Por todo ello, nos propusimos conjugar el glicopéptido **16*** a las AuPNsPEG (Figura 18). Mediante análisis electroforético se confirmó el éxito en la conjugación, ya que al realizarse la unión del glicopéptido a las nanopartículas de oro éstas aumentan su radio hidrodinámico y en consecuencia se reduce la movilidad en el gel electroforético (Figura 18).

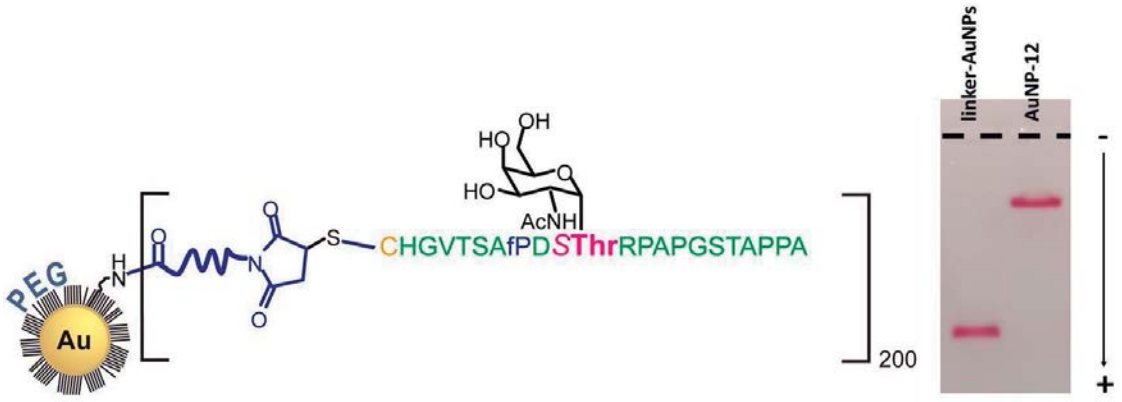



**Figura 18.** *Representación esquemática de la vacuna propuesta junto con la imagen de una electroforesis en gel de agarosa que muestra la funcionalización de las AuPNsPEG.*

.

Posteriormente, se inmunizó a un conjunto de ratones con este derivado con el propósito de ensayar *in vivo* el potencial inmunogénico de la posible vacuna terapéutica sintetizada. Para ello, se administró a un grupo de cinco ratones BALB/c una dosis principal (2 µg de 16-AuPNsPEG) seguida de tres dosis de refuerzo en intervalos de 21 días. Por otro lado, a modo de control, se trató un segundo grupo de ratones con un tampón de fosfato (PBS). Una semana después de la última dosis de refuerzo se sacrificaron los ratones con el fin de extraer el suero. El análisis de dicho suero confirmó que mientras que la vacuna provoca la aparición de anticuerpos IgG anti-MUC1, los ratones tratados con PBS (control negativo) no generaron anticuerpos de este tipo (Figura 19). Además, se determinó que los anticuerpos predominantes en los ratones inmunizados son del tipo IgG1, pudiéndose encontrar también anticuerpos IgG2a, IgG2b e IgG3[44] en una menor cantidad.

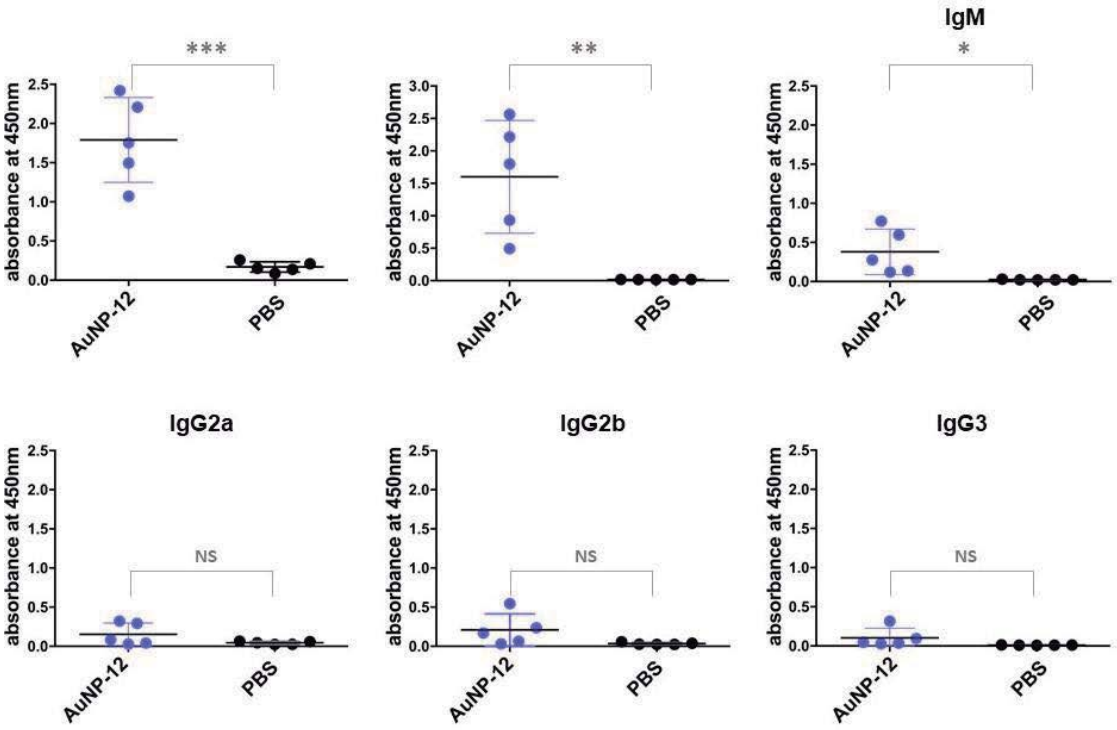



**Figura 19.** *Determinación de la cantidad y tipos de anticuerpos generados por la vacuna x tras la inmunización de ratones BALB/c.*

Con motivo de averiguar si los anticuerpos generados por la vacuna son capaces de reconocer los antígenos naturales presentados por células cancerosas humanas, se procedió a tratar dos líneas celulares tumorales (MCF-7 y T47D) con el suero extraído de los ratones. A modo de control, se estudió también una línea celular de riñón embrionario humano (HEK293T), que no expresa en su superficie MUC1 parcialmente glicosilada. Tras el posterior análisis mediante citometría de flujo, se puede observar que el suero reacciona fuertemente con las líneas celulares MCF-7 y T47D, mientras que no lo hace con la línea celular sana HEK293T (Figura 20).

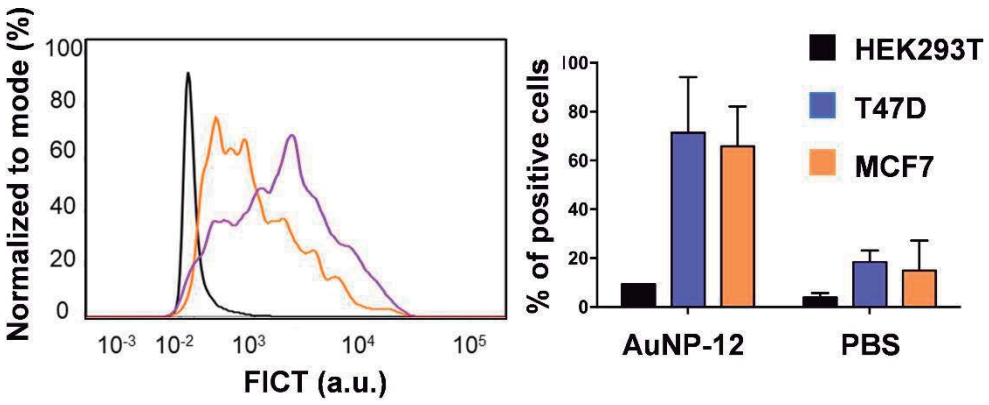



**Figura 20.** *Los estudios de citometría de flujo indican que los anticuerpos generados en los ratones inmunizados con el derivado AuPNsPEG reconocen a MUC1 expresada en células tumorales (T47D y MCF7). La línea celuar HEK293T se utilizó como control negativo.*

Estos resultados están de acuerdo con las imágenes obtenidas por microscopía confocal, ya que se puede observar claramente la presencia de los anticuerpos en la superficie de las líneas celulares MCF-7 y T47D pero no en las células HEK293T (Figura 21).

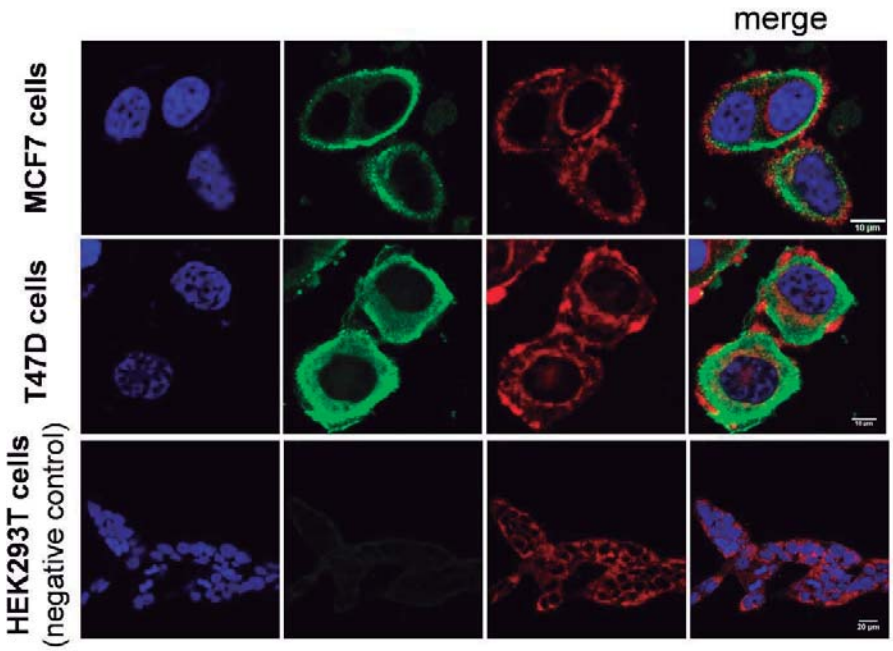



**Figura 21.** *Imágenes de microscopia. En azul se muestran los núcleos, en verde se puede ver el anticuerpo secundario IgG conjugado con Alexa 488 y en rojo las membranas celulares.*

Además de las líneas celulares anteriores, se estudiaron biopsias de pacientes con cáncer de mama. Como se puede ver en la imagen, las células cancerígenas son reconocidas también por los anticuerpos generados por los ratones (Figura 22).

Por lo tanto, estos resultados muestran el potencial como antígeno que posee el glicopéptido no natural **16***.

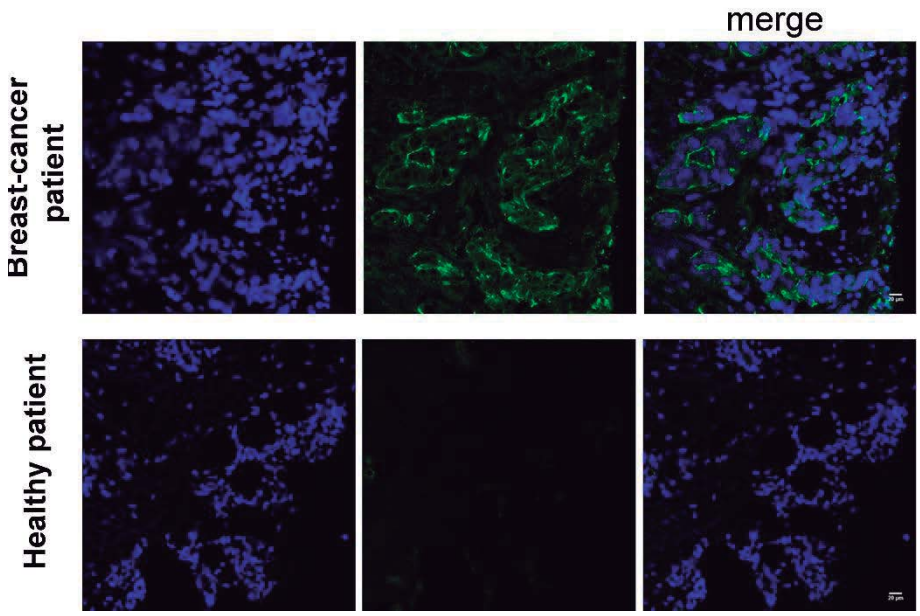



**Figura 22.** *Imágenes de microscopia. En azul se muestran los núcleos mientras que en verde se puede ver el anticuerpo secundario IgG Alexa 488.*

## 6.4. Conclusiones

Se ha logrado diseñar una ruta sintética que permite la obtención de los *building blocks* de *S*-(α-D-GalNAc)-*S*Thr y *Se*-(α-D-GalNAc)-*Se*Thr, los cuales, son aptos para la incorporación de los aminoácidos modificados en secuencias peptídicas mediante el empleo de la técnica de síntesis de péptidos en fase solida (SPPS).

Además, se ha demostrado la posibilidad de modular el modo de presentación de los glicopéptidos que contienen el carbohidrato GalNAc mediante la sustitución del oxígeno glicosídico por un átomo de azufre o selenio. Gracias a este cambio, se consigue una preorganización del péptido modificado, la cual favorece el reconocimiento molecular por el anticuerpo scFv-SM3, mejorando así su afinidad.

Por último, se ha demostrado la capacidad como antígeno del péptido no natural que incorpora el aminoácido *S*-(α-D-GalNAc)-*S*Thr mediante la

formulación de una vacuna terapéutica ensayada en ratones. Es importante señalar que los anticuerpos generados con esta vacuna reconocen de manera *selectiva* los antígenos naturales existentes en las células cancerosas y, por lo tanto, a dichas células.

## 6.5. Parte experimental

**Synthesis of Allyl (*2R,3S*)-2-((tert-butoxycarbonyl)amino)-3-iodobutanoate (7).**

Iodine (2.94 g, 11.6 mmol, 2.1 equiv.) was added in three portions to a solution of triphenylphosphine (2.46 g, 9.38 mmol, 1.7 equiv.) and imidazole (0.713 g, 10.5 mmol, 1.9 equiv.) in $CH_2Cl_2$ (15 mL) at 0 °C and the reaction mixture was stirred at 25 °C for 10 min. A solution of derivative **6** (1.43 g, 5.52 mmol) in $CH_2Cl_2$ (35 mL) was added over 30 min and stirred at 25 °C for 1.5 h. A solution of diethyl ether/hexane (3:2, 100 mL) was then added. The resulting brown slurry was filtered through a plug of silica gel which was subsequently washed with a solution of diethyl ether/hexane (1:1, 200 mL), and the filtrate was concentrated under vacuum to yield a clear pale-yellow oil. The crude product was purified by column chromatography (ethyl acetate/hexane, 1:4) to yield 1.63 g (4.42 mmol, 80% yield) of derivative **7** as a clear colorless oil.

$[\alpha]_D^{25}$ = +49.2 (c 1.00, $CHCl_3$).

HRMS (ESI) m/z: calcd for $C_{12}H_{21}IN_2O_4$ $[M+H]^+$: 370.0510, found 370.0515.

1H NMR (400 MHz, $CDCl_3$) δ (ppm): 5.95 (*ddt*, *J* = 16.4, 10.5, 5.9 Hz, 1H, CH=), 5.28–5.43 (m, 3H, $=CH_2$, NH), 4.65–4.74 (m, 2H, O-$CH_2$), 4.40–4.46

(m, 1H, H3), 4.32 (dd, *J* = 8.4, 3.6 Hz, 1H, H2), 1.97 (d, *J* = 7.2 Hz, 3H, CH3), 1.46 (s, 9H, *t*Bu).

13C NMR (100 MHz, $CDCl_3$) δ (ppm): 168.7 (CO2Allyl), 154.9 (CO2tBu), 131.3 (CH=), 119.4 (=CH2), 80.4 ($C(CH_3)_3$), 66.5 (O-CH2), 60.3 (C2), 28.3 (*tBu*), 26.9 (C3), 25.1 ($CH_3$).

**Synthesis of compound 9.**

A mixture of derivative **8** (0.500 g, 1.28 mmol), Woollins' reagent (1.00 g, 1.87 mmol, 1.46 equiv.), and pyridine (0.41 ml, 5.1 mmol, 3.9 equiv.) in 6 mL of toluene was heated at reflux for 14 h. The resulting solution was allowed to cool to rt, concentrated under reduced pressure and purified by column chromatography on silica gel with hexane/ethyl acetate (8:2→4:6) as the eluent to give derivative **9** as a pale-yellow oil (0.355 g, 0.90 mmol, 70 % yield). *Rf* = 0.25, 2:3 hexane/ethyl acetate.

$[\alpha]_D^{25}$ = +122.7 (c 1.00, $CHCl_3$).

HRMS (ESI) m/z: calcld for $C_{14}H_{20}NO_7Se$ $[M+H]^+$: 394.0405, found 394.0402.

$^{1H}$ NMR (400 MHz, $CDCl_3$) δ 6.77 (d, *J* = 6.1 Hz, 1H, $H_1$), 5.44 (dd, *J* = 3.3, 2.5 Hz, 1H, $H_4$), 5.21 (dd, *J* = 8.8, 3.3 Hz, 1H, $H_3$), 4.34 (ddd, *J* =8.8, 6.1 and *JH*2-Se = 1.1 Hz, 1H, $H_2$), 4.26 (td, *J* = 6.4, 2.5 Hz, 1H, $H_5$), 4.19–4.05 (m, 2H, $H_6$), 2.29 (d, *JH7-Se* = 1.2 Hz, 3H, $H_7$), 2.12 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.03 (s, 3H, OAc).

$^{13}$C NMR (100 MHz, $CDCl_3$) δ 170.5 ($COCH_3$), 170.2 ($COCH_3$), 169.9 ($COCH_3$), 169.0 ($C_7$), 92.0 ($C_1$), 74.6 ($C_2$), 70.8 (2C, $C_3$, $C_5$), 65.9 ($C_4$), 61.2 ($C_6$), 25.9 ($C_8$), 20.9 ($COCH_3$), 20.7 ($COCH_3$), 20.6 ($COCH_3$).

**Synthesis of compound 10.**

To a stirred solution of **9** (0.766 g, 1.95 mmol) in 4 mL 1:1 $CH_2Cl_2$:MeOH at 0 °C, 1 mL trifluoroacetic acid (TFA) was added dropwise followed by 1 mL of $H_2O$. The resulting reaction mixture was stirred at rt for 3.5 h and concentrated under reduced pressure. The crude reaction was purified by column chromatography on silica gel with ethyl acetate/methanol (98:2) as the eluent to give **10** as a yellow foam (0.506 g, 0.58 mmol, 31% yield). *Rf* = 0.5, 98:2 ethyl acetate/methanol.

$[\alpha]_D^{25}$ = +114.2 (c 1.01, $CHCl_3$).

HRMS (ESI) m/z: calcld for $C_{28}H_{40}N_2O_{16}Se_2Na$, $[M+Na]^+$: 843.0606, found 843.0609.

$^1$H NMR (400 MHz, $CDCl_3$): δ 6.33 (d, *J* = 7.7 Hz, NH), 6.03 (d, *J* = 4.9 Hz, 1H, $H_1$), 5.41 (dd, *J* = 3.1, 0.8 Hz, 1H, $H_4$), 5.09 (dd, *J* = 11.7, 3.1 Hz, 1H, $H_3$), 4.68–4.56 (m, 1H, $H_2$), 4.32 (‘t’, *J* = 6.3 Hz, 1H, $H_5$), 4.18 (dd, *J* = 11.3, 6.3 Hz, 1H, $H_6$), 4.10 (dd, *J* = 11.3, 6.5 Hz, 1H, $H_6$), 2.13 (s, 3H, OAc), 2.05 (s, 3H, OAc), 1.98 (s, 3H, OAc), 1.97 (s, 3H, NHAc).

$^{13}$C NMR (100 MHz, $CDCl_3$): δ 171.1 ($COCH_3$), 170.6 ($COCH_3$), 170.4 ($COCH_3$), 170.2 ($COCH_3$), 87.8 ($C_1$), 70.3 ($C_5$), 68.8 ($C_3$), 66.6 ($C_4$), 61.3 ($C_6$), 49.3 ($C_2$), 23.1 (NHAc), 20.8 (OAc), 20.7 (OAc), 20.6 (OAc).

**Synthesis of compound 11.**

$NaBH_4$ (38 mg, 1.0 mmol, 7.0 equiv.) was added portionwise to a solution of **10** (117 mg, 0.142 mmol) in a mixture of dry MeOH/THF (1:1, 6 mL) at 0 °C. The mixture was allowed to warm to rt and stirring was continued for 2 h. The solution turned colorless and gas evolution occurred. The solution of iodide **7** (52 mg, 0.142 mmol, 1 equiv.) in 2 ml of dry THF was added at 0°C. The mixture was allowed to warm to rt and stirring was continued for additional 3 h. The solvents were evaporated under reduced pressure. The

mixture was dissolved in ethyl acetate, washed with water, brine, and dried (MgSO4), filtered, and evaporated under reduced pressure. Purification by column chromatography (hexanes/Ethyl acetate 1:1→0:1) gave product **11** as a yellow foam (60 mg, 0.072 mmol, 51 % yield). *Rf* = 0.31 (hexane/ethyl acetate 4:6).

$[\alpha]_D^{25}$ = +123 (c 1.00, $CHCl_3$).

HRMS (ESI) m/z: calcd for $C_{26}H_{40}N_2O_{12}SeNa$, $[M+Na]^+$: 675.1644, found 675.1644.

$^1$H NMR (400 MHz, $CDCl_3$): δ 5.94 (d, *J* = 5.0 Hz, 1H, $H_{1'}$), 5.93–5.86 (m, 1H, CH=), 5.55 (d, *J* = 9.7 Hz, 1H, NH), 5.47 (d, *J* = 8.5 Hz, 1H, NH), 5.39 (dd, *J* = 3.1, 1.1 Hz, 1H, $H_{4'}$), 5.36–5.25 (m, 2H, =CH2), 4.90 (dd, *J* = 11.7, 3.2 Hz, 1H, $H_{3'}$), 4.68 (m, 1H, $H_{2'}$), 4.63 (m, 1H, $H_2$), 4.59 (m, 2H, O–$CH_2$), 4.49 (‘t’, *J* = 6.4, 1H, $H_{5'}$), 4.18 (‘dq’, *J* = 11.4, 6.4 Hz, 2H, $H_{6'}$), 3.77 (dd, *J* = 7.2, 3.0 Hz, 1H, $H_3$), 2.17 (s, 3H, OAc), 2.07 (s, 3H, OAc), 2.01 (s, 3H, OAc), 1.97 (s, 3H, NHAc), 1.55 (d, *J* = 7.2 Hz, 3H, H4), 1.45 (s, 9H, *t*Bu). $^{13}$C NMR (100 MHz, $CDCl_3$) δ 171.0 (CO2Allyl), 170.4 ($COCH_3$), 170.2 ($COCH_3$), 170.1 ($COCH_3$), 170.0 ($COCH_3$), 155.9 (CO2*t*Bu), 131.3 (CH=), 119.3 (=$CH_2$), 81.3 ($C_{1'}$), 80.2 ($C(CH_3)_3$), 69.6 ($C_{3'}$), 69.0 ($C_{5'}$), 67.0 ($C_{4'}$), 66.4 (O–$CH_2$), 61.8 ($C_{6'}$), 58.9 ($C_2$), 48.8 ($C_{2'}$), 38.7 ($C_3$), 28.3 (*t*Bu), 23.3 (NHAc), 20.8 (OAc), 20.7 (OAc), 20.6 (OAc), 20.3 ($C_4$).

**Synthesis of compound 12.**

*N*-Boc protected compound **11** (146 mg, 0.223 mmol) was dissolved in 3 mL $CH_2Cl_2$ and cooled to 0 °C. 3 mL TFA were added and the solution was allowed to warm to rt. After 2.5 h no more starting material was present (TLC). The solution was concentrated in vacuum and the residue was dissolved in 6 mL CH2Cl2 and cooled to 0 °C. DIEA (78 μL, 0.446 mmol,

2 equiv) and FmocCl (63 mg, 0.245 mmol, 1.1 equiv.) were added and the mixture was allowed to warm to rt. After 1.5 h at rt the starting material was consumed (TLC). The crude was concentrated under vacuum and the residue taken up in Ethyl acetate (20 mL) and washed consecutively with $H_2O$ (20 mL) and brine (20 mL). The organic phase was dried over $MgSO_4$, filtered, and concentrated. Purification by column chromatography (hexane/ethyl acetate 4:6) afforded product **12** as a colorless oil (124 mg, 0.016 mmol, 72% yield). *Rf* = 0.27 (hexane/Ethyl acetate 4:6).

$[\alpha]_D^{25}$ = +99 (c 1.01, $CHCl_3$).

HRMS (ESI) m/z: calcd for $C_{36}H_{42}N_2O_{12}SeNa$, $[M+Na]^+$: 797.1801, found, 797.1798.

$^1$H NMR (400 MHz, $CDCl_3$): δ 7.77 (d, *J* = 7.5 Hz, 2H, Ar), 7.60 (t, *J* = 8.3 Hz, 2H, Ar), 7.43–7.27 (m, 4H, Ar), 5.90 (d, *J* = 4.9 Hz, 1H, $H_{1'}$), 5.89–5.84 (m, 1H, CH=), 5.54 (d, *J* = 8.6 Hz, 1H, NH), 5.39–5.37 (m, 1H, $H_{4'}$), 5.34–5.27 (m, 2H, $=CH_2$), 4.88 (dd, *J* = 11.7, 3.2 Hz, 1H, $H_{3'}$), 4.79– 4.57 (m, 4H, H2, H2', O–$CH_2$), 4.51 (dd, *J* = 10.6, 6.6 Hz, 1H, $CH_2$Fmoc), 4.46 ('t', *J* = 6.4 Hz, 1H, $H_{5'}$), 4.36 (dd, *J* = 10.6, 6.6 Hz, 1H, $CH_2$Fmoc), 4.22 (t, *J* = 6.6 Hz, 1H, CHFmoc), 4.21–4.16 (m, 1H, $H_{6'}$), 4.03 (dd, *J* = 11.4, 7.5 Hz, 1H, $H_{6'}$), 3.77 ('qd', *J* = 7.2, 3.0 Hz, 1H, $H_3$), 2.19 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.01 (s, 3H, OAc), 1.97 (s, 3H, NHAc), 1.47 (d, *J* = 7.2 Hz, 3H, $H_4$). 13C NMR (100 MHz, $CDCl_3$) δ 171.0 ($CO_2$Allyl), 170.4 ($COCH_3$), 170.1 ($COCH_3$), 169.9 ($COCH_3$), 169.8 ($COCH_3$), 156.5 ($NHCO_2$), 143.8 (Cq), 143.5 (Cq), 141.4 (Cq), 141.3 (Ar), 131.3 (CH=), 127.7 (Ar), 127.6 (Ar), 127.1 (Ar), 127.0 (Ar), 125.1 (Ar), 125.0 (Ar), 120.0 ($=CH_2$), 119.4 (Ar), 80.8 ($C_{1'}$), 69.9 ($C_{5'}$), 69.0 ($C_{3'}$), 67.0 ($C_{4'}$), 66.9 ($CH_2$Fmoc), 66.5 (O–$CH_2$), 61.7 ($C_{6'}$), 59.4 ($C_2$), 48.8 ($C_{2'}$), 47.2 (CHFmoc), 38.2 ($C_3$), 23.3 (NHAc), 20.7 (2C, OAc), 20.6 (OAc), 19.9 ($C_4$).

**Synthesis of compound 4.**

To a solution of **12** (124 mg, 0.160 mmol) in 1 mL dry THF under argon at rt, tetrakis(triphenylphosphine)palladium(0) (18.3 mg, 0.016 mmol, 0.1 equiv) and morpholine (55 μL, 0.32 mmol, 2 equiv.) were added sequentially. After 30 min. the solvent was evaporated, and the residue was taken up in 10 mL of $CH_2Cl_2$. The resulting solution was extracted three times with 2 M HCl, dried over $MgSO_4$, and concentrated in vacuum. Purification by column chromatography (ethyl acetate/MeOH/AcOH, 95:5:0.1) provided product **4** as a yellow oil (67 mg, 0.093 mmol, 58% yield). *Rf* = 0.57 (ethyl acetate/MeOH 95:5).

$[\alpha]_D^{25}$ = +67 (c 1.00, $CHCl_3$).

HRMS (ESI) m/z: calcld for $C_{33}H_{38}N_2O_{12}SeNa$, $[M+Na]^+$: 757.1488, found, 797.1487.

$^1$H NMR (400 MHz, $CD_3OD$): δ 7.79 (d, *J* = 7.5 Hz, 2H, Ar), 7.70–7.61 (m, 2H, Ar), 7.38 (t, *J* = 7.5 Hz, 2H, Ar), 7.30 (m, 2H, Ar), 6.33 (br d, *J* = 5.4 Hz, 1H, $H_{1'}$), 5.43 (br d, *J* = 2.6, 1H, $H^{4'}$), 4.84–4.92 (m, $H_{3'}$), 4.64–4.57 (m, 1H, $H_{5'}$), 4.44 (dd, *J* = 10.5, 7.0 Hz, 1H, $CH_2$Fmoc), 4.41–4.37 (m, 1H, $H_{2'}$),4.37–4.33 (m, 1H, $H_2$), 4.31 (dd, *J* = 10.5, 7.0 Hz, 1H, $CH_2$Fmoc), 4.21 (t, *J* = 7.0 Hz, 1H, CHFmoc), 4.17–4.12 (m, 1H, $H_{6'}$), 4.10 (dd, *J* = 11.2, 7.2 Hz, 1H, $H_{6'}$), 3.70–3.77 (m, 1H, $H_3$), 2.11 (s, 3H, OAc), 1.99 (s, 3H, OAc), 1.97 (s, 3H, OAc), 1.94 (s, 3H, NHAc), 1.49 (d, *J* = 7.3 Hz, 1H, $H_4$).

$^{13}$C NMR (100 MHz, $CD_3OD$) δ 176.7 ($CO_2H$), 173.2 (OAc), 170.8 (OAc), 170.5 (OAc), 170.2 (NHAc), 157.2 (NHCO2), 144.0 (Cq), 143.6 (Cq), 141.2 (2C, Ar), 127.4 (Ar), 127.3 (Ar), 126.7 (Ar), 124.7 (Ar), 119.5 (Ar), 81.9 ($C_{1'}$), 68.8 ($C_{3'}$), 68.6 ($C_{5'}$), 66.9 ($CH_2$Fmoc), 66.4 ($C_{4'}$), 61.7 ($C_{6'}$), 60.7 ($C_2$),

49.1 ($C_{2'}$), 49.0 (CHFmoc), 39.3 ($C_3$), 21.2 (OAc), 20.5 (OAc), 19.2 (OAc), 19.1 (NHAc).

$^{77}$Se NMR (76 MHz, CD3OD) δ 375.2.

**Synthesis of compound 14**.

To a stirred solution of compound **7** (500 mg, 1.35 mmol) and DIEA (0.71 mL, 4.1 mmol, 3.0 equiv.) in acetonitrile (15 mL), a solution of compound **13** (1.48 g, 4.1 mmol, 3.0 equiv.) in 10 mL of acetonitrile was added dropwise over 1 h. The resulting solution was stirred further at rt for 3 h. A saturated solution of $NH_4Cl$ (30 mL) were then added to the reaction mixture and acetonitrile was evaporated. The resulting suspension was extracted with $CH_2Cl_2$ (2×30 mL) and the combined organic phases were dried over $Na_2SO_4$, filtered, and concentrated. The residue was purified by column chromatography using ethyl acetate/hexane (7:3) as an eluent, to give compound **14** as a white solid (286 mg, 0.47 mmol, 35% yield).

$[\alpha]_D^{25}$ = 106.4 (c 1.02, $CHCl_3$).

HRMS (ESI) m/z: calcld for $C_{26}H_{41}N_2O_{12}S$, $[M+H]^+$: 605.2375, found, 605.2377.

$^1$H NMR (400 MHz, $D_2O$) δ (ppm): 5.92 (ddt, *J* = 16.4, 10.5, 5.9 Hz, 1H, CH=), 5.60–5.54 (m, 3H, $H_{1'}$, NH, NH), 5.38–5.28 (m, 3H, $H_{4'}$, $=CH_2$), 4.96 (dd, *J* = 11.0, 3.2 Hz, 1H, $H_{3'}$), 4.80–4.74 (m, 1H, $H_{2'}$), 4.70–4.59 (m, 3H, O–$CH_2$, $H_2$), 4.53 ('t', *J* = 6.4, 1H, $H_{5'}$), 4.22–4.08 (m, 2H, $H_{6'}$), 3.65–3.63 (m, 1H, $H_3$), 2.16 (s, 3H, OAc), 2.07 (s, 3H, OAc), 2.0 (s, 3H, OAc), 1.97 (s, 3H, NHAc), 1.46 (s, 9H, *t*Bu), 1.43 (d, *J* = 7.2 Hz, 3H, H4).

$^{13}$C NMR (100 MHz, $D_2O$) δ (ppm): 170.9 ($COCH_3$), 170.4 ($COCH_3$), 170.2 ($COCH_3$), 170.1 ($COCH_3$), 156.0 (CO2*t*Bu), 131.4 (CH=), 119.4 ($=CH_2$),

83.2 ($C_{1'}$), 80.2 (-$C(CH_3)_3$), 68.2 ($C_{3'}$), 68.0 ($C_{5'}$), 67.3 ($C_{4'}$), 66.4 ($C_{6'}$), 61.8 (O–$CH_2$), 58.6 ($C_2$), 48.4 ($C_{2'}$), 43.6 ($C_3$), 28.3 (*t*Bu), 23.3 (NHAc), 20.7 (OAc), 20.7 (OAc), 20.7 (OAc), 19.5 ($C_4$).

**Synthesis of compound 15.**

6 mL TFA were added to a stirred solution of compound **14** (220 mg, 0.364 mmol) in 3 mL $CH_2Cl_2$ at rt. After 2 h, the volatiles were evaporated under vacuum and the residue was dissolved in 3 mL $CH_2Cl_2$. A solution of 9-fluorenylmethyl chloroformate (104 mg, 0.402 mmol, 1.1 equiv.) and DIEA (139 μL, 0.798 mmol, 2.2 equiv.) in 1 mL $CH_2Cl_2$ were added sequentially. The reaction mixture was stirred at rt for 3 h before addition of 5 mL 0.05 M aqueous HCl. The aqueous layer was extracted with $CH_2Cl_2$ (2×5 mL) and the combined organic phases were dried with $Na_2SO_4$, filtered and evaporated. The reaction crude was purified by column chromatography using ethyl acetate/hexane (7:3) as an eluent, to obtain compound **15** as a white solid (251 mg, 0.346 mmol, 95% yield).

$[\alpha]_D^{25}$ = +57 (c 1.00, $CHCl_3$).

HRMS (ESI) m/z: calcd for $C_{36}H_{43}N_2O_{12}S$, $[M+H]^+$: 727.2531, found, 727.2531.

$^1$H NMR (400 MHz, D2O) δ (ppm): 7.80–7.76 (m, 2H, Ar), 7.63–7.60 (m, 2H, Ar), 7.43–7.39 (m, 2H, Ar), 7.37–7.31 (m, 2H, Ar), 5.97–5.86 (m, 2H, CH=, NH), 5.57 (d, *J* = 8.8 Hz, 1H, NH), 5.52 (d, *J* = 4.8 Hz, 1H, $H_{1'}$), 5.38–5.28 (m, 3H, $H_{4'}$, (=$CH_2$), 4.95 (dd, *J* = 11.8, 2.8 Hz, 1H, $H_{3'}$), 4.84–4.78 (m, 1H, $H_{2'}$), 4.69–4.60 (m, 3H, O–$CH_2$, $H_2$), 4.55–4.36 (m, 3H, $H_{5'}$, $CH_2$Fmoc), 4.25–4.01 (m, 3H, $H_{6'}$, CHFmoc), 3.67–3.65 (m, 1H, $H_3$), 2.02 (s, 3H, OAc), 2.02 (s, 3H, OAc), 1.99 (s, 3H, OAc), 1.37 (d, *J* = 7.0 Hz, 3H, H4).

$^{13}$C NMR (100 MHz, $D_2O$) δ (ppm): 171.0 (CO2Allyl), 170.5 ($COCH_3$), 170.2 ($COCH_3$), 170.0 ($COCH_3$), 169.7 ($COCH_3$), 156.0 ($NHCO_2$), 143.9 (Cq), 143.7 (Cq), 141.5 (Cq), 141.4 (Cq), 131.3 (CH=), 127.8 (Ar), 127.7 (Ar), 127.1 (Ar), 125.1 (Ar), 125.0 (Ar), 120.0 (Ar), 119.6 (=$CH_2$), 82.5 ($C_{1'}$), 68.2 ($C_{3'}$, $C_{5'}$), 67.3 ($C_{4'}$), 67.0 ($CH_2$Fmoc), 66.4 ($C_{6'}$), 61.8 (O–$CH_2$), 59.0 ($C_2$), 48.3 ($C_{2'}$), 47.2 (CHFmoc), 43.2 ($C_3$), 30.9 (NHAc), 30.9 (OAc), 20.7 (OAc), 20.6 (OAc), 19.1 ($C_4$).

**Synthesis of compound 5.**

Tetrakis(triphenylphosphine)palladium(0) (30 mg, 0.026 mmol, 0.1 equiv.) and morpholine (90 μL, 0.52 mmol, 2.0 equiv.) were added to a stirred solution of compound **15** (190 mg, 0.262 mmol) in 10 mL dry THF under argon. The resulting solution was stirred at rt for 10 min. The reaction was stopped by addition of 10 mL 0.05 M aqueous HCl and the aqueous layer was extracted with chloroform/2-propanol (3:1, 3×5 mL). The organic phases were combined, dried over $Na_2SO_4$, filtered, and evaporated. The crude product was purified by column chromatography using chloroform/methanol (9:1) as an eluent, to obtain compound **5** as a white solid (162 mg, 0.236 mmol, 90% yield).

HRMS (ESI) m/z: calcd for $C_{33}H_{39}N_2O_{12}S$, $[M+H]^+$: 687.2218, found, 687.2225.

**Solid-phase peptide synthesis (SPPS)**

Glycopeptides were synthesized by stepwise microwave assisted solid-phase synthesis on a Liberty Blue synthesizer using the Fmoc strategy on Rink Amide MBHA resin (0.1 mmol). Fmoc-Thr[GalNAc(Ac)3-α-D]-OH (2.0 equiv) was synthesized as described in the literature or as described in this work. Glycosilated aminoacids were coupled using HBTU [(2-(1*H*-

benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate], while all other Fmoc amino acids (5.0 equiv) were automatically coupled using oxyma pure/DIC (*N*,*N*'-diisopropylcarbodiimide). The *O*-acetyl groups of GalNAc moiety were removed in a mixture of $NH_2NH_2$/MeOH (7:3). (Glyco)peptides were then released from the resin, and all acid sensitive side-chain protecting groups were simultaneously removed using TFA 95%, TIS (triisopropylsilane) 2.5% and H2O 2.5%, followed by precipitation with cold diethyl ether. The crude products were purified by HPLC on a Phenomenex Luna C18(2) column (10 μm, 250 mm × 21.2 mm) and a dual absorbance detector, with a flow rate of 10 mL/min.

**Peptide 1.**

The spectroscopic data are consistent with those described in the literature. HRMS (ESI): m/z 655.3547 $[M+H]^+$; calcd $C_{27}H_{47}N_{10}O_9$: 655.3527.

**Glycopeptide 1*.**

The spectroscopic data are consistent with those described in the literature.

HRMS (ESI): m/z 858.4283 $[M+H]^+$; calcd $C_{35}H_{60}N_{11}O_{14}$: 858.4321.

**Peptide 2**.

HRMS (ESI): m/z 671.3294 $[M+H]^+$; calcd $C_{27}H_{47}N_{10}O_8S^+$: 671.3294.

Rt = 12.5 min (Grad: water 0.1% TFA/acetonitrile (95:5) → (85:15), 13.01 min, λ = 212 nm).

**Glycopeptide 2***.

HRMS (ESI): m/z 874.4068 $[M+H]^+$; calculated $C_{35}H_{60}N_{11}O_{13}S+$: 874.4087.

Rt = 11.53 min (Grad: water 0.1% TFA/acetonitrile (95:5) → (85:15), 13.01 min, λ = 212 nm).

1H NMR (400 MHz, $D_2O$) δ (ppm):

5.53 (d, *J* = 7.2 Hz, 1H, $H_1$), 4.66–4.61 (m, 2H, $H_{\alpha SThr}$, $H_{\alpha Arg}$), 4.45 (dd, *J* = 8.2, 6.3 Hz, 1H, $H_{\alpha Pro2}$), 4.40–4.31 (m, 3H, $H_{\alpha Pro6}$, $H_{\alpha Ala}$, $H_2$), 4.22–4.19 (m, 1H, $H_5$), 3.95 ('d', *J* = 2.8 Hz, 1H, $H_4$), 3.86–3.52 (m, 8H, $H_3$, $2H_6$, $H_{\delta Pro2}$, $H_{\delta Pro6}$, $H_{\beta SThr}$), 3.25–3.14 (m, 2H, $H_{\delta Arg}$), 3.00–2.75 (m, 2H, $H_{\beta Asp}$), 2.35–1.50 (m, 15H, Ac, $H_{\beta Pro2}$, $H_{\beta Pro2}$, $H_{\gamma Pro6}$, $H_{\gamma Pro2}$, $H_{\beta Arg}$, $H_{\gamma Arg}$), 1.51 (d, *J* = 7.0 Hz, 3H, $Ala_\beta$), 1.29 (d, *J* = 7.2 Hz, 3H, $Me_{SThr}$).

$^{13}C$ NMR (100 MHz, $D_2O$) δ (ppm): 176.8, 174.3, 173.8, 173.4, 172.4, 171.0, 169.2, (CO), 156.7 ($C_{\varepsilon Arg}$), 81.5 ($C_1$), 72.2 ($C_5$), 68.4 ($C_4$), 67.6 ($C_3$), 61.2 ($C_6$), 60.3, 60.3 ($C_{\alpha Pro6}$, $C_{\alpha Pro2}$), 57.8($C_{\alpha SThr}$), 51.1 ($C_{\alpha Arg}$), 49.9, 49.9 ($C_{\alpha Ala}$, $C_{\alpha Asp}$), 48.0, 48.0 ($C_{\delta Pro2}$, $C_2$), 47.7 ($C_{\delta Pro6}$), 40.9 ($C_{\beta SThr}$), 40.6 ($C_{\delta Arg}$), 35.0 ($C_{\beta Asp}$), 29.6 ($C_{\beta Pro2}$), 29.4 ($C_{\beta Pro6}$), 27.7 ($C_{\beta Arg}$), 24.7 ($C_{\gamma Pro2}$), 24.6 ($C_{\gamma Pro6}$), 24.2 ($C_{\gamma Arg}$), 22.0 (Ac), 18.3 ($CH_{3SThr}$), 15.1 ($CH_{3Ala}$).

**Glycopeptide 3***.

HRMS (ESI): m/z 922.3489 $[M+H]^+$; calcd $C_{35}H_{60}N_{11}O_{13}Se^+$: 922.3535.

Rt = 11.53 min (Grad: water 0.1% TFA/acetonitrile (95:5) → (89:11), 11.15 min, λ = 212 nm).

$^1H$ NMR (400 MHz, $D_2O$) δ (ppm): 5.96 (d, *J* = 5.4 Hz, 1H, $H_1$), 4.48-4.51 (m, 1H, $H_{\alpha Pro2}$), 4.37-4.43 (m, 3H, $H_{\alpha Pro2}$, $H_{\alpha Ala1}$, $H_2$), 4.16-4.19 (m, 1H, $H_5$), 4.01 ('d', *J* = 3.2 Hz, 1H, $H_4$), 3.61-3.89 (m, 8H, $H_3$, $2H_6$, $H_{\delta Pro2}$, $H_{\delta Pro6}$, $H_{\beta SThr}$), 3.19-3.27 (m, 2H, $H_{\delta Arg}$), 2.74-3.93 (m, 2H, $H_{\beta Asp}$), 1.71-2.37 (m, 15H, Ac, $H\beta_{Pro2}$, $H_{\beta Pro2}$, $H_{\gamma Pro6}$, $H_{\gamma Pro2}$, $H_{\beta Arg}$, $H_{\gamma Arg}$), 1.55 (d, *J* = 7.0 Hz, 3H,

$CH_{3Ala\beta}$), 1.45 (d, *J* = 7.3 Hz, 3H, $CH_{3Se Thr}$). *Note*: some signals are overlapped by the $H_2O$ signal.

**Glycopeptide 16***.

HRMS (ESI): m/z 742.6725 $[M+3H]^{3+}$; calcd $C_{91}H_{148}FN_{28}O_{32}S_2^{3+}$: 742.6741.

Rt = 11.53 min (Grad: water 0.1% TFA/acetonitrile (95:5) → (89:11), 11.15 min, λ = 212 nm).

## 6.6. Bibliografía

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

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

(3) Senapati, S.; Das, S.; Batra, S. K. *Trends Biochem. Sci.* **2010**, *35*, 236–245.

(4) Taylor-Papadimitriou, J.; Burchell, J. M.; Taylor-Papadimitriou, J.; Burchell, J. M. *Mucins and Cancer*, Future Medicine, **2013**.

(5) Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem., Int. Ed.* **2011**, *50*, 1770–1791.

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

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

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

(9) Buskas, T.; Thompson, P.; Boons, G.-J. *Chem. Commun.* **2009**, *0*, 5335–5349.

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

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

(12) Hossain, M. K.; Wall, K. A. *Vaccines* **2016**, *4*, 25.

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

(14) Coltart, D. M.; Royyuru, A. K.; Williams, L. J.; Glunz, P. W.; Sames, D.; Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Danishefsky, S. J.; Live, D. H. *J. Am. Chem. Soc.* **2002**, *124*, 9833–9844.

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

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

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

(18) Burchell, J.; Gendler, S.; Taylor-Papadimitriou, J.; Girling, A.; Lewis, A.; Millis, R.; Lamport, D. *Cancer Res.* **1987**, *47*, 5476–5482.

(19) 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.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem. Int. Ed.* **2015**, *54*, 9830–9834.

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

(21) Jiménez-Moreno, E.; Gómez, A. M.; Bastida, A.; Corzana, F.; Jiménez-Oses, G.; Jiménez-Barbero, J.; Asensio, J. L. *Angew. Chem., Int. Ed.* **2015**, *54*, 4344–4348.

(22) García-Herrero, A.; Montero, E.; Muñoz, J. L.; Espinosa, J. F.; Vián, A.; García, J. L.; Asensio, J. L.; Cañada, F. J.; Jiménez-Barbero, J. *J. Am. Chem. Soc.* **2002**, *124*, 4804–4810.

(23) Gracia-Vitoria, J.; Osante, I.; Cativiela, C. *Tetrahedron: Asymmetry* **2017**, *28*, 215–245.

(24) Chen, S.; Fahmi, N. E.; Nangreave, R. C.; Mehellou, Y.; Hecht, S. M. *Bioorg. Med. Chem.* **2012**, *20*, 2679–2689.

(25) Narayan, R. S.; VanNieuwenhze, M. S. *Org. Lett.* **2005**, *7*, 2655–2658.

(26) Jobron, L.; Hummel, G. *Org. Lett.* **2000**, *2*, 2265–2267.

(27) Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519–538.

(28) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457–479.

(29) Pearlman, D. *J. Biomol. NMR* **1994**, *4*, 1–16.

(30) Fernández-Tejada, A.; Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. *Chem. - Eur. J.* **2009**, *15*, 7297–7301.

(31) Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; Coelho, H.; Valero-González, J.; Castro-López, J.; Asensio, J. L.; Jiménez-

Barbero, J.; Busto, J. H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Corzana, F.; Peregrina, J. M. *ACS Chem. Biol.* **2015**, *10*, 747–756.

(32) Mazeau, K.; Tvaroska, I. *Carbohydr. Res.* **1992**, *225*, 27–41.

(33) Aydillo, C.; Compañón, I.; Avenoza, A.; Busto, J. H.; Corzana, F.; Peregrina, J. M.; Zurbano, M. M. *J. Am. Chem. Soc.* **2014**, *136*, 789–800.

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

(35) Lakshminarayanan, V.; Supekar, N. T.; Wei, J.; McCurry, D. B.; Dueck, A. C.; Kosiorek, H. E.; Trivedi, P. P.; Bradley, J. M.; Madsen, C. S.; Pathangey, L. B.; Hoelzinger, B. D.; Wolfert, M. A.; Boons, G. J.; Cohen, P. A.; Gendler, S. J. *PLoS One* **2016**, *11*, e0145920.

(36) Wu, X.; Yin, Z.; McKay, C.; Pett, C.; Yu, J.; Schorlemer, M.; Gohl, T.; Sungsuwan, S.; Ramadan, S.; Baniel, C.; Allmon, A.; Das, R.; Westerlind, M. G. F.; Huang, X. *J. Am. Chem. Soc.* **2018**, *140*, 16596–16609.

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

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

(39) Nativi, C.; Renaudet, O. *ACS Med. Chem. Lett.* **2014**, *5*, 1176–1178.

(40) Rich, J. J.; Bundle, D. R. *Org. Lett.* **2004**, *6*, 897–900.

(41) Huo, C.-X.; Zheng, X.-J.; Xiao, A.; Liu, C.-C.; Sun, S.; Lv, Z.; Ye, X.-

S. *Org. Biomol. Chem.* **2015**, *13*, 3677–3690.

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

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

(44) Haji-Ghassemi, O.; Blackler, R. J.; Martin Young, N.; Evans, S. V. *Glycobiology* **2015**, *25*, 920–952.

# Colaboraciones basadas en síntesis de (glico)péptidos mediante SPPS

7

## 7.1. Introducción

En este capítulo se pretende realizar un compendio de las colaboraciones que se han realizado de forma paralela a la principal línea de investigación de esta tesis, la cual ha sido comentada en los capítulos anteriores.

Nuestro grupo de investigación posee una gran experiencia en la síntesis y estudio conformacional de moléculas orgánicas, siendo la síntesis de péptidos y glicopéptidos uno de los caballos de batalla mas importantes. En concreto, el auge de la química biológica suscita la necesidad de tener acceso a un elevado número de estos últimos. Es por ello que muchos grupos de investigación, que trabajan en este ámbito, recurren a grupos como el nuestro para la síntesis y posterior estudio conformacional de (glico)péptidos.

En este sentido, durante el desarrollo de esta tesis, se han llevado a cabo varias colaboraciones en el ámbito de la síntesis, purificación, caracterización y estudio conformacional de péptidos y glicopéptidos.

Uno de estos grupos es el liderado por el profesor Ramon Hurtado-Guerrero de la Universidad de Zaragoza, el cual se dedica principalmente al estudio mecanístico de las reacciones de glicosilación llevadas a cabo por la amplia familia de las GalNAc-transferasas (GalNAc-Ts).

Otro grupo de investigación con el que se ha colaborado en varios trabajos es el liderado por el profesor Gonçalo Bernardes de la Universidad de Cambridge. Este grupo de investigación dedica gran parte de su trabajo a la funcionalización de anticuerpos y proteínas con moléculas citotóxicas con el fin de aumentar la eficacia de éstas al ser liberadas en las proximidades de sus dianas biológicas.

Por último, debido a la amplia experiencia adquirida en la síntesis y posterior purificación de (glico)péptidos también se han efectuado numerosas

colaboraciones en varias de las líneas de nuestro propio grupo de investigación.

Las publicaciones originales se incluyen en el Anexo II

## 7.2. Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide N-Acetylgalactosaminyltransferase 2.

Resumen

La GalNAc-T2 es una glicosiltransferasa perteneciente a la amplia familia de las GalNAc-transferasas. Estas enzimas son responsables de la modificación postraduccional de muchas proteínas de la superficie celular. La GalNAc-T2 está involucrada en trastornos del metabolismo de los lípidos por lo que la convierte en un potencial objetivo terapéutico.

En este trabajo se pretende elucidar, a nivel atómico, el mecanismo catalítico que presenta la enzima GalNAc-T2. Para ello se combinan datos cristalográficos de intermedios del ciclo catalítico con cálculos teóricos QM/MM.

Este estudio proporciona información estructural detallada que permite establecer una cinetica *bi-bi* ordenada para el proceso catalítico. Además, revela la existencia de aspectos críticos en el reconocimiento del sustrato, los cuales establecen la especificidad de Thr frente a Ser, forzando una reacción $S_Ni$*-type* en la que el sustrato coordina la transferencia del GalNAc. Los datos obtenidos muestran que la reacción involucra un catión oxocarbenio de corta duración lo cual es compatible con el mecanismo propuesto, siendo muy probable que dicho mecanismo sea común para la mayoría de GalNAc-transferasas.

Finalmente, el establecimiento del mecanismo catalítico de GalNAc-T2 servirá para el diseño de inhibidores de la enzima, los cuales tendrán un elevado potencial terapéutico.

## 7.3. Design of α-*S*-Neoglycopeptides Derived from MUC1 with a Flexible and Solvent-Exposed Sugar Moiety.

Resumen

El uso de glicopéptidos tipo MUC1 como vacunas terapéuticas contra el cancer suscita un elevado interés en el mundo de la glicobiologìa. En este trabajo se presenta la síntesis de diferentes miméticos azufrados del antigeno Tn convenientemente protegidos para su incorporación en secuencias peptídicas mediante la utilización de la técnica SPPS. Dichos antigenos poseen diferentes espaciadores entre la cadena peptídica y el carbohidrato.

Mediante la combinación de técnicas de RMN y simulaciones de dinámica molecular se evaluaron las conformaciones principales en agua de los glicopéptidos sintetizados. Dichos estudios revelan que el espaciador juega un papel clave en la disposición espacial que adoptan estos compuestos. La regulación se lleva a cabo mediante la supresión de los contactos directos entre el esqueleto peptídico y el carbohidrato, promoviendo una conformación de tipo hélice para el residuo glicosilado y favoreciendo la presentación del azúcar en el reconocimiento molecular.

La posibilidad de la utilización de estos compuestos como miméticos del antígeno Tn ha sido validada por la obtención de una estructura de rayos X del complejo formado por uno de estos derivados y un anticuerpo anti-MUC1.

Estas características, junto con la posible falta de supresión inmunológica, hacen que los glicopéptidos no naturales sean candidatos prometedores para el diseño de vacunas terapéuticas contra el cancer.

## 7.4 Oxetane Grafts Installed Site-Selectively on Native Disulfides to Enhance Protein Stability and Activity In Vivo.

Resumen

Este trabajo muestra como el 3,3-bis(bromoetil)oxetano se puede insertar fácilmente en péptidos y proteínas a través de la bisalquilación selectiva de cisteínas que estén involucradas en puentes disulfuro. Además, dicha inserción puede realizarse mediante el empleo de condiciones de reacción biocompatibles.

Los estudios realizados en péptidos cíclicos biológicamente relevantes como la somatostatina y en anticuerpos como el Herceptin y el DesAb-Ab muestran que la introducción selectiva del oxetano aumenta la estabilidad y la actividad de dichas moléculas. La introducción del oxetano en proteínas como CRM197 mejora significativamente la inmunogenicidad de ésta en ratones. Esto último muestra la gran utilidad de esta estrategia a la hora de modular la actividad biológica y aumentar la estabilidad de proteínas terapéuticas que contienen puentes disulfuro en su estructura.

Debido a que muchas proteínas terapéuticas presentan residuos de Cys formando puentes disulfuro, es posible que la inserción del oxetano en ellas pueda usarse como una estrategia general para mejorar su estabilidad in vivo y para modular la estructura que presentan con el fin de conseguir una actividad farmacocinética óptima.

## 7.5. Enhanced Permeability and Binding Activity of Isobutylene-Grafted Peptides.

Resumen

En este trabajo se presenta una nueva estrategia de macrociclación de péptidos mediante la inserción de un isobutileno. Dicha estrategia consiste en la utilización de 3-Bromo-2-(bromometil)prop-1-eno en condiciones suaves, consiguiendo de esta manera la ciclación de forma rápida y eficiente tanto de péptidos lineales como de péptidos cíclicos.

Los péptidos cíclicos generados mediante la inserción de isobutileno poseen una mayor permeabilidad pasiva a la membrana plasmática. Como demuestran las medidas de espectroscopia de RMN y las simulaciones de dinámica molecular, esto es debido a la disminución de la superficie polar del péptido causado por el enmascaramiento de los grupos amida del esqueleto peptídico.

Las estructuras que incorporan el isobutileno son totalmente estables tanto en plasma humano como en presencia de glutatión, pudiéndose aplicar esta estrategia en péptidos cíclicos bioactivos como la somatostatina. Además, es importante destacar que, en el caso de la somatostatina, esta modificación también mejora el reconocimiento por SSTR2.

Debido a todas las ventajas que posee, esta estrategia podría utilizarse de forma general para la obtención de fármacos basados en péptidos cíclicos.

## 7.6. The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer.

Resumen

En este trabajo se describe la obtención de una nueva generación de glicopéptidos asociados a tumores que presentan una mejora en la afinidad frente a anticuerpos anti-MUC1. Dicha mejora es resultado de la incorporación de (4S)-4-fluoroprolina o 4,4-difluoroprolina en el dominio más inmunogénico de la secuencia de la MUC1.

Los ensayos de afinidad realizados mediante *biolayer interferometry* muestran una afinidad de 3 a 10 veces mayor que la encontrada para los glicopéptidos naturales.

De acuerdo con la cristalografía de rayos X y con las simulaciones de dinámica molecular, los residuos fluorados estabilizan la unión antígeno-anticuerpo debido a la mejora en una interacción CH/π clave.

En consecuencia, se ha conseguido mejorar la detección de anticuerpos asociados a tumores en suero de pacientes con cáncer de próstata.

Debido a ello, estos nuevos glicopéptidos se pueden emplear como herramientas para la diagnosis temprana del cáncer de próstata, proporcionando una mayor sensibilidad que los análogos naturales.

## 7.7. Cell-Penetrating Peptides Containing Fluorescent D-Cysteines.

Resumen

En este trabajo se han sintetizado una serie de D-cisteinas fluorescentes cuyas propiedades ópticas han sido evaluadas. El paso clave en la síntesis de estos compuestos es la adición conjugada 1,4 de ariltioles sobre una deshidroalanina bicíclica quiral, la cual ha sido recientemente sintetizada por

nuestro grupo de investigación. La reacción es quimio y estereoselectiva, produciéndose de manera rápida a temperatura ambiente.

Los aductos de Michael son fácilmente derivatizables a sus correspondientes aminoácidos, lo que ha permitido estudiar sus propiedades ópticas y en algunos casos concretos protegerlos con Fmoc con el fin de poder incorporarlos en secuencias peptídicas mediante síntesis de péptidos en fase solida (SPPS).

Para demostrar la gran utilidad que posee esta estrategia, se han sintetizado dos péptidos, con capacidad de penetrar en células, que incorporan cisteinas funcionalizadas con los grupos cumarina y dansilo. Los grupos cromóforos introducidos permiten evaluar la internalización de los péptidos utilizando microscopia de fluorescencia.

## 7.8. Water Sculpts the Distinctive Shapes and Dynamics of the Tn Antigens: Implications for their Molecular Recognition.

Resumen

El antígeno Tn asociado a la presencia de tumores presenta dos variantes, α-O-GalNAc-Thr y y α-O-GalNAc-Ser. Cuando estas moléculas se encuentran en disolución presentan conformaciones diferentes, lo que hace que el reconocimiento molecular por receptores proteicos sea distinto.

En este trabajo se demuestra teórica y experimentalmente que las conformaciones establecidas por las dos moléculas en fase gas son muy similares, lo que evidencia la importancia del papel de las moléculas de agua en disolución. Los estudios teóricos indican que en el caso del antígeno de Thr, la primera esfera de hidratación induce a que el enlace glicosídico gire. Este cambio conformacional fuerza el enmascaramiento del grupo metilo,

permitiendo de esta manera una mejor solvatación de la parte polar del antígeno.

La nueva disposición adoptada por α-O-GalNAc-Thr promueve la aparición de una molécula de agua situada en un bolsillo hidrofílico establecido entre el aminoácido y el carbohidrato. Este mecanismo está avalado por las estructuras de rayos X obtenidas de los antígenos fluorados asociados al anticuerpo anti-MUC1 SM3, en las que se puede ver esa molécula de agua formando parte de la estructura cristalina.

## 7.9. Structural and Mechanistic Insights into the Catalytic-Domain-Mediated Short-Range Glycosylation Preferences of GalNAc-T4.

Resumen

La O-glicosilacion de las mucinas es iniciada por una familia de enzimas polipeptídicas conocidas como GalNAc-transferasas (GalNAc-Ts). Estas enzimas son proteínas transmembrana de tipo II que contienen dominios catalíticos y de tipo lectina, ambos conectados por un linker flexible, en el lumen del aparato de Golgi.

Varias de estas enzimas, incluyendo la GlNAc-T4, llevan a cabo glicosilaciones de largo alcance (dominio lectina) y de corto alcance (dominio catalítico) de una manera específica. Mientras que el mecanismo de glicosilación mediado por el dominio lectina se conoce bien, el mecanismo gobernado por el dominio catalítico no se ha estudiado con suficiente detalle.

Este trabajo muestra la primera estructura cristalina de rayos X del complejo formado por la enzima GalNAc-T4 y el diglicopéptido GAT*GAGAGAGT*-

TPGPG. Esta estructura muestra el modo de unión a GalNAc que presenta el dominio catalítico.

Los estudios cinéticos realizados muestran que la unión a GalNAc del dominio lectina predomina sobre la unión al dominio catalítico siendo posible la supresión de ambas de manera independiente.

Además, se ha observado la presencia de un bucle flexible que sobresale del dominio lectina, el cual resulta esencial para la actividad óptima del dominio catalítico.

## 7.10. Efficient and irreversible antibody–cysteine bioconjugation using carbonylacrylic reagents.

Resumen

En la actualidad, existe un elevado interés en el desarrollo de métodos químicos para la modificación selectiva de anticuerpos con el fin de atribuirles propiedades terapéuticas. En este trabajo se describe una estrategia para la modificación de manera selectiva e irreversible de los residuos de Cys presentes en los anticuerpos.

Esta modificación se lleva a cabo mediante el empleo de la reacción S-Michael sobre reactivos carbonilacrílicos funcionalizados con fármacos citotóxicos.

Esta aproximación es una buena alternativa a la técnica convencional basada en la reacción con maleimida ya que la reacción es irreversible, utiliza reactivos sintéticamente accesibles y puede ser llevada a cabo de manera sencilla sobre un gran número de anticuerpos sin que éstos pierdan afinidad por sus antígenos.

## 7.11. Oxygen by Carbon Replacement at the Glycosidic Linkage Modulates the Sugar Conformation in Tn Antigen Mimics.

Resumen

En este trabajo se ha llevado a cabo la síntesis de diferentes miméticos del antígeno de Thr con el fin de evaluar la influencia que posee el aminoácido en la conformación adoptada por la N-Acetilgalactosamina.

Los cambios efectuados se basan en la sustitución del aminoácido natural Thr por aminoácidos no naturales. Los glicoaminoácidos miméticos del antígeno Tn sintetizados presentan tanto enlaces O-glicosídicos como C-glicosídicos. Además, con motivo de aumentar la rigidez que ya presenta el enlace glicosídico en el derivado natural, se han sintetizado compuestos que incorporan sistemas bicíclicos conformacionalmente restringidos.

Las nuevas moléculas sintetizadas han sido sometidas a un profundo estudio conformacional mediante la utilización de espectroscopia de RMN, cálculos computacionales a nivel cuántico (QM) y simulaciones de dinámica molecular (MD). Estos estudios muestran una preferencia por la silla $^{1}C_{4}$ en el caso de los C-glicósidos, mientras que para los O-glicósidos el carbohidrato adopta una conformación $^{4}C_{1}$.

## 7.12. Quaternization of Vinyl/Alkynyl Pyridine Enables Ultrafast Cysteine-Selective Protein Modification and Charge Modulation.

Resumen

Las vinil- o alquil-piridinas cuaternizadas reaccionan de manera ultrarrápida y selectiva con residuos de cisteína presentes en proteínas, permitiendo la utilización de cantidades estequiométricas de los reactivos.

En este trabajo se demuestra la utilidad que posee esta estrategia a la hora de la funcionalización de anticuerpos con fármacos, ya que, los conjugados presentan una gran estabilidad en suero humano y no pierden la especificidad por sus receptores.

Adicionalmente, la funcionalización de proteínas con estos derivados introduce una carga neta positiva en el conjunto. Esto último permite la modulación de la movilidad de las proteínas funcionalizadas en el gel electroforético con la simple incorporación de estos derivados en residuos de Cys presentes en la secuencia proteica.

## 7.13. Conformational Behavior of D-Lyxose in Gas and Solution Phases by Rotational and NMR Spectroscopies.

Resumen

Para mejorar las propiedades de los carbohidratos y poder usarlos como agentes terapéuticos es crucial entender sus preferencias conformacionales. En este trabajo se estudia de manera experimental el paisaje conformacional del monosacárido D-lixosa, para el cual los resultados obtenidos mediante cálculos computacionales de mecánica cuántica suelen ser muy dependientes

del método. En este estudio se evalúan las preferencias estructurales en fase gas mediante el empleo de espectroscopia rotacional, mientras que para determinar su comportamiento en disolución se emplean cálculos de dinámica molecular apoyados en parámetros experimentales obtenidos de espectros de RMN.

En contraste con los resultados que proporcionan los cálculos de mecánica cuántica, en fase gas, la D-lixosa adopta únicamente la forma piranosa, con una relación entre la conformación $^4C_1$ y $^1C_4$ de 60:40 para el caso del anómero α, mientras que solo aparece la conformación $^4C_1$ en el caso del anómero β. Por el contrario, en disolución, debido a que la solvatación es más efectiva, la configuración de la silla $^1C_4$ esta favorecida.

La combinación de los resultados de MW y RMN proporciona un método poderoso para desentrañar el papel que juega el agua en las preferencias conformacionales de moléculas complicadas como lo son los carbohidratos flexibles.

# Conclusiones

8

## 8.1. Conclusiones

Se ha desarrollado una metodología novedosa que permite la síntesis estereoselectiva de una amplia variedad de derivados de *S*-glicosilcisteína mediante el empleo de la adición sulfa-Michael doblemente estereoselectiva de tiocarbohidratos convenientemente protegidos sobre deshidroalaninas quirales. Estos compuestos son fácilmente derivatizables a los correspondientes *building blocks*, pudiéndose de esta manera emplear en la síntesis de péptidos en fase sólida.

Además, se ha realizado un estudio conformacional del mimético azufrado del antígeno Tn con serina (sulfa-Tn-Ser) mediante experimentos de RMN y cálculos de dinámica molecular. En dicho estudio se han comparado tanto las preferencias conformacionales como la primera esfera de hidratación del derivado azufrado con los *O*-glicosilaminoácidos naturales. Así se ha podido determinar que la disposición espacial que adopta el Tn no natural en disolución es muy similar a la encontrada en el antígeno Tn con treonina. De hecho, en ambas moléculas el carbohidrato se dispone de manera perpendicular con respecto a la cadena peptídica, generando de esta manera el mismo bolsillo de agua entre la parte carbohidrato y la cadena peptídica.

Además, se han estudiado en detalle los elementos claves para el reconocimiento molecular de antígenos tipo MUC1 por el anticuerpo scFv-SM3. El análisis de las estructuras de rayos X de los distintos complejos antígeno-anticuerpo indican que la glicosidación favorece la unión con el anticuerpo gracias a que el GalNAc establece interacciones estabilizantes (enlace de hidrógeno y CH/$\pi$) con los residuos de la superficie de la proteína. También permite explicar la preferencia del anticuerpo SM3 por los derivados de treonina (APDTRP y APDT*RP), ya que la conformación reconocida por el anticuerpo para el correspondiente glicopéptido con serina presenta una población muy minoritaria en disolución.

Estos hallazgos enfatizan las diferencias encontradas entre estos dos antígenos a la hora de ser reconocidos por anticuerpos anti-MUC1, pudiendo tener importantes implicaciones a la hora del diseño de nuevos anticuerpos y biosensores.

Se ha logrado diseñar una ruta sintética que permite la obtención de los *building blocks* de *S*-(α-D-GalNAc)-*S*Thr y *Se*-(α-D-GalNAc)-*Se*Thr, los cuales son aptos para la incorporación en secuencias peptídicas mediante el empleo de la técnica de síntesis de péptidos en fase solida (SPPS).

Además, se ha demostrado la posibilidad de modular el modo de presentación de los glicopéptidos que contienen el carbohidrato GalNAc mediante la sustitución del oxígeno del enlace glicosídico por un átomo de azufre o selenio. Gracias a este cambio, se consigue una preorganización del fragmento peptídico en disolución, la cual favorece el reconocimiento molecular por el anticuerpo scFv-SM3, mejorando así su afinidad por éste.

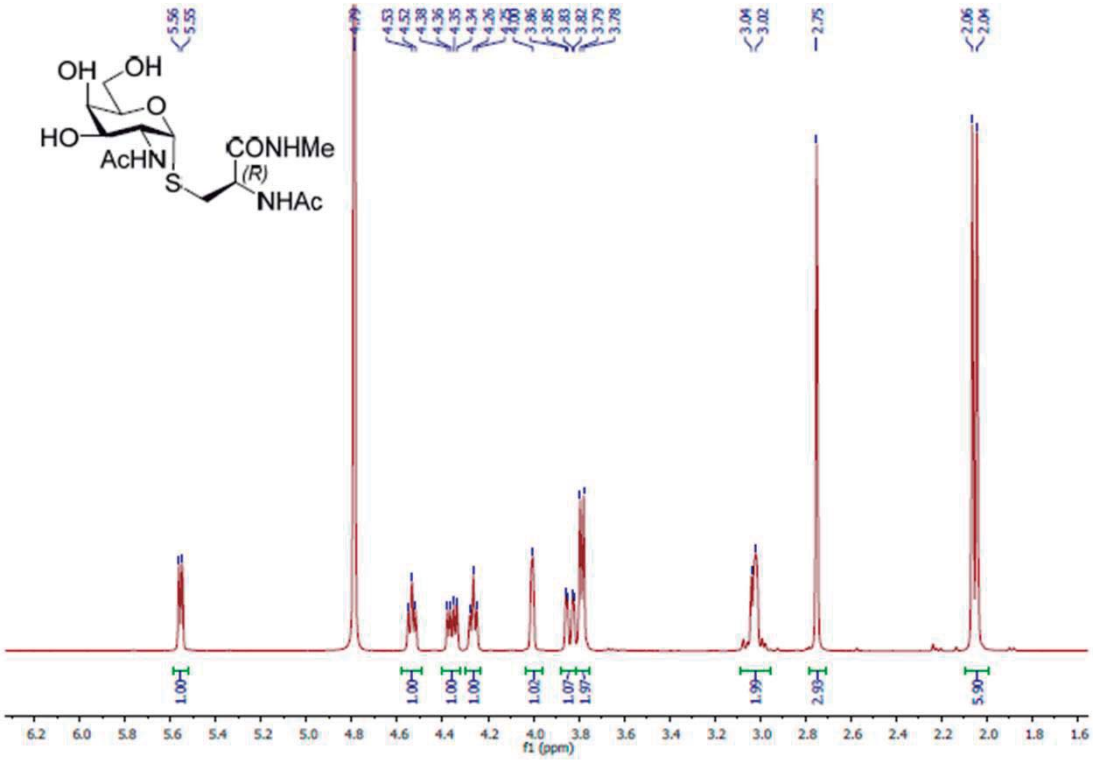
También se ha demostrado la capacidad como antígeno del péptido no natural que incorpora el aminoácido *S*-(α-D-GalNAc)-*S*Thr mediante la formulación de una vacuna terapéutica. Es importante señalar que los anticuerpos generados en ratones con esta vacuna reconocen de manera *selectiva* células cancerosas humanas.

Por último, se han efectuado una serie de colaboraciones con diversos grupos de investigación, a nivel nacional e internacional, basadas en la síntesis, purificación y caracterización de péptidos y glicopéptidos.

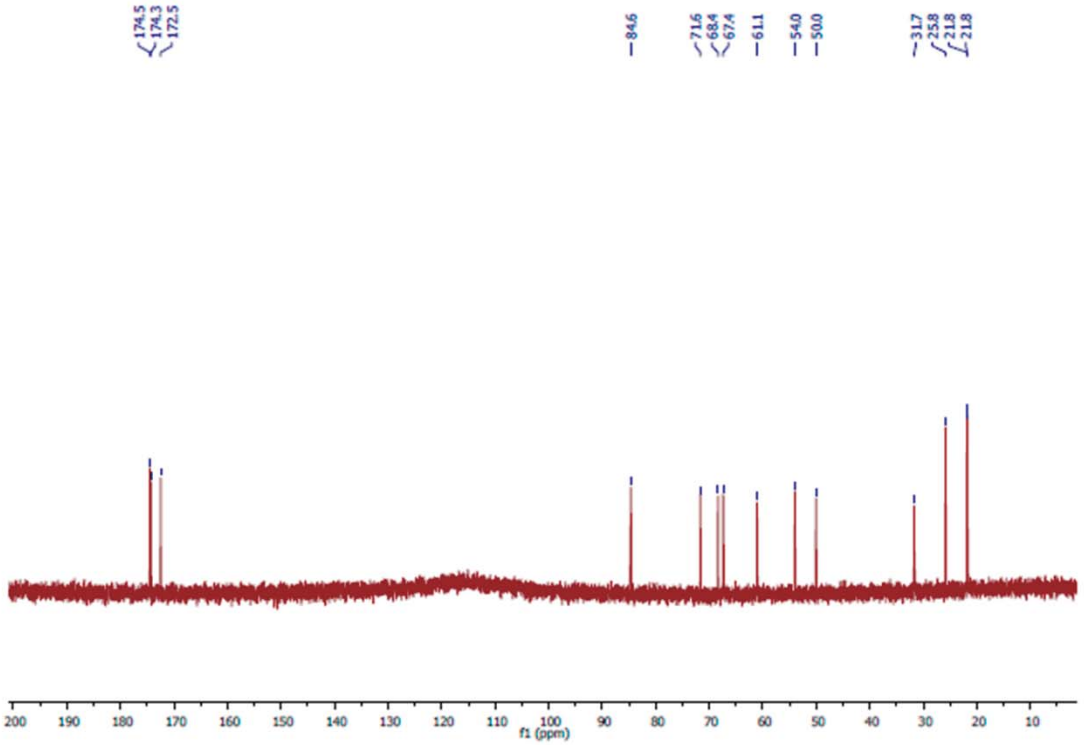
# Anexo I

Compuesto **3f** capítulo 4

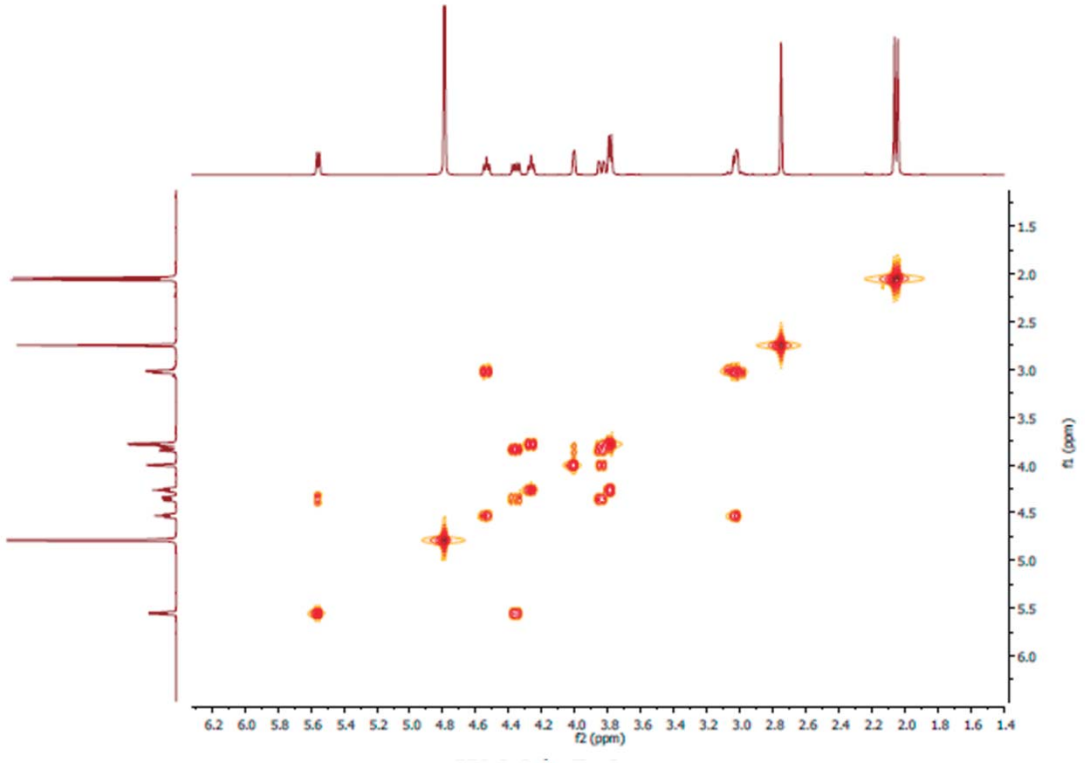
$^{1}$H NMR 400 MHz in $D_2O$



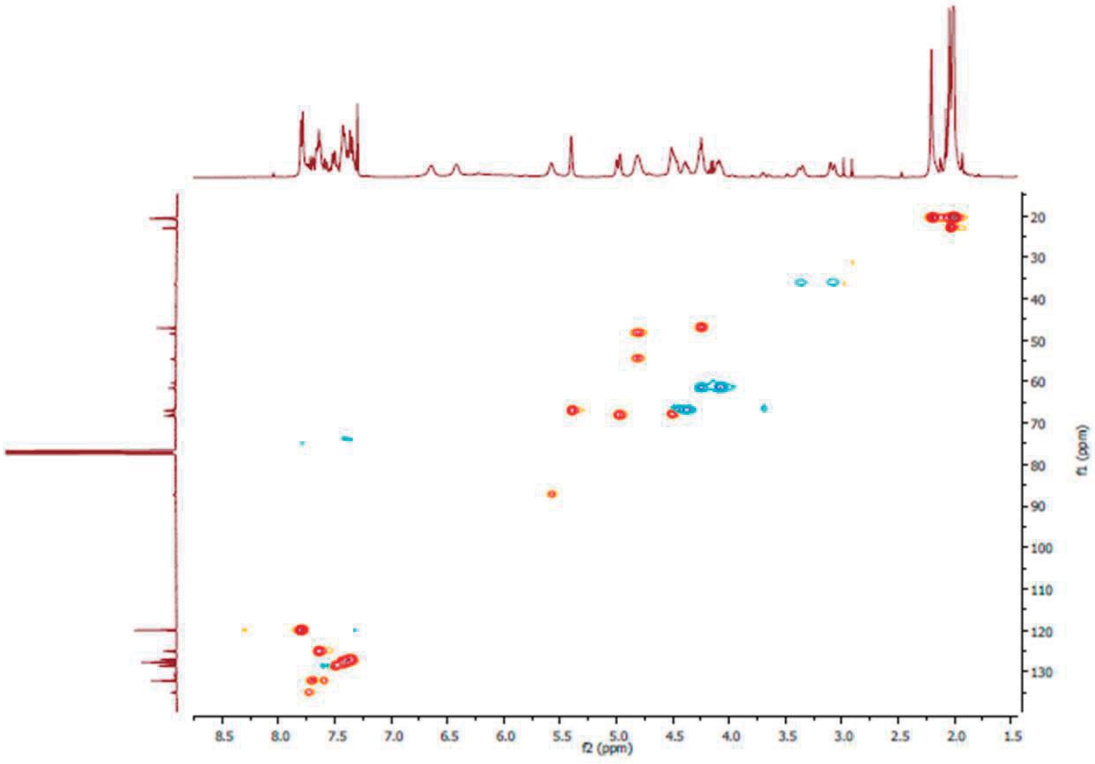

$^{13}$C NMR 100 MHz in $D_2O$



COSY in $D_2O$

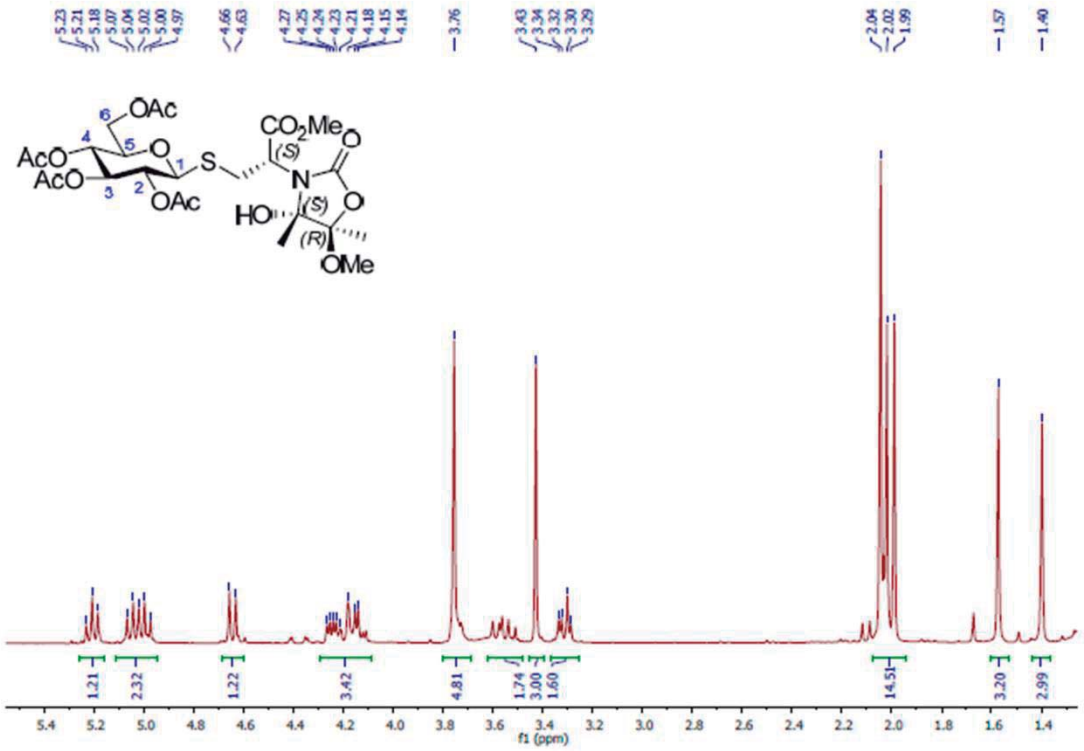



HSQC in $CDCl_3$

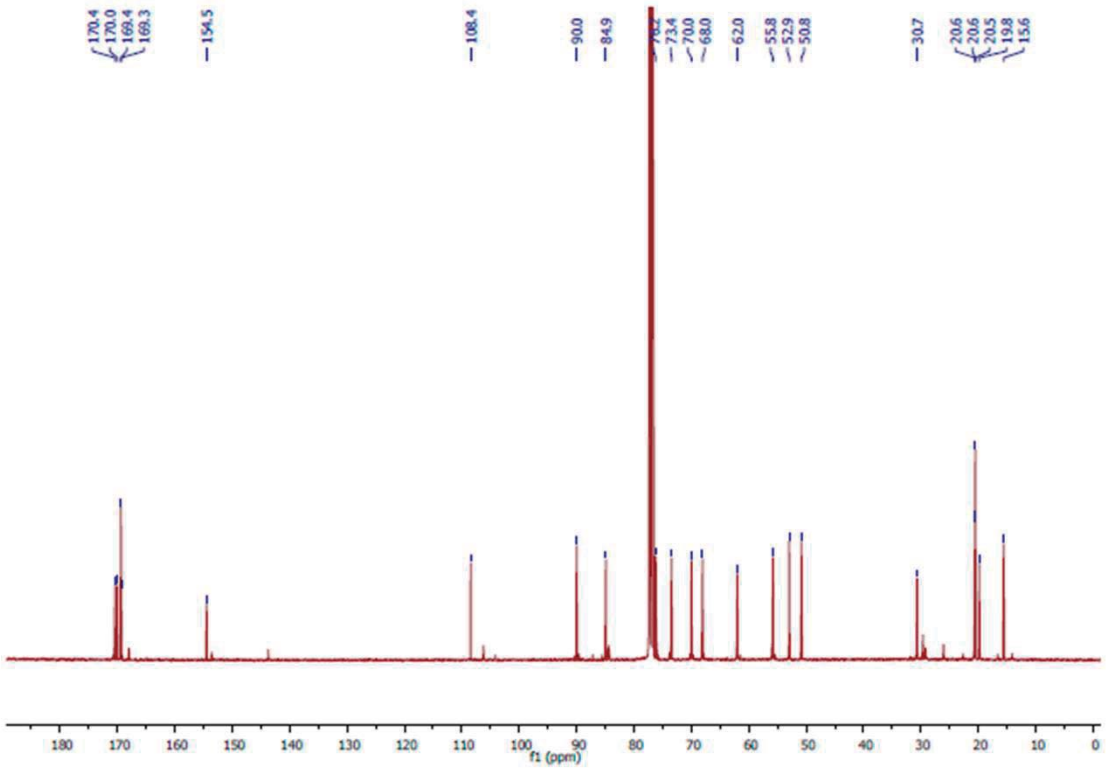


Compuesto **8a** capítulo 4

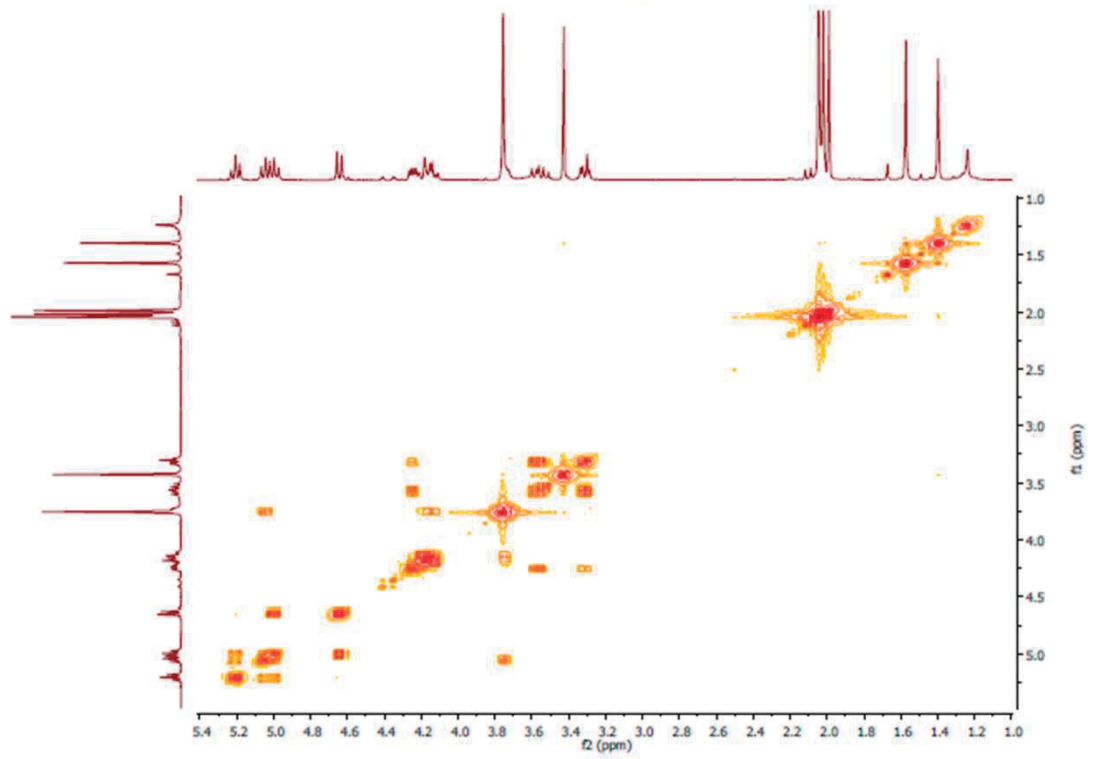
$^{1}$H NMR 400 MHz in $CDCl_3$



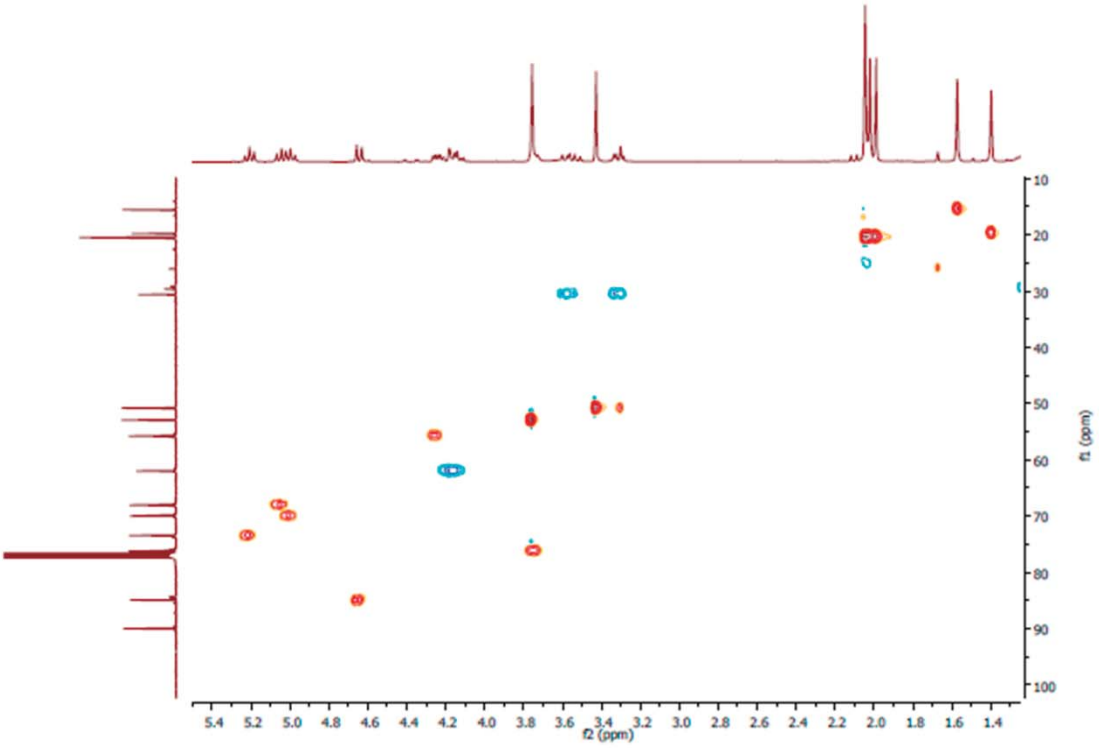

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

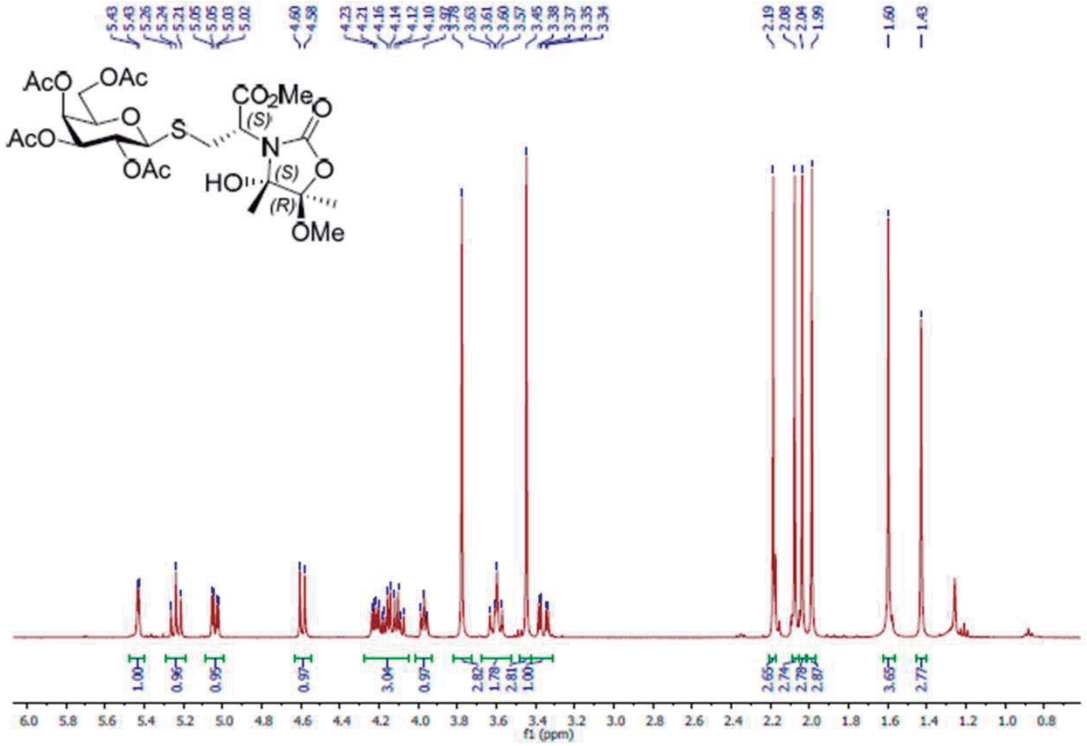



HSQC in $CDCl_3$

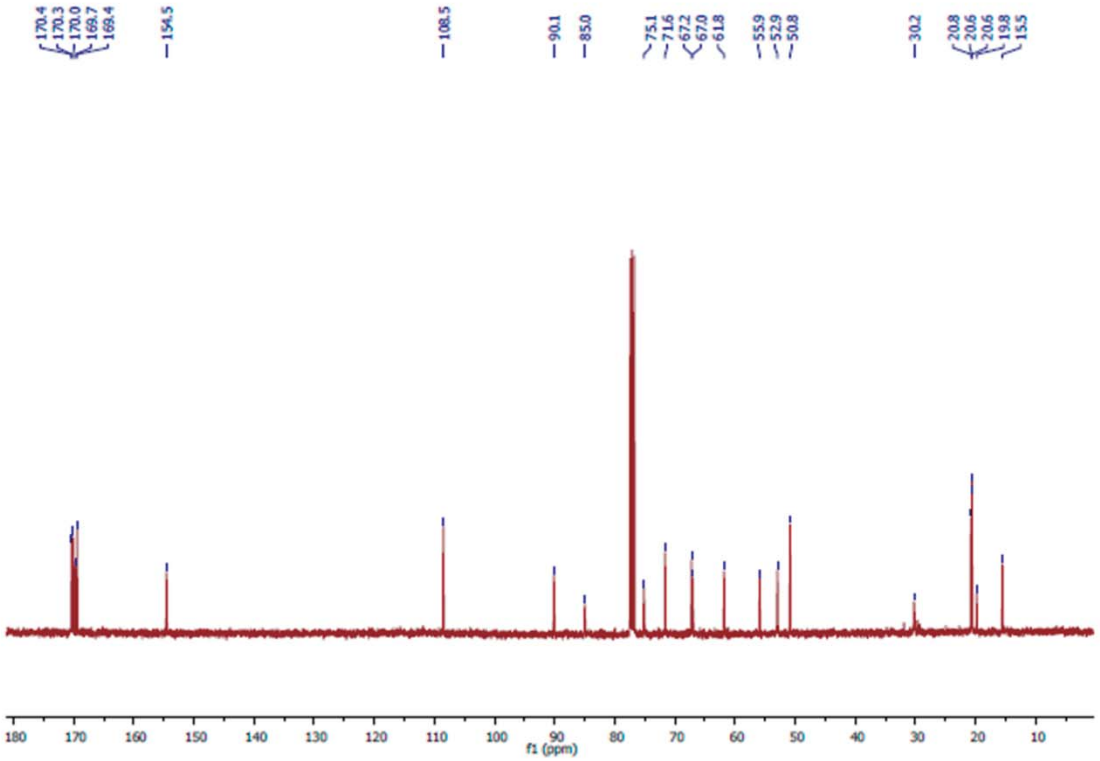


Compuesto **8b** capítulo 4

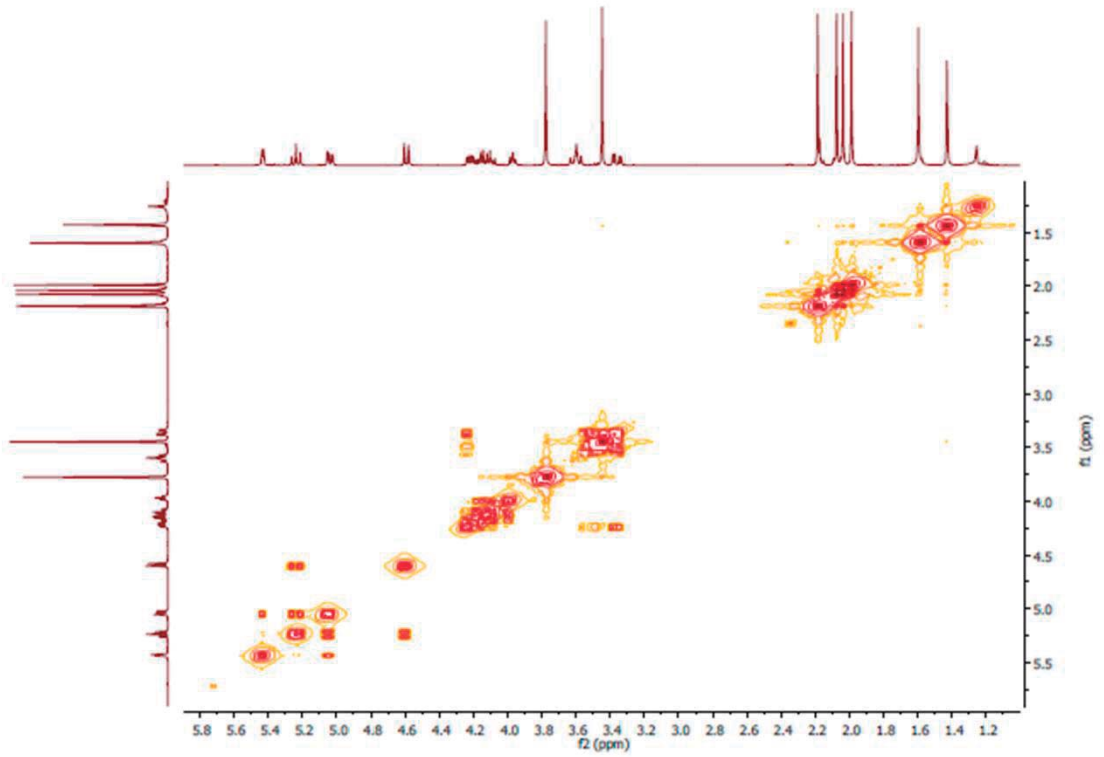
$^{1}$H NMR 400 MHz in $CDCl_3$



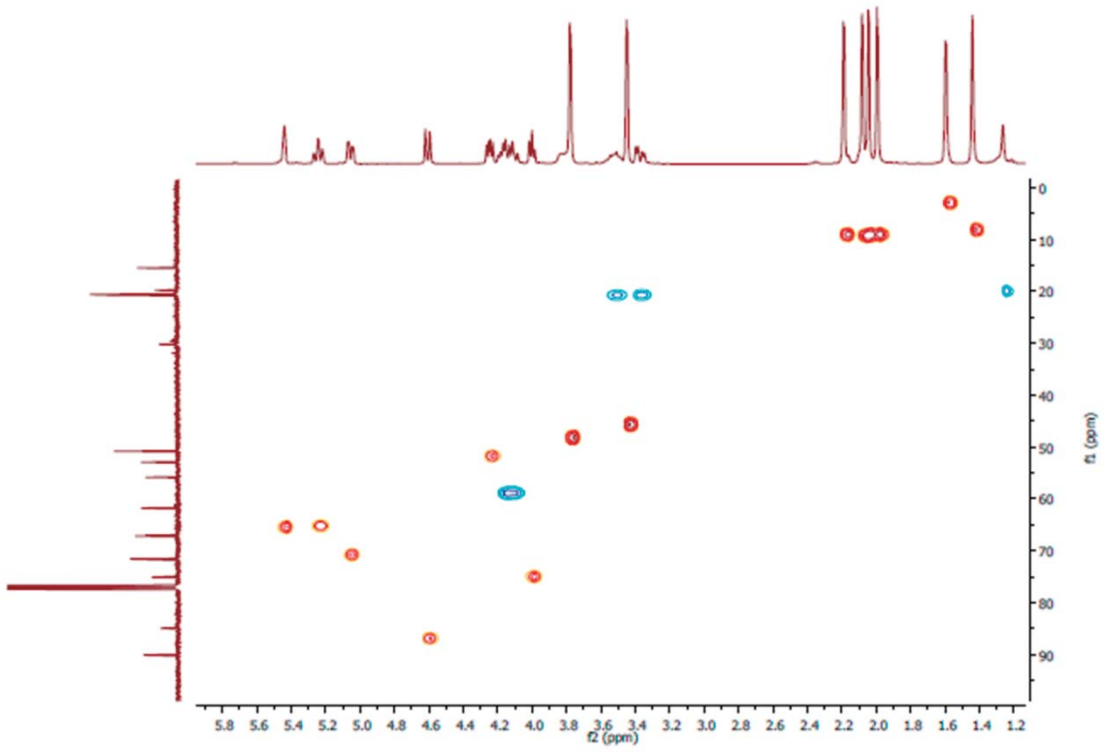

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

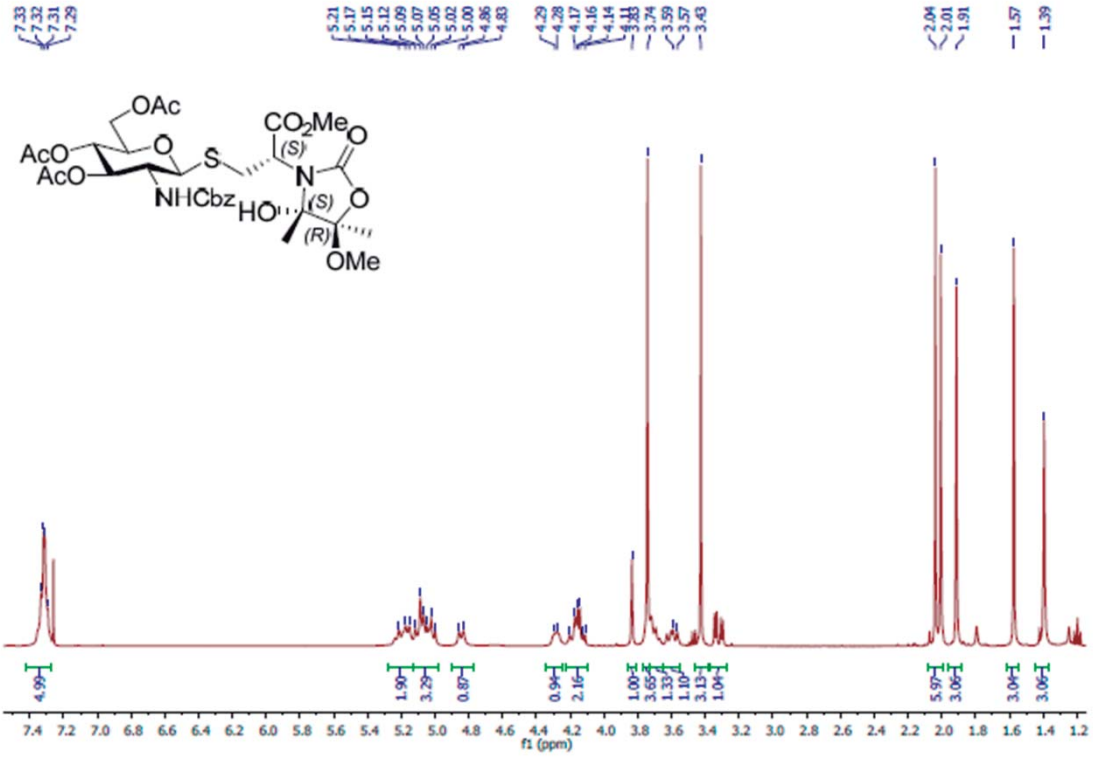



HSQC in $CDCl_3$

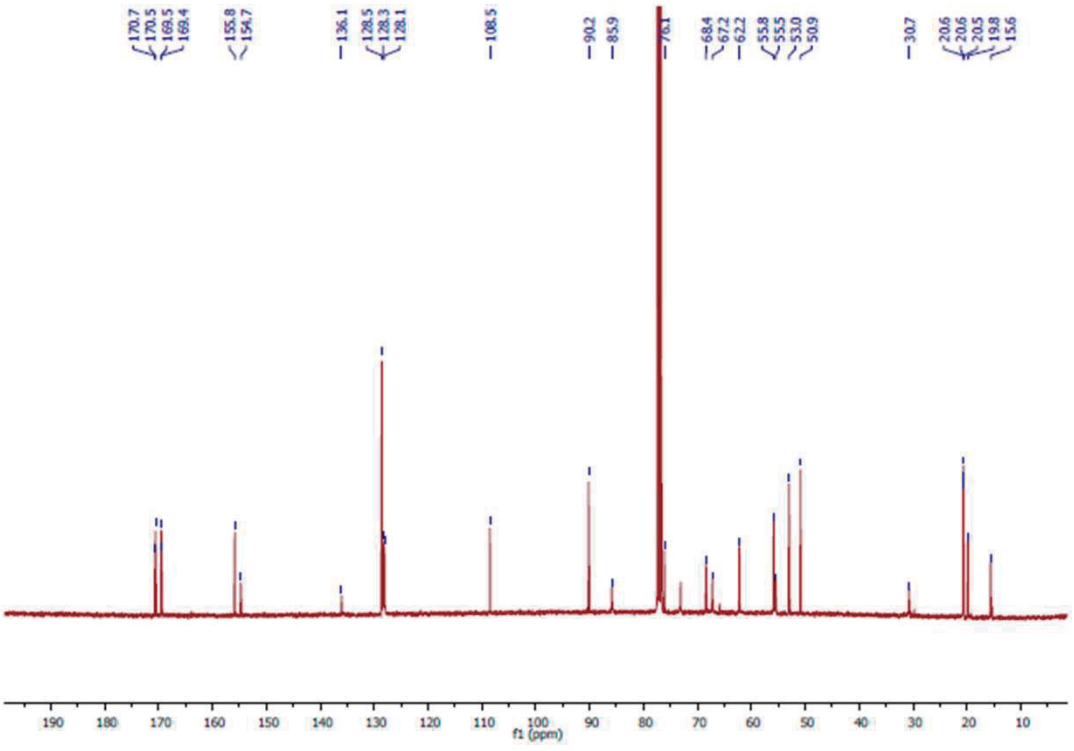


Compuesto **8c** capítulo 4

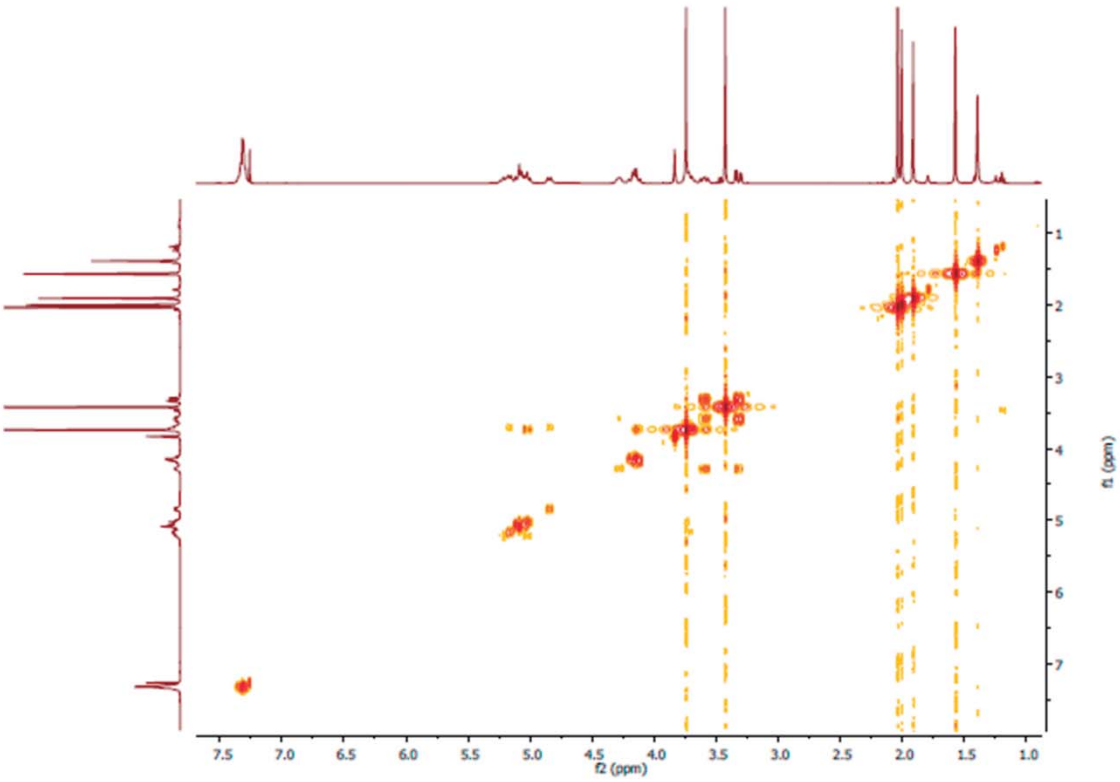
$^{1}$H NMR 400 MHz in $CDCl_3$



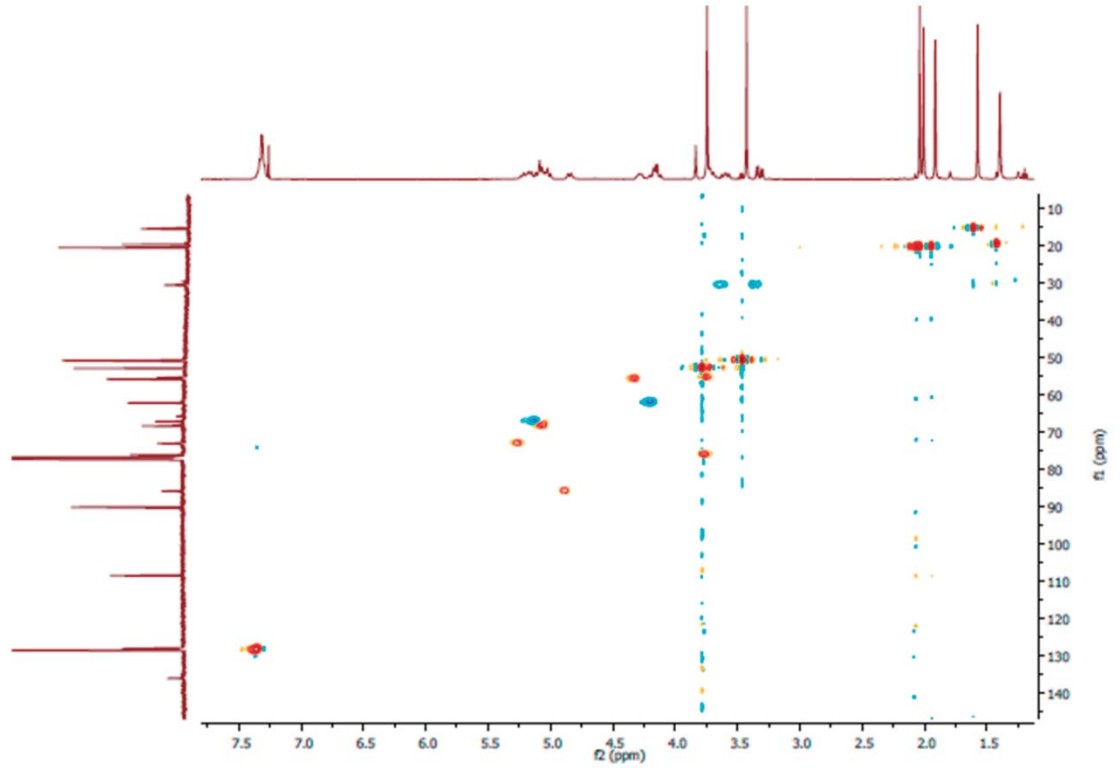

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

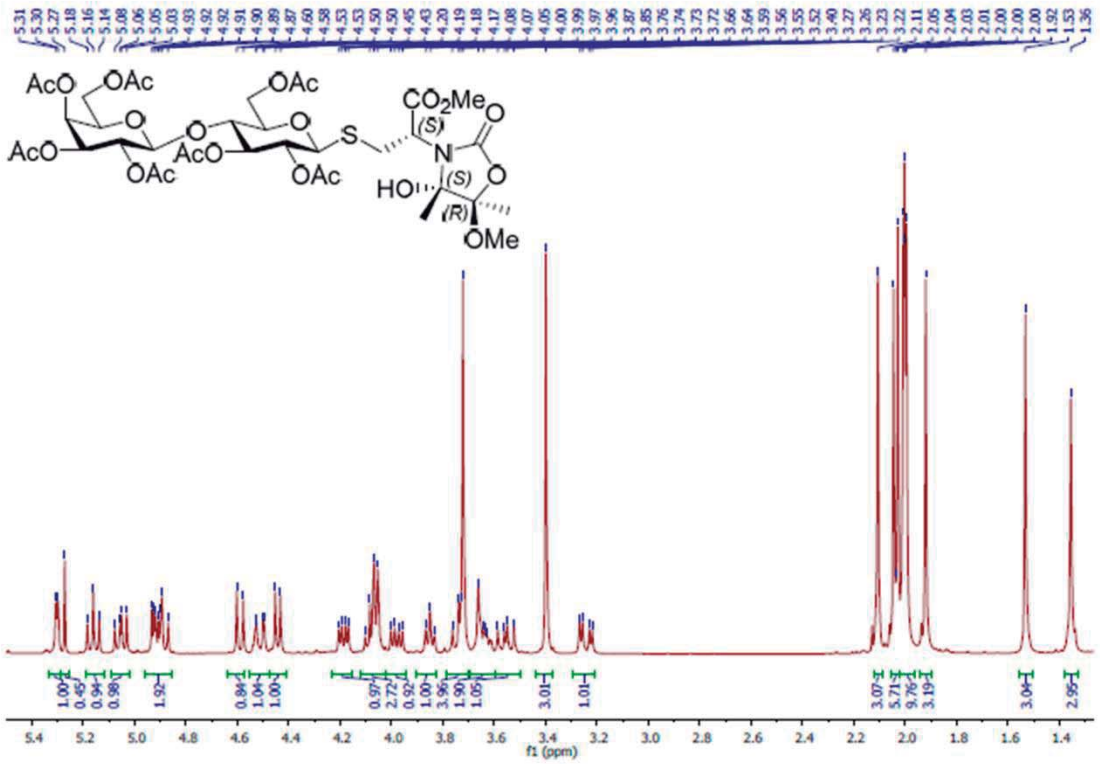



HSQC in $CDCl_3$

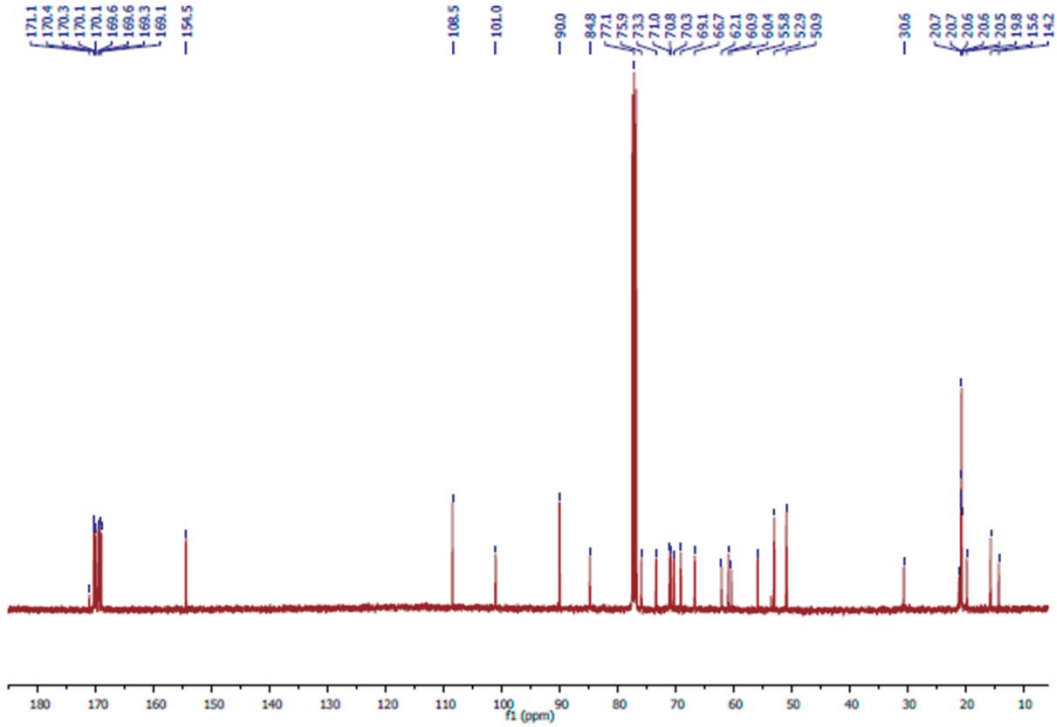


Compuesto **8d** capítulo 4

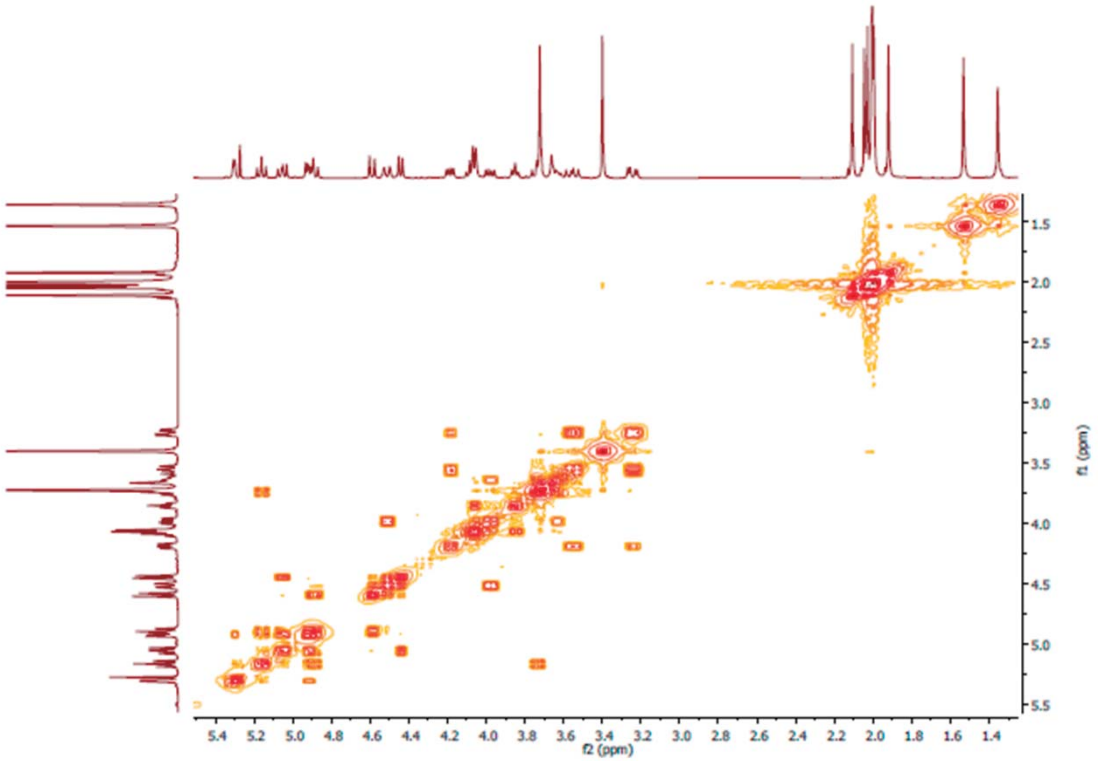
$^{1}$H NMR 400 MHz in $CDCl_3$



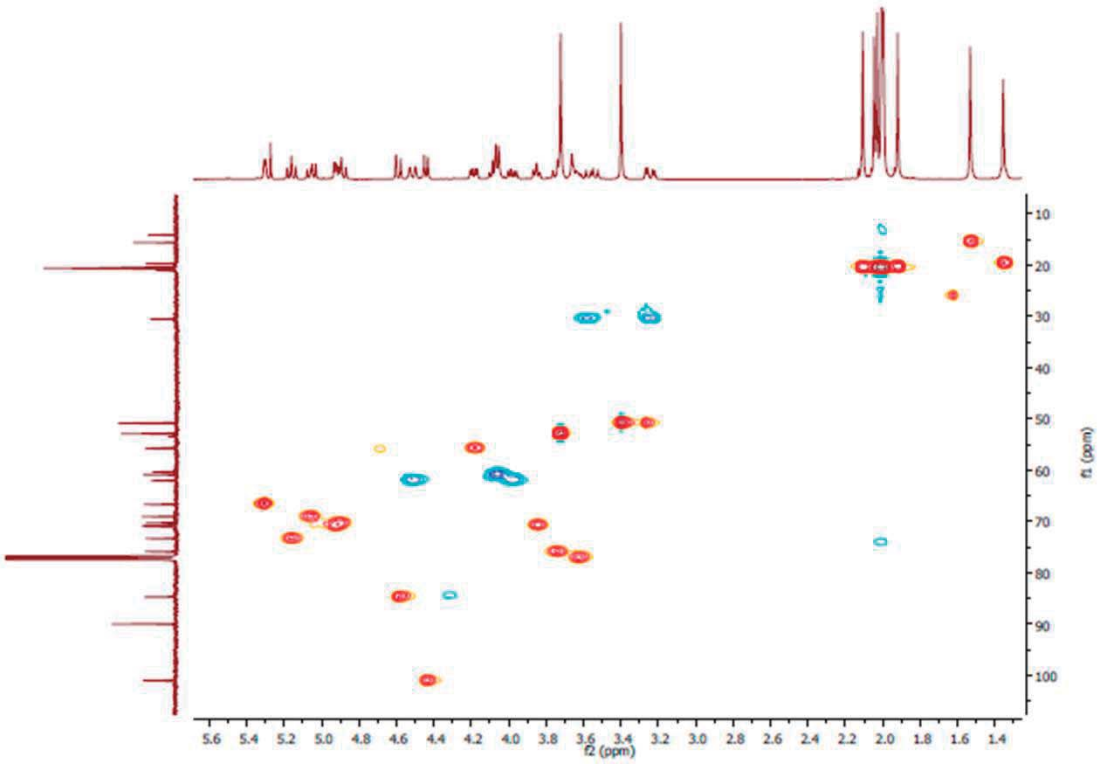

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

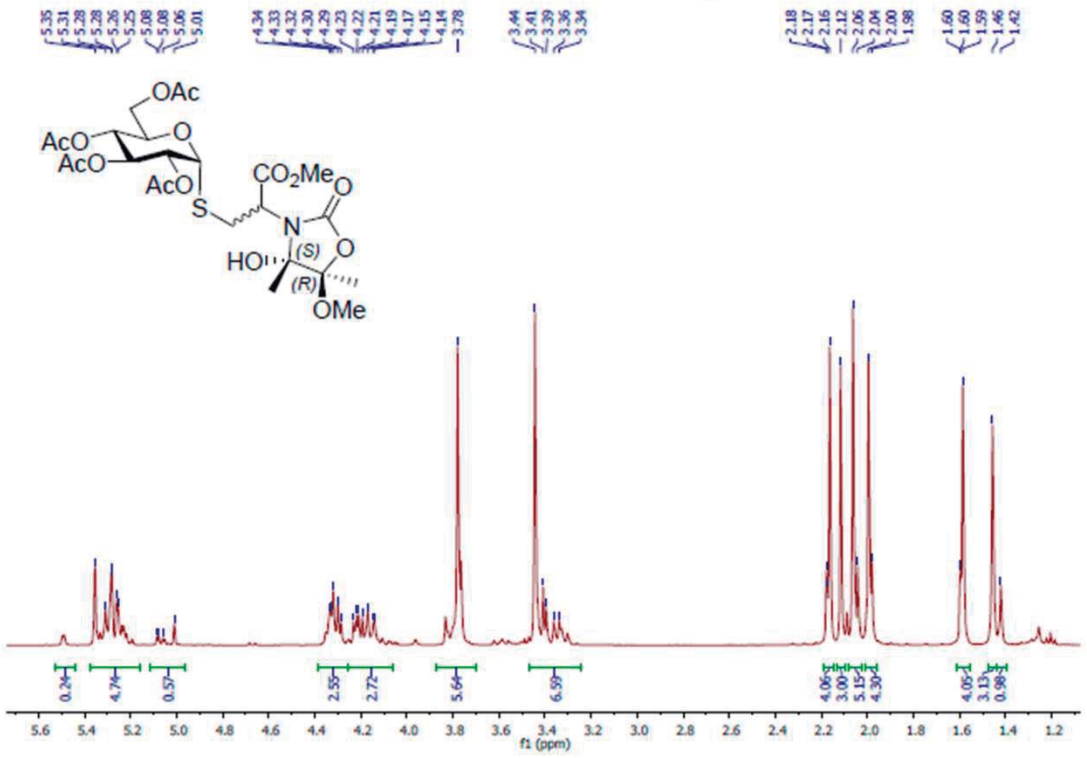



HSQC in $CDCl_3$

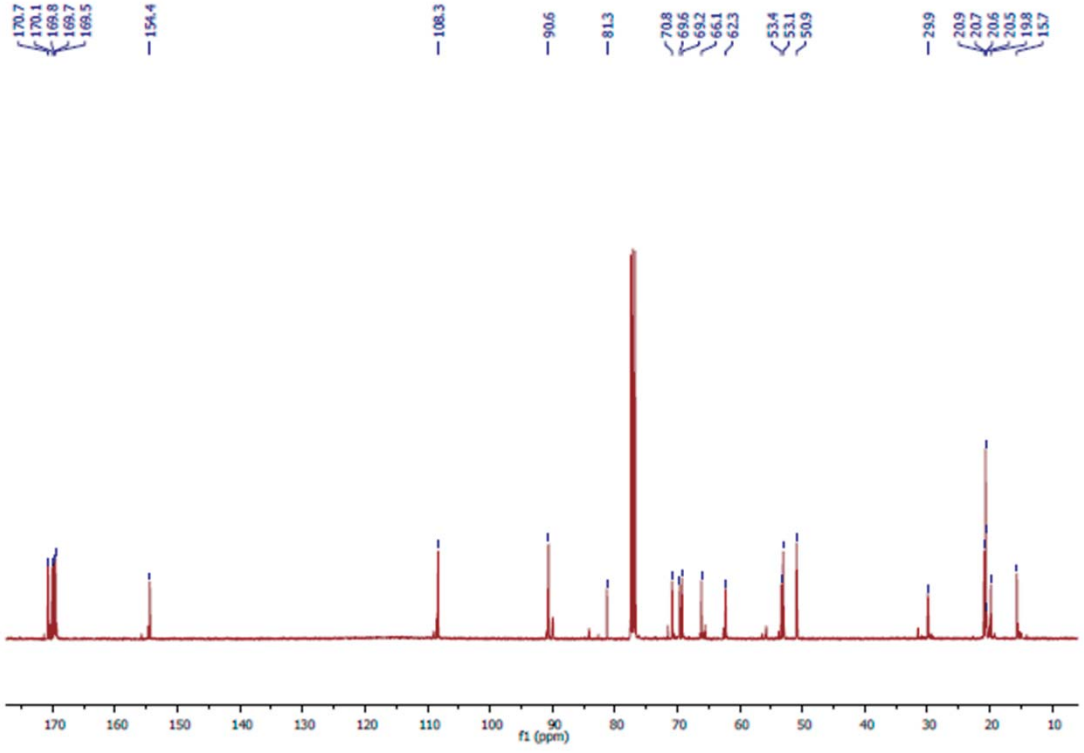


Compuestos **7e**/**8e** capítulo 4

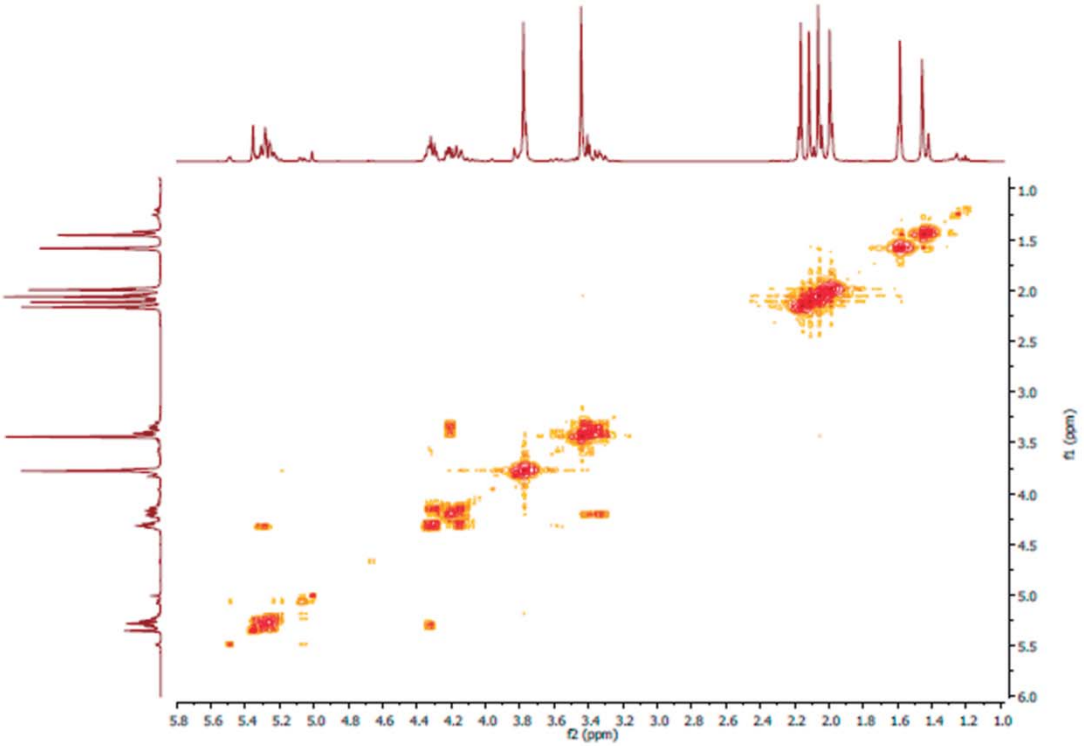
$^{1}$H NMR 400 MHz in $CDCl_3$



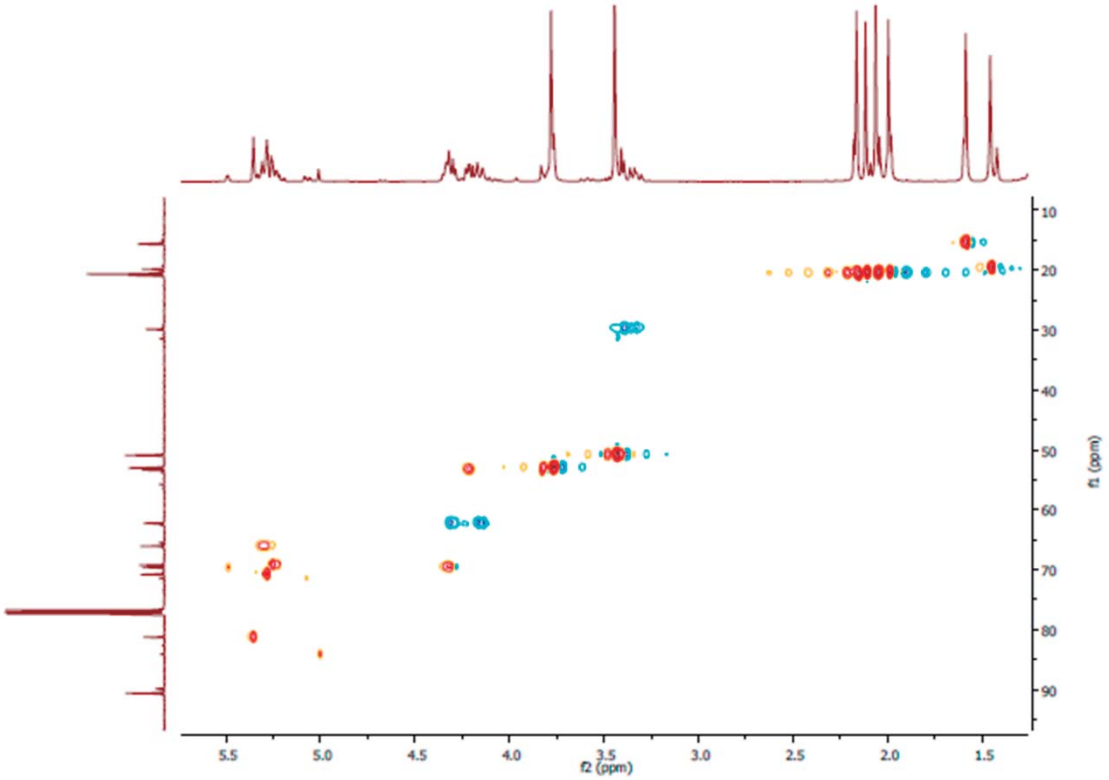

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

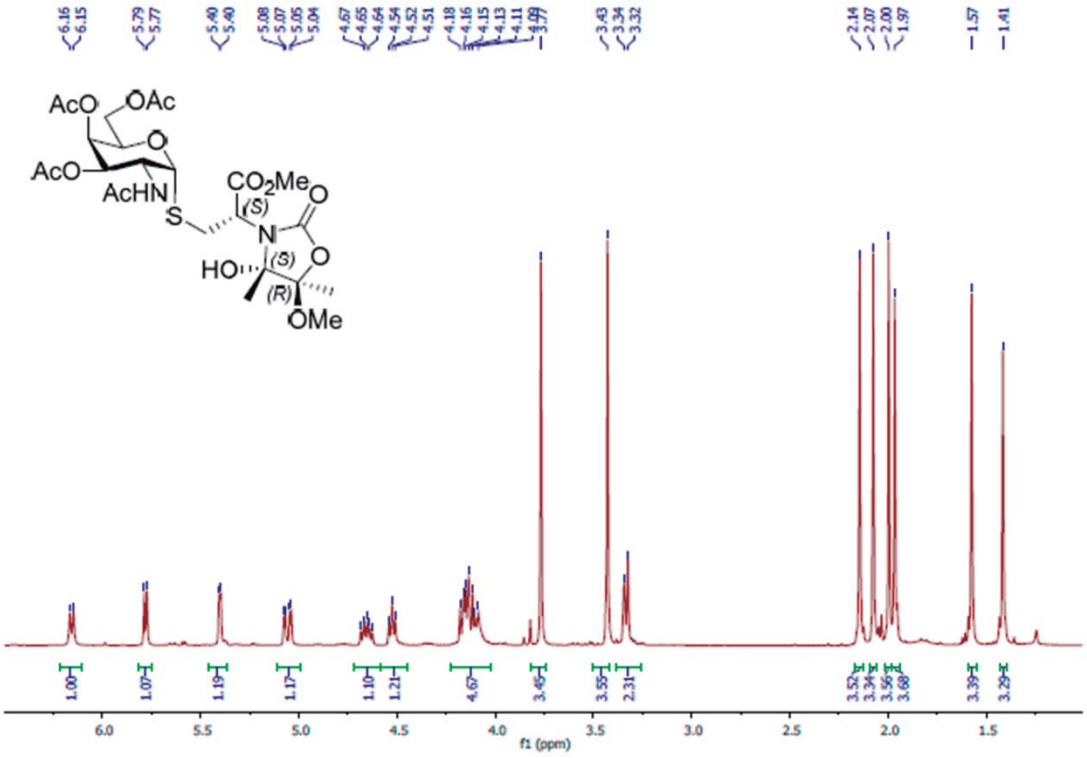



HSQC in $CDCl_3$

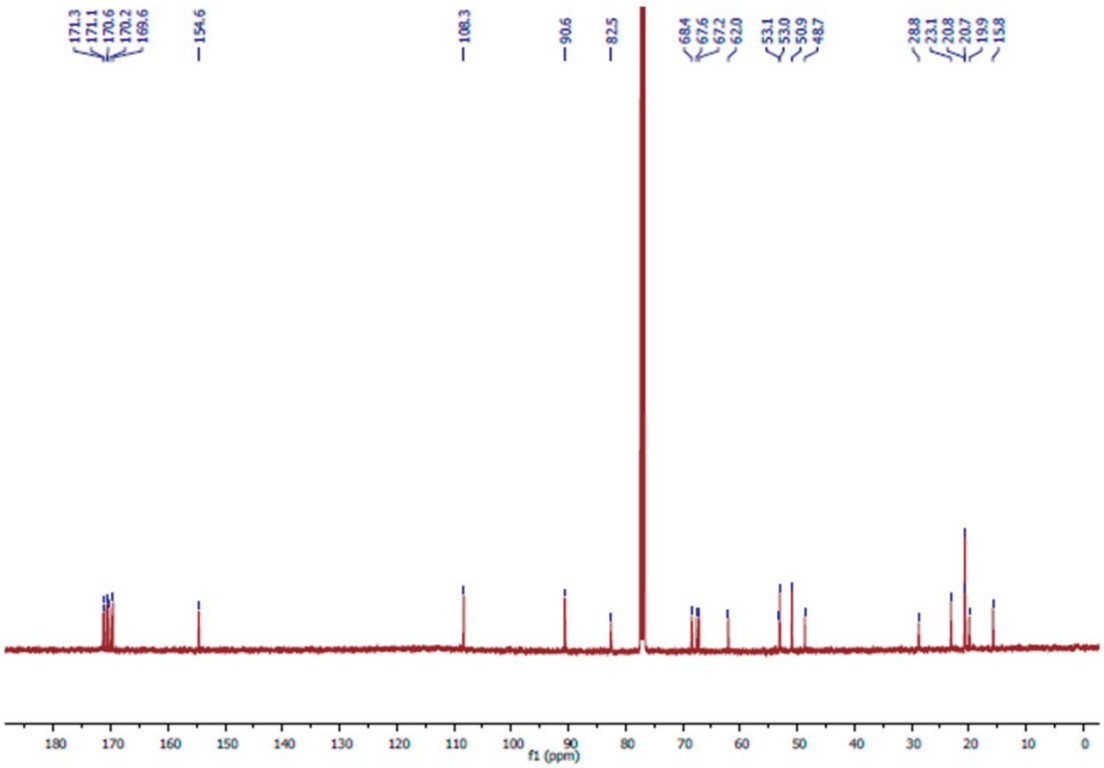


Compuesto **8f** capítulo 4

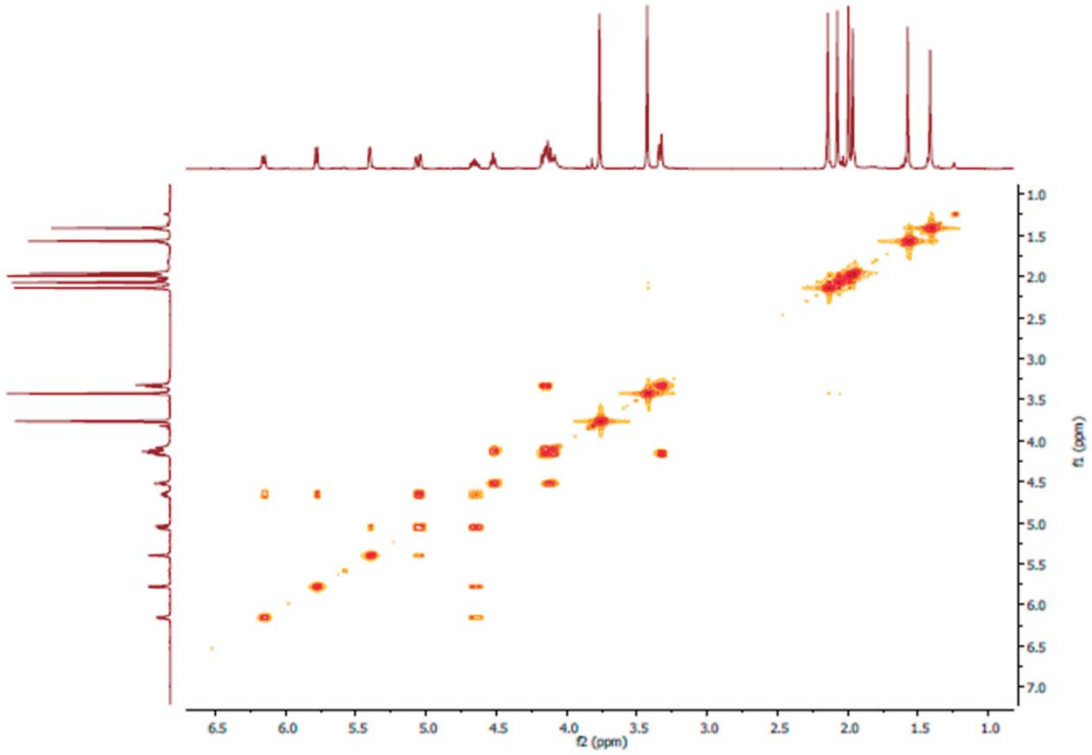
$^{1}$H NMR 400 MHz in $CDCl_3$



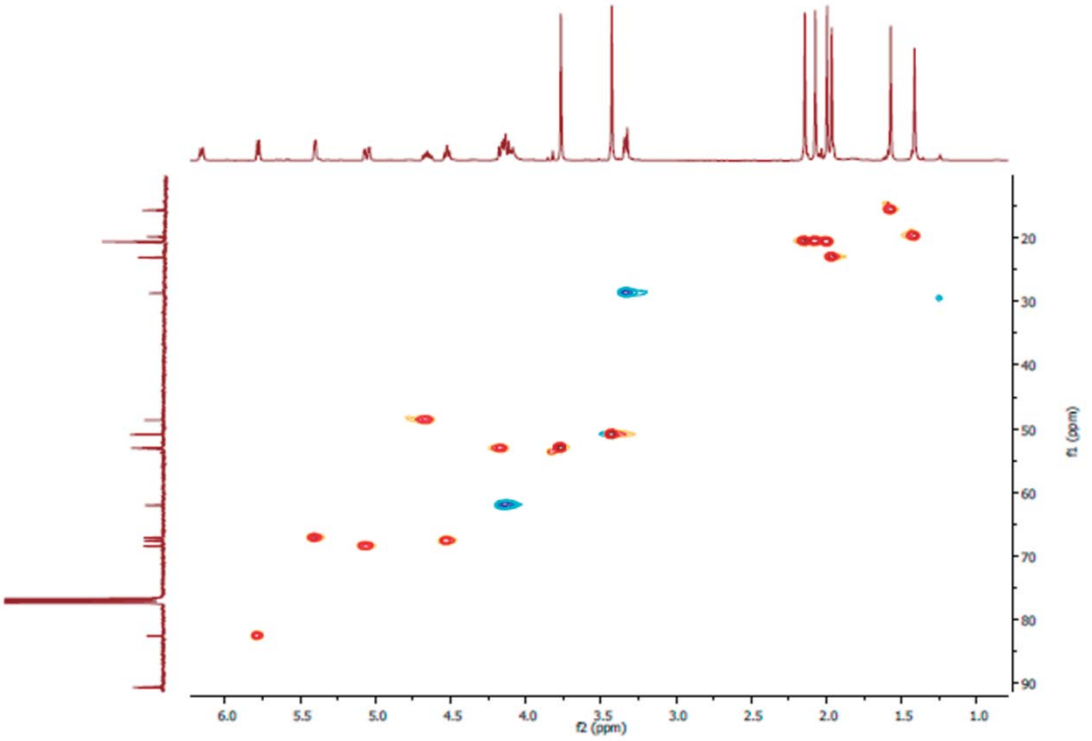

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

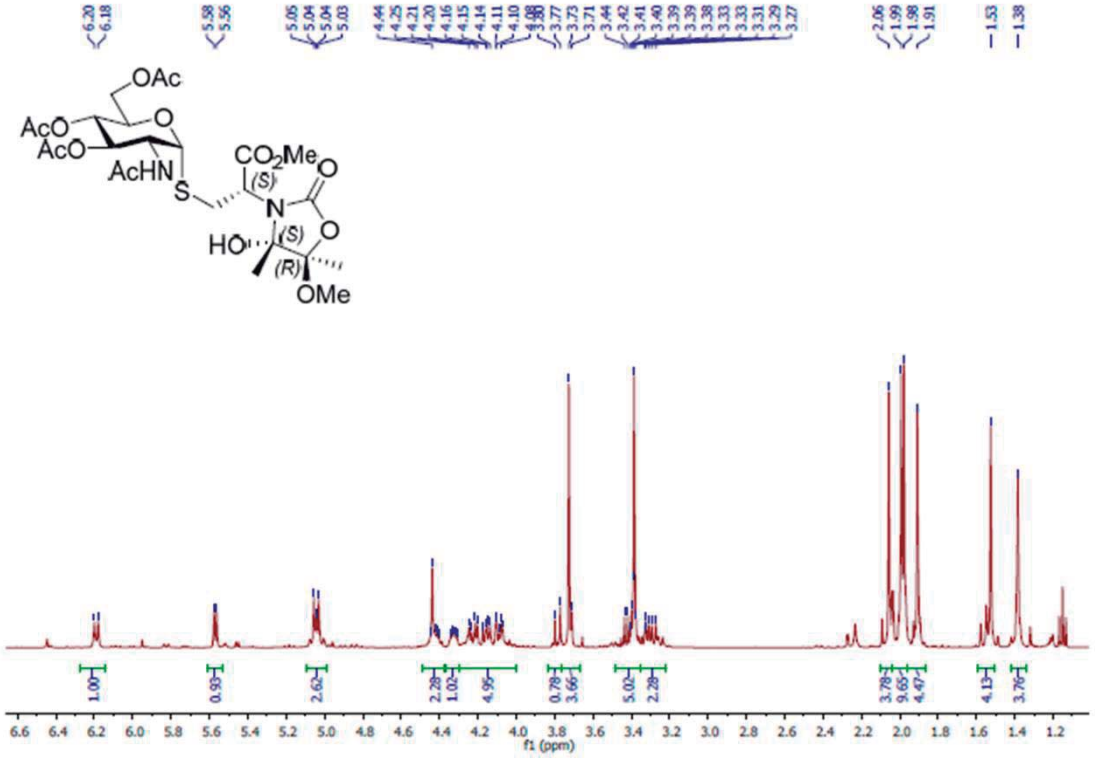



HSQC in $CDCl_3$

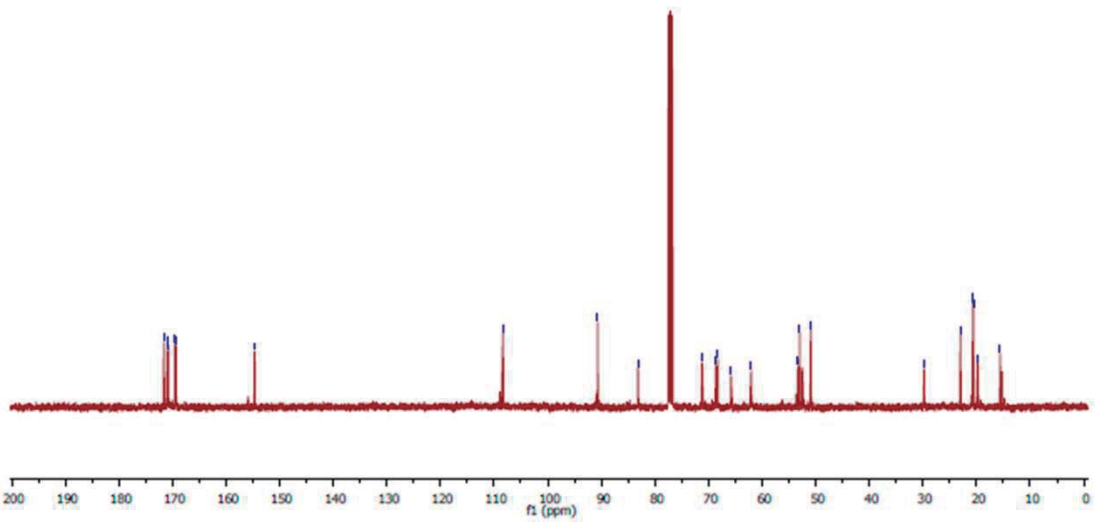


Compuesto **8g** capítulo 4

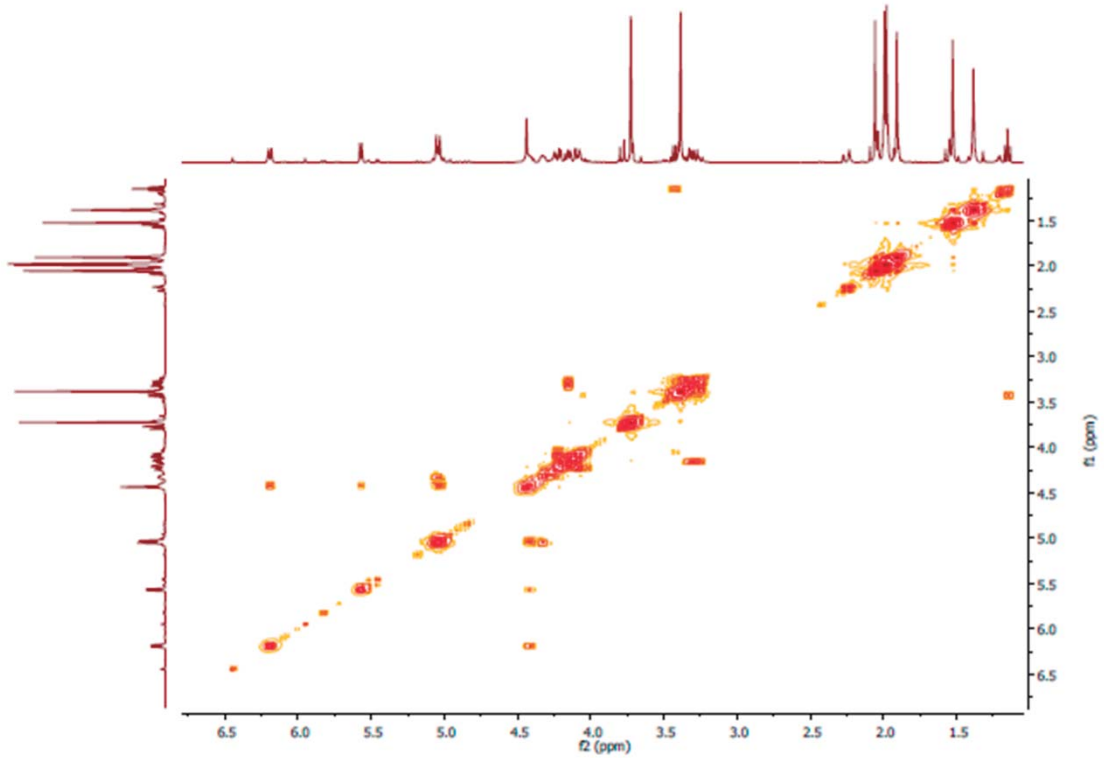
$^{1}$H NMR 400 MHz in $CDCl_3$



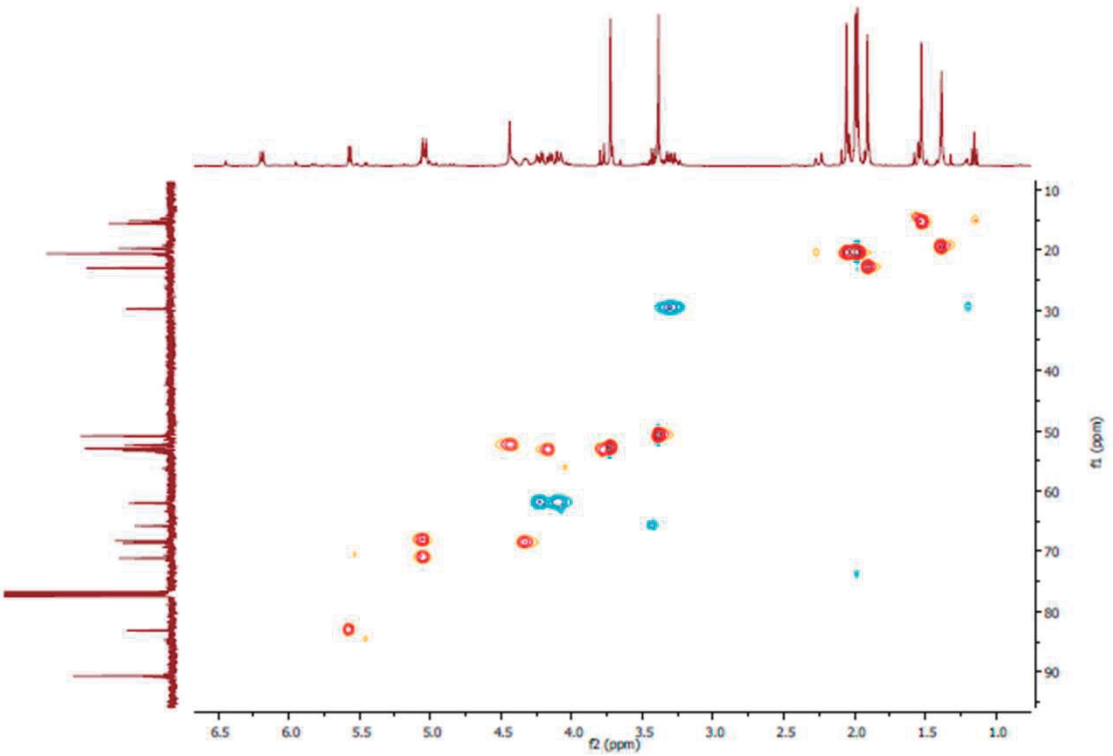

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

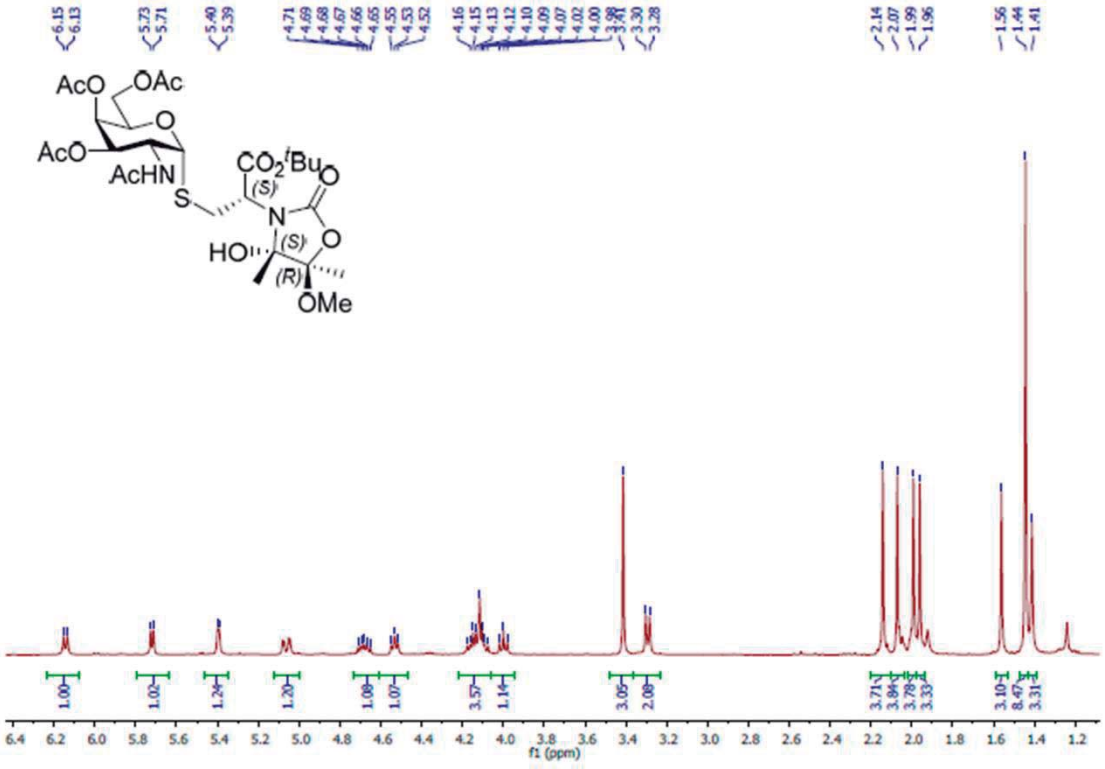



HSQC in $CDCl_3$

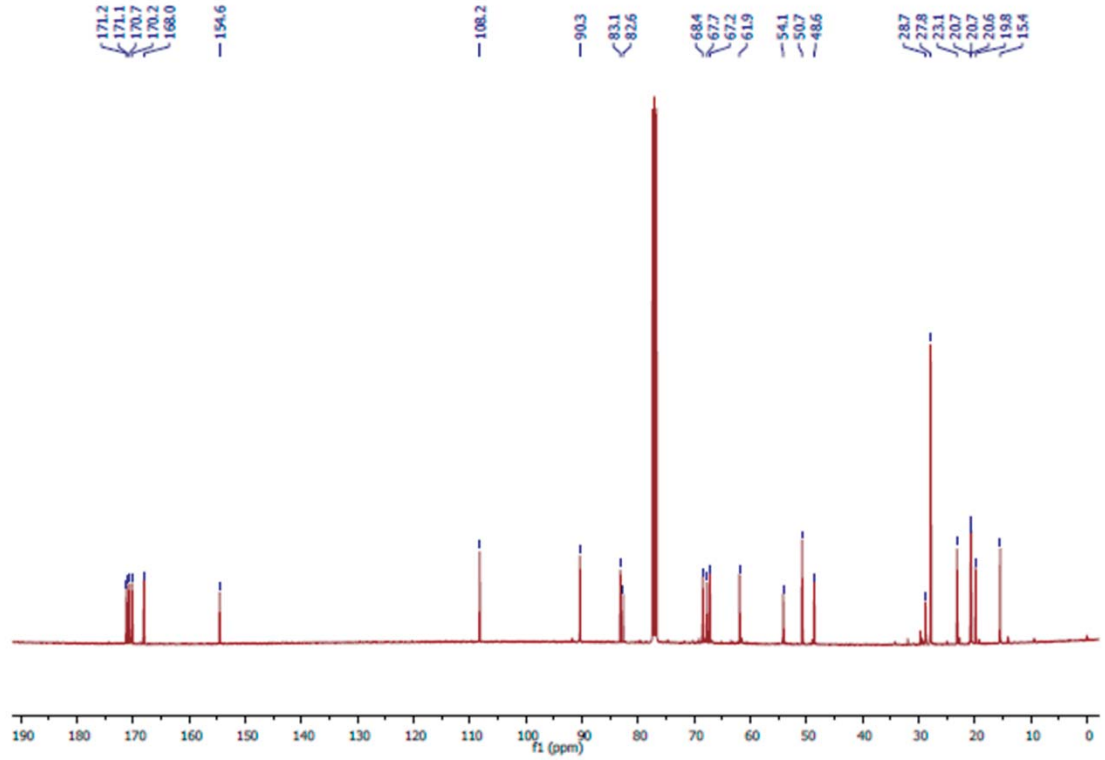


Compuesto **8'f** capítulo 4

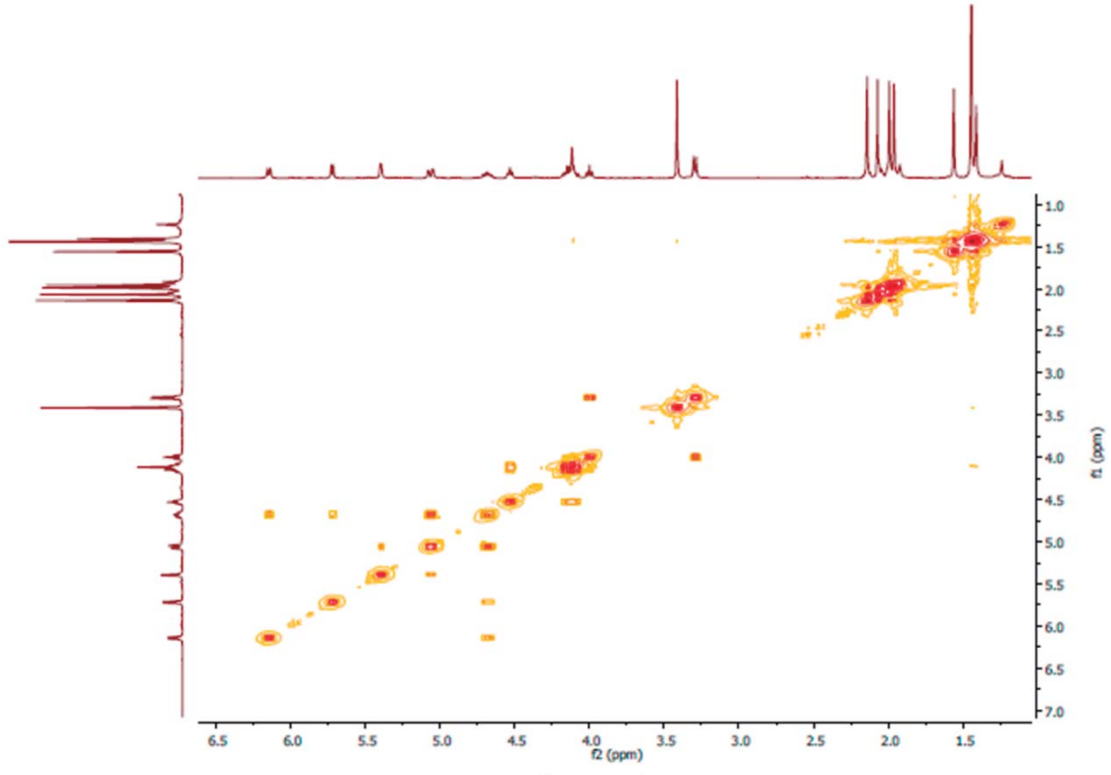
$^{1}$H NMR 400 MHz in $CDCl_3$



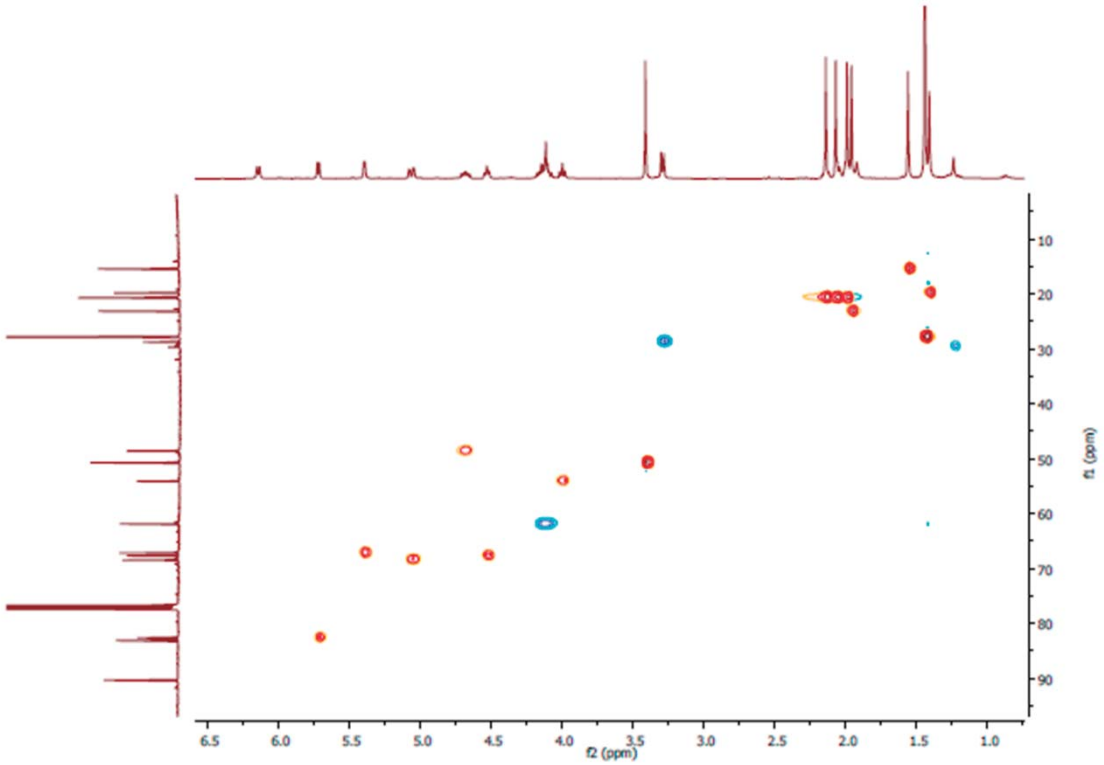

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

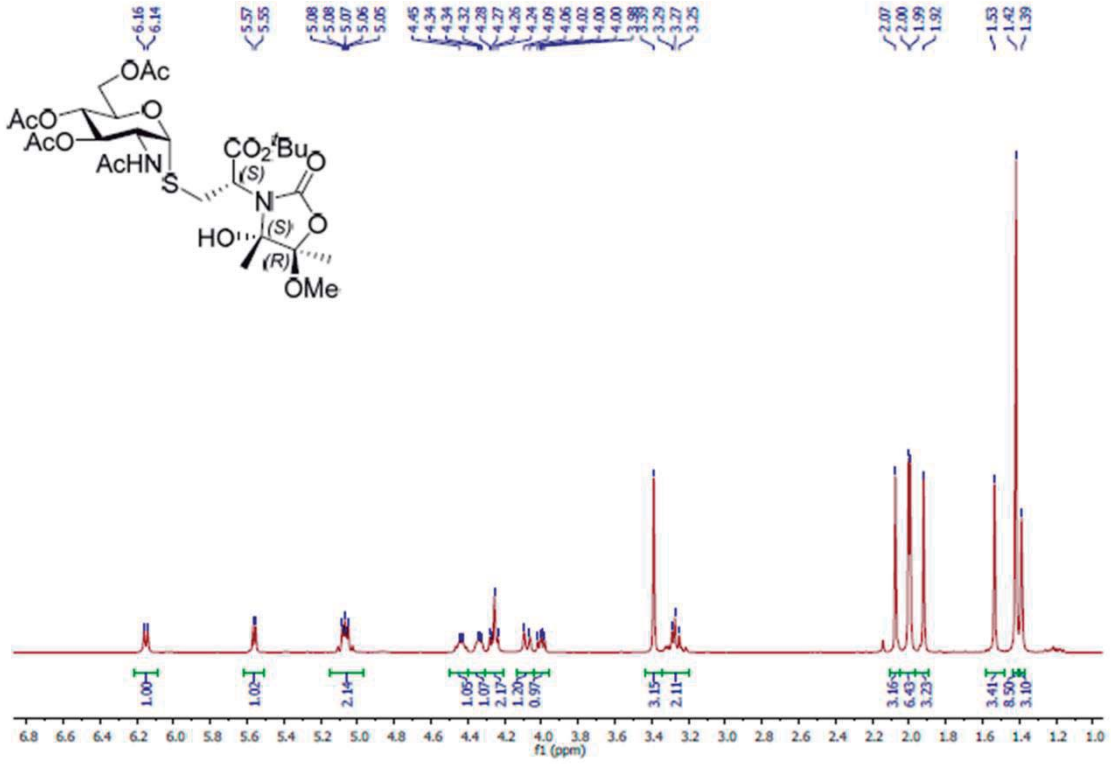



HSQC in $CDCl_3$

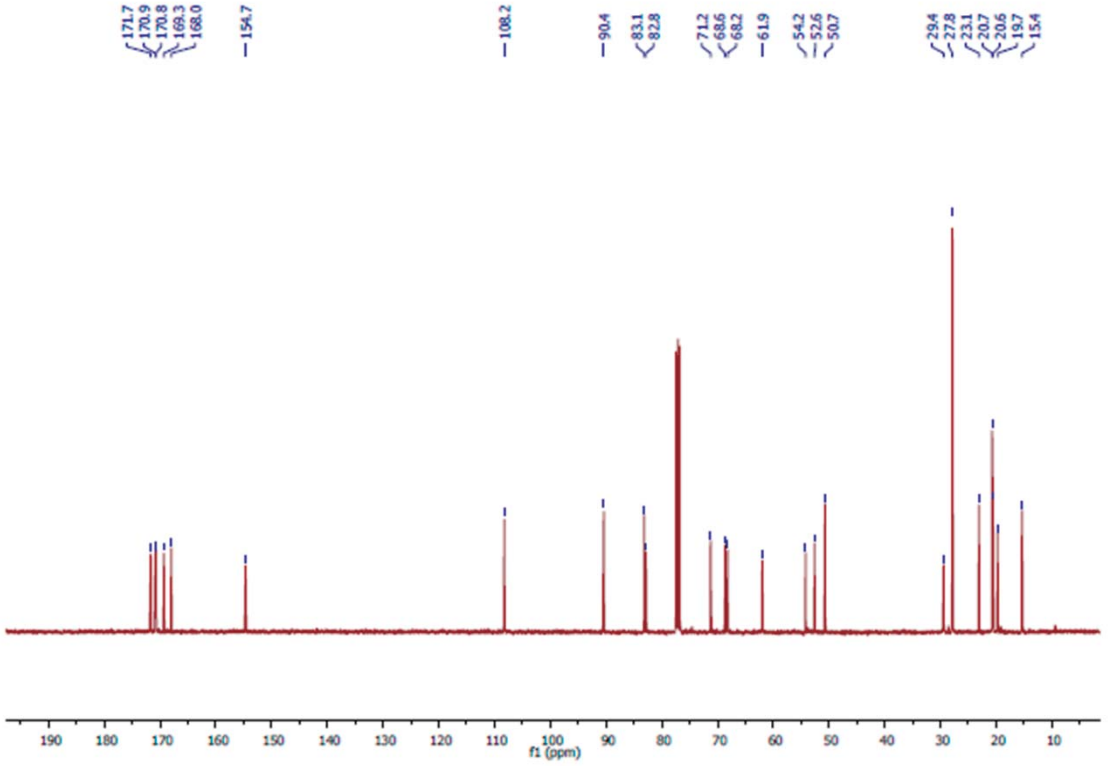


Compuesto **8'g** capítulo 4

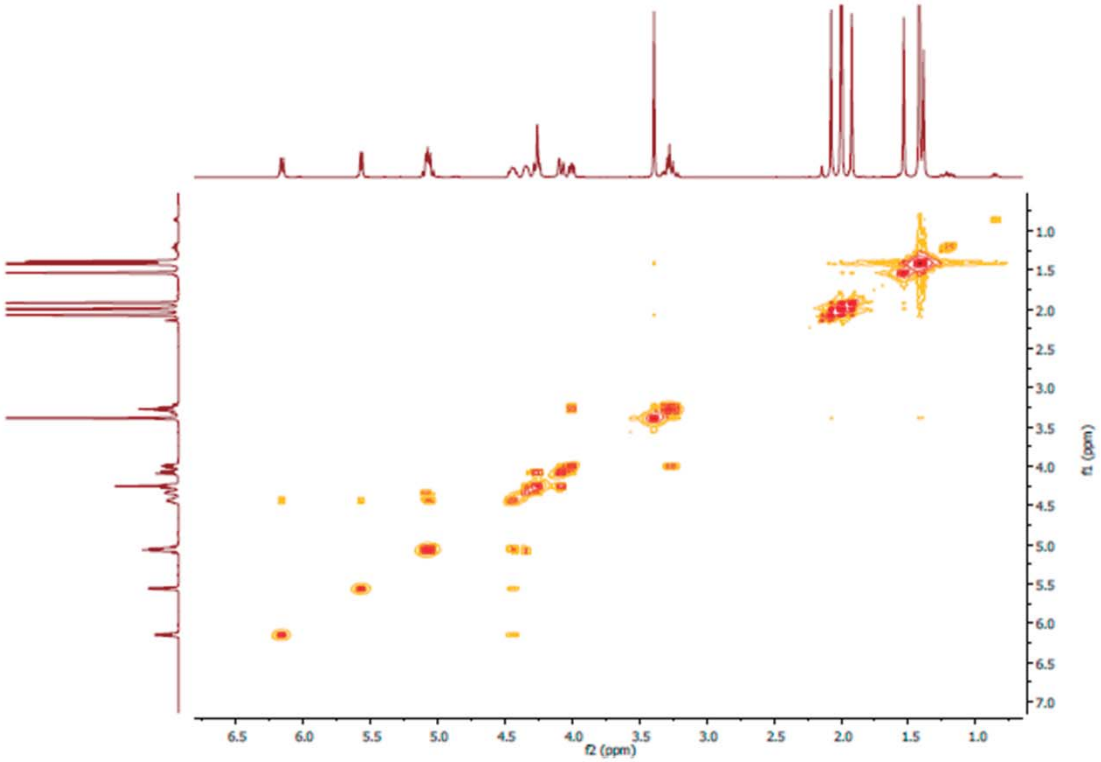
$^{1}$H NMR 400 MHz in $CDCl_3$

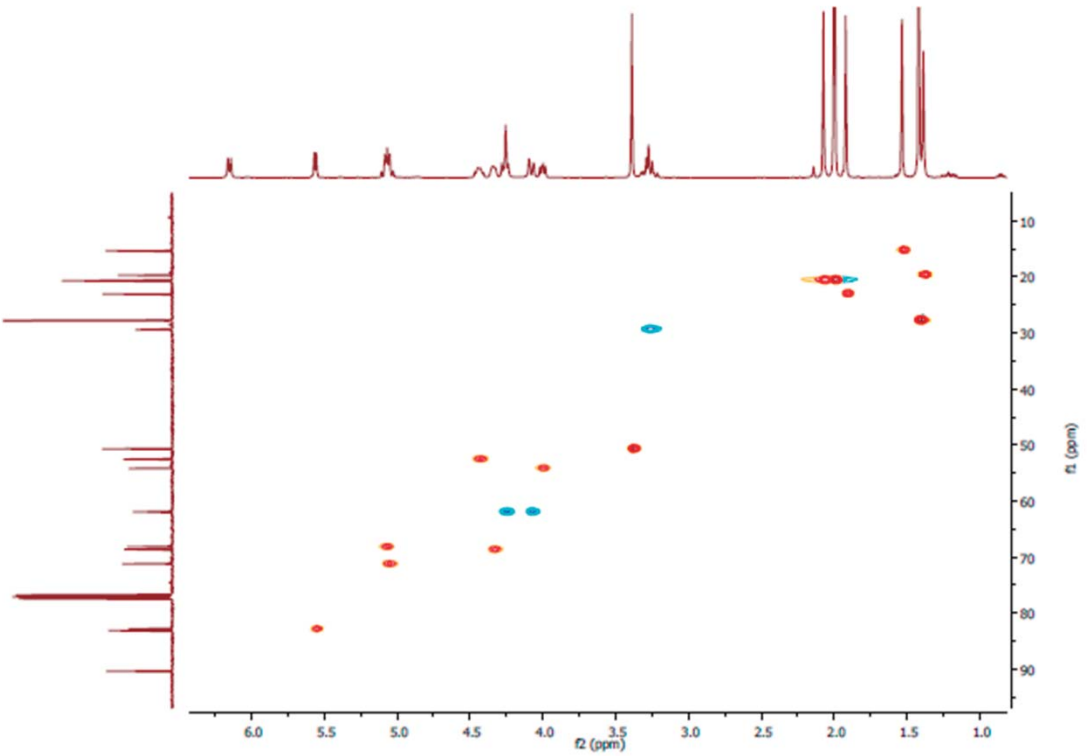



$^{13}$C NMR 100 MHz in $CDCl_3$



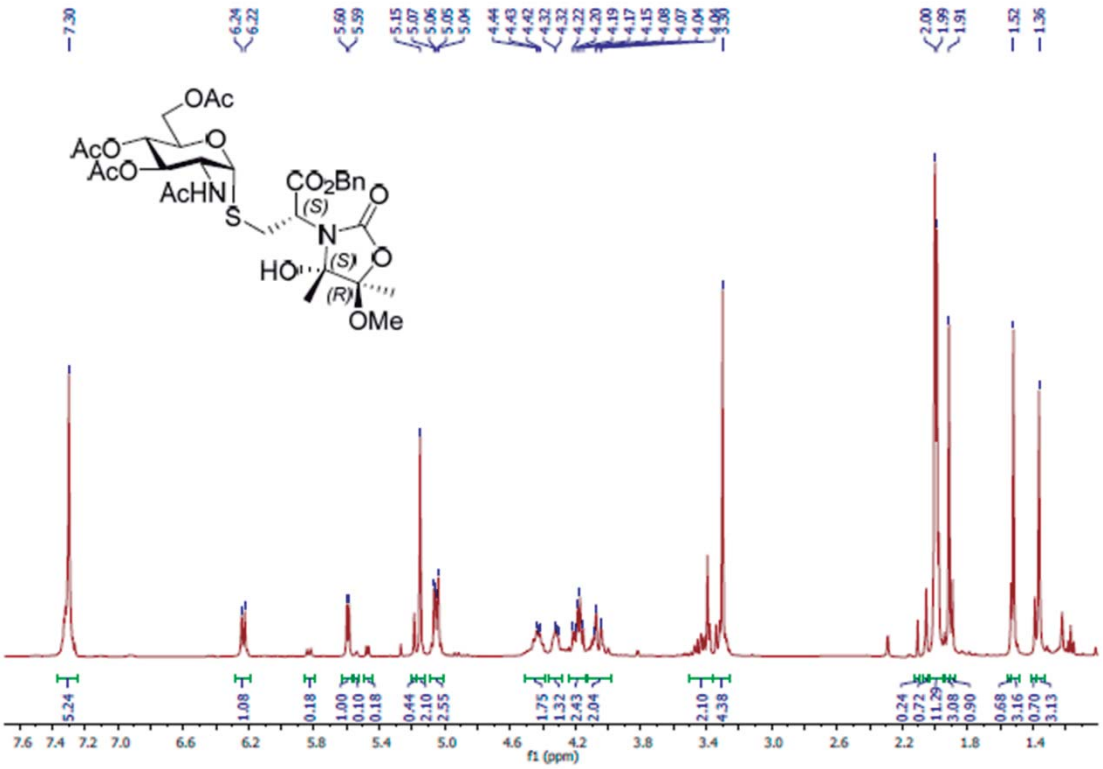
COSY in $CDCl_3$



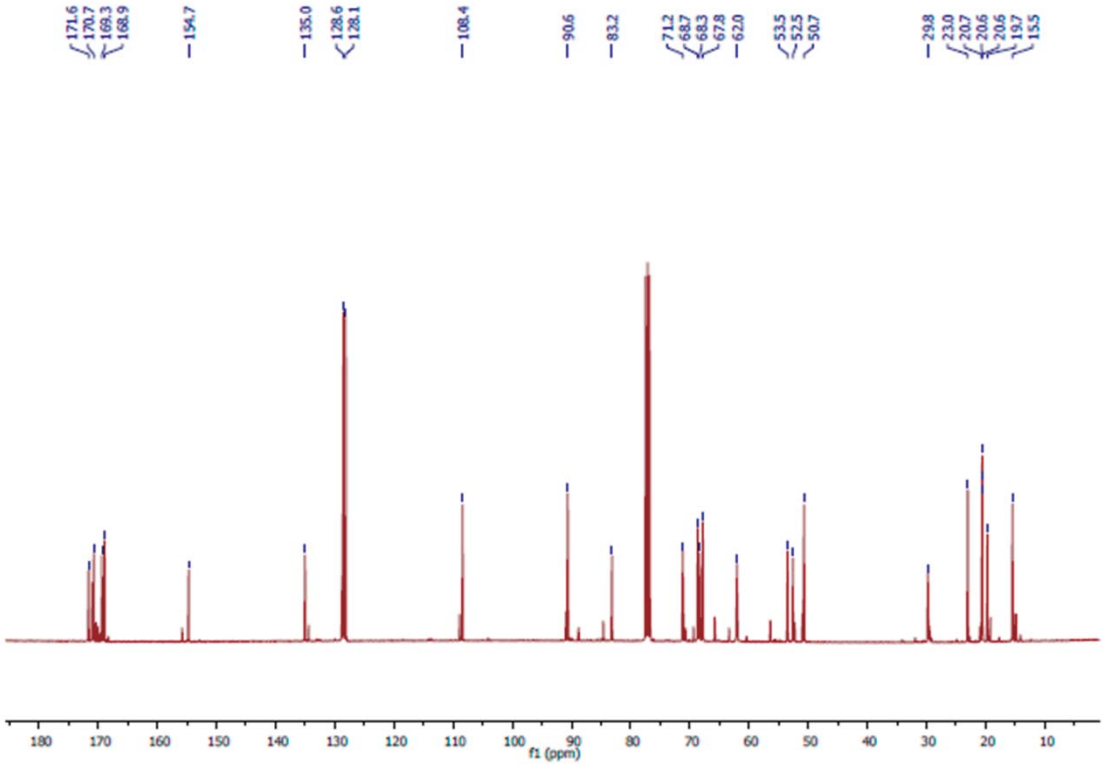



Compuesto **8''g** capítulo 4

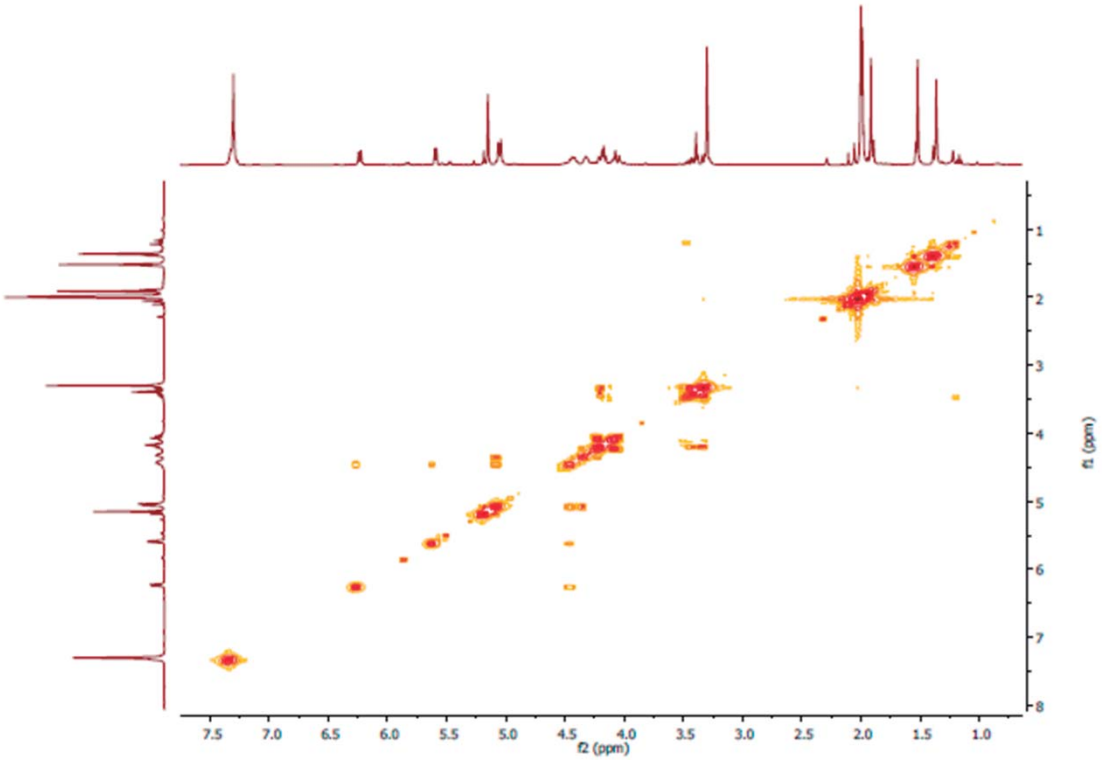
$^{1}$H NMR 400 MHz in $CDCl_3$



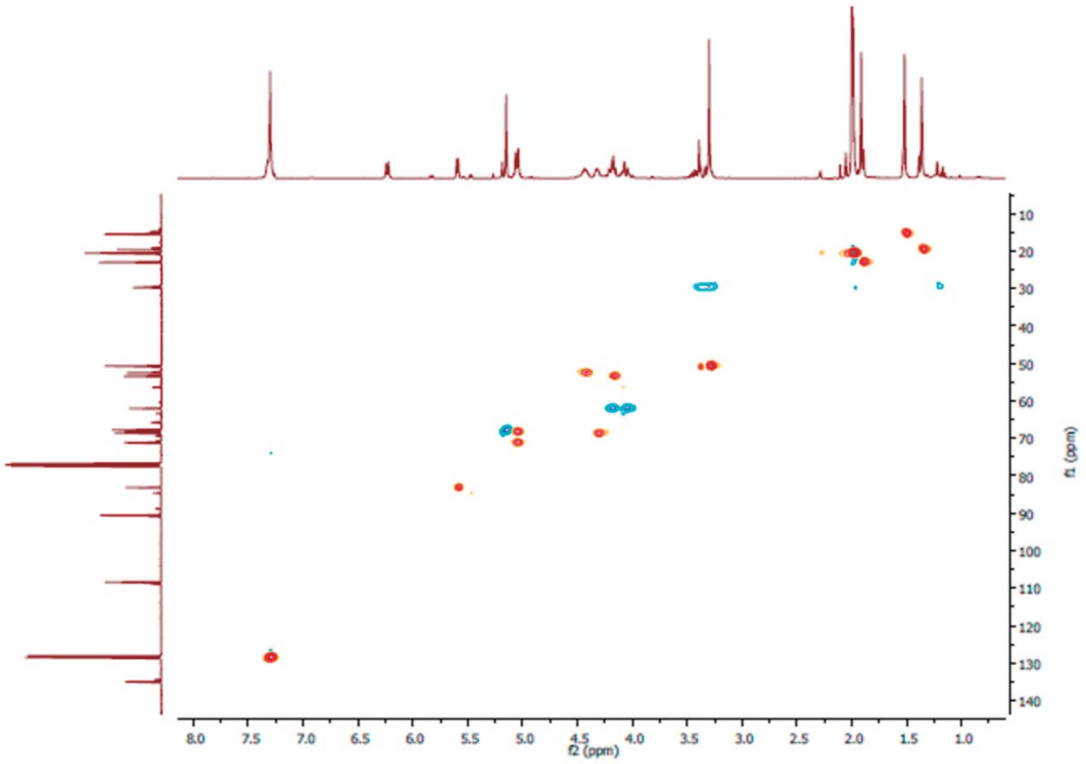

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

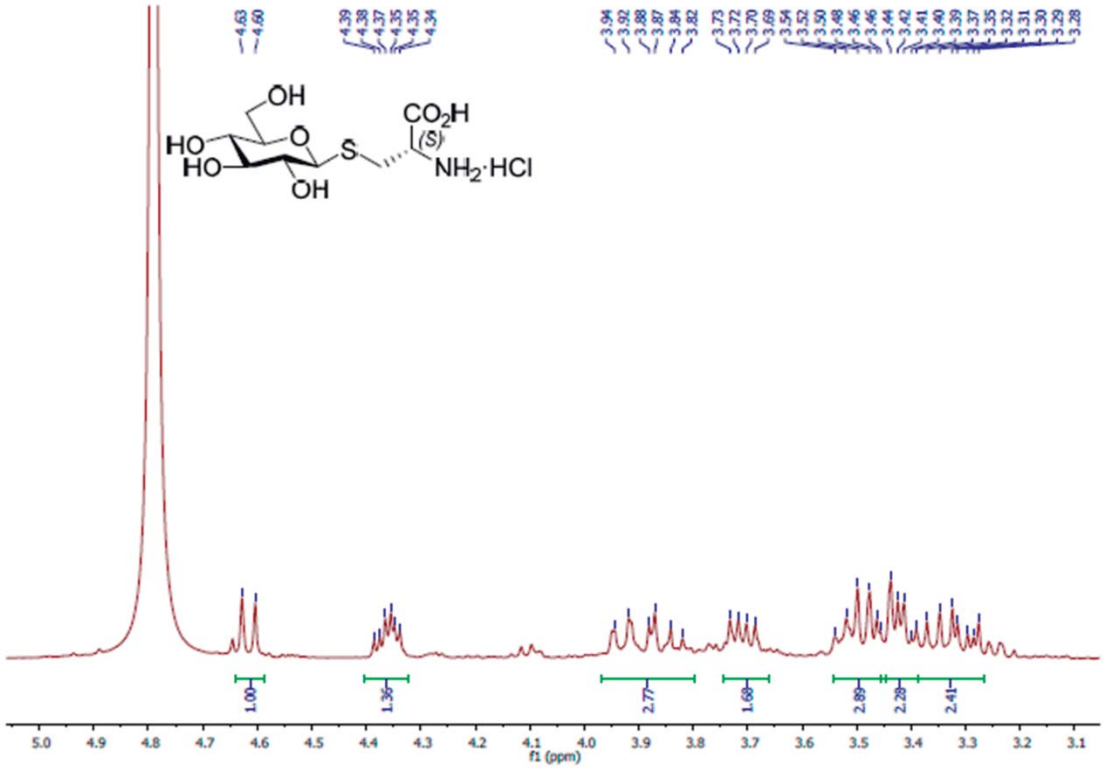



HSQC in $CDCl_3$

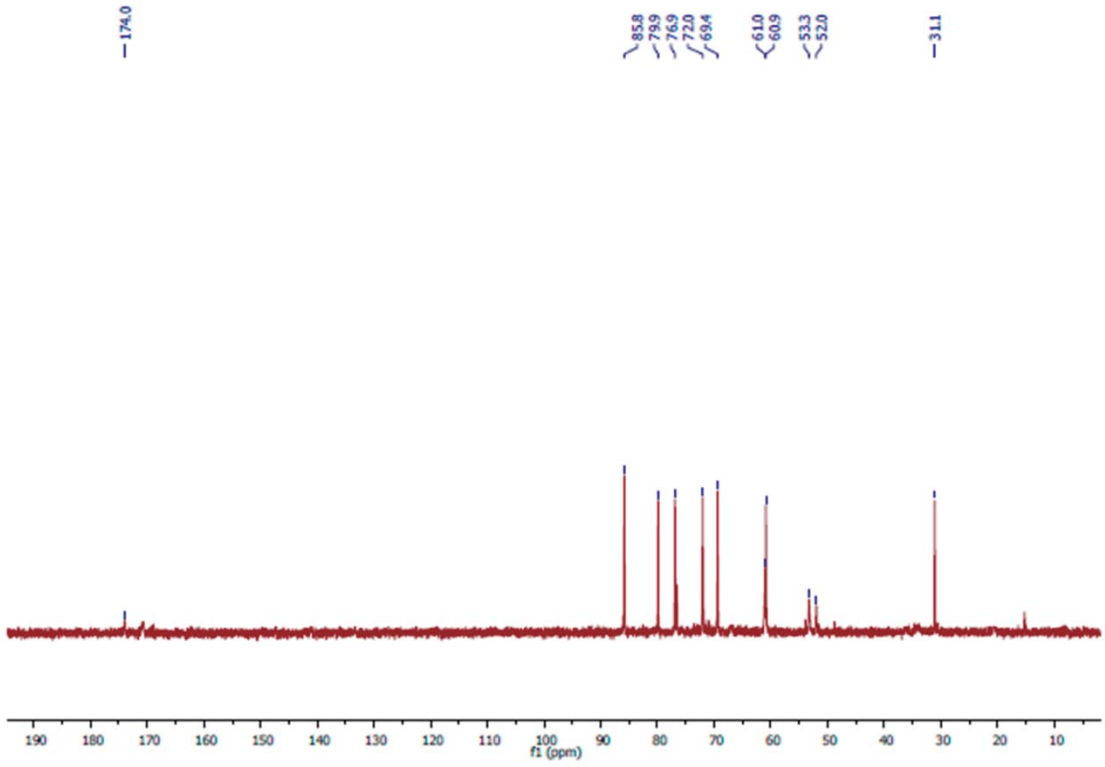


Compuesto **9a** capítulo 4

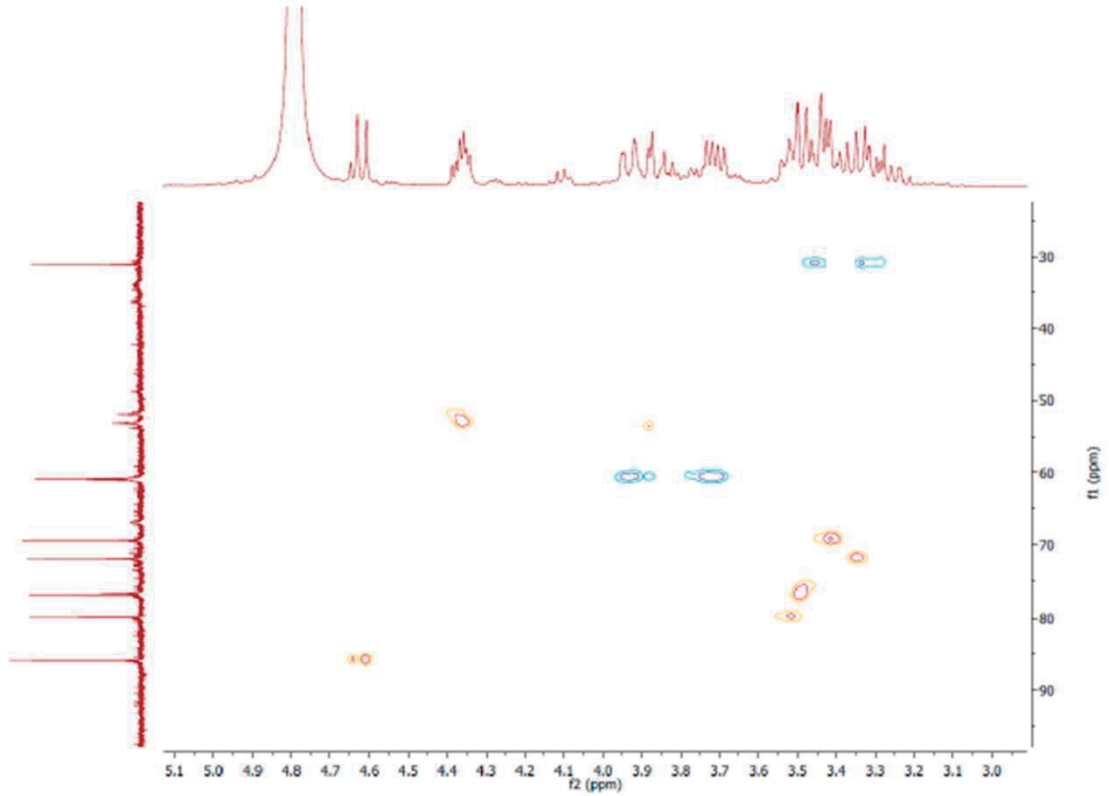
$^{1}$H NMR 400 MHz in $D_2O$



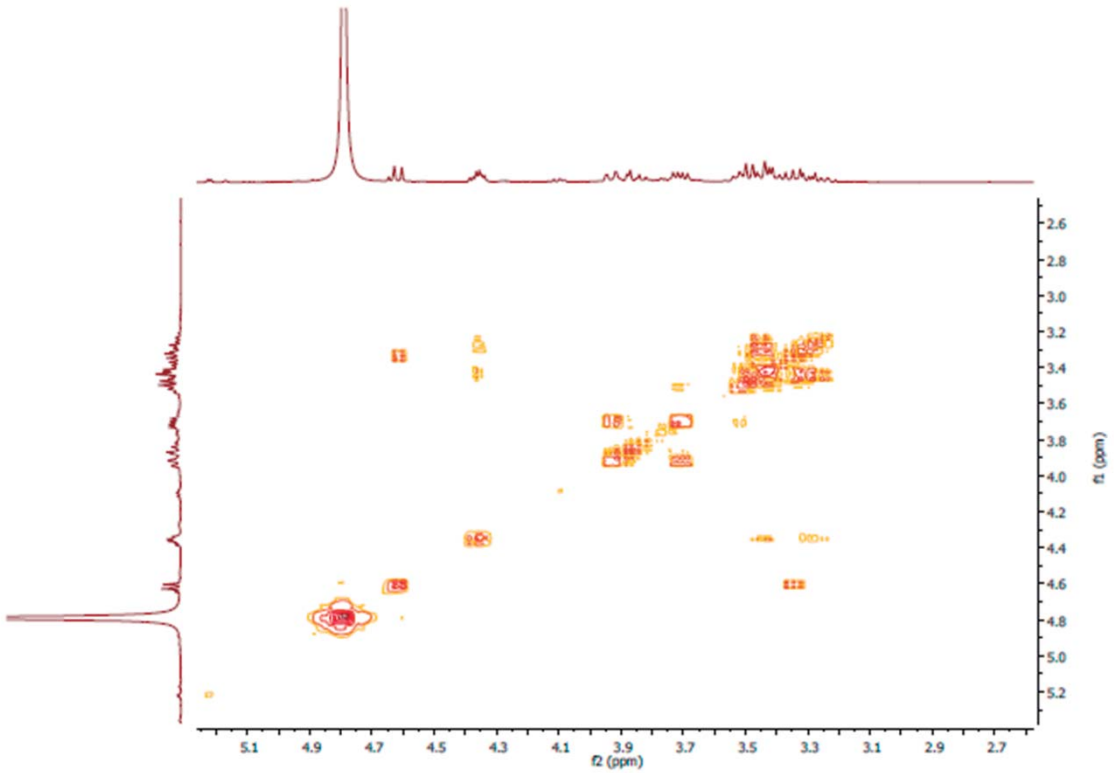

$^{13}$C NMR 100 MHz in $D_2O$



HSQC in $D_2O$

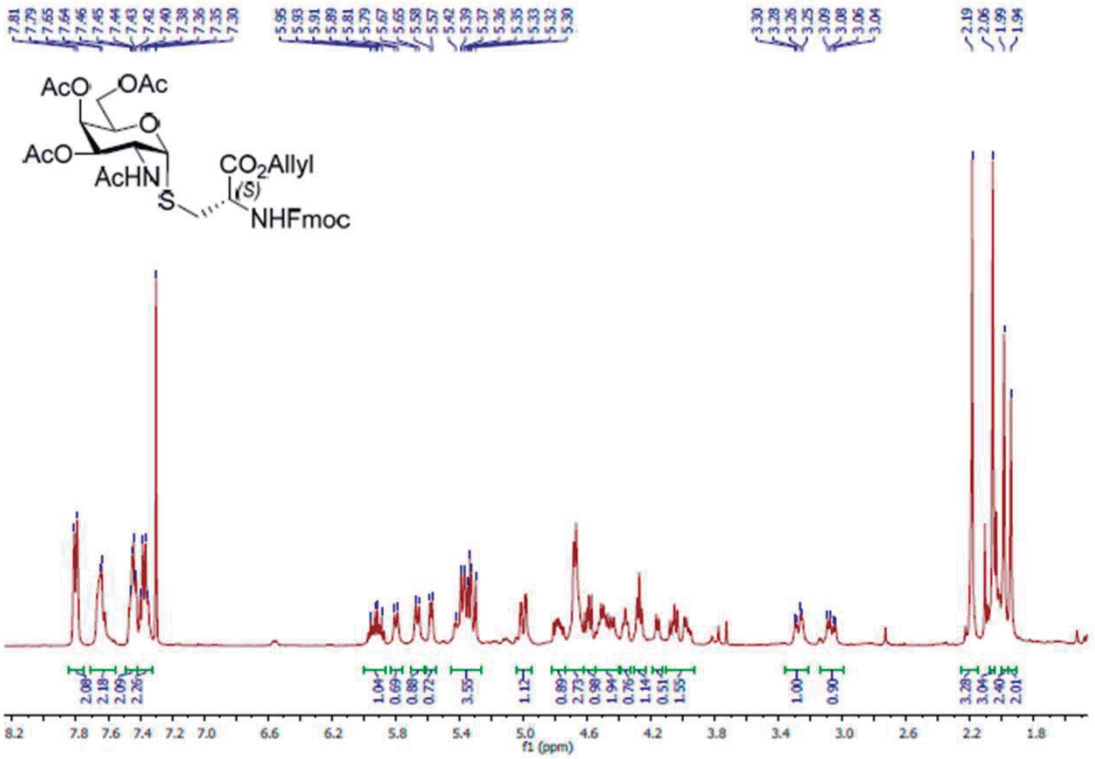



COSY in $D_2O$

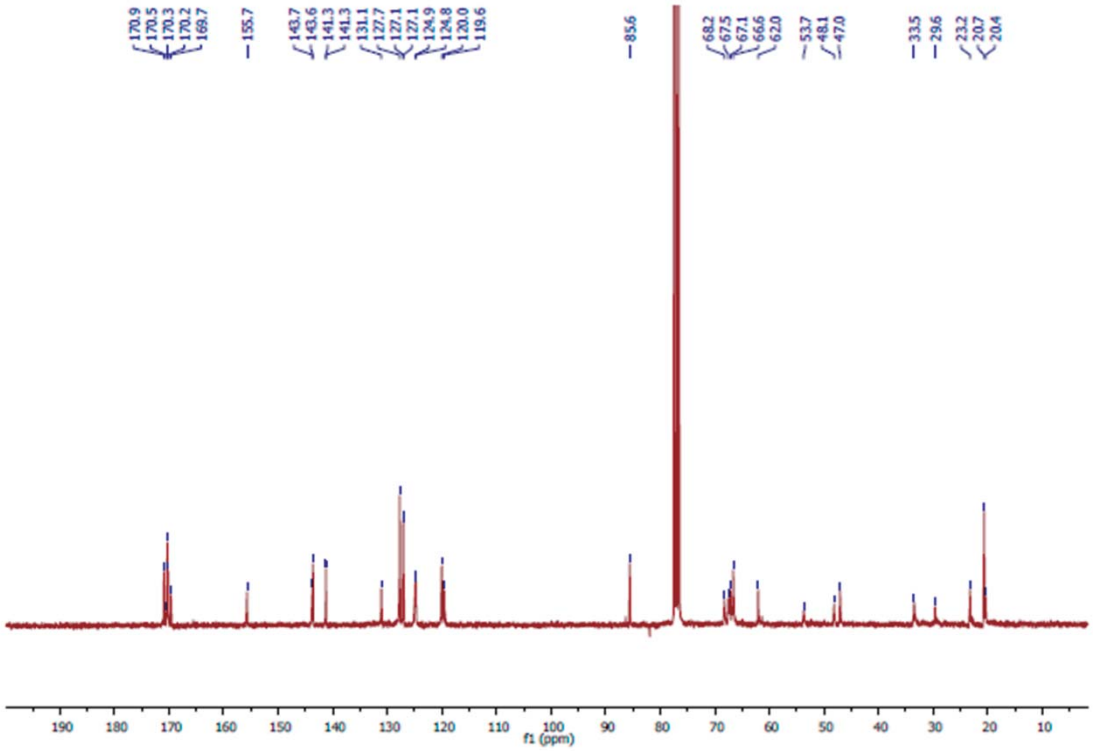


Compuesto **10f** capítulo 4

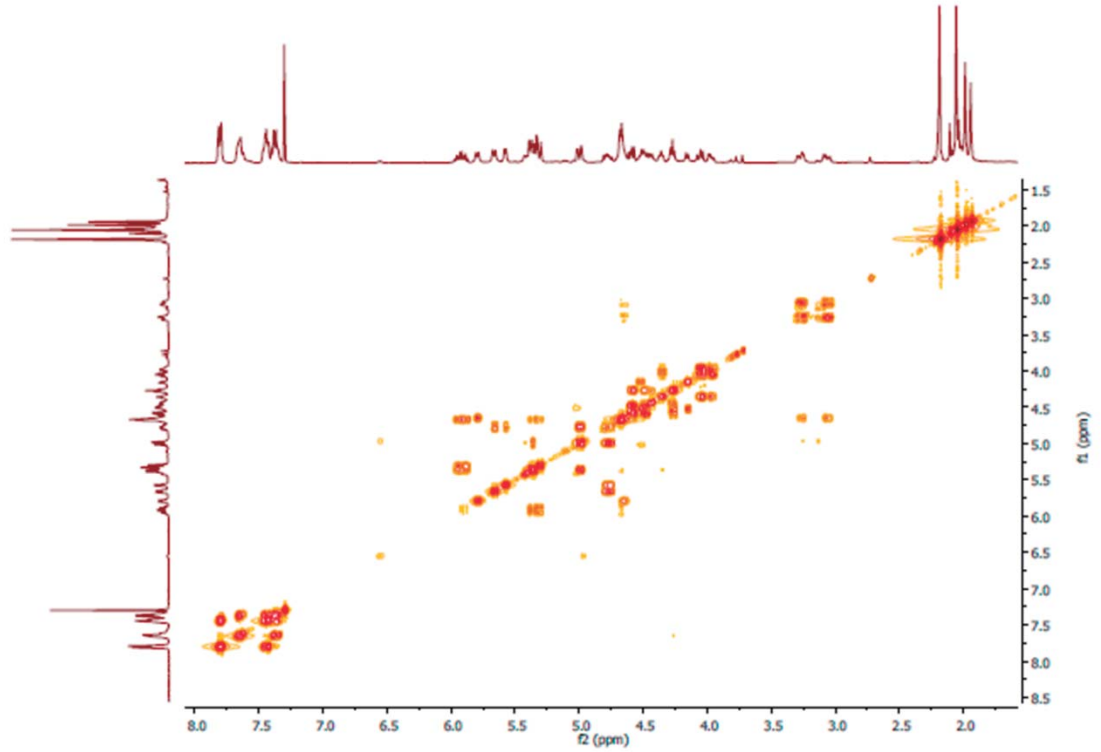
$^{1}$H NMR 400 MHz in $CDCl_3$



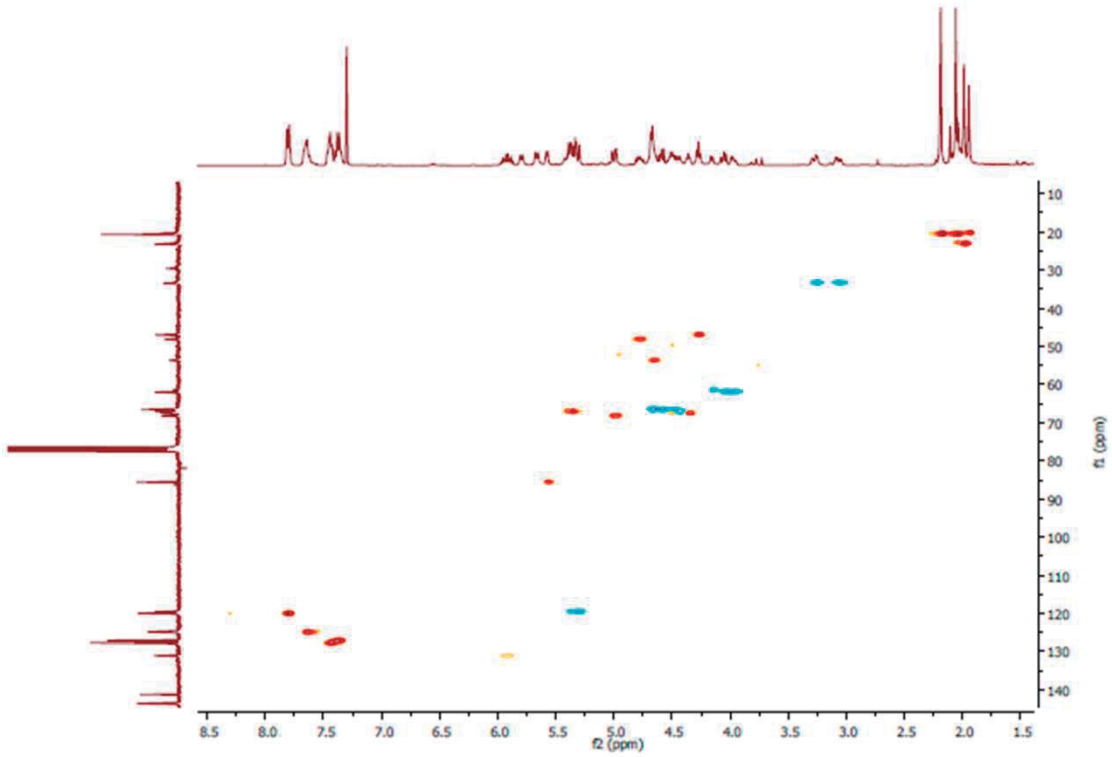

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

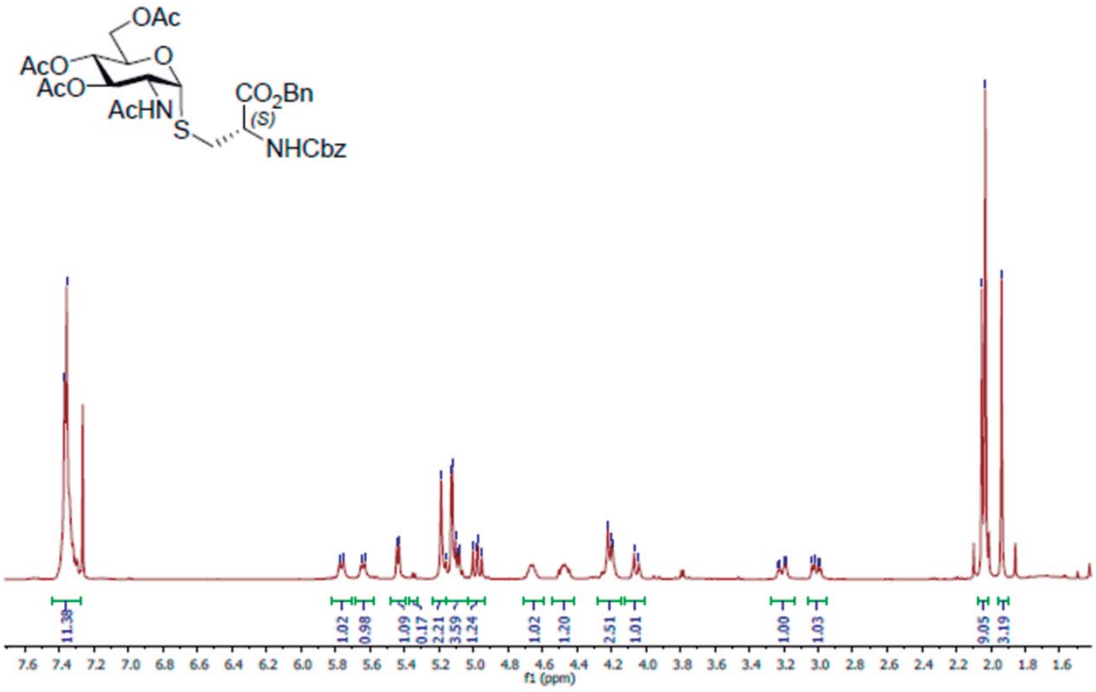



HSQC in $CDCl_3$

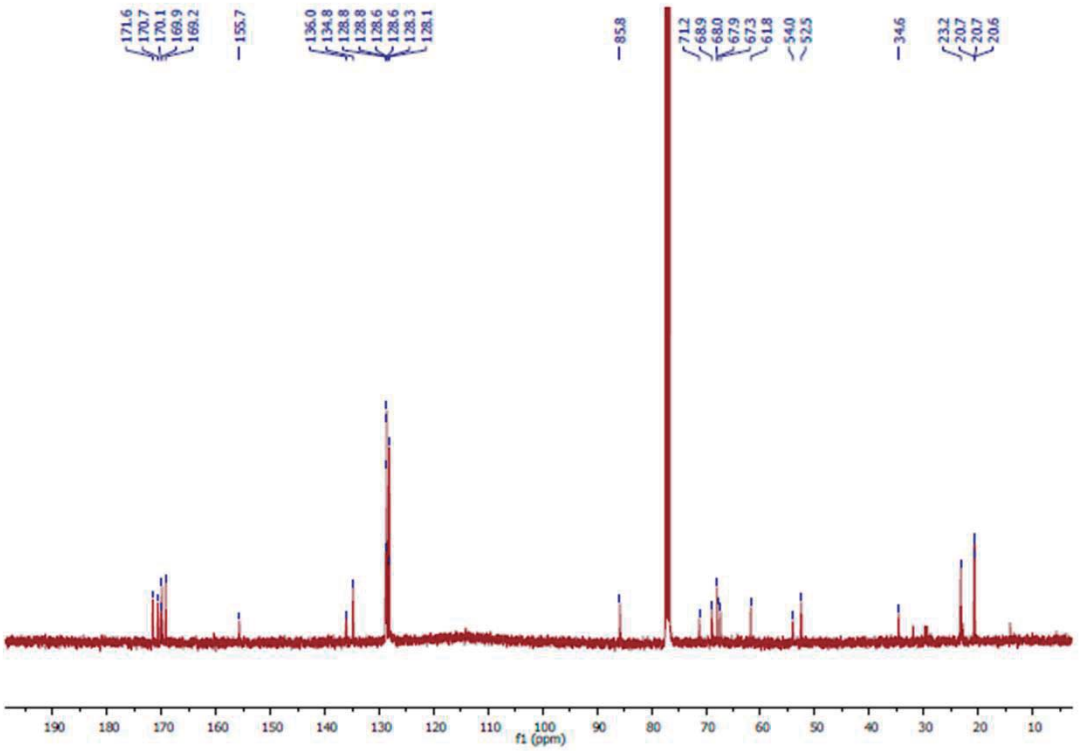


Compuesto **10g** capítulo 4

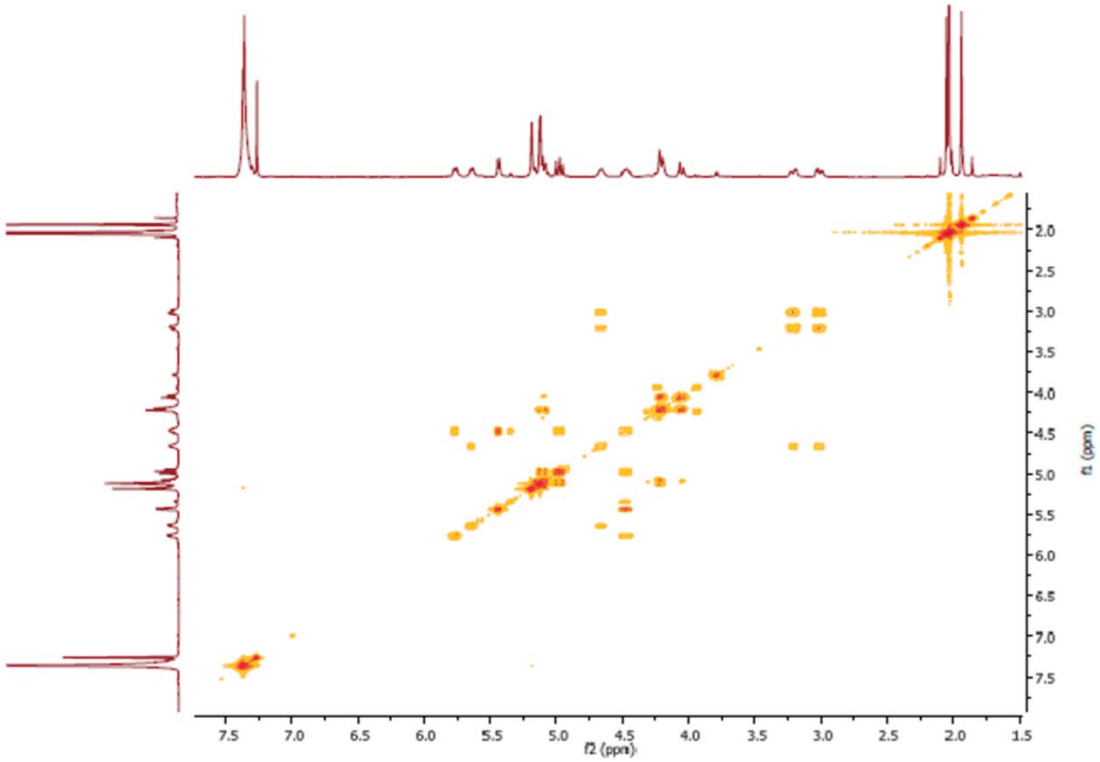
$^{1}$H NMR 400 MHz in $CDCl_3$

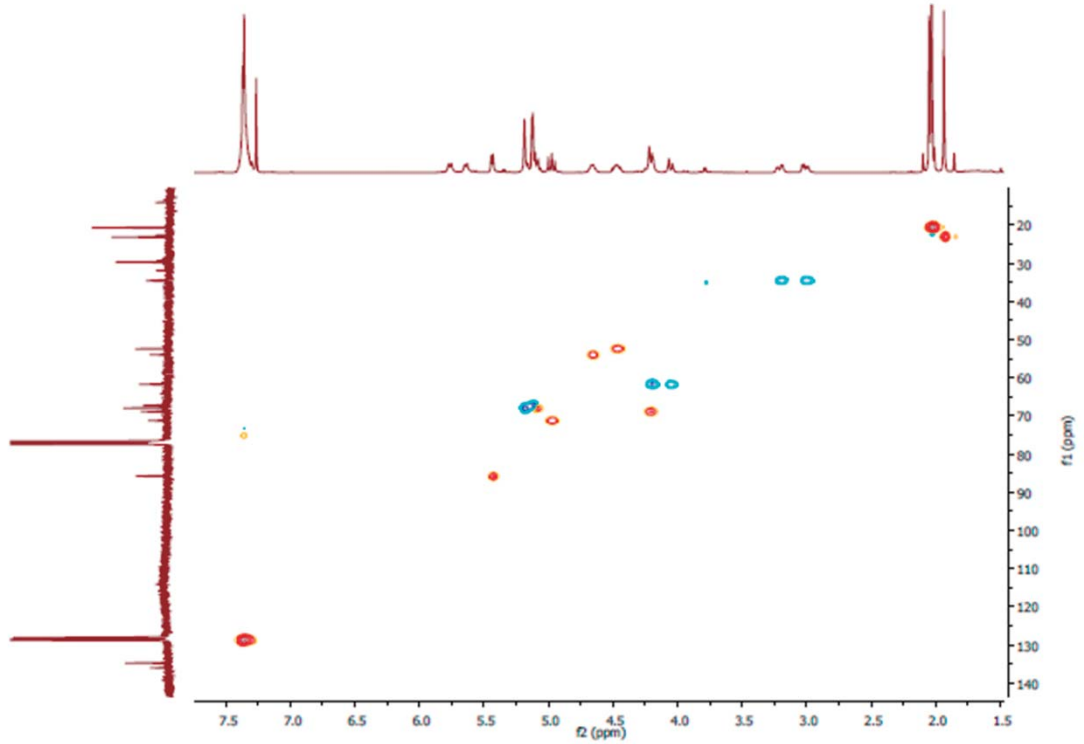



$^{13}$C NMR 100 MHz in $CDCl_3$



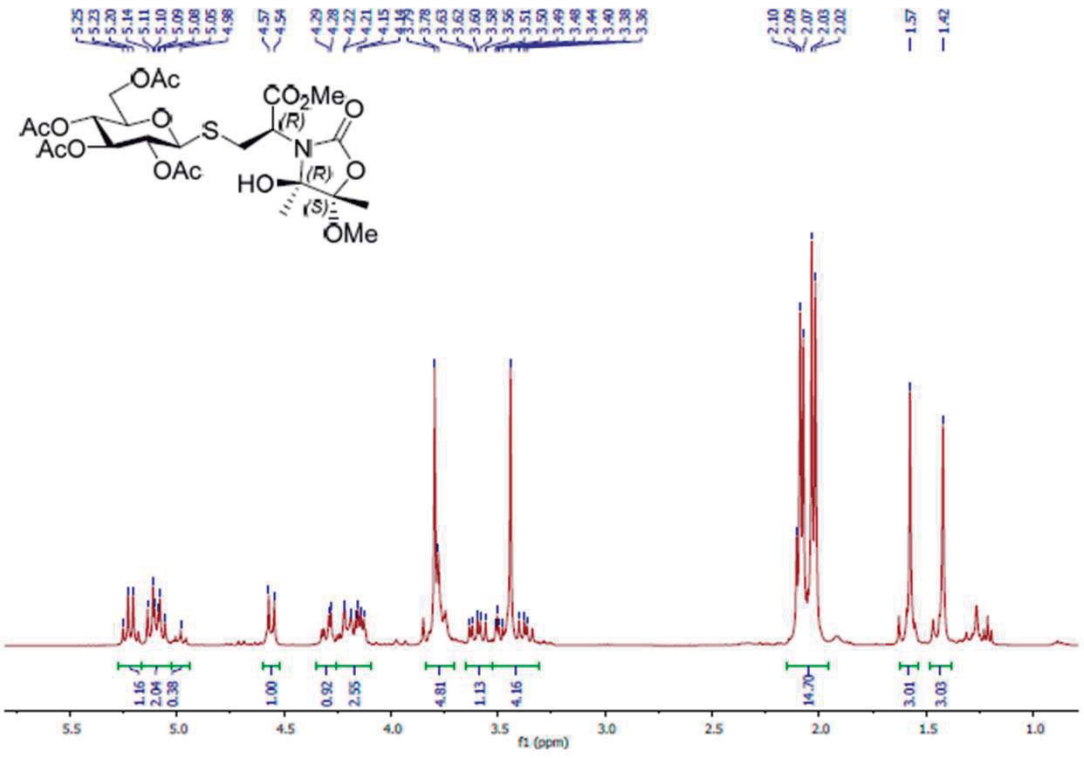
COSY in $CDCl_3$



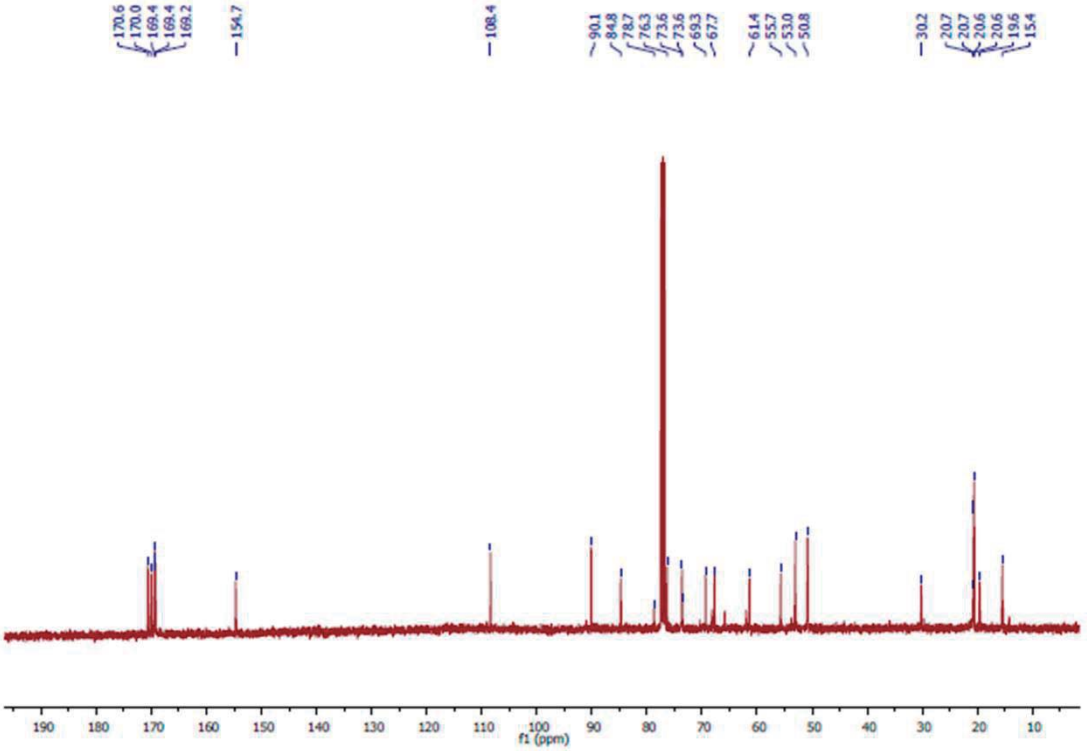



Compuesto **11a** capítulo 4

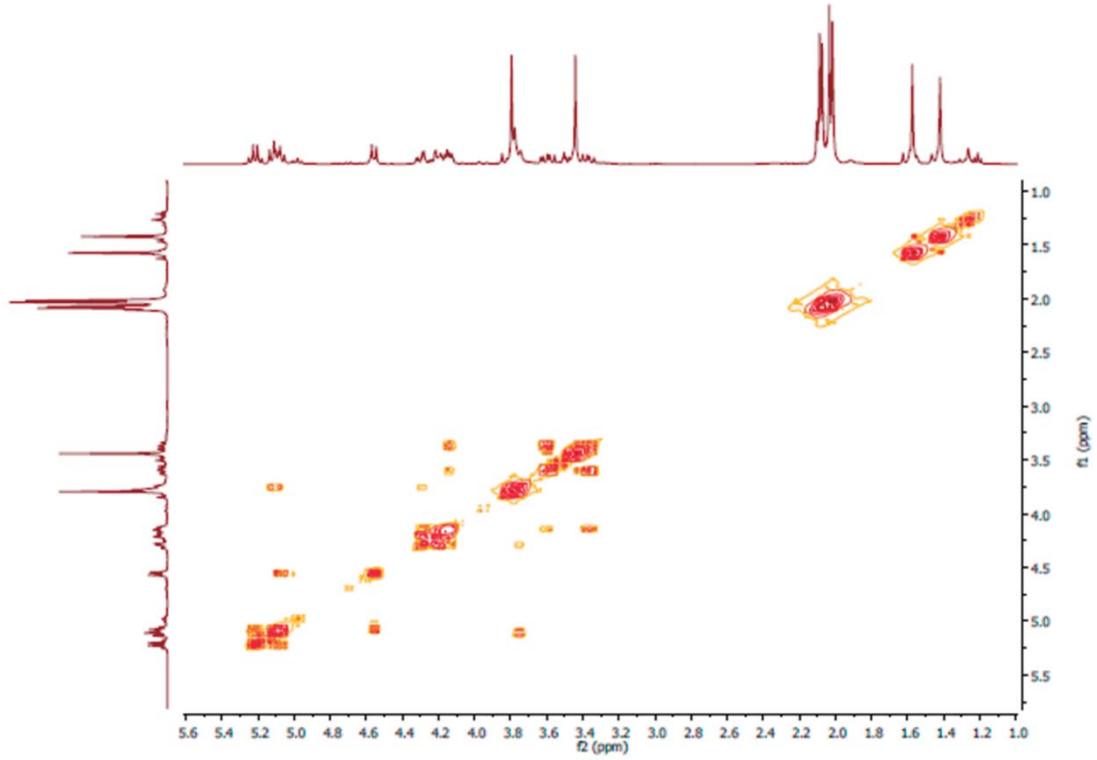
$^{1}$H NMR 400 MHz in $CDCl_3$



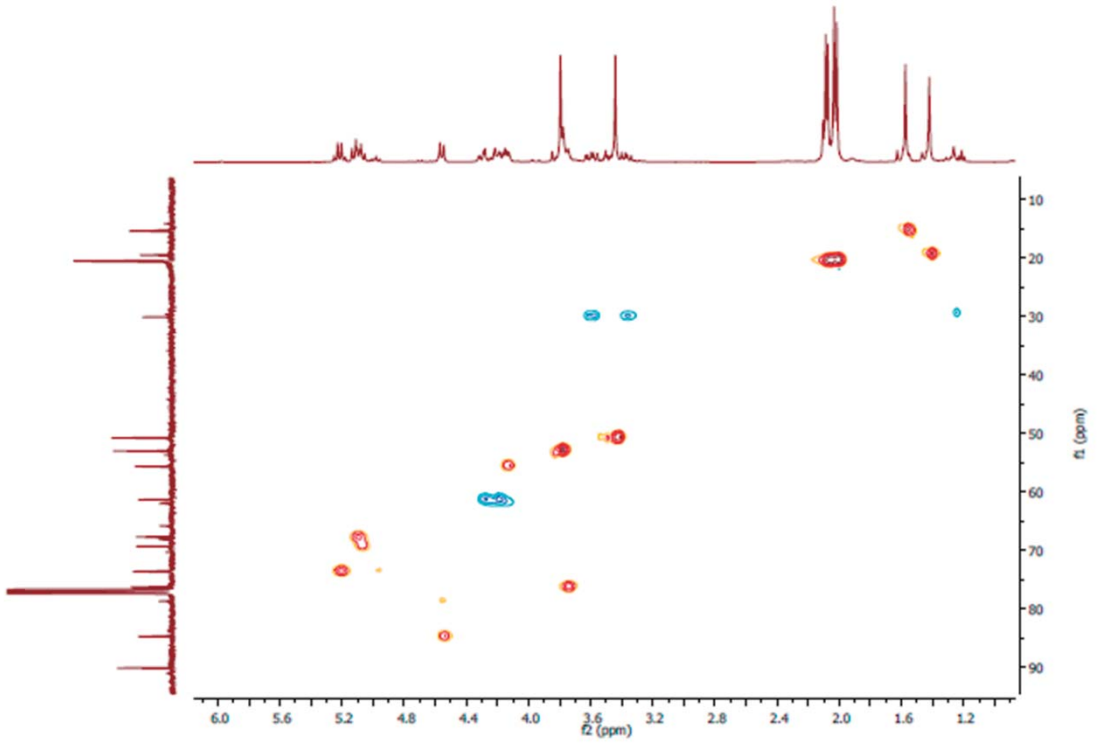

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

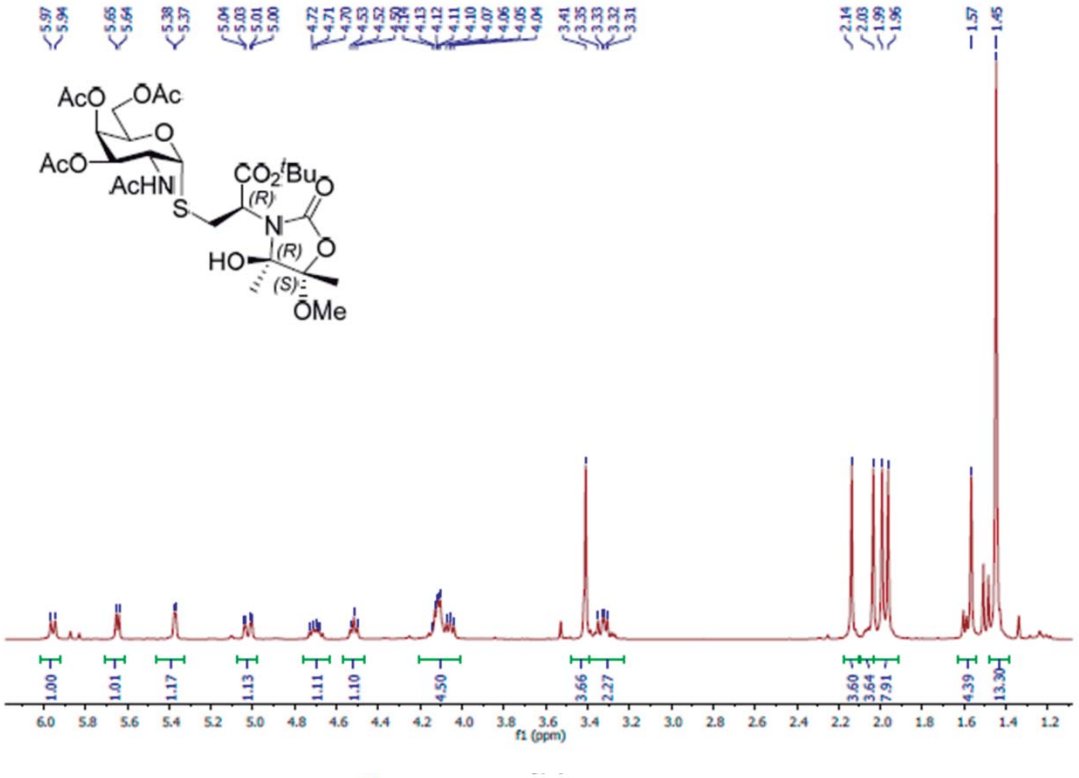



HSQC in $CDCl_3$

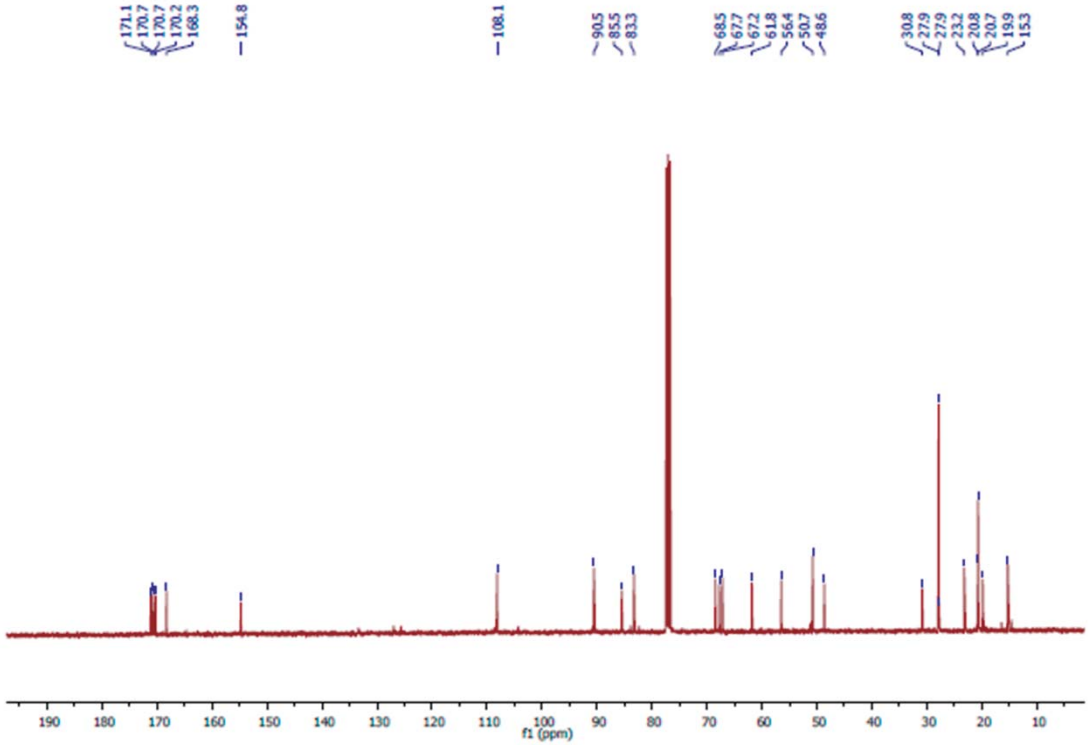


Compuesto **11'f** capítulo 4

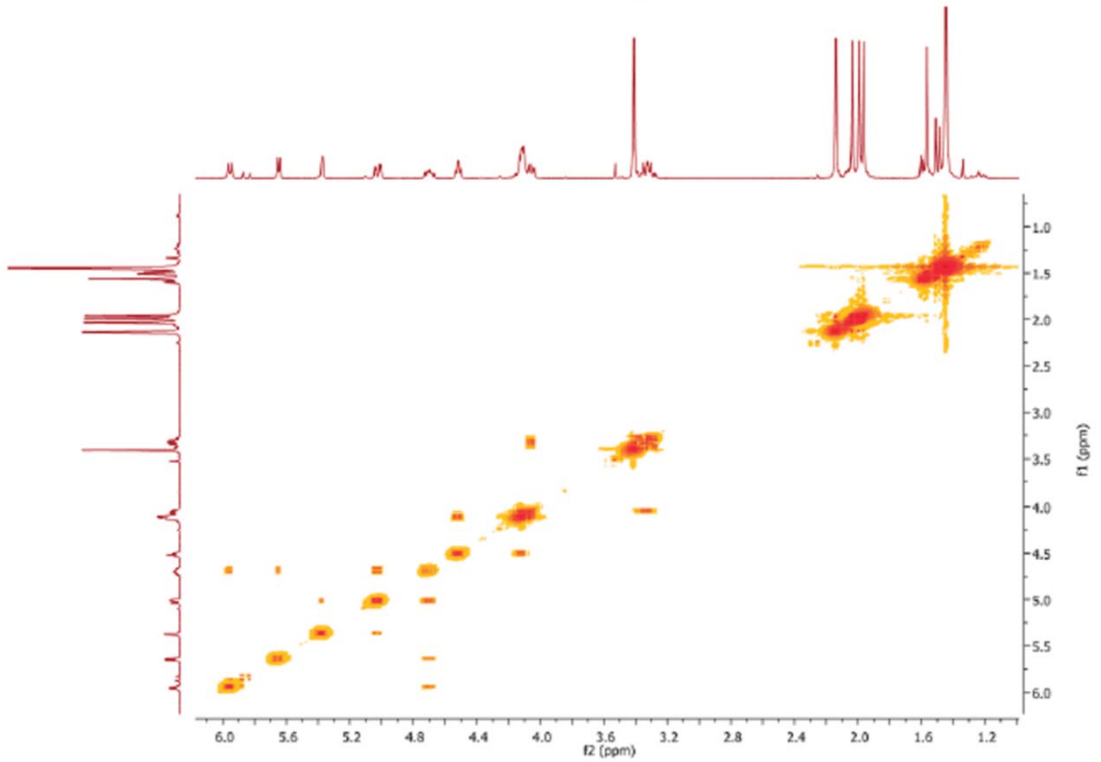
$^{1}$H NMR 400 MHz in $CDCl_3$

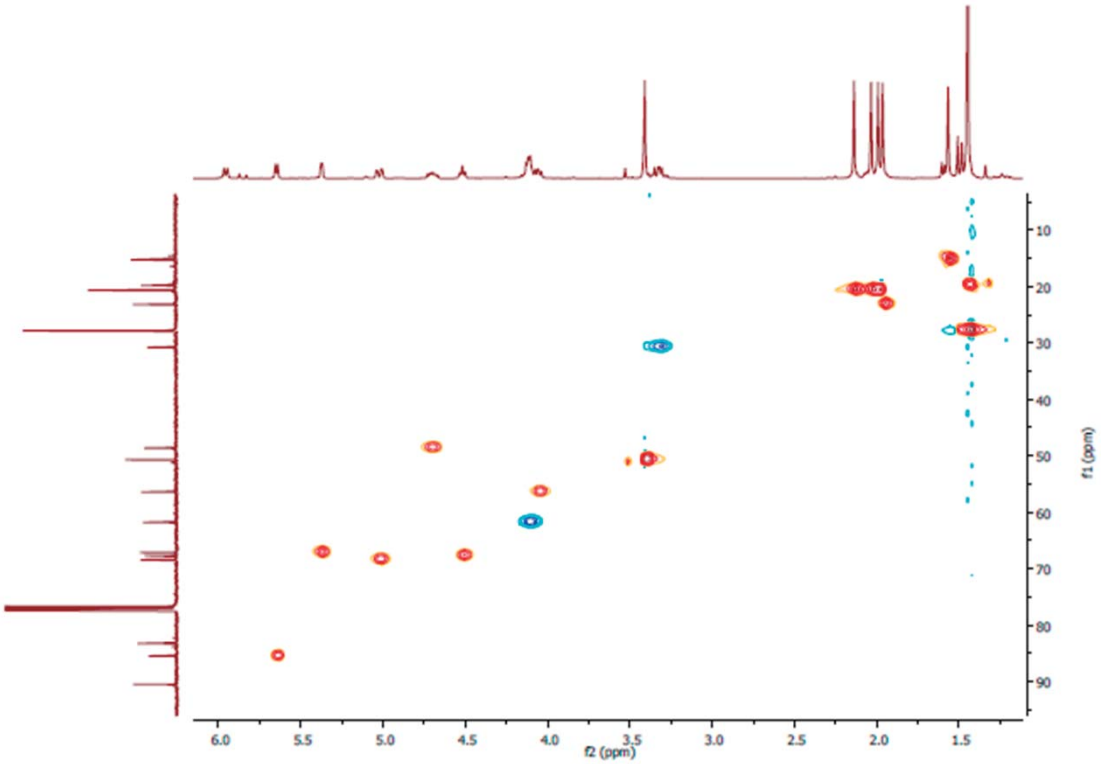



$^{13}$C NMR 100 MHz in $CDCl_3$



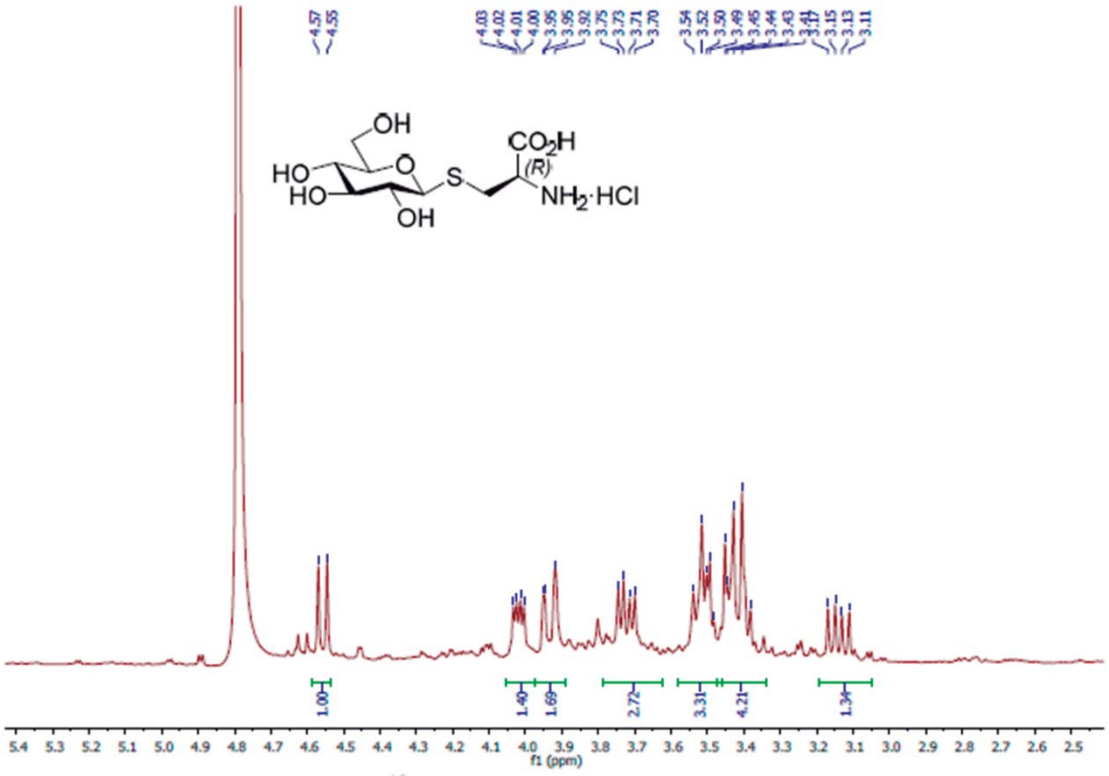
COSY in $CDCl_3$



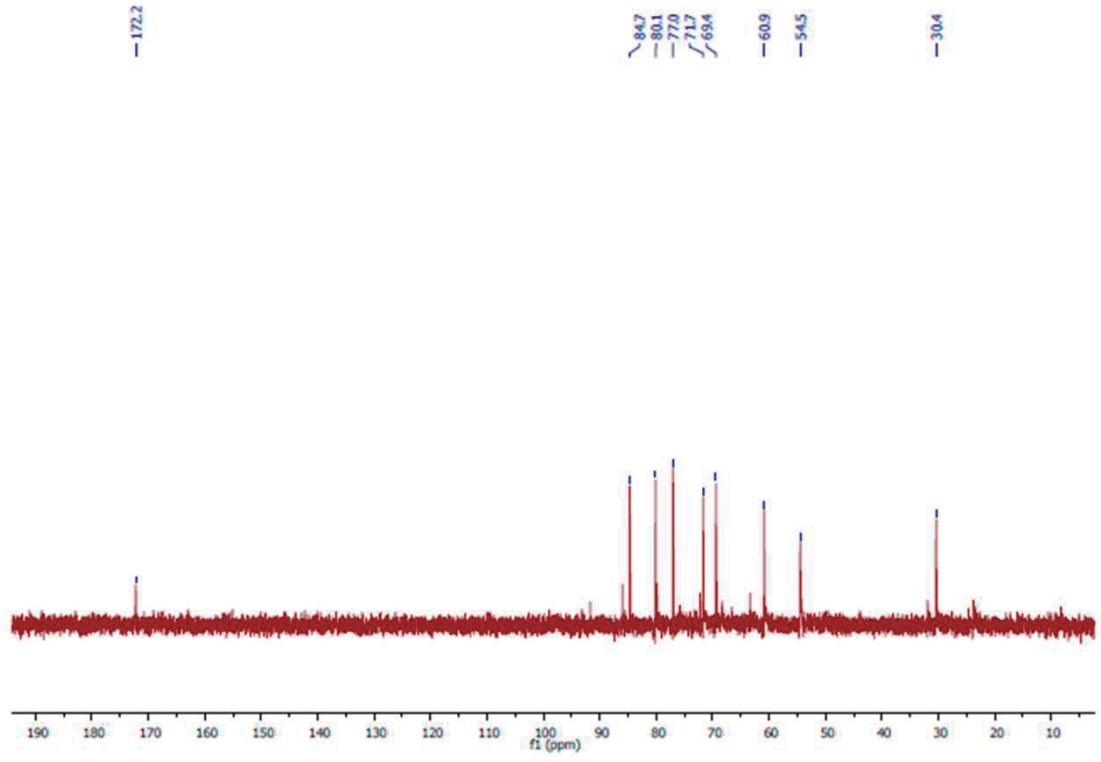



Compuesto **12a** capítulo 4

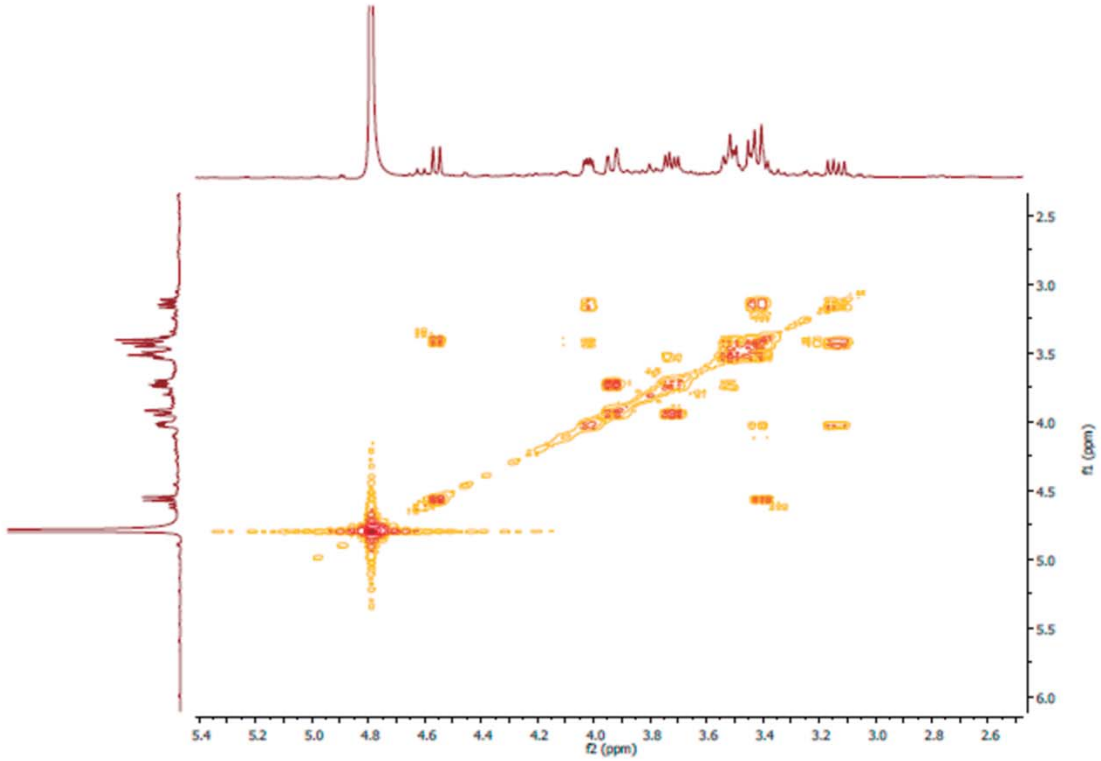
$^{1}$H NMR 400 MHz in $D_2O$



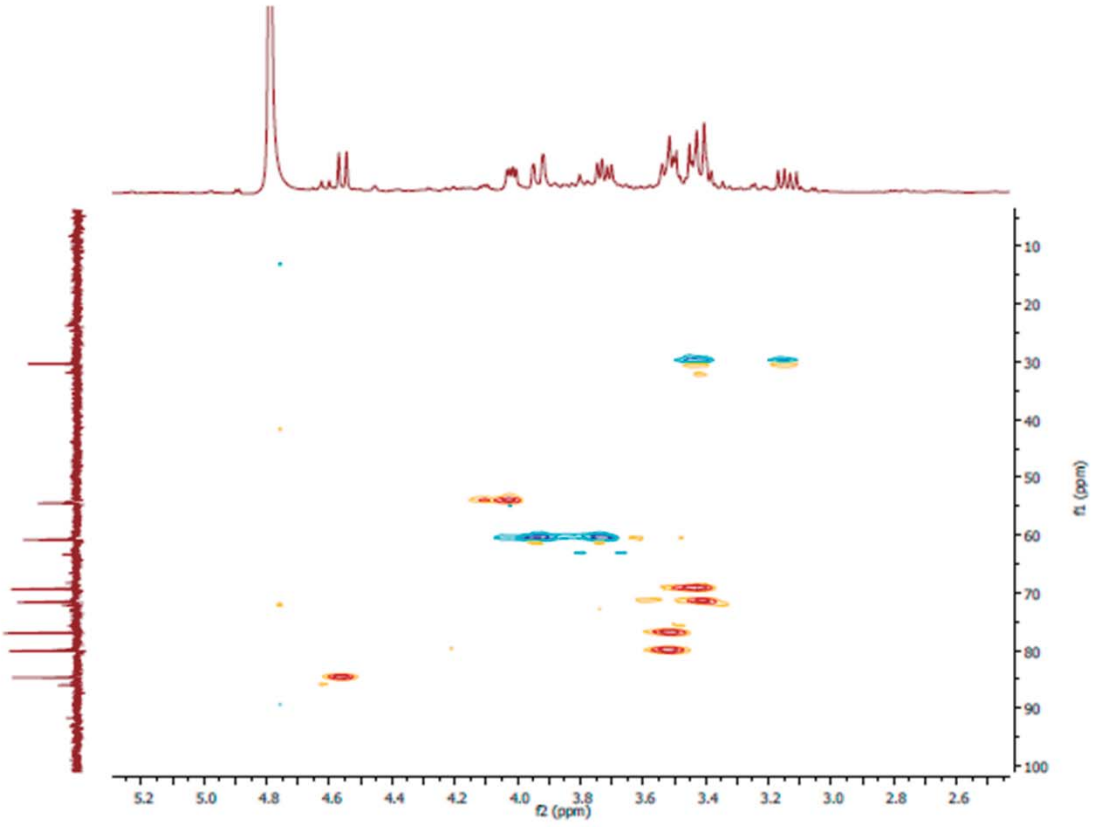

$^{13}$C NMR 100 MHz in $D_2O$



COSY in $D_2O$

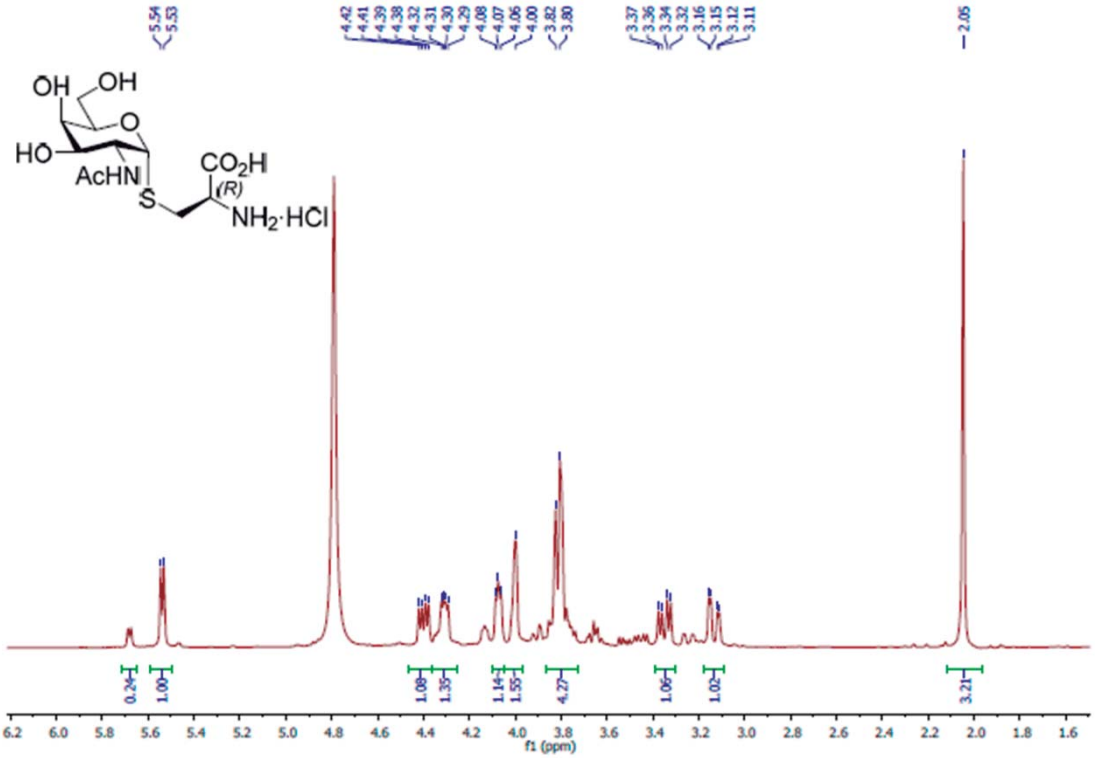



HSQC in $D_2O$

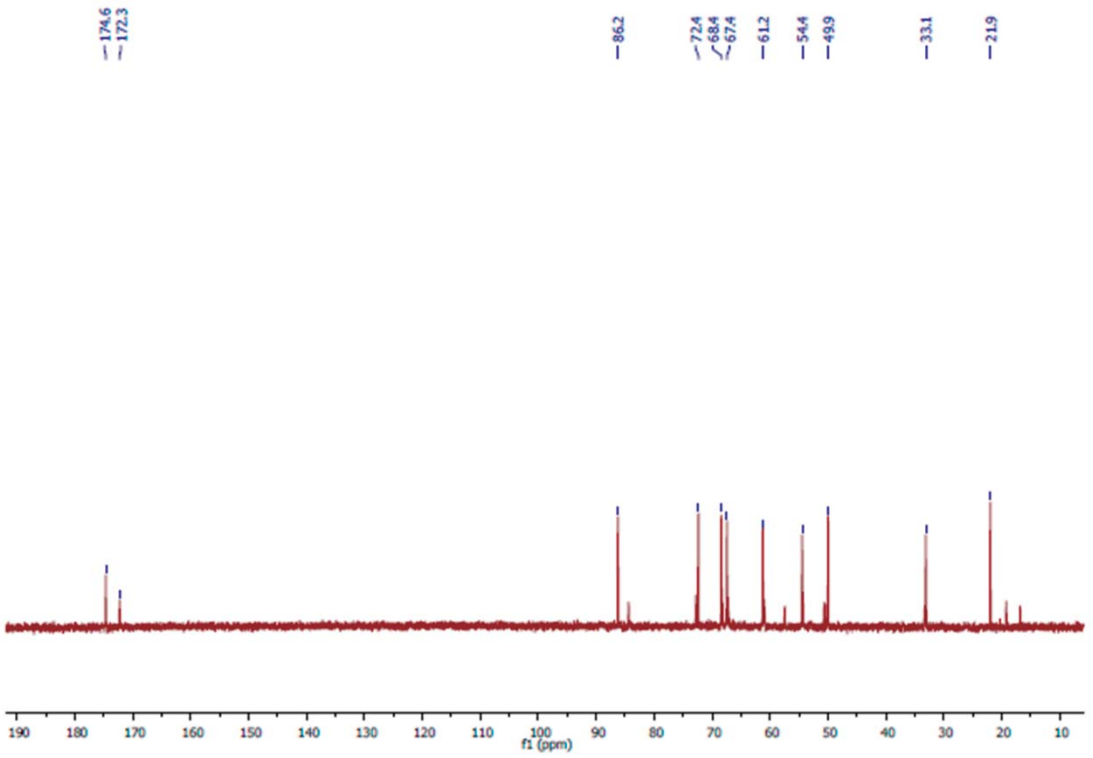


Compuesto **12f** capítulo 4

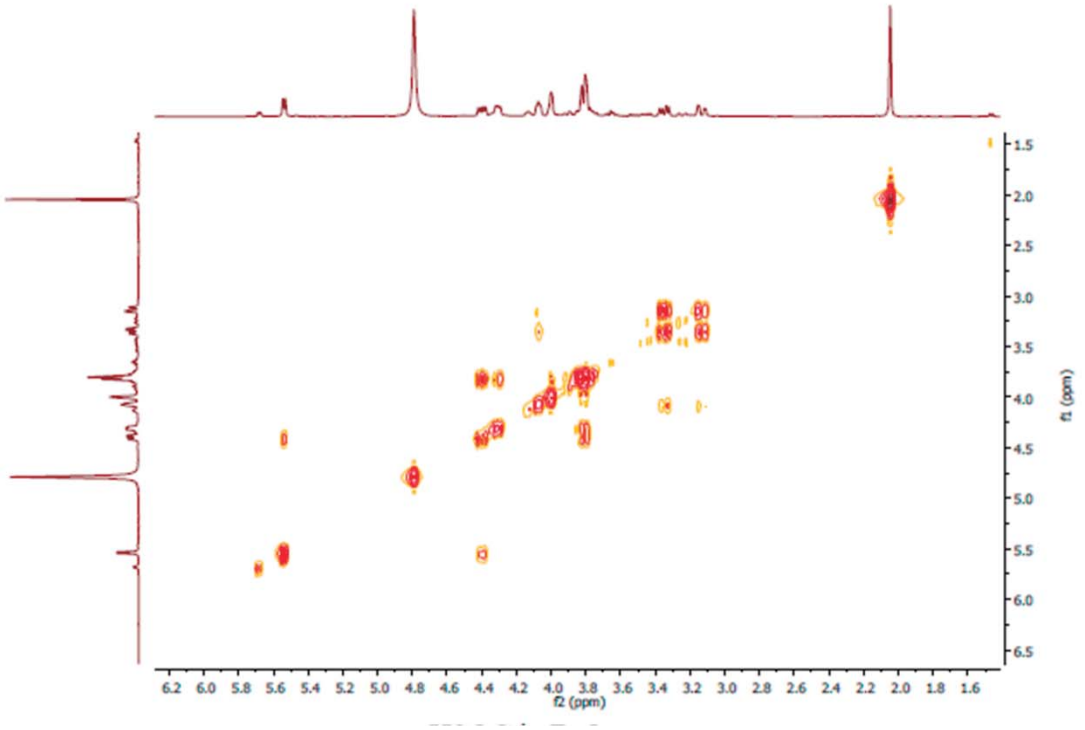
$^{1}$H NMR 400 MHz in $D_2O$



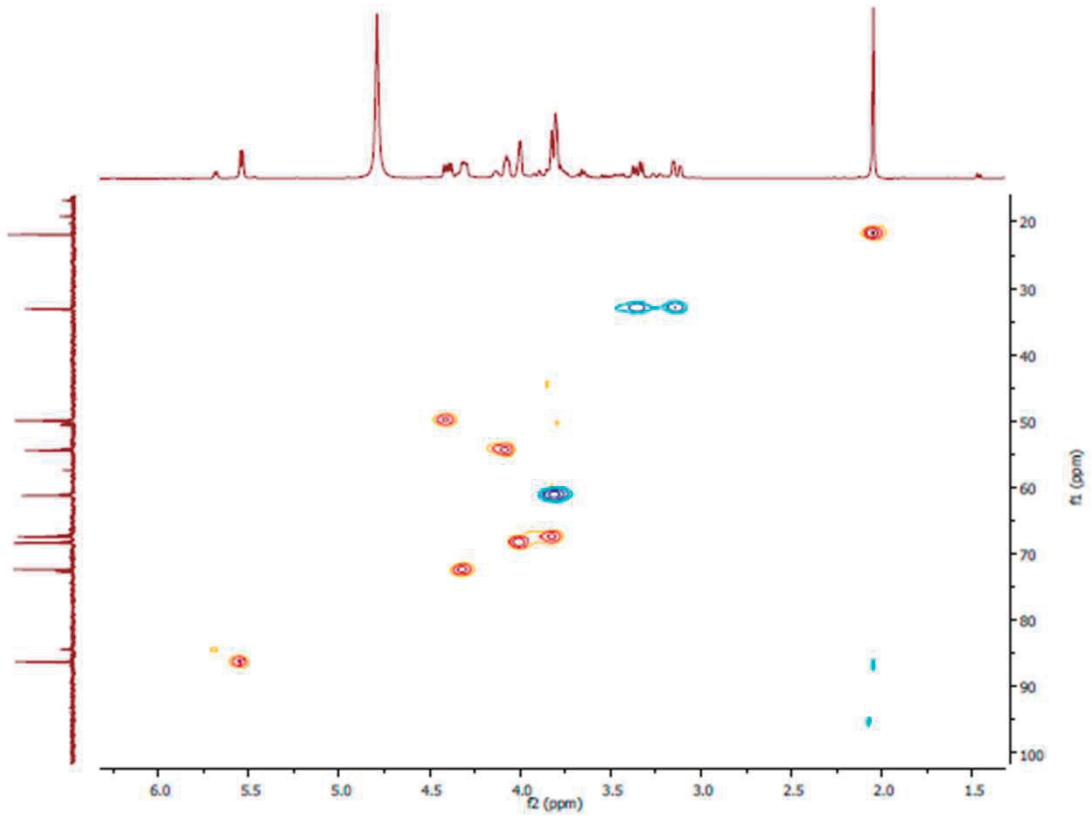

$^{13}$C NMR 100 MHz in $D_2O$



COSY in $D_2O$

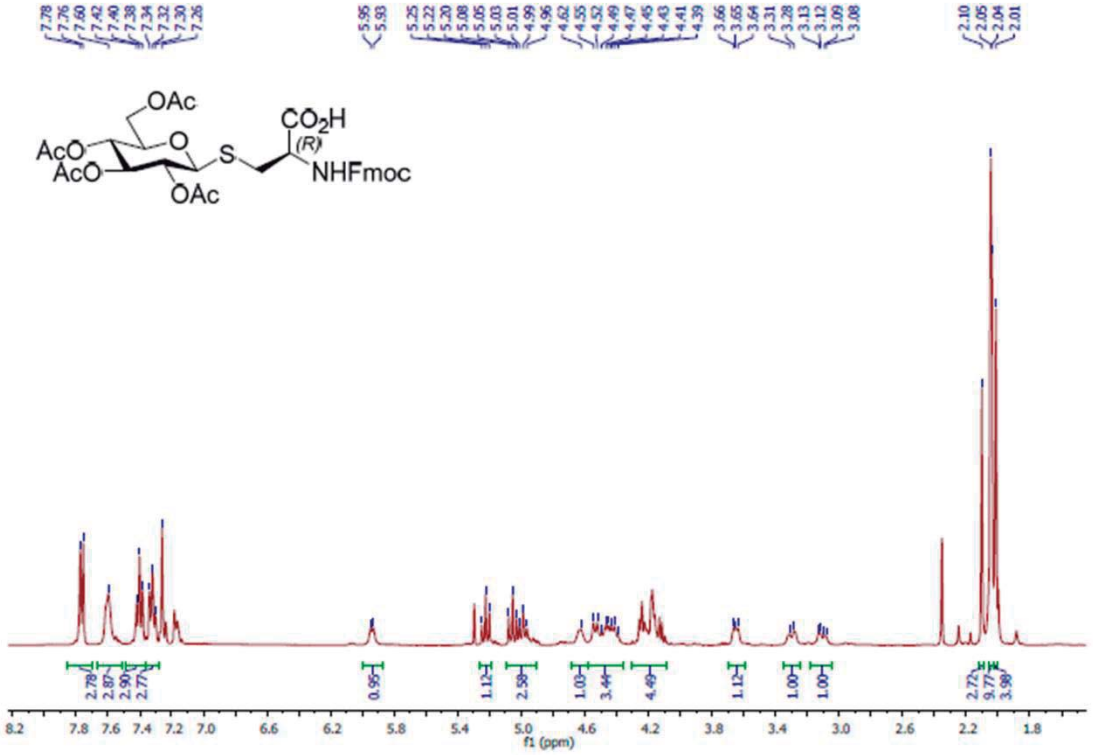



HSQC in $D_2O$

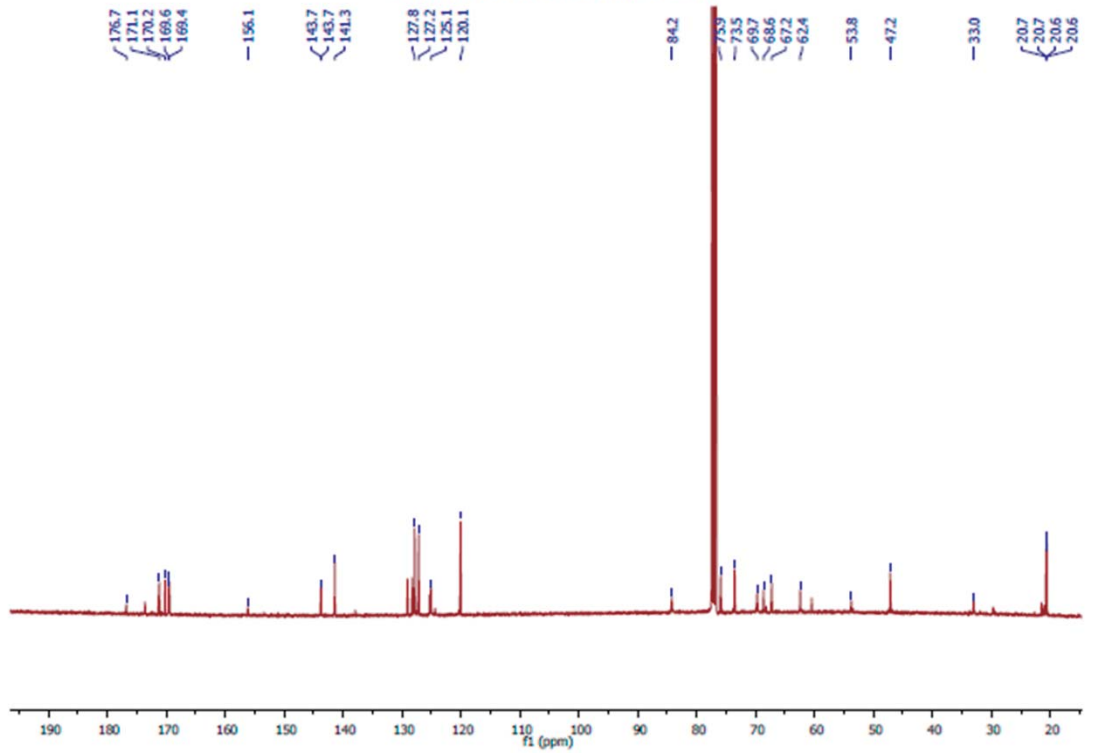


Compuesto **13a** capítulo 4

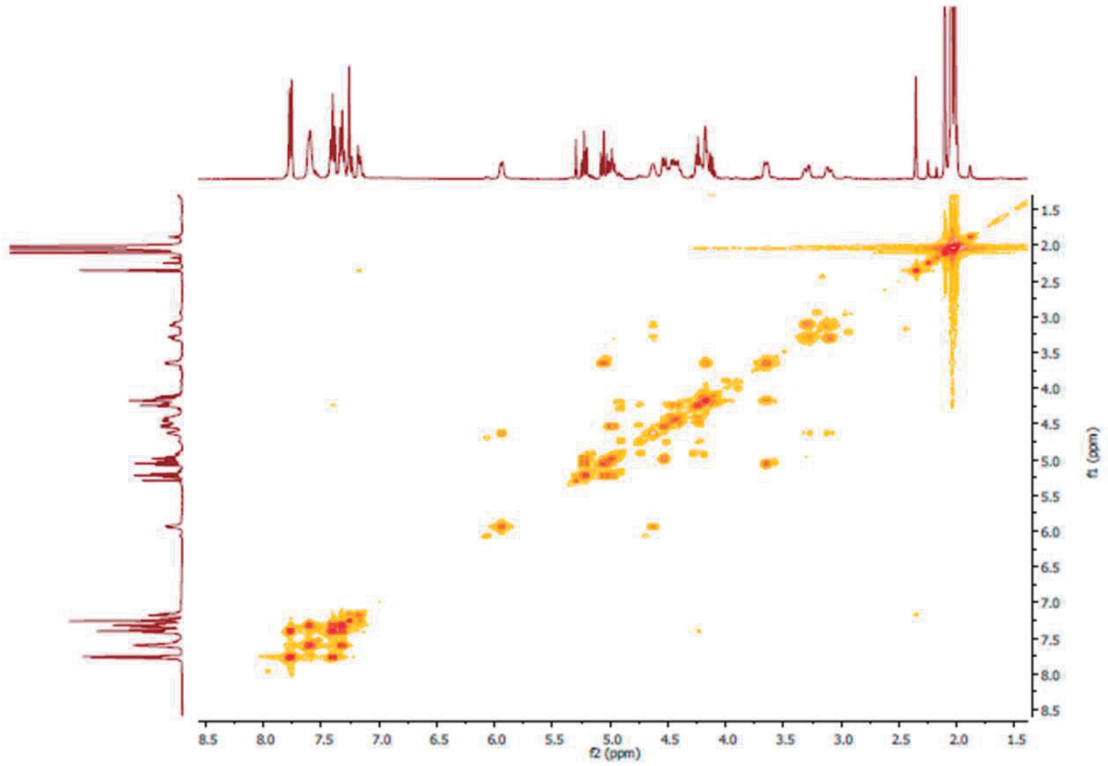
$^{1}$H NMR 400 MHz in $CDCl_3$



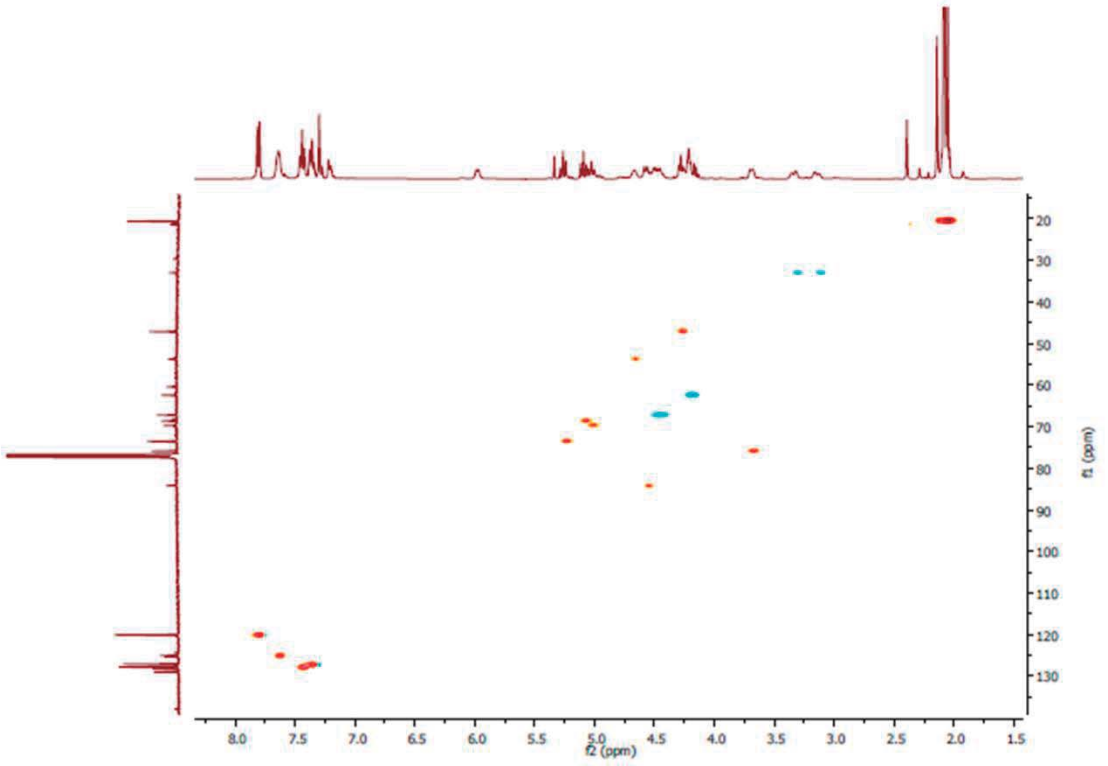

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

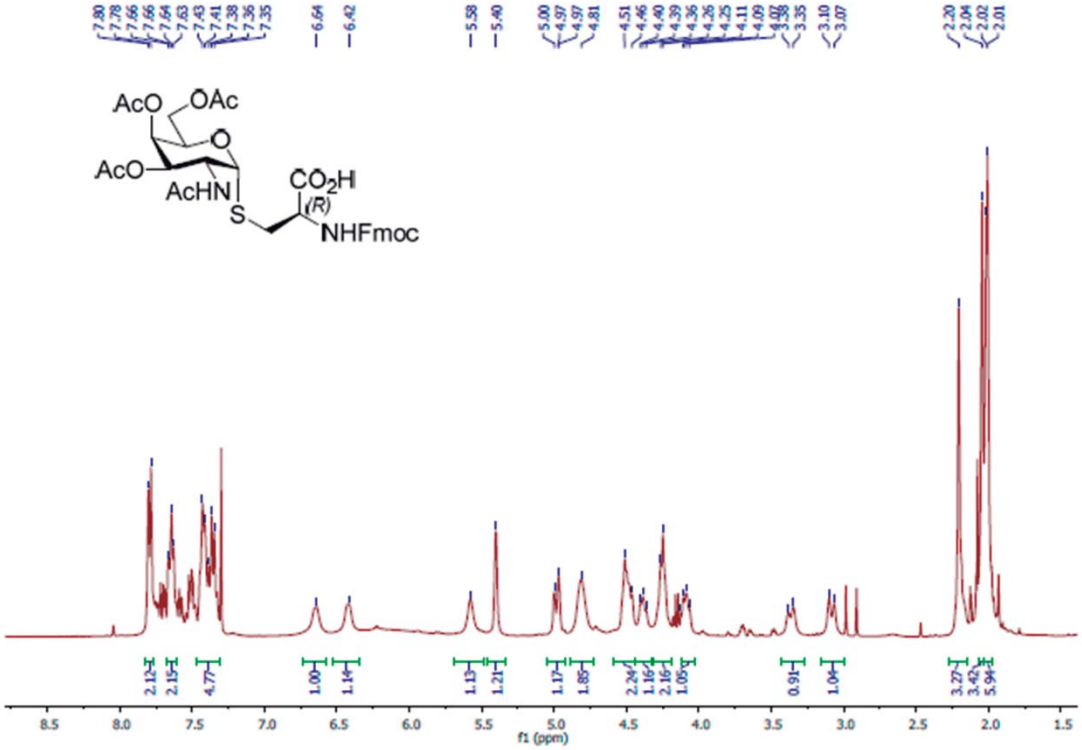



HSQC in $CDCl_3$

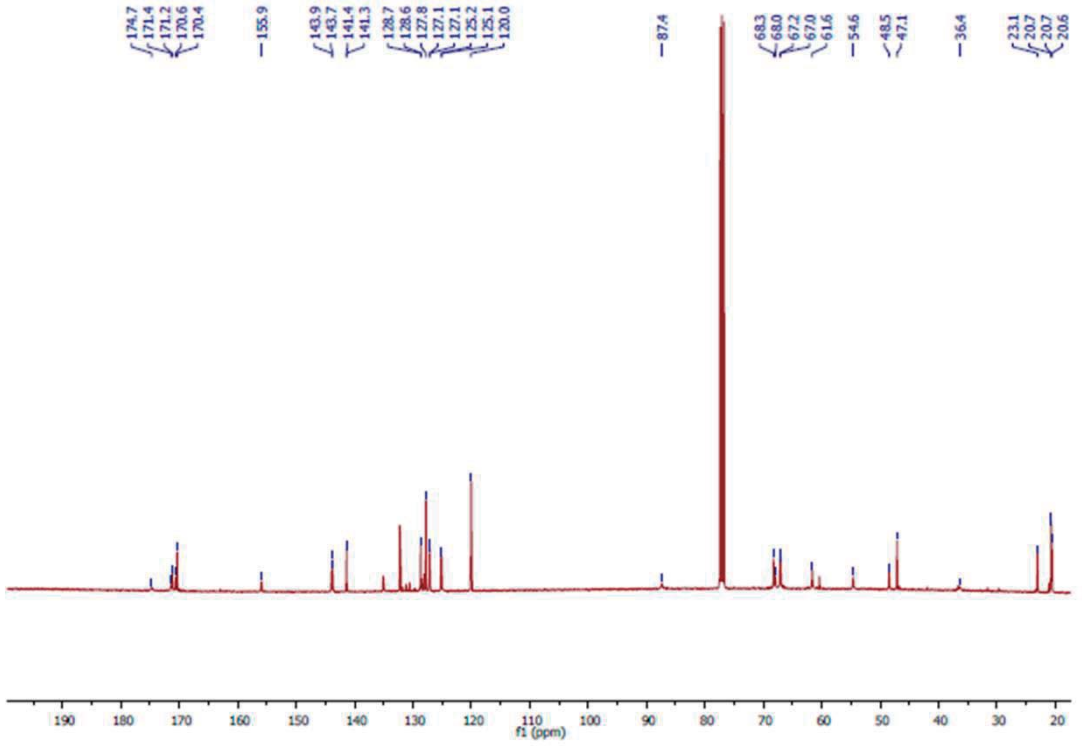


Compuesto **13f** capítulo 4

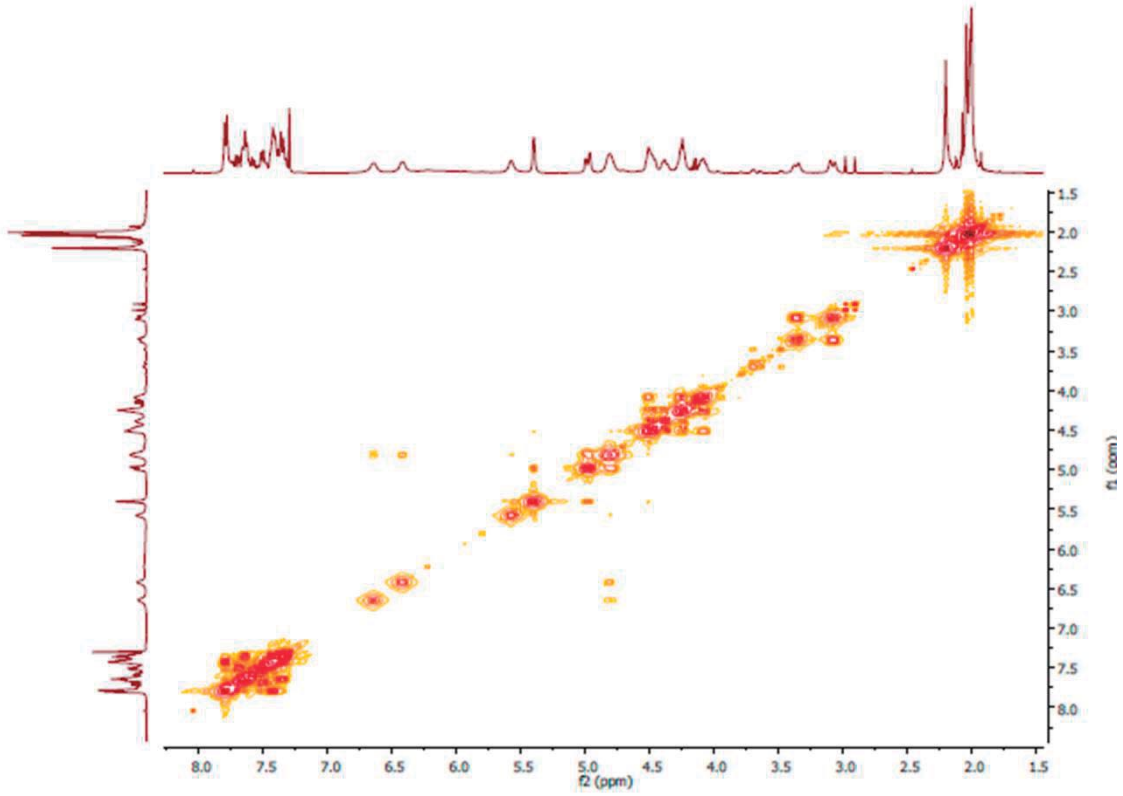
$^{1}$H NMR 400 MHz in $CDCl_3$



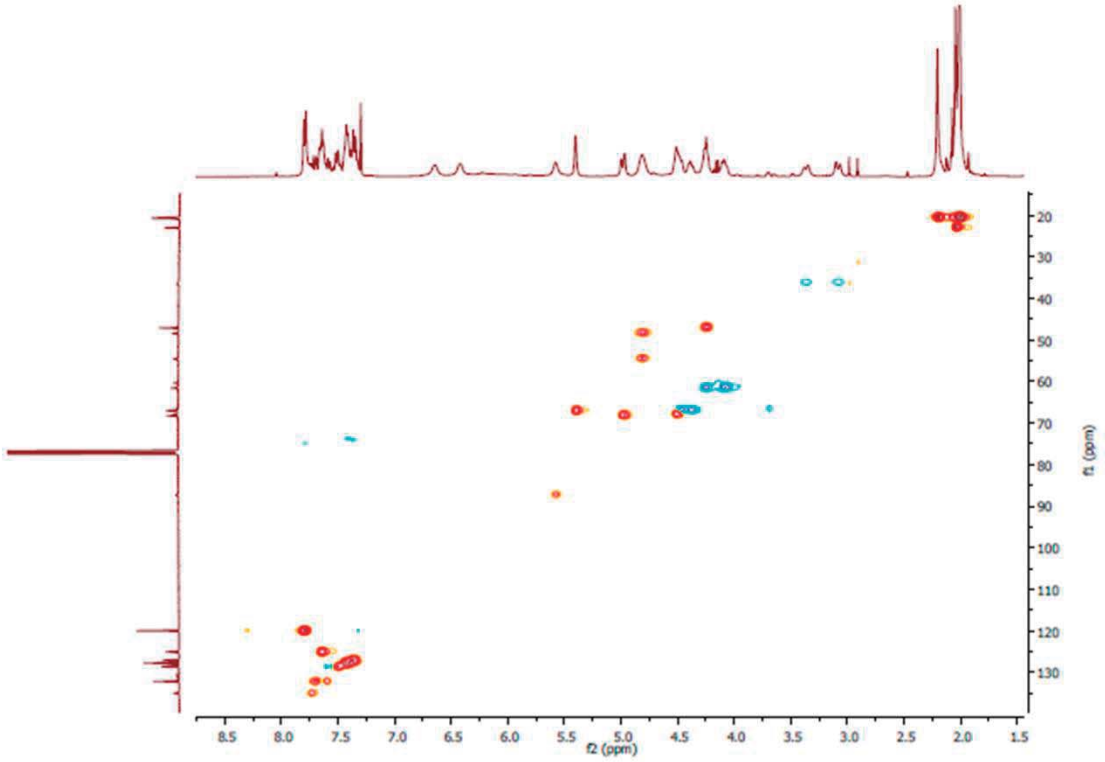

$^{13}$C NMR 100 MHz in $CDCl_3$



COSY in $CDCl_3$

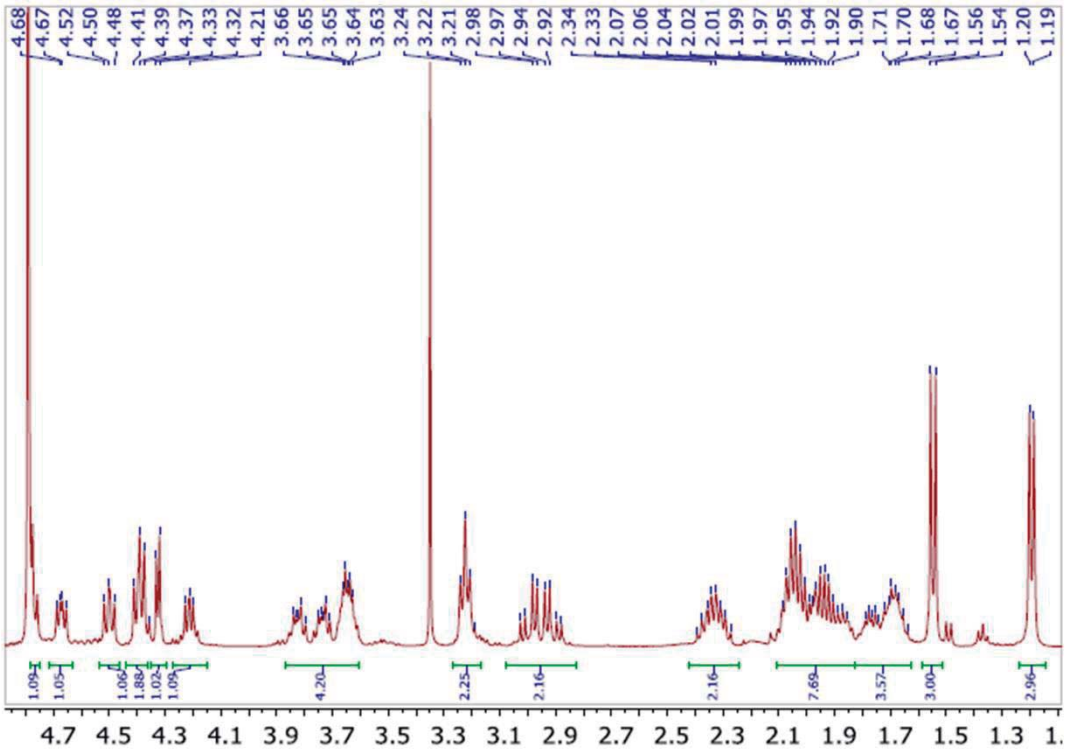



HSQC in $CDCl_3$

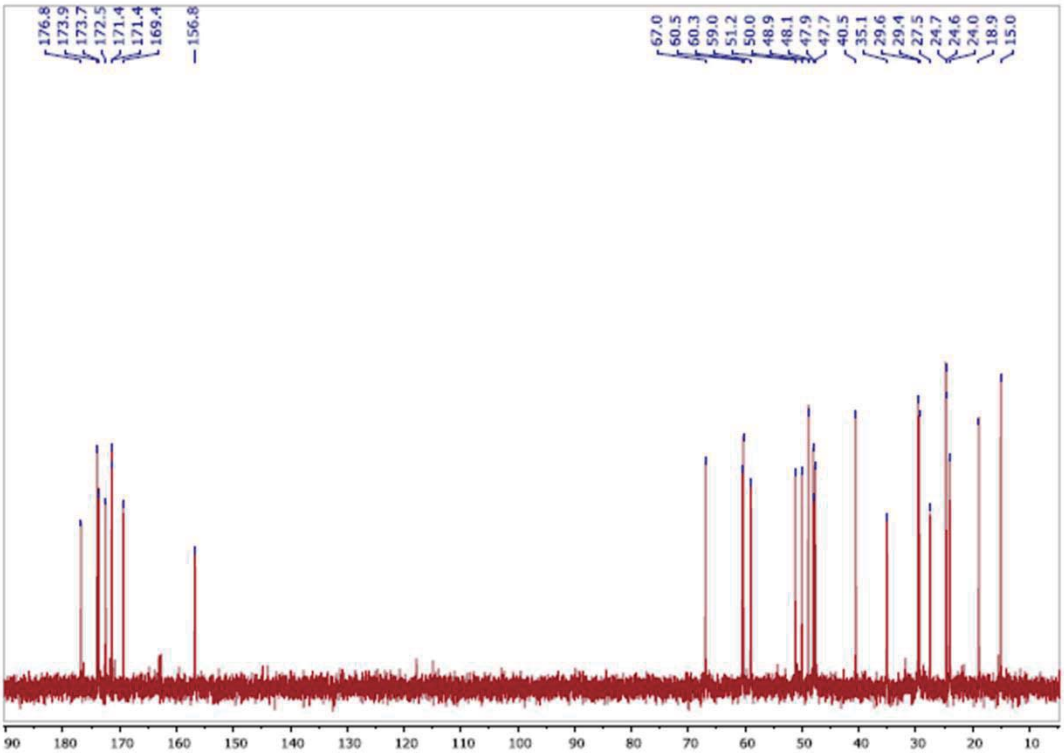


Compuesto **1** capítulo 5

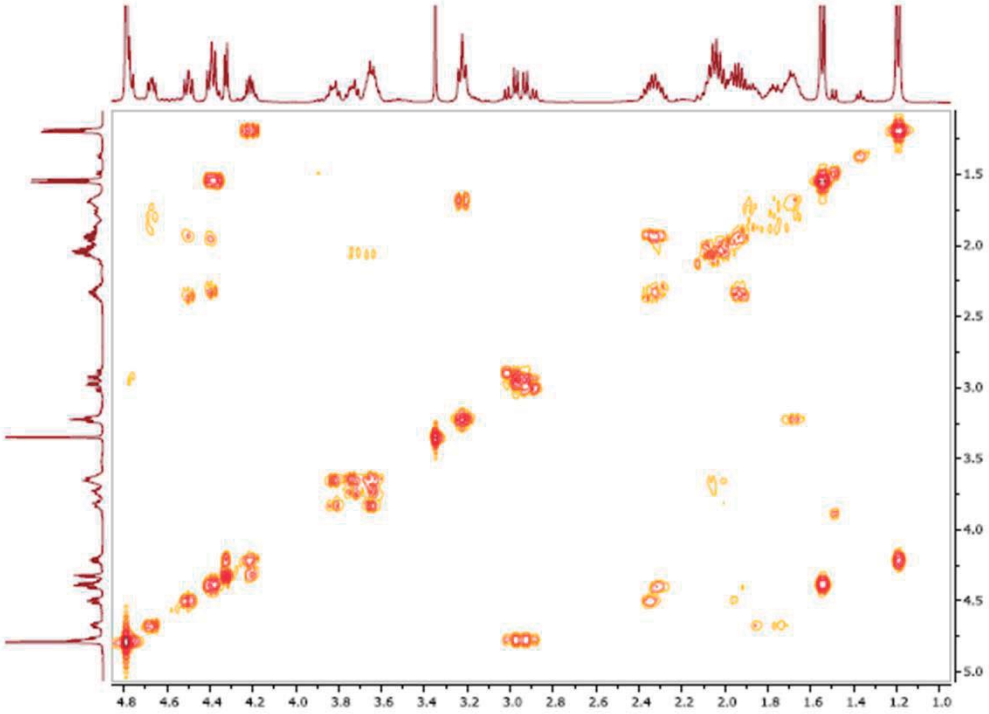
$^{1}$H NMR 400 MHz in $D_2O$



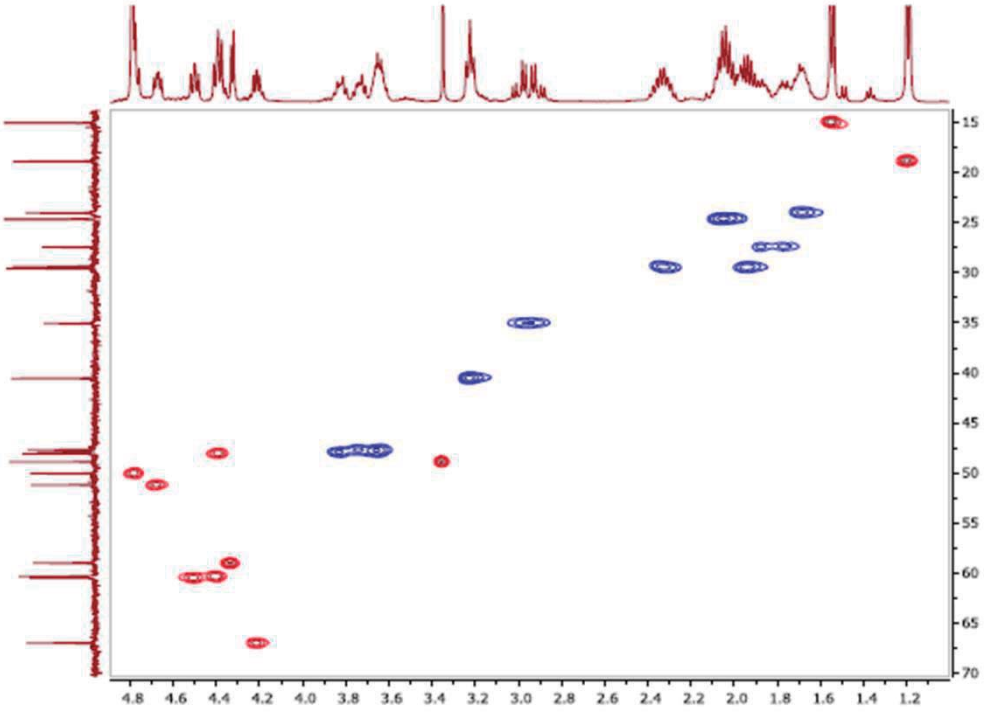

$^{13}$C NMR 100 MHz in $D_2O$



COSY in $D_2O$

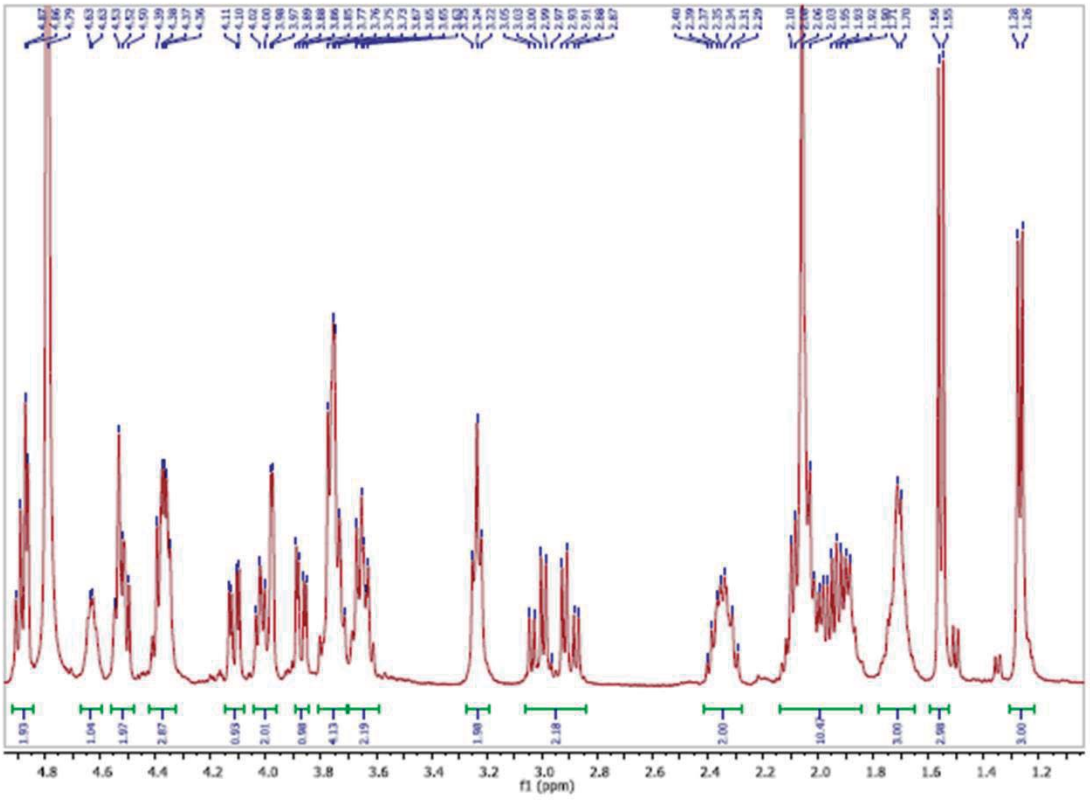



HSQC in $D_2O$

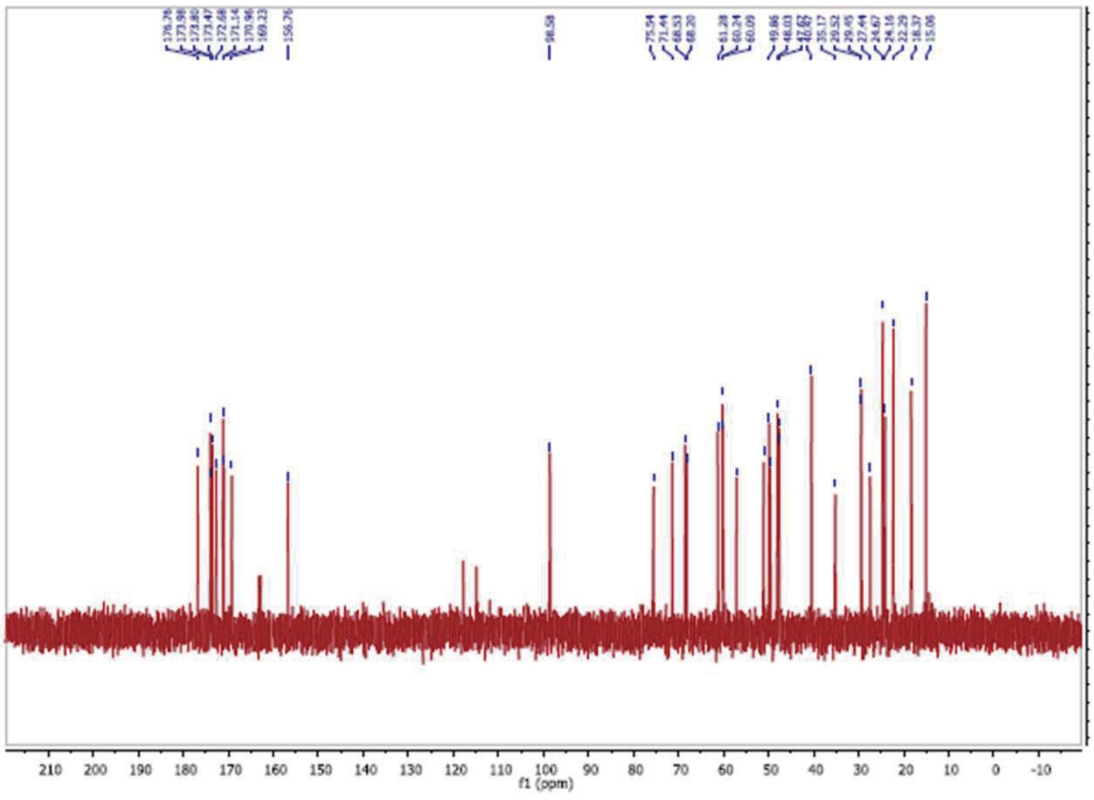


Compuesto **1*** capítulo 5

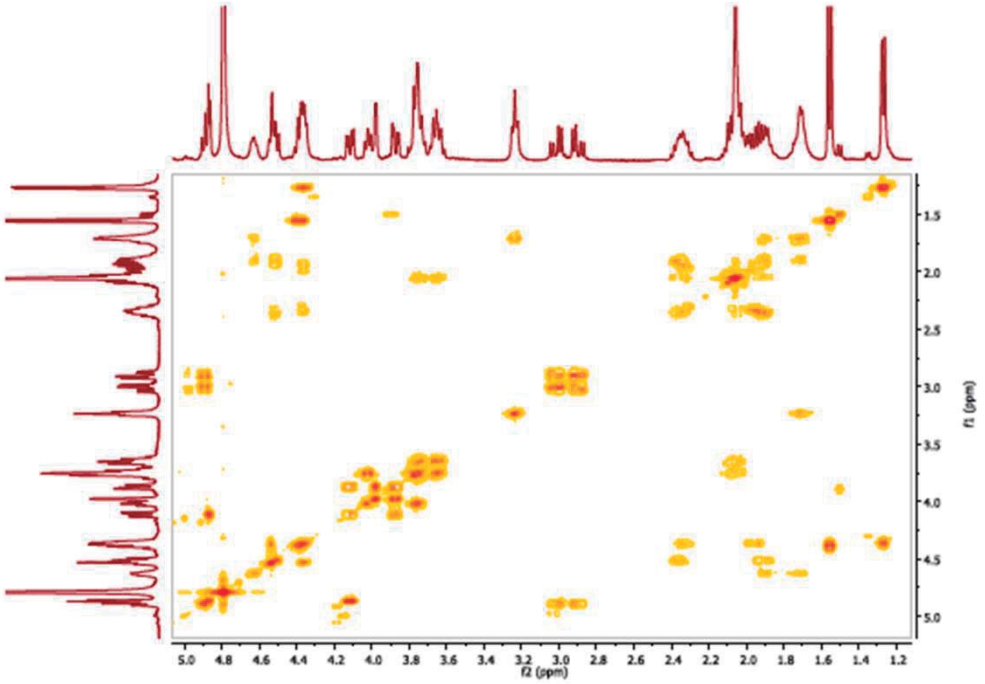
$^{1}$H NMR 400 MHz in $D_2O$



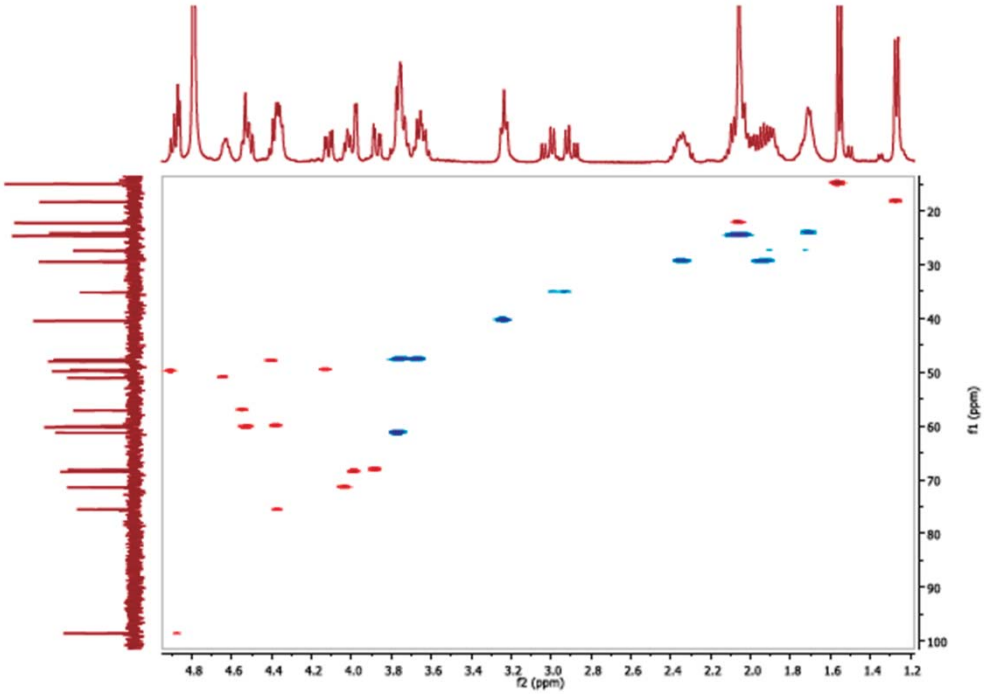

$^{13}$C NMR 100 MHz in $D_2O$



COSY in $D_2O$

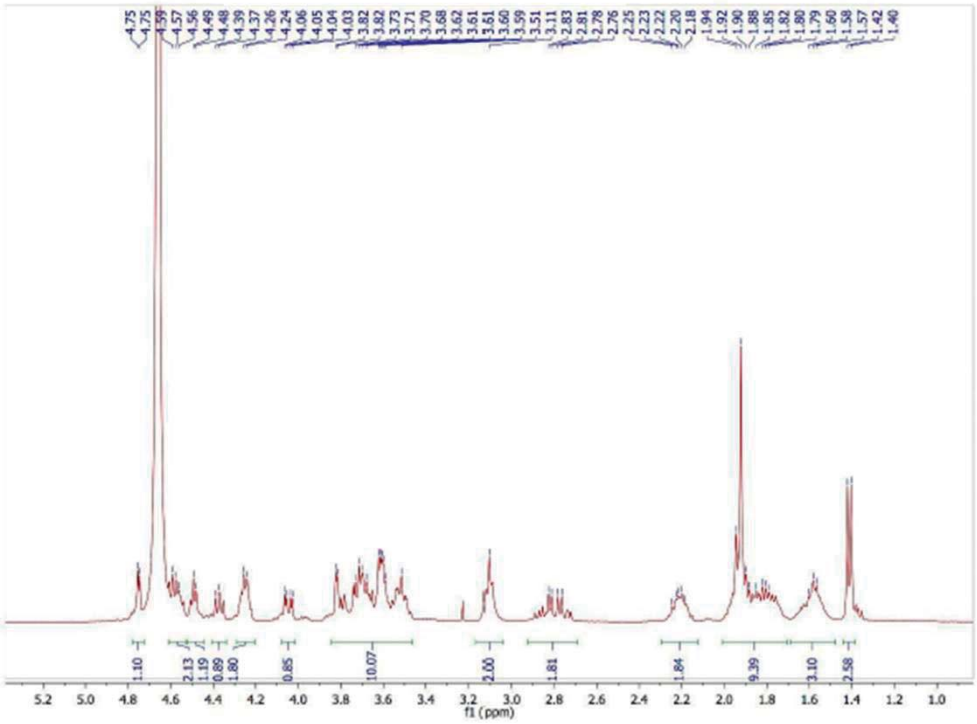



HSQC in $D_2O$

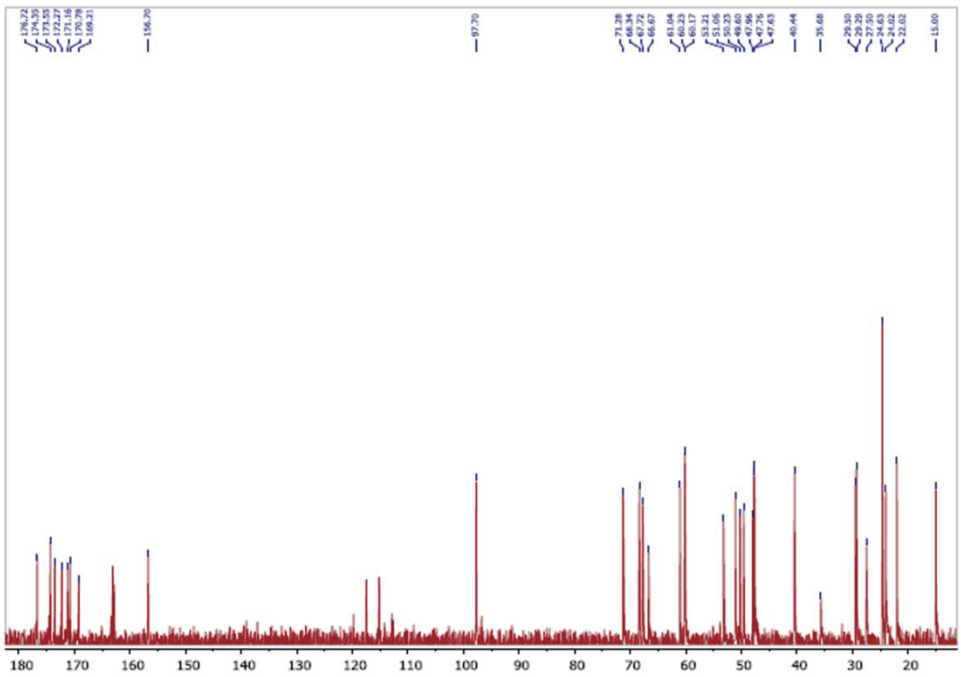


Compuesto **2*** capítulo 5

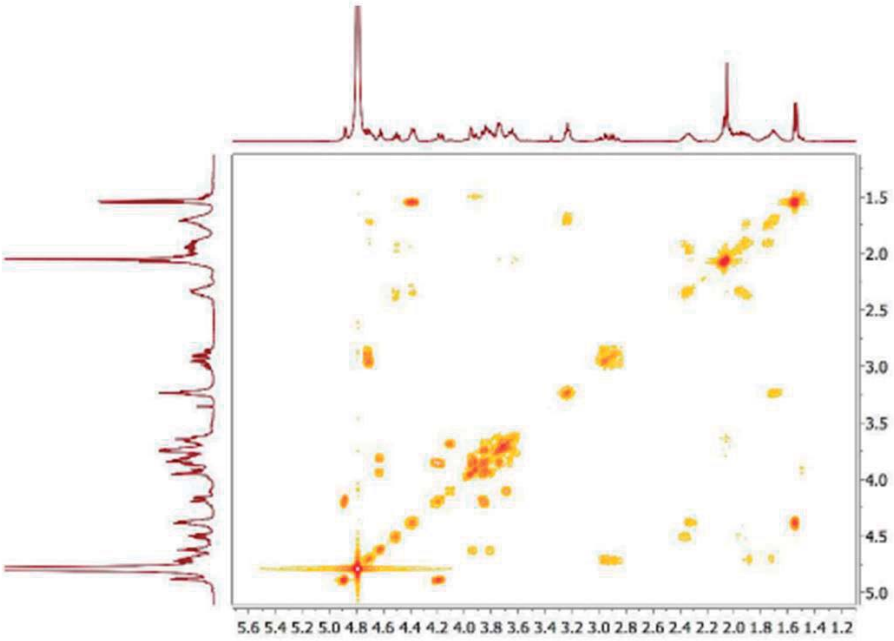
$^{1}$H NMR 400 MHz in $D_2O$



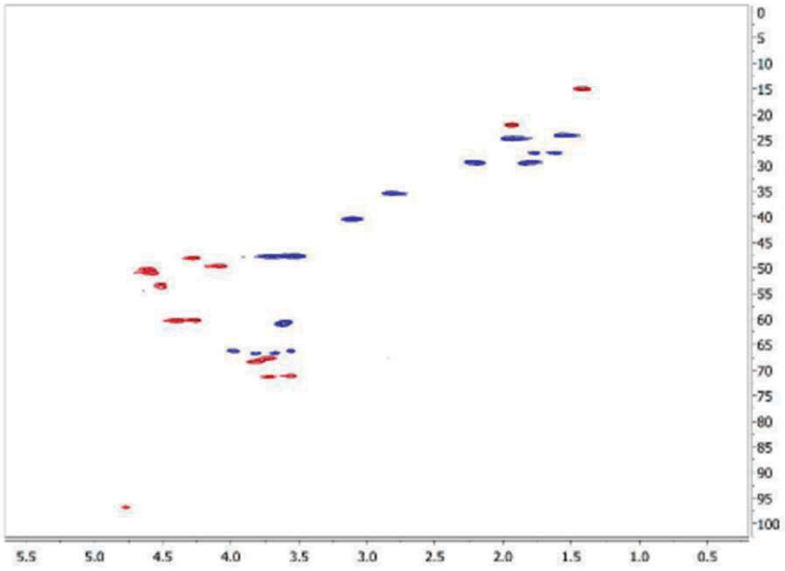
$^{13}$C NMR 125 MHz



COSY in $D_2O$

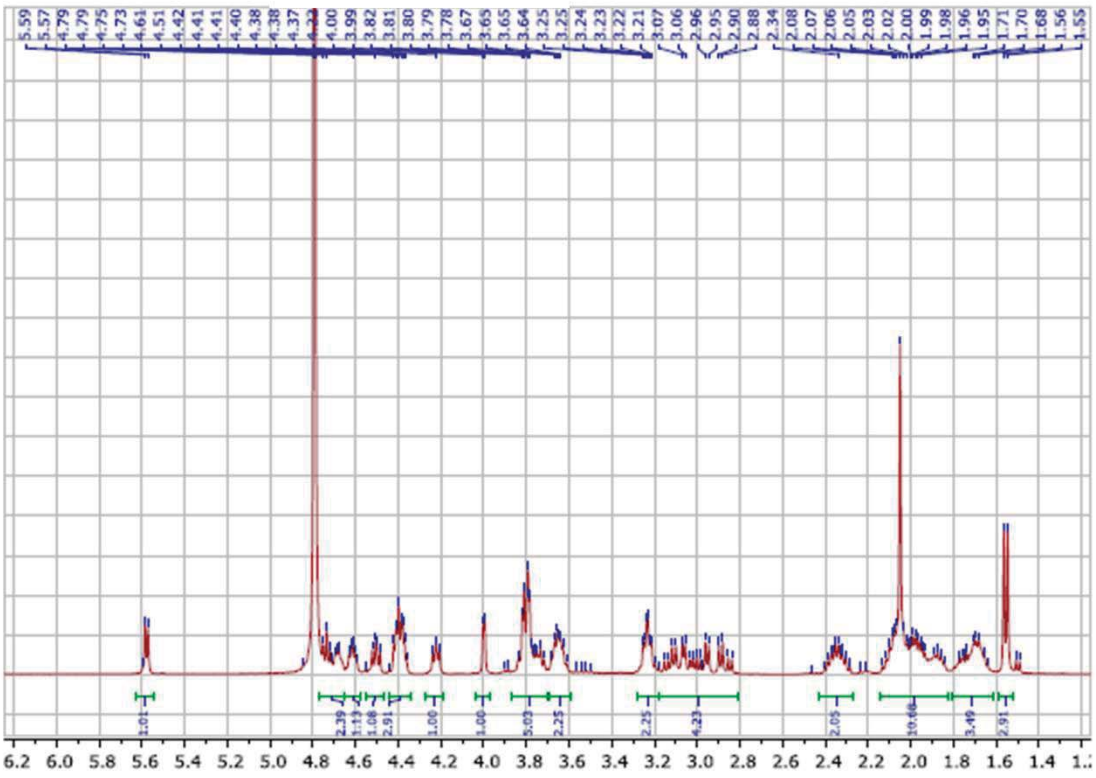



HSQC in $D_2O$

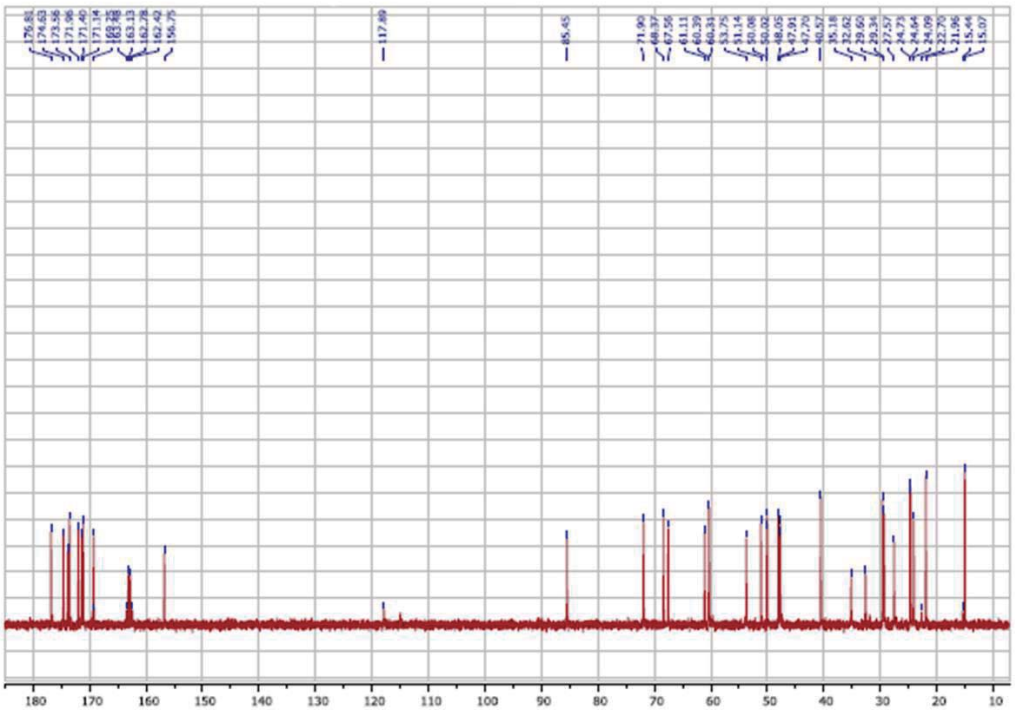


Compuesto **3*** capítulo 5

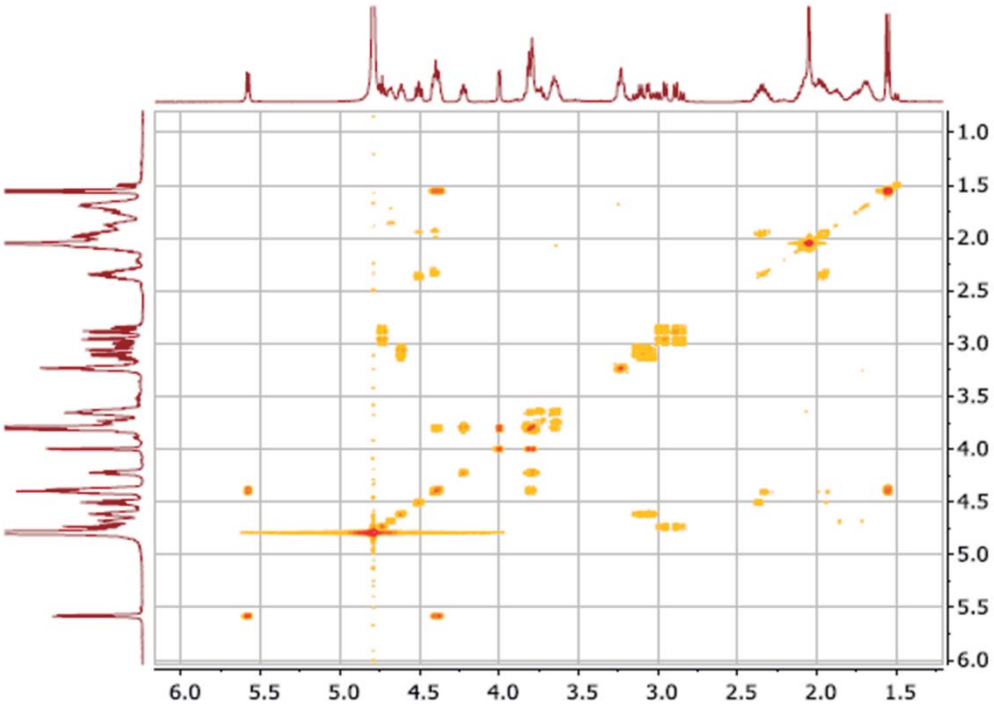
$^{1}$H NMR 400 MHz in $D_2O$



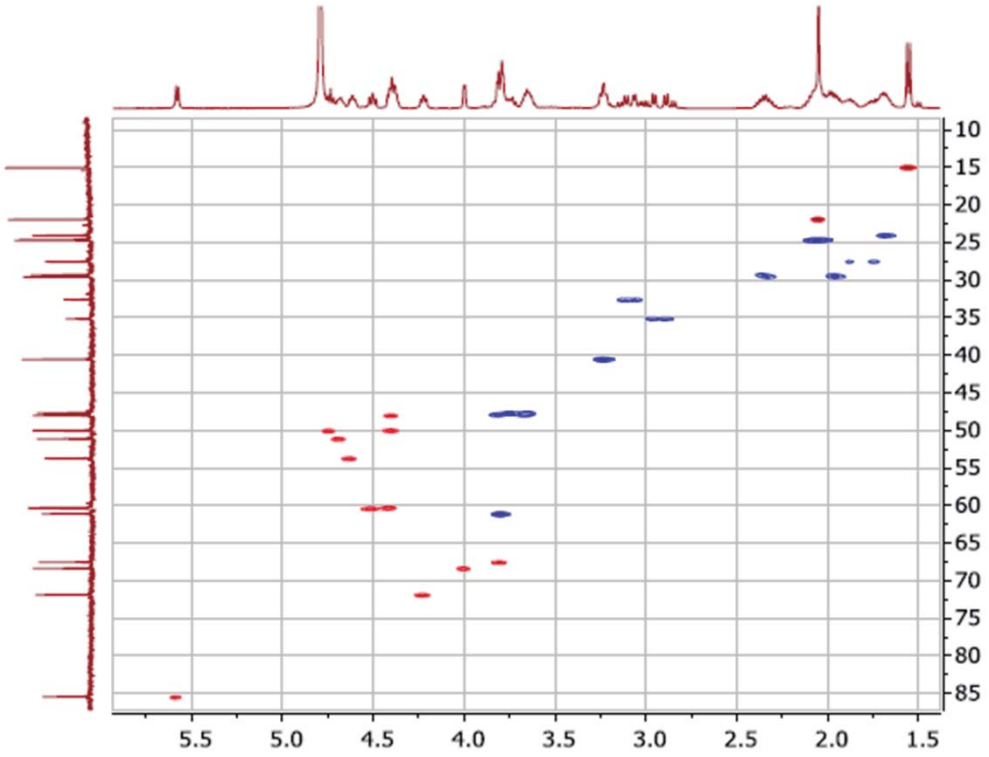
$^{13}$C NMR 100 MHz in $D_2O$



COSY in $D_2O$



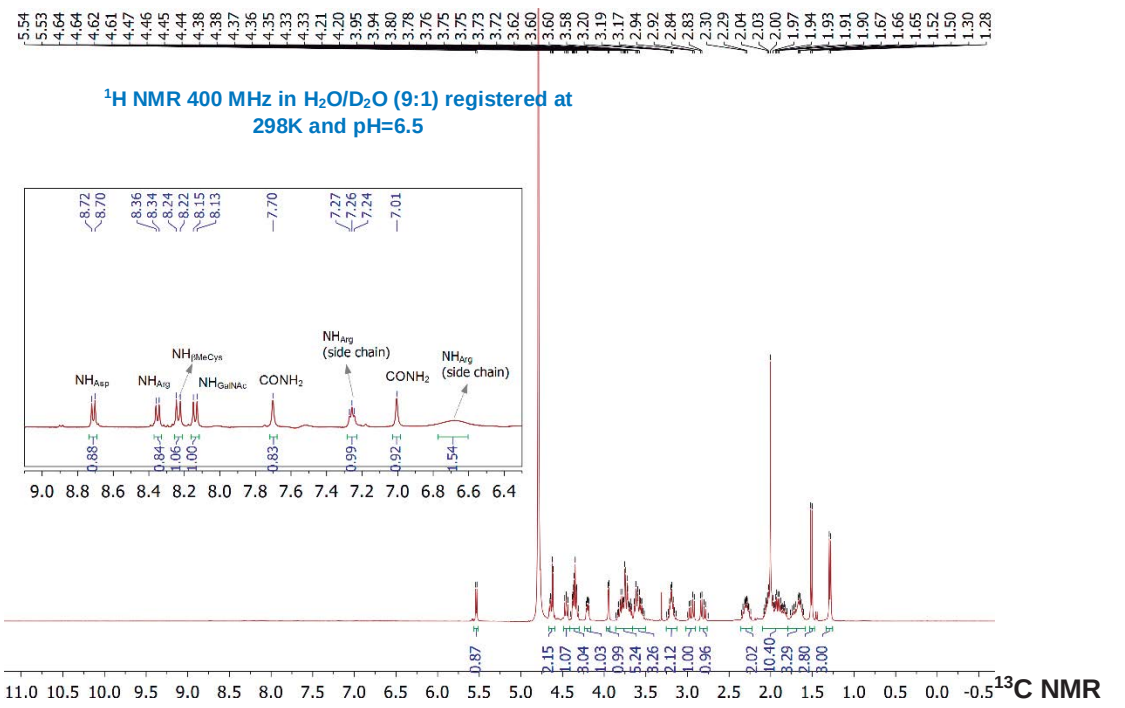

HSQC in $D_2O$



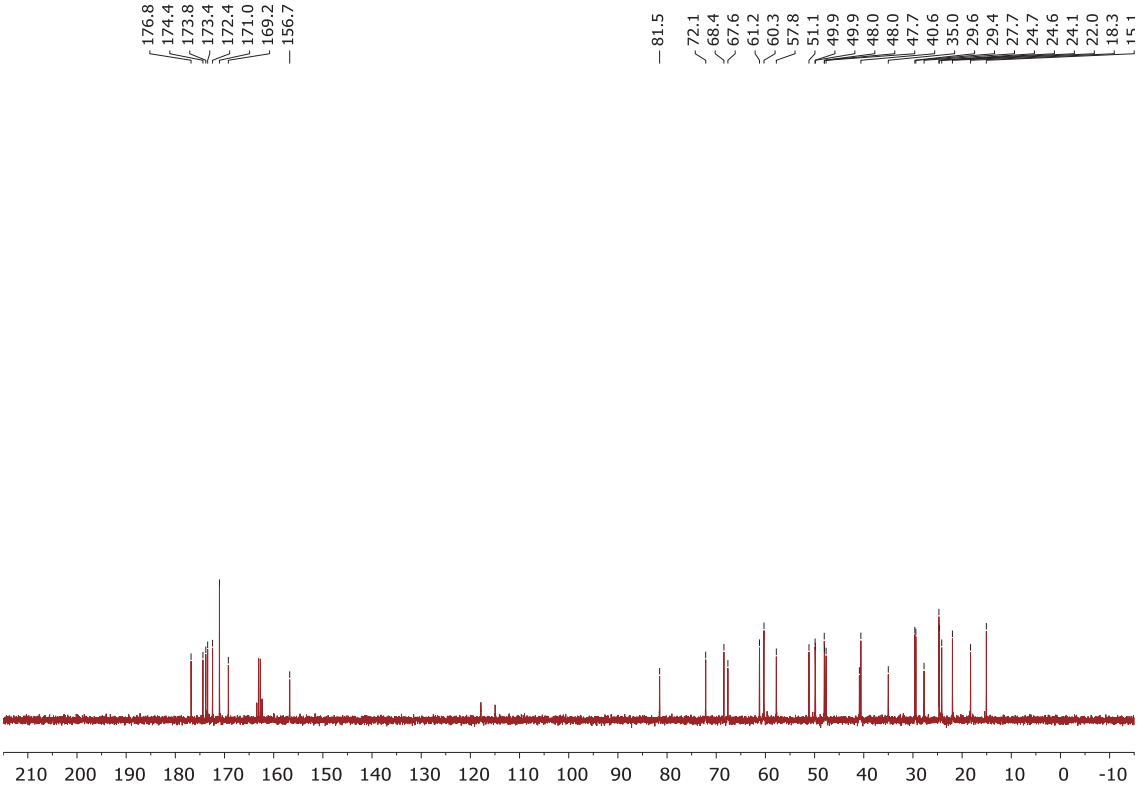
Compuesto **2*** capítulo 6

**$^{1}$H NMR 400 MHz in $D_2O$**

**$^{1}$H NMR 400 MHz in $H_2O/D_2O$ (9:1) registered at 298K and pH=6.5**

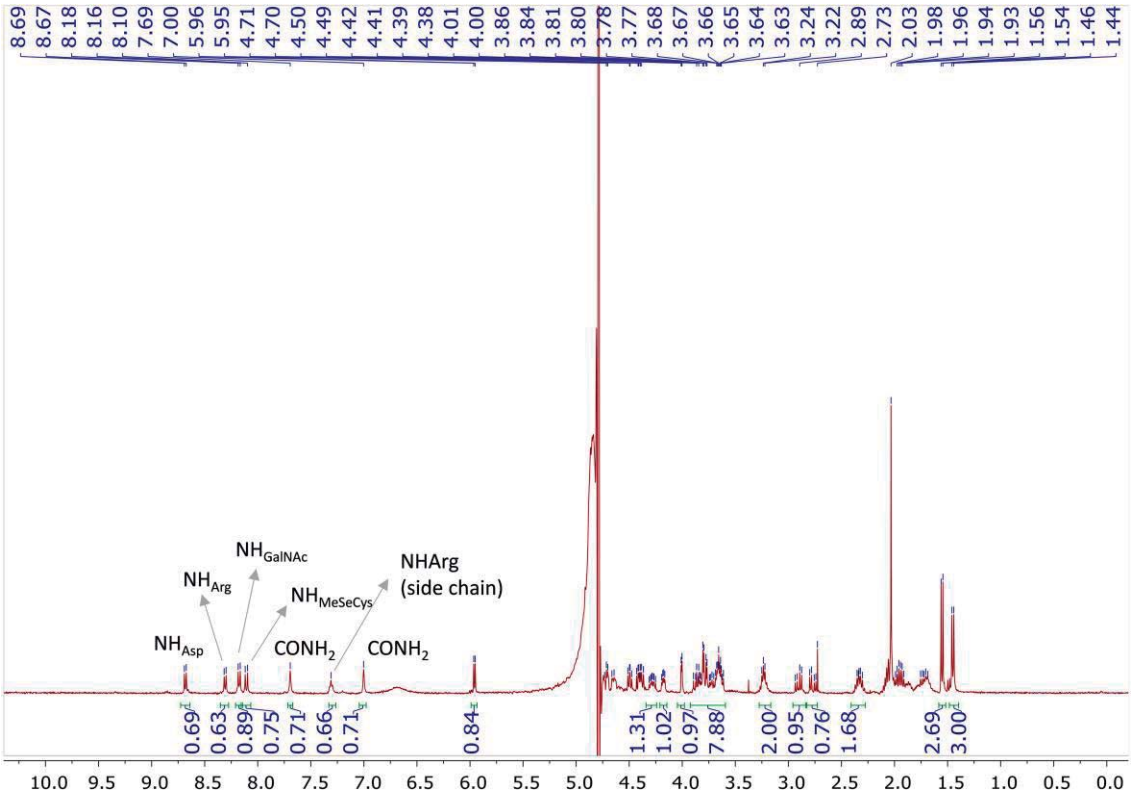



**$^{13}$C NMR 100 MHz in $D_2O$**



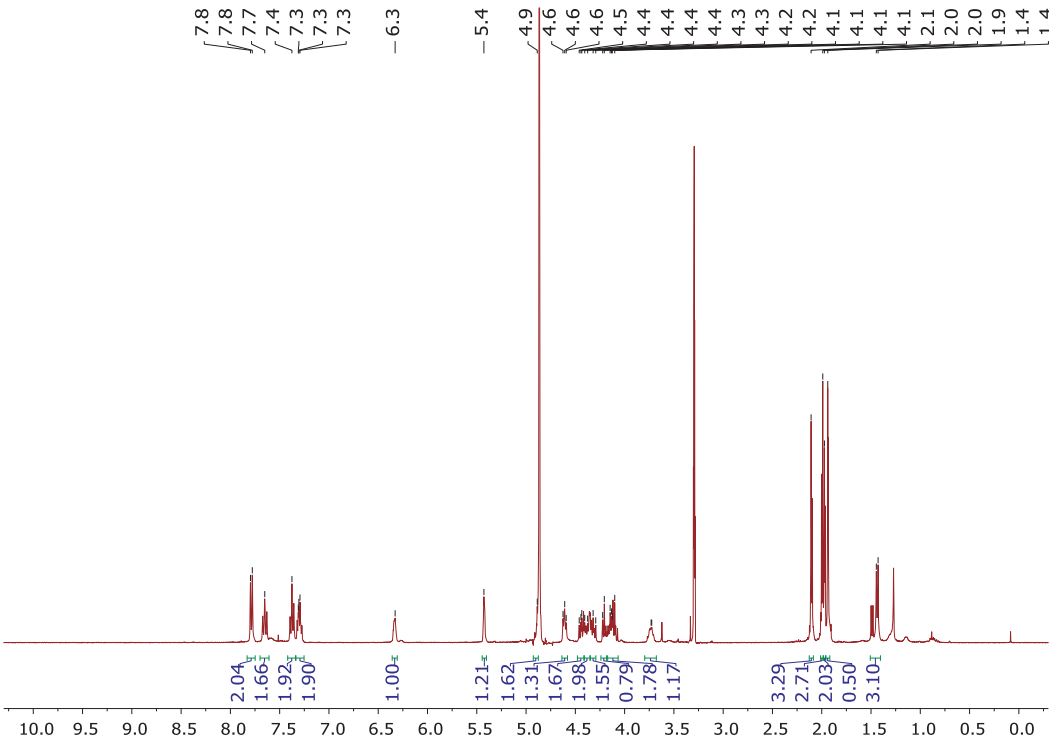
## Compuesto **3*** capítulo 6

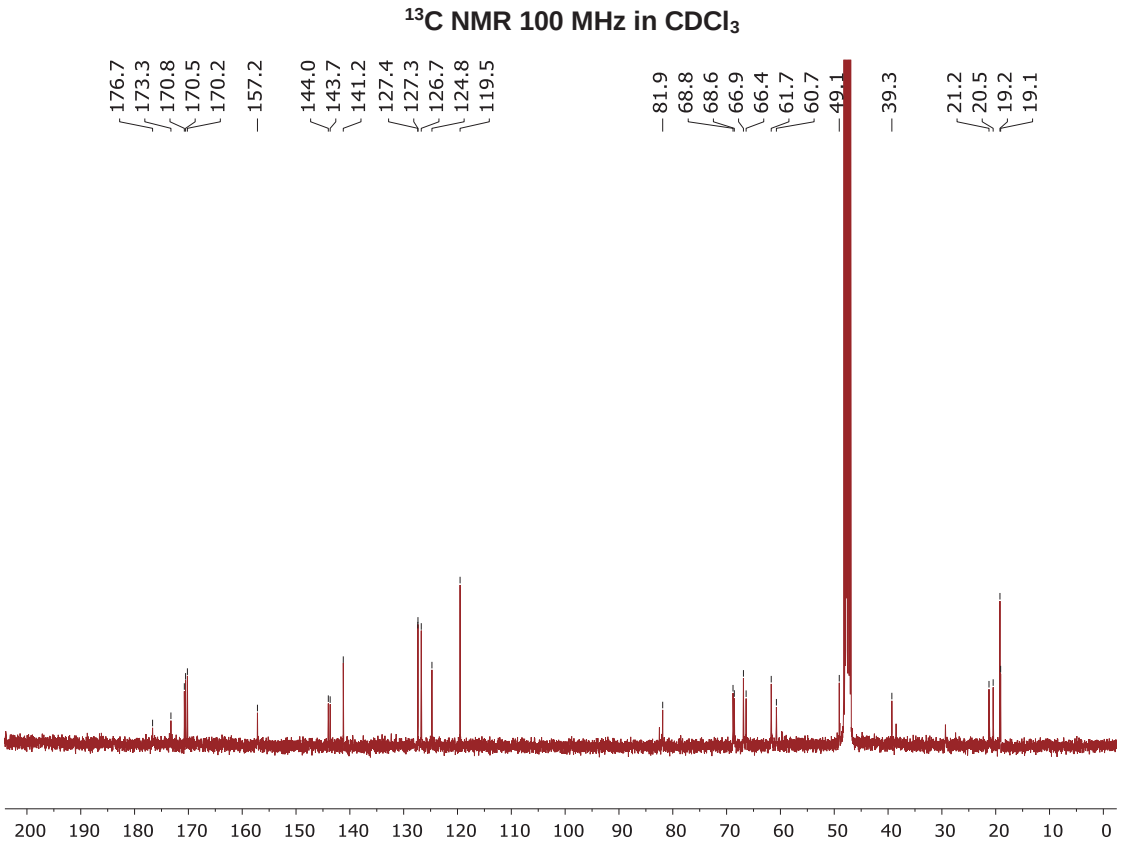
**$^1$H NMR 400 MHz in $H_2O/D_2O$ (9:1)**

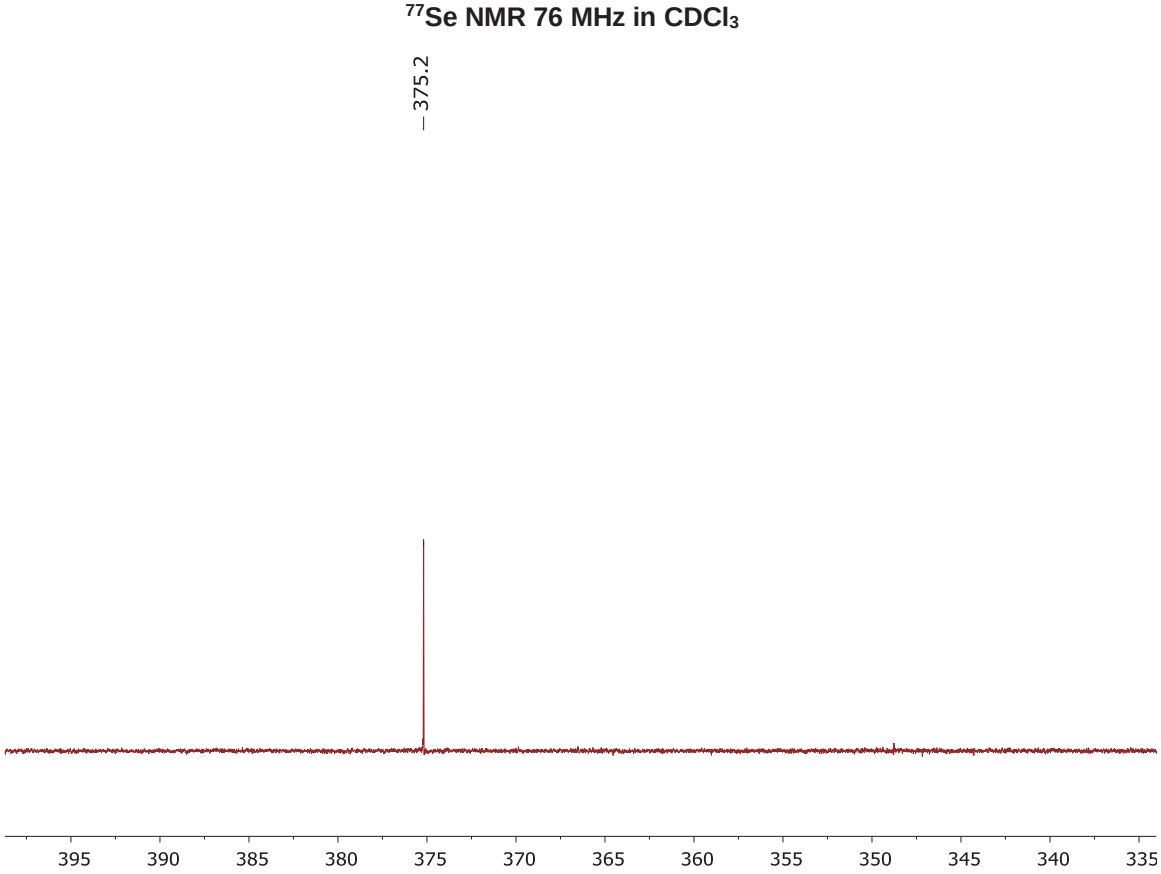



## Compuesto **4** capítulo 6

**$^1$H NMR 400 MHz in $CDCl_3$**

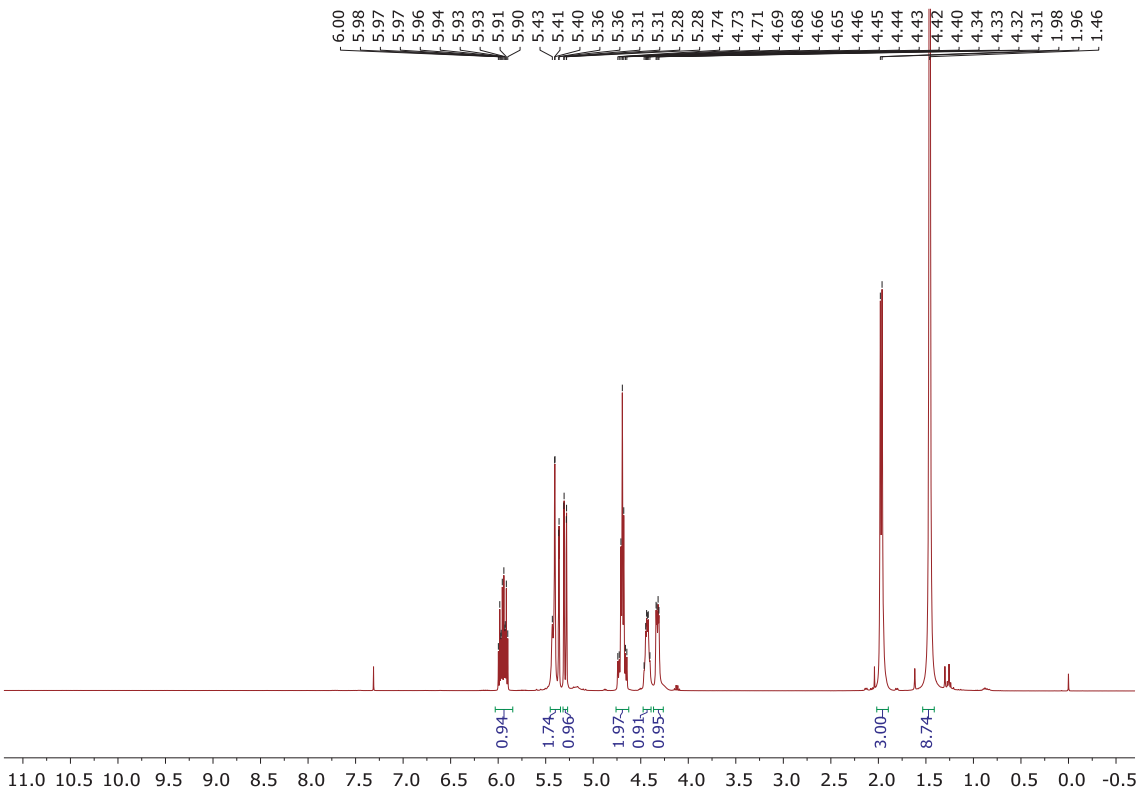




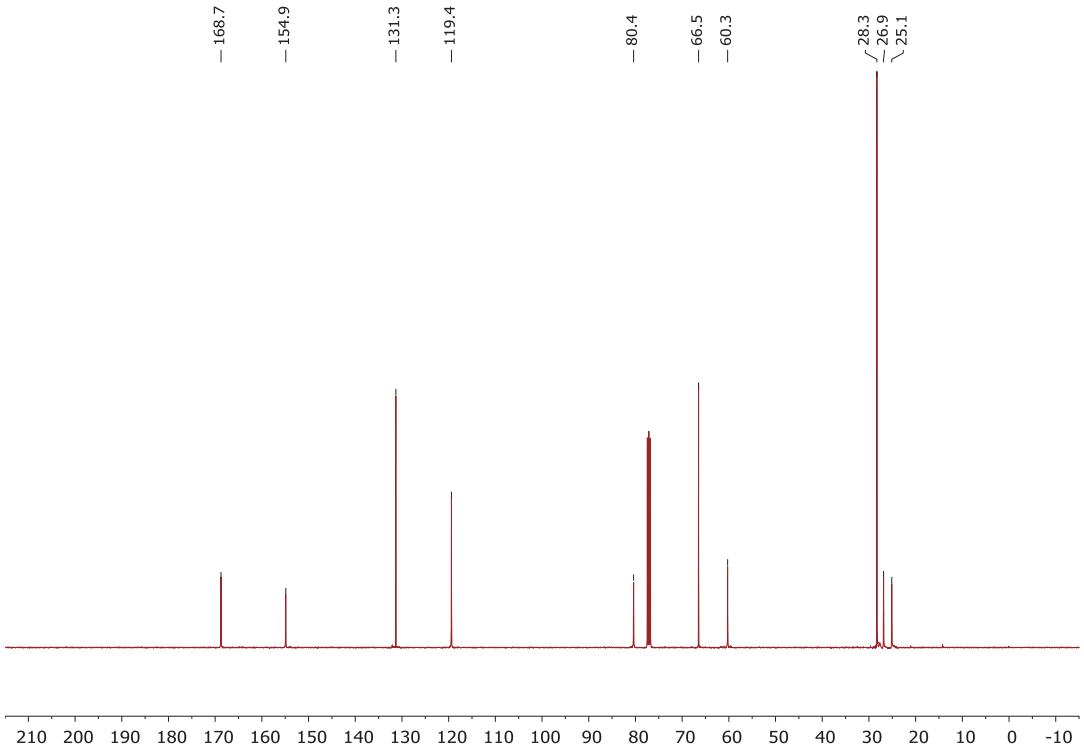




Compuesto **7** capítulo 6

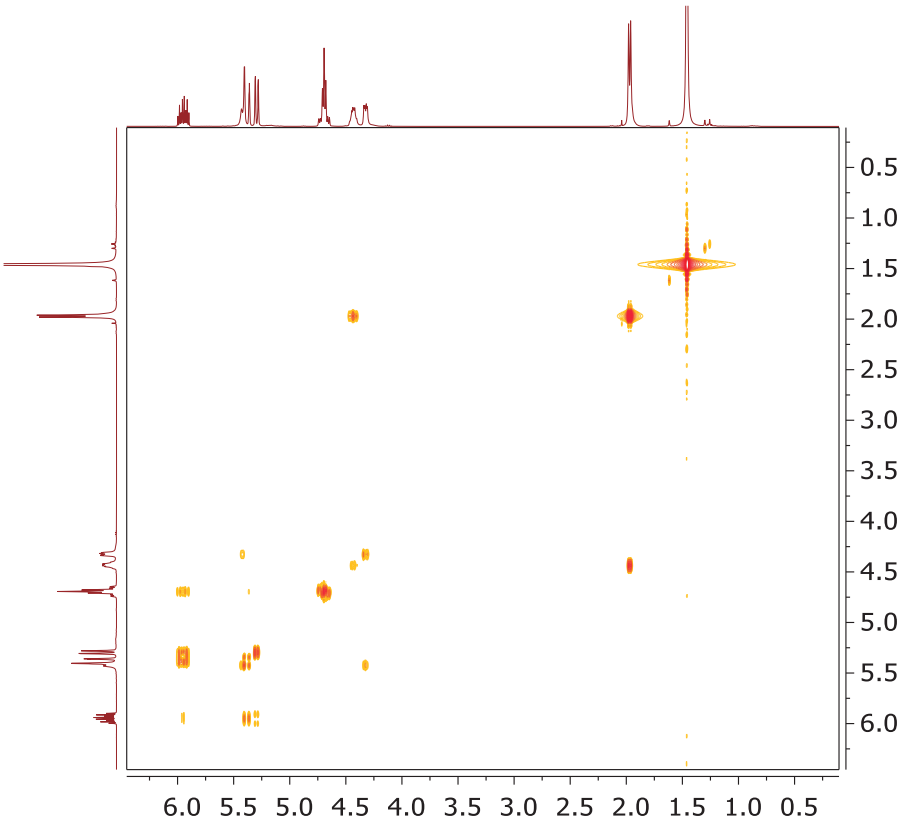
**$^{1}$H NMR 400 MHz in $CDCl_3$**



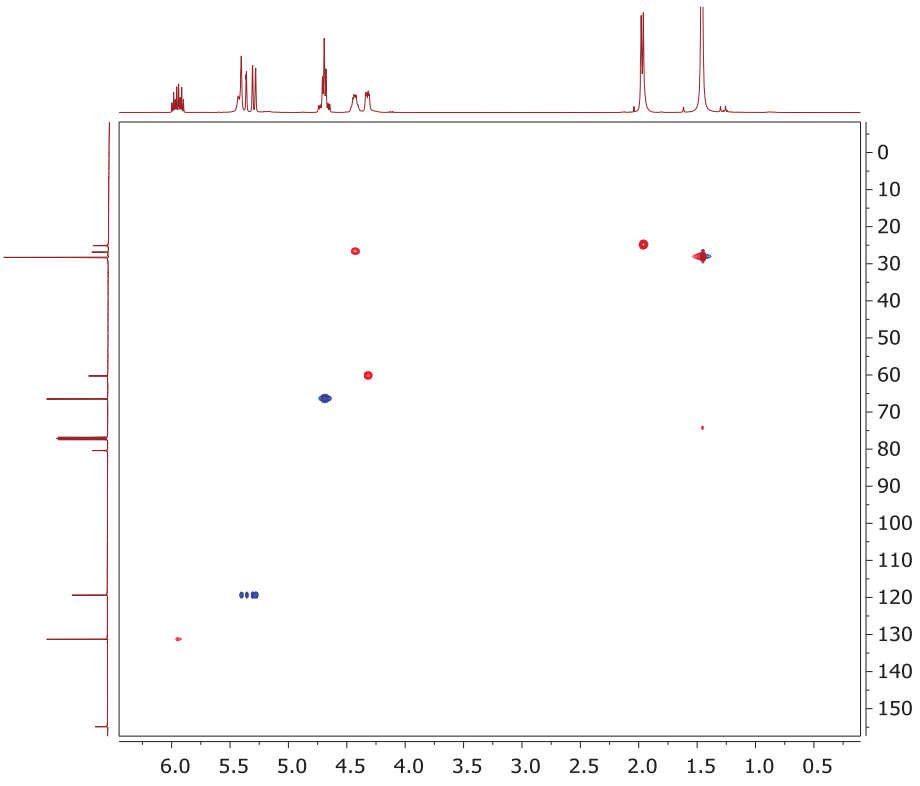

**$^{13}$C NMR 100 MHz in $CDCl_3$**



**COSY in $CDCl_3$**

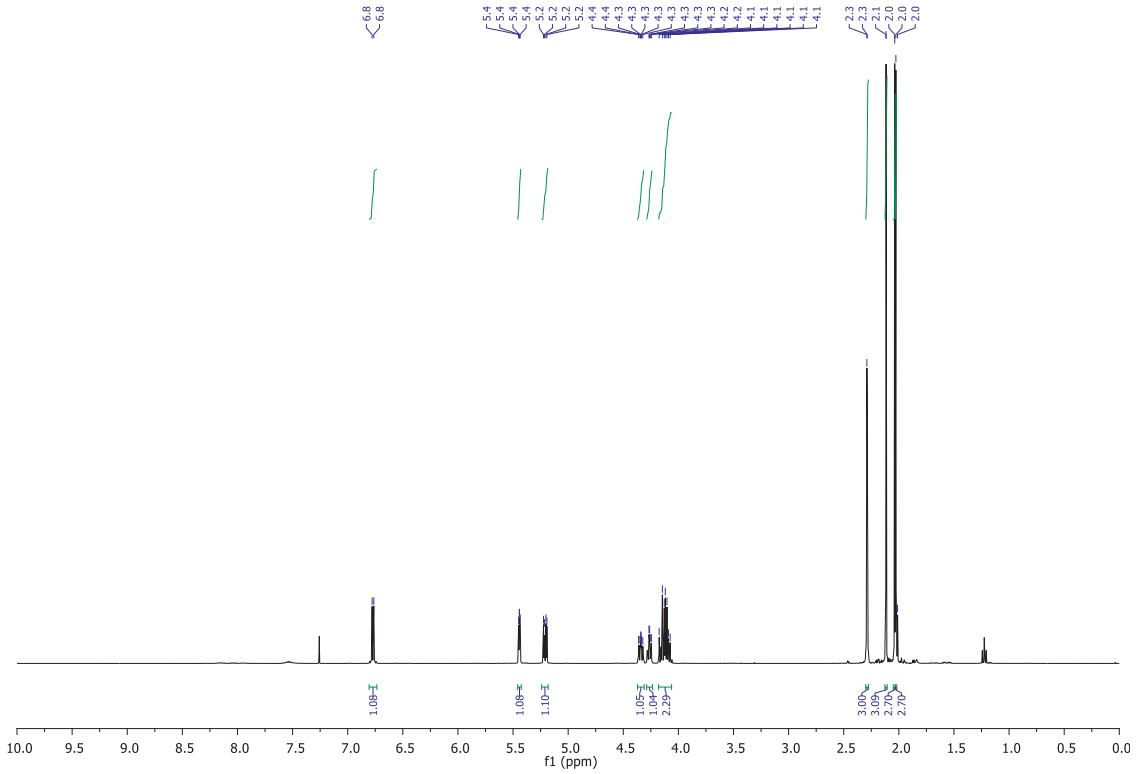



**HSQC in $CDCl_3$**

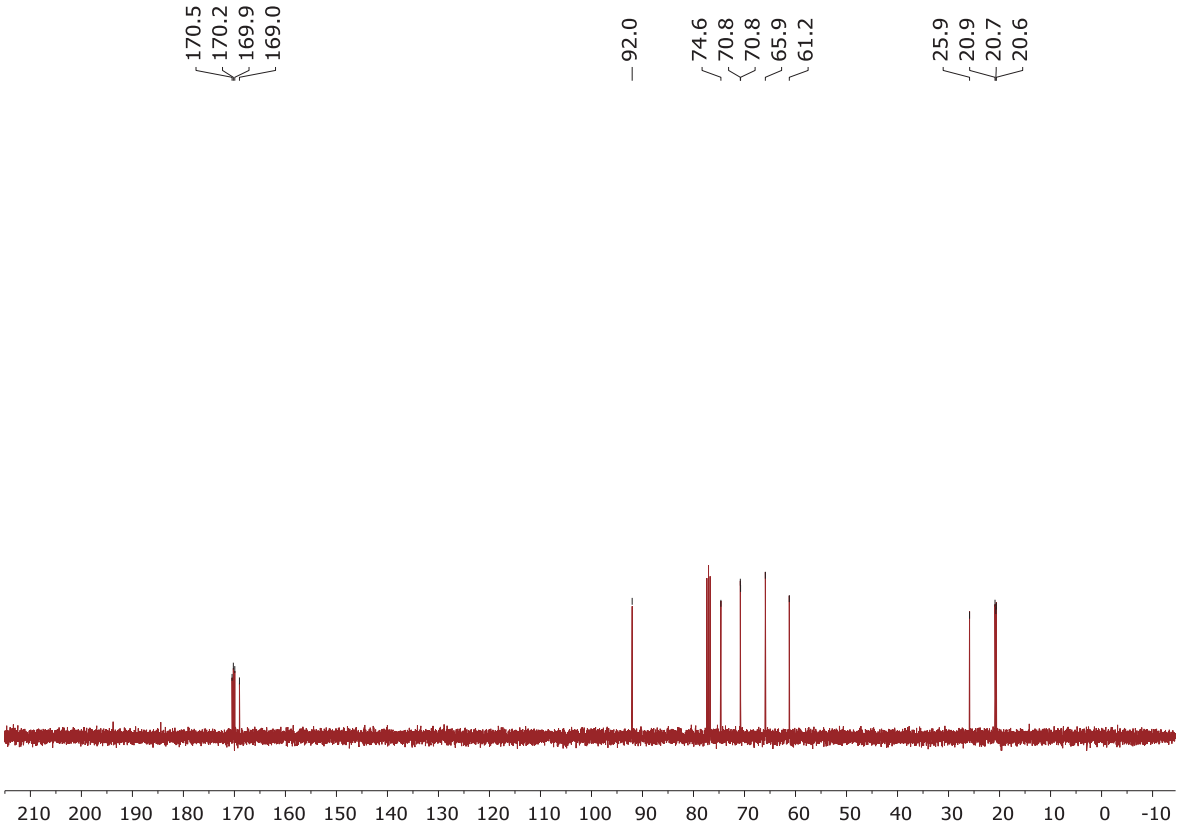


## Compuesto **9** capítulo 6

**$^{1}$H NMR 400 MHz in $CDCl_3$**

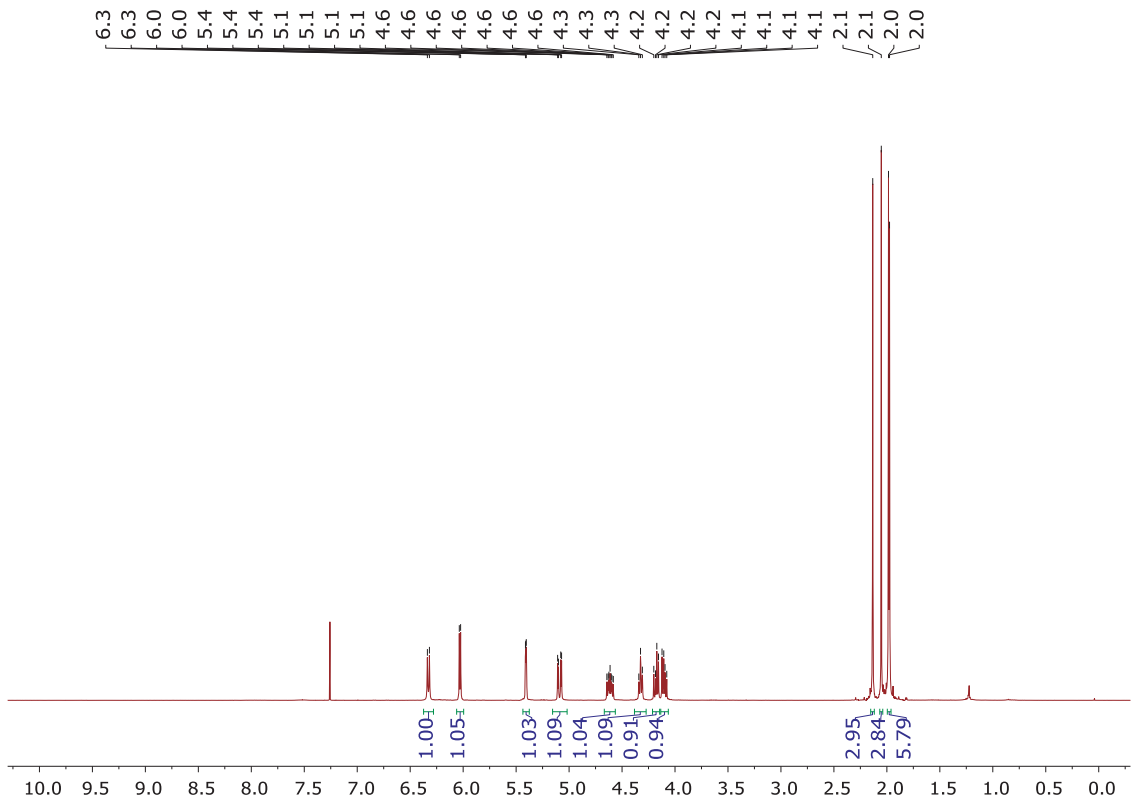



**$^{13}$C NMR 100 MHz in $CDCl_3$ registered at 298K**

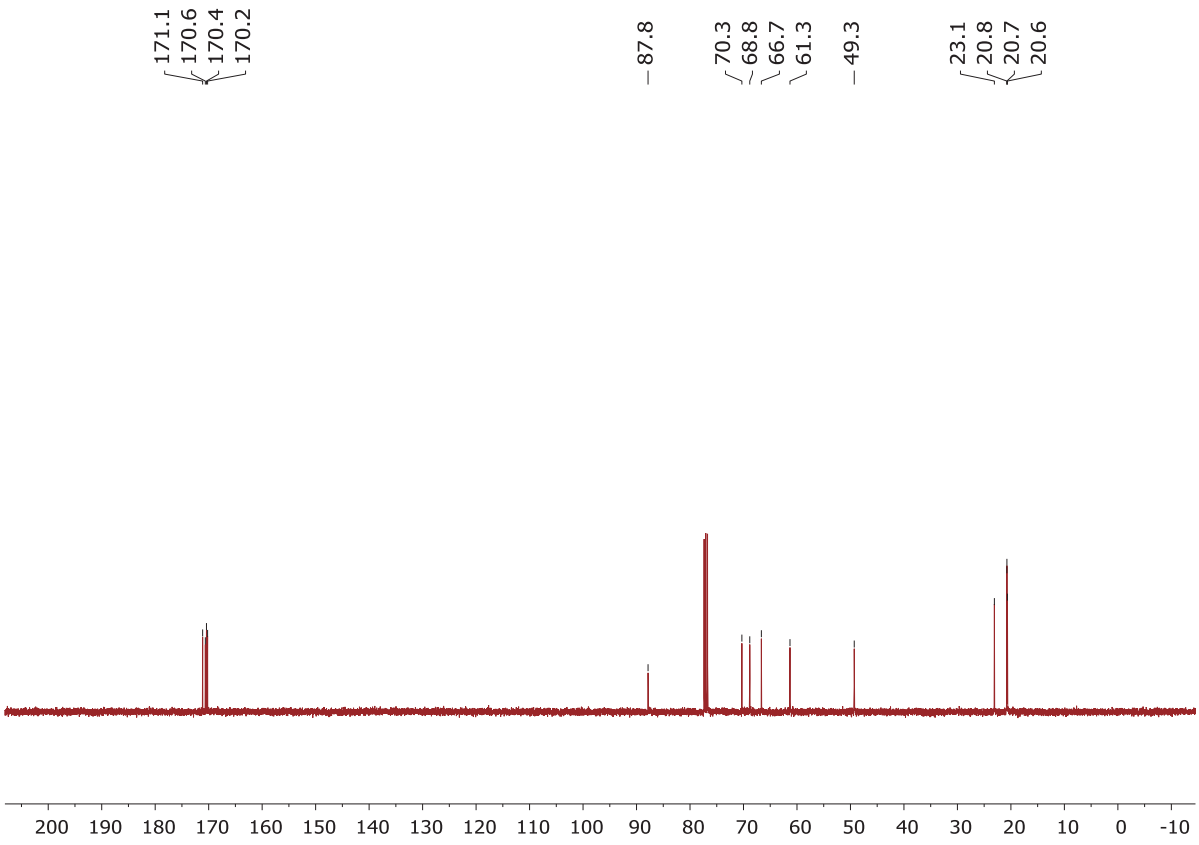


Compuesto **10** capítulo 6

**$^{1}$H NMR 400 MHz in $CDCl_3$**

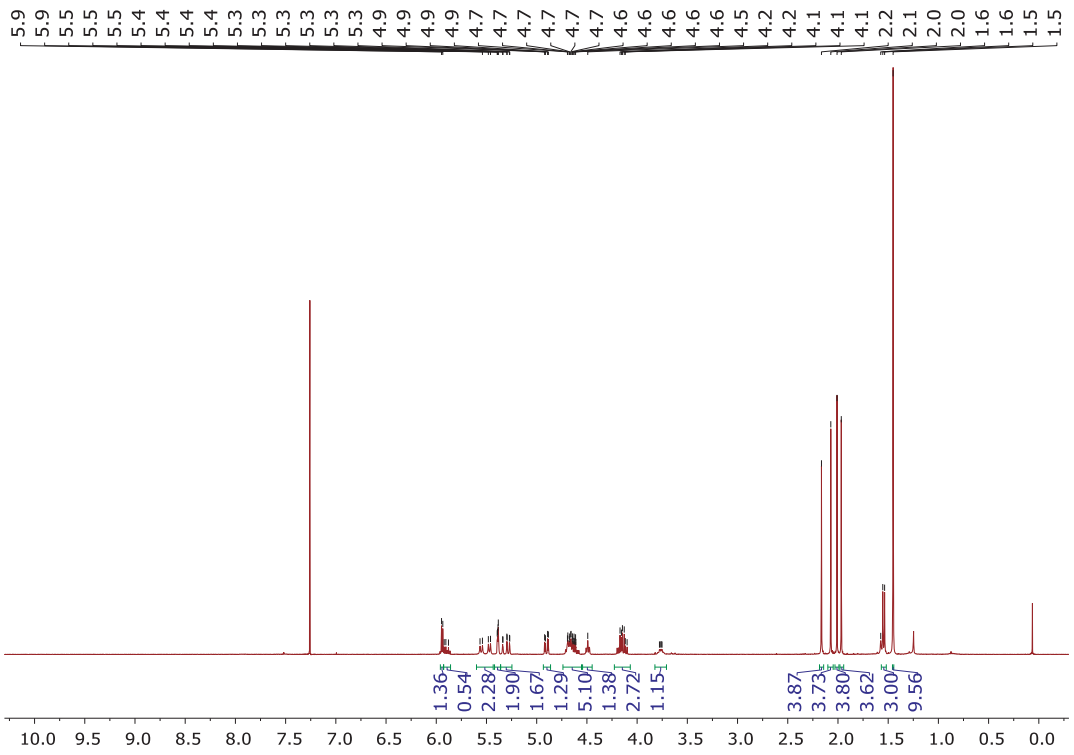



**$^{13}$C NMR 100 MHz in $CDCl_3$**

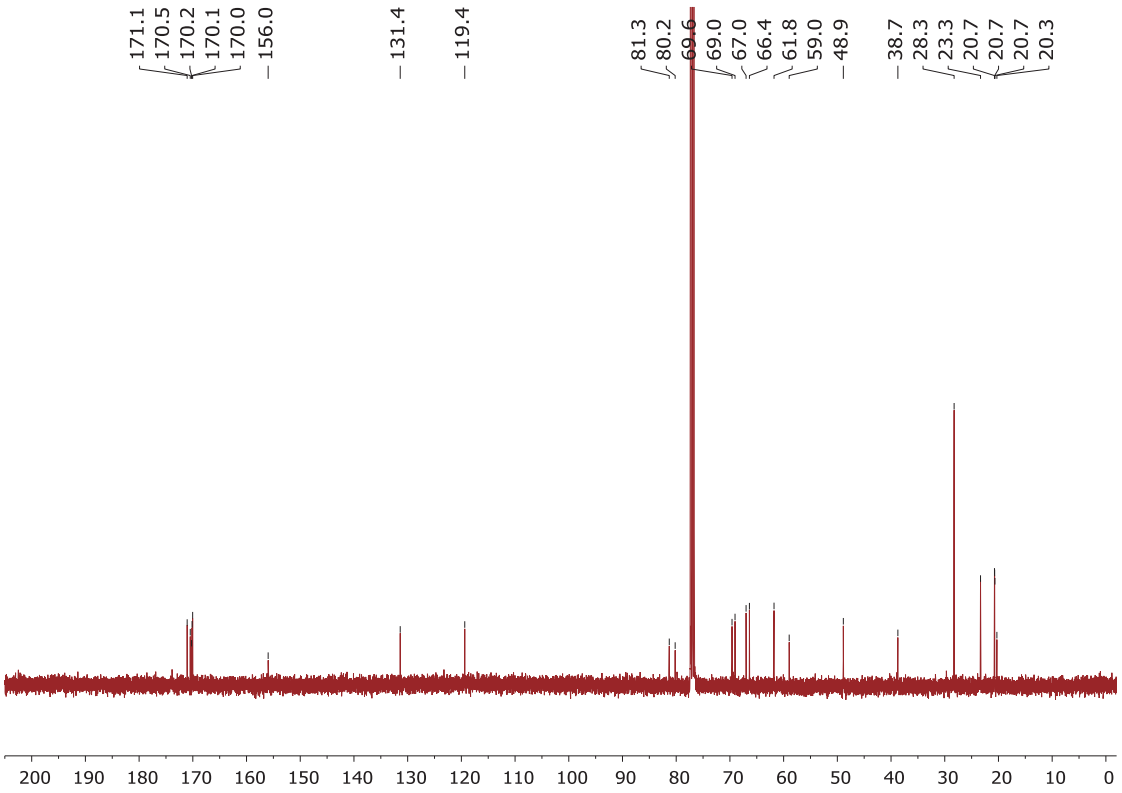


Compuesto **11** capítulo 6

**$^{1}$H NMR 400 MHz in $CDCl_3$**

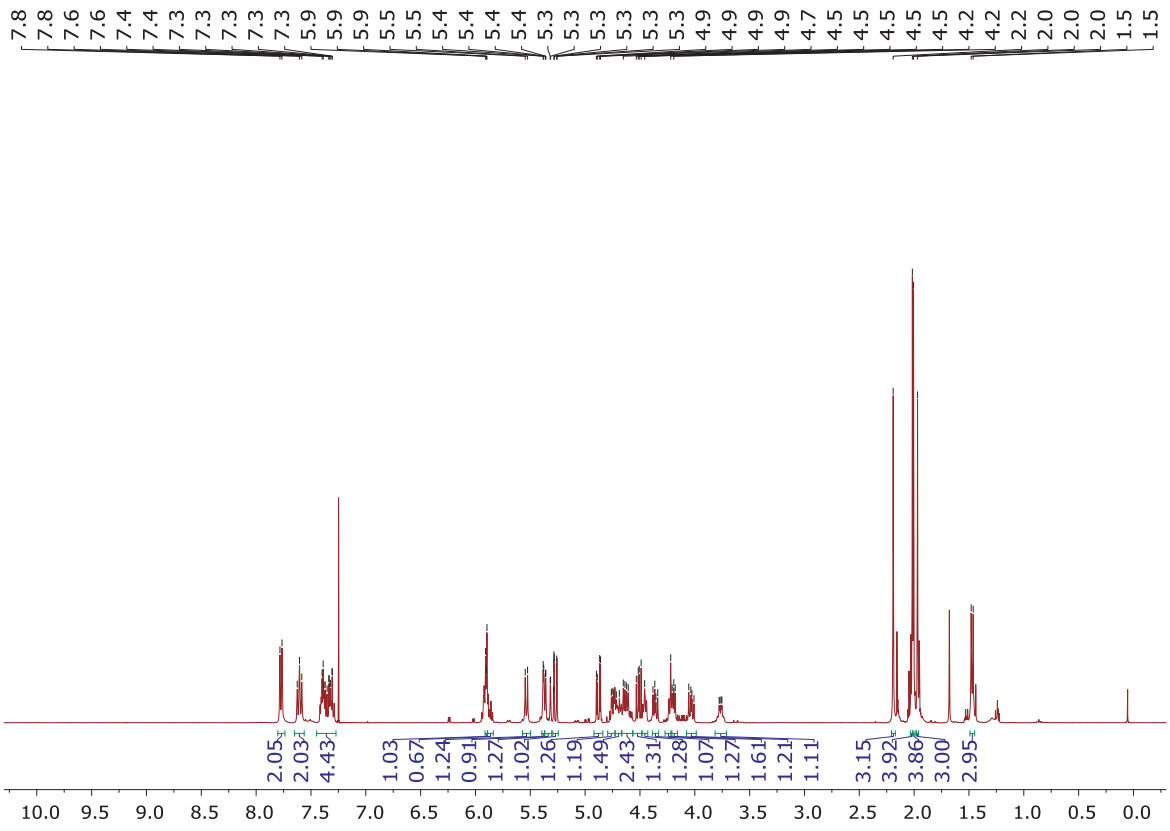


**$^{13}$C NMR 100 MHz in $CDCl_3$**

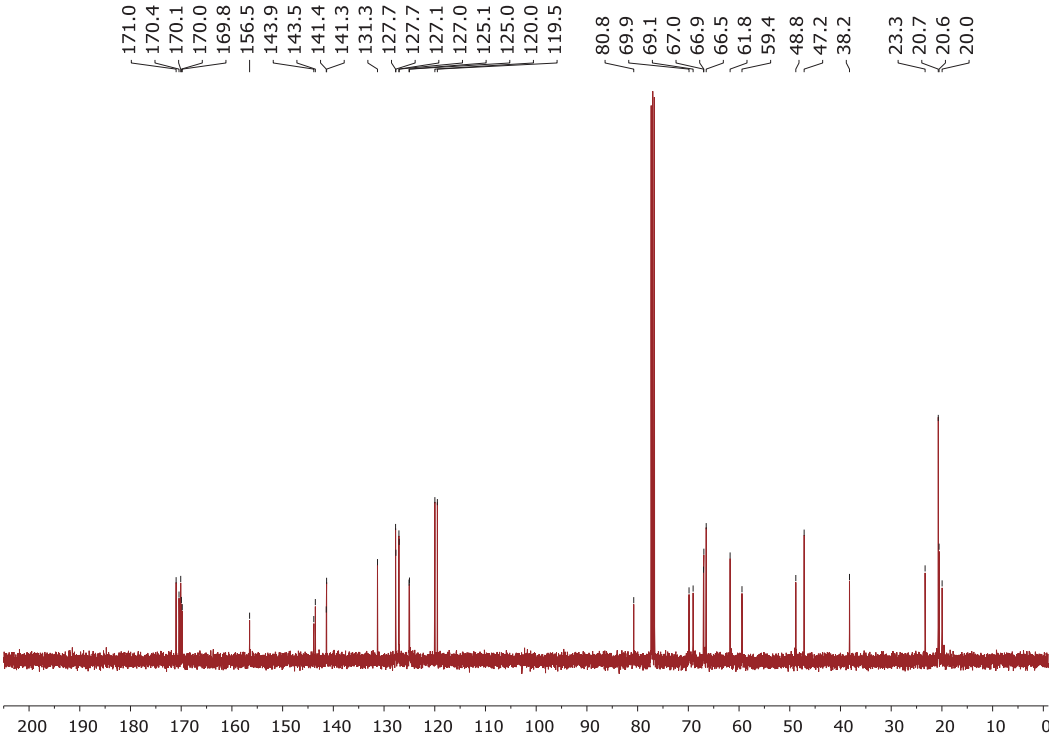


Compuesto **12** capítulo 6

**$^{1}$H NMR 400 MHz in $CDCl_3$**

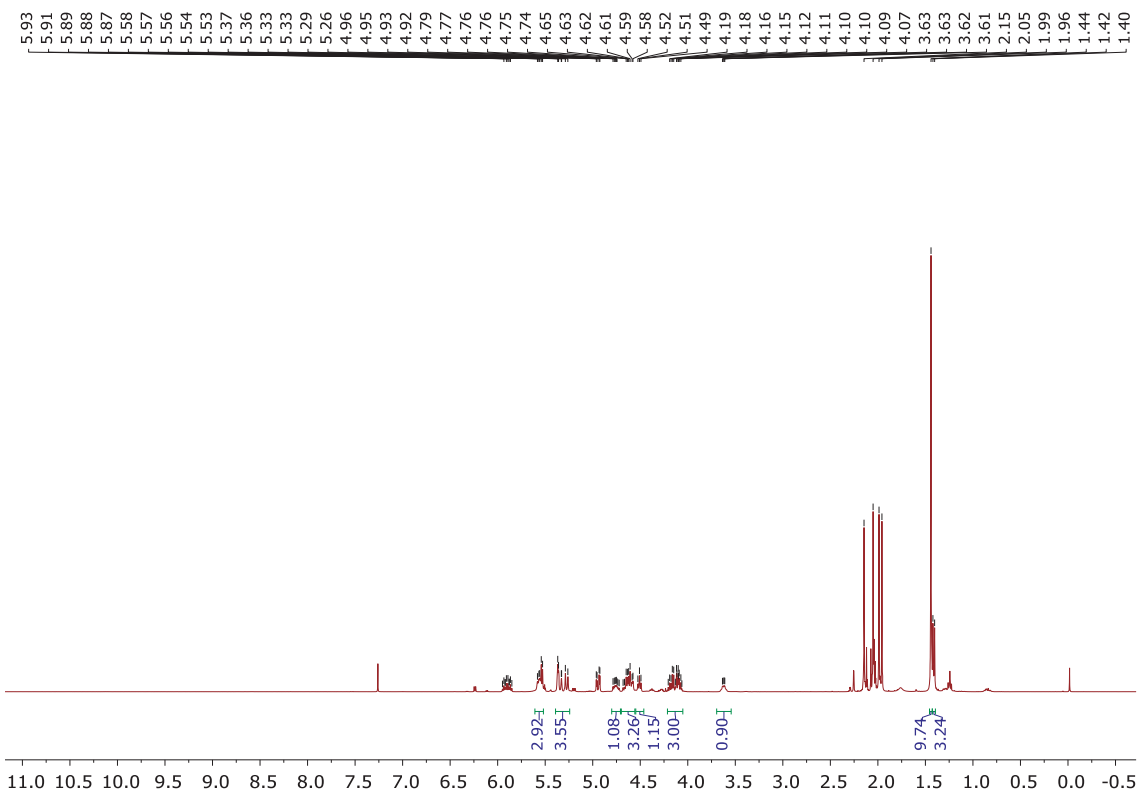


**$^{13}$C NMR 100 MHz in $CDCl_3$**

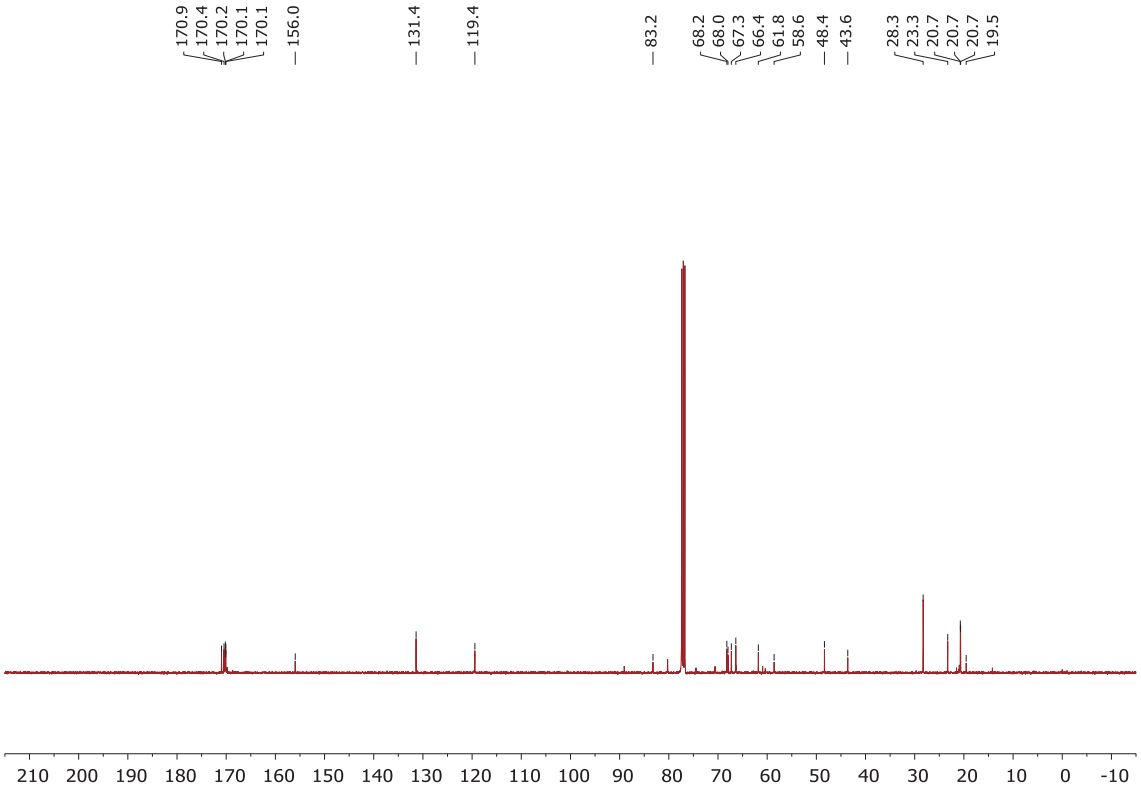


## Compuesto **14** capítulo 6

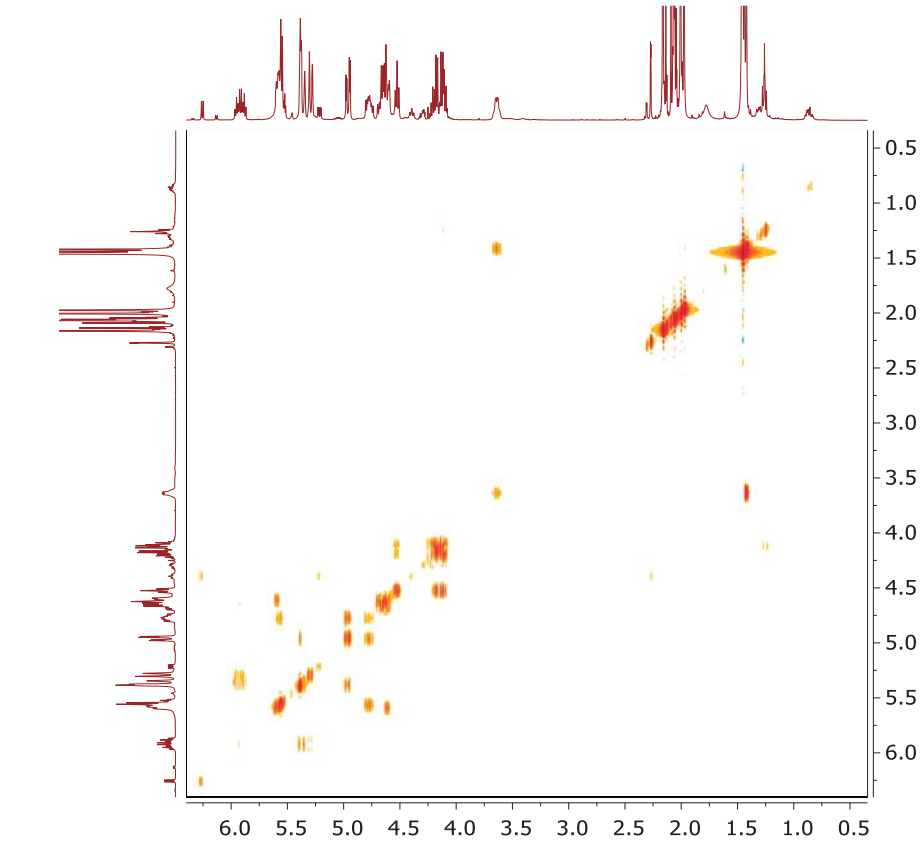
**$^{1}$H NMR 400 MHz in $CDCl_3$**



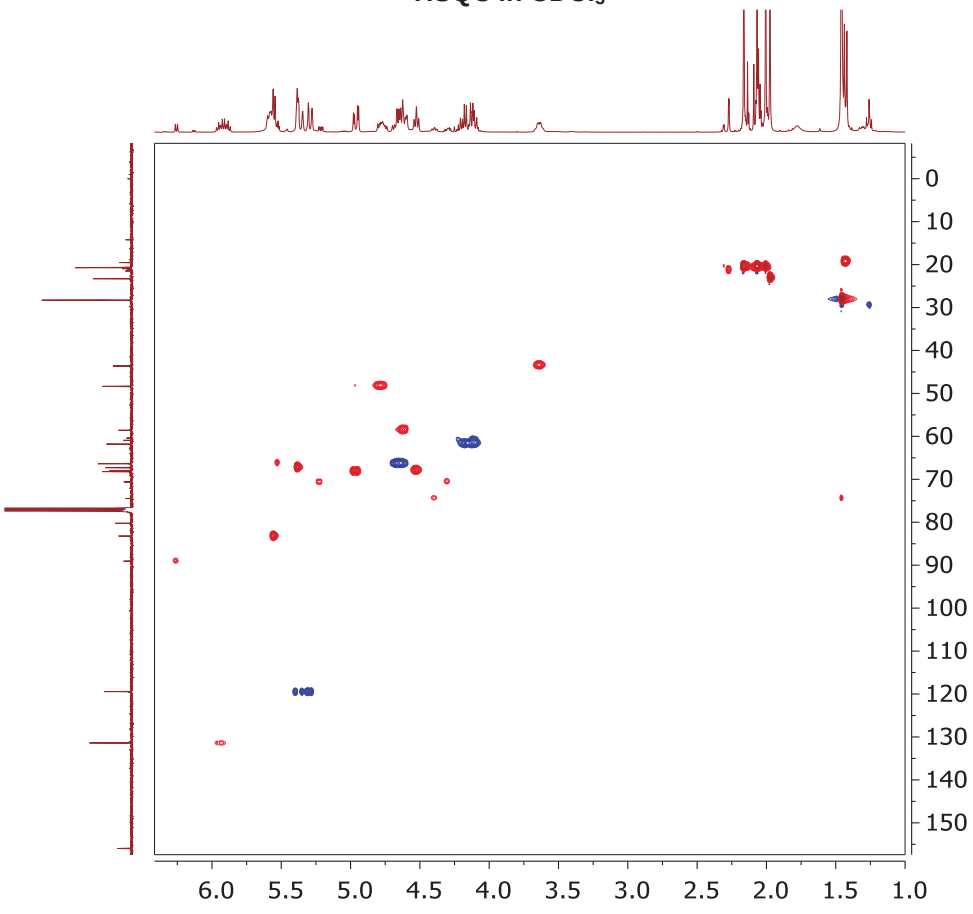

**$^{13}$C NMR 100 MHz in $CDCl_3$**



**COSY in $CDCl_3$ registered at 298K**

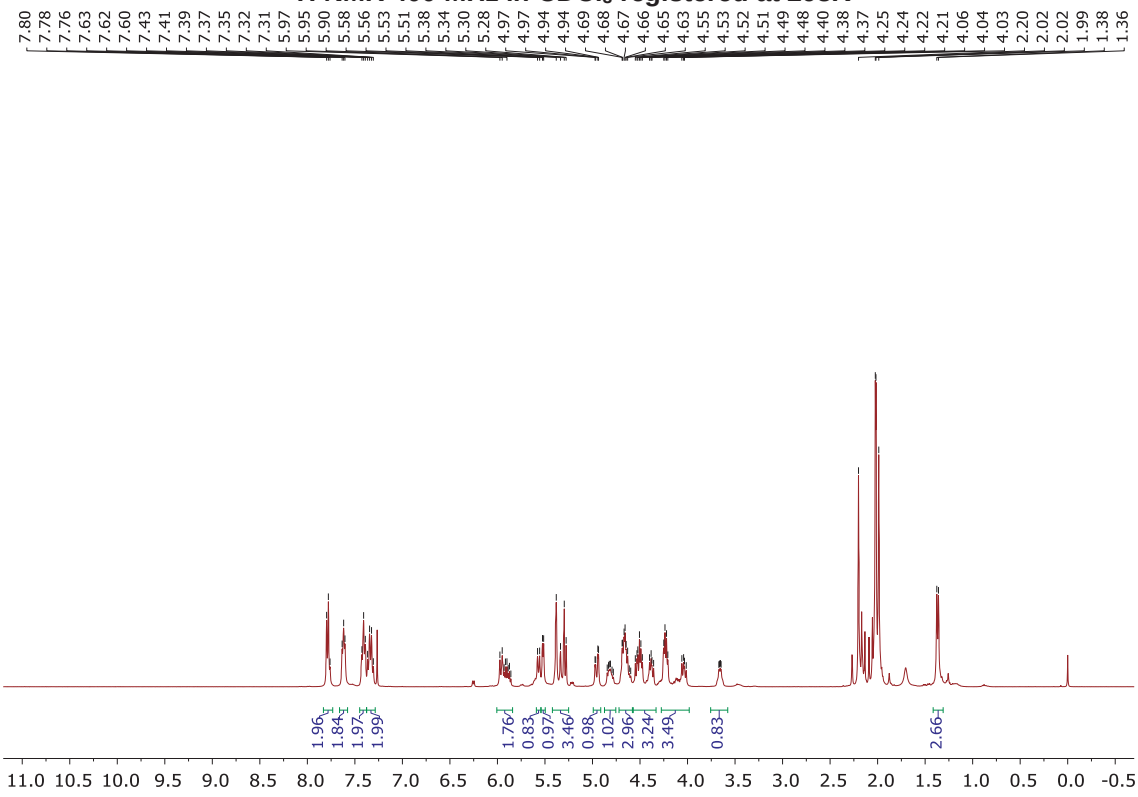



**HSQC in $CDCl_3$**

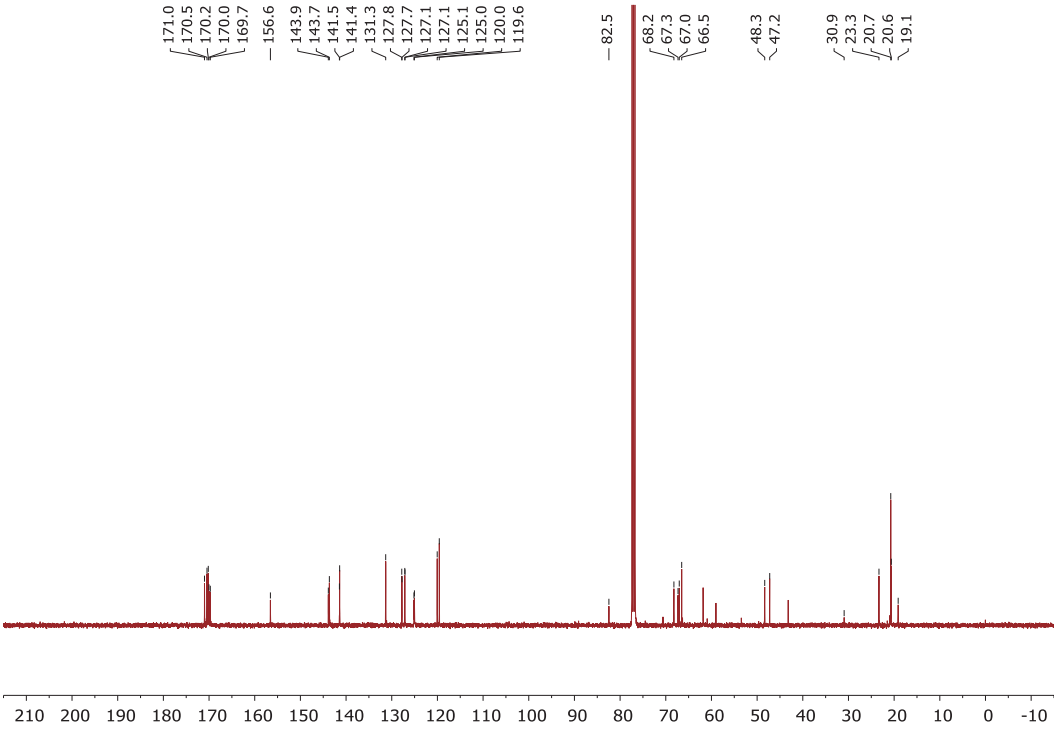


Compuesto **15** capítulo 6

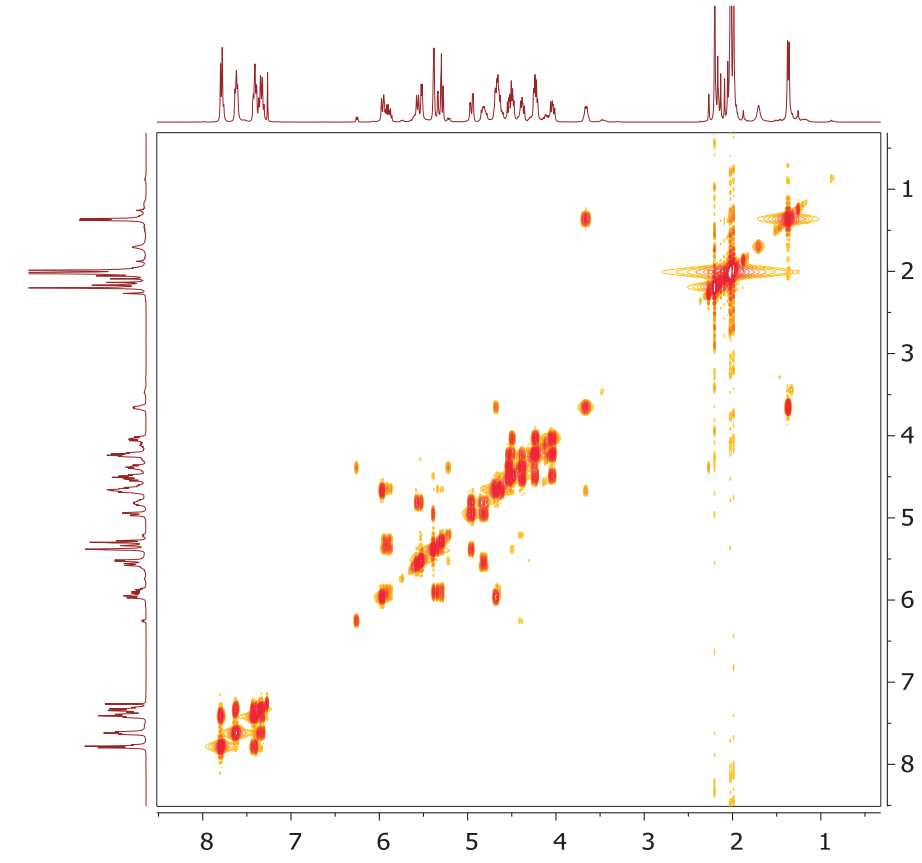
**$^{1}$H NMR 400 MHz in $CDCl_3$ registered at 298K**



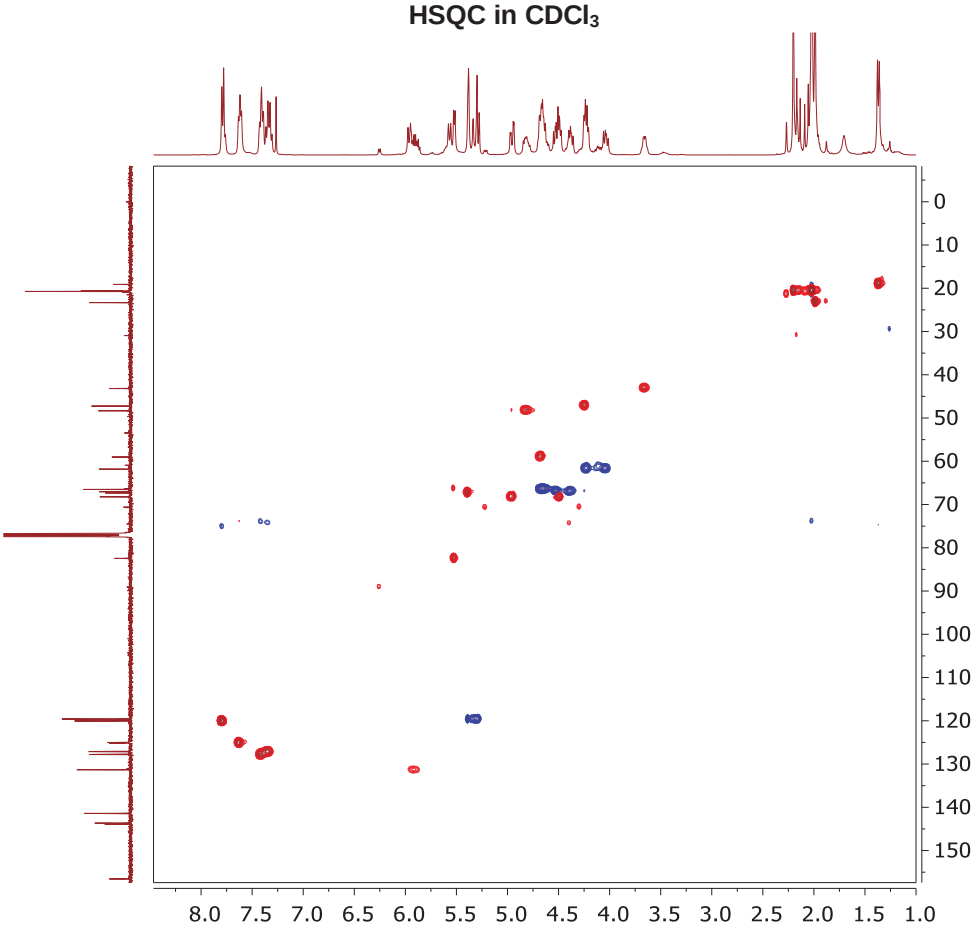
**$^{13}$C NMR 100 MHz in $CDCl_3$**



**COSY in $CDCl_3$**




**HSQC in $CDCl_3$**



Catalytic Mechanisms

DOI: 10.1002/anie.201402781

# Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide *N*-Acetylgalactosaminyltransferase 2**

*Erandi Lira-Navarrete, Javier Iglesias-Fernández, Wesley F. Zandberg, Ismael Compañón, Yun Kong, Francisco Corzana, B. Mario Pinto, Henrik Clausen, Jesús M. Peregrina, David J. Vocadlo, Carme Rovira,* and Ramon Hurtado-Guerrero**

***Abstract:*** *The retaining glycosyltransferase GalNAc-T2 is a member of a large family of human polypeptide GalNAc-transferases that is responsible for the post-translational modification of many cell-surface proteins. By the use of combined structural and computational approaches, we provide the first set of structural snapshots of the enzyme during the catalytic cycle and combine these with quantum-mechanics/molecular-mechanics (QM/MM) metadynamics to unravel the catalytic mechanism of this retaining enzyme at the atomic-electronic level of detail. Our study provides a detailed structural rationale for an ordered bi–bi kinetic mechanism and reveals critical aspects of substrate recognition, which dictate the specificity for acceptor Thr versus Ser residues and enforce a front-face $S_Ni$-type reaction in which the substrate N-acetyl sugar substituent coordinates efficient glycosyl transfer.*

*O*-GalNAc glycosylation[1] is by far the most complex and differentially regulated type of protein glycosylation, and it may also be the most abundant, with over 80% of all proteins passing through the secretory pathway predicted to be O-glycosylated.[2] One of the key enzyme isoforms that control human protein O-glycosylation is polypeptide GalNAc-transferase 2 (GalNAc-T2). This enzyme is part of a large family of retaining isoenzymes that transfer a GalNAc residue from UDP-GalNAc to Ser/Thr side chains and thereby initiate mucin-type or GalNAc-type protein O-glycosylation. How GalNAc-Ts target specific sites on proteins and glycoproteins, and how they catalyze O-GalNAc transfer is not understood. Although the structures of several GalNAc-Ts have been reported,[3] the lack of Michaelis complexes and ternary product complexes has limited mechanistic insight into the glycosyl-transfer reaction. Such information is required for the rational design of inhibitors of these potential drug targets.

Herein we describe the use of a combination of X-ray crystallography and quantum-mechanics/molecular-mechanics (QM/MM) metadynamics to unravel the catalytic mechanism of GalNAc-T2. The structures obtained provide the first structural snapshots of the enzyme in complex with its substrates and products during the catalytic cycle. The findings reveal mobile loops that act to bind and release the substrates and illuminate substrate acceptor preference for threonine over serine. Furthermore, QM/MM metadynamics simulations reveal a front-face reaction mechanism involving a short-lived oxocarbenium-ion-like intermediate.

To obtain a Michaelis complex of GalNAc-T2, we obtained crystals of GalNAc-T2 in complex with UDP-$Mn^{2+}$ (UDP = uridine diphosphate) and soaked them first in

[*] E. Lira-Navarrete,[+] Dr. R. Hurtado-Guerrero
Institute of Biocomputation and Physics of Complex Systems (BIFI)
University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit
Mariano Esquillor s/n, Campus Rio Ebro, Edificio I+D
Fundacion ARAID, Edificio Pignatelli 36 (Spain)
E-mail: rhurtado@bifi.es

J. Iglesias-Fernández,[+] Prof. C. Rovira
Departament de Química Orgànica and Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona
Martí i Franquès 1, Barcelona (Spain)
E-mail: c.rovira@ub.edu

Prof. C. Rovira
Institució Catalana de Recerca i Estudis Avançats (ICREA)
Passeig Lluís Companys 23, 08020 Barcelona (Spain)

Dr. W. F. Zandberg, Prof. B. M. Pinto, Prof. D. J. Vocadlo
Department of Chemistry and Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC (Canada)

I. Compañón, Dr. F. Corzana, Prof. J. M. Peregrina
Departamento de Química, Universidad de La Rioja
Centro de Investigación en Síntesis Química, Logroño (Spain)

Y. Kong, Prof. H. Clausen
Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and School of Dentistry
University of Copenhagen (Denmark)

[+] These authors contributed equally.

[**] We thank the Diamond Light Source (DLS) at Oxford, in particular for the use of beamlines ID24 (experiment number MX8035-14) and I04-1 (experiment number MX8035-17). We thank the Fundación Agencia Aragonesa para la Investigación y el Desarrollo (ARAID, Spain), the Ministerio de Economía y Competitividad (MEC, Spain, grants BFU2010-19504 and CTQ2011-25871), The Danish National Research Foundation (DNRF107), the Diputación General de Aragón (Grupo Protein Targets, B89), and the Generalitat de Catalunya (2009SGR-1309). We acknowledge the computer support, technical expertise, and assistance provided by the Barcelona Supercomputing Center—Centro Nacional de Supercomputación (BSC-CNS). Funding for the research leading to these results was also received from the Seventh Framework Programme (FP7/2007-2013) of the European Community under BioStruct-X (grant agreement No. 283570), and the Natural Sciences and Engineering Research Council of Canada.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201402781.

JOC The Journal of Organic Chemistry

Article

pubs.acs.org/joc

# Design of α-*S*-Neoglycopeptides Derived from MUC1 with a Flexible and Solvent-Exposed Sugar Moiety

Víctor Rojas-Ocáriz,[†,‡] Ismael Compañón,[†,‡] Carlos Aydillo,[†,§] Jorge Castro-López,[∥] Jesús Jiménez-Barbero,[#,∇,○] Ramón Hurtado-Guerrero,[∥,⊥] Alberto Avenoza,[†] María M. Zurbano,[†] Jesús M. Peregrina,[*,†] Jesús H. Busto,[*,†] and Francisco Corzana[*,†]

[†]Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, Madre de Dios 53, 26006 Logroño, Spain

[∥]BIFI, University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit, Mariano Esquillor s/n, Campus Rio Ebro, Edificio I+D, Zaragoza, Spain

[⊥]Fundación ARAID, 50018 Zaragoza, Spain

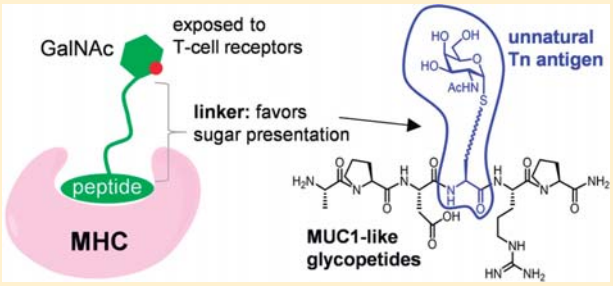
[#]Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia Building 801 A, 48160 Derio, Spain

[∇]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

[○]Department of Chemical and Physical Biology, Centro de Investigaciones Biológicas, CSIC Ramiro de Maeztu 9, 28040 Madrid, Spain

Supporting Information

**ABSTRACT:** The use of vaccines based on MUC1 glycopeptides is a promising approach to treat cancer. We present herein several sulfa-Tn antigens incorporated in MUC1 sequences that possess a variable linker between the carbohydrate (GalNAc) and the peptide backbone. The main conformations of these molecules in solution have been evaluated by combining NMR experiments and molecular dynamics simulations. The linker plays a key role in the modulation of the conformation of these compounds at different levels, blocking a direct contact between the sugar moiety and the backbone, promoting a helix-like conformation for the glycosylated residue and favoring the proper presentation of the sugar unit for molecular recognition events. The feasibility of these novel compounds as mimics of MUC1 antigens has been validated by the X-ray diffraction structure of one of these unnatural derivatives complexed to an anti-MUC1 monoclonal antibody. These features, together with potential lack of immune suppression, render these unnatural glycopeptides promising candidates for designing alternative therapeutic vaccines against cancer.




## INTRODUCTION

Chemistry-based approaches are of paramount importance to battle disease. In the quest of chemical weapons against cancer, MUC1 is a glycoprotein overexpressed in most tumors.[1−4] While in healthy cells the MUC1 backbone presents complex oligosaccharides, in tumor cells the peptide backbone is decorated with simple and truncated carbohydrates. As a consequence, different tumor-associated carbohydrate antigens (TACAs), such as the Tn determinant (α-*O*-GalNAc-Ser/Thr),[5] are exposed to the immune system and can be recognized by different antibodies.[6] For this reason, MUC1 derivatives are attracting great interest as a potential tool in developing therapeutic vaccines for the treatment of cancer.[4,7,8] However, to date, none of these vaccines have succeeded in clinical trials.[9] On one hand, natural TACAs are tolerated by the immune system.[10] To overcome this problem, chemical modifications of these antigens, generating non-natural determinants, have been proposed.[7,11−20] On the other hand, most glycan antigens are B-cell epitopes and produce only short-lived and low-affinity antibodies.[4] Induction of a robust immune response that produces high-affinity IgG antibodies is essential and may be achieved through initiation of a T-cell-dependent pathway.

A key step in this process is the presentation of the antigen through the major histocompatibility (MHC) molecule to T-cell receptors. In that context, recent studies indicate that glycopeptides can arbitrate classical MHC-mediated immune response.[21] The crystal structure of an MHC-I molecule with a glycopeptide containing a linker between the peptide and the sugar[21e] showed that while the aglycone part of the antigen binds to the MHC molecule, the carbohydrate moiety can facilitate the recognition of T-cell receptors (TCR) and therefore stimulate immune response (Figure 1a and 1b).

**Received:** April 13, 2016
**Published:** June 15, 2016





DOI: 10.1021/acs.joc.6b00833
J. Org. Chem. 2016, 81, 5929−5941

Immunogenic Stapled Proteins

International Edition: DOI: 10.1002/anie.201708847
German Edition: DOI: 10.1002/ange.201708847

# Oxetane Grafts Installed Site-Selectively on Native Disulfides to Enhance Protein Stability and Activity In Vivo

*Nuria Martínez-Sáez[+], Shuang Sun[+], Davide Oldrini, Pietro Sormanni, Omar Boutureira, Filippo Carboni, Ismael Compañón, Michael J. Deery, Michele Vendruscolo, Francisco Corzana, Roberto Adamo, and Gonçalo J. L. Bernardes**

***Abstract:*** *A four-membered oxygen ring (oxetane) can be readily grafted into native peptides and proteins through site-selective bis-alkylation of cysteine residues present as disulfides under mild and biocompatible conditions. The selective installation of the oxetane graft enhances stability and activity, as demonstrated for a range of biologically relevant cyclic peptides, including somatostatin, proteins, and antibodies, such as a Fab arm of the antibody Herceptin and a designed antibody DesAb-Aβ against the human Amyloid-β peptide. Oxetane grafting of the genetically detoxified diphtheria toxin $CRM_{197}$ improves significantly the immunogenicity of this protein in mice, which illustrates the general utility of this strategy to modulate the stability and biological activity of therapeutic proteins containing disulfides in their structures.*

The rational modification of the structure of peptides and proteins offers a wide range of opportunities for the modulation of their biological activity.[1] Many efforts have been made to develop strategies that induce such conformational changes and modulation. Towards this end, macrocyclization and stapling have emerged as useful tactics to chemically manipulate peptides and proteins, increasing their proteolytic stability, cell permeability, and producing changes in polarity, binding activity, and pharmacokinetic properties.[2] During the last years, different approaches have been developed for the covalent tethering of the side chains of natural or non-canonical amino acids.[2] Considering natural residues, cysteine (Cys) has been the residue of choice for stapling through alkylation,[3] arylation,[4] cycloaddition,[4b] and disulfide forming reactions both at native[5] or engineered[6] Cys residues. More recently, nitrogen arylation has also been shown to be a useful strategy for macrocyclization of lysine residues on peptides.[7] Otherwise, efficient macrocyclization of linear peptides through the formation of an oxadiazole has also been reported.[8] However, a large number of stapling/macrocyclization/re-bridging strategies consist of the introduction of non-canonical amino acids and their subsequent ligation by ring-closing metathesis,[9] lactamization,[10] or cycloaddition reactions.[2c,4b] Common to many of these strategies is either the requirement for complicated orthogonal protection procedures, sequence engineering, the appendage of bulky/constrained linkers between the two residues, or the use of organic solvents. These conditions have limited, for instance, the application of such methods for the stapling of residues on intact, full-length proteins to impart structural conformational constraints leading to enhanced stability and activity. Thus, there remains a need for simple and robust strategies for stapling native peptides and proteins.

Herein we report a method for site-selective peptide and protein stapling through the bis-alkylation of the sulfhydryl side chain of Cys residues resulting from disulfide reduction, using commercially available 3,3-bis(bromomethyl)oxetane **1** (Figure 1). Oxetanes have become common motifs in drug design due to their ability to modulate parameters including solubility, basicity, lipophilicity, and metabolic stability.[11] While there are examples of the modification of small peptides with oxetanes,[12] their incorporation and modulation of the structure and activity of complex biomolecules[13] remain mostly unexplored.

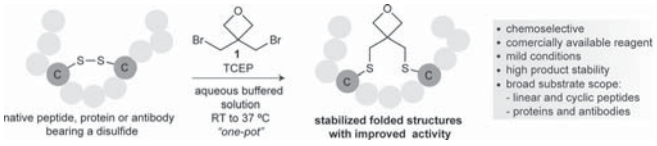



***Figure 1.*** Stabilization of folded structures of peptides and proteins through bis-alkylation of Cys residues present in the form of a native disulfide using an oxetane graft.

[*] Dr. N. Martínez-Sáez,[+] S. Sun,[+] Dr. P. Sormanni, Dr. O. Boutureira, Prof. M. Vendruscolo, Dr. G. J. L. Bernardes
Department of Chemistry, University of Cambridge
Lensfield Road, CB2 1EW Cambridge (UK)
E-mail: gb453@cam.ac.uk
Dr. G. J. L. Bernardes
Instituto de Medicina Molecular, Faculdade de Medicina
Universidade de Lisboa
Avenida Professor Egas Moniz, 1649-028 Lisboa (Portugal)
E-mail: gbernardes@medicina.ulisboa.pt
D. Oldrini, F. Carboni, Dr. R. Adamo
GSK Vaccines
Via Fiorentina 1, 53100 Siena (Italy)
I. Compañón, Dr. F. Corzana
Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja
26006 Logroño (Spain)
M. J. Deery
Cambridge Centre for Proteomics, Cambridge Systems Biology Centre, Department of Biochemistry, University of Cambridge
Tennis Court Road, Cambridge CB2 1QR (UK)
[+] These authors contributed equally to this work.
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201708847.

A requirement for a direct method to graft Cys residues present as disulfides on proteins is the compatibility of the reagent with a reducing agent, such as tris(2-carboxyethyl)-phosphine (TCEP). Importantly, we found that a model pentapeptide bearing two Cys residues reacted with **1** in the presence of TCEP to afford the corresponding stapled cyclic peptide in 75% yield (see the Supporting Information, Figures S2–S5, S34–S39, 48–49 for characterization and discussion of structural features). We then explored the use of **1** to graft two cyclic and biologically relevant peptides: somatostatin **2** and its analogue octreotide **4**, which can be used for imaging and treating neuroendocrine tumors.[14] Peptides **2** and **4** were reacted simultaneously with TCEP and oxetane **1** in a 1:9 mixture of DMF/$H_2O$ at 25 °C for 12 h. After HPLC purification, the grafted cyclic peptides **3** and **5** were obtained in 76% and 81% yield, respectively (Figure 2a). The affinity of these derivatives to the natural

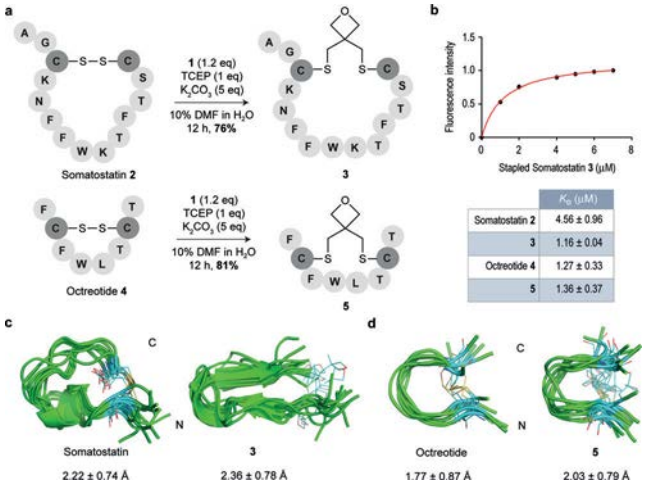



***Figure 2.*** a) Stapling of disulfide-containing cyclic peptides **2** and **4**. b) Binding affinity studies. $K_D$ values were determined by tryptophan fluorescence spectroscopy. c),d) Structural ensembles obtained by 0.5 μs MD simulations. The peptide backbone is in green. Carbon atoms of Cys residues as well as of the oxetane moiety are in cyan. The numbers indicate the root-mean-square deviation (RMSD) for heavy-atom superimposition of the backbone with respect to the average structure.

somatostatin receptor 2 (SSTR2) was experimentally determined by tryptophan fluorescence spectroscopy (Supporting Information, Figure S1). While octreotide **4** and surrogate peptide **5** showed a similar affinity against this receptor, grafted somatostatin **3** displayed improved binding properties, showing a 4-fold enhancement in $K_D$ value (Figure 2b). The improvement of binding activity is a considerable advantage of the incorporation of the oxetane graft when compared, for instance, with the recently reported methylene thioacetal that led to a decrease in binding affinity to SSTR2.[3b] Interestingly, 0.5 μs MD simulations performed on these derivatives in explicit water and using ff14SB amber force-field[15] suggested that octreotide **4** and its stapled derivative **5** presented a similar conformational behavior in solution (Figure 2d; Supporting Information, Figure S8),[16] displaying an equilibrium between antiparallel β-sheet structures and conformations in which the C-terminal residues form a $3_{10}$ helix-like fold, as reported in DMSO solution. In contrast, stapled somatostatin **3** was more rigid and displayed a more defined conformation in solution than **2** (Figure 2c; Supporting Information, Figures S6 and S7).[17] In fact, **3** showed a closely related β-sheet arrangement in solution stabilized by a typical hydrogen bond network, which is apparently ideal for a more efficient binding to the receptor. Finally, analysis of the stability of **3** and **5** both in human plasma as well in the presence glutathione (GSH) showed that the oxetane grafted peptides remain intact under these conditions (Supporting Information, Figures S13–S18).

To initially test the potential of using this one-pot, site-selective bis-alkylation oxetane stapling method directly on proteins, we chose thioredoxin (Trx) as a model protein that features a naturally occurring, solvent-exposed disulfide bond. We could reduce and staple the disulfide bond in a straightforward manner through selective bis-alkylation with **1** in the presence of TCEP and 10% DMF in sodium phosphate buffer at pH 8.5. Complete conversion was achieved after 24 h at 37 °C, as confirmed by HPLC-MS analysis (Supporting Information, Figures S19 and S20). Furthermore, analysis of the CD spectra of the native and stapled Trx-**1** (Supporting Information, Figure S22) indicated that both molecules present very similar conformational preferences in solution. Although this is supported by MD simulations performed on both proteins in explicit water, the calculations indicate a small increase in flexibility for the peptide backbone of stapled Trx-**1** (Supporting Information, Figure S11). This result may be explain attending to the greater S–S distance in Thrx-**1** when compared to the native Trx (4.18 and 2.04 Å, respectively). Finally, we confirmed the suppression of Trx redox activity[18] through the selective and covalent disulfide stapling (Supporting Information, Figure S21).

Next, we demonstrated the utility of the oxetane graft to build stapled antibodies. First, the exposed disulfide bond tethering the heavy and light chains of a Fab fragment of Herceptin (Fab-Her), an antibody currently used to treat Her2+ breast cancer patients,[19] was readily stapled using **1** under aqueous buffered conditions in the presence of TCEP at pH 8.5 and at 37 °C (Figures 3a,b; Supporting Information, Figures S23 and S24). The oxetane stapled Fab-Her-**1**, unlike the disulfide native antibody, was stable under reducing conditions and in human plasma (Supporting Information, Figures S25 and S26). This stability is a key aspect of antibody therapeutics design as thiol-exchange reactions in plasma lowers efficacy and adds side-toxicity.[20] Importantly, a relatively small but significant increase in binding affinity to the Her2 receptor, as determined by bio-layer interferometry (BLI) experiments (Figure 3d; Supporting Information, Figure S27), was observed for Fab-Her-**1** when compared with the native antibody. Next, we extended our stapling strategy to the antibody DesAb-A$\beta_{3\text{-}9}$, which was designed to target the region 3–9 of human Amyloid-β (Aβ42) peptide, the aggregation of which is a hallmark of Alzheimer's disease.[21] This antibody features a challenging, hindered intra-domain disulfide typical of VH domains. Of note, complete conversion into the oxetane grafted antibody DesAb-A$\beta_{3\text{-}9}$-**1** was

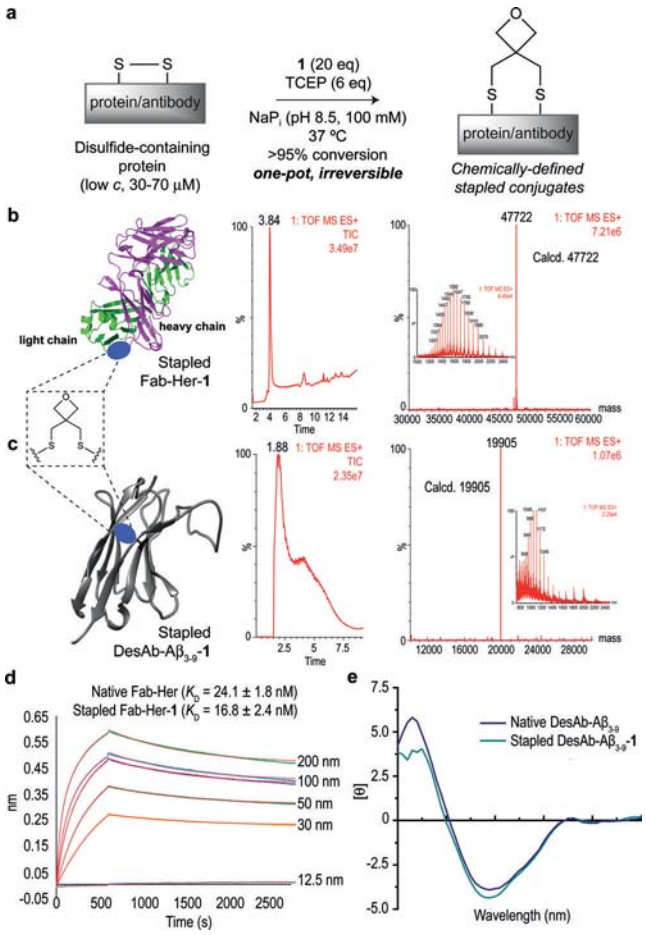



***Figure 3.*** a) Representation of disulfide stapling of native antibody sequences using an oxetane graft. b),c) Total ion chromatogram, combined ion series and deconvoluted mass spectrum reconstructed from the ion series using the MaxEnt algorithm. b) Fab-Her-**1** (pdb ID Fab-Her: 1N8Z) and c) DesAb-Aβ$_{3-9}$-**1** (3D model of DesAb-Aβ$_{3-9}$ generated using ABodyBuilder, see main text). d) BLI and fit curves obtained for Fab-Her-**1**, together with the derived $K_D$ constants for Fab-Her and Fab-Her-**1**. e) CD spectra of DesAb-Aβ$_{3-9}$ and DesAb-Aβ$_{3-9}$-**1**.

achieved using our method (Figure 3c; Supporting Information, Figures S28 and S29). Owing to the fact that the disulfide is deeply buried, an excess of TCEP (40 equiv) and longer reaction times were required (Supporting Information). Unlike the reduced antibody that readily reacts with thiol-specific Elman's reagent, the stapled DesAb-Aβ$_{3-9}$-**1** did not react suggesting complete consumption of the reduced Cys during stapling (Supporting Information). Finally, analysis of secondary structural content by CD showed no significant differences between the original and stapled antibodies (Figure 3e). MD simulations performed on a 3D model of DesAb-Aβ$_{3-9}$, previously generated using ABodyBuilder,[22] and on the stapled derivative DesAb-Aβ$_{3-9}$-**1**, suggest that, although the 3D structure is maintained upon the chemical modification, the oxetane motif provokes a slightly increase in the degree of flexibility (Supporting Information, Figure S31). Collectively, these data demonstrate the suitability of the oxetane motif to staple solvent accessible disulfide bonds on proteins with minimal secondary structure alterations.

To demonstrate the practical application of our method to therapeutic proteins, we investigated the effects of the selective introduction of the oxetane staple into the genetically detoxified diphtheria toxin CRM$_{197}$, which features four Cys residues in the form of two disulfides. Recently, it has been shown that antibodies against CRM$_{197}$ neutralized diphtheria toxin in HIV infected young individuals.[23] Furthermore, CRM$_{197}$ is a clinically used carrier in many glycoconjugate vaccines.[24] Previous structural studies showed that only the disulfide C186-C201 connecting the fragment A (C domain) and B (T/R domain) of CRM$_{197}$ is selectively reduced in the presence of the highly hindered C461–C471 disulfide upon treatment with dithiothreitol.[24b] Addition of a slight excess of TCEP to CRM$_{197}$ under aqueous buffered conditions at pH 8.5 and 37 °C, followed by an excess of **1**, led to the introduction of one oxetane graft (Figure 4a,b and the Supporting Information, Figure S32 for mass spectrometry analysis), presumably at C186–C201 according to our previous findings.[25] The impact of the installation of the oxetane moiety into CRM$_{197}$ on its structure and thermal stability was studied by CD and differential scanning calorimetry (DSC) analysis, respectively, and compared with the native protein (Figure 4c–e). We found that both the far and near UV CD spectra of CRM$_{197}$-**1** were nearly identical to those of CRM$_{197}$, which indicates that the 3D structure is preserved upon the chemical stapling. The DSC curves also corroborate this finding. Although CRM$_{197}$-**1** exhibited a broader DSC peak when compared to the sharp change in

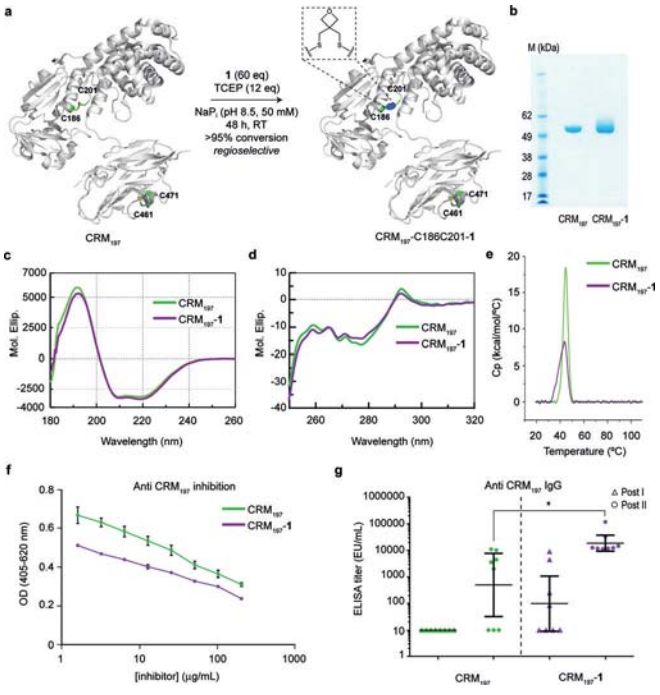



***Figure 4.*** Enhancing the immunogenicity of a protein carrier through disulfide oxetane stapling. a) The functional stapling of CRM$_{197}$ (pdb ID CRM$_{197}$: 4AE0) with **1**. b) SDS-page of native and stapled CRM$_{197}$-**1**. c) Far UV CD spectrum. d) Near UV CD spectrum. e) DCS analysis. f) Competition of anti-CRM$_{197}$ serum binding to the protein with CRM$_{197}$ and its stapled form as inhibitors. g) Anti-CRM$_{197}$ IgG levels of CRM$_{197}$ and CRM$_{197}$-**1** after first and second boost immunizations in mice, 2 weeks apart.

heat capacity, in the range of 40–55 °C for $CRM_{197}$, both proteins present an identical transition midpoint ($T_m$) of 46 °C.

To evaluate the biological effects resulting from the introduction of the oxetane staple into $CRM_{197}$, we first assayed the capacity of competing with the binding of anti-$CRM_{197}$ serum to the proteins. We found that the stapled $CRM_{197}$-**1** induces an inhibition that was slightly lower compared to the unmodified protein (Figure 4 f). In contrast, a much better inhibition of the binding to a commercial anti-diphtheria toxoid human recombinant monoclonal antibody was observed for $CRM_{197}$-**1** compared to the unmodified protein (Supporting Information, Figure S33a). To ascertain that protein epitopes were not impaired by the chemical modification of the disulfide bond, groups of 8 BALB/c mice were immunized with both unmodified $CRM_{197}$ and the stapled $CRM_{197}$-**1** (Figure 4 g). Remarkably, these in vivo experiments demonstrated that $CRM_{197}$-**1** induced a statistically significant higher level of anti-protein antibodies respect to the unmodified protein. The antibodies generated by $CRM_{197}$-**1** had a threefold higher avidity for the protein antigen compared to the anti-$CRM_{197}$ serum (avidity index = $0.8 \pm 0.4$ M for $CRM_{197}$-**1** vs. $0.3 \pm 0.1$ M for $CRM_{197}$), as determined by ELISA using thiocyanate elution (Supporting Information, Figure S33b). These data, together with 200 ns MD simulations performed on both proteins (Supporting Information, Figure S12), suggest that the oxetane bridging of the disulfide bond does not cause relevant structural modifications on the protein but results in improved immunogenic activity in vivo, most likely through chemical stabilization of the antigen against proteases and/or other degradation factors.

In summary, we have presented an efficient method for oxetane stapling of Cys residues present as native disulfides on peptides and proteins under mild and biocompatible aqueous conditions. The four-membered oxetane ring has an ideal distance to enable direct stapling of native disulfides on several protein scaffolds, including antibodies. This approach is however dependent on solvent accessibility of the disulfide within the protein of interest. Furthermore, and unlike current protocols, this method does not require prior sequence engineering neither purification after the disulfide reduction step. The selective installation of the oxetane motif enables stabilization of folded structures and results in disulfide-grafted products with enhanced bioactivity that are stable under biological conditions. We demonstrate the value of oxetane graft installation on protein through the regioselective disulfide stapling of the protein carrier $CRM_{197}$ that showed a significant increase in its immunogenicity in vivo. Because many therapeutic proteins feature Cys residues in the form of disulfide bonds, we anticipate that their direct modulation through oxetane grafting can, in principle, be used as a general strategy to enhance their in vivo stability and to fine-tune their structure for optimal pharmacokinetics and activity.

## Acknowledgements

We thank FCT Portugal, the EU (Marie Curie IEF to O.B.; Marie Sklodowska-Curie ITN *GlycoVax* to G.J.L.B. and R.A.), Cambridge Trust and China Scholarship Council (PhD studentship to S.S.), MINECO (CTQ2015-67727-R and Salvador de Madariaga mobility grant to F.Corzana) and the EPSRC for funding. I.C. thanks Universidad de La Rioja for a FPI fellowship. We thank Dr. Vijay Chudasama for providing the Fab-Her antibody, Dr. Werner Pansegrau (GSK Vaccines) for acquiring CD spectrum and DSC profile, and Dr. Francesco Aprile for providing the construct of the single domain antibody DesAb-$A\beta_{3\text{-}9}$. G.J.L.B. is a Royal Society URF and the recipient of a ERC Grant (*TagIt*).

## Conflict of interest

N.M.S., S.S., O.B., F.Corzana, and G.J.L.B. are listed as inventors on a pending patent application related to the technology described in this work. F.Carboni, D.O., and R.A. are employees of GSK companies.

**Keywords:** antibodies · disulfides · immunogenic proteins · oxetanes · stapling

**How to cite:** *Angew. Chem. Int. Ed.* **2017**, *56*, 14963–14967
*Angew. Chem.* **2017**, *129*, 15159–15163

[1] a) J. E. Bock, J. Gavenonis, J. A. Kritzer, *ACS Chem. Biol.* **2013**, *8*, 488–499; b) N. Krall, F. P. da Cruz, O. Boutureira, G. J. L. Bernardes, *Nat. Chem.* **2016**, *8*, 103–113.

[2] a) M. Góngora-Benítez, J. Tulla-Puche, F. Albericio, *Chem. Rev.* **2014**, *114*, 901–926; b) C. J. White, A. K. Yudin, *Nat. Chem.* **2011**, *3*, 509–524; c) Y. H. Lau, P. De Andrade, S. T. Quah, M. Rossmann, L. Laraia, N. Sköld, T. J. Sum, P. J. E. Rowling, T. L. Joseph, C. Verma, M. Hyvönen, L. S. Itzhaki, A. R. Venkitaraman, C. J. Brown, D. P. Lane, D. R. Spring, *Chem. Sci.* **2014**, *5*, 1804–1809.

[3] a) N. Assem, D. J. Ferreira, D. W. Wolan, P. E. Dawson, *Angew. Chem. Int. Ed.* **2015**, *54*, 8665–8668; *Angew. Chem.* **2015**, *127*, 8789–8792; b) C. M. B. K. Kourra, N. Cramer, *Chem. Sci.* **2016**, *7*, 7007–7012.

[4] a) A. M. Spokoyny, Y. Zou, J. J. Ling, H. Yu, Y.-S. Lin, B. L. Pentelute, *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949; b) S. P. Brown, A. B. Smith, *J. Am. Chem. Soc.* **2015**, *137*, 4034–4037; c) S. Kalhor-Monfared, M. R. Jafari, J. T. Patterson, P. I. Kitov, J. J. Dwyer, J. M. Nuss, R. Derda, *Chem. Sci.* **2016**, *7*, 3785–3790.

[5] S. L. Kuan, T. Wang, T. Weil, *Chem. Eur. J.* **2016**, *22*, 17112–17129.

[6] T. Liu, Y. Wang, X. Luo, J. Li, S. A. Reed, H. Xiao, T. S. Young, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5910–5915.

[7] G. Lautrette, F. Touti, H. G. Lee, P. Dai, B. L. Pentelute, *J. Am. Chem. Soc.* **2016**, *138*, 8340–8343.

[8] J. R. Frost, C. C. G. Scully, A. K. Yudin, *Nat. Chem.* **2016**, *8*, 1105–1111.

[9] a) C. E. Schafmeister, J. Po, G. L. Verdine, *J. Am. Chem. Soc.* **2000**, *122*, 5891–5892; b) H. E. Blackwell, R. H. Grubbs, *Angew. Chem. Int. Ed.* **1998**, *37*, 3281–3284; *Angew. Chem.* **1998**, *110*, 3469–3472.

[10] N. E. Shepherd, H. N. Hoang, G. Abbenante, D. P. Fairlie, *J. Am. Chem. Soc.* **2005**, *127*, 2974–2983.

[11] J. A. Burkhard, G. Wuitschik, M. Rogers-Evans, K. Müller, E. M. Carreira, *Angew. Chem. Int. Ed.* **2010**, *49*, 9052 – 9067; *Angew. Chem.* **2010**, *122*, 9236 – 9251.
[12] a) M. McLaughlin, R. Yazaki, T. C. Fessard, E. M. Carreira, *Org. Lett.* **2014**, *16*, 4070 – 4073; b) N. H. Powell, G. J. Clarkson, R. Notman, P. Raubo, N. G. Martin, M. Shipman, *Chem. Commun.* **2014**, *50*, 8797 – 8800.
[13] O. Boutureira, N. Martínez-Sáez, K. M. Brindle, A. A. Neves, F. Corzana, G. J. L. Bernardes, *Chem. Eur. J.* **2017**, *23*, 6483 – 6489.
[14] E. M. Wolin, *Gastrointest. Cancer Res.* **2012**, *5*, 161 – 168.
[15] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696 – 3713.
[16] G. Melacini, Q. Zhu, M. Goodman, *Biochemistry* **1997**, *36*, 1233 – 1241.
[17] A. Anoop, S. Ranganathan, B. D. Dhaked, N. N. Jha, S. Pratihar, S. Ghosh, S. Sahay, S. Kumar, S. Das, M. Kombrabail, K. Agarwal, R. S. Jacob, P. Singru, P. Bhaumik, R. Padinhateeri, A. Kumar, S. K. Maji, *J. Biol. Chem.* **2014**, *289*, 16884 – 16903.
[18] A. Holmgren, *Annu. Rev. Biochem.* **1985**, *54*, 237 – 271.
[19] C. A. Hudis, *N. Engl. J. Med.* **2007**, *357*, 39 – 51.
[20] a) V. Chudasama, A. Maruani, S. Caddick, *Nat. Chem.* **2016**, *8*, 114 – 119; b) H. Donaghy, *mAbs* **2016**, *8*, 659 – 671.
[21] a) P. Sormanni, F. A. Aprile, M. Vendruscolo, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9902 – 9907; b) F. A. Aprile, P. Sormanni, M. Perni, P. Arosio, S. Linse, T. P. J. Knowles, C. M. Dobson, M. Vendruscolo, *Sci. Adv.* **2017**, *3*, e1700488.
[22] J. Leem, J. Dunbar, G. Georges, J. Shi, C. M. Deane, *mAbs* **2016**, *8*, 1259 – 1268.
[23] G. P. Silva, R. S. Santos, W. F. Pereira-Manfro, B. Ferreira, D. M. Barreto, A. C. C. Frota, C. B. Hofer, L. G. Milagres, *Vaccine* **2017**, *35*, 3803 – 3807.
[24] a) M. Tontini, M. R. Romano, D. Proietti, E. Balducci, F. Micoli, C. Balocchi, L. Santini, V. Masignani, F. Berti, P. Costantino, *Vaccine* **2016**, *34*, 4235 – 4242; b) E. Malito, B. Bursulaya, C. Chen, P. L. Surdo, M. Picchianti, E. Balducci, M. Biancucci, A. Brock, F. Berti, M. J. Bottomley, M. Nissum, P. Costantino, R. Rappuoli, G. Spraggon, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5229 – 5234.
[25] G. Stefanetti, Q.-Y. Hu, A. Usera, Z. Robinson, M. Allan, A. Singh, H. Imase, J. Cobb, H. Zhai, D. Quinn, M. Lei, A. Saul, R. Adamo, C. A. MacLennan, F. Micoli, *Angew. Chem. Int. Ed.* **2015**, *54*, 13198 – 13203; *Angew. Chem.* **2015**, *127*, 13396 – 13401.

Manuscript received: August 28, 2017
Revised manuscript received: September 26, 2017
Accepted manuscript online: October 2, 2017
Version of record online: October 20, 2017

ChemPubSoc Europe

DOI: 10.1002/cbic.201700586

CHEMBIOCHEM Communications

# Enhanced Permeability and Binding Activity of Isobutylene-Grafted Peptides

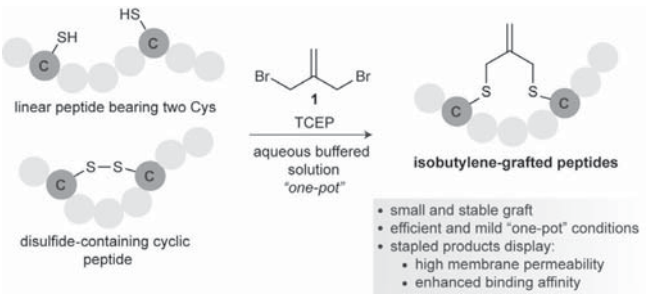
Shuang Sun,[a] Ismael Compañón,[b] Nuria Martínez-Sáez,[a] João D. Seixas,[c] Omar Boutureira,[a] Francisco Corzana,*[b] and Gonçalo J. L. Bernardes*[a, c]

We present a new peptide-macrocyclization strategy with an isobutylene graft. The reaction is mild and proceeds rapidly and efficiently both for linear and cyclic peptides. The resulting isobutylene-grafted peptides possess improved passive membrane permeability due to the shielding of the polar backbone of the amides, as demonstrated by NMR spectroscopy and molecular dynamics simulations. The isobutylene-stapled structures are fully stable in human plasma and in the presence of glutathione. This strategy can be applied to bioactive cyclic peptides such as somatostatin. Importantly, we found that structural preorganization forced by the isobutylene graft leads to a significant improvement in binding. The combined advantages of directness, selectivity, and smallness could allow application to peptide macrocyclization based on this attachment of the isobutylene graft.

Intramolecular side-chain-to-side-chain crosslinking, commonly termed "stapling" or "macrocyclization",[1] is an important technology in the development of bioactive peptide-based therapeutics. Linear peptides are often easily degraded by proteases, and possess low binding affinity and cell permeability.[2] Macrocyclization has evolved as a promising approach to tackling these problems. The cyclized structure often shows enhanced biophysical properties, cellular uptake, and binding affinity of peptides while maintaining high specificity for its biological targets.[3] Over the past decades, the chemical toolbox available for macrocyclization has expanded greatly, and now includes disulfide bond formation,[4] lactam formation,[5] ring-closing metathesis,[1,6] and cycloaddition.[7] Proteinogenic cysteine has attracted significant interest as a convenient handle for stapling, owing to the high nucleophilicity of the thiolate and its unique reactivity.[3a] Chemical approaches for cysteine stapling include S-alkylation,[8] $S_N$-arylation,[9] tetrazine stapling,[10] and radical thiolene[11] reactions. Among the cysteine-stapling methods, S-alkylation is the most flexible approach, as a wide range of bis-thiol-reactive linkers is commercially available. The first investigation of thiol-reactive linkers by using 1,3,5-tris(bromomethyl)benzene was reported in 1985,[12] followed by that of its bis-reactive analogues, 1,2- and 1,3-bis-(bromomethyl)benzene.[13] In recent years, this strategy has also been successfully applied in peptide drug development.[14] Despite the many tools available for peptide macrocyclization, discovering suitable grafts that yield membrane-permeable cyclic peptides with enhanced binding affinities remains a great challenge.

Cysteine residues are easily incorporated into the peptide sequence through solid-phase peptide synthesis. This facile incorporation is an important advantage over other stapling approaches based on nonproteinogenic amino acids. One challenge associated with cysteine macrocyclization strategies is potential oxidation to form disulfides. Thus, an efficient strategy for cysteine stapling should, in principle, be compatible with the presence of reducing agents in a mild, one-pot reaction. Concurrently, the graft should be both small and biologically inert. Here, we described a new cysteine crosslinking strategy that allows the biocompatible and chemoselective installation of a small, rigid isobutylene graft (Scheme 1).

Our investigation commenced with a linear pentamer model peptide CAAAC (peptide **I**, Scheme 2), which had cysteines at the *i* and *i*+4 positions, prepared by Fmoc-based solid-phase peptide synthesis (SPPS). Tris(2-carboxyethyl) phosphine (TCEP) was added to prevent formation of the disulfide, followed by




**Scheme 1.** Schematic representation of the macrocyclization of peptides with cysteine residues by using bis-electrophilic isobutylene.

[a] S. Sun, Dr. N. Martínez-Sáez, Dr. O. Boutureira, Dr. G. J. L. Bernardes
Department of Chemistry, University of Cambridge
Lensfield Road, Cambridge, CB2 1EW (UK)
E-mail: gb453@cam.ac.uk

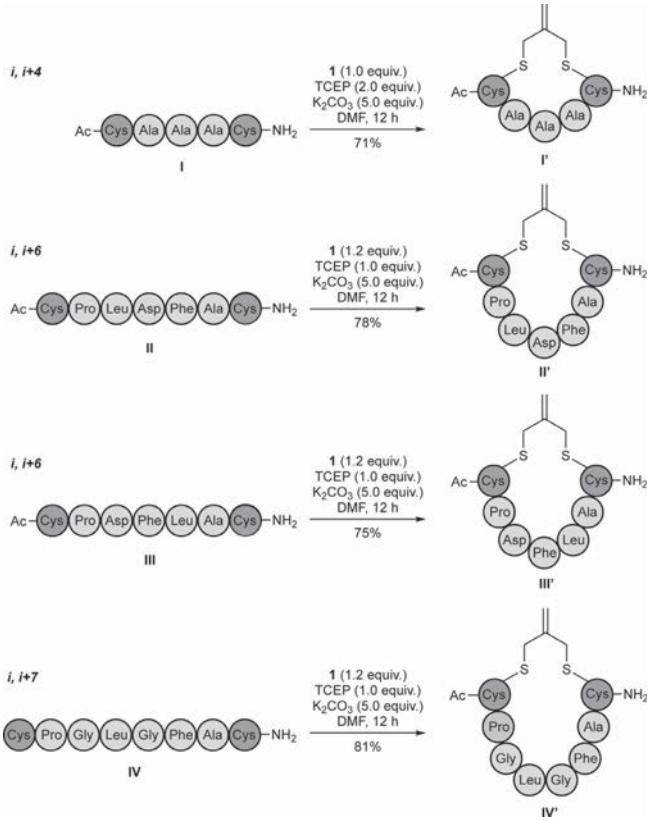
[b] I. Compañón, Dr. F. Corzana
Departamento de Química, Centro de Investigación en Síntesis Química
Universidad de La Rioja
Madre de Dios, 53, 26006 Logroño (Spain)
E-mail: francisco.corzana@unirioja.es

[c] Dr. J. D. Seixas, Dr. G. J. L. Bernardes
Instituto de Medicina Molecular, Faculdade de Medicina
Universidade de Lisboa
Avenida Professor Egas Moniz, 1649-028 Lisboa (Portugal)
E-mail: gbernardes@medicina.ulisboa.pt

Supporting information and the ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201700586.






**Scheme 2.** Macrocyclization of linear peptides with cysteines at the ($i$,$i+4$), ($i$,$i+6$) and ($i$,$i+7$) positions.

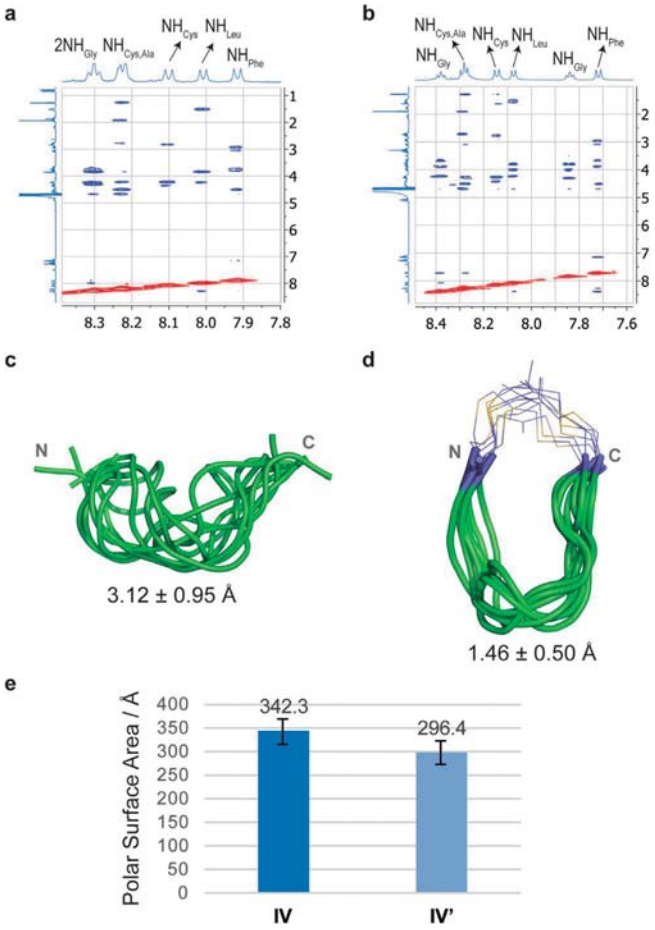



**Figure 1.** Conformation analysis of stapled and unstapled peptides in solution. Sections of the 500 ms ROESY spectra (400 MHz) of peptides A) **IV** and B) **IV′** in $H_2O/D_2O$ (9:1) at pH 6.5 and 20 °C, showing amide–aliphatic crosspeaks. Structural ensembles obtained for C) peptide **IV** and D) stapled peptide **IV′** through 20-ns MD-tar simulations. The backbone is shown in green, and the carbon atoms of isobutylene moiety are in purple. The numbers indicate the rmsd for heavy-atom superimposition of the backbone with respect to the average structure. E) PSA estimated for peptides **IV** and **IV′** through the MD-tar simulations.

the addition of stapling reagent **1**; this afforded exclusively the stapled peptide CAAAC **I′** in 71% yield. The same procedure was then applied successfully to other hepta- and octamer linear peptides with different sequence compositions that containing cysteines at the $i$,$i+6$ or $i$,$i+7$ positions (Scheme 2). In each case, the linear peptides afforded the desired cyclic derivatives in high conversion and yield at room temperature. Finally, because the double bond of the isobutylene graft could potentially be used as a handle for further conjugation, we evaluated its reactivity under thiolene and inverse-electron-demand Diels–Alder conditions. We found that, under the conditions tested, the alkene did not act as a partner for either thiolene or inverse electron demand Diels–Alder (data not shown). This is a significant difference compared to the method described by Dawson and co-workers,[8c] in which the introduced moiety can be used for further conjugation. On the other hand, the isobutylene scaffold is more flexible than the bis(bromomethyl)benzene platform used in the CLIPS strategy;[8b] in some cases this could be a competitive advantage for selecting the bioactive conformation.

Next, we determined the structural changes induced by the incorporation of the isobutylene staple into the large ($i$,$i+7$) peptide **IV** by combining NMR spectroscopy and molecular dynamics (MD) simulations. The 2D ROESY spectra showed substantial differences between the unstapled and stapled peptide in terms of their conformational preferences (Figure 1 A and B, Supporting information). Clear medium-sized NH–NH NOE crosspeaks, which are characteristic of a predominantly folded conformation in solution, were observed for stapled peptide **IV′**; the absence of these NH–NH NOE crosspeaks for unstapled peptide **IV** is in agreement with an extended disposition of the backbone.[15] To obtain an experimentally derived conformational ensemble of compounds **IV** and **IV′**, 20-ns MD simulations with time-averaged restraints (MD-tar)[16] were carried out in explicit water with the key experimental distances included as restraints. The MD-tar simulations were performed by using the AMBER 16[17] package with parm14SB and GAFF force fields.[18] The good agreement found between the experimental and theoretical distances validates the outcome of the MD-tar calculations (see the Supporting Information). According to these calculations, peptide **IV** is reasonably flexible in solution, presenting a random-coil distribution for its backbone. Conversely, stapled peptide **IV′** is rather rigid, showing a folded backbone held by the isobutylene staple (Figure 1 C and D, Supporting Information).

We then estimated the solvent-exposed polar surface area (PSA) of peptides **IV** and **IV′** through the MD simulations. The stapled peptide **IV′** displayed around 15% less PSA than the unstapled variant **IV**; this suggests that the folded conformation forced by the isobutylene fragment promotes shielding of the polar backbone amides (Figure S26). We then decided to determine the passive membrane permeability of peptides **II**–**IV** and their stapled variants experimentally, a key feature for the development of peptide-based therapeutics (Table 1).[19] All

**Table 1.** The parallel artificial membrane permeability assay (PAMPA).[a]

| Compound | Permeability [nm s$^{-1}$] | $-\log P_e$ | Compound | Permeability [nm s$^{-1}$] | $-\log P_e$ |
|---|---|---|---|---|---|
| **II** | 7.6 | 5.12 | **II′** | 12 | 4.96 |
| **III** | < 0.01 | | **III′** | 10 | 4.99 |
| **IV** | 6.0 | 5.22 | **IV′** | 13 | 4.90 |

[a] Permeability was measured at pH 7.4 and at room temperature, the value is reported as an average of quadruplicates.

the isobutylene-grafted peptides had a significant improvement in permeability compared to their linear forms, particularly compound **IV′**. For all stapled derivatives, we observed values of $-\log P_e < 5.0$; this is indicative of highly passively permeable compounds.[20] These data highlight the practicability of the method for developing bioactive peptides with favorable properties.

The feasibility of our stapling approach in aqueous solution, and the impact of the isobutylene scaffold on bioactivity were evaluated further with somatostatin. This peptide inhibits the release of growth hormone, insulin and glucagon, and possesses a disulfide bond between two cysteines at the *i* and *i*+11 positions.[21] Unlike the disulfide bond, which is sensitive to the biological environment, especially in the presence of biological thiols, the stapled form with the isobutylene graft can improve pharmacokinetics and binding affinity. The reaction was conducted in water with 10% DMF as co-solvent, and afforded the stapled somatostatin quantitatively (Figure 2A). This is possible because our method is compatible with TCEP and can be performed in one pot. Notably, the stapled somatostatin was fully stable in the presence of glutathione and human plasma at 37 °C for 48 h (Figures S20 and S21). The affinity of the peptide for the somatostatin receptor type 2 (SSTR2) was experimentally determined by tryptophan fluorescence spectroscopy.[22] As shown in Figure 2B, the fluorescence emission peak of pure SSTR2 solution was at 328 nm. After increasing the concentration of either native or stapled somatostatin, the emission peak of both solutions shifted to 338 nm, with a decrease in intensity. Subsequent addition of somatostatin did not cause any shift in either peak, thus indicating the saturation of SSTR2. The minimum concentration of the somatostatin surrogate required to achieve saturation was 3.5 μM; in contrast, at least 5.5 μM of the native somatostatin was needed. These data suggest that the isobutylene-grafted somatostatin has a higher binding affinity for SSTR2. This improvement in binding activity represents a considerable advantage of the incorporation of the isobutylene graft when compared to other three-carbon grafts, such as the recently reported methylene thioacetal, which led to a decrease in affinity for SSTR2.[8e] Our 0.5-μs unrestrained MD simulations performed on both derivatives indicated that the stapled somatostatin is more rigid in solution (Figure 2C). Although the circular dichroism (CD) spectra

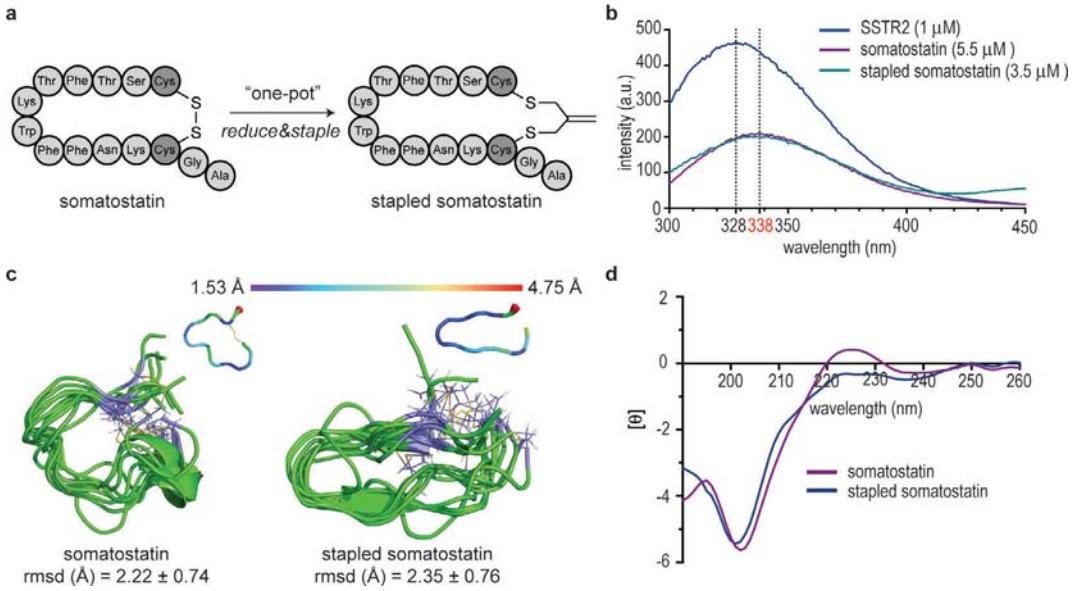



**Figure 2.** Stapling as well as structural and biological evaluation of somatostatin. A) Schematic representation of the stapling of somatostatin; B) Tryptophan fluorescence spectroscopy of somatostatin. Blue: 1 μM SSTR2 in buffer; purple: 5.5 μM native somatostatin and 1 μM SSTR2 in buffer; green: 3.5 μM stapled somatostatin and 1 μM SSTR2 in buffer. C) Structural ensembles obtained for native and stapled somatostatin through 0.5-μs unrestrained MD simulations. The atomic fluctuation (Cα) analysis of both peptides is also shown. The data correspond to the average structure of both molecules throughout the simulations. The backbone is shown in green, and carbon atoms of cysteine isobutylene residues are in purple. The numbers indicate the rmsd for heavy-atom superimposition of the backbone with respect to the average structure. D) CD spectra of native and stapled somatostatin.

of somatostatin and the corresponding stapled peptide (Figure 2D and the Supporting Information) are rather similar, the peak at 225 nm found in somatostatin might be indicative of the presence of a higher degree of polyproline (PPII) conformation for this peptide.[23] MD simulations showed the S–S distance in the isobutylene scaffold to be around $4.2 \pm 0.4$ Å—larger than the conventional S–S disulfide bond length (ca. 2.0 Å) and the S–S distance in methylene thioacetals (close to 3.0 Å);[8e] hypothetically this would allow the required degree of flexibility to adopt a bioactive conformation. The restrained peptide flexibility and structural preorganization within the backbone, favored by the formation of intramolecular hydrogen bonds, might reduce the entropy cost of receptor binding that limits the conformational ensemble and, ultimately, increase the binding affinity compared to those of disulfide cyclized analogues.[24]

In conclusion, we have demonstrated a robust cysteine macrocyclization and stapling strategy in which an isobutylene graft is introduced in a one-pot (with TCEP), biocompatible manner. This method was applied to several linear peptides of various sequence composition and a bioactive disulfide cyclic peptide. The shielding of the polar backbone of the amides promoted by the isobutylene graft led to highly membrane-permeable peptide macrocycles. Enhanced binding activity, resulting from limited flexibility and structural preorganization of the peptide backbone, was also observed. We believe that this access to such a "small" site-selectively introduced isobutylene, which is less disruptive than many current bulky grafts, as demonstrated here for linear and cyclic peptides, is likely to find significant use for the peptide drug discovery community by allowing access to structures with improved properties.

## Experimental Section

**General procedure for peptide stapling with isobutylene:** The linear peptide (0.02 mmol) was dissolved in DMF (10 mL). $K_2CO_3$ (0.10 mmol) and tris(2-carboxyethyl)phosphine (TCEP; 0.02 mmol) were then added. The solution was stirred for 1 h at room temperature. 3-Bromo-2-(bromomethyl)prop-1-ene (0.025 mmol) was then added, and the mixture was stirred for an additional 12 h. The crude peptide was purified by reversed-phase HPLC to obtain the corresponding stapled derivative. In all cases the yield was $\geq 75\,\%$.

## Acknowledgements

*We thank the Cambridge Trust and China Scholarship Council (PhD studentship to S.S.), the European Commission (Marie Skłodowska-Curie ITN ProteinConjugates; Marie Curie IEF to O.B.), MINECO (project CTQ2015-67727-R to F.C.) Universidad de La Rioja (FPI PhD studentship to I.C.), FCT Portugal (FCT Investigator to G.J.L.B.), and the EPSRC for financial support. G.J.L.B. is a Royal Society URF and the recipient of an ERC Starting Grant (TagIt).*

## Conflict of Interest

*S.S., N.M.S., O.B., F.C., and G.J.L.B. are listed as inventors in a patent application related to the work presented here.*

**Keywords:** cyclic peptides · isobutylene · macrocyclization · peptides · stapling

[1] L. D. Walensky, A. L. Kung, I. Escher, T. J. Malia, S. Barbuto, R. D. Wright, G. Wagner, G. L. Verdine, S. J. Korsmeyer, *Science* **2004**, *305*, 1466–1470.
[2] V. Azzarito, K. Long, N. S. Murphy, A. J. Wilson, *Nat. Chem.* **2013**, *5*, 161–173.
[3] a) Y. H. Lau, P. de Andrade, Y. Wu, D. R. Spring, *Chem. Soc. Rev.* **2015**, *44*, 91–102; b) C. J. White, A. K. Yudin, *Nat. Chem.* **2011**, *3*, 509–524.
[4] a) A. Brust, C. I. A. Wang, N. L. Daly, J. Kennerly, M. Sadeghi, M. J. Christie, R. J. Lewis, M. Mobli, P. F. Alewood, *Angew. Chem. Int. Ed.* **2013**, *52*, 12020–12023; *Angew. Chem.* **2013**, *125*, 12242–12245; b) R. J. Clark, H. Fischer, L. Dempster, N. L. Daly, K. J. Rosengren, S. T. Nevin, F. A. Meunier, D. J. Adams, D. J. Craik, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13767–13772; c) M. E. Selsted, A. J. Ouellette, *Nat. Immunol.* **2005**, *6*, 551–557.
[5] a) K. Fujimoto, M. Kajino, M. Inouye, *Chem. Eur. J.* **2008**, *14*, 857–863; b) J. W. Taylor, *Biopolymers* **2002**, *66*, 49–75; c) M. Zhang, B. Wu, J. Baum, J. W. Taylor, *J. Pept. Res.* **2000**, *55*, 398–408.
[6] a) A. J. Huhn, R. M. Guerra, E. P. Harvey, G. H. Bird, L. D. Walensky, *Cell Chem. Biol.* **2016**, *23*, 1123–1134; b) H. E. Blackwell, R. H. Grubbs, *Angew. Chem. Int. Ed.* **1998**, *37*, 3281–3284; *Angew. Chem.* **1998**, *110*, 3469–3472; c) C. E. Schafmeister, J. Po, G. L. Verdine, *J. Am. Chem. Soc.* **2000**, *122*, 5891–5892.
[7] a) H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021; *Angew. Chem.* **2001**, *113*, 2056–2075; b) Y. H. Lau, P. de Andrade, S. T. Quah, M. Rossmann, L. Laraia, N. Skold, T. J. Sum, P. J. E. Rowling, T. L. Joseph, C. Verma, M. Hyvonen, L. S. Itzhaki, A. R. Venkitaraman, C. J. Brown, D. P. Lane, D. R. Spring, *Chem. Sci.* **2014**, *5*, 1804–1809; c) Y. H. Lau, Y. Wu, M. Rossmann, B. X. Tan, P. de Andrade, Y. S. Tan, C. Verma, G. J. McKenzie, A. R. Venkitaraman, M. Hyvonen, D. R. Spring, *Angew. Chem. Int. Ed.* **2015**, *54*, 15410–15413; *Angew. Chem.* **2015**, *127*, 15630–15633.
[8] a) F. Zhang, O. Sadovski, S. J. Xin, G. A. Woolley, *J. Am. Chem. Soc.* **2007**, *129*, 14154–14155; b) H. Jo, N. Meinhardt, Y. Wu, S. Kulkarni, X. Hu, K. E. Low, P. L. Davies, W. F. DeGrado, D. C. Greenbaum, *J. Am. Chem. Soc.* **2012**, *134*, 17704–17713; c) N. Assem, D. J. Ferreira, D. W. Wolan, P. E. Dawson, *Angew. Chem. Int. Ed.* **2015**, *54*, 8665–8668; *Angew. Chem.* **2015**, *127*, 8789–8792; d) S. Bellotto, S. Chen, I. Rentero Rebollo, H. A. Wegner, C. Heinis, *J. Am. Chem. Soc.* **2014**, *136*, 5880–5883; e) C. M. B. K. Kourra, N. Cramer, *Chem. Sci.* **2016**, *7*, 7007–7012.
[9] a) A. J. Rojas, C. Zhang, E. V. Vinogradova, N. H. Buchwald, J. Reilly, B. L. Pentelute, S. L. Buchwald, *Chem. Sci.* **2017**, *8*, 4257–4263; b) A. M. Spokoyny, Y. Zou, J. J. Ling, H. Yu, Y. S. Lin, B. L. Pentelute, *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949; c) E. V. Vinogradova, C. Zhang, A. M. Spokoyny, B. L. Pentelute, S. L. Buchwald, *Nature* **2015**, *526*, 687–691.
[10] a) M. Abdo, S. P. Brown, J. R. Courter, M. J. Tucker, R. M. Hochstrasser, A. B. Smith, *Org. Lett.* **2012**, *14*, 3518–3521; b) M. J. Tucker, J. R. Courter, J. Chen, O. Atasoylu, A. B. Smith, R. M. Hochstrasser, *Angew. Chem. Int. Ed.* **2010**, *49*, 3612–3616; *Angew. Chem.* **2010**, *122*, 3694–3698; c) S. P. Brown, A. B. Smith, *J. Am. Chem. Soc.* **2015**, *137*, 4034–4037.
[11] a) T. N. Grossmann, J. T. Yeh, B. R. Bowman, Q. Chu, R. E. Moellering, G. L. Verdine, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17942–17947; b) Y. Wang, D. H. Chou, *Angew. Chem. Int. Ed.* **2015**, *54*, 10931–10934; *Angew. Chem.* **2015**, *127*, 11081–11084.
[12] D. S. Kemp, P. E. Mcnamara, *J. Org. Chem.* **1985**, *50*, 5834–5838.
[13] Z. Szewczuk, K. L. Rebholz, D. H. Rich, *Int. J. Pept. Protein Res.* **1992**, *40*, 233–242.
[14] P. Timmerman, W. C. Puijk, R. H. Meloen, *J. Mol. Recognit.* **2007**, *20*, 283–299.
[15] F. Corzana, J. H. Busto, S. B. Engelsen, J. Jiménez-Barbero, J. L. Asensio, J. M. Peregrina, A. Avenoza, *Chem. Eur. J.* **2006**, *12*, 7864–7871.
[16] D. A. Pearlman, P. A. Kollman, *J. Mol. Biol.* **1991**, *220*, 457–479.

[17] D. A. Case, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P.A. Kollman, *AMBER 16*, University of California, San Francisco, **2016**.

[18] a) J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696 – 3713; b) J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157 – 1174.

[19] a) J. L. Hickey, S. Zaretsky, M. A. St. Denis, S. Kumar Chakka, M. M. Morshed, C. C. Scully, A. L. Roughton, A. K. Yudin, *J. Med. Chem.* **2016**, *59*, 5368 – 5376; b) E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discovery* **2008**, *7*, 608 – 624; c) T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson, R. S. Lokey, *J. Am. Chem. Soc.* **2006**, *128*, 2510 – 2511; d) T. Rezai, J. E. Bock, M. V. Zhou, C. Kalyanaraman, R. S. Lokey, M. P. Jacobson, *J. Am. Chem. Soc.* **2006**, *128*, 14073 – 14080; e) G. Vistoli, A. Pedretti, B. Testa, *Drug Discovery Today* **2008**, *13*, 285 – 294.

[20] a) G. Ottaviani, S. Martel, P. A. Carrupt, *J. Med. Chem.* **2006**, *49*, 3948 – 3954; b) M. J. Taylor, S. Tanna, T. Sahota, *J. Pharm. Sci.* **2010**, *99*, 4215 – 4227; c) M. Kansy, F. Senner, K. Gubernator, *J. Med. Chem.* **1998**, *41*, 1007 – 1010.

[21] Y. Yamada, S. R. Post, K. Wang, H. S. Tager, G. I. Bell, S. Seino, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 251 – 255.

[22] J. T. Vivian, P. R. Callis, *Biophys. J.* **2001**, *80*, 2093 – 2109.

[23] Y. Tachibana, G. L. Fletcher, N. Fujitani, S. Tsuda, K. Monde, S.-I. Nishimura, *Angew. Chem. Int. Ed.* **2004**, *43*, 856 – 862; *Angew. Chem.* **2004**, *116*, 874 – 880.

[24] a) W. A. Loughlin, J. D. Tyndall, M. P. Glenn, D. P. Fairlie, *Chem. Rev.* **2004**, *104*, 6085 – 6117; b) J. E. DeLorbe, J. H. Clements, M. G. Teresk, A. P. Benfield, H. R. Plake, L. E. Millspaugh, S. F. Martin, *J. Am. Chem. Soc.* **2009**, *131*, 16758 – 16770; c) E. Marsault, M. L. Peterson, *J. Med. Chem.* **2011**, *54*, 1961 – 2004; d) A. K. Yudin, *Chem. Sci.* **2015**, *6*, 30 – 49.

Manuscript received: November 2, 2017

Accepted manuscript online: November 3, 2017

Version of record online: November 22, 2017

JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

Cite This: J. Am. Chem. Soc. 2017, 139, 18255−18261

pubs.acs.org/JACS

# The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer

Víctor J. Somovilla,[†,‡,□] Iris A. Bermejo,[†,□] Inês S. Albuquerque,[§,□] Nuria Martínez-Sáez,[†,‡] Jorge Castro-López,[∥] Fayna García-Martín,[⊥] Ismael Compañón,[†] Hiroshi Hinou,[⊥] Shin-Ichiro Nishimura,[⊥] Jesús Jiménez-Barbero,[#] Juan L. Asensio,[¶] Alberto Avenoza,[†] Jesús H. Busto,[†] Ramón Hurtado-Guerrero,[∥,△] Jesús M. Peregrina,[†] Gonçalo J. L. Bernardes,[*,§,▽] and Francisco Corzana[*,†]

[†]Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, 26006 Logroño, Spain

[‡]Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Universiteitsweg 99, Utrecht, The Netherlands

[§]Instituto de Medicina Molecular, Faculdade de Medicina da, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028, Lisboa, Portugal

[∥]Institute of Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, BIFI-IQFR (CSIC), Zaragoza, Spain

[⊥]Graduate School and Faculty of Advanced Life Science, Field of Drug Discovery Research, Hokkaido University, N21 W11, Sapporo 001-0021, Japan

[#](i) CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain; (ii) Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain; (iii) Department of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Spain

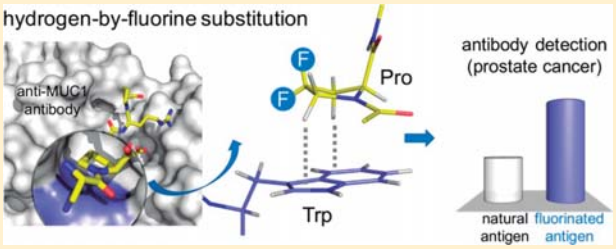
[¶]Instituto de Química Orgánica General, IQOG-CSIC, 28006 Madrid, Spain

[△]Fundación ARAID, 50018 Zaragoza, Spain

[▽]Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, U.K.

Supporting Information

**ABSTRACT:** A structure-based design of a new generation of tumor-associated glycopeptides with improved affinity against two anti-MUC1 antibodies is described. These unique antigens feature a fluorinated proline residue, such as a (4*S*)-4-fluoro-L-proline or 4,4-difluoro-L-proline, at the most immunogenic domain. Binding assays using biolayer interferometry reveal 3-fold to 10-fold affinity improvement with respect to the natural (glyco)peptides. According to X-ray crystallography and MD simulations, the fluorinated residues stabilize the antigen−antibody complex by enhancing key CH/$\pi$ interactions. Interestingly, a notable improvement in detection of cancer-associated anti-MUC1 antibodies from serum of patients with prostate cancer is achieved with the non-natural antigens, which proves that these derivatives can be considered better diagnostic tools than the natural antigen for prostate cancer.




## INTRODUCTION

MUC1 is a glycoprotein overexpressed in around 80% of human cancers.[1−3] It consists of an extracellular domain that comprises a variable number (20 to 125) of tandem repeat regions formed by 20 amino acids (His-Gly-Val-Thr-Ser-Ala-Pro-Asp-Thr-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala-Pro-Pro-Ala). This domain includes five potential *O*-glycosylation sites, with three threonine (Thr) and two serine (Ser) residues. While in healthy cells, the MUC1 backbone displays complex oligosaccharides, in tumors it is decorated with basic, truncated carbohydrates. Consequently, different tumor-associated carbohydrate antigens (TACAs), such as the Tn determinant ($\alpha$-*O*-GalNAc-Ser/Thr), become exposed and are involved in triggering immune responses.[4−7] Because of this unique feature, extensive efforts have been made toward the development of cancer vaccines based on MUC1 fragments.[8−12]

In addition, over the years, several studies have demonstrated that circulating anti-MUC1 antibodies in serum may be used as a favorable prognosis for patients with early breast and pancreatic cancer because these antibodies can limit tumor outgrowth and dissemination.[13−18] Consequently, efforts have

Received: September 4, 2017
Published: November 22, 2017

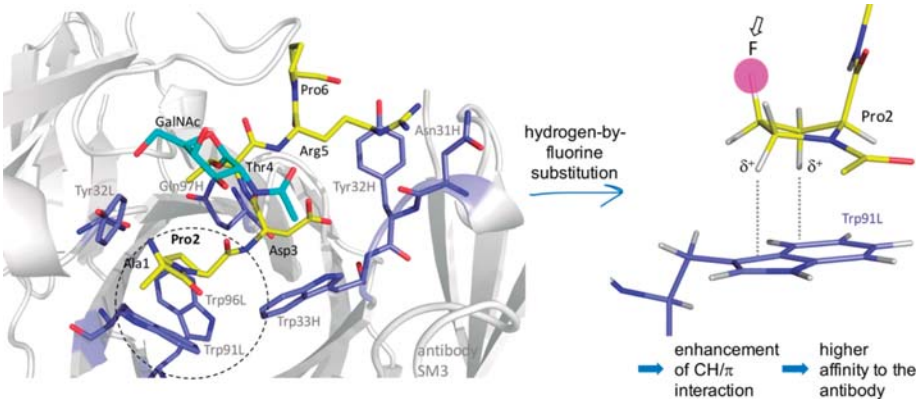
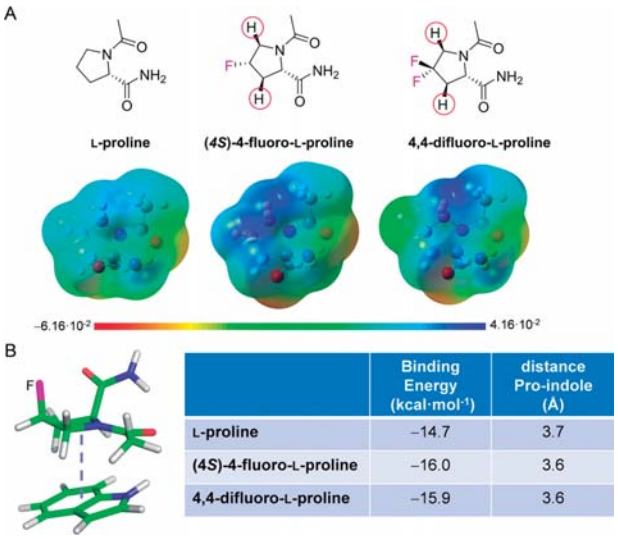



**Figure 1.** Hydrogen-by-fluorine Pro replacement strategy to improve antibody−antigen affinity. Crystal structure of glycopeptide APDT(α-*O*-GalNAc)RP in complex with antibody SM3 (pdb ID: 5a2k) together with the strategy proposed in this work to design effective antigens based on MUC1.




| | Binding Energy (kcal·mol⁻¹) | distance Pro-indole (Å) |
|---|---|---|
| L-proline | −14.7 | 3.7 |
| (4S)-4-fluoro-L-proline | −16.0 | 3.6 |
| 4,4-difluoro-L-proline | −15.9 | 3.6 |

**Figure 2.** Effect of fluorine atoms on the electrostatic potential and on CH/π stability. (A) Electrostatic potential surfaces (in au) calculated at the M06-2X/6-31+G(d,p) level in a vacuum, showing the *re* faces of the Pro derivatives. Blue/red indicates positive/negative potentials. (B) Binding energy calculated at the M06-2X/6-31+G(d,p) level in a vacuum for the three Pro derivatives with an indole residue, together with the optimized distance Pro−indole ring for the complexes.

been devoted toward the rational design of MUC1-based antigens to be used as diagnostic tools for detection of anti-MUC1 antibodies in human serum. Unfortunately, so far, a commercial assay for early cancer diagnosis based on the detection of anti-MUC1 antibodies in human serum remains unavailable. However, significant advances toward this aim were reported by Wang and co-workers,[19] where they described an assay based on a recombinant MUC1 protein that contained six MUC1 tandem repeats and was effective in detecting anti-MUC1 antibodies in serum from patients. More recently, a chimera containing both MUC1 and human epidermal growth factor receptor-2 (HER2), whose overexpression is associated with malignancy in breast cancer, has been developed for detection of antibodies against MUC1 or HER2 in human serum.[20] It should be noted that, in these examples, antibody detection relies on unmodified naturally occurring antigens.

Alternatively, fine-tuning of antibody/antigen interactions by exploiting non-natural, synthetically designed antigen modifications holds great potential in the development of diagnostic detection systems with improved selectivity and sensitivity. This strategy demands a precise understanding of the molecular basis of the antigen−antibody recognition process. In this regard, recent progress unveiled subtle molecular details of the antibody/antigen interaction,[21,22] paving the way to the structure-based design of synthetic antigens with improved potential value in diagnosis and detection. In this context, it has been shown that most anti-MUC1 antibodies display a significant affinity to peptide fragments containing the APDTRP sequence,[23] which consequently represents an attractive target for lead optimization.

Our group has recently described the X-ray structure of a short peptide bearing the sequence APDT(α-*O*-GalNAc)RP in complex with the SM3,[22] which is a therapeutic antibody used in the treatment of cancer.[21] According to these data, the nonterminal Pro residue plays a central role in the stabilization of the antibody/antigen complex, stacking against aromatic units of Trp91L, Trp96L, and Tyr32L (Figure 1). This observation explains why a proline residue at this position of the antigen is indeed essential for the binding of various anti-MUC1 antibodies.[24] Interestingly, recent studies have shown that CH/π bonds can be significantly enhanced by simply increasing the polarization of the interacting CH moieties.[25−28] In the present work, this effect could be achieved by attaching highly electronegative fluorine atoms to specific positions of the proline scaffold. Therefore, we hypothesized that the replacement of the nonterminal proline residue of the antigen by a non-natural proline derivative, such as (4*S*)-4-fluoro-L-proline or 4,4-difluoro-L-proline, should enhance antigen−antibody affinity (Figure 1).

Here, we designed and synthesized various MUC1 antigens that featured a hydrogen-by-fluorine substitution at that proline residue that displayed enhanced affinity to two anti-MUC1 antibodies. By combining molecular dynamic (MD) simulations and X-ray crystallography, we provide an explanation of the superior affinity of our derivatives toward two antibodies, which relies on stronger CH/π interactions. Finally, we demonstrate that these novel derivatives are more efficient than natural antigens in detecting low concentrations of circulating anti-MUC1 antibodies in human serum of patients with prostate cancer (adenocarcinoma and benign prostatic hyperplasia).

## ■ RESULTS AND DISCUSSION

To provide theoretical support for the hypothesis depicted in Figure 1, we calculated the electrostatic potential surface of the


DOI: 10.1021/jacs.7b09447
J. Am. Chem. Soc. 2017, 139, 18255−18261

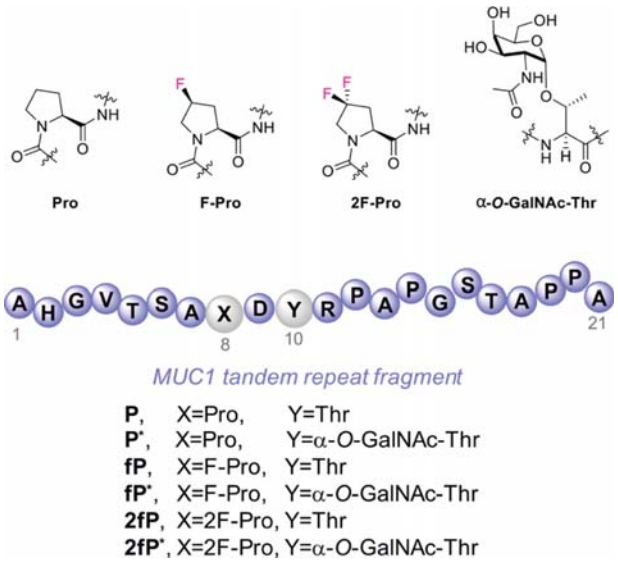



**Figure 3.** MUC1-like peptides and glycopeptides synthesized and studied in this work.

proline and its mono- and difluorinated derivatives shown in Figure 2A. In a second step, their interaction energies with an indole ring were evaluated at the M06-2X/6-31+G(d,p) level of theory.[29] The obtained relative energies were in agreement with our rational design and showed that the fluorinated residues displayed an enhanced positive partial charge on the CH/$\pi$ donor methylene fragments (highlighted with a red circle in Figure 2A). As a result, the interacting proline face displays a significantly larger electrostatic potential even for the mono-substituted derivative (4*S*)-4-fluoro-L-proline. This polarization effect is reflected in the theoretical interaction energies, with complex stabilizations larger than 1 kcal/mol (Figure 2B).

With these data in hand, we decided to synthesize a series of peptides and glycopeptides comprising the tandem repeat sequence of MUC1 (Figure 3) both fluorinated and nonfluorinated at position 8. The synthesis of all these derivatives was conducted using microwave-assisted solid phase peptide synthesis (MW-SPPS), employing a Rink amide MBHA resin and Fmoc-protected amino acids and using our reported protocol[30] (Supporting Information).

We then evaluated the impact of the hydrogen-by-fluorine substitutions on the antigen association to antibody scFv-SM3[22] using biolayer interferometry (BLI). Although both natural peptide **P** and glycopeptide **P*** act as antigens, it has been observed that the latter displays better affinity than the naked peptide (Figure 4A).[22,23] This trend is also noted for the non-natural derivatives. To our delight, surrogates **fP*** and **2fP***, which contain a (4*S*)-4-fluoro-L-proline or a 4,4-difluoro-L-proline at position 8, respectively, showed the best affinity, with ca. 3-fold enhancement with respect to the natural **P*** and 1 order of magnitude regarding natural peptide **P**. In parallel, we performed microarray assays[31,32] to determine the affinity of these MUC1 variants with commercially available antibodies SM3 and VU-3C6[23] (Figure 4B). The results obtained from the microarray experiments were in line with the BLI data, showing a clear increase in binding affinity for the non-natural variants. This result may indicate that both SM3 and VU-3C6 antibodies recognize the antigen with an equivalent binding mode.

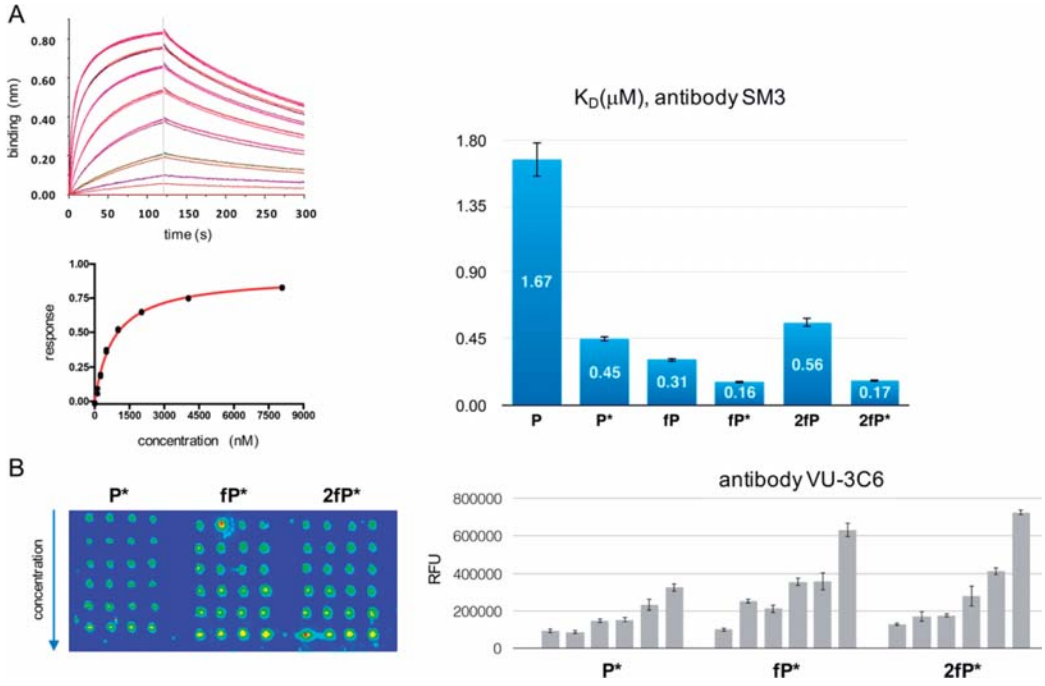



**Figure 4.** Binding of the glycopeptides to anti-MUC1 antibodies. (A) BLI curves and fit obtained for glycopeptide **fP*** and antibody scFv-SM3, together with the $K_D$ constants derived from BLI experiments for all studied MUC1-related compounds. (B) Interaction of the anti-MUC1 antibody VU-3C6 with the glycopeptides using a microarray platform. Compounds were printed onto an aminooxy-functionalized microarray in quadruplicate. Relative fluorescence units (RFU) due to the binding of the Cy3-labeled secondary antibody were measured and represented as mean values in a bar chart (see Supporting Information for experimental details).



DOI: 10.1021/jacs.7b09447
J. Am. Chem. Soc. 2017, 139, 18255−18261

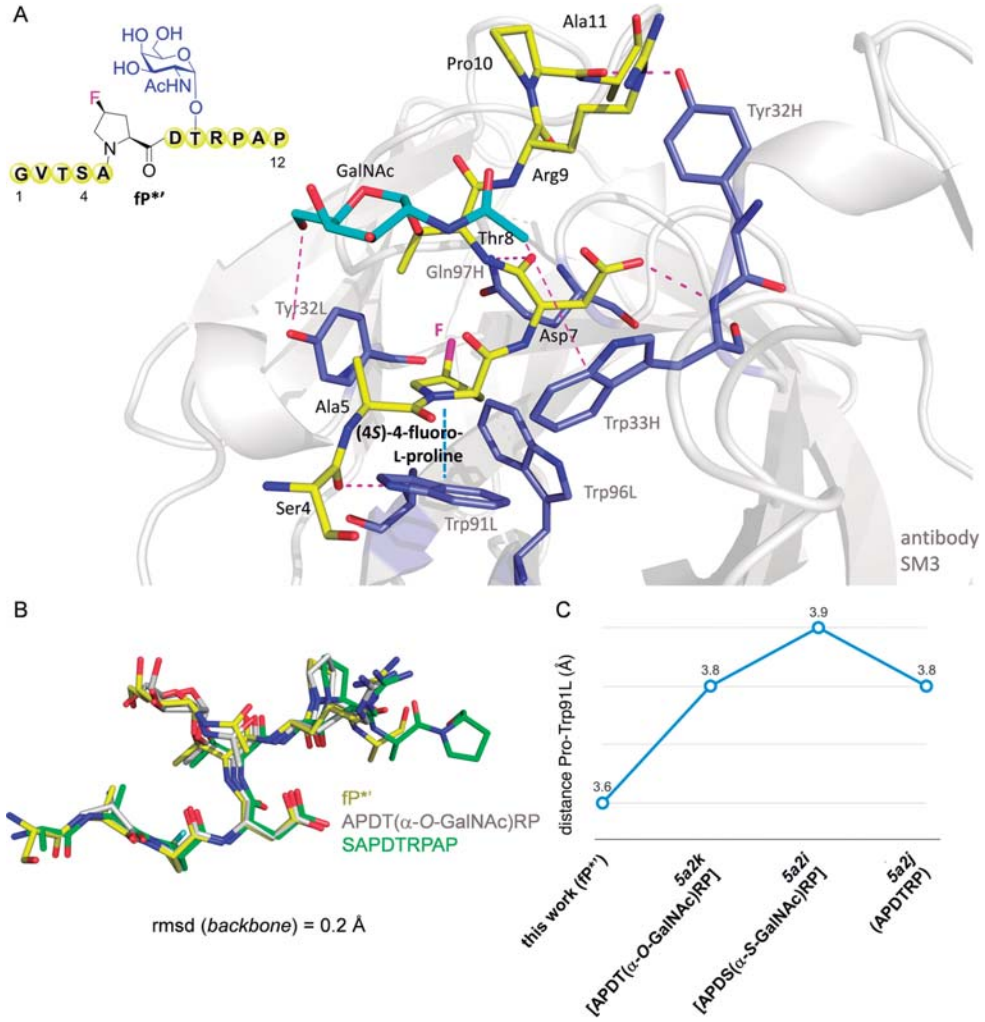



**Figure 5.** Crystal structure of **fP*′** bound to scFv-SM3. (A) Key binding interactions of glycopeptide **fP*′** with scFv-SM3 mAb, as observed in the X-ray crystal structure (pdb ID: 5OWP). Peptide backbone carbon atoms are shown in yellow. GalNAc carbon atoms are shown in cyan. Carbon atoms of key residues of SM3 are in slate. The antibody is shown as gray ribbons. Pink dashed lines indicate hydrogen bonds between peptide backbones and SM3 antibody. It is important to note that only residues SAPDTRP of the peptide could be resolved, presumably due to the higher flexibility of the rest of the amino acids. (B) Superposition of the peptide backbone of glycopeptides **fP*′** and APDT(*α*-*O*-GalNAc)RP, together with SAPDTRPAP peptide[21] in complex with SM3. The root-mean-square deviation (rmsd) of the backbone is shown. (C) Experimental distances Pro−Trp91L obtained from X-ray structures of various MUC1-like derivatives in complex with antibody SM3.

To fully validate our molecular design, we carried out detailed structural studies on our fluorinated antigens in complex with the antibody SM3. To this purpose, we synthesized a simplified variant of glycopeptide **fP***, comprising the sequences GVTSA**f**PDT*RPAP and denoted as **fP*′** throughout the Article. High-quality crystals of **fP*′** in complex with scFv-SM3 were obtained, which enabled the acquisition of a structure at high resolution (<2.0 Å, Figure 5A and Supporting Information). Crystallographic analysis revealed that the conformation of the glycopeptide was almost identical to that found for peptide SAPDTRPAP and for glycopeptide APDT(*α*-*O*-GalNAc)RP in complex with the same antibody.[21,22] This outcome suggests that the incorporation of a fluorine atom at the proline residue does not significantly modify the structure of the peptide in the bound state (Figure 5B). Consequently, the antigen−antibody hydrogen-bonding network is identical to that found in the previously reported complexes.[21,22] The obtained results also indicate that GalNAc glycosylation does not have an influence on the accommodation of the key proline residue (compare *5a2k* and *5a2j* in Figure 5C). In fact, the glycosidic linkage adopts the expected exoanomeric/*syn* conformation, with $\phi$ and $\psi$ values of ∼68 and 91°, respectively. This conformation is similar to that found for a Tn-glycopeptide in complex with SM3 (pdb ID: 5a2k)[22] and allows the formation of an intermolecular hydrogen bond between the hydroxymethyl group of GalNAc and the side chain of Tyr32L of the antibody. Moreover, the *N*-acetyl group of the sugar stacks with the aromatic ring of Trp33H, which provides the driving force for the observed selectivity of SM3 for GalNAc-containing antigens. Despite all these similarities, the obtained crystallographic coordinates revealed a subtle but crucial modification with previously reported structural data. Markedly, the observed distance between the center of the (4*S*)-4-fluoro-L-proline ring and Trp91L was significantly smaller than that observed for the Pro−Trp pairs in other complexes (Figure 5C). This result provides further support to our premise, strongly suggesting that


DOI: 10.1021/jacs.7b09447
J. Am. Chem. Soc. 2017, 139, 18255−18261

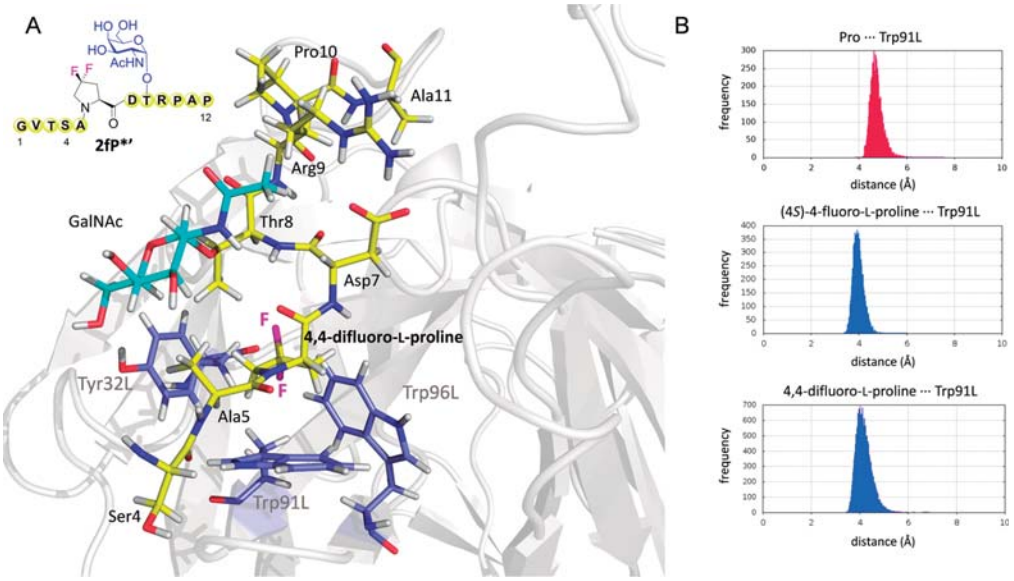



**Figure 6.** MD simulations of glycopeptide **2fP*′** in complex with scFv-SM3. (A) Representative frame of the 200 ns MD simulations performed on **2fP*′** in complex with antibody scFv-SM3. Only peptide fragment Ser4-Ala11 is shown for clarity. Peptide backbone carbon atoms are shown in yellow. GalNAc carbon atoms are shown in cyan. Carbon atoms of key residues of SM3 are in slate. The antibody is shown as gray ribbons. (B) Distance distribution of Pro−Trp91L obtained from 200 ns MD simulations for glycopeptides **P*′** (upper panel), **fP*′** (middle panel), and **2fP*′** (lower panel) bound to the antibody SM3.

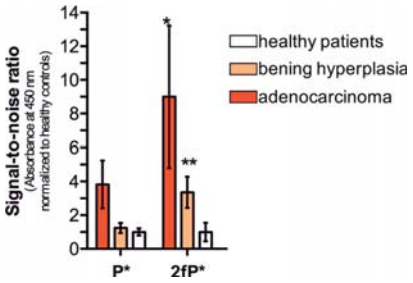



**Figure 7.** Detection of circulating anti-MUC1 antibodies in serum of patients presenting benign and malignant prostate cancer. Binding affinity of human circulating antibodies against MUC1 (**P***) and synthetic variant **2fP***, in the context of prostate adenocarcinoma and benign prostatic hyperplasia. Bars show average ± standard error of the mean of absorbance values normalized to healthy controls. All groups were compared to **P*** using Wilcoxon matched-pairs signed rank test; $*p < 0.05$; $**p < 0.02$ (see Supporting Information for details).

the improvement in the proline/tryptophan stacking accounts for the increased stability of the complexes formed by the non-natural fluorinated antigens.

The influence exerted by proline fluorination on the antigen/SM3 complexes was also analyzed by MD simulations. Thus, we collected 200 ns trajectories on **2fP*′** and **fP*′** bound to scFv-SM3 (Figure 6A). MD simulations on the reduced versions of natural glycopeptide (**P*′**) complexed to the antibody were also conducted for comparative purposes. According to these theoretical data, the three complexes were stable through the simulations. Most importantly, a shorter distance between Pro and Trp91L was found for derivate **fP*′** (with a distance Pro−Trp91L = 4.0 ± 0.4 Å), in agreement with the X-ray structure described in Figure 5, and for compound **2fP*′** (distance Pro−Trp91L = 4.1 ± 0.6 Å). In contrast, the distance found for the natural glycopeptide **P*′** was 4.7 ± 0.4 Å. In summary, both experimental and theoretical data confirm the validity of our design and show that improving proline/tryptophan stacking-type interaction through simple hydrogen-by-fluorine substitutions represents a simple way to stabilize the antigen/antibody complex. These results hint that the tailored fluorinated MUC1 glycopeptides can be employed as potential antigens for the efficient detection of anti-MUC1 antibodies.

With this purpose, we established an indirect ELISA (Supporting Information) assay using both **P*** and non-natural MUC1 **2fP*** as coating antigens to detect anti-MUC1 antibodies from serum samples of patients with benign and malignant prostate tumors. As can be observed in Figure 7, the signal-to-noise ratio was statistically higher when **2fP*** was used for both adenocarcinoma and benign hyperplasia. Our study also shows a higher concentration of anti-MUC1 antibodies in malignant tumors, which agrees with other assays conducted with breast tumor patients[19] and importantly validates our protocol. The results disclosed here demonstrate the potential application of the designed **2fP*** MUC1 variant as a biomarker for improved detection of circulating anti-MUC1 antibodies in the serum of patients.

## ■ CONCLUSIONS

A multidisciplinary approach to unravel key features of MUC1 recognition has been established. In particular, a new set of non-natural MUC1 derivatives comprising a (4*S*)-4-fluoro-L-proline or 4,4-difluoro-L-proline residue at the most immunogenic domain have been designed and synthesized. These compounds present a clear enhancement in the binding affinity against two anti-MUC1 antibodies with respect to the natural antigens. Both experimental X-ray studies and theoretical MD simulations confirm that, in agreement with our expectations, the hydrogen-by-fluorine substitution enhances the key CH/$\pi$ interaction, which is crucial for improving the binding of the antigen to the antibody. Moreover, the obtained glycopeptides


DOI: 10.1021/jacs.7b09447
J. Am. Chem. Soc. 2017, 139, 18255−18261

display a significant potential as diagnostic tools to detect anti-MUC1 antibodies in prostate cancer patients.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b09447.

Characterization of the glycopeptides, biolayer interferometry fit obtained for the glycopeptides, microarray figures, details of the X-ray structure of **fP*′** bound to scFv-SM3, Cartesian coordinates, electronic energies, and lowest frequencies of the DFT calculated structures, additional molecular dynamics simulation figures, protocol of the ELISA, and details of the human sera samples (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Authors

*francisco.corzana@unirioja.es
*gb453@cam.ac.uk; gbernardes@medicina.ulisboa.pt

### ORCID

Víctor J. Somovilla: 0000-0001-5067-6568
Fayna García-Martín: 0000-0001-9118-3874
Shin-Ichiro Nishimura: 0000-0002-6608-8418
Jesús Jiménez-Barbero: 0000-0001-5421-8513
Jesús H. Busto: 0000-0003-4403-4790
Jesús M. Peregrina: 0000-0003-3778-7065
Gonçalo J. L. Bernardes: 0000-0001-6594-8917
Francisco Corzana: 0000-0001-5597-8127


### Author Contributions

□V. J. Somovilla, I. A. Bermejo, and I. S. Albuquerque have contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank the Ministerio de Economía y Competitividad (projects CTQ2015-67727-R, UNLR13-4E-1931, CTQ2013-44367-C2-2-P, CTQ2015-64597-C2-1P, and BFU2016-75633-P). I.A.B. thanks the Asociación Española Contra el Cancer en La Rioja for a grant. I.S.A. and G.J.L.B. thank FCT Portugal (Ph.D. studentship and FCT Investigator, respectively) and EPSRC. G.J.L.B. holds a Royal Society URF and an ERC StG (TagIt). F.C. and G.J.L.B thank the EU (Marie-Sklodowska Curie ITN, Protein Conjugates). R.H-G. thanks Agencia Aragonesa para la Investigación y Desarrollo (ARAID) and the Diputación General de Aragón (DGA, B89) for financial support. The research leading to these results has also received funding from the FP7 (2007-2013) under BioStruct-X (grant agreement no. 283570 and BIOSTRUCTX_5186). We thank synchrotron radiation source DIAMOND (Oxford) and beamline I04 (number of experiment mx10121-19). The Hokkaido University group acknowledges JSPS KAKENHI grant no. 25220206 and JSPS Wakate B KAKENHI grant no. 24710242. We also thank CESGA (Santiago de Compostela) for computer support.

## ■ REFERENCES


(1) Taylor-Papadimitriou, J.; Burchell, J. M. *Mucins and Cancer*; Future Medicine Ltd: Unitec House, London, UK, 2013.
(2) Hollingsworth, M. A.; Swanson, B. J. *Nat. Rev. Cancer* **2004**, *4*, 45−60.
(3) Kufe, D. W. *Nat. Rev. Cancer* **2009**, *9*, 874−885.
(4) Kailemia, M. J.; Park, D.; Lebrilla, C. B. *Anal. Bioanal. Chem.* **2017**, *409*, 395−410.
(5) Adamczyk, B.; Tharmalingam, T.; Rudd, P. M. *Biochim. Biophys. Acta, Gen. Subj.* **2012**, *1820*, 1347−1353.
(6) Varela, J. C.; Atkinson, C.; Woolson, R.; Keane, T. E.; Tomlinson, S. *Int. J. Cancer* **2008**, *123*, 1357−1363.
(7) Rabassa, M. E.; Croce, M. V.; Pereyra, A.; Segal-Eiras, A. *BMC Cancer* **2006**, *6*, 253.
(8) Buskas, T.; Thompson, P.; Boons, G.-J. *Chem. Commun.* **2009**, *105*, 5335−5349.
(9) Wolfert, M. A.; Boons, G.-J. *Nat. Chem. Biol.* **2013**, *9*, 776−784.
(10) Wilson, R. M.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2013**, *135*, 14462−14472.
(11) Gaidzik, N.; Westerlind, U.; Kunz, H. *Chem. Soc. Rev.* **2013**, *42*, 4421−4442.
(12) Richichi, B.; Thomas, B.; Fiore, M.; Bosco, R.; Qureshi, H.; Nativi, C.; Renaudet, O.; BenMohamed, L. *Angew. Chem., Int. Ed.* **2014**, *53*, 11917−11920.
(13) Blixt, O.; Bueti, D.; Burford, B.; Allen, D.; Julien, S.; Hollingsworth, M.; Gammerman, A.; Fentiman, I.; Taylor-Papadimitriou, J.; Burchell, J. M. *Breast Cancer Res.* **2011**, *13*, R25.
(14) Hamanaka, Y.; Suehiro, Y.; Fukui, M.; Shikichi, K.; Imai, K.; Hinoda, Y. *Int. J. Cancer* **2003**, *103*, 97−100.
(15) von Mensdorff-Pouilly, S.; Verstraeten, A. A.; Kenemans, P.; Snijdewint, F. G.; Kok, A.; Van Kamp, G. J.; Paul, M. A.; Van Diest, P. J.; Meijer, S.; Hilgers, J. *J. Clin. Oncol.* **2000**, *18*, 574−583.
(16) Tang, Z.-M.; Ling, Z.-G.; Wang, C.-M.; Wu, Y.-B.; Kong, J.-L. *PLoS One* **2017**, *12*, e0182117.
(17) Chen, H.; Werner, S.; Tao, S.; Zörnig, I.; Brenner, H. *Cancer Lett.* **2014**, *346*, 178−187.
(18) Tang, Y.; Cui, X.; Xiao, H.; Qi, S.; Hu, X.; Yu, Q.; Shi, G.; Zhang, X.; Gu, J.; Yu, Y.; Wang, L.; Li, Y. *Mol. Med. Rep.* **2017**, *15*, 2659−2664.
(19) Tang, Y.; Wang, L.; Zhang, P.; Wei, H.; Gao, R.; Liu, X.; Yu, Y.; Wang, L.; Wang, L. *Clin. Vaccine Immunol.* **2010**, *17*, 1903−1908.
(20) Gheybi, E.; Amani, J.; Salmanian, A. H.; Mashayekhi, F.; Khodi, S. *Tumor Biol.* **2014**, *35*, 11489−11497.
(21) Dokurno, P.; Bates, P. A.; Band, H. A.; Stewart, L. M.; Lally, J. M.; Burchell, J. M.; Taylor-Papadimitriou, J.; Snary, D.; Sternberg, M. J.; Freemont, P. S. *J. Mol. Biol.* **1998**, *284*, 713−728.
(22) Martínez-Sáez, N.; Castro-López, J.; Valero-Gónzalez, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem., Int. Ed.* **2015**, *54*, 9830−9834.
(23) Karsten, U.; Serttas, N.; Paulsen, H.; Danielczyk, A.; Goletz, S. *Glycobiology* **2004**, *14*, 681−692.
(24) Her, C.; Westler, W. M.; Yang, T. *JSM Chem.* **2013**, *1*, 1004.
(25) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. *Acc. Chem. Res.* **2013**, *46*, 946−954.
(26) Jiménez-Moreno, E.; Jiménez-Osés, G.; Gómez, A. M.; Santana, A. G.; Corzana, F.; Bastida, A.; Jiménez-Barbero, J.; Asensio, J. L. *Chem. Sci.* **2015**, *6* (11), 6076−6085.
(27) Hsu, C.-H.; Park, S.; Mortenson, D. E.; Foley, B. L.; Wang, X.; Woods, R. J.; Case, D. A.; Powers, E. T.; Wong, C.-H.; Dyson, H. J.; Kelly, J. W. *J. Am. Chem. Soc.* **2016**, *138*, 7636−7648.
(28) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. *J. Am. Chem. Soc.* **2015**, *137*, 15152−15160.
(29) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(30) Martínez-Sáez, N.; Supekar, N. T.; Wolfert, M. A.; Bermejo, I. A.; Hurtado-Guerrero, R.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Boons, G.-J.; Peregrina, J. M.; Corzana, F. *Chem. Sci.* **2016**, *7*, 2294−2301.

(31) Matsushita, T.; Takada, W.; Igarashi, K.; Naruchi, K.; Miyoshi, R.; Garcia-Martin, F.; Amano, M.; Hinou, H.; Nishimura, S.-I. *Biochim. Biophys. Acta, Gen. Subj.* **2014**, *1840*, 1105−1116.
(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.; Garcia-Martin, F.; Marcelo, F. *J. Am. Chem. Soc.* **2015**, *137*, 12438−12441.


DOI: 10.1021/jacs.7b09447
*J. Am. Chem. Soc.* 2017, 139, 18255−18261

ChemPubSoc Europe

DOI: 10.1002/chem.201800603

CHEMISTRY A European Journal Full Paper

Fluorescent Peptides

# Cell-Penetrating Peptides Containing Fluorescent D-Cysteines

Claudio D. Navo[+],[a] Alicia Asín[+],[a] Eva Gómez-Orte[+],[b] Marta I. Gutiérrez-Jiménez[+],[a] Ismael Compañón,[a] Begoña Ezcurra,[b] Alberto Avenoza,[a] Jesús H. Busto,[a] Francisco Corzana,[a] María M. Zurbano,[a] Gonzalo Jiménez-Osés,[a] Juan Cabello,*[b] and Jesús M. Peregrina*[a]

**Abstract:** A series of fluorescent D-cysteines (Cys) has been synthesized and their optical properties were studied. The key synthetic step is the highly diastereoselective 1,4-conjugate addition of aryl thiols to a chiral bicyclic dehydroalanine recently developed by our group. This reaction is fast at room temperature and proceeds with total chemo- and stereoselectivity. The Michael adducts were easily transformed into the corresponding amino acids to study their optical properties and, in some selected cases, into the corresponding *N*-Fmoc-D-cysteine derivatives to be used in solid-phase peptide synthesis (SPPS). To further demonstrate the utility of these non-natural Cys-derived fluorescent amino acids, the coumaryl and dansyl derivatives were incorporated into cell-penetrating peptide sequences through standard SPPS and their optical properties were studied in different cell lines. The internalization of these fluorescent peptides was monitored by fluorescence microscopy.

## Introduction

Nowadays, fluorescence spectroscopy is routinely used to investigate complex biological processes as well as a diagnostic and research tool in many fields of biomedical sciences.[1] Although natural amino acids such as tryptophan, tyrosine, and phenylalanine have been used as intrinsic fluorescent probes to monitor different biological events, their use is restricted owing to their poor optical properties.[2] Therefore, the design and synthesis of fluorescent non-natural amino acids and their incorporation into peptides or proteins has emerged as a powerful tool to study biological processes by using fluorescence spectroscopy.[3,4] Given the continuous development of imaging techniques and synthetic biology methods, there is an urgent need for expanding the toolbox of available amino acids incorporating a variety of fluorescent moieties with different spectroscopic properties.[3,4]

Although several synthetic protocols have been developed for the synthesis of fluorescent non-natural L-amino acids,[1,3,4] very few are described for D-amino acids.[5] In this context, it has been shown that fluorescent D-amino acids allow quick detection of cell wall biosynthesis in live cells and in numerous bacterial species.[6] This method has also been used to study the synthesis of peptidoglycans in some bacterial species.[7]

Functionalized amino acids such as lysine, cysteine, or tyrosine have been found to be ideal candidates for the development of fluorescent labels through conjugation of the corresponding electrophilic dye. Thus, the aim of this work is to introduce different commercially available fluorophores into the D-Cys scaffold for potential diagnostic or imaging purposes when the resulting amino acids are incorporated into peptides.

L-Cys-derived fluorescent amino acids are normally accessed through nucleophilic attack of the sulfhydryl group of L-Cys to a fluorescent building block equipped with a suitable leaving group (Scheme 1).

However, the formation of C–S bonds involving the opposite approach—the nucleophilic reaction of dyes incorporating a thiol group with electrophilic amino acid derivatives—has received less attention. In this work, we report a new methodology using thiol-containing dyes as nucleophiles, which undergo stereoselective *S*-Michael addition to a chiral dehydroamino acid (Scheme 1).

Dehydroalanine (Dha)[8] is an α,β-unsaturated amino acid of biological and synthetic interest,[9] which acts as a chemical precursor for a range of site-selective post-translational modifications. However, in general, these reactions are poorly stereoselective. We recently reported the synthesis of a new cyclic chiral Dha-derivative **ent-1** derived from D-serine. This compound behaves as a versatile substrate for stereoselective *S*-Michael additions, generally providing high yields and diastereoselectivity (Scheme 2).[10] More recently, we developed a rigidi-

[a] C. D. Navo,[+] A. Asín,[+] M. I. Gutiérrez-Jiménez,[+] I. Compañón, Prof. A. Avenoza, Dr. J. H. Busto, Dr. F. Corzana, Dr. M. M. Zurbano, Dr. G. Jiménez-Osés, Prof. J. M. Peregrina
Dpto. de Química, Centro de Investigación en Síntesis Química
Universidad de La Rioja
C/ Madre de Dios, 53, 26006, Logroño, La Rioja (Spain)
E-mail: jesusmanuel.peregrina@unirioja.es

[b] Dr. E. Gómez-Orte,[+] Dr. B. Ezcurra, Dr. J. Cabello
Center for Biomedical Research of La Rioja (CIBIR)
C/ Piqueras, 98, 26006, Logroño, La Rioja (Spain)
E-mail: juan.cabello@riojasalud.es

[+] These authors contributed equally to this work.

Supporting information and the ORCID numbers for the authors of this article can be found under https://doi.org/10.1002/chem.201800603.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
Cite This: J. Am. Chem. Soc. 2018, 140, 9952−9960
pubs.acs.org/JACS

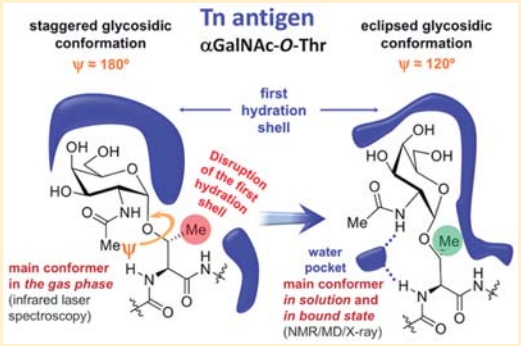
# Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition

Iris A. Bermejo,[†,□] Imanol Usabiaga,[‡,□] Ismael Compañón,[†,□] Jorge Castro-López,[§] Aran Insausti,[‡,¶] José A. Fernández,[‡] Alberto Avenoza,[†] Jesús H. Busto,[†] Jesús Jiménez-Barbero,[∥,⊗,▽] Juan L. Asensio,[⊥] Jesús M. Peregrina,[†] Gonzalo Jiménez-Osés,[†] Ramón Hurtado-Guerrero,[*,§,#] Emilio J. Cocinero,[*,‡,¶] and Francisco Corzana[*,†]

[†]Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, 26006 Logroño, Spain
[‡]Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV-EHU), 48080 Bilbao, Spain
[§]Institute of Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, BIFI-IQFR (CSIC), 50018 Zaragoza, Spain
[∥]CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain
[⊗]Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain
[▽]Department of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Spain
[⊥]Instituto de Química Orgánica General, IQOG-CSIC. 28006 Madrid, Spain
[#]Fundación ARAID, 50018 Zaragoza, Spain
[¶]Biofisika Institute (CSIC, UPV/EHU), University of the Basque Country (UPV/EHU), 48940 Leioa, Spain

Supporting Information

**ABSTRACT:** The tumor-associated carbohydrate Tn antigens include two variants, αGalNAc-*O*-Thr and αGalNAc-*O*-Ser. In solution, they exhibit dissimilar shapes and dynamics and bind differently to the same protein receptor. Here, we demonstrate experimentally and theoretically that their conformational preferences in the gas phase are highly similar, revealing the essential role of water. We propose that water molecules prompt the rotation around the glycosidic linkage in the threonine derivative, shielding its hydrophobic methyl group and allowing an optimal solvation of the polar region of the antigen. The unusual arrangement of αGalNAc-*O*-Thr features a water molecule bound into a "pocket" between the sugar and the threonine. This mechanism is supported by trapping, for the first time, such localized water in the crystal structures of an antibody bound to two glycopeptides that comprise fluorinated Tn antigens in their structure. According to several reported X-ray structures, installing oxygenated amino acids in specific regions of the receptor capable of displacing the bridging water molecule to the bulk-solvent may facilitate the molecular recognition of the Tn antigen with threonine. Overall, our data also explain how water fine-tunes the 3D structure features of similar molecules, which in turn are behind their distinct biological activities.




## INTRODUCTION

The Tn antigens are among the most specific human tumor-associated carbohydrate antigens (TACAs), present in approximately 90% of tumors.[1,2] In general, the aggressiveness of the carcinoma and the occurrence of these antigens has a clear direct correlation,[3] promoting their use as biomarkers and potential therapeutic targets against cancer.[4] Vaccines based on peptide fragments carrying this determinant are able to induce antibodies in mice that discriminate between normal and cancer cells, reducing in some cases, the size of the tumor and increasing the survival of the animals.[5]

In general, the Tn antigen is referred to as GalNAc α-*O*-linked to a serine or a threonine residue (**Tn-Ser** and **Tn-Thr**, respectively), without specifying which of the two amino acids the GalNAc is linked to (Figure 1a). However, despite their structural simplicity and similarity, differing only in a methyl group, they display totally different conformations in solution (Figure 1b), leading to significant biological consequences.[6−12] For instance, while anti-MUC1 antibodies recognize glycopeptides bearing a **Tn-Thr** moiety, they show very low affinity toward derivatives with the **Tn-Ser** residue.[6] On the other hand, several anti-Tn antibodies show a clear preference for glycopeptides containing the **Tn-Ser** antigen.[13] A recent study

Received: May 7, 2018
Published: July 13, 2018






DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

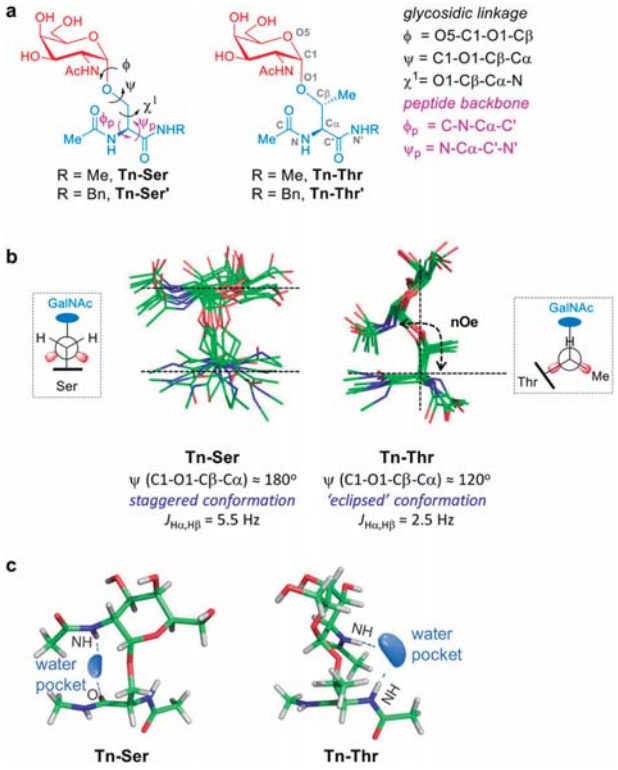



**Figure 1.** Conformational behavior of the antigens **Tn-Ser** and **Tn-Thr** in water. (a) The two Tn-antigens studied in this work, together with the definition of the most relevant torsional angles and atom labels. (b) Major conformations in solution for the Tn antigen with either a serine (**Tn-Ser**) or a threonine (**Tn-Thr**) derived from experiment-guided molecular dynamics (MD) simulations.[7,8] The values of the $\psi$ torsion angle of the glycosidic linkage and of the associated coupling constant $J_{H\alpha,H\beta}$ are shown. The Newman projections for the C$\beta$−O1 bonds are also given, showing the staggered (**Tn-Ser**) and eclipsed (**Tn-Thr**) conformations. (c) Water pockets derived from experiment-guided MD simulations between the peptide fragment and the GalNAc.[7,8] Antigens **Tn-Ser** and **Tn-Thr** accommodate different water pockets owing to their distinct conformational behavior in solution.

conducted by our own group concluded that while some lectins select the **Tn-Thr** determinant, others showed a higher affinity to glycopeptides carrying the **Tn-Ser** epitope.[9] In the context of antifreeze glycoproteins, which consist of the tandem repeating (Thr-Ala-Ala)$_n$ polypeptide glycosylated with Gal$\beta$-(1,3)-GalNAc, the replacement of a threonine by a serine residue eliminates the antifreeze activity.[14] It is important to note that these Tn antigens are also present in the structure of other significant TACAs, such as T, STn, or ST antigens.[4]

Our conformational analysis based on NMR data combined with experiment-guided Molecular Dynamics (MD) simulations,[15] showed that **Tn-Thr** (Figure 1b) is rather rigid in solution,[8] with its *O*-glycosidic linkage in the so-called "eclipsed" conformation ($\phi \approx 80°$, $\psi \approx 120°$), and its side chain fixed at a $\chi^1$ torsional angle value around 60° (Figure 1a). Conversely, **Tn-Ser** (Figure 1b) displays the typical *exo*-anomeric/*syn* conformation for the glycosidic linkage with $\phi$ and $\psi$ values ~80° and ~180°, respectively. This latter system is more flexible and exhibits the three possible staggered conformers for the side chain in solution.[7] In **Tn-Thr**, the carbohydrate lies almost perpendicular to the peptide but in **Tn-Ser** it adopts a parallel orientation. These conformational differences can be attributed to intrinsic steric interactions between the endocyclic oxygen and the methyl group of the threonine, which force the GalNAc unit to be located distant to the threonine, leading to an eclipsed conformation of the $\psi$ torsion angle.[8] Concomitant with these differences, water molecules around both antigens are predicted to be distributed in a different way (Figure 1c). **Tn-Thr** displays a potential water-binding pocket localized between the NH groups of the GalNAc and threonine moieties but in the **Tn-Ser** a bridging water molecule would engage the carbonyl group of the amino acid. Their different locations could play an additional role in stabilizing the observed conformations. Unfortunately, despite various attempts, using $^{15}$N- and $^{13}$C-labeled Tn variants, no evidence of the presence of the proposed water pockets could be directly deduced from NMR experiments to support the computer predictions.[8] Here, by combining synthesis, crystallographic, spectroscopic and computational studies conducted in the solid and gas phases, and comparing these results with our previous and updated NMR and MD simulations data in solution, we elucidate the molecular basis behind the distinct conformations of **Tn-Thr** and **Tn-Ser** in an aqueous environment.

## RESULTS AND DISCUSSION

**Conformational Analysis of the Tn Antigens in the Gas Phase.** First, the factors governing the preferred conformations of the Tn antigens in the gas phase, free of any interference of solvent, were investigated using mass- and conformer-selected infrared laser spectroscopy conducted under molecular beam conditions and coupled with quantum chemical computations. This strategy has been successfully employed to deduce the gas phase conformational preferences of many biomolecules.[16−18] To facilitate their detection through mass-selected ultraviolet (UV) photoionization, the Tn antigens were synthesized as benzylamide derivatives (**Tn-Ser′** and **Tn-Thr′** derivatives in Figure 1a). Of note, these variants displayed a comparable behavior in solution than that observed for the methylated variants, confirmed by 2D-ROESY spectra and experiment-guided MD simulations[15] (see Supporting Information for methods, Table S1, Schemes S1 and S2, and Figures S1, S2, and S11−S42).

The infrared ion-dip (IRID) spectra in the gas phase of **Tn-Thr′** and **Tn-Ser′** are shown in Figure 2 (see also Supporting Information, Figures S3−S7). Remarkably, they are nearly identical, particularly in the regions of the N−H and O−H stretching modes, a region highly sensitive to the presence of specific hydrogen-bonding, which suggests that these entities display the same pattern of hydrogen bonds in the gas phase. Indeed, the lowest free energy conformers calculated for both derivatives using different quantum mechanical methods (Supporting Information, Tables S2−S4) are very similar and show an excellent agreement between experimental and calculated spectra (Figure 2), confirming the observed conformation. In these preferred conformations the amino acid backbone adopts an inverse $\gamma$-turn stabilized by a strong hydrogen bond between the amino acid *C*-terminal amide and the *N*-terminal acetamide carbonyl group (band at ~3370 cm$^{-1}$ for NH). 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: the acetamide carbonyl acts as a H-bond acceptor to


DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

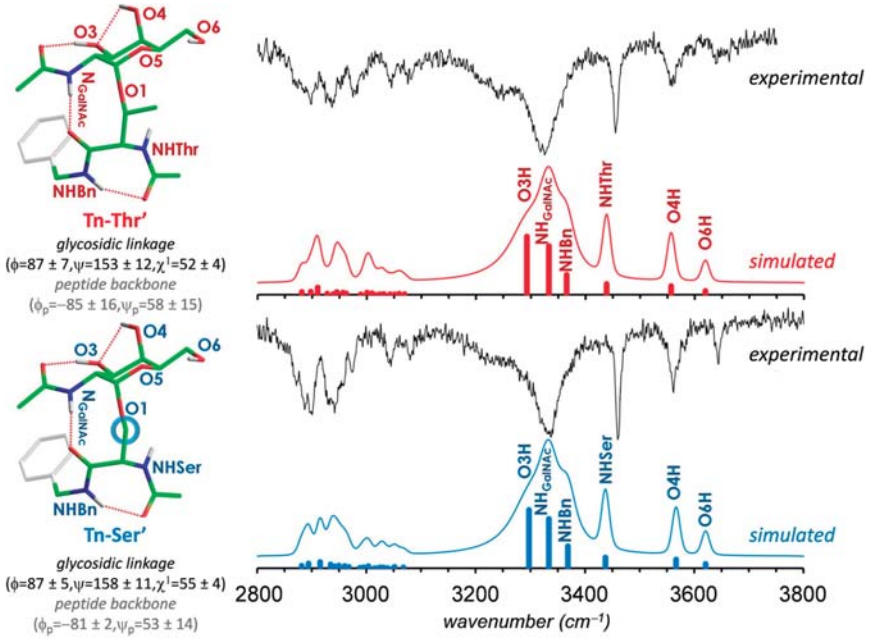



**Figure 2.** Infrared ion-dip (IRID) spectra of antigens **Tn-Ser′** and **Tn-Thr′** in the gas phase. Experimental and simulated infrared ion-dip (IRID) spectra of derivatives **Tn-Thr′** (upper panel) and **Tn-Ser′** (lower panel), together with representative minimum free energy conformers calculated for both compounds at the M06-2X/6-31+G(d) level.[20] These conformers are virtually identical for both derivatives in the gas phase and show the typical staggered conformation for the glycosidic linkage found in solution for the serine derivative. The values for the glycosidic linkages and the peptide backbone dihedrals are an average of the different computational methods used to predict the IRID spectra (Supporting Information, Tables S2−S4).

O3H hydroxyl group and the acetamide NH acts as a H-bond donor to the underlying amino acid *C*-terminal carbonyl group. As a result, the $\psi$ torsion angle of the glycosidic linkage is locked to the value 158 ± 11° for the **Tn-Ser′** antigen and 153 ± 12° for the **Tn-Thr′** analogue, depending on the computational method used (Tables S2−S4). The occurrence of these common stabilizing interactions is in good agreement with the bands observed within the ∼3200−3350 $cm^{-1}$ region for the O3H and the NH of the carbohydrate. In general, the hydroxyl groups that are involved in strong-moderate hydrogen bonds show red-shifted and broad bands. In this particular case, O3—H (strong O3—H···O═C) is displaced about 300 $cm^{-1}$ and it is broadened around 200 $cm^{-1}$. Although NH groups display similar behavior, they display narrower bands and smaller displacements than OH groups.[16b,c] In this case, the acetamide NH of GalNAc (strong $NH_{GalNAc}$···O═C) is displaced around 100 $cm^{-1}$, and it is broadened about 100 $cm^{-1}$. The remaining hydroxyl groups of the sugar are engaged in weak hydrogen bonds, characterized by bands within the 3500−3700 $cm^{-1}$ region: O4H interacts with O3 while O6H is in *gg* conformation and engaged in a hydrogen bond with the endocyclic oxygen O5.

Notably, the theoretical IR spectra derived from the low-energy structures are in excellent agreement with the IRID data, demonstrating the correct prediction of the common hydrogen bond network occurring in both Tn antigen derivatives in the gas phase. These experimentally validated structures also coincide with the one proposed by Csonka and co-workers[19] for the **Tn-Ser** antigen in the gas phase, based on *ab initio* calculations (HF/6-31G(d)). The new spectroscopic evidence conclusively proves that the different behavior observed for the glycosidic linkages in the Tn antigens in *solution* does not *solely* reflect the influence of steric repulsions between the carbohydrate moiety and the $\beta$-methyl group of the threonine derivative.[8] Thus, the dominant population of the eclipsed conformation of **Tn-Thr** found experimentally in solution suggests an important role for differently organized water around this substrate with respect to its **Tn-Ser** analogue.

**Gradual Solvation of the Tn Antigens.** Taking into account that all attempts to experimentally characterize the **Tn-Ser′** and **Tn-Thr′** antigens solvated with a discrete number of water molecules in the gas phase were unsuccessful, a comprehensive theoretical conformational analysis of the Tn antigens (**Tn-Ser** and **Tn-Thr**) with 1 and 20 water molecules was conducted to fulfill the conditions for the first hydration shell of both molecules (Figure 3a and Supporting Information, Tables S5 and S6). This analysis involved, as in the case of the isolated structures, an exhaustive conformational search to find the lowest energy structures for each system and the subsequent minimization through quantum mechanics (Supporting Information and Figure 3a). Although both monohydrated Tn antigens share the staggered conformation around the glycosidic linkage (with $\psi \approx 165°$), the addition of a single water molecule to the **Tn-Ser** derivative disrupts the hydrogen bond between the sugar and the amino acid and promotes a backbone conformational transition from the inverse $\gamma$-turn form observed *in vacuo* toward the extended arrangement populated in solution. In contrast, the **Tn-Thr···$H_2O$** complex retained the folded arrangement for the amino acid fragment, requiring up to 20 discrete water molecules to complete the same conformational shift. However, these water molecules were still not enough to force the **Tn-Thr** antigen to adopt the $\psi \approx 120°$ geometry. Accordingly, when experiment-guided MD simulations on these antigens were conducted in explicit water (Figure 3a, right panel, and Supporting Information, Figure S8), the **Tn-Thr** derivative adopted the "eclipsed" conformation for the


DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

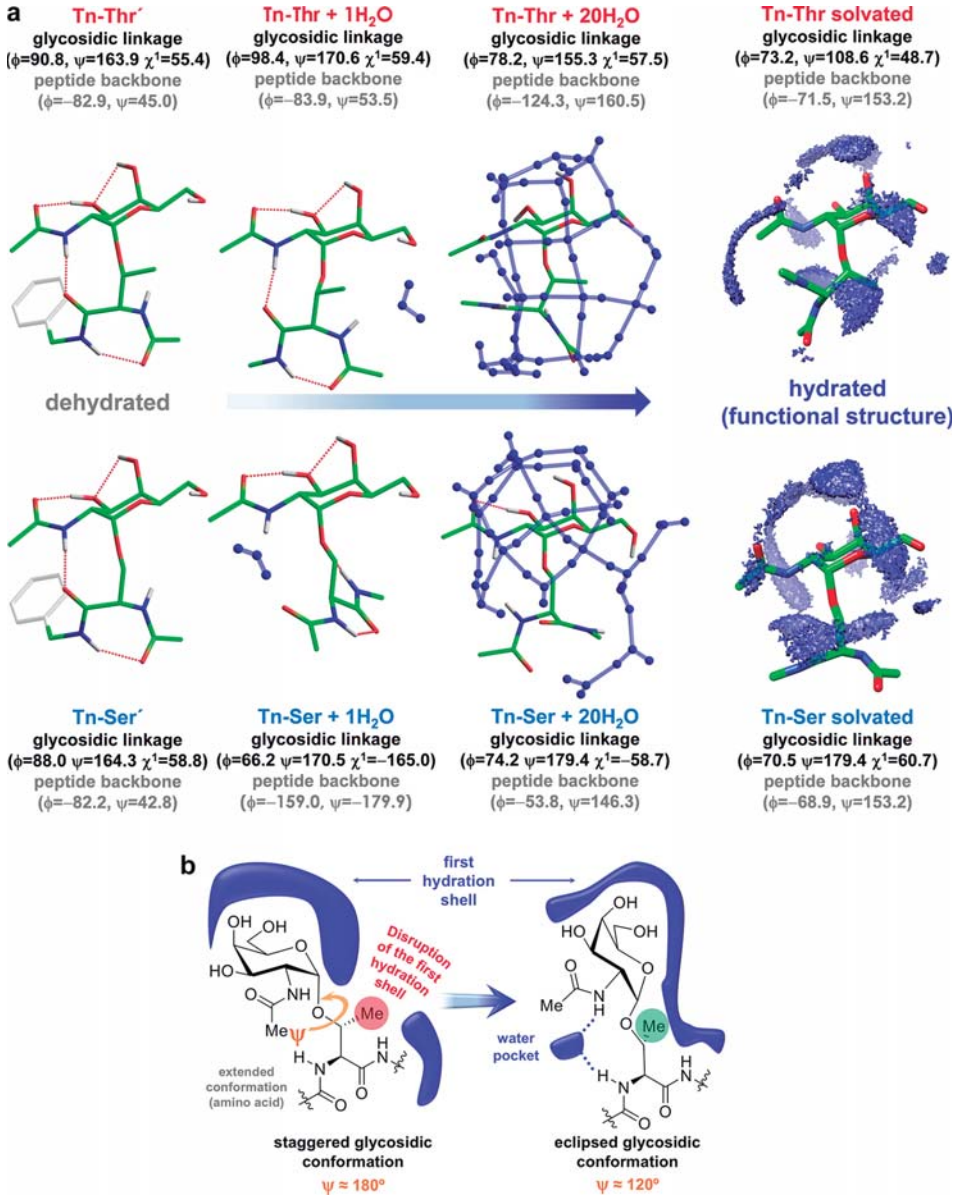



**Figure 3.** Gradual solvation and first hydration shell of the Tn antigens. (a) Lowest energy conformers calculated at the M06-2X/6-31+G(d) level[20] for **Tn-Thr** (upper panel) and **Tn-Ser** (lower panel) with discrete water molecules (1 and 20), together with the averaged first hydration shell derived from the experiment-guided 1 $\mu$s MD simulations. In the **Tn-Ser**, the first hydration shell is not obstructed when the staggered conformer is displayed. On the contrary, in the **Tn-Thr** a conformational shift toward the eclipsed conformer occurs, promoting an efficient solvation of the entire molecule. The geometry of the glycosidic linkage (in black) and the conformation of the peptide backbone (in gray) are also shown. (b) Schematic representation of the proposed role of the water molecules for determining the 3D structure of the **Tn-Thr** antigen.

glycosidic linkage and an extended conformation for the amino acid. Thus, the change in the underlying amino acid backbone from the inverse $\gamma$-turn (in the gas phase) to extended PPII conformations (in water) and complete solvation of the molecule are both crucial to achieve the "eclipsed" glycosidic linkage in **Tn-Thr**.

Full water solvation impairs the key hydrogen bonds involving the *N*-acetyl group of GalNAc described above, disconnecting the sugar and amino acid moieties and exacerbating both the steric and hydrophobic influence of the $\beta$-methyl group of threonine and ultimately unveiling the differences between serine and threonine Tn antigens in solution. In **Tn-Ser**, the contacts between the GalNAc and the serine moieties through water molecules take place without any interference and the more stable staggered form observed in vacuum is retained. Conversely, for **Tn-Thr**, the first solvation shell clashes with the $\beta$-methyl group of the threonine residue (Figure S8), forcing the $\psi$ glycosidic torsion to rotate around 60° to accommodate the complete solvation shell of the antigen. The resulting conformer shows an alternative water pocket between the *N*-acetyl group of the GalNAc and the amino group of the Thr residue. Nevertheless, it is worth mentioning that the entire first hydration shell, and not only the bridging water molecule, causes this particular orientation of the glycosidic linkage in **Tn-Thr** antigen. The proposed mechanism is schematically represented in Figure 3b.

**Analysis of the Crystal Structures of Fluorinated Glycopeptides Bound to an Anti-MUC1 Antibody.** As previously


DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952–9960

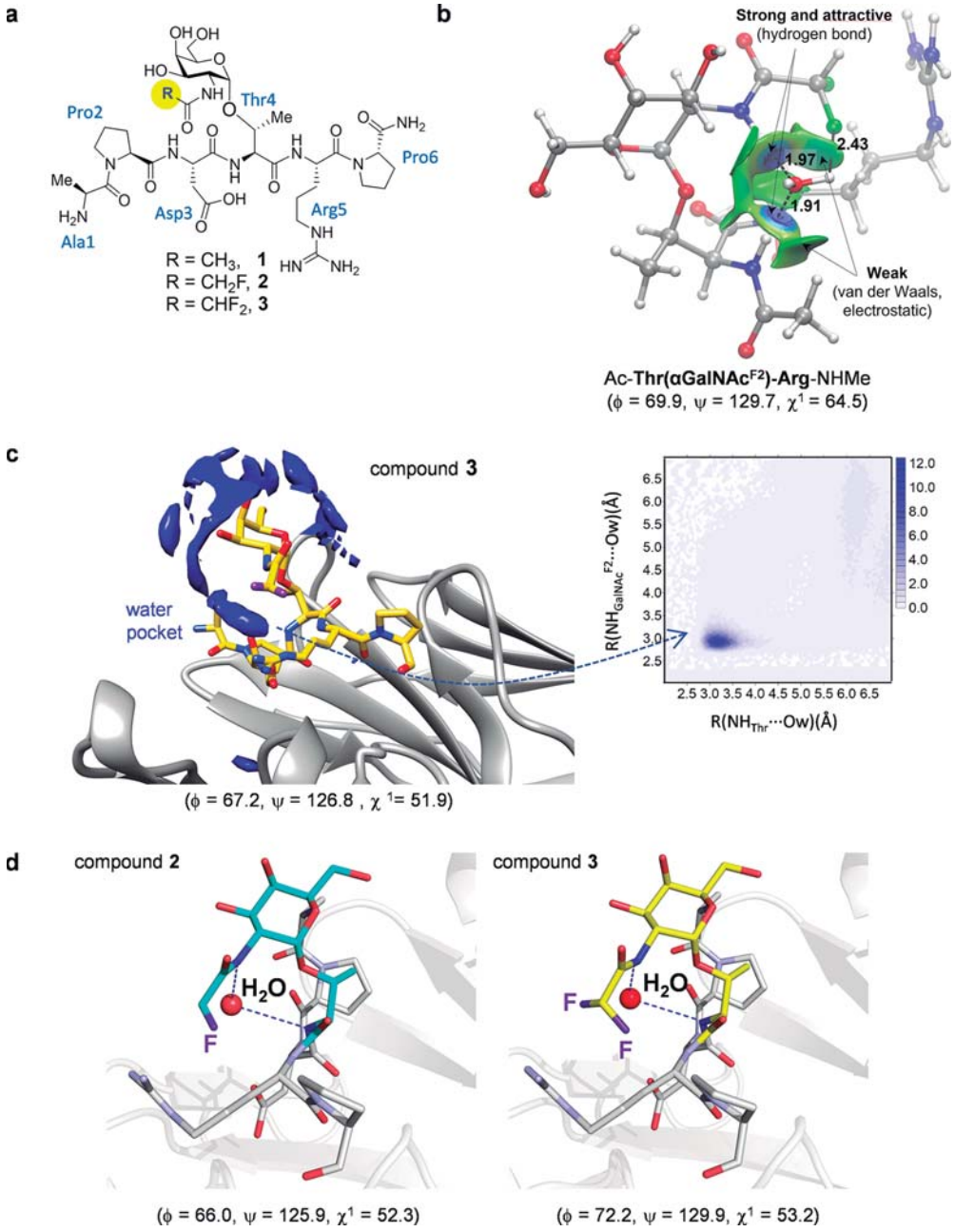



**Figure 4.** Use of fluorinated derivatives to trap bridging water molecules in the solid state. (a) Natural and fluorinated glycopeptides synthesized in this work, comprising the peptide epitope recognized by anti-MUC1 antibodies.[22] (b) Non-covalent interactions for the water molecule bound to the hydrophilic pocket of model glycopeptide **Ac-Thr(*α*GalNAc$^{F2}$)-Arg-NHMe** deduced by QM calculations (M06-2X/6-311G(d,p)). Weak attractive polar interactions between the fluorine atoms and water hydrogens stabilize the water into this pocket. The geometries were fully optimized with the PCM(water)/M06-2X/6-311G(d,p) method.[20,23] (c) Representation of the first hydration shell around the fluorinated Tn antigen derived from 200 ns MD simulations performed on glycopeptide **3** in the SM3-bound state. The 2D radial distribution function[24] calculated for the nitrogen atoms involved in the bridging water molecule is also shown. (d) Views of the binding sites of the complexes between glycopeptides **2** and **3** and the scFv-SM3 antibody (PDB IDs: 6FZR and 6FZQ, respectively), showing the key water molecule between the *N*-fluoroacetyl groups of the sugar and the amino group of the threonine residue. The geometry of the glycosidic linkage is shown in parentheses in (b), (c), and (d).

reported,[6a] the structure of the complex between the Thr-containing glycopeptide **1** (Figure 4a) when bound to an anti-MUC1 antibody (SM3)[21] did not show any bridging water molecule between the sugar and the threonine residue. Probably, the high water exchange rate precluded the experimental detection of the water molecules at this site.[8] In an attempt to detect this relevant water-mediated carbohydrate/amino acid interactions, we hypothesized that a more hydrophilic pocket would be able to bind water molecules more efficiently. Thus, the hydrogen-bonding donor character of the sugar *N*-acetyl fragment could be enhanced by replacing its constituent methyl group by fluoromethyl groups[6b] (compounds **2** and **3** in Figure 4a; see


DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

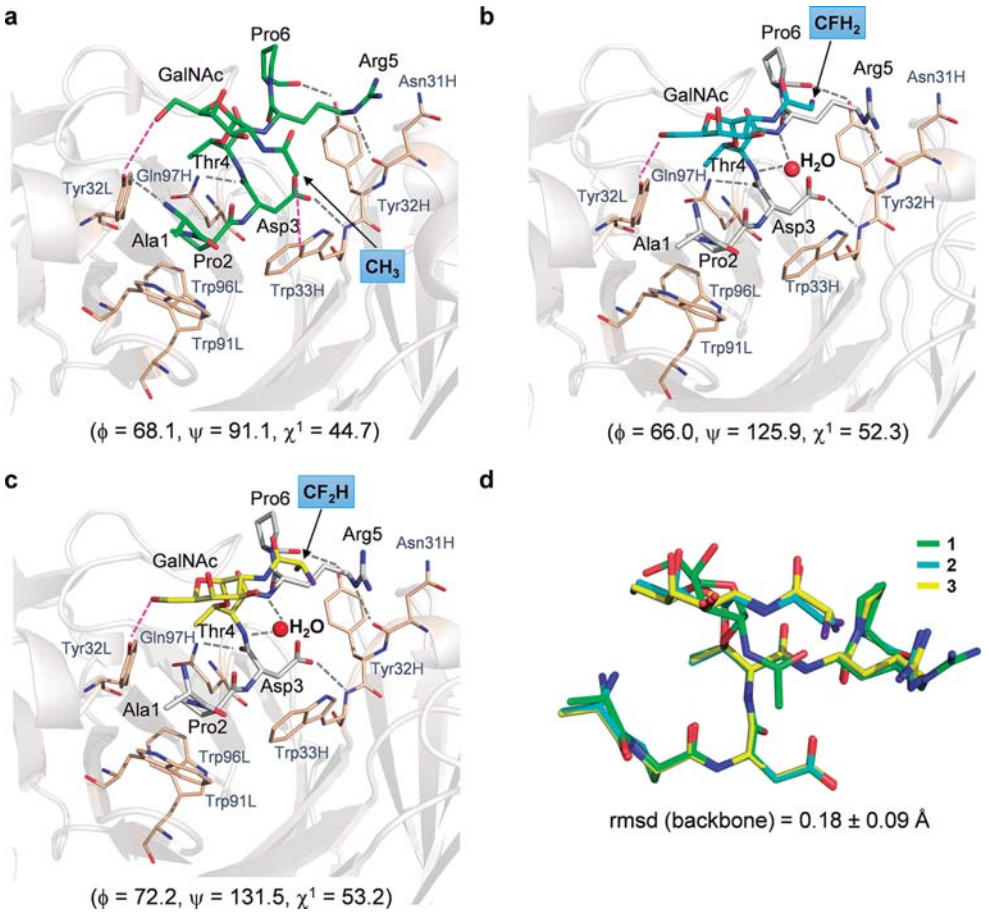



**Figure 5.** Analysis of the X-ray structures of glycopeptides **1** (ref 6), **2**, and **3** in complex with scFv-SM3. Key binding interactions of glycopeptides **1** (a), **2** (b), and **3** (c) with the antibody, as observed in the X-ray crystal structures (PDB IDs: 5A2K, 6FZR, and 6FZQ, respectively). Pink dashed lines indicate hydrophobic and hydrogen bond interactions between GalNAc and SM3 surface, and gray dashed lines indicate hydrogen bonds between peptide backbones and SM3 antibody. (d) Superposition of the peptide backbone of glycopeptides **1**−**3** in complex with SM3.

also Supporting Information, Scheme S3 and Figures S43−S56). In line with this idea, quantum mechanical calculations performed on a reduced model of **3** indicate that the water molecule is stabilized not only by two hydrogen bonds with the NH groups of the sugar and Thr residues, but also through an O—H···F contact, providing a negative water binding free energy compared to the positive value calculated for the non-fluorinated analogue (Figure 4b and Supporting Information, Figure S9).

Moreover, 200 ns MD simulations performed on glycopeptide **3** in the SM3-bound state[6,22] revealed an unusually high water density located between the sugar and the peptide units (Figure 4c), strongly suggesting that the fluorinated *N*-acetyl group indeed configures a more hydrophilic water pocket between the carbohydrate−peptide interface. In light of these theoretical predictions and to provide certain experimental evidence for the proposed solvent-mediated conformational transition in **Tn-Thr** antigen, we determined the X-ray structures of derivatives **2** and **3** bound to the antibody scFv-SM3[22] at high resolution (<2.0 Å, Figures 4a,d and 5; see also Supporting Information, Table S7 and Figure S10). To our delight, these data allowed us, for the first time, to visualize interfacial water molecules. 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.[8,25] To the best of our knowledge, this is the first confirmation of a commonly accepted hypothesis in the field of *O*-glycopeptides.

Crystallographic analysis also revealed that the surface groove of the antibody fits all the peptide residues in the three studied complexes (Figure 5a−c), irrespective of the presence of the natural or fluorinated GalNAc. The overall conformation of the peptide fragment of glycopeptides **1**−**3** is nearly identical, except for the side chain of the arginine reside in the fluorinated variants, and is similar to that found in the crystal structure reported for the naked peptide[21] (Figure 5d). The stabilizing contacts in these complexes between the antigen and the antibody involve several hydrogen bonds, some of them mediated by water molecules, as well as several stacking interactions (Figure 5a−c). For instance, while side chains of Asp3 and Arg5 in all glycopeptides are involved in hydrogen bonds with Trp33H and Asn31H, respectively, the carbonyl group of Thr4 and Pro6 are engaged in a hydrogen bond interaction with Gln97H and Tyr32H. In addition, Pro2 stacks with Trp91L, Trp96L, and Tyr32L, while side chains of Asp3 and Arg5 are engaged in hydrophobic contacts with Trp33H and Tyr32H, respectively. In contrast, while Ala1 in **1** is interacting with Tyr32L through a hydrogen bond, this interaction is not observed in the X-ray structures of the fluorinated glycopeptides. Concerning the glycosidic linkage, it adopts the common "eclipsed" conformer


DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

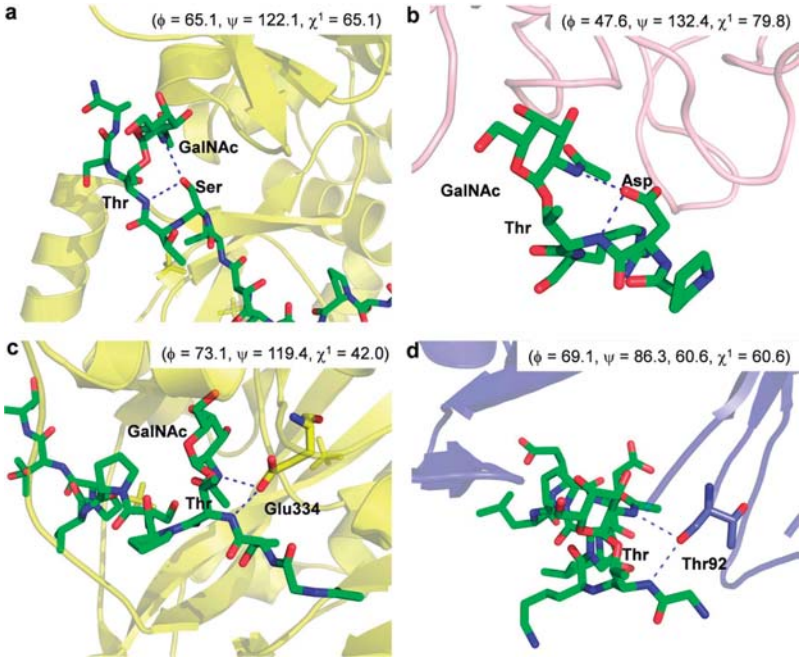



**Figure 6.** Molecular recognition of glycopeptides bearing the **Tn-Thr** antigen. The bridging water molecule is replaced by an atom of either the glycopeptide or the receptor. (a) Zoom-in of the crystal structure of the active form of GalNAc-Transferase 2 (GalNAc-T2) in complex with UDP and the glycopeptide MUC5AC-13, showing the lectin domain (PDB ID: 5AJP).[27] (b) Zoom-in of the crystal structure of soybean agglutinin from *Glycine max* in complex with the glycopeptide PDT($\alpha$GalNAc)R (PDB ID: 4D69).[30] (c) Zoom-in of the crystal structure of the inactive form of GalNAc-T2 in complex with UDP and the glycopeptide MUC5AC-3,13, showing the catalytic domain (PDB ID: 5AJO).[27] (d) Zoom-in of the crystal structure of the antibody 237 in complex with its glycopeptide epitope (PDB ID: 3IET).[28] In all cases, the geometry of the glycosidic linkage is shown in parentheses.

in the solid state for glycopeptides **2** and **3**, with $\psi$ close to 120° (Figure 5b,c), which is stabilized by the occurrence of the bridging water molecule. In the natural glycopeptide, however, the torsional angle $\psi$ takes a value close to 90°. This orientation favors a CH-$\pi$ interaction between the methyl group of GalNAc and Trp33H. In this regard, it is important to note that MD simulations performed on the complex of glycopeptide **1** and scFv-SM3 suggested that the glycosidic linkage of the antigen adopts an "eclipsed" conformation, with a value for $\psi$ around 120° in solution.[6] Moreover, the glycosidic linkage of **Tn-Thr** antigen in complex with other proteins/enzymes can adopt variable values for $\psi \approx 90°$ to 130° in the solid state (Figure 6). Finally, in the three complexes the hydroxymethyl group of GalNAc is engaged in a hydrogen bond with Tyr32L. It is important to note that the density for the fluorine atoms shown in Figures 4d and 5b–d, and Supporting Information Figure S10 is weak, impeding the accurate location of these atoms in the X-ray structures. This is likely due to the rotational mobility of the $CH_2F$ and $CHF_2$ groups. As a consequence, the crystal structures do not confirm the proposed additional O–H···F contact aforementioned and shown in Figure 4b.

**Implications for the Molecular Recognition of the Tn Antigens.** The occurrence of this persistent water pocket in the **Tn-Thr** antigen may have important implications in the stabilization of the eclipsed conformation and, in turn, in the binding of this entity to the corresponding receptors. Indeed, according to the data presented in this work and the crystal structures of receptors complexed with glycopeptides bearing this Tn antigen (Figure 6), two different scenarios are likely:

(1) The water pocket can be retained upon binding. This is the most plausible situation for MUC1-like glycopeptides bound to anti-MUC1 antibodies. In these cases, the bridging water molecules help the antigen to display the bioactive conformation in solution, therefore assisting the binding process.

(2) The "bridging" water molecule can be replaced upon binding to the biological target by an oxygen atom of either the ligand or the receptor. As a result, the eclipsed conformation observed in solution is also maintained in the bound state. According to several reported X-ray structures,[26−28] installing oxygenated amino acids in specific regions of the receptor capable of displacing the bridging water molecule to the bulk-solvent may entropically facilitate the molecular recognition[29] of glycopeptides bearing the **Tn-Thr** antigen. This novel strategy, which resembles the well-known water displacement approach in drug design, although in a reverse manner (i.e., water in the ligand is displaced by the protein receptor) could be valuable for designing receptors with an enhanced affinity toward the **Tn-Thr** antigen.

## ■ CONCLUSIONS

A multidisciplinary approach that includes the experimental and theoretical study of the Tn antigens in the gas, solution, and solid phases has been applied to deduce the key role of water in the modulation of the conformational preferences of these molecules and therefore in their presentations for interacting with protein receptors. In the **Tn-Ser** antigen, water molecules can efficiently solvate the whole molecule in the typical *exo*-anomeric/*syn* conformation also present in the gas phase. However, in the **Tn-Thr** derivative, the methyl group at $C_\beta$ disturbs the proper solvation of the "native" gas-phase


DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952–9960

geometry and $\psi$ rotates around 60° to exhibit an eclipsed conformation. In this geometry, the GalNAc moiety drastically modifies its presentation and displays an almost perpendicular arrangement with respect to the amino acid. Fittingly, this arrangement structure facilitates the efficient accommodation of a water pocket between the NH groups of the sugar and the threonine residues. This mechanism is reinforced by the observation, for the first time, of such structural water in the crystal structures of scFv-SM3 antibody in complex with two fluorinated Tn-Thr antigens. Interestingly, this peculiar arrangement of the **Tn-Thr** antigen is also observed in the bound state of this antigen to different biological receptors, including antibodies,[6,13] enzymes (GalNAc-transferases),[26,27] and lectins.[30] In contrast, for the **Tn-Ser** antigen, different arrangements of the glycosidic linkage may occur in the bound state since the lack of the $\beta$-methyl group renders a more flexible architecture.

In addition, proving the importance of the *O*-GlcNAcylation of threonine and serine residues in different biological events,[31−33] the extension of this mechanism to the $\beta$-*O*-GlcNAc-Ser and $\beta$-*O*-GlcNAc-Thr analogues, is also possible. Overall, our data provide compelling evidence of the molecular basis behind the different conformations of the **Tn-Thr** and **Thr-Ser** antigens in solution and in the enzyme/protein-bound state, which are determinant for their distinct biological functions and outcomes.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b04801.

Synthesis and characterization of **Tn-Ser′**, **Tn-Thr′**, and glycopeptides **2** and **3**; experimental data registered for **Tn-Ser′** and **Tn-Thr′** in the gas phase; computational protocols and Cartesian coordinates of the lowest energy DFT calculated structures; conformational analysis of **Tn-Ser′** and **Tn-Thr′** in solution; data collection and refinement statistics for the X-ray structures of complexes **2**/**scFv-1SM3** and **2**/**scFv-1SM3** (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Authors

*rhurtado@bifi.es
*emiliojose.cocinero@ehu.eus
*francisco.corzana@unirioja.es

### ORCID

Imanol Usabiaga: 0000-0002-1621-8536
José A. Fernández: 0000-0002-7315-2326
Jesús H. Busto: 0000-0003-4403-4790
Jesús Jiménez-Barbero: 0000-0001-5421-8513
Jesús M. Peregrina: 0000-0003-3778-7065
Gonzalo Jiménez-Osés: 0000-0003-0105-4337
Emilio J. Cocinero: 0000-0001-7632-3728
Francisco Corzana: 0000-0001-5597-8127

### Author Contributions

□I.A.B., I.U., and I.C. have contributed equally.


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank MINECO (projects CTQ2015-67727-R and UNLR13-4E-1931 to F.C. and J.M.P; CTQ2013-44367-C2-2-P and BFU2016-75633-P to R.H.-G.; CTQ2015-64597-C2-1P to J.J.-B.; CTQ2015-70524-R and RYC-2013-14706 to G.J.O; SEV-2016-0644 to CIC bioGUNE; CTQ2017-89150-R to E.J.C.; and CTQ2015-68148-C2-1-P to J.A.F.). F.C. thanks the EU (Marie-Sklodowska Curie ITN, *ProteinConjugates*). J.A.F. and E.J.C. thank UPV/EHU (UFI11/23 and PPG17/10) for financial support. I.A.B. thanks the *Asociación Española Contra el Cancer* en La Rioja for a grant. I.C. thanks Universidad de La Rioja for the FPI grant. R.H-G. thanks *Agencia Aragonesa para la Investigación y Desarrollo* (ARAID) and the *Diputación General de Aragón* (DGA, group number E34_R17) for financial support. The research leading to these results has also received funding from the FP7 (2007-2013) under BioStruct-X (grant agreement no. 283570 and BIOSTRUCTX_5186). We thank synchrotron radiation source ALBA (Barcelona) and beamline XALOC. Computational resources of CESGA, Universidad de La Rioja (BERONIA) and UPV/EHU (SGIker and I2Basque) were used in this work. We thank Prof. J. P. Simons (Oxford University) for valuable comments and scientific discussions.

## ■ REFERENCES


(1) Ju, T.; Wang, Y.; Aryal, R. P.; Lehoux, S. D.; Ding, X.; Kudelka, M. R.; Cutler, C.; Zeng, J.; Wang, J.; Sun, X.; Heimburg-Molinaro, J.; Smith, D. F.; Cummings, R. D. *Proteomics: Clin. Appl.* **2013**, *7*, 618−631.

(2) Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem., Int. Ed.* **2011**, *50*, 1770−1791.

(3) Springer, G. F. *J. Mol. Med.* **1997**, *75*, 594−602.

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

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

(6) (a) Martínez-Sáez, N.; Castro-López, J.; Valero-Gónzalez, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem., Int. Ed.* **2015**, *54*, 9830−9834. (b) Unione, L.; Alcalá, M.; Echeverria, B.; Serna, S.; Ardá, A.; Franconetti, A.; Cañada, F. J.; Diercks, T.; Reichardt, N.; Jiménez-Barbero, J. *Chem. - Eur. J.* **2017**, *23*, 3957−3965.

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

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

(9) 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. *Chem. - Eur. J.* **2014**, *20*, 12616−12627.

(10) 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. *Cancer Immunol. Immunother.* **2013**, *62*, 1107−1122.

(11) Zhang, Y.; Li, Q.; Rodriguez, L. G.; Gildersleeve, J. C. *J. Am. Chem. Soc.* **2010**, *132*, 9653−9662.

(12) Kanekura, T.; Sakuraba, H.; Matsuzawa, F.; Aikawa, S.; Doi, H.; Hirabayashi, Y.; Yoshii, N.; Fukushige, T.; Kanzaki, T. *J. Dermatol. Sci.* **2005**, *37*, 15−20.

(13) 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.; Garcia-Martin, F.; Marcelo, F. *J. Am. Chem. Soc.* **2015**, *137*, 12438−12441.




DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

(14) Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S.-I. *Angew. Chem., Int. Ed.* **2004**, *43*, 856−862.
(15) Corzana, F.; Busto, J. H.; Engelsen, S. B.; Jiménez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Avenoza, A. *Chem. - Eur. J.* **2006**, *12*, 7864−7871.
(16) (a) Cocinero, E. J.; Stanca-Kaposta, E. C.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *J. Am. Chem. Soc.* **2009**, *131*, 1282−1287. (b) Barry, C. S.; Cocinero, E. J.; Carçabal, P.; Gamblin, D. P.; Stanca-Kaposta, E. C.; Remmert, S. M.; Fernández-Alonso, M. C.; Rudić, S.; Simons, J. P.; Davis, B. G. *J. Am. Chem. Soc.* **2013**, *135*, 16895−16903. (c) Cocinero, E. J.; Stanca-Kaposta, E. C.; Dethlefsen, M.; Liu, B.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Chem. - Eur. J.* **2009**, *15*, 13427−13434.
(17) Cocinero, E. J.; Carçabal, P.; Vaden, T. D.; Simons, J. P.; Davis, B. G. *Nature* **2011**, *469*, 76−79.
(18) León, I.; Millán, J.; Cocinero, E. J.; Lesarri, A.; Fernández, J. A. *Angew. Chem., Int. Ed.* **2013**, *52*, 7772−7775.
(19) Csonka, G. I.; Schubert, G. A.; Perczel, A.; Sosa, C. P.; Csizmadia, I. G. *Chem. - Eur. J.* **2002**, *8*, 4718−4733.
(20) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(21) Dokurno, P.; Bates, P. A.; Band, H. A.; Stewart, L. M.; Lally, J. M.; Burchell, J. M.; Taylor-Papadimitriou, J.; Snary, D.; Sternberg, M. J.; Freemont, P. S. *J. Mol. Biol.* **1998**, *284*, 713−728.
(22) Karsten, U.; Serttas, N.; Paulsen, H.; Danielczyk, A.; Goletz, S. *Glycobiology* **2004**, *14*, 681−692.
(23) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3094.
(24) Andersson, C.; Engelsen, S. B. *J. Mol. Graphics Modell.* **1999**, *17*, 101−105.
(25) Corzana, F.; Busto, J. H.; García de Luis, M.; Fernández-Tejada, A.; Rodríguez, F.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. *Eur. J. Org. Chem.* **2010**, *2010*, 3525−3532.
(26) Lira-Navarrete, E.; Iglesias-Fernández, J.; Zandberg, W. F.; Compañón, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R. *Angew. Chem., Int. Ed.* **2014**, *53*, 8206−8210.
(27) 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. *Nat. Commun.* **2015**, *6*, 6937.
(28) Brooks, C. L.; Schietinger, A.; Borisova, S. N.; Kufer, P.; Okon, M.; Hirama, T.; Mackenzie, C. R.; Wang, L.-X.; Schreiber, H.; Evans, S. V. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10056−10061.
(29) Persch, E.; Dumele, O.; Diederich, F. *Angew. Chem., Int. Ed.* **2015**, *54*, 3290−3327.
(30) Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; Coelho, H.; Valero-Gónzalez, J.; Castro-López, J.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Corzana, F.; Peregrina, J. M. *ACS Chem. Biol.* **2015**, *10*, 747−756.
(31) Brister, M. A.; Pandey, A. K.; Bielska, A. A.; Zondlo, N. J. *J. Am. Chem. Soc.* **2014**, *136*, 3803−3816.
(32) Rani, L.; Mallajosyula, S. S. *J. Phys. Chem. B* **2017**, *121*, 10618−10638.
(33) Fernández-Tejada, A.; Corzana, F.; Busto, J. H.; Jiménez-Oses, G.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. *Chem. - Eur. J.* **2009**, *15*, 7297−7301.


DOI: 10.1021/jacs.8b04801
*J. Am. Chem. Soc.* 2018, 140, 9952−9960

ACS central science

Research Article

Cite This: ACS Cent. Sci. 2018, 4, 1274−1290

http://pubs.acs.org/journal/acscii

# Structural and Mechanistic Insights into the Catalytic-Domain-Mediated Short-Range Glycosylation Preferences of GalNAc-T4

Matilde de las Rivas,[†,▰] Earnest James Paul Daniel,[‡,▰] Helena Coelho,[§,∥,⊥] Erandi Lira-Navarrete,[#] Lluis Raich,[▽] Ismael Compañón,[○] Ana Diniz,[§] Laura Lagartera,[◆] Jesús Jiménez-Barbero,[∥,⊥,¶] Henrik Clausen,[#] Carme Rovira,[▽,⊡] Filipa Marcelo,[§] Francisco Corzana,[○] Thomas A. Gerken,[*,‡] and Ramon Hurtado-Guerrero[*,†,◧]

[†]BIFI, University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit, Mariano Esquillor s/n, Campus Rio Ebro, Edificio I+D, Zaragoza 50018, Spain

[‡]Departments of Biochemistry, Pediatrics and Chemistry, Case Western Reserve University, Cleveland, Ohio 44106, United States

[§]UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade de Nova de Lisboa, Caparica 2825-149, Portugal

[∥]CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain

[⊥]Departament of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, Leioa, 48940 Bizkaia, Spain

[#]Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, School of Dentistry, University of Copenhagen, Copenhagen 1165, Denmark

[▽]Departament de Química Inorgànica i Orgànica (secció de Química Orgànica) & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

[○]Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, E-26006 Logroño, Spain

[◆]Instituto de Química Médica, IQM-CSIC, 28006 Madrid, Spain

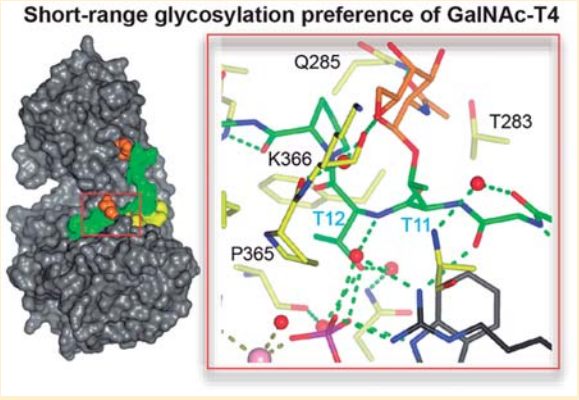
[¶]Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain

[⊡]Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, 08010 Barcelona, Spain

[◧]Fundación ARAID, 50018 Zaragoza, Spain

Supporting Information

**ABSTRACT:** Mucin-type *O*-glycosylation is initiated by a family of polypeptide GalNAc-transferases (GalNAc-Ts) which are type-II transmembrane proteins that contain Golgi luminal catalytic and lectin domains that are connected by a flexible linker. Several GalNAc-Ts, including GalNAc-T4, show both long-range and short-range prior glycosylation specificity, governed by their lectin and catalytic domains, respectively. While the mechanism of the lectin-domain-dependent glycosylation is well-known, the molecular basis for the catalytic-domain-dependent glycosylation of glycopeptides is unclear. Herein, we report the crystal structure of GalNAc-T4 bound to the diglycopeptide GAT*GAGA-GAGT*TPGPG (containing two $\alpha$-GalNAc glycosylated Thr (T*), the PXP motif and a "naked" Thr acceptor site) that describes its catalytic domain glycopeptide GalNAc binding site. Kinetic studies of wild-type and GalNAc binding site mutant enzymes show the lectin domain GalNAc binding activity dominates over the catalytic domain GalNAc binding activity and that these activities can be independently eliminated. Surprisingly, a flexible loop protruding from the lectin domain was found essential for the optimal activity of the catalytic domain. This work provides the first structural basis for the short-range glycosylation preferences of a GalNAc-T.




Mucin-type *O*-glycosylation (*O*-GalNAc-type) is one of the most diverse and complex types of protein *O*-glycosylation found in higher eukaryotes and is likely the most abundant, with over 80% of all proteins passing through the

Received: July 23, 2018
Published: September 14, 2018






DOI: 10.1021/acscentsci.8b00488
ACS Cent. Sci. 2018, 4, 1274−1290

Table 1. Peptide Acceptor Substrates Used in This Study[a]

| Peptide | Sequence |
|---|---|
| 1 (--TT--) | GAGAGAGTTPGPG |
| 2 (-TT--T*-) | AGAGTTPGPGAGAT*GA |
| 3 (-T*--TT-) | GAT*GAGAGAGTTPGPG |
| 4 (-T*T-) | GAGAGAGT*TPGPG |
| 5 (-T*T--T*-) | AGAGT*TPGPGAGAT*GA |
| 6 (-T*--T*T-) | GAT*GAGAGAGT*TPGPG |

[a]Note: T* denotes the Thr-*O*-GalNAc moiety.

secretory pathway predicted to be *O*-glycosylated.[1] This post-translational modification is important in a wide range of cellular/biological processes,[1] including phosphate homeostasis and bone mineralization,[2] regulation of HDL and LDL,[3,4] tumorigenesis and formation of metastasis,[5,6] organogenesis and development,[7] and the fundamental process of ectodomain shedding and cell signaling.[8,9] In addition, alteration of the cellular location and regulation of the activity of the GalNAc-T isoenzymes have clear implications in disease.[1] Mucin-type *O*-glycosylation (henceforth called *O*-glycosylation) is initiated by a large family of GalNAc-T isoenzymes (20 members in humans and similar numbers in most mammals) that are suggested to act in a hierarchical manner. GalNAc-Ts are retaining glycosyltransferases (GTs) that transfer a GalNAc moiety from uridine diphosphate *N*-acetylgalactosamine (UDP-GalNAc) onto Ser/Thr residues of proteins.[1] The GalNAc-Ts are unique among metazoan GTs because in addition to their N-terminal catalytic domain adopting a GT-A fold, they possess a C-terminal GalNAc binding lectin domain with a $\beta$-trefoil fold, which provides additional functions to these enzymes.[10−12] The two domains are linked through a short flexible linker whose motion has been suggested to be responsible for the dynamic conformational landscape of these enzymes.[13] Three distinctive modes of substrate glycosylation have been reported:[10] (1) glycosylation of "naked" unglycosylated peptides; (2) glycosylation of GalNAc-containing glycopeptides where the sites of glycosylation are 1−3 residues away from the prior glycosite, hereafter termed short-range or neighboring glycosylation; and (3) glycosylation of GalNAc glycopeptides where the sites of glycosylation are ∼5−17 residues away from the prior glycosite, termed long-range or remote glycosylation.[10] The latter, long-range glycosylation activity is due to glycopeptide GalNAc binding at the lectin domain, directing the acceptor peptide onto the catalytic domain in an N- and/or C-terminal direction.[10,14] Most isoforms that contain a lectin domain have been found to possess this remote glycosylation activity.[10] On the other hand, sites adjacent and neighboring to an existing GalNAc glycosite are glycosylated in a catalytic-domain-dependent manner, with GalNAc-T4 being part of a small number of GalNAc-Ts (including GalNAc-T7, -T10, and -T12) that have been shown to glycosylate contiguous or nearby sites.[10,11,15] While GalNAc-T4 glycosylates acceptor sites that are directly C-terminal to a prior GalNAc glycosite, GalNAc-T7 and T10 glycosylate sites that are directly N-terminal from a prior site of GalNAc glycosylation. In addition, GalNAc-T12 glycosylates sites located three residues N-terminal from a prior GalNAc glycosite.[10] Thus, it is likely that a GalNAc binding site should exist for those isoenzymes that possess short-range glycosylation preferences on glycopeptides. While the lectin domain GalNAc binding sites have been well-characterized by others and us for several GalNAc-Ts,[16,17] the catalytic domain GalNAc binding site has never been revealed at the structural level. In addition, it is not known how GalNAc-T4, -T7, and -T10 may glycosylate contiguous acceptor sites in glycopeptides exclusively using a catalytic-domain-dependent manner.

Previous structural work from our lab with GalNAc-T2 in complex with "naked" peptides has provided insight into the catalytic-domain-mediated transfer of GalNAc from the UDP-GalNAc donor to the peptide acceptor. In addition, our studies on remotely glycosylated glycopeptides have demonstrated how the GalNAc-T2 lectin domain guides N-terminal acceptor sites into the catalytic domain for catalysis to take place, by binding to a distant prior GalNAc glycosite located at the C-terminus of the substrate.[12,13,18] Recently, we have also demonstrated how other GalNAc-T isoforms such as GalNAc-T3, -T4, -T6, and -T12 achieve the opposite long-range glycosylation preferences.[10,19] In that study we showed that the connecting interdomain flexible linker dictates the orientation of the lectin domain with respect to the catalytic domain. Thus, the different positions of the lectin domain GalNAc binding site of GalNAc-T4 relative to GalNAcT2 readily explained how these two isoenzymes achieved their opposite and distinct long-range glycosylation preferences.[14]

To date however no structural studies have revealed how the catalytic domains of any of the GalNAc-Ts that possess neighboring or adjacent glycopeptide activities actually accommodate an acceptor peptide with a GalNAc residue immediately adjacent to the acceptor site. Such knowledge would help us to understand how these transferases perform their so-called filling-in activities, i.e., completing the glycosylation of heavily *O*-glycosylated mucin domains. The GalNAc-T4 isoform is of particular interest as it is the only isoform shown capable of glycosylating two out of the five acceptor sites in the partially glycosylated MUC1 mucin tandem repeat.[11] The density of glycosylation, together with the structure of *O*-glycans in the mucin tandem repeat regions, are important features for the exploitation of the cancer related glycoforms of MUC1 as diagnostic as well as therapeutic purposes.[20] Generally, GalNAc-T4 glycosylates very few isolated sites in "naked" peptide acceptors.[21]

We report herein a multidisciplinary approach combining different structural and biophysical techniques, with enzyme



DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

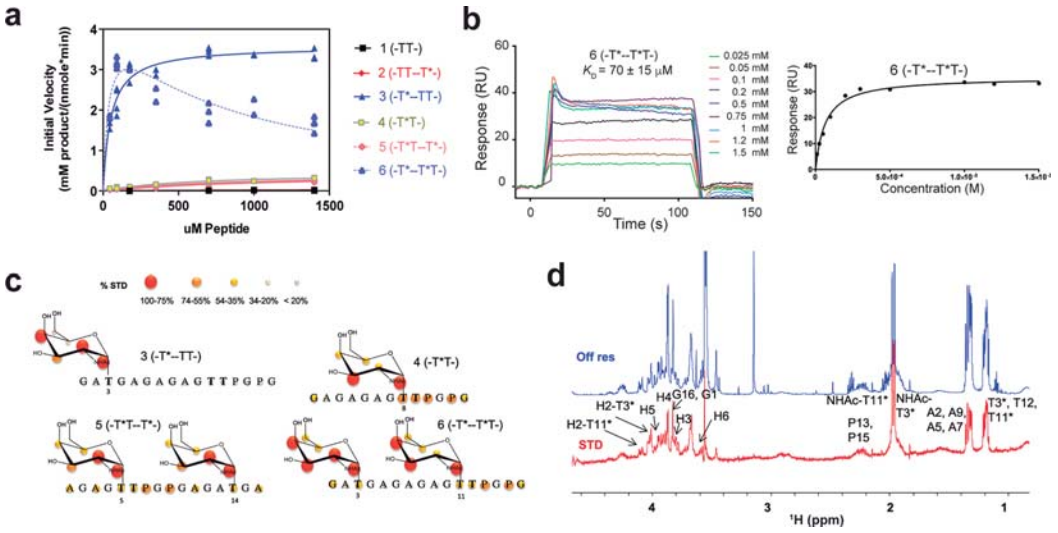



**Figure 1.** Biophysical characterization of GalNAc-T4. (a) Peptide glycosylation kinetics of GalNAc-T4 against (glyco)peptides **1**−**6** (see also Figure 4a). Michaelis−Menten kinetic values, $K_m$, $V_{max}$, and catalytic efficiency ($V_{max}/K_m$) for glycopeptides **3**−**6** were obtained from the nonlinear least-squares fit to the initial rate data, obtained as described in the Methods section and given in Table 2. Peptide substrates **1** and **2** are largely unglycosylated by GalNAc-T4.[14] (b) Left panel: SPR sensogram for binding of glycopeptide **6** to GalNAc-T4. Right panel: Fitting of the SPR binding data giving a $K_d$ of 70 ± 15 $\mu$M. (c) Mapping of substrate binding epitopes by saturation transfer difference (STD) NMR. The size of the colored spheres represents the normalized STD-NMR intensity (i.e., binding) observed for the indicated protons/residues. For sake of clarity, the STD response given for the indicated amino acid residues corresponds to the average of STD for all of the protons in the residue that could be accuracy measured. See Figures S4−S6 for the detailed STD-NMR enhancements of the identified residues/protons. Note that in addition to the GalNAc protons amino acid protons in the **-T*TPGP-** sequence also gave STD-NMR enhancements. (d) Representative 600 MHz $^1$H NMR spectra of glycopeptide **6** (**-T*--T*T-**) at 877 $\mu$M in the presence of 13.5 $\mu$M GalNAc-T4, 75 $\mu$M UDP, and 75 $\mu$M $MnCl_2$ obtained at 298 K. The off resonance reference spectrum (labeled Off res) is displayed in blue, and the on resonance STD spectrum (labeled STD) is in red. Key proton resonances are labeled in the STD spectrum. Note the different STD responses for the identified GalNAc H2 protons of the glycosylated Thr3 and Thr11 found between 4.0 and 4.1 ppm of the STD spectrum.

kinetics analysis and computational studies on GalNAc-T4. Our findings combined with the kinetic characterization of a library of (glyco)peptides, against the wild-type and inactivating catalytic and lectin domain mutant transferases, have begun to reveal the molecular basis of the short-range glycosylation preferences of the GalNAc-T4 on glycopeptide substrates. The experiments further reveal the dominant role of long-range lectin domain glycopeptide binding over short-range catalytic domain glycopeptide binding in the overall glycosylation process of GalNAc-T4. In addition, we determine the role of a flexible loop protruding from the lectin domain as an important structural feature essential for the optimal activity of the catalytic domain.

## RESULTS AND DISCUSSION

**Kinetics of GalNAc-T4 against Glycopeptide Substrates.** We have previously used three model (glyco)peptides (Table 1) to compare the long-range glycosylation preferences and enzyme kinetics of GalNAc-T2 and -T4.[14] Those were the "naked" peptide **1** (denoted **-TT-**, for simplicity), and the C- and N-terminal glycosylated monoglycopeptides, **2** and **3** (denoted **-TT--T*-** and **-T*--TT-**, respectively, where T* represents a GalNAc-glycosylated Thr, see Table 1 and Figure 1a).[14] Here, we have expanded this (glyco)peptide library with monoglycopeptide **4** (denoted **-T*T-**) and diglycopeptides **5** and **6** (denoted **-T*T--T*-** and **-T*--T*T-**, respectively) (Table 1). The diglycopeptides were designed to evaluate the combined effects of having long-range and short-range prior glycosylation in the same substrate. Note that all substrates in Table 1 have one or two potential Thr acceptor sites (i.e., **-T*T-** or **-TT-**) that also contain an adjacent C-terminal PXP motif (as -PGP-), which is recognized by most GalNAc-Ts.[10,13] These peptides also share the same amino acid sequence around the acceptor site, thus eliminating the effects of peptide sequence variation on catalysis (Table 1). The short sequences surrounding the remote prior T* glycosylation site are also identical, thus ensuring nearly identical lectin domain binding properties. Note that the remote prior T* glycosylation site is located 7−8 residues away from the potential Thr acceptor sites, which is in agreement with the optimal distance of 7−11 residues observed for the long-range glycosylation of GalNAc-T4.[10] Likewise the neighboring N-terminal prior glycosylation preference of GalNAc-T4,[10] i.e., **-T*T-**, was included in glycopeptides **5** and **6** in order to examine the combined effects of having both long- and short-range prior glycosylation in one glycopeptide. For simplicity we did not include in our present studies peptides bearing Ser acceptor sites or S*, since Ser is usually less efficiently glycosylated than Thr,[18] and S* displays distinct conformational preferences in contrast to T*, both in the free state and bound to proteins.[22] According to our previous crystal structure of GalNAc-T4 complexed with peptide **3** (**-T*--TT-**),[14] the remote N-terminal glycosites (i.e., T*) of both peptides **3** and **6** (**-T*--TT-** and **-T*--T*T-**) would bind to the GalNAc binding site of the lectin domain, while directing the C-terminal of the peptide acceptor onto the catalytic domain for catalysis. In contrast, the remote C-terminal glycosite of peptides **2** (**-TT—T*-**) and **5** (**-T*T—T*-**), when bound to the lectin domain, would not be expected to correctly orient the N-terminal portion of the acceptor peptide onto the catalytic domain for efficient GalNAc transfer.



DOI: 10.1021/acscentsci.8b00488
ACS Cent. Sci. 2018, 4, 1274−1290

Table 2. Kinetic Parameters for the Wild-Type and Mutant GalNAc-T4

| | | 3(-T*--T-) (Michaelis Menten fit)[a] | 4 (-T*T-) (Michaelis Menten fit)[a] | 5(-T*T--T*-) (Michaelis Menten fit)[a] | 6(-T*--T*T-) (Substrate Inhibition fit)[b] |
|---|---|---|---|---|---|
| Wild-type | Vmax (mM/nmol*min) | 3.6 ± 0.08 | 0.45 ± 0.03 | 0.41 ± 0.07 | 4.4 ± 0.6 |
| | Km (μM) | 50.7 ± 6.3 | 612 ± 99 | 920 ± 309 | 39.4 ± 18 |
| | Ki (μM) | na | na | na | 710 ± 230 |
| | Cat Eff (nmol*min)$^{-1}$ [c] | 71 ± 11 | 0.74 ± 0.17 | 0.45 ± 0.25 | 111 ± 85 |
| | $R^2$ curve fit | 0.89 | 0.95 | 0.93 | 0.75 |
| | UDP-GalNAc Hydrolysis % [d] | 5 | 21 | 20 | 6 |
| Lectin Mutant (D459H) | Vmax (mM/nmol*min) | 0.03 ± 0.01 | 0.92 ± 0.08 | 0.49 ± 0.05 | 0.88 ± 0.3 |
| | Km (μM) | 1009 ± 492 | 644 ± 132 | 617 ± 148 | 878 ± 101 |
| | Ki (μM) | na | na | na | 506 ± 158 |
| | Cat Eff (nmol*min)$^{-1}$ [c] | 0.03 ± 0.03 | 1.43 ± 0.44 | 0.79 ± 0.29 | 1.01 ± 0.46 |
| | $R^2$ curve fit | 0.77 | 0.90 | 0.89 | 0.71 |
| | UDP-GalNAc Hydrolysis % [d] | 70 | 12 | 15 | 22 |
| Catalytic Mutant (T283S-Q285A) | Vmax (mM/nmol*min) | 4.3 ± 0.15 | 0.31 ± 0.06 | 0.18 ± 0.02 | 9.2 ± 1.9 |
| | Km (μM) | 98.4 ± 15 | 1350 ± 535 | 1129 ± 382 | 154.3 ± 46 |
| | Ki (μM) | na | na | na | 435 ± 119 |
| | Cat Eff (nmol*min)$^{-1}$ c | 44 ± 8.2 | 0.23 ± 0.16 | 0.16 ± 0.08 | 60 ± 33 |
| | $R^2$ curve fit | 0.87 | 0.82 | 0.85 | 0.82 |
| | UDP-GalNAc Hydrolysis % [d] | 2 | 25 | 28 | 2 |
| Lectin Flexible Loop (LFL) deletion/mutant ($P_{463}$DNNP$_{467}$ to GGG) | Vmax (mM/nmol*min) | 4.4 ± 0.14 | 0.29 ± 0.01 | 0.19 ± 0.03 | 8.7 ± 0.9 |
| | Km (μM) | 67.5 ± 10 | 720 ± 253 | 821 ± 351 | 23.5 ± 7.7 |
| | Ki (μM) | na | na | na | 381 ± 73 |
| | Cat Eff (nmol*min)$^{-1}$ [c] | 65 ± 12 | 0.40 ± 0.18 | 0.23 ± 0.17 | 370 ± 179 |
| | $R^2$ curve fit | 0.89 | 0.85 | 0.84 | 0.92 |
| | UDP-GalNAc Hydrolysis % [d] | 11 | 48 | 51 | 13 |
| Catalytic/LFL Mutant (T283S-Q285A-D464A) | Vmax (mM/nmol*min) | 5.5 ± 0.1 | 0.07 ± 0.01 | 0.04 ± 0.01 | 9.3 ± 0.8 |
| | Km (μM) | 99.2 ± 11 | 233 ± 45 | 351 ± 86 | 44.1 ± 11 |
| | Ki (μM) | na | na | na | 720 ± 145 |
| | Cat Eff (nmol*min)$^{-1}$ [c] | 55.5 ± 7 | 0.30 ± 0.1 | 0.11 ± 0.05 | 211 ± 78 |
| | $R^2$ curve fit | 0.93 | 0.90 | 0.87 | 0.87 |
| | UDP-GalNAc Hydrolysis % [d] | 8 | 50 | 58 | 11 |
| Lectin/Catalytic Mutant (T283S-Q285A-D459H) | Vmax (mM/nmol*min) | 0.04 ± 0.01 | 0.12 ± 0.01 | 0.1 ± 0.01 | 0.04 ± 0.005 |
| | Km (μM) | 1493 ± 585 | 581 ± 87 | 919 ± 270 | 212 ± 59 |
| | Ki (μM) | na | na | na | 355 ± 355 |
| | Cat Eff (nmol*min)$^{-1}$ [c] | 0.03 ± 0.02 | 0.21 ± 0.05 | 0.11 ± 0.05 | 0.19 ± 0.08 |
| | $R^2$ curve fit | 0.80 | 0.97 | 0.90 | 0.61 |
| | UDP-GalNAc Hydrolysis % [d] | 50 | 18 | 16 | 27 |

[a]Kinetic values and $R^2$ obtained from the GraphPad Michaelis−Menten fit of the plots in Figures 1a and 4. [b]Kinetic values and $R^2$ obtained from the GraphPad Michaelis−Menten fitting with substrate inhibition of the plots in Figure 4. [c]Cat Eff, catalytic efficiency obtained from the $V_{max}/K_m$ ratio. [d]UDP-GalNAc hydrolysis obtained from Sephadex G10 chromatography summarized in Figure S8.

To test the rationale above, detailed enzyme kinetics of GalNAc-T4 against the (glyco)peptides listed in Table 1 was performed using UDP-$^3$H-GalNAc as donor (see the Methods section for details). Plots of initial rates versus substrate concentration are given in Figure 1a, while the obtained kinetic constants are listed in Table 2. Note that the results for peptides **3**, **4**, and **5** could be fitted to a standard Michaelis−Menten model, while peptide **6** was best fitted to a model which included apparent substrate inhibition and will be discussed below. As can be readily seen from Figure 1a, glycopeptides **3** (-**T***--**TT**-) and **6** (-**T***--**T***T-) show the largest activity, which is ∼10-fold higher than that for glycopeptides **4** (-**T***T) and **5** (-**T***T--**T***-). As previously reported, peptides **1** (-**TT**-) and **2** (-**TT**--**T***-) were imperceptibly glycosylated.[14] Note also that only the second Thr is glycosylated in those peptides containing the -**TTPGP**- sequence.[14] These results are consistent with our previous findings that GalNAc-T4 displays a directly neighboring N-terminal prior glycosylation preference, as well as a long-range N-terminal prior glycosylation activity, while showing poor activity against the naked peptide or glycopeptides containing only a C-terminal remote prior glycosylation.[10] Thus, for glycosylation to take place in glycopeptide **4** (-**T***T), it can be concluded that the neighboring glycosylated Thr must directly bind to the catalytic domain as the acceptor Thr is contiguous to this site. The similar plots and kinetic constants obtained for diglycopeptide **5** (-**T***T--**T***-) and monoglycopeptide **4** (-**T***T-), (Figure 1a and Table 2) suggest that the glycosylation of diglycopeptide **5** may be dominated by -**T***T- binding at the catalytic domain rather than binding of the remote C-terminal T* at the lectin domain. This agrees with the imperceptibly low activity of glycopeptide **2** (-**TT**--**T***-), which we take as evidence that the C-terminal glycosite either did not to bind the lectin domain, or if bound, it failed to correctly orient the N-terminal acceptor region of the substrate into the catalytic domain.[14] However, as discussed below, our kinetic studies of the GalNAc-T4 lectin mutant suggest that the lectin domain may play some role in the observed activity of diglycopeptide **5**.

The comparison of the two most active monoglycopeptides **3** (-**T***--**TT**-) and **4** (-**T***T-) reveals the dominant influence of the remote N-terminal prior glycosylation on enzyme activity,



DOI: 10.1021/acscentsci.8b00488
ACS Cent. Sci. 2018, 4, 1274−1290

where peptide **3** has a ~8-fold higher $V_{max}$, ~12-fold lower $K_m$, and a ~100-fold higher catalytic efficiency ($V_{max}/K_m$) compared to peptide **4** (Table 2). This influence of the lectin domain is also observed for diglycopeptide **6** (**-T*--T*T-**), which possesses both remote and neighboring glycosites N-terminal of the acceptor site (Figure 1a). However, the kinetic plot for diglycopeptide **6** was best fitted to a Michaelis–Menten model with substrate inhibition (Figure 1a, Table 2). The $K_m$ of peptide **6** (**-T*--T*T-**) is ~1.3-fold lower than that for peptide **3** (**-T*--TT-**) (~39 vs ~51 $\mu$M) suggesting a weak synergistic effect of the two glycosites in peptide **6**, each binding separately to the lectin and catalytic domains of the enzyme. This fact implies that substrate binding to the enzyme is slightly stronger when both the lectin and catalytic domains bind GalNAc residues with the appropriate N-terminal placement, as found for the diglycosylated substrate. This synergistic effect is clearly observed in the substrate $K_d$ values obtained by surface plasmon resonance (SPR) in the presence of excess of UDP and $MnCl_2$ (see the Methods section). While the $K_d$ values for both monoglycopeptides **4** (**-T*T-**) and **3** (**-T*--TT-**) were relatively weak (in the mM range; see Figure S1 and ref 14), the $K_d$ value for diglycopeptide **6** (**-T*--T*T-**) was 70 ± 15 $\mu$M (Figure 1b). Interestingly, the observed $K_m$ values for these three glycopeptides are lower than their $K_d$ values, especially for glycopeptides **3** and **6**. The discrepancies between GalNAc-T4's higher-affinity $K_m$ values obtained from our kinetics studies (Table 2) and the $K_d$ values obtained from direct binding studies, particularly for peptide **3**, can be attributed to the fact that the UDP-GalNAc donor was present in the enzyme kinetics, but absent in the SPR binding studies.[14] UDP-GalNAc stabilizes the so-called flexible loop of the catalytic domain active site in a "closed" conformation, which completes the formation of the peptide-binding groove and leads to an active transferase.[13,14,18] Thus, our SPR studies in the absence of UDP-GalNAc likely represent the weak binding of peptide substrate to the "open" flexible loop conformation of the enzyme. However, this discrepancy is not as large for peptide **6**, where $K_d$ is only ~2-fold higher than the $K_m$ observed from the kinetic analysis. This suggests that the high binding affinity of peptide **6** (due to GalNAc binding at both domains) might further stabilize/drive the GalNAc-T4's catalytic domain flexible loop in a closed conformation in the presence of either UDP-GalNAc or UDP (as confirmed by the crystal structure discussed below). The synergistic effects observed in the kinetic and $K_d$ values obtained for diglycopeptide **6** (**-T*--T*T-**) compared to monoglycopeptides **3** (**-T*--TT-**) and **4** (**-T*T**) are due to the divalency of peptide **6**, which may simultaneously or consecutively bind to both the lectin and catalytic domains of the transferase. Such synergy is common to multivalent carbohydrate–lectin binding systems.[23,24] In summary, the initial catalytic efficiency ($V_{max}/K_m$) of peptide **6** (**-T*--T*T-**) is ~1.5- and ~150-fold higher than the catalytic efficiency of peptides **3** (**-T*--TT-**) and **4** (**-T*T-**), respectively, showing again that the remote T* binding to the lectin domain dominates overall catalytic efficiency. It is also worth noting that the much higher catalytic efficiencies of glycopeptides **3** and **6** compared to those deduced for glycopeptides **4** and **5** correlate with their extent of nonproductive hydrolysis of UDP-GalNAc by the transferase, where glycopeptides **3** and **6** display only ~5% hydrolysis, while glycopeptides **4** and **5** display ~20% hydrolysis (Table 2 and the Methods section). Likewise, peptides **1** and **2**, with extremely low catalytic efficiencies, display even higher rates of hydrolysis of ~40–50% (Figure S8).[14]

As mentioned above, the kinetic plot for diglycopeptide **6** (**-T*--T*T-**) is best fitted to a Michaelis–Menten kinetic model that includes substrate inhibition (Figure 1a and Table 2). The appearance of substrate inhibition at high diglycopeptide **6** concentration could be due to the nonproductive binding of the neighboring GalNAc in the **-T*TPGP-** sequence at the lectin domain. Such a binding event would compete with the "productive" binding of the remote GalNAc (**-T*-**) at the lectin domain and would lead to the observed decrease in activity at high substrate concentration. This explanation implies that the binding affinity of the **-T*TPGP-** epitope at the lectin domain is weaker than that of the remote T* GalNAc of glycopeptide **6**. In this case, the glycosylation rate at high substrate concentration, would ideally be ~1/2 of the initial maximum rate, as indeed observed (Figure 1a).

**GalNAc-T4–Substrate Interactions with (Glyco)-peptide Substrates by STD-NMR.** To further characterize the different modes of glycopeptide substrate binding to GalNAc-T4, we performed saturation transfer difference (STD) NMR experiments on the substrates in Table 1, in the presence of GalNAc-T4, UDP, and $MnCl_2$, as described in the Methods section. Our previous STD-NMR studies with peptides **1**–**3** (**-TT-**, **-TT--T*-**, **-T*--TT-**)[14] have revealed different features: (1) No significant STD enhancements were observed for peptide **1** (**-TT-**), suggesting that it poorly binds GalNAc-T4, in agreement with its very low glycosylation rate. (2) Poor recognition of the peptide residues was revealed in glycopeptides **2** and **3** (**-TT--T*-**, **-T*--TT-**). (3) Significant, but different, STD intensity patterns for the GalNAc protons of glycopeptides **2** and **3**, suggesting the existence of different binding modes, were probably related to their different rates of glycosylation[14] (Figure S2). These STD-NMR results were further supported by the crystal structure of the complex of GalNAc-T4-UDP-$Mn^{2+}$-glycopeptide **3**, which shows the GalNAc moiety of glycopeptide **3** bound to the GalNAc binding site at the lectin domain.[14]

Here, we continued these STD-NMR studies with the new glycopeptides **4**, **5**, and **6** (**-T*T-**, **-T*T–T*-**, and **–T*--T*T-**). In contrast to monoglycopeptides **2** and **3**, glycopeptides **4**, **5**, and **6** show significant STD-NMR enhancements for the protons of the Thr and Pro residues comprising the **-T*TPGP-** peptide sequence (Figure 1c, Figures S2–S6 and Tables S1–S3). The GalNAc moieties in **4**, **5**, and **6** also gave clear STD-NMR enhancements for the methyl protons of the GalNAc *N*-acetyl group (NHAc) (>75%, Figures S2–S6 and Tables S1–S3). However, differences in the relative STD-NMR enhancements of the GalNAc ring protons are clearly observed between glycopeptides **3**, **4**, **5**, and **6**. When comparing monoglycopeptides **3** and **4**, (**-T*--TT-** and **-T*T-**) the highest STD-NMR responses are observed for the H2, H4, and NHAc methyl protons in **3**, while H3 and NHAc methyl protons give the largest STD-NMR enhancements in **4** (Figure 1c and Figure S4). These differences in the STD-NMR intensities between glycopeptides **3** and **4** suggest different recognition modes of the GalNAc moieties by the transferase, fully consistent with their proposed binding at the lectin and catalytic domains, respectively. The observed STD-NMR enhancements for the protons of the Thr, Pro, and GalNAc moieties in the **-T*TPGP-** acceptor sequence of monoglycopeptide **4** are consistent with the GalNAc residues interacting



DOI: 10.1021/acscentsci.8b00488
ACS Cent. Sci. 2018, 4, 1274–1290

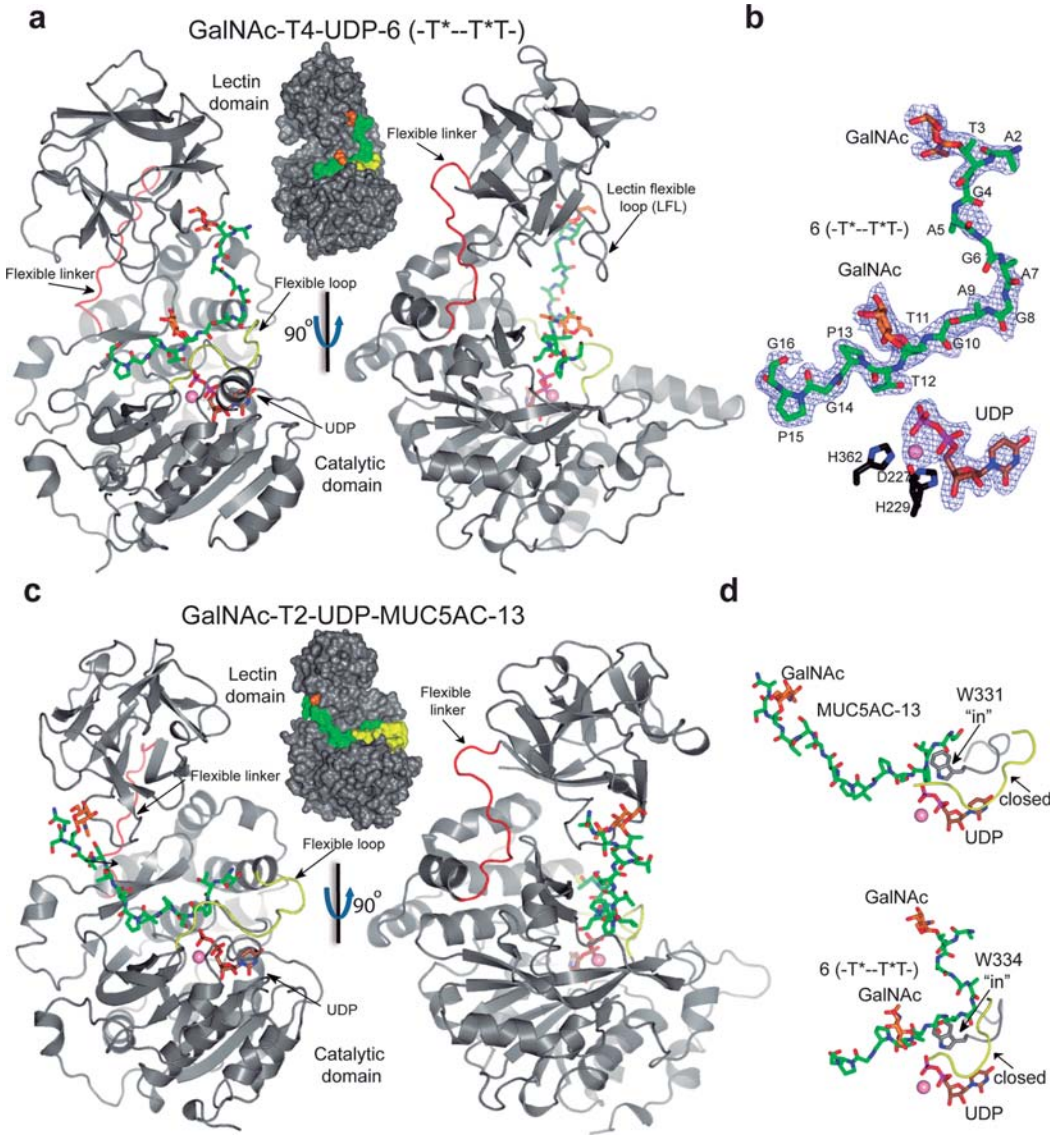



**Figure 2.** Crystal structure of GalNAc-T4 in complex with UDP-$Mn^{+2}$ and glycopeptide **6**. (a) Two different views of GalNAc-T4 in complex with **6**. The catalytic and lectin domains are colored in gray, and the flexible linker and catalytic domain active site loop are depicted in red and yellow, respectively. The lectin domain flexible loop (LFL) is indicated by a black arrow. The GalNAc moiety of the Thr3-GalNAc and Thr11-GalNAc is shown as orange carbon atoms while the rest of the peptide is shown as green carbon atoms. The nucleotide is depicted as brown carbon atoms whereas the manganese atom is shown as a pink sphere. Inserted between the structures is a surface representation of GalNAc-T4 with the same orientation as the cartoon representation of the leftmost structure. (b) Electron density maps are $F_O-F_C$ (blue) contoured at 2.0 $\sigma$ for glycopeptide **6**, UDP, and manganese ion. (c) Two different views of GalNAc-T2 in complex with **MUC5AC-13** (PDB entry 5AJP[14]) again with an inserted surface representation between structures. Atom colors are the same as in part a above. (d) Close-up views of the GalNAc-T2-UDP-**MUC5AC-13** and the GalNAc-T4-UDP-glycopeptide **6** complexes showing the bound glycopeptide and the catalytic domain active site flexible loop (in black). Note that flexible loop residues Trp331 in GalNAc-T2 and Trp334 in GalNAc-T4 adopt an "in" loop conformation.

at the catalytic domain rather than at the lectin domain (see below).

For diglycopeptides **5** and **6** (**-T*T--T*-** and **-T*--T*T-**), given the observed extensive overlapping, most of the observed STD-NMR enhancements in the GalNAc residue represent the combined STD from both GalNAc residues. Nevertheless, the individual NHAc methyl resonances from the different GalNAc residues in **5** and **6** were well-resolved, showing high STD-NMR enhancements (Figures S5 and S6), which suggests that both NHAc groups are bound. In contrast, the STD-NMR response for the resolved H2 protons of the two GalNAc residues in **6** are rather different (Figure 1c and Figures S5 and S6). The H2 proton of the GalNAc at Thr3 (i.e., the remote lectin bound GalNAc) shows a much stronger STD (90%) than that (50%) at Thr11 (i.e., the catalytic domain bound neighboring GalNAc-). This low STD response for the H2 proton of the GalNAc at Thr11 in **6** (**-T*--T*T-**) is similar to that observed for the monoglycopeptide **4** (**-T*T-**) and strongly suggests that the neighboring GalNAc moiety in **6** also preferentially binds to the catalytic domain of GalNAc-T4. This is again supported by the observation of STDs for the



DOI: 10.1021/acscentsci.8b00488
ACS Cent. Sci. 2018, 4, 1274−1290

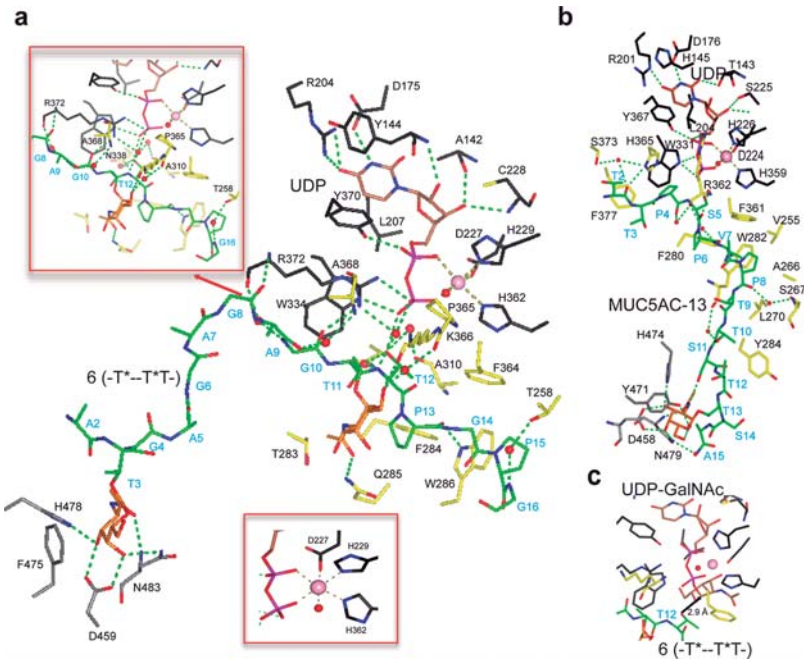



**Figure 3.** Structural features of peptide, UDP, and lectin-domain-binding sites of GalNAc-T4. (a) View of the complete sugar nucleotide, peptide, and lectin-domain-binding sites of the GalNAc-T4-UDP-glycopeptide **6** complex. Upper panel: close-up view of bound glycopeptide. Lower panel: close-up view of the manganese binding site. The residues forming sugar-nucleotide, peptide, and lectin-domain-binding sites are depicted as black, yellow, and gray carbon atoms, respectively. UDP and the glycopeptide are shown as brown and green carbon atoms, respectively. $Mn^{2+}$ and GalNAc moiety are depicted as a pink sphere and orange carbon atoms, respectively. Hydrogen bond interactions are shown as dotted green lines. Water molecules are depicted as red spheres. Note that we only show water-mediated interactions in which only one water molecule acts as a bridge between the residues. (b) View of the sugar nucleotide, glycopeptide, and lectin-domain-binding sites of the GalNAc-T2-UDP-**MUC5AC-13** complex (PDB entry 4D0T). Colors are the same as above. (c) Modeled structure for a GalNAc-T4-UDP-GalNAc-glycopeptide **6** complex. The coordinates of the UDP-GalNAc were obtained by superposing the structure of GalNAc-T2 containing UDP-GalNAc (PDB entry 4D0T) with the GalNAc-T4-UDP-GalNAc-glycopeptide **6** complex. The structure shows that the Thr12 acceptor of glycopeptide **6** is close to the anomeric carbon of UDP-GalNAc.

protons of the peptide residues of the -**T*TPGP**- acceptor sequence in both **4** and **6** (Figure 1c and Figures S4 and S6).

On the other hand, there are no STD enhancements for the peptide backbone of glycopeptides **2** and **3**, which lack the neighboring GalNAc-Thr residue in the acceptor sequence, -**TTPGP**-,[14] while STD enhancements are clearly present in glycopeptides **4**, **5**, and **6**, whose acceptor sequences do contain it. This, taken together with the different STD-NMR patterns for the GalNAc protons and the higher activities of the -**T*T**- containing peptides compared to the -**TT**- analogues (particularly **1** vs **4** and **2** vs **5**), strongly suggests that the neighboring GalNAc in -**T*T**- and the flanking -PGP- motif together increase peptide substrate binding (in the absence of a remote N-terminal prior glycosite). Furthermore, this binding event must specifically occur at the catalytic domain, since the GalNAc moiety is transferred to the free Thr in the -**T*TPGP**- acceptor sequence upon catalysis.

Glycopeptides **3** and **6** (-**T*--TT**- and -**T*--T*T**-), which show the highest $V_{max}$ and catalytic efficiency of all of the peptides (Table 2), nevertheless display dramatically different STD-NMR enhancements at their -**TTPGP**- and -**T*TPGP**- acceptor sequences. In fact, **6** provides high STDs, while **3** does not give any enhancements. This is consistent with a binding event between the remote N-terminal prior glycosite and the lectin domain in both glycopeptides, while only in **6** is there tight binding at the catalytic domain due to the presence of a neighboring glycosite in the acceptor **T*TPGP** sequence. This is in keeping with our earlier structure of **3** (-**T*--TT**-) bound to GalNAc-T4, which revealed no electron density for the peptide bound to the catalytic domain and the lack of STD enhancements for the acceptor -**TTPGP**- sequence. Therefore, the combination of kinetics and STD-NMR results provides strong evidence that a separate peptide GalNAc binding exists in the catalytic domain of GalNAc-T4 that accounts for its N-terminal short-range neighboring glycosylation preference.

**Crystal Structure of GalNAc-T4 in Complex with UDP-$Mn^{2+}$ and Glycopeptide 6 (-T*--T*T-).** To identify the catalytic domain (neighboring) GalNAc-peptide binding site on GalNAc-T4 and to further understand how the transferase recognizes both donor and glycopeptide acceptors, we obtained triclinic crystals of GalNAc-T4 that were subsequently soaked with diglycopeptide **6**, UDP, and $MnCl_2$. The resulting crystal was solved giving a structure of the transferase/diglycopeptide complex at 1.80 Å resolution, containing two independent GalNAc-T4 molecules in the asymmetric unit (Table S4). The obtained structure shows a compact GT-A fold and a lectin domain located at the N- and C-terminal regions of the transferase, respectively. Moreover, a clear structure of diglycopeptide **6** is evident bound to both domains (Figure 2a). Additionally, the structure clearly shows the flexible linker connecting both domains and the flexible loop of the catalytic domain (Figure 2a), which account for the distinct long-range glycosylation preferences and for the inactivation/activation process of these enzymes, respectively.[14,18] In our earlier crystal structure of GalNAc-T4 complexed with **3** (-**T*--TT**-), the flexible loop at the catalytic



DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

domain was disordered, and no density for glycopeptide **3** could be observed at the catalytic domain.[14] In contrast, the present structure of GalNAc-T4-UDP/$Mn^{2+}$ complexed with **6** (**-T*--T*T-**) shows both structural features fully ordered (Figure 2a). This is likely due to the stabilization provided by the binding of both UDP and the GalNAc-containing peptide to the catalytic domain (Figure 2a,b). Herein, a loop protruding from the lectin domain toward the catalytic domain glycopeptide binding site is also evident, which we have called the "lectin flexible loop" (Figure 2a residues 460−472; hereafter named LFL). This LFL motif has not been previously described for other GalNAc-T structures, and its significance will be discussed below.

**Comparison between the GalNAc-T4 and T2 Structures Bound to Glycopeptides.** The GalNAc-T4 structure complexed with **6** (**-T*--T*T-**) and UDP/$Mn^{2+}$ was compared (Figure 2c) to that of GalNAc-T2 bound to with **MUC5AC-13** (GTTPSPVPTTSTT*SAP) and UDP/$Mn^{2+}$ (PDB entry 5AJP). The observed different orientations of the lectin domain account for the distinct long-range/remote prior glycosylation preferences of these transferases, i.e., for N-/C-terminal glycosites, respectively (compare left panels and space filled inserts of Figure 2a,c).[14] In addition, in both crystal structures, the flexible loop of the catalytic domain (residues 363−374 in GalNAc-T4 and residues 360−372 in GalNAc-T2; see Figure S7) adopts the so-called "closed conformation", indicating an active state structure. The active form of the GalNAc-T4-**6** complex is further supported by the presence of the so-called "in conformation" of the catalytic Trp334 (Trp331 in GalNAc-T2, Figure 2d).[13] On this basis, the crystal structure of GalNAc-T4 bound to UDP and **6** describes the active conformation.

*Lectin Domain Substrate Binding.* The remote GalNAc moiety (Thr3-GalNAc of **6**) bound at the lectin domain is tethered by hydrogen bonds by the conserved residues Asp459/His478/Asn483 (Figure 3a) (for GalNAc-T2, Thr13-GalNAc of **MUC5AC-13** interacts with Asp458/His474/Asn479, Figure 3b). In addition, CH−$\pi$ interactions are also observed between the Thr3-GalNAc moiety and Phe475 of GalNAc-T4 (Tyr471 in GalNAc-T2) (Figure 3a,b). Out of the 16 residues of **6**, ∼1/4 (i.e., Gly4-Ala7) are fully exposed to the solvent when bound to GalNAc-T4, while this is not observed for the homologous **MUC5AC-13** bound to GalNAc-T2 (Figure 3a,b). This implies there may be a gap in substrate binding between the lectin domain GalNAc binding site and the catalytic domain (glyco)peptide binding site that has not been observed for GalNAc-T2 (see the surface representations in Figure 2a,b). Although we cannot rule out the role of the particular sequence of the peptide substrate sequence, this observation suggests that the two transferases may indeed differ in their binding mode of glycopeptide substrates in the region spanning the lectin and catalytic domain (Figure 3a,b).

*Catalytic Domain Substrate Binding.* At the catalytic domain peptide-binding site of GalNAc-T4 bound to **6**, most of the interactions with the peptide are through direct and water-mediated hydrogen bonds and to a lesser extent through hydrophobic interactions. Direct hydrogen bonds are observed with the Gly8/Ala9 and Pro13 peptide substrate backbone carbonyls and the side chain NH's of Arg372 and Trp286, respectively, while a CH−$\pi$ interaction takes place between Pro15 and Trp286. The acceptor Thr12 side chain hydroxyl group is hydrogen bonded to the UDP $\beta$-phosphate, showing that this residue is well-located to accept the GalNAc moiety of the UDP-GalNAc (Figure 3a,c). In addition, the Thr12 methyl group is located in a hydrophobic environment formed by the side chains of Phe284, Phe364, Trp286, and Ala310, as earlier described for GalNAc-T2 (residues Phe280, Trp282, Phe361, and Ala307) in complex with the mEA2 peptide (STCPA) (PDB entry 4D0Z).[18] We proposed that this environment helps to correctly position the acceptor Thr hydroxyl group to accept the anomeric carbon of UDP-GalNAc, thus enhancing the turnover and further explaining why Thr residues are better acceptors than Ser residues.[18] Water-mediated hydrogen bonds are observed through (1) Ala9/Gly10 and Gly16 backbone with Ala368/Arg372 backbone and Thr258 side chain hydroxyl, respectively; (2) Thr12 backbone with Arg 372 backbone; and (3) Thr12 side chain with Asn338 side chain and Ala310 backbone (Figure 3a).

In contrast to the observations for **6** bound to GalNAc-T4, most of the interactions of the **MUC5AC-13** glycopeptide (GTTPSPVPTTSTT*SAP) bound to GalNAc-T2 take place through hydrophobic and water-mediated hydrogen bond interactions, with few direct hydrogen bond interactions (Figure 3b).[13] Overall, our comparison of the GalNAc-T2 and -T4 peptide substrate bound structures demonstrates that both employ a well-balanced number of direct and water-mediated interactions that differ in type and number. While the complex of **6** with GalNAc-T4 mainly relies on direct hydrogen bonds, the binding of **MUC5AC-13** to GalNAc-T2 is dominated by multiple hydrophobic interactions. These results may be due to both the difference between the residues forming the peptide binding pocket and the different peptide substrate sequences. However, both enzymes share nearly identical hydrophobic interactions that are used to recognize the -PXP- motif found in both peptides, Pro13 in **6** (GAT*GAGAGAGT*T*PGP*G, italicized), and Pro6 in **MUC5AC-13** (GTTPS*PVP*TTSTT*SAP; Figure 3a,b). The analysis of both crystal structures permits us to infer that the relative promiscuity of both enzymes (outside the -PXP- motif) for recognizing multiple and dissimilar protein/peptide substrates relies on a combination of direct interactions of the peptide substrate amino acid residues with the transferase and with water-mediated hydrogen bond interactions. A similar mechanism of protein substrate recognition has been described for the promiscuous *O*-fucosyltransferase 2 (PoFUT2).[25] Hence, this recognition event may be general for diverse peptide substrates binding to enzymes.

*GalNAc-T4 Catalytic Domain Glycopeptide-GalNAc Binding Site.* A putative GalNAc binding site in the catalytic domain of GalNAc-T4 proposed in our earlier random glycopeptide studies[10] has been confirmed by the kinetics and STD-NMR experiments described above. Such a binding site is not present in GalNAc-T2.[10] The structure of GalNAc-T4 bound to diglycopeptide **6** (**-T*--T*T-**) clearly reveals the GalNAc moiety on Thr11 (directly N-terminal of the acceptor Thr12) tethered to the surface of the catalytic domain (Figure 3a). This binding event is mediated by hydrogen bond interactions between GalNAc-O6 and the Lys336 backbone carbonyl and between the carbonyl group of GalNAc and the Gln285 side chain amide. In addition, water-mediated and $CH_3/CH_3$ hydrophobic interactions are shown between the GalNAc O6 and the Pro365 backbone carbonyl, and the methyl group of GalNAc and the Thr283 methyl, respectively (Figure 3a). As shown below, mutation of Thr283 and Gln285 significantly alters GalNAc-T4's neighboring glycopeptide



DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

activity. All together, these interactions show how the specific GalNAc-Thr binding event at the catalytic domain directly presents the adjacent C-terminal Thr acceptor into the correct orientation for subsequent GalNAc transfer from the bound UDP-GalNAc. Thus, in addition to the remote lectin bound GalNAc residue, the presence of the neighboring substrate GalNAc moiety together provides the specific and essential contacts with the transferase that are required for optimal activity.

*GalNAc-T4 Nucleotide-Sugar Binding Site.* Both GalNAc-T2 and T4 share a very similar nucleotide-sugar binding sites with ∼60% identical residues (Figure 3a,b). The UDP pyrimidine moiety in GalNAc-T4 is sandwiched between Tyr144 and Leu207 (His145 and Leu204 in GalNAc-T2), while the $Mn^{2+}$ ion is hexagonally coordinated by the UDP pyrophosphate group, His362, the $D_{227}XH_{229}$ motif (His359 and $D_{224}XH_{226}$ in GalNAc-T2), and a water molecule (Figure 3a). UDP is mostly tethered by conserved hydrogen bonds with Ala142, Asp175, Arg204, Cys228, Trp334, and Tyr370 that are also present in GalNAc-T2 (Thr143, Asp176, Arg201, Ser225, Trp331, and Tyr367 in GalNAc-T2). Both enzymes also share interactions between Tyr370 (Tyr367 in GalNAc-T2) with the pyrophosphate moiety. The complex of GalNAc-T4 shows a hydrogen bond between the Tyr144 backbone and the uridine ribose, which is not found in that of GalNAc-T2. However, most of the differences at this binding site between both enzymes arise from the nonconserved amino acids in their flexible loops and the differences in peptide substrates. For example, for GalNAc-T4, Arg372 and Pro365/Lys366 recognize the $\beta$−pyrophosphate/glycopeptide **6** and GalNAc on Thr11, respectively, while, for GalNAc-T2, His365 and Arg362 interact with the **MUC5AC-13** substrate (Figure 3a,b). Overall, our results confirm that the recognition of UDP-GalNAc is highly conserved between GalNAc-Ts, whereas the nonconserved flexible loop and the peptide-binding groove are responsible for their different peptide substrate specificity.

*Correlation between Crystal Structure and STD-NMR.* The STD-derived epitope mapping reported in Figure 1c and Figures S2−S6 is consistent with the X-ray crystal structure of **6** bound to GalNAc-T4. For the bound **-T*TPGP-** fragment at the catalytic domain, the observed STDs for methyl group protons of GalNAc and both Thr residues, together with the $CH_2$ protons of both Pro residues of compounds **4**, **5**, and **6**, are readily explained by the X-ray structure since all these protons are in close contact to the transferase surface (Figure S9). The sugar moiety appears relatively far from the catalytic domain. However, the GalNAc moiety bound to the lectin domain fits into a cleft that displays multiple contacts, including the NHAc methyl group protons (Figure S9). These interactions account for the STD enhancements observed for the GalNAc H2, H3, and H4 and NHAc methyl protons in **3**. All together, these results are fully consistent with a model in which the bound GalNAc at the catalytic domain shows poor contacts with the enzyme and therefore weak STDs, while that bound at the lectin domain displays important interactions with large STD enhancements.

**GalNAc Binding Site Mutants of GalNAc-T4.** To further assess the roles of the individual lectin and catalytic domains on the binding of GalNAc glycopeptide substrates, GalNAc binding mutants were expressed for kinetic analysis against the (glyco)peptide substrate library in Table 1. These were designed to (1) eliminate the remote GalNAc binding site at the lectin domain (lectin mutant); (2) attempt to eliminate the neighboring GalNAc binding on the catalytic domain (catalytic mutant); (3) understand the role of the lectin flexible loop (LFL mutant); (4) understand the combined role of the catalytic domain and LFL on GalNAc binding (catalytic/LFL mutant); and finally (5) understand the effect of eliminating both the lectin and catalytic domain GalNAc binding sites (lectin/catalytic mutant). The results of these kinetic studies in comparison to those carried out for wt GalNAc-T4 are given in Figure 4 and Table 2. To summarize, the results clearly show the expected effects of the different mutations. The remote lectin domain glycopeptide activity is eliminated in the lectin mutants (left panels Figure 4), while the neighboring glycopeptide activity is reduced but not completely eliminated in the catalytic domain and LFL mutants (right panels Figure 4). These findings show that it is possible to produce a transferase that lacks one or the other prior GalNAc binding activities or one that is nearly void of any prior glycosylation activity. We discuss the obtained findings for each mutant in more detail below.

**Lectin Domain GalNAc Binding Mutant.** To eliminate the long-range (i.e., **-T*--T*T-** and **-T*--TT-**) glycopeptide binding at the lectin domain, the critical Asp in the classical CysLeuAsp (CLD) GalNAc binding motif was mutated to His (D459H lectin mutant). Such Asp to His mutations in other GalNAc-Ts, including T4, have been shown to eliminate their lectin binding properties.[11,26−29] Accordingly, the GalNAc-T4 D459H mutant shows a dramatic loss in its long-range prior glycosylation activity against glycopeptides **3** and **6** (**-T*--TT-** and **-T*--T*T-**) (Figure 4b, left panel), which can only be observed in the expanded scale plot at the right (Figure 4b). Thus, for **3**, the $K_m$ increased ∼20-fold, and its $V_{max}$ decreased over 1000-fold, approaching the activity of the naked peptide **1**.[14] For **6**, $K_m$ also increased ∼20-fold, but its apparent $V_{max}$ decreased only ∼5-fold, providing half of the activity of glycopeptide **4** (Figure 4b). In addition, the nonproductive UDP-GalNAc hydrolysis significantly increased to ∼70% and ∼20%, for **3** and **6**, respectively, in the lectin mutant, while for the wt transferase hydrolysis is only ∼5% for both substrates (Table 2 and Figure S8). These results clearly demonstrate the role of the lectin domain in directing the long-range prior glycosylation activity of GalNAc-T4 and that this is the dominant activity of GalNAc-T4. Nevertheless, the mutant displays a residual activity for glycopeptide **3** (**-T*--TT-**) compared to peptide **1** (**-TT-**) and a slight substrate inhibition for **6** that suggests that a weak lectin domain binding activity remains in the mutant, probably due to weak binding at its $\beta$- or $\gamma$-lectin subdomains.[10]

Although the GalNAc binding site of the catalytic domain was unchanged, the glycosylation of **4**, containing the neighboring glycosylation motif **-T*T-**, was significantly higher, giving a ∼2-fold higher $V_{max}$ and catalytic efficiency compared to the wt transferase. Moreover, the UDP-GalNAc hydrolysis decreased by ∼1/2 (Figure 4b right, Figure S8, and Table 2). We attribute this ∼2-fold increase in activity to the loss of the nonproductive binding of **-T*TPGP-** at the lectin domain in this mutant compared to the wt GalNAc-T4. Interestingly, both diglycopeptides **5** and **6** (**-T*T--T*-** and **-T*--T*T-**), which contain both the long-range and the neighboring prior GalNAc residues, show about half of the activity of glycopeptide **4** (**-T*T-**) (Figure 4b, right panel, Table 2). This lower activity can be rationalized by recognizing that diglycopeptides **5** and **6** contain two substrate GalNAc residues that compete for binding at the catalytic domain



DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

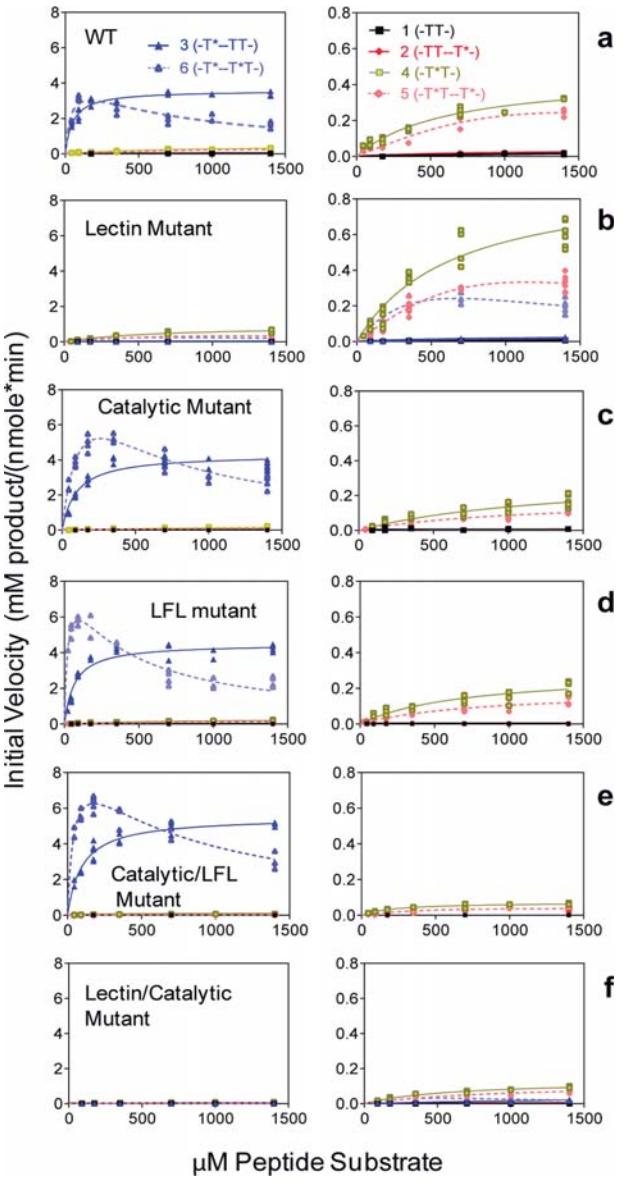



**Figure 4.** Enzyme kinetics of wt and mutant GalNAc-T4 against the (glyco)peptide substrates in Table 1. Note that the left and right panels are plotted with different initial activity scales. Kinetic constants obtained from the plots are given in Table 2. (a) Wild-type GalNAc-T4 showing both long-range (left panel) and short-range (right panel) glycopeptide activities. (b) Lectin mutant (D549H) showing the loss of its long-range glycopeptide activity. (c) Catalytic mutant (T283S, Q285A) showing the partial loss of GalANc-T4's short-range prior glycopeptide activity. (d) Lectin flexible link (LFL) mutant ($P_{463}$DNNP$_{467}$ to GGG) showing a partial loss of the short-range glycopeptide activity while possessing an intact catalytic domain. (e) Catalytic/LFL combined mutant (T283S, Q285A, D464A) showing a more complete loss of the short-range glycopeptide activity. (f) Lectin/catalytic combined mutant (T283S, Q285A, D459H) showing the near complete loss of both the long-range and short-range glycopeptide activities.

GalNAc binding site. This competition process leads to a net reduction in overall transfer, since only the bound **-T*T-** sequence can be glycosylated. Thus, the observed rates of glycosylation of the diglycopeptides should approach half of that of glycopeptide **4** (Figure 4b and Table 2). Hence, the kinetic parameters derived for monoglycopeptide **4** with this "simplified" mutant likely represent the intrinsic rates of GalNAc-T4 for glycosylating the **-T*TPGP-** sequence.

**Catalytic Domain GalNAc Binding Mutant.** Based on the structure of the bound diglycopeptide **6** (**-T*---T*T-**), the side chains of Thr283 and Gln285 of the catalytic domain are the only potentially mutable residues involved in direct interactions with the neighboring peptide GalNAc residue (i.e., **-T*T-**) of the glycopeptide substrate. This catalytic mutant (T283S/Q285A) was therefore expressed to disrupt these interactions. As expected, the catalytic activities ($V_{max}$) against glycopeptide substrates **4** and **5** (**-T*T-**, **-T*T--T*-**), were reduced to ∼1/3 of the wt GalNAc-T4 activity (Figure 4c right, Table 2). This incomplete inactivation suggests that the neighboring GalNAc binding site is only partially abolished in this mutant, presumably due to the remaining GalNAc-peptide backbone binding interactions shown in the crystal structure (i.e., Lys336 and Pro365, Figure 3), which are not eliminated by this mutation. Furthermore, the $K_m$ values for glycopeptides **4** and **5** were ∼2-fold higher than those of the wt and lectin mutant GalNAc-T4. Finally, the small increase in non-productive hydrolysis of UDP-GalNAc observed with these glycopeptides correlates with the lower activity of glycopeptides **4** and **5** compared to the wt enzyme (Table 2 and Figure S8). All together, these observations strongly suggest that the -PGP- motif remains a significant contributor to the binding of substrate in the catalytic mutant.

To further confirm that the catalytic mutant has reduced or eliminated the −1 neighboring glycopeptide activity (i.e., for the **-T*T-** motif), a random glycopeptide GPIID (GA-GAXXXXXT*XXXXXAGAG, where X= GARPNEYV and T) was glycosylated by the catalytic mutant and Edman amino sequenced to determine the sites of $^3$H-GalNAc incorporation.[10] As shown in Figure S10, the catalytic mutant clearly showed a reduction of incorporation of $^3$H-GalNAc at the Thr acceptor contiguous to the T* at the −1 site, compared to the wt GalNAc-T4, further confirming, in a nonambiguous manner, the partial loss of the neighboring glycosylation activity in this mutant.

As expected, glycosylation of **3** and **6** (**-T*--TT-** and **-T*--T*T-**) by the catalytic mutant is dominated by their remote prior glycosylation activity, due to GalNAc binding to the lectin domain. Thus, the obtained kinetic plots are similar to those for the wt GalNAc-T4 (Figure 4a,c), including the presence of substrate inhibition for diglycopeptide **6**. However, **6** displays a ∼4-fold higher $K_m$ and a ∼6-fold higher $K_d$ by SPR (Figure S11) and a ∼2-fold higher initial $V_{max}$ compared to the wt transferase, resulting in about 1/2 the catalytic efficiency, $V_{max}/K_m$ (see Table 2). These higher $K_m$ and $K_d$ values in the catalytic mutant are consistent with a significant loss of GalNAc binding at the catalytic domain and the subsequent loss of divalent binding to the enzyme, as compared to the native GalNAc-T4. Although merely speculative, the observed elevation in apparent $V_{max}$ for **6** could be explained by the loss of nonproductive binding at the catalytic domain for the remote **-T*-** substrate GalNAc residue, along with the weaker binding of the lectin domain to the **-T*TPGP-** sequence (Figure 4c and Table 2). Nevertheless, **6** still displays substrate inhibition at high concentrations, presumably due to the onset of nonproductive binding of the **-T*TPGP-** sequence at the lectin domain competing with the productive binding of the remote **-T*-**. In the catalytic mutant, monoglycopeptide **3**



DOI: 10.1021/acscentsci.8b00488
ACS Cent. Sci. 2018, 4, 1274−1290

(-**T***--**TT**-) does not show any significant increase in rate compared to the wt transferase, likely due to the fact that the catalytic domain GalNAc binding site has not been fully inactivated in this mutation. As shown below, when the catalytic domain GalNAc binding is more fully reduced, the $V_{max}$ for **3** indeed increases. Finally, it should be mentioned that the extent of nonproductive UDP-GalNAc hydrolysis for glycopeptides **3** and **6** for the catalytic mutant are very low, as observed for the wt GalNAc-T4, again consistent with their high activities and lectin domain binding (Table 2 and Figure S8).

**Lectin Domain Flexible Loop Mutant.** As mentioned above, the structure of GalNAc-T4 bound to glycopeptide **6** displays a flexible loop (LFL) protruding from the lectin domain toward the catalytic domain bound GalNAc in the -**T***TPGP- sequence. The superposition of GalNAc-T4 structures in the apo form with those in complex with **3** and **6** revealed conformational changes in the LFL (see below and Figure S12). Thus, a truncated mutant was generated, where the $_{463}$PDNNP$_{467}$ sequence located at the tip of the loop was replaced by -GGG-. Surprisingly, this mutant gave nearly identical reductions in catalytic efficiency against glycopeptides **4** and **5** as those obtained for the catalytic mutant (i.e., ∼50% reductions compared to wt GalNAc-T4, Figure 4d and Table 2). These decreases were driven by $V_{max}$ and not by changes in the $K_m$ values (Table 2). These results confirm that this loop plays a significant role in the binding of the -**T***TPGP- sequence at the catalytic domain. It is also worth noting that the neighboring glycosylation activity of the LFL mutant against the random glycopeptide acceptor GPIID shows the identical −1 glycosylation preference as the wt GalNAc-T4 (see Figure S10). Therefore, the loss of the lectin flexible loop does not alter its intrinsic glycopeptide specificity, which is consistent with the LFL mutant having an intact catalytic domain.

Further evidence that the LFL stabilizes both glycopeptide and peptide substrate binding was deduced by the observation that the nonproductive hydrolysis of UDP-GalNAc for glycopeptides **4** and **5** increased to ∼50%. For peptides **1** and **2** (-**TT**- and -**TT**--**T***), hydrolysis nearly doubled to ∼90%, compared to the wt transferase (Table 2 and Figure S8). These significant decreases in activity and increases in UDP-GalNAc hydrolysis with an intact catalytic domain strongly suggest that the LFL must play a significant role in transient substrate binding or recognition of both glycosylated (-**T***TPGP-) and nonglycosylated (-**TTPGP**-) substrates. This is the first example demonstrating that the lectin domain of a GalNAc-T can directly modulate transferase catalytic activity through its interactions with the substrate bound to the catalytic domain.

As expected, the kinetic plots of the LFL mutant against glycopeptides **3** and **6** (-**T***--**TT**-, -**T***--**T***T-) were very similar to those observed for the catalytic mutant and the wt GalNAc-T4 enzymes, as they all contain an intact lectin domain GalNAc binding site. For **3**, the $K_m$ and $V_{max}$ values are nearly the same with all three enzymes, except for a ∼2-fold higher $K_m$ in the catalytic mutant (Table 2). For **6**, a nearly 2-fold higher apparent $V_{max}$ is observed in the LFL mutant as compared to the wt GalNAc-T4. This number is similar to that observed for the catalytic mutant. Interestingly, the $K_m$ value for glycopeptide **6** in the LFL mutant appears slightly lower, although within experimental error of the wt transferase, while for the catalytic mutant the $K_m$ value is ∼3-fold higher (Table 2). Thus, the apparent catalytic efficiency, $V_{max}/K_m$, of glycopeptide **6** is increased compared to the wt and catalytic mutant GalNAc-T4. Since the LFL loop deletion/mutation is located only 4 residues C-terminal away of the critical Asp459 of the lectin domain GalNAc binding CLD motif, it is therefore possible that the LFL mutant's GalNAc binding properties are altered such that the binding of **6** is enhanced.

To further examine the effects of the LFL mutation on the structure of the LFL, molecular dynamics (MD) simulations were performed. The wild-type GalNAc-T4 and its LFL mutant, in complex with UDP-$Mn^{+2}$ and peptide **6**, were employed as starting geometries for 0.5 $\mu$s MD simulations on each (see MD simulations protocol for details). Both complexes were stable through the complete MD trajectory. Markedly, in the LFL mutant the modified loop (-GGG-) provided an open-like structure and was rather flexible. In contrast, the wild-type enzyme showed a well-defined loop (-PDNNP-), with a closed-like structure (Figure S13). This closed structure was stabilized by the formation of transient hydrogen bonds between O3 of the Thr11 GalNAc residue bound to the catalytic domain and the Asp464 side chain of the lectin loop (Figure S14). These observations provide further support to the hypothesis that the lectin flexible loop, and particularly Asp464, stabilizes the binding of neighboring prior glycosylated peptide substrates containing the -**T***TPGP- motif. The MD simulations were further complemented with a combination of steered molecular dynamics and umbrella sampling simulations (see computational binding simulations) that consisted of pulling out the glycopeptide **6** from the catalytic domain of wt GalNAc-T4 (Figure S15). During this process, the hydrogen bond between Asp464 and the Thr11-GalNAc residue of **6** was lost and replaced with a transient hydrogen bonding with the acceptor hydroxyl of Thr12 (Figure S15). These alternative simulations offer further support for a key role for the LFL and particularly Asp464 in substrate binding.

**Combined Catalytic Domain and LFL Mutant.** To access the combined roles of the catalytic domain GalNAc binding residues, T283 and Q285, and the LFL key D464, the T283S, Q285A, D464A triple mutant was expressed and purified (named as the catalytic/LFL mutant). It was expected that this mutant would show further reduced activity against the -**T***TPGP- motif while retaining its long-range N-terminal -**T***- preferences. Indeed, the apparent activity and $V_{max}$ values of **4** and **5** were reduced to ∼1/4 of the values for the catalytic and LFL mutants (Figure 4e, right panel, Table 2). Interestingly, their $K_m$ values were also reduced compared to the wt and catalytic and LFL mutants; however, this may be an artifact of the low activity of these substrates resulting in an inadequate dynamic range for accurate data fitting. As expected, **3** and **6** displayed similar kinetic plots as the individual mutants with intact lectin domain GalNAc binding sites. However, both glycopeptides gave elevated apparent $V_{max}$ values compared to the wt transferase. These findings are consistent with the previous observation of higher rates of glycosylation for the catalytic domain mutant. Note however that the $K_m$ value for the diglycopeptide **6** in the catalytic/LFL mutant was identical to that for the wt transferase and ∼4-fold lower than that for the catalytic mutant. In contrast, the $K_m$ for monoglycopeptide **3** was ∼2-fold higher than that for the wt, but the same as that observed for the catalytic mutant (Table 2). Although a full explanation for all these observations remains elusive, the catalytic/LFL mutant may nevertheless be



DOI: 10.1021/acscentsci.8b00488
ACS Cent. Sci. 2018, 4, 1274−1290

considered a kinetically simplified version of GalNAc-T4, with an intact N-terminal long-range prior glycosylation preference, while nearly lacking its neighboring preference for GalNAc. Finally, it is worth noting that the patterns of UDP-GalNAc hydrolysis for both the LFL and catalytic/LFL mutants differ from the catalytic mutant, as shown in Figure S8. For both LFL-containing mutants, the degree of hydrolysis in the presence of **1** (-**TT**-), **2** (-**TT**--**T***-), **4** (-**T***T-), and **5** (-**T***T--**T***-) is much higher than that with the catalytic mutant. The fact that hydrolysis is doubled in the presence of both the -**TT**- and -**T***T- containing substrates that lack the N-terminal prior -T* again suggests that the lectin flexible loop plays an important role in productive -**TTPGP**- and -**T***TPGP- substrate binding. Based on our previous experience with the GalNAc-Ts, hydrolysis tends to correlate with poorer substrate activity, which can be considered to represent incomplete or poor binding of substrate. In fact, in the absence of any (glyco)peptide substrate, hydrolysis is typically very low.[10]

**Combined Lectin Domain and Catalytic Domain Mutant.** Next, an attempt to remove both the long- and short-range prior glycosylation preferences of GalNAc-T4 was carried out, by combining the lectin and catalytic domain mutants. Thus, a new triple mutant (T283S, Q285A, D459H) was generated, which will be called the lectin/catalytic mutant. As shown in Figure 4f and Table 2, the target was largely achieved, although a residual activity for those glycopeptides containing the -**T***TPGP- motif still remained. As discussed above, this residual activity may be due to the remaining GalNAc binding residues in the catalytic domain and/or to the presence of Asp464 at the LFL lectin domain. Moreover, UDP-GalNAc hydrolysis for all the substrates, except for glycopeptides **4** and **5**, were significantly increased compared to wt GalNAc-T4 (Figure S8), although hydrolysis in the presence of (glyco)peptides **1**, **2**, **4**, and **5** with this lectin/catalytic mutant were still lower that those observed for both LFL mutants. Together, these results further support the role of the LFL assisting in binding substrates at the catalytic domain and the dominance of the lectin domain binding remote N-terminal glycosylated substrates.

**Peptide Substrate Preferences of GalNAc-T4 and Its Catalytic Domain Mutant.** We have previously developed a series of random peptide substrates, as GAGAXXXXXTXXXX-XAGAG, (where X = randomized amino acids), to characterize GalNAc-T peptide substrate preferences (see random peptide sequence motif determination in the Methods section).[30,31] By using high enzyme and substrate concentrations along with long incubation times, these substrates were glycosylated by GalNAc-T4 and its catalytic mutant. Their substrate preferences were evaluated in terms of the obtained enhancement values (EVs) (Figures S16 and S17). As expected, both wt and catalytic mutant transferases revealed preferences for the TPXP motif, due to the presence of three highly conserved Phe and Trp residues in the catalytic domain of most GalNAc-T family members.[10,13] GalNAc-T4 also revealed high preferences at the −1 position (relative to the site of glycosylation) for Val, Ile, and Met. On this basis, GalNAc-T4 shows peptide substrate preferences fairly close to that of GalNAc-T3.[30] Regarding the question whether mutations in the catalytic domain GalNAc binding site would alter its "naked" peptide substrate specificity, particularly at the −1 position,[10] no significant differences in the preferences were observed between the wt and mutant transferases (Figures S16 and S17). The observed preferences for Thr-*O*-GalNAc, Val, and Ile at the −1 position could be related to the presence of their $\beta$-branched methyl groups. Indeed, intense STDs are observed for the Thr methyl protons in **4**, **5**, and **6** in the presence of GalNAc-T4. The presence of methyl-containing residues at the −1 site are likely related to the reported conformational features provided by these residues.[32]

Thus, additional MD simulations were performed on GalNAc-T4 complexed with UDP-GalNAc and a naked peptide (GAGAGAG**X**TPGPG, where T is the acceptor Thr) in which the residue X8 (−1 with respect to the acceptor Thr) was replaced by either Val or Ala (Figure S18 and Movies S1 and S2). According to the MD calculations, the peptide containing the Val8 is stabilized by hydrophobic interaction between the methyl groups of Val8 and Ala368. In turn, this hydrophobic patch promotes the proximity between the hydroxyl group of Thr9 and the anomeric carbon of GalNAc. Conversely, the interactions between the methyl groups of Ala8 and Ala368 in the Ala8 substrate were negligible, presumably due to the longer distance between their hydrophobic side chains, prompting Thr9 to move away from the active site. These results satisfactorily explain why GalNAc-T4 prefers $\beta$-branched amino acids at position −1 for optimal glycosylation.

## CONCLUSIONS

The GalNAc-Ts comprise a large family of evolutionary conserved glycosyltransferase isoforms that differentially exhibit substrate specificities for peptides and partially glycosylated GalNAc-glycopeptides. All isoforms use their unique C-terminal lectin domains to bind GalNAc-glycopeptides, and here for the first time we demonstrate that a subset of the GalNAc-T isoforms exemplified by GalNAc-T4 also contain a GalNAc binding site in the catalytic domain. We provide the first conclusive evidence for the direct interaction of the catalytic domain with a GalNAc residue immediately adjacent to the acceptor site explaining the observed GalNAc-glycopeptide substrate specificity of GalNAc-T4 and related isoforms. Unambiguous evidence for the two distinct GalNAc binding capabilities was obtained by the structure of GalNAc-T4 bound to a diglycopeptide (i.e., glycopeptide **6**, -**T***--**T***T-). This structure revealed how the catalytic domain of GalNAc-T4 recognizes a glycopeptide substrate, and represents the first structure of a GalNAc-T with a glycopeptide GalNAc residue bound to its catalytic domain.

Interestingly, the key residues at the catalytic domain responsible for the binding of the neighboring GalNAc residue have been identified, showing that the binding process is dominated by rather weak interactions, as further supported by kinetic studies. When both long- and short-range prior glycosylation sites are combined in one substrate, i.e., diglycopeptide **6**, apparent substrate inhibition is observed. Comparing the kinetics of the mutant and wt transferases suggests that the observed inhibition may be due to the weak nonproductive binding of the substrate -**T***TGPG- sequence at the lectin domain. This competition with the binding of the remote GalNAc at high substrate concentrations then leads to a decrease in activity. This observation furthermore suggests the lectin domain of GalNAc-T4 may also possess unique peptide sequence preferences as found for GalNAc-T2.[28]

A unique aspect of this enzyme is the presence of a flexible loop (LFL) protruding from the lectin domain, which assists both GalNAc binding and release. We believe that with this finding we have identified a new structural feature, out of the



DOI: 10.1021/acscentsci.8b00488
ACS Cent. Sci. 2018, 4, 1274−1290

few already described for GalNAc-Ts, that serves to modulate GalNAc-T4 activity and specificity. Thus, alterations outside the catalytic domain and the lectin domain GalNAc binding sites, that would not be predicted to be deleterious, can have profound effects on the catalytic activity of GalNAc-T4.

In addition, the generation of several GalNAc-T4 mutants has allowed us to individually characterize the kinetics and binding features of the different parts of the enzyme. Finally, the "naked" peptide substrate motif for GalNAc-T4 and its catalytic domain mutants reveals that GalNAc-T4 displays the expected TPXP preference and prefers the $\beta$-branched residues Val, Ile, and Met residues preceding its acceptor site.

In summary we have identified the molecular basis for GalNAc-T4's long- and short-range prior GalNAc glycosylation preferences, demonstrating that the long-range specificity greatly dominates the activity/function of this isoform. The combination of long- and short-range GalNAc-glycopeptide substrate specificity makes GalNAc-T4 ideal for performing the proposed role as a follow-up isoenzyme that fills in unoccupied acceptor sites in densely O-glycosylated regions such as in the tandem repeats of mucins.[11,21] This is consistent with GalNAc-T4's high expression levels in mucin secreting tissues such as the colon, lung, and sublingual gland.[1,33] Our newly acquired ability to selectively eliminate the long- and/or short-range glycopeptide activities of GalNAc-T4, based on the crystal structure of GalNAc-T4-UDP-glycopeptide **6** complex, will be an invaluable tool for understanding how GalNAc-T4 performs this important filling-in role.

## METHODS

**Cloning, Site-Directed Mutagenesis, Expression, and Purification.** The expression plasmid pPICZ$\alpha$A*galnact4* (36−578) was used as the template for introducing the following single and multiple amino-acid changes by site-directed mutagenesis as described:[14] T283S-Q285A (catalytic mutant), D459H (lectin mutant), T283S-Q285A-D459H (lectin/catalytic mutant), and T283S-Q285A-D464A (catalytic/LFL mutant). For generating the lectin flexible loop deletion mutant (LFL mutant), the residues $P_{463}$DNNP$_{467}$ were removed and replaced by three Gly residues. Site-directed mutagenesis was performed by GenScript. Wild-type and mutant transferases were purified using the protocol developed for the wild-type enzyme.[14]

**Synthesis of Glycopeptides.** Glycopeptides were synthesized as described previously and confirmed by Edman amino acid sequencing on a Shimadzu PPSQ-53A peptide sequencer.[14] See Table S5 for the HPLC and MS characterization of each glycopeptide.

**Transferase Assays and Kinetics.** GalNAc-T glycosylation reactions against (glyco)peptides **1**−**6** were performed as described.[14] Briefly, reactions consisted of 75 mM sodium cacodylate, pH 6.5, 1 mM 2-mercaptoethanol, 10 mM $MnCl_2$, 0.25 mM [$^3$H]-radiolabeled UDP-GalNAc (~6 × $10^8$ DPM/$\mu$mole, American Radiolabeled Chemicals Inc.), and 0.004−1.4 mM of glycopeptide substrate and variable concentration of transferase (0.02−0.2 $\mu$M) in a final reaction volume of 20 $\mu$L. After incubating at 37 °C reactions were quenched by the addition of 20 $\mu$L of 250 mM EDTA. UDP and nonhydrolyzed UDP-GalNAc were removed by passage over a small Dowex 1 × 8 anion exchange resin. Total UDP-[$^3$H]-GalNAc utilization (transfer to peptide substrate plus transfer to water, i.e., UDP-[$^3$H]-GalNAc hydrolysis) was determined by scintillation counting (Beckman LS5801 scintillation counter) aliquots before and after passage over Dowex. The actual [$^3$H]-GalNAc transfer to peptide and the extent of hydrolysis were determined by gel filtration analysis on Sephadex G10, as described.[19] Example gel filtration chromatograms are given in Figure S8. In all cases the reported transferase activity has been corrected (reduced) for the presence of the nonproductive UDP-[$^3$H]-GalNAc hydrolysis which varied with the glycopeptide substrate and with the transferase mutant. For detailed kinetics studies incubation times (typically 10−180 min) were chosen such that no more than 30% of the UDP-GalNAc donor was depleted while typically giving less than 10% (glyco)peptide glycosylation. Initial velocities/activities were determined using 1 or 3 reaction time points for each substrate concentration and were repeated 2−5 times. Typically 24−30 individual specific activity values were obtained over the entire substrate concentration range. These individual data points were used to calculate the kinetic constants of $K_m$, $V_{max}$, and $K_i$ using the nonlinear Michaelis−Menten and the Michaelis−Menten with substrate inhibition fitting programs found in GraphPad Prism 7.03.

**Random Peptide Sequence Motif Determination.** Peptide sequence motifs for GalNAc-T4 and its mutants were obtained as described for other GalNAc-Ts.[31,34,35] Briefly overnight transferase reactions (0.3 $\mu$M enzyme) were performed with each of the 3 random peptides PVI, PVII, and PVIII (GAGAXXXXXTXXXXXAGAGK, where X = G,A,R,P,E,H,Q,Y,V,L (PVI), G,A,P,R,D,F,I,M,K,N (PVII), and G,A,P,R,E,Y,V,K,N,S (PVIII))[30] at ~6.6 mM in the presence of 200 mM sodium cacodylate, pH 6.9, 1 mM 2-mercaptoethanol, 10 mM $MnCl_2$, 0.1% TritonX100, and 2 mM [$^3$H]-radiolabeled UDP-GalNAc (~6 × $10^8$ DPM/$\mu$mole, American Radiolabeled Chemicals Inc.) in a final volume of 200 $\mu$L. Reactions were quenched with 100 $\mu$L of 250 mM EDTA, passed over Dowex 1 × 8 anion exchange resin and peptide and glycopeptide products isolated by Sephadex G10 gel filtration. Glycopeptide product was isolated by lectin chromatography on a mixed bed lectin column containing immobilized lectins (SJA (*Sophora japonica*), SBA (*Glycine max*), HPA (*Helix pomatia*), and VVA (*Vicia villosa*)) as described.[31] After final purification on Sephadex G10 chromatography the glycopeptide product was Edman amino acid sequenced on a Shimadzu PPSQ-53A peptide sequencer to determine the compositions of the X positions of the random peptides. Enhancement values (EVs) were obtained from the ratio of the mole fractions of each amino acid residue in the product glycopeptide to that in the starting random peptide.[31] Thus, EVs greater than one indicate an enrichment in the glycopeptide while EVs lower than 1 indicate a depletion in the glycopeptide. On this basis the EVs reflect the transferase's preference for a particular amino acid residue at each X position.[30,31] EVs were obtained from triplicate determinations on each random peptide; thus for each amino acid residue there were between 3 and 9 individual EV determinations at each X position depending on their presence in the three different random peptides. The obtained averaged EVs are plotted and compared in Figures S16 and S17 for wt and catalytic mutant GalNAc-T4.

**Determination of Site of Glycosylation.** Substrate glycosylation sites were determined by Edman amino acid sequencing on a Shimadzu PPS-Q53A protein sequencer. Briefly, G10 isolated [$^3$H]-GalNAc glycosylated substrates were spotted on a Polybrene precycled glass fiber disk (GFD) and sequenced using a modified GFD program. The



DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

glycosylated PTH-Thr derivatives (eluting between 2.85 and 3.5 min using the standard PPSQ_HPLC buffers and flow rate) were collected directly into scintillation vials on a Shimadzu FRC-10A fraction collector and scintillation counted for [$^3$H]-GalNAc content (Beckman LS5801 scintillation counter). Note that [$^3$H]-GalNAc lag is commonly observed after a peak of [$^3$H]-GalNAc incorporation in these determinations. This is due to the poor extraction of the glycosylated-PTH residues from the glass fiber disks compared with the nonglycosylated amino acid PTH derivatives.[10,19]

**Surface Plasmon Resonance Experiments.** The SPR experiments for peptides **4** and **6** were performed as described for glycopeptide **3**.[14] As found for peptide **3**, binding saturation was not achieved for glycopeptide **4**; thus its $K_d$ could not be determined. However, saturation was reached for glycopeptide **6** allowing an accurate $K_d$ determination.

**NMR Experiments.** All NMR experiments were recorded on a Bruker Avance 600 MHz spectrometer equipped with a triple channel cryoprobe head. The $^1$H NMR resonances of the glycopeptides **4**, **5**, and **6** were completely assigned through standard 2D-TOCSY (30 and 80 ms mixing time) and 2D-NOESY experiments (400 ms mixing time) obtained at 278 K. Glycopeptides were 1−3 mM in 25 mM perdeuterated tris-d11 (uncorrected pH meter reading 7.4) in $H_2O/D_2O$ (90:10) with 7.5 mM NaCl and 1 mM DTT. The resonance of 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid (TSP) was used as a chemical shift reference ($\delta$ TSP = 0 ppm) in the $^1$H NMR experiments. Peak lists for the 2D-TOCSY and 2D-NOESY spectra were generated by interactive peak picking using CARA software. STD-NMR experiments were performed at 298 K in deuterated water in the presence of 25 mM perdeuterated tris-d11 (uncorrected pH meter reading 7.4), 7.5 mM NaCl, and 1 mM DTT, using ∼877 $\mu$M glycopeptide, 13.5 $\mu$M GalNAc-T4, 75 $\mu$M UDP, and 75 $\mu$M $MnCl_2$. STD-NMR spectra were acquired and the data analyzed as described.[14]

**Crystallization.** Crystals of GalNAc-T4 were grown as described before.[14] The crystals were soaked for 30 min with a mix containing 20 mM glycopeptide **6** and 20 mM UDP in 25 mM Tris pH 7.5 and 2 mM $MnCl_2$. The crystals were subsequently cryoprotected with 25% ethylene glycol, 18% PEG3350, and 0.1 M ammonium nitrate, and frozen in a nitrogen gas stream cooled to 100 K.

**Structure Determination and Refinement.** Diffraction data were collected on the synchrotron beamline I03 of the Diamond Light Source (Harwell Science and Innovation Campus, Oxfordshire, UK) at a wavelength of 0.97 Å and temperature of 100 K. Data were processed and scaled using the XDS package[36] and CCP4[37,38] software. Relevant statistics are given in Table S4. The crystal structure was solved by molecular replacement with Phaser[37,38] using the PDB entry 5NQA of human GalNAc-T4 as template. Initial phases were improved by several cycles of manual model building in Coot[39] and further refined using REFMAC5.[40] The final model of GalNAc-T4 soaked with glycopeptide **6** and UDP was validated with PROCHECK where model statistics are given in Table S4. The asymmetric unit of the triclinic crystal contained 2 molecules of GalNAc-T4 while only one of the monomers contained UDP, $Mn^{+2}$, and glycopeptide **6**. The Ramachandran plot shows that 95.14%, 3.69%, and 1.17% of the amino acids are in most favored, allowed, and disallowed regions, respectively.

**MD Simulations with Peptide and Glycopeptide Substrates.** The wild-type and the LFL mutant of GalNAc-T4, both in complex with UDP-$Mn^{+2}$ and the glycopeptide **6**, were subjected to 500 ns of MD simulation as described previously.[14] Similarly, the wild-type GalNAc-T4 in complex with UDP-GalNAc and a naked peptide (GAGA-GAGXTPGPG where X denotes either Val or Ala and T as the acceptor Thr) was subjected to 200 ns of MD simulations. In all cases mutants were generated using PyMol. The starting coordinates of the UPD-GalNAc in GalNAc-T4 were taken from the X-ray structure of GalNAc-T2 previously reported by our group (PDB ID: 4D0T).

**Glycopeptide Pull-Out Computational Details.** The initial structure for the simulations was taken from the GalNAc-T4-UDP-glycopeptide **6** complex. The UDP substrate was completed adding the GalNAc sugar from the PDB entry 4D0T by superimposition. The protonation states and hydrogen atom positions of all amino acid residues were determined by visual inspection according to protein environment. The system was solvated with a box of 15 Å around the protein surface (31.595 water molecules), and the global charge was neutralized by the addition of 1 sodium ion, leading to a total of 103 442 atoms. Molecular dynamics (MD) simulations using Amber11 software were performed. The protein was modeled with the FF99SB force field, and the carbohydrate substrate and water molecules were described with the GLYCAM06 and TIP3P force fields, respectively. The MD simulation was carried out in several steps. First, the system was minimized, holding the protein and substrate fixed. Then, the entire system was allowed to relax. To gradually reach the desired temperature, weak spatial constraints were initially added to the protein and substrate, while water molecules and sodium ions were allowed to move freely at 100 K. The constraints were then removed, and the working temperature of 300 K was reached after two more 100 K heatings in the NVT ensemble. Afterward, the density was converged up to water density at 300 K in the NPT ensemble, and the simulation was extended to 50 ns in the NVT ensemble. Steered molecular dynamics[41] (SMD) and umbrella sampling[42] (US) simulations were performed to pull out the neighboring glycan from the catalytic domain. The first method was used to generate the initial pathway from which the last method explored the phase-space. One collective variable (CV) was used for the pull-out, defined as the distance between the $\alpha$-carbon of Asn224 (buried in the binding pocket of the catalytic domain) and the $\alpha$-carbon of Thr12 (the acceptor threonine of diglycopeptide **6**). A total of 20 trajectories with random velocities were taken from the reference structure from the equilibration MD, allowing them to relax for 1 ns. Subsequently, a movable harmonic potential of 50 kcal/(mol $A^2$) was used to drive the CV 40 Å apart during 2 ns, with a pulling velocity of 20 Å/ns. The trajectory with the lower energy was taken for the US simulations, and the pathway was divided in 81 windows with a regular separation of 0.5 Å between them. Force constants of 10 kcal/(mol $A^2$) were used for the harmonic potentials. Every window was sampled during 10 ns, leading to a total of ∼0.8 $\mu$s of simulation data. The firsts 2 ns of each window were considered as an equilibration step. Analysis of the trajectories was carried out using standard tools of Amber and VMD.[43] Particularly, the hydrogen bond analysis was performed using the *cpptraj* utility from Amber14, taking into account all the interactions between the substrate (diglycopeptide) and the receptor (GalNAc-T4 bound to UDP-GalNAc), with a



DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

distance cutoff of 3.0 Å between heteroatoms and 135° for the angle that defines the hydrogen bond.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.8b00488.

Additional data and figures including sensograms, SPR data fitting, epitope mappings, STD-NMR spectra, sequence alignment, chromatograms, crystal structures, and atomic fluctuation analysis (PDF)
Movie S1: MD simulation (AVI)
Movie S2: MD simulation (AVI)

### Accession Codes

The coordinate and structure factor has been deposited in the Worldwide Protein Data Bank (wwPDB) with the PDB code 6H0B (see Table S4).

## ■ AUTHOR INFORMATION


### Corresponding Authors

*E-mail: txg2@cwru.edu.
*E-mail: rhurtado@bifi.es.

### ORCID

Jesús Jiménez-Barbero: 0000-0001-5421-8513
Carme Rovira: 0000-0003-1477-5010
Francisco Corzana: 0000-0001-5597-8127
Ramon Hurtado-Guerrero: 0000-0002-3122-9401


### Author Contributions

M. de las Rivas and E. J. Paul Daniel contributed equally to this work. R. Hurtado-Guerrero designed the crystallization construct and solved the crystal structure. E. Lira-Navarrete, M. de las Rivas, and R. Hurtado-Guerrero purified the enzymes, crystallized the complex, and refined the crystal structure. I. Compañón, and F. Corzana synthesized the glycopeptides. F. Corzana, L. Raich, and C. Rovira performed the computational simulations. H. Coelho, A. Diniz, J. Jiménez-Barbero, and F. Marcelo performed the NMR experiments. T. A. Gerken and E. J. Paul Daniel performed the kinetic and peptide substrate motif studies together with the Edman amino acid sequencing. R. Hurtado-Guerrero and T. A. Gerken wrote the article together with main contributions of F. Corzana, H. Clausen, J. Jiménez-Barbero, and F. Marcelo. All authors read and approved the final manuscript.

### Notes

The authors declare no competing financial interest.
Safety statement: no unexpected safety hazards were encountered.

## ■ ACKNOWLEDGMENTS

We thank synchrotron radiation sources DLS (Oxford) and in particular beamline I03 (experiment number MX10121-15). We thank ARAID, MEC (CTQ2013-44367-C2-2-P, BFU2016-75633-P, CTQ2015-67727-R, CTQ2015-70524-R, and CTQ2017-85496-P), AGAUR (SGR2017-1189), the National Institutes of Health (R01-GM113534, and instrument Grant GM113534-01S to T. A. Gerken), the Danish National Research Foundation (DNRF107), the FCT-Portugal [UID/Multi/04378/2013 cofinanced by the FEDER (POCI-01-0145-FEDER-007728)], and the DGA (E34_R17) for financial support. I. Compañón thanks Universidad de La Rioja for the FPI grant. F. Marcelo thanks FCT-Portugal for IF Investigator grant (IF/00780/2015) and PTNMR supported by Project 022161. E. Lira-Navarrete acknowledges her postdoctoral EMBO fellowship ALTF 1553-2015 cofunded by the European Commission (LTFCOFUND2013, GA-2013-609409) and Marie Curie Actions. H. Coelho and J. Jiménez-Barbero thank EU for the TOLLerant project. The research leading to these results has also received funding from the FP7 (2007−2013) under BioStruct-X (Grant agreement 283570 and BIOSTRUCTX_5186). We would also like to acknowledge the assistance of Juwan Lee in obtaining the GalNAc-T4 random peptide motifs.

## ■ REFERENCES


(1) Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **2012**, *22* (6), 736−756.

(2) Kato, K.; Jeanneau, C.; Tarp, M. A.; Benet-Pages, A.; Lorenz-Depiereux, B.; Bennett, E. P.; Mandel, U.; Strom, T. M.; Clausen, H. Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation. *J. Biol. Chem.* **2006**, *281* (27), 18370−18377.

(3) Khetarpal, S. A.; Schjoldager, K. T.; Christoffersen, C.; Raghavan, A.; Edmondson, A. C.; Reutter, H. M.; Ahmed, B.; Ouazzani, R.; Peloso, G. M.; Vitali, C.; Zhao, W.; Somasundara, A. V.; Millar, J. S.; Park, Y.; Fernando, G.; Livanov, V.; Choi, S.; Noe, E.; Patel, P.; Ho, S. P.; Myocardial Infarction Exome Sequencing, S.; Kirchgessner, T. G.; Wandall, H. H.; Hansen, L.; Bennett, E. P.; Vakhrushev, S. Y.; Saleheen, D.; Kathiresan, S.; Brown, C. D.; Abou Jamra, R.; LeGuern, E.; Clausen, H.; Rader, D. J. Loss of Function of GALNT2 Lowers High-Density Lipoproteins in Humans, Nonhuman Primates, and Rodents. *Cell Metab.* **2016**, *24* (2), 234−245.

(4) Pedersen, N. B.; Wang, S.; Narimatsu, Y.; Yang, Z.; Halim, A.; Schjoldager, K. T.; Madsen, T. D.; Seidah, N. G.; Bennett, E. P.; Levery, S. B.; Clausen, H. Low density lipoprotein receptor class A repeats are O-glycosylated in linker regions. *J. Biol. Chem.* **2014**, *289* (25), 17312−17324.

(5) Beaman, E. M.; Brooks, S. A. The extended ppGalNAc-T family and their functional involvement in the metastatic cascade. *Histol Histopathol* **2014**, *29* (3), 293−304.

(6) Nguyen, A. T.; Chia, J.; Ros, M.; Hui, K. M.; Saltel, F.; Bard, F. Organelle Specific O-Glycosylation Drives MMP14 Activation, Tumor Growth, and Metastasis. *Cancer Cell* **2017**, *32* (5), 639−653e6.

(7) Tran, D. T.; Ten Hagen, K. G. Mucin-type O-glycosylation during development. *J. Biol. Chem.* **2013**, *288* (10), 6921−6929.

(8) Goth, C. K.; Vakhrushev, S. Y.; Joshi, H. J.; Clausen, H.; Schjoldager, K. T. Fine-Tuning Limited Proteolysis: A Major Role for Regulated Site-Specific O-Glycosylation. *Trends Biochem. Sci.* **2018**, *43* (4), 269−284.

(9) Goth, C. K.; Halim, A.; Khetarpal, S. A.; Rader, D. J.; Clausen, H.; Schjoldager, K. T. A systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific O-glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (47), 14623−14628.

(10) Revoredo, L.; Wang, S.; Bennett, E. P.; Clausen, H.; Moremen, K. W.; Jarvis, D. L.; Ten Hagen, K. G.; Tabak, L. A.; Gerken, T. A. Mucin-type O-glycosylation is controlled by short- and long-range glycopeptide substrate recognition that varies among members of the polypeptide GalNAc transferase family. *Glycobiology* **2016**, *26* (4), 360−376.

(11) Hassan, H.; Reis, C. A.; Bennett, E. P.; Mirgorodskaya, E.; Roepstorff, P.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. *J. Biol. Chem.* **2000**, *275* (49), 38197−38205.




DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

(12) Raman, J.; Fritz, T. A.; Gerken, T. A.; Jamison, O.; Live, D.; Liu, M.; Tabak, L. A. The catalytic and lectin domains of UDP-GalNAc:polypeptide alpha-N-Acetylgalactosaminyltransferase function in concert to direct glycosylation site selection. *J. Biol. Chem.* **2008**, *283* (34), 22942−22951.
(13) Lira-Navarrete, E.; de Las Rivas, M.; Companon, I.; Pallares, M. C.; Kong, Y.; Iglesias-Fernandez, J.; Bernardes, G. J.; Peregrina, J. M.; Rovira, C.; Bernado, 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.* **2015**, *6*, 6937.
(14) de las Rivas, M.; Lira-Navarrete, E.; Daniel, E. J. P.; Companon, I.; Coelho, H.; Diniz, A.; Jimenez-Barbero, J.; Peregrina, J. M.; Clausen, H.; Corzana, F.; Marcelo, F.; Jimenez-Oses, G.; Gerken, T. A.; Hurtado-Guerrero, R. The interdomain flexible linker of the polypeptide GalNAc transferases dictates their long-range glycosylation preferences. *Nat. Commun.* **2017**, *8* (1), 1959.
(15) Perrine, C. L.; Ganguli, A.; Wu, P.; Bertozzi, C. R.; Fritz, T. A.; Raman, J.; Tabak, L. A.; Gerken, T. A. Glycopeptide-preferring polypeptide GalNAc transferase 10 (ppGalNAc T10), involved in mucin-type O-glycosylation, has a unique GalNAc-O-Ser/Thr-binding site in its catalytic domain not found in ppGalNAc T1 or T2. *J. Biol. Chem.* **2009**, *284* (30), 20387−20397.
(16) Fritz, T. A.; Raman, J.; Tabak, L. A. Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2. *J. Biol. Chem.* **2006**, *281* (13), 8613−8619.
(17) Song, L.; Linstedt, A. D., Inhibitor of ppGalNAc-T3-mediated O-glycosylation blocks cancer cell invasiveness and lowers FGF23 levels. *eLife* **2017**, *6*, 24051.
(18) Lira-Navarrete, E.; Iglesias-Fernandez, J.; Zandberg, W. F.; Companon, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R. Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase 2. *Angew. Chem., Int. Ed.* **2014**, *53* (31), 8206−8210.
(19) Gerken, T. A.; Revoredo, L.; Thome, J. J.; Tabak, L. A.; Vester-Christensen, M. B.; Clausen, H.; Gahlay, G. K.; Jarvis, D. L.; Johnson, R. W.; Moniz, H. A.; Moremen, K. The lectin domain of the polypeptide GalNAc transferase family of glycosyltransferases (ppGalNAc Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type O-glycosylation. *J. Biol. Chem.* **2013**, *288* (27), 19900−19914.
(20) Posey, A. D., Jr.; Schwab, R. D.; Boesteanu, A. C.; Steentoft, C.; Mandel, U.; Engels, B.; Stone, J. D.; Madsen, T. D.; Schreiber, K.; Haines, K. M.; Cogdill, A. P.; Chen, T. J.; Song, D.; Scholler, J.; Kranz, D. M.; Feldman, M. D.; Young, R.; Keith, B.; Schreiber, H.; Clausen, H.; Johnson, L. A.; June, C. H. Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. *Immunity* **2016**, *44* (6), 1444−1454.
(21) Bennett, E. P.; Hassan, H.; Mandel, U.; Mirgorodskaya, E.; Roepstorff, P.; Burchell, J.; Taylor-Papadimitriou, J.; Hollingsworth, M. A.; Merkx, G.; van Kessel, A. G.; Eiberg, H.; Steffensen, R.; Clausen, H. Cloning of a human UDP-N-acetyl-alpha-D-Galactosamine:polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. *J. Biol. Chem.* **1998**, *273* (46), 30472−30481.
(22) Bermejo, I. A.; Usabiaga, I.; Companon, I.; Castro-Lopez, J.; Insausti, A.; Fernandez, J. A.; Avenoza, A.; Busto, J. H.; Jimenez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Jimenez-Oses, G.; Hurtado-Guerrero, R.; Cocinero, E. J.; Corzana, F. Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition. *J. Am. Chem. Soc.* **2018**, *140* (31), 9952−9960.
(23) Dam, T. K.; Gerken, T. A.; Brewer, C. F. Thermodynamics of multivalent carbohydrate-lectin cross-linking interactions: importance of entropy in the bind and jump mechanism. *Biochemistry* **2009**, *48* (18), 3822−3827.
(24) Dam, T. K.; Brewer, C. F. Lectins as pattern recognition molecules: the effects of epitope density in innate immunity. *Glycobiology* **2010**, *20* (3), 270−279.
(25) Valero-Gonzalez, J.; Leonhard-Melief, C.; Lira-Navarrete, E.; Jimenez-Oses, G.; Hernandez-Ruiz, C.; Pallares, M. C.; Yruela, I.; Vasudevan, D.; Lostao, A.; Corzana, F.; Takeuchi, H.; Haltiwanger, R. S.; Hurtado-Guerrero, R. A proactive role of water molecules in acceptor recognition by protein O-fucosyltransferase 2. *Nat. Chem. Biol.* **2016**, *12* (4), 240−246.
(26) Fritz, T. A.; Hurley, J. H.; Trinh, L. B.; Shiloach, J.; Tabak, L. A. The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-T1. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (43), 15307−15312.
(27) Kubota, T.; Shiba, T.; Sugioka, S.; Furukawa, S.; Sawaki, H.; Kato, R.; Wakatsuki, S.; Narimatsu, H. Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J. Mol. Biol.* **2006**, *359* (3), 708−727.
(28) Pedersen, J. W.; Bennett, E. P.; Schjoldager, K. T.; Meldal, M.; Holmer, A. P.; Blixt, O.; Clo, E.; Levery, S. B.; Clausen, H.; Wandall, H. H. Lectin domains of polypeptide GalNAc transferases exhibit glycopeptide binding specificity. *J. Biol. Chem.* **2011**, *286* (37), 32684−32696.
(29) Wandall, H. H.; Irazoqui, F.; Tarp, M. A.; Bennett, E. P.; Mandel, U.; Takeuchi, H.; Kato, K.; Irimura, T.; Suryanarayanan, G.; Hollingsworth, M. A.; Clausen, H. The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation. *Glycobiology* **2007**, *17* (4), 374−387.
(30) Gerken, T. A.; Jamison, O.; Perrine, C. L.; Collette, J. C.; Moinova, H.; Ravi, L.; Markowitz, S. D.; Shen, W.; Patel, H.; Tabak, L. A. Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J. Biol. Chem.* **2011**, *286* (16), 14493−14507.
(31) Gerken, T. A.; Raman, J.; Fritz, T. A.; Jamison, O. Identification of common and unique peptide substrate preferences for the UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates. *J. Biol. Chem.* **2006**, *281* (43), 32403−32416.
(32) Corzana, F.; Busto, J. H.; Jimenez-Oses, G.; Garcia de Luis, M.; Asensio, J. L.; Jimenez-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.* **2007**, *129* (30), 9458−9467.
(33) Young, W. W., Jr.; Holcomb, D. R.; Ten Hagen, K. G.; Tabak, L. A. Expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase isoforms in murine tissues determined by real-time PCR: a new view of a large family. *Glycobiology* **2003**, *13* (7), 549−557.
(34) Perrine, C.; Ju, T.; Cummings, R. D.; Gerken, T. A. Systematic determination of the peptide acceptor preferences for the human UDP-Gal:glycoprotein-alpha-GalNAc beta 3 galactosyltransferase (T-synthase). *Glycobiology* **2009**, *19* (3), 321−328.
(35) Gerken, T. A.; Ten Hagen, K. G.; Jamison, O. Conservation of peptide acceptor preferences between Drosophila and mammalian polypeptide-GalNAc transferase ortholog pairs. *Glycobiology* **2008**, *18* (11), 861−870.
(36) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.
(37) 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.; 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., Sect. D: Biol. Crystallogr.* **2011**, *67*, 235−242.


DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

(38) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 760−763..
(39) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.
(40) Murshudov, G. N.; Skubak, 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., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355−367.
(41) Isralewitz, B.; Gao, M.; Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 224−230.
(42) Park, S.; Im, W. Theory of Adaptive Optimization for Umbrella Sampling. *J. Chem. Theory Comput.* **2014**, *10* (7), 2719−2728.
(43) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.



DOI: 10.1021/acscentsci.8b00488
*ACS Cent. Sci.* 2018, 4, 1274−1290

nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-018-0083-9

# Efficient and irreversible antibody–cysteine bioconjugation using carbonylacrylic reagents

Barbara Bernardim[1,2], Maria J. Matos[1], Xhenti Ferhati[3], Ismael Compañón[3], Ana Guerreiro[4], Padma Akkapeddi[4], Antonio C. B. Burtoloso[5], Gonzalo Jiménez-Osés[3], Francisco Corzana[3], Gonçalo J. L. Bernardes[1,4*]

**There is considerable interest in the development of chemical methods for the precise, site-selective modification of antibodies for therapeutic applications. In this protocol, we describe a strategy for the irreversible and selective modification of cysteine residues on antibodies, using functionalized carbonylacrylic reagents. This protocol is based on a thiol–Michael-type addition of native or engineered cysteine residues to carbonylacrylic reagents equipped with functional compounds such as cytotoxic drugs. This approach is a robust alternative to the conventional maleimide technique; the reaction is irreversible and uses synthetically accessible reagents. Complete conversion to the conjugates, with improved quality and homogeneity, is often achieved using a minimal excess (typically between 5 and 10 equiv.) of the carbonylacrylic reagent. Potential applications of this method cover a broad scope of cysteine-tagged antibodies in various formats (full-length IgGs, nanobodies) for the site-selective incorporation of cytotoxic drugs without loss of antigen-binding affinity. Both the synthesis of the carbonylacrylic reagent armed with a synthetic molecule of interest and the subsequent preparation of the chemically defined, homogeneous antibody conjugate can be achieved within 48 h and can be easily performed by nonspecialists. Importantly, the conjugates formed are stable in human plasma. The use of liquid chromatography–mass spectrometry (LC–MS) analysis is recommended for monitoring the progression of the bioconjugation reactions on protein and antibody substrates with accurate resolution.**

## Introduction

Protein bioconjugation has been widely used in the design of biologically active conjugates for applications in biology and medicine[1–4]. Despite substantial progress in this field[5–8], there is still considerable interest in the development of efficient methods for site-selective chemical installation of modifications into proteins, especially those that can be applied to the covalent attachment of cytotoxic drugs to antibodies[1,9,10]. To date, most approaches for the preparation of antibody–drug conjugates (ADCs) have used methods that typically result in heterogeneous products containing a mixture of species with different drug-to-antibody ratios and, potentially, different pharmacokinetics. Current trends in ADC design are related to achieving site-selective antibody conjugation and a controlled drug-to-antibody ratio while forming products that are stable while in circulation to avoid premature drug release and thus side toxicity.

We have recently reported the design of a new class of Michael-acceptor reagents, carbonylacrylic (caa) derivatives, that undergo very rapid, chemoselective thiol–Michael addition with cysteine residues on proteins[11]. The reaction can proceed to completion even when a single molar equivalent of synthetically accessible reagents is used at low protein concentrations, generating homogeneous conjugates that are fully resistant to degradation in human plasma. The reaction typically results in a homogenous conjugate; most proteins do not have free cysteine residues, and, for most examples, we have engineered the proteins to have a unique cysteine. In examples with multiple cysteines, we have found that all available residues have reacted to form the corresponding thioether. The utility of this reaction has been demonstrated for the labeling of proteins, including antibodies that are available with a free cysteine through large-scale production. The modified proteins (albumin, Annexin-V, C2A domain of synaptotagmin-I (C2Am) and the antibody Thiomab LC-V205C) retained their

[1]Department of Chemistry, University of Cambridge, Cambridge, UK. [2]Departamento de Química, Universidade Federal de São Carlos, São Carlos, Brazil. [3]Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, Logroño, Spain. [4]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal. [5]Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brazil. *e-mail: gb453@cam.ac.uk; gbernardes@medicina.ulisboa.pt

ACS OMEGA

Article

Cite This: ACS Omega 2018, 3, 18142−18152

http://pubs.acs.org/journal/acsodf

# Oxygen by Carbon Replacement at the Glycosidic Linkage Modulates the Sugar Conformation in Tn Antigen Mimics

Claudio D. Navo,*[,†] Iris A. Bermejo,[†] Paula Oroz,[†] Pablo Tovillas,[†] Ismael Compañón,[†] Cristina Matías,[†] Alberto Avenoza,[†] Jesús H. Busto,[†] María M. Zurbano,[†] Gonzalo Jiménez-Osés,[†,‡,§] Francisco Corzana,*[,†] and Jesús M. Peregrina*[,†]

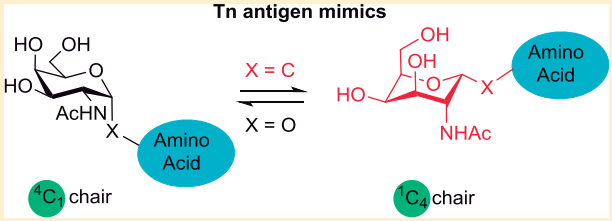
[†]Dept. Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, E-26006 Logroño, Spain

[‡]CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain

[§]Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain

Supporting Information

**ABSTRACT:** *N*-Acetylgalactosamine (GalNAc) α-O-linked to L-threonine (Thr) (Tn antigen) and several mimics of this Tn antigen have been synthesized to explore the impact of the underlying amino acid in the presentation mode of the carbohydrate moiety. The structural changes introduced in the Tn antigen mimics involve the replacement of the natural underlying Thr by non-natural amino acids while maintaining the α-*O*-glycosidic linkage of GalNAc or the substitution of this bond by α-*C*-glycosidic linkages. We also synthesized two bicyclic, conformationally restricted Tn antigen mimics. All of these compounds were subjected to a thorough conformational analysis in solution using NMR data, quantum mechanical (QM) calculations, and molecular dynamics simulations. Interestingly, in *C*-glycosides, the $^1C_4$ chair conformation of the pyranose ring was predicted to be stable by QM calculations and experimentally supported by nuclear Overhauser effect cross-peaks and coupling constants observed in the NMR experiments.




## INTRODUCTION

The Tn antigen is a specific human tumor-associated carbohydrate antigen (TACA) formed by *N*-acetylgalactosamine (GalNAc) α-O-linked to either serine (Ser) or threonine (Thr) residues.[1] Although it is quite small and has a simple structure, the Tn antigen has been attracting a great interest because it has been correlated with many types of tumors,[2−13] including breast carcinoma, where it is highly expressed,[2] and it has also been associated to metastatic behavior and tumor expansion.[1,13] In general, the Tn antigen can be found in mucins,[14−16] which are the most abundant glycoproteins in mucus, playing a key role in several biological processes,[17−19] such as tissue inflammation, immune response, or intercellular recognition. In tumor cells, mucins usually present low glycosylation and abnormal sugar chain extensions as a result of the malfunctioning of some glycosyltransferases.[4] This fact results in an exposure of antigens on the surface of cancer cells.[20−23] Consequently, the immune system recognizes these antigens and can trigger an immune response. Different monoclonal antibodies are able to recognize these exposed antigens, binding specifically to cancer cells.

Thus, TACAs and particularly the Tn antigen are considered promising targets for the development of cancer immunotherapy (carbohydrate-based vaccines).[14−16,24−28] Several groups have reported the synthesis and immunological evaluation of cancer therapeutic vaccines, but,[14,29−42] unfortunately, most of the carbohydrate-based vaccines have failed in clinical trials, probably due to immunotolerance or resistance to carbohydrate antibody efficacy.[43] Consequently, there is a great deal of interest in synthesizing non-natural Tn antigen analogues, which would be more resistant to degradation. The synthesis and biological evaluation of analogues of natural carbohydrates and glycopeptides[44−46] can contribute to understanding different biochemical processes and provide new candidates for biological targets. We have reported several examples of non-natural Tn antigen mimics[47−53] comprising minor structural modifications, some of them being able to imitate the conformational preferences of the natural antigens in solution, enhancing the binding affinity to anti-MUC1 antibodies. Herein, we continue exploring this strategy and report new variants in which the underlying amino acid of the Tn antigen (**1**) has been altered and in some cases covalently linked to the carbohydrate to fix the native conformation of the glycosidic linkage.[54−56] The explored modifications include the substitution of Thr by α-methylserine (mimic **2**)[49,52,57] or isoserine (mimic **3**),[52] two compounds already prepared by us but whose conformational analysis has not been accomplished yet. Additionally, we synthesize and analyze in this work novel

Received: September 28, 2018
Accepted: December 4, 2018
Published: December 24, 2018






DOI: 10.1021/acsomega.8b02576
ACS Omega 2018, 3, 18142−18152

Tn antigen mimics (**4**, **5**, and **6**). A cyclic carbamate connecting the *N*-acetyl moiety of GalNAc and the amino group of Thr was introduced in mimic **4**. Other changes include connecting GalNAc to the C$\alpha$ of Ser, forming a *C*-glycoside derivative (mimic **5**), and incorporating a cyclic urea to generate a bicyclic *C*-glycoside derivative (mimic **6**, Figure 1).

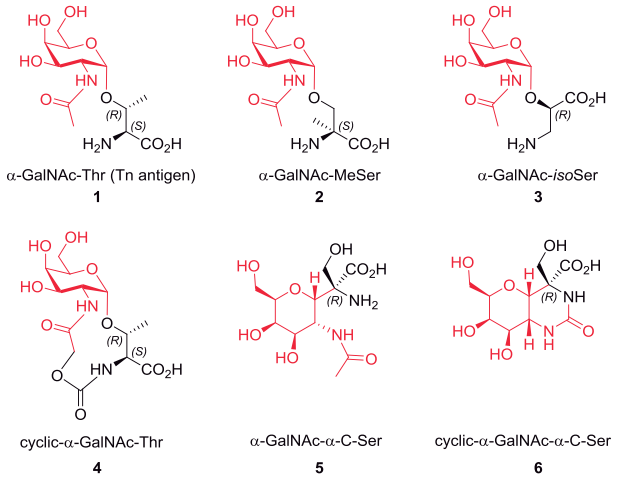



**Figure 1.** Structures of the natural Tn antigen (**1**) and different non-natural mimics (**2**–**6**) studied in this work.

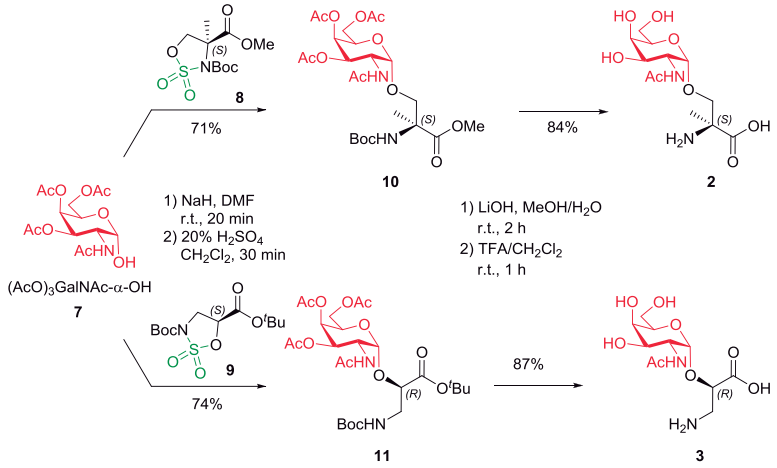
## RESULTS AND DISCUSSION

**Synthesis.** Initially, we synthesized the Tn antigen mimics **2** and **3** through our recently reported strategy involving the ring-opening of chiral cyclic sulfamidates with carbohydrate C1–*O*-nucleophiles.[52] Thus, starting from effectively accessible (*S*)-$\alpha$-methylserine and commercially available (*S*)-isoserine, enantiopure cyclic sulfamidates **8** and **9** were efficiently prepared as chiral building blocks for the synthesis of GalNAc-*O*-glycosylated amino acids (*S*)-$\alpha$-methylserine and (*S*)-isoserine, respectively. The nucleophilic ring-opening reactions of these cyclic sulfamidates **8** and **9** with tri-*O*-acetyl-$\alpha$-D-*N*-acetylgalactosamine (**7**)[52] using sodium hydride as a base in dimethylformamide were highly chemo-, regio-, and stereoselective and they provided good yields of protected glycosyl-amino-acids **10** and **11**, respectively, which were hydrolyzed to obtain the required Tn antigen mimics **2** and **3** (Scheme 1).

The synthesis of glycoside **4** started with the Koenigs–Knorr glycosylation of *N*-Fmoc-threonine *tert*-butyl ester[58] **13** with tri-*O*-acetyl-protected 2-(azido)galactopyranosyl chloride derivative[59] **12** in the presence of silver salts as promotors, giving a mixture of the $\alpha$- and $\beta$-anomers, which could be purified by column chromatography to obtain $\alpha$-anomer[60] **14** in a 30% yield (Scheme 2). The azido group of glycoside **14** was transformed into the amino group using Zn in an acidic medium (AcOH and HCl) to obtain compound **15** in a good yield. This compound was treated with bromoacetic acid to generate the corresponding amide **16** using *N*,*N*′-diisopropylcarbodiimide as a coupling reagent and *N*,*N*-diisopropylethylamine as a base in a mixture of tetrahydrofuran (THF)/$H_2O$ at 0 °C for 1 h. The treatment of compound **16** with 1,8-diazabicyclo[5.4.0]undec-7-ene as a non-nucleophilic base in methylene chloride at 0 °C gave the bicyclic compound **17** in a moderate yield. The proposed mechanism involves the abstraction of hydrogen of Fmoc, indicated in blue in Scheme 2, to give 9-methylene-9*H*-fluorene followed by the intramolecular nucleophilic attack of the generated carbamate to the $\alpha$-bromoacetamide group of the glycosyl moiety. The bromine atom acts as a leaving group to give the corresponding cyclic carbamate **17**. Once purified by column chromatography, the *tert*-butyl ester group was hydrolyzed with trifluoroacetic acid to obtain derivative **18**, which was deacetylated by treatment with sodium methoxide in methanol[61] to give bicyclic glycoside **4** (Scheme 2).

Considering our previous experience in the field of *C*-glycosides,[62,63] we anticipated the synthesis of two *C*-glycosides as new Tn antigen mimics. The starting materials for the synthesis of serine *C*-glycoside **5** were the 2-nitro-tri-*O*-benzyl-D-galactal **19**, obtained following the methodology described by Schmidt and coworkers,[64,65] and the serine-derived *N*,*O*-bicyclic acetals **20a**–**c** synthesized using our

**Scheme 1. Synthesis of Tn Antigen Mimics Featuring (*S*)-$\alpha$-Methylserine (2) or (*S*)-Isoserine (3)**[a]

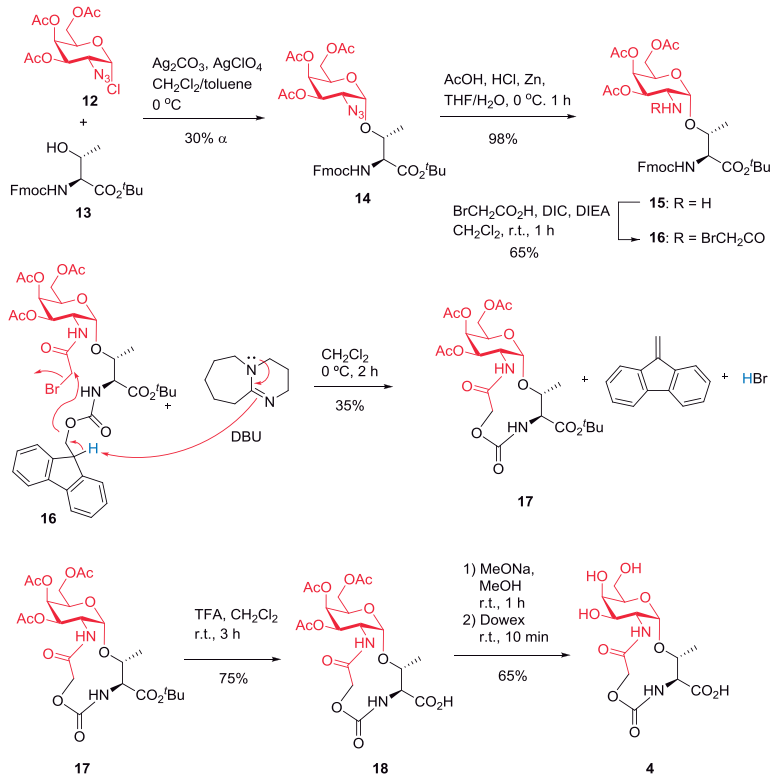



[a]The key step involves the ring-opening reaction of chiral cyclic sulfamidates **8** and **9** with the carbohydrate C1–*O*-nucleophile **7**, followed by hydrolysis of the ring-opened intermediates **10** and **11**.

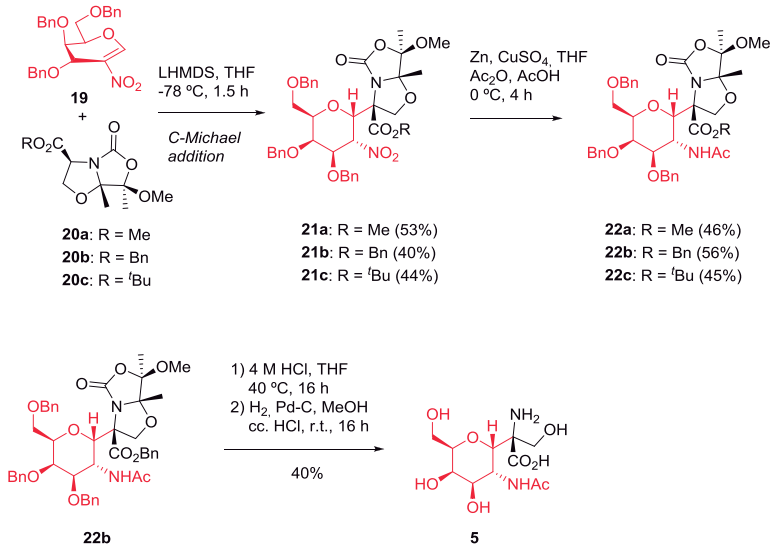


DOI: 10.1021/acsomega.8b02576
ACS Omega 2018, 3, 18142–18152

**Scheme 2. Synthesis of the Tn Antigen Mimic 4**




**Scheme 3. Synthesis of the Tn Antigen Mimic 5 Using a Michael Addition as a Key Step**



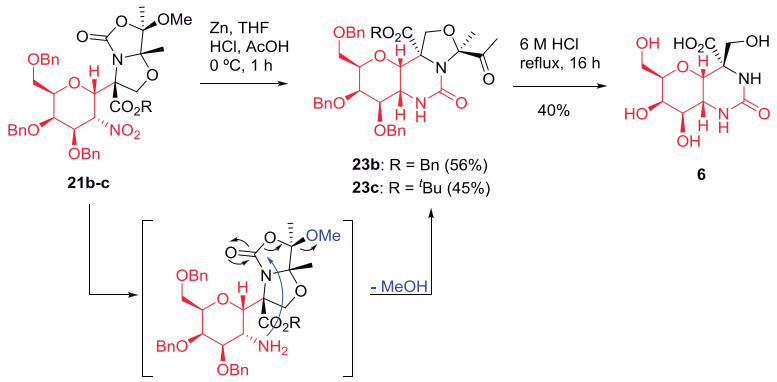

reported methodology[66] from commercially available (*S*)−*N*-Boc-serine methyl ester.

The first step involved a double diastereoselective Michael-type addition of the enolates of bicyclic serine equivalents **20a−c** to the nitrogalactal derivative **19**. THF was used as a solvent, the temperature was set to −78 °C, and lithium bis(trimethylsilyl)amide (LHMDS) was used as a base, obtaining moderate yields of the corresponding Michael adducts **21a−c**, respectively, as previously reported.[62] Notably, in all cases, a unique diastereoisomer **21a**, **21b**, or **21c** was obtained among the eight possible ones. The next step was a selective reduction of the nitro group of compounds **21a−c** and subsequent acetylation of the amines formed. Compounds **21a−c** were treated with Zn and a saturated aqueous solution of $CuSO_4$ in THF−AcOH−$Ac_2O$ (3:2:1) at room temperature for 4 h.[67−69] The corresponding *N*-acetamides **22a−c** were obtained in moderate yields (Scheme 3). Compound **22b** was then reacted with an aqueous 4 M HCl solution in THF at 40 °C for 16 h to afford the corresponding aminoalcohol hydrochloride, which was directly subjected to hydrogenolysis,



DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

**Scheme 4. Synthesis of the Conformationally Restricted Tn Antigen Mimic 6**



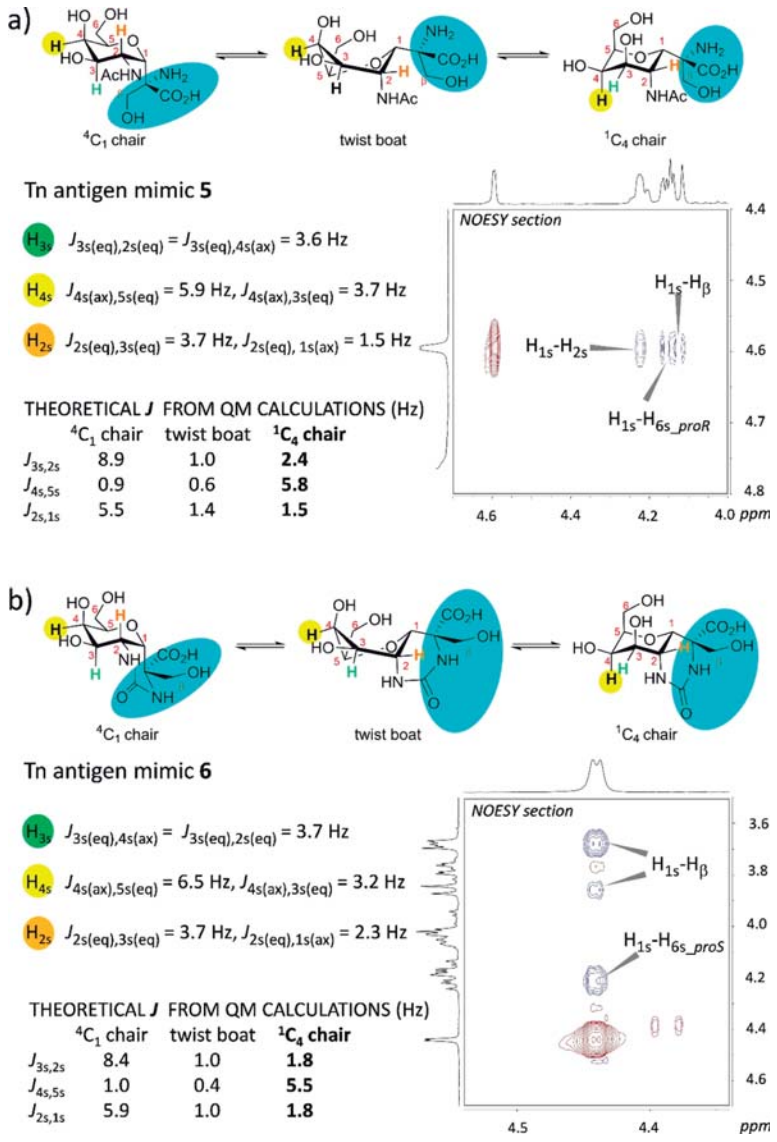




**Figure 2.** Spatial disposition of $H_{3s}$, $H_{4s}$, and $H_{2s}$ protons in $^4C_1$ and $^1C_4$ chair and twist-boat conformations for compounds **5** (a) and **6** (b). According to the coupling constants and the NOESY spectra in $H_2O/D_2O$ (9:1) at 25 °C, in which the most relevant cross-peaks for structural characterization are highlighted, the $^1C_4$ chair conformation should be predominant in solution for compounds **5** and **6**.

using Pd−C as a catalyst in methanol (MeOH) at room temperature for 16 h. Three drops of concentrated HCl were



DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

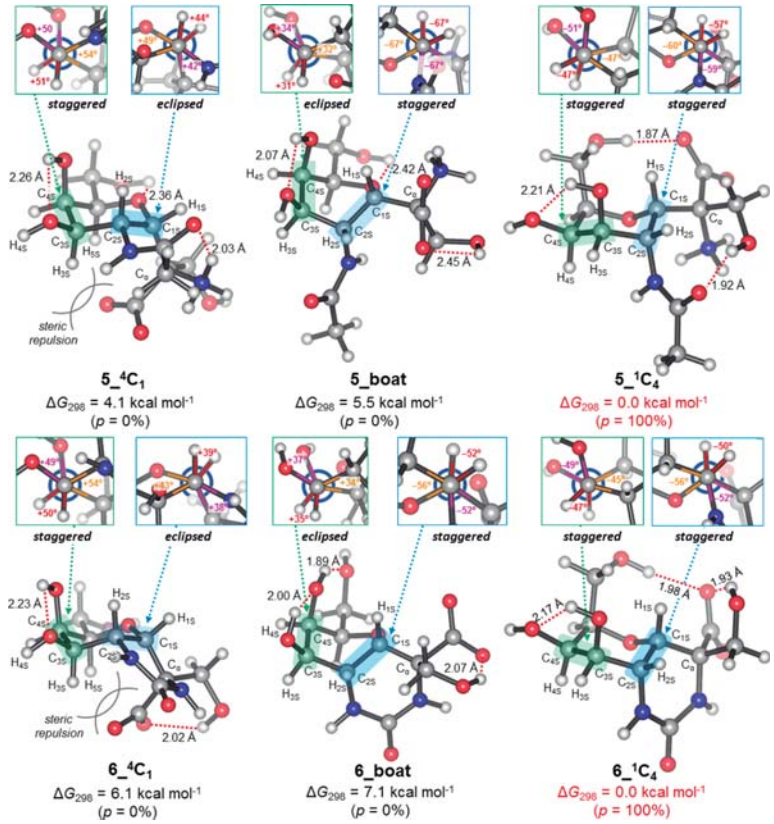



**Figure 3.** Lowest energy structures for the $^4C_1$ (left), twist-boat (middle), and $^4C_1$ (right) conformations for compounds **5** (top) and **6** (bottom) calculated with PCM($H_2O$)/M06-2X/6-31+G(d,p). Free energies at 298 K ($\Delta G_{298}$) are given in kcal $mol^{-1}$ and relative populations ($p$) derived from $\Delta G_{298}$ are shown in parentheses. Torsional strain is represented through the dihedral angles highlighted in three different colors (red, orange, and magenta) at the Newman projections from C 3s to C 4s (carbohydrate, highlighted in green) and from C 2s to C 1s (aglycone, highlighted in blue). Dihedral angles close to ±60° correspond to more staggered conformations. Hydrogen bonds are represented with doted red lines. Distances are given in angstrom.

required to achieve complete deprotection. Purification by reversed-phase high-performance liquid chromatography (RP-HPLC) afforded *C*-glyco-amino-acid **5** (Scheme 3).

The conformationally restricted bicyclic *C*-glycoside **6** was synthesized from Michael adducts **21b** and **21c** (Scheme 4). Treatment of compounds **21b** and **21c** with Zn dust in THF−HCl−AcOH at 0 °C for 1 h reduced their nitro groups into the corresponding amines, which immediately reacted with the oxazolidinone forming bicyclic ureas **23b−c**. Spontaneous carbonyl recovery led to the formation of a pendant methylketone group with concomitant methanol loss. Acidic hydrolysis using an aqueous 6 M HCl solution under reflux for 16 h led to complete deprotection. Purification by RP-HPLC afforded the amino-restricted *C*-glycoside **6** in moderate yield (40%) as a new conformationally restricted Tn antigen mimic (Scheme 4).

**Conformational Study.** The conformational analysis of the Tn antigen **1** and its mimics (**2−6**) in aqueous solution was then performed by NMR and molecular modeling. While in compounds **1−4** the *α*-*O*-GalNAc unit displays the typical $^4C_1$ chair conformation,[70] in *α*-*C*-glycosides **5** and **6**, the absence of the anomeric effect may promote the coexistence of both $^4C_1$ and $^1C_4$ chairs and the twist-boat.[71−73] In agreement with this notion, the signals of $H_{3s}$ proton observed at 3.86 ppm in the $^1H$ NMR spectrum of compound **5** correspond to a degenerated doublet of doublets that collapses in a pseudo-triplet with two similar small coupling constants of an average value of 3.6 Hz, which is indicative of an equatorial position of this proton and suggesting the prevalence of the $^1C_4$ chair conformation for this compound (Figure 2a). Regarding the small $^3J_{2s,1s}$, with an experimental value of 1.5 Hz, the $^4C_1$ chair conformation can be discarded because in this case, this coupling constant should take values close to 6 Hz. However, with these experimental data, we cannot exclude the occurrence of the twist-boat conformation in solution. Fortunately, although the signals corresponding to $H_{4s}$ and $H_{5s}$ protons are collapsed with other protons, we could extract their coupling constants from the spectra. The value of 5.9 Hz for $^3J_{4s,5s}$ agrees with the existence of the $^1C_4$ chair conformation and is incompatible with both the $^4C_1$ chair and twist-boat conformations. Next, we studied the coupling constants $^3J_{5s,6s\_proS}$ and $^3J_{5s,6s\_proR}$, which give relevant information on the rotamer distribution around the C 5s−C 6s bond. According to their values (2.2 and 10.4 Hz, respectively), we deduced the prevalence of gauche−trans conformation for the hydroxymethyl group. Moreover, with the help of 2D nuclear Overhauser enhancement spectroscopy (NOESY) spectra, $H_{6s\_proS}$ and $H_{6s\_proR}$ were correctly assigned. Therefore, the medium-size nuclear Overhauser effect (NOE) cross-peaks observed between the anomeric $H_{1s}$ and $H_{6s\_proR}$ protons also corroborate the presence of the $^1C_4$ chair in solution (Figure 2a). This NOE is not compatible



DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

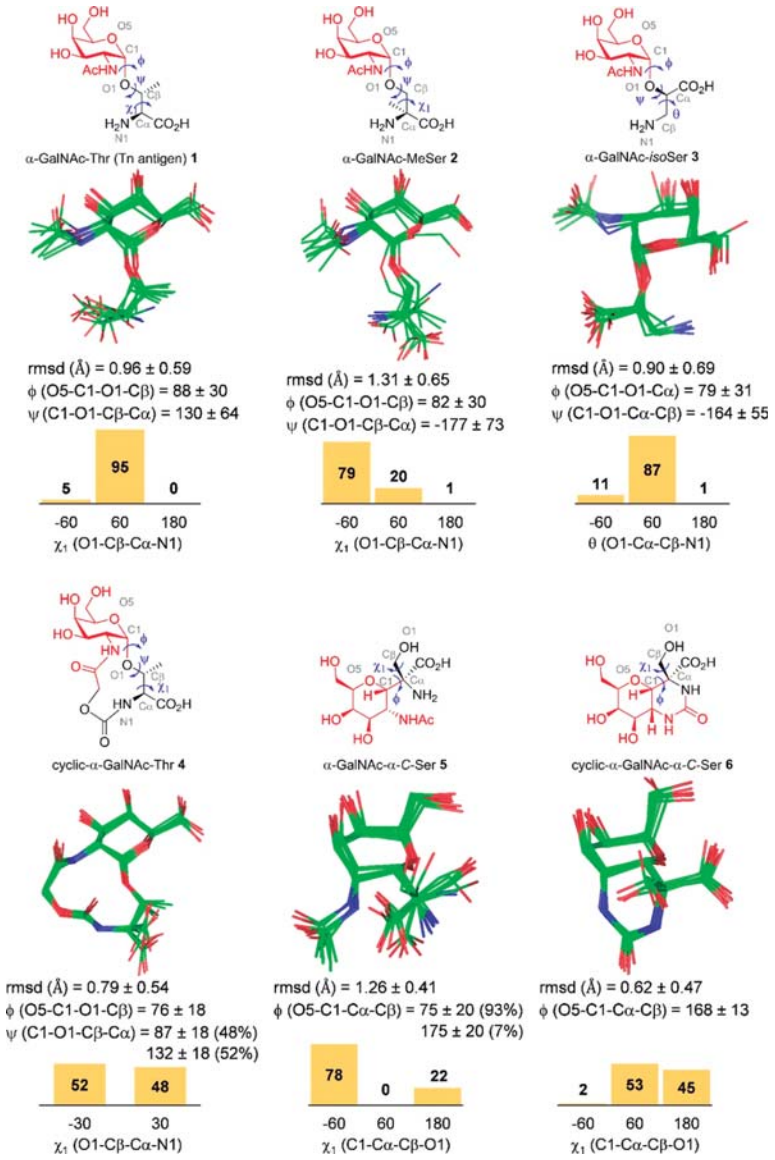



**Figure 4.** Ensembles derived from the 0.5 $\mu$s MD simulations for compound **1** and Tn antigen mimics **2**–**6**. The root-mean-square deviation (Å) values for heavy atom superimposition, together with the main values for the most relevant torsional angles are also indicated. The distribution of $\chi^1$ is shown at the bottom of each derivative.

with the existence of a $^4C_1$ chair or a twist-boat conformer in solution. A similar deduction can be applied to *C*-glycoside **6** (Figure 2b). To conclude the conformational analysis in solution, we accomplished a $^1$H NMR experiment at different temperatures for compound **6**. Under these conditions, the absence of changes in the signals and in the coupling constant agrees with the existence of a unique conformer in water[74] (Figure S30).

Quantum mechanical (QM) conformational analyses were performed for all compounds, and the lowest energy structures for the $^1C_4$ and $^4C_1$ chair and boat conformations in compounds **5** and **6** were calculated with PCM($H_2O$)/M06-2X/6-31+G(d,p). In line with the experimental NMR data, the free energies obtained indicated that the $^1C_4$ chair conformation is the most stable arrangement for the carbohydrate moiety of Tn antigen mimics **5** and **6** (Figure 3 and Supporting Information). In the same way, in the case of the hydroxymethyl group conformation of the saccharide fragment, the gauche−trans arrangement shows prevalence over the trans−gauche or gauche−gauche dispositions in all studied compounds. These data are in agreement with previous published works.[75,76] Exceptionally, in bicyclic system **6**, we observed that the gauche−gauche is isoenergetic with gauche−trans disposition. Additionally, to corroborate our conformational deductions from the NMR data, we calculated the key coupling constants (Figure 2, see *J* values in bold) using the dihedral angles obtained from the optimized geometries for the lowest energy conformers and the Karplus-like equations implemented in the MestReJ software,[77] matching very well with the $^1C_4$ chair conformation.

It is well known that the $^4C_1$ chair conformation is the most stable for *O*-glycosides with a pyranose structure and is determined mainly by the anomeric and exo-anomeric effects.[71−73] This fact occurs in Tn mimics **1**–**4** as we can



DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

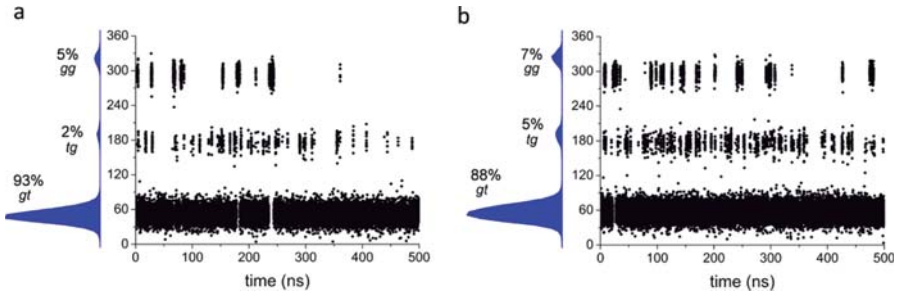



**Figure 5.** Rotamer distribution around the C 5s−C 6s bond derived from 0.5 $\mu$s MD simulations for Tn mimics **5** (a) and **6** (b). The values of the torsional angle $\omega$ (O5−C5−C6−O6) are shown.

see in the structures obtained from molecular dynamics (MD) simulations (see below, Figure 4). However, these electronic effects do not exist when the oxygen is replaced by a carbon atom at the glycosidic linkage in derivatives **5** and **6** (*C*-glycosides).

The greater stability of the $^1C_4$ chair conformations in both $\alpha$-*C*-glycosides **5** and **6** with respect to the $^4C_1$ chair and twist-boat conformations can be attributed to a number of additive contributions. First, the lack of anomeric effects due to the absence of $\alpha$-*O*-glycosidic bonds at the anomeric carbon C 1s precludes the normal stabilization of the $^4C_1$ chair because of donor−acceptor interactions from the O 5s endocyclic oxygen lone pairs ($n_{Oendo}$) to the antibonding orbital of the C 1s−OR bond ($\sigma^*_{C1-Oexo}$) (endo-anomeric effect) and from the OR exocyclic oxygen lone pairs ($n_{Oexo}$) to the antibonding orbital of the C 1s−O 5s bond ($\sigma^*_{C1-Oendo}$) (exo-anomeric effect).[71−73] Second, accumulation of torsional strain at the carbohydrate moiety in the twist-boats and at the aglycone region in the $^4C_1$ chairs (Figure 3, see highlighted dihedral angles) destabilizes these arrangements with respect to the $^1C_4$ conformations, which avoid such torsional strain by placing substituents at C 1s and C 2s in relaxed equatorial and axial dispositions, respectively. On the contrary, the presence of bulky carboxylate groups at C$\alpha$ of the aglycone distorts the C 1s−C$\alpha$ bond from axial to pseudoaxial disposition in the $^4C_1$ chairs,[78] thus increasing torsional strain, particularly in the bicyclic compound **6**. As a result, this compound adopts a much more stable *cis*-decalin-like arrangement which overcomes the 1,3-diaxial strain arising from placing the C 3s−hydroxy and C 5s−hydroxymethyl groups in mutually axial positions. Finally, the aglycone C$\alpha$−carboxylate groups establish strong hydrogen bond networks in the $^1C_4$ chair conformations of both compounds **5** and **6**, which contribute to largely stabilize this unusual arrangement with respect to the $^4C_1$ chair and twist-boat conformations.

Next, 0.5 $\mu$s MD simulations in explicit water were performed on compounds **1**−**6**. The most stable $^1C_4$ chair conformation was used in the starting structures of compounds **5** and **6**. The conformational ensembles obtained for derivatives are shown in Figure 4, and the normalized frequencies of the most relevant dihedral angles for compounds **1**−**6** are depicted in Figure S31. The good agreement between the experimental and theoretical $J_{H\alpha,H\beta}$ coupling constants obtained for compound **1** (2.0 and 2.4 Hz, respectively) validate the MD simulations on this natural derivative.[54−56] Additionally, average $^3J_{H,H}$ values extracted from this MD simulations for compounds **5** and **6** are in agreement with experimental $^3J_{H-H}$ values extracted from NMR experiments in $D_2O$ and with theoretical $^3J_{H-H}$ values for the quantum mechanics-calculated minimum energy structures of the $^1C_4$ conformations (Tables S1 and S2). According to the MD simulations, the $\phi$ torsional angle displays values close to 80° in compounds **1**−**4**, in good accordance with the exo-anomeric effect.[71−73] While in compound **1**, the glycosidic linkage shows the typical "eclipsed" conformation, in derivatives **2** and **3**, the staggered conformation is mainly populated. As previously reported by us,[57] the glycosidic linkage of compound **2** was flexible when the underlying amino acid was presented as a diamide. However, the zwitterion form favors a rigid side chain, with a main value for ($\chi^1$) close to −60°. In compound **4**, although the global structure is rigid, two similarly populated conformers were found for both $\psi$ and $\chi^1$ torsional angles. On the other hand and in good agreement with the QM calculations commented above, **5** and **6** display a rigid *C*-glycosidic linkage. We also analyzed the conformational preferences of the hydroxymethyl group of the sugar moiety. In compounds **1** and **2**, featuring a $^4C_1$ chair conformation, a mixture of the gauche−trans (60%) and trans−gauche (26%) is observed, as deduced experimentally from the values of $^3J_{5s,6s\_proS}$ and $^3J_{5s,6s\_proR}$ (4.7 and 7.5 Hz, respectively). This result agrees with previous studies performed on galactose derivatives.[79] On the contrary, in Tn mimics **5** and **6**, which display the $^1C_4$ chair disposition, the gauche−trans rotamer is mainly populated (87%). This experimental outcome is reproduced by the MD simulations (Figure 5) and agrees with the experimental values of the coupling constants (Tables S1 and S2).

Our MD simulations indicate that the intramolecular hydrogen bonds for compounds **1**−**6** are irrelevant for their 3D disposition, with populations in all cases <20%. This result agrees with the radial distribution functions calculated for all derivatives (Figure S32).[80] Finally, to delimit the role of water in the conformational preferences of these derivatives, we performed the analysis of the first hydration shell. Interestingly, $NH_3^+$ promotes the existence of bridging water molecules between the carbohydrate and amino acid moieties not only in the Tn bearing a threonine residue (Tn−Thr)[55] but also in derivatives **2** and **3** (Figure 6). This feature can contribute to rigidify these derivatives in water. A similar localized water density is found between the NH groups of derivative **4**, mimicking the conformational behavior of the Tn−Thr antigen.

Altogether, these data indicate that mimics **1**−**6** are rather stiff, displaying a main conformation in solution.

## ■ CONCLUSIONS

Several mimics of the Tn antigen have been synthesized by replacing the natural underlying threonine by the non-natural



DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

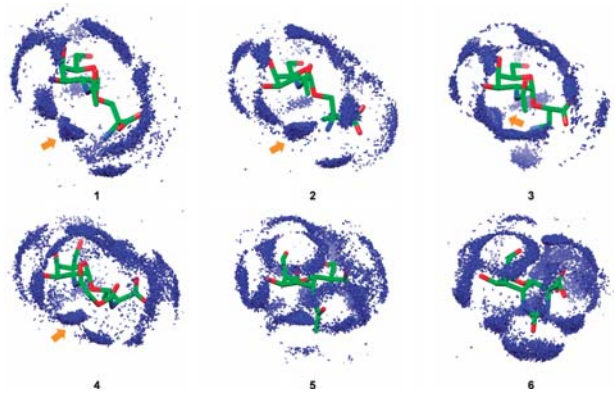



**Figure 6.** Averaged first hydration shell derived from 0.5 $\mu$s MD simulations for compounds **1**−**5**. The arrows show the localized water pockets found in compounds **1**−**4** between the amino acid and carbohydrate moieties.

amino acids isoserine and $\alpha$-methylserine, either maintaining the $\alpha$-*O*-glycosidic linkage with the GalNAc moiety or replacing this bond by $\alpha$-*C*-glycosidic linkages. Additionally, on the basis of our preliminary conformational studies on the Tn antigen,[54−56] in which we demonstrated the rigidity of the glycosidic bond and the water-mediated connection between the carbohydrate and amino acid moieties, we synthesized two bicyclic Tn antigen mimics. The conformational analysis using NMR, QM, and MD of all of the studied compounds suggests that in *C*-glycoside derivatives, the pyranose ring adopts a $^1C_4$ chair conformation. In future works, we will evaluate the binding affinity of these Tn mimics toward different targets of biological interest.

## EXPERIMENTAL SECTION

**Reagent and General Procedures.** We have included in the Supporting Information a section to describe thoroughly the commercial reagents used, the synthetic protocols followed, and the apparatus used to characterize the compounds. Moreover, the synthesis of all compounds is described in the Supporting Information.

**NMR Experiments.** NMR protocols and equipment used to carry out the NMR experiments ($^1$H and $^{13}$C NMR, COSY, HSQC, and NOESY) are detailed in the Supporting Information.

**MD Simulations.** The calculations were performed using the AMBER 16 package[81] implemented with GAFF2[82] force field. More details are shown in the Supporting Information.

**Quantum Mechanics.** Full geometry optimizations were performed using Gaussian 16[83] with the M06-2X hybrid functional[84] and 6-31+G(d,p) basis set. Bulk solvent effects in water were considered implicitly using the IEF-PCM polarizable continuum model.[85] Different conformations were considered for all structures. More information is detailed in the Supporting Information.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b02576.

Additional experimental details, NMR spectra, and computational information including figures (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

*E-mail: claudio-daniel.navon@unirioja.es (C.D.N.).
*E-mail: francisco.corzana@unirioja.es (F.C.).
*E-mail: jesusmanuel.peregrina@unirioja.es (J.M.P.).

### ORCID

Gonzalo Jiménez-Osés: 0000-0003-0105-4337
Francisco Corzana: 0000-0001-5597-8127
Jesús M. Peregrina: 0000-0003-3778-7065

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We are grateful to MINECO (projects CTQ2015-67727-R and UNLR13-4E-1931 to J.M.P. and F.C., CTQ2015-70524-R and RYC-2013-14706 to G.J.-O. and C.D.N.). I.A.B. thanks the Asociación Española Contra el Cáncer en La Rioja (AECC) for a grant. We also thank CESGA (Universidad de Santiago de Compostela) and BERONIA (Universidad de La Rioja) for computer support. F.C. thanks the EU (Marie-Sklodowska Curie ITN, ProteinConjugates).

## REFERENCES

(1) Ju, T.; Otto, V. I.; Cummings, R. D. The Tn antigen-structural simplicity and biological complexity. *Angew. Chem., Int. Ed.* **2011**, *50*, 1770−1791.
(2) Springer, G. F.; Desai, P. R.; Banatwala, I. Blood group MN specific substances and precursors in normal and malignant human breast tissues. *Naturwissenschaften* **1974**, *61*, 457−458.
(3) Davidson, B.; Berner, A.; Nesland, J. M.; Risberg, B.; Kristensen, G. B.; Tropé, C. G.; Bryne, M. Carbohydrate antigen expression in primary tumors, metastatic lesions, and serous effusions from patients diagnosed with epithelial ovarian carcinoma: Evidence of up-regulated Tn and Sialyl Tn antigen expression in effusions. *Hum. Pathol.* **2000**, *31*, 1081−1087.
(4) Costa, C.; Pereira, S.; Lima, L.; Peixoto, A.; Fernandes, E.; Neves, D.; Neves, M.; Gaiteiro, C.; Tavares, A.; da Costa, R. M. G.; Cruz, R.; Amaro, T.; Oliveira, P. A.; Ferreira, J. A.; Santos, L. L. Abnormal protein glycosylation and activated PI3K/Akt/mTOR pathway: Role in bladder cancer prognosis and targeted therapeutics. *PLoS One* **2015**, *10*, e0141253.
(5) Hirao, T.; Sakamoto, Y.; Kamada, M.; Hamada, S.-I.; Aono, T. Tn antigen, a marker of potential for metastasis of uterine cervix cancer cells. *Cancer* **1993**, *72*, 154−159.
(6) Terasawa, K.; Furumoto, H.; Kamada, M.; Aono, T. Expression of Tn and Sialyl-Tn antigens in the neoplastic transformation of uterine cervical epithelial cells. *Cancer Res* **1996**, *56*, 2229−32.
(7) Itzkowitz, S. H.; Yuan, M.; Montgomery, C. K.; Kjeldsen, T.; Takahashi, H. K.; Bigbee, W. L.; Kim, Y. S. Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. *Cancer Res* **1989**, *49*, 197−204.
(8) Byrd, J. C.; Bresalier, R. S. Mucins and mucin binding proteins in colorectal cancer. *Cancer Metastasis Rev.* **2004**, *23*, 77−99.
(9) López-Ferrer, A.; Barranco, C.; de Bolós, C. Differences in the O-glycosylation patterns between lung squamous cell carcinoma and adenocarcinoma. *Am. J. Clin. Pathol.* **2002**, *118*, 749−755.
(10) Ohshio, G.; Imamura, T.; Imamura, M.; Yamabe, H.; Sakahara, H.; Nakada, H.; Yamashina, I. Distribution of Tn antigen recognized by an anti-Tn monoclonal antibody (MLS128) in normal and malignant tissues of the digestive tract. *J. Cancer Res. Clin. Oncol.* **1995**, *121*, 247−252.
(11) Remmers, N.; Anderson, J. M.; Linde, E. M.; DiMaio, D. J.; Lazenby, A. J.; Wandall, H. H.; Mandel, U.; Clausen, H.; Yu, F.; Hollingsworth, M. A. Aberrant expression of mucin core proteins and


DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

O-linked glycans associated with progression of pancreatic cancer. *Clin. Cancer Res.* **2013**, *19*, 1981−1993.
(12) Zhang, S.; Zhang, H. S.; Reuter, V. E.; Slovin, S. F.; Scher, H. I.; Livingston, P. O. Expression of potential target antigens for immunotherapy on primary and metastatic prostate cancers. *Clin Cancer Res* **1998**, *4*, 295−302.
(13) 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* **2016**, *88*, 275−286.
(14) Strous, G. J.; Dekker, J. Mucin-type glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* **1992**, *27*, 57−92.
(15) Hang, H. C.; Bertozzi, C. R. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg. Med. Chem.* **2005**, *13*, 5021−5034.
(16) Fumoto, M.; Hinou, H.; Ohta, T.; Ito, T.; Yamada, K.; Takimoto, A.; Kondo, H.; Shimizu, H.; Inazu, T.; Nakahara, Y.; Nishimura, S.-I. Combinatorial synthesis of MUC1 glycopeptides: Polymer blotting facilitates chemical and enzymatic synthesis of highly complicated mucin glycopeptides. *J. Am. Chem. Soc.* **2005**, *127*, 11804−11818.
(17) Dwek, R. A. Glycobiology: Toward understanding the function of sugars. *Chem. Rev.* **1996**, *96*, 683−720.
(18) Sears, P.; Wong, C.-H. Enzyme action in glycoprotein synthesis. *Cell. Mol. Life Sci.* **1998**, *54*, 223−252.
(19) Johansson, M. E. V.; Hansson, G. C. Immunological aspects of intestinal mucus and mucins. *Nat. Rev. Immunol.* **2016**, *16*, 639−649.
(20) Sell, S. Cancer-associated carbohydrates identified by monoclonal antibodies. *Hum. Pathol.* **1990**, *21*, 1003−1019.
(21) Hakomori, S.-I.; Zhang, Y. Glycosphingolipid antigens and cancer therapy. *Chem. Biol.* **1997**, *4*, 97−104.
(22) Taylor-Papadimitriou, J.; Epenetos, A. A. Exploiting altered glycosylation patterns in cancer: Progress and challenges in diagnosis and therapy. *Trends Biotechnol.* **1994**, *12*, 227−233.
(23) Gabius, H.-J. Animal lectins. *Eur. J. Biochem.* **1997**, *243*, 543−576.
(24) Buskas, T.; Thompson, P.; Boons, G.-J. Immunotherapy for cancer: synthetic carbohydrate-based vaccines. *Chem. Commun.* **2009**, 5335−5349.
(25) Gaidzik, N.; Westerlind, U.; Kunz, H. The development of synthetic antitumour vaccines from mucin glycopeptide antigens. *Chem. Soc. Rev.* **2013**, *42*, 4421−4442.
(26) Wolfert, M. A.; Boons, G.-J. Adaptive immune activation: glycosylation does matter. *Nat. Chem. Biol.* **2013**, *9*, 776−784.
(27) 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.* **2013**, *135*, 14462−14472.
(28) 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.* **2014**, *53*, 1699−1703.
(29) 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.* **2001**, *40*, 366−369.
(30) Brocke, C.; Kunz, H. Synthesis of tumor-associated glycopeptide antigens. *Bioorg. Med. Chem.* **2002**, *10*, 3085−3112.
(31) Dziadek, S.; Kunz, H. Synthesis of tumor-associated glycopeptide antigens for the development of tumor-selective vaccines. *Chem. Rec.* **2004**, *3*, 308−321.
(32) Lai, Z.; Schreiber, J. R. Antigen processing of glycoconjugate vaccines; the polysaccharide portion of the pneumococcal CRM(197) conjugate vaccine co-localizes with MHC II on the antigen processing cell surface. *Vaccine* **2009**, *27*, 3137−3144.
(33) Costantino, P.; Rappuoli, R.; Berti, F. The design of semi-synthetic and synthetic glycoconjugate vaccines. *Expert Opin. Drug Discovery* **2011**, *6*, 1045−1066.
(34) Tarp, M. A.; Clausen, H. Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochim. Biophys. Acta, Gen. Subj.* **2008**, *1780*, 546−563.
(35) Hanisch, F.-G.; Ninkovic, T. Immunology of O-glycosylated proteins: approaches to the design of a MUC1 glycopeptide-based tumor vaccine. *Curr. Protein Pept. Sci.* **2006**, *7*, 307−315.
(36) 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.* **2017**, *46*, 7154−7175.
(37) Danishefsky, S. J.; Allen, J. R. From the laboratory to the clinic: A retrospective on fully synthetic carbohydrate-based anticancer vaccines. *Angew. Chem., Int. Ed.* **2000**, *39*, 836−863.
(38) 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 α-N-acetylgalactosamine-O-serine/threonine conjugate vaccine. *J. Clin. Oncol.* **2003**, *21*, 4292−4298.
(39) Galonić, D. P.; Gin, D. Y. Chemical glycosylation in the synthesis of glycoconjugate antitumour vaccines. *Nature* **2007**, *446*, 1000−1007.
(40) Dube, D. H.; Bertozzi, C. R. Glycans in cancer and inflammation - potential for therapeutics and diagnostics. *Nat. Rev. Drug Discovery* **2005**, *4*, 477−488.
(41) Ragupathi, G.; Coltart, D. M.; Williams, L. J.; Koide, F.; Kagan, E.; Allen, J.; Harris, C.; Glunz, P. W.; Livingston, P. O.; Danishefsky, S. J. On the power of chemical synthesis: immunological evaluation of models for multiantigenic carbohydrate-based cancer vaccines. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13699−13704.
(42) Buskas, T.; Ingale, S.; Boons, G.-J. Towards a fully synthetic carbohydrate-based anticancer vaccine: synthesis and immunological evaluation of a lipidated glycopeptide containing the tumor-associated Tn antigen. *Angew. Chem., Int. Ed.* **2005**, *44*, 5985−5988.
(43) Beatty, G. L.; Gladney, W. L. Immune escape mechanisms as a guide for cancer immunotherapy. *Clin. Cancer Res.* **2014**, *21*, 687−692.
(44) Westerlind, U. Synthetic glycopeptides and glycoproteins with applications in biological research. *Beilstein J. Org. Chem.* **2012**, *8*, 804−818.
(45) Jiménez-Barbero, J.; Dragoni, E.; Venturi, C.; Nannucci, F.; Ardá, A.; Fontanella, M.; André, S.; Cañada, F. J.; Gabius, H.-J.; Nativi, C. Alpha-O-linked glycopeptide mimetics: synthesis, conformation analysis and interactions with viscumin, a galactoside-binding model lectin. *Chem.—Eur. J.* **2009**, *15*, 10423−10431.
(46) Koester, D. C.; Holkenbrink, A.; Werz, D. B. Recent advances in the synthesis of carbohydrate mimetics. *Synthesis* **2010**, 3217−3242.
(47) 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.* **2015**, *54*, 9830−9834.
(48) Rojas-Ocáriz, V.; Compañón, I.; Aydillo, C.; Castro-Lopez, J.; Jiménez-Barbero, J.; Hurtado-Guerrero, R.; Avenoza, A.; Zurbano, M. M.; Peregrina, J. M.; Busto, J. H.; Corzana, F. Design of α-S-neoglycopeptides derived from MUC1 with a flexible and solvent-exposed sugar moiety. *J. Org. Chem.* **2016**, *81*, 5929−5941.
(49) 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.* **2016**, *7*, 2294−2301.
(50) 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


DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

efficient detection of antibodies in patients with prostate cancer. *J. Am. Chem. Soc.* **2017**, *139*, 18255−18261.
(51) 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.* **2016**, *18*, 3890−3893.
(52) 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.* **2018**, *83*, 4973−4980.
(53) 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.* **2014**, *136*, 789−800.
(54) Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. New insights into α-GalNAc−Ser motif: Influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc.* **2006**, *128*, 14640−14648.
(55) 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.* **2007**, *129*, 9458−9467.
(56) 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-Oses, 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.* **2018**, *140*, 9652.
(57) Corzana, F.; Busto, J. H.; Marcelo, F.; de Luis, M. G.; Asensio, J. L.; Martín-Santamaría, S.; Sáenz, Y.; Torres, C.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. Rational design of a Tn antigen mimic. *Chem. Commun.* **2011**, *47*, 5319−5321.
(58) Paulsen, H.; Adermann, K. Synthesis of O-glycopeptides of the N-terminus of interleukin-2. *Liebigs Ann. Chem.* **1989**, *1989*, 751−769.
(59) Plattner, C.; Höfener, M.; Sewald, N. One-pot azido-chlorination of glycals. *Org. Lett.* **2011**, *13*, 545−547.
(60) Koeller, K. M.; Smith, M. E. B.; Wong, C.-H. Chemoenzymatic synthesis of PSGL-1 glycopeptides: Sulfation on tyrosine affects glycosyltransferase-catalyzed synthesis of the O-glycan. *Bioorg. Med. Chem.* **2000**, *8*, 1017−1025.
(61) Zerong, W. *Comprehensive Organic Name Reactions and Reagents*; John, Wiley & Sons, Inc., 2010; pp 3123−3128.
(62) Aydillo, C.; Navo, C. D.; Busto, J. H.; Corzana, F.; Zurbano, M. M.; Avenoza, A.; Peregrina, J. M. A double diastereoselective Michael-type addition as an entry to conformationally restricted Tn antigen mimics. *J. Org. Chem.* **2013**, *78*, 10968−10977.
(63) Navo, C. D.; Corzana, F.; Sánchez-Fernández, E. M.; Busto, J. H.; Avenoza, A.; Zurbano, M. M.; Nanba, E.; Higaki, K.; Mellet, C. O.; Fernández, J. M. G.; Peregrina, J. M. Conformationally-locked C-glycosides: tuning aglycone interactions for optimal chaperone behaviour in Gaucher fibroblasts. *Org. Biomol. Chem.* **2016**, *14*, 1473−1484.
(64) Winterfeld, G. A.; Ito, Y.; Ogawa, T.; Schmidt, R. R. A Novel and efficient route towards α-GalNAc-Ser and α-GalNAc-Thr building blocks for glycopeptide synthesis. *Eur. J. Org. Chem.* **1999**, 1167−1171.
(65) Winterfeld, G. A.; Schmidt, R. R. Nitroglycal concatenation: A broadly applicable and efficient approach to the synthesis of complex O-glycans. *Angew. Chem., Int. Ed.* **2001**, *40*, 2654−2657.
(66) Aydillo, C.; Jiménez-Osés, G.; Busto, J. H.; Peregrina, J. M.; Zurbano, M. M.; Avenoza, A. Theoretical evidence for pyramidalized bicyclic serine enolates in highly diastereoselective alkylations. *Chem.—Eur. J.* **2007**, *13*, 4840−4848.
(67) 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.* **2011**, *9*, 1621−1626.
(68) Meinjohanns, E.; Meldal, M.; Schleyer, A.; Paulsen, H.; Bock, K. Efficient syntheses of core 1, core 2, and core 3 and core 4 building blocks for SPS of mucin O-glycopeptides based on the N-Dts-method. *J. Chem. Soc., Perkin Trans. 1* **1996**, 985−993.
(69) Bourgault, J. P.; Trabbic, K. R.; Shi, M.; Andreana, P. R. Synthesis of the tumor associative α-aminooxy disaccharide of the TF antigen and its conjugation to a polysaccharide immune stimulant. *Org. Biomol. Chem.* **2014**, *12*, 1699−1702.
(70) Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T. How sugars pucker: Electronic structure calculations map the kinetic landscape of five biologically paramount monosaccharides and their implications for enzymatic catalysis. *J. Am. Chem. Soc.* **2014**, *136*, 1008−1022.
(71) Asensio, J. L.; Cañada, F. J.; García-Herrero, A.; Murillo, M. T.; Fernández-Mayoralas, A.; Johns, B. A.; Kozak, J.; Zhu, Z.; Johnson, C. R.; Jiménez-Barbero, J. Conformational Behavior of Aza-C-Glycosides: Experimental Demonstration of the Relative Role of the exo-anomeric Effect and 1,3-Type Interactions in Controlling the Conformation of Regular Glycosides. *J. Am. Chem. Soc.* **1999**, *121*, 11318−11329.
(72) Asensio, J. L.; Cañada, F. J.; Cheng, X.; Khan, N.; Mootoo, D. R.; Jiménez-Barbero, J. Conformational Differences Between O- and C-Glycosides: The α-O-Man-(1−>1)-β-Gal/α-C-Man-(1-−>1)-β-Gal Case − A Decisive Demonstration of the Importance of the exo-Anomeric Effect on the Conformation of Glycosides. *Chem.—Eur. J.* **2000**, *6*, 1035−1041.
(73) Wiberg, K. B.; Bailey, W. F.; Lambert, K. M.; Stempel, Z. D. The anomeric effect: It's complicated. *J. Org. Chem.* **2018**, *83*, 5242−5255.
(74) Unione, L.; Xu, B.; Díaz, D.; Martín-Santamaría, S.; Poveda, A.; Sardinha, J.; Rauter, A. P.; Blériot, Y.; Zhang, Y.; Cañada, F. J.; Sollogoub, M.; Jiménez-Barbero, J. Conformational Plasticity in Glycomimetics: Fluorocarbamethyl-L-idopyranosides Mimic the Intrinsic Dynamic Behaviour of Natural Idose Rings. *Chem.—Eur. J.* **2015**, *21*, 10513−10521.
(75) Stenutz, R.; Carmichael, I.; Widmalm, G.; Serianni, A. S. Hydroxymethyl group conformation in saccharides: structural dependencies of $^{2}J_{HH}$, $^{3}J_{HH}$, and $^{1}J_{CH}$ spin-spin coupling constants. *J. Org. Chem.* **2002**, *67*, 949−958.
(76) Thibaudeau, C.; Stenutz, R.; Hertz, B.; Klepach, T.; Zhao, S.; Wu, Q.; Carmichael, I.; Serianni, A. S. Correlated C−C and C−O Bond Conformations in Saccharide Hydroxymethyl Groups: Parametrization and Application of Redundant $^{1}H-^{1}H$, $^{13}C-^{1}H$, and $^{13}C-^{13}C$ NMR J-Couplings. *J. Am. Chem. Soc.* **2004**, *126*, 15668−15685.
(77) Navarro-Vázquez, A.; Cobas, J. C.; Sardina, F. J.; Casanueva, J.; Díez, E. A Graphical Tool for the Prediction of Vicinal Proton-Proton $^{3}J_{HH}$ Coupling Constants. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1680−1685.
(78) Leclerc, E.; Pannecoucke, X.; Ethève-Quelquejeu, M.; Sollogoub, M. Fluoro-C-glycosides and fluoro-carbasugars, hydrolytically stable and synthetically challenging glycomimetics. *Chem. Soc. Rev.* **2013**, *42*, 4270−4283.
(79) Kirschner, K. N.; Woods, R. J. Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541−10545.
(80) Soper, A. K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **2000**, *258*, 121−137.
(81) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; 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.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.;



DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. *AMBER 2016*; University of California: San Francisco, 2016.
(82) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(83) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford CT, 2016.
(84) 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 functionals. *Theor. Chem. Acc.* **2007**, *120*, 215−241.
(85) Scalmani, G.; Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **2010**, *132*, 114110.



DOI: 10.1021/acsomega.8b02576
*ACS Omega* 2018, 3, 18142−18152

Protein Modification

International Edition: DOI: 10.1002/anie.201901405
German Edition: DOI: 10.1002/ange.201901405

# Quaternization of Vinyl/Alkynyl Pyridine Enables Ultrafast Cysteine-Selective Protein Modification and Charge Modulation

*Maria J. Matos[+], Claudio D. Navo[+], Tuuli Hakala[+], Xhenti Ferhati[+], Ana Guerreiro, David Hartmann, Barbara Bernardim, Kadi L. Saar, Ismael Compañón, Francisco Corzana,* Tuomas P. J. Knowles,* Gonzalo Jiménez-Osés,* and Gonçalo J. L. Bernardes**

**Abstract:** *Quaternized vinyl- and alkynyl-pyridine reagents were shown to react in an ultrafast and selective manner with several cysteine-tagged proteins at near-stoichiometric quantities. We have demonstrated that this method can effectively create a homogenous antibody–drug conjugate that features a precise drug-to-antibody ratio of 2, which was stable in human plasma and retained its specificity towards Her2 + cells. Finally, the developed warhead introduces a +1 charge to the overall net charge of the protein, which enabled us to show that the electrophoretic mobility of the protein may be tuned through the simple attachment of a quaternized vinyl pyridinium reagent at the cysteine residues. We anticipate the generalized use of quaternized vinyl- and alkynyl-pyridine reagents not only for bioconjugation, but also as warheads for covalent inhibition and as tools to profile cysteine reactivity.*

Chemical site-selective modification offers a means to diversify the function and properties of the protein.[1] For example, by using the targeting capabilities of an antibody, it is possible to covalently attach a very potent drug to the antibody through a precise chemical reaction to shuttle this drug to a specific tissue.[2] The toolbox of reactions for protein modification has expanded significantly in the last decade.[3] Of these, reactions that target proteinogenic amino acids seem particularly suitable to modify native proteins in the test tube. Lysine,[4] methionine,[5] tryptophan,[6] and the N-[7] and the C-terminus[8] may now be targeted by using a variety of approaches. However, cysteine[9] remains perhaps the residue of choice to produce functional and, in particular, clinically useful protein conjugates, namely antibody–drug conjugates (ADCs). This choice is a result of the high nucleophilicity of the sulfhydryl side-chain combined with the low abundance of free cysteine residues, since many cysteines are paired as structural disulfides. Thus, many research groups have focused on developing efficient methods to chemoselectively modify cysteine-tagged proteins. For example, electrophiles, such as carbonylacrylic acid reagents for Michael addition,[10] arylation reactions based on transition metals,[11] or a selective amino acid sequence,[12] or conjugate additions at dehydroalanine[13] formed from cysteine or thiol-yne reactions using cyclooctynes[14] have been developed. Each of these methods has relative advantages and disadvantages, but a method based on the simple attachment of a warhead-like structure whose utility would go beyond bioconjugation and for example be used to tune protein pharmacokinetics, through changes in the overall net charge of a protein, or be used to design cysteine covalent inhibitors is missing from the current toolbox.

Herein, we report the computational chemistry assisted discovery of quaternization of the nitrogen of vinyl- and alkynyl pyridines to convert otherwise non-reactive reagents into ultrafast and chemoselective cysteine-modifying reagents. The utility of these reagents was demonstrated for bioconjugation and for modulation of electrophoretic mobility through charge incorporation by using microfluidics.

Based on computational predictions, we identified promising alkenes and alkynes through modelling the nucleophilic addition reaction between a simple thiolate or an amine with various electrophiles. As a result of the lower $pK_a$ of solvent-exposed thiols relative to amines, it is expected that at neutral or slightly basic pH a larger proportion of thiols would be deprotonated. Activation free energies ($\Delta G^{\ddagger}$) for the corresponding addition reactions were predicted by using quantum mechanical calculations (Figure 1a and b). 2-Vinyl- and 2-ethynyl pyridines (compounds **1** and **3**) were calculated to be poorly reactive toward thiolates ($\Delta G^{\ddagger} > 24\ \text{kcal mol}^{-1}$) and unreactive toward primary amines ($\Delta G^{\ddagger} > 31\ \text{kcal mol}^{-1}$).

[*] Dr. M. J. Matos,[+] T. Hakala,[+] D. Hartmann, Dr. B. Bernardim, K. L. Saar, Prof. T. P. J. Knowles, Dr. G. J. L. Bernardes
Department of Chemistry, University of Cambridge
Lensfield Road, CB2 1EW Cambridge (UK)
E-mail: tpjk2@cam.ac.uk
gb453@cam.ac.uk

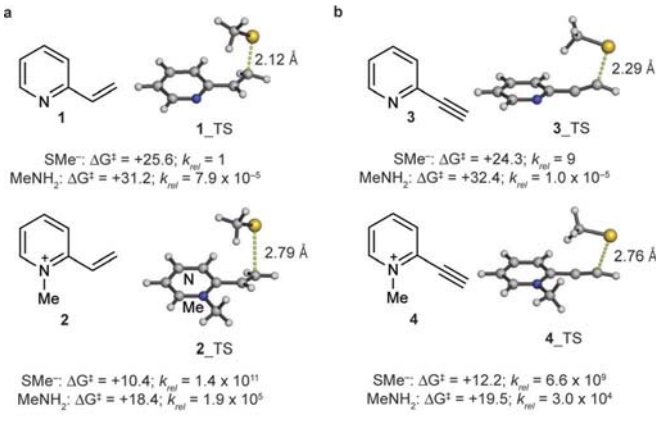
A. Guerreiro, Dr. G. J. L. Bernardes
Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa
Avenida Professor Egas Moniz, 1649-028 Lisboa (Portugal)
E-mail: gbernardes@medicina.ulisboa.pt

Dr. C. D. Navo,[+] X. Ferhati,[+] I. Compañón, Dr. F. Corzana, Dr. G. Jiménez-Osés
Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química
26006 Logroño (Spain)
E-mail: francisco.corzana@unirioja.es

Dr. C. D. Navo,[+] Dr. G. Jiménez-Osés
CIC bioGUNE, Bizkaia Technology Park
Building 801A, 48170 Derio (Spain)
E-mail: gjoses@cicbiogune.es

[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201901405.




c

| Entry | Electrophile | Nucleophile | Conversion (%) 4 min | 1 h | 24 h |
|---|---|---|---|---|---|
| 1 | **1** | *n*-PrSH | 0 | 10 | 62 |
| 2 | **2** | *n*-PrSH | 83 | 99 | – |
| 3 | **3** | *n*-PrSH | 0 | 0 | 6 |
| 4 | **4** | *n*-PrSH | 84 | 99 | – |
| 5 | **1** | Ac-Cys-NH$_2$ | 2 | 36 | 89 |
| 6 | **2** | Ac-Cys-NH$_2$ | 93 | 93 | – |
| 7 | **2** | *n*-PrNH$_2$ | 0 | 1 | 3 |
| 8 | **2** | Ac-Lys-NH$_2$ | 0 | 1 | 6 |

***Figure 1.*** Computer-assisted evaluation and experimental validation of a) vinyl- and b) alkynyl pyridinium electrophile reagents **1**–**4** for cysteine modification. c) Reaction of **1**–**4** with stoichiometric amounts of various nucleophiles.

However, when quaternized, the resulting *N*-methylpyridinium derivatives were calculated to be exceedingly reactive towards nucleophilic addition, with dramatically decreased activation free energies. For instance, 2-vinyl- and 2-ethynylpyridiniums **2** and **4** were calculated to be more than a billion times more reactive with thiolates ($\Delta G^{\neq} = 10$–$12$ kcal mol$^{-1}$) than their non-quaternized analogues **1** and **3**, which shows predicted reactivities similar to those of maleimide[15] and carbonylacrylic derivatives[10] previously used for cysteine ligation. In all cases, a complete selectivity for cysteine over lysine addition was predicted, as reflected by the nearly 8 kcal mol$^{-1}$ higher activation barrier calculated for the latter.

We validated our computational predictions by reacting stoichiometric amounts of minimal cysteine and lysine side-chain model compounds (that is 1-propanethiol and 1-propylamine, respectively) with commercially available **1**–**3** and synthesized reagent **4** (Figure 1c), and monitoring reaction with $^1$H NMR spectroscopy (3 mM, phosphate buffer in $D_2O$, pH 7.6; see the Supporting Information for additional data and a detailed discussion on the effect of using a slightly more acidic pH). In agreement with theoretical predictions, the reactions with non-quaternized reagents **1** and **3** were very slow and gave 62 and 6% conversions, respectively, of addition products after 24 hours (Figure 1c, Entries 1 and 3). On the contrary, pyridine *N*-methylated reagents **2** and **4** gave a conversion of 99% of addition product after 1 hour with only 1 equivalent of electrophile (Figure 1c, Entries 2 and 4). 2-Alkynyl reagent **4** gave a 6:4 mixture of isomers as products, namely the *cis* and *trans* alkene adducts (See the Supporting Information for their structural characterization and thermal stability). The kinetic profiles obtained with a protected cysteine (*N*-acetylcysteine amide, Ac-Cys-NH$_2$) were similar, albeit slightly faster than with 1-propanethiol (89% adduct formation in 24 hours with non-quaternized **1**, and 93% adduct formation in $<4$ min with quaternized **2**; Figure 1c, Entries 5 and 6). Also, in agreement with computational predictions, the fast kinetics observed with quaternized pyridinium reagents **2** and **4** were comparable to those of *N*-ethylmaleimide,[15] for which the reaction with 1-propanethiol and Ac-Cys-NH$_2$ were complete within the 4 min necessary to obtain the first $^1$H NMR spectroscopic measurements, although the formation of an undetermined alkene by-product was observed in these reactions with maleimide (see the Supporting Information). With lysine side-chain mimic 1-propylamine (Figure 1c, Entry 7) and *N*-acetyllysine amide (Ac-Lys-NH$_2$; Figure 1c, Entry 8) the reactions with quaternized 1-methyl-2-vinylpyridinium (**2**) were exceedingly slow ($<6$% adduct formation after 24 hours with 1 equivalent of electrophile). This result agrees with the computationally predicted orthogonality of these electrophiles towards cysteines, with no expected reaction on typically solvent-exposed lysines. It is interesting to note that the terminal alkyne hydrogen in compounds **3** and **4** quickly exchanged to deuterium in $D_2O$ at pH 7.6, particularly in the case of quaternized **4**, for which deuteration takes place within 4 min. H/D exchange at the vinyl hydrogens of quaternized **2** was also observed when [D$_7$]DMF was used as a co-solvent.

Accurate reaction rate constants could not be obtained for very reactive quaternized pyridiniums **2** and **4**, and *N*-ethylmaleimide as a result of the ultrafast kinetics observed under the reaction conditions used (3 mM was the detection limit for $^1$H NMR spectroscopy in our experiments). Also, the high level of sensitivity to small reagent concentration changes on the observed rate for second-order reactions, further complicates the derivation of the rate constants under stoichiometric conditions. Furthermore, although alkynyl pyridinium **4** was selective towards mono-functionalization under strict stoichiometric conditions, it was found to undergo double thiol addition to the external position of the triple bond to give a stable dithioacetal when a slight excess of the thiol was used (see the Supporting Information). Once the dramatic effect of pyridine quaternization in boosting the electrophilicity of reagents **1** and **3** was demonstrated and given the very similar and ultrafast reactivity of quaternized reagents **2** and **4** towards thiols, we decided to use more convenient 1-methyl-2-vinylpyridinium **2** for cysteine-selective modification on proteins (to avoid terminal alkyne deuteration and formation of mixtures of potentially reactive alkene isomers upon nucleophilic addition).

Having demonstrated the superior reactivity of **2** towards thiols on small molecule models, we decided to test the use of vinyl pyridinium reagents towards cysteine-tagged proteins (Figure 2a). We selected four representative proteins, the C2A domain of Synaptotagmin-I, Annexin-V, ubiquitin, and albumin, which display either natural or engineered single surface-exposed cysteines (see the Supporting Information). In all cases, and with stoichiometric amounts or only a small excess of **2** (1 equiv for ubiquitin and albumin and 10 equiv for annexin-V and C2Am), under buffered (NaP$_i$ pH 8.0,

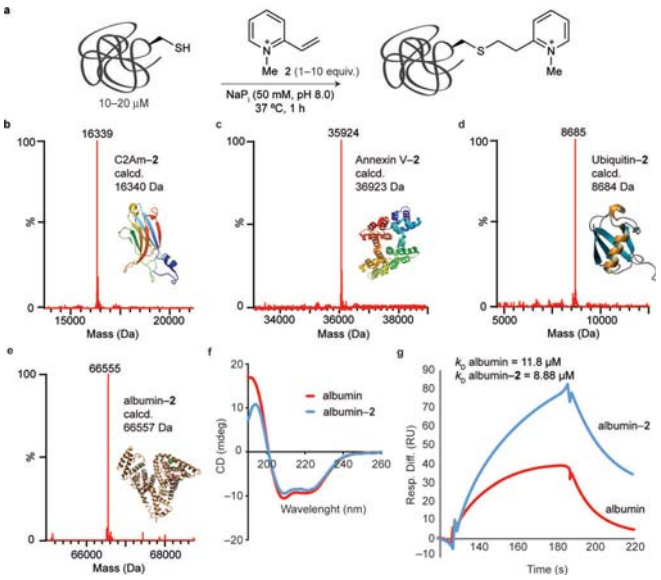



***Figure 2.*** Cysteine-selective protein modification. a) The reaction of a cysteine-tagged protein with **2**. General conditions: a cysteine-tagged protein (10–20 μM) is reacted with **2** (1–10 equiv) in $NaP_i$ (50 mM, pH 8.0) at 37 °C for 1 hour. b–e) ES-MS spectra of the product of the reaction between b) C2Am, c) Annexin V, d) Ubiquitin, and e) albumin with **2** (deconvoluted spectra; expected increase of 118 Da in the total protein mass). f) Comparative CD analysis of albumin and albumin-**2**. g) Surface plasma resonance of binding to the FcRn receptor. $k_D$ for rHSA and rHSA-**2**.

50 mM) conditions for 1 hour at 37 °C, complete conversion into the corresponding thioether adduct was observed, as confirmed by LC-MS analysis, and in high yield (> 95 %), as shown by Bradford protein assay (Figure 2 b-2e; see also the Supporting Information; no reaction was detected with either **1** or **2** under slightly acidic pH 5.5). Despite the use of 10 equiv to ensure efficient modification of even sterically crowded, and thus poorly deprotonated, cysteine residues, such as in the case of Annexin-V, no non-specific reactions at lysine were detected. Other studies that use vinyl-substituted pyridine derivatives achieved a maximum of only 85 % conversion after 24 hours,[16] which demonstrates well the effect of quaternization of the pyridine on the kinetics of the bioconjugation reaction. Furthermore, the final conjugates did not react with Ellman's reagent [5,5′-dithiobis(2-nitrobenzoic acid)] that showed that all cysteine had been consumed (see the Supporting Information). Importantly, we verified that the formed thioether bond is stable towards potential thiol exchange reactions, and the conjugates were also fully stable in human plasma in buffered pH 5–8 solutions (see the Supporting Information). Next, we confirmed that the conditions used are mild and do not induce changes in both the structure or function of the native proteins. In the case of albumin, conjugate albumin-**2** showed no significant changes in its secondary structural content as determined by circular dichroism (CD; Figure 2 f) and retained its ability to bind to the neonatal Fc receptor (FcRn) as determined by Surface Plasma Resonance ($K_D = 11.8$ μM for albumin versus $K_D = 8.88$ μM for albumin-**2**; Figure 2 g). Together, our data demonstrates the generalization of a method for cysteine-selective protein modification based simply on the use of quaternized vinyl pyridinium reagents. The products are formed in high conversion and yield, are stable in human plasma, and both native structure and function activities are retained.

Next, we decided to demonstrate that vinyl pyridinium reagents may be easily functionalized first through quaternization of the pyridine nitrogen with an alkyne linker followed by subsequent derivatization through $Cu^I$-catalyzed azide–alkyne cycloaddition reaction with a suitable synthetic motif that displays an azido group (see the Supporting Information). We chose to build a quaternized vinyl pyridinium reagent that features cytotoxic drug monomethyl auristatin E (MMAE; Figure 3 a). By connecting the cysteine-selective

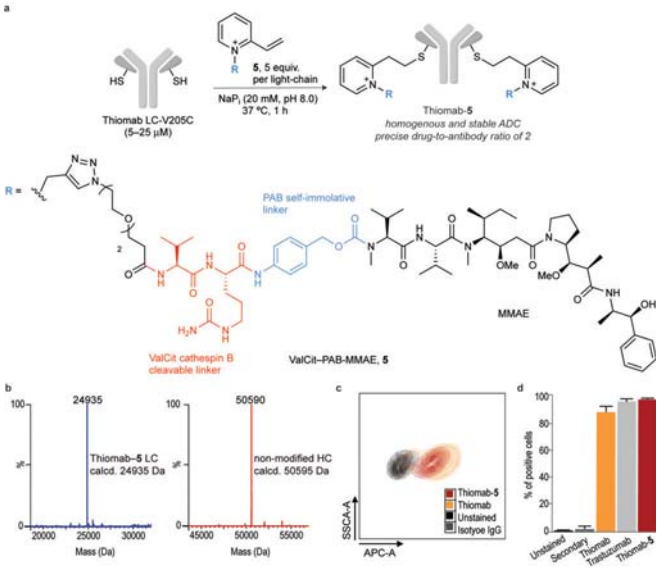



***Figure 3.*** Cysteine-selective antibody modification. a) Schematic for the bioconjugation of **5** to Thiomab and chemical structure of **5**. b) ES-MS spectra of the light- and heavy-chain of Thiomab after conjugation with **5**. c) Counter plot indicates the binding affinity of Thiomab-**5** to Her2 expressing SKBR3 cells at 50 nM. d) Percentage of SKBR3 cells bound to Thiomab-**5** at 50 nM. Trastuzumab and Thiomab were used as positive controls and IgG Isotype was used as negative control.

reagent and the drug, we introduced the dipeptide valine–citrulline (ValCit), which is known to cleave on exposure to cathepsin B,[17] and the self-immolating spacer *p*-aminobenzyl carbamate (PABC).[18] This reagent **5** was then conjugated to Thiomab, an HER2 targeting antibody that has been engineered to contain an additional cysteine residue at position 205 in each light-chain (Thiomab LC-V205C).[19] To our delight, the conjugation reaction proceeded to completion with 5 equiv per light-chain of **5**, within 1 hour at 37 °C in $NaP_i$ pH 8.0, 20 mM buffer (Figure 3 a). LC-MS analysis showed a single modification in each-light chain and no modifications in the heavy-chain (Figure 3 b) to form a conjugate with a precise drug-to-antibody ratio of 2. Importantly, Thiomab-**5** conjugate retained its specificity towards Her2 positive cells, as demonstrated by flow cytometry analysis (Figure 3 c and d). Our data demonstrate the use of quaternized vinyl pyridinium

reagents to efficiently generate homogenous and functional antibody conjugates.

The conjugation of a protein can be monitored directly by following changes in its electrophoretic mobility because the quaternized vinyl pyridinium reagent introduces an additional charge of +1 per cysteine residue to the net charge. For instance, albumin features one single free cysteine and thus conjugation of **2** adds +1 charge to the overall net charge of the protein (charge ca. −7 at pH 7). We decided to test this hypothesis by using a microfluidic free-flow electrophoresis device,[20] which allows the charge of proteins to be determined under native conditions in solution and requires only low volumes of sample. The samples were introduced into the device with a flanking buffer, and the deflection of the protein samples was observed while a transverse electric field was applied by increasing the potential between electrodes in the device from 0–120 V (Figure 4a). When the measured

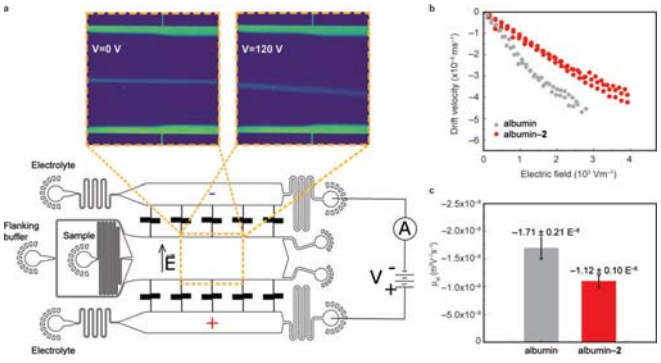



***Figure 4.*** a) Representation and b),c) results from microfluidic determination of the electrophoretic mobilities of native albumin and albumin-**2**. The device is operated by withdrawing the sample and flanking buffer from the outlets to create a laminar stream of sample that flows between two electrolyte streams, separated by the buffer. When a potential difference is applied between the electrolyte streams, a transverse electric field is created in the separation chamber and the deflection of the sample can be monitored by using the intrinsic fluorescence of the protein at 280 nm.[21] Both samples have a unique response to the electric field and this can be seen by plotting electric field (*E*) versus the drift velocity ($V_{drift}$) b), which is used to determine the representative mobilities (c).

deflection in the field was converted into drift velocity ($V_{drift}$) and the electric field inside the devices was calibrated (Figure 4b), a clear difference in the electrophoretic behavior of albumin and albumin-**2** was observed. Furthermore, the electrophoretic mobilities were estimated to be: $-1.71 \pm 0.21 \times 10^{-8}\,m^2V^{-1}s^{-1}$ to give a charge of about $\approx -6.84 \pm 0.58$ and $-1.12 \pm 0.10 \times 10^{-8}\,m^2V^{-1}s^{-1}$ to give a charge of about $-4.64 \pm 0.12$ for albumin and albumin-**2**, respectively. The decrease in mobility can be explained by the direct proportionality of the charge ($q$) to the electrophoretic mobility [$\mu_{el}$; $q = (\mu_{el} \times k_B T)/D$], which implies that if the charge is decreased, the mobility should also decrease as confirmed by the data. Moreover, because we are adding positive charge through conjugation of the quaternized vinyl pyridinium reagent **2** into a negatively charged albumin, the net charge of conjugate albumin-**2** should be smaller, as observed in the results obtained with the microfluidic electrophoresis. These data thus demonstrate the application of site-selective protein modification to change the overall net charge of a protein in a controlled manner, and consequently modulate its electrophoretic mobility.

In summary, an efficient and irreversible cysteine-selective bioconjugation method is reported. This method is enabled by the discovery that quaternization of the nitrogen of vinyl- and alkynyl pyridines transforms these molecules into extremely reactive electrophiles towards thiols. We demonstrate the utility of these reagents for cysteine-selective bioconjugation of five different protein scaffolds, including a clinically used antibody. Importantly, the conjugates formed are resistant towards thiol exchange reactions and retain their native activity. By using a microfluidic setup, we show that the electrophoretic mobility of a protein may be modulated through simple attachment of the quaternized vinyl pyridinium reagent at cysteine to introduce a +1 charge to the overall net charge of the protein. This property of the quaternized vinyl pyridinium may prove useful to enhance the hydrophilicity, half-life in circulation, or internalization rates of ADCs. The simplicity of these reagents combined with their excellent reactivity and selectivity towards thiols, makes them ideal for protein bioconjugation, but also for use as warheads in the design of cysteine covalent ligands[22] or to probe hyper-reactive cysteines within the human proteome in chemical proteomic approaches.[23]

## Acknowledgements

Funded under the EU Horizon 2020 Programme, Marie Skłodowska-Curie ITN GA No. 675007, the Royal Society (UF110046 and URF\R\180019 to G.J.L.B.), FCT Portugal (iFCT IF/00624/2015 to G.J.L.B. and PhD studentship SFRH/BD/115932/2016 to A.G.), Xunta de Galicia (Galician Plan of research, innovation and growth 2011–2015, ED481B 2014/086-0 and ED481B 2018/007 to M.J.M.), D.G.I. MINECO/FEDER (grants CTQ2015-70524-R and RYC-2013–14706 to G.J.-O. and C.D.N and CTQ2015-67727-R to F.C.), Universidad de la Rioja (FPI PhD studentship to I.C.), FAPESP (BEPE 2015/07509-1 and 2017/13168-8 to B.B.), and by an ERC StG (GA No. 676832). We also thank Genentech for providing the Thiomab antibody, Albumedix for providing rHSA, Dr. André Neves and Prof. Kevin Brindle for providing C2Am, and Dr. Vikki Cantrill for her help with the editing of this manuscript.

## Conflict of interest

The authors declare no conflict of interest.

**Keywords:** antibody–drug conjugates · bioconjugation · cysteine · microfluidics · protein modification

**How to cite:** *Angew. Chem. Int. Ed.* **2019**, *58*, 6640–6644
*Angew. Chem.* **2019**, *131*, 6712–6716

[1] N. Krall, F. P. da Cruz, O. Boutureira, G. J. L. Bernardes, *Nat. Chem.* **2015**, *8*, 103 – 113.
[2] V. Chudasama, A. Maruani, S. Caddick, *Nat. Chem.* **2016**, *8*, 114 – 119.
[3] a) C. D. Spicer, B. G. Davis, *Nat. Commun.* **2014**, *5*, 4740; b) O. Boutureira, G. J. L. Bernardes, *Chem. Rev.* **2015**, *115*, 2174 – 2195; c) E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira, G. J. L. Bernardes, *Nat. Rev. Chem.* **2019**, *3*, 147 – 171; d) L. Xue, I. A. Karpenko, J. Hiblot, K. Johnsson, *Nat. Chem. Biol.* **2015**, *11*, 917 – 923.
[4] a) M. J. Matos, B. L. Oliveira, N. Martínez-Sáez, A. Guerreiro, P. M. S. D. Cal, J. Bertoldo, M. Maneiro, E. Perkins, J. Howard, M. J. Deery, J. M. Chalker, F. Corzana, G. Jiménez-Osés, G. J. L. Bernardes, *J. Am. Chem. Soc.* **2018**, *140*, 4004 – 4017; b) S. R. Adusumalli, D. G. Rawale, U. Singh, P. Tripathi, R. Paul, N. Kalra, R. K. Mishra, S. Shukla, V. Rai, *J. Am. Chem. Soc.* **2018**, *140*, 15114 – 15123.
[5] a) S. Lin, X. Yang, S. Jia, A. M. Weeks, M. Hornsby, P. S. Lee, R. V. Nichiporuk, A. T. Iavarone, J. A. Wells, F. D. Toste, C. J. Chang, *Science* **2017**, *355*, 597 – 602; b) M. T. Taylor, J. E. Nelson, M. G. Suero, M. J. Gaunt, *Nature* **2018**, *562*, 563 – 568.
[6] Y. Seki, T. Ishiyama, D. Sasaki, J. Abe, Y. Sohma, K. Oisaki, M. Kanai, *J. Am. Chem. Soc.* **2016**, *138*, 10798 – 10801.
[7] C. B. Rosen, M. B. Francis, *Nat. Chem. Biol.* **2017**, *13*, 697 – 705.
[8] S. Bloom, C. Liu, D. K. Kölmel, J. X. Qiao, Y. Zhang, M. A. Poss, W. R. Ewing, D. W. C. MacMillan, *Nat. Chem.* **2017**, *10*, 205 – 211.
[9] a) J. M. Chalker, G. J. L. Bernardes, Y. A. Lin, B. G. Davis, *Chem. Asian J.* **2009**, *4*, 630 – 640; b) S. B. Gunnoo, A. Madder, *ChemBioChem* **2016**, *17*, 529 – 553.
[10] a) B. Bernardim, P. M. S. D. Cal, M. J. Matos, B. L. Oliveira, N. Martínez-Sáez, I. S. Albuquerque, E. Perkins, F. Corzana, A. C. B. Burtoloso, G. Jiménez-Osés, G. J. L. Bernardes, *Nat. Commun.* **2016**, *7*, 13128; b) B. Bernardim, M. J. Matos, X. Ferhati, I. Compañón, A. Guerreiro, P. Akkapeddi, A. C. B. Burtoloso, G. Jiménez-Osés, F. Corzana, G. J. L. Bernardes, *Nat. Protoc.* **2019**, *14*, 86 – 99.
[11] E. V. Vinogradova, C. Zhang, A. M. Spokoyny, B. L. Pentelute, S. L. Buchwald, *Nature* **2015**, *526*, 687 – 691.
[12] C. Zhang, M. Welborn, T. Zhu, N. J. Yang, M. S. Santos, T. Van Voorhis, B. L. Pentelute, *Nat. Chem.* **2015**, *8*, 120 – 128.
[13] a) G. J. L. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, *J. Am. Chem. Soc.* **2008**, *130*, 5052 – 5053; b) A. M. Freedy, M. J. Matos, O. Boutureira, F. Corzana, A. Guerreiro, P. Akkapeddi, V. J. Somovilla, T. Rodrigues, K. Nicholls, B. Xie, G. Jiménez-Osés, K. M. Brindle, A. A. Neves, G. J. L. Bernardes, *J. Am. Chem. Soc.* **2017**, *139*, 18365 – 18375.
[14] C. Zhang, P. Dai, A. A. Vinogradov, Z. P. Gates, B. L. Pentelute, *Angew. Chem. Int. Ed.* **2018**, *57*, 6459 – 6463; *Angew. Chem.* **2018**, *130*, 6569 – 6573.
[15] J. M. J. M. Ravasco, H. Faustino, A. Trindade, P. M. P. Gois, *Chem. Eur. J.* **2019**, *25*, 43 – 59.
[16] F.-H. Ma, J.-L. Chen, Q.-F. Li, H.-H. Zuo, F. Huang, X.-C. Su, *Chem. Asian J.* **2014**, *9*, 1808 – 1816.
[17] G. M. Dubowchik, R. A. Firestone, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3341 – 3346.
[18] P. L. Carl, P. K. Chakravarty, J. A. Katzenellenbogen, *J. Med. Chem.* **1981**, *24*, 479 – 480.
[19] J. R. Junutula, H. Raab, S. Clark, S. Bhakta, D. D. Leipold, S. Weir, Y. Chen, M. Simpson, S. P. Tsai, M. S. Dennis, Y. Lu, Y. G. Meng, C. Ng, J. Yang, C. C. Lee, E. Duenas, J. Gorrell, V. Katta, A. Kim, K. McDorman, K. Flagella, R. Venook, S. Ross, S. D. Spencer, W. Lee Wong, H. B. Lowman, R. Vandlen, M. X. Sliwkowski, R. H. Scheller, P. Polakis, W. Mallet, *Nat. Biotechnol.* **2008**, *26*, 925.
[20] K. L. Saar, Y. Zhang, T. Müller, C. P. Kumar, S. Devenish, A. Lynn, U. Łapińska, X. Yang, S. Linse, T. P. J. Knowles, *Lab Chip* **2018**, *18*, 162 – 170.
[21] P. K. Challa, Q. Peter, M. A. Wright, Y. Zhang, K. L. Saar, J. A. Carozza, J. L. P. Benesch, T. P. J. Knowles, *Anal. Chem.* **2018**, *90*, 3849 – 3855.
[22] M. Forster, A. Chaikuad, T. Dimitrov, E. Döring, J. Holstein, B. T. Berger, M. Gehringer, K. Ghoreschi, S. Müller, S. Knapp, S. A. Laufer, *J. Med. Chem.* **2018**, *61*, 5350 – 5366.
[23] E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. D. Dillon, D. A. Bachovchin, K. Mowen, D. Baker, B. F. Cravatt, *Nature* **2010**, *468*, 790 – 795.

Manuscript received: January 31, 2019
Revised manuscript received: March 11, 2019
Accepted manuscript online: March 21, 2019
Version of record online: April 9, 2019

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

Cite This: J. Phys. Chem. Lett. 2019, 10, 3339−3345

pubs.acs.org/JPCL

# Conformational Behavior of D-Lyxose in Gas and Solution Phases by Rotational and NMR Spectroscopies

Camilla Calabrese,[†,‡] Patricia Écija,[†] Ismael Compañón,[§] Montserrat Vallejo-López,[†] Álvaro Cimas,[¶] Maider Parra,[†,‡] Francisco J. Basterretxea,[†] José I. Santos,[⊥] Jesús Jiménez-Barbero,[#,▽,☆] Alberto Lesarri,[⊕] Francisco Corzana,*[,§] and Emilio J. Cocinero*[,†,‡]

[†]Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain

[‡]Instituto Biofisika (CSIC, UPV/EHU), 48080 Bilbao, Spain

[§]Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, 26006 Logroño, Spain

[¶]Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, LAMBE UMR8587, Université d'Évry val d'Essonne, 91025 Évry, France

[⊥]SGIker UPV/EHU, Centro Joxe Mari Korta, Tolosa Hiribidea 72, 20018 Donostia, Spain

[#]Departamento de Química Orgánica II, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain

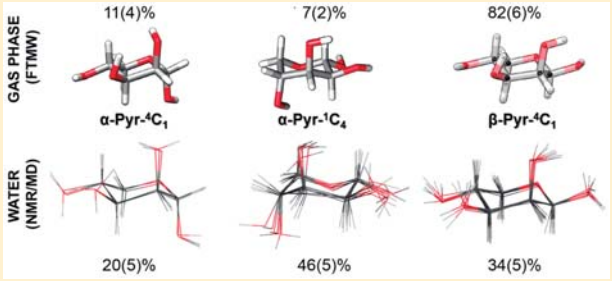
[▽]IKERBASQUE, Basque Foundation for Science, 48009 Bilbao, Spain

[☆]Chemical Glycobiology Laboratory, CIC bioGUNE, Bizkaia Technology Park, Building 800, 48160 Derio, Spain

[⊕]Departamento de Química Física y Química Inorgánica−IU CINQUIMA, Facultad de Ciencias, Universidad de Valladolid, 47011 Valladolid, Spain

Supporting Information

**ABSTRACT:** Understanding the conformational preferences of carbohydrates is crucial to explain the interactions with their biological targets and to improve their use as therapeutic agents. We present experimental data resolving the conformational landscape of the monosaccharide D-lyxose, for which quantum mechanical (QM) calculations offer model-dependent results. This study compares the structural preferences in the gas phase, determined by rotational spectroscopy, with those in solution, resolved by nuclear magnetic resonance (NMR) and molecular dynamics (MD) simulations. In contrast to QM calculations, D-lyxose adopts only pyranose forms in the gas phase, with the $\alpha$-anomer exhibiting both the $^4C_1$ and $^1C_4$ chairs (60:40). The predominantly populated $\beta$-anomer shows the $^4C_1$ form exclusively, as determined experimentally by isotopic substitution. In aqueous solution, the pyranose forms are also dominant. However, in contrast to the gas phase, the $\alpha$-anomer as $^1C_4$ chair is the most populated, and its solvation is more effective than for the $\beta$ derivative. Markedly, the main conformers found in the gas phase and solution are characterized by the lack of the stabilizing anomeric effect. From a mechanistic perspective, both rotational spectroscopy and solid-state nuclear magnetic resonance (NMR) corroborate that $\alpha \leftrightarrow \beta$ or furanose $\leftrightarrow$ pyranose interconversions are prevented in the gas phase. Combining microwave (MW) and NMR results provides a powerful method for unraveling the water role in the conformational preferences of challenging molecules, such as flexible monosaccharides.




## INTRODUCTION

Carbohydrates are the most abundant class of biomolecules in living organisms. Their molecular diversity encodes structural information that is essential in many physiological processes, such as protein folding and cell signaling, proliferation, and differentiation.[1,2] Glycan-mediated interactions are also involved in bacterial adhesion, viral infection, inflammation, and immune system activation. Therefore, understanding the conformational choices of carbohydrates is crucial for the elucidation of these interactions and, in turn, the rational design of therapeutic agents based on carbohydrates.[3,4]

It is well-known that water molecules play a pivotal role in the conformational preferences of saccharides.[5] For example, Woods and co-workers[6] found an increment of the $\omega$ angles in the solvated *trans*−*gauche* conformers of galactopyranosides, in

Received: April 5, 2019
Accepted: May 29, 2019
Published: May 29, 2019





DOI: 10.1021/acs.jpclett.9b00978
J. Phys. Chem. Lett. 2019, 10, 3339−3345